STATISTICAL METHODS FOR SIGNAL PROCESSING

Alfred O. Hero

December 20, 2001

This set of notes is the primary source material for the course "Estimation, filtering and detection" taught by Prof. Alfred Hero over the period 1998-2002 at the University of Michigan Ann Arbor. The course pack is the property of Alfred O. Hero III and is copyright protected. Author can be reached at
Dept. EECS, University of Michigan, Ann Arbor, MI 48109-2122
Tel: 734-763-0564.
email hero@eecs.umich.edu;
`http://www.eecs.umich.edu/~hero/hero.html`.

Contents

1 INTRODUCTION

1.1 STATISTICAL SIGNAL PROCESSING

Many engineering applications require extraction of a signal or parameter of interest from degraded measurements. To accomplish this it is often useful to deploy fine-grained statistical models; diverse sensors which acquire extra spatial, temporal, or polarization information; or multi-dimensional signal representations, e.g. time-frequency or time-scale. When applied in combination these approaches can be used to develop highly sensitive signal estimation, detection, or tracking algorithms which can exploit small but persistent differences between signals, interferences, and noise. Conversely, these approaches can be used to develop algorithms to identify a channel or system producing a signal in additive noise and interference, even when the channel input is unknown but has known statistical properties.

Broadly stated, statistical signal processing is concerned with the reliable estimation, detection and classification of signals which are subject to random fluctuations. Statistical signal processing has its roots in probability theory, mathematical statistics and, more recently, systems theory and statistical communications theory. The practice of statistical signal processing involves: (1) description of a mathematical and statistical model for measured data, including models for sensor, signal, and noise; (2) careful statistical analysis of the fundamental limitations of the data including deriving benchmarks on performance, e.g. the Cramèr-Rao, Ziv-Zakai, Barankin, Rate Distortion, Chernov, or other lower bounds on average estimator/detector error; (3) development of mathematically optimal or suboptimal estimation/detection algorithms; (4) asymptotic analysis of error performance establishing that the proposed algorithm comes close to reaching a benchmark derived in (2); (5) simulations or experiments which compare algorithm performance to the lower bound and to other competing algorithms. Depending on the specific application, the algorithm may also have to be adaptive to changing signal and noise environments. This requires incorporating flexible statistical models, implementing low-complexity real-time estimation and filtering algorithms, and on-line performance monitoring.

1.2 PERSPECTIVE ADOPTED IN THIS BOOK

This book is at the interface between mathematical statistics and signal processing. The idea for the book arose in 1986 when I was preparing notes for the engineering course on detection, estimation and filtering at the University of Michigan. There were then no textbooks available which provided a firm background on relevant aspects of mathemetical statistics and multivariate analysis. These fields of statistics formed the backbone of this engineering field in the 1940's 50's and 60's when statistical communication theory was first being developed. However, more recent textbooks have downplayed the important role of statistics in signal processing in order to accommodate coverage of technological issues of implementation and data acquisition for specific engineering applications such as radar, sonar, and communications. The result is that students finishing the course would have a good notion of how to solve focussed problems in these applications but would find it difficult either to extend the theory to a moderately different problem or to apply the considerable power and generality of mathematical statistics to other applications areas.

The technological viewpoint currently in vogue is certainly a useful one; it provides an essential engineering backdrop to the subject which helps motivate the engineering students. However, the disadvantage is that such a viewpoint can produce a disjointed presentation of the component parts of statistical signal processing making it difficult to appreciate the commonalities between detection, classification, estimation, filtering, pattern recognition, confidence intervals and other useful tools. These commonalities are difficult to appreciate without adopting a proper statistical perspective. This book strives to provide this perspective by more thoroughly covering elements of mathematical statistics than other statistical signal processing textbooks. In particular we cover point estimation, interval estimation, hypothesis testing, time series, and multivariate analysis. In adopting a strong statistical perspective the book provides a unique viewpoint on the subject which permits unification of many areas of statistical signal processing which are otherwise difficult to treat in a single textbook.

The book is organized into 21 chapters listed in the attached table of contents. The first group of chapters reviews the fundamentals of matrix algebra, systems theory, and probability and random processes. The second group of chapters introduces the fundamentals of mathematical statistics, point estimation, hypothesis testing, and interval estimation in the standard context of independent identically distributed observations. Specific topics include: least squares techniques; likelihood ratio tests of hypotheses; e.g. testing for whiteness, independence, in single and multi-channel populations of measurements; iterative implementation of estimators and detectors via steepest ascent, Monte Carlo Markov Chain simulation, and the EM algorithm; and the construction of confidence envelopes on signals. These chapters provide the conceptual backbone for the rest of the book. Each subtopic is introduced with a set of one or two examples for illustration. Many of the topics here can be found in other graduate textbooks on the subject, e.g. those by Van Trees, Kay, and Srinath etal. However, the coverage here is broader with more depth and mathematical detail which is necessary for the sequel of the textbook. For example in the section on hypothesis testing and interval estimation the full theory of sampling distributions is used to derive the form and null distribution of the standard statistical tests of shift in mean, variance and correlation in a Normal sample.

The third group of chapters treats the extension of the theory in the previous chapters to non i.i.d. sampled Gaussian waveforms. This group contains extensions of the detection and estimation theory to single and multiple channels. As before, special emphasis is placed on the sampling distributions of the decision statistics. This group starts with offline methods; least squares and Wiener filtering; and culminates in a compact introduction of on-line Kalman filtering methods. A feature not found in other treatments is the separation principle of detection and estimation which is made explicit via Kalman and Wiener filter implementations of the generalized likelihood ratio test for model selection, reducing to a whiteness test of each the innovations produced by a bank of Kalman filters.

The book then turns to a set of more concrete applications areas arising in radar, communications, acoustic and radar signal processing, imaging, and other areas of signal processing. Topics include: testing for independence; parametric and non-parametric testing of a sample distribution; testing of binary Markov sequences and applictions to internet traffic monitoring; extensions to complex valued and continuous time observations; optimal coherent and incoherent receivers for digital and analog communications; spatio-temporal signal processing with multi-sensor sensor arrays; CFAR (constant false alarm rate) detection strategies for Electro-optical (EO) and Synthetic Aperture Radar (SAR) imaging and channel equalization.

1.2.1 PREREQUISITES

Readers are expected to possess a background in basic probability and random processes at the level of Stark&Woods [34], Ross [28] or Papoulis [23], exposure to undergraduate vector and matrix algebra at the level of Noble and Daniel [21] or Shilov [31] , and basic undergraduate course on signals and systems at the level of Oppenheim and Willsky [22]. These notes have been used to teach a first year graduate level course (42 hours) in the Department of Electrical Engineering and Computer Science at the University of

Michigan in 1997, 1998, 2000, and 2001 and a one week long short course (40 hours) given at EG&G in Las Vegas in 1998.

The author would like to thank Hyung Soo Kim, Robby Gupta, and Mustafa Demirci for their help with drafting the figures for these notes. He would also like to thank the numerous students who comments led to an improvement of the presentation.

End of chapter

2 PRELIMINARIES

* Vector and matrix concepts

* Signals and systems concepts

REFERENCES

* Noble and Daniel [21]

* Golub and Van Loan [5]

* Harville [8]

* Graybill [6]

* Oppenheim and Willsky [22]

* Proakis and Manolakis [26]

2.1 VECTORS AND MATRICES

2.1.1 ROW AND COLUMN VECTORS

A vector is an ordered list of n values:

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

which resides in $\mathbb{R}^n$.

Convention: in this course $\underline{x}$ is (almost) always a column vector. Its transpose is the row vector

$$\underline{x}^T = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$$

When the elements $x_i = u + jv$ are complex (u, v real valued) the Hermitian transpose is defined as

$$\underline{x}^H = \begin{bmatrix} x_1^* & \cdots & x_n^* \end{bmatrix}$$

where $x_i^* = u - jv$ is the complex conjugate of x_i.

Some common vectors we will see:

$$\underline{1} = [1, \ldots, 1]^T, \qquad \underline{e}_j = [0, \ldots, 0, \underbrace{1}_{j-th}, 0, \ldots 0]^T$$

2.1.2 VECTOR/VECTOR MULTIPLICATION

For 2 vectors $\underline{x}$ and $\underline{y}$ of the same length n their "inner product" is the scalar

$$\underline{x}^T \underline{y} = \sum_{i=1}^{n} x_i y_i$$

For 2 vectors $\underline{x}$ and $\underline{y}$ of possibly different lengths n, m their "outer product" is the $n \times m$ matrix

$$\begin{aligned} \underline{x}\underline{y}^T &= (x_i y_j)_{i,j=1}^{n,m} \\ &= [\underline{x}y_1, \ldots, \underline{x}y_m] \\ &= \begin{bmatrix} x_1y_1 & \cdots & x_1y_m \\ \vdots & \ddots & \vdots \\ x_ny_1 & \cdots & x_ny_m \end{bmatrix} \end{aligned}$$

2.1.3 VECTOR/MATRIX MULTIPLICATION

Let A be an $m \times n$ matrix with columns $\underline{a}_{*1}, \ldots, \underline{a}_{*n}$ and $\underline{x}$ be any n-element vector.
The (compatible) product $A\underline{x}$ is a (column) vector composed of linear combinations of the columns of A

$$A\underline{x} = \sum_{j=1}^{n} x_j \, \underline{a}_{*j}$$

For $\underline{y}$ an m-element vector the product $\underline{y}^T A$ is a (row) vector composed of linear combinations of the rows of A

$$\underline{y}^T A = \sum_{i=1}^{m} y_i \, \underline{a}_{i*}$$

2.1.4 RANK OF A MATRIX

The (column) rank of a matrix A is equal to the number its columns which are linearly independent
* If A has full rank then

$$0 = A\underline{x} = \sum_i x_i \underline{a}_{*i} \;\Leftrightarrow\; \underline{x} = \underline{0}$$

* If in addition A is square then it is said to be non-singular.

2.1.5 MATRIX INVERSION

If A is non-singular square matrix then it has an inverse A^{-1} which satisfies the relation $AA^{-1} = I$
Special case of 2×2 matrix (Cramèr's formula)

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad \text{if} \;\; ad \neq bc$$

Frequently, the inverse of a matrix with special structure has a simple form. The books by Graybill [6] and Golub and VanLoan [5] give many interesting and useful examples. General results which we will encounter in this text are: the *Woodbury-Sherman-Morissey identity*

$$[A + UV^T]^{-1} = A^{-1} - A^{-1}U[I + V^T A^{-1} U^T]^{-1} V^T A^{-1}, \tag{1}$$

where A, U, V are compatible matrices, $[A + UV]^{-1}$ and A^{-1} exist; and the *partitioned matrix identity*

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} [A_{11} - A_{12}A_{22}^{-1}A_{21}]^{-1} & -A_{11}^{-1}A_{12}[A_{22} - A_{21}A_{11}^{-1}A_{12}]^{-1} \\ -A_{22}^{-1}A_{21}[A_{11} - A_{12}A_{22}^{-1}A_{21}]^{-1} & [A_{22} - A_{21}A_{11}^{-1}A_{12}]^{-1} \end{bmatrix}, \tag{2}$$

assuming that all the indicated inverses exist.

2.1.6 EIGENVALUES OF A SYMMETRIC MATRIX

If R is arbitrary $n \times n$ **symmetric matrix**,
* $R^T = R$

then there exist a set of n orthonormal eigenvectors $\underline{\nu}_i$
* $\underline{\nu}_i^T \underline{\nu}_j = \Delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$

and a set of associated eigenvectors λ_i such that:

$$R\underline{\nu}_i = \lambda_i \underline{\nu}_i, \qquad i = 1, \ldots, n$$

Furthermore:
* $\underline{\nu}_i^T R \underline{\nu}_i = \lambda_i$
* $\underline{\nu}_i^T R \underline{\nu}_j = 0, \quad i \neq j$

2.1.7 MATRIX DIAGONALIZATION AND EIGENDECOMPOSITION

Let $U = [\underline{\nu}_1, \ldots, \underline{\nu}_n]$ be the $n \times n$ matrix formed from the eigenvectors of a symmetric matrix R. Then

$$U^T U = I, \qquad (U \text{ is orthogonal matrix})$$

and

$$\begin{aligned} U^T R U &= \Lambda, && \text{(diagonalization of } R) \\ R &= U \Lambda U^T, && \text{(eigendecomposition of } R) \end{aligned}$$

where

$$\Lambda = \mathrm{diag}(\lambda_i) = \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{bmatrix}$$

2.1.8 QUADRATIC FORMS

For a square symmetric matrix R and a compatible vector $\underline{x}$, a quadratic form is the scalar defined by $\underline{x}^T R \underline{x}$ The matrix R is non-negative definite if for any $\underline{x}$

$$\underline{x}^T R \underline{x} \geq 0$$

Examples of positive definite matrices:

* $R = B^T B$ for arbitrary matrix B

* R symmtric with only non-negative eigenvalues

Rayleigh Theorem: If A is a non-negative definite $n \times n$ matrix with eigenvalues $\{\lambda_i\}_{i=1}^n$ the quadratic form

$$\min(\lambda_i) \leq \frac{\underline{u}^T A \underline{u}}{\underline{u}^T \underline{u}} \leq \max(\lambda_i)$$

where the lower bound is attained when $\underline{u}$ is the eigenvector of A associated with the minimum eigenvalue of A and the upper bound is attained by the eigenvector associated with the maximum eigenvalue of A.

2.1.9 DETERMINANT OF A MATRIX

If A is any square matrix its determinant is

$$|A| = \prod_i \lambda_i$$

Note: a square matrix is non-singular iff its determinint is non-zero

If A is partitioned as in (2) and A_{11}^{-1} exists then

$$|A| = |A_{11}||A_{22} - A_{21} A_{11}^{-1} A_{12}| \tag{3}$$

2.1.10 TRACE OF A MATRIX

For any square matrix $A = ((a_{ij}))$ the trace of A is defined as

$$\text{trace}\{A\} = \sum_i a_{ii} = \sum_i \lambda_i$$

Important identity:

$$\text{trace}\{AB\} = \text{trace}\{BA\}$$

Implication for quadratic forms:

$$\underline{x}^T R \underline{x} = \text{trace}\{\underline{x}\underline{x}^T\ R\}$$

2.2 SIGNALS AND SYSTEMS

2.2.1 GEOMETRIC SERIES

$$\sum_{i=0}^{n} a^n = \frac{1-a^{n+1}}{1-a}, \quad \text{if } a \neq 1; \qquad \sum_{i=0}^{\infty} a^n = \frac{1}{1-a}, \quad \text{if } |a| < 1$$

2.2.2 LAPLACE AND FOURIER TRANSFORMS

If $h(t)$, $-\infty < t < \infty$, a square integrable function of a continuous variable t (usually time) then its Laplace and Fourier transforms are defined as follows.

The Laplace transform of h is

$$\mathcal{L}\{h\} = H(s) = \int_{-\infty}^{\infty} h(t)e^{-st}\, dt$$

where $s = \sigma + j\omega \in \mathbb{C}$ is a complex variable.

The Fourier transform of h is

$$\mathcal{F}\{h\} = H(\omega) = \int_{-\infty}^{\infty} h(t)e^{-j\omega t}\, dt$$

Note: $\mathcal{F}\{h\} = \mathcal{L}\{h\}|_{s=j\omega}$.

Example: if $h(t) = e^{-at}u(t)$, for $a > 0$, then the Laplace transform is

$$H(s) = \int_0^{\infty} e^{-at}e^{-st}\, dt = \int_0^{\infty} e^{-(a+s)t}\, dt = \frac{-1}{a+s}e^{-(a+st}\bigg|_0^{\infty} = \frac{1}{a+s}$$

2.2.3 Z-TRANSFORM AND DISCRETE-TIME FOURIER TRANSFORM (DTFT)

If h_k, $k = \ldots, -1, 0, 1, \ldots$, is a square summable function of a discrete variable then its Z-transfrom and discrete-time Fourier transform (DTFT) are defined as follows.

The Z-transform is

$$\mathcal{Z}\{h\} = H(z) = \sum_{k=-\infty}^{\infty} h_k z^{-k}$$

The DTFT is

$$\mathcal{F}\{h\} = H(\omega) = \sum_{k=-\infty}^{\infty} h_k e^{-j\omega k}$$

Note: $H(\omega)$ really means $H(e^{j\omega})$ and is an abuse of notation

- $\mathcal{F}\{h\} = \mathcal{Z}\{h\}|_{z=e^{j\omega}}$
- the DTFT is always periodic in ω with period 2π.

Example: if $h_k = a^{|k|}$, then for $|az^{-1}| < 1$, the Z-transform is

$$H(z) = \sum_{k=-\infty}^{\infty} a^{|k|} z^{-k} = \sum_{k=-\infty}^{-1} a^{-k} z^{-k} + \sum_{k=0}^{\infty} a^k z^{-k}$$

$$= \sum_{k=1}^{\infty} (az)^k + \sum_{k=0}^{\infty} (az^{-1})^k = \frac{az^{-1}}{1 - az^{-1}} + \frac{1}{1 - az^{-1}}$$

Likewise the DTFT is (for $|a| < 1$):

$$H(\omega) = H(z)|_{z=e^{j\omega}} = \frac{1 - a^2}{1 - 2a\cos\omega + a^2}$$

2.2.4 CONVOLUTION: CONTINUOUS TIME

If $h(t)$ and $x(t)$ are square integrable functions of a continuous variable t then the convolution of x and h is defined as

$$(h * x)(t) = \int_{-\infty}^{\infty} h(t - \tau) x(\tau)\, d\tau$$

Note: The convolution of h and x is a waveform indexed by time t. $(h * x)(t)$ is this waveform evaluated at time t and is frequently denoted $h(t) * x(t)$.

Example: $h(t) = e^{-at} u(t)$, for $a > 0$, (the filter) and $x(t) = e^{-bt} u(t)$, for $b > 0$, (the filter input) then

$$(h * x)(t) = \int_{-\infty}^{\infty} e^{-a(t-\tau)} e^{-b\tau} u(t - \tau) u(\tau)\, d\tau = \left(\int_0^t e^{-a(t-\tau)} e^{-b\tau}\, d\tau \right) u(t)$$

$$= e^{-at} \left(\int_0^t e^{-(b-a)\tau}\, d\tau \right) u(t) = e^{-at} \left(\frac{-1}{b-a} e^{-(b-a)\tau} \Big|_0^t \right) u(t) = \frac{e^{-at} - e^{-bt}}{b - a} u(t)$$

2.2.5 CONVOLUTION: DISCRETE TIME

If h_k and x_k are square integrable sequences then

$$h_n * x_n = \sum_{j=-\infty}^{\infty} h_j x_{n-j} = \sum_{j=-\infty}^{\infty} h_{n-j} x_j$$

h_k is a called a "causal" filter if it is zero for negative indices:

$$h_k = 0, \quad k < 0$$

2.2.6 CORRELATION: DISCRETE TIME

For time sequences $\{x_k\}_{k=1}^n$ and $\{y_k\}_{k=1}^n$ their temporal correlation is

$$z_n = \sum_{j=1}^{n} x_k y_k^*$$

2.2.7 RELATION BETWEEN CORRELATION AND CONVOLUTION

$$z_n = \sum_{j=1}^{n} x_k y_k^* = \sum_{j=-\infty}^{\infty} x_k h_{n-k} = h_n \star x_n$$

where

$$h_k = \begin{cases} y_{n-k}^*, & k = 1, \dots, n \\ 0, & o.w. \end{cases}$$

2.2.8 CONVOLUTION AS A MATRIX OPERATION

Let h_k be a causal filter and let x_k be an input starting at time $k = 0$. Arranging n outputs z_k in a vector $\underline{z}$ it is easy to see that

$$\begin{aligned}
\underline{z} = \begin{bmatrix} z_n \\ \vdots \\ z_1 \end{bmatrix} &= \begin{bmatrix} \sum_{j=1}^{n} h_{n-j} x_j \\ \vdots \\ \sum_{j=1}^{n} h_{1-j} x_j \end{bmatrix} \\
&= \begin{bmatrix} h_0 & h_1 & \cdots & h_{n-1} \\ 0 & h_0 & \ddots & h_{n-2} \\ \vdots & \ddots & h_0 & h_1 \\ 0 & \cdots & 0 & h_0 \end{bmatrix} \begin{bmatrix} x_n \\ \vdots \\ x_1 \end{bmatrix}
\end{aligned}$$

2.3 Exercises

1. Let $\underline{a}$, $\underline{b}$ be $n \times 1$ vectors and let C be an invertible $n \times n$ matrix. Assuming α is not equal to $-1/(\underline{a}^T C^{-1} \underline{b})$ show the following identity

$$[C + \alpha \underline{a}\underline{b}^T]^{-1} = C^{-1} - C^{-1}\underline{a}\underline{b}^T C^{-1} \alpha/(1 + \alpha \underline{a}^T C^{-1} \underline{b}).$$

2. A discrete time LTI filter $h(k)$ is causal when $h(k) = 0,\ k < 0$ and anticausal when $h(k) = 0,\ k > 0$. Show that if $|h(k)| < \infty$ for all k, the transfer function $H(z) = \sum_{k=-\infty}^{\infty} h(k) z^{-k}$ of a causal LTI has no singularities outside the unit circle, i.e. $|H(z)| < \infty$, $|z| > 1$ while an anticausal LTI has no singularities inside the unit circle, i.e. $|H(z)| < \infty$, $|z| < 1$. (Hint: generalized triangle inequality $|\sum_i a_i| \leq \sum |a_i|$)
3. A discrete time LTI filter $h(k)$ is said to be BIBO stable when $\sum_{k=-\infty}^{\infty} |h(k)| < \infty$. Define the transfer function (Z-transform) $H(z) = \sum_{k=-\infty}^{\infty} h(k) z^{-k}$, for z a complex variable.
   (a) Show that $H(z)$ has no singularites on the unit circle, i.e $|H(z)| < \infty$, $|z| = 1$.
   (b) Show that if a BIBO stable $h(k)$ is causal then $H(z)$ has all its singularities (poles) strictly inside the unit circle, i.e $|H(z)| < \infty$, $|z| \geq 1$.
   (c) Show that if a BIBO stable $h(k)$ is anticausal, i.e. $h(k) = 0,\ k > 0$, then $H(z)$ has all its singularities (poles) strictly outside the unit circle, i.e $|H(z)| < \infty$, $|z| \leq 1$.

4. If you are only given the mathematical form of the transfer function $H(z)$ of an LTI, and not told whether it correponds to an LTI whichis causal, anticausal, or stable, then it is not possible to uniquely specify the impulse response $\{h_k\}_k$. This simple example illustration this fact.
   Let $H(z)$ be
   $$H(z) = \frac{1}{1 - az^{-1}}$$
   (a) Show that if the LTI is causal, then for $|z| > |a|$ you can write $H(z)$ as the convergent series
   $$H(z) = \sum_{k=0}^{\infty} a^k z^{-k}, \quad |z| > |a|$$
   which corresponds to $h_k = a^k,\ k = 0, 1, \ldots$ and $h_k = 0,\ k < 0$.
   (b) Show that if the LTI is anticausal, then for $|z| < |a|$ you can write $H(z)$ as the convergent series
   $$H(z) = -\sum_{k=0}^{\infty} a^{-k} z^{k+1}, \quad |z| < |a|$$
   which corresponds to $h_k = -a^{-k},\ k = 1, 2 \ldots$ and $h_k = 0,\ k \geq 0$.
   (c) Show that if $|a| < 1$ then the causal LTI is BIBO stable while the anti-causal LTI is BIBO unstable while if $|a| > 1$ then the reverse is true. What happens to stability when $|a| = 1$?
5. An LTI has transfer function
   $$H(z) = \frac{3 - 4z^{-1}}{1 - 3.5z^{-1} + 1.5z^{-2}}$$
   (a) If you are told that the LTI is stable specify the region of convergence (ROC) in the z-plane, i.e. specify the range of values of $|z|$ for which $|H(z)| < \infty$, and specify the impulse response.
   (b) If you are told that the LTI is causal specify the region of convergence (ROC) in the z-plane, and specify the impulse response.
   (c) If you are told that the LTI is anticausal specify the region of convergence (ROC) in the z-plane, and specify the impulse response.

End of chapter

3 STATISTICAL MODELS

Estimation, detection and classification can be grouped under the broad heading of statistical inference which is the process of inferring properties about the distribution of a random variable X given a realization x, which is also called a data sample, a measurement, or an observation. A key concept is that of the statistical model which is simply a hypothesized probability distribution or density function $f(x)$ for X. Broadly stated statistical inference explores the possibility of fitting a given model to the data x. To simplify this task it is common to restrict $f(x)$ to a class of *parameteric models* $\{f(x;\underline{\theta})\}_{\underline{\theta}\in\Theta}$, where $f(x;\bullet)$ is a known function and $\underline{\theta}$ is a vector of unknown parameters taking values in a parameter space Θ. In this special case statistical inference boils down to inferring properties of the *true value* of $\underline{\theta}$ parameterizing $f(x;\underline{\theta})$ that generated the data sample x.

In this chapter we discuss several models that are related to the ubiquitous Gaussian distribution, the more general class of exponential families of distributions, and the important concept of a sufficient statistic for infering properties about $\underline{\theta}$.

REFERENCES:

* Rao [27]

* Manoukian [18]

* Mood, Graybill and Boes [19]

* Johnson *etal* [13]

3.1 DISTRIBUTIONS RELATED TO GAUSSIAN

The Gaussian distribution and its close relatives play a major role in parameteric statistical inference due to the relative simplicity of the Gaussian model and its broad applicability (recall the Central Limit Theorem!). In referring to such models in this book the following compatible notations are used:

* "x is distributed as a Gaussian random variable with mean μ and variance σ^2"

$$x \sim \mathcal{N}(\mu,\sigma^2)$$

* "x is equal to a scaled and shifted standard Gaussian random variable"

$$x = a\,\mathcal{N}(0,1)+b \Leftrightarrow x \sim \mathcal{N}(b,a^2)$$

* "The CDF of $\mathcal{N}(0,1)$ equals α when evaluated at a point $v\in\mathbb{R}$"

$$\mathcal{N}(v)=\alpha$$

* "The inverse CDF of $\mathcal{N}(0,1)$ equals v when evaluated at a point $\alpha\in[0,1]$"

$$\mathcal{N}^{-1}(\alpha)=v$$

* "$\underline{x}$ is distributed as an n dimensional Gaussian random vector with mean $\underline{\mu}$ covariance matrix R"

$$\underline{x} \sim \mathcal{N}_n(\underline{\mu},\ R)$$

In the following $x_1, \ldots, x_n$ are i.i.d. $\mathcal{N}(0,1)$
Shorthand notation

$$\sum_{i=1}^n \mathcal{N}(0,1) = \sum_{i=1}^n x_i$$

* The CDF $\mathcal{N}(x)$ is related to the error function or error integral [1]: $\mathrm{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-t^2} dt,\ x \geq 0$, through the relation

$$\mathcal{N}(x) = \begin{cases} \frac{1}{2}[1 + \mathrm{erf}(|x|/\sqrt{2})] & x \geq 0 \\ \frac{1}{2}[1 - \mathrm{erf}(|x|/\sqrt{2})], & x < 0 \end{cases} .$$

3.1.1 CHI-SQUARE

The sum of the squares of n i.i.d. $\mathcal{N}(0,1)$'s is (central) **Chi-square** with n degrees of freedom

$$\sum_{i=1}^n [\mathcal{N}(0,1)]^2 = \chi_n$$

* $E[\chi_n] = n, \quad \mathrm{var}(\chi_n) = 2n$
* Asymptotic relation for large n:

$$\chi_n = \sqrt{2n}\mathcal{N}(0,1) + n$$

* χ_2 an exponential r.v. with mean 2, i.e. $X = \chi_2$ is a non-negative r.v. with probability density $f(x) = \frac{1}{2}e^{-x/2}$.
* $\sqrt{\chi_2}$ is a Rayleigh random variable.

3.1.2 NON-CENTRAL CHI SQUARE

The sum of squares of independent Gaussian r.v.s with unit variances but non-zero means is called a **non-central Chi-square** with n degrees of freedom and non-centrality parameter $\delta = \sum_{i=1}^n \mu_i^2$

$$\sum_{i=1}^n [\mathcal{N}(0,1) + \mu_1]^2 = \chi_{n,\delta} = \sum_{i=1}^n [\mathcal{N}(\mu_i, 1)]^2 = \chi_{n,\delta}$$

* $E[\chi_{n,\delta}] = n + \delta, \quad \mathrm{var}(\chi_{n,\delta}) = 2(n + 2\delta)$
* $\sqrt{\chi_{2,\mu_1^2+\mu_2^2}}$ is a Rician r.v.

3.1.3 CHI-SQUARE MIXTURE

The following is called a (central) **Chi-square mixture** (Chi-bar square) with n degrees of freedom and mixture parameter $\underline{c} = [c_1, \ldots, c_n]^T$

$$\sum_{i=1}^{n} \frac{c_i}{\sum_j c_j} [\mathcal{N}(0,1)]^2 = \overline{\chi}_{n,\underline{c}}$$

* $E[\overline{\chi}_{n,\underline{c}}] = 1$, , var$(\overline{\chi}_{n,\underline{c}}) = 2\sum_{i=1}^{N} \left(\frac{c_i}{\sum_j c_i}\right)^2$

* Property: $\overline{\chi}_{n,c\underline{1}} = \frac{1}{n}\, \chi_n$ for any $c \neq 0$

3.1.4 STUDENT-T

For independent $\mathcal{N}(0,1)$ and χ_n r.v.s the following is called a **Student-t** with n degrees of freedom

$$\frac{\mathcal{N}(0,1)}{\sqrt{\chi_n/n}} = \mathcal{T}_n$$

* $E[\mathcal{T}_n] = 0$, , var$(\mathcal{T}_n) = \frac{n}{n-2}$ $(n > 2)$

* Asymptotic relation for large n:

$$\mathcal{T}_n \approx \mathcal{N}(0,1)$$

3.1.5 FISHER-F

For independent χ_m and χ_n r.v.s the following is called a **Fisher-F** r.v. with m, n degrees of freedom

$$\frac{\chi_m/m}{\chi_n/n} = \mathcal{F}_{m,n}$$

* moments of order $> n/2$ do not exist

* Asymptotic relation for n large and $n \gg m$:

$$\mathcal{F}_{m,n} \approx \chi_m$$

3.1.6 CAUCHY

For independent $\mathcal{N}(0,1)$ r.v.'s U and V the following has a **Cauchy** distribution

$$U/V = \mathcal{C}$$

* Cauchy distribution has no moments of order greater than 1.

* Cauchy is the same as a Student-t with 1 d.f.

3.1.7 BETA

For independent χ_m and χ_n r.v.s Y_m and Y_n the following ratio has a **Beta** distribution

$$\frac{\chi_m}{\chi_m + \chi_n} = \mathcal{B}(m/2, n/2)$$

where $B(p,q)$ is a r.v. with a Beta density $f(x) = c_{p,q} x^{p-1}(1-x)^{q-1}$, $x \in [0,1]$ and $c_{p,q} > 0$ a normalizing constant. The special case of $m = n = 1$ gives rise to an **arcsin** distributed r.v.

3.2 FISHER-COCHRAN THEOREM

Let

* $\underline{x} = [x_1, \ldots, x_n]^T$ be a vector of i.i.d. $\mathcal{N}(0,1)$'s

* A be a symmetric idempotent matrix ($AA = A$) of rank p

Then

$$\underline{x}^T A \underline{x} \;=\; \chi_p$$

Proof: Let $A = U^T \Lambda U$ be eigendecomposition of A. Then

* All eigenvalues λ_i of A are either 0 or 1

$$\begin{aligned} AA &= U^T \Lambda \underbrace{U U^T}_{=I} \Lambda U \\ &= U^T \Lambda^2 U = U^T \Lambda U \end{aligned}$$

and therefore

$$\begin{aligned} \underline{x}^T A \underline{x} &= \underline{x}^T U^T \Lambda \underbrace{U \underline{x}}_{\underline{z} = \mathcal{N}_n(0,I)} \\ &= \sum_{i=1}^n \lambda_i z_i^2 = \sum_{i=1}^p [\mathcal{N}(0,1)]^2 \end{aligned}$$

3.3 SAMPLE MEAN AND SAMPLE VARIANCE

Let x_i's be i.i.d. $\mathcal{N}(0,1)$ r.v.'s

* Sample mean: $\overline{x_i} = n^{-1} \sum_{i=1}^n x_i$

* Sample variance: $\mathsf{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x_i})^2$

Then

1. $\overline{x_i} \;=\; \mathcal{N}(0, 1/n)$

2. $\mathsf{s}^2 \;=\; \frac{1}{n-1}\,\chi_{n-1}$

3. $\overline{x_i}$ and s^2 are independent r.v.s.

Implication:

$$\frac{\overline{x_i}}{\mathsf{s}/\sqrt{n}} = \mathcal{T}_{n-1}$$

Proof: Express $\overline{x_i}$ and s^2 in vector form

$$\begin{aligned} \overline{x_i} &= \frac{1}{n}\,\underline{1}^T\underline{x} \\ \mathsf{s}^2 &= \frac{1}{n-1}\,[\underline{x}-\underline{1}\overline{x_i}]^T[\underline{x}-\underline{1}\overline{x_i}] \end{aligned}$$

$$\underline{x}-\underline{1}\overline{x_i} = [x_1-\overline{x_i},\ldots,x_n-\overline{x_i}]^T$$

$$\underline{1} = [1,\ldots,1]^T$$

Observation 1: $[\underline{x}-\overline{x_i}\underline{1}]$ and $\overline{x_i}\underline{1}$ are orthogonal random vectors

$$E[\overline{x_i}\underline{1}^T(\underline{x}-\underline{1}\overline{x_i})] \quad = \quad E[\overline{x_i}(\underbrace{\underline{1}^T\underline{x}}_{n\overline{x_i}} - \underbrace{\underline{1}^T\underline{1}}_{=n}\,\overline{x_i})] = 0$$

$\Rightarrow x_1-\overline{x_i},\ldots,x_n-\overline{x_i}$ and $\overline{x_i}$ are actually independent r.v.s!

Observation 2: Since independence is preserved under arbitrary transformations

$\Rightarrow \sum_{j=1}^n (x_j-\overline{x_i})^2$ and $\overline{x_i}$ are also independent r.v.s!

Observation 3: using $\overline{x_i} = n^{-1}\,\underline{1}^T\underline{x}$

$$\begin{aligned} [\underline{x}-\underline{1}\overline{x_i}]^T[\underline{x}-\underline{1}\overline{x_i}] &= \underline{x}^T\underbrace{[I-\underline{1}\underline{1}^T\frac{1}{n}][I-\underline{1}\underline{1}^T\frac{1}{n}]}_{\text{idempotent}}\underline{x}^T \\ &= \underline{x}^T\underbrace{[I-\underline{1}\underline{1}^T\frac{1}{n}]}_{\text{orth. proj.}}\underline{x} \end{aligned}$$

* rank$[I-\underline{1}\underline{1}^T\frac{1}{n}] = n-1$

$\Rightarrow$ By Fisher-Cochran $[\underline{x}-\underline{1}\overline{x_i}]^T[\underline{x}-\underline{1}\overline{x_i}] = (n-1)\,\mathsf{s}^2$ is χ_{n-1} ◇

3.4 SUFFICIENT STATISTICS

Common structure of many detection/estimation/classification problems

* $\underline{x} = [x_1, \ldots, x_n]^T$: measurement or observation

* $\mathcal{X}$ is sample or measurement space

* $\theta \in \Theta$ is an unknown parameter vector of interest

* Θ is parameter space

* $f_\theta(\underline{x}) = f(\underline{x}; \theta)$ is p.d.f. of $\underline{x}$ for given θ (a known function)

OBJECTIVE: Given $\underline{x}$ infer properties of θ

3.4.1 SUFFICIENT STATISTICS AND THE COMPRESSION RATIO

Definition: A statistic $T = T(\underline{x})$ is any function of the data

Interpretation in terms of memory storage requirements: If function $\underline{x} \to T(\underline{x})$ not invertible, then T is a "compression" of the data sample into a lower dimension.

$$\text{Compression of } \underline{x} \;\Leftrightarrow\; \text{Function of } \underline{x}$$

* Compression ratio = $r_c = \dfrac{\text{\# bytes of storage required for } T(\underline{x})}{\text{\# bytes of storage required for } \underline{x}}$

Note that each element x_i of the data vector $\underline{x} = [x_1, \ldots, x_n]^T$ consists of a value (voltage, current, counter value, etc.) and a "time stamp" or index i both of which must be stored. Therefore

* # bytes$\{\underline{x}\}$ = # bytes$[x_1, \ldots, x_n]^T$ = # bytes$\{indices\}$ + # bytes$\{values\}$ $= 2n$

Examples: ($x_{(i)} = i$-th largest element of $\underline{x}$, i.e. $x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(n)}$)

$T(\underline{x}) = [x_1, \ldots, x_n]^T$,	entire data sample	$r_c = 1$
$T(\underline{x}) = [x_{(1)}, \ldots, x_{(n)}]^T$,	rank ordered sample (time stamp removed)	$r_c = 1/2$
$T(\underline{x}) = \overline{x_i}$,	sample mean	$r_c = 1/(2n)$
$T(\underline{x}) = [\overline{x_i}, \mathsf{s}^2]^T$,	sample mean,var	$r_c = 1/n$

A natural question is: what is the maximal compression ratio one can get away with without loss of information about θ? The answer is: compression to a minimal sufficient statistic. We first need to define a sufficient statistic.

3.4.2 DEFINITION OF SUFFICIENCY

Definition: $T = T(\underline{x})$ is **sufficient statistic** for θ if it captures all the information about θ in the data sample.

Specifically, T is a suff. stat. (SS) if it satisfies:

$$f_\theta(\underline{x}|T(\underline{x})) = f(\underline{x}|T(\underline{x})), \qquad \text{indep. of } \theta$$

Terminology: if $T(\underline{x}) = [T_1(\underline{x}), \ldots, T_K(\underline{x})]^T$ satisfies the above we say that T_k's are **jointly sufficient** for θ

An alternative (equivalent) definition of a sufficient statistic [27]:

Fisher factorization (FF): $T = T(\underline{x})$ is a sufficient statistic for θ if the distribution $f_\theta(\underline{x})$ has the representation

$$f_\theta(\underline{x}) = g(T, \theta)\ h(\underline{x}) \tag{4}$$

for some non-negative functions g and h.

An important implication of the Fisher Factorization is that when the density function of a sample $\underline{x}$ satisfies (4) then the density $f_T(t,\theta)$ of the sufficient statistic T is equal to $g(t,\theta)$ up to a θ-independent constant $q(t)$ (see exercises at end of this chapter):

$$f_T(t;\theta) = g(t,\theta)q(t).$$

Examples of sufficient statistics:

Ex 1: $\underline{x} = [x_1, \ldots, x_n]^T$ is always sufficient

Ex 2: $x_{(1)}, \ldots, x_{(n)}$ is sufficient when x_i's i.i.d.

Proof: Since x_i's are i.i.d., we have

$$f(x_1, \ldots, x_n) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n f(x_{(i)})$$

hence sufficiency of the rank ordered sample $x_{(1)}, \ldots, x_{(n)}$ follows from Fisher factorization.

Ex 3: For $\Theta = \{0, 1\}$ the "likelihood ratio" $\Lambda(\underline{x}) = f_1(\underline{x})/f_0(\underline{x})$ is sufficient for θ.

Proof: Express $f_\theta(\underline{x})$ as function of θ, f_0, f_1, factor out f_0, identify Λ, and invoke FF

$$\begin{aligned} f_\theta(\underline{x}) &= \theta f_1(\underline{x}) + (1-\theta) f_0(\underline{x}) \\ &= \left(\underbrace{\theta\Lambda(\underline{x}) + (1-\theta)}_{g(T,\theta)} \right) \underbrace{f_0(\underline{x})}_{h(\underline{x})} \end{aligned}$$

$\Rightarrow$ to discriminate between two values of θ_1 and θ_2 in Θ we can throw away all data except for $\Lambda(\underline{x})$

Ex 4: For $\Theta = \{\theta_0, \ldots, \theta_K\}$ the vector of K likelihood ratios

$$\left[\frac{f_{\theta_1}(\underline{x})}{f_{\theta_0}(\underline{x})}, \ldots, \frac{f_{\theta_K}(\underline{x})}{f_{\theta_0}(\underline{x})} \right]$$

is sufficient for θ

Ex. 5: For any Θ the likelihood ratio trajectory

$$\Lambda(\theta) = \frac{f_\theta(\underline{x})}{f_{\theta_0}(\underline{x})}, \quad \theta \in \Theta$$

is sufficient for θ. Here θ_0 is arbitrary reference point in Θ.

3.4.3 MINIMAL SUFFICIENCY

What is the maximum possible amount of compression one can apply to the data sample without losing information concerning how the model depends on θ? The answer to this question lies in the notion of a minimal sufficient statistic. Such a statistic cannot be compressed any further without loss in information. In other words, any other sufficient statistic can be compressed down to a minimal sufficient statistic.

Definition: a sufficient statistic $T(\underline{x})$ is said to be **minimal** if it is a compression of any other sufficient statistic, i.e. T_{min} is minimally sufficient if for any sufficient statistic T there exists a function q such that $T_{min} = q(T)$.

Another important concept in practical applications is that of finite dimensionality.

Definition: a sufficient statistic $T(\underline{x})$ is said to be **finite dimensional** if its dimension is not a function of the number of data samples n.

Frequently, but not always (see Cauchy example below), minimal sufficient statistics are finite dimensional.

REMARKS:

* Minimal suff. stats are not unique: e.g., $g(T_{min})$ is minimally suff. if g is any invertible function.

* The likelihood ratio trajectory $\{\Lambda(\theta)\}_{\theta\in\Theta}$ is minimally sufficient

Proof:

Let T be any suff. stat. Then by FF $f_\theta(\underline{x}) = g(T, \theta)\ h(\underline{x})$.

Now pick arbitrary point $\theta = \theta_0$ out of Θ. Then

$$\Lambda(\theta) = \frac{f_\theta(\underline{x})}{f_{\theta_0}(\underline{x})} = \frac{g(T,\theta)}{g(T,\theta_0)} = \text{fun. of } T$$

Example 1 *Sufficient statistic for Gaussian mean and variance*

Assume $x \sim \mathcal{N}(\mu, \sigma^2)$. Find Suff. Stat. for $\theta = [\mu, \sigma^2]^T$

$$\begin{aligned}
f_\theta(\underline{x}) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + \mu^2\right)} \\
&= \underbrace{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{\mu^2}{2\sigma^2}}\ e^{[\mu/\sigma^2,\ -1/(2\sigma^2)]\overbrace{\left[\sum_{i=1}^n x_i,\ \sum_{i=1}^n x_i^2\right]^T}^{\underline{T}(\underline{x})}}}_{g(\underline{T},\theta)}\ \underbrace{1}_{h(\underline{x})}
\end{aligned}$$

Thus

$$\underline{T} = \left[\underbrace{\sum_{i=1}^n x_i}_{T_1},\ \underbrace{\sum_{i=1}^n x_i^2}_{T_2}\right]$$

is a (jointly) sufficient statistic for μ, σ^2.

Furthermore, as $q(\underline{T}) = [n^{-1}T_1, T_2 - n^{-1}T_1^2]$ is a 1-1 function of $\underline{T}$ ($\underline{T} = [T_1, T_2]^T$)

$$\underline{S} = \left[\overline{x_i},\ \mathsf{s}^2\right]$$

is an equivalent sufficient statistic

Example 2 *Cauchy distribution*

$x \ \sim \ f(x;\theta) = \frac{1}{\pi} \frac{1}{1+(x-\theta)^2}$

Then

$$f(\underline{x};\theta) = \prod_{i=1}^n \frac{1}{\pi} \frac{1}{1+(x_i-\theta)^2} = \frac{1}{\pi^n} \frac{1}{\prod_{i=1}^n (1+(x_i-\theta)^2)}$$

Difficulty: denominator is a $2n$-degree polynomial in θ whose roots depend on all cross products $x_{i_1} \ldots x_{i_p}$, $p = 1, 2, \ldots, n$, of $x_i's$.

$\Rightarrow$ Minimal suff. stat. is order statistic $[x_{(1)}, \ldots, x_{(n)}]^T$

$\Rightarrow$ No finite dimensional sufficent statistic exists

Example 3 *Exponential family of distributions*

Let $\underline{\theta} = [\theta_1, \ldots, \theta_p]^T$. The distribution of a r.v. x is a member of the p parameter exponential family if

$$f_{\underline{\theta}}(x) = a(\underline{\theta}) b(x) e^{-\underline{c}^T(\underline{\theta}) \underline{t}(x)}, \quad -\infty < x < \infty$$

for some scalar functions a, b and some p-element vector functions $\underline{c}, \underline{t}$. Note that for any $f_{\underline{\theta}}$ in the exponential family its support set $\{x : f_{\underline{\theta}}(x) > 0\}$ does not depend on $\underline{\theta}$.

Examples include: Gaussian with unknown mean or variance, Poisson with unknown mean, exponential with unknown mean, gamma, Bernoulli with unknown success probability, binomial with unknown success probability, multinomial with unknown cell probabilities.

Examples do not include: Cauchy with unknown median, uniform of unknown extent, Fisher-F with unknown degrees of freedom.

Sufficient statistics for exponential family members have a particularly simple form:

$$\begin{aligned} f_{\underline{\theta}}(\underline{x}) &= \prod_{i=1}^n a(\underline{\theta}) b(x_i) e^{-\underline{c}^T(\underline{\theta}) \underline{t}(x_i)} \\ &= \underbrace{a^n(\underline{\theta})\ e^{-\underline{c}^T(\underline{\theta}) \overbrace{\sum_{i=1}^n \underline{t}(x_i)}^{\underline{T}}}}_{g(\underline{T},\underline{\theta})} \underbrace{\prod_{i=1}^n b(x_i)}_{h(\underline{x})} \end{aligned}$$

Therefore, the following is a p-dimensional sufficient statistic for $\underline{\theta}$

$$\sum_{i=1}^n \underline{t}(x_i) = \left[\sum_{i=1}^n t_1(x_i), \ldots, \sum_{i=1}^n t_p(x_i)\right]^T$$

In fact this is a finite dimensional suff. statistic which is minimal.

3.5 Exercises

1. Show that the matrix $\Pi = I_n - \underline{1}\underline{1}^T/n$ is symmetric and idempotent, where I_n is the $n \times n$ identity matrix and $\underline{1} = [1, \ldots, 1]^T$ is an n-element column vector of 1's. Show that for $\underline{x} \in \mathbb{R}^n$, $\pi\underline{x}$ is the vector of residuals $[x_1 - \overline{x_i}, \ldots, x_n - \overline{x_i}]^T$ where $\overline{x_i}$ is the sample mean of elements of $\underline{x}$. Finally show that if $\underline{x}$ has the decomposition $\underline{y} + c\underline{1}$ where $\underline{y}$ has zero (sample) mean and c is an arbitrary scalar, then $\Pi\underline{x} = \underline{y}$, i.e the matrix Π extracts the zero (sample) mean component of $\underline{x}$. It is in this sense that Π is an orthogonal projection matrix onto the space of zero (sample) mean vectors in $\mathbb{R}^n$.

2. Assume that a random vector $\underline{X} = [X_1, \ldots, X_n]^T$ has a density $p_{\underline{\theta}}(\underline{x})$ which depends on an unknown parameter vector $\underline{\theta}$. In this exercise you will show that if a statistic $\underline{S} = \underline{S}(\underline{X}) = [S_1(\underline{X}), \ldots, S_k(\underline{X})]^T$ satisfies the Fisher Factorization theorem then the conditional density $p_{\underline{\theta}}(\underline{X}|\underline{S})$ is not a function of $\underline{\theta}$ and thus $\underline{S}$ is a sufficient statistic for $\underline{\theta}$. In the following you should assume that $\underline{X}$ is a discrete random vector and that its joint density $p_{\underline{\theta}}(\underline{x}) = P_{\underline{\theta}}(\underline{X} = \underline{x})$ is a probability mass function (i.e. $p_{\underline{\theta}}(\underline{x}) = 0$ except for a countable number of points $\underline{x} \in \{\underline{x}_1, \underline{x}_2, \ldots\}$ where $p_{\underline{\theta}}(\underline{x}_i) > 0$, and $\sum_{\underline{x}_i} p_{\underline{\theta}}(\underline{x}_i) = 1$).

   (a) Use Bayes rule to establish that

   $$p_\theta(\underline{x}|\underline{s}) \stackrel{\text{def}}{=} P_{\underline{\theta}}(\underline{X} = \underline{x}|\underline{S} = \underline{s}) = \frac{P_{\underline{\theta}}(\underline{S} = \underline{s}|\underline{X} = \underline{x}) p_{\underline{\theta}}(\underline{x})}{\sum_{\underline{x}_i\,:\,\underline{S}(\underline{x}_i) = \underline{s}} p_{\underline{\theta}}(\underline{x}_i)},$$

   where the summation of $p_{\underline{\theta}}(\underline{x})$ is over all possible realizations $\{\underline{x}_i\}$ of the vector $\underline{X}$ such that $\underline{S}(\underline{x}_i) = \underline{s}$.

   (b) Show that $P_{\underline{\theta}}(\underline{S} = \underline{s}|\underline{X} = \underline{x})$ is equal to one or zero depending on whether $\underline{S}(\underline{x}) = \underline{s}$ or $\underline{S}(\underline{x}) \neq \underline{s}$, respectively. (Hint: express the conditional probability as a ratio and use the definition $\underline{S} = \underline{S}(\underline{X})$ to evaluate the intersection of the events $\underline{S} = \underline{s}$ and $\underline{X} = \underline{x}$).

   (c) Using the Fisher Factorization $p_{\underline{\theta}}(\underline{x}) = g_{\underline{\theta}}(\underline{s}) \cdot h(\underline{x}) \cdot m_{\underline{\theta}}$ show that

   $$p_\theta(\underline{x}|\underline{s}) = \begin{cases} \frac{h(\underline{x})}{\sum_{\underline{x}_i\,:\,\underline{S}(\underline{x}_i) = \underline{s}} h(\underline{x}_i)}, & \underline{S}(\underline{x}) = \underline{s} \\ 0, & o.w. \end{cases},$$

   which, as claimed, does not depend on $\underline{\theta}$.

3. Show that the Poisson distribution $p_\lambda(x) = P_\lambda(X = x) = \frac{\lambda^x}{x!}\exp(-\lambda)$, $x = 0, 1, 2, \ldots$ is a member of the one-parameter exponential family. For an i.i.d. sample $\underline{X} = [X_1, \ldots, X_n]^T$ of these Poisson r.v.s find a one dimensional sufficient statistic for λ.

4. Let $\underline{X} = [X_1, \ldots, X_n]^T$ be a vector of i.i.d. r.v.s X_i which are uniformly distributed over the interval (θ_1, θ_2), $\theta_1 < \theta_2$. Show that $\underline{S}(\underline{X}) = [\min_i\{X_i\}, \max_i\{X_i\}]^T$ is a sufficient statistic for $\underline{\theta} = [\theta_1, \theta_2]^T$.

5. Let Z_i, $i = 1, \ldots, n$, be a set of i.i.d. random variables each with the *alpha density*

$$p_\theta(z) = \frac{\beta}{\sqrt{2\pi}\Phi(\alpha) z^2} \exp\left(-\frac{1}{2}\left[\alpha - \beta/z\right]^2\right),$$

where $\beta > 0$ is unknown, α is known and $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$ is the standard normal CDF. The alpha distribution is often used to model tool wear for rotating machinery.

   (a) Is the joint density $p_\theta(\underline{z})$ a member of the exponential family of densities?

   (b) using the Fisher Factorization find a two dimensional sufficient statistic for estimating the parameter β based on the observation $\underline{Z} = [Z_1, \ldots, Z_n]^T$. Show that this reduces to a one dimensional (scalar) statistic when $\alpha = 0$.

6. Let $\underline{X} = [X_1, \ldots, X_n]^T$ be a vector of i.i.d. Gaussian r.v.s with mean μ and variance $\sigma^2 = \mu^2$ $(X_i \sim \mathcal{N}(\mu, \mu^2))$.

   (a) Show that the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is *not* a sufficient statistic for μ by demonstrating that the conditional jpdf of $\underline{X}$ given $\overline{X}$ is a function of μ.

   (b) Find a two dimensional sufficient statistic.

   (c) Let $T = T(\underline{x})$ be a sufficient statistic for θ, where $\underline{x} \sim f(\underline{x}; \theta) = g(T(\underline{x}), \theta) h(\underline{x})$ is a discrete random variable. Show that T has probability mass function

$$f_T(t; \theta) = g(t, \theta) q(t),$$

   where

$$q(t) = \sum_{\{\underline{x}: T(\underline{x}) = t\}} h(\underline{x}).$$

End of chapter

4 FUNDAMENTALS OF ESTIMATION

* Mathematical formulation of estimation problems
* Estimation of random parameters: Bayes estimates
* MSE, MAE, Min P_e estimators
* Estimation of non-random parameters
* Method of moments, maximum likelihood
* Optimal unbiased estimators and the CR bound

REFERENCES:

INTRODUCTORY TEXTS:

Mood, Graybill and Boes [19]

Van Trees [36]

Srinath, Rajasekaran and Viswanathan [33]

Scharf [29]

MORE ADVANCED:

Bickel and Doksum [2]

Lehmann [16]

Ibragimov and Has'minskii [12]

Poor [24]

4.1 PROBLEM DESCRIPTION

$X \in \mathcal{X}$ is a measurement or observation

$\mathcal{X}$ is the sample or measurement space

$\theta \in \Theta$ is an unknown parameter vector of interest

$\Theta \subset \mathbb{R}^p$ is the parameter space

$f(x;\theta)$ is p.d.f. of X for given θ (a known function)

Estimation objective: design an estimator function

$$\hat{\theta} = \hat{\theta}(x)$$

which maps $\mathcal{X}$ to $\mathbb{R}^p \supset \Theta$

Terminology:

* the function $\hat{\theta}$ is an *estimator*
* the point $\hat{\theta}(x)$ is an *estimate*

Q. What estimator design criterion should be used?

Define $c(\hat{\theta};\theta)$ a cost or risk function of $\hat{\theta}$ for given θ

A. We will seek to minimize average cost $E[c]$

We must treat cases of random and non-random θ differently

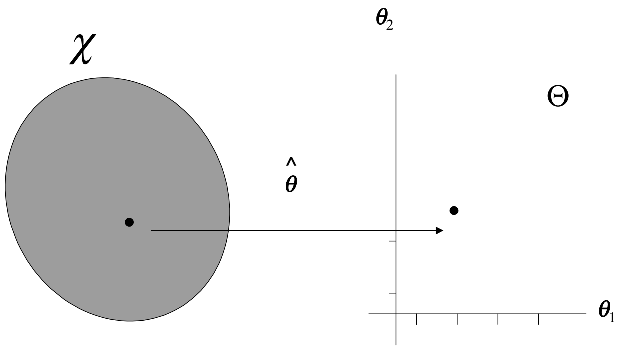


Figure 1: *Estimator maps $\mathcal{X}$ to $\mathbb{R}^p$*

4.2 RANDOM PARAMETERS

For this case we have following information:

$f(\theta)$: is a prior p.d.f. for θ

$f(x;\theta) = f(x|\theta)$: is a conditional p.d.f.

$f(\theta|x)$: is a posterior p.d.f. for θ

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)}$$

$f(x)$ is marginal p.d.f.

$$f(x) = \int_\Theta f(x|\theta)f(\theta)d\theta$$

Average cost (risk) is

$$E[c] = \int_\Theta \int_\mathcal{X} c(\hat{\theta}(x),\theta)f(x|\theta)f(\theta)\ dxd\theta$$

Definition 1 *An estimator $\hat{\theta}$ which minimizes the average cost is called a Bayes estimator*

Some reasonable cost functions for scalar θ

$c(\hat{\theta};\theta) = |\hat{\theta} - \theta|^2$: squared error

$c(\hat{\theta};\theta) = |\hat{\theta} - \theta|$: absolute error

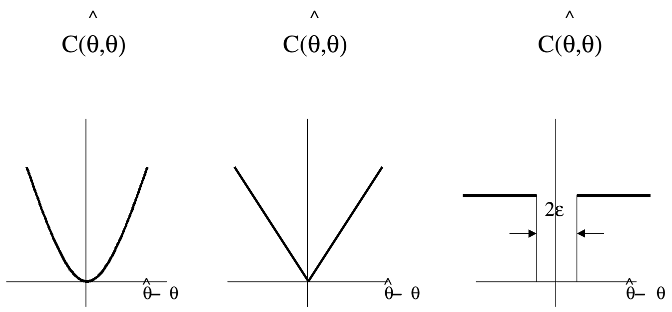


Figure 2: *(a) squared error, (b) absolute error, (c) uniform error*

$c(\hat{\theta};\theta) = I(|\hat{\theta}-\theta| > \epsilon)$: uniform error

Gives the average costs:

Estimator MSE:

$$\mathrm{MSE}(\hat{\theta}) = E[|\hat{\theta}-\theta|^2]$$

Estimator MAE:

$$\mathrm{MAE}(\hat{\theta}) = E[|\hat{\theta}-\theta|]$$

Error Probability:

$$P_e(\hat{\theta}) = P(|\hat{\theta}-\theta| > \epsilon)$$

4.2.1 SQUARED ERROR: $|\hat{\theta}-\theta|^2$

Optimal minimum mean squared error estimator (MMSEE) is conditional mean estimator (CME)

$$\hat{\theta}(x) = E[\theta|x] = \mathrm{mean}_{\theta\in\Theta}\{f(\theta|x)\}$$

where

$$\mathrm{mean}_{\theta\in\Theta}\{f(\theta|x)\} = \int_{-\infty}^{\infty} \theta f(\theta|x)d\theta$$

CME satisfies orthogonality condition (stated for general case of complex r.v.'s):

$$E[(\theta - \hat{\theta}(x))g(x)^*] = 0$$

for any linear or non-linear function g of x.

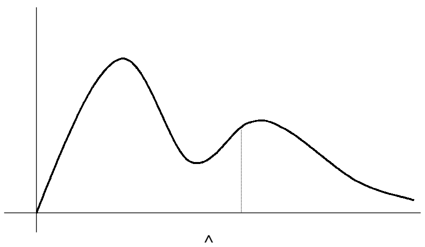


Figure 3: *Conditional mean estimator minimizes MSE*

Proof

Write MSE as

$$E[|\hat{\theta}-\theta|^2] = E[|(\hat{\theta}-E[\theta|x]) - (\theta - E[\theta|x])|^2]$$

$$= E[|\hat{\theta}-E[\theta|x]|^2] + E[|\theta - E[\theta|x]|^2]$$

$$-E[g(x)^*(\theta - E[\theta|x])] - E[g(x)(\theta - E[\theta|x])^*]$$

where $g(x) = \hat{\theta} - E[\theta|x]$ is a function of x only.

Step 1: show orthogonality condition

$$\begin{aligned} E[g(x)(\theta - E[\theta|x])] &= E\left[\ E[g(x)(\theta - E[\theta|x])^H \big|\ x]\ \right] \\ &= E\left[g(x)\ \underbrace{E\ [\theta - E[\theta|x]\ |\ x]}_{=0}\right] = 0 \end{aligned}$$

Step 2: Next show $E[\theta|x]$ minimizes MSE

$$\begin{aligned} E[|\hat{\theta}-\theta|^2] &= E[|\hat{\theta}-E[\theta|x]|^2] + E[|\theta - E[\theta|x]|^2] \\ &\geq E[|\theta - E[\theta|x]|^2] \end{aligned}$$

where "=" occurs iff $\hat{\theta} = E[\theta|x]$ $\diamond$

4.2.2 ABSOLUTE ERROR: $|\hat{\theta} - \theta|$

Optimal minimal mean absolute error estimator (MMAEE) is conditional median estimator (CmE) (real valued θ only)

$$\hat{\theta}(x) = \mathrm{median}_{\theta\in\Theta}\{f(\theta|x)\}$$

where

$$\begin{aligned} \mathrm{median}_{\theta\in\Theta}\{f(\theta|x)\} &= \{u : \int_{-\infty}^{u} f(\theta|x)d\theta = 1/2\} &(5)\\ &= \left\{u : \int_{-\infty}^{u} f(x|\theta)f(\theta)d\theta = \int_{u}^{\infty} f(x|\theta)f(\theta)d\theta\right\} &(6)\end{aligned}$$

For $f(\theta|x)$ continuous in θ CmE satisfies orthogonality condition:

$$E[\mathrm{sgn}(\theta - \hat{\theta}(x))g(x)] = 0$$

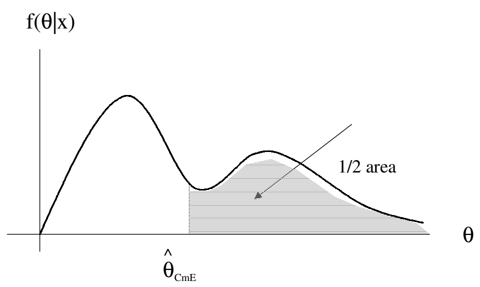


Figure 4: *Conditional median estimator minimizes MAE*

Proof

Let $\hat{\theta}_m$ = median of $f(\theta|x)$.

Then by definition of median for cts densities

$$E[\mathrm{sgn}(\theta - \hat{\theta}_m) \mid x] = \int_{\Theta} \mathrm{sgn}(\theta - \hat{\theta}_m(x))\ f(\theta|x)d\theta$$

$$
\begin{aligned}
&= \int_{\theta > \hat{\theta}_m(x)} f(\theta|x)d\theta - \int_{\theta \leq \hat{\theta}_m(x)} f(\theta|x)d\theta \\
&= 0
\end{aligned}
$$

Step 1: show orthogonality condition:

$$
E[\mathrm{sgn}(\theta - \hat{\theta}_m)g(x)] = E[\underbrace{E[\mathrm{sgn}(\theta - \hat{\theta}_m)|x]}_{=0}\ g(x)]
$$

Step 2: for $\hat{\theta}$ arbitrary we have (apply "useful formula" below)

$$
\begin{aligned}
\mathrm{MAE}(\hat{\theta}) &= E[|\underbrace{\theta - \hat{\theta}_m}_{a} + \underbrace{\hat{\theta}_m - \hat{\theta}}_{\Delta}|] \\
&= E[|\theta - \hat{\theta}_m|] + \underbrace{E[\mathrm{sgn}(\theta - \hat{\theta})\Delta]}_{=0} \\
&\quad + E\underbrace{[\mathrm{sgn}(a+\Delta) - \mathrm{sgn}(a)](a+\Delta)}_{\geq[\mathrm{sgn}(a+\Delta)-1](a+\Delta)\geq 0} \\
&\geq E[|\theta - \hat{\theta}_m|]
\end{aligned}
$$

Useful formula: $|a + \Delta| = |a| + \mathrm{sgn}(a)\Delta + [\mathrm{sgn}(a + \Delta) - \mathrm{sgn}(a)](a + \Delta)$

4.2.3 UNIFORM ERROR: $I(|\hat{\theta} - \theta| > \epsilon)$

For small ϵ the optimal estimator is the **maximum a posteriori** (MAP) estimator, which is also called the **posterior mode** estimator

$$
\begin{aligned}
\hat{\theta}(x) &= \mathrm{argmax}_{\theta \in \Theta}\{f(\theta|x)\} \\
&= \mathrm{argmax}_{\theta \in \Theta}\left\{\frac{f(x|\theta)f(\theta)}{f(x)}\right\} \\
&= \mathrm{argmax}_{\theta \in \Theta}\{f(x|\theta)f(\theta)\}
\end{aligned}
$$

Proof of optimality of MAP estimator:

$$
\begin{aligned}
P_e(\hat{\theta}) &= 1 - P(|\theta - \hat{\theta}| \leq \epsilon) \\
&= 1 - \int_{\mathcal{X}} dx f(x) \int_{|\theta - \hat{\theta}| \leq \epsilon} f(\theta|x)d\theta
\end{aligned}
$$

Important remarks:

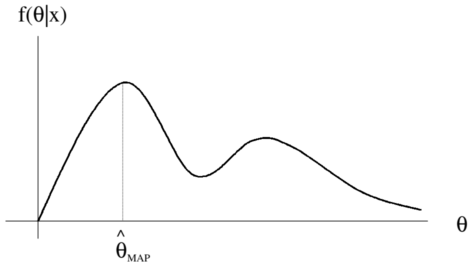


Figure 5: *Maximum a posteriori estimator minimizes* P_e

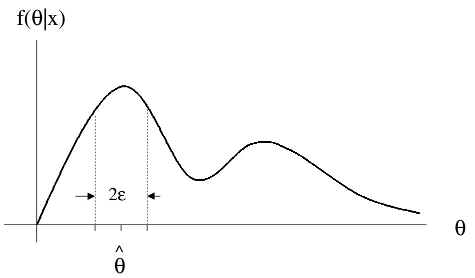


Figure 6: *Posterior density integrated over window of length* ϵ

- Only the CME requires difficult computation of $f(x)$
- Each of these estimators depends on x only through posterior $f(\theta|x)$
- When posterior is continuous, unimodal, and symmetric then each of the above estimators are identical (VanTrees [36])!
- If $T = T(x)$ is a sufficient statistic, posterior depends on x only through T.
  Indeed, if $f(x|\theta) = g(T;\theta)h(x)$, then by Bayes rule

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int_\Theta f(x|\theta)f(\theta)d\theta} = \frac{g(T;\theta)f(\theta)}{\int_\Theta g(T;\theta)f(\theta)d\theta}$$

  which is a function of T only. Thus, in terms MAP-optimal estimation performance, one loses nothing by compressing $\underline{x}$ to a sufficient statistic.

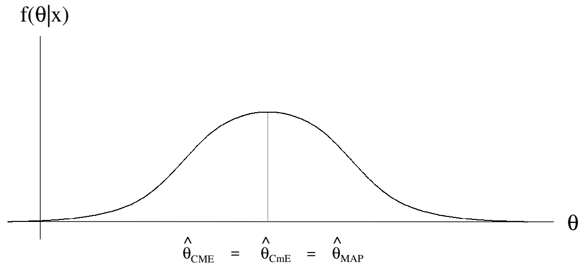


Figure 7: *Symmetric and continuous posterior density*

4.2.4 EXAMPLES

* Estimation of width of uniform density

* Estimation of a Gaussian signal

* Estimation of magnitude of Gaussian signal

* Estimation of a binary signal in Gaussian noise

Example 4 *ESTIMATION OF WIDTH OF UNIFORM PDF*

$x_1, \ldots, x_n$ conditionally i.i.d. uniform sample, θ a scalar

$f(x_1|\theta) = \frac{1}{\theta}I_{[0,\theta]}(x_1)$

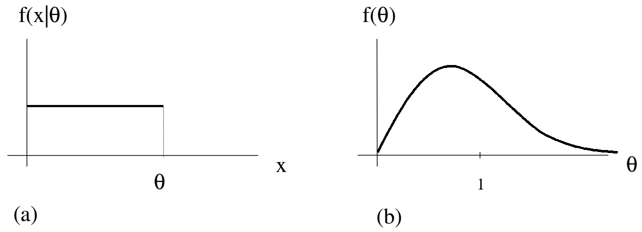


Figure 8: *(a) Uniform density of unknown width θ, (b) prior on θ*

with prior

$f(\theta) = \theta\, e^{-\theta}, \;\; \theta > 0$

Objective: compute CME, CmE, MAP estimators

Step 1: Find posterior $f(\theta|\underline{x}) = f(\underline{x}|\theta)f(\theta)/f(\underline{x})$

$$\begin{aligned} f(\underline{x}|\theta)f(\theta) &= \prod_{i=1}^{n} \frac{1}{\theta} I_{[x_i,\infty)}(\theta)\; \theta e^{-\theta} \\ &= \frac{e^{-\theta}}{\theta^{n-1}} \underbrace{\prod_{i=1}^{n} I_{[x_i,\infty)}(\theta)}_{I_{[x_{(1)},\infty)}(\theta)} \\ &= \frac{e^{-\theta}}{\theta^{n-1}} I_{[x_{(1)},\infty)}(\theta) \end{aligned}$$

where $x_{(1)} = \max\{x_i\}$.

Observe: function $\frac{e^{-\theta}}{\theta^{n-1}}$ is monotone decreasing over $\theta > 0$ (derivative of its $\ln$ is negative).

Furthermore,

$$\begin{aligned} f(\underline{x}) &= \int_0^\infty f(\underline{x}|\theta)f(\theta)d\theta \\ &= q_{-n+1}(x_{(1)}) \end{aligned}$$

where q_n is the monotone decreasing function

$$q_n(x) \stackrel{\text{def}}{=} \int_x^\infty \theta^n e^{-\theta} d\theta$$

Recursive formula: $q_{-n-1}(x) = \frac{1}{n}\left(\frac{1}{x^n}e^{-x} - q_{-n}(x)\right)$, $n = 0, -1, -2, \ldots$.

Step 2: find estimators:

$$\begin{aligned}
\hat{\theta}_{\mathrm{MAP}} &= x_{(1)} \\
\hat{\theta}_{\mathrm{CME}} &= q_{-n+2}(x_{(1)})/q_{-n+1}(x_{(1)}) \\
\hat{\theta}_{\mathrm{CmE}} &= q_{-n+1}^{-1}\left(\tfrac{1}{2}q_{-n+1}(x_{(1)})\right)
\end{aligned}$$

Note that only MAP estimator is explicit while the two others are implicit reuiring computation of intergral q_n and/or inverse function q_n^{-1}.

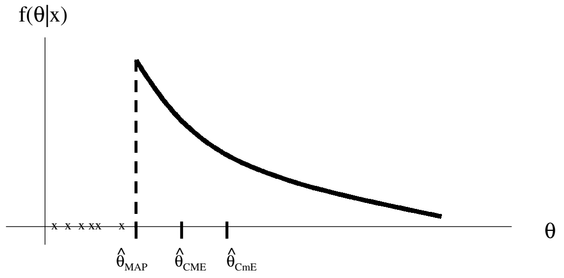


Figure 9: *Three estimators for width θ of uniform density.*

Example 5 *ESTIMATION OF GAUSSIAN AMPLITUDE*

Two jointly Gaussian r.v.s: s, x

$$\begin{aligned}
&E[s] = \mu_s,\ E[x] = \mu_x, \\
&\mathrm{var}(s) = \sigma_s^2,\ \mathrm{var}(x) = \sigma_x^2 \\
&\mathrm{cov}(s,x) = \rho\ \sigma_s\sigma_x
\end{aligned}$$

Objective: estimate s given measured x

Solution:

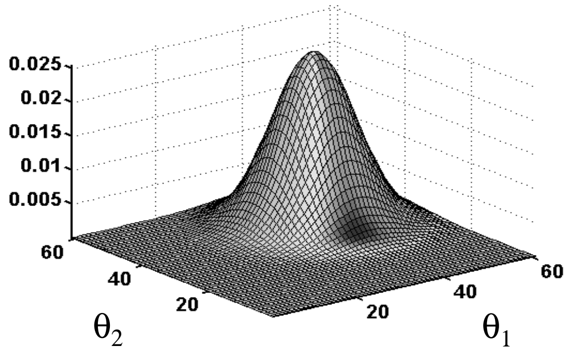


Figure 10: *s and x follow jointly Gaussian density*

Step 1: find posterior density

$$
\begin{aligned}
f(s|x) &= \frac{f(x|s)f(s)}{f(x)} \\
&= \frac{1}{\sqrt{2\pi\sigma^2_{s|x}}} \exp\left\{-\frac{\left(s-\mu_{s|x}\right)^2}{2\sigma^2_{s|x}}\right\}
\end{aligned}
$$

where

$$
\begin{aligned}
\mu_{s|x} &= \mu_s + \rho\frac{\sigma_s}{\sigma_x}(x-\mu_x) \\
\sigma^2_{s|x} &= (1-\rho^2)\sigma^2_s
\end{aligned}
$$

Step 2: find estimators

As posterior is continuous, symmetric and unimodal MAP, CME, and CmE estimators are identical:

$$
\hat{s} = \mu_{s|x} = \text{ linear in } x
$$

Special case: indep. additive noise model $x = s + v$

Then

$\sigma^2_x = \sigma^2_s + \sigma^2_v$

$\rho^2 = \sigma^2_s/(\sigma^2_s + \sigma^2_v)$

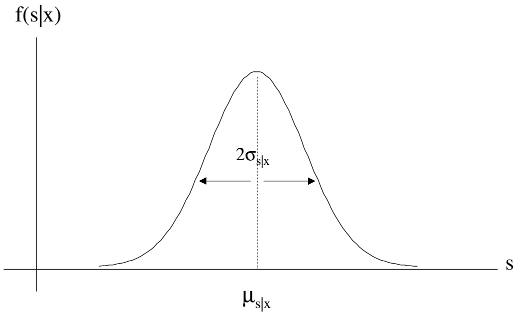


Figure 11: *Posterior $f(s|x)$ is Gaussian*

and

$$\hat{s} = \mu_s + \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2}(x - \mu_x)$$

PERFORMANCE:

Note that $E[\hat{s}] = E[E[s|x]] = \mu_s$ and

Minimum MSE: $E[(s - \hat{s})^2] = (1 - \rho^2)\sigma_s^2$

Minimum MAE: $E[|s - \hat{s}|] = \sqrt{(1 - \rho^2)\sigma_s^2}\sqrt{\frac{2}{\pi}}$

Minimum Pe: $P(|s - \hat{s}| > \epsilon) = 1 - 2\text{erf}\left(\epsilon/\sqrt{(1 - \rho^2)\sigma_s^2}\right)$

Example 6 *ESTIMATION OF MAGNITUDE OF GAUSSIAN SIGNAL*

Two jointly Gaussian r.v.s: s, x

$$E[s] = \mu_s,\ E[x] = \mu_x,$$
$$\text{var}(s) = \sigma_s^2,\ \text{var}(x) = \sigma_x^2$$
$$\text{cov}(s, x) = \rho\ \sigma_s\sigma_x$$

Objective: estimate $y = |s|$ based on x.

Note: no longer a jointly Gaussian model.

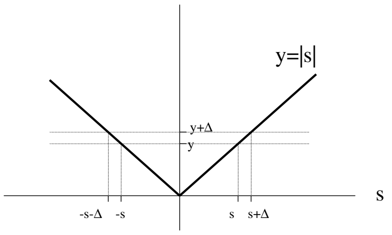


Figure 12: *Method of differentials for finding joint density of* $y = |s|$, x

Solution:

Using simple transformation of variables find posterior density $f_{y|x}(y|x) = f(y|x)$

$$f_{y|x}(y|x)\Delta = f_{s|x}(y|x)\Delta + f_{s|x}(-y|x)\Delta, \quad y \geq 0$$

or more explicitly

$$\begin{aligned} f(y|x) &= \\ & \frac{1}{\sqrt{2\pi\sigma_{s|x}^2}} \left(\exp\left\{ -\frac{\left(y-\mu_{s|x}\right)^2}{2\sigma_{s|x}^2} \right\} + \exp\left\{ -\frac{\left(y+\mu_{s|x}\right)^2}{2\sigma_{s|x}^2} \right\} \right) \; I_{[0,\infty)}(y) \end{aligned}$$

As density is no longer symmetric, CME, CmE, and MAP will differ

CME is in explicit form

$$\hat{y}_{\text{CME}} = E[y|x] \quad = \quad \left|\mu_{s|x}\right| \; \text{erf}\left(\frac{|\mu_{s|x}|}{\sigma_{s|x}\sqrt{2}} \right) + \sqrt{\frac{2}{\pi}} \sigma_{s|x} \; e^{-\mu_s^2/2\sigma_{s|x}^2}$$

CmE is solution $\hat{y} = \hat{y}_{\text{CmE}}$ of equation

$$\text{erf}\left(\frac{\hat{y} - \mu_{s|x}}{\sigma_{s|x}\sqrt{2}} \right) + \text{erf}\left(\frac{\hat{y} + \mu_{s|x}}{\sigma_{s|x}\sqrt{2}} \right) = \frac{1}{2}$$

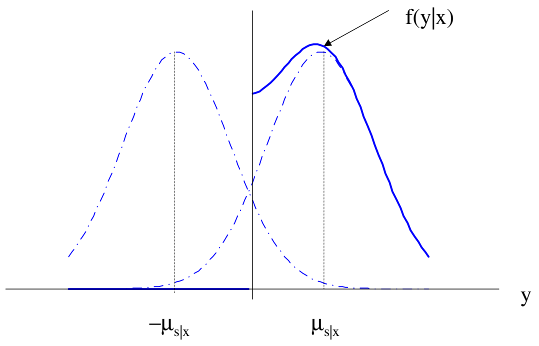


Figure 13: *Posterior density of* $y = |s|$ *given* x

As $f(y|x)$ is convex and smooth, MAP estimator $\hat{y} = \hat{y}_{\text{MAP}}$ occurs at stationary point given by "MAP equation"

$$0 = \frac{\partial f(y|x)}{\partial y}$$

which is equivalent to

$$\frac{\hat{y}}{\mu_{s|y}} \frac{\exp\left\{-\frac{(\hat{y}-\mu_{s|x})^2}{2\sigma_{s|x}}\right\} - \exp\left\{-\frac{(\hat{y}+\mu_{s|x})^2}{2\sigma_{s|x}}\right\}}{\exp\left\{-\frac{(\hat{y}-\mu_{s|x})^2}{2\sigma_{s|x}}\right\} + \exp\left\{-\frac{(\hat{y}+\mu_{s|x})^2}{2\sigma_{s|x}}\right\}} = 1$$

Note: as $\mu_{s|x}/\sigma_{s|x} \to \infty$ all three estimators converge to identical limit

$$\hat{y} \to \left|\mu_{s|y}\right|$$

Example 7 *Linear observation model (*$\theta = y = \text{sgn}(s)$ *in previous example)*

$$x = \theta + w$$

w zero mean Gaussian noise with variance σ^2

θ binary random variable: $P(\theta = 1) = P(\theta = -1) = \frac{1}{2}$

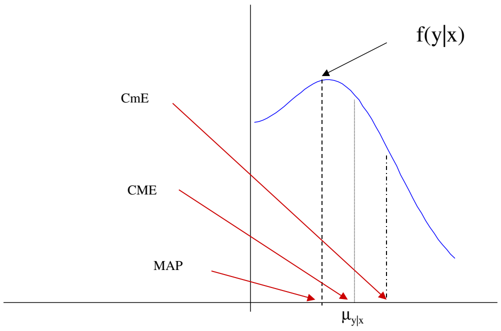


Figure 14: *Three estimators of* $y = |s|$

Posterior density:

$$f(\theta|x) = p(\theta|x) = \frac{f(x|\theta)\frac{1}{2}}{f(x)}$$

Now

$$f(x|\theta) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x-1)^2}{2\sigma^2}\right), & \theta = 1 \\ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x+1)^2}{2\sigma^2}\right), & \theta = -1 \end{cases}$$

and

$$f(x) = f(x|\theta = 1)\tfrac{1}{2} + f(x|\theta = -1)\tfrac{1}{2}$$

MAP estimator is minimum distance decision rule:

$$\begin{aligned} \hat{\theta}_{\mathrm{MAP}} &= \text{argmax}_{\theta=1,-1} p(x|\theta) \\ &= \text{argmin}_{\theta=1,-1}\{(x-\theta)^2\} \\ &= \begin{cases} 1, & x > 0 \\ -1, & x < 0 \end{cases} \end{aligned}$$

CME estimator is continuous over $\hat{\theta} \in (-1, 1)$

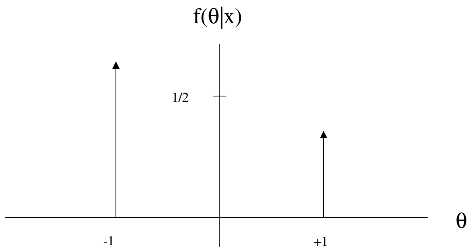


Figure 15: *The posterior density $f(\theta|x)$ is a pair of delta functions*

$$\begin{aligned}\hat{\theta}_{CME} &= (1)P(\theta=1|x) + (-1)P(\theta=-1|x)\\ &= \frac{\exp\left(-\frac{(x-1)^2}{2\sigma^2}\right) - \exp\left(-\frac{(x+1)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(x-1)^2}{2\sigma^2}\right) + \exp\left(-\frac{(x+1)^2}{2\sigma^2}\right)}\end{aligned}$$

REMARKS:

1. Different error criteria usually give different estimators

2. Estimators of random parameters are not invariant to functional transformations

$$\widehat{g(\theta)} \neq g(\hat{\theta})$$

3. CmE and MAP estimators always take values in Θ. CME may not if Θ is discrete.

4. "MAP equation" stationary point condition $\partial f(\theta|x)/\partial\theta = 0$ at $\theta = \hat{\theta}_{MAP}$ is only useful for smooth and convex densities.

4.3 VECTOR VALUED PARAMETERS

Define norm squared of $\underline{\theta} \in \Theta \subset \mathbb{R}^p$, $\underline{\theta} = [\theta_1, \ldots, \theta_p]^T$

$$\|\underline{\theta}\|^2 = \sum_{i=1}^{p} |\theta_i|^2$$

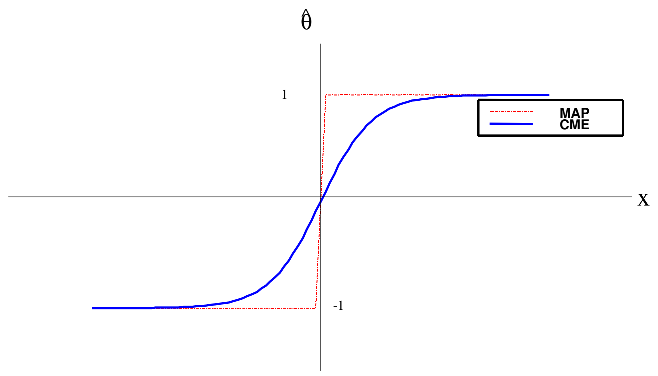


Figure 16: *MAP (light font sign function) and CME (heavy font "S" curve) as function of measurement* x

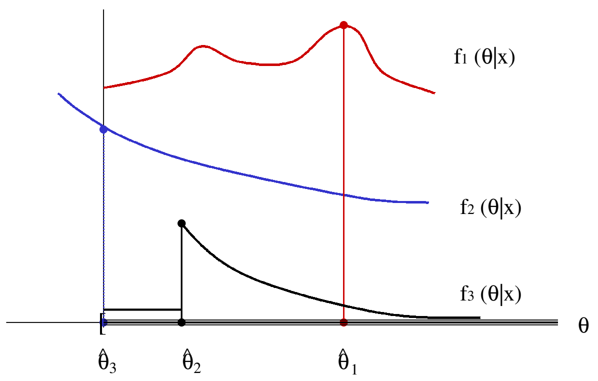


Figure 17: *In general there may exist no stationary points or there may be multiple stationary points which may or may not correspond to the MAP estimator.*

Generalized estimation criteria for vectors

Estimator MSE:

$$\mathrm{MSE}(\hat{\theta}) = E[\|\hat{\theta} - \theta\|^2]$$

Estimator MAE:

$$\mathrm{MAE}(\hat{\theta}) = E[\|\hat{\theta} - \theta\|]$$

Error Probability:

$$P_e(\hat{\theta}) = P(\|\hat{\theta} - \theta\| > \epsilon)$$

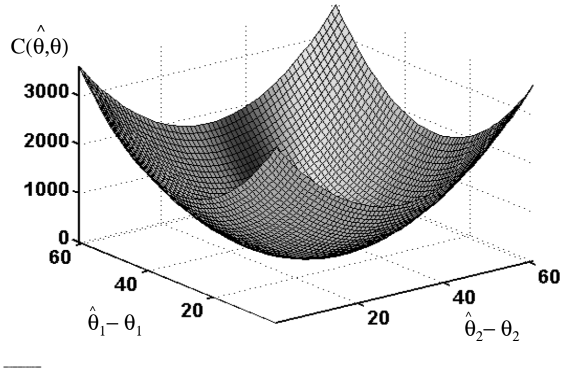


Figure 18: *Squared error criterion*

Only MSE and P_e give useful optimal vector estimators:

Note: Minimum MAE vector estimator is not uniquely defined and is seldom used

4.3.1 VECTOR SQUARED ERROR: $\|\hat{\theta} - \theta\|^2$

As $\mathrm{MSE}(\hat{\theta}) = \sum_{i=1}^p \mathrm{MSE}(\hat{\theta}_i)$ the minimum MSE vector estimator attains minimum of each component MSE. Hence vector MMSEE is the vector of scalar CME's:

$$\underline{\hat{\theta}}_{\mathrm{CME}} = E[\underline{\theta}|x] = \begin{bmatrix} E[\theta_1|x] \\ \vdots \\ E[\theta_p|x] \end{bmatrix}$$

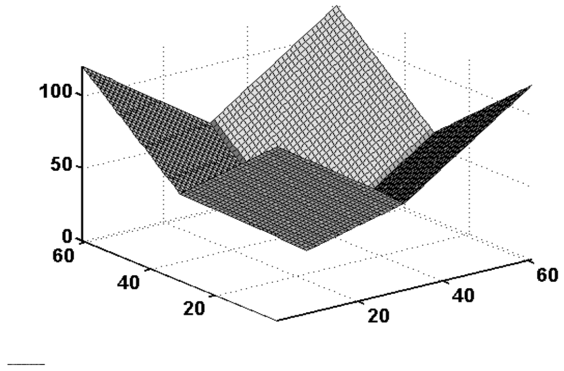


Figure 19: *Absolute error criterion*

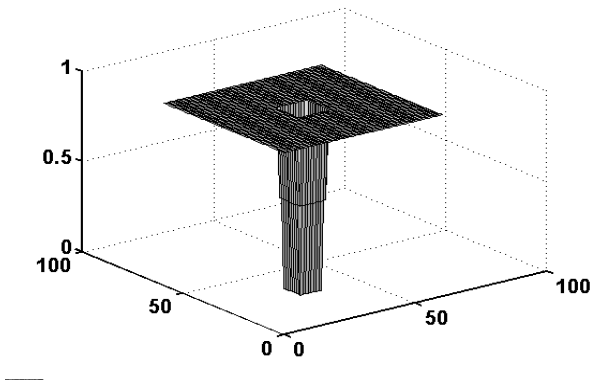


Figure 20: *Uniform error criterion*

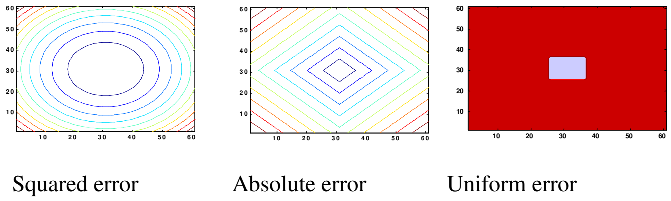


Figure 21: *Constant contours of three error criteria*

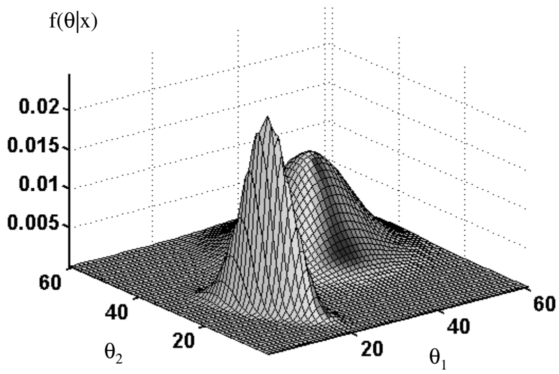


Figure 22: *Vector conditional mean estimator*

4.3.2 VECTOR UNIFORM ERROR: $I(\|\hat{\theta}-\theta\|^2 > \epsilon)$

For small ϵ the minimum P_e is attained by the vector MAP estimator which has form similar to the MAP for scalar θ studied before

$$\underline{\hat{\theta}}_{\text{MAP}} = \text{argmax}_{\underline{\theta}\in\Theta} f(\underline{\theta}|x)$$

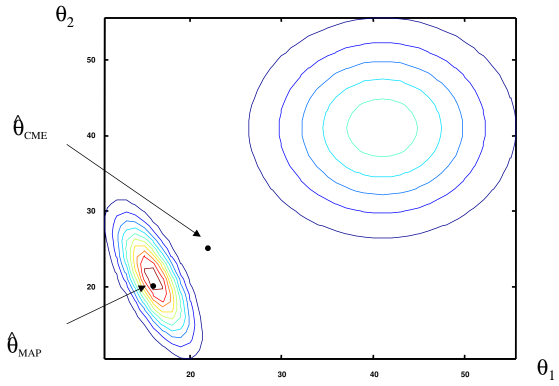


Figure 23: *Vector MAP estimator*

4.4 ESTIMATION OF NON-RANDOM PARAMETERS

Now no prior $f(\theta)$ available

Can only interpret $f(x|\theta)$ as $f(x;\theta)$

Possible cost criteria for scalar real θ?

Q. How about MSE?

$$\text{MSE}_\theta = E_\theta[(\hat{\theta}-\theta)^2]$$

A. Difficulty: trivial estimator $\hat{\theta} = \theta_o$ attains 0 MSE when $\theta = \theta_0$.

4.4.1 SCALAR ESTIMATION CRITERIA FOR NON-RANDOM PARAMETERS

1. Minimize worst case MSE:

$$\max_\theta \text{MSE}_\theta(\hat{\theta}) = \max_\theta E_\theta[(\hat{\theta}-\theta)^2]$$

2. Minimize maximum estimator error probability:

$$\max P_e = \max_\theta P_\theta(|\hat{\theta}-\theta| > \epsilon)$$

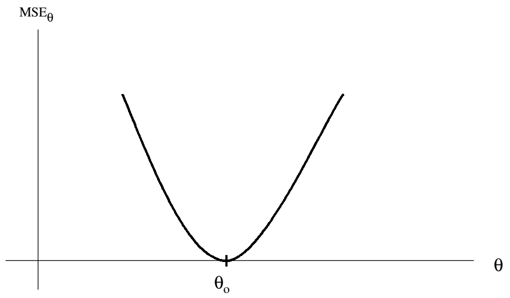


Figure 24: *MSE curve as function of* θ *for estimator* $\hat{\theta} = \theta_o$

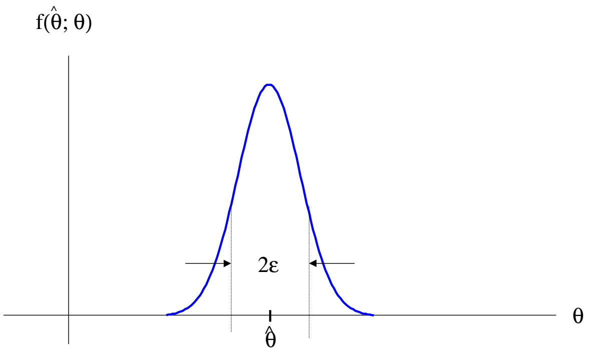


Figure 25: *Density* $f(\hat{\theta};\theta)$ *of* $\hat{\theta}$ *measures concentration of* $\hat{\theta}$ *about true parameter* θ

Note: Min-max MSE and min-max P_e estimators are related through Tchebychev

$$P_\theta(|\hat{\theta} - \theta| \geq \epsilon) \leq \frac{E_\theta[|\hat{\theta} - \theta|^2]}{\epsilon^2}$$

Definition: $\hat{\theta} = \hat{\theta}(\underline{x})$ is said to be consistent if for all θ

$$\lim_{n\to\infty} P_\theta(|\hat{\theta} - \theta| > \epsilon) = 0$$

i.e. p.d.f. of estimator concentrates about θ.

Remark:

by Tchebychev, if MSE goes to zero as $n \to \infty$ then $\hat{\theta}$ is consistent.

A third estimation criterion places unbiasedness restriction on the class of estimators

3. Constrain estimator to be unbiased, $b_\theta(\hat{\theta}) = 0$ for all θ, and try to minimize variance

Estimator Bias:

$$b_\theta(\hat{\theta}) = E_\theta[\hat{\theta}] - \theta$$

Estimator variance:

$$\text{var}_\theta(\hat{\theta}) = E_\theta[(\hat{\theta} - E_\theta[\hat{\theta}])^2]$$

Definition: $\hat{\theta}$ is said to be a uniform minimum variance unbiased (UMVU) if for all $\theta \in \Theta$ it has less variance than any other unbiased estimator $\hat{\hat{\theta}}$

$$\text{var}_\theta(\hat{\theta}) \leq \text{var}_\theta(\hat{\hat{\theta}}), \quad \theta \in \Theta$$

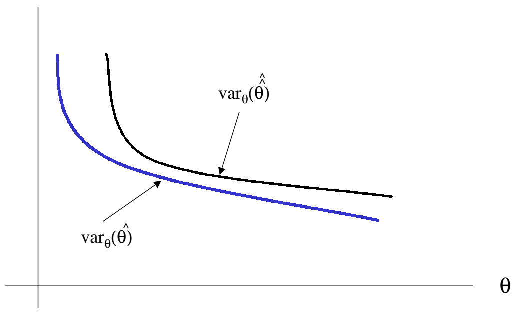


Figure 26: *A UMVU estimator $\hat{\theta}$ has lower variance than any other unbiased estimator $\hat{\hat{\theta}}$*

Relation between MSE, variance and bias:

$$
\begin{aligned}
\mathrm{MSE}_\theta(\hat{\theta}) &= E_\theta[(\hat{\theta}-\theta)^2] \\
&= E_\theta[\Big((\hat{\theta}-E_\theta[\hat{\theta}]) + (E_\theta[\hat{\theta}]-\theta)\Big)^2] \\
&= \underbrace{E_\theta[(\hat{\theta}-E_\theta[\hat{\theta}])^2]}_{\mathrm{var}_\theta(\hat{\theta})} + \underbrace{\Big(E_\theta[\hat{\theta}]-\theta\Big)^2}_{b_\theta(\hat{\theta})} \\
&\quad +2\underbrace{E_\theta[\hat{\theta}-E_\theta[\hat{\theta}]]}_{=0}\; b_\theta(\hat{\theta}) \\
&= \mathrm{var}_\theta(\hat{\theta}) + b_\theta^2(\hat{\theta})
\end{aligned}
$$

Note: in general, for specified MSE, there always exists a "bias-variance tradeoff"

Two important classes of estimators:

* method of moments

* maximum likelihood

4.4.2 METHOD OF MOMENTS (MOM) SCALAR ESTIMATORS

Let $m_k = m_k(\theta)$ be the k-th order moment of $f(x;\theta)$:

$$m_k = \int x^k f(x;\theta)dx.$$

Then find K moments such that the vector function

$$g(\theta) = [m_1(\theta), \ldots, m_K(\theta)]$$

can be inverted for θ

$$\theta = g^{-1}(m_1, \ldots, m_K)$$

The MOM estimator is then simply obtained by replacing m_k with its empirical estimate

$$\hat{\theta} = g^{-1}(\hat{m}_1, \ldots, \hat{m}_K)$$

$\hat{m}_k = \frac{1}{n}\sum_{i=1}^n x_i^k$

IMPORTANT PROPERTIES OF MOM (Serfling [30])

When the moments m_k are smooth functions of the parameter θ

* MOM estimators are asymptotically unbiased as $n \to \infty$

* MOM estimators are consistent

Difficulties with MOM:

1. MOM does not always give unique estimator

2. Inapplicable to cases where moments do not exist (e.g. Cauchy p.d.f.) or are unstable

Example 8 $\underline{x}$ *i.i.d. Bernoulli random variables*

$$x_i \sim f(x;\theta) = \theta^x(1-\theta)^{1-x}, \quad x = 0, 1$$

* $\theta = P(x_i = 1), 1-\theta = P(x_i = 0)$

* $E[x_i] = P(x_i = 1) = \theta$

Objective: find a MOM estimator of θ

Note that

$$m_k = E[x^k] = (1)\theta + (0)(1-\theta) = \theta$$

A MOM estimator of θ is sample mean

$$\hat{\theta} = \overline{x_i}$$

$\hat{\theta}$ is unbiased and has variance having maximum at $\theta = \frac{1}{2}$

$$E_\theta(\hat{\theta}_1) = m_1 = \theta, \qquad \text{var}_\theta(\hat{\theta}_1) = (m_2 - m_1^2)/n = \theta(1-\theta)/n$$

Example 9 $\underline{x}$ *i.i.d. Poisson random variables*

$$x_i \sim f(x;\theta) = \frac{\theta^x}{x!}e^{-\theta}, \quad x = 0, 1, 2, \ldots$$

Objective: find a MOM estimator of rate parameter θ

Note that $m_1 = \theta$ so a MOM estimator of θ is sample mean

$$\hat{\theta}_1 = \overline{x_i}$$

Alternatively, as $m_2 = \theta + \theta^2$, another MOM estimator is the positive $\hat{\theta}_2$ which satisfies the equation : $\hat{\theta} + \hat{\theta}^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 := \overline{x_i^2}$, i.e.

$$\hat{\theta}_2 = \frac{\sqrt{1 + 4\overline{x_i^2}} + -1}{2}$$

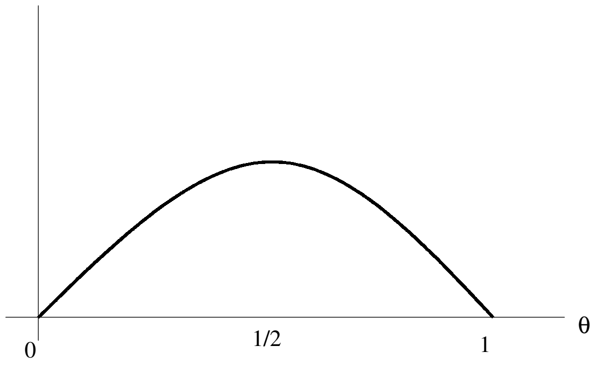


Figure 27: *Variance of MOM estimator of probability of success of Bernoulli r.v.*

As yet another example, we can express m_2 as $m_2 = \theta + m_1^2$ or $\theta = m_2 - m_1^2$ so another MOM estimator is

$$\hat{\theta}_3 = \overline{x^2} - \overline{x_i}^2 = n^{-1} \sum_{i=1}^{n} (x_i - \overline{x_i})^2$$

Only sample mean is unbiased:

$$E_\theta(\hat{\theta}_1) = \theta, \qquad \mathrm{var}_\theta(\hat{\theta}_1) = \theta/n$$

$$E_\theta(\hat{\theta}_3) = \frac{n-1}{n}\theta, \qquad \mathrm{var}_\theta(\hat{\theta}_3) = (2\theta^2 + \theta)/n$$

(Closed form expressions for bias and variance of $\hat{\theta}_2$ do not exist).

