# A STATISTICIAN'S PERSPECTIVE

## Cultural Differences

The subject matter of this course is very important in the field of <u>statistics</u>. The subjects of Estimation and Detection are likely to be covered in a course on "statistical inference," while Filtering and Spectral Estimation will be taught in the context of "time series analysis."

A statistician will emphasize certain topics moreso than an engineer (and vice versa). A sucessfull statistician/engineer should have command of the material/terminology in both fields. This set of notes is intended to help the engineer bridge that gap.

The two fields often use different terminology. Here are some examples. Often both terms will be used but one is usually preferred.

| **Engineering** | **Statistics** |
|---|---|
| Statistical signal processing | Statistical data analysis |
| Estimation theory | Point estimation |
| Detection theory | Hypothesis testing |
| Filtering | Time series analysis |
| Gaussian distribution | Normal distribution |
| False alarm | Type I error, false positive |
| Miss | Type II error, false negative |
| Detector | Test |
| False alarm rate | Size, significance level |
| Detection rate | Power |
| GLRT | LRT |

# The Multivariate Gaussian: Composite Testing and Sampling Distributions

One sample problems

$$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

Two sample problems

$$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu_x, \sigma_x^2)$$

$$Y_1, \ldots, Y_m \overset{iid}{\sim} N(\mu_y, \sigma_y^2)$$

| COMPOSITE HYPOTHESES IN THE UNIVARIATE GAUSSIAN MODEL | GLRT |
|---|---|
| TESTS ON THE MEAN: $\sigma^2$ KNOWN | |
| CASE III: $H_0 : \mu = \mu_o$, $H_1 : \mu \neq \mu_o$ | $N$ |
| TESTS ON THE MEAN: $\sigma^2$ UNKNOWN | |
| CASE I: $H_0 : \mu = \mu_o$, $\sigma^2 > 0$, $H_1 : \mu > \mu_o$, $\sigma^2 > 0$ | $t$ |
| CASE II: $H_0 : \mu \leq \mu_o$, $\sigma^2 > 0$, $H_1 : \mu > \mu_o$, $\sigma^2 > 0$ | $t$ |
| CASE III: $H_0 : \mu = \mu_o$, $\sigma^2 > 0$, $H_1 : \mu \neq \mu_o$, $\sigma^2 > 0$ | $t$ |
| TESTS ON VARIANCE: KNOWN MEAN | |
| CASE I: $H_0 : \sigma^2 = \sigma_o^2$, $H_1 : \sigma^2 > \sigma_o^2$ | $\chi^2$ |
| CASE II: $H_0 : \sigma^2 \leq \sigma_o^2$, $H_1 : \sigma^2 > \sigma_o^2$ | $\chi^2$ |
| CASE III: $H_0 : \sigma^2 = \sigma_o^2$, $H_1 : \sigma^2 \neq \sigma_o^2$ | $\chi^2$ |
| TESTS ON VARIANCE: UNKNOWN MEAN | |
| CASE I: $H_0 : \sigma^2 = \sigma_o^2$, $H_1 : \sigma^2 > \sigma_o^2$ | $\chi^2$ |
| CASE II: $H_0 : \sigma^2 < \sigma_o^2$, $\mu \in \mathbb{R}$, $H_1 : \sigma^2 > \sigma_o^2$, $\mu \in \mathbb{R}$ | $\chi^2$ |
| CASE III: $H_0 : \sigma^2 = \sigma_o^2$, $\mu \in \mathbb{R}$, $H_1 : \sigma^2 \neq \sigma_o^2$ $\mu \in \mathbb{R}$ | $\chi^2$ |
| TESTS ON EQUALITY OF MEANS: UNKNOWN VARIANCE | |
| CASE I: $H_0 : \mu_x = \mu_y$, $\sigma^2 > 0$, $H_1 : \mu_x \neq \mu_y$, $\sigma^2 > 0$ | $t$ |
| CASE II: $H_0 : \mu_y \leq \mu_x$, $\sigma^2 > 0$, $H_1 : \mu_y > \mu_x$, $\sigma^2 > 0$ | $t$ |
| TESTS ON EQUALITY OF VARIANCES | |
| CASE I: $H_0 : \sigma_x^2 = \sigma_y^2$, $H_1 : \sigma_x^2 \neq \sigma_y^2$ | $F$ |
| CASE II: $H_0 : \sigma_x^2 = \sigma_y^2$, $H_1 : \sigma_y^2 > \sigma_x^2$ | $F$ |
| TESTS ON CORRELATION | |
| CASE I: $H_0 : \rho = \rho_o$, $H_1 : \rho \neq \rho_o$ | $t$ |
| CASE II: $H_0 : \rho = 0$, $H_1 : \rho > 0$ | $t$ |

These basic tests are widely used in applied statitistics, and therefore it is customary to catalogue them and study them together. The tests are usually named according to the sampling distribution, that is, the distribution of the test statistic.

