Lower Bounds on Estimator Performance for Energy-Invariant Parameters of Multidimensional Poisson Processes

ALFRED O. HERO, MEMBER, IEEE

Abstract —Using rate distortion theory, lower bounds are developed for the mean-square error of estimates of a random parameter of an M-dimensional inhomogeneous Poisson process, with respect to which the energy, i.e., the average number of points, is invariant. The bounds are derived without stringent assumptions on either the form of the intensity or the prior distribution of the parameter, and they can handle random nuisance parameters. The derivation makes use of a side-information averaging principle applied to the distortion-rate function and a maximum entropy property of energy constrained Poisson processes. Under the additional assumption of conditional entropy invariance of the point process with respect to the parameter of interest, an explicit bound is given which depends on the information discrimination between the inhomogeneous conditionally Poisson process and a nearly homogeneous Poisson process. The application of the explicit bound is illustrated through a treatment of the problems of time-shift estimation and relative time-shift estimation for Poisson streams.

I. Introduction

A PROBLEM of importance in optical communications is the etection and demodulation of a pulse-position modulation (PPM) optical signal. This entails the estimation of time shifts of the point process intensity function that governs the average rate of detected photons. A related problem is the estimation of the relative group delay between two photon packets. This problem arises in nuclear particle detection systems, such as positron-emission tomography (PET) medical imaging systems where the group delay is related to the differential time-of-flight between a pair of positrons, and in neural response systems where packets are generated by two associated neural firings processes. Finally, in astronomical imaging of weak stellar sources, absolute and relative positional information such as Doppler frequency is to be extracted from an observed photon spatial process. The aforementioned estimation problems can all be grouped under a common model: a multidimensional inhomogeneous Poisson process N for which the "process energy" $E[N]$ is independent of the parameter of interest but may depend on unknown "nuisance parameters" which are not of interest. The lower bounding of the mean-square error (mse) is important in that it provides an indication of the fundamental limits on estimator performance and also a benchmark for comparison to the known performance of a particular estimator.

Manuscript received November 2, 1987; revised September 19, 1988. This work was supported in part by the National Institutes of Health under Contract R01-CA46622-01.

The author is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109.

IEEE Log Number 8929753.

Most of the available performance results have been for the one-dimensional case of single-detector timing estimation. In [1] the local, or small error, mean-square error of several *ad hoc* estimators was studied under the assumption that the unknown time delay τ is a nonrandom constant, and the intensity function is smooth and essentially observable with a small random error. In [3] exact and approximate expressions for the mse were derived for the case of conditional mean estimators of random τ and very simple functional forms for the intensity. In [2] the mse for the maximum *a posteriori* timing estimator was characterized by an upper bound for optical communications signals. More recently, in [4] the mse of the maximum likelihood timing estimator has been studied for PET systems.

One of the simpler lower bounds on timing mse is the Cramer–Rao (CR) lower bound. The CR bound has been studied in the context of optical communications [2] and in the context of PET systems [4]. It was observed, however, that the Cramer–Rao lower bound becomes useless for sharp intensity functions. In particular, nearly nondifferentiable intensities give the unachievable CR bound of zero mse. Important examples of nondifferentiable intensities are the rectangular pulse train, a common model for a PPM optical communications signal, and the bi-exponential pulse, a common model for scintillator pulse shape in particle detection systems. For the special case of a narrow rectangular pulse, an approximation to an mse lower bound for random time shift was derived in [3] using an information theoretic approach, and in [6] the parameter estimation Ziv–Zakai lower bound for nonrandom time shift was computed. While this special case is important for the optical communications problem, the approximation techniques in [3] and [6] cannot easily be extended to the more general framework considered in this paper.

Rate distortion theory has been applied to bound the mse for both static [7] and dynamic [8] nonlinear estimation problems. In this paper an exact rate distortion bound on the mse of estimators of random parameters, with respect to which the point process is energy invariant, is

0018-9448/89/0700-0843$01.00 ©1989 IEEE

derived. The bound can handle multiple observations and nuisance parameters, and it is derived under broad assumptions on the intensity function of the process. These assumptions are that 1) the intensity function is totally supported in the observation interval, and 2) the intensity function is bounded. Hence the bound is more generally applicable than previous lower bounds. The foundation for the derivation is rate distortion theory. The general bound is developed via an important extremal property: the differential entropy of an energy constrained point process is maximized by an inhomogeneous Poisson process.

In Section VI the paper focuses on the case where the conditional entropy of the point process is independent of the probability density function (pdf) of the parameter of interest. This includes the case of shift estimation in the presence of nuisance parameters, e.g., time delay and relative time delay estimation in Poisson streams for which the intensity function is a randomly shifted pulse over time. For these estimation problems, the mse bound is shown to decrease exponentially as a function of the average energy of the point process $E[N]$ where the rate of decrease is controlled by the information discrimination between the intensity density of the observed points and a uniform intensity density over the *a priori* pulse interval. Since the information discrimination is a measure of the difference between two densities, the bound displays the expected relationship between the accuracy of the estimate of the pulse location and the sharpness of the pulse. Furthermore, unlike the CR bound, our bound is always nontrivial for finite valued intensity functions. For the case of a rectangular PPM pulse, the rate distortion lower bound on the mse of time-shift estimators is tighter than the approximate bound of [3] in the low pulse amplitude regime. Next an mse bound for time-delay estimation is derived for the case of unknown intensity pulse amplitude and time width. For this case the mse bound decreases exponentially in the mean pulse amplitude and increases exponentially in the mean of the logarithm of the pulsewidth. Finally, the paper treats the case of relative time delay appropriate for time-of-flight estimation in PET.

Based on the form of the CR and rate-distortion-type lower bounds presented in this paper, we expect the rate-distortion bound to be useful for predicting optimum estimator performance in cases where the process energy $E[N]$ is in the low to moderate range. On the other hand, when the CR bound is nondegenerate, the CR bound is expected to be tighter than the rate-distortion bound for higher ranges of process energy.

The organization of the paper is as follows. In Section II the statistical model for the point process observation is presented, and all of the necessary assumptions for the derivation of the rate-distortion bound are stated. In Section III the estimation problem is stated, a CR bound on estimator mse for conditionally Poisson processes with random parameters is derived, and the bound [3] for the PPM problem is given. In Section IV some information properties of Poisson process are established to derive the rate-distortion bound. The main results of the paper are given in Section V. In Section V-A the general bound is given implicitly in terms of an undetermined channel capacity for which necessary and sufficient conditions are given. In Section V-B the property of conditional entropy invariance is used to derive an explicit mse bound. In Section VI the bound in Section V-B is specialized to the problems of shift and relative shift estimation for spatial point processes and for conditionally Poisson streams.

II. Conditionally Poisson Observation Model

Let $N = \{dN(\underline{t}): \underline{t} \in I\}$ be an inhomogeneous Poisson process whose index space is the finite M-dimensional rectangular region $I \triangleq \times_{i=1}^{M}[-T_i/2, T_i/2], 0 < T_i < \infty$. As a matter of terminology, N will be referred to as an M-dimensional (inhomogeneous) Poisson process over I. The notation $N(B)$, B a Borel subset of $\mathbb{R}^M$ contained in I, denotes the total number of points contained in B: $N(B) = \int_B dN(\underline{t})$. The point process N is described through the set of points in I where $dN(\underline{t}) = 1$. We refer to a specific relative ordering of these points by the labels $\underline{t}_1, \cdots, \underline{t}_n$ where $n = N(I)$ is the total number of observed points over I. By contrast, the population of points, without regard to ordering, is denoted by $\{\underline{t}_{(1)}, \cdots, \underline{t}_{(n)}\}$.

Let $\underline{\tau}$ be an L-dimensional random vector with pdf $p_{\underline{\tau}}$ whose support set is $J = \times_{i=1}^{L} J_i \triangleq \{\underline{\tau}: p_{\underline{\tau}}(\underline{\tau}) > 0\}$. In what follows τ_1 is the parameter of interest to be estimated, and $\tau_2, \cdots, \tau_L$ are nuisance parameters. We assume that, conditioned on $\underline{\tau} = \underline{\tau}$, the point process is Poisson with intensity $\lambda(\underline{t}; \underline{\tau})$: $\lambda(\underline{t}; \underline{\tau}) d\underline{t} \triangleq E[dN(\underline{t}) | \underline{\tau} = \underline{\tau}]$. Specifically, if the intensity $\lambda(\underline{t}, \underline{\tau})$ is integrable over $\underline{t} \in I$, the Poisson conditional distribution of N, denoted by $P(N|\underline{\tau})$, is specified by the set of joint pdf's of the points $\underline{t}_1, \cdots, \underline{t}_n$ for $n = 0, 1, \cdots$ [9]:

$$p_{\underline{t}_1, \cdots, \underline{t}_n | \underline{\tau}}(\underline{t}_1, \cdots, \underline{t}_n | \underline{\tau}) = \begin{cases} \prod_{i=1}^{n} \lambda(\underline{t}_i; \underline{\tau}) e^{-\Lambda(\underline{\tau})}, & n > 0 \\ e^{-\Lambda(\underline{\tau})}, & n = 0 \end{cases} \tag{1}$$

where $\Lambda(\underline{\tau})$ is the integral of the intensity

$$\Lambda(\underline{\tau}) \triangleq E[N|\underline{\tau}] = \int_I \lambda(\underline{t}; \underline{\tau})\, d\underline{t}. \tag{2}$$

It is hereafter implicitly assumed that $\Lambda(\underline{\tau})$ is finite (with probability 1).