Observations:

1. $\hat{\theta}_1$ is unbiased for all n
2. $\hat{\theta}_2, \theta_3$ are asymptotically unbiased as $n \to \infty$
3. $\hat{\theta}_1$ and $\hat{\theta}_3$ are consistent.

4.4.3 MAXIMUM LIKELIHOOD (ML) SCALAR ESTIMATORS

For a fixed measurement x we define the "likelihood function" for θ

$$L(\theta) = f(x; \theta)$$

and the log-likelihood function

$$l(\theta) = \ln f(x;\theta)$$

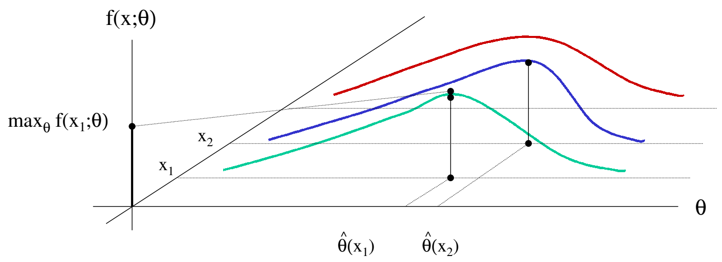


Figure 28: *The likelihood function for θ*

The ML estimator is defined as

$$\begin{aligned}\hat{\theta} &= \text{argmax}_{\theta\in\Theta} f(x;\theta) \\ &= \text{argmax}_{\theta\in\Theta} L(\theta) \\ &= \text{argmax}_{\theta\in\Theta} l(\theta)\end{aligned}$$

Important properties of ML estimators for smooth likelihoods (Ibragimov and Has'minskii [12], Serfling [30])

Property 1. MLE's are asymptotically unbiased

Property 2. MLE's are consistent

Property 3. MLE's are invariant to arbitrary transformations of the parameters

$$\varphi = g(\theta) \quad \Rightarrow \quad \hat{\varphi} = g(\hat{\theta})$$

Property 4: MLE's are asymptotically UMVU:

$$\lim_{n\to\infty} n\text{var}_\theta(\hat{\theta}) = \frac{1}{F_1(\theta)}$$

so that the variance of $\hat{\theta}$ decays with fastest possible asymptotic rate.

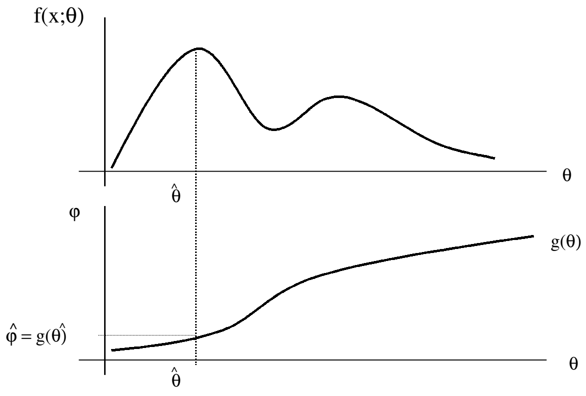


Figure 29: *Invariance of MLE to functional transformation g*

Property 5: MLE's are asymptotically Gaussian in the sense

$$\sqrt{n}(\hat{\theta}_n - \theta) \to z, \quad (i.d.)$$

where $z \sim \mathcal{N}(0, 1/F_1(\theta))$

Property 6: MLE is equivalent to MAP for uniform prior on θ

Example 10 $\underline{x}$ *i.i.d. Bernoulli random variables*

Objective: find MLE of $\theta = P(x_i = 1)$

Likelihood function is:

$$f(\underline{x};\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

Rewrite this in form

$$\begin{aligned} f(\underline{x};\theta) &= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \\ &= \theta^{n\overline{x_i}}(1-\theta)^{n-n\overline{x_i}} \end{aligned}$$

As smooth and convex in θ,

$\Rightarrow$ differentiation wrt θ yields stationary point condition for $\theta = \hat{\theta}$

$$0 = \frac{d}{d\theta} f(\underline{x};\theta) = n \left[\frac{(1-\theta)\overline{x_i} - \theta(1-\overline{x_i})}{\theta(1-\theta)} \right] f(\underline{x};\theta)$$

or MLE is identical to MOM estimator

$$\hat{\theta} = \overline{x_i}$$

Example 11 $\underline{x}$ *i.i.d. Poisson random variables*

Objective: find MLE of rate parameter θ

$$f(\underline{x};\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta}$$

Solution: First note that as $\ln(\cdot)$ is a monotone increasing function

$$\hat{\theta}_{ml} = \text{argmax}_{\theta>0} \ln f(\underline{x};\theta)$$

Thus we can equivalently maximize the loglikelihood

$$\begin{aligned} l(\theta) &= \ln f(\underline{x};\theta) \\ &= \ln \prod_{k=1}^{n} \frac{\theta^{x_k}}{x_k!} e^{-\theta} \\ &= \sum_{k=1}^{n} x_k \ln\theta - n\theta - \underbrace{\sum_{k=1}^{n} \ln x_k!}_{\text{constant in } \theta} \\ &= \overline{x_i} n \;\ln\theta - n\theta + c \end{aligned}$$

where c is an irrelevant constant

As loglikelihood is smooth convex function of θ, MLE is the unique solution $\theta = \hat{\theta}$ of equation

$$0 = \partial \ln f / \partial\theta = \frac{n\overline{x_i}}{\theta} - n$$

or we obtain ML identical to MOM

$$\hat{\theta} = \overline{x_i}$$

Check asymptotic Gaussian property

$$\begin{aligned}\sqrt{n}(\overline{x_i}-\theta) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n (x_i-\theta)\right)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (x_i-\theta)\end{aligned}$$

By CLT, this converges in distribution to a Gaussian r.v. as predicted with

$$\begin{aligned}E_\theta[\sqrt{n}(\overline{x_i}-\theta)] &= 0\\ \mathrm{var}_\theta(\sqrt{n}(\overline{x_i}-\theta)) &= \theta\end{aligned}$$

4.4.4 SCALAR CRAMÈR-RAO BOUND (CRB) ON ESTIMATOR VARIANCE

Let $\theta \in \Theta$ be a *non-random* scalar and assume:

1. Θ is an open subset, e.g. (a, b), of $\mathbb{R}$

2. $f(x;\theta)$ is smooth (Ibragimov and Has'minskii [12]) and differentiable in θ

CRAMER-RAO BOUND for scalar θ

For any *unbiased* estimator $\hat{\theta}$ of θ

$$\mathrm{var}_\theta(\hat{\theta}) \geq 1/F(\theta), \qquad (CRB)$$

where "=" is attained iff for some non-random scalar k_θ

$$\frac{\partial}{\partial\theta}\ln f(\underline{x};\theta) = k_\theta(\hat{\theta}-\theta)$$

$\Rightarrow$ When CRB is attainable it is said to be a "tight bound"

In CRB $F(\theta)$ is the Fisher information

$$\begin{aligned}F(\theta) &= E\left[\left(\frac{\partial}{\partial\theta}\ln f(\underline{x};\theta)\right)^2\right]\\ &= -E\left[\frac{\partial^2}{\partial\theta^2}\ln f(\underline{x};\theta)\right]\end{aligned}$$

(Derivation of scalar CRB will be done in context of general vector case in next subsection).

Example 12 *CRB for the Poisson rate*

$\underline{x}$ i.i.d. Poisson random variables

$$x_i \sim f(x;\theta) = \frac{\theta^x}{x!}e^{-\theta}, \quad x = 0, 1, 2, \ldots$$

Compute Fisher information

$$\ln f(\underline{x};\theta) \quad = \quad \sum_{k=1}^n x_k \ln\theta - n\theta - \underbrace{\sum_{k=1}^n \ln x_k!}_{\text{constant in } \theta}$$

hence

$$\begin{aligned} \partial \ln f/\partial\theta &= \frac{1}{\theta}\sum_{k=1}^n x_k - n \\ \partial^2 \ln f/\partial\theta^2 &= -\frac{1}{\theta^2}\sum_{k=1}^n x_k \end{aligned}$$

Therefore, as $E[\sum_{k=1}^n x_k] = n\theta$

$$F_n(\theta) = \frac{n}{\theta}$$

The CRB asserts that for any unbiased estimator

$$\mathrm{var}_\theta(\hat\theta) \geq \frac{\theta}{n}$$

Observation 1: the sample mean $\overline{x_i}$ is unbiased and has $\mathrm{var}_\theta(\overline{x_i}) = \theta/n$ that equals the CRB

Observation 2: we could have found that $\overline{x_i}$ attains the CRB directly from condition for CRB equality

$$\partial \ln f/\partial\theta \quad = \quad \frac{1}{\theta}\sum_{k=1}^n x_k - n = \underbrace{\frac{n}{\theta}}_{k_\theta}(\underbrace{\overline{x_i}}_{\hat\theta} - \theta)$$

GENERAL PROPERTIES OF CR BOUND

Property 1. Fisher information is a measure of the average curvature of the log likelihood function $\ln f(x;\theta)$ near the true θ (Hero [11])

Property 2. Let $F_n(\theta)$ be the Fisher information for a sample of n i.i.d. measurements $x_1, \ldots, x_n$. Then

$$F_n(\theta) = nF_1(\theta)$$

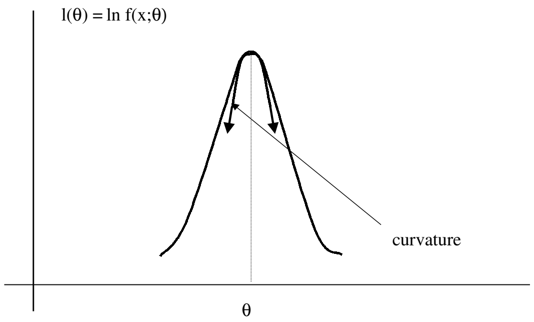


Figure 30: *The curvature of the log likelihood function* $\ln f(x;\theta)$ *in the vicinity of true* θ

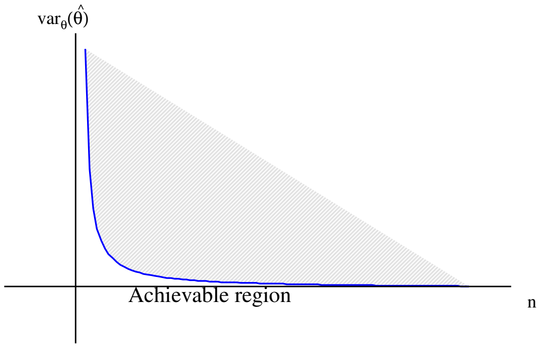


Figure 31: *CRB defines unachievable region which is under the CRB curve. Good unbiased estimators of continous parameters have variance that decays as* $1/n$.

Hence for continuous parameters $\mathrm{var}_\theta(\hat{\theta})$ cannot decay faster than order $1/n$

$\mathrm{var}_\theta(\hat{\theta}) = O(1/n)$ is expected for good estimators $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$.

Proof of $1/n$ variance decay rate:

Since $\underline{x} = [x_1, \ldots, x_n]^T$ are i.i.d.

$$f(\underline{x}; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

so that

$$\begin{aligned} F_n(\theta) &= -E\left[\frac{\partial^2}{\partial\theta^2} \ln f(\underline{x}; \theta)\right] \\ &= -E\left[\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2} \ln f(x_i; \theta)\right] \\ &= \sum_{i=1}^n \underbrace{-E\left[\frac{\partial^2}{\partial\theta^2} \ln f(x_i; \theta)\right]}_{F_1(\theta)} \end{aligned}$$

Property 3. If $\hat{\theta}$ is unbiased and $\mathrm{var}_\theta(\hat{\theta})$ attains the CRB $\hat{\theta}$ is said to be an *efficient estimator*

Efficient estimators are always UMVU (but not conversely)

Furthermore, if asymptotically unbiased and variance decays with optimal rate constant

$$\lim_{n\to\infty} b_\theta(\hat{\theta}) = 0, \qquad \lim_{n\to\infty} n\mathrm{var}_\theta(\hat{\theta}) = F_1(\theta)$$

then $\hat{\theta}$ is said to be *asymptotically efficient*.

Property 4. Efficient estimators can only exist when density comes from an exponential family

$$f(x; \theta) = a(\theta)b(x)e^{-c(\theta)t(x)}$$

Proof of efficiency for exponential family:

Without loss of generality we specialize to the case of a single sample $n = 1$. Recall the condition for equality in the CR bound to be achieved by an estimator $\hat{\theta}$ is that the p.d.f. be expressible as

$$\frac{\partial}{\partial\theta} \ln f(x; \theta) = k_\theta(\hat{\theta} - \theta). \tag{7}$$

For fixed θ_o, integrate the LHS of condition (7) over $\theta \in [\theta_o, \theta']$

$$\int_{\theta_o}^{\theta'} \frac{\partial}{\partial \theta} \ln f(x;\theta) = \ln f(x;\theta') - \ln f(x;\theta_o)$$

On the other hand, integrating the RHS of the condition

$$\begin{aligned}
\int_{\theta_0}^{\theta} k_\theta(\hat{\theta}-\theta)\, d\theta &= \int_{\theta_o}^{\theta'} k_\theta(\hat{\theta}-\theta) d\theta \\
&= \hat{\theta} \underbrace{\int_{\theta_o}^{\theta'} k_\theta d\theta}_{c(\theta')} - \underbrace{\int_{\theta_o}^{\theta'} k_\theta \theta d\theta}_{d(\theta')}
\end{aligned}$$

Or combining the integrals of RHS and LHS of (7)

$$f(x;\theta) = \underbrace{e^{-d(\theta)}}_{a(\theta)} \underbrace{f(x;\theta_o)}_{b(x)} \; e^{c(\theta)} \; \overbrace{\hat{\theta}}^{t(x)}$$

Therefore, $\hat{\theta}$ is an efficient estimator of θ only when

1. p.d.f. is of form $f(x;\theta) = a(\theta)b(x)e^{-c(\theta)t(x)}$
2. $\hat{\theta}$ can be identified with $t(x)$
3. $E[t(x)] = \theta$

Example 13 *Condition for CRB achievability for Poisson rate*

$\underline{x}$ i.i.d. Poisson random variables

$$x_i \sim f(x;\theta) = \frac{\theta^x}{x!} e^{-\theta}, \quad x = 0, 1, 2, \ldots$$

The p.d.f. of sample $\underline{x}$ can be put into the form

$$\begin{aligned}
f(\underline{x};\theta) &= \prod_{i=1}^n \frac{\theta^{x_i}}{x_i!} e^{-\theta} \\
&= \underbrace{e^{-n\theta}}_{a(\theta)} \underbrace{\prod_{i=1}^n \frac{1}{x_i!}}_{b(\underline{x})} \; e^{\overbrace{n \ln(\theta)}^{c(\theta)} \; \overbrace{n^{-1} \sum_{i=1} x_i}^{t(\underline{x})}}
\end{aligned}$$

Conclude:

1. $f(\underline{x};\theta)$ in the exponential family

2. $t(x) = \overline{x_i}$ is an efficient estimator of $E[t(x)] = \theta$.

NOTE: cannot conclude that $1/t(x)$ is efficient (or even unbiased) estimator of $1/\theta$.

Example 14 $\underline{x}$ *i.i.d.,* $x_i \sim \mathcal{N}(\theta, \sigma^2)$

Objective: find MLE of mean θ of univariate Gaussian with known variance σ^2

First find likelihood function

$$f(\underline{x};\theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^n (x_k-\theta)^2\right\}$$

Or

$$\ln f(\underline{x};\underline{\theta}) = -\frac{n}{2}\ \ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^n (x_k-\theta)^2 + c$$

Go right to the CRB equality condition to see if we can identify efficient estimator

$$\begin{aligned}\partial \ln f/\partial\theta &= \frac{1}{\sigma^2}\sum_{k=1}^n (x_k-\theta) \\ &= \underbrace{\frac{n}{\sigma^2}}_{k_\theta}\,(\overline{x_i}-\theta)\end{aligned}$$

CRB attainability condition is satisfied so that sample mean $\overline{x_i}$ is again optimal estimator

Conclude:

1. sample mean is efficient and has minimum variance among unbiased estimators of mean in Gaussian model with known variance

2. Take another derivative of loglikelihood w.r.t. θ and invert to find CRB = variance of $\overline{x_i}$

$$\mathrm{var}_\theta(\overline{x_i}) = \sigma^2/n$$

3. Leading factor in tight CRB condition is: $k_\theta = \mathrm{var}_\theta^{-1}(\overline{x_i})$

4.5 ESTIMATION OF MULTIPLE NON-RANDOM PARAMETERS

Arrange all unknown parameters in a vector: $\underline{\theta} = [\theta_1, \ldots, \theta_p]^T$
Density for measurement is:

$$f(x; \theta_1, \ldots, \theta_p) = f(x; \underline{\theta})$$

VECTOR ESTIMATOR PERFORMANCE CRITERIA
Estimator bias vector:

$$b_{\underline{\theta}}(\hat{\underline{\theta}}) = E_{\underline{\theta}}[\hat{\underline{\theta}}] - \underline{\theta}$$

Estimator covariance matrix (Hermitian symmetric):

$$\begin{aligned}
\mathrm{cov}_{\underline{\theta}}(\hat{\underline{\theta}}) &= E_{\underline{\theta}}[(\hat{\underline{\theta}} - E[\hat{\underline{\theta}}])(\hat{\underline{\theta}} - E[\hat{\underline{\theta}}])^T] \\
&= \begin{bmatrix} \mathrm{var}_{\underline{\theta}}(\hat{\theta}_1) & \mathrm{cov}_{\underline{\theta}}(\hat{\theta}_1, \hat{\theta}_2) & \ldots & \mathrm{cov}_{\underline{\theta}}(\hat{\theta}_1, \hat{\theta}_p) \\ \mathrm{cov}_{\underline{\theta}}(\hat{\theta}_2, \hat{\theta}_1) & \mathrm{var}_{\underline{\theta}}(\hat{\theta}_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathrm{cov}_{\underline{\theta}}(\hat{\theta}_p, \hat{\theta}_1) & \cdots & \cdots & \mathrm{var}_{\underline{\theta}}(\hat{\theta}_p) \end{bmatrix}
\end{aligned}$$

Estimator concentration:

$$\begin{aligned}
P_{\underline{\theta}}(\|\hat{\underline{\theta}} - \underline{\theta}\| > \epsilon) &= \int_{\|\hat{\underline{\theta}} - \underline{\theta}\| > \epsilon} f(\hat{\theta}; \theta) d\hat{\theta} \\
&= \int_{\{\underline{x}: \|\hat{\underline{\theta}}(\underline{x}) - \underline{\theta}\| > \epsilon\}} f(\underline{x}; \theta) d\underline{x}
\end{aligned}$$

4.5.1 MATRIX CRAMÈR-RAO BOUND (CRB) ON COVARIANCE

Let $\underline{\theta} \in \Theta$ be a $p \times 1$ vector and assume:
1. Θ is an open subset of $\mathbb{R}^p$
2. $f(x; \underline{\theta})$ is smooth [12] and differentiable in $\underline{\theta}$
3. $\mathrm{cov}_{\underline{\theta}}(\hat{\underline{\theta}})$ and $F(\underline{\theta})$ (defined below) are non-singular matrices

CRAMER-RAO BOUND FOR VECTOR PARAMETERS
For any *unbiased* estimator $\hat{\underline{\theta}}$ of $\underline{\theta}$

$$\mathrm{cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq F^{-1}(\underline{\theta}), \qquad \text{(CRB)}$$

where "=" is attained iff for some non-random matrix $K_{\underline{\theta}}$

$$\nabla_{\underline{\theta}} \ln f(x; \underline{\theta}) = K_{\underline{\theta}}(\hat{\underline{\theta}} - \underline{\theta})$$

In CRB $F(\underline{\theta})$ is the Fisher information matrix

$$
\begin{aligned}
F(\underline{\theta}) &= E\left[\left(\nabla_{\underline{\theta}} \ln f(x;\underline{\theta})\right)\left(\nabla_{\underline{\theta}} \ln f(x;\underline{\theta})\right)^T\right] \\
&= -E\left[\nabla^2_{\underline{\theta}} \ln f(x;\underline{\theta})\right]
\end{aligned}
$$

where we have defined the gradient operator

$$
\nabla_{\underline{\theta}} = \left[\frac{\partial}{\partial\theta_1}, \ldots, \frac{\partial}{\partial\theta_p}\right]^T
$$

and the symmetric Hessian (curvature) operator

$$
\nabla^2_{\underline{\theta}} = \begin{bmatrix}
\frac{\partial^2}{\partial\theta_1^2} & \frac{\partial^2}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2}{\partial\theta_1\partial\theta_p} \\
\frac{\partial^2}{\partial\theta_2\partial\theta_1} & \frac{\partial^2}{\partial\theta_2^2} & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
\frac{\partial^2}{\partial\theta_p\partial\theta_1} & \cdots & \cdots & \frac{\partial^2}{\partial\theta_p^2}
\end{bmatrix}
$$

Important Properties of matrix CR Bound

$$
\mathrm{cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq F^{-1}(\underline{\theta})
$$

Property 1: The inequality in the bound is in the sense of positive definiteness.

Specifically if A, B are $p \times p$ matrices

$$
A \geq B \iff A - B \geq 0
$$

where $A - B \geq 0$ means $A - B$ is non-negative definite, i.e.:

* $\underline{z}^T(A-B)\underline{z} \geq 0$ for any vector $\underline{z} \in \mathbb{R}^p$

* all eigenvalues of $A - B$ are non-negative

In particular, choosing $\underline{z} = [1, 0, \ldots, 0]^T$ and $\underline{z} = [1, \ldots, 1]^T$, respectively, $A \geq B$ implies

$$
a_{ii} \geq b_{ii}, \quad \text{and} \quad \sum_{i,j} a_{ij} \geq \sum_{ij} b_{ij}
$$

However, $A \geq B$ does NOT mean $a_{ij} \geq b_{ij}$ in general.

Counterexample: Let $0 < \rho < 1$ and consider

$$
\underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}}_{A} - \underbrace{\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}}_{B} = \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}
$$

which has two eigenvalues $1-\rho > 0$ and $1+\rho > 0$.

Hence $A - B > 0$ while clearly $a_{12} = 0 \not> \rho$.

Property 2. Fisher information matrix is a measure of the average curvature profile of the log likelihood near $\underline{\theta}$

Property 3. Let $F_n(\underline{\theta})$ be the Fisher information for a sample of n i.i.d. measurements $x_1, \ldots, x_n$. Then

$$F_n(\underline{\theta}) = nF_1(\underline{\theta})$$

so that $\text{var}_{\underline{\theta}}(\hat{\underline{\theta}}) = O(1/n)$ is also expected for multiple unknown parameters and smooth $f(x;\theta)$.

Property 4. Efficient vector estimators only exist for multiparameter exponential families

$$f(x;\underline{\theta}) = a(\underline{\theta})b(x)e^{-[\underline{c}(\underline{\theta})]^T[\underline{t}(x)]}$$

Property 5. If an estimator $\hat{\underline{\theta}}$ satisfies

$$\nabla_{\underline{\theta}} \ln f = K_{\underline{\theta}}(\hat{\underline{\theta}} - \underline{\theta})$$

Then we can immediately conclude:

1. $\hat{\underline{\theta}}$ is unbiased since, as shown in proof of CRB

$$E_{\underline{\theta}}[\nabla_{\underline{\theta}} \ln f(x;\underline{\theta})] = 0$$

2. $\hat{\underline{\theta}}$ is efficient and thus UMVUE

3. The covariance of $\hat{\underline{\theta}}$ is given by the inverse Fisher information $F(\underline{\theta})$

4. If $K_{\underline{\theta}} = K$ (not a function of $\underline{\theta}$) then $F(\underline{\theta}) = K_{\underline{\theta}}$ and

$$\text{cov}_{\underline{\theta}}(\hat{\underline{\theta}}) = K^{-1}$$

Proof of Matrix CR bound:

Recall: $\hat{\underline{\theta}}$ is unbiased and

$$E_{\underline{\theta}}[g(x)] := \int_{\mathcal{X}} g(x) f(x;\underline{\theta}) dx$$

3 steps in derivation:

* gradient of loglikelihood is zero mean

* correlation between gradient of loglikelihood and estimator bias is constant

* covariance matrix of gradient and bias gives relation between Fisher info and estimator covariance

Step 1. Show $E_{\underline{\theta}}\left[\nabla_{\underline{\theta}} \ln f(x;\underline{\theta})\right] = 0$

$$
\begin{aligned}
\Rightarrow \quad &= E_{\underline{\theta}}\left[\frac{1}{f(x;\underline{\theta})}\ \nabla_{\underline{\theta}} f(x;\underline{\theta})\right] = \int_{\mathcal{X}} \nabla_{\underline{\theta}} f(x;\underline{\theta}) dx \\
&= \underbrace{\nabla_{\underline{\theta}} \int_{\mathcal{X}} f(x;\underline{\theta}) dx}_{=1} = 0
\end{aligned}
$$

Step 2. $E_{\underline{\theta}}\left[\nabla_{\underline{\theta}} \ln f(x;\underline{\theta})\ (\underline{\hat{\theta}} - \underline{\theta})^T\right] = I$

First observe

$$
\begin{aligned}
E_{\underline{\theta}}\left[\nabla_{\underline{\theta}} \ln f(x;\underline{\theta})\ \underline{\hat{\theta}}^T\right] &= E_{\underline{\theta}}\left[\frac{1}{f(x;\underline{\theta})}\ \nabla_{\underline{\theta}} f(x;\underline{\theta})\underline{\hat{\theta}}^T\right] \\
&= \int_{\mathcal{X}} \nabla_{\underline{\theta}} f(x;\underline{\theta}) \hat{\theta}^T(x) dx \\
&= \nabla_{\underline{\theta}} \underbrace{\int_{\mathcal{X}} f(x;\underline{\theta}) \hat{\theta}^T(x) dx}_{E_{\underline{\theta}}[\underline{\hat{\theta}}^T] = \underline{\theta}^T} \\
&= I
\end{aligned}
$$

Now putting this together with result of previous step

$$
\begin{aligned}
&E_{\underline{\theta}}\left[\nabla_{\underline{\theta}} \ln f(x;\underline{\theta})\ (\underline{\hat{\theta}} - \underline{\theta})^T\right] \\
&= \underbrace{E_{\underline{\theta}}\left[\nabla_{\underline{\theta}} \ln f(x;\underline{\theta})\ \underline{\hat{\theta}}^T\right]}_{=I} + \underbrace{E_{\underline{\theta}}\left[\nabla_{\underline{\theta}} \ln f(x;\underline{\theta})\right]}_{=0}\ \underline{\theta}^T
\end{aligned}
$$

Step 3. Define a $2p \times 1$ vector U

$$
U = \begin{bmatrix} \underline{\hat{\theta}} - \underline{\theta} \\ \nabla_{\underline{\theta}} \ln f \end{bmatrix} \tag{8}
$$

Then, since outer product matrices are always non-negative definite

$$
E_{\underline{\theta}}\left[UU^T\right] \geq 0
$$

using results of steps 1 and 2, we have

$$
E_{\underline{\theta}}\left[UU^T\right] = \begin{bmatrix} \mathrm{cov}_{\underline{\theta}}(\underline{\hat{\theta}}) & I \\ I & F(\underline{\theta}) \end{bmatrix} \geq 0
$$

Next, let $\underline{w}$ and $\underline{y}$ be arbitrary p-vectors and define

$\underline{v} = \left[\begin{array}{c} \underline{w} \\ \underline{y} \end{array} \right]$

Then, as $\underline{v}^T \; E_{\underline{\theta}} \left[UU^T\right] \; \underline{v} \geq 0$

$$\underline{w}^T \text{cov}_{\underline{\theta}}(\hat{\underline{\theta}})\underline{w} + 2\underline{w}^T\underline{y} + \underline{y}^T F(\underline{\theta})\underline{y} \geq 0$$

Finally, take $\underline{y} = -F^{-1}(\underline{\theta}) \; \underline{w}$ in the above to obtain

$$\underline{w}^T [\text{cov}_{\underline{\theta}}(\hat{\underline{\theta}}) - F^{-1}(\underline{\theta})]\underline{w} \geq 0$$

It only remains to look at condition for equality in the CRB. Note first that if $\text{cov}_\theta(\hat{\theta}) = F^{-1}$ then $E_\theta[UU^T]$ necessarily has rank p (see exercises at end of chapter). This can only happen if U has p linearly dependent components:

$$U = \left[\begin{array}{c} \hat{\underline{\theta}} - \underline{\theta} \\ \nabla_{\underline{\theta}} \ln f \end{array} \right]$$

* As $\text{cov}_{\underline{\theta}}(\underline{\theta})$ and $F(\underline{\theta})$ have been assumed non-singular

$\Rightarrow \hat{\underline{\theta}} - \underline{\theta}$ has no linear dependencies and neither does $\nabla_{\underline{\theta}} \ln f$.

Hence it can only be that

$$\nabla_{\underline{\theta}} \ln f = K_{\underline{\theta}}(\hat{\underline{\theta}} - \underline{\theta})$$

for some non-random matrix $K_{\underline{\theta}}$ ◇

4.5.2 METHODS OF MOMENTS (MOM) VECTOR ESTIMATION

Let $m_k = m_k(\underline{\theta})$ be the k-th order moment of $f(x; \underline{\theta})$. Then find K moments such that the vector function of $\underline{\theta} \in \mathbb{R}^p$

$$\underline{g}(\underline{\theta}) = [m_1(\theta), \ldots, m_K(\underline{\theta})]$$

can be inverted for $\underline{\theta}$

$$\underline{\theta} = \underline{g}^{-1}(m_1, \ldots, m_K)$$

The MOM estimator is then simply obtained by replacing m_k with its empirical estimate

$$\hat{\underline{\theta}} = \underline{g}^{-1}(\hat{m}_1, \ldots, \hat{m}_K)$$

where $\hat{m}_k = \frac{1}{n}\sum_{i=1}^n x_i^k$

4.5.3 MAXIMUM LIKELIHOOD (ML) VECTOR ESTIMATION

Vector ML estimator is direct generalization of scalar MLE

$$\hat{\theta} = \text{argmax}_{\theta \in \Theta} f(x; \theta)$$

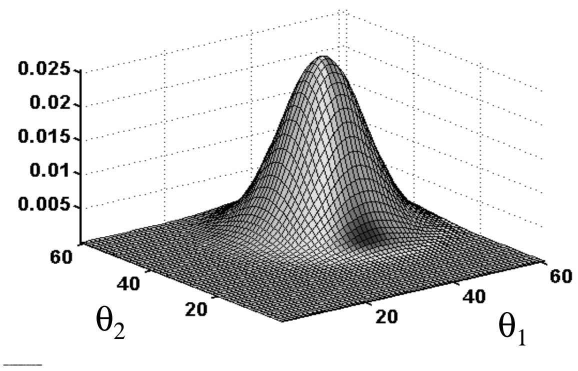


Figure 32: 2-*dimensional likelihood function*

Important properties of ML estimators for smooth likelihoods ([12]):

1. Vector MLE's are asymptotically unbiased

2. Vector MLE's are consistent

3. Vector MLE's are invariant to arbitrary vector transformations

$$\underline{\varphi} = \underline{g}(\underline{\theta}) \quad \Rightarrow \quad \hat{\underline{\varphi}} = \underline{g}(\hat{\underline{\theta}})$$

4: Vector MLE's are asymptotically efficient and UMVU

5. Vector MLE's are asymptotically Gaussian in the sense

$$\sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}) \to \underline{z}, \quad (i.d.)$$

where $\underline{z} \sim \mathcal{N}_p(0, F_1^{-1}(\underline{\theta}))$

$F_1(\underline{\theta})$ is the single sample Fisher information matrix

$$F_1(\underline{\theta}) \quad = \quad -E_{\underline{\theta}}\left[\nabla^2_{\underline{\theta}} \log f(x_1; \underline{\theta})\right]$$

Example 15 *Joint estimation of mean and variance in Gaussian*

* $\underline{x} = [x_1, \ldots, x_n]$, x_i's i.i.d., $x_i \sim \mathcal{N}(\mu, \sigma^2)$,
* $\underline{\theta} = [\mu, \sigma^2]$ (here $p = 2$).

$$\ln f(\underline{x}; \underline{\theta}) = -\frac{n}{2}\ \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{k=1}^{n} (x_k - \mu)^2 + c$$

A. MOM approach

We know

* $m_1 = \mu$, $m_2 = \sigma^2 + \mu^2$

and thus

$$\mu = m_1, \quad \sigma^2 = m_2 - m_1^2$$

hence a MOM estimator is:

$$\begin{aligned} \hat{\underline{\theta}} &= [\hat{\mu}, \hat{\sigma^2}] \\ &= [\hat{m}_1, \hat{m}_2 - \hat{m}_1^2] \\ &= \left[\overline{x_i}, \overline{x^2} - \overline{x_i}^2\right] \\ &= \left[\overline{x_i}, \overline{(x - \overline{x_i})^2}\right] \end{aligned}$$

where as usual

$$\begin{aligned} \overline{x_i} &= n^{-1} \sum_{k=1}^{n} x_k \\ \overline{(x - \overline{x_i})^2} &= n^{-1} \sum_{k=1}^{n} (x_k - \overline{x_i})^2 = \frac{n-1}{n} \mathsf{s}^2 \end{aligned}$$

and

$$\mathsf{s}^2 = (n-1)^{-1} \sum_{k=1}^{n} (x_k - \overline{x_i})^2$$

is sample variance.

B. ML approach

$$\ln f(x; \underline{\theta}) = -\frac{n}{2} \ln \theta_2 - \frac{1}{2\theta_2} \sum_{k=1}^{n} (x_k - \theta_1)^2 + c$$

As this is convex function in both θ_1 and θ_2 we can use likelihood equation (stationary point condition) for $\underline{\theta} = \underline{\hat{\theta}}$

$$0 = \nabla_{\underline{\theta}} \ln f = \left[\begin{array}{c} \frac{1}{\theta_2} \sum_{k=1}^n (x_k - \theta_1) \\ \\ \frac{n/2}{\theta_2} - \frac{1}{2\theta_2^2} \sum_{k=1}^n (x_k - \theta_1)^2 \end{array} \right]$$

or we have the same result as obtained using MOM

$$\hat{\theta}_1 = \hat{\mu} = \overline{x_i}, \qquad \hat{\theta}_2 = \hat{\sigma^2} = \frac{n-1}{n}\mathsf{s}^2$$

Next look at performance of MLE:

Bias:

$$\underbrace{E_{\underline{\theta}}[\hat{\mu}] = \mu,}_{unbiased} \qquad \underbrace{E_{\underline{\theta}}[\hat{\sigma^2}] = \left(\frac{n-1}{n}\right)\sigma^2}_{biased}$$

Covariance:

Find:

$$\mathrm{var}_{\underline{\theta}}(\overline{x_i}) = \sigma^2/n$$

and

$$\mathrm{var}_{\underline{\theta}}(\hat{\sigma^2}) = \left(\frac{n-1}{n}\right)^2 \mathrm{var}_{\underline{\theta}}(\mathsf{s}^2) = 2\sigma^4/n \ \left(\frac{n-1}{n}\right)$$

(to obtain $\mathrm{var}_{\underline{\theta}}(\mathsf{s}^2) = 2\sigma^4/(n-1)$ use fact that $\mathsf{s}^2 = \frac{1}{n-1}\ \chi_{n-1}$ for i.i.d. zero mean unit variance Gaussian x_i (see homework exercise).)

Since sample mean and sample variance are uncorrelated $\hat{\mu}$ and $\hat{\sigma^2}$ are also uncorrelated and

$$\mathrm{cov}_{\underline{\theta}}(\underline{\hat{\theta}}) = \left[\begin{array}{cc} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n\ \left(\frac{n-1}{n}\right) \end{array} \right]$$

Next compute CR bound:

Fisher information matrix can be found from expectation of Hessian $-\nabla^2_{\underline{\theta}} \ln f$

$$F(\underline{\theta}) = \left[\begin{array}{cc} n/\sigma^2 & 0 \\ 0 & n/(2\sigma^4) \end{array} \right]$$

Giving the CR bound

$$\mathrm{cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}$$

OBSERVATIONS

Observation 1. MOM and ML estimators derived above have covariance which violates the CR bound. This is not a contradiction since the ML variance estimator is not unbiased.

Observation 2. Consider the bias-corrected estimator of $[\mu, \sigma^2]^T$

$$\hat{\hat{\underline{\theta}}} = [\overline{x_i}, \mathsf{s}^2]^T$$

This estimator is unbiased.

Now, as $\mathsf{s}^2 = \left(\frac{n}{n-1}\right)\hat{\sigma^2}$

$$\mathrm{var}_{\underline{\theta}}(\mathsf{s}^2) = \left(\frac{n}{n-1}\right)^2 \mathrm{var}_{\underline{\theta}}(\hat{\sigma^2})$$

$$\mathrm{cov}_{\underline{\theta}}(\hat{\hat{\underline{\theta}}}) = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n\left(\frac{n}{n-1}\right) \end{bmatrix} \geq \; F^{-1}(\underline{\theta})$$

* estimator covariance no longer violates CRB

* $\overline{x_i}$ is efficient estimator of μ since

$$\mathrm{var}_{\underline{\theta}}(\hat{\mu}) = [F^{-1}]_{11} = \sigma^2/n$$

* s^2 is not efficient estimator of σ^2 since

$$\mathrm{var}_{\underline{\theta}}(\hat{\mathsf{s}}^2) > [F^{-1}]_{22}$$

Observation 3. ML estimator is asymptotically efficient as $n \to \infty$.

$$n\mathrm{cov}_{\underline{\theta}}(\hat{\underline{\theta}}) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4\;\left(\frac{n-1}{n}\right) \end{bmatrix} \to \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} = F_1^{-1}(\underline{\theta})$$

Observation 4. $[\hat{\mu}, \hat{\sigma^2}]$ asymptotically Gaussian

a) $\hat{\mu}$ and $\hat{\sigma^2}$ are independent r.v.s

b) $\sqrt{n}(\hat{\mu} - \mu) = \mathcal{N}(0, \sigma^2)$

c) $\sqrt{n}(\mathsf{s}^2 - \sigma^2) = \sigma^2\sqrt{n}(\chi^2_{n-1}/(n-1) - 1)$

d) $\chi^2_\nu \;\sim\; \mathcal{N}(\nu, 2\nu)$, $\nu \to \infty$.

Observation 5. We can easily manipulate the condition for equality in the CR bound to find

$$\nabla_{\underline{\theta}} \ln f \;\; = \;\; K_{\underline{\theta}} \begin{bmatrix} \overline{x_i} - \mu \\ \overline{x_i^2} - (\sigma^2 + \mu^2) \end{bmatrix}$$

where

$$K_{\underline{\theta}} := \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2\mu & 1 \end{bmatrix}^{-1}$$

Conclude:

* $\overline{x_i}, \overline{x_i^2}$ are unbiased estimator of of the first and second (non-central) moments $E[x] = \mu$ and $E[x^2] = \sigma^2 + \mu^2$, respectively.

* $[\overline{x_i}, \overline{x_i^2}]$ is efficient and thus UMVUE estimator of $\left[E[x_i], E[x_i^2]\right]$

Example 16 $\underline{N} = [N_1, \ldots, N_p]^T$ *a multinomial random vector*

$$f(\underline{N}; \underline{\theta}) = \frac{n!}{N_1! \cdots N_p!} \theta^{N_1} \ldots \theta^{N_p}$$

where

* $N_i \geq 0$ are integers satisfying $\sum_{i=1}^p N_i = n$

* $\theta_i \in [0, 1]$ are cell probabilities satisfying $\sum_{i=1}^p \theta_i = 1$

Objective: find MLE of cell probabilities

CAUTION: p parameters $\underline{\theta}$ live in a $p-1$ subspace of $\mathbb{R}^p$ due to total cell probability constraint

Can find MLE either by reparameterization or using Lagrange multipliers

Will use Lagrange multipliers: MLE maximizes

$$J(\underline{\theta}) = \ln f(\underline{N}; \underline{\theta}) - \lambda \left(\sum_{i=1}^p \theta_i - 1 \right)$$

where λ is to be determined.

Step 1: As J is smooth and convex we set gradient of $J(\underline{\theta})$ to zero

$$\begin{aligned} 0 = \nabla_{\underline{\theta}} J(\underline{\theta}) &= \nabla_{\underline{\theta}} \left[\sum_{i=1}^p N_i \ln \theta_i - \lambda \theta_i \right] \\ &= \left[\frac{N_1}{\theta_1} - \lambda, \ldots, \frac{N_p}{\theta_p} - \lambda \right] \end{aligned}$$

Thus

$$\hat{\theta}_i = N_i / \lambda, \qquad i = 1, \ldots, p$$

Step 2: find λ by forcing $\hat{\underline{\theta}}$ to satisfy constraint

$$\sum_{i=1}^p N_i / \lambda = 1 \quad \Rightarrow \quad \lambda = \sum_{i=1}^p N_i = n$$

Hence MLE is

$$\underline{\hat{\theta}} = \left[\frac{N_1}{n}, \ldots, \frac{N_p}{n}\right]$$

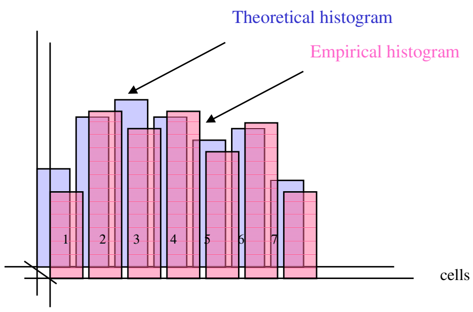


Figure 33: *MLE of cell probabilities from a multinomial mesurement is the empirical histogram*

4.6 HANDLING NUISANCE PARAMETERS

In many cases only a single parameter θ_1 is of direct interest

Other unknowns $\theta_2, \ldots, \theta_p$ are merely nuisance parameters.

CASE I: HANDLING RANDOM NUISANCE PARAMETERS:

Here average cost only penalizes for $\hat{\theta}_1$'s errors:

$$E[c(\hat{\theta}_1, \theta_1)] = \int_{\Theta_1} d\theta_1 \int_{\mathcal{X}} dx\ c(\hat{\theta}_1(x), \theta_1) f(x|\theta_1) f(\theta_1)$$

Prior on θ_1

$$f(\theta_1) = \int d\theta_2 \ldots \int d\theta_p\ f(\theta_1, \theta_2, \ldots, \theta_p)$$

Likelihood on θ_1

$$f(x|\theta_1) = \int d\theta_2 \ldots \int d\theta_p\ f(x|\theta_1, \theta_2, \ldots, \theta_p) f(\theta_2, \ldots, \theta_p|\theta_1)$$

Posterior on θ_1

$$f(\theta_1|x) = \int d\theta_2 \ldots \int d\theta_p \; f(\theta_1, \ldots, \theta_p|x)$$

OBSERVE:

* estimation of $\theta_2, \ldots, \theta_p$ is not required

* integration (marginalization) over $\theta_2, \ldots, \theta_p$ is required

CASE II: HANDLING NON-RANDOM NUISANCE PARAMETERS:

Average cost only penalizes for $\hat{\theta}_1$ errors but nonetheless depends on all unknowns:

$$E_{\underline{\theta}}[c(\hat{\theta}_1, \theta_1)] = \int_{\mathcal{X}} c(\hat{\theta}_1(x), \theta_1) f(x; \underline{\theta}) \; dx$$

Maximum Likelihood Estimator of θ_1

$$\hat{\theta}_1 = \text{argmax}_{\theta_1} \left(\max_{\theta_2, \ldots, \theta_p} \log f(x|\theta_1, \theta_2, \ldots, \theta_p) \right)$$

* joint maximum likelihood required

* to implement MLE need to estimate all parameters

CR BOUND PREDICTIONS FOR NON-RANDOM NUISANCE PARAMETERS

CRB for unbiased estimator $\hat{\theta}_1$ is equal to inverse Fisher matrix entry in the upper left hand corner

$$\text{var}_{\underline{\theta}}(\hat{\underline{\theta}}) \quad \geq \quad [[F^{-1}(\underline{\theta})]]_{11}$$

where as before

$$F(\underline{\theta}) = -E \begin{bmatrix} \frac{\partial^2 l(\underline{\theta})}{\partial \theta_1^2} & \frac{\partial^2 l(\underline{\theta})}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 l(\underline{\theta})}{\partial \theta_1 \partial \theta_p} \\ \frac{\partial^2 l(\underline{\theta})}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 l(\underline{\theta})}{\partial \theta_2^2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial^2 l(\underline{\theta})}{\partial \theta_p \partial \theta_1} & \cdots & \cdots & \frac{\partial^2 l(\underline{\theta})}{\partial \theta_p^2} \end{bmatrix}$$

$l(\underline{\theta}) = \ln f(x; \underline{\theta})$

Let the Fisher matrix be partitioned as

$$F(\underline{\theta}) = \begin{bmatrix} a & \underline{b}^T \\ \underline{b} & C \end{bmatrix}$$

* $a = -E_{\underline{\theta}}[\partial^2 \ln f / \partial \theta_1^2]$ = Fisher info for θ_1 w/o nuisance

* $\underline{b} = -E_{\underline{\theta}}[\partial \nabla_{\theta_2, \ldots, \theta_p} \ln f / \partial_1]$ = Fisher coupling of θ_1 to nuisance

* $C = -E_{\underline{\theta}}[\nabla^2_{\theta_2, \ldots, \theta_p} \ln f]$ = Fisher info for nuisance

Using matrix partition inverse identities the RHS of CRB can be expressed as

$$[[F^{-1}(\underline{\theta})]]_{11} = \frac{1}{a - \underline{b}^T C^{-1} \underline{b}}$$

OBSERVATIONS:

Observation 1: $[[F^{-1}(\underline{\theta})]]_{11} \geq 1/a = 1/[[F(\underline{\theta})]]_{11}$

Thus presence of nuisance parameters can only degrade estimator performance

Observation 2: the amount of degradation is directly proportional to the amount of information coupling between θ_1 and $\theta_2, \ldots, \theta_p$.

Observation 3: no degradation occurs when Fisher matrix is block diagonal

4.7 Exercises

1. Prove the useful formula $|a + \Delta| = |a| + \text{sgn}(a)\Delta + [\text{sgn}(a + \Delta) - \text{sgn}(a)](a + \Delta)$ in Sec. 4.2.2.
2. Show the equivalence of the two espression on the right hand side of the inequality (6).
3. Let $\underline{X} = [X_1, \ldots, X_n]^T$ be a vector of i.i.d. r.v.s X_i which are uniformly distributed over the interval (θ_1, θ_2), $\theta_1 < \theta_2$. Find the maximum likelihood estimator of $\underline{\theta}$.
4. Let Z_i, $i = 1, \ldots, n$, be a set of i.i.d. random variables each with the *alpha density*

$$p_\theta(z) = \frac{\beta}{\sqrt{2\pi}\Phi(\alpha) z^2} \exp\left(-\tfrac{1}{2}\left[\alpha - \beta/z\right]^2\right),$$

   where $\beta > 0$ is unknown, α is known and $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$ is the standard normal CDF. Assuming that $\alpha = 0$ and that β has an exponential prior density: $p(\beta) = \frac{1}{\sigma_\beta} e^{-\beta/\sigma_\beta}$, where $\sigma_\beta > 0$ is known. Find an expression for the MAP estimate of β. What does the MAP estimate reduce to as $\sigma_\beta \to \infty$ (least informative prior)?
5. Let W_i, $i = 1, \ldots, n$, be a set of zero mean i.i.d. Gaussian random variables with variance σ_w^2. Let a be a zero mean Gaussian random variable with variance σ_a^2 which is independent of W_i. The objective is to estimate the value of a given the observation

$$X_i = a + W_i, \quad i = 1, \ldots, n$$

   (a) Find the MMSE estimator of a. How does this estimator compare to the MAP and MMAE estimators of a?
   (b) Compute the MSE of the MMSE estimator (Hint: express error as a sum of two independent r.v.'s to simplify algebra). What happens to the MSE as $n \to \infty$ or as SNR $= \sigma_a^2/\sigma_w^2 \to \infty$?
6. Let $\underline{X} = [X_1, \ldots, X_n]^T$ be a vector of i.i.d. Gaussian r.v.s with mean μ and variance $\sigma^2 = \mu^2$ ($X_i \sim \mathcal{N}(\mu, \mu^2)$).
   (a) Find a method of moments (MOM) estimator of μ based on the first moment.
   (b) Find the maximum likelihood estimate of μ.
7. Let X_i, $i = 1, \ldots, n$, be an i.i.d. sample from the shifted exponential density $f(x; \theta) = e^{-(x-\theta)}$, $x \geq \theta$, where θ is an unknown parameter $-\infty < \theta < \infty$. Assume that $n > 1$.

(a) Find a MOM estimator of θ.

(b) Find the ML estimator of θ.

(c) Assuming the exponential prior for θ, $f(\theta) = e^{-\theta}$, $\theta \geq 0$, find the MAP estimator, the MMSE estimator, and the MMAE estimator of θ given the i.i.d. sample (be careful with your limits of integration in computing $f(\theta|\underline{x})$!). What happens to these estimators as $n \to \infty$?

(d) Calculate the MSE of each of the estimators derived in part (c) (assume large n). Verify that the MMSE estimator has the lowest MSE.

8. The mean square error of a certain unbiased estimator $\hat{\theta}(x)$ of the mean of a measured random variable is equal to $\sigma^2/2$ where $\sigma^2 = \text{var}(x)$. What if anything does this tell you about the distribution of x (Hint: what does the CR bound say about distributions that are impossible)?

9. Available are n i.i.d. samples of a random variable x with density

$$f(x;\theta) = \tfrac{1}{2}\frac{1 + 3\theta x^2}{1 + \theta}$$

where $-1 \leq x \leq 1$ and $\theta \geq 0$.

(a) Is this density in the exponential family?

(b) Is the sample mean a sufficient statistic? If so, prove it for general n. If not, give a counterexample, e.g. specialize to $n = 2$.

(c) Find a MOM estimator of θ.

(d) Find the CR bound on estimator variance for any unbiased estimator of θ.

(e) Using either numerical integration (MATLAB) or analysis find the bias and variance of the MOM estimator and compare to the CR bound for large n (e.g. $n = 100$).

10. Let the observation X have conditionally uniform density

$$f(x|\theta) = \begin{cases} \frac{1}{\theta}, & 0 < x \leq \theta \\ 0, & o.w. \end{cases}$$

where θ is a random variable with density

$$f_\theta(\theta) = \begin{cases} \theta \exp(-\theta), & \theta \geq 0 \\ 0, & o.w. \end{cases}$$

A useful formula ($v \geq 0$): $\int_v^\infty u e^{-u} du = (v+1)e^{-v}$

(a) Find the MAP estimator of θ.

(b) Find the minimum mean squared error estimator of θ.

(c) Find the minimum mean absolute error estimator of θ.

11. Let Z be a single observation having density function

$$p_\theta(z) = (2\theta z + 1 - \theta), \qquad 0 \leq z \leq 1$$

where $-1 \leq \theta \leq 1$.

(a) Assuming that θ is a nonrandom parameter, find and plot the maximum likelihood estimator of θ as a function of Z.

(b) Is the ML estimator unbiased? If so does it achieve the CR bound?

(c) Now assume that θ is a random variable with uniform prior density: $p_\theta(\theta) = \frac{1}{2}, \quad \theta \in [-1, 1]$. Find and plot the minimum mean square error estimator of θ as a function of Z.

(d) Compute the conditional bias $E[\hat{\theta}|\theta] - \theta$ and the conditional MSE $E[(\hat{\theta} - \theta)^2|\theta]$ given θ for each of the estimators of part a and c. Plot the two conditional MSE functions obtained and compare the MSE's of the two estimators. Does one estimator perform uniformly better than the other in terms of MSE?

12. $\underline{X} = [X_1, \ldots, X_n]^T$ is an i.i.d. observation from the Gamma density

$$X_i \sim f(x|\theta) = \frac{1}{\Gamma(\theta)} x^{\theta-1} e^{-x}, \quad x \geq 0$$

where θ is an unknown non-negative parameter and $\Gamma(\theta)$ is the Gamma function. You should note the useful formulii [1, Sec. 6.1]

$$\Gamma(\theta) = \int_0^\infty x^{\theta-1} e^{-x} dx \quad and \quad \frac{\Gamma(\theta + k)}{\Gamma(\theta)} = \theta(\theta + 1) \ldots (\theta + k - 1)$$

.

(a) Find the CR bound on unbiased estimators of θ. You may leave your answer in terms of the first and second derivatives of the Gamma function.

(b) Find the first order MOM estimator of θ by matching ensemble mean to sample mean. Is your estimator unbiased?

13. In this exercise you will establish that UMVUE's do not always exist. Let Z be a r.v. with probability mass function

$$p_\theta(z) = \begin{cases} \theta, & z = -1 \\ (1 - \theta)^2 \theta^z, & z = 0, 1, 2, \ldots \end{cases}$$

where $\theta \in (0, 1)$.

(a) Define the estimator

$$\hat{\theta}_o(z) = \begin{cases} 1, & z = -1 \\ 0, & z = 0, 1, 2, \ldots \end{cases} .$$

Show that $\hat{\theta}_o$ is an unbiased estimator of θ.

(b) Note that any unbiased estimator $\hat{\theta}$ can be expressed in the form $\hat{\theta} = \hat{\theta}_o + U$ where $U = U(Z)$ is a statistic satisfying $E_\theta[U] = 0$ (any U satisfying this condition is called an *ancillary statistic*). Using this condition and the form for the pmf of Z given above, establish that U must be of the form $U(Z) = aZ$ for some non-random constant a (Hint: Z-transform tables may be helpful).

(c) Now find an expression for the variance of an unbiased $\hat{\theta}$ and show that the value a which minimizes the variance is a function of θ. Hence no single unbiased estimator can achieve minimum variance for all $\theta \in (0, 1)$ and therefore no UMVUE for θ exists.

(d) Show that a UMVUE for $\phi = (1 - \theta)^2$ *does* exist even though a UMVUE for θ does not exist (Hint: define $\hat{\phi}_o(z) = 1$ for $z = 0$ and $\hat{\phi}_o(z) = 0$, otherwise and repeat the steps in part a through c).

14. The observation consists of $x_1, \ldots, x_n$ i.i.d. samples where $x_i \sim f(x|\theta)$ and

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} x^{\frac{1}{\theta} - 1}, & 0 \leq x \leq 1 \\ 0, & o.w. \end{cases}$$

where θ, $0 < \theta < \infty$ is an unknown parameter.

(a) Compute the CR bound on unbiased estimators of θ. Is there an estimator that achieves the bound?

(b) Find the maximum likelihood estimator of θ.

(c) What is the mean and variance of the maximum likelihood estimator? Specify a class of functions $\varphi = g(\theta)$ for which the maximum likelihood estimator of φ is efficient.

(d) From one of your answers to parts a-c you should be able to derive the following formula

$$\int_0^1 u^\beta \ln\left(\frac{1}{u}\right) du = \frac{1}{(1+\beta)^2}, \qquad \beta > -1.$$

15. The measurement $\underline{x} = [x_1, \ldots, x_n]^T$ is i.i.d. Gaussian with unknown mean μ and variance σ^2.

    (a) Show that the sample mean $\overline{x_i} = n^{-1}\sum_{i=1}^n x_i$ and sample variance $\mathsf{s}^2 = (n-1)^{-1}\sum_{k=1}^n (x_k - \overline{x_i})^2$ are unbiased estimators and that they are uncorrelated *and* independent random variables (**Hint**: show that the Gaussian random variables $x_i - \overline{x_i}$ and $\overline{x_i}$ are uncorrelated for $i = 1, \ldots, n$).

    (b) Using the results of part (a) derive the covariance matrix for the estimator $\hat{\underline{\theta}} = [\overline{x_i}, \mathsf{s}^2]^T$. (Hint: to save yourself lots of algebra you should represent $\mathsf{s}^2 = \mathsf{s}^2(\underline{x})$ in terms of σ^2 and the sample variance $\mathsf{s}^2(\underline{z})$ for $\underline{z}$ a vector of n i.i.d. zero mean unit variance Gaussian variables. Then use the representation (ch. 3 of course notes) $\mathsf{s}^2(\underline{z}) = \frac{1}{n-1}\chi_{n-1}$ and properties of the Chi square r.v. to find the expression for variance of s^2).

    (c) Derive the CR bound on the covariance matrix of any unbiased estimator $\hat{\underline{\theta}}$ of $\underline{\theta} = [\theta_1, \theta_2]^T = [\mu, \sigma^2]^T$. Compare to the result of part (b).

16. Available are n i.i.d. samples of a random variable x with density

    $$f(x;\theta) = \tfrac{1}{2}\frac{1+3\theta x^2}{1+\theta}$$

    where $-1 \leq x \leq 1$ and $\theta \geq 0$.

    (a) Is this density in the exponential family?

    (b) Is the sample mean a sufficient statistic? If so, prove it for general n. If not, give a counterexample, e.g. specialize to $n = 2$.

    (c) Find a MOM estimator of θ.

    (d) Find the CR bound on estimator variance for any unbiased estimator of θ.

17. Show that if the CR bound is attained with equality then $E_\theta[UU^T]$ has rank p, where U is given by (8). (Hint: show that the matrix

    $$E_{\underline{\theta}}\left[UU^T\right] = \begin{bmatrix} F^{-1}(\underline{\theta}) & I \\ I & F(\underline{\theta}) \end{bmatrix}$$

    has rank p.)

End of chapter

5 LINEAR LEAST SQUARES ESTIMATION

* Min MSE scalar, linear and affine estimation
* Projection theorem and orthogonality condition of linear estimation
* Optimality of affine estimator for jointly Gaussian case
* Nonstatistical linear least squares (LLS) estimation (regression)
* Optimality of LLS for Gaussian model

REFERENCES

Scharf [29] Kailath [14]

Rao [27]

Soderstrom and Stoica [32]

Measurement vector: $\underline{x} = [x_1, \ldots, x_n]^T$

Parameter vector: $\underline{\theta} = [\theta_1, \ldots, \theta_p]^T$

Two cases to consider:

Random vs non-random $\underline{x}$, $\underline{\theta}$

5.1 MIN MSE SCALAR, LINEAR, AND AFFINE ESTIMATION

Assume that $\underline{x}$ and $\underline{\theta}$ are random vectors

5.1.1 BEST CONSTANT ESTIMATOR OF RANDOM PARAMETERS

Quantities which must be known:

* Mean vector: $E[\underline{\theta}]$

Objective: find a constant

$$\hat{\underline{\theta}} = \underline{c}$$

which minimizes MSE:

$$\text{MSE}(\hat{\underline{\theta}}) = E[\|\underline{\theta} - \underline{c}\|^2] = \int_\Theta \|\underline{\theta} - \underline{c}\|^2 f(\underline{\theta})\, d\underline{\theta}$$

Solution:

$$\begin{aligned}
\text{MSE}(\hat{\underline{\theta}}) &= E[\|(\underline{\theta} - E[\underline{\theta}]) - (\underline{c} - E[\underline{\theta}])\|^2 \\
&= E[\|\underline{\theta} - E[\underline{\theta}]\|^2] + \|E[\underline{\theta}] - \underline{c}\|^2
\end{aligned}$$

$$-(E[\underline{\theta}] - \underline{c})^T \underbrace{E[\hat{\underline{\theta}} - E[\underline{\theta}]]}_{=0} - \underbrace{E[(\hat{\underline{\theta}} - E[\underline{\theta}])^T]}_{=0}(E[\underline{\theta}] - \underline{c})$$

Hence: $\underline{c} = E[\underline{\theta}]$ is the best constant

5.2 BEST LINEAR ESTIMATOR OF A SCALAR R.V. θ

Quantities which must be known:

* Measurement vector: $\underline{x} = [x_1, \ldots, x_n]^T$

* Means: $E[\underline{x}]$ and $E[\theta]$

* Second moment matrix:

$$M_x = E[\underline{x}\underline{x}^T]$$

* Cross moment vector

$$\underline{m}_{x\theta} = E[\underline{x}\theta]$$

Objective: find $\underline{h}$ to give optimal linear estimate $\hat{\theta}$

$$\hat{\theta} = \underline{h}^T \underline{x}$$

$\underline{h} = [h_1, \ldots, h_n]^T$ a non-random vector of coefficients

Optimality criterion: MSE

$$\begin{aligned}
\text{MSE}(\hat{\theta}) &= E[|\theta - \hat{\theta}|^2] = E[|\theta - \underline{h}^T\underline{x}|^2] \\
&= \underline{h}^T E[\underline{x}\underline{x}^T]\underline{h} + E[|\theta|^2] \\
&\quad -\underline{h}^T E[\underline{x}\theta] - E[\theta\underline{x}^T]\underline{h}
\end{aligned}$$

Note: Quadratic in unknowns $\underline{h}$

Optimal linear estimator can be found by differentiation:

$$\begin{aligned}
\underline{0}^T &= \nabla_{\underline{h}}\text{MSE}(\hat{\theta}) = \left[\frac{\partial}{\partial h_1}, \ldots, \frac{\partial}{\partial h_n}\right] \text{MSE}(\hat{\theta}) \\
&= 2\left(\underline{h}^T E[\underline{x}\underline{x}^T] - E[\theta\underline{x}^T]\right)
\end{aligned}$$

Therefore optimal $\underline{h}$ satisfies "Normal" equation:

$$E[\underline{x}\underline{x}^T]\underline{h} = E[\underline{x}\theta]$$

When $M_x = E[\underline{x}\underline{x}^T]$ is non-singular this is equivalent to

$$\underline{h} = M_x^{-1}\underline{m}_{x\theta}$$

and optimal linear estimator is

$$\begin{aligned} \hat{\theta} &= \underline{m}_{\theta x} M_x^{-1}\,\underline{x} \\ &= \underline{m}_{x\theta}^T M_x^{-1}\,\underline{x} \end{aligned}$$

which attains minimum MSE

$$\mathrm{MSE}_{\min} = E[|\theta|^2] - \underline{m}_{x\theta}^T M_x^{-1}\underline{m}_{x\theta}$$

5.3 BEST AFFINE ESTIMATOR OF A SCALAR R.V. θ

Quantities which must be known:

* Measurement vector: $\underline{x} = [x_1, \ldots, x_n]$
* Means: $E[\underline{x}]$ and $E[\theta]$
* Covariances:

$$R_x = \mathrm{cov}(\underline{x}), \quad \underline{r}_{x,\theta} = \mathrm{cov}(\underline{x}, \theta)$$

Objective: find optimal affine estimate $\hat{\theta}$ of a r.v. θ

$$\hat{\theta} = \underline{h}^T\underline{x} + b = \underline{h}^T(\underline{x} - E[\underline{x}]) + c$$

Optimality criterion: MSE

$$\begin{aligned} \mathrm{MSE}(\hat{\theta}) &= E[|\theta - \hat{\theta}|^2] = E[|(\theta - c) - \underline{h}^T(\underline{x} - E[\underline{x}])|^2] \\ &= \underline{h}^T \underbrace{E[(\underline{x} - E[\underline{x}])(\underline{x} - E[\underline{x}])^T]}_{R_x} \underline{h} + E[|\theta - c|^2] \\ &\quad -\underline{h}^T \underbrace{E[(\underline{x} - E[\underline{x}])(\theta - c)]}_{\underline{r}_{x,\theta}} - \underbrace{E[(\theta - c)(\underline{x} - E[\underline{x}])^T]}_{\underline{r}_{\theta,x}} \underline{h} \end{aligned}$$

Note: quadratic in unknowns $\underline{h}$, c.

Two observations:

1. As before we see that $c = E[\underline{\theta}]$ minimizes $\mathrm{MSE}(\hat{\theta})$
2. For $\underline{h}$, optimal solution is found by differentiation:

$$0 = \nabla_{\underline{h}}\mathrm{MSE} = \underline{h}^T R_x - \underline{r}_{\theta,x}$$

Or $\underline{h}$ satisfies equation:

$$R_x \underline{h} = \underline{r}_{x,\theta}$$

When R_x is non-singular this is equivalent to

$$\underline{h} = R_x^{-1}\underline{r}_{x,\theta}$$

and optimal affine estimator is

$$\hat{\theta} = E[\underline{\theta}] + \underline{r}_{x,\theta}^T R_x^{-1}(\underline{x} - E[\underline{x}])$$

which attains minimum MSE

$$\mathrm{MSE}_{\min} = \mathrm{var}(\theta) - \underline{r}_{x,\theta}^T R_x^{-1}\underline{r}_{x,\theta}$$

5.3.1 SUPERPOSITION PROPERTY OF LINEAR/AFFINE ESTIMATORS

Let ψ and ϕ be two random variables. Then, as statistical expectation is a linear operator, the best linear (affine) estimator of the sum $\theta = \psi + \phi$ given $\underline{x}$ is

$$\hat{\theta} = \hat{\psi} + \hat{\phi}, \tag{9}$$

where $\hat{\psi}$ and $\hat{\phi}$ are the best linear (affine) estimators of ψ given $\underline{x}$ and of ϕ given $\underline{x}$, respectively.

5.4 ORTHOGONALITY CONDITION AND PROJECTION THEOREM

Definition: $\mathcal{H}$ is a vector space over a scalar field $\mathcal{F}$ if for any elements $x, y, z \in \mathcal{H}$ and scalars $\alpha, \beta \in \mathcal{F}$

1. $\alpha \cdot x + \beta \cdot y \;\in\; \mathcal{H}$ (Closure)
2. $x + (y + z) = (x + y) + z$
3. $\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y$
4. $(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x$
5. There is a vector $\phi \in \mathcal{H}$ s.t.: $x + \phi = x$
6. There are scalars $1, 0$ s.t.: $1 \cdot x = x$, $0 \cdot x = \phi$

A normed vector space $\mathcal{H}$ has an inner product $< \cdot, \cdot >$ and a norm $\| \cdot \|$ which satsify

1. $< x, y > = < y, x >^*$

2. $< \alpha \cdot x, y >= \alpha^* < x, y >$
3. $< x + y, z >=< x, z > + < y, z >$
4. $\|x\|^2 =< x, x >$
5. $\|x\| \geq 0$
6. $\|x\| = 0$ iff $x = \phi$
7. $\|x + y\| \leq \|x\| + \|y\|$ (Triangle ineq)
8. $| < x, y > | \leq \|x\| \; \|y\|$ (Cauchy-Schwarz ineq)
9. Angle between x, y: $\psi = \cos^{-1}\left(\frac{<x,y>}{\|x\| \; \|y\|}\right)$
10. $< x, y >= 0$ iff x, y are orthogonal
11. $| < x, y > | = \|x\| \; \|y\|$ iff $x = \alpha \cdot y$ for some α
12. The linear span of vectors $\{x_1, \ldots, x_k\}$ is

$$span\,\{x_1, \ldots, x_k\} := \left\{ y : y = \sum_{i=1}^{k} \alpha_i \cdot x_i, \;\; \alpha_i \in \mathcal{F} \right\}$$

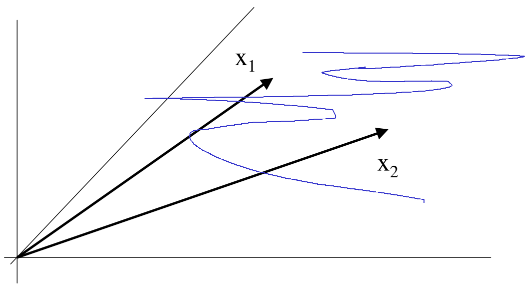


Figure 34: *Illustration of linear span of two vectors in* $\mathbb{R}^3$

13. A basis for $\mathcal{H}$ is any set of linearly independent vectors $x_1, \ldots x_k$ such that $span\{x_1, \ldots, x_k\} = \mathcal{H}$
14. The dimension of $\mathcal{H}$ is the number of elements in any basis for $\mathcal{H}$
15. A linear subspace $\mathcal{S}$ is any subset of $\mathcal{H}$ which is itself a vector space.
16. The projection x of a vector y onto a subspace $\mathcal{S}$ is a vector x that satisfies

$$< y - x, u >= 0, \qquad \text{for all } u \in \mathcal{S}$$

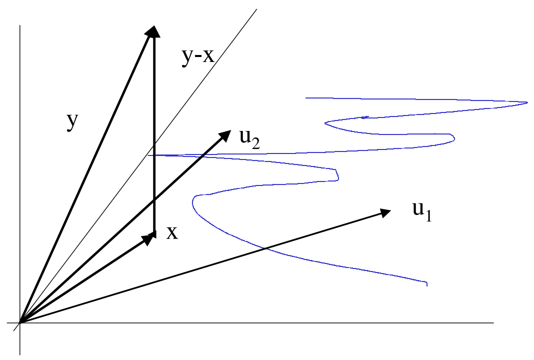


Figure 35: *The projection of a vector x onto a subspace $\mathcal{S}$ in the plane*

Examples of vector spaces:

1. Euclidean p-dimensional space $\mathbb{R}^p$:

$$< x, y >= x^T y = \sum_{i=1}^{p} x_i y_i$$

A one dimensional subspace: the line

$$\mathcal{S} = \{y : y = a\underline{v},\ a \in \mathbb{R}\}$$

where $\underline{v} \in \mathbb{R}^p$ is any fixed vector.

2. Complex p-space: $\underline{x} = [x_1, \ldots, x_p]$, $\underline{y} = [y_1, \ldots, y_p]$,

$$< \underline{x}, \underline{y} >= \underline{x}^H \underline{y} = \sum_{i=1}^{p} x_i^* y_i$$

An n-dimensional subspace:

$$\begin{aligned} \mathcal{S} &= \{y : y = \sum_{i=1}^{n} a_i \underline{v}_i,\ a_i \in \mathbb{C}\} \\ &= span\{\underline{v}_1, \ldots, \underline{v}_n\} \end{aligned}$$

where $\underline{v}_i \in \mathbb{C}^p$ are any linearly independent vectors in $\mathcal{H}$.

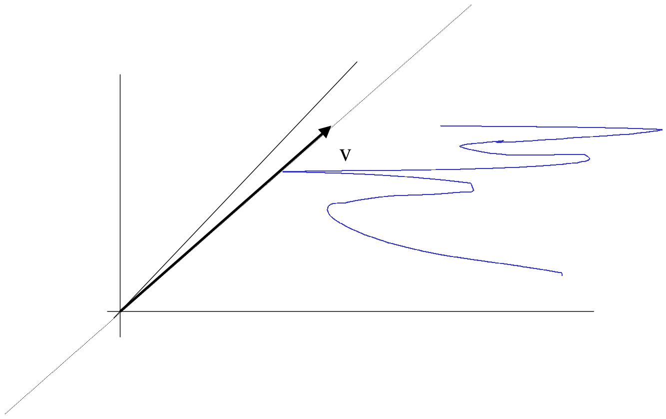


Figure 36: *A line is a one dimensional subspace of* $\mathcal{H} = \mathbb{R}^p$

3. The space of square integrable cts. time functions $x(t)$

$$< x, y >= \int x(t)y(t)\ dt$$

A one dimensional subspace: scales of a given function

$$\mathcal{S} = \{g : g(t) = a\ f(t),\ a \in \mathbb{R}\}$$

where $f = f(t)$ is any fixed function in $\mathcal{H}$.

4. The space of second order real random variables x defined on a sample space:

$$< x, y >= E[xy] = \int_\Omega x(\omega)y(\omega)f(\omega)\ d\omega$$

Ω: sample space of elementary outcomes ω

Q. How to use vector spaces for estimation?

A. Identify $\mathcal{H} = \{y : y \text{ a r.v. with } E[|y|^2] < \infty\}$.

Inner product beween two "vectors" in $\mathcal{H}$ is defined as

$$< x, y >:= E[xy]$$

(x, y real r.v.s)

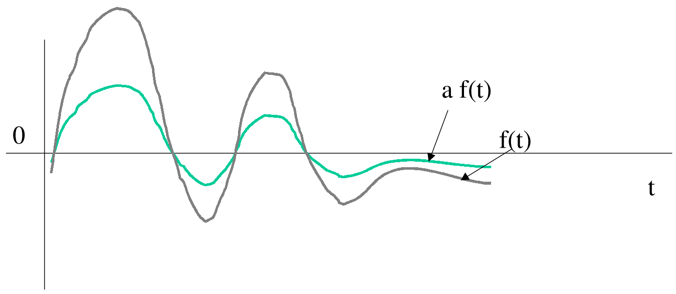


Figure 37: *All scalings of a fixed function is a one dimensional subspace of $\mathcal{H}$*

5.4.1 LINEAR MINIMUM MSE ESTIMATION REVISITED

Cast the linear estimation problem into $\mathcal{H}$

Observe r.v.s: $\underline{x} = [x_1, \ldots, x_n]^T$

Find linear estimator:

$$\hat{\theta} = \underline{h}^T \underline{x}$$

Identify "solution subspace"

$$\mathcal{S} := span\{x_1, \ldots, x_n\}$$

and norm

$$\|\theta - \hat{\theta}\|^2 = E[|\theta - \hat{\theta}|^2] = \text{MSE}(\hat{\theta})$$

Hence: best linear estimate of θ is a vector $\hat{\theta} \in \mathcal{S}$ which minimizes the norm squared $\|\theta - \hat{\theta}\|^2$.

Linear estimator projection theorem: the best linear estimator of θ based r.v.s $x_1, \ldots, x_n$ is the projection of θ onto $span\{x_1, \ldots, x_n\}$.

Orthogonality Condition: the best linear estimator $\hat{\theta}$ satisfies

$$< \theta - \hat{\theta}, u >= 0, \qquad \text{for all } u \in \mathcal{S}$$

Equivalently, if $u_1, \ldots, u_{n'}$ is any basis for $\mathcal{S}$

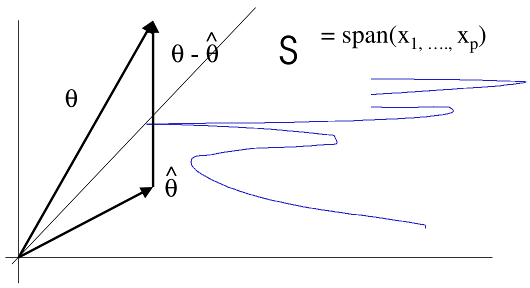


Figure 38: *The orthogonality condition for the best linear estimator $\hat{\theta}$ of a random variable θ given $x_1, \ldots, x_n$*

$$< \theta - \hat{\theta}, u_i >= 0, \qquad i = 1, \ldots, n'$$

Simple expression for minimum MSE:

$$\begin{aligned}
\|\theta - \hat{\theta}\|^2 &= < \theta - \hat{\theta}, \theta - \hat{\theta} > \\
&= < \theta - \hat{\theta}, \theta > - < \theta - \hat{\theta}, \underbrace{\hat{\theta}}_{\in \mathcal{S}} > \\
&= < \theta - \hat{\theta}, \theta >
\end{aligned}$$

5.4.2 AFFINE MINIMUM MSE ESTIMATION

Cast the affine estimation problem into $\mathcal{H}$

Observe r.v.s: $\underline{x} = [x_1, \ldots, x_n]^T$

Re-express estimator equation to identify solution subspace

$$\begin{aligned}
\hat{\theta} &= \underline{h}^T \underline{x} + b \\
&= [\underline{h}^T, b] \begin{bmatrix} \underline{x} \\ 1 \end{bmatrix}
\end{aligned}$$

Identify solution subspace and error norm

$$\mathcal{S} := span\{x_1, \ldots, x_n, 1\}$$

$$\|\theta - \hat{\theta}\|^2 = \text{MSE}(\hat{\theta})$$

Affine projection theorem: the best affine estimator of θ based r.v.s $x_1, \ldots, x_n$ is the projection of θ onto $span\{x_1, \ldots, x_n, 1\}$.

5.4.3 OPTIMALITY OF AFFINE ESTIMATOR FOR LINEAR GAUSSIAN MODEL

Assume $\underline{x}$, θ are jointly Gaussian

Then min MSE estimator $\underline{\hat{\theta}} = \underline{\hat{\theta}}(\underline{x})$ is in fact affine:

$$E[\theta|\underline{x}] = E[\theta] + \underline{r}_{x,\theta}^T(\underline{x} - E[\underline{x}])$$

Proof:

Let $\hat{\theta}_l$ be the affine min MSE estimator.

From projection theorem we know that error is orthogonal to measurements

$$E[(\theta - \hat{\theta}_l)\underline{x}] = \underline{0}$$

Note: Since $\theta - \hat{\theta}_l$ is a linear combination of Gaussian r.v.s it is itself Gaussian.

Since Gaussian r.v.s that are orthogonal are in fact independent r.v.'s

$$E[(\theta - \hat{\theta}_l)|\underline{x}] = E[(\theta - \hat{\theta}_l)] = 0$$

Therefore, as $\hat{\theta}_l$ is a function of $\underline{x}$ we have

$$0 = E[(\theta - \hat{\theta}_l)|x] = E[\theta|x] - \hat{\theta}_l$$

or

$$E[\theta|\underline{x}] = \hat{\theta}_l$$

Which establishes the desired result. $\diamond$

5.5 BEST AFFINE ESTIMATION OF A VECTOR

Quantities which must be known:

* Measurement vector: $\underline{x} = [x_1, \ldots, x_n]$
* Measurement mean vector: $E[\underline{x}]$
* Prior Mean Vector: $E[\underline{\theta}]$
* Covariances:

$$R_x = \text{cov}(\underline{x})$$

$$R_{x,\underline{\theta}} = \text{cov}(\underline{x}, \underline{\theta}) = [\underline{r}_{x,\theta_1}, \ldots, \underline{r}_{x,\theta_p}]$$

Objective: estimate $\underline{\theta} = [\theta_1, \ldots, \theta_p]^T$

Sum of MSE estimation criterion:

$$\begin{aligned} \text{MSE}(\hat{\underline{\theta}}) &= \sum_{i=1}^{p} \text{MSE}(\hat{\theta}_i) \\ &= \sum_{i=1}^{p} E|\theta_i - \hat{\theta}_i|^2 = \text{trace}\left(E[(\underline{\theta} - \hat{\underline{\theta}})\ (\underline{\theta} - \hat{\underline{\theta}})^T]\right) \end{aligned}$$

Affine estimator $\hat{\theta}_i$ is defined by

$$\hat{\theta}_i = \underline{h}_i^T \underline{x} + b_i, \quad i = 1, \ldots, p$$

and

$$\begin{aligned} \hat{\underline{\theta}} &= [\hat{\theta}_1, \ldots, \hat{\theta}_p]^T = H^T \underline{x} + \underline{b} \\ H &= [\underline{h}_1, \ldots, \underline{h}_p] \end{aligned}$$

SOLUTION TO MINIMUM SUM OF MSE PROBLEM:

Since $\underline{h}_i$'s appear seperately in each of the summands of $\text{MSE}(\hat{\underline{\theta}})$, the minimization of MSE is equivalent to the uncoupled minimization of each $\text{MSE}(\hat{\theta}_i)$.

$$\min_{\hat{\underline{\theta}}} \text{MSE}(\hat{\underline{\theta}}) = \sum_{i=1}^{p} \min_{\hat{\theta}_i} \text{MSE}(\hat{\theta}_i)$$

Therefore the minimum MSE solution is simply the concatenation of the scalar estimators of each θ_i:

$$\begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_p \end{bmatrix} = \begin{bmatrix} E[\theta_1] \\ \vdots \\ E[\theta_p] \end{bmatrix} + \begin{bmatrix} \underline{r}_{\theta_1,x} R_x^{-1}(\underline{x} - E[\underline{x}]) \\ \vdots \\ \underline{r}_{\theta_p,x} R_x^{-1}(\underline{x} - E[\underline{x}]) \end{bmatrix},$$

Or equivalently:

$$\underline{\hat{\theta}} = E[\underline{\theta}] + R_{\underline{\theta},x} R_x^{-1}(\underline{x} - E[\underline{x}])$$

resultant minimum sum of MSE

$$\mathrm{MSE}_{\min} = \text{trace}\left(R_\theta - R_{\underline{\theta},x} R_x^{-1} R_{x,\underline{\theta}}\right)$$

Example 17 *Min MSE Linear Prediction*

Linear p-th order 1-step predictor:

$$\hat{x}_k = \sum_{i=1}^{p} a_i x_{k-i}$$

$\{x_i\}$ a zero mean w.s.s. random sequence with a.c.f.

$$r(k) := E[x_i x_{i-k}]$$

Objective: find min MSE predictor coefficients $\{a_i\}$

Note: identify $\theta = x_k$ as random scalar "parameter," and $h_i = a_i$ coefficient vector to be determined

Solution:

Step 1. Rewrite predictor equation in vector form

$$\hat{x}_k = \underline{a}^T \underline{x}$$

where

$$\underline{x} = [x_{k-1}, \ldots, x_{k-p}]^T, \quad \underline{a} = [a_1, \ldots, a_p]^T$$

Step 2. orthogonality condition becomes

$$E[(x_k - \underline{a}^T \underline{x}) x_{k-i}] = 0, \quad i = 1, \ldots, p$$

or concatenation into column vector gives

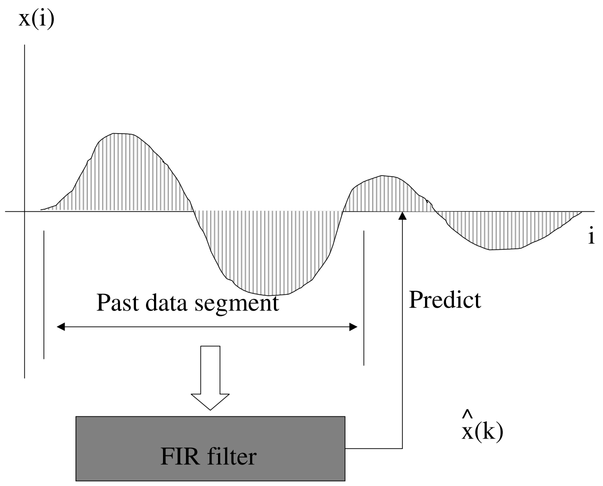


Figure 39: *Linear predictor as a FIR filter*

$$\underline{0} = \begin{bmatrix} E[(x_k - \underline{a}^T \underline{x}) x_{k-1}] \\ \vdots \\ E[(x_k - \underline{a}^T \underline{x}) x_{k-p}] \end{bmatrix} = E[(x_k - \underline{a}^T \underline{x}) \underline{x}]$$

Which specifies optimal predictor coefficients $\underline{a} = \hat{\underline{a}}$ as

$$\hat{\underline{a}} = R^{-1} \underline{r}$$

where we have defined the correlation vector:

$$\underline{r}^T = [r_1, \ldots, r_p] = E[\underline{x} x_k]$$

and Toeplitz covariance matrix

$$R = ((r_{(i-j)}))_{i,j=1,p} = E[\underline{x}\, \underline{x}^T]$$

The min MSE is

$$\begin{aligned} \text{MSE}_{\min} &= < x_k - \hat{\underline{a}}^T \underline{x}, x_k > \\ &= r_0 - \hat{\underline{a}}^T \underline{r} \\ &= r_0 - \underline{r}^T R^{-1} \underline{r} \end{aligned}$$

Relation: Linear prediction and AR(p) model for $\{x_i\}$.

Define residual prediction error $v_k = x_k - \hat{x}_k$. Then, obviously

$$x_k = \hat{x}_k + v_k = \sum_{i=1}^{p} a_i x_{k-i} + v_k$$

When v_k is w.s.s. white noise the above representation is called an *p-th order autoregressive*, or AR(p), model for $\{x_i\}$

Example 18 *An Inverse Problem*

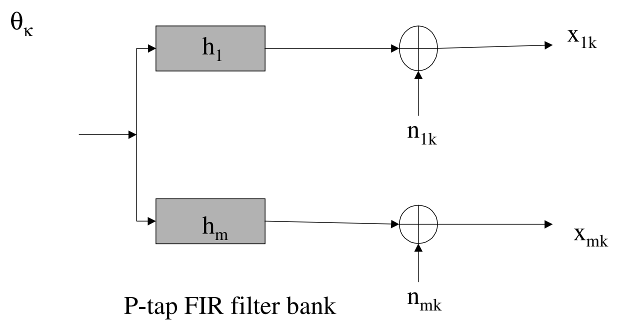


Figure 40: *Block diagram for inverse problem*

Measurement model:

$$\underline{x} = H\underline{\theta} + \underline{N}$$

* $\underline{x} = [x_1, \ldots, x_m]^T$: random measurements
* $\underline{\theta} = [\theta_1, \ldots, \theta_p]^T$: unknown random parameters
* $\underline{N} = [n_1, \ldots, n_m]^T$: zero mean measurement noise
* $\underline{\theta}, \underline{N}$ uncorrelated
* H: a known $m \times p$ matrix

Objective: find affine min MSE estimator $\hat{\underline{\theta}}$

Solution: directly follows from vector min MSE estimation results

$$\hat{\underline{\theta}} = E[\underline{\theta}] + R_{\underline{\theta},x} R_x^{-1} (\underline{x} - E[\underline{x}])$$

Find

$$\begin{aligned} E[\underline{x}] &= E[H\underline{\theta} + \underline{N}] = HE[\underline{\theta}] \\ R_x &= \text{cov}(\underbrace{H\underline{\theta} + \underline{N}}_{\text{uncorrelated}}) = HR_\theta H^T + R_N \\ R_{x,\underline{\theta}} &= \text{cov}((H\underline{\theta} + \underline{N}), \underline{\theta}) = HR_{\underline{\theta}} \end{aligned}$$

Final result:

$$\hat{\underline{\theta}} = E[\underline{\theta}] + R_{\underline{\theta}} H^T [HR_{\underline{\theta}} H^T + R_N]^{-1}(\underline{x} - HE[\underline{\theta}])$$

and min sum of MSE is

$$\text{MSE}_{\min} = \text{trace}\left(R_\theta - R_\theta H^T [HR_\theta H^T + R_N]^{-1} HR_\theta\right)$$

Remarks:

1. When R_N dominates $HR_\theta H^T$: $\text{MSE}_{\min} \approx \text{trace} R_{\underline{\theta}}$
2. When $HR_{\underline{\theta}} H^T$ dominates R_N *and H is full rank*: $\text{MSE}_{\min} \approx 0$.

5.6 NONSTATISTICAL LEAST SQUARES (LINEAR REGRESSION)

Pair of measurements available ($n \geq p$)

$$y_i, \ \ \underline{x}_i = [x_{i1}, \ldots, x_{ip}]^T, \quad i = 1, \ldots, n$$

(x_{ip} could be equal to x_{i-p} here, but not necessary).
Postulate an "input-output" relation:

$$y_i = \underline{x}_i^T \underline{a} + v_i, \quad i = 1, \ldots n$$

* y_i is response or output or dependent variable
* $\underline{x}_i$ is treatment or input or independent variable
* $\underline{a}$ is unknown $p \times 1$ coefficient vector to be estimated

$$\underline{a} = [a_1, \ldots, a_p]^T$$

Objective: find linear least squares estimator $\hat{\underline{a}}$ of $\underline{a}$ that minimizes sum of squared errors

$$\text{SSE}(\underline{a}) = \sum_{i=1}^n (y_i - \underline{x}_i^T \underline{a})^2$$

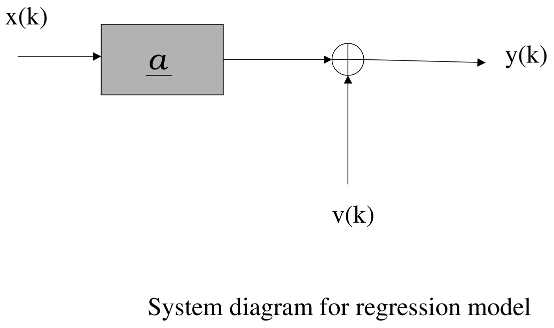


Figure 41: *System identification block diagram for linear regression*

Equivalent $n \times 1$ vector measurement model:

$$\begin{aligned} \begin{bmatrix} y_1, \\ \vdots \\ y_n \end{bmatrix} &= \begin{bmatrix} \underline{x}_1^T, \\ \vdots \\ \underline{x}_n^T \end{bmatrix} \underline{a} + \begin{bmatrix} v_1, \\ \vdots \\ v_n \end{bmatrix} \\ \underline{y} &= X\underline{a} + \underline{v}, \end{aligned}$$

where X is a $n \times p$ input matrix.

The estimation criterion is

$$\mathrm{SSE}(\underline{a}) = (\underline{y} - X\underline{a})^T(\underline{y} - X\underline{a})$$

Solution to LLSE of $\underline{a}$:

Step 1. Identify vector space containing $\underline{y}$: $\mathcal{H} = \mathbb{R}^n$

Inner product: $<\underline{y}, \underline{z}>= \underline{y}^T \ \underline{z}$

Step 2. Identify solution subspace containing $X\underline{a}$

$$\mathcal{S} = span\{\text{columns of X}\}$$

which contains vectors of form

$$X\underline{a} = \sum_{k=1}^{p} a_k \left[\ x_{1k}, \ldots, x_{nk} \ \right]^T$$

Step 3. apply projection theorem

Orthogonality Condition: the best linear estimator $\underline{\hat{a}}$ satisfies

$$< (\underline{y} - X\underline{\hat{a}}), \underline{u}_i >= 0, \qquad i = 1, \ldots, n$$

where $\underline{u}_i$ are columns of X, or equivalently

$$\underline{0}^T = (\underline{y} - X\underline{\hat{a}})^T X$$

$$= \underline{y}^T \underline{X} - \underline{\hat{a}}^T \; X^T X$$

or, if X has full column rank p then $X^T X$ is invertible and

$$\begin{aligned} \underline{\hat{a}} &= [X^T X]^{-1} \; X^T \underline{y} \\ &= \hat{R}_x^{-1} \; \underline{\hat{r}}_{xy} \end{aligned}$$

where

$$\hat{R}_x \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n \underline{x}_i \; \underline{x}_i^T, \qquad \underline{\hat{r}}_{xy} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n \underline{x}_i \; y_i$$

Projection operator form of predicted output response:

$$\underline{\hat{y}} = X\underline{\hat{a}}$$

which, using above, can be represented as the orthogonal projection of $\underline{y}$ onto $\mathcal{H}$

$$\begin{aligned} \underline{\hat{y}} &= X\underline{\hat{a}} \\ &= X[X^T X]^{-1} \; X^T \underline{y} \\ &= \underbrace{X[X^T X]^{-1} X^T}_{\text{orthog. projection}} \; \underline{y} \end{aligned}$$

Properties of orthogonal projection operator:

$$\Pi_X = X[X^T X]^{-1} X^T$$

Property 1. Π_X projects vectors onto column space of X

Define decomposition of $\underline{y}$ into components $\underline{y}_X$ in column space of X and $\underline{y}_X^{\perp}$ orthogonal to column space of X

$$\underline{y} = \underline{y}_X + \underline{y}_X^{\perp}$$

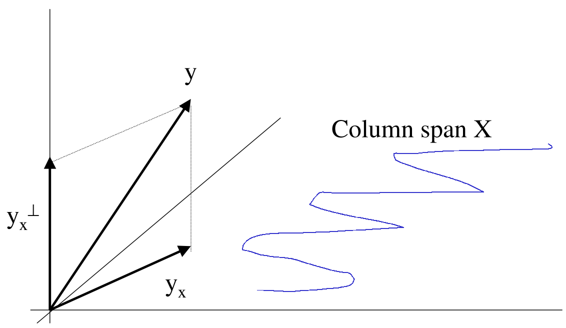


Figure 42: *Column space decomposition of a vector* $\underline{y}$

Then for some vector $\underline{\alpha} = [\alpha_1, \ldots, \alpha_n]^T$

$$\underline{y}_X = X\underline{\alpha}, \qquad X^T \underline{y}_X^{\perp} = \underline{0}$$

We have:

$$\begin{aligned}
\Pi_X \underline{y} &= \Pi_X(\underline{y}_X + \underline{y}^{\perp}) \\
&= X \underbrace{[X^T X]^{-1} X^T X}_{=I} \underline{\alpha} + X[X^T X]^{-1} \underbrace{X^T \underline{y}_X^{\perp}}_{=0} \\
&= X\underline{\alpha} \\
&= \underline{y}_X
\end{aligned}$$

so that Π_X extracts the column space component of $\underline{y}$.