# Gaussian Sampling Distributions

The GLRT test statistics have one of four different sampling distributions.

## Definitions

1. **Chi-square**: If $Z_i \overset{iid}{\sim} N(0,1)$, $i=1,\dots,r$ and $Y = \sum_{i=1}^{r} Z_i^2$ then $Y \sim \chi_r^2$

2. **Student t**: If $Z \sim N(0,1)$ and $Y \sim \chi_r^2$ (independent), and $X = \dfrac{Z}{\sqrt{Y/r}}$ then $X \sim t_r$.

3 **Fisher F**: If $U \sim \chi_p^2$ and $V \sim \chi_q^2$ (independent) and $W = \dfrac{U/p}{V/q}$, then $W \sim F_{p,q}$

It turns out that the GLRTs are very intuitive, as is usually the case when Gaussianity is assumed. In particular, they tend to involve the sample mean

$$\bar{x} = \frac{1}{n} \Sigma x_i$$

the sample variance

$$s^2 := \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

or the sample correlation coefficient $\hat{\rho}$.

This is why the distributions just discussed are important, because

$$\bar{X} \sim N\left(0, \frac{\sigma^2}{n}\right)$$

$$\frac{(n-1) S^2}{\sigma^2} \sim \chi^2_{n-1}$$

and $\bar{X}$, $S^2$ are independent.

Example | One-sample $t$-test

Suppose a company produces 5 pound bags of sugar for retail. They have a new packaging process and want to test whether their bags have the correct weight. So they measure the weights of $n$ bags, $X_1, \dots, X_n$, selected at random. Assume $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ with $\mu, \sigma^2$ unknown.

$$H_0 : \mu = \mu_0 \qquad (\mu_0 = 5)$$
$$H_1 : \mu \neq \mu_0$$

It can be shown (through routine steps) that the GLRT reduces to

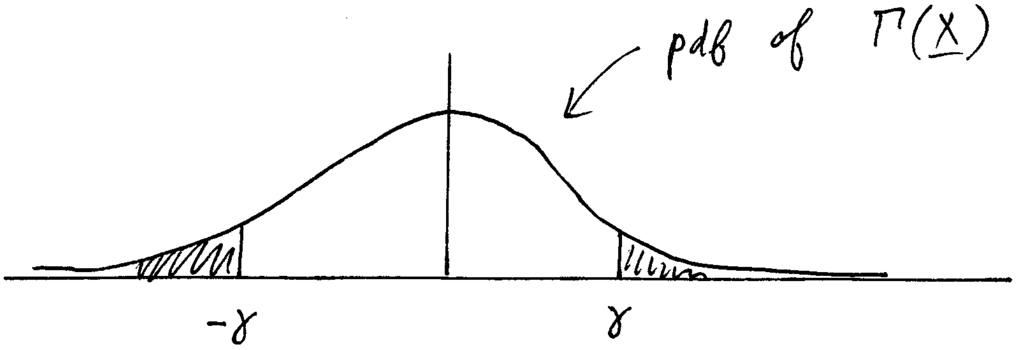
$$|\Gamma(\underline{x})| = \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

Claim: $\Gamma(\underline{X}) \sim t_{n-1}$ under $H_0$.

To see this, note

$$\Gamma(\underline{x}) = \frac{(\bar{x} - \mu_0)}{s/\sqrt{n}} = \frac{\dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)s^2}{\sigma^2} \Big/ (n-1)}}$$

$$= \frac{\mathcal{N}(0,1)}{\sqrt{\chi^2_{n-1}/(n-1)}}$$



$$Q_{t_{n-1}}(\gamma) = \frac{\alpha}{2} \implies \gamma = Q^{-1}_{t_{n-1}}\left(\frac{\alpha}{2}\right)$$

Example | Two sample unpaired t-test.

To test the affect of two treatments for blood pressure, $n_1$ patients are given one treatment and $n_2$ patients are given another. Their blood pressures are measured

$$X_1, \dots, X_{n_1} \sim N(\mu_X, \sigma^2)$$

$$Y_1, \dots, Y_{n_2} \sim N(\mu_Y, \sigma^2)$$

$$H_0 : \mu_X = \mu_Y$$

$$H_1 : \mu_X \neq \mu_Y$$

The GLRT can be reduced to

$$|\Gamma(\underline{x}, \underline{y})| = \left| \frac{\bar{y} - \bar{x}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

where

$$s_p^2 = \frac{1}{n-2} \left( \sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{i=1}^{n_2} (y_i - \bar{y})^2 \right)$$

$$= \frac{(n_1 - 1) s_X^2 + (n_2 - 2) s_y^2}{n-2}$$

Exercise | Use the fact that $\frac{n-2}{\sigma^2} s_p^2 \sim \chi^2_{n-2}$ to show $\Gamma(\underline{X},\underline{Y}) \sim t_{n-2}$ under $H_0$.