The unconditional distribution of N, denoted by $P(N)$, is specified by the expectation of (1) over $\underline{\tau} \in J$:

$$p_{\underline{t}_1, \cdots, \underline{t}_n}(\underline{t}_1, \cdots, \underline{t}_n) = \int_J d\underline{\tau}\, p_{\underline{t}_1, \cdots, \underline{t}_n | \underline{\tau}}(\underline{t}_1, \cdots, \underline{t}_n | \underline{\tau})\, p_{\underline{\tau}}(\underline{\tau}). \tag{3}$$

Likewise, the conditional distribution of N given τ_1, $P(N|\tau_1)$, is specified by the conditional expectation over $\underline{\tau}_2^L \triangleq (\tau_2, \cdots, \tau_L) \in \times_{i=2}^{L} J_i$:

$$p_{\underline{t}_1, \cdots, \underline{t}_n | \tau_1}(\underline{t}_1, \cdots, \underline{t}_n | \tau_1) = \int_{\times_{i=2}^{L} J_i} d\underline{\tau}_2^L\, p_{\underline{t}_1, \cdots, \underline{t}_n | \underline{\tau}}(\underline{t}_1, \cdots, \underline{t}_n | \underline{\tau})\, p_{\underline{\tau}_2^L | \tau_1}(\underline{\tau}_2^L | \tau_1). \tag{4}$$

Two important properties of Poisson processes are the following

Lemma 1: Let N^o be an M-dimensional Poisson point process with integrable intensity function λ^o: i.e., $\lambda^o = \{\lambda^o(\underline{t}): \underline{t} \in I\}$, $\Lambda \triangleq \int_I \lambda^o(\underline{t})\, d\underline{t} < \infty$. Define n^o as the total number of points over I and $\{\underline{t}_1, \cdots, \underline{t}_{n^o}\}$ the set of these points. If g is any function of $\underline{t}$ such that $\int_I |g(\underline{t})|^2 \lambda^o(\underline{t})\, d\underline{t} < \infty$, then

$$E\left[\sum_{i=1}^{n} g(\underline{t}_i)\right] = \int_I g(\underline{t})\lambda^o(\underline{t})\, d\underline{t}. \quad (5)$$

If g_1 and g_2 are two such functions, then

$$E\left[\sum_{i=1}^{n} g_1(\underline{t}_i) \sum_{j=1}^{n} g_2(\underline{t}_j)\right] = \int_I g_1(\underline{t}) g_2(\underline{t}) \lambda^o(\underline{t})\, d\underline{t} + \int_I g_1(\underline{t})\lambda^o(\underline{t})\, d\underline{t} \int_I g_2(\underline{t})\lambda^o(\underline{t})\, d\underline{t}. \quad (6)$$

Recall the definition of the differential entropy $H(X)$ of a random quantity X:

$$H(X) \triangleq E[-\ln p_X(X)] \quad (7)$$

where the expectation in (7) is with respect to the density p_X.

Lemma 2: Let N^o be a Poisson process as in Lemma 1. Assume that λ^o is bounded almost everywhere. Then the differential entropy $H(N^o)$ of N^o is given by

$$H(N^o) = \int_I \lambda^o(\underline{t})\, d\underline{t} - \int_I \lambda^o(\underline{t}) \ln \lambda^o(\underline{t})\, d\underline{t}. \quad (8)$$

Proof of Lemma 1: The proof follows straightforwardly from the form of the characteristic function of the inhomogeneous Poisson process. Define the random variable $Z \triangleq \sum_{i=1}^{n^o} g(\underline{t}_i)$. The characteristic function ϕ_Z of Z is [9, theorem 4.1]

$$\phi_Z(u) \triangleq E[e^{juZ}] = \exp\left(\int_I \lambda^o(\underline{t})[e^{jug(\underline{t})} - 1]\, d\underline{t}\right). \quad (9)$$

Hence from the moment generating property of ϕ_Z:

$$\begin{aligned} E[Z] &= \frac{1}{j} \lim_{u \to 0} \frac{d}{du}\phi_Z(u) \\ &= \frac{1}{j} \lim_{u \to 0} \left(\frac{d}{du}\int_I \lambda^o(\underline{t})[e^{jug(\underline{t})} - 1]\, dt\right)\phi_Z(u). \quad (10) \end{aligned}$$

Since $\int_I \lambda^o|g| \leq (\Lambda \int_I \lambda^o |g|^2)^{1/2} < \infty$ (Schwarz inequality), the limiting and integral operations in (10) are interchangeable (Lebesgue dominated convergence lemma [10]); hence

$$E[Z] = \int_I \lambda^o(\underline{t}) g(\underline{t}) \lim_{u \to 0} e^{jug(\underline{t})}\, d\underline{t} = \int_I \lambda^o(\underline{t}) g(\underline{t})\, d\underline{t}, \quad (11)$$

which is (5).

Define the random variables $Z_1 \triangleq \sum_{i=1}^{n^o} g_1(\underline{t}_i)$ and $Z_2 \triangleq \sum_{i=1}^{n^o}(\underline{t}_i)$. The joint characteristic function ϕ_{Z_1, Z_2} of Z_1 and Z_2 is

$$\begin{aligned} \phi_{Z_1, Z_2}(u_1, u_2) &\triangleq E[e^{j(u_1 Z_1 + u_2 Z_2)}] \\ &= E[e^{juZ'}] \quad (12) \end{aligned}$$

where $Z' \triangleq \sum_{i=1}^{n^o} g_1(\underline{t}_i) + (u_2/u_1) g_2(\underline{t}_i)$ and $u \triangleq u_1$. The identification of g in (9) with $g_1 + (u_2/u_1)g_2$ yields the identity

$$\phi_{Z_1; Z_2}(u_1, u_2) = \exp\left(\int_I \lambda^o(\underline{t})[e^{j[u_1 g_1(\underline{t}) + u_2 g_2(\underline{t})]} - 1]\, d\underline{t}\right). \quad (13)$$

Now

$$\begin{aligned} E[Z_1 Z_2] &= \frac{1}{j^2} \lim_{u_1, u_2 \to 0} \frac{\partial^2}{\partial u_1 \partial u_2}\phi_{Z_1, Z_2}(u_1, u_2) \\ &= \lim_{u_1, u_2 \to 0} \phi_{Z_1, Z_2}(u_1, u_2) \\ &\quad \cdot \left[\int_I \lambda^o(\underline{t}) g_1(\underline{t}) g_2(\underline{t}) e^{j[u_1 g_1(\underline{t}) + u_2 g_2(\underline{t})]}\, d\underline{t} \right. \\ &\quad + \int_I \lambda^o(\underline{t}) g_1(\underline{t}) e^{j[u_1 g_1(\underline{t}) + u_2 g_2(\underline{t})]}\, d\underline{t} \\ &\quad \left. \cdot \int_I \lambda^o(\underline{t}) g_2(\underline{t}) e^{j[u_1 g_1(\underline{t}) + u_2 g_2(\underline{t})]}\, d\underline{t}\right]. \quad (14) \end{aligned}$$

Since $\int_I \lambda^o |g_1 g_2| \leq (\int_I \lambda^o |g_1|^2 \int_I \lambda^o |g_2|^2)^{1/2} < \infty$, and $\int_I \lambda^o |g_i| \leq (\Lambda \int_I \lambda^o |g_i|^2)^{1/2} < \infty$, the limiting operation can be taken inside the integrals on the right side of (14) to yield (6). This completes the proof of Lemma 1.

Proof of Lemma 2: The distribution $P(N^o)$ of the Poisson process is specified by the density (1) with the identification $\lambda^o(\underline{t}) = \lambda(\underline{t}; \underline{\tau})$. Hence, with the definition of entropy (7):

$$\begin{aligned} H(N^o) &= \Lambda - E\left[\ln \prod_{i=1}^{n^o} \lambda^o(\underline{t}_i)\right] \\ &= \Lambda - E\left[\sum_{i=1}^{n^o} \ln \lambda^o(\underline{t}_i)\right]. \quad (15) \end{aligned}$$

In (15), $\Lambda \triangleq \int_I \lambda^o(\underline{t})\, d\underline{t}$. Note that $\int_I \lambda^o(\underline{t}) |\ln \lambda^o(t)|^2\, d\underline{t} < \infty$ since λ^o is bounded and I is a finite region. Hence (5) of Lemma 1 can be applied to (15) to obtain the expression (8). This establishes Lemma 2.

III. The Estimation Problem

In what follows we focus on the problem of lower bounding the mse of an estimator of the single parameter τ_1. In this context the $L-1$ remaining parameters $\underline{\tau}_2^T = [\tau_2, \cdots, \tau_L]^T$ are regarded as nuisance parameters which confound the single parameter estimator.

A. Energy Invariance

For the derivation of the general bound in Section V, we make one major assumption concerning the dependency of

$P(N|\underline{\tau})$ on τ_1: the total process energy, i.e., the conditional rate $E[N(I)|\underline{\tau}]$, is not influenced by the value of τ_1. This property will be called *energy variance* with respect to τ_1. More specifically, it will be assumed that the following property holds.

Property 1: $\int_I \lambda(\underline{t};\underline{\tau})\,d\underline{t} = \Lambda(\underline{\tau}_2^L)$, independent of τ_1 for $\underline{\tau} \in J$.

In the applications section of this paper, Section VI, Property 1 will be invoked for the special case of shift invariance for which $\lambda(\underline{t};\underline{\tau}) = \lambda(t_1 - \tau_1, \underline{t}_2^M; 0, \underline{\tau}_2^L)$.

B. Lower Bounds on MSE

Let $\hat{\tau}_1 = \hat{\tau}_1(N)$ be an estimator of the parameter τ_1. We are interested in lower-bounding the mse, $E(\tau_1 - \hat{\tau}_1)^2$, associated with $\hat{\tau}_1$ in the presence of the nuisance parameters $\underline{\tau}_2^L$. We first discuss Cramer–Rao (CR) lower bounds for this problem.