Thus we can identify $\underline{y}_X = \Pi_X \underline{y}$ so that we have the representation

$$\underline{y} = \Pi_X \underline{y} + \underbrace{(I - \Pi_X)\underline{y}}_{\underline{y}_X^\perp}$$

It follows immediately that

2. $I - \Pi_X$ projects onto spac eorthogonal to $span\{\text{cols}X\}$

3. Π_X is symmetric and idempotent: $\Pi_X^T \Pi_X = \Pi$

4. $(I - \Pi_X)\Pi_X = 0$

Projection operator form of LS estimator gives alternative expression for minimum SSE

$$\begin{aligned} \mathrm{S}SE_{\min} &= (\underline{y} - \hat{\underline{y}})^T(\underline{y} - \hat{\underline{y}}) \\ &= \underline{y}^T[I - \Pi_X]^T \; [I - \Pi_X]\underline{y} \\ &= \underline{y}^T[I - \Pi_X]\underline{y} \end{aligned}$$

Example 19 *LS optimality of sample mean*

Measure $\underline{x} = [x_1, \ldots, x_n]^T$

Objective: Find best constant c which minimizes the sum of squares

$$\sum_{k=1}^{n}(x_i - c)^2 = (\underline{x} - c\underline{1})^T(\underline{x} - c\underline{1})$$

where $\underline{1} = [1, \ldots, 1]^T$

Step 1: identify solution subspace

$\mathcal{S}$ is diagonal line: $\{\underline{y} : \underline{y} = a\underline{1},\ a \in \mathbb{R}\}$

Step 2. apply orthogonality condition

$$(\underline{x} - c\underline{1})^T\underline{1} = 0 \iff c = \frac{\underline{x}^T\underline{1}}{\underline{1}^T\underline{1}} = n^{-1}\sum_{k=1}^{n} x_i$$

Example 20 *LLS linear prediction from training sample*

Measurement sequence $\{z_i\}$

Training sequence of $n + p$ samples of z_i

$$\{z_i\}_{i=1}^{p+n}, \quad i = 1, \ldots, n$$

Fit an AR(p) model to training sequence

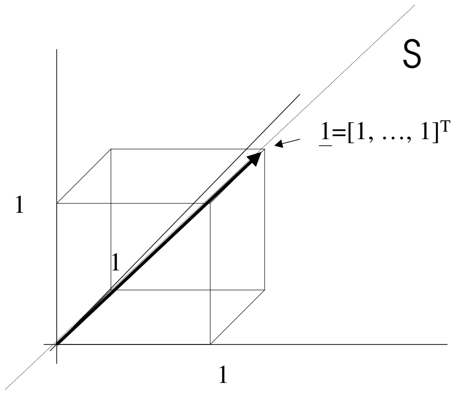


Figure 43: *Diagonal line is solution subspace for LS scalar*

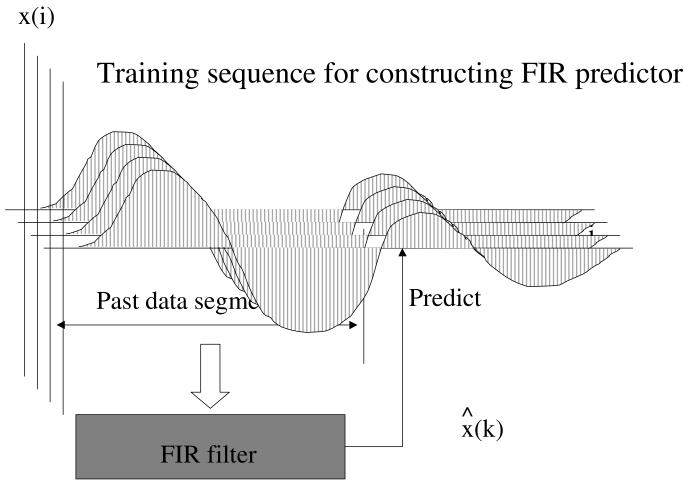


Figure 44: *Construction of LLS predictor from training sequence*

$$z_k = \sum_{i=1}^{p} a_i z_{k-i} + v_k, \quad k = p+1, \ldots, n$$

such that SSE is minimized

$$\mathrm{S}SE(n) = \sum_{k=1}^{n} (z_{k+p} - \sum_{i=1}^{p} a_i z_{k+p-i})^2$$

Solution

Step 1. Identify response variables $y_k = z_k$ and input vectors $\underline{x}_k = [z_{k-1}, \ldots, z_{k-p}]^T$.

$$\begin{aligned} \begin{bmatrix} z_{n+p}, \\ \vdots \\ z_{p+1} \end{bmatrix} &= \begin{bmatrix} \underline{z}_{n+p}^T, \\ \vdots \\ \underline{z}_{p+1}^T \end{bmatrix} \underline{a} + \begin{bmatrix} v_{n+p} \\ \vdots \\ v_{p+1} \end{bmatrix} \\ \underline{y} &= X\underline{a} + \underline{v}, \end{aligned}$$

Step 2. Apply orthogonality condition

The LLS p-th order linear predictor is of the form:

$$\hat{z}_k = \sum_{i=1}^{p} \hat{a}_i z_{k-i}$$

where $\underline{\hat{a}} = [\hat{a}_1, \ldots, \hat{a}_p]^T$ is obtained from formula

$$\underline{\hat{a}} = [X^T X]^{-1} X^T \underline{y} = \hat{R}^{-1}\, \underline{\hat{r}}$$

and we have defined the sample correlation quantities:

$$\hat{r} = [\hat{r}_1, \ldots, \hat{r}_p]^T$$

$$\hat{R} = ((\hat{r}(i-j)))_{i,j=1,p}$$

$$\hat{r}_j := n^{-1} \sum_{i=1}^{n} z_{i+p} z_{i+p-j}, \quad j = 0, \ldots, p$$

5.7 WEIGHTED LINEAR LEAST SQUARES

As before assume linear model for input and response variables

$$\begin{bmatrix} y_1, \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \underline{x}_1^T, \\ \vdots \\ \underline{x}_n^T \end{bmatrix} \underline{a} + \begin{bmatrix} v_1, \\ \vdots \\ v_n \end{bmatrix}$$

$$\underline{y} = X\underline{a} + \underline{v},$$

Weighted linear least squares estimator $\hat{\underline{a}}$ of $\underline{a}$ minimizes

$$\mathrm{S}SE(\underline{a}) = (\underline{y} - X\underline{a})^T \; W \; (\underline{y} - X\underline{a})$$

where W is a symmetric positive definite $n \times n$ matrix

Solution to WLLS problem:

Step 1. Identify vector space containing $\underline{y}$: $\mathcal{H} = \mathbb{R}^n$

Inner product: $< \underline{y}, \underline{z} >= \underline{y}^T \; W\underline{z}$

Step 2. Identify solution subspace $\mathcal{S}$

$$X\underline{a} = span\{\text{columns of X}\}$$

Step 3. apply projection theorem

Orthogonality Condition: the best linear estimator $\hat{\underline{a}}$ satisfies

$$0 = (\underline{y} - X\hat{\underline{a}})^T W X$$

$$= \underline{y}^T W \underline{X} - \hat{\underline{a}}^T \; X^T W X$$

or, if X has full column rank p then $X^T W X$ is invertible and

$$\hat{\underline{a}} = [X^T W X]^{-1} X^T W \underline{y}$$

5.7.1 PROJECTION OPERATOR FORM OF WLS PREDICTOR

The vector $\hat{\underline{y}}$ of LS predictors $\hat{y}_i = \underline{x}_i^T \hat{\underline{a}}$ of the actual output $\underline{y}$ is

$$\hat{\underline{y}} = X\hat{\underline{a}}$$

which can be represented as the "oblique" projection of $\underline{y}$ onto $\mathcal{H}$

$$\underline{\hat{y}} = \underbrace{X[X^T W X]^{-1} X^T W}_{\text{oblique projection } \Pi_{X,W}} \; \underline{y}$$

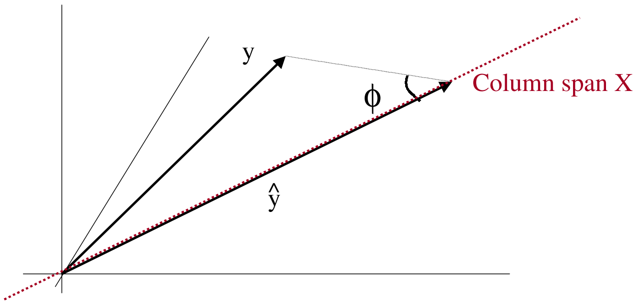


Figure 45: *Oblique projection interpretation of WLS estimator*

Resultant weighted sum of square error:

$$\begin{aligned} & \mathrm{W}SSE_{\min} \\ = \; & \underline{y}^T \, [I - X[X^T W X]^{-1} X^T W][I - X[X^T W X]^{-1} X^T W]^T \; \underline{y} \\ = \; & \underline{y}^T [I - \Pi_{X,W}]^T [I - \Pi_{X,W}] \underline{y} \end{aligned}$$

ALTERNATIVE INTERPRETATION: WLS predictor as LS predictor with preprocessing and postprocessing:

As W is positive definite we have square root factorization

$$W = W^{\frac{1}{2}} W^{\frac{1}{2}}$$

and

$$\begin{aligned} \underline{\hat{y}} &= W^{-\frac{1}{2}} \underbrace{W^{\frac{1}{2}} X [X^T W^{\frac{1}{2}} \, W^{\frac{1}{2}} X]^{-1} X^T W^{\frac{1}{2}}}_{\text{orthog. projector } \Pi_{W^{\frac{1}{2}} X}} [W^{\frac{1}{2}} \underline{y}] \\ &= W^{-\frac{1}{2}} \, \Pi_{X W^{\frac{1}{2}}} \, W^{\frac{1}{2}} \underline{y} \end{aligned}$$

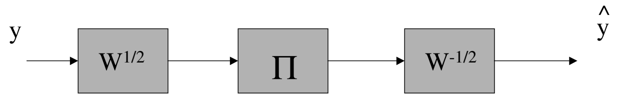


Figure 46: *Interpretation of WLS estimator as pre- and postprocessing with orthogonal projection*

Example 21 *Adaptive Linear Prediction*

Now want to fit AR(p) model

$$z_k = \sum_{i=1}^{p} a_i z_{k-i} + v_k, \quad k = 1, 2, \ldots$$

such that at time n we minimize WLS criterion

$$\mathrm{W}SSE(n) = \sum_{k=1}^{n} \rho^{n-k} (z_{k+p} - \sum_{i=1}^{p} a_i z_{k+p-i})^2$$

$\rho \in [0, 1]$ is an exponential forgetting factor

Solution of WLS problem:

As before, identify response variables $y_k = z_k$ and input vectors $\underline{x}_k = [z_{k-1}, \ldots, z_{k-p}]^T$.

Also identify weight matrix

$$W = \begin{bmatrix} \rho^0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \rho^{n-1} \end{bmatrix}$$

In this way we obtain WLLS predictor coefficients as

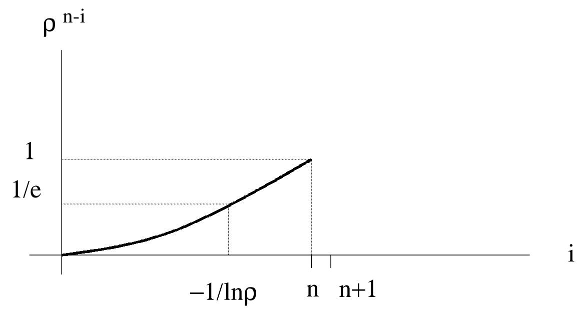


Figure 47: *Exponential forgetting factor applied to past errors for adaptive prediction*

$$\begin{aligned}\underline{\hat{a}} &= [X^T W X]^{-1} X^T W \underline{y} \\ &= \hat{\tilde{R}}^{-1} \underline{\hat{\tilde{r}}}\end{aligned}$$

and we have defined the smoothed sample correlation quantities:

$$\hat{\tilde{r}} = [\hat{\tilde{r}}_1, \ldots, \hat{\tilde{r}}_p]^T$$

$$\hat{\tilde{R}} = ((\hat{\tilde{r}}(i-j)))_{i,j=1,p}$$

$$\hat{\tilde{r}}_j := \sum_{i=1}^{n} \rho^{n-i} z_{i+p} z_{i+p-j}, \quad j = 0, \ldots, p$$

Minimum WSSE is:

$$\mathrm{W}SSE_{\min} = \hat{\tilde{r}}_0 - \hat{\tilde{r}}^T \hat{\tilde{R}}^{-1} \hat{\tilde{r}}$$

5.8 OPTIMALITY OF WLLS IN THE GAUSSIAN MODEL

Model

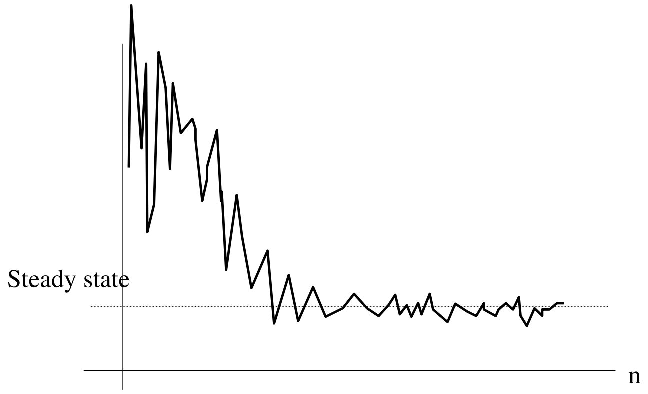


Figure 48: *Typical trajectory of the error criterion for predicting a stationary AR(1) process*

$$\underline{y} = X\ \underline{a} + \underline{v}$$

* $\underline{v} \sim \mathcal{N}_n(0, R)$
* covariance matrix R is known
* X is known

Density function of $\underline{y}$

$$f(\underline{y}; \underline{a}) = \frac{1}{\sqrt{(2\pi)^n |R|}}\ \exp\left(-\frac{1}{2\sigma^2}(\underline{y} - X\underline{a})^T R^{-1}(\underline{y} - X\underline{a})\right)$$

ML estimator of $\underline{a}$ is identical to WLLS estimator

$$\begin{aligned} \hat{\underline{a}}_{ml} &= \text{argmax}_{\underline{a}} \ln f(\underline{y}; \underline{a}) \\ &= \text{argmin}_{\underline{a}} (\underline{y} - X\underline{a})^T R^{-1}(\underline{y} - X\underline{a}) \end{aligned}$$

so that

$$\hat{y} = X\hat{\underline{a}}_{ml} \quad = \quad X[X^T R^{-1} X]^{-1} R^{-1} X^T \underline{y} = \Pi_{X,W}\ \underline{y}$$

PERFORMANCE: look at condition for equality in CRB

$$
\begin{aligned}
(\nabla_{\underline{a}} \ln f)^T &= \left(\underline{y} - X\underline{a}\right)^T R^{-1} X \, \frac{1}{\sigma^2} \\
&= \left(\underbrace{\underline{y}^T R^{-1} X [X^T R^{-1} X]^{-1}}_{\hat{\underline{a}}^T} - \underline{a}^T \right) \underbrace{X^T R^{-1} X \, \sigma^{-2}}_{K_{\underline{a}}}
\end{aligned}
$$

Hence, When noise covariance R is known and equals W^{-1}, conclude

* the WLLS estimator $\hat{\underline{a}}$ is unbiased

* the WLLS estimator is efficient and therefore UMVUE

* recalling property 5 of CRB in Section 4.5.1, as $K_{\underline{a}}$ is not a function of $\underline{a}$ the estimator covariance is

$$
\mathrm{cov}_{\underline{a}}(\hat{\underline{a}}) = K_{\underline{a}}^{-1} = [X^T R^{-1} X]^{-1} \, \sigma^2 = \hat{\hat{R}}^{-1} \, \frac{\sigma^2}{n}
$$

5.9 Exercises

1. Give a concrete example of two zero mean random variables X and Z for which the linear minimum mean square error estimator of X given Z is equal to 0. Give another concrete example where the overall (linear or non-linear) minimum mean square error estimator is 0.
2. Several problems on min MSE estimation will be added.
3. The least squares method can also be applied to multiple-response linear observation models of the form

$$
y_{1k} + \alpha_1 y_{2k} \ldots + \alpha_{p-1} y_{pk} = \beta_1 x_{1k} + \ldots + \beta_q x_{qk} + v_k, \quad k = 1, \ldots, n
$$

where $\{y_{1k}, \ldots, y_{pk}\}_k$ are n different observed waveforms (responses) and the α_i. and β_i coefficients are to be determined by minimizing the least squares criterion

$$
\mathrm{SSE}(\underline{\alpha}, \underline{\beta}) = \sum_{k=1}^{n} (y_k + \alpha_1 y_{2k} + \ldots + \alpha_{p-1} y_{pk} - \beta_1 x_{1k} - \ldots - \beta_q x_{qk})^2
$$

(a) Show that the above observation model is equivalent to the vector model

$$
Y[1, \underline{\alpha}^T]^T = X\underline{\beta} + \underline{v}
$$

where Y and X are $n \times p$ and $n \times q$ matrices, respectively, $\underline{v}$ is a $n \times 1$ vector of residuals v_k, $\underline{\alpha} = [\alpha_1, \ldots, \alpha_{p-1}]^T$ and $\underline{\beta} = [\beta_1, \ldots, \beta_q]^T$.

(b) Assuming that X has linearly independent columns (full rank) find the least squares estimates $\hat{\underline{\beta}}$ and $\hat{\underline{\alpha}}$ and the resulting minimum SSE. (Hint: first minimize over β then over α)

(c) Assume that the vector $\underline{u} = [1, \underline{\alpha}]^T$ is constrained to have length $\|\underline{u}\| = c$, where $c \geq 1$ is a specified constant. Derive an explicit form for the LS estimators. (Hint: Rayleigh theorem (Ch. 2 or [5]) on minimizing quadratic forms).

(d) The "sensor selection problem" is the following. Fix $p^{'}$ and consider choosing the subset of $p^{'}$ response waveforms $\{y_{i_1,k}\}_{k=1}^n, \ldots, \{y_{i_{p'},k}\}_{k=1}^n$, $i_1, \ldots, i_{p'}$ distinct integres in $1 \ldots, p$, out of the p responses which provide the best fit, i.e. minimize the residuals. Show that the algorithm for solving this sensor selection problem requires solving $n!/(p^{'}!(n-p^{'}))!$ separate least squares problems.

(e) The optimal sensor selection algorithm which you obtained in the previous part of this exercise is of high computational complexity, in the worst case it requires solving approximately $n^{p^{'}}$ least squares problems. Comment on how the solutions to parts (b) or (c) of this exercise could be used to approximate the optimal solution.

4. Let the observation have the standard linear model $y_k = \underline{x}_k^T \underline{a} + v_k, \quad k = 1, \ldots, n$. We saw in this chapter that when y_k and $\underline{x}_k$ are known and v_k is Gaussian the MLE of $\underline{a}$ is equivalent to the WLSE with weight matrix equal to the covariance matrix of the vector $\underline{v} = [v_1, \ldots, v_n]^T$. In many applications there exist outliers, i.e. a small number of unusually large residual errors v_k, and the Gaussian assumption is not appropriate. Here we treat the case of heavy-tailed distributions of v_k which leads to an estimate of $\underline{a}$ which is more robust to outliers.

(a) Assume that v_k are i.i.d. r.v.s with marginal density $f_v(v)$. Show that the MLE of $\underline{a}$ is

$$\hat{\underline{a}} = \text{argmin}_{\underline{a}} \left\{ \sum_{k=1}^n \log f_v(y_k - \underline{x}_k^T \underline{a}) \right\}$$

(b) Assuming f_v is a smooth function, derive the CR bound on unbiased estimators of $\underline{a}$. Under what conditions is the bound attainable?

(c) Show that for Laplacian noise with $f_v(v) = \frac{\beta}{2} \exp(-\beta|v|)$, $\beta \geq 0$, the MLE reduces to the minimizer of the sum of the absolute errors $|y_k - \underline{x}_k^T \underline{a}|$.

(d) Consider the noise density $f_v(v) = c(\alpha, b) \exp(-v^2/(\alpha^2 + v^2))$, $v \in [-b, b]$, b and α fixed known parameters and c a normalizing constant. Show that the MLE $\hat{\underline{a}}$ can be interpreted as a non-linearly weighted LSE in the sense that it satisfies the "orthogonality condition"

$$\sum_{k=1}^n \lambda_k(\hat{\underline{a}})(y_k - \underline{x}_k^T \hat{\underline{a}})\underline{x}_k = 0$$

where

$$\lambda_k(\hat{\underline{a}}) = \frac{1}{\alpha^2 + (y_k - \underline{x}_k^T \hat{\underline{a}})^2}$$

(e) The solution to the non-linearly weighted LSE above can be approximated using an "iterative reweighted least squares" technique which consists of approximating the above "orthogonality condition" by implementing the following procedure

i. Initialize $\hat{\underline{a}}_0 = \hat{\underline{a}}$ equal to the standard unweighted LS estimate $\hat{\underline{a}} = [X^T X] X^T \underline{y}$.

ii. Repeat until convergence:

$$\hat{\underline{a}}_{i+1} = [X^T W_i X]^{-1} X^T W_i \underline{y}, i = 1, 2, \ldots$$

where W_i is a diagonal weight matrix with diagonal entries $\lambda_1(\hat{\underline{a}}_i), \ldots, \lambda_n(\hat{\underline{a}}_i)$.

Implement this algorithm in MATLAB and study its convergence for various values of α, b.

5. In many applications involving fitting a model to a set of input and output measurements X $(n \times p)$ and y $(n \times 1)$, not only are the output measurements noisy but the input measurements may also be noisy. In this case the method of Total Least Squares (TLS) [5] is applicable. One formulation of TLS is to model the measurements by

$$y_k = (\underline{x}_k + \underline{\epsilon}_k)^T \underline{a} + v_k, \quad k = 1, \ldots, n$$

where v_k is a zero mean white Gaussian noise with variance σ_v^2 and $\underline{\epsilon}_k$ is an i.i.d. sequence of zero mean Gaussian $p \times 1$ random vectors with diagonal covariance matrix $\sigma_\epsilon^2 I_p$.

(a) Find the likelihood equation which must be satisfied by the MLE of $\underline{a}$ when σ_v and σ_ϵ are known. To what does your equation reduce when σ_v^2 dominates σ_ϵ^2?.

(b) Show that the MLE of $\underline{a}$ is identical to the standard LS estimator for unknown σ_ϵ.

(c) Find the CR bound on unbiased estimator covariance for the case of known σ_v and σ_ϵ. Repeat for the case of unknown σ_ϵ.

6. It is desired to find the linear least sum of squares (LLSS) fit of a complex valued vector $\underline{a}$ to the model

$$y_k = \underline{x}_k^T \underline{a} + v_k, \quad k = 1, \ldots, n$$

where y_k and $\underline{x}_k = [x_{k1}, \ldots, x_{kp}]^T$ are observed. Defining the vector space $\mathcal{H}$ of complex valued n-dimensional vectors with norm $< \underline{y}, \underline{z} >= \underline{y}^H \underline{z}$ ("H" denotes complex conjugate transpose) and vector $\underline{y} = [y_1, \ldots, y_n]^T$ and matrix $X = [\underline{x}_1, \ldots, \underline{x}_n]^T$ (analogously to the case studied in sec. 5.6 of notes). Assume that X is full rank. Using the projection theorem show that the solution to the LLSE problem $\min_{\underline{a}} \|\underline{y} - X\underline{a}\|^2$ is of the form

$$\hat{\underline{a}} = [X^H X]^{-1} X^H \underline{y}$$

with minimum LLSS residual error squared

$$\|\underline{y} - X\hat{\underline{a}}\|^2 = \underline{y}^H [I - X[X^H X]^{-1} X^H] \underline{y}.$$

7. This problem applies the solution to the previous exercise. Let the complex observations be given as $X = \{X(0), \ldots, X(N-1)\}$. Hypothesize that $X(k)$ is a damped sinusoid in additive noise:

$$X(k) = a e^{-\alpha k} e^{j2\pi f_0 k} + Z(k), \quad k \geq 0,$$

where $a \in \mathbb{C}$ is an unknown complex scale factor, $\alpha \geq 0$ is an unknown decay constant, and $f_0 \in [0, \frac{1}{2}]$ is an unknown frequency.

(a) For known α and f_0 show that the least-squares estimator of a which minimizes the sum of the squared residuals $SSE(\underline{a}) = \sum_{k=0}^{N-1} |X(k) - a e^{-\alpha k} e^{j2\pi f_0 k}|^2$ over a has the form (large N):

$$\hat{a} = \mathcal{X}(z_0)\,(1 - e^{-2\alpha}),$$

where $\mathcal{X}(z_0) = \sum_{k=0}^{N-1} X(k) z_0^{-k}$ is the Z-transform of X evaluated at the point $z = z_0 = e^{\alpha + j2\pi f_0}$ outside the unit circle. Note that for $\alpha = 0$ this is just the DFT of $X(k)$.

(b) Now for known α but unknown a show that the (non-linear) least-squares estimator for f_0 which minimizes the sum of the squared residuals $SSE(\hat{a}) = \sum_{k=0}^{N-1} |X(k) - \hat{a}e^{-\alpha k}e^{j2\pi f_0 k}|^2$ over f_o is obtained by maximizing the Z-transform of X over the radius e^α circle $|z| = e^\alpha$:

$$\hat{f}_0 = \text{argmax}_{f_0} |\mathcal{X}(z_0)|^2,$$

and:

$$\hat{a} = \mathcal{X}(e^{\alpha + j2\pi \hat{f}_0})\ (1 - e^{-2\alpha}).$$

Note that $\hat{f}_0$ reduces to the location of the highest peak in the magnitude "frequency spectrum" $S(f) = \left|\mathcal{X}(e^{2\pi f})\right|$ of $X(k)$ when α is known to be equal to 0.

(c) Finally for unknown a, α, f_0 show that the non-linear least-squares estimator of α, f_0 is obtained by maximizing the scaled Z-transform of X over the exterior of the unit disk:

$$\hat{f}_0, \hat{\alpha} = \text{argmax}_{f_0, \alpha \geq 0} |\mathcal{X}(z_0)|^2\ (1 - e^{-2\alpha}),$$

and:

$$\hat{a} = \mathcal{X}(e^{\hat{\alpha} + j2\pi \hat{f}_0})\ (1 - e^{-2\hat{\alpha}}).$$

8. It is desired to fit the coefficients α and β to the linear model for the measurements $y_k = \alpha + \beta k + v_k$, $k = 1, \ldots, N$, where v_k is the model error residual to be minimized by suitable chice of α, β. Find the linear least squares estimator for these coeffcents (you can leave your solution in the form of a pair of simultaneous equations if you wish).

9. It is hypothesized that the relation between a pair of measured variables y_k and x_k is non-linear. A reasonable model for this is

$$y_k = a_0 + a_1 x_k + \ldots + a_p x_k^p + v_k, \quad k = 1, \ldots, n$$

where v_k is an error residual.

(a) Find the set of coefficients $a_0, \ldots, a_p$ which minimizes the mean squared error $E[(y - [a_0 + a_1 x + \ldots + a_p x^p])^2]$ under the assumption that y and x are r.v.'s with known moments $E[yx^l]$, $l = 0, \ldots, p$, and $E[x^l]$, $l = 0, \ldots, 2p$.

(b) Repeat part (a) for the non-statistical least squares estimation error criterion for $n \geq p$ samples $\{x_k, y_k\}_{k=1}^n$: $\sum_{k=1}^n (y_k - [a_0 + a_1 x_k + \ldots + a_p x^p])^2$.

(c) Under the assumption of i.i.d. samples show that the the estimators found in (b) becomes equivalent to that found in (a) as $n \to \infty$.

End of chapter

6 OPTIMAL LINEAR FILTERING AND PREDICTION

* Wiener-Hopf Equations of min MSE filtering for w.s.s. processes

* Non-causal filtering, estimation, and prediction

* Causal LTI prewhitening: spectral factorization

* Causal LTI prediction: the Wiener filter

* Causal TV prewhitening: the innovations filter

* Causal TV prediction: the Kalman-Bucy filter

REFERENCES

Kailath [14]

Haykin [9]

Thomas [35]

6.1 WIENER-HOPF EQUATIONS OF OPTIMAL FILTERING

Two zero mean w.s.s. discrete time random processes

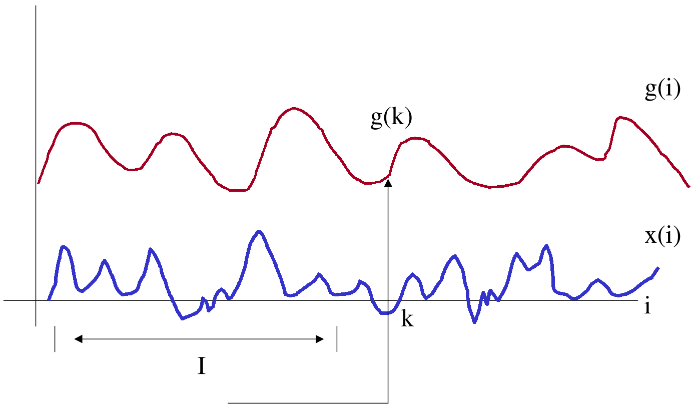


Figure 49: *Two random processes over time. Objective is to predict g_k as a linear function of observation x_i over a time interval $\mathcal{I}$.*

$x = \{x_k : k \in \mathcal{I}\}$: observed over an index set $\mathcal{I}$

$g = \{g_k : -\infty < k < \infty\}$: unobserved

Objective: estimate a time sample of g as linear function of x

$$\hat{g}_k = \sum_{j \in \mathcal{I}} h(k, j) x_j$$

such that we achieve minimum MSE

$$\mathrm{MSE}(\hat{g}_k) = E[|g_k - \hat{g}_k|^2]$$

Solution: by orthogonality condition

$$E[(g_k - \hat{g}_k)u_i^*] = 0, \quad i \in I$$

for any basis set $\{u_i\}$ spanning $span\{x_i : i \in \mathcal{I}\}$.
Basis $u_i = x_i$, $i \in \mathcal{I}$, will suffice here.
Obtain Wiener-Hopf (WH) equation for optimal filter $h(k,j)$ from orthog. condition

$$\begin{aligned} 0 &= E[(g_k - \hat{g}_k)x_i^*] \\ &= E[g_k x_i^*] - \sum_{j\in\mathcal{I}} h(k,j)E[x_j x_i^*] \\ &= r_{gx}(k-i) - \sum_{j\in\mathcal{I}} h(k,j) r_x(j-i), \quad i \in \mathcal{I} \end{aligned}$$

When $\mathcal{I}$ is finite this is equivalent to a matrix equation which can be solved as in previous section.
Two cases of interest here:
Case I: $\mathcal{I} = \{\infty, \ldots, \infty\}$: non-causal estimation (smoothing)
Case II: $\mathcal{I} = \{\infty, \ldots, k\}$: Causal estimation (filtering)

6.2 NON-CAUSAL ESTIMATION

For $\mathcal{I} = \{\infty, \ldots, \infty\}$ WH equation becomes

$$r_{gx}(k-i) - \sum_{j=-\infty}^{\infty} h(k,j) r_x(j-i) = 0, \quad -\infty < i < \infty$$

Fact: solution $h(k,j)$ is LTI filter $h(k-j)$
Take Z-transform to obtain:

$$\mathcal{P}_{gx}(z) - H(z)\mathcal{P}_x(z) = 0$$

or optimum filter has frequency domain transfer function

$$H(e^{j\omega}) = \frac{\mathcal{P}_{gx}(e^{j\omega})}{\mathcal{P}_x(e^{j\omega})}$$

Minimum MSE is

$$\begin{aligned}\text{MSE}_{\min} &= E[(g_k - h_k * x_k)g_k^*] \\ &= r_g(0) - h_k * r_{xg}(k)|_{k=0}\end{aligned}$$

Or in frequency domain (recall notation $\mathcal{P}(\omega)$ for $\mathcal{P}(e^{j\omega})$)

$$\begin{aligned}\text{MSE}_{\min} &= \int_{-\infty}^{\infty} [\mathcal{P}_g(\omega) - H(\omega)\mathcal{P}_{xg}(\omega)]\, d\omega \\ &= \int_{-\infty}^{\infty} \left[\mathcal{P}_g(\omega) - \frac{\mathcal{P}_{gx}(\omega)\mathcal{P}_{xg}(\omega)}{\mathcal{P}_x(\omega)}\right] d\omega\end{aligned}$$

Example 22 *Signal in additive noise*

Measurement model

$$x_k = s_k + w_k$$

* s, w are uncorrelated
* $g_k = s_k$ to be estimated

Obtain min MSE filter for estimation of s_k and its min MSE

$$H = \frac{\mathcal{P}_s}{\mathcal{P}_s + \mathcal{P}_w} = \begin{cases} 1, & \mathcal{P}_s/\mathcal{P}_w \gg 1 \\ 0, & \mathcal{P}_s/\mathcal{P}_w \ll 1 \end{cases}$$

$$\text{MSE}_{\min} = \int_{-\infty}^{\infty} \frac{\mathcal{P}_s\mathcal{P}_w}{\mathcal{P}_s + \mathcal{P}_w} d\omega = \begin{cases} 0, & \mathcal{P}_s/\mathcal{P}_w \gg 1 \\ \text{var}(s), & \mathcal{P}_s/\mathcal{P}_w \ll 1 \end{cases}$$

Problem: $H(\omega)$ is non-negative real so that h_k is symmetric impulse response. This implies that $\hat{s}_k = \sum_{j=-\infty}^{\infty} h(j)x_{k-j}$ depends on future measurements (non-causal)!

6.3 CAUSAL ESTIMATION

Objective: find linear min MSE estimate of g_k based only on past measurements

$$\hat{g}_k = \sum_{j=-\infty}^{k} h(k-j)x_j$$

where h satsifies Wiener-Hopf equations

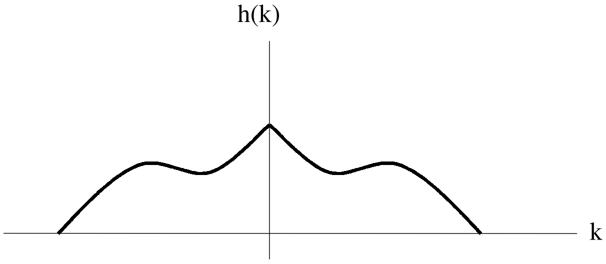


Figure 50: *Min MSE filter for estimating w.s.s. signal in noise has symmetric impulse response.*

$$0 = r_{gx}(k-i) - \sum_{j=-\infty}^{k} h(k-j) r_x(j-i), \quad -\infty < i \leq k \tag{10}$$

Let's explicitly constrain filter h_j to be causal: $h(j) = 0, j < 0$. Then we have after change of variable (see homework exercises)

$$0 = r_{gx}(l) - \sum_{j=-\infty}^{\infty} h(l-j) r_x(j), \quad l \geq 0 \tag{11}$$

Difficulty: cannot just take z-transform as we did before due to restriction on l. Indeed the equation (11) *does not specify* the value of the difference on the LHS for negative values of l; these values can be *arbitrary* as long as RHS = 0 for $l \geq 0$.

6.3.1 SPECIAL CASE OF WHITE NOISE MEASUREMENTS

One case where solution to WH is simple: $x_i = w_i$ = white noise of unit variance:

$$r_w(k) = \delta_k = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

Wiener-Hopf Eqn becomes

$$r_{gw}(l) - \sum_{j=-\infty}^{\infty} h(l-j) r_w(j) = r_{gw}(l) - h(l), \quad l \geq 0$$

Hence we can specify optimal causal filter as

$$h(k) = \begin{cases} r_{gw}(k), & k \geq 0 \\ 0, & o.w. \end{cases}$$

Or in z-transform domain:

$$H(z) = \{\mathcal{P}_{gw}(z)\}_+ \tag{12}$$

where we have defined truncated z-transform of a time function $b(k)$ with z-transform $B(z)$

$$\{B(z)\}_+ := \sum_{k=0}^{\infty} b(k) z^{-k}$$

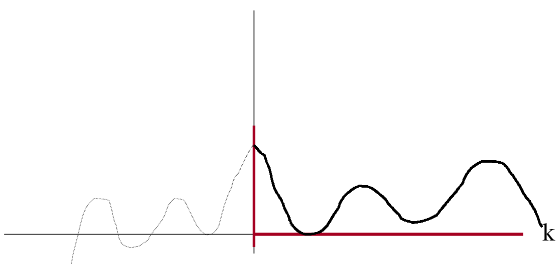


Figure 51: *Truncated z-transform of a function* b_k.

6.3.2 GENERAL CASE OF NON-WHITE MEASUREMENTS

Main idea: if we could "prewhiten" x with a filter h_w then we could follow with optimal filter of form (12).

This suggests a "prewhitening approach" to solving general problem. However, not just any old whitening filter will do. In keeping with the original objective of causal linear minimum mean square error estimation

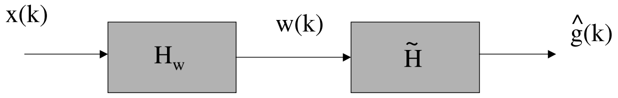


Figure 52: *Solution of the causal estimation problem by a cascade of a prewhitening filter h_w and an optimal wiener filter $\tilde{h}$ for white noise.*

we the prewhitening filter must itself be causal *and*, to ensure that we lose no information about $\{x_i\}_{i=-\infty}^{k}$, it must be causally invertible, i.e. we must be able to recover past values of the input $\{x_i\}_{i=-\infty}^{k}$ from past values of the output.

Thus the optimal filter $\tilde{H}$ for whitened measurements is specified by

$$\begin{aligned}\tilde{H}(z) &= \{\mathcal{P}_{gw}(z)\}_+ \\ &= \left\{\mathcal{P}_{gx}(z)H_w(z^{-1})\right\}_+\end{aligned}$$

The filter H_w must satisfy conditions

1. h_w whitens the input process

$$\mathcal{P}_w(z) = H_w(z)H_w(z^{-1})\mathcal{P}_x(z) = 1$$

2. h_w is causal

3. h_w is causally invertible

Definition: A filter (discrete time) h_w with transfer function $H_w(z)$ is causal and causally invertible iff $H_w(z)$ and $1/H_w(z)$ have no singularities outside the unit circle, i.e.,

$$|H_w(z)| < \infty, \quad |z| > 1$$

and similarly,

$$1/|H_w(z)| < \infty, \quad |z| > 1$$

.

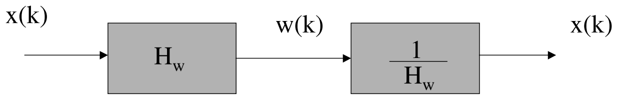


Figure 53: *The cascade of causal filters H_w and $1/H_w$ is causal and all pass thus implying that $\{w_i\}_{i=-\infty}^{k}$ contains same information as $\{x_i\}_{i=-\infty}^{k}$, i.e. w_i is white sufficient statistic.*

6.4 CAUSAL PREWHITENING VIA SPECTRAL FACTORIZATION

Assume x_k is w.s.s. with

* $r_x(k)$ positive definite and summable $\left(\sum_{k=-\infty}^{\infty} |r_x(k)| < \infty\right)$

* rational PSD

$$\mathcal{P}_x(z) = \sum_k r_x(k) z^{-k} = \frac{b(z)}{a(z)}$$

where

$r_x(k) = E[x_i x_{i-k}]$ is acf of x

$b(z) = b_q z^q + \cdots + b_1 z + b_0$

$a(z) = a_p z^p + \cdots + a_1 z + a_0$

For h_w to satisfy whitening condition require

$$H_w(z) H_w(z^{-1}) = \frac{1}{\mathcal{P}_x(z)} \tag{13}$$

Next we deal with causality conditions

Important properties of rational PSD $\mathcal{P}_x(z)$:

Pole zero constellationof rational PSD

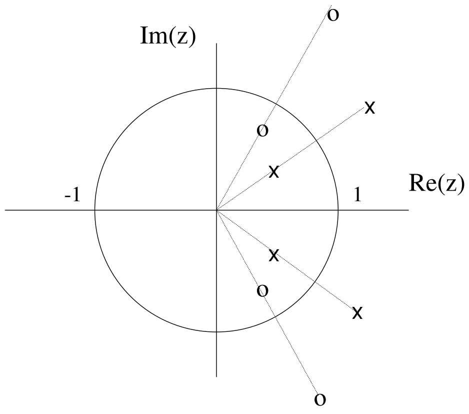


Figure 54: *Pole zero constellation of a rational PSD.*

$\mathcal{P}_x(z^{-1}) =$	$\mathcal{P}_x(z)$,	(symmetric r_k)
$\mathcal{P}_x(z^*) =$	$\mathcal{P}^*_x(z)$,	(real r_k)
$\mathcal{P}_x(z) \rightarrow$	no poles on unit circle	(bounded r_k)
$\mathcal{P}_x(z) \rightarrow$	zeros on unit circle occur in pairs	(p.d. r_k)

These conditions imply that there exist positive square root factors $\mathcal{P}^+_x(z)$ and $\mathcal{P}^-_x(z)$ which satisfy:

$$\mathcal{P}_x(z) = \mathcal{P}^+_x(z)\mathcal{P}^-_x(z)$$

$$\mathcal{P}^+_x(z^{-1}) = \mathcal{P}^-_x(z)$$

and

$\mathcal{P}^+_x(z)$ has all poles and zeros inside unit circle

$\mathcal{P}^-_x(z)$ has all poles and zeros outside unit circle

Therefore, conclude that the assignment

$$H_w(z) = 1/\mathcal{P}^+_x(z)$$

satisfies whitening condition (13) and $H(z)$, $1/H(z)$ have all their poles inside unit circle

Can identify $1/H_w(z)$ causal synthesis filter for measurement process x_k

$$x_k = h_w^{-1}(k) * w_k$$

where $h_w^{-1}(k)$ is the inverse Z-transform of $1/H_w(z) = \mathcal{P}_x^+(z)$.

This can be useful for simulation of x_k with arbitrary rational PSD from pseudo-random white noise samples w_k.

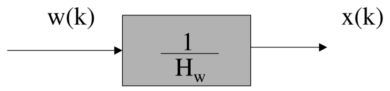


Figure 55: *A representation of x_k with PSD $\mathcal{P}_x$ as output of LTI causal filter $H_w(z) = \mathcal{P}_x^+$ driven by white noise*

6.5 CAUSAL WIENER FILTERING

Putting H_w and $\tilde{H}$ together we obtain formula for the causal Weiner filter

$$\begin{aligned} H(z) &= H_w(z)\,\tilde{H}(z) = \frac{1}{\mathcal{P}_x^+(z)}\left\{\mathcal{P}_{gx}(z)H_w(z^{-1})\right\}_+ \\ &= \frac{1}{\mathcal{P}_x^+(z)}\left\{\frac{\mathcal{P}_{gx}(z)}{\mathcal{P}_x^-(z)}\right\}_+ \end{aligned} \tag{14}$$

Example 23 *Causal prewhitening an AR(1) noise process*

$$x(k) = -ax(k-1) + u(k), \quad (-1 < a < 1)$$

where $u(k)$ is white noise with variance 1.

Find PSD

$$\mathcal{P}_x(z) = \underbrace{\frac{1}{(1+az^{-1})}}_{\mathcal{P}_x^+(z)}\underbrace{\frac{1}{(1+az)}}_{\mathcal{P}_x^-(z)}, \qquad r_x(k) = \frac{a^k}{1-a^2}$$

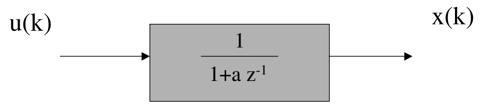


Figure 56: *Synthesis filter of AR(1) process is a single pole IIR.*

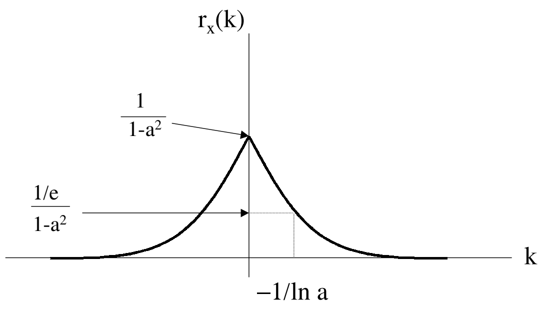


Figure 57: *Auto-correlation function of AR(1) process with AR coefficient a is slowly decaying double sided exponential for $-1 \ll a \leq 0$. (figure k-axis label should be $-1/\ln|a|$).*

The causal prewhitening filter is FIR

$$H_w(z) = 1 + az^{-1} \iff h_w(k) = \delta(k) + a\delta(k-1)$$

Can be implemented even without access to infinite past

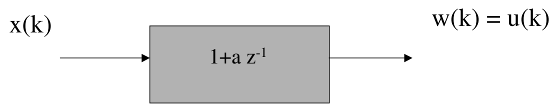


Figure 58: *Causal prewhitening filter for AR(1) process with AR coeffient a is a single tap FIR filter.*

Example (ctd.) *Prediction of an AR(1) from noiseless observations*

Now we let $g_k = x_{k+\alpha}$ where α is a positive integer. When $\alpha > 0$ this is a prediction problem. In light of the fact that we have just found the prewhitening filter, it remains to find the quantity

$$\left\{\frac{\mathcal{P}_{gx}(z)}{\mathcal{P}_x^-(z)}\right\}_+$$

to specify the Wiener filter-predictor (14).

As $r_{gx}(k) = E[x_{i+\alpha}x_{i-k}] = r_x(k+\alpha)$: $\mathcal{P}_{gx}(z) = z^\alpha \mathcal{P}_x(z)$. Hence

$$\begin{aligned}\left\{\frac{\mathcal{P}_{gx}(z)}{\mathcal{P}_x^-(z)}\right\}_+ &= \left\{z^\alpha \mathcal{P}_x^+(z)\right\}_+ \\ &= \left\{z^\alpha/(1+az^{-1})\right\}_+\end{aligned}$$

Now, using the identity (32) derived in the exercises, we see that

$$\left\{z^\alpha/(1+az^{-1})\right\}_+ = (-a)^\alpha/(1+az^{-1})$$

Hence, the Wiener filter-predictor is simply

$$H(z) = \frac{1}{\mathcal{P}_x^+(z)}\left(\frac{(-a)^\alpha}{1+az^{-1}}\right) = (-a)^\alpha$$

which in the time domain gives the optimal predictor as $\hat{x}_{k+\alpha} = (-a)^\alpha x_k$. This just corresponds to scaling the most recent observation and is consistent with x_k being a 1st order Markov sequence so that, for predicting the future, past information is not useful given present information.

Example 24 *Causally prewhitening an AR(1) plus white noise process*

$$x(k) = v_{\mathrm{A}R(1)}(k) + V(k)$$

where

* $v_{\mathrm{A}R(1)}(k)$ is AR(1) with $a = -0.8$
* $V(k)$ is white noise of variance $1/0.36$
* $v_{\mathrm{A}R(1)}(k)$ and $V(k)$ are uncorrelated

Using result (31) derived in the Exercises, find PSD as a rational function with double pole and double zero

$$\begin{aligned}\mathcal{P}_x(z) &= \frac{1}{(1-0.8z^{-1})}\frac{1}{(1-0.8z)} + 1/0.36 \\ &= \underbrace{\frac{d(1+bz^{-1})}{(1+az^{-1})}}_{\mathcal{P}_x^+(z)} \underbrace{\frac{d(1+bz)}{(1+az)}}_{\mathcal{P}_x^-(z)}\end{aligned}$$

where $a = -0.8, b = -0.5$ and $d = 1/\sqrt{0.225}$.

Unlike previous example, the causal prewhitening filter $h_w(k)$ is now IIR

$$H_w(z) = 1/d\frac{(1+az^{-1})}{(1+bz^{-1})}$$

and thus prewhitening cannot be implemented without access to infinite past.

Note that the synthesis filter $1/H_w(z)$ can be applied to white noise w_k to obtain recursion for x_k with both an autoregressive (AR) component (LHS) and a moving average (MA) component (RHS):

$$x_k + ax_{k-1} = b_1 w_k + b_2 w_{k-1}$$

where $b_1 = d$ and $b_2 = db$. Random processes x_k that satisfy the recursion above are "ARMA(1,1)" process.

Example (ctd.) *Prediction of AR(1) from noisy observations*

Similarly to the previous example we let $\hat{g}(k) = v_{\mathrm{A}R(1)}(k+\alpha)$ where α is a non-negative integer. As the measurement noise $u(k)$ and the AR(1) process $v_{\mathrm{A}R(1)}(k)$ are uncorrelated $\mathcal{P}_{gx}(z) = \mathcal{P}_{gv}(z) = z^\alpha \mathcal{P}_v(z)$ where $\mathcal{P}_v(z)$ is the PSD of $v_{\mathrm{A}R(1)}$ and $\mathcal{P}_{gv}(z)$ is the cross spectral density of g and $v_{\mathrm{A}R(1)}$. Therefore, after substitution of the expression for $\mathcal{P}_x^-$ obtained above,

$$\left\{\frac{\mathcal{P}_{gx}(z)}{\mathcal{P}_x^-(z)}\right\}_+ = \frac{1}{d}\left\{z^\alpha \frac{1}{1+bz}\frac{1}{1+az^{-1}}\right\}_+. \qquad (15)$$

Before proceeding further, we will need to express the product of two ratios in $\{\cdot\}_+$ as a sum of two ratios in order to apply the identities (32) and (33). To do this, observe

$$\begin{aligned}\frac{1}{1+bz}\frac{1}{1+az^{-1}} &= \frac{z^{-1}}{b+z^{-1}}\frac{1}{1+az^{-1}}\\ &= \frac{A}{b+z^{-1}}+\frac{B}{1+az^{-1}}\end{aligned}$$

where A and B are to be determined. Comparing the LHS of the top line to the bottom line of this last equation it is obvious that

$$\begin{aligned}A &= \lim_{z^{-1}\to -b}\frac{z^{-1}}{1+az^{-1}} = -b/(1-ab)\\ B &= \lim_{z^{-1}\to -1/a}\frac{z^{-1}}{b+z^{-1}} = 1/(1-ab)\end{aligned}$$

Thus we have from (15)

$$\begin{aligned}\left\{\frac{\mathcal{P}_{gx}(z)}{\mathcal{P}_x^-(z)}\right\}_+ &= \frac{1}{d(1-ab)}\left\{\frac{z^\alpha}{1+az^{-1}}-\frac{z^{\alpha+1}b}{1+bz}\right\}_+\\ &= \frac{1}{d(1-ab)}\frac{(-a)^\alpha}{1+az^{-1}}\end{aligned}$$

where we have used the identity (33) which shows that only the first additive term in $\{\cdot\}_+$ survives (the second term corresponds to an anticausal component).

Hence, using (14) the Wiener filter-predictor is simply

$$H(z) = \frac{q}{1+bz^{-1}}$$

where $q = \frac{(-a)^\alpha}{d^2(1-ab)}$, which can be implemented in the time domain as the single pole IIR filter recursion

$$\hat{g}(k) = -b\hat{g}(k-1) + qx_k.$$

with $\hat{g}(k) = \hat{v}_{AR(1)}(k+\alpha)$. It can be readily verified that in the limit as the measurement noise var$(u(k))$ goes to zero, $b \to 0$, $d^2 \to 1$, and $q \to (-a)^\alpha$ so that this IIR predictor filter reduces to the simple Wiener predictor filter of the previous example having no measurement noise.

6.6 CAUSAL TIME VARYING ESTIMATION

The Wiener filter is limited to the cases where the processes g_k and x_k are jointly w.s.s. and the estimating filter is LTI, i.e. access to infinite past is available. In practice, however, this is not the case and we will need to handle the situation for which

1. $g_k x_k$ may not be jointly w.s.s.

2. estimator filter is turned on at time $k = 0$ with initial conditions and is not LTI.

Objective: find linear min MSE estimate of g_k based only on *finite* past measurements

$$\hat{g}_k = \sum_{j=0}^{k} h(k,j) x_j$$

We know that optimal h satisfies the $k \times k$ system of Wiener-Hopf equations.

$$0 = r_{gx}(k,i) - \sum_{j=0}^{k} h(k,j) r_x(j,i), \quad 0 \leq i \leq k$$

Or, since summation is over finite number of indices we can express this in the familiar matrix form

$$\underline{h}_k = R_x^{-1} \underline{r}_{xg}$$

where $\underline{h}_k = [h(k,k), h(k,k-1), \ldots, h(k,1)]^T$, R_x is the $k \times k$ covariance matrix of the measurements, and $\underline{r}_{gx}$ is the k-element vector of cross correlations.

Difficulty: standard matrix inverse approach has growing memory and computation as k increases: not suitable for real time.

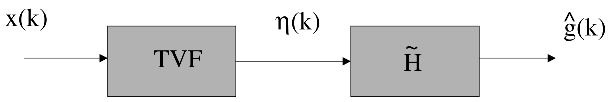


Figure 59: *Decomposition of min MSE filter by prefiltering with time-varying innovations filter.*

6.6.1 SPECIAL CASE OF UNCORRELATED MEASUREMENTS

As before we first convert to a case where solution to WH is simple:

$\Rightarrow x_i = \eta_i$ = non-stationary white noise

$$r_\eta(j,i) = \sigma_\eta^2(i)\ \delta_{j-i}$$

Solution to WH equation is now immediate

$$0 = r_{g\eta}(k,i) - \sum_{j=0}^{k} h(k,j) r_\eta(j,i) = r_{g\eta}(k,i) - h(k,i)\sigma^2_\eta(i), \quad 0 \le i \le k$$

and gives optimal filter as a projection coefficient associeted with projecting g_k onto the i-th noise component η_i

$$\begin{aligned} h(k,i) &= \frac{< g_k, \eta_i >}{< \eta_i, \eta_i >} \\ &= r_{g\eta}(k,i)/\sigma^2_\eta(i), \quad 0 \le i \le k \end{aligned}$$

6.6.2 CORRELATED MEASUREMENTS: THE INNOVATIONS FILTER

Q. How to "prewhiten" x_k?

A. A time-varying "prewhitening filter" has to yield output variables $\{\eta_i\}$ which are uncorrelated, which are causally generated from past of $\{x_i\}$, and from which the past $\{x_i\}$ can be recovered

Required conditions on η_i:

1. cov$(\eta_i, \eta_j) = 0, i \neq j$

2. $span\{\eta_k, \eta_{k-1}, \ldots, \eta_1\} = span\{x_k, x_{k-1}, \ldots, x_1\}$, $k = 1, 2, \ldots$

Recursive construction of $\{\eta_i\}$:

Let $\hat{x}_{k|k-1}$ be the optimal 1-step linear predictor of x_k given past $\{x_{k-1}, \ldots, x_0\}$

$$\hat{x}_{k|k-1} = \sum_{i=0}^{k-1} a_{k,i}\, x_i$$

Equivalently,

$$\hat{x}_{k|k-1} = \sum_{i=0}^{k-1} \alpha_{k,i}\, \eta_i$$

Recall orthogonality condition

$$E[(x_k - \hat{x}_{k|k-1})x_i] = 0, \quad i = 0, \ldots, k-1$$

Suggests following algorithm for η_i's

$$\begin{aligned} \eta_0 &= x_0 \\ \eta_1 &= x_1 - \hat{x}_{1|0} \\ \vdots & \quad \vdots \\ \eta_k &= x_k - \hat{x}_{k|k-1} \end{aligned}$$

or, more explicitly, in matrix form

$$\begin{bmatrix} \eta_0 \\ \vdots \\ \eta_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -a_{10} & \ddots & 0 & \vdots \\ \vdots & \ddots & \ddots & 0 \\ -a_{k,0} & \cdots & -a_{k,k-1} & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_k \end{bmatrix}$$

$$\underline{\eta}_k = \mathcal{A}\, \underline{x}_k$$

* η_i is the ”innovations process”
* Rows of $\mathcal{A}$ specify causal invertible “innovations filter”
* Note: as $\mathcal{A}$ is invertible $\{\eta_i\}$ is a sufficient statistic

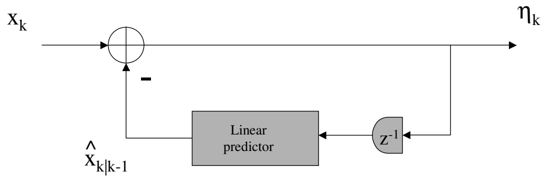


Figure 60: *The innovations filter produces equivalent uncorrelated measurement sequence η_i.*

6.6.3 INNOVATIONS AND CHOLESKY DECOMPOSITION

The innovations representation gives decomposition for covariance R_x of $\underline{x}$

$$\begin{aligned} R_x &= E[\underline{x}\underline{x}^T] \\ &= E[\mathcal{A}^{-1}\underline{\eta}\underline{\eta}^T\mathcal{A}^{-T}] \end{aligned}$$

$$= \mathcal{A}^{-1} E[\underline{\eta\eta}^T] \mathcal{A}^{-T}$$

$$= \mathcal{A}^{-1} \, R_\eta \, \mathcal{A}^{-T}$$

FORWARD CHOLESKY DECOMPOSITION

Any symmetric positive definite matrix B has a decomposition of the form

$$B = L_f \; P_f \; L_f^T$$

where

* P_f diagonal matrix of “forward prediction error variances”
* L_f is lower triangular matrix of “forward prediction coefficients”
* P_f and L_f are non-singular

BACKWARD CHOLESKY DECOMPOSITION

Any symmetric positive definite matrix B has a decomposition of the form

$$B = L_b^T \; P_b \; L_b$$

where

* P_b diagonal matrix of “backwards prediction error variances”
* L_b is lower triangular matrix of “backwards prediction coefficients”
* P_b and L_b are non-singular

6.7 TIME VARYING ESTIMATION/PREDICTION VIA THE KALMAN FILTER

Since $\{\eta_i\}$ is equivalent to $\{x_i\}$ we have equivalent representation for optimal estimator of g

$$\hat{g}_k = \hat{g}_{k|k} = \sum_{j=0}^{k} h_w(k,j)\eta_j$$

where h_w is projection coefficent

$$h_w(k,j) = \frac{< g_k, \eta_j >}{< \eta_j, \eta_j >} = \frac{r_{g\eta}(k,j)}{\sigma_\eta^2(j)}$$

We can now write a “pseudo recursion” for $\hat{g}_{k|k}$

$$\hat{g}_{k|k} = \underbrace{\sum_{j=0}^{k-1} h_w(k,j)\eta_j}_{\hat{g}_{k|k-1}} + h_w(k,k) \; \eta_k \tag{16}$$

This is not a "true recursion" since we do not yet know how to compute the update $h_w(k-1,j) \to h_w(k,j)$ for the projection coefficient nor the update $\eta_{k-1} \to \eta_k$ of the innovations.

To obtain true recursion $\hat{g}_{k|k-1} \to \hat{g}_{k+1|k}$ will need to assume x_k is given by the model

$$\begin{aligned} x_k &= s_k + v_k \\ s_k &= \underline{c}_k^T \underline{\xi}_k \\ \underline{\xi}_{k+1} &= A_{k+1,k}\, \underline{\xi}_k + B_k\, \underline{w}_k, \quad \underline{\xi}_0 = \underline{\xi}_o \end{aligned} \tag{17}$$

where:

s_k is a (scalar) signal and v_k is a (scalar) measurement noise,

$\underline{\xi}_k$ is a p dimensional "state vector" ("internal state" of system generating $\{\underline{s}_k\}$),

$\underline{c}_k$ is a p-element vector describing how the state affects the signal component s_k of the measurement x_k,

$\underline{w}_k$ is a state noise vector $(q \times 1)$

B_k is the state noise input matrix $(p \times q)$

$A_{k+1,k}$ is a state transition matrix $(p \times p)$ describing the signal dynamics which are due soley to the initial condition $\underline{\xi}_o$ (in the absence of driving noise $\underline{w}_k$).

We make the following simplifying statistical assumptions:

State Model (SM) Assumptions A1-A4

- A1 v_k: uncorrelated sequence with zero mean and variance $E[v_k^2] = \sigma_v^2(k)$
- A2 $\underline{w}_k$: uncorrelated sequence with zero mean and covariance matrix $E[\underline{w}_k \underline{w}_k^T] = R_w(k)$ $(q \times q)$
- A3 $\underline{\xi}_o$ has zero mean and covariance matrix $E[\underline{\xi}_o \underline{\xi}_o^T] = R_\xi(0)$ $(p \times p)$
- A4 $v_k, \underline{\xi}_0, \underline{w}_j$ are mutually uncorrelated for all j, k

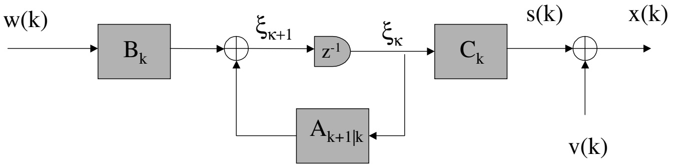


Figure 61: *State space model for observation.*

6.7.1 KALMAN FILTER FOR STATE ESTIMATION

Under the assumptions A1-A4 we obtain the following *Kalman filter recursions for state estimation* which generate the innovations η_k from $\{x_j\}_{j=0}^{k-1}$

Kalman Filter Recursions

$$\eta_k = x_k - \underline{c}_k^T \, \hat{\underline{\xi}}_{k|k-1} \tag{18}$$

$$\hat{\underline{\xi}}_{k+1|k} = A_{k+1,k} \, \hat{\underline{\xi}}_{k|k-1} + \Gamma_{k+1,k} \, \eta_k, \quad \hat{\underline{\xi}}_{0|-1} = 0 \tag{19}$$

where $\Gamma_{k+1,k}$ is the Kalman Gain, computed offline as function of state predictor error covariance

$$R_{\tilde{\xi}}(k|k-1) = E[(\underline{\xi}_k - \hat{\underline{\xi}}_{k|k-1})(\underline{\xi}_k - \hat{\underline{\xi}}_{k|k-1})^T]$$

$$\Gamma_{k+1,k} = A_{k+1,k} R_{\tilde{\xi}}(k|k-1) \underline{c}_k \, \frac{1}{\sigma_\eta^2(k)} \tag{20}$$

where $\sigma_\eta^2(k)$ is the innovations variance

$$\sigma_\eta^2(k) = \underline{c}_k^T R_{\tilde{\xi}}(k|k-1) \underline{c}_k + \sigma_v^2(k) \tag{21}$$

and the covariance matrix $R_{\tilde{\xi}}(k|k-1)$ is updated according to

$$\begin{aligned} R_{\tilde{\xi}}(k+1|k) &= [A_{k+1,k} - \Gamma_{k+1,k}\underline{c}_k^T] R_{\tilde{\xi}}(k|k-1) [A_{k+1,k} - \Gamma_{k+1,k}\underline{c}_k]^T \qquad (22) \\ &\quad + B_k R_w(k) B_k^T + \Gamma_{k+1,k}\Gamma_{k+1,k}^T \sigma_\eta^2(k). \\ R_{\tilde{\xi}}(0|-1) &= R_\xi(0) \qquad (23) \end{aligned}$$

6.7.2 KALMAN FILTER DERIVATION

Step 1: Gather Together Key Properties

The simplifying assumptions (A1-A4) imply the following properties

Properties P1-P3

P1 $\{\underline{\xi}_j\}_{j=1}^k$ is uncorrelated with $\underline{w}_k$ (A2,A4) and with v_k (A1,A4), and therefore

P2 $\{x_j\}_{j=1}^k$ is uncorrelated with $\underline{w}_k$ and v_{k+1}.

P3 $< x_k, \eta_j > = < s_k + v_k, \eta_j > = < s_k, \eta_j >$, $j < k$, since $< v_k, \eta_j > = 0$, $j < k$, as $\eta_j \in span\{x_j, x_{j-1}, \ldots, x_0\}$ and v_k is uncorrelated sequence (A1).

Step 2: establish relation between $\hat{x}_{k|k-1}$ and $\hat{\underline{\xi}}_{k|k-1}$

Putting P2 and P3 together we obtain from representation (16) for $g_k = s_k$:

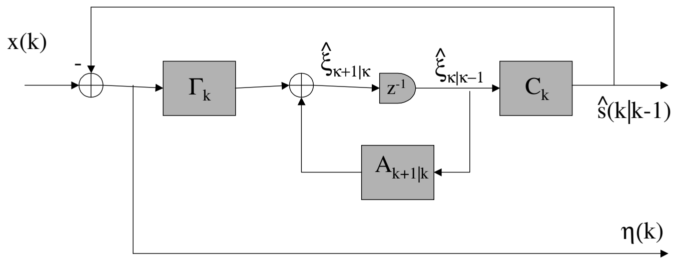


Figure 62: *Kalman filter block diagram for generation of* η_k *and* $\hat{s}(k)$

$$\begin{aligned}\hat{x}_{k|k-1} &= \sum_{j=0}^{k-1} \frac{< s_k, \eta_j >}{\|\eta_j\|^2}\eta_j \\ &= \underline{c}_k^T \underbrace{\sum_{j=0}^{k-1} \frac{< \underline{\xi}_k, \eta_j >}{\|\eta_j\|^2}\eta_j}_{\hat{\underline{\xi}}_{k|k-1}} .\end{aligned}$$

Step 3: Find expression for variance of innovation

Define the state estimator error vector

$$\tilde{\underline{\xi}}_{k|k-1} = \underline{\xi}_{k|k-1} - \hat{\underline{\xi}}_{k|k-1}.$$

Then, again using the state model for x_k,

$$\eta_k = x_k - \hat{x}_{k|k-1} = \underline{c}_k^T \underline{\xi}_k + v_k - \underline{c}_k^T \hat{\underline{\xi}}_{k|k-1} = \underline{c}_k^T \tilde{\underline{\xi}}_{k|k-1} + v_k \tag{24}$$

We can use this result to find the innovations variance $\|\eta_k\|^2 = \sigma_\eta^2(k)$ which is required for computing the projection coefficients $h_w(k,j) =< g_k, \eta_j > /\|\eta_j\|^2$ in (16). As $\tilde{\underline{\xi}}_{k|k-1} \in span\{\xi_k, x_{k-1}, \ldots, x_0\}$, $\tilde{\underline{\xi}}_{k|k-1}$ is uncorrelated with v_k, from (24)

$$\sigma_\eta^2(k) = \underline{c}_k^T R_{\tilde{\xi}}(k|k-1)\underline{c}_k + \sigma_v^2(k) \tag{25}$$

where $R_{\tilde{\xi}}(k|k-1)$ is the state estimator error covariance matrix. To evaluate this we will need to establish a recursion for this error covariance matrix. We first find recursions for the state estimator $\hat{\underline{\xi}}_{k|k-1}$.

Step 4: Establish update formula for $\hat{\underline{\xi}}_{k|k-1} \rightarrow \hat{\underline{\xi}}_{k+1|k}$

Recall that the linear minimum mean square estimator for a random vector is simply the concatenation of the linear minimum mean square estimators for each element of the random vector. Thus, with abuse of notation, denoting $< \underline{\xi}_{k+1}, \eta_k >$ the vector composed of inner products $< (\xi_{k+1})_i, \eta_k >, i = 1, \ldots, p$,

$$\begin{aligned} \hat{\underline{\xi}}_{k+1|k} &= \sum_{j=0}^{k} \frac{< \underline{\xi}_{k+1}, \eta_j >}{\|\eta_j\|^2} \eta_j \\ &= \sum_{j=0}^{k-1} \frac{< \underline{\xi}_{k+1}, \eta_j >}{\|\eta_j\|^2} \eta_j + \frac{< \underline{\xi}_{k+1}, \eta_k >}{\|\eta_k\|^2} \eta_k \end{aligned} \tag{26}$$

Define the *Kalman gain* vector

$$\Gamma_{k+1,k} = \frac{< \underline{\xi}_{k+1}, \eta_k >}{\|\eta_k\|^2}, \tag{27}$$

and note that, from the state equation (17) for $\underline{\xi}_k$ and the fact that $\underline{w}_k$ and η_j are uncorrelated for $j \leq k$,

$$\begin{aligned} \frac{< \underline{\xi}_{k+1}, \eta_j >}{\|\eta_j\|^2} &= \frac{< A_{k+1,k}\underline{\xi}_k, \eta_j >}{\|\eta_j\|^2} + \frac{< B_k \underline{w}_k, \eta_j >}{\|\eta_j\|^2} \\ &= A_{k+1,k} \frac{< \underline{\xi}_k, \eta_j >}{\|\eta_j\|^2}, \quad j \leq k \end{aligned}$$

which is $A_{k+1,k}$ times the projection coefficient for projecting $\underline{\xi}_k$ onto η_j.

Substitution of the above back into (26) gives the desired recursion

$$\begin{aligned} \hat{\underline{\xi}}_{k+1|k} &= A_{k+1,k} \sum_{j=0}^{k-1} \frac{< \underline{\xi}_k, \eta_j >}{\|\eta_j\|^2} \eta_j + \Gamma_{k+1,k}\eta_k \\ &= A_{k+1,k} \hat{\underline{\xi}}_{k|k-1} + \Gamma_{k+1,k}\eta_k \end{aligned} \tag{28}$$

with initial condition $\hat{\underline{\xi}}_{0|-1} = 0$.

Step 5: Express Kalman gain $\Gamma_{k+1,k}$ in terms of state estimator error covariance $R_{\tilde{\xi}}(k|k-1)$

The Kalman gain vector $\Gamma_{k+1,k}$ (27) can be related to the state estimator error covariance by the following steps

1. $< \underline{\xi}_{k+1}, \eta_k >= A_{k+1,k} < \underline{\xi}_k, \eta_k > + B_k < \underline{w}_k, \eta_k >= A_{k+1,k} < \underline{\xi}_k, \eta_k >$ (from whiteness of $\underline{w}_k$ and fact that $\eta_k \in span\{x_k, \ldots, x_0\}$).
2. $< \underline{\xi}_k, \eta_k >=< \underline{\xi}_k - \hat{\underline{\xi}}_{k|k-1}, \eta_k >=< \tilde{\underline{\xi}}_{k|k-1}, \eta_k >$ (noting that $< \hat{\underline{\xi}}_{k|k-1}, \eta_k >= 0$ from orthogonality principle of linear estimation)
3. $< \tilde{\underline{\xi}}_{k|k-1}, \eta_k >= E[\tilde{\underline{\xi}}_{k|k-1}\eta_k] = E[\tilde{\underline{\xi}}_{k|k-1}\tilde{\underline{\xi}}^T_{k|k-1}]\underline{c}_k$ ($\eta_k = \underline{c}_k^T \tilde{\underline{\xi}}_{k|k-1} + v_k$ and v_k is white noise uncorrelated with $\tilde{\underline{\xi}}_{k|k-1}$)

Putting the above together, and recalling that $\|\eta_k\|^2 = \sigma_\eta^2(k)$ calculated in (25), we obtain

$$
\begin{aligned}
\Gamma_{k+1,k} &= \frac{<\underline{\xi}_{k+1}, \eta_k>}{\|\eta_k\|^2} = A_{k+1,k}\frac{<\underline{\xi}_k, \eta_k>}{\|\eta_k\|^2} \\
&= A_{k+1,k} R_{\tilde{\xi}}(k|k-1)\underline{c}_k \, \frac{1}{\underline{c}_k^T R_{\tilde{\xi}}(k|k-1)\underline{c}_k + \sigma_v^2(k)} \qquad (29)
\end{aligned}
$$

Step 6: Find recursive update for estimator error covariance $R_{\tilde{\xi}}(k|k-1) \;\rightarrow\; R_{\tilde{\xi}}(k+1|k)$

First find update equation for $\tilde{\underline{\xi}}_k$ by subtracting the state estimator update equation (28) from the actual state update equation (17)

$$
\underline{\xi}_{k+1} - \hat{\underline{\xi}}_{k+1|k} = A_{k+1,k}(\underline{\xi}_k - \hat{\underline{\xi}}_{k|k-1}) + B_k \underline{w}_k - \Gamma_{k+1,k}\eta_k.
$$

Identifying $\tilde{\underline{\xi}}_{k+1} = \underline{\xi}_{k+1} - \hat{\underline{\xi}}_{k+1|k}$ and using $\eta_k = \underline{c}_k^T \tilde{\underline{\xi}}_k + v_k$ in the above

$$
\begin{aligned}
\tilde{\underline{\xi}}_{k+1|k} &= A_{k+1,k}\tilde{\underline{\xi}}_{k|k-1} + B_k \underline{w}_k - \Gamma_{k+1,k}(\underline{c}_k^T \tilde{\underline{\xi}}_k + v_k) \\
&= [A_{k+1,k} - \Gamma_{k+1,k}\underline{c}_k^T]\tilde{\underline{\xi}}_{k|k-1} + B_k \underline{w}_k - \Gamma_{k+1,k} v_k \qquad (30)
\end{aligned}
$$

Now properties P1-P3 imply that the three additive terms in (30) are mutually uncorrelated. Therefore, the covariance of the RHS is the sum of three covariance matrices and therefore we obtain (23)

$$
\begin{aligned}
R_{\tilde{\xi}}(k+1|k) &= [A_{k+1,k} - \Gamma_{k+1,k}\underline{c}_k^T] R_{\tilde{\xi}}(k|k-1)[A_{k+1,k} - \Gamma_{k+1,k}\underline{c}_k^T]^T \\
&\quad + B_k R_w(k) B_k^T + \Gamma_{k+1,k}\Gamma_{k+1,k}^T \sigma_v^2(k).
\end{aligned}
$$

Some simplification is possible by using (21) and (27) which makes $R_{\tilde{\xi}}(k|k-1)$ explicit in the quantity $\Gamma_{k+1,k}$, respectively. After some algebra this produces the equivalent update equation

$$
\begin{aligned}
R_{\tilde{\xi}}(k+1|k) &= A_{k+1,k} R_{\tilde{\xi}}(k|k-1) A_{k+1,k} + B_k R_w(k) B_k^T \\
&\quad - A_{k+1,k} R_{\tilde{\xi}}(k|k-1)\underline{c}_k^T \underline{c}_k R_{\tilde{\xi}}(k|k-1) A_{k+1,k}^T / \sigma_\eta^2(k)
\end{aligned}
$$

where $\sigma_\eta^2(k) = \underline{c}_k^T R_{\tilde{\xi}}(k|k-1)\underline{c}_k + \sigma_v^2(k)$, as defined above.

6.8 KALMAN FILTERING: SPECIAL CASES

The Kalman filter equation (19) generates the innovations sequence η_k which is needeed to compute the estimate $\hat{g}_{k|k}$ defined in Sec. 6.7 by the equation (16). Also needed are the projection coefficients $h_w(k,j)$, $j = 1, \ldots, k$. We discuss two special cases for which these coefficients are simply computed, Kalman prediction and Kalman filtering

6.8.1 KALMAN PREDICTION

The linear prediction problem is to predict future value of the observation x_{k+1} from a linear combination of past and present observations $\{x_j\}_{j=0}^k$, or, equivalently, from the past and present innovations $\{\eta_j\}_{j=0}^k$. Recalling the measurement model (17), $x_{k+1} = s_{k+1} + v_{k+1}$ is the sum of two uncorrelated components.

Hence, denoting the predictor by $\hat{x}_{k+1|k}$ and applying the superposition property (9) of linear estimators of a sum of random variables

$$\hat{x}_{k+1|k} = \hat{s}_{k+1|k} + \hat{v}_{k+1|k} = \hat{s}_{k+1|k}$$

where $\hat{v}_{k+1|k} = 0$ due to the fact that v_k is white and thus uncorrelated with the past innovations, i.e. unpredictable. Finally, as $s_{k+1} = \underline{c}_{k+1}^T \underline{\xi}_{k+1}$

$$\hat{s}_{k+1|k} = \underline{c}_{k+1}^T \hat{\underline{\xi}}_{k+1|k}$$

which can be computed from the Kalman filter (19) for state estimation discussed in the previous subsection.

6.8.2 KALMAN FILTERING

The filtering problem is to estimate the signal component s_k in $x_k = s_k + v_k$ from past and present measurements $\{x_j\}_{j=0}^k$ (equivalently $\{\eta_j\}_{j=0}^k$). Let $\hat{s}_{k|k}$ denote this estimate. Set $g_k = s_k$ and from the general recursion (16) we obtain

$$\hat{s}_{k|k} = \hat{s}_{k|k-1} + h_w(k,k)\eta_k,$$

where $\hat{s}_{k|k-1}$ is the linear predictor derived in the last subsection and

$$h_w(k,k) = \frac{E[s_k \eta_k]}{\text{var}(\eta_k)} = \frac{\underline{c}_k^T E[\underline{\xi}_k \eta_k]}{\text{var}(\eta_k)}.$$

Recall that in the process of showing the expression (29) for the Kalman gain $\Gamma_{k+1,k}$ we established $E[\underline{\xi}_k \eta_k] = R_{\tilde{\xi}}(k|k-1)\underline{c}_k$. Putting this together with the expression (21) we obtain

$$h_w(k,k) = \frac{\underline{c}_k^T R_{\tilde{\xi}}(k|k-1)\underline{c}_k}{\underline{c}_k^T R_{\tilde{\xi}}(k|k-1)\underline{c}_k + \sigma_v^2(k)}.$$

All of the above quantites are available from the Kalman filter recursions (18) and (23).

6.9 SUMMARY OF STATISTICAL PROPERTIES OF THE INNOVATIONS

As the observation noise v_k is uncorrelated with the signal s_k, we have three equivalent expressions for the innovations

$$\eta_k = x_k - \hat{x}_{k|k-1}$$

$$\eta_k = x_k - \hat{s}_{k|k-1}$$

$$\eta_k = \underline{c}_k^T(\underline{\xi}_k - \hat{\underline{\xi}}_{k|k-1}) + v_k$$

Furthermore:

$$E[\eta_i] = 0$$
$$\text{cov}(\eta_i, \eta_j) = 0, \quad i \neq j$$

and, as shown above, the innovations variance is

$$\begin{aligned}\text{var}(\eta_k) &= \sigma_\eta^2(k) \\ &= \underline{c}_k^T R_{\tilde{\xi}}(k)\underline{c}_k + \sigma_v^2(k)\end{aligned}$$

6.10 KALMAN FILTER FOR SPECIAL CASE OF GAUSSIAN STATE AND NOISE

Assume:

* v_k, $\underline{\xi}_o$ and $\underline{w}_k$ are jointly Gaussian

Then:

* $x_k = s_k + v_k$ is a Gaussian random process

* Kalman filter yields min MSE state predictor

$$\hat{\underline{\xi}}_{k|k-1} = E[\underline{\xi}_k | x_{k-1}, \ldots, x_1]$$

* $\{\eta_k\}$ is an equivalent uncorrelated Gaussian measurement

6.11 STEADY STATE KALMAN FILTER AND WIENER FILTER

Assume

* $A_{k+1,k}, b_k, c_k$ and $R_w(k)$ are time-invariant

* $R_v(k)$ is time-invariant

Then:

* s_k is w.s.s. as $k \to \infty$

* x_k is w.s.s. as $k \to \infty$

$\Rightarrow$ Steady state innovations filter is equivalent to Wiener prewhitening filter

In particular, in steady state, η_k becomes a (w.s.s.) white noise with

$$\sigma_\eta^2(k) \to \underline{c}^T R_{\tilde{\xi}}(\infty)\underline{c} + \sigma_v^2$$

6.12 Exercises

1. As we know, the optimal filter $h(k,j)$ for estimating the sample $g(k)$ of the process $\{g(k)\}_{i=-\infty}^{\infty}$ from zero mean process $\{x(i)\}_{i=-\infty}^{\infty}$ satisfies the Wiener-Hopf equation

$$r_{gx}(k,i) - \sum_{j=-\infty}^{\infty} h(k,j) r_x(i,j) = 0, \quad -\infty < i < \infty$$

Show that when g and x are jointly w.s.s. random processes, i.e. $r_x(i,j) = r_x(i-j)$ and $r_{gx}(k,i) = r_{gx}(k-i)$, $h(k,j)$ can be assumed to be linear time invariant (LTI), i.e. $h(k,j) = h(k-j)$ will satisfy the WH equation. Now show that the same holds for the causal optimal filter $h(k,j)$ which satsifies

$$r_{gx}(k,i) - \sum_{j=-\infty}^{\infty} h(k,j) r_x(i,j) = 0, \quad -\infty < i \leq k.$$

(Hint: Find Weiner-Hopf equations for estimating $y(k+l)$ where l is an arbitrary time shift, make a change of index i and a change of variable j and show by comparison to the Weiner-Hopf equations for estimating $y(k)$ that $h(k,j) = h(k-j,0)$).

2. Derive the equation (11) for the Causal Wiener filter $0 = r_{gx}(l) - \sum_{m=-\infty}^{\infty} h(l-m) r_x(m)$ from the original equation (10) by making two changes of variable in sequence: reindex i by $l = k - i$ and reindex j by $m = k - j$.

3. For constants a, c and the definitions $q = 1 + c + ca^2$, $r = ca$, derive the following identity

$$\frac{1}{(1+az^{-1})} \frac{1}{(1+az)} + c = \frac{d(1+bz^{-1})}{(1+az^{-1})} \frac{d(1+bz)}{(1+az)} \tag{31}$$

where

$$b = \frac{q/r \pm \sqrt{(q/r)^2 - 4}}{2}, \qquad d = q/(1+b^2)$$

Observe that when c is positive real, one of the roots in the equation for b satisfies $|b| \leq 1$ while the other satisfies $|b| \geq 1$.

4. In the development of the causal and non-causal Weiner estimators we have assumed that all processes were zero mean. Here we deal with the case of non-zero mean w.s.s. processes for which the *affine Weiner estimator* is appropriate.

   Assume the measurement process $\{x_k\}_k$ and the target process $\{g_k\}_k$ to be estimated are w.s.s. and have non-zero means $E[x_k] = \mu_x(k)$ and $E[g_k] = \mu_g(k)$.

   (a) The affine non-causal Wiener estimator of g_k is defined by the filter $h(k,j)$ and sequence of constants s_k as

$$\hat{g}_k = a_k + \sum_{j=-\infty}^{\infty} h(k,j) x_j$$

   where $h(k,j)$ and a_k are selected to minimize the mean square estimation error $\text{MSE}(h,a) = E[(g_k - \hat{g}_k)^2]$. Show that the optimal filter satisfies $h(k,j) = h(k-j)$ where $h(j)$ is the optimal linear time invariant non-causal Weiner filter for estimating the zero mean process $g_k - \mu_g(k)$ from the centered zero mean measurements $x_k - \mu_x(k)$ and that the optimal sequence a_k is $\mu_g(k) + \sum_{j=-\infty}^{\infty} h(k-j)\mu_x(j)$. Further show that the minimum MSE of the affine non-causal Wiener estimator is functionally independent of the means μ_g and μ_x.

(b) Repeat question (a) for the case of the causal affine Wiener estimator which satisfies the additional restriction that $h(k,j) = h(k-j) = 0$, $k < j$.

5. You have the measurement model

$$x_k = s_k^2 + w_k$$

where w_k is zero mean white noise of variance σ^2 and s_k is a w.s.s. zero mean Gaussian random sequence with acf $r_s(k) = a^k/(1-a^2)$, $k \geq 0$. w_k is independent of s_k.

(a) Find the quantities $E[s_i^2 s_{i-k}^2]$ and $E[s_i^2 s_{i-k}]$ (Hint: use the property that for any zero mean jointly Gaussian r.v.s W, X, Y, Z: $E[XYZ] = 0$ and $E[WXYZ] = E[WX]E[YZ] + E[WZ]E[XY] + E[WY]E[XZ]$)

(b) Using the results of (a) find the optimal affine non-causal Wiener estimator for s_k. Explain your result.

(c) Using the results of (a) find the optimal affine non-causal Wiener estimator for s_k^2.

6. Assume the measurement model

$$x_k = a s_k + v_k$$

where s_k and v_k are zero mean jointly w.s.s. and uncorrelated, and a is a random variable uncorrelated with s_k having mean μ_a and variance σ_a^2.

(a) Find the non-causal Weiner filter for estimating s_k.

(b) Find the MSE of the output of the non-causal Weiner filter. How does it behave as a function of μ_a and σ_a^2?

(c) Find the causal Weiner filter for estimating s_k.

7. For $|a| < 1$, use the geometric series formula $\sum_{k=0}^{\infty} c^k z^k = 1/(1-cz)$, $\quad |cz| < 1$ to derive the following two results:

$$\left\{\frac{z^l}{1+az^{-1}}\right\}_+ = \begin{cases} (-a)^l \frac{1}{1+az^{-1}}, & l \geq 0 \\ \frac{z^l}{1+az^{-1}}, & l < 0 \end{cases} \tag{32}$$

and

$$\left\{\frac{z^l}{1+az}\right\}_+ = \begin{cases} 1, & l = 0 \\ 0, & l > 0 \end{cases} \tag{33}$$

Now apply these results to compute the Z-domain quantity (for $l \geq 0$)

$$\left\{\frac{z^l}{(1+az^{-1})(1+bz)}\right\}_+ \qquad |a|, |b| < 1$$

.

8. Let the measurement model be

$$x_k = h_k * s_k + v_k$$

where s_k and v_k are zero mean jointly w.s.s. and uncorrelated, and h_k is a causal and causally invertible filter with Z-transform $H(z)$.

(a) Find the non-causal Weiner filter for estimating s_k.

(b) Find the causal Weiner filter for estimating s_k.

(c) Compare the results for (a) and (b) to the estimator $h_k^{-1} * \hat{s}_k$ where $\hat{s}_k$ is the output of the standard Weiner filter for estimating s_k using the measurement model $x_k = s_k + v_k$ and h_k^{-1} is the inverse Z-transform of $1/H(z)$.

9. Let the measurement model be as in Example 24

$$x_k = s_k + v_k$$

where s_k and v_k are zero mean jointly w.s.s. and uncorrelated. It is desired to estimate $g_k = h_k * s_k$ where h_k is the causal IIR filter with transfer function

$$H(z) = 1/(1 + \alpha z^{-1}), \quad \alpha \in (-1, 1)$$

.

(a) Find the non-causal Weiner filter for estimating g_k.
(b) Find the causal Weiner filter for estimating g_k.
(c) Compare the results for (a) and (b) to the estimator $\hat{g}_k = h_k * \hat{s}_k$ where $\hat{s}_k$ is alternatively the output of the standard non-causal and causal Weiner filters, respectively, for estimating s_k.

10. Let the measurement model be as in Example 24

$$x_k = s_k + v_k$$

where s_k and v_k are zero mean jointly w.s.s. and uncorrelated. It is desired to estimate $g_k = h_k * s_k$ where h_k is the causal FIR filter with transfer function

$$H(z) = 1 + \alpha z^{-1}, \quad \alpha \in (-1, 1)$$

.

(a) Find the non-causal Weiner filter for predicting g_{k+d} where $d > 0$.
(b) Find the causal Weiner filter for predicting g_{k+d} where $d > 0$.
(c) Compare the results for (a) and (b) to the estimator $\hat{g}_k = h_k * \hat{s}_k$ where $\hat{s}_k$ is alternatively the output of the standard non-causal and causal Weiner filters, respectively, for predicting s_{k+d}.

11. The process s_k is a zero mean AR(2) process following the recursion

$$s_k = 0.8 s_{k-1} - 0.15 s_{k-2} + w_k$$

where w_k is zero mean white noise of variance 1.5 uncorrelated with $s_{k-1}, s_{k-2}, \ldots$. The observation is

$$x_k = s_k + v_k$$

where v_k is zero mean white noise with variance 0.5 independent of s_k.

(a) Express the AR(2) recursion in Z-transform domain as $\mathcal{Z}\{s_k\} = H(z)\mathcal{Z}\{w_k\}$ and use the input/output PSD relation $\mathcal{P}_s(z) = H(z)H(z^{-1})\mathcal{P}_w(z)$ to determine the PSD $\mathcal{P}_s(z)$ of s_k.
(b) Find the non-causal Wiener filter for estimating s_k.
(c) Find the causal Wiener filter for estimating s_k.

12. A common multipath model for a communications receiver is that the direct path signal plus an attenuated and delayed indirect path version of the signal are received in additive white noise:

$$x_k = s_k + b s_{k-1} + w_k$$

The objective is to estimate the signal s_k given a set of measurements $\{x_k\}_k$. In the following assume that w_k is zero mean white with variance σ_w^2, s_k is zero mean white with variance σ_s^2, b is a constant $|b| < 1$, and s_k, w_k are uncorrelated. You can assume that $(\sigma_s^2(1+b^2)+\sigma_w^2)/(\sigma_s^2 b) = 5/2$ if that helps simplify your answers to the following.

 (a) Find the power spectral density (PSD) $\mathcal{P}_x(e^{j\omega})$ of x_k, the cross PSD $\mathcal{P}_{sx}(e^{j\omega})$ of s_k and x_k, and the spectral factorization of $\mathcal{P}_x(z)$, $z \in \mathbb{C}$.
 (b) Find the optimal non-causal Wiener filter for estimating s_k.
 (c) Find the optimal causal Wiener filter for estimating s_k.

13. In the derivation of the discrete time Kalman filter we assumed that the state noise $\underline{w}_k$ was uncorrelated with the measurement noise $v(k)$. In this problem we generalize the Kalman filter to the case where $E[\underline{w}_k v(l)] = V_{wv}\delta_{kl}$ where δ_{kl} is the kronecker delta function. Derive the Kalman filter equations.

14. The measurement equation is given by

$$x_k = s_k + v_k$$

where s_k satisfies the dynamic model ($|a| < 1$)

$$s_{k+1} = -a s_k + w_k, \quad s_0 = s_o$$

and v_k, w_k, s_o are uncorrelated, v_k and w_k are zero mean white noises with variances σ_v^2 and σ_w^2, respectively.

 (a) Derive the Kalman filter equations.
 (b) Derive the steady state state error covariance (variance) $R_{\tilde{s}}(\infty) \stackrel{\text{def}}{=} \lim_{k\to\infty} R_{\tilde{s}}(K|K-1)$ by setting $R_{\tilde{s}}(k+1|k) = R_{\tilde{s}}(k|k-1) = R_{\tilde{s}}(\infty)$ in the Kalman error covariance update formula and solving explicitly for $R_{\tilde{s}}(\infty)$. Find the corresponding steady state Kalman gain.
 (c) By taking the Z-transform of the steady state Kalman state recursion show that the Kalman predictor $\hat{s}_{k+1|k}$ is the output of a LTI with input x_k.
 (d) Compare the steady state Kalman predictor previously derived to the causal Wiener predictor based on the infinite past.

End of chapter

7 FUNDAMENTALS OF DETECTION

* Optimal detection theory

* Bayesian approach to detection

* Frequentist approach to detection

* Reciever Operating Characteristic (ROC) curves

* Classification theory

REFERENCES

Van Trees [36]

Scharf [29]

Mood, Graybill and Boes [19]

Bickel and Doksum [2]

Srinath, Rajasekaran and Viswanathan [33]

MORE ADVANCED REFERENCES

Poor [24]

Helstrom [10]

Lehmann [15]

Ferguson [4]

Example 25 *A motivating radar example*

$$x(t) = \theta s(t-\tau) + w(t), \quad 0 \leq t \leq T$$

* $s(t)$ is a known signal

* $w(t)$ is a zero mean Gaussian white noise with known level $N_o/2$

* $0 \leq \tau \ll T$, τ known

* $\int_0^T |s(t-\tau)|^2 dt = \int_0^T |s(t)|^2 dt$

* $\theta \in \{0, 1\}$ unknown

Detection objective: decide whether signal is present

Notation:

$$\begin{array}{lcl} H_0 : x(t) = w(t) & & H_0 : \theta = 0 \\ & \Leftrightarrow & \\ H_1 : x(t) = s(t-\tau) + w(t) & & H_1 : \theta = 1 \end{array}$$

Two methods of detection could be considered:

Energy-threshold detector:

$$y = \int_0^T |x(t)|^2 dt \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

Filter-threshold detector

$$y = \int_0^T h(T-t)x(t)dt \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

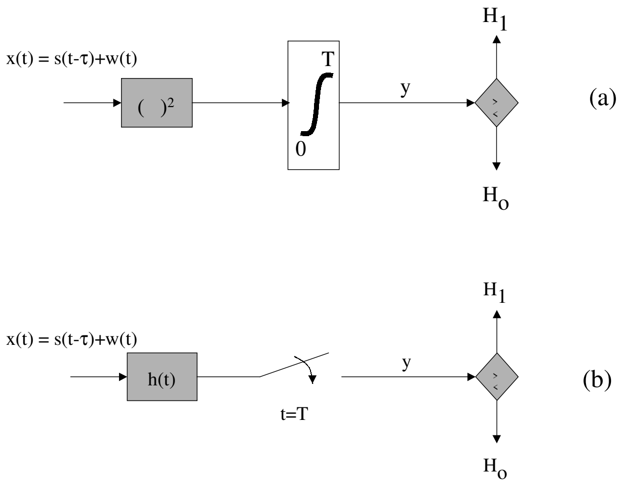


Figure 63: *(a) energy-threshold detector, (b) filter-threshold detector*

Filter-threshold detector design criterion:

Chose $h(\cdot)$ to maximize *deflection*

$$d^2 = \frac{|E[y|\text{signal}] - E[y|\text{no signal}]|^2}{\text{var}(y|\text{no signal})}$$

Find: signal produces shift in mean only

$$\begin{aligned}
E[y|\text{no signal}] &= 0 \\
E[y|\text{signal}] &= \int_0^T h(T-t)s(t-\tau)dt \\
\text{var}[y|\text{no signal}] &= N_o/2 \int_0^T |h(t)|^2 dt
\end{aligned}$$

Then by Cauchy-Schwarz inequality

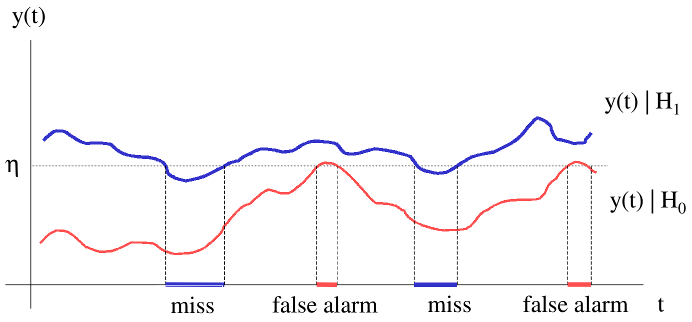


Figure 64: *Repeated tests of radar produce sequence of y_i's. One sequence has signal present (H_1) and one has signal absent (H_0).*

$$d^2 = \frac{2}{N_o} \frac{\left|\int_0^T h(T-t)s(t-\tau)dt\right|^2}{\int_0^T |h(T-t)|^2 dt} \leq \frac{2}{N_o} \underbrace{\int_0^T |s(t-\tau)|^2 \, dt}_{\int_0^T |s(t)|^2 = \|s\|^2}$$

with "=" if and only if $h(T-t) = as(t-\tau)$ for some constant a

$\Rightarrow$ obtain "matched filter" solution as optimal deflection filter

$$h(t) = s(T+\tau-t)$$

CASE of $s(\tau)$ = a short duration "pulse"

* $\int_0^T |s(t-\tau)|^2 dt$ does not depend on τ

* optimal detector can be implemented as:

$$\begin{aligned} y &= \int_0^T s(t-\tau)x(t)dt \\ &= \int_{-\infty}^{\infty} s(t-\tau)x(t)dt \\ &= s(-t) * x(t)|_{t=\tau} \end{aligned}$$

ERROR ANALYSIS:

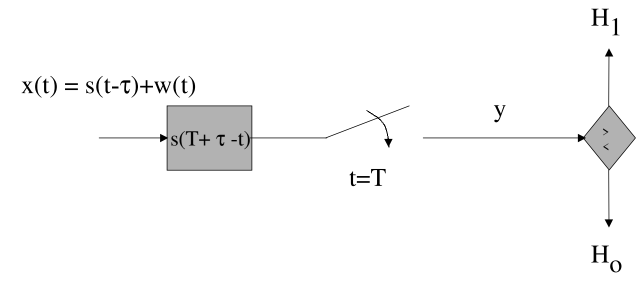


Figure 65: *SNR optimal receiver implemented as a matched filter reciever for delayed signal in noise.*

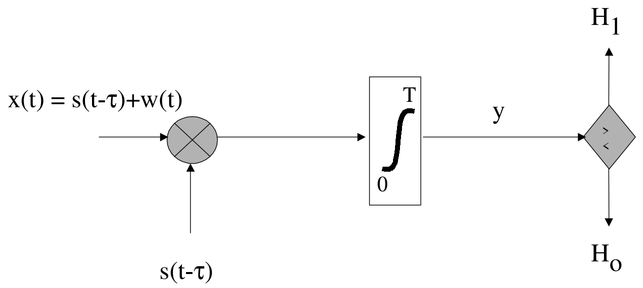


Figure 66: *SNR optimal receiver implemented as a correlator reciever for delayed signal in noise.*

Two kinds of decision error to be concerned about:

FALSE ALARM: $y > \eta$ when no signal

MISS: $y < \eta$ when signal

Compute the conditional probabilities of these errors:

$$P_F = P(\text{say signal}|\text{no signal}) = \int_{y>\eta} f(y|\text{no signal})dy$$

$$P_M = P(\text{say no signal}|\text{signal}) = \int_{y\leq\eta} f(y|\text{signal})dy$$

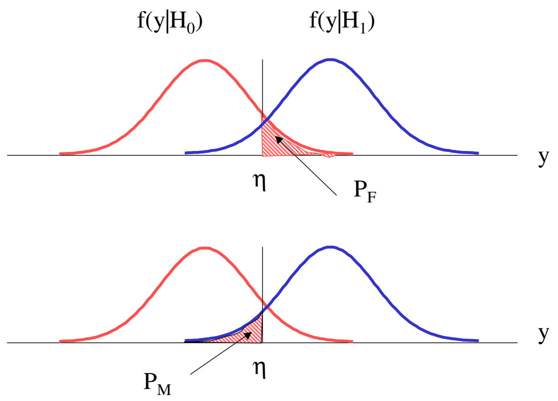


Figure 67: *Miss probability P_M and false alarm probability P_F for the radar example (h in figure should be η). Note that decreasing one type of error is accompanied by increase in other type of error.*

TWO QUESTIONS

Q: Is there an optimal way of trading off P_M for P_F?