Solution $\quad \bar{X} \sim N\left(\mu_x, \frac{\sigma^2}{n_1}\right), \quad \bar{Y} \sim N\left(\mu_y, \frac{\sigma^2}{n_2}\right)$

$$\Longrightarrow \quad \bar{Y} - \bar{X} \sim N\left(\mu_y - \mu_x, \ \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

So under $H_0$

$$\frac{\bar{Y} - \bar{X}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1).$$

Thus, under $H_0$,

$$\Gamma(\underline{X}, \underline{Y}) = \frac{\bar{Y} - \bar{X}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{\bar{Y} - \bar{X} \Big/ \left(\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)}{\sqrt{\frac{(n-2)\, s_p^2}{\sigma^2} \Big/ (n-2)}}$$

$$= \frac{N(0,1)}{\sqrt{\chi^2_{n-2} / (n-2)}} \sim t_{n-2}$$

$$\Longrightarrow \quad \gamma = Q^{-1}_{t_{n-2}}\left(\frac{\alpha}{2}\right)$$

Example | Two-sample paired t-test.

Suppose we measure a patient's blood pressure before and after a treatment

$$\left.\begin{array}{l} X_1, \dots, X_n \sim N(\mu_x, \sigma^2) \\ Y_1, \dots, Y_n \sim N(\mu_y, \sigma^2) \end{array}\right\} \text{dependent!}$$

We may be able to gain information from the natural pairing of the measurements.

$$H_0: \mu_x = \mu_y$$
$$H_1: \mu_x \neq \mu_y$$

This leads to the paired t-test

$$|\Gamma(\underline{x}, \underline{y})| = \left| \frac{(\bar{y} - \bar{x})}{s_d / \sqrt{n}} \right| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

where

$$s_d^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left( y_i - x_i - (\bar{y} - \bar{x}) \right)^2$$

is the sample variance of the pairwise differences.

It can be shown that $\Gamma \sim t_{n-1}$ under $H_0$.

In essence, the paired t-test is a one-sample t-test with $\mu_0 = 0$ for the differences $Z_i := Y_i - X_i$.

Remark | Both two sample problems assumed $\sigma_x^2 = \sigma_y^2$. If this cannot be assumed, the problem becomes more challenging.

For the unpaired problem (independent samples), the problem is called the Behrens-Fisher problem. The natural statistic is

$$\Gamma = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_x^2}{n_1} + \frac{S_y^2}{n_2}}}$$

but it's distribution depends on $\sigma_x / \sigma_y$ under $H_0$!

The most common solution is Welch's approximation,

$$\Gamma \sim t_\nu$$

where

$$\nu = \frac{\left(\frac{S_x^2}{n_1} + \frac{S_y^2}{n_2}\right)^2}{\left(\frac{S_x^2}{n_1}\right)^2 / (n_1 - 1) + \left(\frac{S_y^2}{n_2}\right)^2 / (n_2 - 1)}$$

# Hypothesis Falsification

In our discussion of hypothesis testing thus far, we have said we need to either <u>choose</u> $H_0$ or <u>chose</u> $H_1$.

However, in many situations, our real objective is to prove the alternative hypothesis $H_1$ to be true. For this purpose we introduce the null hypothesis $H_0$, and our goal is to <u>falsify</u> $H_0$.

This philosophical distinction reflects the scientific viewpoint that it is easier to falsify a hypothesis ($H_0$) than to prove another ($H_1$).

$H_0$: coin is fair

$H_1$: coin is unfair

If we toss the coin 100 times and observe only 7 heads, we choose $H_1$ <u>because the data falsifies $H_0$</u>

For this reason, we can say that the outcome of a test is either to <u>accept $H_0$</u> or to <u>reject $H_0$</u>.

There is a distinction between accepting $H_0$ and proving $H_0$ / rejecting $H$.

If we observe 49 heads of 100 it doesn't prove $H_0$, we just allow that $H_0$ might be true.

Said another way, the possible outcomes of a test are

- There is enough evidence to reject $H_0$
- There is insufficient evidence to reject $H_0$.