Cramer–Rao Lower Bounds: Under some regularity assumptions [11], [12], the following bound on the mse of an estimator $\hat{\tau}_1$ of τ_1 holds:

$$\text{mse} \triangleq E(\tau_1 - \hat{\tau}_1)^2 \geq \frac{1}{E\left(\dfrac{\partial}{\partial \tau_1} \ln p_{\underline{t}_1,\cdots,\underline{t}_n,\tau_1}(\underline{t}_1,\cdots,\underline{t}_n,\tau_1)\right)^2}. \tag{16}$$

The CR bound (16) is dependent on the joint distribution of N and τ_1:

$$p_{\underline{t}_1,\cdots,\underline{t}_n,\tau_1}(\underline{t}_1,\cdots,\underline{t}_n,\tau_1) = \int \cdots \int p_{\underline{t}_1,\cdots,\underline{t}_n,\underline{\tau}}(\underline{t}_1,\cdots,\underline{t}_n,\underline{\tau}) \cdot p_{\tau_2,\cdots,\tau_L}(\tau_2,\cdots,\tau_L)\,d\tau_2\cdots d\tau_L. \tag{17}$$

A frequently more tractable lower bound on estimators of τ_1 can be obtained via the multiparameter CR bound:

$$\text{mse} = \underline{e}_1^T E(\underline{\tau} - \hat{\underline{\tau}})(\underline{\tau} - \hat{\underline{\tau}})^T \underline{e}_1 \geq \underline{e}_1^T F^{-1} \underline{e}_1 = \left[a - \underline{b}^T F_s^{-1} \underline{b}\right]^{-1} \tag{18}$$

where $\hat{\underline{\tau}}$ estimates the vector parameter $\underline{\tau}$, $\underline{e}_1 = [1,0,\cdots,0]^T$, and F is the Fisher matrix defined by:

$$F \triangleq \begin{bmatrix} a & \underline{b}^T \\ \underline{b} & F_s \end{bmatrix} = E\left(\nabla_{\underline{\tau}} \ln\{p_{\underline{t}_1,\cdots,\underline{t}_n,\underline{\tau}}(\underline{t}_1,\cdots,\underline{t}_n,\underline{\tau})\}\right)^T \cdot \left(\nabla_{\underline{\tau}} \ln\{p_{\underline{t}_1,\cdots,\underline{t}_n,\underline{\tau}}(\underline{t}_1,\cdots,\underline{t}_n,\underline{\tau})\}\right). \tag{19}$$

In (19) $\nabla_{\underline{\tau}}$ is the gradient operator ($\nabla_{\underline{\tau}} p_{\underline{t}_1,\cdots,\underline{t}_n,\underline{\tau}}$ is a row vector), and a, $\underline{b}$ and F_s are 1×1, $(L-1)\times 1$ and $(L-1)\times(L-1)$ vector quantities, respectively. The quantity a is the (positive) Fisher information associated with τ_1, F_s is the (positive definite) Fisher information matrix for the nuisance parameters, $\underline{\tau}_2^L$, and $\underline{b}$ is the Fisher information vector which couples τ_1 to the nuisance parameters.

The bound (18) is identical to the bound (16) in the absence of random nuisance parameters ($L = 1$). However, when $L > 1$ it has been shown [15] that in general (16) is tighter than (18). On the other hand, since the bound (16) involves averaging over the nuisance parameters within the argument of the natural logarithm, an explicit expression for (16) is generally more difficult to obtain.

A CR Bound for Conditionally Poisson Processes: Here we state and prove a result which specifies the form of the bound (18) for the case of observations which are conditionally Poisson given the random parameters $\underline{\tau}$.

Proposition 1: Let $N = \{dN(\underline{t}): \underline{t} \in I\}$ be an M-dimensional conditionally Poisson process, with intensity $\lambda = \{\lambda(\underline{t};\underline{\tau}): \underline{t} \in I\}$, given the L-dimensional random parameter vector $\underline{\tau} = \tau \in J$. Assume that for $i = 1,\cdots,L$ the intensity satisfies the following condition:

$$E\left(\int_I \left|\frac{\partial \ln \lambda(\underline{t};\underline{\tau})}{\partial \tau_i}\right|^2 \lambda(\underline{t};\underline{\tau})\,dt\right) < \infty \tag{20}$$

and that $\partial \ln p_{\underline{\tau}}(\underline{\tau})/\partial \tau_i$ has finite second moment. Then the Fisher matrix (19) associated with $\underline{\tau}$ has the form

$$F = \left(\left(\int_I d\underline{t}\, E\left[\frac{\partial \ln \lambda(\underline{t};\underline{\tau})}{\partial \tau_i} \frac{\partial \ln \lambda(\underline{t};\underline{\tau})}{\partial \tau_j} \lambda(\underline{t};\underline{\tau})\right] + E\left[\frac{\partial \ln p_{\underline{\tau}}(\underline{\tau})}{\partial \tau_i} \frac{\partial \ln p_{\underline{\tau}}(\underline{\tau})}{\partial \tau_j}\right]\right)\right)_{i,j}. \tag{21}$$

Proof of Proposition 1: Using the form for the conditionally Poisson distribution (1), the partial derivative factor associated with the ijth element of the Fisher matrix (19) is

$$\begin{aligned}
\frac{\partial}{\partial \tau_i} \ln p_{\underline{t}_1,\cdots,\underline{t}_n,\underline{\tau}}(\underline{t}_1,\cdots,\underline{t}_n,\underline{\tau})
&= \frac{\partial}{\partial \tau_i} \ln\{p_{\underline{t}_1,\cdots,\underline{t}_n|\underline{\tau}}(\underline{t}_1,\cdots,\underline{t}_n|\underline{\tau})\,p_{\underline{\tau}}(\underline{\tau})\} \\
&= \frac{\partial}{\partial \tau_i} \ln\left\{\left(\prod_{l=1}^{n} \lambda(\underline{t}_l;\underline{\tau})\right) e^{-\Lambda(\underline{\tau})} p_{\underline{\tau}}(\underline{\tau})\right\} \\
&= \frac{\partial}{\partial \tau_i} \sum_{l=1}^{n} \ln \lambda(\underline{t}_l;\underline{\tau}) - \frac{\partial}{\partial \tau_i}(\Lambda(\underline{\tau}) - \ln p_{\underline{\tau}}(\underline{\tau})) \\
&= \sum_{l=1}^{n} g_i(\underline{t}_l) - c_i
\end{aligned} \tag{22}$$

where we have defined

$$g_i(\underline{t}) \triangleq \frac{\partial}{\partial \tau_i} \ln \lambda(\underline{t};\underline{\tau}), \tag{23}$$

and

$$c_i \triangleq \frac{\partial}{\partial \tau_i} \Lambda(\underline{\tau}) - \frac{\partial}{\partial \tau_i} \ln p_{\underline{\tau}}(\underline{\tau}). \tag{24}$$

Substitution of (22) into (19) gives the ijth element of the

when the CR bound is nondegenerate, the CR bound is expected to be tighter than the rate-distortion bound for higher ranges of process energy.

probability 1).

The unconditional distribution of N, denoted by $P(N)$, is specified by the expectation of (1) over $\underline{\tau} \in J$:

Fisher matrix:

$$\begin{aligned} F_{ij} &= E\left(\sum_{l=1}^{n} g_i(\underline{t}_l) - c_i\right)\left(\sum_{l=1}^{n} g_j(\underline{t}_l) - c_j\right) \\ &= E\left(E\left(\sum_{l=1}^{n} g_i(\underline{t}_l) \sum_{m=1}^{n} g_j(\underline{t}_m)\middle|\underline{\tau}\right)\right) \\ &\quad - E\left(c_i E\left(\sum_{l=1}^{n} g_j(\underline{t}_l)\middle|\underline{\tau}\right)\right) - E\left(c_j E\left(\sum_{l=1}^{n} g_i(\underline{t}_l)\middle|\underline{\tau}\right)\right) \\ &\quad + E(c_i c_j). \end{aligned} \tag{25}$$

In (25) the smoothing property of conditional expectation and the functional dependence of c_i (24) on $\underline{\tau}$ have been used. Since, conditioned on $\underline{\tau}$, N is a Poisson process with intensity $\lambda(\underline{t};\underline{\tau})$ satisfying (20), (6) and (5) of Lemma 1 can be applied to the inner expectations contained in the leading term and also those contained in the middle two terms of the final equality in (25). This yields

$$\begin{aligned} F_{ij} &= E\left(\int_I g_i(\underline{t}) g_j(\underline{t}) \lambda(\underline{t};\underline{\tau})\, d\underline{t}\right) \\ &\quad + E\left(\int_I g_i(\underline{t}) \lambda(\underline{t};\tau)\, d\underline{t} \int_I g_j(\underline{t}) \lambda(\underline{t};\underline{\tau})\, d\underline{t}\right) \\ &\quad - E\left(c_i \int_I g_j(\underline{t}) \lambda(\underline{t};\underline{\tau})\, d\underline{t}\right) - E\left(c_j \int_I g_i(\underline{t}) \lambda(\underline{t};\underline{\tau})\, d\underline{t}\right) \\ &\quad + E(c_i c_j) \\ &= E\left(\int_I g_i(\underline{t}) g_j(\underline{t}) \lambda(\underline{t};\underline{\tau})\, d\underline{t}\right) \\ &\quad + E\left(\int_I g_i(\underline{t}) \lambda(\underline{t};\underline{\tau})\, d\underline{t} - c_i\right)\left(\int_I g_j(\underline{t}) \lambda(\underline{t};\underline{\tau})\, d\underline{t} - c_j\right). \end{aligned} \tag{26}$$

Next, using the definition (2) in (24):

$$\begin{aligned} c_i &= \frac{\partial}{\partial \tau_i} \int_I \lambda(\underline{t};\underline{\tau})\, d\underline{t} - \frac{\partial}{\partial \tau_i} \ln p_{\underline{\tau}}(\underline{\tau}) \\ &= \int_I \frac{\partial \ln \lambda(\underline{t};\underline{\tau})}{\partial \tau_i} \lambda(\underline{t};\underline{\tau})\, d\underline{t} - \frac{\partial \ln p_{\underline{\tau}}(\underline{\tau})}{\partial \tau_i} \\ &= \int g_i(\underline{t}) \lambda(\underline{t};\underline{\tau})\, d\underline{t} - \frac{\partial \ln p_{\underline{\tau}}(\underline{\tau})}{\partial \tau_i}. \end{aligned} \tag{27}$$

In (27) the boundedness condition (20) must be used to justify the interchange of integration and differentiation. Finally, substitution of (27) into (26) yields (21). This completes the proof of Proposition 1.