Q: Does matched filter give optimal tradeoff?

7.1 THE GENERAL DETECTION PROBLEM

$x \in \mathcal{X}$ is a measurement or observation

$\theta \in \Theta$ are unknown parameters

$f(x;\theta)$ is p.d.f. of x (a known function)

Two distinct hypotheses on θ

$$\theta \in \Theta_0, \quad or \quad \theta \in \Theta_1$$

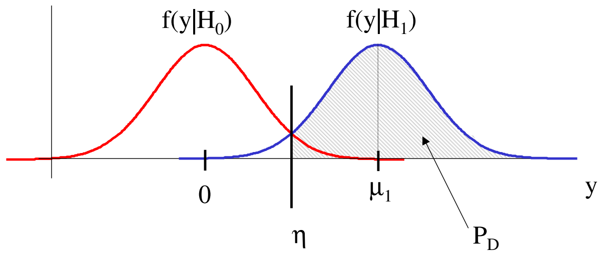


Figure 68: *Detection probability* $P_D = 1 - P_M$ *for the radar example.*

Θ_0, Θ_1 is partition of Θ into two disjoint regions

$$\Theta_0 \cup \Theta_1 = \Theta, \quad \Theta_0 \cap \Theta_1 = \{\text{empty}\}$$

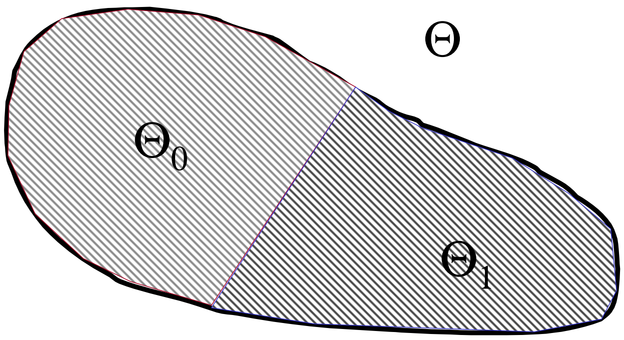


Figure 69: *The detector must decide on the region of* Θ *that contains the unknown parameter* θ.

NOTATION:

$$\begin{array}{lcl} H_0 : \theta \in \Theta_0 & & H_0 : x \sim f(x;\theta), \;\; \theta \in \Theta_0 \\ & \Leftrightarrow & \\ H_1 : \theta \in \Theta_1 & & H_1 : x \sim f(x;\theta), \;\; \theta \in \Theta_1 \end{array}$$

H_0: the null hypothesis, noise alone hypothesis

H_1: the alternative hypothesis, signal present hypothesis

7.1.1 SIMPLE VS COMPOSITE HYPOTHESES

When θ can take on only two values the hypotheses are said to be *simple*.

$\Theta = \{\theta_0, \theta_1\}, \quad \Theta_0 = \{\theta_0\}, \quad \Theta_1 = \{\theta_1\}$.

In this case p.d.f. $f(x;\theta)$ is known given H_0 and H_1

Otherwise hypotheses are said to be composite.

7.1.2 THE DECISION FUNCTION

Detection objective: design a decision rule (test function)

$$\phi = \phi(x) = \begin{cases} 1, & \text{decide } H_1 \\ 0, & \text{decide } H_0 \end{cases}$$

which maps $\mathcal{X}$ to decision space $\{0, 1\}$

Test function ϕ is equivalent to partition of $\mathcal{X}$ into decision regions

$$\mathcal{X}_0 = \{x : \phi(x) = 0\}, \quad \mathcal{X}_1 = \{x : \phi(x) = 1\}$$

FALSE ALARM AND MISS ERRORS

False alarm and miss probabilities associated with a test function ϕ are

$$P_F(\theta) = E_\theta[\phi], \quad \theta \in \Theta_0$$

$$P_M(\theta) = E_\theta[1 - \phi], \quad \theta \in \Theta_1$$

Or equivalently

$$\begin{aligned} P_F(\theta) &= \int_{\mathcal{X}_1} f(x|\theta)dx, \quad \theta \in \Theta_0 \\ P_M(\theta) &= 1 - \int_{\mathcal{X}_1} f(x|\theta)dx, \quad \theta \in \Theta_1 \end{aligned}$$

The probability of correctly deciding H_1 is called the (correct-) detection probability:

$$1 - P_M(\theta) = P_D(\theta) = E_\theta[\phi], \quad \theta \in \Theta_1$$

We give separate treatment for case of random and non-random θ

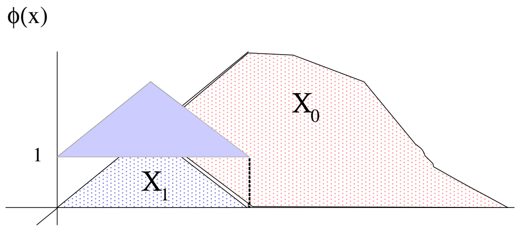


Figure 70: *Test function separates measurement space into two decision regions $\mathcal{X}_0$ and $\mathcal{X}_1$ (the region under the raised platform).*

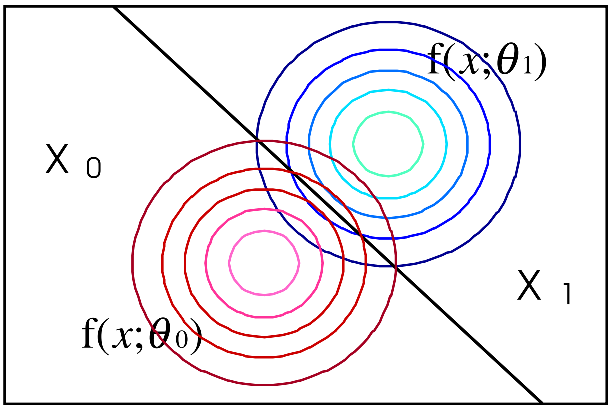


Figure 71: *Illustration of decision regions χ_0 and χ_1 for deciding H_0 and H_1 for an observation x in the plane. Also shown are constant contours of the H_0 and H_1 densities $f(x;\theta_0)$ $f(x;\theta_1)$. False alarm probability P_F is integral of $f(x;\theta_0)$ over χ_1, miss probability P_M is integral of $f(x;\theta_1)$ over χ_0, and detection probability P_D is integral of $f(x;\theta_1)$ over χ_1.*

7.2 BAYES APPROACH TO DETECTION

Three elements:

1. Assign a prior $f(\theta)$ density for θ

2. Assign a cost or risk to wrong decisions

* c_{ij} = cost of deciding H_i when H_j is true

3. Find and implement decision rule which has minimum average risk

7.2.1 PRIOR PROBABILITIES

Obtain prior probabilities on H_0, H_1

$$
\begin{aligned}
P(H_0) &= P(\theta \in \Theta_0) = \int_{\Theta_0} f(\theta) d\theta \\
P(H_1) &= P(\theta \in \Theta_1) = \int_{\Theta_1} f(\theta) d\theta
\end{aligned}
$$

with $P(H_0) + P(H_1) = 1$

In this case we can compute conditional p.d.f.'s given H_0 and H_1 by integrating over θ

$$
\begin{aligned}
f(x|H_0) &= \frac{\int_{\Theta_0} f(x|\theta) f(\theta) d\theta}{P(H_0)} \\
f(x|H_1) &= \frac{\int_{\Theta_1} f(x|\theta) f(\theta) d\theta}{P(H_1)}
\end{aligned}
$$

$\Rightarrow$ in Bayesian detection θ hypotheses are always simple!

7.2.2 MINIMIZE AVERAGE RISK

Cost or risk matrix:

$$
C = \begin{bmatrix} c_{11} & c_{10} \\ c_{01} & c_{00} \end{bmatrix}
$$

NOTE: usually $c_{ii} \leq c_{ij}$

Design Criterion: Select ϕ, equivalently $\mathcal{X}_0$ and $\mathcal{X}_1$, to minimize average risk

$$
\begin{aligned}
\overline{c} &= c_{11} P(\text{say } H_1 | H_1) P(H_1) + c_{00} P(\text{say } H_0 | H_0) P(H_0) \\
&\quad + c_{10} P(\text{say } H_1 | H_0) P(H_0) + c_{01} P(\text{say } H_0 | H_1) P(H_1)
\end{aligned}
$$

Express in equivalent form

$$\begin{aligned} \overline{c} &= c_{00}P(H_0) + c_{11}P(H_1) \\ &+[c_{01} - c_{11}]P(H_1)P_M + [c_{10} - c_{00}]P(H_0)P_F \end{aligned}$$

Observe: $\overline{c}$ linear in $P_M, P_F, P(H_1), P(H_0)$ for fixed decision rule ϕ.

7.2.3 OPTIMAL BAYES TEST MINIMIZES $\overline{c}$

First we note

$$\begin{aligned} P_F &= \int_{\mathcal{X}_1} f(x|H_0)dx \\ P_M &= 1 - \int_{\mathcal{X}_1} f(x|H_1)dx \end{aligned}$$

This allows us to rewrite $\overline{c}$ explicitly as function of decision region $\mathcal{X}_1$

$$\overline{c} = c_{00}P(H_0) + c_{01}P(H_1)$$

$$+\int_{\mathcal{X}_1} ([c_{10} - c_{00}]P(H_0)f(x|H_0) - [c_{01} - c_{11}]P(H_1)f(x|H_1)) \; dx$$

Solution is now obvious: assign x to $\mathcal{X}_1$ when

$$[c_{10} - c_{00}]P(H_0)f(x|H_0) < [c_{01} - c_{11}]P(H_1)f(x|H_1)$$

and assign x to $\mathcal{X}_0$ otherwise.
Optimal test is the Bayes likelihood ratio test (BLRT)

$$\Lambda_B(x) := \frac{f(x|H_1)}{f(x|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

where η is the optimal Bayes threshold

$$\eta = \frac{[c_{10} - c_{00}]P(H_0)}{[c_{01} - c_{11}]P(H_1)}$$

and Λ_B is called the BLR test statistic

7.2.4 MINIMUM PROBABILITY OF ERROR TEST

Special case: $c_{00} = c_{11} = 0$ and $c_{01} = c_{10} = 1$ gives the prob. error criterion

$$\overline{c} = P_M P(H_1) + P_F P(H_0) = P_e$$

which is minimized by the LR test

$$\frac{f(x|H_1)}{f(x|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)}{P(H_1)}$$

which, by Bayes rule, is equivalent to the MAP test

$$\frac{P(H_1|x)}{P(H_0|x)} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

"Maximum a posteriori test"

7.2.5 PERFORMANCE OF BLRT

Let minimum of $\overline{c}$ be denoted $\overline{c}^*$

$$\begin{aligned} \overline{c}^* &= c_{00}P(H_0) + c_{11}P(H_1) \\ &\quad + [c_{01} - c_{11}]P(H_1)P_M^*(\eta) + [c_{10} - c_{00}]P(H_0)P_F^*(\eta) \end{aligned}$$

where

$$P_F^*(\eta) = P(\Lambda_B > \eta|H_0), \qquad P_M^*(\eta) = P(\Lambda_B \leq \eta|H_1)$$

This describes a performance curve as a function of $p = P(H_0)$

7.2.6 MIN-MAX BAYES DETECTOR

BLRT requires $p = P(H_0)$ which may be unknown.
Minmax design: select a test to minimize worst case error

$$\overline{c}_{minmax} = \max_{p \in [0,1]} \overline{c}(p)$$

Min-max test is the optimal Bayes test having a threshold η^* which makes $\overline{c}$ a horizontal line
This specifies minmax optimality condition giving minimax values of $p = p^*$ and $\eta^* = \eta(p^*)$

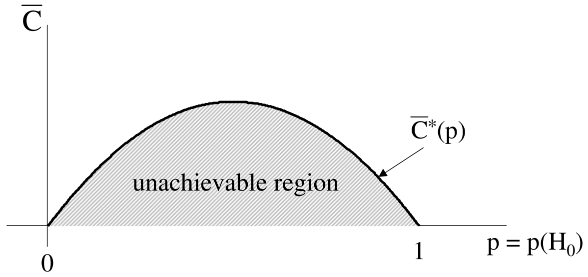


Figure 72: *The Bayes optimal test has a power curve which specifies an achievable lower bound on average risk.*

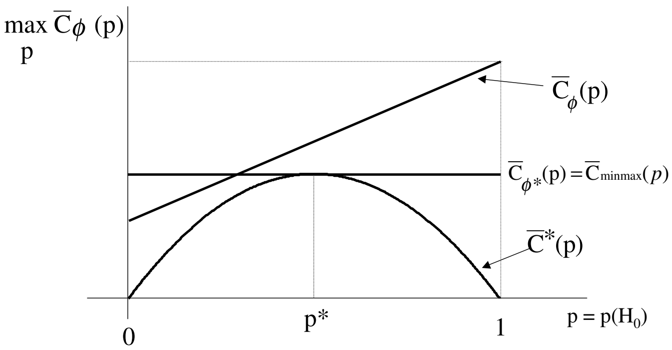


Figure 73: *Power curve of any fixed test ϕ is a straight line. The min-max optimal test ϕ^* has a horizontal power curve which is tangent to $\overline{c}^*(p)$ at its maximum.*

$$\begin{aligned}\overline{c} &= \underbrace{[c_{00}(1-P_F^*) + c_{10}P_F^* - c_{11}(1-P_M^*) - c_{01}P_M^*]}_{=0}\ p \\ &\quad + c_{11}(1-P_M^*) + c_{01}P_M^*\end{aligned}$$

Minmax probability error test: $\overline{c} = P_e$: $c_{00} = c_{11} = 0$, $c_{10} = c_{01} = 1$.

$$\overline{c} = \underbrace{[P_F^* - P_M^*]}_{=0}\ p + P_M^*$$

$\Rightarrow$ minmax prob. error test is BLRT with η selected to satisfy "equalization" condition

$$P_F^*(\eta) = P(\Lambda_B > \eta | H_0) = P(\Lambda_B \leq \eta | H_1) = P_M^*(\eta)$$

7.2.7 EXAMPLES

Example 26 *Radar example revisited*

Objective: Given matched filter output y find the Bayes optimal detector.

1. Assume that $P(H_0) = P(H_1) = \frac{1}{2}$
2. Recall that y is

$$y = \int_0^T s(t)x(t)dt$$

which is a Gaussian random variable having means and variances

$$\begin{array}{ll} E[y|H_0] = 0, & \text{var}[y|H_0] = N_o/2\ \int_0^T |s(t)|^2 dt = \sigma_0^2 \\ E[y|H_1] = \int_0^T |s(t)|^2 dt = \mu_1, & \text{var}[y|H_1] = N_o/2\ \int_0^T |s(t)|^2 dt = \sigma_0^2 \end{array}$$

Bayes LR test is

$$\begin{aligned}\Lambda_B &= \frac{\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(y-\mu_1)^2}}{\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}y^2}} \\ &= e^{y\mu_1/\sigma_0^2 - \frac{1}{2}\mu_1^2/\sigma_0^2} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta = 1\end{aligned}$$

LR test statistic is a monotone function of y since $\mu_1 > 0$.

Equivalent test is filter-threshold detector

$$y \overset{H_1}{\underset{H_0}{\gtrless}} \gamma = \tfrac{1}{2}\mu_1$$

Performance of Bayes LRT for radar example

$$\begin{aligned} P_F &= P(y > \gamma | H_0) \\ &= P(\underbrace{y/\sigma_0}_{\mathcal{N}(0,1)} > \gamma/\sigma_0 | H_0) \\ &= \int_{\gamma/\sigma_0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \\ &= 1 - \mathcal{N}(\gamma/\sigma_0) := Q(\gamma/\sigma_0) \end{aligned}$$

and

$$\begin{aligned} P_M &= P(y < \gamma | H_1) \\ &= P(\underbrace{(y-\mu_1)/\sigma_0}_{\mathcal{N}(0,1)} > (\gamma - \mu_1)/\sigma_0 | H_1) \\ &= \mathcal{N}((\gamma - \mu_1)/\sigma_0) \end{aligned}$$

Note: since standard Gaussian p.d.f. is symmetric

$\mathcal{N}(-u) = 1 - \mathcal{N}(u)$

and thus, since $\gamma = \mu_1/2$

$$P_M = P_F = 1 - \mathcal{N}(\mu_1/\sigma_0)$$

Conclude: the Bayes threshold γ is actually minmax!

Therefore the probability of error reduces to

$$\begin{aligned} P_e &= P_M \tfrac{1}{2} + P_F \tfrac{1}{2} \\ &= P_F \end{aligned}$$

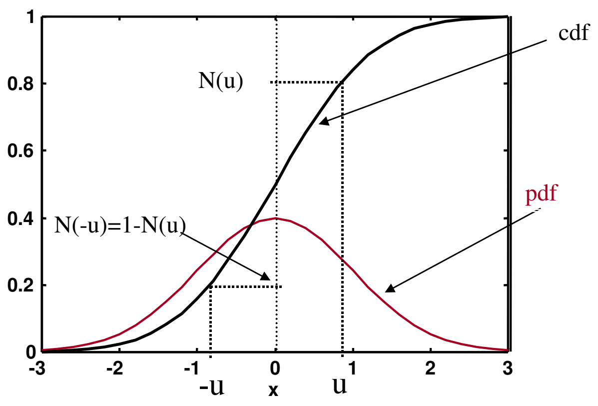


Figure 74: *CDF $\mathcal{N}(u)$ of symmetric Gaussian density*

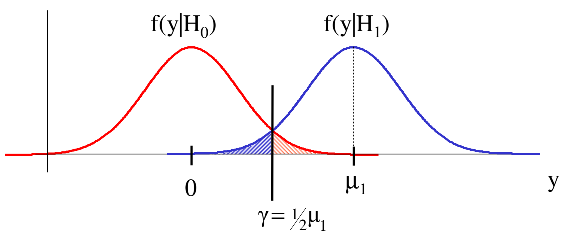


Figure 75: *Equally likely hypotheses have minmax threshold $\gamma = \frac{1}{2}\mu_1$ for problem of detection of shift in mean of a Gaussian r.v.*

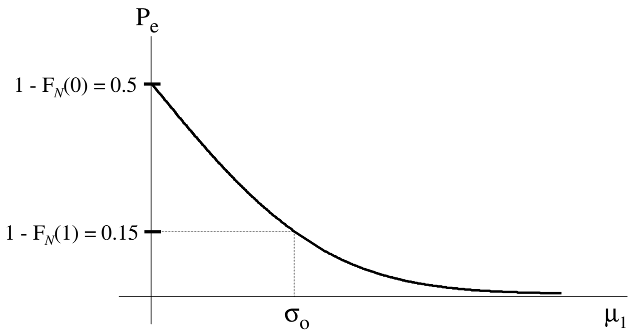


Figure 76: *Error probability curve of Bayes LRT as function of* $\mu_1 = \|s\|^2$.

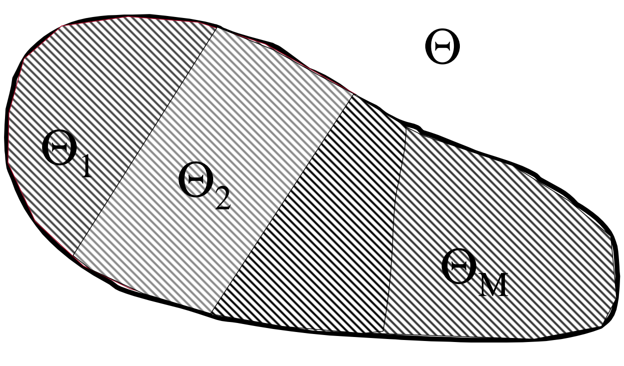


Figure 77: *Partition of* Θ *into* M *different regions*

7.3 TESTING MULTIPLE HYPOTHESES

We measure x having conditional p.d.f. $f(x|\theta)$

* $\theta \in \Theta$

Consider partition $\Theta_1, \ldots, \Theta_M$ of parameter space

OBJECTIVE: To test M hypotheses on θ

$$\begin{aligned} H_1 : \theta &\in \Theta_1 \\ \vdots \quad &\vdots \quad \vdots \\ H_M : \theta &\in \Theta_M \end{aligned}$$

DECISION FUNCTION:

$$\underline{\phi}(x) = [\phi_1(x), \ldots, \phi_M(x)]^T$$

where

$$\phi_i(x) \in \{0, 1\}, \qquad \sum_{i=1}^{M} \phi_i(x) = 1$$

NOTE: decision function specifies a partition of measurement space $\mathcal{X}$ into M decision regions

$$\mathcal{X}_i = \{x : \phi_i(x) = 1\}, \quad i = 1, \ldots, M$$

BAYES APPROACH: has three elements:

1. Assign a prior $f(\theta)$ density for θ

2. Assign costs to wrong decisions

c_{ij} = cost of deciding H_i when H_j is true

3. Find and implement decision rule which has minimum average cost

7.3.1 PRIOR PROBABILITIES

Obtain prior probabilities on $H_i, i = 1, \ldots, M$

$$P(H_i) = P(\theta \in \Theta_i) = \int_{\Theta_i} f(\theta) d\theta$$

with $\sum_{i=1}^{M} P(H_i) = 1$

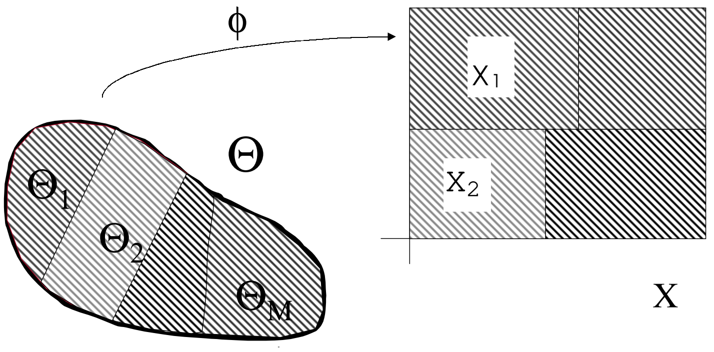


Figure 78: *Partition of parameter space into M hypotheses is equivalent (via the test function $\underline{\phi}(x)$) to partition of $\mathcal{X}$ into M regions.*

We now have conditional p.d.f.s

$$f(x|H_i) = \frac{\int_{\Theta_i} f(x|\theta) f(\theta) d\theta}{P(H_i)}$$

⇒ Thus we have reduced the composite hypotheses above to the following simple hypotheses

$$\begin{array}{rcl} H_1 : x & \sim & f(x|H_1) \\ \vdots & \vdots & \vdots \\ H_M : x & \sim & f(x|H_M) \end{array}$$

where H_i has prior probability $P(H_i)$

7.3.2 MINIMIZE AVERAGE RISK

Cost or risk matrix:

$$C = \begin{bmatrix} c_{11} & \cdots & c_{1M} \\ \vdots & \ddots & \vdots \\ c_{M1} & \cdots & c_{MM} \end{bmatrix}$$

Design Criterion: Select $\underline{\phi}$, equivalently $\{\mathcal{X}_i\}_{i=1}^M$, to minimize average risk

$$\overline{c} = \sum_{i,j=1}^{M} c_{ij} P(\text{say } H_i | H_j) P(H_j)$$

Specialize to case

* $c_{ii} = 0$

* $c_{ij} = 1, i \neq j$

Then we have $\overline{c} = P_e$

$$\begin{aligned}
\overline{c} &= \sum_{i,j:i\neq j} P(\text{say } H_i | H_j) P(H_j) \\
&= 1 - \sum\nolimits_{i,j:i=j} P(\text{say } H_i | H_j) P(H_j) \\
&= 1 - \sum\nolimits_{i,j:i=j} P(x \in \mathcal{X}_i | H_i) P(H_i) \\
&= 1 - \sum_{i=1}^{M} \int_{\mathcal{X}_i} f(x|H_i) dx \; P(H_i)
\end{aligned}$$

Observe: to make $\overline{c}$ as small as possible

$$x \in \mathcal{X}_i \quad \Leftrightarrow \quad f(x|H_i)P(H_i) \geq f(x|H_j)P(H_j), \quad j \neq i$$

Or in terms of decision function:

$$\phi_i(x) = \begin{cases} 1, & f(x|H_i)P(H_i) \geq f(x|H_j)P(H_j) \\ 0, & o.w. \end{cases}$$

Shorthand notation

$$\hat{H}_i = \hat{H}_i(x) = \text{argmax}_{H_j} \{f(x|H_j)P(H_j)\}$$

This is equivalent to the "MAP" rule

$$\hat{H}_i = \text{argmax}_{H_j} \{P(H_j|x)\}$$

REMARKS:

* MAP decision rule minimizes average P_e

* Minimum average P_e is equal to

$$P_e^* = 1 - \sum_{i=1}^{M} E[\phi_i(x)|H_i]P(H_i)$$

* MAP decision rule on x depends only through LR = sufficient statistic
* For equally likely H_i, $P(H_i) = 1/M$ and MAP test is of form

$$\hat{H}_i = \text{argmax}_{H_j} \{f(x|H_j)\}$$

= "Maximum likelihood" estimate of true hypothesis H_j

Example 27 *Classifier of Gaussian Means*

* $\underline{x} = [x_1, \ldots, x_n]^T$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$
* σ^2 is known
OBJECTIVE: classify μ among three possible values

$$\begin{aligned} H_1 : \mu &= \mu_1 \\ H_2 : \mu &= \mu_2 \\ H_3 : \mu &= \mu_3 \end{aligned}$$

Assume equally likely hypotheses
We know that MAP classifier depends on $\underline{x}$ only through sufficient statistic for μ:

$$\overline{x_i} = n^{-1} \sum_{i=1}^{n} x_i$$

which is Gaussian with mean μ and variance σ^2/n.
Therefore MAP test is of form:
Decide H_k iff

$$f(\overline{x_i}|H_k) \geq f(\overline{x_i}|H_j)$$

where

$$f(\overline{x_k}|H_k) = \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left(-\frac{1}{2\sigma^2/n} (\overline{x_i} - \mu_k)^2\right)$$

or eliminating common factors and taking logarithm

$$\overline{x_i}\mu_k - \tfrac{1}{2}\mu_k^2 \geq \overline{x_i}\mu_j - \tfrac{1}{2}\mu_j^2$$

$\Rightarrow$ Linear decision regions!

Concrete example: $\mu_1 = -1$, $\mu_2 = +1$, $\mu_3 = 2$

Plot 3 lines as function of $\overline{x_i}$ to find the regions.

$$\begin{aligned}
\mathcal{X}_1 &= \{x : \overline{x_i} \leq 0\} \\
\mathcal{X}_2 &= \{x : 0 < \overline{x_i} \leq 3/2\} \\
\mathcal{X}_3 &= \{x : \overline{x_i} \geq 3/2\}
\end{aligned}$$

These are regions separated by hyperplanes in $\mathcal{X} = \mathbb{R}^n$.

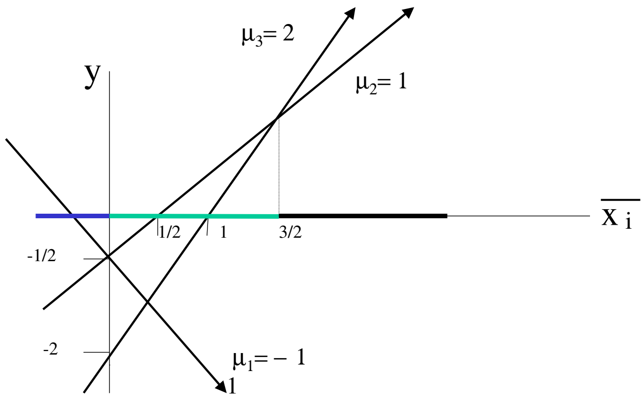


Figure 79: *For three hypotheses on Gaussian mean the decision regions are specified by intersections of three lines $y = \overline{x}_i\mu_k - \frac{1}{2}\mu_k^2$, $k = 1, 2, 3$, which are graphed here over domain $\overline{x_i}$. (Note, the labels "$\mu_1 = 2$" and "$\mu_2 = 1$" should be switched in figure).*

7.3.3 DEFICIENCIES OF BAYES APPROACH

* Requires assigning prior to θ, H_0, $H_1, \ldots$

* Only ensures best average performance w.r.t. selected prior

* Provides no guaranteed protection against FA, M

7.4 FREQUENTIST APPROACH TO DETECTION

Alternative criterion: constrain FA and minimize M.

We will need to extend the test function to allow randomization:

$$\phi(x) = \begin{cases} 1, & \text{say } H_1 \\ q(x), & \text{flip a coin w/ prob Heads } (H_1) = q(x) \\ 0, & \text{say } H_0 \end{cases}$$

Note, we have interpretation:

$$\phi(x) = P(\text{say } H_1 | \text{observe } x)$$

False alarm probability and detection probability are functions of θ

$$E_\theta[\phi] = \int_{\mathcal{X}} \phi(x) f(x;\theta) dx = \begin{cases} P_F(\theta), & \theta \in \Theta_0 \\ P_D(\theta), & \theta \in \Theta_1 \end{cases}$$

Definition: A test ϕ is said to be of (FA) level $\alpha \in [0,1]$ if

$$\max_{\theta \in \Theta_0} P_F(\theta) \leq \alpha$$

Definition: The **power function**' of a test ϕ is

$$\beta(\theta) = P_D(\theta) = 1 - P_M(\theta), \quad \theta \in \Theta_1$$

7.4.1 CASE OF SIMPLE HYPOTHESES: $\theta \in \{\theta_0, \theta_1\}$

$$H_0 : x \sim f(x;\theta_0)$$

$$H_1 : x \sim f(x;\theta_1)$$

Neyman-Pearson Strategy: find most powerful (MP) test ϕ^* of level α:

$$E_{\theta_1}[\phi^*] \geq E_{\theta_1}[\phi]$$

for any other test satisfying $E_{\theta_0}[\phi] \leq \alpha$.

Lemma 1 Neyman Pearson Lemma: *The MP test of level $\alpha \in [0,1]$ is a randomized LRT of the form*

$$\phi^*(x) = \begin{cases} 1, & f(x;\theta_1) > \eta f(x;\theta_0) \\ q, & f(x;\theta_1) = \eta f(x;\theta_0) \\ 0, & f(x;\theta_1) < \eta f(x;\theta_0) \end{cases} \tag{34}$$

where η and q are selected to satisfy

$$E_{\theta_0}[\phi^*] = \alpha$$

Proof 1 of NPL: uses Kuhn-Tucker theory [17] of constrained maximization

The MP test maximizes power $E_{\theta_1}[\phi(x)]$ subject to constraint $E_{\theta_0}[\phi(x)] \leq \alpha$.

This constrained estimation problem is equivalent to maximizing unconstrained objective function

$$L(\phi) = E_{\theta_1}[\phi(x)] + \eta\left(\alpha - E_{\theta_0}[\phi(x)]\right)$$

where $\eta > 0$ is Lagrange multiplier selected so that solution ϕ^* satisfies constraint $E_{\theta_0}[\phi] = \alpha$.

Now the power can be expressed via the LR transformation for expectation

$$E_{\theta_1}[\phi(x)] = E_{\theta_0}\left[\phi(x)\frac{f(x;\theta_1)}{f(x;\theta_0)}\right]$$

and hence:

$$L(\phi) = E_{\theta_0}\left[\phi(x)\left(\frac{f(x;\theta_1)}{f(x;\theta_0)} - \eta\right)\right] + \eta\alpha$$

Which yields the randomized LRT (34) of NPL. ◇

Proof 2 of NPL: more elementary

Need show that for ϕ arbitrary, ϕ^* satisfies

$$E_{\theta_1}[\phi^*] \geq E_{\theta_1}[\phi], \quad \text{when} \quad E_{\theta_0}[\phi^*] = \alpha, \ \ E_{\theta_0}[\phi] \leq \alpha$$

Two steps:

Step 1: Show by enumerating all possible cases of $>, <$ and $=$ between the terms on RHS and LHS

$$\phi^*(x)[f(x;\theta_1) - \eta f(x;\theta_0)] \geq \phi(x)[f(x;\theta_1) - \eta f(x;\theta_0)] \tag{35}$$

Step 2: integrate (35) over all x

$$\int_{\mathcal{X}} \phi^*(x)[f(x;\theta_1) - \eta f(x;\theta_0)]dx \geq \int_{\mathcal{X}} \phi(x)[f(x;\theta_1) - \eta f(x;\theta_0)]dx$$

$$= \underbrace{\int_{\mathcal{X}} \phi^*(x) f(x;\theta_1)dx}_{E_{\theta_1}[\phi^*]} - \eta \underbrace{\int_{\mathcal{X}} \phi^*(x) f(x;\theta_0)dx}_{E_{\theta_0}[\phi^*]}$$

$$\geq \underbrace{\int_{\mathcal{X}} \phi(x) f(x;\theta_1)dx}_{E_{\theta_1}[\phi]} - \eta \underbrace{\int_{\mathcal{X}} \phi(x) f(x;\theta_0)dx}_{E_{\theta_0}[\phi]}$$

Hence

$$E_{\theta_1}[\phi^*] - E_{\theta_1}[\phi] \geq \eta(\underbrace{E_{\theta_0}[\phi^*]}_{=\alpha} - \underbrace{E_{\theta_0}[\phi]}_{\leq\alpha}) \geq 0$$

Which establishes NPL. $\diamond$

RESOURCE ALLOCATION INTERPRETATION OF MP TEST

$f(x;\theta_0)$ = current value of security x

$f(x;\theta_1)$ = future value of security x

$\phi(x)$ = decision whether or not to invest in security x

α = total available dollars for investment

β = total future value of investment

NPL says simply: it is best to invest α\$ in the securities which have the highest returns $f(x;\theta_1)/f(x;\theta_0)$.

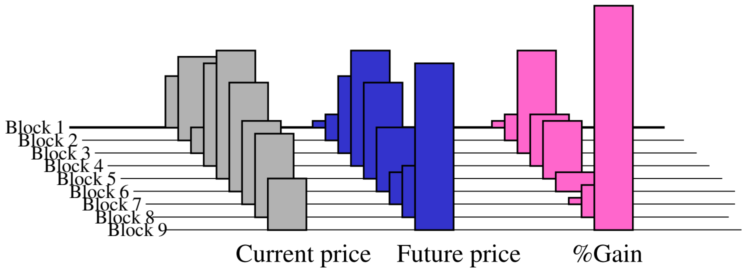


Figure 80: *Future value, current value, and relative return of a set of securities $\mathcal{X}$*

GENERAL REMARKS CONCERNING MP TESTS

Remark 1. shorthand LRT notation

$$\Lambda(x) = f(x;\theta_1)/f(x;\theta_0) \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

Remark 2. P_F of MP test is

$$P_F = E_{\theta_0}[\phi^*(x)] = \underbrace{P_{\theta_0}(\Lambda > \eta)}_{1-F_\Lambda(\eta|H_0)} + qP_{\theta_0}(\Lambda = \eta)$$

Randomization must be performed only if $P_{\theta_0}(\Lambda > \eta)$ has jump discontinuites.

Otherwise q can be set to zero and randomization is not necessary.

Jump discontinuities occur when $\Lambda = \Lambda(x)$ is not a cts random variable

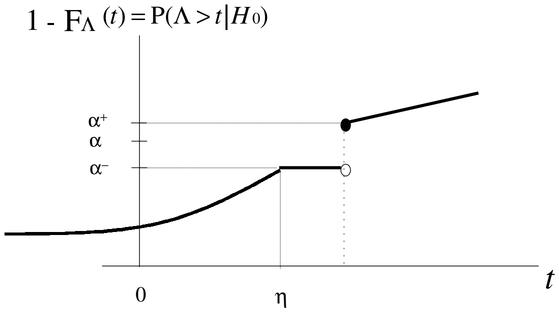


Figure 81: *Randomization is necessary to attain a level α that is not in the range of 1-CDF of Λ.*

Remark 3. LR is identical to Bayes LR for simple hypotheses

Remark 4. Unlike BLRT threshold η is specified by only one quantity α.

Remark 5. If $T = T(x)$ is a sufficient statistic for θ, LRT depends on x only through $T(x)$

Indeed if $f(x;\theta) = g(T,\theta)h(x)$ then

$$\Lambda(x) = g(T,\theta_1)/g(T,\theta_0) = \Lambda(T)$$

Conclude: can formulate the LRT based on p.d.f. of T instead of the p.d.f. of entire data sample x.

7.5 ROC CURVES FOR THRESHOLD TESTS

All threshold tests have P_F and P_D indexed by a parameter η.

The Receiver Operating Characteristic (ROC) is simply the plot of the parametric curve $\{P_F(\eta,q), P_D(\eta,q)\}_{\eta,q}$.

Equivalently, ROC is the plot of $\beta = P_D$ vs $\alpha = P_F$.

PROPERTIES OF ROC'S

1. ROC for coin flip detector ($\phi(x) = q$ independent of data) is a diagonal line with slope =1

$$\begin{aligned} \alpha = P_F &= E_{\theta_0}[\phi] = q \\ \beta = P_D &= E_{\theta_1}[\phi] = q \end{aligned}$$

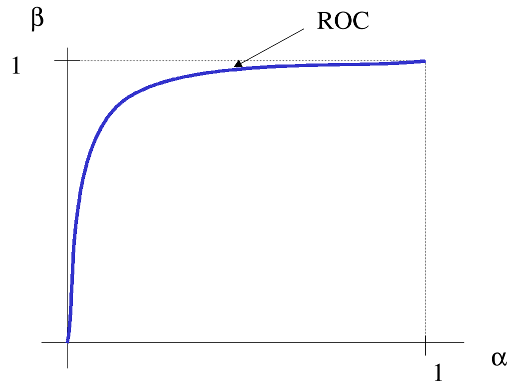


Figure 82: *A typical ROC curve.*

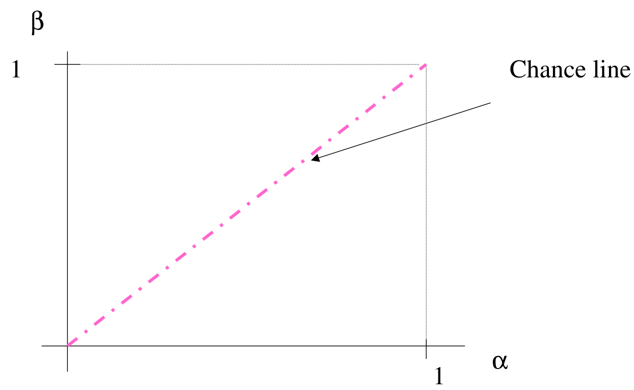


Figure 83: *ROC curve for coin flip detector.*

2. ROC of any LRT always lies above diagonal: LRT is "unbiased" test

Definition: a test ϕ is unbiased if its detection probability β is at least as great as its false alarm α: $\beta \geq \alpha$.

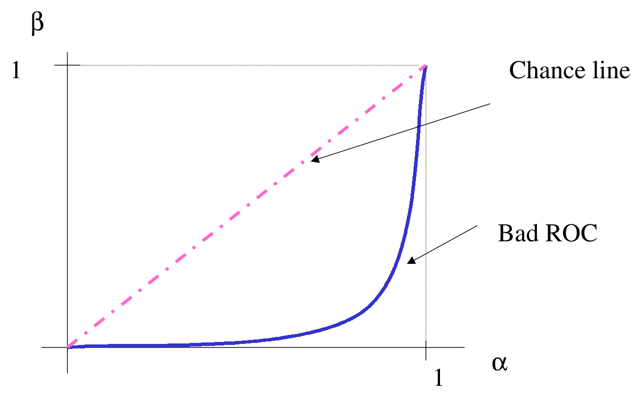


Figure 84: *ROC curve for LRT always lies above diagonal.*

3. ROC of any LRT is always convex cap (concave).

To see concavity, let (α_1, β_1) be the level and power of a test ϕ_1 and (α_1, β_1) be the level and power of a test ϕ_2. Then, with z a random variable independent of x having $P(z = 1) = p = 1 - P(z = 0)$, define the test

$$\phi_{12} = z\phi_1 + (1 - z)\phi_2$$

This test, selects ϕ_1 and ϕ_2 at random with probability p and $1 - p$, respectively. As z is independent of ϕ_1, ϕ_2, the level of this test is

$$\alpha_{12} = E_o[\phi_{12}] = E_o[z]E_o[\phi_1] + E_o[(1 - z)]E_o[\phi_2] = p\alpha_1 + (1 - p)\alpha_2$$

and its power is similarly

$$\beta_{12} = E_o[\phi_{12}] = p\beta_1 + (1 - p)\beta_2$$

Thus, as p varies between 0 and 1, $\phi_1 2$ has performance $(\alpha_{12}, \beta_{12})$ which varies on a straight line connecting the points (α_1, β_1) and (α_2, β_2).

4. If ROC curve is differentiable, MP-LRT threshold needed for attaining any pair $(\alpha, P_D(\alpha))$ on ROC can be found graphically as slope of ROC at the point α.

$$\eta = \frac{d}{d\alpha} P_D(\alpha)$$

5. The minmax P_e LRT threshold can also be found graphically by intersection of line $P_M = 1 - P_D = P_F$ and ROC.

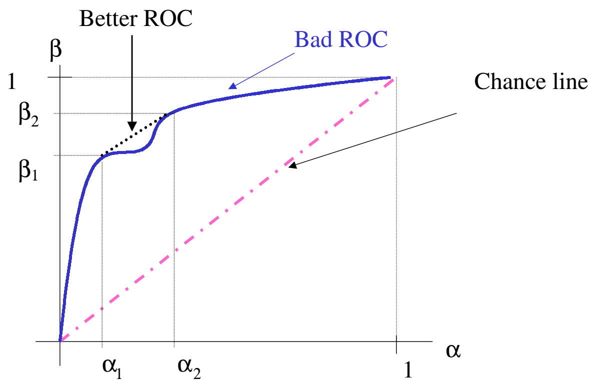


Figure 85: *ROC of any LRT is always convex cap. A test with non-convex ROC (thick line) can always be improved by randomization which has effect of connecting two endpoints (*(α_1, β_1) *and* (α_2, β_2) *on ROC by straight line.*

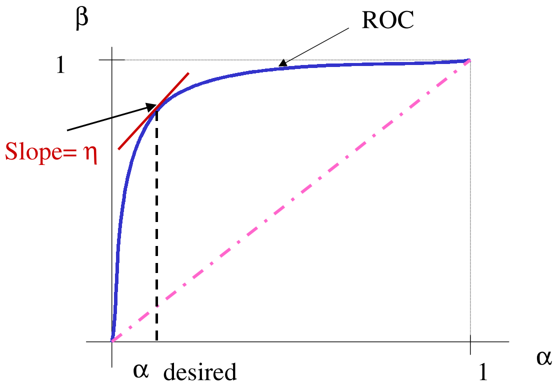


Figure 86: *Threshold of LRT can be found by differentiation of ROC curve.*

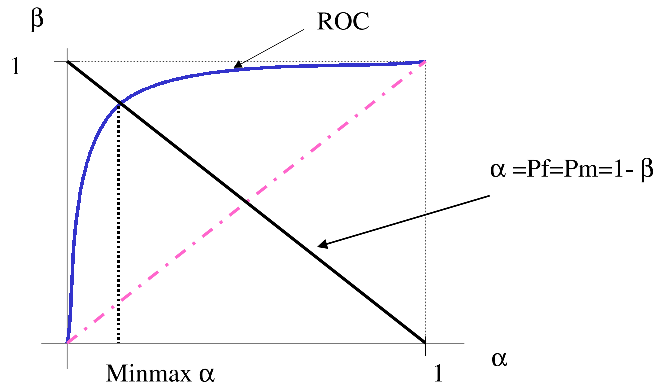


Figure 87: *Threshold of min-max Bayes LRT can be found by intersection.*

Example 28 *Test against uniform density*

Two hypotheses on a scalar r.v. x

$$H_0 : f(x) = f_0(x)$$

$$H_1 : f(x) = f_1(x)$$

where f_0 and f_1 are two densities shown below

Objective: find the MP-LRT

Solution:

LRT is

$$\Lambda = \frac{f_1(x)}{f_0(x)} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

or equivalently

$$f_1(x) \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta f_0(x)$$

From figure it is obvious that for a given η the H_1 decision region is

$$\mathcal{X}_1 = \begin{cases} \{\eta/4 < x < 1 - \eta/4\}, & 0 \leq \eta \leq 2 \\ \text{empty}, & o.w. \end{cases}$$

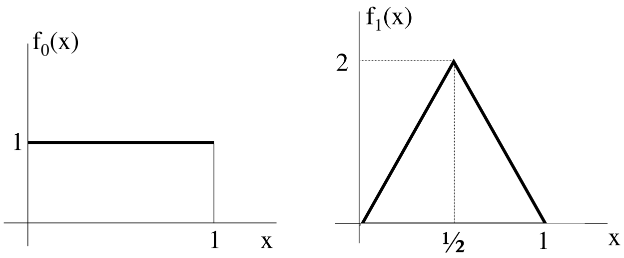


Figure 88: *Two densities to be tested*

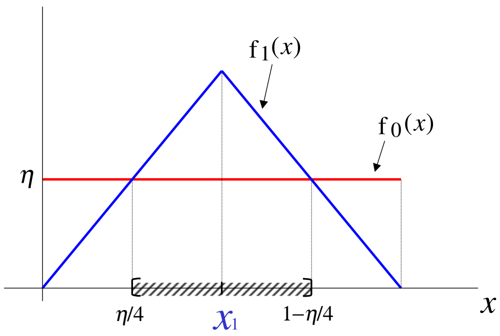


Figure 89: *Region $\mathcal{X}_1$ for which MP-LRT decides H_1 are set of values x for which triangle exceeds horizontal line of height η.*

Setting threshold

Select η to meet constraint $P_F = \alpha$.

FIRST: attempt to set η without randomization $(q = 0)$.

Assume $\eta \in [0, 2]$

$$\begin{aligned}
\alpha &= P(x \in \mathcal{X}_1 | H_0) = \int_{\eta/4}^{1-\eta/4} f_0(x)dx \\
&= 1 - \eta/2
\end{aligned}$$

Hence required η is simply

$$\eta = 2(1 - \alpha)$$

and we see that no randomization is required.

Power of MP-LRT is:

$$\begin{aligned}
P_D &= P(x \in \mathcal{X}_1 | H_1) = \int_{\eta/4}^{1-\eta/4} f_1(x)dx \\
&= 2\int_{\eta/4}^{\frac{1}{2}} f_1(x)dx = 2\int_{\eta/4}^{\frac{1}{2}} 4xdx \\
&= 1 - \eta^2/4
\end{aligned}$$

Plug in level α threshold $\eta = 2(1 - \alpha)$ to power expression to obtain the ROC curve

$$\beta = 1 - (1 - \alpha)^2$$

Example 29 *Testing for an increase in rate θ of Poisson arrivals*

$$x \;\sim\; f(x; \theta) = \frac{\theta^x}{x!}e^{-\theta}, \quad x = 0, 1, \ldots$$

Objective: to find MP-LRT of

$$\begin{aligned}
H_0 : \theta &= \theta_0 \\
H_1 : \theta &= \theta_1
\end{aligned}$$

for $\theta_1 > \theta_0 > 0$

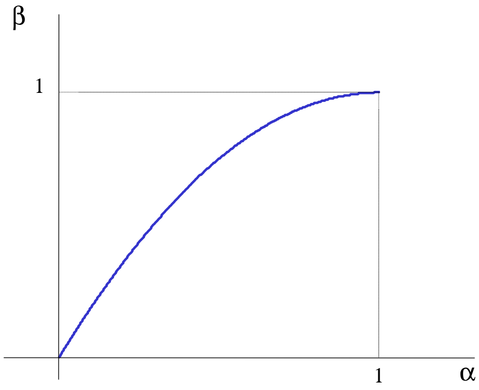


Figure 90: *ROC curve for uniform vs. triangle pdf example.*

Solution: MP test is LRT

$$\Lambda(x) = \left(\frac{\theta_1}{\theta_0}\right)^x e^{\theta_0 - \theta_1} \quad \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \quad \eta$$

Since "ln" is monotone increasing, and $\theta_1 > \theta_0$, MP-LRT is equivalent to linear test

$$x \quad \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \quad \gamma$$

where (immaterial for MP-LRT) $\gamma = \frac{\ln \eta + \theta_1 - \theta_0}{\ln(\theta_1/\theta_0)}$
First try to set threshold γ without randomization:

$$\alpha = P_{\theta_0}(x > \gamma) = 1 - \mathcal{P}\text{o}_{\theta_0}(\gamma)$$

where $\mathcal{P}\text{o}_{\theta}(\cdot)$ is the CDF of a Poisson r.v. with rate θ.
Note: as Poisson CDF is not continuous only the following values are attainable without randomization

$$\alpha \in \{\alpha_i\}_{i=1}^{\infty}, \qquad \alpha_i = 1 - \mathcal{P}\text{o}_{\theta_0}(i)$$

If $\alpha \neq \alpha_i$ for some i we need randomize by selecting γ, q to satisfy:

$$\alpha = P_{\theta_0}(x > \gamma) + qP_{\theta_0}(x = \gamma)$$

Solution to randomization threshold for Poisson example:

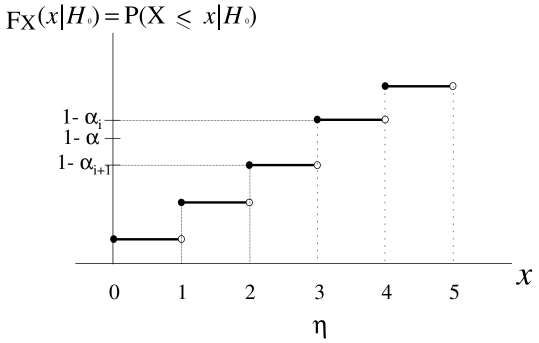


Figure 91: *CDF of LR test statistic for testing increase in Poisson rate is staircase function*

Assume $\alpha \in (\alpha_i, \alpha_{i+1})$. Set

$$\gamma = \gamma^* := \mathcal{P}\text{o}_{\theta_0}^{-1}(1 - \alpha_i)$$

which gives $P_{\theta_0}(x > \gamma^*) = \alpha_i$
and $P_{\theta_0}(x = \gamma^*) = \alpha_{i+1} - \alpha_i$
giving

$$\alpha = \alpha_i + q(\alpha_{i+1} - \alpha_i)$$

remains to solve this equation for q:

$$q = q^* := \frac{\alpha - \alpha_i}{\alpha_{i+1} - \alpha_i}$$

Power of randomized MP-LRT is:

$$P_D = P_{\theta_1}(x > \gamma^*) + q^* P_{\theta_1}(x = \gamma^*)$$

Example 30 *On Off keying (OOK) in Gaussian noise*

$$\begin{aligned} H_0 : x &= w \\ H_1 : x &= 1 + w \end{aligned}$$

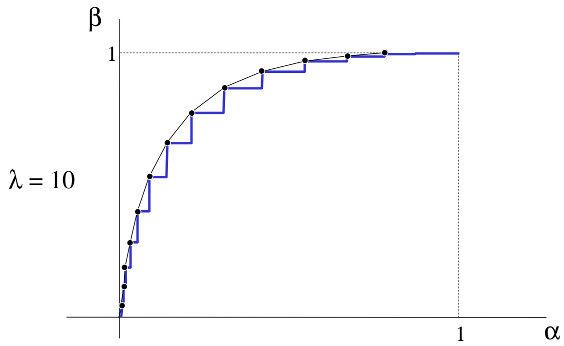


Figure 92: *Power curves of LRT for detecting an increase in rate of a Poisson r.v. The smooth curve is the (randomized) MP test while the staircase curve is the non-randomized LRT.*

where $w \sim \mathcal{N}_1(0,1)$

LR statistic is

$$\Lambda(x) = \frac{\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x-1)^2}}{\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}} = e^{x-\frac{1}{2}}$$

We will compare Bayes LRT , MP-LRT and minmax LRT

1. Min P_e (Bayes) test for equally likely H_0, H_1 ($\eta = 1$):

$$x \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \ln\eta + \tfrac{1}{2} = \tfrac{1}{2}$$

2. Minmax test:

$$x \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \ln\eta + \tfrac{1}{2} := \lambda$$

Where λ is chosen to satisfy

$$P_F = 1 - \mathcal{N}(\lambda) = \mathcal{N}(\lambda - 1) = P_M$$

Solution $\lambda = \frac{1}{2}$ is obvious since $\mathcal{N}(-x) = 1 - \mathcal{N}(x)$:

3. NP test:

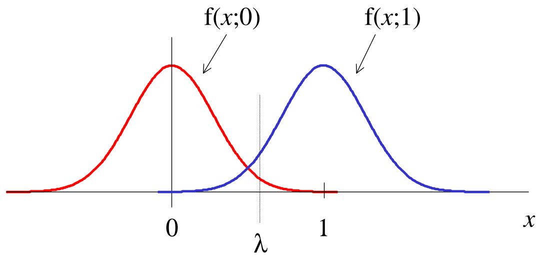


Figure 93: *Densities under H_0 and H_1 for on-off keying detection.*

$$x \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

where $\alpha = P(x > \lambda | H_0) = 1 - \mathcal{N}(\lambda)$ or

$$\lambda = \mathcal{N}^{-1}(1-\alpha)$$

ROC curve is obtained from $P_D = P(x > \lambda | H_1) = P_D(x - 1 > \lambda - 1 | H_1) = 1 - \mathcal{N}(\lambda - 1)$ so that, substituting λ expression above,

$$\beta = P_D = 1 - \mathcal{N}(\lambda - 1) = 1 - \mathcal{N}(\mathcal{N}^{-1}(1-\alpha) - 1)$$

Performance comparison:

Specifying FA level $\alpha = 0.001$ we obtain for MP-LRT: $\lambda = 2.329$

x	P_F	P_D	P_e
Bayes	0.31	0.69	0.31
Minmax	0.31	0.69	0.31
NP	0.001	0.092	0.5

7.6 Exercises

1. A proprietary binary hypothesis test ϕ is implemented in a software package which you are considering purchasing based on a trial examination period. You run several experiments and obtain the following table of probabilities of detection β vs. false alarm α

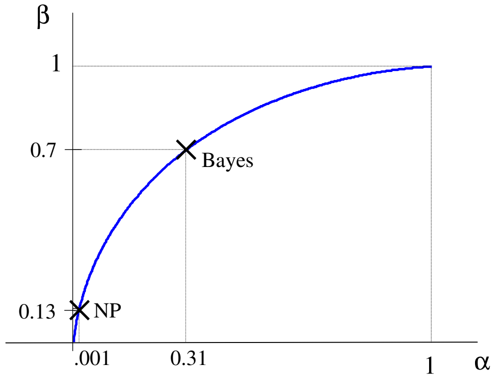


Figure 94: *ROC curve for Gaussian OOK example.*

α	β
0.1	0.2
0.3	0.4
0.5	0.8
0.7	0.9

Comment on the quality of this test. Could you improve on this test? If so specify the improved test and compute its ROC curve.

2. Let Z be a random variable with values in the interval $[-1, 1]$ having density function

$$p_\theta(z) = \frac{1}{2}\frac{3}{3+\theta}\left(\theta z^2 + 1\right)$$

where $\theta > 0$. Note θ controls the deviation of p_θ from the uniform density p_0. You are to test Z against non-uniformity given a *single sample* Z.

   (a) Assuming priors $p = P(H_1) = 1 - P(H_0)$ derive the minimum probability of error (MAP) test for the simple hypotheses $H_0 : \theta = 0$ vs. $H_1 : \theta = \theta_1$, where θ_1 is a fixed and known positive value.

   (b) Find an expression for the ROC curve and plot for $\theta_1 = 0, 1, 10$.

   (c) Now find the form of the min-max test. Show how you can use your answer to part b) to graphically determine the min-max threshold.

   (d) Derive the MP test of level α for the same simple hypotheses as in part (a). Does the form of the test change as a function of θ?

3. Let Z be a single observation having density function

$$p_\theta(z) = (2\theta z + 1 - \theta), \qquad 0 \leq z \leq 1$$

where $-1 \leq \theta \leq 1$.

(a) Find the most powerful test between the hypotheses

$$\begin{aligned} H_0 &: \theta = 0 \\ H_1 &: \theta = 1 \end{aligned}$$

Be sure to express your test in terms of the false alarm level $\alpha \in [0,1]$. Plot the ROC curve for this test.

(b) repeat part (a) for $H_1 : \theta = -1$.

4. It is desired to test the following hypotheses based on a single sample x:

$$\begin{aligned} H_0 &: x \sim f_0(x) = \frac{3}{2}\, x^2, -1 \le x \le 1 \\ H_1 &: x \sim f_1(x) = \frac{3}{4}\, (1 - x^2), -1 \le x \le 1 \end{aligned}$$

(a) Under the assumption that the prior probabilities of H_0 and H_1 are identical, find the minimum probability of error (Bayes) test.

(b) Find the Most Powerful test of level $\alpha \in [0,1]$.

(c) Derive and plot the ROC curve for these tests.

5. Let $f(x|H_0)$ and $f(x|H_1)$ be densities of an observed r.v. x and assume that the likelihood ratio $\Lambda = f(x|H_1)/f(x|H_0)$ has corresponding densities $f_\Lambda(\lambda|H_0)$ and $f_\Lambda(\lambda|H_1)$ under H_0 and H_1, respectively. Show that the slope $d\beta/d\alpha$ at a point α of the ROC of the LRT is equal to the threshold η attaining level α. (Hint: show that $d\beta/d\alpha = f_\Lambda(\eta|H_1)/f_\Lambda(\eta|H_0)$ and then apply $f_\Lambda(u|H_k) = \int_{\{x:\Lambda(x)=u\}} f(x|H_k)dx$, $k = 0, 1$.)

6. Let a detector have the ROC curve $\{(\alpha, \beta) : \alpha \in [0,1]\}$ where the power function $\beta = \beta(\alpha)$ is a function of the false alarm level α. The area under the ROC is defined as

$$\mathrm{AUC} = \int_0^1 \beta(\alpha) d\alpha$$

The AUC is frequently used as an alternative to the power function to assess the performance of various detectors. Assume simple hypotheses and invoke properties of ROCs in answering the following questions.

(a) Show that among all tests the MP LRT maximizes AUC.

(b) Show the following inequalities for the AUC of a MP LRT

$$\frac{1}{2} \le \mathrm{AUC} \le \beta(\tfrac{1}{2}) \le 1$$

(c) Show that for any LRT whose ROC $\beta(\alpha)$ is differentiable in α

$$\mathrm{AUC} = 1 - \int_0^1 \alpha\eta(\alpha) d\alpha$$

where $\eta = \eta(\alpha)$ is the LRT's threshold attaining the false alarm level α. When combined with (b) this implies the interesting result for LRT's: as the integral is bounded $\lim_{\alpha\to 0}(\alpha\eta) = 0$, i.e. α decreases to zero faster than $\eta(\alpha)$ increases to ∞.

End of chapter

8 DETECTION FOR COMPOSITE HYPOTHESES

* Uniformly most powerful (UMP) tests
* Min-max tests
* Locally best tests,
* CFAR tests
* Unbiased tests
* GLR tests

REFERENCES

Van Trees [36]

Lehmann [15]

Ferguson [4]

8.1 UMP TESTS

First we consider a simple null hypothesis with composite alternative

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &\in \Theta_1 \end{aligned}$$

Note that for good detectors $P_D = P_D(\theta_1)$ varies as function of $\theta_1 \in \Theta_1$

Recall that ϕ is of level α if

$$P_F = E_{\theta_0}[\phi] \leq \alpha$$

Definition: a test ϕ^* is a uniformly most powerful (UMP) test of level α if for any other level α test ϕ

$$\beta^*(\theta) = E_\theta[\phi^*] \geq E_\theta[\phi] = \beta(\theta), \quad \text{for all } \theta \in \Theta_1$$

Two steps to discovering UMPs when they exist:

Step 1: find MP test of level α for fixed θ

Step 2: if decision regions do not vary with $\theta \in \Theta_1$ then UMP

Example 31 *Tests of mean in Gaussian sample with known variance*

$x = [x_1, \ldots, x_n]^T$ i.i.d., $x_1 \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 is known.

Three cases of interest:

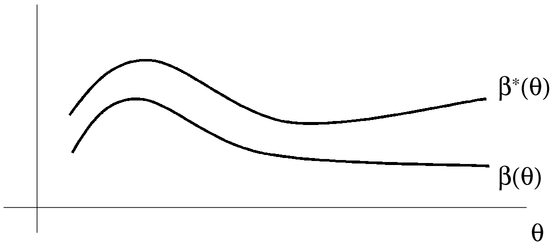


Figure 95: *Power curve $\beta^*(\theta)$, $\theta \in \Theta_1$ of a UMP test is uniformly higher than that of any other test.*

$$\underbrace{\begin{array}{c} H_0 : \mu = 0 \\ H_1 : \mu > 0 \end{array}}_{\text{Case I}} \quad \underbrace{\begin{array}{c} H_0 : \mu = 0 \\ H_1 : \mu < 0 \end{array}}_{\text{Case II}} \quad \underbrace{\begin{array}{c} H_0 : \mu = 0 \\ H_1 : \mu \neq 0 \end{array}}_{\text{Case III}}$$

Step 1: find LRT for fixed μ under H_1

Enough to work problem based on sufficient statistic $T = \overline{x_i}$ for μ.

Know:

$$\begin{aligned} \overline{x_i} &\sim \mathcal{N}(0, \sigma^2/n), \quad \text{under } H_0 \\ \overline{x_i} &\sim \mathcal{N}(\mu, \sigma^2/n), \quad \text{under } H_1 \end{aligned}$$

so

$$\begin{aligned} \Lambda(\mu) &= \frac{f(\overline{x_i};\mu)}{f(\overline{x_i};0)} = \frac{\exp\left(-\frac{(\overline{x_i}-\mu)^2}{2\sigma^2/n}\right)}{\exp\left(-\frac{\overline{x_i}^2}{2\sigma^2/n}\right)} \\ &= \exp\left(\frac{n\mu}{\sigma^2}\overline{x_i} - \frac{n\mu^2}{2\sigma^2}\right) \overset{H_1}{\underset{H_0}{\gtrless}} \eta \end{aligned}$$

Note that $\Lambda(\mu)$ is monotone increasing in $\mu\overline{x_i}$ so that one form of the MP-LRT is

$$\mu \left(\frac{\sqrt{n}\,\overline{x_i}}{\sigma}\right) \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

CASE I: Single sided alternative $H_1 : \mu > 0$

In this case μ can be absorbed into RHS without changing inequalities:

$$T(\underline{x}) = \frac{\sqrt{n}\,\overline{x_i}}{\sigma} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma^+$$

or equivalently, MP-LRT is the linear detector

$$\sum_{i=1}^n x_i \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma' = \gamma^+ \sqrt{n}\,\sigma^2$$

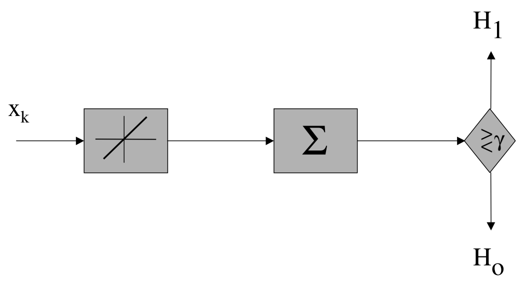


Figure 96: *Optimal detector for positive Gaussian mean is a memoryless linear device followed by a summer and decision mechanism.*

Next we must set threshold:

Since we know $\overline{x_i} \sim \mathcal{N}(0, \sigma^2/n)$ under H_0:

$$\alpha = P_0(\underbrace{\sqrt{n}\,\overline{x_i}/\sigma}_{\mathcal{N}(0,1)} > \gamma^+) \quad = \quad 1 - \mathcal{N}(\gamma^+)$$

Or

$$\gamma^+ = \mathcal{N}^{-1}(1 - \alpha)$$

Final form of MP-LRT for $H_1 : \mu > 0$ reveals that it is UMP against unknown positive μ

$$y \stackrel{\text{def}}{=} \frac{\sqrt{n}\,\overline{x_i}}{\sigma} \underset{H_0}{\overset{H_1}{\gtrless}} \mathcal{N}^{-1}(1-\alpha)$$

Equivalent form in terms of sample mean statistic

$$\overline{x_i} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma}{\sqrt{n}}\mathcal{N}^{-1}(1-\alpha)$$

Equivalent form in terms of sum statistic

$$\sum_{i=1}^{n} x_i \underset{H_0}{\overset{H_1}{\gtrless}} \sqrt{n}\,\sigma\,\mathcal{N}^{-1}(1-\alpha)$$

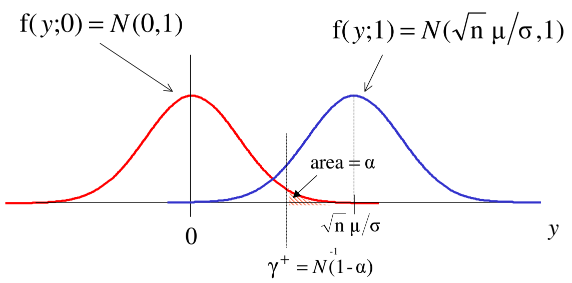


Figure 97: *Threshold γ^+ of MP-LRT for $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$ in i.i.d. Gaussian with known variance. $f(y;0)$ and $f(y;1)$ denote the densities of $y = \frac{\sqrt{n}\,\overline{x_i}}{\sigma}$ under H_0 and H_1, respectively.*

Power of single sided test:

Since $\overline{x_i} \sim \mathcal{N}(\mu, \sigma^2/n)$ under H_1

$$\begin{aligned}\beta &= P_1(\underbrace{\sqrt{n}\,\overline{x_i}/\sigma}_{\mathcal{N}(\sqrt{n}\,\mu/\sigma,\,1)} > \gamma^+) \\ &= 1-\mathcal{N}\left(\gamma^+ - \frac{\sqrt{n}\mu}{\sigma}\right) \\ &= 1-\mathcal{N}\left(\mathcal{N}^{-1}(1-\alpha) - d\right)\end{aligned}$$

where d is the **positive** *detectability index*

$$\begin{aligned} d &= \frac{\sqrt{n}\mu}{\sigma} \\ &= \frac{|E[T|H_1] - E[T|H_0]|}{\sqrt{\mathrm{var}_0(T)}} \end{aligned}$$

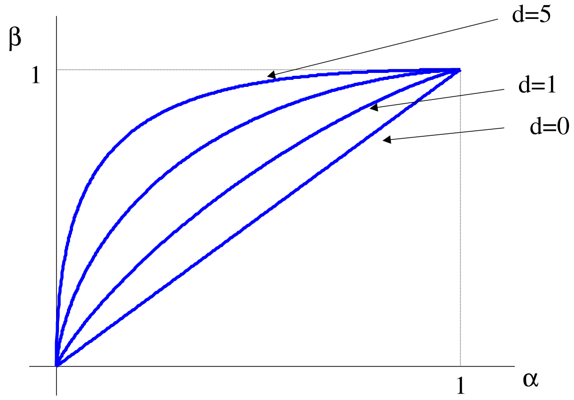


Figure 98: *The ROC curve of MP-LRT for $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$ for n i.i.d. Gaussian with known variance for various values of d.*

CASE II: Single sided alternative $H_1 : \mu < 0$

Recall that the MP LRT for fixed μ has the form

$$\mu \, \frac{\sqrt{n}\,\overline{x_i}}{\sigma} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

This is now equivalent to

$$\frac{\sqrt{n}\,\overline{x_i}}{\sigma} \quad \overset{H_0}{\underset{H_1}{\gtrless}} \quad \gamma^-$$

Setting threshold:

$$\alpha = P_0(\underbrace{\sqrt{n}\,\overline{x_i}/\sigma}_{\mathcal{N}(0,1)} \leq \gamma^-) \quad = \quad \mathcal{N}(\gamma^-)$$

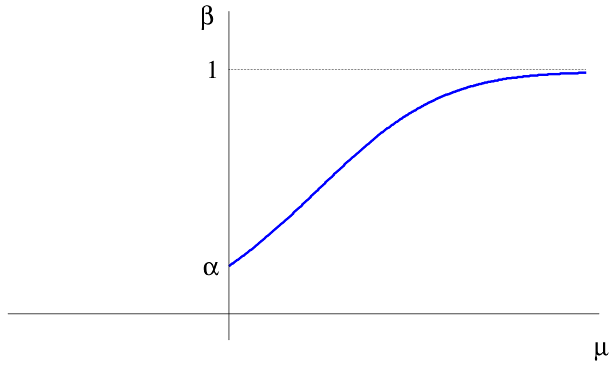


Figure 99: *The power curve of MP-LRT for* $H_0 : \mu = 0$ *vs.* $H_1 : \mu > 0$ *for* n *i.i.d. Gaussian with known variance plotted as a function of* $d > 0$

or now

$$\gamma^- = \mathcal{N}^{-1}(\alpha) = -\mathcal{N}^{-1}(1-\alpha)$$

Again we see MP-LRT is UMP against unknown negative μ

Power curve for $\mu < 0$ case can be derived similarly

$$\beta = 1 - \mathcal{N}\left(\mathcal{N}^{-1}(1-\alpha) + \underbrace{\frac{\sqrt{n}\,\mu}{\sigma}}_{-|d|}\right)$$

where d is now negative valued

$$d = \frac{\sqrt{n}\mu}{\sigma}$$

CASE III: Double sided alternative $H_1 : \mu \neq 0$

Recall again the form of MP LRT for fixed μ

$$\mu\,\frac{\sqrt{n}\,\overline{x_i}}{\sigma} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

Now cannot absorb μ into threshold without affecting inequalities

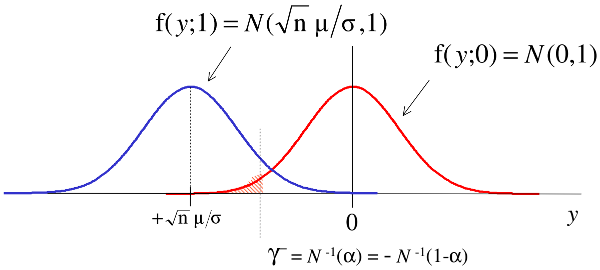


Figure 100: *Threshold determination for MP-LRT of* $H_0 : \mu = 0$ *vs.* $H_1 : \mu < 0$ *for* n *i.i.d. Gaussian observations with known variance*

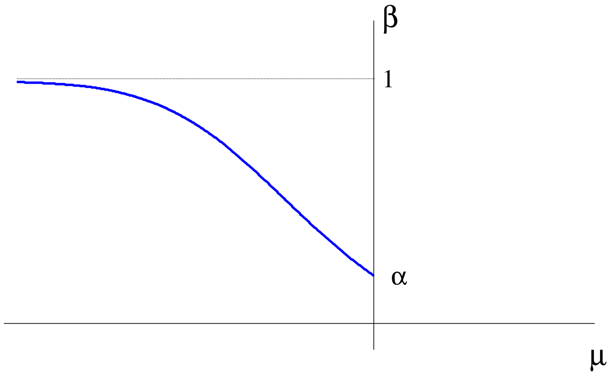


Figure 101: *The power curve of MP-LRT for* $H_0 : \mu = 0$ *vs.* $H_1 : \mu < 0$ *in i.i.d. Gaussian with known variance plotted as a function of d*

$\Rightarrow$ Decision region varies depending on sign of μ

Therefore no UMP exists.

NOTE: use of either of the previous UMP tests for one sided alternative gives biased test!

If use single sided test from CASE I then

$$\beta = 1 - \mathcal{N}\left(\mathcal{N}^{-1}(1-\alpha) - d\right)$$

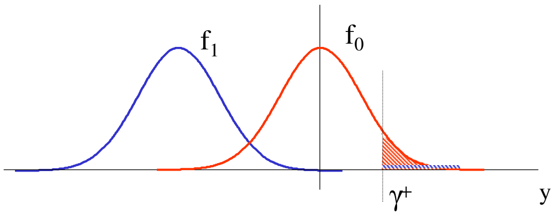


Figure 102: *The single sided MP-LRT for $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$ fails to detect negative signal.*

If use single sided test from CASE II then

$$\beta = 1 - \mathcal{N}\left(\mathcal{N}^{-1}(1-\alpha) + d\right)$$

Example 32 *Test of variance in Gaussian sample with known mean*

$x = [x_1, \ldots, x_n]^T$ i.i.d., $x_1 \sim \mathcal{N}(\mu, \sigma^2)$,

Again three cases of interest:

$$\underbrace{\begin{array}{l} H_0 : \sigma^2 = \sigma_o^2 \\ H_1 : \sigma^2 > \sigma_o^2 \end{array}}_{\text{Case I}} \quad \underbrace{\begin{array}{l} H_0 : \sigma^2 = \sigma_o^2 \\ H_1 : \sigma^2 < \sigma_o^2 \end{array}}_{\text{Case II}} \quad \underbrace{\begin{array}{l} H_0 : \sigma^2 = \sigma_o^2 \\ H_1 : \sigma^2 \neq \sigma_o^2 \end{array}}_{\text{Case III}}$$

Solution

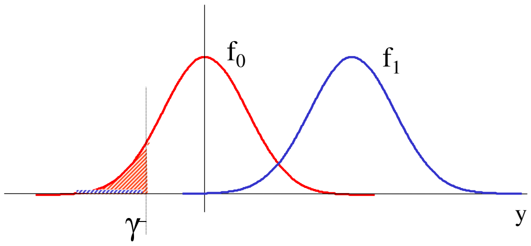


Figure 103: *The single sided MP-LRT for $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$ fails to detect positive signal.*

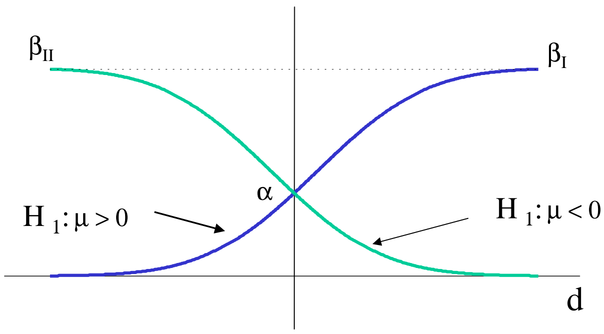


Figure 104: *The power curve of Case I or Case II MP-LRT's for double sided hypotheses is biased over range $-\infty < d < \infty$.*

STEP 1: find MP-LRT for fixed σ^2

Approach 1: work problem directly from entire data sample $\underline{x}$

$$\begin{aligned}
\Lambda(\sigma^2) &= \frac{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^n (x_k-\mu)^2\right)}{\left(\frac{1}{\sqrt{2\pi\sigma_o^2}}\right)^n \exp\left(-\frac{1}{2\sigma_o^2}\sum_{k=1}^n (x_k-\mu)^2\right)} \\
&= \left(\frac{\sigma_o^2}{\sigma^2}\right)^{n/2} \exp\left(\frac{\sigma^2-\sigma_o^2}{2\sigma^2\sigma_o^2}\sum_{k=1}^n (x_k-\mu)^2\right) \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta
\end{aligned}$$

Which is monotone increasing in the quantity

$$(\sigma^2-\sigma_o^2)\sum_{k=1}^n (x_k-\mu)^2$$

Thus we obtain MP-LRT

$$(\sigma^2-\sigma_o^2)\ \frac{n\ \hat{\sigma}_\mu^2}{\sigma_o^2} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

where

$$\hat{\sigma}_\mu^2 = n^{-1}\sum_{k=1}^n (x_k-\mu)^2$$

Approach 2: work problem based on sufficient statistic for σ^2

$$T(\underline{x}) = \sum_{i=1}^n (x_i-\mu)^2$$

The density of $T(\underline{x})/\sigma^2$ is Chi-square with n d.f.

$\Rightarrow$ using standard transformation of variables formula, p.d.f. of T is:

$$\begin{aligned}
f(T;\sigma^2) &= \sigma^{-2}\ f_\chi(T\ \sigma^{-2}) \\
&= \sigma^{-2}\ \frac{1}{2^{n/2}\Gamma(n/2)}\ e^{-T/(2\sigma^2)}(T/\sigma^2)^{n/2-1}
\end{aligned}$$

Hence MP-LRT is

$$\Lambda(\sigma^2) = \left(\frac{\sigma_o^2}{\sigma^2}\right)^{n/2} \exp\left\{\frac{\sigma^2-\sigma_o^2}{2\sigma^2\sigma_o^2}\ T\right\} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

Which is monotone in $(\sigma^2 - \sigma_o^2)\, T$.

Thus we obtain MP-LRT

$$(\sigma^2 - \sigma_o^2)\, \frac{n\, \hat{\sigma}_\mu^2}{\sigma_o^2} \overset{H_1}{\underset{H_0}{\gtrless}} \gamma$$

where again

$$\hat{\sigma}_\mu^2 = n^{-1} \sum_{k=1}^{n} (x_k - \mu)^2$$

CASE I: Single sided alternative $H_1 : \sigma^2 > \sigma_o^2$

In this case MP-LRT is simply

$$T(\underline{x}) = \frac{n\, \hat{\sigma}_\mu^2}{\sigma_o^2} \overset{H_1}{\underset{H_0}{\gtrless}} \gamma^+$$

or equivalently, we have a square law detector

$$T(\underline{x}) = \frac{1}{\sigma_o^2} \sum_{i=1}^{n} (x_i - \mu)^2 \overset{H_1}{\underset{H_0}{\gtrless}} \gamma^+$$

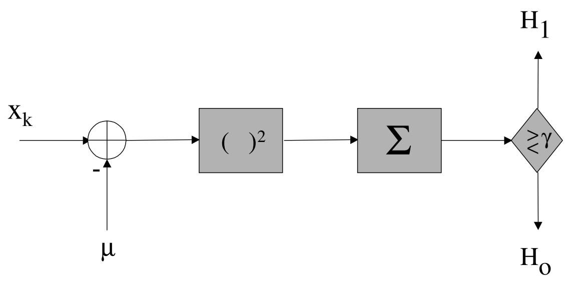


Figure 105: *Optimal detector for increase in Gaussian variance is a memoryless non-linear device (squarer) followed by a summer and decision mechanism.*

Under H_0, $x_k \sim \mathcal{N}(\mu, \sigma_o^2)$ so the test statistic $T(\underline{x})$ is Chi-square with n d.f.

Therefore

$$\alpha = P_0(T(\underline{x}) > \gamma^+) = 1 - \chi_n(\gamma^+)$$

and

$$\gamma^+ = \chi_n^{-1}(1-\alpha)$$

Hence MP-LRT is

$$\frac{n\ \hat{\sigma}_\mu^2}{\sigma_o^2} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \chi_n^{-1}(1-\alpha)$$

which is UMP against any $\sigma^2 > \sigma_o^2$ for known μ.

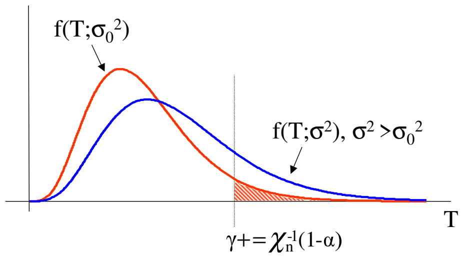


Figure 106: *Density functions under H_0 and H_1 of optimal UMP test statistic for testing against $\sigma^2 > \sigma_o^2$ for known mean μ. Threshold γ^+ is determined by the $1-\alpha$ quantile of the H_0 density.*

Power: since $\frac{\sigma_o^2}{\sigma^2}\ \frac{n\hat{\sigma}_\mu^2}{\sigma_o^2} = \chi_n$

$$\beta = P_1\left(n\ \hat{\sigma}_\mu^2/\sigma_o^2 > \gamma^+\right) = 1 - \chi_n\left(\frac{\sigma_o^2}{\sigma^2}\ \chi_n^{-1}(1-\alpha)\right)$$

CASE II: Single sided alternative $H_1 : \sigma^2 < \sigma_o^2$

Find that MP-LRT has form

$$\frac{n\ \hat{\sigma}_\mu^2}{\sigma_o^2} \begin{array}{c} H_0 \\ > \\ < \\ H_1 \end{array} \gamma^-$$

where now

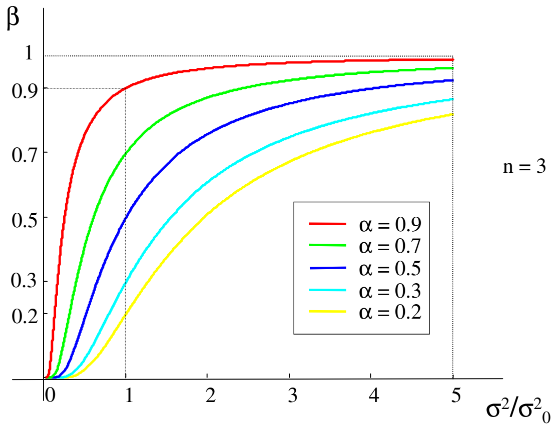


Figure 107: *Power curves for one sided test of variance $\sigma^2 > \sigma_o^2$ for known mean μ with i.i.d. Gaussian observations for various values of σ^2/σ_o^2 and $n = 3$.*

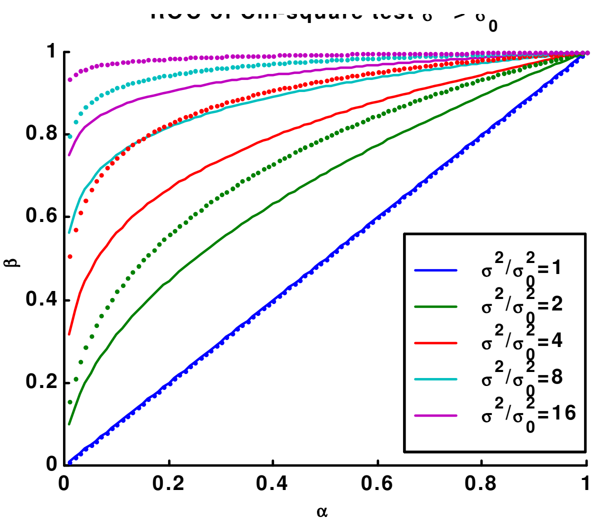


Figure 108: *ROC curves for one sided test of variance $\sigma^2 > \sigma_o^2$ with i.i.d. Gaussian observations for various values of σ^2/σ_o^2 and $n = 3$.*

$$\gamma^- = \chi_n^{-1}(\alpha)$$

So that we have UMP test against $\sigma^2 < \sigma_o^2$ for known μ

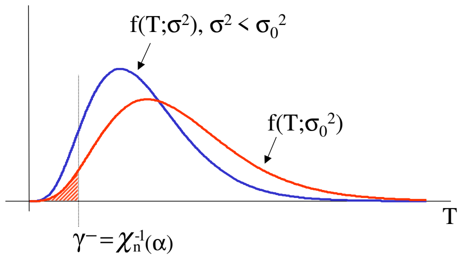


Figure 109: *Density functions under H_0 and H_1 of optimal UMP test statistic for testing against $\sigma^2 < \sigma_o^2$ for known mean μ. Threshold γ^+ is determined by the $1-\alpha$ quantile of the H_0 density.*

Power:

$$\beta = 1 - \chi_n\left(\frac{\sigma_o^2}{\sigma^2}\, \chi_n^{-1}(\alpha)\right)$$

Case III: Double sided alternative $H_1 : \sigma^2 \neq \sigma_o^2$

No UMP exists.

Example 33 *One sided test on median of Cauchy density*

Assume $x_1, \ldots, x_n$ i.i.d. with marginal density

$$f(x_1; \theta) = \frac{1}{\pi} \frac{1}{1 + (x_i - \theta)^2}$$

Objective: investigate existence of UMP test for

$$H_0 : \theta = 0$$
$$H_1 : \theta > 0$$

Step 1: First find LRT for fixed $\theta > 0$

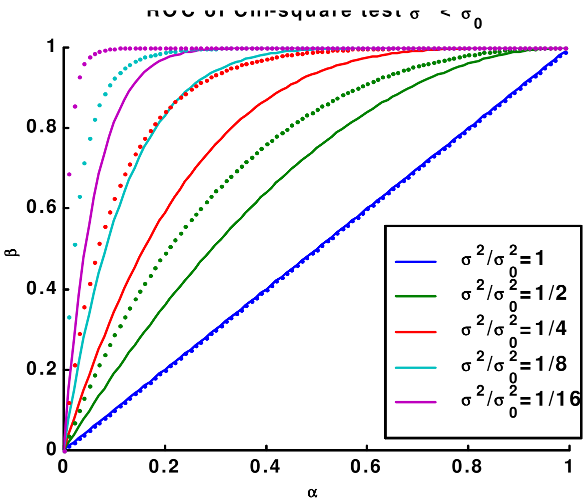


Figure 110: *ROC curves for one sided test of variance* $\sigma^2 < \sigma_o^2$ *with i.i.d. Gaussian observations for various values of* σ^2/σ_o^2 *and* $n = 3$.

$$\Lambda(\theta) = \frac{f(x;\theta)}{f(x;0)} = \prod_{i=1}^{n} \frac{1+x_i^2}{1+(x_i-\theta)^2} \quad \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \quad \eta$$

Special case of a single sample $n = 1$

$$\Lambda(\theta) = \frac{1+x_1^2}{1+(x_1-\theta)^2} \quad \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \quad \eta$$

Decision region depends on θ even if $\theta > 0$.

$\Rightarrow$ no UMP exists

8.2 GENERAL CONDITIONS FOR EXISTENCE OF UMP TESTS

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &\in \Theta_1 \end{aligned}$$

Monotone LR condition (1D parameter θ)

Let $f(x;\theta)$ have Fisher Factorization

$$f(x;\theta) = g(T,\theta)h(x)$$

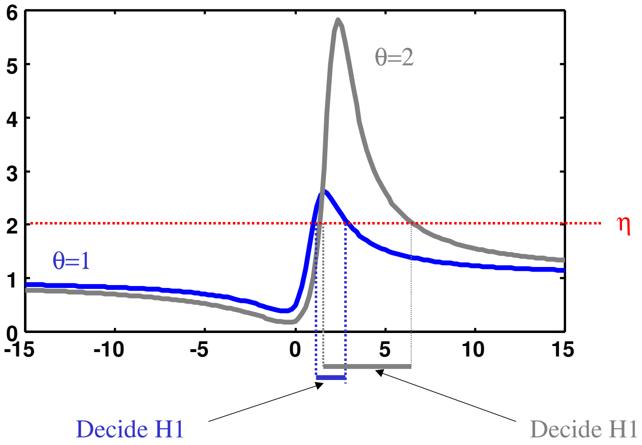


Figure 111: *Plot of LR statistic for two values $\theta = 1, 2$ for Cauchy example and a single sample.*

T a sufficient statistic.

Then an UMP test of any level α exists if the likelihood ratio is either monotone increasing or monotone decreasing in T for all $\theta \in \Theta_1$

$$\Lambda = \frac{f(x;\theta)}{f(x;\theta_0)} = \frac{g(T;\theta)}{g(T;\theta_0)} = \Lambda_\theta(T)$$

To see this note that MP test for $H_1 : \theta = \theta_1$ is

$$\Lambda_{\theta_1}(T) \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \eta$$

which is equivalent to test on T with threshold $\gamma = \Lambda^{-1}_{\theta_1}(\eta)$

$$\begin{array}{cccc} T & \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} & \gamma & \text{(increasing } \Lambda) \\ T & \begin{array}{c} H_0 \\ > \\ < \\ H_1 \end{array} & \gamma & \text{(decreasing } \Lambda) \end{array}$$

Densities satisfying monotone LR condition for one sided H_1:

$$H_1 : \theta > \theta_0, \quad \text{or} \quad H_1 : \theta < \theta_0$$

1. x i.i.d. sample from 1D exponential family,
2. In particular: Gaussian, Bernoulli, Exponential, Poisson, Gamma, Beta

3. x i.i.d. sample from a Uniform density $\mathcal{U}(0, \theta)$

4. x i.i.d. sample from noncentral-t, noncentral Fisher F

5. x i.i.d. sample from shifted Laplace, logistic

In fact: monotone LR condition guarantees MP-LRT is UMP wrt $H_o : \theta < \theta_0$ too!

Situations where monotone LR does not hold

1. Gaussian density with single sided H_1 on mean but having unknown variance

2. Cauchy density with single sided H_1

3. Exponential family with double sided H_1

8.3 COMPOSITE HYPOTHESIS DETECTION STRATEGIES

Here it is desired to test doubly composite

$$H_0 : \theta \in \Theta_0$$

$$H_1 : \theta \in \Theta_1$$

Now, most fixed detectors will have both P_F and P_D varying as functions of $\theta \in \Theta_0$ and $\theta \in \Theta_1$, respectively.

Recall that for composite H_0 we say that test ϕ is of level α if

$$\max_{\theta_0 \in \Theta_0} P_F(\theta_0) \leq \alpha$$

where $P_F(\theta) = E_{\theta_0}[\phi]$

Two classes of strategies:

1. Optimize alternative detection criterion

2. Constrain form of detector to a class for which UMP may exist

8.4 MIN-MAX NP DETECTION

Objective: find level α test which maximizes worst case power

Constraint:

$$\max_{\theta \in \Theta_0} E_\theta[\phi] \leq \alpha$$

Maximize:

$$\min_{\theta \in \Theta_1} E_\theta[\phi]$$

METHOD OF SOLUTION: find “least favorable” densities

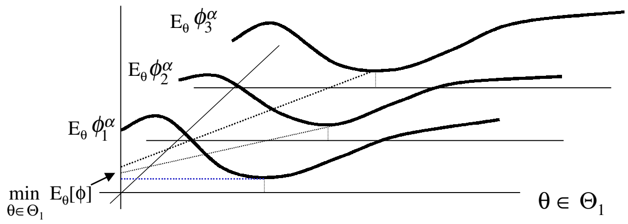


Figure 112: *Various power curves for different test functions and their minima over the unknown parameter θ varying over H_1 parameter space Θ_1. Minimax NP test ϕ_3^α maximizes minimum power.*

Simplifying assumption: Θ discrete parameter space:

Fundamental identity on the mean [7]:

For any summable sequence $\{a(k)\}_k$ and any probability distribution $\{p(k)\}_k$

$$\min_k a(k) \leq \sum_k a(k)p(k) \leq \max_k a(k)$$

with equality when $p(k) =$ delta function concentrated on $\mathrm{argmin}_k a(k)$ and $\mathrm{argmax}_k a(k)$.

and therefore

$$\min_k a(k) = \min_{\{p(k)\}} \sum_k a(k)p(k), \quad \max_k a(k) = \max_{\{p(k)\}} \sum_k a(k)p(k)$$

* Let $\{p_0(\theta)\}$ be an arbitrary probability distribution on Θ_0

* Let $\{p_1(\theta)\}$ be an arbitrary probability distribution on Θ_1

Then the worst case $P_F(\theta)$ and $P_M(\theta)$ can be expressed as worst case average P_F and P_M

$$\max_{\theta\in\Theta_0} E_\theta[\phi] = \max_{p_0} \sum_{\theta\in\Theta_0} E_\theta[\phi]p_0(\theta) = \sum_{\theta\in\Theta_0} E_\theta[\phi]p_0^*(\theta)$$

$$\min_{\theta\in\Theta_1} E_\theta[\phi] = \min_{p_1} \sum_{\theta\in\Theta_1} E_\theta[\phi]p_1(\theta) = \sum_{\theta\in\Theta_1} E_\theta[\phi]p_1^*(\theta)$$

where

* p_0^* maximizes the false alarm probability

* p_1^* minimizes the detection probability (power)

Define "least favorable pair" of densities

$$\begin{aligned} f_0^*(x) &= \sum_{\theta \in \Theta_0} f(x;\theta) p_0^*(\theta) \\ f_1^*(x) &= \sum_{\theta \in \Theta_1} f(x;\theta) p_1^*(\theta) \end{aligned}$$

Then the minmax objective reduces to

Constraint:

$$E_0^*[\phi] = \int_{\mathcal{X}} \phi(x)\ f_0^*(x) dx \leq \alpha$$

Maximize:

$$E_1^*[\phi] = \int_{\mathcal{X}} \phi(x)\ f_1^*(x) dx$$

Which, corresponds to finding NP MP test of level α for the derived simple hypotheses

$$H_0^* : x \ \sim \ f_0^*$$

$$H_1^* : x \ \sim \ f_1^*$$

Hence min-max NP test is the LRT

$$\frac{f_1^*(x)}{f_0^*(x)} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \eta$$

where threshold η is chosen to satisfy:

$$\int_{\mathcal{X}} \phi^*(x) f_0^*(x) dx = \alpha$$

Observations

* Minmax NP test is an optimal Bayes test for random θ over Θ_1 and Θ_0 but without prior probabilities on H_0 and H_1.

* Performance of minimax NP test can be overly conservative, especially if least favorable proirs concentrate on atypical values of θ.

* Least favorable priors p_1^*, p_0^* may be difficult to find in practice

⇒ Helpful facts concerning f_1^*, f_0^* [4]:

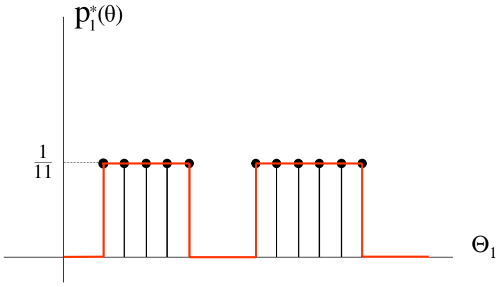


Figure 113: *Least favorable density $p_1^*(\theta)$ is piecewise constant over Θ_1.*

o p_0^* and p_1^* make H_0^* and H_1^* the most difficult to discriminate

o p_1^* and p_0^* can each assume at most two values over Θ_1 and Θ_0

Specifically, there exists a subset Θ_0^+ of Θ_0 such that

$$p_0^*(\theta) = \begin{cases} q, & \theta \in \Theta_0^+ \\ 0, & \theta \in \Theta_0 - \Theta_0^+ \end{cases}$$

where q is equal to the volume of Θ_0^+

$$q = \begin{cases} \int_{\Theta_0^+} d\theta, & \Theta_0 \text{ cts.} \\ \sum_{\theta \in \Theta_0^+}, & \Theta_0 \text{ discrete} \end{cases}$$

and similarly for p_1^*.

Examples of min-max test: to be given in section on simultaneous detection, estimation, and classification.

8.5 LOCALLY MOST POWERFUL (LMP) SINGLE SIDED TEST

Main idea: find MP test for detecting small perturbation from H_0

First consider single sided case and 1D parameter θ

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &> \theta_0 \end{aligned}$$

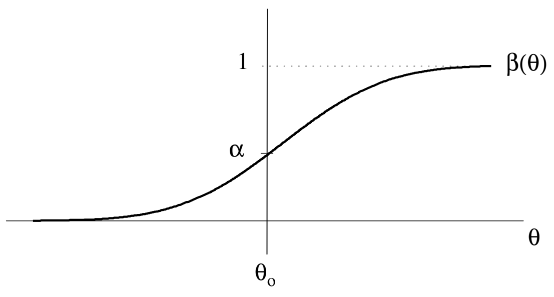


Figure 114: *Typical power curve $\beta(\theta)$. LMP test of $H_1 : \theta > \theta_o$ seeks to maximize the slope of the power curve at point $\theta = \theta_o$.*

Definition 2 *Locally most powerful test ϕ of level α has power curve which maximizes slope of $\beta(\theta)$ at $\theta = \theta_0$*

Constraint: $E_{\theta_0}[\phi] \leq \alpha$

Maximize: $\frac{d}{d\theta_0} E_{\theta_0}[\phi]$

Similarly to derivation of NPL, we obtain differential LRT

$$\phi_{\mathrm{LMP}} = \begin{cases} 1, & df(x;\theta_0)/d\theta_0 > \eta f(x;\theta_0) \\ q, & df(x;\theta_0)/d\theta_0 = \eta f(x;\theta_0) \\ 0, & df(x;\theta_0)/d\theta_0 < \eta f(x;\theta_0) \end{cases}$$

Or for short

$$\Lambda_{LMP} = \frac{df(x;\theta_0)/d\theta_0}{f(x;\theta_0)} \overset{H_1}{\underset{H_0}{\gtrless}} \eta$$

where η is selected to satisfy constraint (possibly with randomization)

$$E_{\theta_0}[\phi] \leq \alpha$$

Connection between LMP and maximum likelihood estimation:

Can write LMP test in equivalent form

$$\Lambda_{LMP} = \frac{d}{d\theta_o} \ln f(x;\theta_0) \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

Thus we decide H_0 if θ_0 is near a stationary point, e.g. the MLE, of $f(x;\theta)$ over Θ.