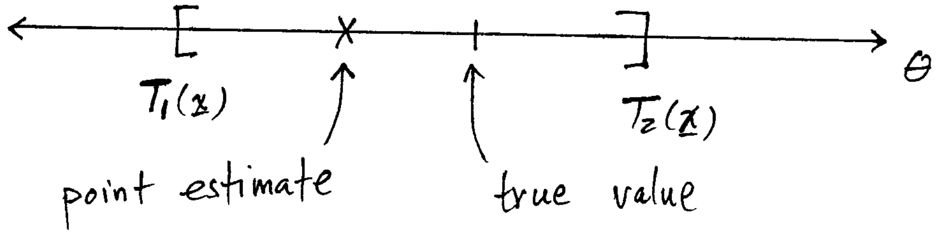
In other words, the null hypothesis is "innocent until proven guilty."

# Confidence Intervals

Our study of estimation theory thus far has centered on "point estimation." An alternative is interval estimation.

Rather that outputing a single "point" in the parameter space, an interval estimator outputs an entire interval, called a <u>confidence interval</u>.

$$\underline{x} \longmapsto [T_1(\underline{x}), T_2(\underline{x})]$$



<u>Definition</u> A $100(1-\alpha)\%$ confidence interval for a scalar parameter $\theta$ is defined by endpoints $T_1(\underline{x}), T_2(\underline{x})$ such that

$$P\{\theta \in [T_1(\underline{\mathbf{X}}), T_2(\underline{\mathbf{X}})]\} = 1-\alpha.$$

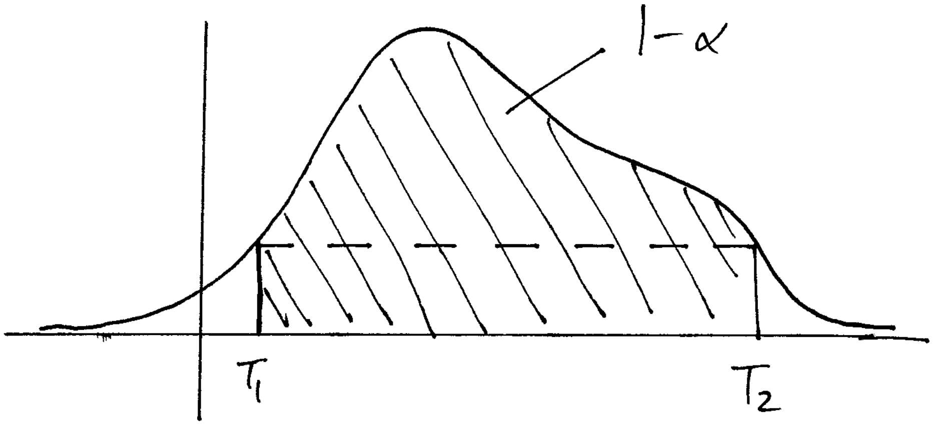
# Bayesian confidence intervals

Bayesian estimation specifies a prior and likelihood and returns a posterior distribution.

A Bayesian confidence interval should satisfy

$$\int_{T_1(\underline{x})}^{T_2(x)} f(\theta/\underline{x})\, d\theta = 1-\alpha.$$

However, there are many such intervals. Therefore, we can impose an additional restriction, for example requiring the confidence interval to have minimal length.



The corresponding interval is a level set of the posterior:

$$[T_1, T_2] = \{\theta : f(\theta|\underline{x}) \geq \lambda\} \quad \text{for some } \lambda.$$

## Classical confidence intervals

In classical estimation, the parameter is nonrandom, so what does it even mean to say

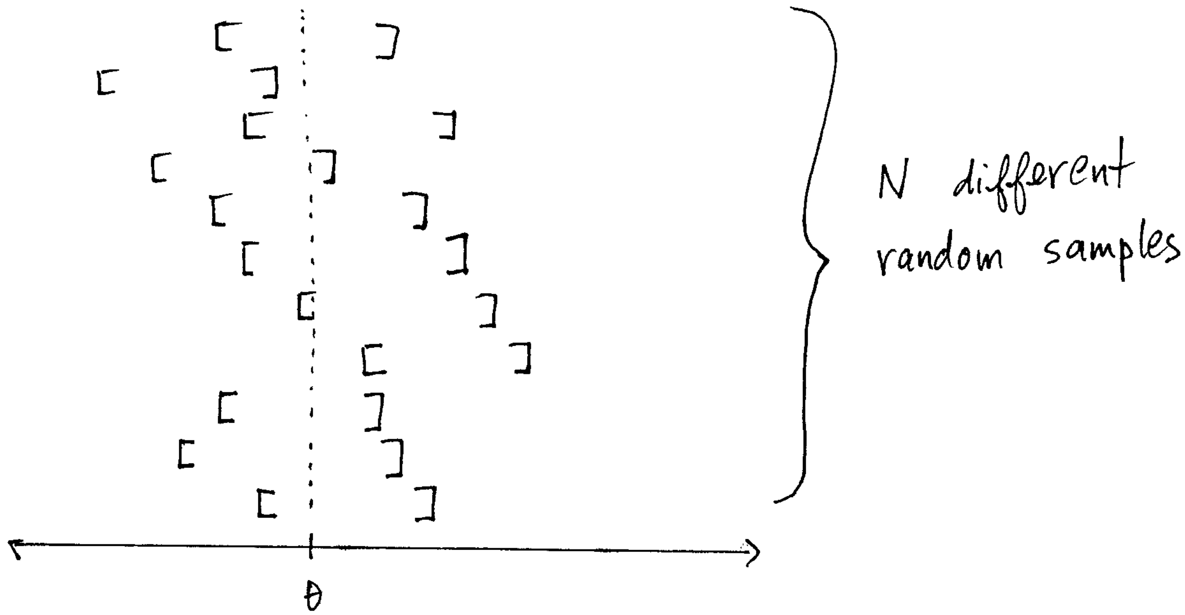
"I'm 95% sure that $\theta \in [T_1, T_2]$"?