The CR bounds given by (16) and (18) have a major deficiency which has partially motivated the present work: they can become trivial bounds even for finite λ. For concreteness, consider the case $M = L = 1$, appropriate for one-dimensional time shift estimation in the absence of nuisance parameters: $\lambda(\underline{t};\underline{\tau}) = \lambda(t-\tau)$ and $\nabla_{\underline{\tau}}\lambda(\underline{t};\underline{\tau}) = \lambda'(t-\tau)$. For this case the CR bounds (16) and (18) are equivalent and can be found from (21):

$$\text{mse} \geq \frac{1}{\displaystyle\int_{-T/2}^{T/2} \frac{(\lambda'(t))^2}{\lambda(t)}\, dt + E\left(\frac{p'_\tau(\tau)}{p_\tau(\tau)}\right)^2}. \tag{28}$$

In (28) it has been assumed that the support of $\lambda(t-\tau)$ is contained in the observational interval $[-T/2, T/2]$ (with probability 1). Observe that for an intensity with sharp edges, e.g., with $|\lambda'(t)| \gg 1$ and $\lambda(t) \ll 1$ over some finite interval of time, the integral in the denominator of (28) becomes large, and the CR bound approaches the degenerate case: mse ≥ 0. This suggests that the CR bound is not tight for rapidly varying intensity functions. To overcome this severe sensitivity to small but abrupt changes in λ an approximation was developed in [5].

An Approximate Lower Bound for Rectangular λ: For the special case of one-dimensional time-shift estimation, rectangular λ:

$$\lambda(\underline{t};\underline{\tau}) = \lambda(t-\tau) = \begin{cases} \dfrac{\Lambda}{T_\lambda}, & \tau - \dfrac{T_\lambda}{2} \leq t \leq \tau + \dfrac{T_\lambda}{2}, \\ 0, & \text{other } t \end{cases} \tag{29}$$

T_λ much less than the observation time $T_1 = T$, and uniform $p_{\tau_1}(\tau_1)$, an approximation to a nontrivial lower bound was given in [5] using an information theoretic approach:

$$\text{mse} \geq \begin{cases} 0.143 T_\lambda^2 \Lambda^{-2} & \Lambda \geq 4 \\ 0.059 T_\lambda^2 e^{-\Lambda/2}, & \frac{1}{2} \leq \Lambda < 4. \\ 0.059 T_\lambda^2 e^{-\Lambda^2/2}, & 0 \leq \Lambda < \frac{1}{2} \end{cases} \tag{30}$$

In the derivation of the approximation (30) a number of technical manipulations and approximations were necessary. In what follows we will derive an exact lower bound under more general assumptions than in [5] using special properties of Poisson processes and elements of rate distortion theory.

IV. Information Theory Framework

In this section the information theoretic framework for the Poisson parameter estimation problem is presented, and the rate distortion principle is outlined. Recall the definition of the mutual information $I(X;Y)$ between two random quantities X and Y:

$$I(X;Y) \triangleq E\left[\ln \frac{dP(X,Y)}{dP(X)\,dP(Y)}\right] \tag{31}$$

where $dP(X,Y)/dP(X)\,dP(Y)$ is the likelihood ratio (Radon–Nikodym derivative) of the joint distribution $P(X,Y)$ with respect to the product distribution $P(X)P(Y)$. If P is absolutely continuous with respect to some reference measure μ, then the mutual information has the entropy decomposition:

$$I(X;Y) = H(Y) - H(Y|X) \tag{32}$$

where $H(Y) = E[-\ln dP(Y)/d\mu]$ and $H(Y|X) = E[-\ln dP(Y|X)/d\mu]$ are the unconditional and conditional entropies. We first give a brief outline of the rate-distortion theory that will be relevant to the bounding procedure.

A. Rate Distortion Theory

The starting point is Shannon theory. This gives a bound on a source's rate-distortion function $R_{\bar{\rho}}(d)$ in terms of the channel capacity C [13], [14]:

$$\inf_{P(Y|X):\,\bar{\rho}(X,Y)\le d} I(X;Y) = R_{\bar{\rho}}(d) \le C = \sup_{p_X} I(X;Y). \tag{33}$$

In (33), $I(X;Y)$ is the mutual information between the source symbol X and the destination symbol Y, p_X is the source distribution, $P(Y|X)$ are the forward transition probabilities associated with the channel, and $\bar{\rho}(X,Y)$ is an average distortion metric between X and Y. In the present context the source symbol X is identified with the random variable τ_1, the destination symbol Y is identified with the observation N, and $\bar{\rho}(X;Y)$ is the mse

$$\bar{\rho}(\tau_1;N) \triangleq \text{mse} = E(\tau_1 - \hat{\tau}_1(N))^2. \tag{34}$$

For a general source distribution p_{τ_1}, the rate-distortion function $R_{\bar{\rho}}$ is computable from a parametric formula [13]. It is known that $R_{\bar{\rho}}(d)$ is continuous and strictly monotone decreasing as a function of d over $d \le d_{\max}$ and constant for $d > d_{\max}$. Furthermore, the infimum of the mutual information over $\bar{\rho} \le d$ is achieved for $\bar{\rho} = d$. Hence we have the bound on mse:

$$\begin{aligned} &\text{mse} \ge d \ge R^{-1}_{\text{mse}}(C), \qquad d \le d_{\max} \\ &\text{mse} \ge d_{\max}, \qquad d > d_{\max}. \end{aligned} \tag{35}$$

The bound (35) is not of analytical form for general p_{τ_1}. Weaker lower bounds are obtained in the following proposition via the Shannon lower bound on R_{mse} given in (36) [13], and by adjoining the "side information" $\underline{\tau}_2^L$ to the destination symbols and averaging the resultant mse bound over $\underline{\tau}_2^L$ given (37).

Proposition 2: The following bounds on the mse of an estimator $\hat{\underline{\tau}}_1$ hold:

$$\text{mse} \ge \frac{1}{2\pi e} e^{2H(\tau_1)} e^{-2C} \tag{36}$$

$$\text{mse} \ge \frac{1}{2\pi e} e^{2H(\tau_1|\underline{\tau}_2^L)} e^{-2C^c}. \tag{37}$$

In (36) and (37),

$$C^c \triangleq \sup_{p_{\tau_1|\underline{\tau}}} \left\{ H(N|\underline{\tau}_2^L) - H(N|\underline{\tau}) \right\} \tag{38}$$

and $H(\tau_1|\underline{\tau}_2^L) \triangleq E[-\ln p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)]$ is the conditional differential entropy of τ_1 given $\underline{\tau}_2^L$.

Proof: We first show (36). Fix the source distribution equal to p_{τ_1}. Then the Shannon inequality (33) becomes

$$\inf_{P(N|\tau_1):\,\bar{\rho}(\tau_1,N)\le d} I(\tau_1;N) = R_{\text{mse}}(d) \le C = \sup_{p_{\tau_1}} I(\tau_1;N). \tag{39}$$

Next we recall the Shannon lower bound on R_L on the rate distortion function $R_{\bar{\rho}}$ which is derived for $\bar{\rho}$ equal to mse in [13]:

$$R_{\text{mse}}(d) \ge R_L(d) \triangleq E\left[-\ln p_{\tau_1}(\tau_1)\right] - \frac{1}{2}\ln(2\pi e d). \tag{40}$$

Since the mse can be no less than d for a source of rate $R_{\text{mse}}(d)$ [13] we obtain the first inequality (36) by application of the inverse of R_L to the Shannon inequality $R_L \le C$ as in (35).