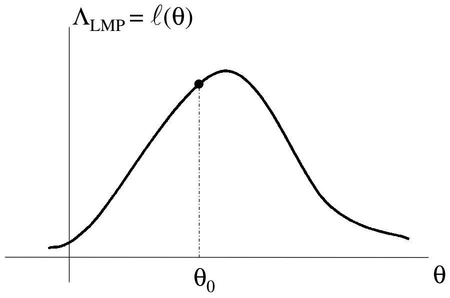


Figure 115: *LMP test $H_1 : \theta > \theta_o$ decides H_0 if θ_0 is near stationary point of log likelihood function $l(\theta)$.*

Example 34 *Gaussian one sided test against zero mean*

Find: differential LR has the form

$$\begin{aligned} \frac{df(x;\theta)/d\theta}{f(x;\theta)} &= \frac{d}{d\theta} \ln f(x;\theta) \\ &= \frac{\sum_{i=1}^n (x_i - \theta)}{\sigma^2} \end{aligned}$$

LMP for testing $\theta = \theta_0 = 0$ vs. $\theta > 0$ is therefore:

$$\sum_{i=1}^n x_i \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

or level α LMP is the linear UMP test obtained before

$$\frac{\sqrt{n}\, \overline{x_i}}{\sigma} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

Example 35 *Cauchy one sided test against zero median (ctd)*

Find: differential LR has the form

$$\frac{df(x;\theta)/d\theta}{f(x;\theta)} = 2\sum_{i=1}^n \frac{x_i-\theta}{1+(x_i-\theta)^2}$$

For $\theta_0 = 0$ LMP test is therefore:

$$T(\underline{x}) = \sum_{i=1}^n \frac{x_i}{1+x_i^2} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma$$

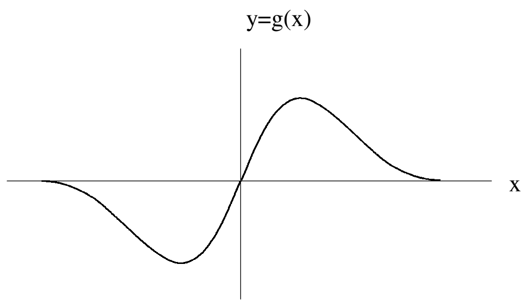


Figure 116: *Memoryless non-linearity $g(x) = x/(1+x^2)$ input-output characteristic for LMP test of one sided test against zero median for a Cauchy r.v.*

Test statistic $T(\underline{x})$ is sum of i.i.d. r.v.s with mean 0 and variance $1/8$ under H_0. Therefore threshold γ can be found via CLT for large n:

$$\gamma = \sqrt{n/8}\,\mathcal{N}^{-1}(1-\alpha)$$

Example 36 *Testing for positive mean of Laplace distribution*

* $\underline{x} = [x_1, \ldots, x_n]$ i.i.d,

$$x_i \sim f(x;\theta) = \frac{a}{2}e^{-a|x-\theta|}, \quad a > 0$$

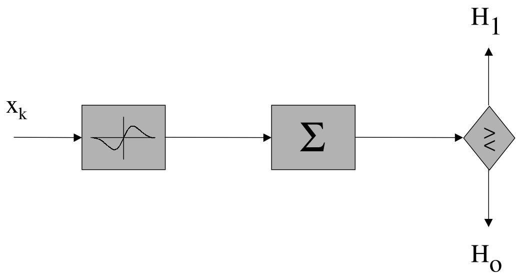


Figure 117: *Optimal detector for positive Cauchy median is a memoryless non-linearity followed by a summer and decision mechanism.*

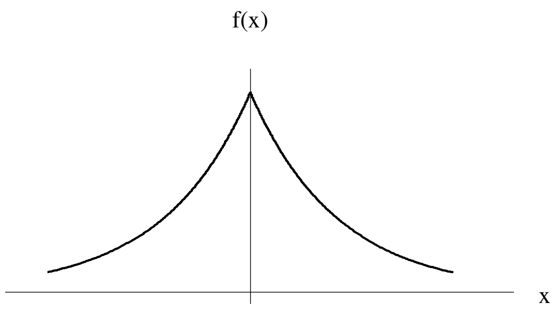


Figure 118: *Laplace density* $f(x) = ae^{-a|x-\theta|}/2$*. Width, as measured by where* $f(x)$ *falls to* $1/e$ *of its peak, is* $2/a$*.*

Log-likelihood function takes the form:

$$\begin{aligned}\ln f(\underline{x};\theta) &= -a\sum_{i=1}^{n}|x_i-\theta| + n\ln\frac{a}{2} \\ &= -a\sum_{x_i>\theta}(x_i-\theta) + a\sum_{x_i<\theta}(x_i-\theta) + c \\ &= a\theta\,(n_+ - n_-) + b(\theta)\end{aligned}$$

where

$$n_+ = \,\#\, x_i > \theta, \qquad n_- = \,\#\, x_i < \theta = n - n_+$$

Note: $b(\theta)$ is piecewise constant function

Find: differential LR has the form

$$\frac{df(x;\theta)/d\theta}{f(x;\theta)} = a(n_+ - n_-)$$

LMP is therefore:

$$T(\underline{x}) = n_+ - n_- \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

or equivalently, in a form more comparable to Cauchy and Gaussian example

$$T(\underline{x}) = \sum_{i=1}^{n}\mathrm{sgn}(x_i) \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

PERFORMANCE:

$T(\underline{x})$ is a discrete shifted Binomial r.v.

$$T(\underline{x}) = \sum_{i=1}^{n}(2b_i - 1) = 2B(n,p) - n$$

where b_i are i.i.d. Bernoulli r.v.'s with parameter

$$p = P_\theta(b_i = 1) = P_\theta(x_i > 0)$$

$\Rightarrow$ Randomized test is necessary to set false alarm.

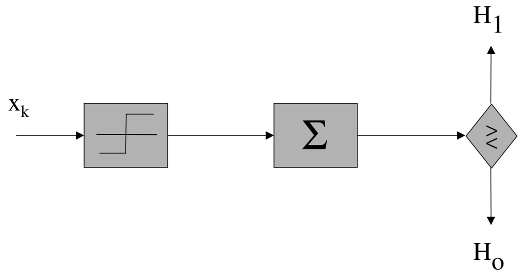


Figure 119: *LMP detector for testing positive mean $\theta > 0$ for a Laplace r.v. is composed of a summer and memoryless non-linearity.*

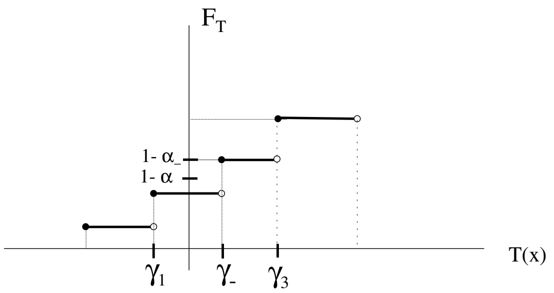


Figure 120: *The CDF of test statistic is step function (value of F_T over $\gamma_1 \leq T(x) < \gamma_-$ is $1 - \alpha_+$). Randomization is necessary for meeting FA constraint.*

$$\alpha = P_0(T(\underline{x}) > \gamma_-) + q(\alpha_+ - \alpha_-)$$

where α_- and γ_- are related by

$$\alpha_- = P_0(\underbrace{T(\underline{x})}_{2B(n,\frac{1}{2})-n} > \gamma_-) = 1 - B_{n,p}\left(\frac{\gamma_- + n}{2}\right)$$

and the randomization paremeter q is as usual

$$q = \frac{\alpha - \alpha_-}{\alpha_+ - \alpha_-}$$

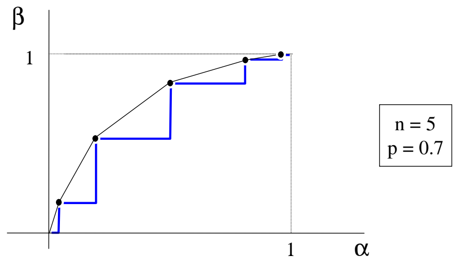


Figure 121: *ROC curve of LMP detector for testing positive mean $\theta > 0$ for a Laplace r.v.*

8.6 MOST POWERFUL UNBIASED (MPU) TESTS

Recall: a test ϕ of level α is an unbiased test if

$$E_\theta[\phi] \geq \alpha, \qquad \text{all } \theta \in \Theta_1$$

A test ϕ of level α is uniformly MPU (UMPU) if for all $\theta \in \Theta_1$ its power function dominates that of all other unbiased tests of level α.

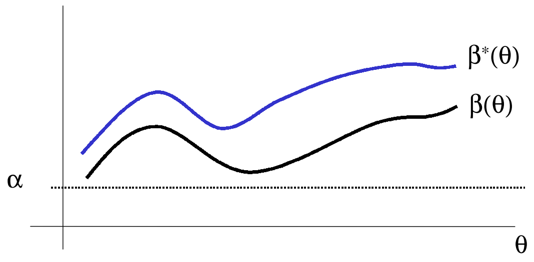


Figure 122: *Power curve of most powerful unbiased test (MPU) dominates that of all other unbiased tests of the same FA level.*

8.7 LOCALLY MOST POWERFUL UNBIASED DOUBLE SIDED TEST

Consider double sided hypotheses:

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &\neq \theta_0 \end{aligned}$$

Observe: Power function of a good unbiased level α test ϕ should have global minimum at $\theta = \theta_0$.

Locally unbiased test optimization for 1D parameter θ

Constraints:

$$E_{\theta_0}[\phi] \leq \alpha, \quad \frac{d}{d\theta_0} E_{\theta_0}[\phi] = 0. \tag{36}$$

Subject to these constraints want to maximize curvature at θ_0

$$\frac{d^2}{d\theta_0^2} E_{\theta_0}[\phi].$$

Using Lagrange multipliers it is easily shown that the test function ϕ which solves this constrained maximization problem has the form:

$$\phi(x) = \begin{cases} 1, & d^2 f(x;\theta_0)/d\theta_0^2 > \eta(f(x;\theta_0) + \rho\, df(x;\theta_0)/d\theta_0) \\ q & d^2 f(x;\theta_0)/d\theta_0^2 = \eta(f(x;\theta_0) + \rho\, df(x;\theta_0)/d\theta_0) \\ 0 & d^2 f(x;\theta_0)/d\theta_0^2 < \eta(f(x;\theta_0) + \rho\, df(x;\theta_0)/d\theta_0) \end{cases} \tag{37}$$

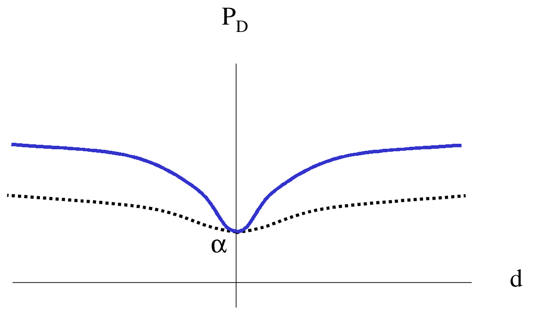


Figure 123: *Power curve of a good locally unbiased test has minimum at α with maximum curvature.*

where ρ, η, q are selected to satisfy the two constraints.

In some cases, one can meet the constraints by selecting $\rho = 0$ and varying only $q \in [0,1]$ and $\eta \in [0, \infty)$. In this situation, the locally optimal test (37) reduces to the simpler (randomized) LRT form

$$\frac{d^2 f(x;\theta_0)/d\theta_0^2}{f(x;\theta_0)} \overset{H_1}{\underset{H_0}{\gtrless}} \eta.$$

Example 37 *Double sided test against zero mean of Gaussian sample with known variance*

Step 1: Find derivatives of p.d.f. of sufficient statistic $\overline{x_i}$

$$\begin{aligned} f(\overline{x_i};\mu) &= \frac{1}{\sqrt{2\pi\sigma^2/n}} e^{-\frac{(\overline{x_i}-\mu)^2}{2\sigma^2/n}} \\ df(\overline{x_i};\mu)/d\mu &= n/\sigma^2 \;\; (\overline{x_i}-\mu) f(x;\mu) \\ d^2 f(\overline{x_i};\mu) d\mu^2 &= n/\sigma^2 \;\; [n/\sigma^2(\overline{x_i}-\mu)^2 - 1] f(x;\mu) \end{aligned}$$

Thus LMPU LRT is

$$\frac{\overline{x_i}^2 - \sigma^2/n}{\sigma^2/n + \rho\overline{x_i}} \overset{H_1}{\underset{H_0}{\gtrless}} \eta$$

Step 2: Select ρ, η to satisfy constraints

First we attempt to satisfy constraints with $\rho = 0$ and η a free variable.

For this case LMPU LRT reduces to

$$|\overline{x_i}| \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma \tag{38}$$

Since

$$\overline{x_i} \sim \mathcal{N}(0, \sigma^2/n), \quad \text{under } H_0$$

we have

$$\alpha = 1 - P_0(-\gamma < \overline{x_i} \le \gamma) = 2(1 - \mathcal{N}(\gamma\sqrt{n}/\sigma))$$

Or

$$\gamma = \frac{\sigma}{\sqrt{n}}\, \mathcal{N}^{-1}(1 - \alpha/2)$$

Success! We can set threshold γ to achieve arbitrary $P_F = \alpha$ with $\rho = 0$ and without randomization. Of course it still must be verified that the test (38) sataisfies the second constraint in (36) which is that $d/d\mu P_D(\mu)|_{\mu=0} = 0$. This is left as an exercise.

Equivalent form of locally-unbiased test

$$\frac{1}{n}\,|\sum_{i=1}^{n} x_i| \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma$$

Power:

Since

$$\overline{x_i} \sim \mathcal{N}(\mu, \sigma^2/n), \quad \text{under } H_1$$

$$\begin{aligned} P_D &= 1 - P_\mu(-\gamma < \overline{x_i} \le \gamma) \\ &= 1 - [\mathcal{N}(\sqrt{n}\,(\gamma - \mu)/\sigma) - \mathcal{N}(\sqrt{n}\,(-\gamma - \mu)/\sigma)] \end{aligned}$$

$$= 1 - \mathcal{N}(\mathcal{N}^{-1}(1 - \alpha/2) - d) - \mathcal{N}(\mathcal{N}^{-1}(1 - \alpha/2) + d)$$

where as usual:

* $d = \sqrt{n}\,\mu/\sigma$ is detectability index

Remark:

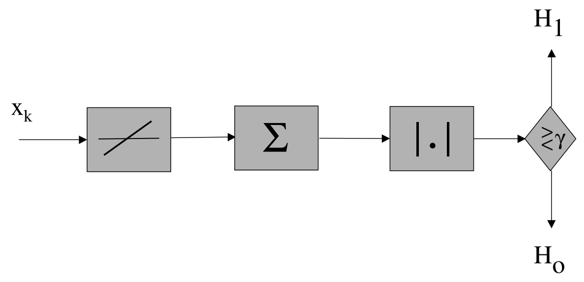


Figure 124: *Locally best unbiased double-sided test for non-zero Gaussian mean is a memoryless non-linearity followed by a summer and decision device.*

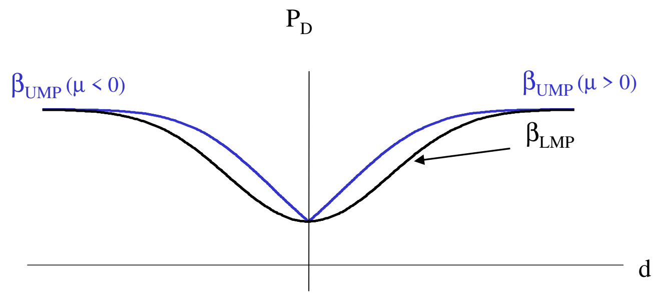


Figure 125: *Power curve of LMPU test for for non-zero Gaussian mean with known variance as a function of values of d.*

* It can be shown that in the Gaussian example above the LMPU test is actually UMPU. See Ferguson [4].

LMPU strategy can easily be extended to multiple parameters:

Assume

$$\underline{\theta} = [\theta_1, \ldots, \theta_p]^T$$

and let's test hypotheses:

$$H_0 : \underline{\theta} = \underline{\theta}_0$$
$$H_1 : \underline{\theta} \neq \underline{\theta}_0$$

Constraints:

$$E_{\theta_0}[\phi] \leq \alpha, \quad \|\nabla_{\theta_0} E_{\theta_0}[\phi]\| = 0$$

Maximize:

$$\text{trace}\left\{\nabla^2_{\theta_0} E_{\theta_0}[\phi]\right\}$$

where trace $\{A\}$ denotes trace of matrix A.

Optimal test can immediately be found via application of NPL

$$\frac{\text{trace}\left\{\nabla^2_{\theta_0} f(x;\theta_0)\right\}}{f(x;\theta_0) + \rho\,\|\nabla_{\theta_0} f(x;\theta_0)\|} \overset{H_1}{\underset{H_0}{\gtrless}} \eta$$

where ρ, η are selected to satisfy the two constraints (possibly with randomization).

8.8 CFAR DETECTION

A sometimes reasonable condition is to require that tests have constant false alarm probability over Θ_0.

The objective is to find a UMP CFAR test:

Constraint:

$$E_\theta[\phi] = \alpha, \quad \theta \in \Theta_0$$

Maximize:

$$E_\theta[\phi], \quad \theta \in \Theta_1$$

8.9 INVARIANT TESTS

Consider general case where we have partition of $\underline{\theta}$

$$\underline{\theta} = [\varphi_1, \ldots, \varphi_p, \quad \underbrace{\xi_1, \ldots, \xi_q}_{\text{nuisance parameters}} \quad]^T$$

and $x \sim f(x; \underline{\varphi}, \underline{\xi})$

It is desired to test single sided hypotheses

$$H_0 : \underline{\varphi} = 0, \quad \underline{\xi} = \underline{\xi}_o$$

$$H_1 : \underline{\varphi} > 0, \quad \underline{\xi} = \underline{\xi}_o$$

where $\underline{\xi}_o \in \mathbb{R}^q$ is *unknown* $\Rightarrow$ UMP does not usually exist.

Invariant tests seek to find a transformation (compression) of the data

$$Z = Z(x)$$

which satisfies:

Property 1. $Z(x)$ contains (almost) as much information concerning φ as x

Property 2. Distribution of $Z(x)$ is not a function of $\underline{\xi}$.

Due to Property 2, if we throw away x and retain only

$Z \sim f(Z; \varphi)$

then we are back to testing simpler hypotheses for which an UMP may exist

$$H_0 : \underline{\varphi} = 0$$
$$H_1 : \underline{\varphi} > 0$$

We will treat theory of optimal invariant tests in detail in a section to follow in the context of testing for presence of signal φ when entries of noise covariance matrix $\underline{\xi}$ are unknown.

For now we concentrate on a particular suboptimal "invariant" approach

8.9.1 GENERALIZED LIKELIHOOD RATIO TEST

General composite hypotheses

$$H_0 : \theta \in \Theta_0$$
$$H_1 : \theta \in \Theta_1$$

General ad hoc "estimate-and-plug" procedure to test H_0 vs. H_1:

Step 1: Find good estimates $\hat{\theta}_0$ and $\hat{\theta}_1$ of θ under H_0 and H_1

Step 2: substitute these estimates into the LR statistic

Obtain the GLRT

$$\Lambda = \frac{f(x;\hat{\theta}_1)}{f(x;\hat{\theta}_0)} \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \eta$$

where η is selected to give FA level α

Most common case:

$\Rightarrow$ $\hat{\theta}_1$ and $\hat{\theta}_0$ are "maximum likelihood" estimates:

$$\Lambda_{GLR} = \frac{\max_{\theta\in\Theta_1} f(x;\theta)}{\max_{\theta\in\Theta_0} f(x;\theta)} \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \eta$$

Note: the MLE's are conditioned on $\theta \in \Theta_0$ and $\theta \in \Theta_1$, respectively.

For case of simple H_0 this reduces to

$$\begin{aligned} \Lambda_{GLR} &= \frac{\sup_{\theta\in\Theta_1} f(x;\theta)}{f(x;\theta_0)} \\ &= \sup_{\theta\in\Theta_1} \Lambda(\theta) = \Lambda(\hat{\theta}_1) \end{aligned}$$

where

$$\Lambda(\theta) = \frac{f(x;\theta)}{f(x;\theta_0)}$$

8.9.2 ASYMPTOTICS OF GLRT

1. Since MLE $\hat{\theta}$ is consistent estimator as $n \to \infty$

$\Rightarrow$ GLRT is asymptotically UMP

2. GLR statistic has Chi-square limiting distribution:

Assume the partition

$$\underline{\theta} = [\varphi_1, \ldots, \varphi_p, \quad \underbrace{\xi_1, \ldots, \xi_q}_{\text{nuisance parameters}} \quad]^T$$

and consider the GLRT for double sided hypotheses

$$H_0 : \underline{\varphi} = \underline{\varphi}_0, \quad \underline{\xi} = \underline{\xi}_o$$

$$H_1 : \underline{\varphi} \neq \underline{\varphi}_0, \quad \underline{\xi} = \underline{\xi}_o$$

where $\underline{\xi}_o$ is unknown
FACT: when the density $f(x; \underline{\theta})$ is smooth under H_0 it can be shown that for large n

$$2 \ln \Lambda_{GLR} \sim \chi_p, \quad \text{under } H_0$$

(See Bickel and Doksum [2])

8.10 Exercises

1. The observations $\{x_i\}_{i=1}^n$ are i.i.d. exponential $x_i \sim f_\theta(x) = \beta e^{-\beta x}$, where $x, \beta \geq 0$. Consider testing the following single sided hypotheses

$$H_0 : \beta = \beta_0$$
$$H_1 : \beta > \beta_0$$

   (a) First find the MP test of level α for the simple alternative $H_1 : \beta = \beta_1$ where $\beta_1 > \beta_0$. Express the threshold in terms of the Gamma distribution (distribution of n i.i.d. exponential r.v.s). Next establish that your test is UMP for the single sided composite H_1 above.
   (b) Specialize the results of (a) to the case of a single observation $n = 1$ and derive the ROC curve. Plot your curve for $\beta_1/\beta_0 = 1, 5, 10$.
   (c) Derive the locally most powerful test (LMPT) for the single sided hypotheses (maximize slope of power subject to FA constraint) and verify that it is identical to the UMP test.
   (d) Now consider testing the double sided hypotheses

$$H_0 : \beta = \beta_0$$
$$H_1 : \beta \neq \beta_0$$

   Derive the LMPT (maximize curvature of power subject to FA constraint and zero slope condition). Derive the ROC for $n = 1$ and compare to the ROC of part (b) over the region $\beta > \beta_0$.
   (e) Derive the GLRT for the double sided hypotheses of part (d). Compare to your answer obtained in part (d).

2. Let Z be a single observation having density function

$$p_\theta(z) = (2\theta z + 1 - \theta), \qquad 0 \leq z \leq 1$$

   where $-1 \leq \theta \leq 1$.

(a) is there a uniformly most powerful test between the composite hypotheses

$$H_0 : \theta = 0$$
$$H_1 : \theta \neq 0$$

If so what is it?

(b) Find the generalized likelihood ratio test for these hypotheses.

(c) Now assume that under H_1 the parameter θ has prior density $p(\theta) = |\theta| I_{[-1,1]}(\theta)$ so that under H_1 the density of Z is $f_1(z) = \int f(z|\theta) p_\theta(\theta) d\theta$, where $f(z|\theta) = p_\theta(z)$. Find the MP test between hypotheses H_0 and this new H_1. What if the prior density were the assymetric $p(\theta) = \frac{1}{2}(\theta + 1) I_{[-1,1]}(\theta)$?

3. A random variable X has density

$$f(x; \theta) = \frac{1 + \theta x}{2}, \quad -1 \leq x \leq 1$$

where $\theta \in [-1, 1]$.

(a) Find the MP test of level α for testing the simple hypotheses

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned}$$

based on a single sample x, where $\theta_0 \in [-1, 0]$ and $\theta_1 \in (0, 1]$ are known. Derive and plot the ROC when $\theta_0 = 0$.

(b) Is there a UMP test of level α, and if so what is it, for the following hypotheses?

$$\begin{aligned} H_0 &: \theta = 0 \\ H_1 &: \theta > 0 \end{aligned}$$

(c) Now consider testing the doubly composite hypotheses

$$\begin{aligned} H_0 &: \theta \leq 0 \\ H_1 &: \theta > 0 \end{aligned}$$

Find the GLRT for the above hypotheses. Derive the threshold of the GLRT that ensures the level α condition $\max_{\theta \in [-1,0]} P_{FA}(\theta) \leq \alpha$.

4. Available is an i.i.d. sample of a Poisson r.v. with distribution $p_\theta(k) = P_\theta(x_i = k) = \frac{\theta^k}{k!} e^{-\theta}$, $k = 0, 1, 2, \ldots$.

(a) Find the GLRT for testing the hypotheses

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

Do not attempt to set the exact threshold for level α.

In the following parts of this exercise you will show how to set the GLRT threshold under the large n Chi-square approximation to the GLRT test statistic $\Lambda = max_{\theta \neq \theta_o} p_\theta(\underline{x}) / p_{\theta_o}(\underline{x})$.

(b) Directly show that under H_0 the statistic $2\log\Lambda$ is asymptotically Chi-square with 1 d.f. by expanding $\Lambda = \Lambda(\overline{x_i})$ about the sample mean $\overline{x_i} = \theta_o$, neglecting all terms of order $(\overline{x_i} - \theta_o)^3$ and higher, and recalling that $(\mathcal{N}(0,1))^2$ is Chi-square with 1 d.f.

(c) Using the result of part (b) set the threshold of your GLRT in part (a).

(d) Using the asymptotic results of part (b) find the GLRT between

$$H_0 : \theta \le \theta_0$$
$$H_1 : \theta > \theta_0$$

with threshold.

5. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with the marginal density $X_i \sim f(x) = \epsilon g(x) + (1-\epsilon)h(x)$, where $\epsilon \in [0,1]$ is a non-random constant and $g(x)$ and $h(x)$ are known density functions. It is desired to test the composite hypotheses

$$H_0 \;:\; \epsilon = 1/2 \tag{39}$$
$$H_1 \;:\; \epsilon > 1/2 \tag{40}$$

(a) Find the most powerful (MP) test between H_0 and the simple hypothesis $H_1 : \epsilon = \epsilon_1$, where $\epsilon_1 > 1/2$ (you needn't solve for the threshold). Is your MP test a UMP test of the composite hypotheses (40)?

(b) Find the locally most powerful (LMP) test for (40). Show how you can use the CLT to set the threshold for large n.

(c) Find the generalized LRT (GLRT) test for (40) in the case of $n = 1$. Compare to your answer in part (b).

End of chapter

9 GLRT FOR THE I.I.D. GAUSSIAN MODEL

* Tests on mean of a single population: σ^2 known
* Tests on mean of a single population: σ^2 unknown
* Tests on variance of a single population: μ known
* Tests on variance of a single population: μ unknown
* Tests on equality of means in two populations
* Tests on equality of variances in two populations
* Tests on correlation between two populations

References:

Bickel and Doksum [2]

Morrison [20]

Recall the form of the density of an i.i.d. Gaussian vector $\underline{x} = [x_1, \ldots, x_n]^T$ with mean μ and variance σ^2.

$$f(\underline{x};\ \mu, \sigma) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right)$$

9.1 TESTS ON THE MEAN: σ^2 KNOWN

Case I: $H_0 : \mu = \mu_o$, $H_1 : \mu > \mu_o$

Case II: $H_0 : \mu \leq \mu_o$, $H_1 : \mu > \mu_o$

Case III: $H_0 : \mu = \mu_o$, $H_1 : \mu \neq \mu_o$

We have already established that UMP test exists for Case I. Can show that same test is UMP for case II by checking monotone likelihood condition [15].

9.1.1 CASE III: $H_0 : \mu = \mu_o$, $H_1 : \mu \neq \mu_o$

$\underline{x} = [x_1, \ldots, x_n]^T$ i.i.d., $x_i \sim \mathcal{N}(\mu, \sigma^2)$

Here $\theta = \mu$, $\Theta = \mathbb{R}$ and we want to test

$$\begin{aligned} H_0 : \mu &= \mu_o \\ H_1 : \mu &\neq \mu_o \end{aligned}$$

where μ_o is fixed and known.

$$\Lambda_{GLR} = \max_{\mu \neq \mu_o} \Lambda(\mu) = \frac{\max_{\mu \neq \mu_o} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_o)^2\right)}$$

Must consider two cases

1. $\hat{\mu}_{ml} \neq \mu_o$

2. $\hat{\mu}_{ml} = \mu_o$

Case 1. $\hat{\mu}_{ml} \neq \mu_o$:

In this case it is obvious that

$$\max_{\mu \neq \mu_o} \Lambda(\mu) = \Lambda(\hat{\mu}_{ml})$$

Case 2. $\hat{\mu}_{ml} = \mu_o$:

Since $\Lambda(\mu)$ is a continuous function with maximum at $\mu = \hat{\mu}_{ml}$ we have again

$$\max_{\mu \neq \mu_o} \Lambda(\mu) = \lim_{\epsilon \to 0} \Lambda(\hat{\mu}_{ml} + \epsilon) = \Lambda(\hat{\mu}_{ml})$$

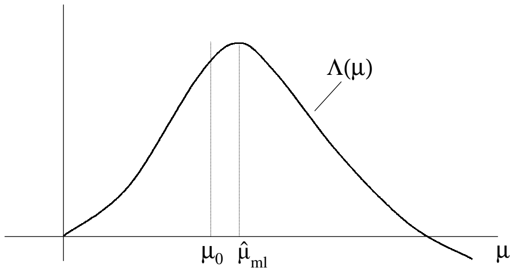


Figure 126: *For a continuous LR density $f(x;\mu)$ the maximum of the likelihood ratio test statistic $\lambda(\mu)$ occurs at the MLE $\mu = \hat{\mu}_{ml}$*

Thus, since we know $\hat{\mu}_{ml}$ is sample mean for Gaussian model

$$\Lambda_{GLR} = \frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^n (x_j - \overline{x_i})^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^n (x_j - \mu_o)^2\right)}$$

Now $\sum_{i=1}^n (x_i - \overline{x_i}) = 0$ ($\overline{x_i}$ is best constant LLS estimator!)

$$\sum_{j=1}^n (x_j - \mu_o)^2 = \sum_{j=1}^n (x_j - \overline{x_i})^2 + n(\overline{x_i}^2 - \mu_o)^2$$

Hence,

$$\Lambda_{GLR} = \exp\left(\frac{n}{2\sigma^2}(\overline{x_i} - \mu_o)^2\right)$$

and GLRT is simply

$$\frac{\sqrt{n}\ |\overline{x_i} - \mu_o|}{\sigma} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma = \mathcal{N}^{-1}(1-\alpha/2)$$

Identical to the LMPU (UMPU) test!

Note: as predicted by result on asymptotic distribution of GLRT

$$\begin{aligned} 2\ln\Lambda_{GLR} &= 2\ln\left\{\exp\left(\frac{n}{2\sigma^2}(\overline{x_i} - \mu_o)^2\right)\right\} \\ &= \left(\underbrace{\frac{\overline{x_i} - \mu_o}{\sigma/\sqrt{n}}}_{\mathcal{N}(0,1)}\right)^2 \end{aligned}$$

which is distributed as a central Chi-square with 1 d.f.

A general lesson learned for GLRT's:

$\Rightarrow$ for testing double sided hypotheses of form

$$\begin{aligned} H_0 : \underline{\theta} &= \underline{\theta}_o \\ H_1 : \underline{\theta} &\neq \underline{\theta}_o \end{aligned}$$

if LR $\Lambda(\underline{\theta})$ is a continous function of $\underline{\theta}$ then

$$\max_{\underline{\theta} \neq \underline{\theta}_o} \Lambda(\underline{\theta}) = \max_{\underline{\theta}} \Lambda(\underline{\theta}) = \Lambda(\hat{\underline{\theta}}_{ml})$$

9.2 TESTS ON THE MEAN: σ^2 UNKNOWN

Case I: $H_0 : \mu = \mu_o,\ \sigma^2 > 0$, $H_1 : \mu > \mu_o,\ \sigma^2 > 0$

Case II: $H_0 : \mu \leq \mu_o,\ \sigma^2 > 0$, $H_1 : \mu > \mu_o,\ \sigma^2 > 0$

Case III: $H_0 : \mu = \mu_o,\ \sigma^2 > 0$, $H_1 : \mu \neq \mu_o,\ \sigma^2 > 0$

9.2.1 CASE I: $H_0 : \mu = \mu_o,\ \sigma^2 > 0$, $H_1 : \mu > \mu_o,\ \sigma^2 > 0$

From properties of the MLE for Gaussian mean and variance we can easily show

$$\begin{aligned}
\Lambda_{GLR} &= \frac{\max_{\mu>\mu_o,\sigma^2>0} f(\underline{x};\mu,\sigma^2)}{\max_{\sigma^2>0} f(\underline{x};\mu_o,\sigma^2)} \\
&= \begin{cases} \frac{f(\underline{x};\, \overline{x_i},\, \overline{(x_i-\overline{x_i})^2})}{f(\underline{x};\, \mu_o,\, \overline{(x_i-\mu_o)^2})}, & \overline{x_i} > \mu_o \\ 1, & \overline{x_i} \le \mu_o \end{cases}
\end{aligned}$$

where $f(x;\mu,\sigma^2)$ is the $\mathcal{N}(\mu,\sigma^2)$ density and (as usual)

$$\begin{aligned}
\overline{x_i} &= n^{-1}\sum_{i=1}^n x_i \\
\overline{(x_i-t)^2} &= n^{-1}\sum_{i=1}^n (x_i-t)^2 = \hat{\sigma}_t^2
\end{aligned}$$

Now

$$\begin{aligned}
\frac{f(\underline{x};\ \overline{x_i},\ \overline{(x_i-\overline{x_i})^2})}{f(\underline{x};\ \mu_o,\ \overline{(x_i-\mu_o)^2})} &= \left(\frac{\overline{(x_i-\mu_o)^2}}{\overline{(x_i-\overline{x_i})^2}}\right)^{n/2} \\
&= \left(1+\frac{(\overline{x_i}-\mu_o)^2}{\overline{(x_i-\overline{x_i})^2}}\right)^{n/2} \\
&= \left(1+T^2(\underline{x})\right)^{n/2}
\end{aligned}$$

where

$$T(\underline{x}) = \frac{\overline{x_i}-\mu_o}{\sqrt{\overline{(x_i-\overline{x_i})^2}}}$$

Now, since $T^2(\underline{x})$ is monotone in $T(\underline{x})$ for $\overline{x_i} > \mu_o$
GLRT is

$$T(\underline{x}) = \frac{\overline{x_i}-\mu_o}{\sqrt{\overline{(x_i-\overline{x_i})^2}}} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

which is equivalent to the one sided t-test:

$$\frac{\sqrt{n}\,(\overline{x_i}-\mu_o)}{\mathsf{s}} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma'$$

where recall that s^2 is the *unbiased* variance estimate for unknown mean

$$\mathsf{s}^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i-\overline{x_i})^2$$

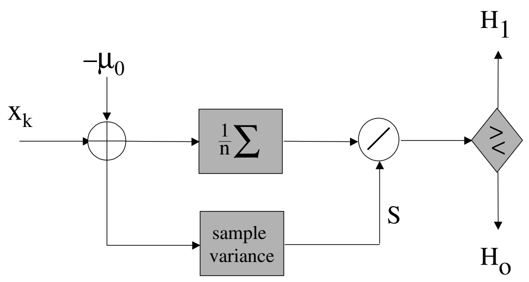


Figure 127: *The one sided t-test for detection of mean exceeding μ_o when variance is unknown*

PERFORMANCE:

Under H_0 we have

$$\begin{aligned} T(x) &= \frac{\overbrace{\sqrt{n}\,(\overline{x_i}-\mu_o)}^{\mathcal{N}(0,1)\cdot\sigma}}{\sqrt{\underbrace{\sum_{i=1}^{n}(x_i-\overline{x_i})^2}_{\chi_{n-1}\cdot\sigma^2}/(n-1)}} \\ &= \frac{\mathcal{N}(0,1)}{\sqrt{\chi_{n-1}/(n-1)}} = \mathcal{T}_{n-1} \end{aligned}$$

where $\mathcal{T}_{n-1}$ is Student-t r.v. with $n-1$ d.f.

Thus

$$\alpha = P_0(T(x) > \gamma) = 1 - \mathcal{T}_{n-1}(\gamma)$$

so that

$$\gamma = \mathcal{T}_{n-1}^{-1}(1-\alpha)$$

Under H_1 we have:

* $\overline{x_i} - \mu_o$ has mean $\mu - \mu_o$ which is no longer zero.

$\Rightarrow T(x)$ follows the non-central Student-t distribution

$\mathcal{T}_{n,d}$ with $n-1$ d.f. and non-centrality parameter

$$d = \frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma}$$

Hence, the power of the one sided t-test is

$$\beta = 1 - \mathcal{T}_{n-1,d}(\mathcal{T}_{n-1}^{-1}(1-\alpha))$$

Note: for large n, $\mathcal{T}_{n-1,d} \rightarrow \mathcal{N}_1(d,1)$

$\Rightarrow$ power function goes to that of single sided hypothesis on μ with σ^2 known.

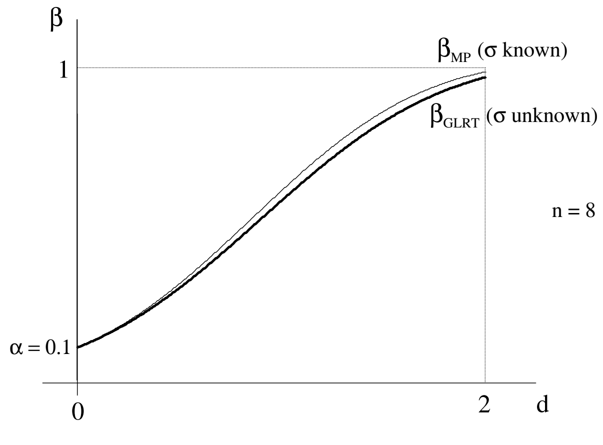


Figure 128: *Power curve for one sided t-test*

9.2.2 CASE II: $H_0 : \mu \leq \mu_o,\ \sigma^2 > 0$, $H_1 : \mu > \mu_o,\ \sigma^2 > 0$

Can show that GLRT has identical one-sided t-test form as in Case I. (Exercise)

$$\frac{\sqrt{n}\ (\overline{x_i} - \mu_o)}{s} \overset{H_1}{\underset{H_0}{\gtrless}} \mathcal{T}_{n-1}^{-1}(1-\alpha)$$

9.2.3 CASE III: $H_0 : \mu = \mu_o,\ \sigma^2 > 0$, $H_1 : \mu \neq \mu_o,\ \sigma^2 > 0$

This case is very similar to the double sided case for known σ^2.
We obtain GLRT as double sided t-test

$$\frac{\sqrt{n}|\overline{x_i} - \mu_o|}{\mathsf{s}} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \mathcal{T}_{n-1}^{-1}(1-\alpha/2)$$

with power curve

$$\beta = 1 - \mathcal{T}_{n-1,d}(\mathcal{T}_{n-1}^{-1}(1-\alpha/2)) + \mathcal{T}_{n-1,-d}(\mathcal{T}_{n-1}^{-1}(1-\alpha/2))$$

For large n this converges to the power curve derived in the case of known variance.

9.3 TESTS ON VARIANCE: KNOWN MEAN

CASE I: $H_0 : \sigma^2 = \sigma_o^2$, $H_1 : \sigma^2 > \sigma_o^2$
CASE II: $H_0 : \sigma^2 \leq \sigma_o^2$, $H_1 : \sigma^2 > \sigma_o^2$
CASE III: $H_0 : \sigma^2 = \sigma_o^2$, $H_1 : \sigma^2 \neq \sigma_o^2$

9.3.1 CASE I: $H_0 : \sigma^2 = \sigma_o^2$, $H_1 : \sigma^2 > \sigma_o^2$

Similarly to the case studied for one sided tests of the mean
* the continuity of the Gaussian p.d.f. $f(\underline{x}; \mu, \sigma^2)$ as a function of σ^2 gives:

$$\begin{aligned}
\Lambda_{GLR} &= \frac{\max_{\sigma^2 > \sigma_o^2} f(\underline{x}; \mu, \sigma^2)}{f(\underline{x}; \mu, \sigma_o^2)} \\
&= \begin{cases} \frac{f(\underline{x}; \mu,\, \hat{\sigma}_\mu^2)}{f(\underline{x};\, \mu, \sigma_o^2)}, & \hat{\sigma}_\mu^2 > \sigma_o^2 \\ 1, & \hat{\sigma}_\mu^2 \leq \sigma_o^2 \end{cases}
\end{aligned}$$

where $\hat{\sigma}_\mu^2$ is the *unbiased* estimate of the variance for known mean

$$\hat{\sigma}_\mu^2 = n^{-1} \sum_{i=1}^n (x_i - \mu)^2$$

After some simple manipulation the GLRT takes the form

$$\Lambda_{GLR} = \left(\underbrace{\frac{1}{\max\{\hat{\sigma}_\mu^2/\sigma_o^2,\ 1\}}}_{u} \, e^{\max\{\hat{\sigma}_\mu^2/\sigma_o^2,\ 1\} - 1} \right)^{n/2} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

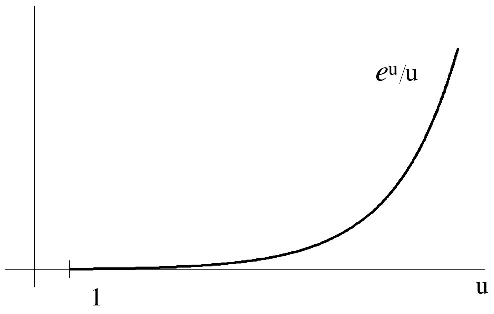


Figure 129: *The function* e^u/u *is monotone increasing over* $u \geq 1$.

As the function e^u/u is monotone increasing over $u \geq 1$, the GLRT reduces to

$$\max\{\hat{\sigma}_\mu^2/\sigma_o^2, 1\} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma$$

Now if $\gamma \leq 1$ then false alarm $\alpha = 1$.

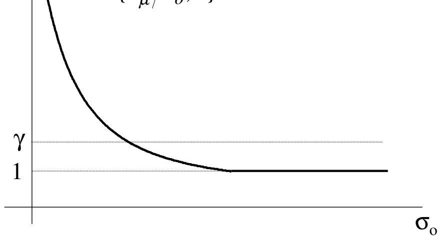


Figure 130: *The GLRT always chooses* H_1 *for* $\gamma < 1$.

Hence we can select $\gamma > 1$ and GLRT reduces to the single sided Chi-square test

$$T(\underline{x}) = \frac{n\hat{\sigma}_\mu^2}{\sigma_o^2} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \chi_n^{-1}(1-\alpha)$$

which we know from previous work is actually an UMP test.

Alternative form of UMP test is a single sided energy detector

$$\sum_{i=1}^n (x_i - \mu)^2 \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma$$

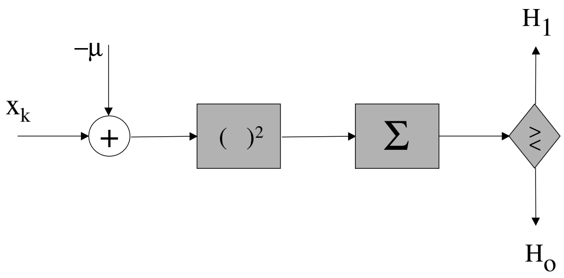


Figure 131: *GLRT for one sided test of positive shift in variance is an energy detector.*

9.3.2 CASE II: $H_0 : \sigma^2 \le \sigma_o^2, H_1 : \sigma^2 > \sigma_o^2$

Now we have

$$\begin{aligned} \Lambda_{GLR} &= \frac{\max_{\sigma^2 > \sigma_o^2} f(\underline{x};\ \mu,\ \sigma^2)}{\max_{\sigma^2 \le \sigma_o^2} f(\underline{x};\ \mu,\ \sigma^2)} \\ &= \left\{ \begin{array}{ll} \frac{f(\underline{x};\ \mu,\ \hat{\sigma}_\mu^2)}{f(\underline{x};\ \mu,\ \sigma_o^2)}, & \hat{\sigma}_\mu^2 > \sigma_o^2 \\ \\ \frac{f(\underline{x};\ \mu,\ \sigma_0^2)}{f(\underline{x};\ \mu,\ \hat{\sigma}_\mu^2)}, & \hat{\sigma}_\mu^2 \le \sigma_o^2 \end{array} \right. \end{aligned}$$

or

$$\Lambda_{GLR} = \begin{cases} \left(\frac{1}{\hat{\sigma}_\mu^2/\sigma_o^2}\, e^{\hat{\sigma}_\mu^2/\sigma_o^2 - 1}\right)^{n/2}, & \hat{\sigma}_\mu^2 > \sigma_o^2 \\ \left(\hat{\sigma}_\mu^2/\sigma_o^2\, e^{1-\hat{\sigma}_\mu^2/\sigma_o^2}\right)^{n/2}, & \hat{\sigma}_\mu^2 \le \sigma_o^2 \end{cases}$$

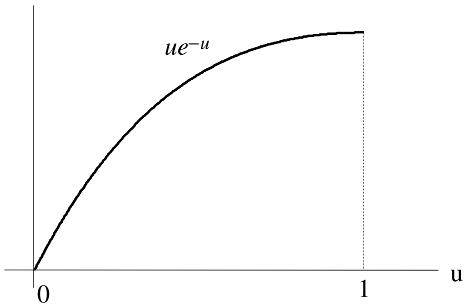


Figure 132: *The function ue^{-u} is monotone increasing over $0 \le u \le 1$.*

As e^u/u is monotone increasing over $u > 1$ and ue^{-u} is monotone increasing over $0 \le u \le 1$, the GLRT reduces to the same form as derived for Case I:

$$\frac{n\,\hat{\sigma}_\mu^2}{\sigma_o^2} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

and the rest of the analysis is identical to before.

9.3.3 CASE III: $H_0 : \sigma^2 = \sigma_o^2$, $H_1 : \sigma^2 \neq \sigma_o^2$

Now, we have

$$\begin{aligned} \Lambda_{GLR} &= \frac{\max_{\sigma^2 \neq \sigma_o^2} f(\underline{x}; \mu, \sigma^2)}{f(\underline{x}; \mu, \sigma_o^2)} \\ &= \frac{f(\underline{x}; \mu,\ \hat{\sigma}_\mu^2)}{f(\underline{x};\ \mu,\ \sigma_o^2)} \\ &= \left(\frac{1}{\hat{\sigma}_\mu^2/\sigma_o^2}\ e^{\hat{\sigma}_\mu^2/\sigma_o^2 - 1}\right)^{n/2} \end{aligned}$$

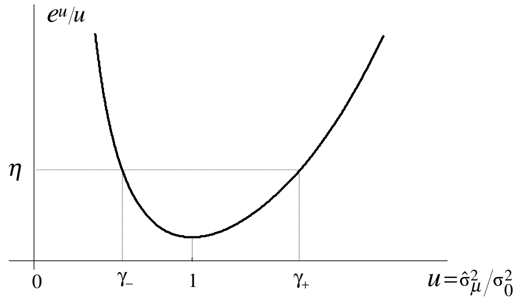


Figure 133: *As e^u/u is convex the decision H_0 region of double sided Chi-square test is an interval $[\gamma_-, \gamma_+]$.*

As the function e^u/u is convex over $u \geq 0$ the H_0 decision region can be written in the form

$$\gamma_- \;\leq\; \frac{\hat{\sigma}^2_\mu}{\sigma^2_o} \;\leq\; \gamma_+,$$

where γ_- and γ_+ are selected to give $P_F = \alpha$.
A common choice of thresholds is ($\alpha \leq \frac{1}{2}$):

$$\begin{aligned}\gamma_- &= 1/n\ \chi_n^{-1}(\alpha/2)\\ \gamma_+ &= 1/n\ \chi_n^{-1}(1-\alpha/2)\end{aligned}$$

which gives equal area ($\alpha/2$) to the upper and lower tails of the χ_n distribution corresponding to a total FA probability $P_F = \alpha$ (see Fig. 134).

<u>Power of double sided GLRT of variance:</u>

Assume that the true value of $\sigma^2 > \sigma^2_o$ under H_1 is $\sigma^2 = \sigma^2_1$. Then

$$\begin{aligned}\beta = 1 - P(n\gamma_- \leq \frac{n\hat{\sigma}^2_\mu}{\sigma^2_o} \leq n\gamma_+|H_1)\\ &= 1 - P(n\gamma_- \leq \underbrace{\frac{n\hat{\sigma}^2_\mu}{\sigma^2_1}}_{\chi_n \text{ under } H_1}\left(\frac{\sigma^2_1}{\sigma^2_o}\right) \leq n\gamma_+|H_1)\\ &= 1 - \chi_n\left(n\gamma_+\frac{\sigma^2_o}{\sigma^2_1}\right) + \chi_n\left(n\gamma_-\frac{\sigma^2_o}{\sigma^2_1}\right).\end{aligned}$$

9.4 TESTS ON VARIANCE: UNKNOWN MEAN

CASE I: $H_0 : \sigma^2 = \sigma^2_o,\ \mu \in \mathbb{R}$, $H_1 : \sigma^2 > \sigma^2_o,\ \mu \in \mathbb{R}$

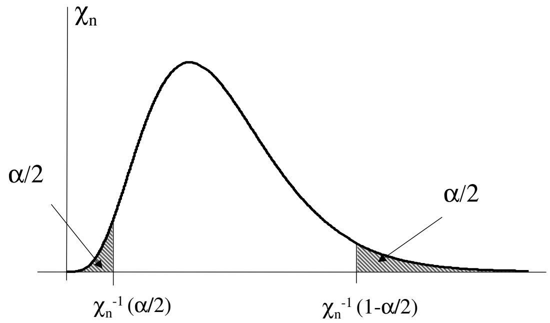


Figure 134: *Quantiles of Chi-square specify the thresholds γ_- and γ_+ for double sided test of variance.*

CASE II: $H_0 : \sigma^2 < \sigma_o^2,\ \mu \in \mathbb{R}$, $H_1 : \sigma^2 > \sigma_o^2,\ \mu \in \mathbb{R}$
CASE III: $H_0 : \sigma^2 = \sigma_o^2,\ \mu \in \mathbb{R}$, $H_1 : \sigma^2 \neq \sigma_o^2\ \mu \in \mathbb{R}$

9.4.1 CASE I: $H_0 : \sigma^2 = \sigma_o^2$, $H_1 : \sigma^2 > \sigma_o^2$

We now have

$$\begin{aligned}
\Lambda_{GLR} &= \frac{\max_{\sigma^2 > \sigma_o^2,\, \mu} f(\underline{x}; \mu, \sigma^2)}{\max_\mu f(\underline{x}; \mu, \sigma_o^2)} \\
&= \left\{ \begin{array}{ll} \frac{f(\underline{x};\, \overline{x_i},\, \hat{\sigma}^2)}{f(\underline{x};\, \overline{x_i}, \sigma_o^2)}, & \hat{\sigma}^2 > \sigma_o^2 \\ 1, & \hat{\sigma}^2 \leq \sigma_o^2 \end{array} \right.
\end{aligned}$$

where now

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (x_i - \overline{x_i})^2$$

This is identical to the case of known μ with μ replaced by $\overline{x_i}$.
Hence, referring to work done for that case, we immediately obtain the single sided GLRT

$$T(\underline{x}) = \frac{n\hat{\sigma}^2}{\sigma_o^2} = \frac{(n-1)s^2}{\sigma_o^2} \mathop{\gtrless}_{H_0}^{H_1} \gamma$$

PERFORMANCE

Under H_0,

* $T(\underline{x})$ is a Chi-square r.v. with $n-1$ d.f. and thus

$$\gamma = \chi^{-1}_{n-1}(1-\alpha)$$

Under H_1,

* $T(\underline{x})$ is Chi-square with $n-1$ d.f. (scaled by σ^2/σ_o) so

$$\begin{aligned}\beta &= 1-\chi^{-1}_{n-1}(\gamma\ \sigma_o^2/\sigma^2)\\ &= 1-\chi_{n-1}(\chi^{-1}_{n-1}(1-\alpha)\ \sigma_o^2/\sigma^2)\end{aligned}$$

9.4.2 CASE II: $H_0: \sigma^2 < \sigma_o^2,\ \mu \in \mathbb{R}$, $H_1: \sigma^2 > \sigma_o^2,\ \mu \in \mathbb{R}$

GLRT is dentical to Case I.

9.4.3 CASE III: $H_0: \sigma^2 = \sigma_o^2,\ \mu \in \mathbb{R}$, $H_1: \sigma^2 \neq \sigma_o^2\ \mu \in \mathbb{R}$

The derivation of the GLRT is completely analogous to Case III for known mean.

The H_0 decision region is identical to before except that sample variance replaces $\hat{\sigma}^2_\mu$, and the test statistic now has only $n-1$ d.f.

$$\chi_{n-1}(\alpha/2)\ \leq\ \frac{(n-1)\mathsf{s}^2}{\sigma_o^2}\ \leq\ \chi_{n-1}(1-\alpha/2)$$

The power function is identical to previous Case III for known μ except that χ_n CDF is replaced by χ_{n-1} CDF.

9.5 TESTS ON EQUALITY OF MEANS: UNKNOWN VARIANCE

Two i.i.d. independent samples

$\underline{x} = [x_1, \ldots, x_{n_1}]^T, x_i\ \sim\ \mathcal{N}(\mu_x, \sigma^2)$

$\underline{y} = [y_1, \ldots, y_{n_2}]^T, y_i\ \sim\ \mathcal{N}(\mu_y, \sigma^2)$

Case I: $H_0: \mu_x = \mu_y,\ \sigma^2 > 0$, $H_1: \mu_x \neq \mu_y,\ \sigma^2 > 0$

Case II: $H_0: \mu_x \leq \mu_y,\ \sigma^2 > 0$, $H_1: \mu_x > \mu_y,\ \sigma^2 > 0$

$\underline{x}, \underline{y}$ has the density

$$
\begin{aligned}
f(\underline{x}, \underline{y};\ \mu_x, \mu_y, \sigma_x, \sigma_y) &= \left(\frac{1}{2\pi\sigma_x^2}\right)^{n_1/2} \left(\frac{1}{2\pi\sigma_y^2}\right)^{n_2/2} \\
&\cdot \exp\left(-\frac{1}{2\sigma_x^2}\ \sum_{i=1}^{n}(y_i-\mu_x)^2 - \frac{1}{2\sigma_y^2}\ \sum_{i=1}^{n}(y_i-\mu_y)^2\right)
\end{aligned}
$$

where $n = n_1 + n_2$.

9.5.1 CASE I: $H_0 : \mu_x \leq \mu_y,\ \sigma^2 > 0$, $H_1 : \mu_x > \mu_y,\ \sigma^2 > 0$

GLR statistic is

$$
\begin{aligned}
\Lambda_{GLR} &= \frac{\max_{\mu_x \neq \mu_y, \sigma^2>0} f(\underline{x}, \underline{y}; \mu_x, \mu_y, \sigma^2)}{\max_{\mu, \sigma^2>0} f(\underline{x}, \underline{y}; \mu, \mu, \sigma^2)} \\
&= \frac{\max_{\mu_x, \mu_y, \sigma^2>0} f(\underline{x}, \underline{y}; \mu_x, \mu_y, \sigma^2)}{\max_{\mu, \sigma^2>0} f(\underline{x}, \underline{y}; \mu, \mu, \sigma^2)}
\end{aligned}
$$

Now note: MLE for case $\mu_x \neq \mu_y$ is

$$
\begin{aligned}
\hat{\mu}_x &= \overline{x_i} = n_1^{-1}\sum_{i=1}^{n_1} x_i \\
\hat{\mu}_y &= \overline{x_i} = n_2^{-1}\sum_{i=1}^{n_2} y_i \\
\hat{\sigma}_1^2 &= n^{-1}\sum_{i=1}^{n_1}(x_i - \overline{x_i})^2 + n^{-1}\sum_{i=1}^{n_2}(y_i - \overline{y_i})^2 \\
&= \frac{n_1}{n}\hat{\sigma}_x^2 + \frac{n_2}{n}\hat{\sigma}_y^2
\end{aligned}
$$

while MLE for case $\mu_x = \mu_y = \mu$ is

$$
\begin{aligned}
\hat{\mu} &= n^{-1}\sum_{i=1}^{n_1} x_i + n^{-1}\sum_{i=1}^{n_2} y_i = \frac{n_1}{n}\hat{\mu}_x + \frac{n_2}{n}\hat{\mu}_y \\
\hat{\sigma}_0^2 &= n^{-1}\sum_{i=1}^{n_1}(x_i - \hat{\mu})^2 + n^{-1}\sum_{i=1}^{n_2}(y_i - \hat{\mu})^2 \\
&= \hat{\sigma}_1^2 + \frac{n_1}{n}(\hat{\mu} - \overline{x_i})^2 + \frac{n_2}{n}(\hat{\mu} - \overline{y_i})^2
\end{aligned}
$$

Plugging these MLE's into numerator and denominator of LR statistic, obtain after some algebra

$$\begin{aligned}\Lambda_{GLR} &= \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2}\right)^{n/2} c \\ &= \left(\frac{\hat{\sigma}_1^2 + \frac{n_1}{n}(\hat{\mu} - \overline{x_i})^2 + \frac{n_2}{n}(\hat{\mu} - \overline{y_i})^2}{\hat{\sigma}_1^2}\right)^{n/2} c\end{aligned}$$

So that one form of GLRT test is

$$\frac{\frac{n_1}{n}(\hat{\mu} - \overline{x_i})^2 + \frac{n_2}{n}(\hat{\mu} - \overline{y_i})^2}{\frac{n_1}{n}\hat{\sigma}_x^2 + \frac{n_2}{n}\hat{\sigma}_y^2} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma$$

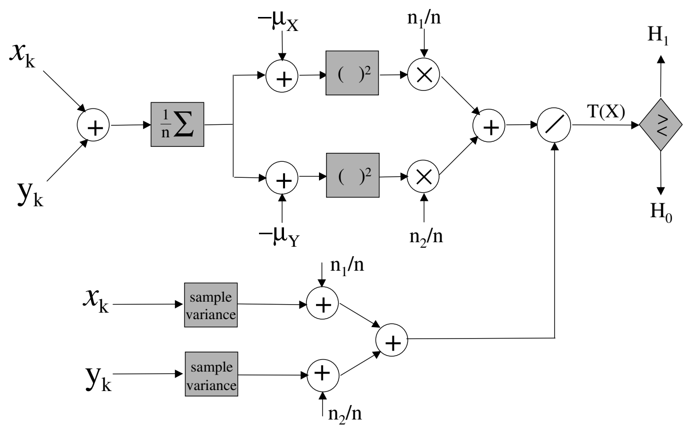


Figure 135: *Block diagram of test of equality of means of two populations.*

To reduce this to a well known test statistic we use the identities

$$\hat{\mu} - \overline{x_i} = \frac{n_1}{n}(\overline{y_i} - \overline{x_i})$$

$$\hat{\mu} - \overline{y_i} = -\frac{n_2}{n}(\overline{y_i} - \overline{x_i})$$

to obtain final form of GLRT

$$T(\underline{x}, \underline{y}) = \frac{\left|\sqrt{\frac{n_1 n_2}{n}}(\overline{y_i} - \overline{x_i})\right|}{\mathsf{s}_2} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma \qquad (41)$$

where we have defined the pooled sample variance

$$\mathsf{s}_2^2 = \frac{1}{n-2}\left(\sum_{i=1}^{n_1}(x_i - \hat{\mu})^2 + \sum_{i=1}^{n_2}(y_i - \hat{\mu})^2\right)$$

The test (41) is the well known *paired t-test*.

PERFORMANCE

Under H_0:

$$\overline{y_i} - \overline{x_i} = \mathcal{N}(0, \sigma^2) \cdot \sqrt{(1/n_1 + 1/n_2)}$$

and the test statistic is of the form of a |Student-t| with $n_1 + n_2 - 2 = n - 2$ d.f.

$$T(\underline{x}, \underline{y}) = \left|\frac{\mathcal{N}(0,1)}{\sqrt{\chi_{n-2}/n-2}}\right|$$

Threshold setting

$$\alpha = P_0(-\gamma < \mathcal{T}_{n-2} \leq \gamma)$$

$\Rightarrow \gamma = \mathcal{T}_{n-2}^{-1}(1 - \alpha/2)$
so that we have level α test

$$\frac{\left|\sqrt{\frac{n_1 n_2}{n}}(\overline{y_i} - \overline{x_i})\right|}{\mathsf{s}_2} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \mathcal{T}_{n-2}^{-1}(1 - \alpha/2)$$

Under H_1 the test statistic is | non-central Student-t | with $n-2$ d.f and non-centrality $d = \sqrt{n_1 n_2/n}|\mu_y - \mu_x|/\sigma$.

9.5.2 CASE II: $H_0 : \mu_y \leq \mu_x,\ \sigma^2 > 0$, $H_1 : \mu_y > \mu_x,\ \sigma^2 > 0$

In an analogous manner to before: find that the GLRT reduces to the one sided t-test

$$\frac{\frac{\sqrt{n_1 n_2}}{n}(\overline{y_i} - \overline{x_i})}{\mathsf{s}_2} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma = \mathcal{T}_{n-2}^{-1}(1 - \alpha)$$

9.6 TESTS ON EQUALITY OF VARIANCES

Two i.i.d. independent samples
* $\underline{x} = [x_1, \ldots, x_{n_1}]^T$, $x_i \sim \mathcal{N}(\mu_x, \sigma_x^2)$
* $\underline{y} = [y_1, \ldots, y_{n_2}]^T$, $y_i \sim \mathcal{N}(\mu_y, \sigma_x^2)$
* μ_x, μ_y unknown

Case I: $H_0 : \sigma_x^2 = \sigma_y^2$, $H_1 : \sigma_x^2 \neq \sigma_y^2$
Case II: $H_0 : \sigma_x^2 = \sigma_y^2$, $H_1 : \sigma_x^2 > \sigma_y^2$

9.6.1 CASE I: $H_0 : \sigma_x^2 = \sigma_y^2$, $H_1 : \sigma_x^2 = \sigma_y^2$

The GLR statistic is

$$\begin{aligned}\Lambda_{GLR} &= \frac{\max_{\sigma_x^2 \neq \sigma_y^2, \mu_x, \mu_y} f(\underline{x}, \underline{y}; \mu_x, \mu_y, \sigma_x^2, \sigma_y^2)}{\max_{\sigma_x^2 = \sigma_y^2, \mu_x, \mu_y} f(\underline{x}, \underline{y}; \mu, \mu, \sigma_x^2, \sigma_y^2)} \\ &= \frac{f(\underline{x}, \underline{y}; \overline{x_i}, \overline{y_i}, \hat{\sigma}_x^2, \hat{\sigma}_y^2)}{f(\underline{x}, \underline{y}; \overline{x_i}, \overline{y_i}, \hat{\sigma}^2, \hat{\sigma}^2)}\end{aligned}$$

which is easily shown to reduce to

$$\Lambda_{GLR} = \sqrt{\frac{(\hat{\sigma}^2)^{n_1+n_2}}{(\hat{\sigma}_x^2)^{n_1}\,(\hat{\sigma}_y^2)^{n_2}}} \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \eta$$

Or, simplifying further by expressing the pooled variance estimate $\hat{\sigma}^2$ as

$$\hat{\sigma}^2 = \frac{n_1}{n}\hat{\sigma}_x^2 + \frac{n_2}{n}\hat{\sigma}_y^2$$

we obtain the equivalent test:

$$\sqrt{\underbrace{\left(1 + \overbrace{\frac{n_2\hat{\sigma}_y^2}{n_1\hat{\sigma}_x^2}}^{u}\right)^{n_1} \cdot \left(1 + \overbrace{\frac{n_1\hat{\sigma}_x^2}{n_2\hat{\sigma}_y^2}}^{1/u}\right)^{n_2}}_{g(u)}} \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \eta$$

By investigating stationary points of $g(u)$
$\Rightarrow$ function $g(u) = (1+u)^{n_1}(1+1/u)^{n_2}$ is convex and has a single minimum over the range $u \geq 0$
Specifically, note that:

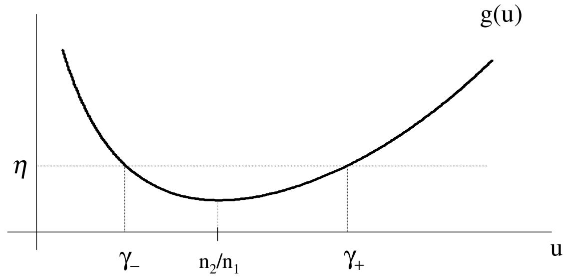


Figure 136: *The double sided test statistic is of the form $g(u)$ which is convex over $u \geq 0$ with minimum at $u = n_2/n_1$.*

$$g^{'}(u) = \frac{n_1}{u^2}(1+u)^{n_1-1}(1+1/u)^{n_2-1}\ \left(u^2 + (1-n_2/n_1)u - n_2/n_1\right)$$

has only one positive root which occurs at $u = n_2/n_1$.
Hence the H_0 decision region for the GLRT is of the form

$$\gamma_- \leq \frac{\hat{\sigma}_y^2}{\hat{\sigma}_x^2} \leq \gamma_+$$

which is equivalent to a Fisher F-test

$$\gamma_-^{'} \leq \frac{\mathsf{s}_x^2}{\mathsf{s}_y^2} \leq \gamma_+^{'}$$

Where the thresholds can be set according to

$$\begin{aligned}
\gamma_-^{'} &= \mathcal{F}^{-1}_{n_1-1,n_2-1}(\alpha/2) \\
\gamma_+^{'} &= \mathcal{F}^{-1}_{n_1-1,n_2-1}(1-\alpha/2)
\end{aligned}$$

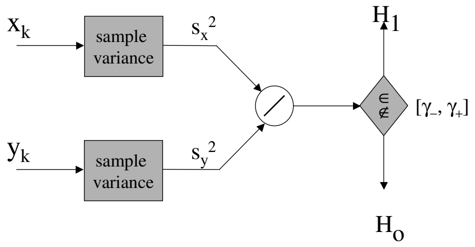


Figure 137: *Block diagram of test of equality of two variances.*

9.6.2 CASE II: $H_0 : \sigma_x^2 = \sigma_y^2$, $H_1 : \sigma_y^2 > \sigma_x^2$

We now have

$$\begin{aligned}\Lambda_{GLR} &= \frac{\max_{\sigma_y^2 > \sigma_x^2, \mu_x, \mu_y} f(\underline{x}, \underline{y}; \mu_x, \mu_y, \sigma_x^2, \sigma_y^2)}{\max_{\sigma_y^2 = \sigma_x^2, \mu_x, \mu_y} f(\underline{x}, \underline{y}; \mu, \mu, \sigma_x^2, \sigma_y^2)} \\ &= \begin{cases} \sqrt{\frac{(\hat{\sigma}^2)^{n_1+n_2}}{(\hat{\sigma}_x^2)^{n_1} (\hat{\sigma}_y^2)^{n_2}}}, & \hat{\sigma}_y^2 > \hat{\sigma}_x^2 \\ 1, & \hat{\sigma}_y^2 = \hat{\sigma}_x^2 \end{cases}\end{aligned}$$

From study of the previous case we know that a certain function $g(u)$ is convex with minimum at n_2/n_1. This implies that Λ_{GLR} is monotone increasing in $\hat{\sigma}_y^2/\hat{\sigma}_x^2$ over the range $\hat{\sigma}_y^2 > \hat{\sigma}_x^2$
Thus we obtain that the GLRT is a single sided F-test

$$\frac{\mathsf{s}_y^2}{\mathsf{s}_x^2} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

where

$$\gamma = \mathcal{F}^{-1}_{n_2-1, n_1-1}(1-\alpha)$$

9.7 TESTS ON CORRELATION

Two i.i.d. samples
$\underline{x} = [x_1, \ldots, x_n]^T$,
$\underline{y} = [y_1, \ldots, y_n]^T$,
$\underline{z}_i = [x_i, y_i]^T \ \sim \ \mathcal{N}_2(\underline{\mu},\ R)$

$$\underline{\mu} = [\mu_x, \mu_y]^T, \quad R = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}$$

Correlation coefficient ρ

$$\rho = \frac{\sigma_{xy}}{\sigma_x\ \sigma_y}$$

Case I: $H_0 : \rho = 0$, $H_1 : \rho \neq 0$
Case II: $H_0 : \rho = 0$, $H_1 : \rho > 0$
The sequence of i.i.d. vectors $\{\underline{z}_i\}_{i=1}^n$ has the bivariate density

$$f(\underline{x}, \underline{y};\ \underline{\mu}, R) =$$
$$\left(\frac{1}{2\pi\sqrt{|\det R|}}\right)^n \exp\left(-\frac{1}{2}\sum_{i=1}^n (\underline{z}_i - \underline{\mu})^T R^{-1} (\underline{z}_i - \underline{\mu})\right)$$

9.7.1 CASE I: $H_0 : \rho = \rho_o$, $H_1 : \rho \neq \rho_o$

Will show later that

* the maximum of the bivariate Gaussian p.d.f. $f(\underline{x}, \underline{y};\ \mu_x, \mu_y, R)$ over μ_x, μ_y and R is equal to

$$\max_{R, \mu_x, \mu_y} f(\underline{x}, \underline{y};\ \mu_x, \mu_y, R) = \left(\frac{1}{2\pi|\det R|}\right)^{n/2} e^{-n/2}$$

and maximum is attained by the joint ML estimates

$$\begin{aligned}
\hat{\mu}_x &= \overline{x} \\
\hat{\mu}_y &= \overline{y} \\
\hat{R} &= \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} x_i \\ y_i \end{bmatrix} [x_i, y_i] = \begin{bmatrix} \hat{\sigma}_x^2 & \hat{\sigma}_{xy} \\ \hat{\sigma}_{yx} & \hat{\sigma}_y^2 \end{bmatrix}
\end{aligned}$$

where

$$\hat{\sigma}_{xy} = n^{-1} \sum_{i=1}^{n} (x_i y_i - \overline{xy})$$

Using this we can easily find GLR statistic

$$\begin{aligned}
\Lambda_{GLR} &= \frac{\max_{R,\mu_x,\mu_y} f(\underline{x}, \underline{y}; \mu_x, \mu_y, R)}{\max_{\text{diagonal } R,\mu_x,\mu_y} f(\underline{x}, \underline{y}; \mu_x, \mu_y, R)} \\
&= \left(\frac{\hat{\sigma}_x^2 \hat{\sigma}_y^2}{\hat{\sigma}_x^2 \hat{\sigma}_y^2 - \hat{\sigma}_{xy}^2} \right)^{n/2} \\
&= \left(\frac{1}{1 - \hat{\rho}^2} \right)^{n/2}
\end{aligned}$$

where we have defined *sample correlation coefficient*

$$\hat{\rho} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \, \hat{\sigma}_y}$$

As Λ_{GLR} is monotonic increasing in $|\hat{\rho}|$ we have one simple form of the GLRT

$$|\hat{\rho}| \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

The distribution under H_0 of $\hat{\rho}$ is analytical but non-standard [27].

A more convenient test:

Note that the function

$$g(u) = u^2/(1 - u^2)$$

is monotone increasing in $|u|$.

Therefore, GLRT is equivalent to

$$\frac{(n-2) \, |\hat{\rho}|}{\sqrt{1 - \hat{\rho}^2}} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

This test statistic can be shown to follow the student-t distribution with $n - 2$ d.f. [27].

$\Rightarrow$ Level α threshold is determined

$$\gamma = \mathcal{T}_{n-2}^{-1}(1 - \alpha/2)$$

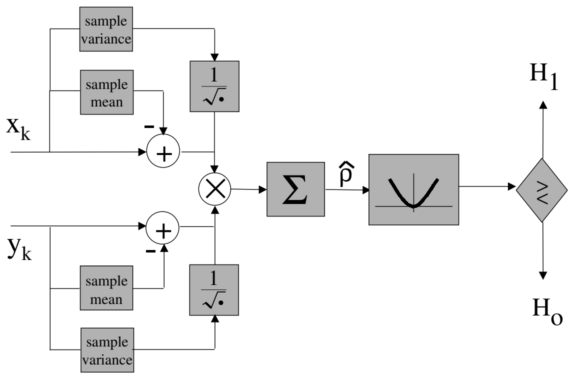


Figure 138: *Block diagram of test of nonzero correlation between two populations.*

9.7.2 CASE II: $H_0 : \rho = 0$, $H_1 : \rho > 0$

By analogous methods as used before the GLRT is a one sided test of the form

$$\hat{\rho} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

or alternatively

$$\frac{(n-2)\,\hat{\rho}}{\sqrt{1-\hat{\rho}^2}} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \mathcal{T}_{n-2}^{-1}(1-\alpha)$$

9.8 Exercises

1. n i.i.d. realizations of a bivariate Gaussian random vector $\underline{z} = [z_1, z_2]^T$ are observed where the mean of $\underline{z}$ is $\underline{s}$ and the covariance is of the form:

$$R_z = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \sigma^2, \qquad \text{Note}: R_z^{-1} = \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \frac{1}{\sigma^2(1-\rho^2)}$$

where the component variances σ^2 and the correlation coefficient $\rho \in [-1, 1]$ are known.

(a) Derive the MP LRT (with threshold) of the simple hypotheses

$$\begin{aligned} H_0 &: \underline{s} = \underline{s}_0 \\ H_1 &: \underline{s} = \underline{s}_1. \end{aligned}$$

For $\underline{s}_0 = 0$ is your test UMP for $H_1 : \underline{s} \neq 0$? when $|\rho| > 0$? How about for $\rho = 0$?.

(b) Find the ROC of the MP LRT of part (a) and show that it is specified by the detectability index

$$d^2 = \frac{|E[T|H_1] - E[T|H_0]|^2}{\text{var}(T|H_0)} = n(\underline{s}_1 - \underline{s}_0)^T R_z^{-1}(\underline{s}_1 - \underline{s}_0)$$

where T is a test statistic linear in $\underline{z}$.

(c) Now assume that $\underline{s}_0$ and $\underline{s}_1$ satisfy the "power" constraints $\underline{s}_0^T \underline{s}_0 = \underline{s}_1^T \underline{s}_1 = 1$. For fixed ρ show that the optimal signal pair $\underline{s}_1, \underline{s}_0$ which maximizes d^2 must satisfy $\underline{s}_1 - \underline{s}_0 = c[1, 1]$ when $\rho < 0$ while it must satisfy $\underline{s}_1 - \underline{s}_0 = c[1, -1]$ when $\rho > 0$, where c is a constant ensuring the power constraint.

(d) Assuming the optimal signal pair derived in part (b), for what value of ρ is the detectability index the worst (smallest) and for what value is it the best? Does this make sense?

2. The test on equality of a pair of means in Sec. 9.5 required equal variances and led to the paired t-test of (41. Extend this to the case where the variances on each population may be unequal.

End of chapter

10 CONFIDENCE INTERVALS

OUTLINE

* Confidence intervals via pivots

* Confidence intervals and double sided tests

* Confidence interval for mean, variance, correlation

* Confidence regions for vector mean, covariance variance

REFERENCES

Bickel and Doksum [2]

Mood, Graybill and Boes [19]

Morrison [20]

10.1 DEFINITION OF CONFIDENCE INTERVAL

Let $\theta \in \Theta$ be an unknown scalar parameter

Let $x \sim f(x;\theta)$

As opposed to a point estimator $\hat{\theta}(x)$ which is a (random) point in the parameter space Θ, a confidence interval $T_1(x), T_2(x)$ is a (random) interval in parameter space. Confidence interevals are also called set or interval estimators.

OBJECTIVE: find two statistics $T_1(x)$ and $T_2(x)$, $T_1 < T_2$, which specify endpoints of a random interval

$$[T_1,\ T_2] = [T_1(x),\ T_2(x)]$$

that contains θ with high probability.

CONSTRUCTION OF CONFIDENCE INTERVALS

1. Fix $\alpha \in [0,1]$

2. For all $\theta \in \Theta$ we require

$$P_\theta(T_1(x) \leq \theta \leq T_2(x)) \geq 1-\alpha$$

* $[T_1, T_2]$ is a $100(1-\alpha)\%$ confidence interval

Equivalently:

$\Rightarrow$ $1-\alpha$ is confidence level of statement "$\theta \in [T_1, T_2]$"

$\Rightarrow$ $[T_1, T_2]$ is a set estimate of θ

* $P_\theta(T_1(x) \leq \theta \leq T_2(x))$ is coverage probability of the confidence interval

* $1-\alpha$ is lower bound on coverage probability of a $100(1-\alpha)\%$ conf. interval

* $T_2 - T_1$ is the size of confidence interval

* Sometimes a level α GLRT or LMPT of double sided hypotheses can be useful for finding confidence intervals.

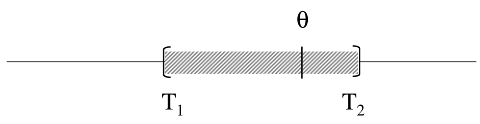


Figure 139: *Confidence interval covers θ with probability at least* $1 - \alpha$.

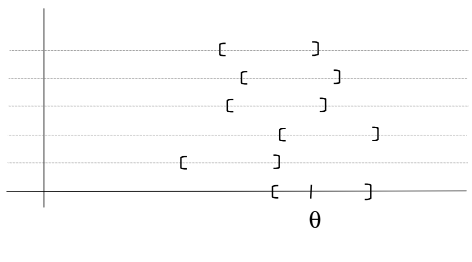


Figure 140: *In n-trials a $(1 - \alpha)\%$ confidence interval $[T_1, T_2]$ covers θ approximately $(1 - \alpha)n$ times (or more) for n large. Here shown is an 85% confidence interval.*

10.2 CONFIDENCE ON MEAN: KNOWN VAR

Objective: find confidence interval on the mean μ based on i.i.d. Gaussian sample $\underline{x} = [x_1, \ldots, x_n]$ with known variance σ^2.

APPROACH 1: Use Tchebychev inequality:

$$P_\mu(|\overline{x_i} - \mu| \geq \epsilon) \leq \frac{\overbrace{E_\mu[(\overline{x_i} - \mu)^2]}^{\sigma^2/n}}{\epsilon^2}$$

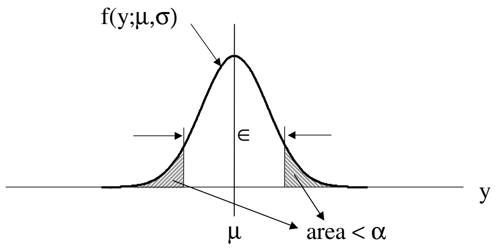


Figure 141: *Tchebychev inequality specifies an interval containing the mean with at least probability* $1 - \sigma^2/n\epsilon^2$. *In figure* $f(y;\mu,\sigma)$ *denotes the density of the sample mean* $y = \overline{x_i}$.

or, setting $\epsilon = c\,\sigma/\sqrt{n}$

$$P_\mu(|\overline{x_i} - \mu| \geq c\,\sigma/\sqrt{n}) \leq \frac{1}{c^2}$$

or equivalently

$$P_\mu(|\overline{x_i} - \mu| \leq c\,\sigma/\sqrt{n}) \geq 1 - \frac{1}{c^2}$$

i.e.

$$P_\mu(\overline{x_i} - c\,\sigma/\sqrt{n} \leq \mu \leq \overline{x_i} + c\,\sigma/\sqrt{n}) \geq 1 - \frac{1}{c^2}$$

Finally take $c = 1/\sqrt{\alpha}$ to obtain $100(1-\alpha)\%$ confidence interval for μ

$$\left[\overline{x_i} - \frac{\sigma}{\sqrt{\alpha n}},\ \overline{x_i} + \frac{\sigma}{\sqrt{\alpha n}}\right]$$

OBSERVATIONS:

* Tchebychev interval is symmetric about sample mean

* Size $2\frac{\sigma}{\sqrt{\alpha n}}$ of interval increases in σ^2/n

* There is a tradeoff between coverage probability $\geq 1-\alpha$ and small size

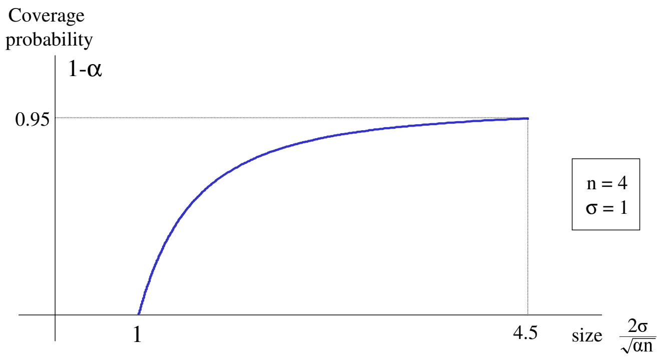


Figure 142: *Coverage vs. size of Tchebychev confidence interval for mean for a Gaussian sample having known variance.*

* Tchebychev interval is “distribution free”

* Actual coverage probability may be $\gg$ desired confidence $1-\alpha$

* Tchebychev intervals are usually excessively large

APPROACH 2: Find exact confidence interval via PIVOT

Recall problem of testing double sided hypotheses for i.i.d. Gaussian r.v.s having known variance

$$H_0: \mu = \mu_o$$

$$H_1: \mu \neq \mu_o$$

we found level α GLRT

$$|T(\underline{x}, \mu_o)| = \frac{\sqrt{n}|\overline{x_i} - \mu_o|}{\sigma} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma = \mathcal{N}^{-1}(1-\alpha/2)$$

where we have defined the function

$$T(\underline{x}, \mu_o) = \frac{\sqrt{n}(\mu_o - \overline{x_i})}{\sigma}$$

Note when $\mu = \mu_0$:

1) $T(\underline{x}, \mu_o)$ is:

– monotone function of μ_o

– has a probability distribution independent of μ_o:

$\Rightarrow$ such a function $T(\underline{x}, \mu_o)$ is called a PIVOT.

2) By design of the threshold $\gamma = \mathcal{N}^{-1}(1 - \alpha/2)$ the false alarm probability of the test is

$$P_{\mu_o}\left(|T(\underline{x}, \mu_o)| > \gamma\right) = \alpha$$

for arbitrary μ_o.

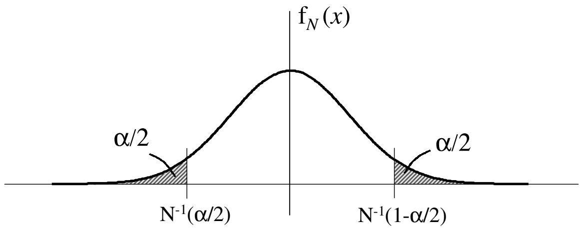


Figure 143: *The false alarm level setting for double sided test mean gives an exact* $1 - \alpha$ *confidence interval.*

As P_F is independent of μ_o, μ_o is just a dummy variable which can be replaced by the generic μ and

$$P_\mu\left(-\gamma \leq T(\underline{x}, \mu) \leq \gamma\right) = 1 - \alpha.$$

As $T(\underline{x}, \mu) = \sqrt{n}(\overline{x_i} - \mu)/\sigma$ is monotone in μ the T-inequalities $-\gamma \leq T(\underline{x}, \mu) \leq \gamma$ are equivalent to the μ-inequalities $\overline{x_i} - \frac{\sigma}{\sqrt{n}}\,\gamma \leq \mu \leq \overline{x_i} + \frac{\sigma}{\sqrt{n}}\,\gamma$. Summarizing, we have found the following EXACT $100(1-\alpha)\%$ conf. interval for μ

$$\left[\overline{x_i} - \frac{\sigma}{\sqrt{n}}\,\mathcal{N}^{-1}(1 - \alpha/2),\; \overline{x_i} + \frac{\sigma}{\sqrt{n}}\,\mathcal{N}^{-1}(1 - \alpha/2)\right]$$

OBSERVATIONS

1. By CLT this interval is accurate for large n even for non-Gaussian case

$$\sqrt{n}(\overline{x_i} - \mu)/\sigma \ \rightarrow \ \mathcal{N}(0,1), \quad (i.d.)$$

2. Exact interval is symmetric about $\overline{x_i}$

3. Exact interval is significantly smaller than Tchebychev

$$[T_2 - T_1]_{Tchby} = 2\frac{\sigma}{\sqrt{\alpha n}} > 2\frac{\sigma}{\sqrt{n}}\,\mathcal{N}^{-1}(1-\alpha/2) = [T_2 - T_1]_{Exact}$$

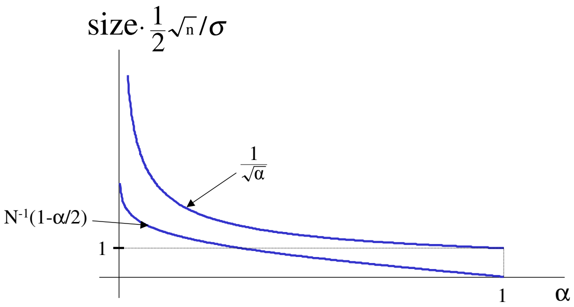


Figure 144: *Size vs. (1-confidence level) for exact and Tchebychev intervals.*

10.3 CONFIDENCE ON MEAN: UNKNOWN VAR

Objctive: find conf. inerval on the mean μ based on i.i.d. Gaussian sample $\underline{x} = [x_1, \ldots, x_n]$ with unknown variance σ^2.

APPROACH: Exact confidence interval via pivot

Solution: Motivated by previous example we go back to the double sided hypothesis GLRT for unknown variance

$$\begin{aligned} H_0 : \mu &= \mu_o, \quad \sigma^2 > 0 \\ H_1 : \mu &\neq \mu_o, \quad \sigma^2 > 0 \end{aligned}$$

we found level α t-test for Gaussian x_i's:

$$|T(\underline{x}, \mu_o)| = \frac{\sqrt{n}|\overline{x_i} - \mu_o|}{\mathsf{s}} \overset{H_1}{\underset{H_0}{\gtrless}} \gamma = \mathcal{T}_{n-1}^{-1}(1-\alpha/2)$$

Therefore, an exact $(1-\alpha)\%$ confidence interval for μ is

$$\left[\overline{x_i} - \frac{\mathsf{s}}{\sqrt{n}}\, \mathcal{T}_n^{-1}(1-\alpha/2),\ \overline{x_i} + \frac{\mathsf{s}}{\sqrt{n}}\, \mathcal{T}_{n-1}^{-1}(1-\alpha/2)\right]$$

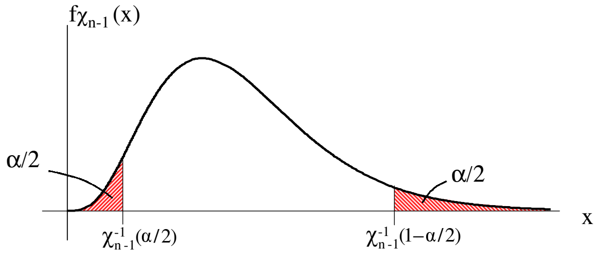


Figure 145: *Exact confidence interval for mean for unknown variance in a Gaussian sample is given by quantiles of student-t distribution with* $n-1$ *d.f.*

10.4 CONFIDENCE ON VARIANCE

Objective: find conf. interval on variance σ^2 based on i.i.d. Gaussian sample $\underline{x} = [x_1, \ldots, x_n]$ with unknown mean μ.

Solution: Recall the double sided hypothesis GLRT for variance

$$H_0 : \sigma^2 = \sigma_o^2,$$

$$H_1 : \sigma^2 \neq \sigma_o^2,$$

we found level α Chi-square test for Gaussian x_i's:

$$\chi_{n-1}^{-1}(\alpha/2) \leq \frac{(n-1)\mathsf{s}^2}{\sigma_o^2} \leq \chi_{n-1}^{-1}(1-\alpha/2)$$

Therefore, an exact $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\left[\frac{(n-1)\mathsf{s}^2}{\chi_{n-1}^{-1}(1-\alpha/2)}, \frac{(n-1)\mathsf{s}^2}{\chi_{n-1}^{-1}(\alpha/2)}\right]$$

Note: confidence interval for variance is not symmetric about s^2

10.5 CONFIDENCE ON DIFFERENCE OF TWO MEANS

Objective: find conf. interval on the difference $\Delta = \mu_x - \mu_y$ of means in two i.i.d. Gaussian samples $\underline{x} = [x_1, \ldots, x_{n_1}]$, $\underline{y} = [y_1, \ldots, y_{n_2}]$
Solution: consider the hypotheses

$$\begin{aligned} H_0 : \Delta &= \Delta_o \\ H_1 : \Delta &\neq \Delta_o \end{aligned}$$

A GLRT of level α would give a confidence interval for Δ similiarly to before.

Recall: the t test

$$\frac{\sqrt{\frac{n_1 n_2}{n}}|\overline{x_i} - \overline{y_i}|}{\mathsf{s}_2} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \mathcal{T}_{n-2}^{-1}(1-\alpha/2)$$

was previously derived for double sided hypotheses

$$\begin{aligned} H_0^{'} : \mu_x &= \mu_y \\ H_1^{'} : \mu_x &\neq \mu_y \end{aligned}$$

Difficulty: Δ does not appear any where in the test statistic

* $\overline{x_i} - \overline{y_i}$ has mean Δ under $H_0 : \Delta = \Delta_o$

* therefore distribution of t-test statistic above depends on Δ and is not a pivot

However, as $\overline{x_i} - \overline{y_i} - \Delta$ has mean zero under H_0 and same variance as $\overline{x_i} - \overline{y_i}$, we can immediately identify pivot

$$\frac{\sqrt{\frac{n_1 n_2}{n}}(\overline{x_i} - \overline{y_i} - \Delta)}{\mathsf{s}_2} \sim \mathcal{T}_{n-2}$$

Thus, the left and right endpoints of a $100(1-\alpha)\%$ conf. interval on Δ are given by

$$\overline{x_i} - \overline{y_i} \mp \sqrt{\frac{n}{n_1 n_2}}\, \mathsf{s}_2\, \mathcal{T}_{n-2}^{-1}(1-\alpha/2)$$

10.6 CONFIDENCE ON DIFFERENCE OF TWO VARIANCES

Objective: find conf. interval on the ratio

$$c = \sigma_x^2/\sigma_y^2$$

of variances in two i.i.d. Gaussian samples

* $\underline{x} = [x_1, \ldots, x_{n_1}]$, $\underline{y} = [y_1, \ldots, y_{n_2}]$

Solution: Recall that the GLRT for double sided $H_1 : \sigma_x^2 \neq \sigma_y^2$ was F-test

$$\mathcal{F}^{-1}_{n_1-1,n_2-1}(\alpha/2) \leq \frac{\mathsf{s}_x^2}{\mathsf{s}_y^2} \leq \mathcal{F}^{-1}_{n_1-1,n_2-1}(1-\alpha/2)$$

Difficulty: distribution of test statistic depends on $c = \sigma_x^2/\sigma_y^2$

However, as

$$\frac{1}{c}\frac{\mathsf{s}_x^2}{\mathsf{s}_y^2} = \frac{\mathsf{s}_x^2/\sigma_x^2}{\mathsf{s}_y^2/\sigma_y^2} \sim \mathcal{F}_{n_1-1,n_2-1}$$

we have identified a pivot.

Therefore, a $(1-\alpha)\%$ conf. interval on variance ratio $c = \sigma_x^2/\sigma_y^2$ is given by

$$[T_1, T_2] = \left[\frac{\mathsf{s}_x^2}{\mathsf{s}_y^2\ \mathcal{F}^{-1}_{n_1-1,n_2-1}(1-\alpha/2)}, \frac{\mathsf{s}_x^2}{\mathsf{s}_y^2\ \mathcal{F}^{-1}_{n_1-1,n_2-1}(\alpha/2)}\right]$$

Figure 146: *Confidence interval on variance ratio in a pair of Gaussian samples depends on quantiles of F-distribution* $\mathcal{F}_{n_1-1,n_2-1}$

10.7 CONFIDENCE ON CORRELATION COEFFICIENT

Objective: find conf. interval on correlation coefficient ρ between two i.i.d. Gaussian samples $\underline{x} = [x_1, \ldots, x_n]$, $\underline{y} = [y_1, \ldots, y_n]$ with unknown means and variances.

NOTE: not obvious how to obtain pivot from previously derived GLRT test statistic for testing $H_1 : \rho \neq 0$.

Solution: Fisher Transformation.

Let $\hat{\rho}$ be sample correlation coefficient. Then

$$\upsilon = \tfrac{1}{2} \ln\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right) = \tanh^{-1}(\hat{\rho})$$

has an asymptotic normal dsn with

$$E_\theta[\upsilon] = \tanh^{-1}(\rho)$$
$$\mathrm{var}_\theta(\upsilon) = \frac{1}{n-3}$$

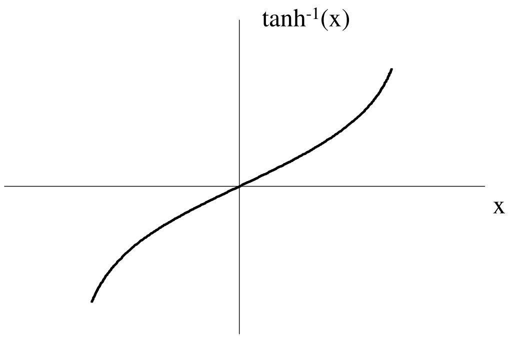


Figure 147: *Inverse tanh function is monotone increasing.*

Hence we have a pivot

$$T(\underline{x}) = \frac{\tanh^{-1}(\hat{\rho}) - \tanh^{-1}(\rho)}{1/\sqrt{n-3}} \sim \mathcal{N}(0,1)$$

This gives $100(1-\alpha)\%$ conf. interval on $\tanh^{-1}(\rho)$

$$\left[\upsilon - \frac{1}{\sqrt{n-3}}\,\mathcal{N}^{-1}(1-\alpha/2),\ \upsilon + \frac{1}{\sqrt{n-3}}\,\mathcal{N}^{-1}(1-\alpha/2)\right]$$

Since $\tanh^{-1}(\cdot)$ is monotone, left and right endpoints T_1, T_2 of $(1-\alpha)\%$ conf. interval $[T_1, T_2]$ on ρ are

$$\tanh\left(\upsilon \mp \frac{1}{\sqrt{n-3}}\,\mathcal{N}^{-1}(1-\alpha/2)\right)$$

OBSERVATIONS:

1. Conf. interval is not symmetric
2. Conf. interval prescribes a level α test of double sided hypotheses

$$\begin{aligned} H_0 : \rho &= \rho_o, \\ H_1 : \rho &\neq \rho_o \end{aligned}$$

which is

$$\phi(\underline{x}) = \begin{cases} 1, & \rho_o \notin [T_1, T_2] \\ 0, & \rho_o \in [T_1, T_2] \end{cases}$$

Indeed:

$$\begin{aligned} E_0[\phi] &= 1 - P_{\rho_o}(T_1 \leq \rho_o \leq T_2) \\ &= 1 - (1-\alpha) = \alpha \end{aligned}$$

10.8 Exercises

1. Let $\{X_i\}_{i=1}^n$ be an i.i.d. sample with marginal p.d.f. $f(x;\theta) = \theta e^{-\theta x},\ x > 0$.
   (a) Show that the maximum likelihood estimator (MLE) for θ is $\frac{1}{\bar{X}}$, where $\bar{X}$ is the sample mean.
   (b) Show that the CR bound $I^{-1}(\theta)$ for θ given $\{X_i\}_{i=1}^n$ is of the form: $I^{-1}(\theta) = \frac{\theta^2}{n}$.
   (c) Now, using the fact that for large n the MLE is distributed as approximately $\mathcal{N}(\theta, I^{-1}(\theta))$, show that $\left(\frac{1/\bar{X}}{1+Z(1-\frac{\alpha}{2})/\sqrt{n}}, \frac{1/\bar{X}}{1-Z(1-\frac{\alpha}{2})/\sqrt{n}}\right)$ is a $(1-\alpha)\cdot 100\%$ confidence interval for θ, where $Z(p)$ is the $p-th$ quantile of $\mathcal{N}(0,1)$.
2. Let $\underline{\mathbf{X}} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]^T$ be a Gaussian random vector with mean $\underline{\mu} = [\underline{\mu}_1, \ldots, \underline{\mu}_n]^T$ and covariance matrix R_X.
   (a) Show that the distribution of $\mathbf{W} \stackrel{\text{def}}{=} (\underline{\mathbf{X}} - \underline{\mu})^T R_X^{-1} (\underline{\mathbf{X}} - \underline{\mu})$ is Chi-Square with n degrees of freedom (Hint: use square root factor $R_X^{\frac{1}{2}}$ to represent $(\underline{\mathbf{X}} - \underline{\mu})$ in terms of a vector of uncorrelated standard Gaussian variates).