Well, we must adjust our thinking. We view the measurement $\underline{X}$ as random so that

$$P\left\{ \theta \in [T_1(\underline{X}), T_2(\underline{X})] \right\}$$

is defined with respect to the randomness of $\underline{X}$.

What does this even mean?



For large $N$, we expect that at least $N(1-\alpha)$ of the interval estimates contain the true $\theta$.

It turns out that we may derive confidence intervals using results about hypothesis tests.

Example | Suppose $X_1, \dots, X_n \overset{iid}{\sim} N(\mu_0, \sigma^2)$ with $\mu_0, \sigma^2$ unknown. Find a $100(1-\alpha)\%$ confidence interval for $\mu_0$.

Recall the testing problem

$$H_0: \mu = \mu_0, \quad \sigma^2 > 0$$
$$H_1: \mu \neq \mu_0, \quad \sigma^2 > 0$$

We saw that the test

$$\left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_\alpha := Q^{-1}_{t_{n-1}}\left(\frac{\alpha}{2}\right)$$

has size $\alpha$.

In other words

$$P\left\{ -\gamma_\alpha \le \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \le \gamma_\alpha \right\} = 1 - \alpha.$$

when $\mu = \mu_0$.

That is,

$$1-\alpha = P\left\{-\gamma_\alpha \le \frac{\bar{X}-\mu_0}{s/\sqrt{n}} \le \gamma_\alpha\right\}$$

$$= P\left\{-\gamma_\alpha \frac{s}{\sqrt{n}} \le \bar{X}-\mu_0 \le \gamma_\alpha \frac{s}{\sqrt{n}}\right\}$$

$$= P\left\{-\gamma_\alpha \frac{s}{\sqrt{n}} \le \mu_0-\bar{X} \le \gamma_\alpha \frac{s}{\sqrt{n}}\right\}$$

$$= P\left\{\bar{X}-\gamma_\alpha \frac{s}{\sqrt{n}} \le \mu_0 \le \bar{X}+\gamma_\alpha \frac{s}{\sqrt{n}}\right\}$$

Therefore,

$$\left[\bar{X}-\gamma_\alpha \frac{s}{\sqrt{n}}\ ,\ \bar{X}+\gamma_\alpha \frac{s}{\sqrt{n}}\right]$$

is a $100(1-\alpha)\%$ confidence interval for the mean.

Example | $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\mu, \sigma^2$ unknown.

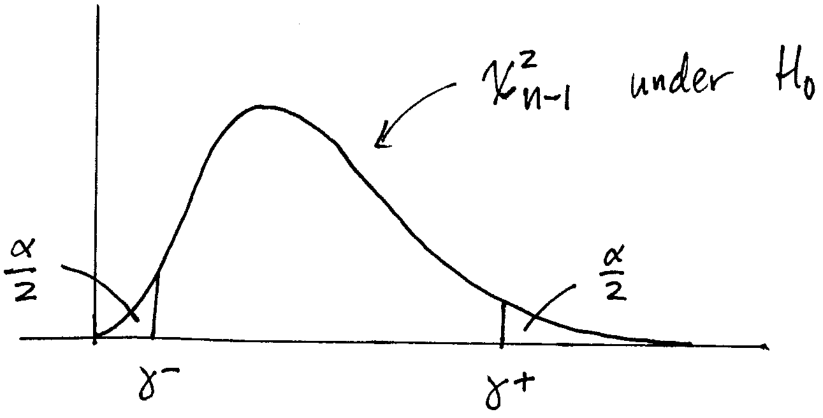
Find $100(1-\alpha)\%$ CI for $\sigma^2$.