The second inequality (37) is obtained by finding the rate distortion bound on the conditional mean squared error $\hat{\tau}_1$ given $\underline{\tau}_2^L$ and then averaging the result over these parameters. In this case the relevant source distribution is given by the conditional probability density $p_{\tau_1|\underline{\tau}_2^L}$ and, in view of (32), the capacity of the channel is the quantity C^c given in (38). Hence, using $E[-\ln p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)|\underline{\tau}_2^L]$ in place of $H(\tau_1)$ in (40) we obtain the inequality

$$R_{\text{mse}|\underline{\tau}_2^L}(d) \ge R_L^c(d) \triangleq E\left[-\ln p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)|\underline{\tau}_2^L\right] - \frac{1}{2}\ln(2\pi e d) \tag{41}$$

or, equivalently, the conditional expectation of the squared error of $\hat{\tau}_1$ given $\underline{\tau}_2^L$ has the lower bound:

$$\begin{aligned} E&\left[(\tau_1 - \hat{\tau}_1)^2 [\underline{\tau}_2^L\right] \ge d \\ &\ge \frac{1}{2\pi e}\exp\left(2\left[E\left[-\ln p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)|\underline{\tau}_2^L\right]\right.\right. \\ &\qquad \left.\left. - \sup_{p_{\tau_1|\underline{\tau}_2^L}} E\left[\frac{dP(N|\underline{\tau})}{dP(N|\underline{\tau}_2^L)}\bigg|\underline{\tau}_2^L\right]\right]\right). \end{aligned} \tag{42}$$

Applying the smoothing property of conditional expectation, $E[X] = E[E[X|\underline{\tau}_2^L]]$ and the following special case of Jensen's inequality: $E[\exp X] \ge \exp E[X]$, to (42):

$$\begin{aligned} \text{mse} &= E\left[E\left[(\tau_1 - \hat{\tau}_1)^2|\underline{\tau}_2^L\right]\right] \\ &\ge \frac{1}{2\pi e}\exp\left(2\left[E\left[-\ln p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\tau_2^L)\right]\right.\right. \\ &\qquad \left.\left. - E\sup_{p_{\tau_1|\underline{\tau}_2^L}} E\left[\frac{dP(N|\underline{\tau})}{dP(N|\underline{\tau}_2^L)}\bigg|\underline{\tau}_2^L\right]\right]\right). \end{aligned} \tag{43}$$

Since the outer expectation in the second term of the exponent in (43) is with respect to the joint pdf of $\underline{\tau}_2^L$, it is independent of the conditional pdf $\bar{p}_{\tau_1|\underline{\tau}_2^L}$. Hence the order of the "E" and "sup" operations can be interchanged, giving (37). This completes the proof of Proposition 2.

Observe that Proposition 2 gives lower bounds that involve two major factors: the first is independent of the observation statistics, involving only the source entropy $H(\tau_1)$, or conditional entropy $H(\tau_1|\underline{\tau}_2^L)$, and the second is independent of the source statistics, involving the capacity C, or conditional capacity C^c. It can be shown that when τ_1 is independent of $\underline{\tau}_2^L$ the two bounds (36) and (37) are equivalent. For the general case it appears difficult to establish any dominance conditions of one bound relative to the other. This is despite the fact that (37) is calculated via introduction of the side information $\underline{\tau}_2^L$ at the receiver. On the other hand, it is straightforward to show that

$$\text{bound}|_{\text{no side information}} \geq \text{bound}|_{\text{side information}} \cdot \exp\left[I\left(\tau_1; \underline{\tau}_2^L\right) - \sup_{p_{\tau_1|\underline{\tau}_2^L}} I\left(\tau_1; \underline{\tau}_2^L\right)\right],$$

so that as the mutual information between τ_1 and $\underline{\tau}_2^L$ approaches its maximum value, (36) becomes at least as tight as (37).

B. An Upper Bound on Channel Capacity

To evaluate the rate-distortion lower bounds presented in Proposition 2, one must be able to compute C or C^c. Since the computation of C involves the difficult joint marginal distribution of N and τ_1, we will concentrate on the latter. For the present estimation problem, C^c is obtained by maximizing the information over positive normed functions $p_{\tau_1|\underline{\tau}_2^L}$. This functional maximization is quite difficult in general. In what follows we will obtain a tractable but weaker rate-distortion bound by upper bounding C^c. The bound is obtained by upper-bounding the conditional entropy $H(N|\underline{\tau}_2^L)$ by $H(N^o|\underline{\tau}_2^L)$, whence

$$C^c \leq \sup_{p_{\tau_1|\underline{\tau}_2^L}} \left\{H\left(N^o|\underline{\tau}_2^L\right) - H\left(N|\underline{\tau}\right)\right\} \tag{44}$$

where N^o is an inhomogeneous Poisson process. Specifically, using the energy invariance property, Property 1, we will show that a Poisson process model for N, unconditioned on τ_1, maximizes the conditional entropy. We give this maximum entropy property in the following proposition.

Proposition 3: Let $N = \{dN(\underline{t})\colon \underline{t} \in I\}$ be an M-dimensional conditionally Poisson process, with almost everywhere bounded intensity $\lambda = \{\lambda(\underline{t}; \underline{\tau})\colon \underline{t} \in I\}$, given the L-dimensional random parameter vector $\underline{\tau} = \underline{\tau} \in J$. Assume that N is energy invariant in the sense that the function λ satisfies Property 1 with respect to τ_1. Then the conditional differential entropy of N given $\underline{\tau}_2^L$ is always inferior to the conditional differential entropy of a Poisson process with intensity $E[\lambda(\underline{t}; \underline{\tau})|\underline{\tau}_2^L]$. Specifically,

$$H\left(N|\underline{\tau}_2^L\right) \leq E\left[\Lambda\left(\underline{\tau}_2^L\right)\right] - E\left[\int_I \bar{\lambda}\left(\underline{t}; \underline{\tau}_2^L\right) \ln \bar{\lambda}\left(\underline{t}; \underline{\tau}_2^L\right) d\underline{t}\right], \tag{45}$$

where

$$\bar{\lambda}\left(\underline{t}; \underline{\tau}_2^L\right) \triangleq E\left[\lambda\left(\underline{t}; \underline{\tau}\right)|\underline{\tau}_2^L\right] = \int_{J_1} \lambda\left(\underline{t}; \tau_1, \underline{\tau}_2^L\right) p_{\tau_1|\underline{\tau}_2^L}\left(\tau_1|\underline{\tau}_2^L\right) d\tau_1 \tag{46}$$

and J_1 is the support of p_{τ_1}.

Proof: The Proposition will be proven by showing that, conditioned on the nuisance parameters $\underline{\tau}_2^L$ and the total number of points of N over I, $\boldsymbol{n}$, the difference between the entropy of a Poisson process and the entropy of N is identically the mutual information of the population of points, $\{\underline{t}_{(1)}, \cdots, \underline{t}_{(n)}\}$, which is nonnegative.

Since N is conditionally Poisson, the form (1) for the conditionally Poisson process distribution can be used:

$$\begin{aligned}
H&\left(N|\underline{\tau}_2^L\right)\\
&= E\left[-\ln p_{\underline{t}_1,\cdots,\underline{t}_n|\underline{\tau}_2^L}\left(\underline{t}_1, \cdots, \underline{t}_n|\underline{\tau}_2^L\right)\right]\\
&= E\left[-\ln \int_{J_1} d\tau_1\, p_{\underline{t}_1,\cdots,\underline{t}_n|\underline{\tau}}\left(\underline{t}_1, \cdots, \underline{t}_n|\tau_1, \underline{\tau}_2^L\right) p_{\tau_1|\underline{\tau}_2^L}\left(\tau_1|\underline{\tau}_2^L\right)\right]\\
&= E\left[-\ln E\left[e^{-\int_I \lambda(\underline{t}, \underline{\tau})\, d\underline{t}}|\underline{\tau}_2^L\right]\Big|\boldsymbol{n} = 0\right] P(\boldsymbol{n} = 0) \qquad (47)\\
&\quad + E\left[-\ln\left(\int_{J_1} d\tau_1\, p_{\tau_1|\underline{\tau}_2^L}\left(\tau_1|\underline{\tau}_2^L\right) e^{-\int_I \lambda(\underline{t}, \underline{\tau})\, d\underline{t}} \cdot \prod_{i=1}^{n} \lambda\left(\underline{t}_i; \tau_1, \underline{\tau}_2^L\right)\right)\Bigg|\boldsymbol{n} > 0\right] P(\boldsymbol{n} > 0)\\
&= E\left[\Lambda\left(\underline{\tau}_2^L\right)\right] + E\left[-\ln\left(\int_{J_1} d\tau_1\, p_{\tau_1|\underline{\tau}_2^L}\left(\tau_1|\underline{\tau}_2^L\right) \cdot \prod_{i=1}^{n} \lambda\left(\underline{t}_i; \tau_1, \underline{\tau}_2^L\right)\right)\right]. \qquad (48)
\end{aligned}$$

Equality (47) is simply the representation of the outer expectation by an iterated conditional expectation given the events $\boldsymbol{n} = 0$ and its complement $\boldsymbol{n} > 0$. To obtain (48), Property 1 has been used to move the integrated intensity $\int_I \lambda(\underline{t}; \underline{\tau})\, d\underline{t} = \Lambda(\underline{\tau}_2^L)$ outside of the expectation operator over τ_1. In (48) the convention is that $\prod_{i=1}^{n} = 1$ for $n = 0$.