(b) Since $\mathbf{W}$ has a distribution which is independent of $\underline{\mu}$ and R_X, $\mathbf{W}$ is similar to a pivot for scaler μ which can be used to generate confidence regions on $\underline{\mu}$. Assume $n = 2$ and let R_X be a fixed and known diagonal matrix with eigenvalues λ_1 and λ_2. Show that $\mathcal{R} \stackrel{\text{def}}{=} \{\underline{\mu} : (\underline{\mathbf{X}} - \underline{\mu})^T R_X^{-1} (\underline{\mathbf{X}} - \underline{\mu}) \leq -2 \ln \alpha\}$ is a $100(1-\alpha)\%$ confidence region for the vector $\underline{\mu}$ in the sense that: $P(\underline{\mu} \in \mathcal{R}) = P(W \leq -2 \ln \alpha) = 1 - \alpha$. Draw a concise picture of this confidence region for the case $\underline{\mu} \in \mathbb{R}^2$. Label and identify all quantities in your picture. What happens to the confidence region as $\lambda_1 \to 0$? Does this make sense?

3. *This exercise establishes that a pivot always exists when the marginal CDF is strictly increasing.* Let $\{X_i\}_{i=1}^n$ be an i.i.d. sample with marginal p.d.f. $f(x;\theta)$ and a CDF $F(x;\theta)$ which is strictly increasing: $F(x + \Delta; \theta + \delta) > F(x;\theta)$, $\Delta, \delta > 0$.

   (a) Show that the random variable $F(\mathbf{X}_i;\theta)$ has a uniform distribution over the interval $[0,1]$, and that therefore $-\log F(\mathbf{X}_i;\theta)$ has an exponential distribution $f(u) = e^{-u}$, $u > 0$.

   (b) Show that the CDF of the entire sample, $\prod_{i=1}^n F(\mathbf{X}_i;\theta)$ is a pivot for θ. (Hint: the sum of n i.i.d. exponential r.v.s with distribution $f(u) = e^{-u}$, $u > 0$ has a Gamma density).

   (c) Show that a $(1-\alpha)\cdot 100\%$ confidence interval for θ can be constructed since $F(x;\theta)$ is monotone in θ using the result of part (b). (Hint: the product of monotone functions is monotone).

4. Use the approach of the previous problem to construct $(1-\alpha)100\%$ confidence intervals for the following parameters.

   (a) θ is the parameter in the density $f(x;\theta) = 2\theta x + 1 - \theta$, $0 \leq x \leq 1$, $-1 \leq \theta \leq 1$. Verify your results by numerical simulation for $n = 10$ using Matlab. Note it may be helpful to use Matlab's polynomial rooting procedure `roots.m` to find the interval endpoints. Note for this example you *cannot* use double sided GLRT since the GLRT is degenerate.

   (b) θ is the median of the Cauchy distribution $f(x_i, \theta) = (1 + (x - \theta)^2)/\pi$. Note that numerical integration may be required. Verify your results by numerical simulation for $n = 10$ using Matlab.

6. Let $\{x1, \ldots, x_n\}$ be an i.i.d. sample of a Poisson r.v. with distribution $p_\theta(k) = P_\theta(x_i = k) = \frac{\theta^k}{k!} e^{-\theta}$, $k = 0, 1, 2, \ldots$. Use the GLRT derived in Exercise 8.2 to specify a $(1-\alpha)\%$ confidence interval on θ.

End of chapter

11 DETECTION FOR SAMPLED GAUSSIAN WAVEFORMS

1. Offline methods:

* General vector Gaussian problem

* Detection of non-random signals in noise: matched-filter

* Detection of random signals in noise: filter-squarer and estimator-correlator

2. Online methods

* On line detection of non-random signals: causal matched-filter

* On-line detection for nonstationary signals: kalman filter detector

References

Van Trees [36]

Srinath, Rajasekaran and Viswanathan [33]

Whalen [37]

11.1 OFFLINE METHODS

Observation: $\underline{x} = [x_1, \ldots, x_n]^T \sim \mathcal{N}_n(\underline{\mu}, R)$
mean: $\underline{\mu} = E[\underline{x}] = [\mu_1, \ldots, \mu_n]^T$
covariance: $R = ((\mathrm{cov}(x_i, x_j)))_{i,j=1,\ldots,n}$

$$R = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \ldots & \sigma_n^2 \end{bmatrix}$$

Joint density

$$f(\underline{x};\ \underline{\mu}, R) = \frac{1}{(2\pi)^{n/2}\sqrt{|R|}} \exp\left(-\tfrac{1}{2}(\underline{x} - \underline{\mu})^T R^{-1} (\underline{x} - \underline{\mu})\right)$$

Consider simple detection problem

$$\begin{aligned} H_0 &: \underline{\mu} = \underline{\mu}_0, \quad R = R_0 \\ H_1 &: \underline{\mu} = \underline{\mu}_1, \quad R = R_1 \end{aligned}$$

Likelihood ratio is

$$\Lambda(\underline{x}) = \frac{\sqrt{|R_0|} \exp\left(-\tfrac{1}{2}(\underline{x} - \underline{\mu}_1)^T R_1^{-1} (\underline{x} - \underline{\mu}_1)\right)}{\sqrt{|R_1|} \exp\left(-\tfrac{1}{2}(\underline{x} - \underline{\mu}_0)^T R_0^{-1} (\underline{x} - \underline{\mu}_0)\right)}$$

Giving LRT

$$T(\underline{x}) =$$

$$\tfrac{1}{2}(\underline{x}-\underline{\mu}_0)^T R_0^{-1}(\underline{x}-\underline{\mu}_0) - \tfrac{1}{2}(\underline{x}-\underline{\mu}_1)^T R_1^{-1}(\underline{x}-\underline{\mu}_0) \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

where

$$\gamma = \log\eta + \tfrac{1}{2}\log\frac{|R_1|}{|R_0|}$$

Interpretation of LRT as distorted "Mahalanobis" distance test:

Define two norms in $\mathbb{R}^n$

$$\|\underline{z}\|_{R_0} = \underline{z}^T R_0^{-1}\underline{z}, \qquad \|\underline{z}\|_{R_1} = \underline{z}^T R_1^{-1}\underline{z},$$

Norm $\|\underline{z}\|_R$ emphasizes components of $\underline{z}$ which are colinear to eigenvectors of R associated with small eigenvalues

Then MP-LRT takes form of a comparison between weighted distances of $\underline{x}$ to $\underline{\mu}_0$ vs. $\underline{\mu}_1$

$$\|\underline{x}-\underline{\mu}_0\|^2_{R_0} - \|\underline{x}-\underline{\mu}_1\|^2_{R_1} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma'$$

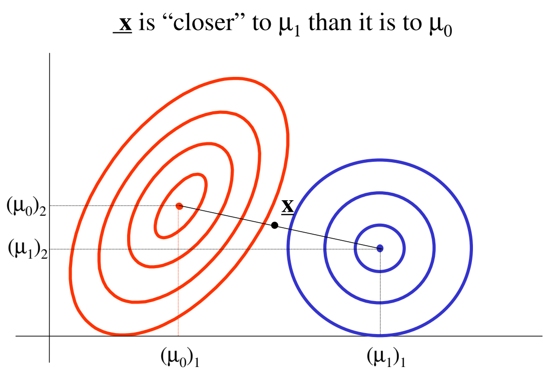


Figure 148: *LRT for general Gaussian problem compares "closeness" of $\underline{x}$ to distorted neighborhoods of the means μ_0 and μ_1.*

Alternative form of LRT as quadratic test

$$T'(\underline{x}) = \tfrac{1}{2}\underline{x}^T[R_0^{-1} - R_1^{-1}]\underline{x} + (\underline{\mu}_1^T R_1^{-1} - \underline{\mu}_0^T R_0^{-1})\underline{x} \overset{H_1}{\underset{H_0}{\gtrless}} \gamma'$$

$$\gamma' = \log \eta + \tfrac{1}{2}\log\frac{|R_1|}{|R_0|} + \underline{\mu}_1^T R_1^{-1}\underline{\mu}_1 - \underline{\mu}_0^T R_0^{-1}\underline{\mu}_0$$

11.1.1 GENERAL CHARACTERIZATION OF LRT DECISION REGIONS

Divide treatement into four cases:

1. $R_0 = R_1$,
2. $R_0 - R_1 > 0$,
3. $R_0 - R_1 < 0$,
4. $R_0 - R_1$ non-singular

Case 1. $R_0 = R_1 = R$:

In this case $T'(\underline{x}) = \underline{a}^T\underline{x}$ is linear function

$$\underline{a} = R^{-1}(\underline{\mu}_1 - \underline{\mu}_0),$$

and decision regions are separated by a hyperplane.

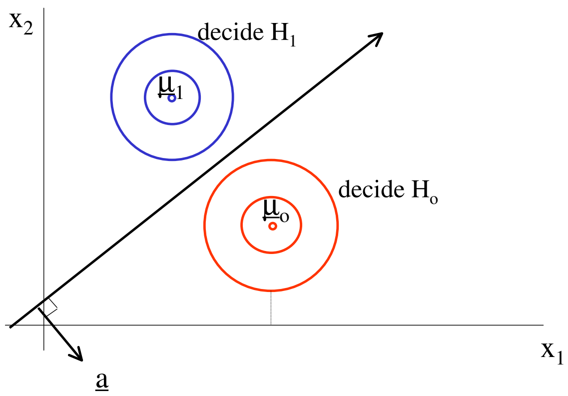


Figure 149: *For equal covariances of a multivariate Gaussian sample the decision regions are separated by a hyperplane (here shown for $\gamma' = 0$ for which a is orthogonal to separating hyperplane).*

Case 2. $R_0 - R_1 > 0$: (p.d.)
In this case, as $R_0 > R_1$ implies $R_0^{-1} < R_1^{-1}$,
$R_1^{-1} - R_0^{-1} = \Delta_{10} R^{-1} > 0$ (p.d.):
and

$$T^{'}(\underline{x}) = -\tfrac{1}{2}(\underline{x} - \underline{b})^T \Delta_{10} R^{-1} (\underline{x} - \underline{b}) + c$$

$$\underline{b} \quad = \quad (\Delta_{10} R^{-1})^{-1} [R_1^{-1} \underline{\mu}_1 - R_0^{-1} \underline{\mu}_0]$$

Hence the H_1 decision region is an ellipsoid

$$\mathcal{X}_1 = \{\tfrac{1}{2}(\underline{x} - \underline{b})^T [\Delta_{10} R^{-1}] (\underline{x} - \underline{b}) < \gamma^{''}\}$$

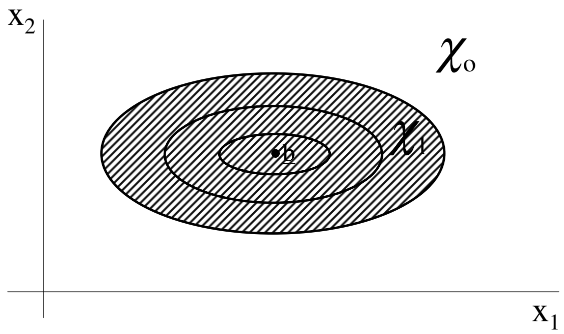


Figure 150: *For $R_0 > R_1$ the H_1 decision region is the interior of an ellipsoid for testing covariance of a multivariate Gaussian sample.*

Case 3. $R_0 < R_1$
In this case
$R_0^{-1} - R_1^{-1} = \Delta_{01} R^{-1} > 0$ (p.d.):
and

$$T^{'}(\underline{x}) = \tfrac{1}{2}(\underline{x} - \underline{b})^T \ [\Delta_{01} R^{-1}] \ (\underline{x} - \underline{b}) + c$$

$$\underline{b} \quad = \quad (\Delta_{01} R^{-1})^{-1} [R_0^{-1} \underline{\mu}_0 - R_1^{-1} \underline{\mu}_1]$$

So now the H_0 decision region is an ellipsoid

$$\mathcal{X}_0 = \{\tfrac{1}{2}(\underline{x} - \underline{b})^T \Delta_{01} R^{-1} (\underline{x} - \underline{b}) < \gamma''\}$$

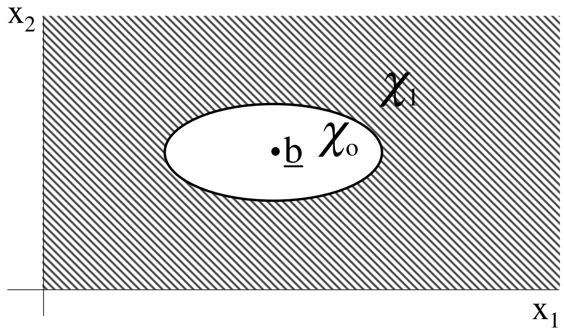


Figure 151: *For $R_0 < R_1$ the H_1 decision region for testing the covariance of a multivariate Gaussian sample is the exterior of an ellipsoid.*

4. $R_0 - R_1$ not p.d, n.d., or singular

Let $\Delta_{01} R^{-1}$ be defined as above

$$R_0^{-1} - R_1^{-1} =: \Delta_{01} R^{-1}$$

Let $\{\lambda_i\}_{i=1}^n$ denote the eigenvalues of this matrix

Let the *signature* of $\Delta_{01} R^{-1}$ be the binary vector

$$\underline{b} = [b_1, \ldots, b_n]^T$$

Definition: The signature of a non-singular symmetric matrix B are the signs of its eigenvalues arranged in decreasing order of magnitude.

⇒ If the signature $[b_1, \ldots, b_n]^T$ equals $[1, \ldots, 1]$ then all eigenvalues are positive and B is positive definite.

⇒ If $[b_1, \ldots, b_n]^T$ equals $[-1, \ldots, -1]$ then $-B$ is positive definite.

General representation:

$$[b_1, \ldots, b_n] = [\mathrm{sgn}(\lambda_1), \ldots, \mathrm{sgn}(\lambda_n)]$$

where

$$\mathrm{sgn}(u) := \begin{cases} 1, & u > 0 \\ 0, & u = 0 \\ -1 & u < 0 \end{cases}$$

We can rewrite the LRT test statistic as

$$\begin{aligned} T'(\underline{x}) &= \tfrac{1}{2}(\underline{x} - \underline{b})^T [\Delta_{01} R^{-1}](\underline{x} - \underline{b}) + c \\ &= b_1\ z_1^2 + \ldots + b_n\ z_n^2 + c \end{aligned}$$

where

$$z_i = \lambda_i(\underline{x} - \underline{b})^T \underline{\nu}_i$$

and $\underline{\nu}_i$'s are eigenvectors of $\Delta_{01} R^{-1}$.

Thus each decision region is hyperbolic.

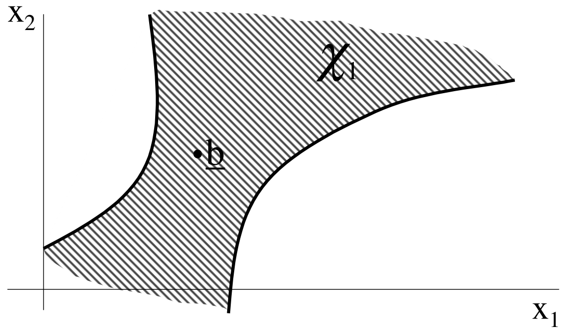


Figure 152: *For $R_0 - R_1$ non-singular but neither p.d. nor n.d. the H_1 decision region for testing the covariance of a multivariate Gaussian sample is a hyperboloid.*

11.1.2 CASE OF EQUAL COVARIANCES

Here $R_0 = R_1 = R$ and LRT collapses to linear test

$$T(\underline{x}) = \Delta\underline{\mu}^T R^{-1} \underline{x} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma_1$$

where $\Delta\underline{\mu} = (\underline{\mu}_1 - \underline{\mu}_0)$
and
$\gamma_1 = \log\eta + \underline{\mu}_0^T R^{-1}\underline{\mu}_0 - \underline{\mu}_1^T R^{-1}\underline{\mu}_1$
DETECTOR PERFORMANCE
As test statistic $T(\underline{x})$ is Gaussian suffices to find means

$$\begin{aligned} E_0[T] &= \Delta\underline{\mu}^T R^{-1}\underline{\mu}_0 \\ E_1[T] &= \Delta\underline{\mu}^T R^{-1}\underline{\mu}_1 \end{aligned}$$

and variances

$$\begin{aligned} \text{var}_0(T) &= \text{var}_0\left(\Delta\underline{\mu}^T R^{-1}\underline{x}\right) \\ &= \Delta\underline{\mu}^T R^{-1} \underbrace{\text{cov}_0(\underline{x})}_{R} R^{-1}\Delta\underline{\mu} \\ &= \Delta\underline{\mu}^T R^{-1}\Delta\underline{\mu} \\ \text{var}_1(T) &= \text{var}_0(T) \end{aligned}$$

Thus we find

$$P_F = \alpha = 1 - \mathcal{N}\left(\frac{\gamma_1 - E_0[T]}{\sqrt{\text{var}_0(T)}}\right)$$

so that the NP MP-LRT test is

$$\Delta\underline{\mu}^T R^{-1}\underline{x} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \sqrt{\Delta\underline{\mu}^T R^{-1}\Delta\underline{\mu}}\,\mathcal{N}^{-1}(1-\alpha) \;+\; \Delta\underline{\mu}^T R^{-1}\underline{\mu}_0$$

or equivalently

$$\frac{\Delta\underline{\mu}^T R^{-1}(\underline{x} - \underline{\mu}_0)}{\sqrt{\Delta\underline{\mu}^T R^{-1}\Delta\underline{\mu}}} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \mathcal{N}^{-1}(1-\alpha)$$

NOTES:
1. For $\underline{\mu}_0 \neq 0$: MP test is not UMP w.r.t unknown parameter variations
2. For $\underline{\mu}_0 = 0$: MP test is UMP w.r.t. constant scaling of $\underline{\mu}_1$
Next find power:

$$P_D = \beta = 1 - \mathcal{N}\left(\frac{\gamma_1 - E_1[T]}{\sqrt{\mathrm{var}_1(T)}}\right)$$

giving ROC curve:

$$\beta = 1 - \mathcal{N}\left(\mathcal{N}^{-1}(1-\alpha) - d\right)$$

where d is detectability index

$$\begin{aligned} d &= \frac{E_1[T] - E_0[T]}{\sqrt{\mathrm{var}_0(T)}} \\ &= \sqrt{\Delta\underline{\mu}^T R^{-1} \Delta\underline{\mu}} \end{aligned}$$

Example 38 *Detection of known signal in white noise*

$$\begin{aligned} H_0 : x_k &= w_k \\ & \qquad\qquad\qquad k = 1, \ldots, n \\ H_1 : x_k &= s_k + w_k \end{aligned}$$

* $\underline{w} \sim \mathcal{N}_n(0, \sigma^2 I)$,
* $\underline{s}$ and σ^2 known
Identify:

$$\underline{\mu}_0 = 0, \quad \Delta\underline{\mu} = \underline{s}, \quad R = \sigma^2 I$$

LRT takes form of a matched filter

$$T(x) = \underline{s}^T \underline{x} = \sum_{k=1}^n s_k x_k \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

$\gamma = \log \eta - \frac{1}{\sigma^2}\|\underline{s}\|^2$
GEOMETRIC INTERPRETATION
LRT can be expressed geometrically as a signal "projection" detector
Projection of $\underline{x}$ onto $\underline{s}$ is

$$\begin{aligned}
\underline{\hat{x}} &= \underbrace{\left[\frac{\underline{s}\,\underline{s}^T}{\|\underline{s}\|^2}\right]}_{\Pi_s} \underline{x} \\
&= \underline{s}\,\frac{\underline{s}^T\underline{x}}{\|\underline{s}\|^2} \\
&= \underline{s}\,\underbrace{\frac{<\underline{s},\underline{x}>}{\|\underline{s}\|^2}}_{\text{Proj. coef.}}
\end{aligned}$$

Length of this projection is

$$\begin{aligned}
\|\underline{\hat{x}}\| &= \|\underline{s}\| \left|\frac{\underline{s}^T\underline{x}}{\|\underline{s}\|^2}\right| \\
&= |T(\underline{x})|\,\frac{1}{\|\underline{s}\|}
\end{aligned}$$

Conclude:

* LRT is a threshold test on the projection coefficient of the orthogonal projection of $\underline{x}$ onto $\underline{s}$

* LRT is threshold test on "signed length" of $\underline{\hat{x}}$

* LRT is related to LLS estimator $\underline{\hat{x}}$ of $\underline{x}$ given $\underline{s}$

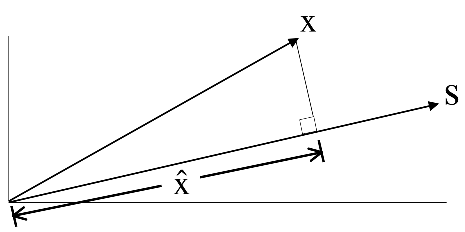


Figure 153: *MP detector applies a threshold to projection coefficient $< \underline{x}, \underline{s} > /\|\underline{s}\|^2$ of orthogonal projection of $\underline{x}$ onto $\underline{s}$, shown here for the case of $n = 2$.*

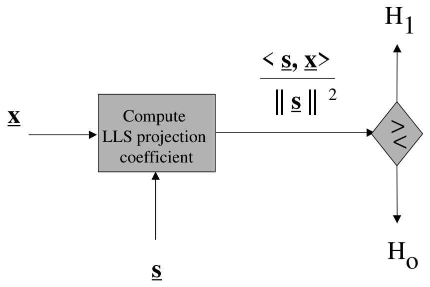


Figure 154: *MP detector block diagram implemented with a LLS estimator of $\underline{x}$.*

PERFORMANCE

Equivalently, for MP test of level α we have

$$\frac{\underline{s}^T \ \underline{x}}{\|\underline{s}\| \ \sigma} \ \overset{H_1}{\underset{H_0}{\gtrless}} \ \mathcal{N}^{-1}(1-\alpha)$$

This test is UMP relative to signal energy $\|\underline{s}\|^2$

Now compute detectability index:

$$d^2 = \|\underline{s}\|^2/\sigma^2 = \frac{\sum_{k=1}^n s_k^2}{\sigma^2} =: \mathrm{SNR} \tag{42}$$

NOTE:

* detection index is invariant to shape of waveform $\underline{s}$.

* detector power only depends on total signal-to-noise ratio (SNR).

Note: two ways to implement optimal detector

Cross-correlator:

Matched Filter:

Example 39 *Detection of known signal in non-white noise:*

$$\begin{aligned} H_0 : x_k &= w_k \\ & \qquad\qquad\qquad k = 1, \ldots, n \\ H_1 : x_k &= s_k + w_k \end{aligned}$$

* $\underline{w} \sim \mathcal{N}_n(0, R)$, R not scaled identity.

Optimal detector

$$T(\underline{x}) = \frac{\underline{s}^T R^{-1} \underline{x}}{\sqrt{\underline{s}^T R^{-1} \underline{s}}} \overset{H_1}{\underset{H_0}{\gtrless}} \mathcal{N}^{-1}(1-\alpha)$$

Q. How to modularize detector?

A. transform to the white noise case via preprocessing with matrix H

Produces white noise measurements

$$\underline{\tilde{x}} = H \cdot \underline{x}$$

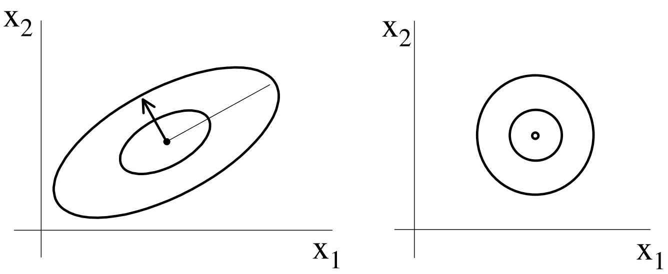


Figure 155: *Matrix prewhitener H applied to $\underline{x}$ renders the transforms contours of the multivariate Gaussian density to concentric circles (spherically symmetric).*

We will require matrix filter H to have properties:

1. $\mathrm{cov}_0(\underline{\tilde{x}}) = \mathrm{cov}_1(\underline{\tilde{x}}) = I$: $\Rightarrow$ whitening property

2. H is invertible matrix: $\Rightarrow$ output remains sufficient statistic

MATRIX FACTORIZATION

For any positive definite covariance matrix R there exists symmetric positive definite square root factors $R^{\frac{1}{2}}$ and $R^{-\frac{1}{2}}$ which satisfy:

$$R = R^{\frac{1}{2}} R^{\frac{1}{2}}, \text{ and } R^{-1} = R^{-\frac{1}{2}} R^{-\frac{1}{2}}$$

Furthermore

$$R^{\frac{1}{2}} R^{-\frac{1}{2}} = I$$

Proof via eigendecomposition: $R = UDU^T$

* $D = \text{diag}(\lambda_i)$ are (positive) eigenvalues of R

* $U = [\underline{\nu}_1, \ldots, \underline{\nu}_p]$ are (orthogonal) eigenvectors of R

Hence: as $U^TU = I$

$$R = UDU^T = UD^{\frac{1}{2}}D^{\frac{1}{2}}U^T = UD^{\frac{1}{2}}U^T\, UD^{\frac{1}{2}}U^T$$

Therefore we identify

$$R^{\frac{1}{2}} = UD^{\frac{1}{2}}U^T$$

Furthermore, since $U^{-1} = U^T$ we have

$$R^{-\frac{1}{2}} = UD^{-\frac{1}{2}}U^T$$

which satisisfy all of the desired properties of square root factors.

◇

Using square root factors the test statistic can be rewritten as

$$\begin{aligned} T(\underline{x}) &= \frac{\underline{s}^T R^{-\frac{1}{2}} R^{-\frac{1}{2}} \underline{x}}{\sqrt{\underline{s}^T R^{-\frac{1}{2}} R^{-\frac{1}{2}} \underline{s}}} \\ &= \frac{\underline{\tilde{s}}^T \underline{\tilde{x}}}{\|\underline{\tilde{s}}\|} \end{aligned}$$

Where $\underline{x}$, $\underline{s}$ are transformed vectors

$$\underline{\tilde{x}} = R^{-\frac{1}{2}}\underline{x}, \quad \underline{\tilde{s}} = R^{-\frac{1}{2}}\underline{s}$$

Now we see that

$E_0[\underline{\tilde{x}}] = 0, \quad E_1[\underline{\tilde{x}}] = \underline{\tilde{s}}, \quad \text{cov}_0(\underline{\tilde{x}}) = \text{cov}_1(\underline{\tilde{x}}) = I$

so that problem is equivalent to testing for a signal in white noise of unit variance.

Detectability index for non-white noise:

Note: $d^2 = \|\underline{\tilde{s}}\|^2 = \underline{s}^T R^{-1} \underline{s}$.

Remark

No longer is detection performance independent of shape of $\underline{s}$

OPTIMAL SIGNAL DESIGN FOR NON-WHITE NOISE:

Constraint: $\|\underline{s}\|^2 = 1$

Maximize: $d^2 = \underline{s}^T R^{-1} \underline{s}$

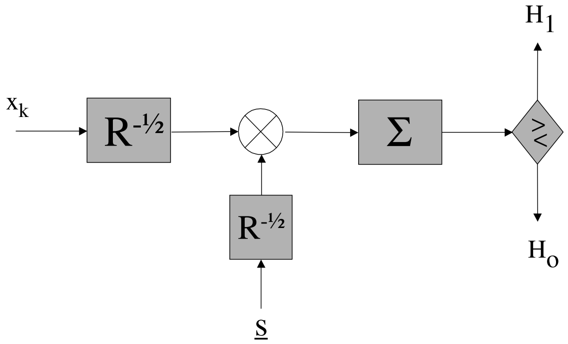


Figure 156: *Matrix prewhitener $H = R^{-\frac{1}{2}}$ is applied prior to optimal matched filter detection.*

Solution: Rayleigh quotient theorem specifies:

$$\frac{\underline{s}^T R^{-1} \underline{s}}{\underline{s}^T \underline{s}} \leq \frac{1}{\min_i \lambda_i^R}$$

$\lambda_i^R =$ an eigenvalue of R.
Furthermore

$$\frac{\underline{s}^T R^{-1} \underline{s}}{\underline{s}^T \underline{s}} = \frac{1}{\min_i \lambda_i^R}$$

when $\underline{s}$ is (any) minimizing eigenvector of R (there will be multiple minimizing eigenvectors if more than one eigenvalue $\{\lambda_k\}$ equals $\min_i \lambda_i^R$).

Example 40 *Application: (Real) Signal detection in a sensor array*

k-th snapshot of p-sensor array output is a *multi-channel* measurement:

$$\underline{x}_k = \underline{a}\, s + \underline{v}_k, \quad k = 1, \ldots, n$$

or equivalently we have $p \times n$ measurement matrix

$$X = [\underline{x}_1, \ldots, \underline{x}_n]$$

* $\underline{a}$: array response vector

* $\underline{v}_k$: Gaussian $\mathcal{N}_p(0, R)$, known spatial covariance R

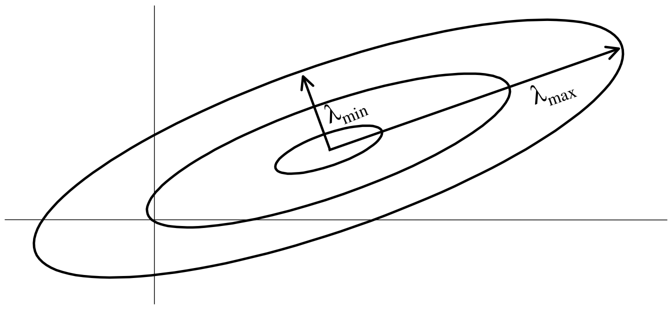


Figure 157: *The optimal signal which maximizes detectability is the eigenvector of noise covariance R with minimum eigenvalue.*

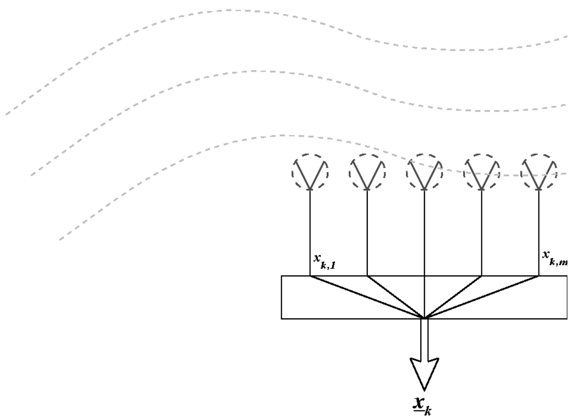


Figure 158: *Sensor array receiving signal wavefront generates spatio-temporal measurement.*

* s: deterministic signal amplitude

Three cases of interest:

1. Detection of known signal amplitude
2. Detection of positive signal amplitude
3. Detection of non-zero signal amplitude

Case 1: Known signal amplitude

$$\begin{aligned} H_0 : s &= 0, \quad k = 1, \ldots, n \\ H_1 : s &= s_1, \quad k = 1, \ldots, n \end{aligned}$$

Approach: reduce to single-channel problem via coordinate rotation

As $\underline{a}$, R are known, we can transform the array to one with

* spatially uncorrelated noise (R diagonal)

* signal energy present only in first channel.

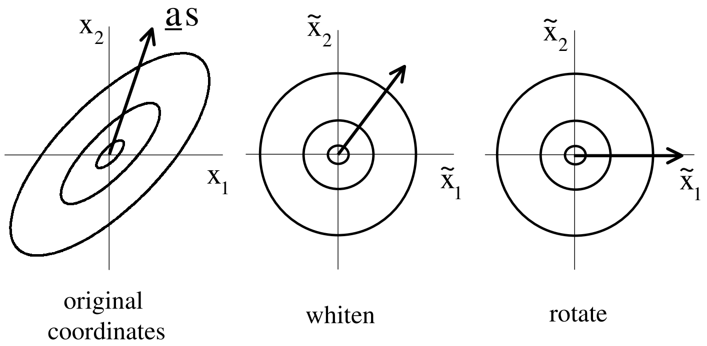


Figure 159: *Transformation of Gaussian multichannel problem to a Gaussian single channel problem is a two step procedure. First a whitening coordinate transformation $R^{-\frac{1}{2}}$ is applied to measurements $\underline{x} = \underline{a}s + \underline{n}$ (joint density in original coordinates is shown in left panel) which makes noise component $\underline{n}$ i.i.d. (transformed measurements have joint density with spherically symmetric constant contours shown in middle panel). Then a pure rotation (unitary matrix) H is applied to the transformed measurements $\tilde{\underline{x}}$ which aligns its signal component $R^{-\frac{1}{2}}\underline{a}s$ with the first coordinate axis (right panel).*

Define $p \times p$ matrix H:

$$H = \left[\frac{1}{\tilde{a}} R^{-\frac{1}{2}} \underline{a}, \underline{\nu}_2, \ldots, \underline{\nu}_p \right]$$

where

* $\tilde{a} = \sqrt{\underline{a}^T R^{-1} \underline{a}}$

* $\underline{\nu}_i$ orthonormal vectors orthogonal to $R^{-\frac{1}{2}}\underline{a}$ (found via Gramm-Schmidt)

Then

$$\underbrace{H^T R^{-\frac{1}{2}}}_{W} \underline{a} = \begin{bmatrix} \tilde{a} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \tilde{a}\, \underline{e}_1$$

Now as W is invertible, the following is equivalent measurement

$$\begin{aligned} \tilde{\underline{x}}_k &= W\underline{x}_k, \\ &= sW\underline{a} + W\underline{v}_k, \\ &= s_1\tilde{a}\, \underline{e}_1 + \tilde{\underline{v}}_k \end{aligned}$$

where $\tilde{\underline{v}}_k$'s are i.i.d. zero mean Gaussian with identity covariance

$$\text{cov}(\tilde{\underline{v}}_k) = WRW^T = H^T\ R^{-\frac{1}{2}} R R^{-\frac{1}{2}}\ H = H^T H = I$$

Matrix representation

$$\tilde{X} = s_1\ \tilde{a}\underline{1}^T + V$$

where V is a $p \times n$ matrix of i.i.d. $\mathcal{N}(0,1)$'s

Note properties:

* all rows of $\tilde{X} = WX$ are independent

* only first row $\tilde{X}_{1*} = \underline{e}_1^T \tilde{X}$ depends on s_1

Therefore LR only depends on the first row $\tilde{X}$

$$\begin{aligned} \Lambda(\tilde{X}) &= \frac{f(\tilde{X}_{1*}, \tilde{X}_{2*}, \ldots, \tilde{X}_{n*};\ s = s_1)}{f(\tilde{X}_{1*}, \tilde{X}_{2*}, \ldots, \tilde{X}_{n*};\ s = 0)} \\ &= \frac{f(\tilde{X}_{1*};\ s = s_1)}{f(\tilde{X}_{1*};\ s = 0)} \underbrace{\prod_{i=2}^{p} \frac{f(\tilde{X}_{i*};\ s = s_1)}{f(\tilde{X}_{i*};\ s = 0)}}_{=1} \\ &= \frac{f(\tilde{X}_{1*};\ s = s_1)}{f(\tilde{X}_{1*};\ s = 0)} = \Lambda(\tilde{X}_{1*}) \end{aligned}$$

Thus we have reduced the problem to equivalent hypotheses that a (row) vector measurement $\underline{z}^T = \tilde{X}_{1*}$ contains a constant signal in i.i.d. Gaussian noise of variance 1

$$\begin{array}{lcl} H_0 : \underline{z} = \underline{\tilde{v}} & & H_0 : z_k = \tilde{v}_k \\ & \Leftrightarrow & \\ H_1 : \underline{z} = s_1\, \tilde{a}\, \underline{1} + \underline{\tilde{v}} & & H_1 : z_k = s_1\, \tilde{a} + \tilde{v}_k \end{array}$$

The LRT follows immediately from our previous work in detection of constant signal $\mu = s_1 \tilde{a}$

$$s_1 \tilde{a}\, \overline{z_i} \overset{H_1}{\underset{H_0}{\gtrless}} \gamma$$

Or, as $\tilde{a}$ is positive, the final form of the LRT is

$$\begin{array}{rcl} T(z) = \sqrt{n}\, \overline{z_i} & \overset{H_1}{\underset{H_0}{\gtrless}} & \mathcal{N}^{-1}(1-\alpha) \quad s_1 > 0, \\ \sqrt{n}\, \overline{z_i} & \overset{H_0}{\underset{H_1}{\gtrless}} & -\mathcal{N}^{-1}(1-\alpha) \quad s_1 < 0, \end{array}$$

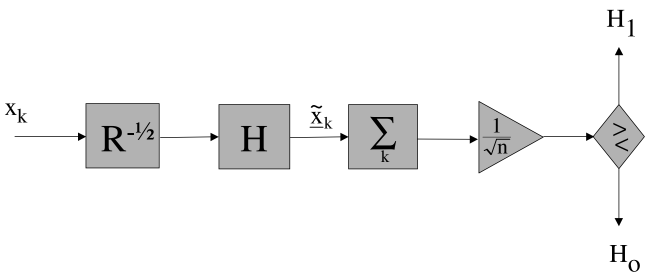


Figure 160: *LRT for detecting presence of a spatio-temporal signal implemented with whitening and coordinate rotation preprocessing.*

The power of the test is determined by the detectibility indez

$$d = \frac{|E[\overline{z_i}|H_1]|}{\sqrt{\text{var}(\overline{z_i}|H_0)}} = \sqrt{n}\, |s_1 \tilde{a}| = \sqrt{n}\, |s_1|\, \sqrt{\underline{a}^T R^{-1} \underline{a}}$$

We can express LRT in original coordinates by identifying

$$
\begin{aligned}
\underline{z}^T &= \tilde{X}_{1*} = \underline{e}_1^T \tilde{X} = \underline{e}_1^T \underbrace{W}_{H^T R^{-\frac{1}{2}}} X \\
&= \underbrace{\frac{1}{\sqrt{\underline{a}^T R^{-1} \underline{a}}}\, \underline{a}^T R^{-\frac{1}{2}}}_{\underline{e}_1^T H^T} R^{-\frac{1}{2}} X \\
&= \frac{1}{\sqrt{\underline{a}^T R^{-1} \underline{a}}}\, \underline{a}^T R^{-1} X
\end{aligned}
$$

and the identity

$$
\overline{z_i} = (\underline{z}^T \, \underline{1}) \, \frac{1}{n}
$$

to obtain $(s_1 > 0)$

$$
T(z) = \frac{1}{\sqrt{n \underline{a}^T R^{-1} \underline{a}}}\, \underline{a}^T R^{-1} X \underline{1} \overset{H_1}{\underset{H_0}{\gtrless}} \mathcal{N}^{-1}(1-\alpha),
$$

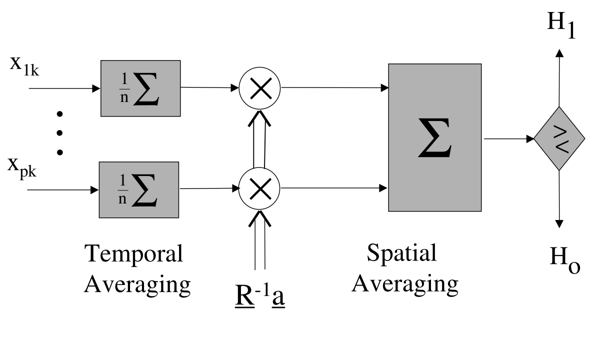


Figure 161: *LRT for detecting presence of a spatio-temporal signal implemented without coordinate transformation preprocessing.*

OBSERVATIONS

1. The LRT above is UMP w.r.t. any positive amplitude s_1

2. A modified LRT is UMP w.r.t. any negative amplitude s_1

3. The detectibility index

$$d = \sqrt{n}|s_1| \underbrace{\sqrt{\underline{a}^T R^{-1} \underline{a}}}_{\mathrm{ASNR}}$$

depends on normalized array SNR = ASNR

$\Rightarrow$ ASNR depends only on $\|\underline{a}\|$ when noise $\underline{v}_k$ is spatially white ($R = \sigma^2 I$).

4. Coherent interferers can severely degrade performance

Case 2: Unknown signal amplitude

$$H_0 : s = 0, \quad k = 1, \ldots, n$$

$$H_1 : s \neq 0, \quad k = 1, \ldots, n$$

No UMP exists!

Solution: double sided GLRT

$$|T(z)| = \sqrt{n}|\overline{z_i}| = \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \mathcal{N}^{-1}(1 - \alpha/2)$$

1. Implementation of GLRT via signal subspace projection:

Projection of $\underline{z}$ onto $\underline{s} = n^{-1}\ \underline{1}$ is

$$\begin{aligned} \hat{\underline{z}} &= \underbrace{\left[\frac{\underline{s}\ \underline{s}^T}{\|\underline{s}\|^2}\right]}_{\Pi_s} \underline{z} \\ &= \underline{s}\ \frac{\overbrace{\underline{s}^T \underline{z}}^{\overline{z_i}}}{\|\underline{s}\|^2} \end{aligned}$$

* Π_s = signal subspace projection operator

Length of $\hat{\underline{z}}$ is

$$\begin{aligned} \|\hat{\underline{z}}\| &= \|\underline{s}\| \left|\frac{\underline{s}^T \underline{z}}{\|\underline{s}\|^2}\right| \\ &= |\overline{z_i}|\ \frac{1}{\|\underline{s}\|} \\ &= |\overline{z_i}| \end{aligned}$$

Conclude:

* GLRT is a threshold test on the length of the orthogonal projection of $\underline{z}$ onto $span(\underline{s})$

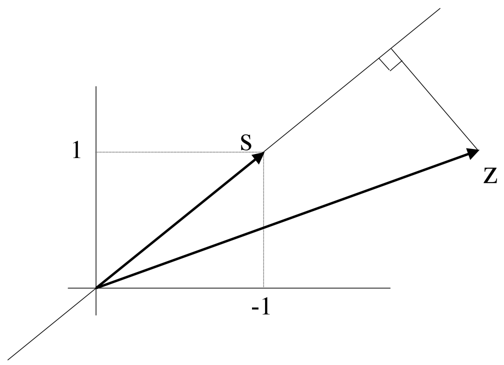


Figure 162: *GLRT detector thresholds length of orthogonal projection of $\underline{z}$ onto $\underline{s}$, shown here for the case of $n = 2$.*

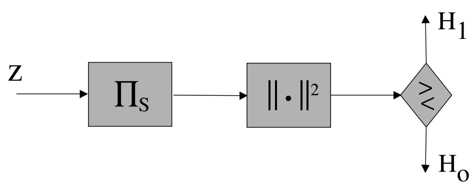


Figure 163: *GLRT detector block diagram implemented via signal subspace projection.*

2. Implementation of GLRT via "noise subspace" projection:

Recall orthogonal decomposition

$$\underline{z} = \Pi_s z + [I - \Pi_s]\underline{z}$$

* Π_s = signal subspace projection operator
* $I - \Pi_s$ = noise subspace projection operator

With this we can express GLRT as

$$|\overline{z_i}|^2 = \|\Pi_s \underline{z}\| = \|\underline{z}\|^2 - \| \underbrace{[I - \Pi_s]\underline{z}}_{\underline{z}-\hat{\underline{z}}} \|^2 \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma'$$

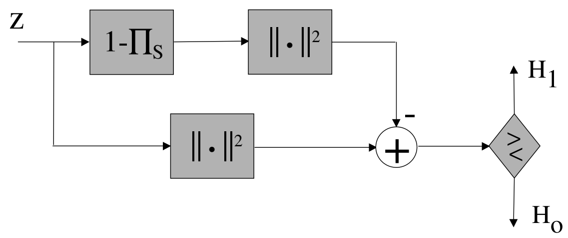


Figure 164: *GLRT detector block diagram implemented via noise subspace projection.*

11.1.3 CASE OF EQUAL MEANS, UNEQUAL COVARIANCES

Here $\underline{\mu}_0 = \underline{\mu}_1 = \underline{\mu}$ and LRT collapses to purely quadratic test

$$T(\underline{x}) = (\underline{x} - \underline{\mu})^T [R_0^{-1} - R_1^{-1}](\underline{x} - \underline{\mu}) \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

where $\gamma = 2 \ln \eta$.

Analysis will be simplified by prefiltering to diagonalize $R_0^{-1} - R_1^{-1}$

$\Rightarrow$ Require prefilter to perform simultaneous diagonalization

PERFORMANCE ANALYSIS

Under H_0 reexpress $(\underline{x} - \underline{\mu})^T [R_0^{-1} - R_1^{-1}](\underline{x} - \underline{\mu})$ as

$$T(\underline{x}) \quad = \quad \underbrace{(\underline{x} - \underline{\mu})^T R_0^{-\frac{1}{2}}}_{=\underline{z}^T \sim \mathcal{N}_n(0,I)} [I - R_0^{\frac{1}{2}} R_1^{-1} R_0^{\frac{1}{2}}] \underbrace{R_0^{-\frac{1}{2}} (\underline{x} - \underline{\mu})}_{=\underline{z} \sim \mathcal{N}_n(0,I)}$$

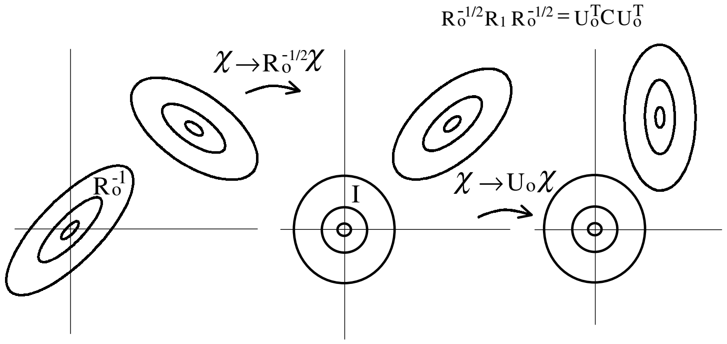


Figure 165: *Illustration of simultaneous whitening of two Gaussian data sets, or equivalently simultaneous diagonalization of two p.d. matrices as a two stage procedure. First one of the matrices is diagonalized and scaled to have identical diagonal elements via appropriate coordinate transformation (superposition of the constant contours of the two densities is shown on left panel along with the result of the coordinate transformation in middle panel). Then a unitary transformation is applied to diagonalize the other matrix without affecting the transformed first matrix (constant contours of the two densities shown on right panel).*

Now let $R_0^{\frac{1}{2}} R_1^{-1} R_0^{\frac{1}{2}}$ have eigendecomposition

$$R_0^{\frac{1}{2}} R_1^{-1} R_0^{\frac{1}{2}} = U_0^T C U_0$$

* U_0 orthogonal matrix of eigenvectors

* $C = \text{diag}(c_1, \ldots, c_n)$ diagonal matrix of eigenvalues.

$$\begin{aligned} T(\underline{x}) &= \underbrace{(U_0 \underline{z})^T}_{\mathcal{N}_n(0,I)} [I - C] \underbrace{(U_0 \underline{z})}_{\mathcal{N}_n(0,I)} \\ &= n\overline{(1-c)} \underbrace{\sum_i \frac{z_i^2 (1-c_i)}{\sum_j (1-c_j)}}_{\text{Chi-sq.-mixture}} \end{aligned}$$

where

$$\overline{(1-c)} = n^{-1} \sum_i (1 - c_i)$$

Two cases to consider: $0 \le c_i < 1$ vs $c_i > 1$.

CASE 1: $0 \leq c_i < 1$

Here

$$\overline{(1-c)} > 0$$

so we can absorb it into into threshold γ

This gives MP level α test in terms of orthogonalized measurements z_i

$$\sum_i \frac{z_i^2(1-c_i)}{\sum_j(1-c_j)} \overset{H_1}{\underset{H_0}{\gtrless}} \overline{\chi}^{-1}_{n,1-c}(1-\alpha)$$

$\overline{\chi}_{n,1-c}$ is CDF of Chi-square-mixture r.v. with n degrees of freedom and mixture parameter vector

$$1 - c = [1 - c_1, \ldots, 1 - c_n]^T$$

(Johnson, Kotz and Balakrishnan [13, Sec. 18.8]).

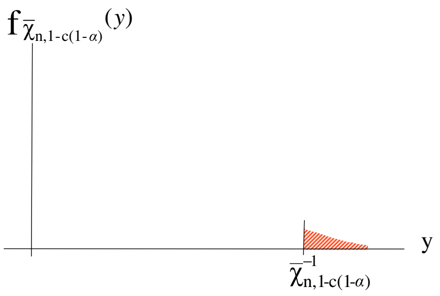


Figure 166: *For $c \leq c_i < 1$, threshold of test between two multivariate Gaussian models with identical means but unequal covariances is determined by quantile of Chi-square-mixture p.d.f.*

Next find power:

In a similar manner, under H_1 we can express

$$\begin{aligned} T(\underline{x}) &= (\underline{x} - \underline{\mu})^T R_1^{-\frac{1}{2}} [R_1^{\frac{1}{2}} R_0^{-1} R_1^{\frac{1}{2}} - I] R_1^{-\frac{1}{2}} (\underline{x} - \underline{\mu}) \\ &= (U_1 \underline{z})^T [C^{-1} - I](U_1 \underline{z}) \end{aligned}$$

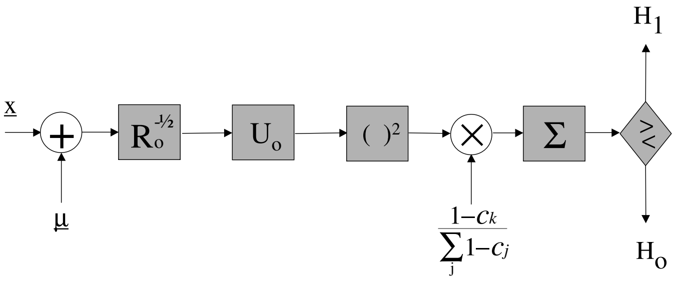


Figure 167: *An implementation of the MP-LRT for equal means unequal covariances using orthogonal prefilter U_0 obtained from eigendecomposition: $R_0^{\frac{1}{2}} R_1^{-1} R_0^{\frac{1}{2}} = U_0^T C U_0$, where C is diagonal.*

$$= \quad n\overline{(1/c-1)} \underbrace{\sum_i \frac{z_i^2 (1/c_i - 1)}{\sum_j (1/c_j - 1)}}_{Chi-sq.-mixture}$$

where U_1 in the above is an orthogonal matrix.

As $\overline{(1/c-1)} > 0$, we easily obtain power as:

$$\beta = 1 - \overline{\chi}_{n,1/c-1}(\rho\, \overline{\chi}^{-1}_{n,1-c}(1-\alpha))$$

where

$$\rho = \overline{(1-c)}/\overline{(1/c-1)}$$

CASE 2: $c_i > 1$

Here we have $\overline{(1-c)} < 0$ and the constant in the test can be absorbed into threshold only with change of the inequalities.

Obtain the MP level α test in z_i coordinates

$$\sum_i \frac{z_i^2 (1-c_i)}{\sum_j (1-c_j)} \quad \overset{H_0}{\underset{H_1}{\gtrless}} \quad \overline{\chi}^{-1}_{n,1-c}(\alpha)$$

and, using similar arguments as before, we obtain power curve

$$\beta = \overline{\chi}_{n,1/c-1}(\rho\, \overline{\chi}^{-1}_{n,1-c}(\alpha))$$

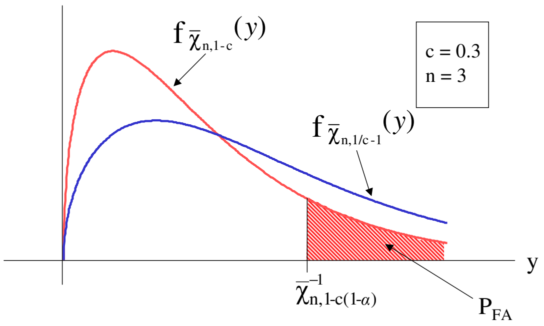


Figure 168: *For $c \leq c_i < 1$ ROC of test between two multivariate Gaussian models with identical means but unequal covariances is determined by upper quantiles of pair of Chi-square-mixture p.d.f.'s*

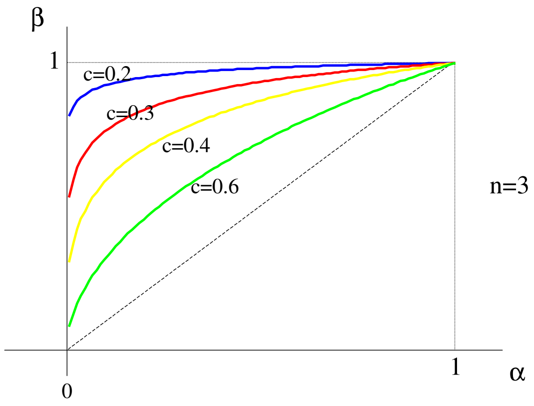


Figure 169: *ROC curve corresponding to Fig. 168 with variaous parameters c_1, c_2, c_3 for $n = 3$.*

11.2 APPLICATION: DETECTION OF RANDOM SIGNALS

Example 41 *Detection of shift in variance of white noise*

$$\begin{aligned} H_0 : x_k &= w_k \\ H_1 : x_k &= s_k + w_k \end{aligned}$$

$w_k \sim \mathcal{N}(0, \sigma_w^2)$: zero mean white noise
$s_k \sim \mathcal{N}(0, \sigma_s^2)$: zero mean white noise
w_k, s_k uncorrelated
Now

$$R_0 = \sigma_w^2 I, \quad R_1 = (\sigma_s^2 + \sigma_w^2) I$$

and

$$R_0^{-1} - R_1^{-1} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_w^2} \frac{1}{\sigma_w^2} I$$

Hence, defining

$$\mathrm{SNR} = \sigma_s^2 / \sigma_w^2$$

c_i is the constant

$$\begin{aligned} c_i &= \sigma_w^2 / (\sigma_w^2 + \sigma_s^2) = 1/(1 + \mathrm{SNR}) \\ 1 - c_i &= \mathrm{SNR}/(1 + \mathrm{SNR}) \\ 1/c_i - 1 &= 1/\mathrm{SNR} \\ \rho &= \overline{(1 - c)} / \overline{(1/c - 1)} = 1/(1 + \mathrm{SNR}) \end{aligned}$$

Note the SNR is defined differently in the case of a zero mean stochastic signal and a non-zero mean deterministic signal (42).

INTERPRETATION: $1 - c_i$ is the temporal "coherency function" (SNR normalized to interval $[0, 1]$)) of the signal w.r.t. the measurement

$$\kappa := \sigma_s^2 / (\sigma_w^2 + \sigma_s^2) = \mathrm{SNR}/(1 + \mathrm{SNR})$$

Thus LRT reduces to

$$T(\underline{x}) = \frac{\kappa}{\sigma_w^2} \sum_{k=1}^{n} x_k^2 \quad \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \quad \gamma$$

Which reduces to the Chi-squared test ("energy detector")

$$T'(\underline{x}) = \sum_{k=1}^{n} x_k^2/\sigma_w^2 \underset{H_0}{\overset{H_1}{\gtrless}} \chi_n^{-1}(1-\alpha)$$

NOTE: relation between Chi-square-mixture and Chi-square CDF's when $1 - c_i$ = constant

$$\overline{\chi}_{n,\overline{(1-c)}} = n^{-1}\chi_n$$

Power curve reduces to

$$\beta = 1 - \chi_n\left(\frac{1}{1+\mathrm{SNR}}\,\chi_n^{-1}(1-\alpha)\right)$$

Example 42 *Detection of uncorrelated non-stationary signal in noise*

$$\begin{aligned} H_0 : x_k &= w_k \\ H_1 : x_k &= s_k + w_k \end{aligned}$$

$w_k \sim \mathcal{N}(0, \sigma_w^2(k))$: uncorrelated noise samples
$s_k \sim \mathcal{N}(0, \sigma_s^2(k))$: uncorrelated signal samples
w_k, s_k uncorrelated
In this case

$$R_0 = \mathrm{diag}(\sigma_w^2(i)), \quad R_1 = \mathrm{diag}(\sigma_s^2(i) + \sigma_w^2(i))$$

and

$$\begin{aligned} R_0^{-1} - R_1^{-1} &= \mathrm{diag}\left(\frac{\sigma_s^2(i)}{\sigma_s^2(i)+\sigma_w^2(i)}\,\frac{1}{\sigma_w^2(i)}\right) \\ &= \mathrm{diag}\left(\kappa_i/\sigma_w^2(i)\right) \end{aligned}$$

where κ_i is time varying coherency function

$$\kappa_i = \frac{\sigma_s^2(i)}{\sigma_s^2(i)+\sigma_w^2(i)}$$

is

Hence, MP-LRT of level α reduces to

$$\frac{1}{\overline{\kappa}} \sum_{k=1}^{n} \kappa_k \, \frac{x_k^2}{\sigma_w^2(k)} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} = \overline{\chi}_{n,\kappa}^{-1}(1-\alpha)$$

or equivalently in terms of the original $T(\underline{x})$

$$T(\underline{x}) = \sum_{k=1}^{n} \kappa_k \, \frac{x_k^2}{\sigma_w^2(k)} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma = \overline{\kappa}\, \overline{\chi}_{n,\kappa}^{-1}(1-\alpha)$$

Special case of white noise: $\sigma_w^2(k) = N_o/2$

$$\sum_{k=1}^{n} \kappa_k \, x_k^2 \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma = \frac{N_o}{2}\, \overline{\kappa}\, \overline{\chi}_{n,\kappa}^{-1}(1-\alpha)$$

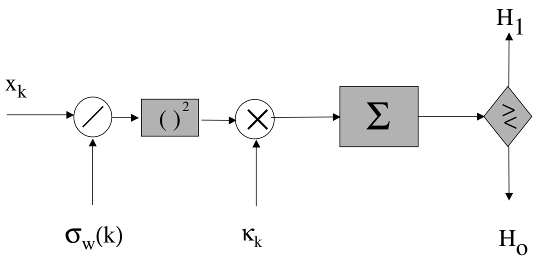


Figure 170: *LRT for detecting independent zero mean non-stationary Gaussian signal in non-stationary Gaussian noise.*

TWO USEFUL INTERPRETATIONS

Assume white noise for simplicity (we know that we can simply prewhiten by $1/\sigma_w(k)$ if non-white w_k).

1. "MEMORYLESS" ESTIMATOR CORRELATOR IMPLEMENTATION

Rewrite test statistic as

$$\sum_{k=1}^{n} \hat{s}_k \, x_k \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma$$

where $\hat{s}_k$ is LLMSE estimator of s_k given x_k

$$\hat{s}_k = \frac{\sigma_s^2(k)}{\sigma_s^2(k) + \sigma_w^2}\, x_k = \kappa_k\, x_k$$

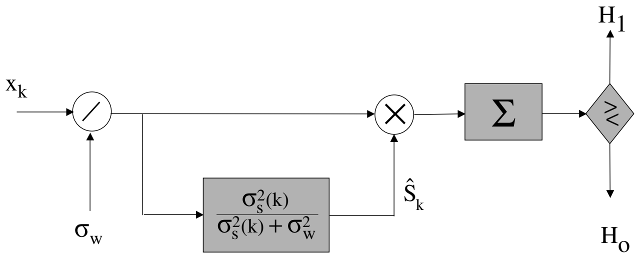


Figure 171: *Memoryless estimator correlator implementation of LRT for non-stationary uncorrelated signal in white noise. Note prewhitening operation* $1/\sigma_w^2$ *preceeds the estimator correlator.*

2. "MEMORYLESS" FILTER-SQUARER IMPLEMENTATION

Rewrite test statistic as

$$\sum_{k=1}^{n} y_k^2 \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

where y_k is defined as

$$y_k = \sqrt{\frac{\sigma_s^2(k)}{\sigma_s^2(k) + \sigma_w^2}}\, x_k = \sqrt{\kappa_i}\, x_k$$

POWER OF MEMORYLESS ESTIMATOR CORRELATOR:

as above

$$\beta = 1 - \overline{\chi}_{n,1/c-1}(\rho \overline{\chi}_{n,1-c}^{-1}(1-\alpha))$$

where

$$c_i = \frac{\sigma_w^2}{(\sigma_s^2(i) + \sigma_w^2)} = 1 - \kappa_i$$

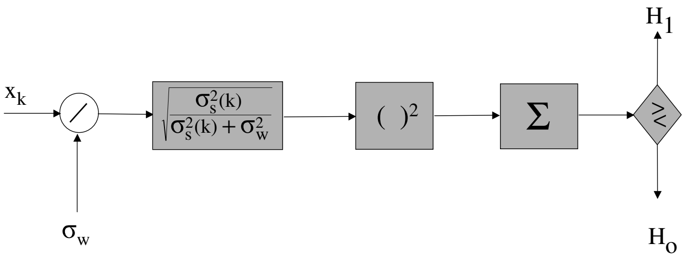


Figure 172: *Memoryless filter squarer implementation of LRT for non-stationary uncorrelated signal in white noise.*

$$\rho = \frac{\sum_i \kappa_i}{\sum_i \kappa_i/(1-\kappa_i)}$$

To a good approximation, at high SNR it can be shown that ROC depends on n, N_o, $\sigma_s^2(i)$ only through the following three SNR moments

$$\begin{aligned}
SNR^{(1)} &= \frac{1}{\sigma_w^2} \sum_{i=1}^n \sigma_s^2(i) \\
SNR^{(2)} &= \left(\frac{1}{\sigma_w^2}\right)^2 \sum_{i,j=1}^n \sigma_s^2(i)\sigma_s^2(j) \\
SNR^{(3)} &= \left(\frac{1}{\sigma_w^2}\right)^3 \sum_{i,j,k=1}^n \sigma_s^2(i)\sigma_s^2(j)\sigma_s^2(k)
\end{aligned}$$

Example 43 *Offline detection of w.s.s. signal in white noise*

Assume a window of n samples of a zero mean w.s.s. process x_k are available to test

$$\begin{aligned}
H_0 : x_k &= w_k \\
& \quad k = 0, \ldots, n-1 \\
H_1 : x_k &= s_k + w_k
\end{aligned}$$

where

* w_k: Gaussian white noise with PSD $\mathcal{P}_w(\omega) = N_o/2$

* s_k: zero mean w.s.s. Gaussian signal with known autocorrelation function $r_s(k) = E[s_l s_{l-k}]$ and PSD $\mathcal{P}_s(\omega) = \text{DFT}\{r_s(k)\}$

* w_k, s_k uncorrelated

The $n \times n$ covariance matrices under H_0 and H_1 are

$$R_0 = R_w = \sigma_w^2 I, \quad R_1 = R_s + \sigma_w^2 I$$

where $R_s = ((\, r_s(l-m)\,))_{l,m=1}^n$ is an $n \times n$ p.d. Toeplitz matrix and $\sigma^2 = N_o/2$.

We know that MP-LRT is of form

$$T(\underline{x}) = \underline{x}^T(R_0^{-1} - R_1^{-1})\underline{x} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

However, in this form, the detector is unimplementable for large n due to R_s matrix inversion. An alternative, for large n is via "spectral decomposition" of the Toeplitz R_s.

Step 1: diagonalize $R_0^{-1} - R_1^{-1}$

Let R_s have eigendecomposition

$$R_s = E D_s E^H$$

where $D_s = \text{diag}(\lambda_s(i))$ is a diagonal matrix of eigenvalues and $E = [\underline{e}_1, \ldots, \underline{e}]$ are the n associated eigenvectors of R_s. Now as eigenvectors are orthonormal, E is a unitary matrix: $E^H E = E E^H = I$ and $E^{-1} = E^H$. Thus, R_0 and R_1 have the eigendecompositions

$$R_0 = E D_0 E^H$$

and

$$R_1 = E D_1 E^H = E D_s E^H + E D_0 E^H = E(D_s + D_0)E^H$$

where $D_0 = \sigma_w^2 I$ and $D_1 = \text{diag}(\lambda_s(i) + \sigma_w^2)$. Furthermore,

$$R_0^{-1} - R_1^{-1} = E(D_0^{-1} - D_1^{-1})E^H = E\left[\frac{1}{\sigma_w^2} I - \text{diag}\left(\frac{1}{\lambda_s(i) + \sigma_w^2}\right)\right] E^H$$

so that the MP-LRT has the alternative representation (recall that $\underline{x}$ is real)

$$\underline{x}^T(R_0^{-1} - R_1^{-1})\underline{x} = [E^H \underline{x}]^H (D_0^{-1} - D_1^{-1})[E^H \underline{x}] \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta \tag{43}$$

Step 2: Use large n spectral approximation to D_s and E matrices.

Since s_k is wss R_s is Toeplitz and it can be shown that for large n eigendecomposition is [3]

$$R_s = E \ \text{diag}(\mathcal{P}_s(\omega_k))\ E^H$$

where

* $\omega_k = \frac{k}{n} 2\pi \in [0, 2\pi)$, $k = 0, \ldots, n-1$, is k-th radian frequency.

* E is $n \times n$ unitary "DFT matrix" with n columns $\underline{e}_k$:

$$\underline{e}_k = [1, e^{\omega_k}, \ldots, e^{\omega_k(n-1)}]^T / \sqrt{n}$$

Proof of spectral approximation:

It suffices to show that DFT matrix E diagonalizes R_s, i.e.

$$\underline{e}_k^H R_s \underline{e}_l = \begin{cases} \mathcal{P}_s(\omega_k), & k = l \\ 0, & o.w. \end{cases}$$

So let's write out the quadratic form explicitly

$$\underline{e}_k^H R_s \underline{e}_l \qquad = n^{-1} \sum_{p=0}^{n-1} \sum_{m=0}^{n-1} e^{-\omega_k p} e^{\omega_l m} r_s(p-m)$$

Next we rearrange the summation a bit

$$\underline{e}_k^H R_s \underline{e}_l \qquad = n^{-1} \sum_{m=0}^{n-1} e^{(\omega_l - \omega_k)m} \sum_{p=0}^{n-1} e^{-\omega_k(p-m)} r_s(p-m)$$

Next we make a change of indices in the summations $m \to t \in \{1, \ldots, n\}$ and $m - p \to \tau \in \{-n, \ldots, n\}$ to obtain

$$\underline{e}_k^H R_s \underline{e}_l \qquad = \sum_{\tau=-n+1}^{n-1} r_s(\tau) e^{-\omega_k(\tau)} \underbrace{n^{-1} \sum_{t=0}^{n-1} e^{(\omega_l - \omega_k)t}}_{g_n(\omega_l - \omega_k)}$$

where $g_n(u) = n^{-1}(e^{jun/2} \sin(u(n+1)/2)/\sin(u/2) - 1)$. Now, as $n \to \infty$, the term $g_n(\omega_l - \omega_k)$ converges to a discrete delta function:

$$\lim_{n \to \infty} g_n(\omega_l - \omega_k) = \delta_{k-l}$$

and so, assuming appropriate conditions allowing us to bring the limit under the summation, we have the large n approximation

$$\underline{e}_k^H R_s \underline{e}_l \qquad = \sum_{\tau=-n+1}^{n-1} r_s(\tau) e^{-\omega_k(\tau)} \; \delta_{k-l} = \mathcal{P}_s(\omega_k) \delta_{k-l}$$

which establishes the spectral approximation.

Step 3: substitute spectral approximation back into form of MP-LRT test statistic.

As we now know that $\lambda_s(i) = \mathcal{P}_s(\omega_i)$ we have from (43) the following form of the MP-LRT test statistic (recall that $\sigma_w^2 = N_o/2$)

$$T(\underline{x}) = [E^H \underline{x}]^H [D_0^{-1} - D_1^{-1}][E^H \underline{x}] \quad (44)$$

$$= \frac{2}{N_o} [E^H \underline{x}]^H \operatorname{diag}(\kappa(\omega_k)) \; [E^H \underline{x}] \quad (45)$$

where $\kappa(\omega)$ is the spectral coherency function

$$\kappa(\omega) = \frac{\mathcal{P}_s(\omega)}{\mathcal{P}_s(\omega) + N_o/2}$$

Next consider k-th row of $E^H \underline{x}$

$$[E^H \underline{x}]_k = \underline{e}_k^H \underline{x} = \sum_{l=0}^{n-1} x_l e^{-j\omega_{k-1} l} \,/\sqrt{n} = X(\omega_k)\; e^{-j\omega_{k-1}}/\sqrt{n}$$

where $\{X(\omega_k)\}_{k=0}^{n-1}$ is the DFT of $\underline{x}$.

Therefore, substituting this last expression into (45) we have the equivalent large n form for the MP-LRT

$$T(\underline{x}) = \frac{2}{N_o} \sum_{k=0}^{n-1} \frac{\mathcal{P}_s(\omega_k)}{\mathcal{P}_s(\omega_k) + N_o/2} \; |X(\omega_k)|^2 \;\; 1/n \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

where, as before, γ is the level α threshold

$$\gamma = \overline{\kappa}\, \overline{\chi}_{n,\kappa}(1-\alpha)$$

.

Note, as we take window length $n \to \infty$ and let $2\pi/n \to d\omega$, we see that the MP-LRT statistic converges to an integral over the continous ω domain

$$T(\underline{x}) \;\to\; \frac{2}{N_o} \frac{1}{2\pi} \int_0^{2\pi} \frac{\mathcal{P}_s(\omega)}{\mathcal{P}_s(\omega) + N_o/2} \; |X(\omega)|^2 \, d\omega$$

IMPLEMENTATION ISSUES

Using the duality between convolution in the time domain and multiplication in the frequency domain, identify the test statistic as:

$$T(\underline{x}) = \frac{2}{N_o}\, n^{-1} \sum_{k=0}^{n-1} \underbrace{\frac{\mathcal{P}_s(\omega_k)}{\mathcal{P}_s(\omega_k) + N_o/2} X^*(\omega_k)}_{(\hat{S}(\omega_k))^*} \; X(\omega_k)\; e^{-j\omega_k n}|_{n=0} = \frac{2}{N_o} \sum_{k=0}^{n-1} \hat{s}_{n+k} x_k|_{n=0},$$

where $\hat{s}_k$ is the inverse DFT of $\hat{S}(\omega_k)$.

Implementation 1: Estimator correlator:

Absorbing $N_o/2$ into the threshold, the MP-LRT can be written as

$$\sum_{k=0}^{n-1} \hat{s}_k x_k \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma = \frac{N_o}{2}\, \overline{\kappa}\, \overline{\chi}_{n,\kappa}(1-\alpha)$$

where

$$\hat{s}_k = h_{\mathrm{M}MSE}(k) * x_k$$

and $h_{\mathrm{M}MSE}(k)$ is the Wiener filter with transfer function

$$H_{\mathrm{M}MSE}(\omega) = \frac{\mathcal{P}_s(\omega)}{\mathcal{P}_s(\omega) + N_o/2}$$

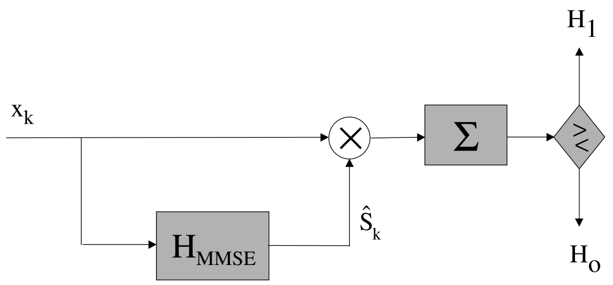


Figure 173: *Estimator correlator implementation of LRT for w.s.s. signal in white noise.*

Alternatively use

Parseval's theorem: if $f(k) \;\Leftrightarrow\; F(\omega_k)$ are DFT pair then

$$n^{-1} \sum_{k=-\infty}^{\infty} |F(\omega_k)|^2 = \sum_{k=0}^{n-1} |f(k)|^2$$

Implementation 2: filter-squarer

$$\sum_{k=0}^{n-1} y_k^2 \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma = \frac{N_o}{2}\, \overline{\kappa}\, \overline{\chi}_{n,\kappa}(1-\alpha)$$

where

$$y_k = h_k * x_k$$

and h_k has transfer function

$$H(\omega) = \sqrt{\frac{\mathcal{P}_s(\omega)}{\mathcal{P}_s(\omega) + N_o/2}}$$

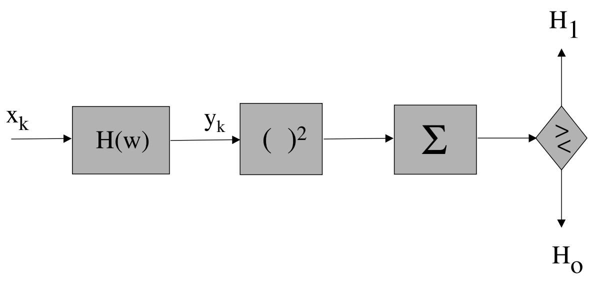


Figure 174: *Filter squarer implementation of LRT for w.s.s. signal in white noise.*

ROC: identically to previous example

$$\beta = 1 - \overline{\chi}_{n,1/c-1}(\rho \overline{\chi}^{-1}_{n,1-c}(1-\alpha))$$

except now $c = [c_1, \ldots, c_n]$ is

$$c_i = (N_o/2)/(\mathcal{P}_s(\omega_i) + N_o/2)$$

11.3 DETECTION OF NON-ZERO MEAN NON-STATIONARY SIGNAL IN WHITE NOISE

Now consider

$$\begin{aligned} H_0 : x_k &= w_k \\ H_1 : x_k &= s_k + w_k \end{aligned}$$

* $w_k \sim \mathcal{N}(0, \sigma_w^2)$: white noise
* $s_k \sim \mathcal{N}(\mu_k, \sigma_s^2(k))$: uncorrelated signal samples
* w_k, s_k uncorrelated

Recall general formula for nonequal means and covariances for LRT

$$T(\underline{x}) \;=\; \tfrac{1}{2}\underline{x}^T[R_0^{-1} - R_1^{-1}]\underline{x} + (\underline{\mu}_1^T R_1^{-1} - \underline{\mu}_0^T R_0^{-1})\underline{x} \;\; \overset{H_1}{\underset{H_0}{\gtrless}} \;\; \gamma$$

For present case $\underline{\mu}_0 = 0$, R_1 and R_0 are diagonal and LRT statistic reduces to

$$\begin{aligned} T(\underline{x}) &= \frac{1}{2\sigma_w^2} \sum_{k=1}^{n} \frac{\sigma_s^2(k)}{\sigma_s^2(k) + \sigma_w^2}\, x_k^2 + \sum_{k=1}^{n} \frac{1}{\sigma_s^2(k) + \sigma_w^2}\, \mu_k x_k \\ &= \frac{1}{2\sigma_w^2} \left(\sum_{k=1}^{n} \kappa_k x_k^2 + 2\sum_{k=1}^{n} (1-\kappa_k)\mu_k x_k \right) \end{aligned}$$

It is easily shown (see exercises) that this LRT is equivalent to the test

$$\sum_{k=1}^{n} \frac{\sigma_s^2(k)}{\sigma_s^2(k) + \sigma_w^2}\, (x_k - \mu_k)^2 + 2\sum_{k=1}^{n} \mu_k x_k \;\; \overset{H_1}{\underset{H_0}{\gtrless}} \;\; \gamma' \tag{46}$$

This test can be implemented by a combination of estimator-correlator and matched filter.

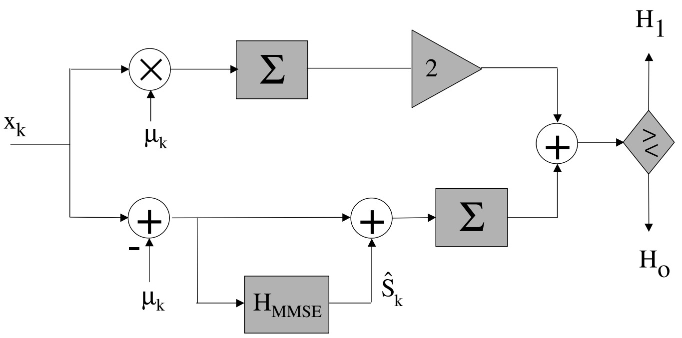


Figure 175: *Estimator correlator plus matched filter implementation of LRT for non-zero mean w.s.s. signal in white noise. (Note: the "summer" with inputs $x_k - \mu_k$ and $\hat{s}_k$ should be a multiplier)*

PERFORMANCE:

The test statistic is now distributed as a noncentral Chi-square-mixture under H_0 and H_1 and analysis is somewhat more complicated (Johnson, Kotz and Balakrishnan [13, Sec. 18.8]).

11.4 ONLINE IMPLEMENTATIONS OF OPTIMAL DETECTORS

Objective: perform optimal detection at each sampling time $n = 1, 2, \ldots$ based only on past observations $0 < k \leq n$

$$H_0 : x_k = w_k$$

$$H_1 : x_k = s_k + w_k$$

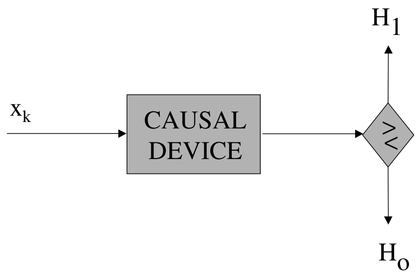


Figure 176: *Online detection seeks to develop optimal detector for each time instant* n *based only on past measurements.*

11.4.1 ONLINE DETECTION FOR NON-STATIONARY SIGNALS

Objective: decide between the presence of either of two random signals based on finite past $0 < k \leq n$

$$H_0 : x_k = s_0(k) + w_k$$

$$H_1 : x_k = s_1(k) + w_k$$

where

v_k: non-stationary zero mean Gaussian noise

$s_0(k),\ s_1(k)$: non-stationary zero mean Gaussian signals with known state space representations

Recall: general MP-LRT is of form

$$T(\underline{x}) = \underline{x}^T[R_0^{-1} - R_1^{-1}]\underline{x} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

Difficulty: growing memory in n makes computation of $T(x)$ impractical

Solution 1: Online dual Kalman signal selector

Solution 2: Online signal detector via Cholesky

11.4.2 ONLINE DUAL KALMAN SIGNAL SELECTOR

Let $\underline{\eta}_0$ and $\underline{\eta}_1$ denote vectors of innovations generated by Kalman filters matched to H_0 and H_1, respectively.

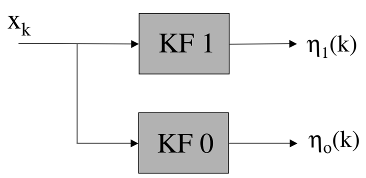


Figure 177: *Dual Kalman filters generate innovations processes η_1 and η_0*

We know that

$$\underline{\eta}_0 = \mathcal{A}_0\underline{x}, \qquad \underline{\eta}_1 = \mathcal{A}_1\underline{x}$$

$$R_0 \quad = \quad \mathcal{A}_0^{-1}R_{\eta_0}\mathcal{A}_0^{-T}, \qquad R_1 = \mathcal{A}_1^{-1}R_{\eta_1}\mathcal{A}_1^{-T}$$

where

* $\mathcal{A}_0, \mathcal{A}_1$ are lower triangular matrices of prediction coefficients

* R_{η_0}, R_{η_1} are diagonal matrices of prediction error variances

Recall important property of innovations:

$\eta(k) = x(k) - \hat{x}(k|k-1) = x(k) - \hat{s}(k|k-1)$

* $E[\eta(k)] = 0$

* $\text{var}(\eta(k)) = \underbrace{\underline{c}^T R_{\hat{\xi}}(k, k-1)\underline{c}}_{\sigma^2_{\hat{s}}(k)} + \sigma_v^2$ is minimum prediction error variance

* $\left\{ \frac{\eta(k)}{\sqrt{\text{var}(\eta(k))}} \right\}$ is white

* $[\eta_1, \ldots, \eta_n]^T \sim \mathcal{N}_n(0, \text{diag}(\text{var}(\eta(k)))I)$

Using innovations representation we can re-express LR statistic

$$
\begin{aligned}
T(\underline{x}) &= \underline{x}^T[R_0^{-1} - R_1^{-1}]\underline{x} \\
&= \underline{x}^T[\ \mathcal{A}_0^T R_{\eta_0}^{-1} \mathcal{A}_0 \ - \ \mathcal{A}_1^T R_{\eta_1}^{-1} \mathcal{A}_1\ ]\underline{x} \\
&= [\mathcal{A}_0\underline{x}]^T\ R_{\eta_0}^{-1}\ \underbrace{[\mathcal{A}_0\underline{x}]}_{\underline{\eta}_0} - [\mathcal{A}_1\underline{x}]^T\ R_{\eta_1}^{-1}\ \underbrace{[\mathcal{A}_1\underline{x}]}_{\underline{\eta}_1}
\end{aligned}
$$

Or, LRT reduces to

$$
T(\underline{x}) = \sum_{i=1}^n \frac{\eta_0^2(i)}{\text{var}(\eta_0(i))} \ - \ \sum_{i=1}^n \frac{\eta_1^2(i)}{\text{var}(\eta_1(i))} \ \overset{H_1}{\underset{H_0}{\gtrless}} \ \gamma
$$

where, level α threshold is time varying. For example if $R_0^{-1} > R_1^{-1}$

$$
\gamma = n\overline{(1-c)}\ \overline{\chi}^{-1}_{n,1-c}(1-\alpha)
$$

Special Case: SIGNAL DETECTION IN WHITE NOISE

Here $s_0(k)$ is zero and v_k is white

$$
\begin{aligned}
H_0 : x_k &= v_k \\
H_1 : x_k &= s_1(k) + v_k
\end{aligned}
$$

and

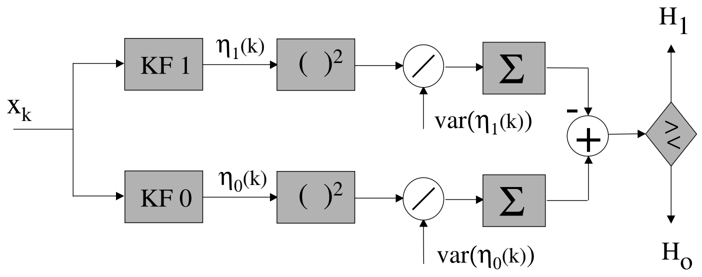


Figure 178: *Dual Kalman filter implementation of state space signal selector.*

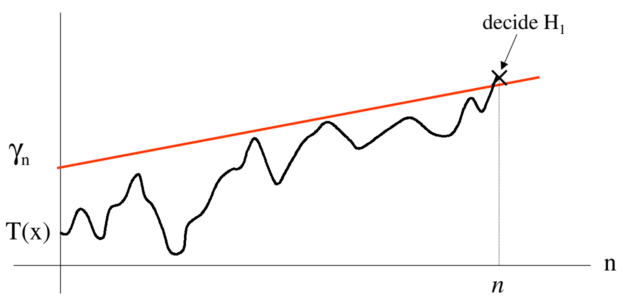


Figure 179: *Trajectory of dual Kalman filter implementation of state space signal selector. Note that the threshold is a function of time. If the number of samples n is random then the threshold of the test must be set by the method of repeated tests of significance.*

* $\hat{s}_0(k|k-1) = 0$

* $\eta_0(k) = x_k$

* $\text{var}_0(\eta_0(k)) = \sigma_v^2$

Thus MP-LRT simplifies to a "measured energy" vs. "Kalman residual" detector

$$T(\underline{x}) = \frac{1}{\sigma_v^2}\sum_{i=1}^n x_i^2 - \sum_{i=1}^n \frac{\eta_1^2(i)}{\text{var}(\eta_1(i))} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

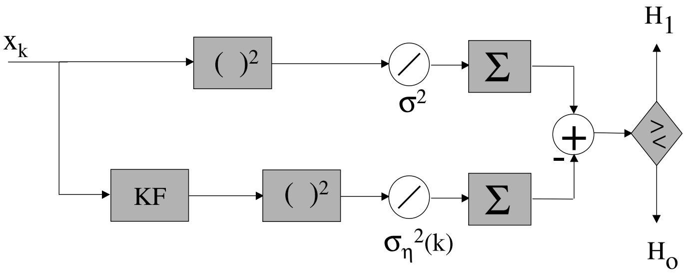


Figure 180: *The optimal detector of a single state space signal in noise.*

11.4.3 ONLINE SIGNAL DETECTOR VIA CHOLESKY

Again assume $s_0(k)$ is zero so that

$$\begin{aligned} H_0 : x_k &= v_k \\ H_1 : x_k &= s_1(k) + v_k \end{aligned}$$

Solution: apply Cholesky decomposition to $R_0^{-1} - R_1^{-1}$

Note

$$R_0^{-1} - R_1^{-1} = R_v^{-1} - [R_s + R_v]^{-1}$$

$$\begin{aligned} &= [R_s + R_v]^{-\frac{1}{2}} R_s^{\frac{1}{2}} R_v^{-1} R_s^{\frac{1}{2}} [R_s + R_v]^{-\frac{1}{2}} \\ &> 0 \end{aligned}$$

Hence we can apply the Cholesky decomposition

$$R_0^{-1} - R_1^{-1} = L^T \; P \; L$$

* L is lower triangular matrix of "backward predictor coefficients"

* P is diagonal matrix of "backward predictor error variances"

Now apply Cholesky decomposition to $T(\underline{x})$

$$\begin{aligned} T(\underline{x}) &= \tfrac{1}{2} \underline{x}^T [R_0^{-1} - R_1^{-1}] \underline{x} \\ &= \tfrac{1}{2} \underline{x}^T [L^T \; P \; L] \underline{x} \\ &= \tfrac{1}{2} [L\underline{x}]^T \; P \; \underbrace{[L\underline{x}]}_{\underline{y}} \end{aligned}$$

or we have representation

$$T(\underline{x}) = \tfrac{1}{2} \sum_{i=1}^{n} \sigma_i^2 \; y_i^2$$

The MP level α test is simply

$$\sum_{i=1}^{n} \sigma_i^2 \; y_i^2 \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma_n = n\overline{(1-c)} \; \overline{\chi}_{n,\overline{\kappa}}^{-1}(1-\alpha)$$

where σ_i^2 and $\kappa = n\overline{(1-c)}$ are computed offline. Note that while we can generate a recursion for the test statistic, a recursion for the threshold γ_n is not available.

In many cases y_i can be generated by a "lumped" Kalman filter matched to a state observation model

$$\begin{aligned} x_k' &= s_k' + v_k' \\ s_k' &= \underline{c}_k^T \underline{\nu}_k \\ \underline{\nu}_{k+1} &= D_k \underline{\nu}_k + E_k \underline{w}_k' \end{aligned}$$

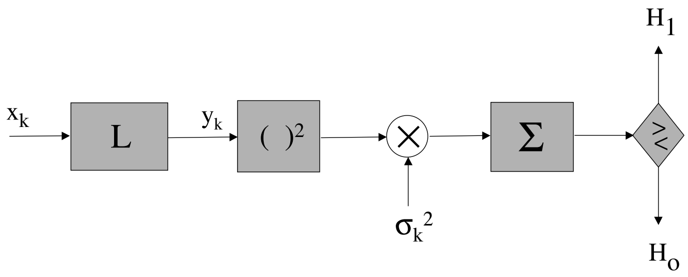


Figure 181: *On-line implementation of non-stationary signal detector via using Cholesky factor L.*

synthesized such that the measurement covariance satsifies

$$R_{x'} = R_0^{-1} - R_1^{-1}$$

11.5 STEADY-STATE STATE-SPACE SIGNAL DETECTOR

Assume:

* State model for s_1 is LTI

* measurement noise v_k is w.s.s.

* Kalman filter is in steady state (n large)

Then, as innovations are w.s.s., the MP-LRT statistic can be written

$$T(\underline{x}) = \frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 \; - \; \frac{1}{2\text{var}(\eta_1)}\sum_{i=1}^n \eta_1^2(i) \quad \mathop{\gtrless}_{H_0}^{H_1} \quad \gamma$$

Or, using asymptotic innovations variance, we have MP test

$$\sum_{i=1}^n x_i^2 \; - \; \frac{\sigma_v^2}{\sigma_{\tilde{s}}^2 + \sigma_v^2}\sum_{i=1}^n \eta_1^2(i) \quad \mathop{\gtrless}_{H_0}^{H_1} \quad \gamma$$

APPROXIMATE GLRT FOR UNKNOWN MEASUREMENT NOISE VARIANCE

We can implement an approximate GLRT to handle the case where the variance of the observation noise σ_v^2 is unknown. For this case the GLR statistic is

$$\Lambda_{GLR}(\underline{x}) = \frac{\max_{\sigma_v^2>0} \, (\sigma_{\tilde{s}}^2 + \sigma_v^2)^{-n/2} \exp\left(-\frac{1}{2(\sigma_{\tilde{s}}^2+\sigma_v^2)} \, \sum_{i=1}^n \eta_1^2(i)\right)}{\max_{\sigma_v^2>0} \, (\sigma_v^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_v^2} \, \sum_{i=1}^n x_i^2\right)}$$

Of course the denominator is maximized for

$$\sigma_v^2 = n^{-1} \sum_{i=1}^n x_i^2$$

and, as for the numerator, we proceed by an iterative approximation. First neglect the dependence of of η_1 on σ_v^2. Then the numerator is maximized for

$$\hat{\sigma_v}^2(n) = n^{-1} \sum_{i=1}^n \eta_1^2(i) - \sigma_s^2$$

Now generate $\eta_1(n+1)$ from the Kalman Filter having parameters A, b, c and $\hat{\sigma_v}^2$. In this way we obtain an approximate GLRT which is implemented by comparing the ratio of two variance estimators to a threshold. Note that the numerator and denominator of the test statistic are dependent so this is not an F-test.

$$\frac{\hat{\sigma}_v^2}{\hat{\sigma}_{\eta_1}^2} = \frac{\sum_{i=1}^n x^2(i)}{\sum_{i=1}^n \eta_1^2(i)} \quad \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \quad \gamma$$

In analogous manner, for the GLRT signal selector we obtain

$$\frac{\sum_{i=1}^n \eta_0^2(i)}{\sum_{i=1}^n \eta_1^2(i)} \quad \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \quad \gamma$$

11.6 Exercises

1. Let $\underline{x} = [x_1, \ldots, x_n]^T$ be n samples of a waveform. It is of interest to test the two hypotheses

$$\begin{aligned} H_0 : \underline{x} &= a\underline{y} + \underline{w} \\ H_1 : \underline{x} &= \underline{s} + a\underline{y} + \underline{w} \end{aligned}$$

   where $\underline{w}$ is zero mean Gaussian white noise, $\mathrm{cov}(\underline{w}) = \sigma^2 I$, $\underline{s}$ and $\underline{y}$ are known waveforms, and the scalar constant a is unknown.

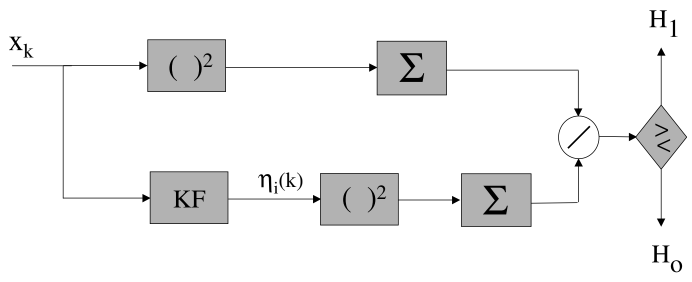


Figure 182: *Approximate steady-state GLRT signal detector for unknown measurement noise*

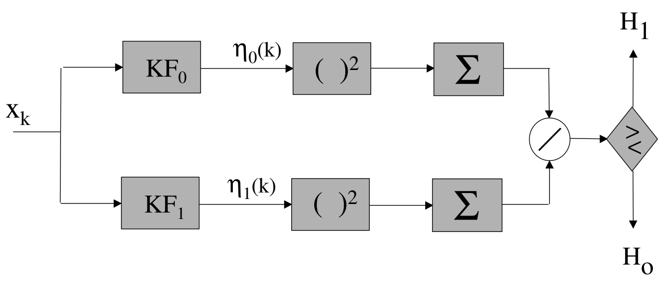


Figure 183: *GLRT signal selector for unknown measurement noise*

(a) Assuming that a is a Gaussian r.v. with zero mean and variance σ_a^2 derive the MP LRT (with threshold) to test H_0 vs. H_1. Assume that a is independent of $\underline{w}$. Is this a UMP test for the case that the signal shape $\underline{s}/\|\underline{s}\|$ is known but its energy $\|\underline{s}\|^2$ is unknown? How about when signal shape $\underline{y}/\|\underline{y}\|$ is known but $\|\underline{y}\|^2$ is unknown?

(b) Under the assumption on a of part (a) find the detectibility index d which controls the ROC curve. Assume that $\|\underline{s}\| \leq 1$. Show that the ROC curve is optimized (maximum d) when the signal $\underline{s}$ is orthogonal to the interferer $\underline{y}$ but is otherwise arbitrary (Hint: you might want to use the matrix identity $[I + UU^T]^{-1} = I - U[I + U^TU]^{-1}U^T$).

(c) Assuming that a is a deterministic unknown constant, repeat parts (a) and (b) for the GLRT of H_0 vs. H_1.

2. Let x_k, $k = 1, \ldots, n$ be a segment of a discrete time random process. It is desired to test whether x_k contains a harmonic component (sinusoidal signal) or not

$$\begin{aligned} H_0 : x_k &= w_k \\ H_1 : x_k &= A\cos(\omega_o k + \psi) + w_k \end{aligned}$$

where w_k is zero mean Gaussian white noise with acf $r_w(k) = N_0/2\delta_k$, $\omega_o = 2\pi l/n$ for some integer l, A is a deterministic amplitude, and ψ is a uniform phase over $[0, 2\pi]$. The random phase of the sinusoid and the noise samples are independent of each other.

(a) Show that under H_1 the auto-correlation function of x_k is $E[x_i x_{i-k}] = r_x(k) = A^2/2\cos(\omega_o k) + N_0/2\delta_k$ and derive the PSD $\mathcal{P}_x$.

(b) Derive the MP LRT with threshold and implement the MP LRT as an estimator correlator and a filter squarer. (Hint: as ψ is uniform and $f_1(\underline{x}|\psi)$ is a Gaussian p.d.f. $f_1(\underline{x}) = (2\pi)^{-1}\int_0^{2\pi} f_1(\underline{x}|\psi)d\psi$ is a Bessel function which is monotone in a test statistic which under H_0 is distributed as a Chi-square with 2 df, i.e exponential.)

(c) Show that the MP LRT can also be implemented as a test on the magnitude of the *periodogram spectral estimator* $\mathcal{P}_{per}(\omega_o) = \frac{1}{n}|DFT\{x_k\}_{\omega=\omega_o}|^2$ where $DFT\{x_k\}_\omega = \sum_{k=1}^n x_k e^{-j\omega}$ is the DTFT of $\{x_k\}_{k=1}^n$.