For the testing problem

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 \neq \sigma_0^2$$

the GLRT reduces to

$$\text{declare } H_0 \iff \gamma^- \leq \frac{(n-1)s^2}{\sigma_0^2} \leq \gamma^+$$



$$\gamma_\alpha^+ = Q^{-1}_{\chi^2_{n-1}}\left(\frac{\alpha}{2}\right)$$

$$\gamma_\alpha^- = Q^{-1}_{\chi^2_{n-1}}\left(1-\frac{\alpha}{2}\right)$$

↖ "equal tail" thresholds; other choices are possible, e.g. minimal length

Exercise | Find a $100(1-\alpha)\%$ CI for $\sigma^2$.

Solution

$$1-\alpha = P\left\{ \gamma_\alpha^- \le \frac{(n-1)s^2}{\sigma_0^2} \le \gamma_\alpha^+ \right\}$$

$$= P\left\{ \frac{(n-1)s^2}{\gamma_\alpha^+} \le \sigma_0^2 \le \frac{(n-1)s^2}{\gamma_\alpha^-} \right\}$$

$\Longrightarrow$ $\left[ \frac{(n-1)s^2}{\gamma_\alpha^+} , \frac{(n-1)s^2}{\gamma_\alpha^-} \right]$ is a $100(1-\alpha)\%$ CI.

The basic mechanism for constructing CI's presented here can be generalized using the concept of pivots.

## p-values

The size or false alarm rate of a test is sometimes called the significance level.

Our general approach has been to set the significance level $\alpha$ in advance, and to make a hard binary decision ($H_0$ or $H_1$) depending on $\alpha$.

Such "hard decisions" do not convey how close the observation was to the opposite decision.

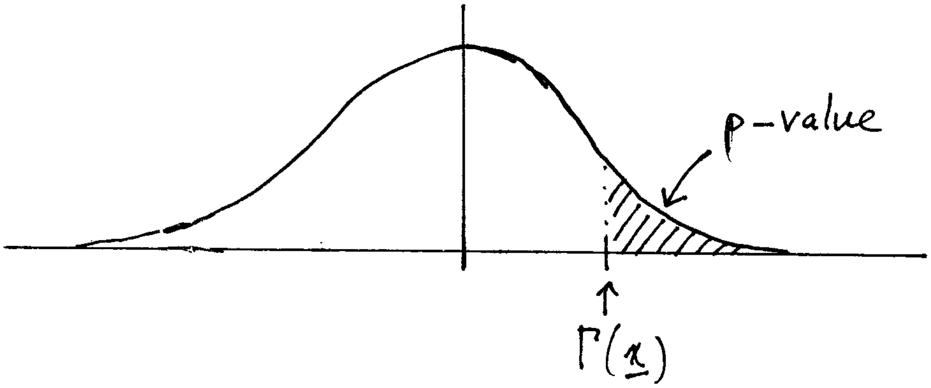
Definition Consider testing a simple null hypothesis against some alternative. The p-value of a measurement $\underline{x}$ is the probability, under the null hypothesis, of observing a measurement at least as extreme as $\underline{x}$

Example | $X_1, \dots, X_n \sim N(\mu, \sigma^2)$, $\sigma^2$ known

$$\left.\begin{array}{ll} H_0: & \mu = \mu_0 \\ H_1: & \mu > \mu_0 \end{array}\right\} \xrightarrow{\text{GLRT}} \Gamma(\underline{x}) = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

Ignore the threshold for the moment.

The distribution of $\Gamma(\underline{X})$ is $N(0,1)$ under $H_0$.



The probability (under $H_0$) of $\Gamma(\underline{X})$ being more extreme than $\Gamma(\underline{x})$ is

ⓐ

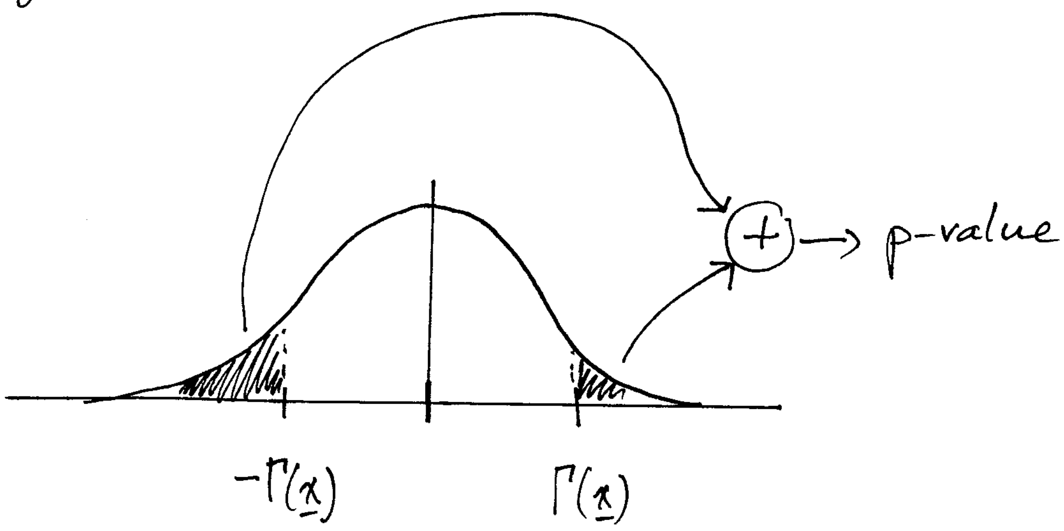
If we want a hard decision at level $\alpha$ we can express this in terms of the p-value as