Consider the second expectation in (48):

$$\begin{aligned}
&E\left[-\ln\left(\int_{J_1} d\tau_1\, p_{\tau_1|\underline{\tau}_2^L}\left(\tau_1|\underline{\tau}_2^L\right) \prod_{i=1}^{n} \lambda\left(\underline{t}_i; \tau_1, \underline{\tau}_2^L\right)\right)\right] &(49)\\
&= E\left[-\ln \prod_{i=1}^{n}\left(\int_{J_1} d\tau_1\, \lambda\left(\underline{t}_i; \tau_1, \underline{\tau}_2^L\right) p_{\tau_1|\underline{\tau}_2^L}\left(\tau_1|\underline{\tau}_2^L\right)\right)\right] &(50)\\
&\quad - E\left[\ln \frac{\int_{J_1} d\tau_1\, p_{\tau_1|\underline{\tau}_2^L}\left(\tau_1|\underline{\tau}_2^L\right) \prod_{i=1}^{n} \lambda\left(\underline{t}_i; \tau_1, \underline{\tau}_2^L\right)}{\prod_{i=1}^{n} \int_{J_1} d\tau_1\, \lambda\left(\underline{t}_i; \tau_1, \underline{\tau}_2^L\right) p_{\tau_1|\underline{\tau}_2^L}\left(\tau_1|\underline{\tau}_2^L\right)}\right]. &(51)
\end{aligned}$$

The term (50) is the second term on the right of the inequality (45):

$$
\begin{aligned}
&E\left[-\ln\prod_{i=1}^{n}\left(\int_{J_1} d\tau_1\,\lambda(\underline{t}_i;\tau_1,\underline{\tau}_2^L)\,p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)\right)\right]\\
&\quad=E\left[-\sum_{i=1}^{n}\ln\bar{\lambda}(\underline{t}_i;\underline{\tau}_2^L)\right]\\
&\quad=E\left[E\left[-\sum_{i=1}^{n}\ln\bar{\lambda}(\underline{t}_i;\underline{\tau}_2^L)|\underline{\tau}\right]\right] &(52)\\
&\quad=E\left[-\int_I\lambda(\underline{t};\underline{\tau})\ln\bar{\lambda}(\underline{t};\underline{\tau}_2^L)\,d\underline{t}\right] &(53)\\
&\quad=E\left[-\int_I E\left[\lambda(\underline{t};\underline{\tau})|\underline{\tau}_2^L\right]\ln\bar{\lambda}(\underline{t};\underline{\tau}_2^L)\,d\underline{t}\right]\\
&\quad=E\left[-\int_I\bar{\lambda}(\underline{t};\underline{\tau}_2^L)\ln\bar{\lambda}(\underline{t};\underline{\tau}_2^L)\,d\underline{t}\right]. &(54)
\end{aligned}
$$

To go from (52) to (53), Lemma 1 has been applied to compute the expectation of the sum of Poisson shifted functions, $\Sigma\ln\bar{\lambda}(\underline{t}_i;\underline{\tau}_2^L)$. In view of the identities (48), (49), and (54), we need only show that the term (51) is nonnegative to establish the inequality (45). By the smoothing property of conditional expectation, we obtain the equivalent expression for (51):

$$
E\left[\ln\frac{\int_{J_1}d\tau_1\,p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)\prod_{i=1}^{n}\lambda(\underline{t}_i;\tau_1,\underline{\tau}_2^L)}{\prod_{i=1}^{n}\int_{J_1}d\tau_1\,\lambda(\underline{t}_i;\tau_1,\underline{\tau}_2^L)\,p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)}\right]=E\left[E\left[\ln\frac{\int_{J_1}d\tau_1\,p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)\prod_{i=1}^{n}\lambda(\underline{t}_i;\tau_1,\underline{\tau}_2^L)}{\prod_{i=1}^{n}\int_{J_1}d\tau_1\,\lambda(\underline{t}_i;\tau_1,\underline{\tau}_2^L)\,p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)}\middle|\boldsymbol{n},\underline{\tau}_2^L\right]\right]. \tag{55}
$$

Recall the following facts from Poisson process theory [9]. Let N^o be an inhomogeneous Poisson process with intensity λ^o having an integral $\Lambda^o\triangleq\int_I\lambda^o$. Then, conditioned on the total number of points $\boldsymbol{n}^o=n>0$, we have 1) the joint density of the (unordered) population of Poisson points $\underline{t}_{(1)},\cdots,\underline{t}_{(n)}$ is given by

$$
p_{\underline{t}_{(1)},\cdots,\underline{t}_{(n^o)}|\boldsymbol{n}^o}(\underline{t}_1,\cdots,\underline{t}_n|n)=\prod_{i=1}^{n}\frac{\lambda^o(\underline{t}_i)}{\Lambda^o} \tag{56}
$$

and 2) the marginal density of $\underline{t}_{(i)}$ is given by

$$
p_{\underline{t}_{(i)}|\boldsymbol{n}^o}(\underline{t}_i|n)=\frac{\lambda^o(\underline{t}_i)}{\Lambda^o}. \tag{57}
$$

Let $\boldsymbol{n}=n>0$ and $\underline{\tau}_2^L=\underline{\tau}_2^L$ be fixed. Then, since N is conditionally Poisson, from (56) the conditional joint density of the unordered points $\underline{t}_{(1)},\cdots,\underline{t}_{(n)}$ is

$$
\begin{aligned}
&p_{\underline{t}_{(1)},\cdots,\underline{t}_{(n)}|\boldsymbol{n},\underline{\tau}_2^L}(\underline{t}_1,\cdots,\underline{t}_n|n,\underline{\tau}_2^L)\\
&\quad=\int_{J_1}d\tau_1\prod_{i=1}^{n}\frac{\lambda(\underline{t}_i;\tau_1,\underline{\tau}_2^L)}{\Lambda(\underline{\tau}_2^L)}p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L). &(58)
\end{aligned}
$$

On the other hand, the marginal density of any one of the points $\underline{t}_{(i)}$ is

$$
p_{\underline{t}_{(i)}|\boldsymbol{n},\underline{\tau}_2^L}(\underline{t}_i|n,\underline{\tau}_2^L)=\int_{J_1}d\tau_1\frac{\lambda(\underline{t}_i;\tau_1,\underline{\tau}_2^L)}{\Lambda(\underline{\tau}_2^L)}p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L). \tag{59}
$$

In view of (55), (58), (59), and the τ_1 independence of $\Lambda(\underline{\tau}_2^L)$, recognize the inner expectation in (55) as the quantity:

$$
\begin{aligned}
&E\left[\ln\frac{\int_{J_1}d\tau_1\,p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)\prod_{i=1}^{n}\lambda(\underline{t}_i;\tau_1,\underline{\tau}_2^L)}{\prod_{i=1}^{n}\int_{J_1}d\tau_1\,\lambda(\underline{t}_i;\tau_1,\underline{\tau}_2^L)\,p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)}\middle|\boldsymbol{n},\underline{\tau}_2^L\right]\\
&\quad=E\left[\ln\frac{\int_{J_1}d\tau_1\,p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)\prod_{i=1}^{n}\frac{\lambda(\underline{t}_i;\tau_1,\underline{\tau}_2^L)}{\Lambda(\underline{\tau}_2^L)}}{\prod_{i=1}^{n}\int_{J_1}d\tau_1\frac{\lambda(\underline{t}_i;\tau_1,\underline{\tau}_2^L)}{\Lambda(\underline{\tau}_2^L)}p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)}\middle|\boldsymbol{n},\underline{\tau}_2^L\right]\\
&\quad=E\left[\ln\frac{p_{\underline{t}_{(1)},\cdots,\underline{t}_{(n)}|\boldsymbol{n},\underline{\tau}_2^L}(\underline{t}_1,\cdots,\underline{t}_n|\boldsymbol{n},\underline{\tau}_2^L)}{\prod_{i=1}^{n}p_{\underline{t}_{(i)}|\boldsymbol{n},\underline{\tau}_2^L}(\underline{t}_i|\boldsymbol{n},\underline{\tau}_2^L)}\middle|\boldsymbol{n},\underline{\tau}_2^L\right]\\
&\quad=I(\underline{t}_{(1)},\cdots,\underline{t}_{(n)}|\boldsymbol{n},\underline{\tau}_2^L) &(60)
\end{aligned}
$$

where $I(\underline{t}_{(1)},\cdots,\underline{t}_{(n)}|\boldsymbol{n},\underline{\tau}_2^L)$ is the conditional mutual information of the population of points $\underline{t}_{(1)},\cdots,\underline{t}_{(n)}$ given $\boldsymbol{n}$ and $\underline{\tau}_2^L$. The nonnegativity of the mutual information [14] establishes Proposition 3.

The combination of Proposition 3 and (38) gives the lemma.

Lemma 3: Under the hypotheses of Proposition 3 we have

$$
C^c\le C^*\triangleq\sup_{p_{\tau_1|\underline{\tau}_2^L}}E\left[\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau})\cdot\ln\frac{\lambda(\underline{t};\underline{\tau})}{\int_{J_1}\lambda(\underline{t};\tau_1,\underline{\tau}_2^L)\,p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L)\,d\tau_1}\right]. \tag{61}
$$

Furthermore, the argument of the "sup," $E[\ ]$, in (61) is a concave function of the conditional density $p_{\tau_1|\underline{\tau}_2^L}$.

Proof: Recall the identity (38)

$$C^c = \sup p_{\tau_1|\underline{\tau}_2^L}\left\{ H\left(N|\underline{\tau}_2^L\right) - H\left(N|\underline{\tau}\right)\right\}.$$

Since N is conditionally a Poisson process and is energy invariant with respect to τ_1, we have from Lemma 2.

$$\begin{aligned} H(N|\underline{\tau}) &= E\left[E\left[-\ln p_{\underline{t}_1,\cdots,\underline{t}_n|\underline{\tau}}(\underline{t}_1,\cdots,\underline{t}_n|\underline{\tau})|\underline{\tau}\right]\right] \\ &= E\left[\Lambda(\underline{\tau}) - \int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau})\ln\lambda(\underline{t};\underline{\tau})\right] \\ &= E\left[\Lambda\left(\underline{\tau}_2^L\right)\right] - E\left[\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau})\ln\lambda(\underline{t};\underline{\tau})\right]. \qquad (62)\end{aligned}$$

In (62) Lemma 1 has been applied to the inner expectation. Hence, from (62) and Proposition 3, we have the upper bound

$$\begin{aligned} C^c &= \sup_{p_{\tau_1|\underline{\tau}_2^L}}\left\{H\left(N|\underline{\tau}_2^L\right)\right. \\ &\qquad \left. + E\left[\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau})\ln\lambda(\underline{t};\underline{\tau})\right] - E\left[\Lambda\left(\underline{\tau}_2^L\right)\right]\right\} \\ &\le \sup_{p_{\tau_1|\underline{\tau}_2^L}}\left\{E\left[-\int_I d\underline{t}\,\bar{\lambda}\left(\underline{t};\underline{\tau}_2^L\right)\ln\bar{\lambda}\left(\underline{t};\underline{\tau}_2^L\right)\right]\right. \\ &\qquad \left. + E\left[\int_I d\underline{t}\,\lambda(\underline{t};\tau)\ln\lambda(\underline{t};\underline{\tau})\right]\right\} \\ &= \sup_{p_{\tau_1|\underline{\tau}_2^L}}\left\{E\left[\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau})\left[-\ln\bar{\lambda}\left(\underline{t};\underline{\tau}_2^L\right) + \ln\lambda(\underline{t};\underline{\tau})\right]\right]\right. . \end{aligned} \qquad (63)$$

which is equivalent to (61).