3. Find the GLRT for the previous problem under the assumption that both A and ω_o are unknown (Hint: as no closed form solution exists for the MLE's of A and ω_o you can leave your answer in the form of a "peak detector" block diagram).

4. Derive the "completion of the square" result in section 11.3 of these notes.

5. A sensor is placed on a North Atlantic oil derick at a particular spatial location to monitor the mechanical state of the structure. When the mechanical state is "normal" the sensor produces a measurement which follows the state space model:

$$\begin{aligned} x_k &= s_k + v_k \\ s_{k+1} &= a s_k + w_k \end{aligned}$$

$k = 0, 1, \ldots$. A model for impending failure of the mechanical structure is that a shift in the damping constant a occurs. Assuming the standard Gaussian assumptions on the dynamical model under both normal and failure modes, the detection of impending failure can be formulated as testing between

$$\begin{aligned} H_0 : a &= a_o \\ H_1 : a &\neq a_o \end{aligned}$$

where $a_o \in (-1, 1)$ is known.

(a) Implement the MP test of level α for the simple alternative $H_1 : a = a_1$ where $a_1 \neq a_0$ with a pair of Kalman filters. Give explicit forms for your filters using the results of problem 5 of Hwk 4.

(b) Now treat the general composite case above with your favorite method, e.g. LMP or GLRT. Take this problem as far as you can by making simplifying assumptions starting with assuming steady state operation of the Kalman filters.

6. Available for observation are n time samples $X(k)$,

$$X(k) = \sum_{i=1}^{p} \alpha_i g_i(k - \tau_i) + W(k), \quad k = 1, \dots, n$$

where $W(k)$ is a zero mean Gaussian white noise with variance var$(W(k)) = \sigma_w^2$, α_i, $i = 1, \dots, p$, are p i.i.d. zero mean Gaussian random variables with variance σ_a^2, and $g_i(u)$, $i = 1, \dots, p$, are p known time functions over $u \in (-\infty, \infty)$. The α_i and $W(k)$ are uncorrelated and p is known. Define K as the $p \times p$ matrix of inner products of the g_i's, i.e. K has entries $\kappa_{ij} = \sum_{k=1}^{n} g_i(k - \tau_i) g_j(k - \tau_j)$.

(a) Show that the ML estimator of the τ_i's involves maximizing a quadratic form $\underline{y}^T[I + \rho K]^{-1}\underline{y} - b$ where $\underline{y} = [y_1, \dots, y_p]^T$ is a vector of p correlator outputs $y_i(\tau_i) = \sum_{k=1}^{n} x(k) g_i(k - \tau_i)$, $i = 1, \dots, p$, $b = b(\underline{\tau})$ is an observation independent bias term, and $\rho = \sigma_a^2/\sigma_w^2$ is the SNR (Hint: express log-likelihood function in vector-matrix form and use a matrix inverse (Woodbury) identity). Draw a block diagram of your ML estimator implemented as a peak picker, i.e. a variable filter applied to the data over which you seek to maximize the output.

(b) Now consider the detection problem

$$\begin{aligned} H_0 &: X(k) = W(k) \\ H_1 &: X(k) = \sum_{i=1}^{p} \alpha_i g_i(k - \tau_i) + W(k) \end{aligned}$$

For known τ_i's derive the LRT and draw a block diagrom of the detector. Is the LRT UMP for unknown τ_i's? How about for known τ_i's but unknown SNR σ_a^2/σ_w^2?

(c) Now assume that the τ_i's are unknown and that the α_i's are also unknown and non-random. Show that in the GLRT the maximization over the α_i's can be performed explicitly. Draw a block diagram of the GLRT implemented with a thresholded peak picker over the τ_i's.

End of chapter

12 MULTI-CHANNEL GAUSSIAN GLRT

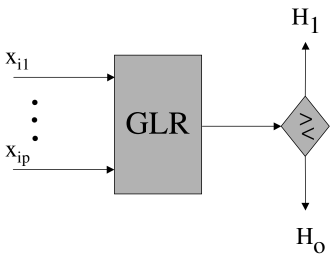


Figure 184: *GLR detector from multi-channel Gaussian measurements.*

OUTLINE

* Double sided GLRT for equality of vector mean

* Double sided GLRT for equality two vector means

* Double sided GLRT for independence of samples

* Double sided GLRT for whiteness of samples

* Confidence regions for vector mean

REFERENCES

Morrison [20]

Measurements are i.i.d. p-dimensional Gaussian vectors:

$$\underline{x}_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix}, \qquad i = 1, \ldots n$$

or equivalently, we have a $p \times n$ *measurement matrix*

$$X = [\underline{x}_1, \ldots, \underline{x}_n]$$

* $\{\underline{x}_i\}_{i=1}^n$: independent Gaussian columns ($n \geq p$)

* $\underline{\mu} = E_\theta[\underline{x}_i]$: mean vector

* $R = \text{cov}_\theta(\underline{x}_i)$: covariance matrix ($p \times p$)

Density of Gaussian matrix X with i.i.d. columns

$$
\begin{aligned}
& f(X;\ \underline{\mu}, R) \\
&= \left(\frac{1}{(2\pi)^{np}\,|R|}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\underline{x}_i - \underline{\mu})^T R^{-1}(\underline{x}_i - \underline{\mu})\right) \\
&= \left(\frac{1}{(2\pi)^{np}\,|R|}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\text{trace}\left\{(\underline{x}_i - \underline{\mu})(\underline{x}_i - \underline{\mu})^T R^{-1}\right\}\right) \\
&= \left(\frac{1}{(2\pi)^{np}\,|R|}\right)^{n/2} \exp\left(-\frac{n}{2}\,\text{trace}\{\hat{R}_\mu\, R\}\right)
\end{aligned}
$$

where we have defined the $p \times p$ covariance estimator

$$
\begin{aligned}
\hat{R}_{\underline{\mu}} &= n^{-1}\sum_{i=1}^{n}(\underline{x}_i - \underline{\mu})(\underline{x}_i - \underline{\mu})^T \\
&= \frac{1}{n}(X - \underline{\mu}\underline{1}^T)(X - \underline{\mu}\underline{1}^T)^T
\end{aligned}
$$

12.1 DOUBLE SIDED TEST OF VECTOR MEAN

$$
\begin{aligned}
H_0 : \underline{\mu} &= \underline{\mu}_o, \quad R > 0 \\
H_1 : \underline{\mu} &\neq \underline{\mu}_o, \quad R > 0
\end{aligned}
$$

$$
\Lambda_{GLR} = \frac{\max_{\underline{\mu}, R>0} f(X; \underline{\mu}, R)}{\max_{R>0} f(X; \underline{\mu}_o, R)}
$$

Now, it is easily seen that

$$
\max_{\underline{\mu}, R>0} f(X; \underline{\mu}, R) = \max_{R>0} f(X;\ \overline{\underline{x}_i}, R)
$$

where the column sample mean is

$$\overline{\underline{x}_i} = n^{-1}\sum_{i=1}^n \underline{x}_i = X\underline{1}\,\frac{1}{n}$$

Therefore, we can rewrite GLRT as

$$\begin{aligned}
\Lambda_{GLR} &= \frac{\max_{\underline{\mu},R>0} f(X;\underline{\mu},R)}{\max_{R>0} f(X;\underline{\mu}_o,R)} \\
&= \frac{\max_{R>0}\ |R|^{-n/2}\ \exp\left(-\frac{1}{2}\ \text{trace}\left\{\hat{R}_{\overline{\underline{x}_i}} R^{-1}\right\}\right)}{\max_{R>0}\ |R|^{-n/2}\ \exp\left(-\frac{n}{2}\ \text{trace}\left\{\hat{R}_{\underline{\mu}} R^{-1}\right\}\right)}
\end{aligned}$$

FACT: for any vector $\underline{t} = [t_1,\ldots,p]^T$

$$\max_{R>0}\left\{|R|^{-n/2}\ \exp\left(-\frac{n}{2}\ \text{trace}\left\{\hat{R}_{\underline{t}} R^{-1}\right\}\right)\right\} = |\hat{R}_{\underline{t}}|^{-n/2}\ e^{-n^2/2}$$

and the maximum is attained by

$$R = \hat{R}_{\underline{t}} = n^{-1}\sum_{i=1}^n (\underline{x}_i - \underline{t})(\underline{x}_i - \underline{t})^T$$

Proof:

The maximizing R also maximizes

$$l(R) = \ln f(X;\underline{t},R) = \frac{n}{2}\ \ln|R|\ -\frac{n}{2}\ \text{trace}\left\{\hat{R}_{\underline{t}} R^{-1}\right\}$$

Define the transformed covariance $\tilde{R}$

$$\tilde{R} = \hat{R}_{\underline{t}}^{-1/2}\ R\ \hat{R}_{\underline{t}}^{-1/2}$$

Then, since

$$\text{trace}\{AB\} = \text{trace}\{BA\},\quad |AB| = |BA| = |B|\ |A|$$

$$\begin{aligned}
l(R) &= -\frac{n}{2}\ \left(\ln|\hat{R}_{\underline{t}}^{1/2}\tilde{R}\hat{R}_{\underline{t}}^{1/2}|\ +\ \text{trace}\left\{\hat{R}_{\underline{t}}^{1/2}\tilde{R}^{-1}\hat{R}_{\underline{t}}^{-1/2}\right\}\right) \\
&= -\frac{n}{2}\ \left(\ln|\hat{R}_{\underline{t}}\tilde{R}|\ +\ \text{trace}\left\{\tilde{R}^{-1}\right\}\right)
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{n}{2}\left(\ln|\hat{R}_{\underline{t}}| + \ln|\tilde{R}| \ + \ \text{trace}\left\{\tilde{R}^{-1}\right\}\right) \\
&= -\frac{n}{2}\left(\ln|\hat{R}_{\underline{t}}| + \sum_{j=1}^{p}\ln\tilde{\lambda}_j + \sum_{j=1}^{p}\frac{1}{\tilde{\lambda}_j}\right)
\end{aligned}
$$

Where $\{\tilde{\lambda}_j\}$ are the eigenvalues of $\tilde{R}$

Hence the maximizing R satisfies for $j = 1, \ldots, p$

$$
\begin{aligned}
0 &= \frac{d}{d\tilde{\lambda}_j} l(R) \\
&= -\frac{n}{2}\left(\frac{1}{\tilde{\lambda}_j} - \frac{1}{\tilde{\lambda}_j^2}\right)
\end{aligned}
$$

so that the maximizing $\tilde{R}$ has identical eigenvalues

$$
\tilde{\lambda}_j = 1, \quad j = 1, \ldots, p
$$

Implication: the maximizing $\tilde{R}$ is an orthogonal (unitary) matrix U.

But since $\tilde{R}$ is also symmetric the only possible solution is the identity I![o]

Therefore

$$
I = \tilde{R} = \hat{R}_{\underline{t}}^{-1/2} R \hat{R}_{\underline{t}}^{-1/2}
$$

giving the maximizing R as

$$
R = \hat{R}_{\underline{t}}
$$

$\diamond$

Note: We have just shown that

1. MLE of R for known $\underline{\mu} = \underline{\mu}_o$ is

$$
\hat{R}_{\underline{\mu}} = n^{-1}\sum_{i=1}^{n}(\underline{x}_i - \underline{\mu}_o)(\underline{x}_i - \underline{\mu}_o)^T
$$

2. MLE of R for unknown $\underline{\mu}$ is

[0] If U is orthogonal then $U^H = U^{-1}$. If in addition U is symmetric then $U = U^T = U^{-1}$ so that $U = I$.

$$\hat{R}_{\overline{\underline{x}_i}} = \hat{R} = n^{-1}\sum_{i=1}^{n}(\underline{x}_i - \overline{\underline{x}_i})(\underline{x}_i - \overline{\underline{x}_i})^T$$

Plugging these ML solutions back into GLRT statistic

$$\begin{aligned} \Lambda_{GLR} &= \left(\frac{|\hat{R}_{\underline{\mu}_o}|}{|\hat{R}|}\right)^{n/2} \\ &= \left(\left|\hat{R}_{\underline{\mu}_o}\,\hat{R}^{-1}\right|\right)^{n/2} \end{aligned}$$

Now using

$$\hat{R}_{\underline{\mu}_o} = \hat{R} + (\overline{\underline{x}_i} - \underline{\mu}_o)(\overline{\underline{x}_i} - \underline{\mu}_o)^T$$

we have the equivalent GLRT

$$\begin{aligned} T(X) &= \left|I + (\overline{\underline{x}_i} - \underline{\mu}_o)(\overline{\underline{x}_i} - \underline{\mu}_o)^T\hat{R}^{-1}\right| \\ &= \left|I + \underbrace{\hat{R}^{-\frac{1}{2}}(\overline{\underline{x}_i} - \underline{\mu}_o)}_{\underline{u}}\underbrace{(\overline{\underline{x}_i} - \underline{\mu}_o)^T\hat{R}^{-\frac{1}{2}}}_{\underline{u}^T}\right| \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma \end{aligned}$$

SIMPLIFICATION OF GLRT

Observe: $T(x)$ is the determinant of the sum of a rank 1 matrix and the identity matrix

$$\begin{aligned} T(X) &= \left|I + \underbrace{\underline{u}\,\underline{u}^T}_{\text{rank = 1}}\right| \\ &= \prod_{j=1}^{p}\lambda_j \end{aligned}$$

where λ_j are the eigenvalues of the matrix.

IMPORTANT FACTS:

1. Eigenvectors of $I + A$ are identical to eigenvectors of A

2. Eigenvectors of $A = \underline{u}\,\underline{u}^T$ are

$$\begin{aligned}
\underline{\nu}_1 &= \underline{u}\frac{1}{\|\underline{u}\|} = \hat{R}^{-1/2}(\overline{\underline{x}_i} - \underline{\mu}_o)\,\frac{1}{\sqrt{(\overline{\underline{x}_i} - \underline{\mu}_o)^T \hat{R}^{-1}(\overline{\underline{x}_i} - \underline{\mu}_o)}} \\
\underline{\nu}_2, \ldots, \underline{\nu}_p &= \text{determined via Gramm-Schmidt}
\end{aligned}$$

3. Eigenvalues of $I + A$ are

$$\begin{aligned}
\lambda_1 &= \underline{\nu}_1^T (I + A)\underline{\nu}_1 = 1 + (\overline{\underline{x}_i} - \underline{\mu}_o)^T \hat{R}^{-1}(\overline{\underline{x}_i} - \underline{\mu}_o) \\
\lambda_2 &= \ldots = \lambda_p = 1
\end{aligned}$$

Thus we have GLRT

$$T(x) = \prod_{j=1}^{p} \lambda_j = 1 + (\overline{\underline{x}_i} - \underline{\mu}_o)^T \hat{R}^{-1}(\overline{\underline{x}_i} - \underline{\mu}_o) \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \gamma$$

Or, equivalently, the GLRT has form of *Hotelling's* T^2 *test*

$$T^2 := n(\overline{\underline{x}_i} - \underline{\mu}_o)^T \mathsf{S}^{-1}(\overline{\underline{x}_i} - \underline{\mu}_o) \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \gamma$$

where S is the (unbiased) sample covariance

$$\mathsf{S} = \frac{1}{n-1} \sum_{k=1}^{n} (\underline{x}_k - \overline{x_i})(\underline{x}_k - \overline{x_i})^T$$

FACT: Under H_0, Hotelling's T^2 is distributed as a $\mathcal{T}^2$ distributed r.v. with $(p, n-p)$ d.f. [20]

Level α test is therefore

$$T^2 := n(\overline{\underline{x}_i} - \underline{\mu}_o)^T \mathsf{S}^{-1}(\overline{\underline{x}_i} - \underline{\mu}_o) \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \mathcal{T}^{-2}_{p,n-p}(1-\alpha)$$

REMARKS

1. The Hotelling T^2 test is CFAR since under H_0 its distribution is independent of R

2. The T^2 statistic is equal to a F-statistic within a scale factor

$$T^2 = \frac{p(n-1)}{n-p} \, F_{p,n-p}$$

3. An equivalent test is therefore

$$T^2 := n(\overline{\underline{x}_i} - \underline{\mu}_o)^T \mathsf{S}^{-1}(\overline{\underline{x}_i} - \underline{\mu}_o) \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \frac{p(n-1)}{n-p} \, F^{-1}_{p,n-p}(1-\alpha)$$

12.2 TEST OF EQUALITY OF TWO MEAN VECTORS

Two i.i.d. independent vector samples $n_1 + n_2 = n$

$\underline{x} = [\underline{x}_1, \ldots, \underline{x}_{n_1}], \ \ \underline{x}_i \ \sim \ \mathcal{N}_p(\underline{\mu}_x, R)$

$\underline{y} = [\underline{y}_1, \ldots, \underline{y}_{n_2}], \ \ \underline{y}_i \ \sim \ \mathcal{N}_p(\underline{\mu}_y, R)$

To test

$$\begin{aligned} H_0 : \underline{\mu}_x - \underline{\mu}_y &= \underline{\Delta}, \ \ R > 0 \\ H_1 : \underline{\mu}_x - \underline{\mu}_y &\neq \underline{\Delta}, \ \ R > 0 \end{aligned}$$

Combining elements of the derivation of scalar two population case and the previous vector mean example:

Find GLRT

$$\sqrt{\frac{n_1 n_2}{n}}(\overline{\underline{y}_i} - \overline{\underline{x}_i} - \underline{\Delta})^T \mathsf{S}_2^{-1}(\overline{\underline{y}_i} - \overline{\underline{x}_i} - \underline{\Delta}) \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \mathcal{T}^{-2}_{p,n-p-1}(1-\alpha) \qquad (47)$$

where we have defined the pooled sample covariance

$$\mathsf{S}_2 = \frac{1}{n-2} \left(\sum_{i=1}^{n_1} (\underline{x}_i - \underline{\hat{\mu}})(\underline{x}_i - \underline{\hat{\mu}})^T + \sum_{i=1}^{n_2} (\underline{y}_i - \underline{\hat{\mu}})(\underline{y}_i - \underline{\hat{\mu}})^T \right)$$

In analogy to the paired t-test of Sec. 9.5, the test (47) is called the *multivariate paired t-test*.

12.3 TEST OF INDEPENDENCE

n i.i.d. vector samples

$\underline{x} = [\underline{x}_1, \ldots, \underline{x}_n], \underline{x}_i \ \sim \ \mathcal{N}_p(\underline{\mu}, R)$

To test

$$
\begin{aligned}
H_0 : R &= \operatorname{diag}(\sigma_j^2) \\
H_1 : R &\neq \operatorname{diag}(\sigma_j^2)
\end{aligned}
$$

with mean vector $\underline{\mu}$ unknown

$$
\Lambda_{GLR} = \frac{\max_{R \neq \text{diag}, \underline{\mu}} f(X; \underline{\mu}, R)}{\max_{R = \text{diag}, \underline{\mu}} f(X; \underline{\mu}, R)} =
$$

$$
\frac{\max_{R>0} \ |R|^{-n/2} \ \exp\left(-\frac{1}{2} \ \sum_{k=1}^n (\underline{x}_k - \overline{x_i})^T R^{-1} (\underline{x}_k - \overline{x_i})\right)}{\max_{\sigma_j^2 > 0} \ \left(\prod_{k=1}^p \sigma_k^2\right)^{-n/2} \ \exp\left(-\frac{1}{2} \ \sum_{k=1}^n \frac{1}{\sigma_k^2} \ \|\underline{x}_k - \overline{x_i})\|^2\right)}
$$

Using previous results

$$
\Lambda_{GLR} = \left(\frac{\prod_{j=1}^p \hat{\sigma}_j^2}{|\hat{R}|}\right)^{n/2} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma
$$

where we have the variance estimate for each channel (row) of X

$$
\hat{\sigma}_j^2 := \frac{1}{n} \sum_{k=1}^n (\underline{x}_k - \overline{x_i})_j^2
$$

For n large we can set γ by using Chi-square asymptotics

$$
2 \ln \Lambda_{GLR} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma' = \chi^{-1}_{p(p-1)/2}(1 - \alpha)
$$

Calculation of degrees of freedom ν:

ν = # parameters fixed under H_0

1. $p^2 - p = p(p-1)$ off diagonals in R

2. $1/2$ of these off diagonals are identical

$\Rightarrow \nu = p(p-1)/2$

12.4 TEST OF WHITENESS

n i.i.d. vector samples

$\underline{x} = [\underline{x}_1, \ldots, \underline{x}_n]$, $\underline{x}_i \sim \mathcal{N}_p(\underline{\mu}, R)$

To test

$$
\begin{aligned}
H_0 : R &= \sigma^2 I \\
H_1 : R &\neq \sigma^2 I
\end{aligned}
$$

with mean vector $\underline{\mu}$ unknown

$$
\begin{aligned}
\Lambda_{GLR} &= \frac{\max_{R \neq \sigma^2 I, \underline{\mu}} f(X; \underline{\mu}, R)}{\max_{R = \sigma^2 I, \underline{\mu}} f(X; \underline{\mu}, R)} \\
&= \frac{\max_{R>0} \ |R|^{-n/2} \ \exp\left(-\frac{1}{2} \ \sum_{k=1}^n (\underline{x}_k - \overline{x_i})^T R^{-1} (\underline{x}_k - \overline{x_i})\right)}{\max_{\sigma^2>0} \ (\sigma^{2p})^{-n/2} \ \exp\left(-\frac{1}{2\sigma^2} \ \sum_{k=1}^n \|\underline{x}_k - \overline{x_i}\|^2\right)}
\end{aligned}
$$

Or we have (similarly to before)

$$
\Lambda_{GLR} = \left(\frac{\hat{\sigma}^{2p}}{|\hat{R}|}\right)^{n/2} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma
$$

where

$$
\begin{aligned}
\hat{\sigma}^2 &:= \frac{1}{np} \underbrace{\sum_{k=1}^n \|\underline{x}_k - \overline{x_i}\|^2}_{n\text{trace}\{\hat{R}\}} \\
&= \frac{1}{p} \text{trace}\left\{\hat{R}\right\}
\end{aligned}
$$

and we have defined the covariance estimate

$$
\hat{R} := \frac{1}{n} (X - \overline{x_i}\underline{1}^T)(X - \overline{x_i}\underline{1}^T)^T
$$

PERFORMANCE

For n large set γ using Chi-square asymptotics

$$2\ln\Lambda_{GLR} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma' = \chi^{-1}_{p(p+1)/2-1}(1-\alpha)$$

Calculation of degrees of freedom ν:

ν = # parameters fixed under H_0

1. $p(p+1)/2$ elements on diagonal and upper triangle of R

2. All p diagonal elements are identical under H_0

$\Rightarrow \nu = p(p+1)/2 - 1$

Equivalent form of test

Test of "ratio of arithmetic mean to geometric mean":

$$\begin{aligned}(\Lambda_{GLR})^{2/(np)} &= \frac{\hat{\sigma}^2}{|\hat{R}|^{1/p}} \\ &= \frac{p^{-1}\sum_{i=1}^p \lambda_i^{\hat{R}}}{\prod_{i=1}^p \left(\lambda_i^{\hat{R}}\right)^{1/p}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma\end{aligned}$$

12.5 CONFIDENCE REGIONS ON VECTOR MEAN

Recall: from the level α double sided test of vector mean we know

$$P_{\underline{\theta}}\left(n(\overline{\underline{x}_i} - \underline{\mu}_o)^T \mathsf{S}^{-1}(\overline{\underline{x}_i} - \underline{\mu}_o) \ > \ \mathcal{T}^{-2}_{p,n-p}(1-\alpha)\right) = \alpha$$

where $\underline{\theta} = [\underline{\mu}, R]$.

Equivalently

$$P_{\underline{\theta}}\left(n(\overline{\underline{x}_i} - \underline{\mu}_o)^T \mathsf{S}^{-1}(\overline{\underline{x}_i} - \underline{\mu}_o) \ \leq \ \mathcal{T}^{-2}_{p,n-p}(1-\alpha)\right) = 1-\alpha$$

This is a "simultaneous confidence statement" on all elements of mean vector $\underline{\mu}$ for unknown covariance R given measurement X

$\Rightarrow$ $(1-\alpha)$% confidence region on $\underline{\mu}$ is the ellipsoid

$$\left\{\underline{\mu} : n(\overline{\underline{x}_i} - \underline{\mu})^T \mathsf{S}^{-1}(\overline{\underline{x}_i} - \underline{\mu}) \ \leq \ \mathcal{T}^{-2}_{p,n-p}(1-\alpha)\right\}$$

.

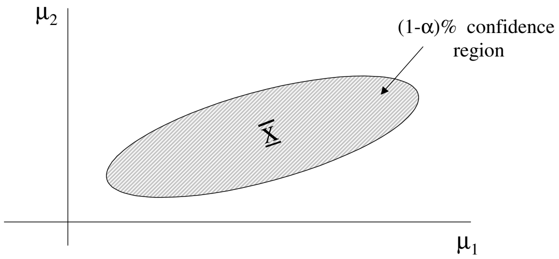


Figure 185: *Confidence region for all elements of mean vector* $\underline{\mu}$ *is an ellipsoid*

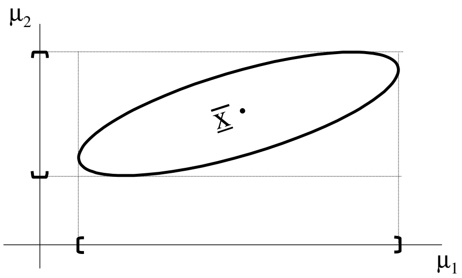


Figure 186: *Confidence ellipsoid gives "marginal" confidence intervals on each element of* $\underline{\mu} = [\mu_1, \ldots, \mu_p]^T$

12.6 EXAMPLES

Example 44 *Confidence band on a periodic signal in noise*

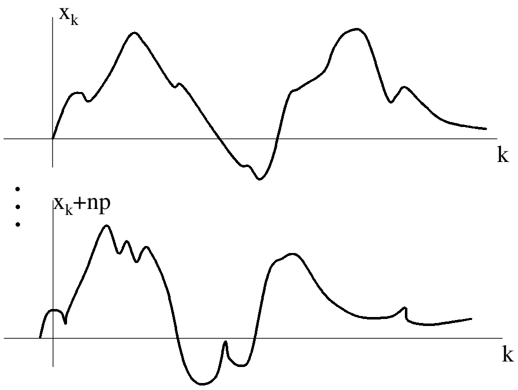


Figure 187: *Multiple uncorrelated measurements of a segment of a periodic signal.*

$$x_k = s_k + v_k$$

* $s_k = s_{k+nT_p}$: unknown periodic signal with known period T_p

* v_k: zero mean w.s.s. noise of bandwidth $1/(MT_p)$ Hz

Step 1: construct measurement matrix

$$\underline{x}_i = [x_{1+(i-1)MT_p}, \ldots, x_{T_p+(i-1)MT_p}]^T$$

Step 2: find conf. intervals on each s_k from ellipsoid

$$[(\overline{\underline{x}_i})_k - l_k \ \leq s_k \leq (\overline{\underline{x}_i})_k + u_k]$$

Example 45 *CFAR signal detection in narrowband uncalibrated array*

k-th snapshot of p-sensor array output:

$$\underline{x}_k = \underline{a}\, s + \underline{v}_k, \quad k = 1, \ldots, n. \tag{48}$$

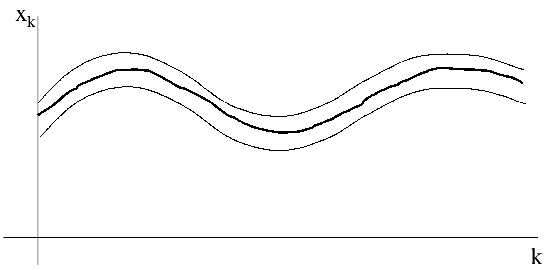


Figure 188: *Confidence band on signal over one signal period.*

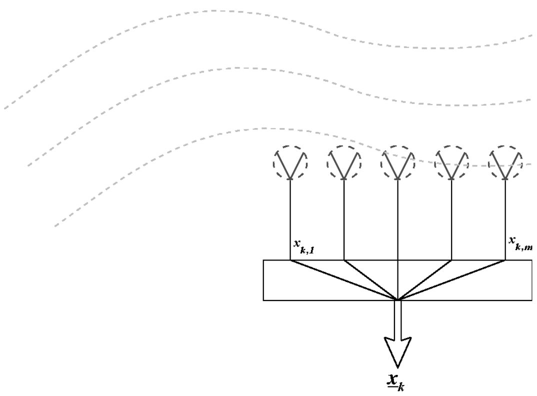


Figure 189: *Sensor array generates spatio-temporal measurement.*

* $\underline{a}$: unknown array response (steering) vector

* $\underline{v}_k$: Gaussian $\mathcal{N}_p(0, R)$ array noise vector with unknown spatial covariance R

* s: unknown deterministic signal amplitude

Objective: detect presence of any non-zero signal amplitude at level α

$$H_0 : s = 0, \quad k = 1, \ldots, n$$

$$H_1 : s \neq 0, \quad k = 1, \ldots, n$$

This is equivalent to

$$H_0 : E[\underline{x}_i] = \underline{\mu} = 0, \quad R > 0$$

$$H_1 : E[\underline{x}_i] = \underline{\mu} \neq 0, \quad R > 0$$

For which we know:

* level α GLRT is the Hotelling T^2 test

* confidence region for $\underline{\mu} = \underline{a}s$ is an ellipsoid.

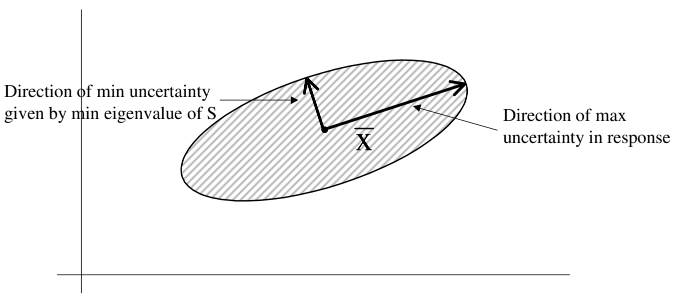


Figure 190: *Confidence region for array response vector $\underline{a}s$ is an ellipse in 2D.*

12.7 Exercises

1. Extend the multivariate paired-t test derived in Sec. 12.2 to the case where $\underline{x}_i \sim \mathcal{N}(\underline{\mu}_x, R_x)$ and $\underline{y}_i \sim \mathcal{N}(\underline{\mu}_y, R_y)$ for the case that the two covariance matrices R_x and R_y may be unequal and are unknown. How many degrees of freedom does the the asymptotic Chi-square distribution have?
2. In Example 45 the optimal CFAR detector for a scalar signal s viewed from a p-sensor array output with array response $\underline{a}$ and noise $\underline{v}_k$. In this problem we extend this to CFAR detection of multiple (m) scalar signals $\underline{s} = [s_1, \ldots, s_m]$ following the observation model:

$$\underline{x}_k = A\underline{s} + \underline{v}_k, \quad k = 1, \ldots, n \tag{49}$$

   where $A = [\underline{a}_1, \ldots, \underline{a}_p]$ is an unknown $p \times m$ matrix and $\underline{v}_k$ are i.i.d. $\mathcal{N}(0, R)$ random vectors with unknown covariance R. Derive the GLRT for this problem. How many degrees of freedom does the the asymptotic Chi-square distribution have?
3. Consider the same model as (48) but assume that s is a Gaussian distributed random variable and $\underline{a}$ and R are unknown. Derive the GLRT.
4. Consider the same scalar model as (48) but now assume that $\underline{a}$ is known while the noise covariance R is unknown. Derive the GLRT.
5. Extend the analysis of the previous problem to the multiple signal case (49) when A has columns of sinusoidal form:

$$\underline{a}_k = [1, \cos(2\pi f_k), \ldots, \cos(2\pi f_k(p-1))]^T, \quad k = 1, \ldots, m$$

   while the noise covariance R is unknown. Derive the GLRT (you may assume that the $\underline{a}_k$'s are orthogonal if you wish).

End of chapter

13 DETECTION AND ESTIMATION IN COMPLEX GAUSSIAN CASE

OUTLINE

* Complex Gaussian random variables

* Complex Gaussian random vectors

13.1 COMPLEX GAUSSIAN SCALARS

Let x_R, x_I be bivariate Gaussian r.v.'s with mean μ and covariance R

Consider the complex r.v.

$$x = x_R + jx_I$$

Mean is complex

$$\mu = E[x] = E[x_R] + jE[x_I]$$

Variance is positive real

$$\text{var}(x) := E[|x - E[x]|^2] = E[(x - E[x])(x - E[x])^*]$$

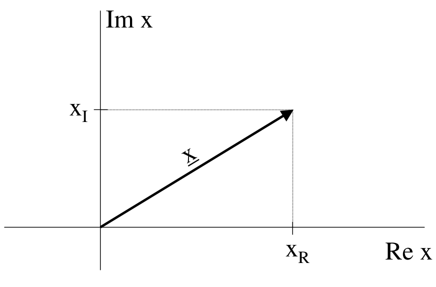


Figure 191: *A complex r.v. in terms of its real and imaginary components.*

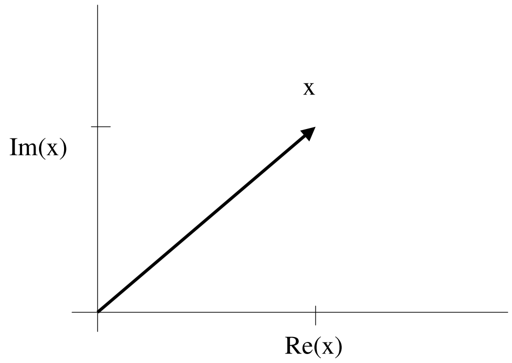


Figure 192: *A complex r.v. as a vector in the complex plane.*

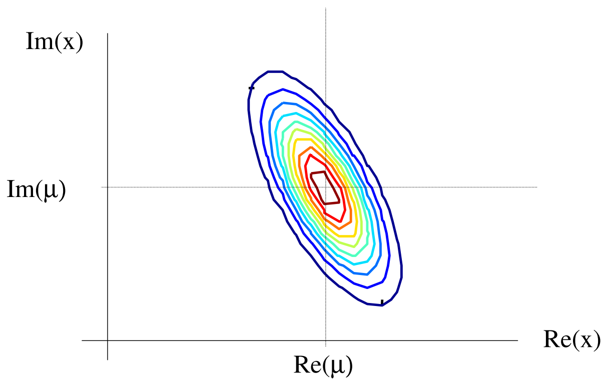


Figure 193: *Constant contours of joint density of a complex Gaussian r.v.*

Definition 3 *z is said to be a standard circularly complex Gaussian (CG) r.v. when x_R, x_I are uncorrelated and have identical variance.*

As for real case we say z is standard CG r.v. when $E[x] = 0 + j0$ and $\text{var}(x) = 1$.

SPECIAL CASE: standard circular CG r.v.

Condition 1: $E[x_R] = E[x_I] = 0$

Condition 2: $\text{var}(x_R) = \text{var}(x_I) = 1/2$

Then

$$E[x] = 0, \qquad \text{var}(x) = \text{var}(x_R) + \text{var}(x_I) = 2\ \text{var}(x_R) = 1$$

As in real Gaussian case, can generate general circular Gaussian z by affine transformation

$$z = a \cdot x + b$$

$\Rightarrow E[z] = a \cdot E[x] + b = b, \quad \text{var}(z) = a \cdot a^* \text{var}(x) = |a|^2$

Circular CG r.v. x with mean $\mu = [\mu_R,\ \mu_I]^T$ variance σ^2 has p.d.f.

$$\begin{aligned} f(x; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi(\sigma^2/2)}} \exp\left(-\frac{1}{2(\sigma^2/2)}(x_R - \mu_R)^2\right) \\ &\quad \cdot \frac{1}{\sqrt{2\pi(\sigma^2/2)}} \exp\left(-\frac{1}{2(\sigma^2/2)}(x_I - \mu_I)^2\right) \\ &= \frac{1}{\pi\sigma^2} \exp\left(-\frac{1}{\sigma^2}|x-\mu|^2\right) \end{aligned}$$

13.2 COMPLEX GAUSSIAN VECTORS

Measurement: $\underline{x} = [x_1, \ldots, x_n]^T$ a CG random vector

mean: $E[\underline{x}] = \underline{\mu}_x = [\mu_1, \ldots, \mu_n]^T$

covariance: $\text{cov}(\underline{x}) = E[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^H] = R$

For circular Gaussian case joint density for $\underline{x}$ is

$$f(\underline{x};\ \underline{\mu}, R) = \frac{1}{\pi^n |R|} \exp\left(-(\underline{x} - \underline{\mu})^H R^{-1} (\underline{x} - \underline{\mu})\right)$$

CG has properties analogous to real Gaussian case

PROPERTY 1. If $\underline{x}$ is CG vector, A, $\underline{b}$ arbitrary complex matrix and vector

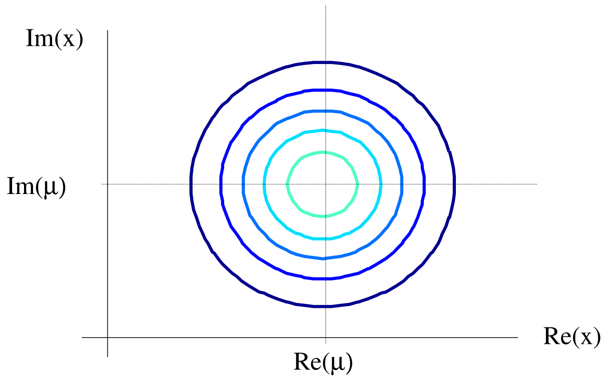


Figure 194: *Constant contours of joint density of a circular complex Gaussian r.v.*

$$\underline{z} = A\underline{x} + \underline{b} \quad \text{is CG vector}$$

* $E[\underline{z}] = A\mu + \underline{b}$

* $\text{cov}(\underline{z}) = A\text{cov}(\underline{x})A^H$

PROPERTY 2. If $\underline{x}$ and $\underline{y}$ are uncorrelated CG

$$E[\underline{x}\ \underline{y}^H] = E[\underline{x}]\ E[\underline{y}^H]$$

$\Rightarrow$ $\underline{x}$ and $\underline{y}$ are independent

PROPERTY 3: $\underline{x}$, $\underline{y}$ CG implies that

$a\underline{x} + b\underline{y}$ is CG for any $a, b \in \mathbb{C}$

PROPERTY 4. If $\underline{z}$ CG then $\text{Re}(\underline{x})$ is Gaussian with

* mean = $E[\text{Re}\underline{x}] = \text{Re}E[\underline{x}]$

* variance =$\text{var}(\text{Re}\underline{x}) = \frac{1}{2}\text{var}(\underline{x})$

PROPERTY 5: $|\mathcal{CN}(0, \sigma^2)|^2 = \chi_2\ \sigma^2/2$

Derivation

$$|\mathcal{CN}(0, \sigma^2)|^2 \quad = \quad |\mathcal{N}(0,1)\sqrt{\sigma^2/2} + j\mathcal{N}(0,1)\sqrt{\sigma^2/2}|^2$$

$$
\begin{aligned}
&= |\mathcal{N}(0,1)|^2\ \sigma^2/2 + |\mathcal{N}(0,1)|^2\ \sigma^2/2 \\
&= (\underbrace{|\mathcal{N}(0,1)|^2 + |\mathcal{N}(0,1)|^2}_{\chi_2})\sigma^2/2
\end{aligned}
$$

PROPERTY 6: $|\mathcal{CN}(\mu,\sigma^2)|^2 = \chi_{2,|d|^2}\ \sigma^2/2$

where $d = \frac{\mu}{\sqrt{\sigma^2/2}}$

Derivation

$$|\mathcal{CN}(\mu,\sigma^2)|^2 = |\mathcal{N}(\mu,\sigma^2/2)|^2 + |\mathcal{N}(\mu,\sigma^2/2)|^2 =$$

$$= |\mathrm{Re}\{\mu\} + \mathcal{N}(0,1)\sqrt{\sigma^2/2}|^2 + |\mathrm{Im}\{\mu\} + \mathcal{N}(0,1)\sqrt{\sigma^2/2}|^2$$

$$= |\mathrm{Re}\{d\} + \mathcal{N}(0,1)|^2\ \sigma^2/2 + |\mathrm{Im}\{d\} + \mathcal{N}(0,1)|^2\ \sigma^2/2$$

$$= \chi_{2,|d|^2}\ \sigma^2/2$$

13.3 GENERAL COMPLEX GAUSSIAN DETECTION PROBLEM

Measurement: $\underline{x} = [x_1, \ldots, x_n]^T$ is a complex Gaussian random vector

$\underline{x} \sim \mathcal{CN}_n(\underline{\mu}, R)$

mean: $\underline{\mu} = E[\underline{x}] = [\mu_1, \ldots, \mu_n]^T$

covariance: $R = E[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^H]$

Consider simple detection problem

$$
\begin{array}{l}
H_0 : \underline{\mu} = \underline{\mu}_0, \quad R = R_0 \\
H_1 : \underline{\mu} = \underline{\mu}_1, \quad R = R_1
\end{array}
$$

Likelihood ratio is

$$\Lambda = \frac{|R_0| \exp\left(-(\underline{x} - \underline{\mu}_1)^H R_1^{-1}(\underline{x} - \underline{\mu}_1)\right)}{|R_1| \exp\left(-(\underline{x} - \underline{\mu}_0)^H R_0^{-1}(\underline{x} - \underline{\mu}_0)\right)}$$

Giving an LRT similar to real case

$$T(\underline{x}) \quad = \quad (\underline{x} - \underline{\mu}_0)^H R_0^{-1}(\underline{x} - \underline{\mu}_0) - (\underline{x} - \underline{\mu}_1)^H R_1^{-1}(\underline{x} - \underline{\mu}_0) \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

where

$$\gamma = \log \eta + \log \frac{|R_1|}{|R_0|}$$

13.4 CASE OF EQUAL COVARIANCES

Assume $R_0 = R_1 = R$

Then, defining $\Delta\underline{\mu} = \underline{\mu}_1 - \underline{\mu}_0$

$$T(\underline{x}) = (\underline{x} - \underline{\mu}_0)^H R^{-1}(\underline{x} - \underline{\mu}_0) - (\underline{x} - \underline{\mu}_1)^H R^{-1}(\underline{x} - \underline{\mu}_0)$$

$$= \underbrace{\underline{x}^H R^{-1}\Delta\underline{\mu} + \Delta\underline{\mu}^H R^{-1}\underline{x}}_{2\mathrm{R}e\{\Delta\underline{\mu}^H R^{-1}\underline{x}\}} - \underbrace{(\underline{\mu}_1^H R^{-1}\underline{\mu}_1 - \underline{\mu}_0^H R^{-1}\underline{\mu}_0)}_{const}$$

Thus LRT is simply

$$T'(\underline{x}) = \mathrm{R}e\{\Delta\underline{\mu}^H R^{-1}\underline{x}\} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

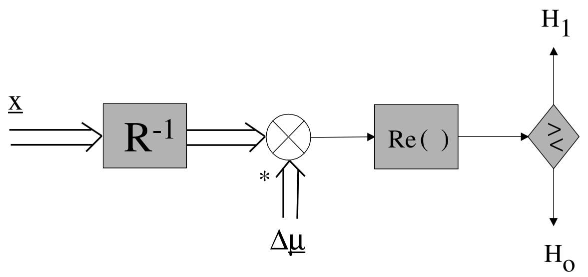


Figure 195: *Block diagram of LRT for testing between two different vector means μ_1 and μ_0 of a circular complex Gaussian vector with known covariance R^{-1}. Here $\Delta\mu = \mu_1 - \mu_0$.*

Performance:

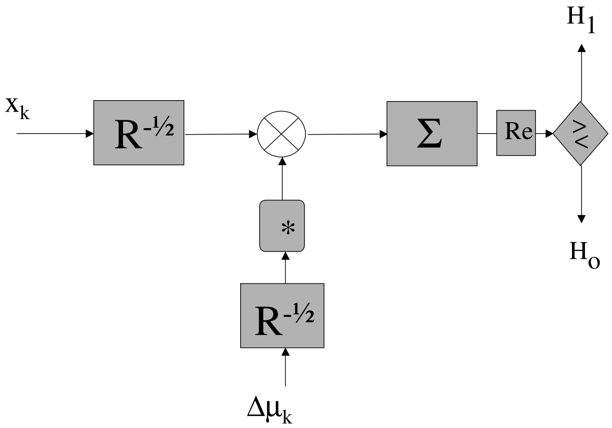


Figure 196: *Prewhitener implementation of LRT in Fig. 195. In this block diagram $R^{-\frac{1}{2}}$ denotes a whitening filter acting on the sequences x_k and $\Delta\mu_k$ (elements of the vectors $\underline{x}$ and $\Delta\underline{\mu}$) such that, with $\underline{x} = [x_1, \ldots, x_n]$, the matrix-vector product $R^{-\frac{1}{2}}\underline{x}$ is circular complex Gaussian with covariance I.*

Under H_0: T' is real Gaussian

$$
\begin{aligned}
E[T'(\underline{x})|H_0] &= \mathrm{Re}\{E[\,\Delta\underline{\mu}^H R^{-1}\underline{x}\,|H_0]\} \\
&= \mathrm{Re}\{\Delta\underline{\mu}^H R^{-1}\,\underbrace{E[\underline{x}\,|H_0]}_{\underline{\mu}_0}\} \\
&= \mathrm{Re}\{\Delta\underline{\mu}^H R^{-1}\underline{\mu}_0\}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{var}(T'(\underline{x})|H_0] &= \tfrac{1}{2}\mathrm{var}(\Delta\underline{\mu}^H R^{-1}\underline{x}) \\
&= \tfrac{1}{2}\Delta\underline{\mu}^H R^{-1}\mathrm{cov}(\underline{x})R^{-1}\Delta\underline{\mu} \\
&= \tfrac{1}{2}\Delta\underline{\mu}^H R^{-1}\Delta\underline{\mu}
\end{aligned}
$$

Thus level α test is simply obtained by setting

$$
\gamma = \sqrt{\tfrac{1}{2}\Delta\underline{\mu}^H R^{-1}\Delta\underline{\mu}}\,\mathcal{N}^{-1}(1-\alpha) + \mathrm{Re}\{\Delta\underline{\mu}^H R^{-1}\underline{\mu}_0\}
$$

SPECIAL CASE $n = 1$

$$T^{'}(\underline{x}) = \mathrm{Re}\{\Delta\underline{\mu}^H R^{-1}\underline{x}\} = \frac{\mathrm{R}e\{\Delta\mu^* x\}}{\sigma^2} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma$$

Or equivalently

$$T^{''}(x) = \frac{\mathrm{R}e\{\Delta\mu^* x\}}{\|\Delta\mu\|} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma^{'}$$

Let

* $\Delta\mu = \Delta_R + j\Delta_I$

* $x = x_R + jx_I$

Then

$$\mathrm{R}e\{\Delta\mu^* x\} = x_R\Delta_R + x_I\Delta_I = [x_R, x_I]\begin{bmatrix} \Delta_R \\ \Delta_I \end{bmatrix}$$

so the detector is identical to real case with $n = 2$

$$\frac{< [x_R, x_I],\ [\Delta_R, \Delta_I] >}{\|[\Delta_R, \Delta_I]\|} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma$$

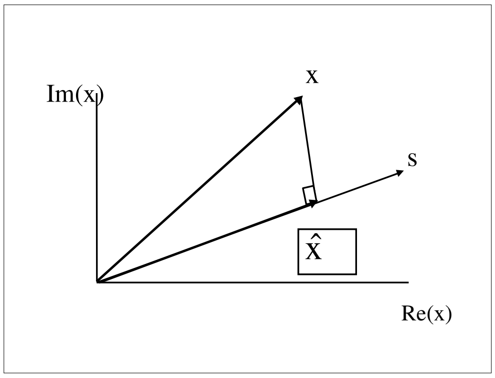


Figure 197: *Complex LRT for $n = 1$ is identical to the real LRT for $n = 2$: i.e. a test on the amplitude of the projection of complex observation $\underline{x}$ onto complex signal vector $\underline{s}$.*

Example 46 *Detection of spatio temporal complex Gaussian signal*

k-th snapshot of p-sensor array output is a *multi-channel* complex measurement:

$$\underline{x}_k = \underline{a}\, s + \underline{v}_k, \quad k = 1, \ldots, n$$

or equivalently we have $p \times n$ space-time measurement matrix

$$X = [\underline{x}_1, \ldots, \underline{x}_n] = s\, \underline{a}\, \underline{1}^T + V$$

* $\underline{a}$: complex array response ("steering" vector)

* $V = [\underline{v}_1, \ldots, \underline{v}_n]$: i.i.d. complex Gaussian $\underline{v}_k \sim \mathcal{CN}_p(0, R)$

* known spatial covariance R

* s: deterministic complex signal amplitude

CASE 1: $s = s_1$ known

$$\begin{aligned} H_0 : s &= 0, \quad k = 1, \ldots, n \\ H_1 : s &= s_1, \quad k = 1, \ldots, n \end{aligned}$$

Using prewhitening and coordinate rotations as in real case:

* Reduce the problem to testing a single vector observation $\underline{z}$

$$\begin{array}{lcl} H_0 : \underline{z} = \underline{\tilde{v}} & & H_0 : z_k = \tilde{v}_k \\ & \Leftrightarrow & \\ H_1 : \underline{z} = s_1\, \tilde{a}\, \underline{1} + \underline{\tilde{v}} & & H_1 : z_k = s_1\, \tilde{a} + \tilde{v}_k \end{array}$$

where

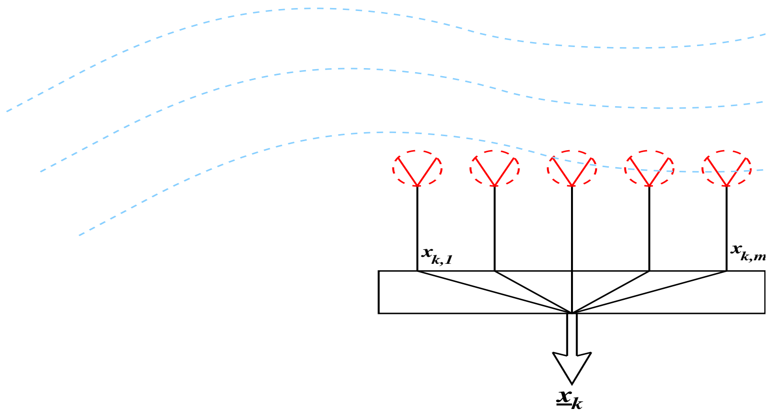
* $\underline{\tilde{v}}$ are i.i.d. $\mathcal{CN}(0, 1)$ r.v.'s

* $\tilde{a} = \sqrt{\underline{a}^H R^{-1} \underline{a}}$

The LRT is therefore

$$T'(X) \;=\; \mathrm{Re}\{\underbrace{\Delta\underline{\mu}^H}_{s_1^*\, \tilde{a}^* \underline{1}^T}\, \overbrace{R^{-1}}^{I}\, \underbrace{\underline{x}}_{\underline{z}}\} \overset{H_1}{\underset{H_0}{\gtrless}} \sqrt{\tfrac{1}{2}\, \underbrace{\Delta\underline{\mu}^H R^{-1} \Delta\underline{\mu}}_{|s_1\tilde{a}|^2 \underline{1}^T \underline{1} = n|s_1\tilde{a}|^2}}\, \mathcal{N}^{-1}(1-\alpha)$$

or, equivalently, since $\tilde{a} > 0$



(a) (NEW)

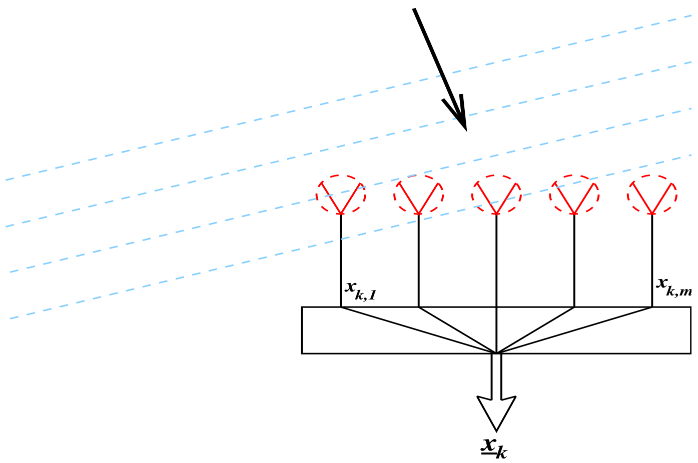


Figure 198: Narrowband sensor array generates spatio-temporal measurement.

$$\sqrt{2n}\ \mathrm{Re}\left\{\frac{s_1^*}{|s_1|}\ \ \overline{z_i}\right\} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \mathcal{N}^{-1}(1-\alpha)$$

NOTE: as contrasted to case of real spatio-temporal signal, the LRT is not UMP w.r.t. signal amplitude s

$\Rightarrow$ Need to know signal phase offset $s_1^*/|s_1|$!

CASE 2: unknown signal with complex amplitude

$$\begin{aligned} H_0 : s &= 0, \quad k=1,\ldots,n \\ H_1 : s &\neq 0, \quad k=1,\ldots,n \end{aligned}$$

Solution: GLRT

The LR is

$$\begin{aligned} \Lambda(X;s) &= \exp\left(2\mathrm{R}e\{\Delta\underline{\mu}^H R^{-1}\underline{x}\} - \underline{\mu}_1^H R^{-1}\underline{\mu}_1\right) \\ &= \exp\left(2\mathrm{R}e\{s^*\tilde{a}^*\ \underline{1}^T\underline{z}\} - n|s\tilde{a}|^2\right) \end{aligned}$$

Complete square in exponent

$$\ln\Lambda(X;s_1) = -n|s\tilde{a} - \overline{z_i}|^2 + n|\overline{z_i}|^2$$

and conclude that:

* $\hat{s} = \overline{z_i}/\tilde{a}$

* GLRT test is

$$\ln\Lambda(X;\hat{s}) = n|\overline{z_i}|^2 \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

PERFORMANCE

Under H_0: $\overline{z_i}\ \sim\ \mathcal{CN}(0,1/n)$

Use Property 5: $|\mathcal{CN}(0,\sigma^2)|^2 = (\chi_2)\ \sigma^2/2$

$\Rightarrow |\overline{z_i}|^2\ =\ (\chi_2)\ \frac{1}{2n}$

giving level α GLRT

$$2n\ |\overline{z_i}|^2 \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \chi_2^{-1}(1-\alpha)$$

Under H_1: $\overline{z_i}\ \sim\ \mathcal{CN}(s\tilde{a}, 1/n)$

Use Property 6: $|\mathcal{CN}(\mu,\sigma^2)|^2 = (\chi_{2,|d|^2})\ \sigma^2/2$,

where $|d|^2 = |\mu|^2/(\sigma^2/2)$

$\Rightarrow 2n|\overline{z_i}|^2\ =\ \chi_{2,\underbrace{2n|s|^2\tilde{a}^2}_{|d|^2}}$

Power curve:

$$\begin{aligned} \beta &= 1 - \chi_{2,|d|^2}(\chi_2^{-1}(1-\alpha)) \\ &= Q\left(|d|, \sqrt{1/\ln\alpha}\right) \end{aligned}$$

where $Q(a,b)$ is the Marcum-Q function

$$Q(a,b) = 1 - \int_0^b v \exp\left(\frac{-(v^2+a^2)}{2}\right)\ I_o(av)dv$$

REMARKS on: $2n|\overline{z_i}|^2 \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \chi_2^{-1}(1-\alpha)$

1. GLRT detects length of projection of $\overline{z_i}$ onto ”whitened signal vector” $n^{-1}\underline{1}$

2. An alternative form compares a Rayleigh distributed statistic

$$\sqrt{2n}\ |\overline{z_i}| \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \mathcal{R}^{-1}(1-\alpha)$$

3. Compare $2n|\overline{z_i}|^2 \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \chi_2^{-1}(1-\alpha)$

to the GLRT for the real case derived previously:

$$\sqrt{n}\ |\overline{z_i}| = \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \mathcal{N}^{-1}(1-\alpha/2)$$

End of chapter

14 CONTINUOUS TIME EXTENSIONS

* Binary classification of non-random signals in colored noise: KLE

* Prewhitening receivers

* Coherent and incoherent detection of narrowband signals

References

Whalen [37]

Helstrom [10]

Van Trees [36]

Poor [24]

14.1 BINARY SIGNAL CLASSIFICATION

Assume have waveform measurements $x(t), t \in [0, T]$

To test hypotheses

$$
\begin{aligned}
H_0 : x(t) &= s_0(t) + n(t) \\
H_1 : x(t) &= s_1(t) + n(t)
\end{aligned}
$$

* s_0, s_1 are known functions

* $s_0 = 0$ reduces to signal detection problem

* $n(t)$ is Gaussian

$$
E[n(t)] = 0, \qquad E[n(t)n^*(\tau)] = r_n(t, \tau)
$$

We know that LRT is "optimal" test

$$
\Lambda(x) = \frac{f(x|H_1)}{f(x|H_0)} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta
$$

Difficulty: how to find p.d.f.'s $f(x|H_k)$?

Solution: Convert problem into "discrete time"

Two approaches:

1. Time sample $x(t)$ at "Nyquist rate" to obtain equivalent measurements $\{x(kT_s)\}_k$.

2. Expand $x(t)$ in a series of orthonormal basis functions

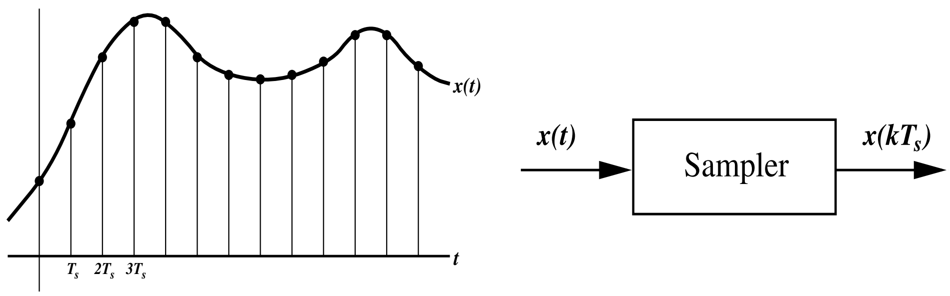


Figure 199: *Nyquist rate sampling provides a sequence* $\{x(kT_s)\}_{k=-\infty}^{\infty}$ *which is equivalent to continuous time bandlimited Gaussian signal* $x(t)$*,*

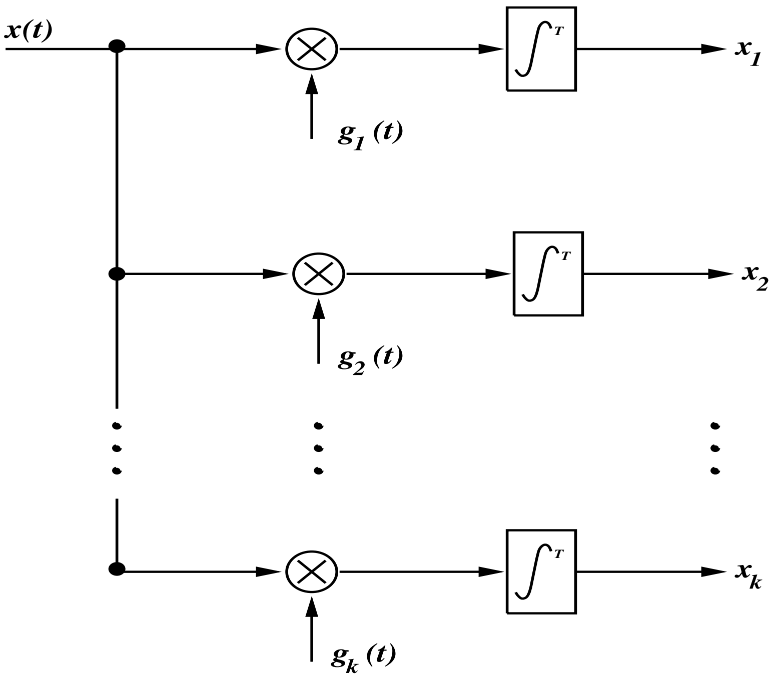


Figure 200: *Orthogonal sampling produces a set of r.v.s* $\{x_i\}_{i=1}^{k}$ *which become equivalent to continuous time Gaussian signal* $x(t)$ *as* $k \to \infty$*.*

14.2 KARHUNEN LOEVE EXPANSION

Let $g_1, g_2, \ldots$ be a c.o.n. basis for time functions over $t \in [0, T]$

$$< g_i, g_j >= \int_0^T g_i(t) g_j^*(t) dt = \Delta_{ij}$$

Then

$$x(t) = \sum_{i=1}^{\infty} x_i g_i(t)$$

where

$$x_i = \int_0^T x(t) g_i^*(t) dt$$

Properties of $\{x_i\}$

* $\{x_i\}$ is an equivalent "discrete time" measurement

* $\{x_i\}$ is a Gaussian random sequence

* For $k = 0, 1$

$$E[x_i | H_k] = \int_0^T E[x(t)|H_k] g_i^*(t) dt$$

$$\begin{aligned} \text{cov}(x_i, x_j | H_k) &= \int_0^T dt \int_0^T d\tau \; \text{cov}(x(t), x(\tau)|H_k) \; g_i(t) g_j^*(\tau) \\ &= \int_0^T dt \int_0^T d\tau \; r_n(t, \tau) \; g_i(t) g_j^*(\tau) \qquad \text{indep. of } H_k \end{aligned}$$

Q: Is it possible to choose g_i such that x_i's are independent?

A. Yes if $\{g_i\}_i$ are chosen as "eigenfunctions" of $r_n(t, \tau)$

Rewrite $\text{cov}(x_i, x_j)$ as

$$\text{cov}(x_i, x_j) = \int_0^T d\tau \underbrace{\int_0^T dt \; r_n(t, \tau) \; g_i(t)}_{q_i(\tau)} \; g_j^*(\tau)$$

If could choose

$$q_i(\tau) = \lambda_i g_i(\tau)$$

then would have uncorrelated x_i's

$$\begin{aligned}\mathrm{cov}(x_i, x_j) &= \lambda_i \underbrace{\int_0^T g_i(\tau)\ g_j^*(\tau)\ d\tau}_{\Delta_{ij}} \\ &= \begin{cases} \lambda_i, & i = j \\ 0, & i \neq j \end{cases}\end{aligned}$$

Required condition $q_i(\tau) = \lambda_i g_i(\tau)$ is "eigenfunction equation"

$$\int_0^T r_n(t,\tau) g_i(t)\ dt \ = \ \lambda_i\ g_i(\tau), \qquad \tau \in [0, T]$$

where λ_i can be recovered by integrating equation

$$\begin{aligned}\lambda_i &= \int_0^T (\lambda_i\ g_i(\tau))\ \ g_i^*(\tau)\ d\tau \\ &= \int_0^T d\tau \int_0^T dt g_i^*(\tau) r_n(t,\tau) g_i(t)\ dt = \lambda_i\end{aligned}$$

* $g_i(t)$'s are eigenfunctions of $r_n(t,\tau)$

* λ_i's are eigenvalues of $r_n(t,\tau)$

* resultant series expansion is the Karhunen Loeve expansion

$$x(t) = \sum_{i=1}^{\infty} x_i g_i(t)$$

Mercer's Theorem:

$$r_n(t,\tau) = \sum_{i=1}^{\infty} \lambda_i g_i(t) g_i^*(\tau)$$

* This is an SVD for continuous time a.c.f.

* square root and inverse operators can be defined as

$$
\begin{aligned}
r_n^{\frac{1}{2}}(t,\tau) &= \sum_{i=1}^{\infty} \sqrt{\lambda_i} g_i(t) g_i^*(\tau) \\
r_n^{-1}(t,\tau) &= \sum_{i=1}^{\infty} \frac{1}{\lambda_i} g_i(t) g_i^*(\tau)
\end{aligned}
$$

Obtain identities

$$\delta(t-\tau) = \int_0^T r_n(t,u) r_n^{-1}(u,\tau)\ du$$

$$r_n^{-1}(t,\tau) = \int_0^T r_n^{-\frac{1}{2}}(t,u) r_n^{-\frac{1}{2}}(u,\tau)\ du$$

Special case of white noise: $r_n(t,\tau) = \frac{N_o}{2}\ \delta(t-\tau)$

* Eigenfunction equation becomes

$$
\begin{aligned}
\lambda_i\ g_i(\tau) &= \int_0^T r_n(t,\tau) g_i(t)\ dt \\
&= \frac{N_o}{2} g_i(\tau), \qquad \tau \in [0,T]
\end{aligned}
$$

* Identify:

$$
\begin{aligned}
\lambda_i &= \frac{N_o}{2} \\
g_i(t) &= \text{arbitrary c.o.n.}
\end{aligned}
$$

Apply Mercer's thm to get representation for dirac delta

$$\delta(t-\tau) = \sum_{i=1}^{\infty} g_i(t) g_i^*(\tau)$$

14.3 APPLICATION OF KLE

$$
\begin{array}{lcl}
H_0 : x(t) = s_0(t) + n(t) & & H_0 : x_i = s_{0i} + n_i \\
 & \Leftrightarrow & \\
H_1 : x(t) = s_1(t) + n(t) & & H_1 : x_i = s_{1i} + n_i
\end{array}
$$

where

$$s_{0i} = \int_0^T s_0(t) g_i(t) dt, \qquad s_{1i} = \int_0^T s_1(t) g_i(t) dt$$

and

$$\begin{aligned} n_i &= \text{i.i.d. Gaussian under } H_0, H_1 \\ E[n_i|H_k] &= 0 \\ \text{var}(n_i|H_k) &= \lambda_i \end{aligned}$$

Thus we can now write down the LRT statistic:

$$\Lambda(x) = \lim_{k\to\infty} \Lambda(x_1, \ldots, x_k)$$

where

$$\begin{aligned} \Lambda(x_1, \ldots, x_k) &= \frac{f(x_1, \ldots, x_k|H_1)}{f(x_1, \ldots, x_k|H_0)} \\ &= \prod_{i=1}^k \frac{f(x_i|H_1)}{f(x_i|H_0)} \end{aligned}$$

As x_i is Gaussian with mean s_{ki} and variance λ_i under H_k

$$\begin{aligned} \frac{f(x_i|H_1)}{f(x_i|H_0)} &= \exp\left(-\frac{1}{2\lambda_i}(x_i - s_{1i})^2 + \frac{1}{2\lambda_i}(x_i - s_{0i})^2\right) \\ &= \exp\left(\frac{1}{\lambda_i}\,(s_{1i} - s_{0i})x_i - \frac{1}{2\lambda_i}\,(s_{1i}^2 - s_{0i}^2)\right) \end{aligned}$$

(real x)

Therefore LRT statistic is

$$\Lambda(x) = \exp\left(\overbrace{\sum_{i=1}^\infty \frac{1}{\lambda_i}\,(s_{1i} - s_{0i})x_i}^{T(x)} - \overbrace{\sum_{i=1}^\infty \frac{1}{2\lambda_i}\,(s_{1i}^2 - s_{0i}^2)}^{B}\right)$$

or, as B is independent of x, LRT is

$$T(x) \overset{H_1}{\underset{H_0}{\gtrless}} \gamma = \ln \eta + B$$

It remains to express LRT statistic in time domain

$$\begin{aligned}
T(x) &= \sum_{i=1}^{\infty} \frac{1}{\lambda_i} (s_{1i} - s_{0i}) x_i \\
&= \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \left(\int_0^T (s_1(t) - s_0(t)) g_i(t) dt \right) \left(\int_0^T x(\tau) g_i(\tau) d\tau \right) \\
&= \int_0^T dt \int_0^T d\tau (s_1(t) - s_0(t)) x(\tau) \underbrace{\sum_{i=1}^{\infty} \frac{1}{\lambda_i} g_i(t) g_i(\tau)}_{r_n^{-1}(t,\tau)}
\end{aligned}$$

Thus

$$T(x) = \int_0^T dt \int_0^T d\tau (s_1(t) - s_0(t)) x(\tau) r_n^{-1}(t,\tau)$$

Likewise

$$B = \frac{1}{2} \int_0^T dt \int_0^T d\tau (s_1^2(t) - s_0^2(t)) r_n^{-1}(t,\tau)$$

For w.s.s. noise $n(t)$ the a.c.f. is $r_n(t,\tau) = r_n(t - \tau)$ and

$$\begin{aligned}
A(x) &= \int_0^T dt \; (s_1(t) - s_0(t)) \underbrace{\int_0^T r_n^{-1}(t-\tau) x(\tau) d\tau}_{r_n^{-1} * x} \\
&= (s_1(T-t) - s_0(T-t)) * r_n^{-1}(t) * x(t) \big|_{t=T}
\end{aligned}$$

PREWHITENING FILTER IMPLEMENTATION OF LRT

Use (square root factor) identity

$$r_n^{-1}(t,\tau) = \int_0^T r_n^{-\frac{1}{2}}(t,u) r_n^{-\frac{1}{2}}(u,\tau) \; du$$

and define

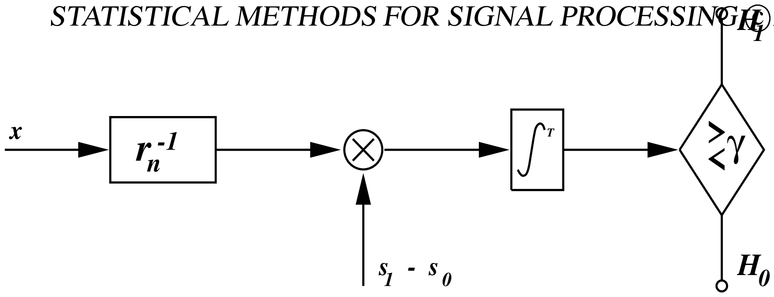


Figure 201: *The LRT for deciding between one of two known signals $s_1(t)$ and $s_0(t)$ in additive noise for known noise auto-correlation function $r_n(t,\tau)$.*

$$\begin{aligned}
\tilde{x}(t) &= \int_0^T r_n^{-\frac{1}{2}}(t,u)\; x(u)\; du \\
\tilde{s}_0(t) &= \int_0^T r_n^{-\frac{1}{2}}(t,u)\; s_0(u)\; du \\
\tilde{s}_1(t) &= \int_0^T r_n^{-\frac{1}{2}}(t,u)\; s_1(u)\; du
\end{aligned}$$

to obtain the equivalent expressions

$$\begin{aligned}
T(x) &= \int_0^T (\tilde{s}_1(t) - \tilde{s}_0(t))\tilde{x}(t)\; dt = <(\tilde{s}_1 - \tilde{s}_0),\; \tilde{x}> \\
B &= \tfrac{1}{2}\int_0^T \left(\tilde{s}_1^2(t) - \tilde{s}_0^2(t)\right)\; dt = \tfrac{1}{2}\left(\|\tilde{s}_1\|^2 - \|\tilde{s}_0\|^2\right)
\end{aligned}$$

Obtain Master Formula for LR statistic:

$$\Lambda(x) = e^{<(\tilde{s}_1 - \tilde{s}_0),\, \tilde{x}> - \frac{1}{2}\,(\|\tilde{s}_1\|^2 - \|\tilde{s}_0\|^2)} \;\begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array}\; \eta$$

Alternative implementation involving "minimum distance" decoder for $\tilde{s}_0$, $\tilde{s}_1$

$$\Lambda(x) = e^{\frac{1}{2}\|\tilde{x} - \tilde{s}_0\| - \frac{1}{2}\|\tilde{x} - \tilde{s}_1\|} \;\begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array}\; \eta$$

PERFORMANCE

As test statistic $T(x)$ is Gaussian

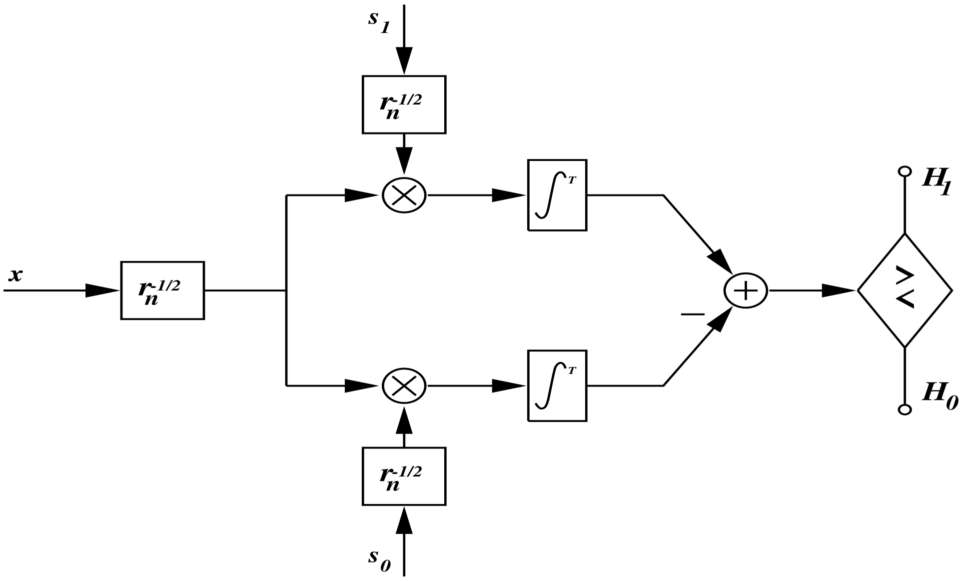


Figure 202: *Prewhitening filter correlator-detector implementation of Fig. for non-random continuous time Gaussian signals* $s_1(t)$ *and* $s_0(t)$.

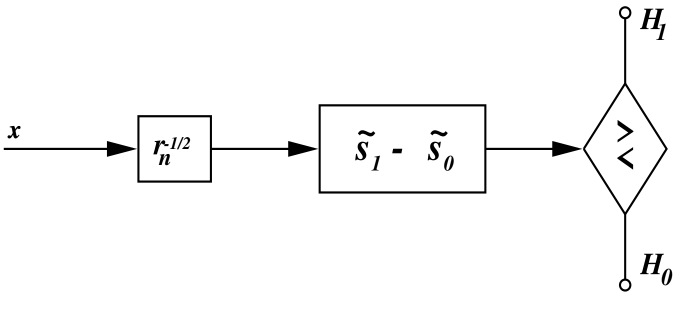


Figure 203: *Prewhitening filter matched filter detector implementation of Fig. for non-random continuous time Gaussian signals* $s_1(t)$ *and* $s_0(t)$. *Here* $\tilde{s}_1 - \tilde{s}_0$ *denotes a filter* $h(t) = \tilde{s}_1(T-t) - \tilde{s}_0(T-t)$ *where* $\tilde{s}_i(t) = \int^T s_i(\tau - t) r_n^{-\frac{1}{2}}(t,\tau) d\tau$, $i = 1, 0$

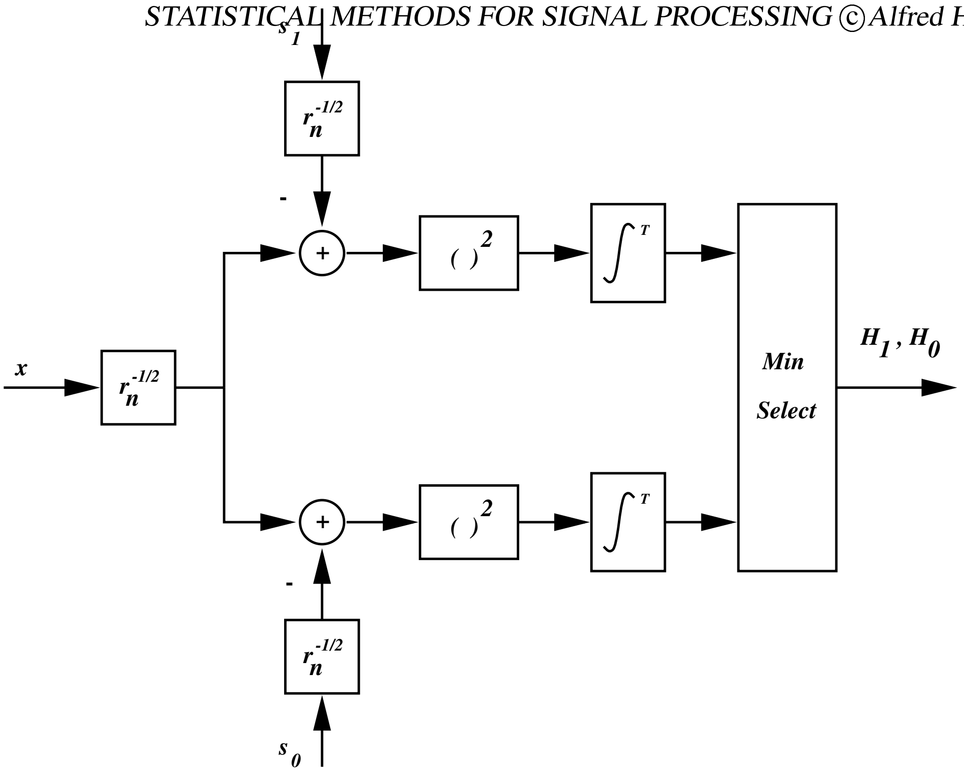


Figure 204: *Prewhitening filter minimum distance detector implementation of Fig. for non-random continuous time Gaussian signals* $s_1(t)$ *and* $s_0(t)$.

$\alpha = P_F$ is

$$\alpha = P(T(x) > \gamma | H_0) = 1 - \mathcal{N}\left(\frac{\gamma - E[T|H_0]}{\sqrt{\text{var}(T|H_0)}}\right)$$

and $\beta = P_D$ is

$$\beta \;\; = \;\; P(T(x) > \gamma | H_1) = 1 - \mathcal{N}\left(\frac{\gamma - E[T|H_1]}{\sqrt{\text{var}(T|H_1)}}\right)$$

$$\beta \;\; = \;\; 1 - \mathcal{N}\left(\frac{\sqrt{\text{var}(T|H_0)}}{\sqrt{\text{var}(T|H_1)}}\mathcal{N}^{-1}(1-\alpha) - \frac{E[T|H_1] - E[T|H_0]}{\sqrt{\text{var}(T|H_1)}}\right)$$

Here we easily compute:

$$E[T|H_0] \;\; = \;\; E[< \tilde{s}_1 - \tilde{s}_0, \tilde{x} > |H_0] =< \tilde{s}_1 - \tilde{s}_0, \tilde{s}_0 >$$

$$
\begin{aligned}
E[T|H_1] &= E[<\tilde{s}_1 - \tilde{s}_0, \tilde{x}> |H_1] = <\tilde{s}_1 - \tilde{s}_0, \tilde{s}_1> \\
\mathrm{var}(T|H_0) &= \mathrm{var}(T|H_1) = \|\tilde{s}_1 - \tilde{s}_0\|^2
\end{aligned}
$$

Hence

$$
\begin{aligned}
\frac{E[T|H_1] - E[T|H_0]}{\sqrt{\mathrm{var}(T|H_1)}} &= \frac{\overbrace{<\tilde{s}_1 - \tilde{s}_0, \tilde{s}_1> - <\tilde{s}_1 - \tilde{s}_0, \tilde{s}_0>}^{\|\tilde{s}_1 - \tilde{s}_0\|^2}}{\|\tilde{s}_1 - \tilde{s}_0\|} \\
&= \|\tilde{s}_1 - \tilde{s}_0\|
\end{aligned}
$$

we therefore have power curve

$$
\beta = 1 - \mathcal{N}\left(\mathcal{N}^{-1}(1-\alpha) - \underbrace{\|\tilde{s}_1 - \tilde{s}_0\|}_{d}\right)
$$

Interpretations of detectibility index d

* Assume white noise $r_n(t,\tau) = \frac{N_o}{2}\delta(t-\tau)$

$$
\begin{aligned}
\Rightarrow \ \tilde{s}_k &= \sqrt{\frac{2}{N_o}}\, s_k(t), \quad k = 0, 1 \\
\Rightarrow \ d^2 &= \frac{2\|s_1 - s_0\|^2}{N_o} = SDNR
\end{aligned}
$$

(SDNR = signal-difference to noise ratio)

ORTHOGONAL VS ANTIPODAL SIGNALS

$$
\begin{aligned}
d^2 &= \|\tilde{s}_1 - \tilde{s}_0\|^2 \\
&= \|\tilde{s}_1\|^2 + \|\tilde{s}_0\|^2 - 2 <\tilde{s}_0, \tilde{s}_1> \\
&= \|\tilde{s}_1\|^2 + \|\tilde{s}_0\|^2 - 2\rho_{01}\, \|\tilde{s}_1\|\, \|\tilde{s}_2\|
\end{aligned}
$$

where ρ_{01} is the signal correlation coefficient

$$
\rho_{01} = \frac{<\tilde{s}_0, \tilde{s}_1>}{\|\tilde{s}_0\|\, \|\tilde{s}_1\|}
$$

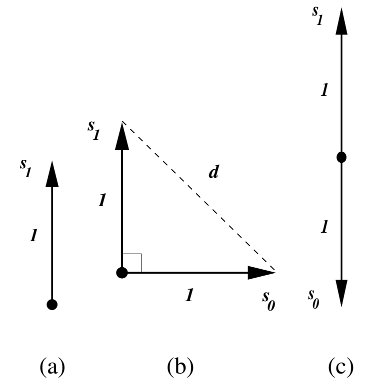


Figure 205: Antipodal signaling gives better detectibility index d than orthogonal signaling or on-off-keying signaling. (a) OOK : $d = 1$, (b) orthogonal : $d = \sqrt{2}$, (c) antipodal : $d = 2$.

Let $\|\tilde{s}_0\|^2 = \|\tilde{s}_1\|^2 = 2E/N_o$

Case 1: For orthogonal signals $\rho_{01} = 0$ and

$$d^2_{orth} = \|\tilde{s}_1\|^2 + \|\tilde{s}_1\|^2 = 2\|\tilde{s}_1\|^2 = 4E/N_o$$

Case 2: For antipodal signals $\rho_{01} = -1$

$$d^2_{anti} = \|\tilde{s}_1 + \tilde{s}_1\|^2 = 4\|\tilde{s}_1\|^2 = 8E/N_o$$

CONCLUDE: 3dB advantage to antipodal signaling as compared with orthogonal signaling

Example 47 *Radar detection at range bin τ:*

Observe $x(t), \quad t \in [0, T]$. Decide between

$$\begin{aligned} H_0 : x(t) &= w(t) \\ H_1 : x(t) &= s(t-\tau) + w(t) \end{aligned}$$

where $w(t)$ is Gaussian white noise

Identify

* $s_0 = 0, \qquad s_1 = s(t-\tau)$

LRT is therefore a correlator

$$T(x) = \frac{2}{N_o} < s(t-\tau), x(t) > \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma$$

or, equivalently a matched filter

$$T(x) = \frac{2}{N_o} s(-t) * x(t)\Big|_{t=\tau} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma$$

Example 48 *Coherent detection of OOK signals*

$$\begin{aligned} H_0 : x(t) &= w(t) \\ H_1 : x(t) &= A\cos(\omega_c t + \psi) + w(t) \end{aligned}$$

* $A > 0, \omega_c, \psi \in [0, 2\pi]$ are known

* $T \gg 2\pi/\omega_c$

* $w(t)$ Gaussian white noise

LRT is

$$T(x) = \frac{2A}{N_o} < \cos(\omega_c t + \psi),\ x(t) > \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma$$

or, as $A > 0$, equivalently

$$< \cos(\omega_c t + \psi),\ x(t) >= \int_0^T \cos(\omega_c t + \psi)\ x(t) dt \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma^{'}$$

PERFORMANCE

$$\begin{aligned} d^2 &= \frac{2\|s_1 - s_0\|^2}{N_o} \\ &= \frac{2}{N_o} A^2 \underbrace{\int_0^T \cos^2(\omega_c t + \psi) dt}_{T/2} \\ &= \frac{TA^2}{N_o} = \frac{2E}{N_o} \end{aligned}$$

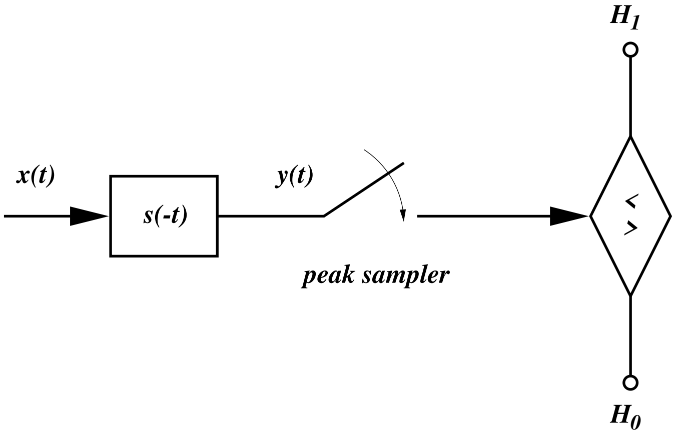


Figure 206: *For unknown τ, GLRT is MF peak detector if* $\|s(t-\tau)\| = \|s(t)\|, \tau \in [0, T]$ *(NEW)*

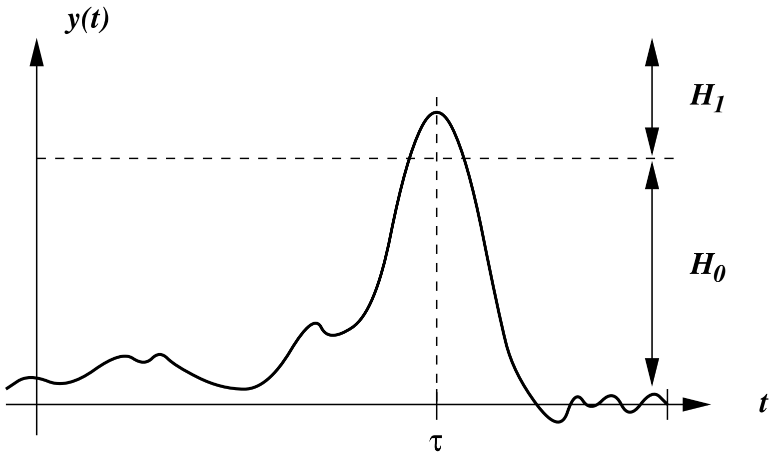


Figure 207: *Example of a matched filter output* $y(\tau) = \int_0^T s(t'-\tau)x(t')dt'$ *with a detection threshold.*

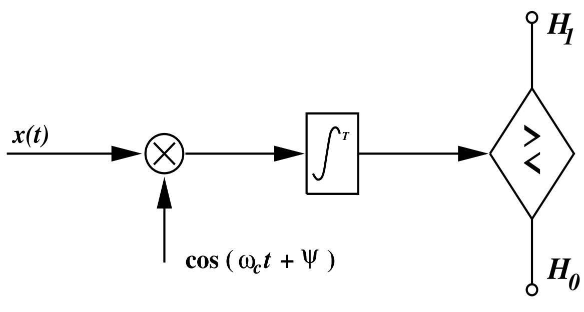


Figure 208: *Coherent detector of AM-OOK signal*

Example 49 *Coherent detection of PSK signals*

$$
\begin{aligned}
H_0 : x(t) &= A\cos(\omega_c t + \psi_0) + w(t) \\
H_1 : x(t) &= A\cos(\omega_c t + \psi_1) + w(t)
\end{aligned}
$$

* $A > 0, \omega_c, \psi_0, \psi_1 \in [0, 2\pi]$ are known

* $T \gg 2\pi/\omega_c$

* $w(t)$ Gaussian white noise

LRT is

$$
< \cos(\omega_c t + \psi_1) - \cos(\omega_c t + \psi_0),\ x(t) > \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma
$$

Know that best performance occurs when $\psi_1 = \psi_0 + \pi$ (antipodal)

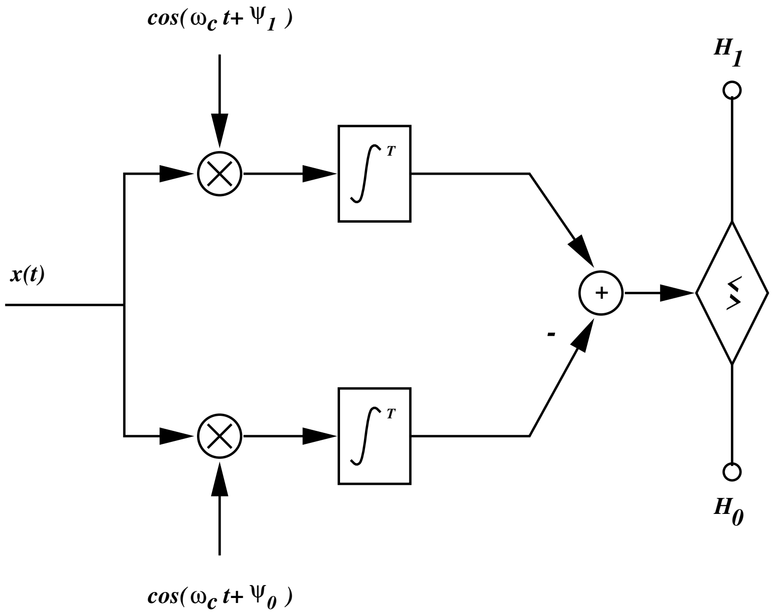


Figure 209: *Coherent detector of PSK signal*

PERFORMANCE

Assume antipodal case:

$$
d^2 = \frac{2\|s_1 - s_0\|^2}{N_o}
$$

$$= \frac{2}{N_o} 4A^2 \underbrace{\|\cos(\omega_c t + \psi_0)\|^2}_{T/2}$$

$$= \frac{4TA^2}{N_o}$$

* 6dB improvement in d^2 w.r.t. OOK

Example 50 *Coherent classification of binary FSK*

$$H_0 : x(t) = A\cos(\omega_0 t + \psi_0) + w(t)$$

$$H_1 : x(t) = A\cos(\omega_1 t + \psi_1) + w(t)$$

* $A > 0$, $\omega_k, \psi_k \in [0, 2\pi]$ are known, $k = 1, 0$

* $w(t)$ Gaussian white noise

LRT is

$$< \cos(\omega_1 t + \psi_1) - \cos(\omega_0 t + \psi_0),\ x(t) > \ \overset{H_1}{\underset{H_0}{\gtrless}} \ \gamma$$

PERFORMANCE

$$d^2 = \frac{2\|s_1 - s_0\|^2}{N_o}$$

$$= \frac{2}{N_o} A^2 \underbrace{\|\cos(\omega_1 t + \psi_1) - \cos(\omega_0 t + \psi_0)\|^2}_{T(1-\rho_{01})}$$

where ρ_{01} is the "ambiguity function" (see discussion below)

$$\rho_{01} = \frac{2}{T}\int_0^T \cos(\omega_0 t + \psi)\cos(\omega_1 t + \psi)dt$$

$$= \cos(T\Delta\omega/2 + \Delta\psi)\,\frac{\sin(T\Delta\omega/2)}{T\Delta\omega/2}$$

and

$$\Delta\omega = \omega_0 - \omega_1, \qquad \Delta\psi = \psi_0 - \psi_1$$

* FSK is equivalent to orthogonal signaling if, for n integer

$$\Delta\omega = 2\pi n/T$$

* FSK performance generally lies between OOK and PSK

* $\rho_{01} = \rho(\Delta\omega, \Delta\psi)$ is an example of an "ambiguity function" (see below)

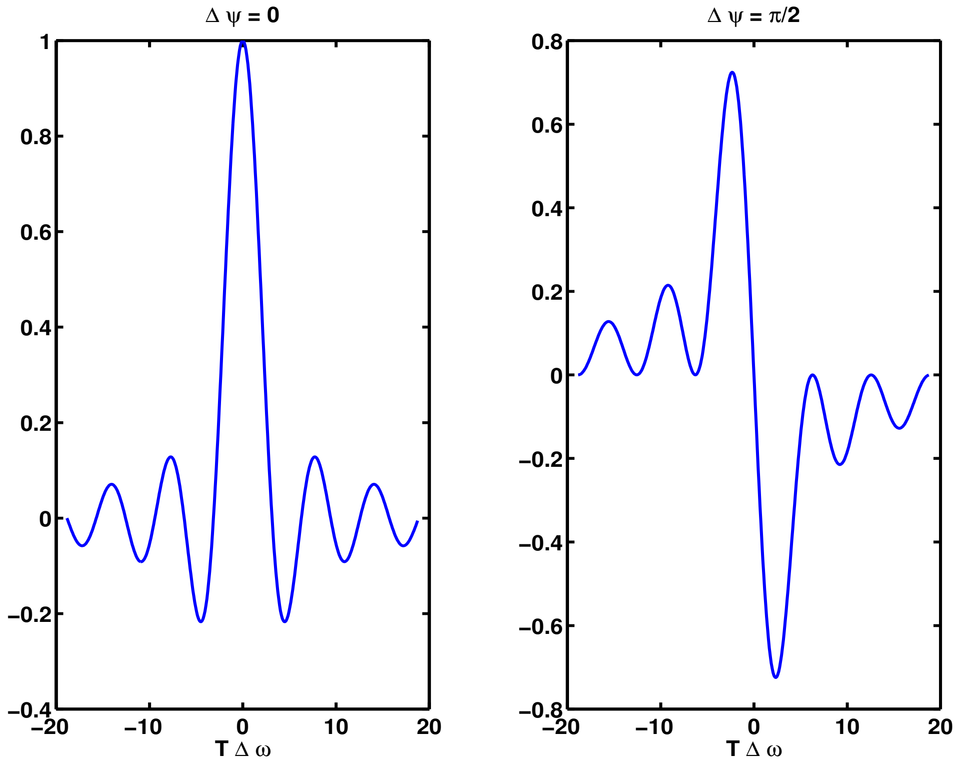


Figure 210: *Plot of ambiguity function $\rho_{01} = \rho(\Delta\omega, \Delta\psi)$ as a function of $T\Delta\omega$ for coherent FSK example. Note that ambiguity function is sensitive to unknown phase.*

14.4 M-ARY CLASSIFICATION OF NON-RANDOM SIGNALS

$$\begin{aligned} H_1 : x(t) &= s_1(t) + w(t) \\ \vdots \quad & \vdots \quad \vdots \\ H_M : x(t) &= s_M(t) + w(t) \end{aligned}$$

* $s_1, \ldots, s_M$ are known signals

* $w(t)$ is Gaussian white noise

* Priors $P(H_1), \ldots, P(H_M)$ are known

We know that minimum P_e is obtained by using MAP decision rule

$$\begin{aligned}\hat{H}_i &= \text{argmax}_{H_i}\{P(H_i|x)\} \\ &= \text{argmax}_{H_i}\left\{\frac{f(x|H_i)P(H_i)}{f(x)}\right\}\end{aligned}$$

Simplification: reduce problem to LRT statistic already derived

$\Rightarrow$ multiply posterior $P(H_i|x)$ by $f(x)/f(x|H_0)$

* $f(x|H_0)$ is p.d.f. under artificial "noise alone" hypothesis

$$H_o : x(t) = w(t)$$

Then

$$\begin{aligned}\hat{H}_i &= \text{argmax}_{H_i}\left\{\frac{f(x|H_i)P(H_i)}{f(x)}\;\frac{f(x)}{f(x|H_0)}\right\} \\ &= \text{argmax}_{H_i}\left\{\underbrace{\frac{f(x|H_i)}{f(x|H_0)}}_{\Lambda(x)}\;P(H_i)\right\} \\ &= \text{argmax}_{H_i}\left\{\exp\left(\frac{2}{N_o}\;< s_i, x > -\frac{1}{N_o}\;\|s_i\|^2\right)\;P(H_i)\right\}\end{aligned}$$

Or equivalently, we can express this in terms of distance between x, s_i

$$\hat{H}_i = \text{argmax}_{H_i}\left\{\exp\left(-\frac{1}{N_o}\;\|s_i - x\|^2\right)\;P(H_i)\right\}$$

For equally likely H_i we have two equivalent implementations

Implementation 1: maximum correlation with energy offset

$$\text{argmax}_{H_i}\left\{< s_i, x > -\tfrac{1}{2}\;\|s_i\|^2\right\}$$

Implementation 2: minimum distance decision rule

$$\hat{H}_i = \text{argmin}_{H_i} \|s_i - x\|$$

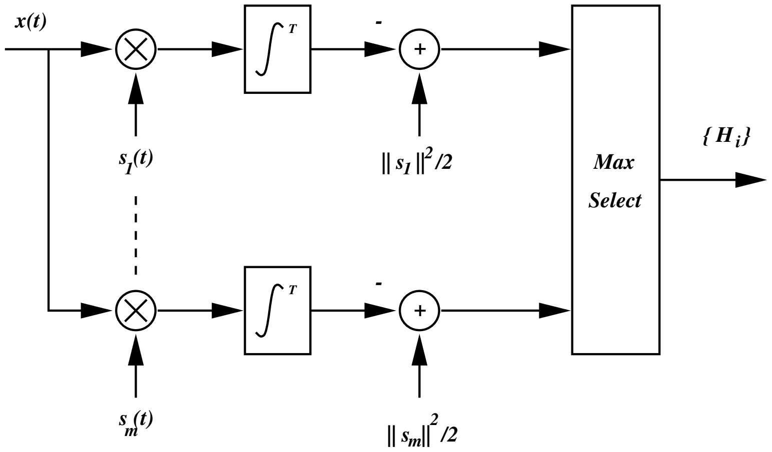


Figure 211: *Implementation of optimal classifier as correlation receiver*

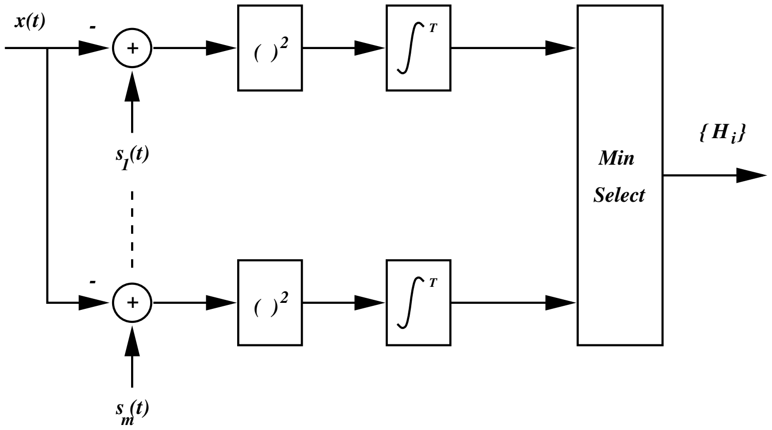


Figure 212: *Implementation of optimal classifier as minimum distance receiver*

PROBABILITY OF ERROR (Equally likely case)

$$P_e \;=\; 1 - \sum_{i=1}^{M} P(\text{say } H_i | H_i) P(H_i)$$

NOTE: correlator decision statistics $\{< s_i, x >\}_i$ are jointly Gaussian

* $E[< s_i, x > | H_j] =< s_i, s_j >$

* var$(< s_i, x > | H_j) = \frac{2\|s_i\|^2}{N_o} = \sigma_{ij}^2$, (indep. of j)

Hence

$$P(\text{say } H_i|H_i) =$$

$$= P\left(\underbrace{< s_i, x > -\tfrac{1}{2}\,\|s_i\|^2}_{T_i} \;\geq\; \underbrace{\max_{j\neq i}\{< s_j, x > -\tfrac{1}{2}\,\|s_j\|^2\}}_{T_{j\neq i}^{max}} \middle| H_i\right)$$

$$= \int_{-\infty}^{\infty} P(T_{j\neq i}^{max} \;\leq\; T_i|H_i, T_i)\; \frac{1}{\sqrt{2\pi\sigma_{ii}^2}} e^{-\frac{1}{2\sigma_{ii}^2}(T_i - \frac{1}{2}\|s_i\|^2)^2} dT_i$$

SPECIAL CASE: orthogonal equal energy signals

$\Rightarrow \{T_j\}_j$ are uncorrelated (independent) r.v.s and

$$\int_{-\infty}^{\infty} P(T_{j\neq i}^{max} \;\leq\; T_i|H_i, T_i)\; \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2\sigma_i^2}(T_i - \frac{1}{2}\|s_i\|^2)^2} dT_i$$

$$= \int_{-\infty}^{\infty} \prod_{j\neq i}\left[\mathcal{N}\left(\frac{T_i - < s_i, s_j > + \frac{1}{2}\|s\|^2}{\sigma}\right)\right]^{M-1} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(T_i - \frac{1}{2}\|s\|^2)^2} dT_i$$

Or

$$P(\text{say } H_i|H_i) \;=\; \int_{-\infty}^{\infty} \prod_{j\neq i}\left[\mathcal{N}\left(T_i + \frac{\|s_i - s_j\|^2}{2\sigma}\right)\right]^{M-1} \;\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}T_i^2} dT_i$$

REMARKS:

1. P_e is decreasing in the SDNR $\|s_i - s_j\|^2/(2\sigma)$
2. P_e inversely proportional to distances $\|s_i - s_j\|$
3. Worst case probability of error dominated by minimum distance

Example 51 *Coherent detection of Q-ary FSK*

$$H_1 : x(t) \;=\; \underbrace{A\cos(\omega_1 t + \psi_1)}_{s_1(t)} + w(t)$$

$$
\begin{array}{rcl}
\vdots & \vdots & \vdots \\
H_M : x(t) & = & \underbrace{A\cos(\omega_M t + \psi_M)}_{s_M(t)} + w(t)
\end{array}
$$

* $\omega_1, \ldots, \omega_M$ are known, equally likely

* $A > 0$ and ϑ_i's are known

* $T \gg 2\pi/\omega_k$, $k = 1, \ldots, M$

* $w(t)$ Gaussian white noise

Min-P_e receiver is

$$
\begin{aligned}
\hat{H}_i &= \text{argmax}_{H_i}\{P(H_i|x)\} \\
&= \text{argmax}_{H_i}\{f(x|H_i)\} \\
&= \text{argmax}_{H_i}\{\|x - s_i\|^2\}
\end{aligned}
$$

For FSK all signals have equal energy

$$
\|s_i\|^2 = A^2T/2
$$

Therefore, optimal receiver is

$$
\begin{aligned}
\hat{H}_i &= \text{argmax}_{H_i}\{< s_i, x >\} \\
&= \text{argmax}_{H_i}\left\{\int_0^T \cos(\omega_i t + \psi_i)x(t)dt\right\}
\end{aligned}
$$

14.5 DETECTION AND ESTIMATION OF SIGNALS WITH UNKNOWN PARAMETERS

$$
\begin{aligned}
H_0 : x(t) &= s(t, \theta_0) + w(t) \\
H_1 : x(t) &= s(t, \theta_1) + w(t)
\end{aligned}
$$

* θ_0, θ_1 are unknown

* s is a known function

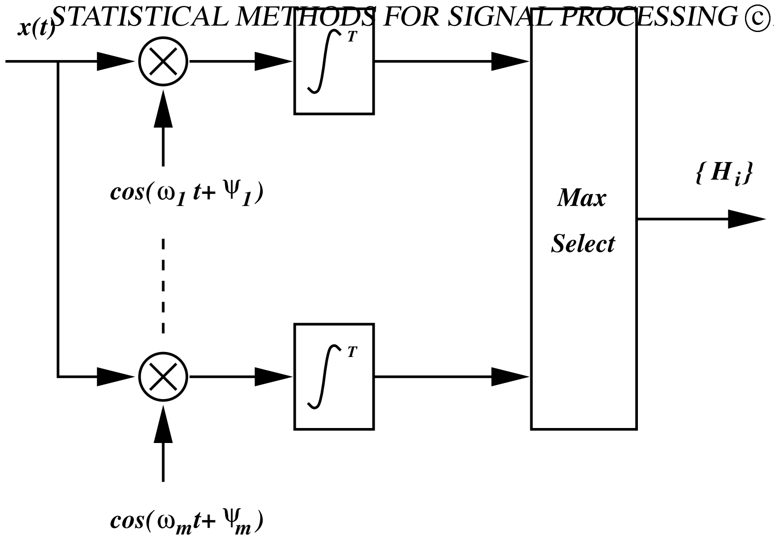


Figure 213: *Optimal coherent receiver for Q-ary FSK*

* w is Gaussian white noise

Approaches:

1. First check for UMP test w.r.t. some of the unknowns
2. Find GLRT
3. Assign p.d.f. to θ and find LRT

14.5.1 DETECTION PERFORMANCE AND THE AMBIGUITY FUNCTION

For unknown θ_0, θ_1 the power surface of any detector of level α is upper bounded by the "clairvoyant" power surface which governs the case of perfectly known θ_0, θ_1.

$$\beta(\theta_0, \theta_1) = 1 - \mathcal{N}\left(\mathcal{N}^{-1}(1-\alpha) - d(\theta_0, \theta_1)\right),$$

* $\theta_0 \in \Theta_0, \; \theta_1 \in \Theta_1$

where

$$\begin{aligned} d^2(\theta_0, \theta_1) &= \frac{2}{N_0} \|s(\theta_1) - s(\theta_0)\|^2 \\ &= \frac{2}{N_0} \left(\|s(\theta_1)\|^2 + \|s(\theta_0)\|^2 - 2 < s(\theta_0), \; s(\theta_1) >\right) \end{aligned}$$

ASSUMPTION: θ_0 and θ_1 do not affect the energy of s (e.g. delays, phase and frequency shifts)

IMPLICATION: the upper bound $\{\beta(\theta_0, \theta_1)\}_{\theta_0,\theta_1}$ is decreasing in the the signal correlation

$$< s(\theta_0),\ s(\theta_1) >= \int_0^T s(t, \theta_0)\ s(t, \theta_1) dt$$

The correlation coefficient is more useful for evaluating performance of a specific signal design s

$$\rho(\theta_0, \theta_1) \quad = \quad \frac{< s(\theta_1),\ s(\theta_0) >}{\|s(\theta_1)\|\ \|s(\theta_0)\|}$$

This is called the **ambiguity function** or **ambiguity surface**

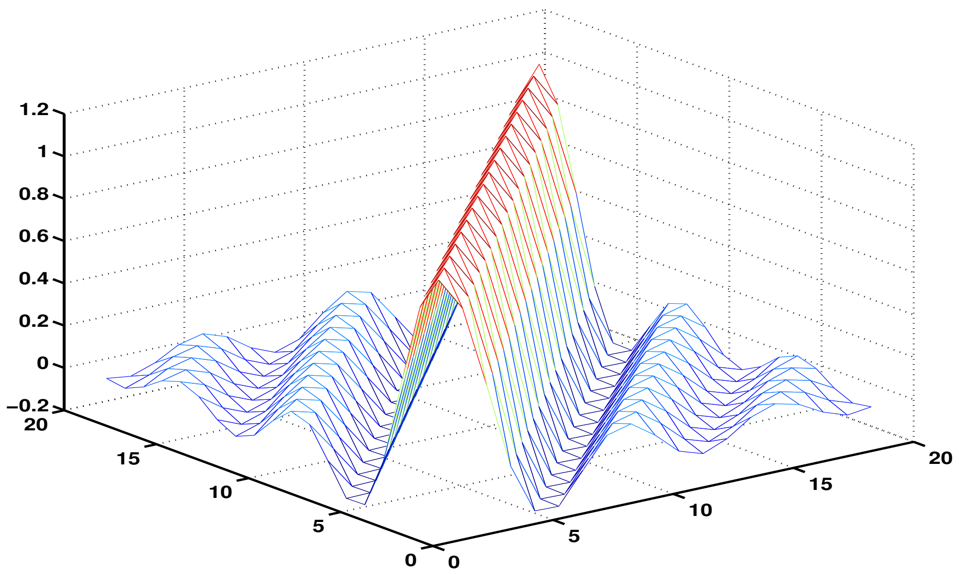


Figure 214: *Prototypical ambiguity surface*

Properties of ambiguity function:

1. $-1 \leq \rho(\theta_0, \theta_1) \leq 1$

2. $\rho(\theta_0, \theta_1) = 1$ is achieved when $\theta_0 = \theta_1$ $(d = 0)$

3. $\rho(\theta_0, \theta_1) = -1$ is achieved when $s(\theta_1)$ and $s(\theta_0)$ are orthogonal.