ⓑ

Now consider the two-sided problem

$$\left.\begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{array}\right\} \xRightarrow{GLRT} |\Gamma(\underline{x})| = \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

Again, $\Gamma(\underline{X}) \sim N(0,1)$ under $H_0$, but now "extreme" takes on a new meaning



© $\Longrightarrow$ p-value =

Typically a p-value $\leq .05$ is considered grounds for rejecting the null hypothesis. p-values are especially useful for addressing the multiple testing problem.

# Multiple Testing

Suppose you want to decide whether a certain coin is fair.

$$H_0 : \quad \theta = \frac{1}{2}$$

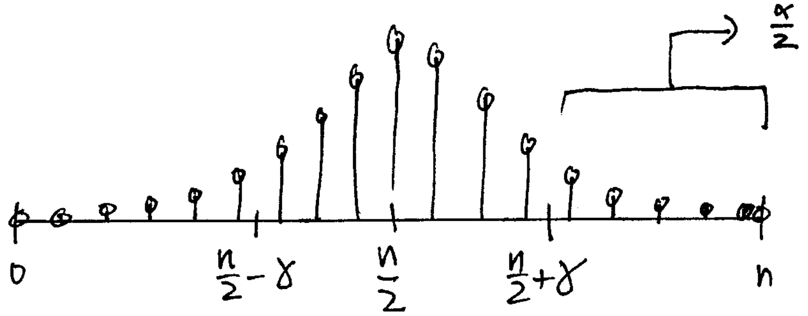
$$H_1 : \quad \theta \neq \frac{1}{2}.$$

So you toss the coin n times and observe the number x of heads. A natural test is

$$\left| x - \frac{n}{2} \right| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

To ensure a false alarm rate of $\alpha$, we know to choose $\gamma$ such that

$$2 \cdot \sum_{k=\lceil \frac{n}{2}+\gamma \rceil}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^{n} \approx \alpha$$



Now suppose you are presented with $N$ different coins, and you must determine which of them are fair.

If you perform the test we just discussed, even if all the coins are fair, we expect to "discover" about $N \cdot \alpha$ unfair coins.

Example | If $N = 1000$, $\alpha = .05$, and we discover 50 unfair coins, would you really believe those coins are unfair?

A solution to this conundrum is to adopt an alternative notion of "size." A common choice is the family-wise error rate

$$\text{FWER} = P\left( \geq 1 \ H_0 \text{ rejected} \mid \text{all } H_0 \text{ true} \right)$$

# Sidak correction

Suppose all measurements are independent.

Denote $\Omega_i$ = event that ith

$$\Omega = \bigcup_{i=1}^{N} \Omega_i$$

Then

$$\begin{aligned}
\text{FWER} &= P_{\text{all } H_0}(\Omega) \\
&= 1 - P_{\text{all } H_0}(\Omega^c) \\
&= 1 - P_{\text{all } H_0}\left(\bigcap_{i=1}^{N} \Omega_i^c\right) \qquad \boxed{\text{by independence}} \\
\boxed{\begin{array}{l}\alpha = \text{size of} \\ \text{individual} \\ \text{test}\end{array}} \qquad &= 1 - \prod_{i=1}^{N} P_{H_0^i}(\Omega_i^c) \\
&= 1 - (1-\alpha)^N
\end{aligned}$$

Thus, if we desire $\text{FWER} \le \alpha'$, it suffices to set $\alpha = 1 - (1-\alpha')^{1/N}$ in each individual test.

# Bonferroni Correction

If $\{\Omega_i\}_{i=1}^N$ are not independent, the union bound implies

$$\begin{aligned} FWER &= P_{\text{all } H_0}(\Omega) \\ &= P_{\text{all } H_0}\left(\bigcup_{i=1}^N \Omega_i\right) \\ &\le \sum_{i=1}^N P_{H_0^i}(\Omega_i) \\ &= N \cdot \alpha \end{aligned}$$

So $\alpha = \frac{\alpha'}{N} \Longrightarrow FWER \le \alpha'$.

This "adjustment" is more conservative than the Sidak correction, but also more general.