To establish concavity of the expectation in (61) as a function of $p_{\tau_1|\underline{\tau}_2^L}$, it is sufficient to show that the following function L is concave:

$$\begin{aligned} L\left(p_{\tau_1|\underline{\tau}_2^L}\right) &\triangleq E\left[\int_I d\underline{t}\,\lambda(\underline{t};\tau) \cdot \ln \frac{\lambda(\underline{t};\tau)}{\int_{J_1}\lambda\left(\underline{t};\tau_1,\underline{\tau}_2^L\right)p_{\tau_1|\underline{\tau}_2^L}\left(\tau_1|\underline{\tau}_2^L\right)d\tau_1}\Bigg|_{\underline{\tau}_2^L = \underline{\tau}_2^L}\right] \\ &= \int_{J_1} d\tau_1\, p_{\tau_1|\underline{\tau}_2^L}\left(\tau_1|\underline{\tau}_2^L\right)\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau}) \\ &\qquad \cdot \ln\frac{\lambda(\underline{t};\underline{\tau})}{\int_{J_1}\lambda\left(\underline{t};\tau_1,\underline{\tau}_2^L\right)p_{\tau_1|\underline{\tau}_2^L}\left(\tau_1|\underline{\tau}_2^L\right)d\tau_1}. \end{aligned} \qquad (64)$$

Let $p_{\tau_1|\underline{\tau}_2^L} = p^* = \alpha p_1 + (1-\alpha)p_2$, $0 \le \alpha \le 1$, and p_1, p_2 are density functions of τ_1. Then using the inequality $\ln x \le x - 1$:

$$\begin{aligned} &L(p^*) - \alpha L(p_1) - (1-\alpha)L(p_2) \\ &\quad = -\alpha\int_{J_1} d\tau_1\, p_1(\tau_1)\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau}) \\ &\qquad \cdot\ln\frac{\int_{J_1} d\tau\,\lambda\left(\underline{t};\tau,\underline{\tau}_2^L\right)p^*(\tau)}{\int_{J_1} d\tau\,\lambda\left(\underline{t};\tau,\underline{\tau}_2^L\right)p_1(\tau)} \\ &\qquad -(1-\alpha)\int_{J_1} d\tau_1\, p_2(\tau_1)\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau}) \\ &\qquad \cdot\ln\frac{\int_{J_1} d\tau\,\lambda\left(\underline{t};\tau,\underline{\tau}_2^L\right)p^*(\tau)}{\int_{J_1} d\tau\,\lambda\left(\underline{t};\tau,\underline{\tau}_2^L\right)p_2(\tau)} \end{aligned} \qquad (65)$$

$$\begin{aligned} &\quad \ge -\alpha\int_{J_1} d\tau_1\left(p_1(\tau_1) - p^*(\tau_1)\right)\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau}) \\ &\qquad -(1-\alpha)\int_{J_1} d\tau_1\left(p_2(\tau_1) - p^*(\tau_1)\right)\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau}) \\ &\quad = -\Lambda\left(\underline{\tau}_2^L\right)\left[\alpha\int_{J_1} d\tau_1\left(p_1(\tau_1) - p^*(\tau_1)\right)\right. \\ &\qquad \left. +(1-\alpha)\int_{J_1} d\tau_1\left(p_2(\tau_1) - p^*(\tau_1)\right)\right] = 0. \end{aligned} \qquad (66)$$

Hence L is concave. To obtain the equality (66), the fact that Λ is independent of τ_1, Property 1, has been used. This completes the proof of Lemma 3.

The next result deals with the maximizing conditional probability density $p_{\tau_1|\underline{\tau}_2^L}$ *of Lemma* 3.

Lemma 4: Let $N = \{dN(\underline{t})\colon \underline{t} \in I\}$ be an M-dimensional conditionally Poisson process, with almost everywhere bounded intensity $\lambda = \{\lambda(\underline{t};\underline{\tau})\colon \underline{t} \in I\}$, given the L-dimensional random parameter vector $\underline{\tau} = \underline{\tau} \in J$. Assume that N is energy invariant in the sense that the function λ satisfies Property 1 with respect to τ_1. Then with C^* as defined in Lemma 3:

$$C^* = E\left[\int_{J_1} d\tau_1\, p^o\left(\tau_1;\underline{\tau}_2^L\right)\int_I d\underline{t}\,\lambda\left(\underline{t};\tau_1,\underline{\tau}_2^L\right)\ln\frac{\lambda\left(\underline{t};\tau_1,\underline{\tau}_2^L\right)}{\bar{\lambda}^o\left(\underline{t};\underline{\tau}_2^L\right)}\right] \qquad (67)$$

$$\bar{\lambda}^o\left(\underline{t};\underline{\tau}_2^L\right) \triangleq \int_{J_1} p^o\left(\tau_1;\underline{\tau}_2^L\right)\lambda(\underline{t};\underline{\tau})\,d\tau_1 \qquad (68)$$

where, in (68), p^o is a (density) function of τ_1 which satisfies for all $\tau_1 \in J_1$:

$$\int_I d\underline{t}\, \lambda(\underline{t}; \tau_1, \underline{\tau}_2^L) \ln \frac{\lambda(\underline{t}; \tau_1, \underline{\tau}_2^L)}{\int_{J_1} d\tau\, \lambda(\underline{t}; \tau, \underline{\tau}_2^L) p^o(\tau; \underline{\tau}_2^L)} \cdot \begin{cases} = c(\underline{\tau}_2^L), & p^o(\tau_1; \underline{\tau}_2^L) > 0 \\ \le c(\underline{\tau}_2^L), & p^o(\tau_1; \underline{\tau}_2^L) = 0 \end{cases} \tag{69}$$

and

$$\int_{J_1} p^o(\tau_1; \underline{\tau}_2^L)\, d\tau_1 = 1. \tag{70}$$

In (69) $c(\underline{\tau}_2^L)$ is a constant independent of τ_1.

Proof: Lemma 4 is established by the Kuhn–Tucker theory of constrained maximization [16]. We start from the identity in Lemma 3 (see (61)), re-expressed to bring out explicit dependence on $p_{\tau_1|\underline{\tau}_2^L}$:

$$C^* = \int_{J_2} d\tau_2 \cdots \int_{J_L} d\tau_L\, p_{\underline{\tau}_2^L}(\underline{\tau}_2^L) \cdot \sup_{p_{\tau_1|\underline{\tau}_2^L}} \int_{J_1} d\tau_1\, p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L) \int_I d\underline{t}\, \lambda(\underline{t}; \underline{\tau}) \cdot \ln \frac{\lambda(\underline{t}; \underline{\tau})}{\int_{J_1} \lambda(\underline{t}; \tau, \underline{\tau}_2^L) p_{\tau_1|\underline{\tau}_2^L}(\tau|\underline{\tau}_2^L)\, d\tau}. \tag{71}$$

In (71) the inner iterated integral over $J_1 \times I$ is a concave function of $p_{\tau_1|\underline{\tau}_2^L}$, by Lemma 3, which is to be maximized over the convex, closed domain:

$$\mathscr{D} \triangleq \left\{ p:\ p(\tau) \ge 0;\ \forall \tau, \text{ and } \int_{J_1} p(\tau_1)\, d\tau_1 = 1 \right\}. \tag{72}$$

The objective functional to be maximized is the continuous function:

$$J(g) \triangleq \int_{J_1} d\tau_1\, g(\tau_1; \underline{\tau}_2^L) \int_I d\underline{t}\, \lambda(\underline{t}; \underline{\tau}) \cdot \ln \frac{\lambda(\underline{t}; \underline{\tau})}{\int_{J_1} d\tau\, \lambda(\underline{t}; \tau, \underline{\tau}_2^L) g(\tau; \underline{\tau}_2^L)} + \eta \left[\int_{J_1} d\tau_1\, g(\tau_1; \underline{\tau}_2^L) - 1 \right]. \tag{73}$$

In (73) the unit normalization constraint (70) on the τ_1 density g has been introduced via the undetermined multiplier η. Let ri$\{\mathscr{D}\}$ be the relative interior of $\mathscr{D}$, and let $\partial\mathscr{D}$ be the boundary of $\mathscr{D}$. Since the domain (72) is closed, a maximizing $p_{\tau_1|\underline{\tau}_2^L}$ must exist in $\mathscr{D}$. Furthermore, by the concavity and continuity of J (73), the Kuhn–Tucker theorem gives the following necessary and sufficient conditions for the maximizing density $p_{\tau_1|\underline{\tau}_2^L} = p^o \in \mathscr{D}$:

$$\left. \frac{\partial J(p^o + \epsilon \delta g)}{\delta \epsilon} \right|_{\epsilon = 0} \begin{cases} = 0, & p^o \in \text{ri}\{\mathscr{D}\} \\ \le 0, & p^o \in \partial\mathscr{D} \end{cases}. \tag{74}$$