4. Curvature of $\rho(\theta_0, \theta_1)$ is a measure of resolvability of θ_0 and θ_1 (see below)

5. Curvature of $\rho(\theta_0, \theta_1)$ at $\theta_0 = \theta_1$ determines ability of an optimal detector to discriminate between "close signals"

14.5.2 AMBIGUITY AND THE LOG LR

The mean of the log-likelihood ratio under H_1, θ_1, is

$$
\begin{aligned}
&E_{\theta_1}[\ln \frac{f(x;\theta_1)}{f(x;\theta_0)}] \\
&= E_{\theta_1}\left[\ln\left(e^{\frac{2}{N_o}<s(\theta_1)-s(\theta_0),x>-\frac{1}{N_o}(\|s(\theta_1)\|^2-\|s(\theta_0)\|^2)}\right)\right] \\
&= \frac{2}{N_o} < s(\theta_1)-s(\theta_0), s(\theta_1) > -\frac{1}{N_o}(\|s(\theta_1)\|^2 - \|s(\theta_0)\|^2) \\
&= \frac{1}{N_o}\|s(\theta_1)-s(\theta_0)\|^2 = d^2(\theta_0,\theta_1)/2
\end{aligned}
$$

while the mean under H_0, θ_0, is

$$
E_{\theta_0}[l(\theta_0,\theta_1)] \quad = \quad -\frac{1}{N_o}\|s(\theta_1)-s(\theta_0)\|^2 = -d^2(\theta_0,\theta_1)/2
$$

Conclude:

1. $d^2(\theta_0,\theta_1) = E_{\theta_1}[l(\theta_0,\theta_1)] - E_{\theta_0}[l(\theta_0,\theta_1)]$

2. under the assumption that θ does not affect signal energy $\|s\|^2$, we have

$$
\begin{aligned}
d^2(\theta_0,\theta_1) &= -\frac{2}{N_o}\underbrace{< s(\theta_0),\ s(\theta_1) >}_{\rho(\theta_0,\theta_1)\ \|s\|^2} + \frac{2}{N_o}\|s\|^2 \\
&= \frac{2\|s\|^2}{N_o}(1-\rho(\theta_0,\theta_1))
\end{aligned}
$$

Therefore: ambiguity function describes ability of the LR statistic to discriminate between θ_1 and θ_0.

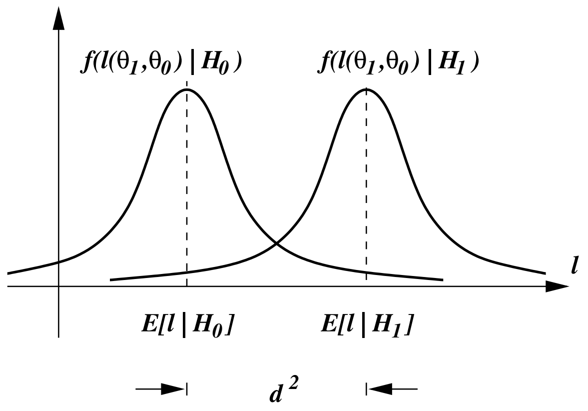


Figure 215: *The distance d between the conditional means of the log-likelihood ratio $l(\theta_1,\theta_0)$ under H_0 and under H_1 is the detectability index providing measure of discriminability between the two hypotheses.*

14.5.3 AMBIGUITY FUNCTION AND CRB

Under H_1 consider the CRB on unbiased estimators of θ_1

$$\text{cov}_{\theta_1}(\hat{\theta}_1) = F^{-1}(\theta_1)$$

$\Rightarrow F(\theta_1)$ is Fisher info matrix

$$\begin{aligned} F(\theta_1) &= E_{\theta_1}\left[-\nabla^2_{\theta_1} \ln f(x;\theta_1)\right] \\ &= E_{\theta_1}\left[-\nabla^2_{\theta_1} \ln \frac{f(x;\theta_1)}{f(x;\theta_0)}\right] \\ &= -\nabla^2_{\theta} E_{\theta_1}\left[\ln \frac{f(x;\theta)}{f(x;\theta_0)}\right]\Bigg|_{\theta=\theta_1} \end{aligned}$$

Now substitute previous "ln LR" expression and use assumption that $\|s(\theta)\|^2$ is indep. of θ

$$-\nabla^2_{\theta} E_{\theta_1}\left[\ln \frac{f(x;\theta)}{f(x;\theta_0)}\right]\Bigg|_{\theta=\theta_1} = -\nabla^2_{\theta}\left[\frac{2}{N_o} < s(\theta)-s(\theta_0), s(\theta_1) > -\frac{1}{N_o}(\|s(\theta)\|^2 - \|s(\theta_0)\|^2)\right]\Bigg|_{\theta=\theta_1}$$

Or

$$F(\theta_1) = -\nabla^2_{\theta} \frac{2}{N_o} < s(\theta), s(\theta_1) > \Big|_{\theta=\theta_1}$$

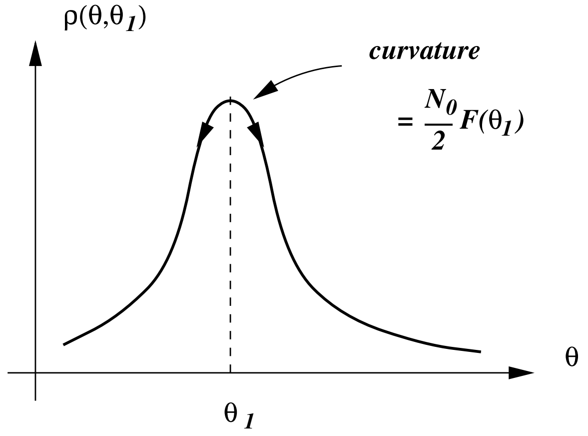


Figure 216: *Curvature of ambiguity function is proportional to Fisher information and therefore gives measure of estimability of the parameter θ.*

Conclude:

1. Fisher info matrix is directly proportional to curvature of ambiguity function $\{\rho(\theta,\theta_1)\}_\theta$ near its peak $\theta = \theta_1$

$$F(\theta_1) = -\nabla^2_\theta \left. \underbrace{\frac{2}{N_o} < s(\theta), s(\theta_1) >}_{\rho(\theta,\theta_1)\cdot 2/\|s\|^2} \right|_{\theta=\theta_1}$$

2. Ambiguity function is a measure of estimability of θ

14.5.4 BANDPASS SIGNALS AND NARROWBAND REPRESENTATIONS

Consider the real valued narrowband bandpass signal:

$$s(t) = Aa(t)\cos(\omega_c t + \psi(t)), \quad t \in [0,T]$$

* $\omega_c >> |d\psi/dt|$

* $A, a(t) \geq 0$

* $a(t)$ has bandwidth $B_a \ll \omega_c$

* $\|a\|^2 = \int_0^T a^2(t)dt = 1$

* $|d\psi(t)/dt| \ll \omega_c$

Under these assumptions:

1. Instantaneous frequency (Hz) of $s(t)$ satisfies

$$\frac{\omega_c + d\psi(t)/dt}{2\pi} \gg B_a$$

2. Energy of $s(t)$ is

$$\|s\|^2 = \int_0^T s^2(t)dt = \frac{A^2}{2}$$

ENVELOPE AND PHASE OF BANDPASS SIGNALS

Rewrite $s(t)$ as real part of complex signal

$$s(t) = \mathrm{Re}\left\{s_a(t)e^{j\omega_c t}\right\}$$

where $s_a(t)$ is the analytic signal of $s(t)$

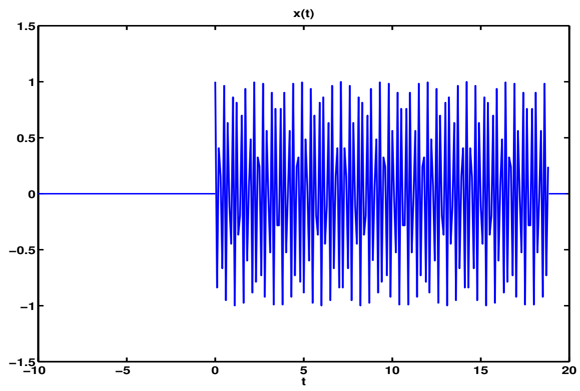


Figure 217: *A burst of a narrowband bandpass signal waveform*

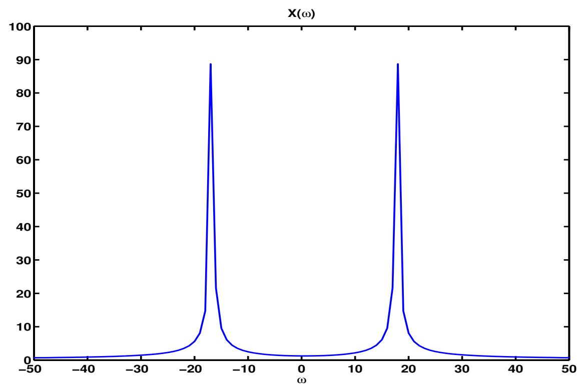


Figure 218: *Spectrum of narrowband bandpass signalwaveform.*

$$s_a(t) = Aa(t)e^{j\psi(t)} = y_c(t) + jy_s(t)$$

* s_a is called the complex "envelope", "baseband signal", or "phasor"

* y_c is "in phase" and y_s is "quadrature" component of complex envelope

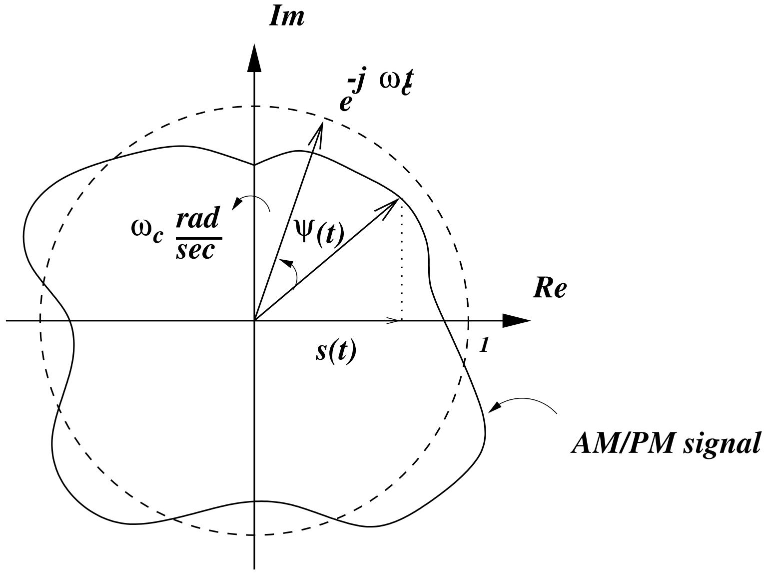


Figure 219: *Complex representation of narrowband signal $s(t)$ as the projection of the complex signal $s_a(t)e^{j\omega_c t}$ onto the real axis where $s_a(t) = Aa(t)e^{j\psi(t)}$ is the complex envelopw.*

RELATIONSHIPS BETWEEN y_c, y_s AND $s(t)$

$$s(t) = y_c(t)\cos(\omega_c t) - y_s(t)\sin(\omega_c t)$$

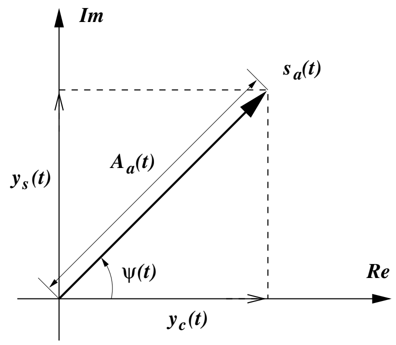
$$\begin{aligned} y_c(t) &= Aa(t)\cos(\psi(t)) \\ y_s(t) &= Aa(t)\sin(\psi(t)) \end{aligned}$$

* Real envelope of $s(t)$ is defined as:

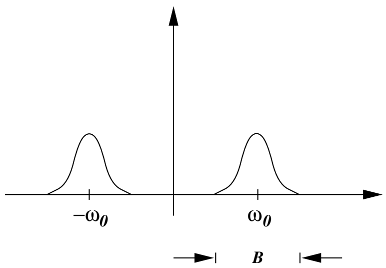
$$|s_a(t)| = Aa(t) = \sqrt{y_c^2 + y_s^2}$$

* phase of $s(t)$ is

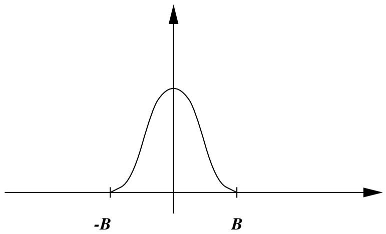
$$\psi(t) = \tan^{-1}[y_s(t)/y_c(t)]$$



(a) Complex envelope s_a of $s(t)$



(b) Spectrum of $s(t)$



(c) Spectra of y_s, y_c

Figure 220: (NEW)

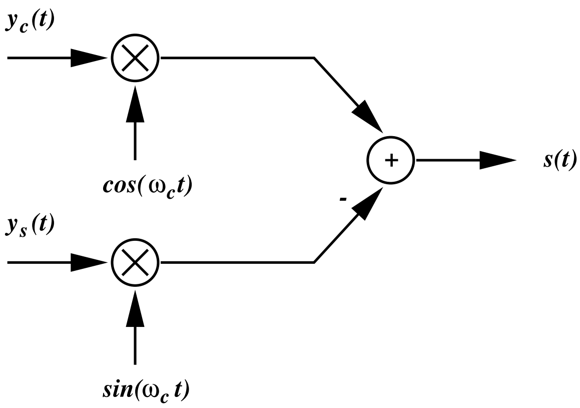


Figure 221: *System to recover* $s(t)$ *from IQ where* $s(t) = Re\{(y_c(t) + jy_s(t))e^{j\omega_c t}\} = y_c(t)cos(\omega_c t) - y_s(t)sin(\omega_c t)$.

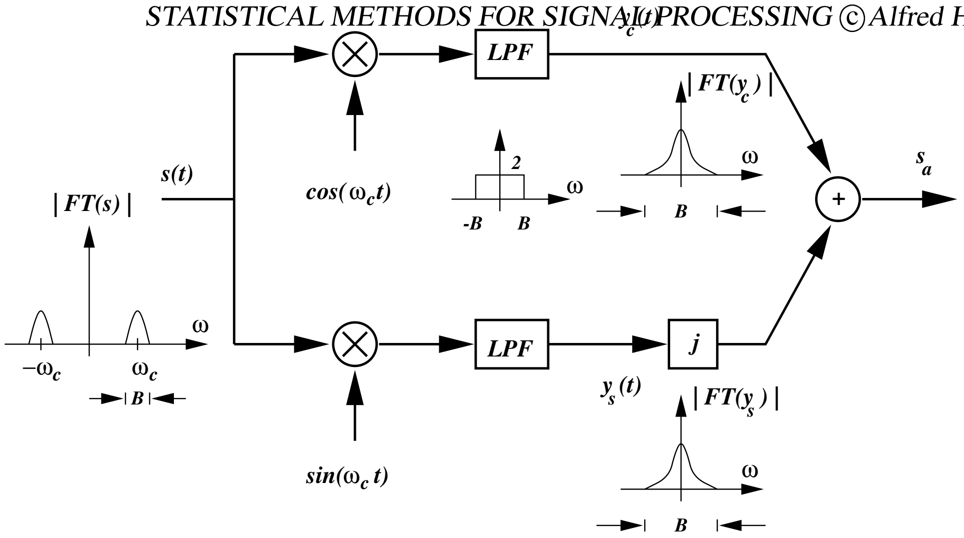


Figure 222: *System to generate IQ components from a signal* $y(t) = s(t)$.

14.5.5 GENERAL COHERENT BANDPASS SIGNAL DETECTION

$$\begin{aligned} H_0 : x(t) &= w(t), \quad t \in [0,T] \\ H_1 : x(t) &= Aa(t)\cos(\omega_c t + \psi(t)) + w(t), \quad t \in [0,T] \end{aligned}$$

* $A > 0$ unknown

* ω_c, $a(t)$, $\psi(t)$ are known

* $\|a\|^2 = 1$

* $T \gg 2\pi/\omega_c$,

* bandwidth of $a(t) \ll \omega_c$

* $|d\psi(t)/dt| \ll \omega_c$

* $w(t)$ Gaussian white noise

We know from before that LRT for known A is

$$T(x) = \frac{2A}{N_o} < a(t)\cos(\omega_c t + \psi(t)),\ x(t) > \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

or, as $A > 0$

$$< a(t)\cos(\omega_c t + \psi(t)),\ x(t) >= \int_0^T a(t)\cos(\omega_c t + \psi(t))\, x(t) dt \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma'$$

Alternatively, in terms of normalized s_a: $\bar{s}_a(t) = a(t)e^{j\psi(t)}$

$$\mathrm{Re}\left\{< \bar{s}_a(t,\theta),\ e^{-j\omega_c t}x(t) >\right\} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma'$$

NOTE:

* LRT is UMP w.r.t. unknown A but *not* w.r.t. unknown $\psi(t)$

Complex correlator representation of LRT

$$\mathrm{Re}\left\{\int_0^T x(t)a(t)e^{j(\omega_c t+\psi(t))}dt\right\} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma'$$

Complex matched filter representation of LRT

$$\mathrm{Re}\left\{h(T) * x(T)\right\} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma'$$

where

$$h(t) = a(T-t)e^{j(\omega_c(T-t)+\psi(T-t))}$$

PERFORMANCE

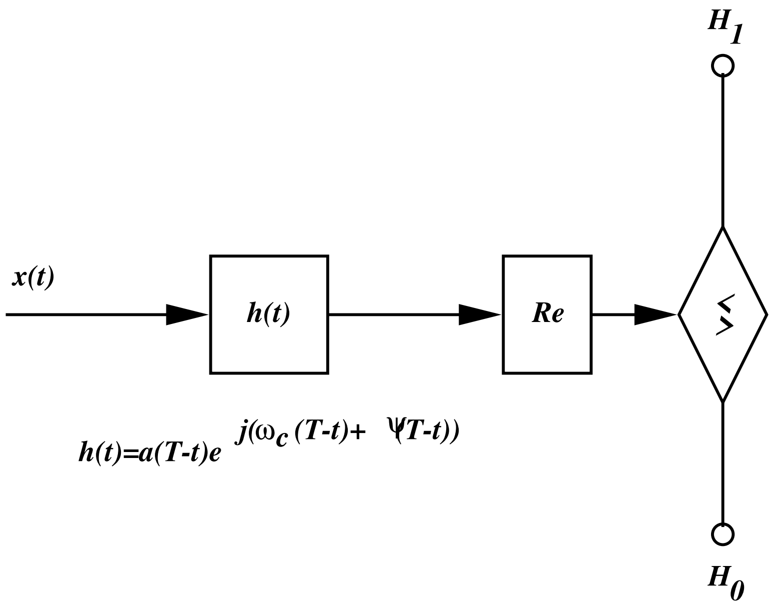
Performance is governed by Gaussian ROC depending only on detectibility index

$$d^2 = \frac{2\|s\|^2}{N_o} = \frac{A^2}{N_o}$$

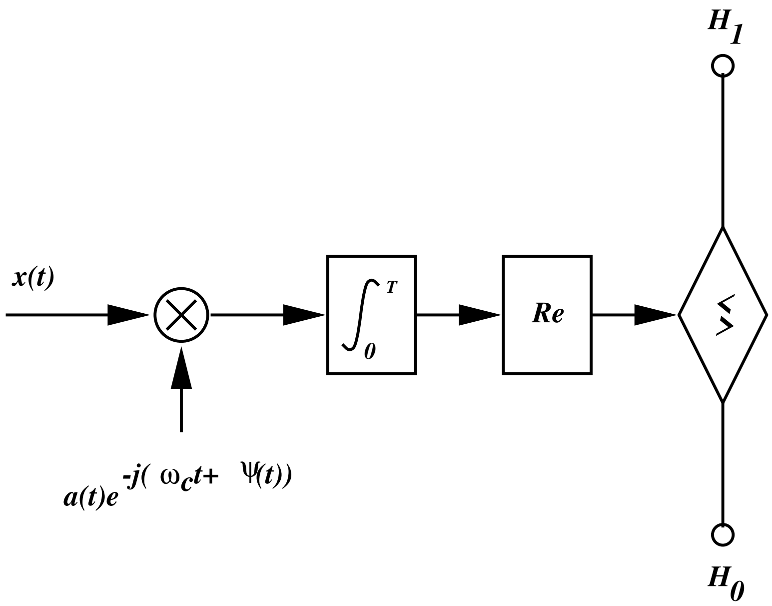
BINARY CLASSIFICATION OF COHERENT NARROWBAND SIGNALS

Pose two hypotheses:

$$\begin{aligned} H_0 : x(t) &= \underbrace{Aa(t,\theta_0)\cos(\omega_c t + \psi(t,\theta_0))}_{s(t,\theta_0)} + w(t) \\ H_1 : x(t) &= \underbrace{Aa(t,\theta_1)\cos(\omega_c t + \psi(t,\theta_1))}_{s(t,\theta_1)} + w(t) \end{aligned}$$



(a)



(b)

Figure 223: (a) Complex matched filter representation of coherent LRT for signal detection. (b) Complex correlator representation of coherent LRT for signal detection.

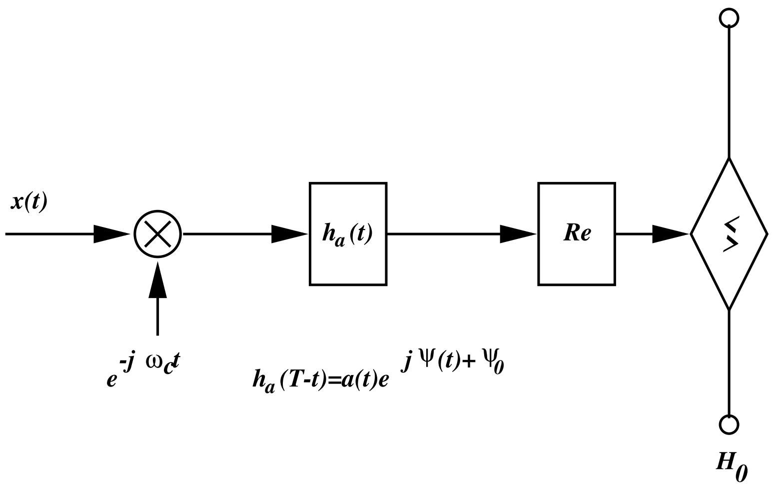


Figure 224: *Heterodyne and matched filter representation of coherent LRT for signal detection.*

Optimal detector is simply

$$\mathrm{Re}\left\{< \bar{s}_a(t,\theta_1) - \bar{s}_a(t,\theta_0),\ e^{-j\omega_c t}x(t) >\right\} \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \gamma'$$

which is UMP wrt A

Complex matched filter representation

$$\mathrm{Re}\left\{h_{10}(T) * x(t)\right\} \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \gamma'$$

$$h_{10}(T-t) = a(t,\theta_1)e^{j(\omega_c t + \psi(t,\theta_1))} - a(t,\theta_0)e^{j(\omega_c t + \psi(t,\theta_0))}$$

Using previous results we know that ambiguity function is

$$\begin{aligned} \rho(\theta_0,\theta_1) &= \frac{< s(\theta_1),\ s(\theta_0) >}{\|s(\theta_1)\|\ \|s(\theta_0)\|} \\ &= 2\ < a(t,\theta_0)\cos(\omega_c t + \psi(t,\theta_0)), a(t,\theta_1)\cos(\omega_c t + \psi(t,\theta_1)) > \\ &= \int_0^T a(t,\theta_0)a(t,\theta_1)\cos\left(\psi(t,\theta_0 - \psi(t,\theta_1)\right) dt \\ &\qquad + \underbrace{\int_0^T a(t,\theta_0)a(t,\theta_1)\cos(2\omega_c t + \vartheta)dt}_{=0} \end{aligned}$$

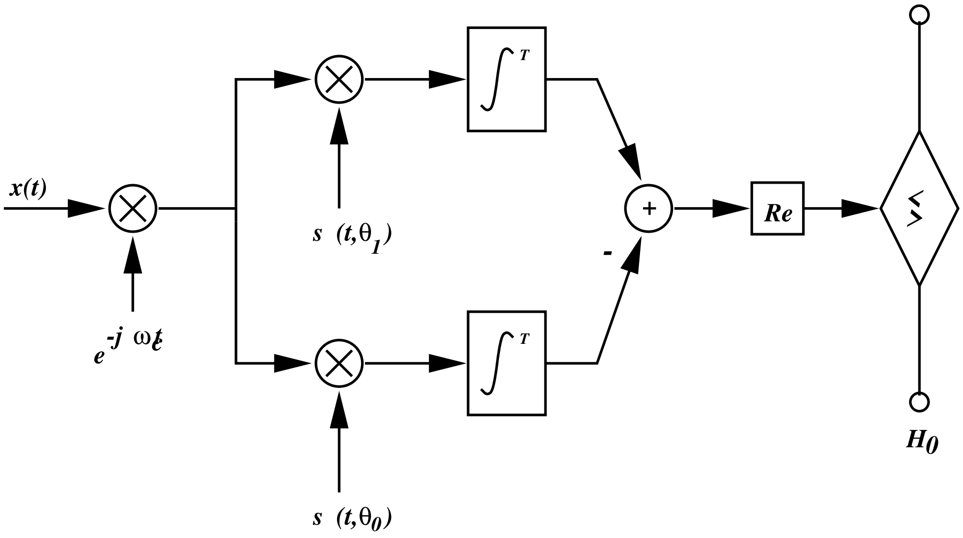


Figure 225: *Correlator implementation of LRT for binary signal classification.*

A convenient representation for coherent ambiguity function in terms of analytic signal $s_a(t,\theta) = Aa(t)e^{j\psi(t,\theta)}$:

$$
\begin{aligned}
\rho(\theta_0,\theta_1) &= \mathrm{Re}\left\{\int_0^T a(t,\theta_0)a(t,\theta_1)e^{\psi(t,\theta_0 - \psi(t,\theta_1)}dt\right\} \\
&= \mathrm{Re}\left\{\int_0^T s_a(t,\theta_0)\ s_a^*(t,\theta_1)dt\right\}\ \frac{1}{A^2} \\
&= \mathrm{Re}\left\{\frac{< s_a(\theta_0),\ s_a(\theta_1) >}{\|s_a(\theta_0)\|\ \|s_a(\theta_1)\|}\right\}
\end{aligned}
$$

14.5.6 DETECTION OF INCOHERENT BANDPASS SIGNALS

$$
\begin{aligned}
H_0 : x(t) &= w(t) \\
H_1 : x(t) &= \underbrace{Aa(t)\cos(\omega_c t + \psi(t) + \psi_o)}_{s_1 = s(t,A,\psi_o)} + w(t)
\end{aligned}
$$

* $A > 0$, $\psi_o \in [0, 2\pi]$ are unknown

* ω_c, $a(t)$, $\psi(t)$ are known

* $T \gg 2\pi/\omega_c$,

* bandwidth of $a(t) \ll \omega_c$

* $|d\psi(t)/dt| \ll \omega_c$

* $w(t)$ Gaussian white noise

APPROACH I. BAYES: Assume ψ_o is uniform over $[0, 2\pi]$

Appropriate LRT is now

$$\Lambda(x) = \frac{1}{2\pi} \int_0^{2\pi} \Lambda(x; \psi) d\psi_o = \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \eta$$

where

$$\Lambda(x; \psi_o) = \frac{f(x|A, \psi_o, H_1)}{f(x|H_0)}$$

is the old LRT statistic (for known ψ_o)

Thus we have:

$$\begin{aligned} \Lambda(x) &= \frac{1}{2\pi} \int_0^{2\pi} e^{\frac{2}{N_o} <s_1, x> - \frac{1}{N_o}\|s_1\|^2} d\psi_o \\ &= \frac{1}{2\pi} \int_0^{2\pi} e^{\frac{2A}{N_o} <a(t)\cos(\omega_c t + \psi(t)\psi_o), x(t)> - \frac{\|s_1\|^2}{N_o}} d\psi_o \\ &= \frac{e^{-\|s_1\|^2/N_o}}{2\pi} \int_0^{2\pi} e^{\frac{2A}{N_o} <a(t)\cos(\overbrace{\omega_c t + \psi(t)}^{a} + \overbrace{\psi_o}^{b}) x(t)>} d\psi_o \end{aligned}$$

Next, to find expression for integral we use trig identity

$$\cos(a + b) = \cos(a)\cos(b) - \sin(a)\sin(b)$$

to obtain

$$\begin{aligned} &< a(t)\cos(\omega_c t + \psi(t) + \psi_o), x(t) > \\ &= \underbrace{< a(t)\cos(\omega_c t + \psi(t)), x(t) >}_{z_c} \cos(\psi_o) \\ &\quad - \underbrace{< a(t)\sin(\omega_c t + \psi(t)), x(t) >}_{z_s} \sin(\psi_o) \\ &= \sqrt{z_c^2 + z_s^2}\ \cos(\psi_o - \vartheta) \end{aligned}$$

where

* $\vartheta = \tan^{-1}(z_s/z_c)$

Now substitute back into LRT statistic

$$
\begin{aligned}
\Lambda(x) &= \frac{e^{-\|s_1\|^2/N_o}}{2\pi} \int_0^{2\pi} e^{\overbrace{\frac{2A}{N_o}\sqrt{z_c^2+z_s^2}}^{Y} \cos(\psi_o - \vartheta)} d\psi_o \\
&= e^{-\|s_1\|^2/N_o} \underbrace{\frac{1}{2\pi} \int_0^{2\pi} e^{Y \cos(\psi_o + \vartheta)} d\psi_o}_{I_o(Y)} \\
&= e^{-\|s_1\|^2/N_o}\, I_o(Y)
\end{aligned}
$$

* $I_o(Y)$ is modified Bessel function of zero'th order

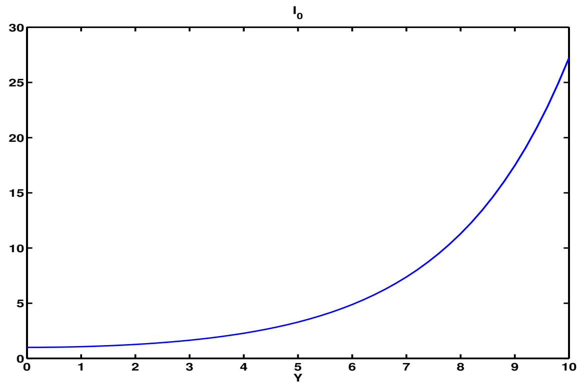


Figure 226: *Bessel function is monotonic*

IMPORTANT FACT: $I_o(Y)$ is monotone increasing function of Y

Hence we have equivalent LRT

$$
Y = \frac{2A}{N_o} \sqrt{z_c^2 + z_s^2} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma
$$

Implementation 1: Quadrature receiver

$$
z_c^2 + z_s^2 \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \gamma'
$$

Implementation 2: incoherent matched filter + envelope detector

$$\sqrt{z_c^2 + z_s^2} = |h(T) * x(T)| \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \gamma'$$

where

$$h(t) = a(T-t)e^{j(\omega_c(T-t)+\psi(T-t))}, \quad t \in [0,T]$$

or equivalently, as any constant phase term is immaterial

$$h(t) = a(T-t)e^{-j(\omega_c t-\psi(T-t))}, \quad t \in [0,T]$$

VERIFY:

$$\begin{aligned} |h(T) * x(T)| &= |\mathrm{Re}\{h(T)\} * x(T) + j\mathrm{Im}\{h(T)\} * x(T)| \\ &= \sqrt{(\mathrm{Re}\{h(T)\} * x(T))^2 + (\mathrm{Im}\{h(T)\} * x(T))^2} \\ &= \sqrt{z_c^2 + z_s^2} \end{aligned}$$

where we have used Euler's relation:

$$e^{j(\omega_c(T-t)+\psi(T-t))} = \cos(\omega_c(T-t)+\psi(T-t)) + j\sin(\omega_c(T-t)+\psi(T-t))$$

Implementation 3: heterodyne and matched filter to analytic signal

$$\left| < \bar{s}_a(t,\theta),\ e^{-j\omega_c t}x(t) > \right| \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \gamma'$$

APPROACH II. GLRT: ψ_o is non-random

Appropriate LRT is now

$$\Lambda_{GLR}(x) = \max_{\psi_o} \left\{ \frac{f(x|A,\psi_o,H_1)}{f(x|H_0)} \right\} \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \eta$$

Using the same in-phase and quadrature representation as in Bayes derivation, and again defining

$$Y = \frac{2A}{N_o}\sqrt{z_c^2 + z_s^2}, \quad \vartheta = \tan^{-1}(z_s/z_c)$$

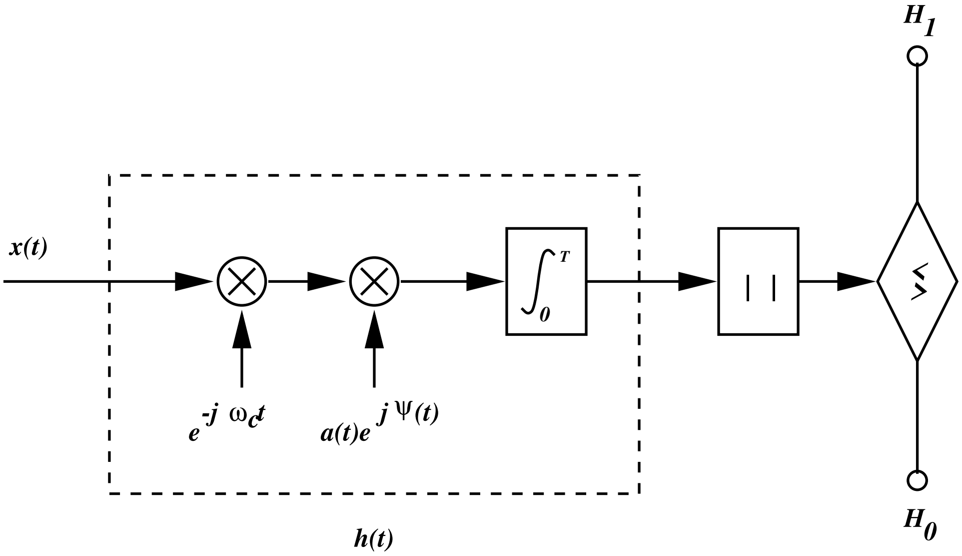


Figure 227: *Incoherent detector via incoherent MF and envelope reciever*

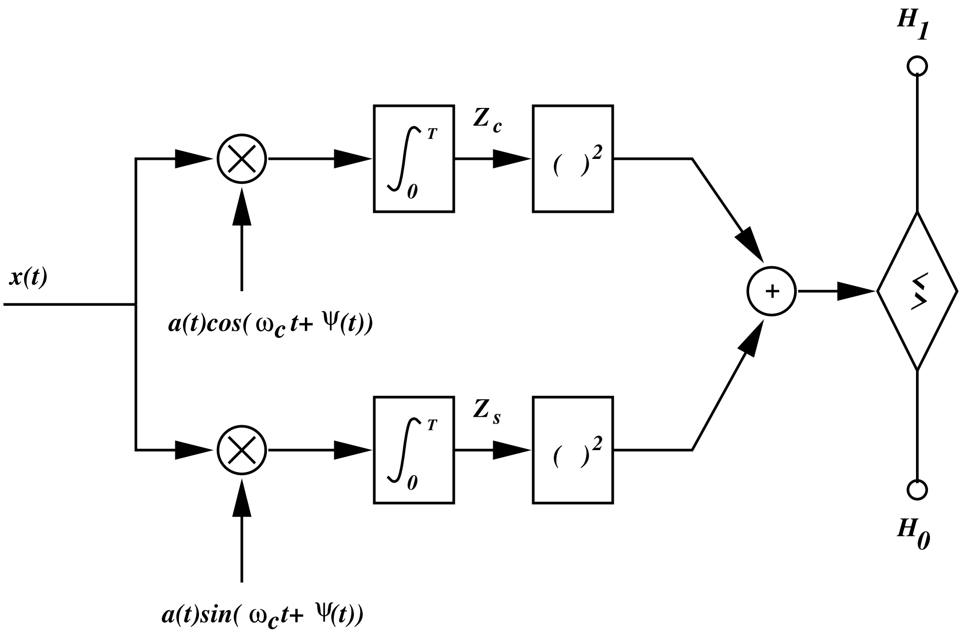


Figure 228: *Incoherent detector via quadrature receiver*

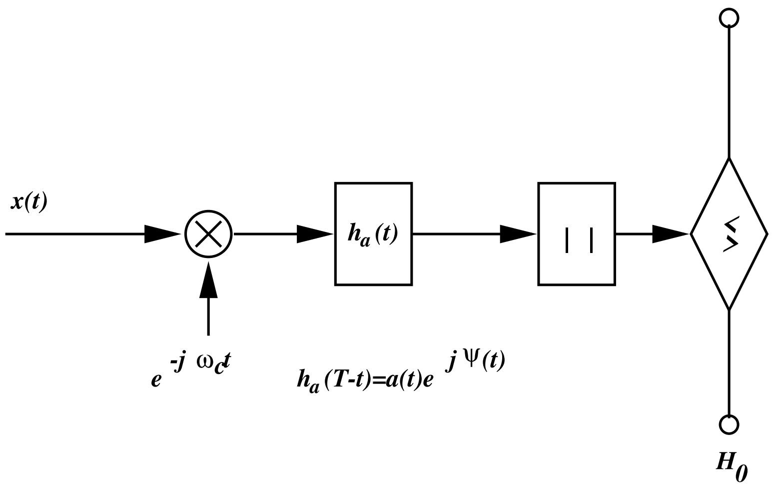


Figure 229: *Incoherent detector via heterodyne, analytic signal MF, and envelope detection*

we have

$$\frac{f(x|A,\psi_o,H_1)}{f(x|H_0)} = e^{-\|s_1\|^2/N_o}\; e^{Y\cos(\psi_o-\vartheta)}$$

OBSERVE: Maximum obviously occurs at $\hat{\psi}_o = \vartheta$ and thus we obtain GLRT

$$\Lambda_{GLR}(x) = e^{-\|s_1\|^2/N_o}\; e^{Y} \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \eta$$

which (except for factor $1/2\pi$) is identical to the LRT for case of random ψ_o!

$$Y \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

PERFORMANCE

Under H_0: z_c, z_s are i.i.d. $\mathcal{N}(0,\sigma_z^2)$

$\Rightarrow (z_c^2+z_s^2)/\sigma_z^2 = \chi_2$ (= Rayleigh)^2

Under H_1: given ψ_o, r.v.s z_c, z_s are indep Gaussian

* $z_c \;\sim\; \mathcal{N}(\mu_c,\; \sigma_z^2)$

* $z_s \;\sim\; \mathcal{N}(\mu_s,\; \sigma_z^2)$

$$\mu_c \quad = \quad < a(t)\cos(\omega_c t + \psi(t)),\; s(t,\theta)) > = \frac{\|s(\theta)\|^2}{A}\; \cos(\psi_o)$$

$$\mu_s \quad = \quad < a(t)\sin(\omega_c t + \psi(t)),\ s(t,\theta) > = \frac{\|s(\theta)\|^2}{A}\ \sin(\psi_o)$$

$\Rightarrow (z_c^2 + z_s^2)/\sigma_z^2 = \chi^2_{2,d^2} = (\text{Rician})^2$

where the common variance of IQ components is

$$\begin{aligned}
\sigma_z^2 &= \text{var}\left(< a(t)\cos(\omega_c t + \psi(t)),\ x(t) >\right) \\
&= \int_0^T dt \int_0^T d\tau a(t)\cos(\omega_c t + \psi(t)) \underbrace{E[x(t)x(\tau)]}_{N_o/2\ \delta(t-\tau)} a(\tau)\cos(\omega_c \tau + \psi(\tau)) \\
&= \frac{N_o}{2} \int_0^T a^2(t)\cos^2(\omega_c t + \psi(t)) dt \\
&= \frac{N_o}{2} \frac{\|s(\theta)\|^2}{A^2}
\end{aligned}$$

and non-centrality parameter is the detectibility

$$d^2 = (\mu_c^2 + \mu_s^2)/\sigma_z^2 = \frac{2\ \|s(\theta)\|^2}{N_o}$$

SIMPLER DERIVATION: USE PROPERTIES OF COMPLEX GAUSSIAN

Consider alternative form of LRT

$$\left|< s_a(t,\theta),\ e^{-j\omega_c t}x(t) >\right| \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma'$$

Under H_0: $z =< s_a(t,\theta),\ e^{-j\omega_c t}x(t) >$ is $\mathcal{CN}(0, \sigma_z^2)$.

$\Rightarrow |z|^2/(\sigma^2/2) = \chi_2$

Under H_1: $z =< s_a(t,\theta),\ e^{-j\omega_c t}x(t) >$ is $\mathcal{CN}(\mu_z, \sigma_z^2)$ where

$$\begin{aligned}
\mu_z &= < s_a(t,\theta), e^{-j\omega_c t} \underbrace{E_1[x(t)]}_{s_a(t,\theta)e^{j\omega_c t}} > \\
&= \|s_a(\theta)\|^2 = 2\|s(\theta)\|^2
\end{aligned}$$

$\Rightarrow |z|^2/(\sigma_z^2/2) = \chi_{2,d^2}$

and under both hypotheses we have common variance

$$\begin{aligned}
\sigma^2/2 &= \tfrac{1}{2}\int_0^T dt \int_0^T d\tau \; s_a(t,\theta)e^{j\omega_c t}\underbrace{\mathrm{cov}(x^*(t),x(\tau))}_{N_o/2\delta(t-\tau)}e^{-j\omega_c\tau}s_a^*(\tau,\theta) \\
&= \frac{N_o\|s_a(\theta)\|^2}{4} = \frac{N_o\|s(\theta)\|^2}{2}
\end{aligned}$$

Explicit form for power curve

$$\beta = Q\left(\sqrt{|d|^2}, \sqrt{1/\ln\alpha}\right)$$

where $Q(a,b)$ is the Marcum-Q function.

BINARY CLASSIFICATION OF INCOHERENT BANDPASS SIGNALS

Under same assumptions on A, $a(t,\theta)$, $\psi(t,\theta)$ as before we pose the following two hypotheses

$$\begin{aligned}
H_0 : x(t) &= \underbrace{Aa(t,\theta_0)\cos(\omega_c t + \psi(t,\theta_0) + \psi_0)}_{s(t,\theta_0)} + w(t) \\
H_1 : x(t) &= \underbrace{Aa(t,\theta_1)\cos(\omega_c t + \psi(t,\theta_1) + \psi_1)}_{s(t,\theta_1)} + w(t)
\end{aligned}$$

* Bayes and GLRT are again identical

GLRT: (non-random ψ)

$$\Lambda_{GLR}(x) = \frac{\max_{\psi_1} f(x|\psi_1,H_1)}{\max_{\psi_0} f(x|\psi_0,H_0)} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \eta$$

Using in-phase and quadrature representations on both $s(t,\theta_0)$ and $s(t,\theta_1)$, straightforward to see that GLRT is

$$\Lambda_{GLR}(x) = \frac{I_o(Y_1)}{I_o(Y_0)} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \eta$$

or equivalently

$$\ln I_o(Y_1) - \ln I_o(Y_0) \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \eta$$

where

$$Y_0 = \frac{2A}{N_o} \left|< \bar{s}_a(t,\theta_0),\ e^{-j\omega_c t} x(t) >\right|$$

$$Y_1 = \frac{2A}{N_o} \left|< \bar{s}_a(t,\theta_1),\ e^{-j\omega_c t} x(t) >\right|$$

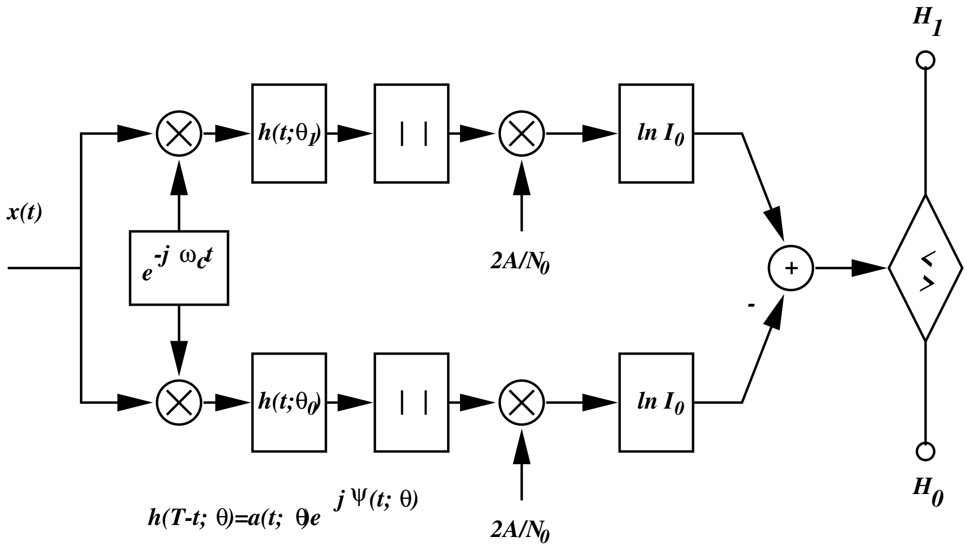


Figure 230: *GLRT for incoherent binary classification of narrowband signals*

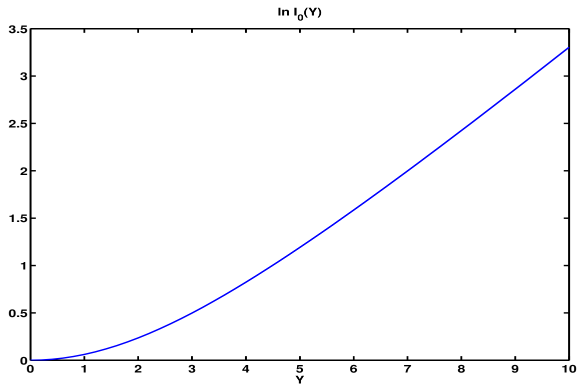


Figure 231: *Non-linearity* $\ln I_o(Y)$ *is approximately quadratic for small* Y *and approximately linear for large* Y.

Performance analysis is difficult and requires asymptotic approximations (see Whalen or Helstrom).

Note: No longer UMP w.r.t. A!

INCOHERENT AMBIGUITY FUNCTION:

Recalling form of ambiguity function in coherent case:

$$\rho(\theta_0, \theta_1) = \mathrm{Re}\left\{\frac{< s_a(\theta_0),\ s_a(\theta_1) >}{\|s_a(\theta_0)\|\ \|s_a(\theta_1)\|}\right\}$$

for incoherent case the appropriate ambuiguity function is the envelope

$$\rho(\theta_0, \theta_1) = \frac{|< s_a(\theta_1),\ s_a(\theta_0) >|}{\|s_a(\theta_1)\|\ \|s_a(\theta_0)\|}$$

14.5.7 INCOHERENT OOK DETECTION WITH RAYLEIGH FADING

$$\begin{aligned} H_0 : x(t) &= w(t) \\ H_1 : x(t) &= \sum_{i=1}^{\#paths} A_i \cos(\omega_c t + \psi_i) + w(t) \end{aligned}$$

Simplification: use **Swerling I model** for signal amplitude and phase

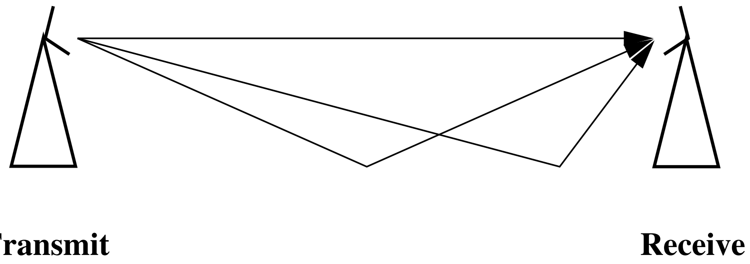


Figure 232: *Slow Rayleigh fading is caused by multipath propagation due to several reflections along transmission path.*

For large # of paths Swerling I model is

$$\begin{aligned} H_0 : x(t) &= w(t) \\ H_1 : x(t) &= A\cos(\omega_c t + \psi_o) + w(t) \end{aligned}$$

* ω_c known

* $T \gg 2\pi/\omega_c$,

* $w(t)$ Gaussian white noise

* $\psi_o \in [0, 2\pi]$ is uniform

* $A > 0$ is Rayleigh random variable with $E[A^2] = A_0^2/2$

$$f(A) = (A/A_o)^2 \exp(-A^2/A_o^2),$$

RECALL: LRT is UMP w.r.t. A

$\Rightarrow$ need only specialize envelope detector to $a(t) = 1$, $\psi(t) = 0$

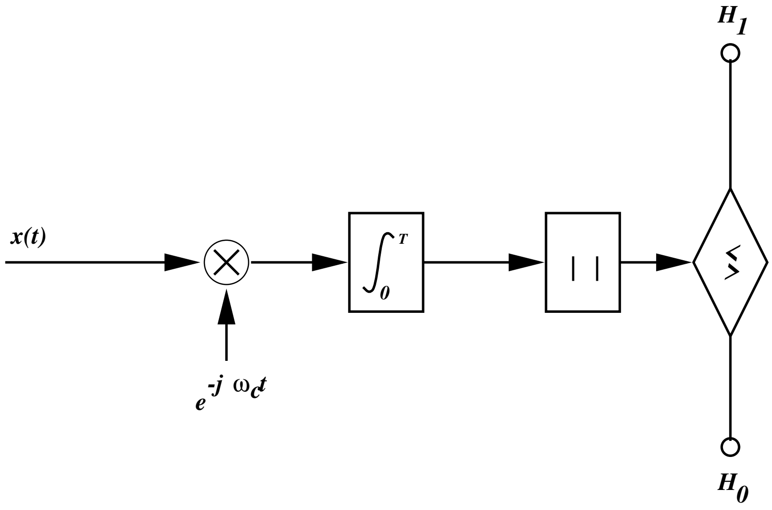


Figure 233: *Optimal incoherent reciever for OOK in Rayleigh fading*

PERFORMANCE

While form of detector remains the same as for non-random A the power curve is different [37]

$$\begin{aligned} \beta &= \int_0^\infty f(A) Q\left(\sqrt{A^2T/N_o}, \sqrt{\ln(1/\alpha)}\right) dA \\ &= \alpha^{(1+A_o^2T/N_o)^{-1}} \end{aligned}$$

14.5.8 INCOHERENT DETECTION OF Q-ary FSK

$$\begin{aligned} H_1 : x(t) &= \underbrace{A\cos(\omega_1 t + \psi_1)}_{s_1(t)} + w(t), \quad t \in [0,T] \\ \vdots \quad &\vdots \quad \vdots \\ H_M : x(t) &= \underbrace{A\cos(\omega_M t + \psi_M)}_{s_M(t)} + w(t), \quad t \in [0,T] \end{aligned}$$

* $\omega_1, \ldots, \omega_M$ are known, equally likely

* $T \gg 2\pi/\omega_k$, $k = 1, \ldots, M$

* $w(t)$ Gaussian white noise

* A and ψ_i's are indep. Rayleigh and uniform r.v.s

For arbitrary fixed A min-P_e receiver is

$$\begin{aligned} \hat{H}_i &= \text{argmax}_{H_i}\{f(H_i|x, A)\} \\ &= \text{argmax}_{H_i}\{f(x|H_i, A)\} \\ &= \text{argmax}_{H_i}\left\{e^{-\|s_i\|^2/N_o}\ I_o(Y_i)\right\} \end{aligned}$$

where

$$Y_i = \frac{2A}{N_o}|h_i(T) * x(T)|$$

and h_i is a bandpass filter with center frequency ω_i

$$h_i(t) = e^{-j\omega_i t}, \quad t \in [0, T]$$

As $I_o(Y_i)$ is monotone and $\|s_i\|^2 = A^2/2$ is constant

Final form of Min-P_e receiver is

$$\begin{aligned} \hat{H}_i &= \text{argmax}_{H_i}\{I_o(Y_i)\} \\ &= \text{argmax}_{H_i}|h_i(T) * x(T)| \end{aligned}$$

Again the decision regions do not depend on A!

PERFORMANCE:

Assuming orthogonal s_i and no Rayleigh fading the P_e can calculated (Proakis [25]).

Rayleigh fading case is somewhat more complicated [37]

INCOHERENT AMBIGUITY FUNCTION

Define: $\omega_c := M^{-1}\sum_{i=1}^{M}\omega_i$

As $s(t, \omega_i) = A\cos(\omega_i t + \psi_i)$,, $t \in [0, T]$ we have analytic signal w.r.t. ω_c

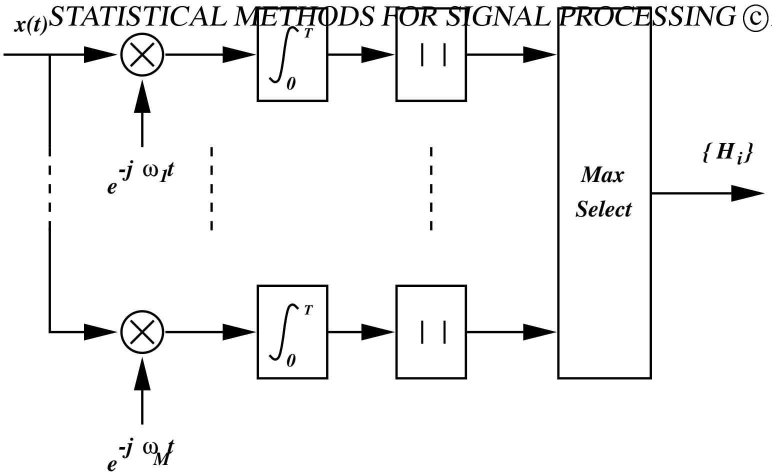


Figure 234: *Incoherent Q-ary FSK detector implemented with a bank of envelope detectors each matched to different frequency*

$$s_a(t, \omega_i) = Ae^{j[(\omega_i - \omega_c)t + \psi_i]}$$

Hence ambiguity function governing resolvability of two frequencies ω_1, ω_0 separated by $\Delta\omega = \omega_0 - \omega_1$ is

$$\begin{aligned}
\rho(\omega_1, \omega_0) &= \frac{|< s_a(\omega_1),\ s_a(\omega_0) >|}{\| < s_a(\omega_1)\| \ \|s_a(\omega_0)\|} \\
&= \frac{1}{A^2 T} \left| \int_0^T \underbrace{s_a(\omega_1) s_a^*(\omega_0)}_{A^2 \exp(j[\Delta\omega t + \Delta\psi])} > dt \right| \\
&= \frac{\sin(\pi \Delta\omega T)}{\Delta\omega T} = \mathrm{sinc}(\Delta\omega T)
\end{aligned}$$

14.5.9 DIVERSITY IMPROVEMENT

Consider receiving FSK signal over N different independent channels $i = 1, \ldots, N$

$$\begin{aligned}
H_1 : x_k(t) &= \underbrace{A_k \cos(\omega_1 t + \psi_{k1})}_{s_1(t)} + w_k(t), \qquad t \in [0, T] \\
\vdots \quad &\vdots \quad \vdots
\end{aligned}$$

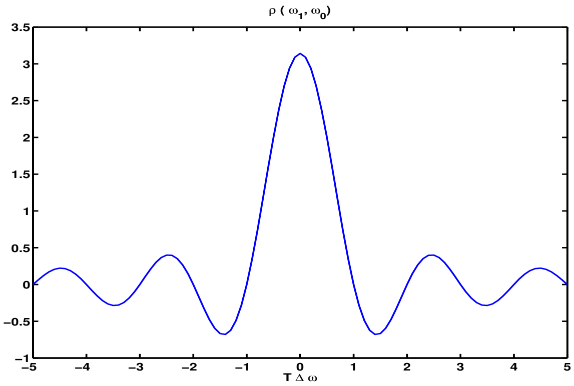


Figure 235: *Ambiguity function for discriminating between two frequencies*

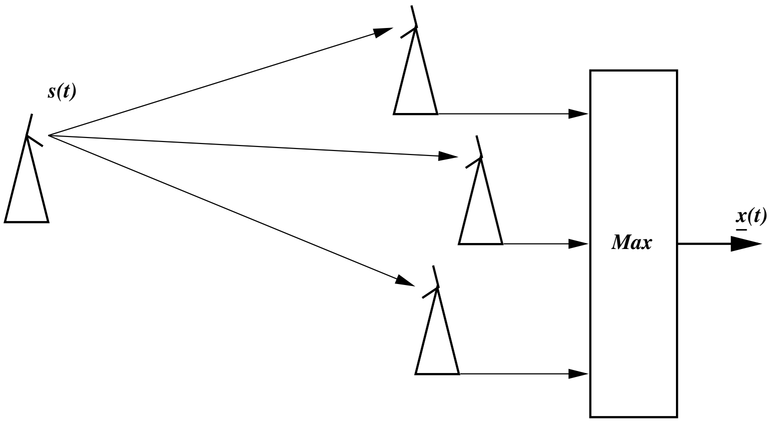


Figure 236: *Spatial diversity*

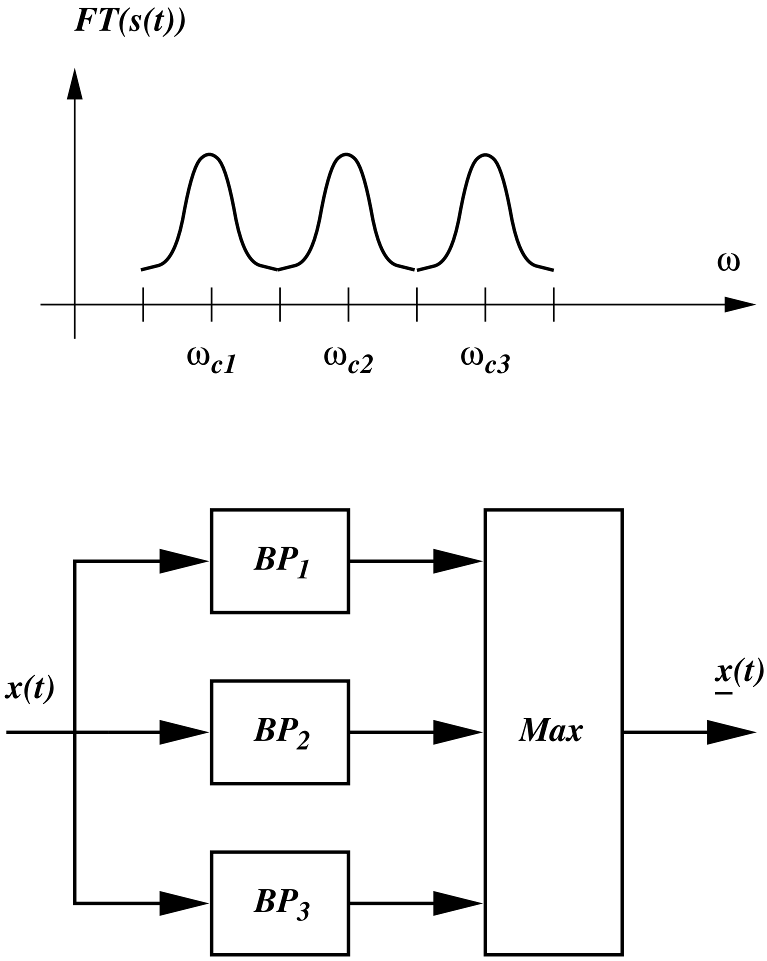


Figure 237: *Spectral diversity*

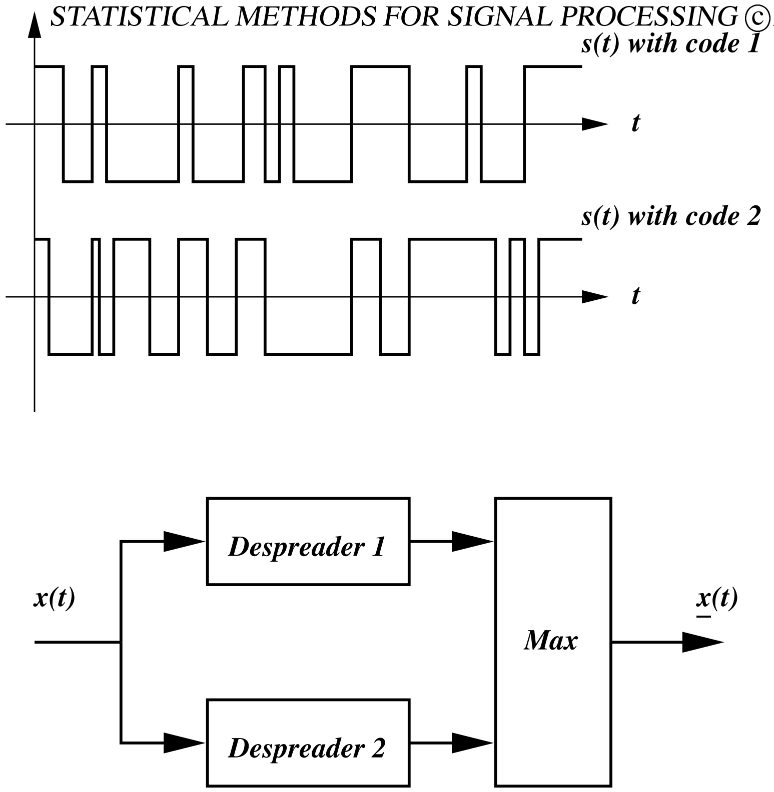


Figure 238: *Code diversity*

$$H_M : x_k(t) = \underbrace{A_k \cos(\omega_M t + \psi_{kM})}_{s_M(t)} + w_k(t), \qquad t \in [0,T]$$

* $\{A_k\}_{k=1}^N$ i.i.d. Rayleigh with $E[A^2] = A_o^2/2$

* $\{\psi_{ki}\}_{ki=1}^{NM}$ i.i.d. uniform over $[0, 2\pi]$

* $\{w_k(t)\}_{k=1}^N$ independent Gaussian white noises

* ω_i's known, equally likely

First find min P_e receiver for fixed A_k's

$$\begin{aligned}
\hat{H}_i &= \text{argmax}_{H_i}\{f(x|H_i, \{A_k\})\} \\
&= \text{argmax}_{H_i}\{\prod_{k=1}^N f(x_k|H_i, \{A_k\})\} \\
&= \text{argmax}_{H_i}\left\{ e^{-\sum_{k=1}^N \|s_i\|^2/N_o} \prod_{k=1}^N I_o(Y_i(k)) \right\}
\end{aligned}$$

where

$$Y_i(k) = \frac{2A_i}{N_o} \underbrace{|h_i(T) * x_k(T)|}_{T_i(k)}$$

$$\|s_i\|^2 = A_i^2 T/2$$

Decision regions are now function of amplitudes A_i!

To find optimal receiver

$\Rightarrow$ need average $f(x_k|H_i, \{A_k\})$ over $\{A_k\}$'s

This is straightforward but tedious (Whalen):

$$\begin{aligned} f(x|H_i) &= \left(\prod_{k=1}^N \int_0^\infty dA_k f(A_k)\right) f(x|H_i, \{A_k\}) \\ &= \prod_{k=1}^N \frac{N_o}{N_o + A_o^2 T} \exp\left(\frac{A_o\, T_i(k)}{N_o + A_o^2 T}\right) \end{aligned}$$

Therefore, we obtain optimal receiver as an incoherent diversity combiner

$$\begin{aligned} \hat{H}_i &= \text{argmax}_{H_i} \left\{\sum_{k=1}^N T_i(k)\right\} \\ \hat{H}_i &= \text{argmax}_{H_i} \left\{\sum_{k=1}^N |h_i(T) * x_k(T)|\right\} \end{aligned}$$

PERFORMANCE: P_e experiences exponential decrease in N [37].

14.5.10 DETECTION WITH UNKNOWN FREQUENCY OFFSET

Recieve narrowband signal with unknown (Doppler) offset $\delta\omega$

$$\begin{aligned} H_0 : x(t) &= w(t) \\ H_1 : x(t) &= \underbrace{Aa(t)\cos((\omega_c + \delta\omega)t + \psi_o)}_{s(t,\delta\omega)} + w(t) \end{aligned}$$

Assumption: $\delta\omega \ll \omega_c$ so that $\|s(\delta\omega)\|^2 = \|s\|^2$

INCOHERENT GLRT (w.r.t. unknown $\psi_o, \delta\omega$)

$$\begin{aligned}\Lambda_{GLR}(x) &= \max_{\psi_o,\delta\omega}\left\{\frac{f(x|A,\psi_o,\delta\omega,H_1)}{f(x|H_0)}\right\}\\ &= \max_{\delta\omega}\left\{e^{-\|s\|^2/N_o}\; e^{Y(\delta\omega)}\right\}\end{aligned}$$

Thus incoherent GLRT reduces to

$$\left| < \bar{s}_a(t, \hat{\delta\omega}_{ml}),\ e^{-j\omega_c t}x(t) > \right| \quad \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \quad \gamma$$

Two possible implementations:

1. Detection via spectrum analyzer of envelope

2: Detection via bandpass filter banks

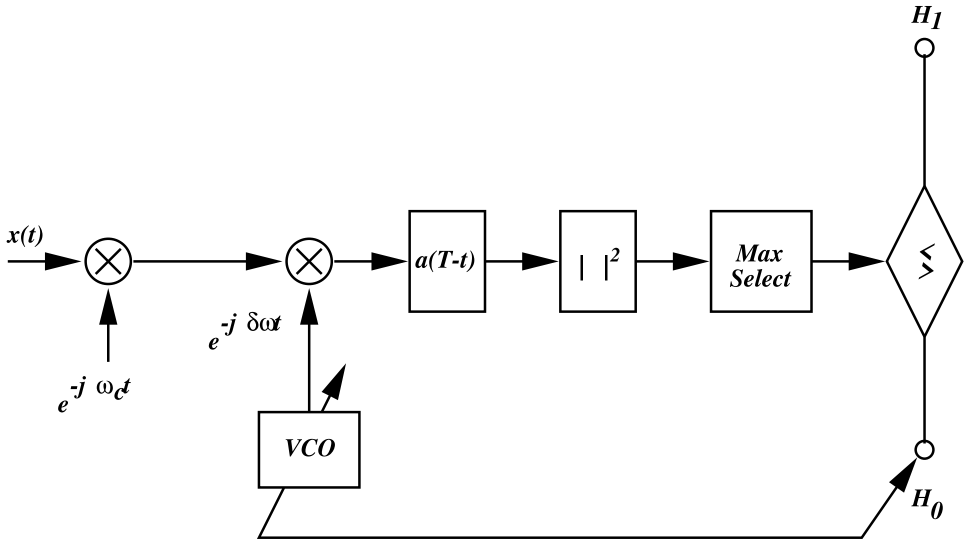


Figure 239: *Spectrum analyzer implementation for incoherent narrowband detector with unknown Doppler*

14.5.11 GLRT FOR UNKNOWN DELAY

Here the narrowband signal has unknown delay $\tau \in [0, T]$

$$\begin{aligned}H_0 : x(t) &= w(t)\\ H_1 : x(t) &= Aa(t-\tau)\cos(\omega_c(t-\tau)+\psi_o)+w(t)\end{aligned}$$

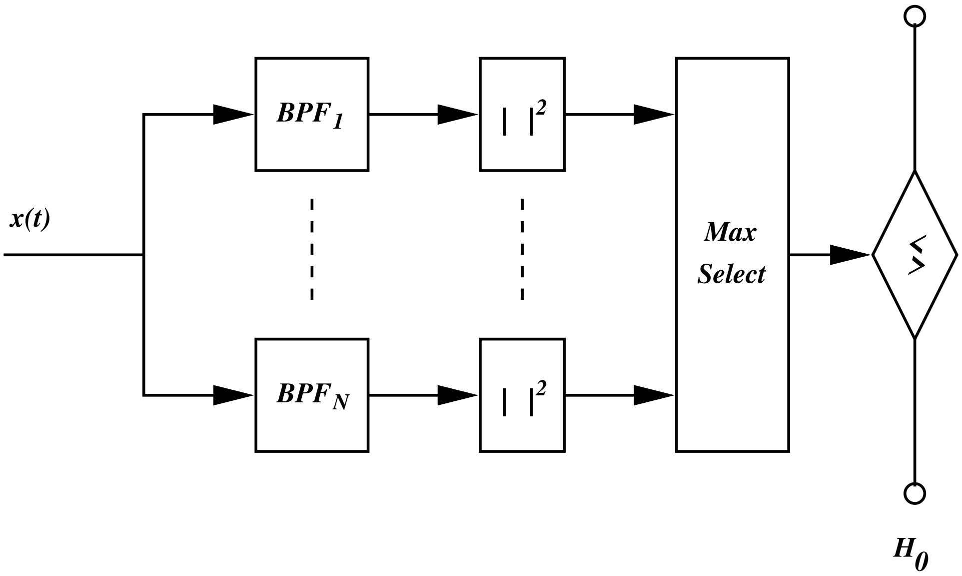


Figure 240: *Bandpass filterbank implementation*

ASSUMPTIONS:

* $a(t)$ is pulse with duration T_a, $\|a\|^2 = 1$

* $\tau \ll T_a$ so that $\|s_1(\tau)\|^2 = \|s_1\|^2$

SOLUTION: GLRT (w.r.t. unknown ψ_o, τ)

$$\begin{aligned}\Lambda_{GLR}(x) &= \max_{\psi_o,\tau}\left\{\frac{f(x|A,\psi_o,\tau,H_1)}{f(x|H_0)}\right\}\\ &= \max\tau\left\{e^{-\|s_1\|^2/N_o}\ e^{Y(\tau)}\right\}\end{aligned}$$

Thus, similarly to Doppler case, GLRT reduces to

$$\left|< \bar{s}_a(t,\hat{\tau}_{ml}),\ x(t)e^{-j\omega_c t} >\right| \quad \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \quad \gamma$$

INCOHERENT AMBIGUITY FUNCTION

Analytic signal is

$$s_a(t,\tau) = Aa(t-\tau)e^{j[-\omega_c\tau+\psi_o]}$$

Hence ambiguity function governing resolvability of two pulse arrival times τ_1, τ_0 separated by $\Delta\tau = \tau_0 - \tau_1$ is

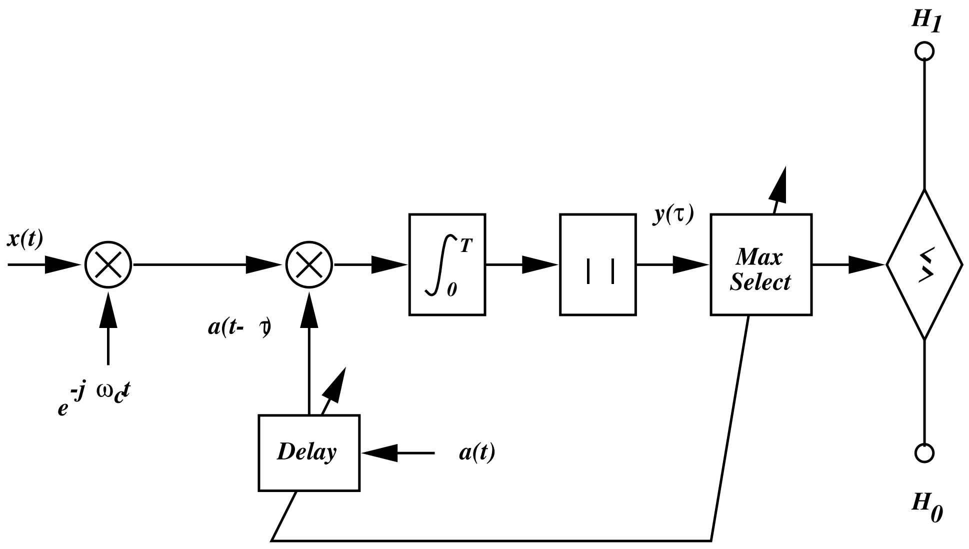


Figure 241: *Maximum threshold crossing implementation of GLRT for narrowband delayed pulse.*

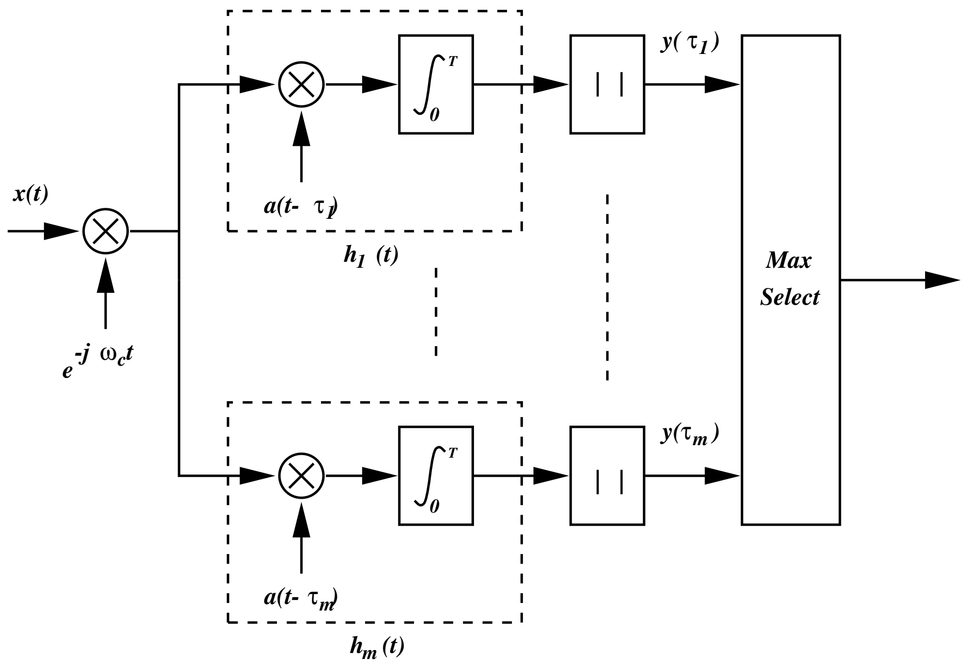


Figure 242: *Matched filter bank implementation of detection GLRT for narrowband delayed pulse.*

$$\begin{aligned}
\rho(\tau_0, \tau_1) &= \frac{|< s_a(\tau_1),\ s_a(\tau_0) >|}{\| < s_a(\tau_1)\| \, \|s_a(\tau_0)\|} \\
&= \frac{1}{A^2} \left| \int_0^T \underbrace{s_a(\tau_1) s_a^*(\tau_0)}_{A^2 a(t-\tau_1) a(t-\tau_2) e^{-j\omega\Delta\tau}} dt \right| \\
&= \left| \int_0^T a(t-\tau_1) a(t-\tau_2) dt \right| \\
&= \left| \int_0^T a(t-\Delta\tau) a(t) dt \right|
\end{aligned}$$

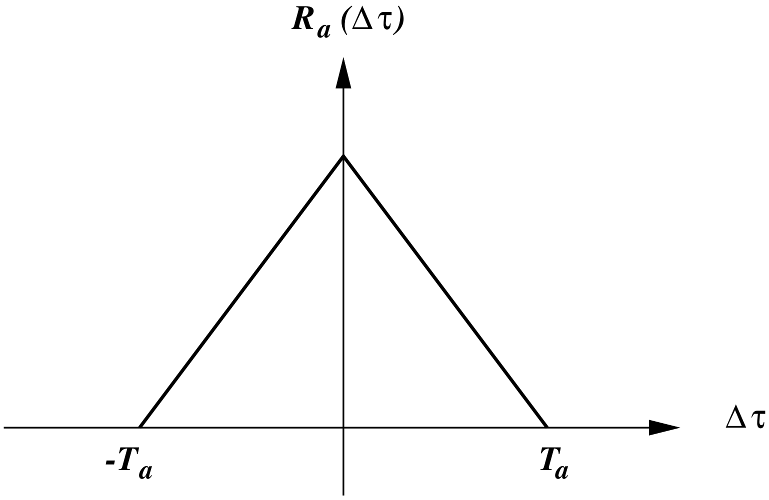


Figure 243: *Incoherent ambiguity function for delay*

14.5.12 GLRT FOR UNKNOWN DELAY AND FREQUENCY OFFSET

Here both delay and doppler are unknown parameters

$$\begin{aligned}
H_0 : x(t) &= w(t) \\
H_1 : x(t) &= Aa(t-\tau)\cos((\omega_c + \delta\omega_c)t + \psi_o) + w(t)
\end{aligned}$$

GLRT derivation is analogous to before

$$\left| < \bar{s}_a(t, \hat{\tau}_{ml}, \hat{\delta\omega}),\ x e^{-j\omega_c t} > \right| \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad \gamma$$

INCOHERENT AMBIGUITY FUNCTION

$$\rho(\tau_0, \tau_1, \delta\omega_0, \delta\omega_1) = \rho(\Delta\tau, \Delta\omega)$$

$$= \frac{|< s_a(\tau_0, \delta\omega_0),\ s_a(\tau_1, \delta\omega_1) >|}{\|s_a(\tau_0, \delta\omega_0)\|\ \|s_a(\tau_1, \delta\omega_1)\|}$$

$$= \left|\int_0^T a(t-\tau_0)a(t-\tau_1)e^{j(\Delta\omega t + \omega_c \Delta\tau)}dt\right|$$

$$= \left|\int_0^T a(t)a(t-\Delta\tau)e^{j\Delta\omega t}dt\right|$$

Resolution is limited by pulse duration $a(t)$

Compare this to COHERENT AMBIGUITY FUNCTION for this problem

$$\frac{\mathrm{Re}\,\{< s_a(\tau_0, \delta\omega_0),\ s_a(\tau_1, \delta\omega_1) >\}}{\|s_a(\tau_0, \delta\omega_0)\|\ \|s_a(\tau_1, \delta\omega_1)\|}$$

$$= \mathrm{Re}\left\{\int_0^T a(t-\tau_0)a(t-\tau_1)e^{j(\Delta\omega t + \omega_c\Delta\tau + \Delta\psi}dt\right\}$$

$$= \cos(\omega_c\Delta\tau + \Delta\psi)\ \int_0^T a(t)a(t-\Delta\tau)\cos(\Delta\omega t)dt$$

Resolution is no longer limited by pulse duration

$\Rightarrow$ much higher τ resolution is possible if know $\Delta\psi$!

14.6 Exercises

1. When a known signal $y(t)$ is transmitted through a certain multipath channel the output consists of a direct path component $y(t)$ plus an attenuated and delayed indirect path component $ay(t-\tau)$. A receiver measures the output of the channel in white additive Gaussian noise $v(t)$, $\phi_v(t) = \frac{N_o}{2}\delta(t)$:

$$z(t) = y(t) + ay(t-\tau) + v(t), \quad 0 \leq t \leq T$$

where a and τ are unknown nonrandom parameters, $0 < \tau \ll T$.

(a) Write down an expression for the likelihood function for a and τ involving $z(t)$ and $y(t)$.

(b) Find a closed form expression for the maximum likelihood estimate of a for known τ.

(c) Now, using the result of part b, find an expression for the maximum likelihood estimate of τ and draw a block diagram of your maximum likelihood receiver.

2. A model for TCP/IP network activity for heavy internet traffic is the superposition of time reversed exponentials: $y(t) = A\sum_{i=1}^{n} e^{\alpha(t-\tau_i)}u(\tau_i - t)$ where A is a peak traffic value, α is a traffic explosion factor, and τ_i are reset times. The objective is decide between light vs. heavy internet traffic based on a noisy observation of $y(t)$. For the purposes of this problem assume $\alpha = 1$ and that we capture only the first explosion occuring at time $\tau = \tau_1$ over the observation interval $t_o \leq t \leq t_f$. This gives the following

$$\begin{aligned} H_0 : z(t) &= \rho A + v(t) \\ H_1 : z(t) &= Ae^{t-\tau}u(\tau - t) + v(t), \end{aligned}$$

where ρ is the light-to-heavy traffic activity ratio and v is a Gaussian zero mean white noise with power level $N_o/2$. In the following you should make the simplifying assumption that $t_o \ll \tau \leq t_f$.

   (a) Assume that all parameters are known. Derive the MP test and draw a block diagram.

   (b) Is your test UMP for unknown τ? How about for unknown A or unknown ρ?

   (c) Now assume that τ is unknown but all other parameters are known. Derive the GLRT and draw a block diagram.

End of chapter

End of chapter

15 BIBLIOGRAPHY

References

[1] M. Abromowitz and I. Stegun, *Handbook of Mathematical Functions*, Dover, New York, N.Y., 1965.

[2] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, San Francisco, 1977.

[3] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Springer-Verlag, New York, 1981.

[4] T. S. Ferguson, *Mathematical Statistics - A Decision Theoretic Approach*, Academic Press, Orlando FL, 1967.

[5] G. H. Golub and C. F. Van Loan, *Matrix Computations (2nd Edition)*, The Johns Hopkins University Press, Baltimore, 1989.

[6] F. A. Graybill, *Matrices with Applications in Statistics*, Wadsworth Publishing Co., Belmont CA, 1983.

[7] G. Hardy, J. Littlewood, and G. Pólya, *Inequalities (2nd Edition)*, Cambridge Univ. Press, 1951.

[8] D. A. Harville, *Matrix Algebra from a Statistician's Perspective*, Springer, 1997.

[9] S. Haykin, *Modern filters*, Macmillan, New York, 1989.

[10] C. Helstrom, *Elements of signal detection and estimation*, Prentice-Hall, Englewood Cliffs, 1995.

[11] A. O. Hero, "A Cramer-Rao type lower bound for essentially unbiased parameter estimation," Technical Report 890, M.I.T. Lincoln Laboratory, Lexington, MA, Lexington, MA 02173-0073, Jan., 1992. DTIC AD-A246666.

[12] I. A. Ibragimov and R. Z. Has'minskii, *Statistical estimation: Asymptotic theory*, Springer-Verlag, New York, 1981.

[13] N. L. Johnson, S. Kotz, and A. W. N. Balakrishnan, *Continuous univariate distributions*, Wiley, New York, 1994.

[14] T. Kailath, *Lectures on Wiener and Kalman Filtering*, Springer-Verlag, New York, 1981.

[15] E. L. Lehmann, *Testing Statistical Hypotheses*, Wiley, New York, 1959.

[16] E. L. Lehmann, *Theory of Point Estimation*, Wiley, New York, 1983.

[17] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley, New York, 1969.

[18] E. B. Manoukian, *Modern Concepts and Theorems of Mathematical Statistics*, Springer-Verlag, New York N.Y., 1986.

[19] A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics*, McGraw Hill, New York, 1976.

[20] D. F. Morrison, *Multivariate statistical methods*, McGraw Hill, New York, 1967.

[21] B. Noble and J. W. Daniel, *Applied Linear Algebra*, Prentice Hall, Englewood Cliffs, NJ, 1977.

[22] A. V. Oppenheim and A. S. Willsky, *Signals and Systems*, Prentice-Hall, Englewood Cliffs, N.J., 1983.

[23] A. Papoulis, *Probability, random variables, and stochastic processes (3rd ed)*, McGraw Hill, New York, N.Y., 1991.

[24] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1988.

[25] J. G. Proakis, *Digital communications*, Prentice-Hall, NY, 1983.

[26] J. G. Proakis and D. G. Manolakis, *Digital signal processing: principles, algorithms, and applications*, Prentice-Hall, NY, 1996.

[27] C. R. Rao, *Linear Statistical Inference and Its Applications*, Wiley, New York, 1973.

[28] S. Ross, *A first course in probability*, Prentice-Hall, Englewood Cliffs, N.J., 1998.

[29] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, Reading, MA, 1991.

[30] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.

[31] G. E. Shilov, *Linear Algebra*, Dover Publications, New York, N.Y., 1977.

[32] T. Soderstrom and P. Stoica, *System identification*, Prentice Hall, Englewood Cliffs NJ, 1989.

[33] M. D. Srinath, P. K. Rajasekaran, and R. Viswanathan, *Introduction to statistical signal processing with applications*, Prentice Hall, Englewood Cliffs, 1996.

[34] H. Stark and J. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice-Hall, Englewood Cliffs, N.J., 1986.

[35] J. B. Thomas, *An Introduction to Statistical Communications Theory*, Wiley, New York, 1969.

[36] H. L. Van-Trees, *Detection, Estimation, and Modulation Theory: Part I*, Wiley, New York, 1968.

[37] A. D. Whalen, *Detection of Signals in Noise*, Academic Press, Orlando, 1971.