Equivalently, if $p_1, \dots, p_N$ are the p-values of the $N$ tests, we may define the "adjusted p-values" $p_i' = N \cdot p_i$ and decide by comparing $p_i'$ to $\alpha'$.

# False Discovery Rate

Many think the FWER is too conservative. It is often worthwhile to allow a few false alarms if the number of correct detections increases significantly.

This led Benjamini and Hochberg to study the false discovery rate (FDR)

$$\text{FDR} = E\left[\frac{FD}{D}\right]$$

where

$D$ = # of "discoveries", i.e. $H_0$ rejected

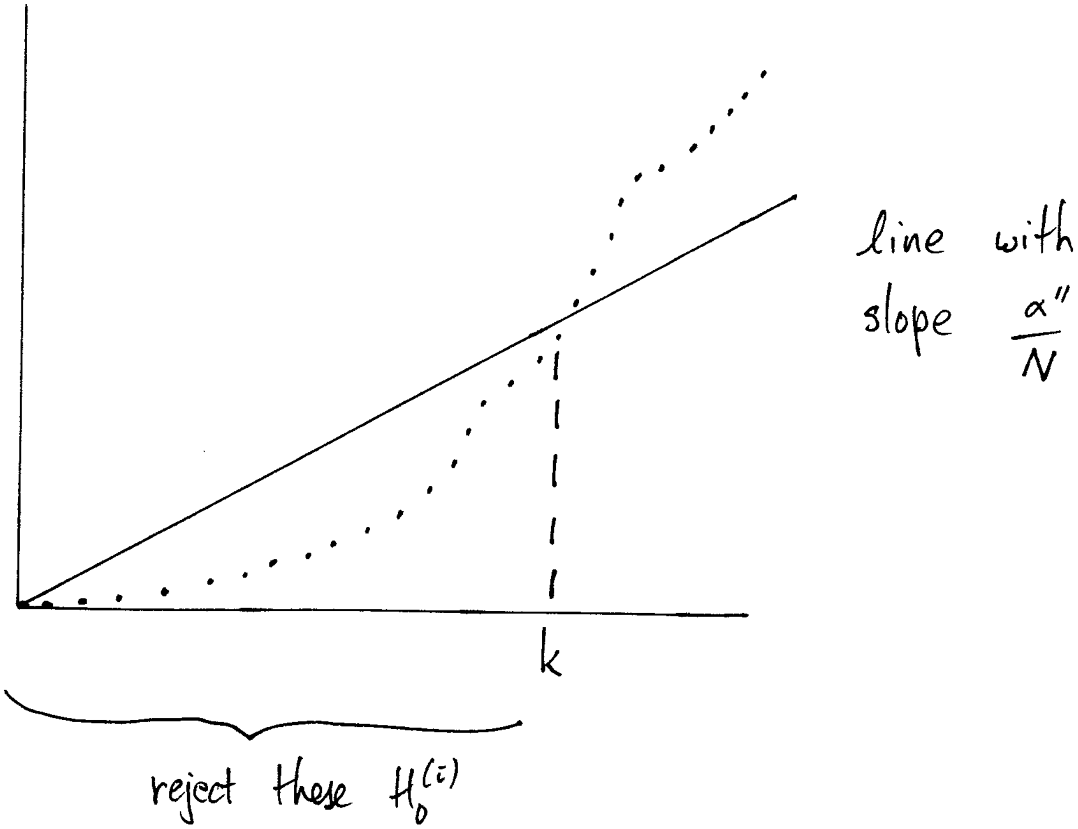
$FD$ = # of "false discoveries", i.e. $H_0$ incorrectly rejected.

B&H showed how to ensure $\text{FDR} \leq \alpha$"

Let $p_{(1)} \le p_{(2)} \le \cdots \le p_{(N)}$ be the ordered p-values, and let $H_0^{(i)}$ be the hypothesis corresponding to $p_{(i)}$. Let $k$ be the largest $i$ such that

$$p_{(i)} \le \frac{i}{N} \alpha''.$$

Reject all $H_0^{(i)}$, $i = 1, 2, \ldots, k$.



$\Longrightarrow$ **F**DR $\le \alpha''$

## Summary

- Statisticians study similar problems to those encountered in statistical signal processing, but often with different terminology and emphases.
- Composite testing with multivariate normal data:
  - GLRT yields intuitive tests
  - tests named after distribution of test stat. under $H_0$
- Classical confidence intervals: equivalence with classical hypothesis testing
- p-values: hypothesis testing with "soft" decisions, convenient for addressing multiple testing problem

Key

a. $Q(\Gamma(\underline{x}))$

b. $\text{p-value}(\underline{x}) = Q(\Gamma(\underline{x})) \underset{H_1}{\overset{H_0}{\gtrless}} \alpha$

c. $2Q(\Gamma(\underline{x}))$