In (74) δg is an arbitrary integrable function such that $p^o + \epsilon \delta g \in \mathscr{D}$, i.e., the mixture remains a density over $\tau_1 \in J_1$. The derivative is given by

$$\begin{aligned} \left. \frac{\partial J(p^o + \epsilon \delta g)}{\partial \epsilon} \right|_{\epsilon = 0} &= \int_{J_1} d\tau_1\, \delta g(\tau_1; \underline{\tau}_2^L) \int_I d\underline{t}\, \lambda(\underline{t}; \underline{\tau}) \cdot \ln \frac{\lambda(\underline{t}; \underline{\tau})}{\int_{J_1} d\tau\, \lambda(\underline{t}; \tau, \underline{\tau}_2^L) p^o(\tau; \underline{\tau}_2^L)} \\ &\quad - \int_{J_1} d\tau_1\, p^o(\tau_1; \underline{\tau}_2^L) \int_I d\underline{t}\, \lambda(\underline{t}; \underline{\tau}) \cdot \frac{\int_{J_1} d\tau\, \delta g(\tau; \underline{\tau}_2^L) \lambda(\underline{t}; \tau, \underline{\tau}_2^L)}{\int_{J_1} d\tau\, \lambda(\underline{t}; \tau, \underline{\tau}_2^L) p^o(\tau; \underline{\tau}_2^L)} \\ &\quad + \eta \int_{J_1} d\tau_1\, \delta g(\tau_1; \underline{\tau}_2^L). \end{aligned} \tag{75}$$

Note that the second iterated integral on the right of (75) integrates to the quantity: $\int_{J_1} d\tau_1\, \delta g \int_I d\underline{t}\, \lambda$. Hence, collecting terms in (75), we get the equivalent expression to the right side of (75)

$$= \int_{J_1} d\tau_1\, \delta g(\tau_1; \underline{\tau}_2^L) \cdot \left\{ \int_I d\underline{t}\, \lambda(\underline{t}; \underline{\tau}) \ln \frac{\lambda(\underline{t}; \underline{\tau})}{\int_{J_1} d\tau\, \lambda(\underline{t}; \tau, \underline{\tau}_2^L) p^o(\tau; \underline{\tau}_2^L)} - \int_I d\underline{t}\, \lambda(\underline{t}; \underline{\tau}) + \eta \right\}. \tag{76}$$

Finally, since by Property 1 the second additive term in (76) is the τ_1-independent term $\Lambda(\underline{\tau}_2^L)$,

$$\left. \frac{\partial J(p^o + \epsilon \delta g)}{\partial \epsilon} \right|_{\epsilon = 0} = \int_{J_1} d\tau_1\, \delta g(\tau_1; \underline{\tau}_2^L) \left\{ \int_J d\underline{t}\, \lambda(\underline{t}; \underline{\tau}) \cdot \ln \frac{\lambda(\underline{t}; \underline{\tau})}{\int_{J_1} d\tau\, \lambda(\underline{t}; \tau, \underline{\tau}_2^L) p^o(\tau; \underline{\tau}_2^L)} - c(\underline{\tau}_2^L) \right\} \tag{77}$$

where $c(\underline{\tau}_2^L) \triangleq \Lambda(\underline{\tau}_2^L) - \eta$. In the present case it is clear that p^o is on the boundary of $\mathscr{D}$ (72) if and only if $p^o(\tau_1) \triangleq p^o(\tau_1; \underline{\tau}_2^L) = 0$ for some $\tau_1 \in J_1$. By the Kuhn–Tucker condition (74), (77) must be equal to zero for those values of τ_1 where $p^o(\tau_1) > 0$ and less than or equal to zero for those τ_1 for which $p^o(\tau_1)=0$. This implies that the integrand of (77) must be identically zero if $p^o(\tau_1)>0$. On the other hand, if $p^o(\tau_1)=0$, the perturbation $\delta g(\tau_1)$ must be positive so that $p^o+\epsilon\delta g\in\mathscr{D}$, which implies that the integrand must be less than or equal to zero for this case. This completes the proof of Lemma 4.

V. Main Theorems

In this section the results obtained in the previous section are combined to yield the main results of this paper.

A. A General Bound

Theorem 1: Let $N=\{dN(\underline{t}):\ \underline{t}\in I\}$ be an M-dimensional conditionally Poisson process, with almost everywhere bounded intensity $\lambda=\{\lambda(\underline{t};\underline{\tau}):\ \underline{t}\in I\}$, given the L-dimensional random parameter vector $\underline{\tau}=\underline{\tau}\in J$. Assume that N is energy invariant in the sense that the function λ satisfies Property 1 with respect to τ_1. Then the mse of any estimator $\hat{\tau}_1$ of τ_1 has the lower bound

$$\text{mse}\geq\frac{1}{2\pi e}e^{2H(\tau_1|\underline{\tau}_2^L)}\exp\left(-2E\left[\int_{J_1}d\tau_1\,p^o(\tau_1;\underline{\tau}_2^L)\cdot\int_I d\underline{t}\,\lambda(\underline{t};\tau_1,\underline{\tau}_2^L)\ln\frac{\lambda(\underline{t};\tau_1,\underline{\tau}_2^L)}{\overline{\lambda}^o(\underline{t};\underline{\tau}_2^L)}\right]\right)\tag{78}$$

where p^o and $\overline{\lambda}^o$ are the functions given by (68) and (69) of Lemma 4.

Theorem 1 applies to estimation of τ_1 for a general M-dimensional conditionally Poisson point processes which is energy invariant. It involves the solution of the integral equation (69) of Lemma 4 for the maximizing conditional density $p_{\tau_1|\underline{\tau}_2^L}=p^o$. Finding this solution is as difficult as solving for the capacity of a general continuous channel [14].

B. Conditional Entropy Invariance

The lower bound of Theorem 1 was obtained by establishing necessary and sufficient conditions for the maximizing conditional density, $p_{\tau_1|\underline{\tau}_2^L}=p^o$, in $C^*=\sup p_{\tau_1|\underline{\tau}_2^L}[H(N^o|\underline{\tau}_2^L)-H(N|\underline{\tau})]$, (61) of Lemma 3. Thus the bound of Theorem 1 is not explicit as a function of λ and $p_{\underline{\tau}}$. The next theorem deals with the special case where N satisfies the additional property of *conditional entropy invariance*. For this case, a weaker but more explicit lower bound is obtained via maximization of the entropy difference (61) in Lemma 3 directly over $\overline{\lambda}$.

Let the differential conditional entropy of N given $\underline{\tau}$ be functionally independent of τ_1. Then N is said to satisfy the conditional entropy invariance property with respect to τ_1. More specifically, taking into account the form of the conditional entropy of N in (2), the conditional entropy invariance property is given by the following property.

Property 2: $H(N|\underline{\tau})\triangleq E[-\ln P(N|\underline{\tau})|\underline{\tau}=\underline{\tau}]=\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau})\ln\lambda(\underline{t};\underline{\tau})$, is independent of $\tau_1\in J_1$.

At this point it is appropriate to give two definitions.

Definition 1: A set of τ_1-insensitive components of $\underline{t}=[t_1,\cdots,t_M]$ is any set of components, $[t_{i_1},\cdots,t_{i_K}]$, $\{i_j\}$ distinct indices, such that the integral of λ over the remaining components of $\underline{t}$ is independent of τ_1.

Without any loss in generality we will assume that the components of $\underline{t}$ have been ordered so that any τ_1-insensitive components are successive components: $\underline{t}_{m+1}^M\triangleq[t_{m+1},\cdots,t_M]$ for some m. Then, by definition,

$$\int_{-T_1/2}^{T_1/2}dt_1\cdots\int_{-Tm/2}^{Tm/2}dt_m\,\lambda(\underline{t};\underline{\tau})\text{ is independent of }\tau_1.\tag{79}$$

By the terminology "maximal set of τ_1-insensitive components of $\underline{t}$" is meant either $\min_m\underline{t}_{m+1}^M$, if the set of τ_1-insensitive components of $\underline{t}$ is nonempty, or the empty set, if no components of $\underline{t}$ are τ_1-insensitive. Associated with the maximal set of "τ_1-insensitive" components are what will be called the "τ_1-sensitive" components: $\underline{t}_1^m$. As an illustrative example consider the case:

$$\lambda(\underline{t};\underline{\tau})=\lambda_1(\underline{t}_1^l;\tau_1)\lambda_2(\underline{t}_{l+1}^M;\underline{\tau}_2^L)\tag{80}$$

where λ_1 satisfies energy invariance Property 1. For (80), the integral (79) with $m=l$ is independent of τ_1, and the $\underline{t}_{l+1}^M$ are thus τ_1-insensitive components.

Definition 2: The support of λ over $\underline{t}_1^m\triangleq[t_1,\cdots,t_m]$, denoted $\text{supp}_{\underline{t}_1^m}\lambda(\underline{t};\underline{\tau})$, is the m-dimensional set:

$$\text{supp}_{\underline{t}_1^m}\lambda(\underline{t};\underline{\tau})\triangleq\{\underline{t}_1^m:\ \lambda(\underline{t};\underline{\tau})>0\}.\tag{81}$$

Observe that $\text{supp}_{\underline{t}_1^m}\lambda(\underline{t};\underline{\tau})$ is generally a function of $\underline{t}_{m+1}^M\triangleq[t_{m+1},\cdots,t_M]$ and $\underline{\tau}$.

Lemma 5: Let $\underline{\tau}$ be a random vector with support $\underline{\tau}\in J=\times_{i=1}^L J_i$. Assume that $\lambda=\{\lambda(\underline{t};\underline{\tau}):\ \underline{t}\in I,\ \underline{\tau}\in J\}$ is an intensity function which satisfies the energy and conditional entropy invariance properties, Property 1, and Property 2, with respect to τ_1. Let $\underline{t}_{m+1}^M$ be the maximal set of τ_1-insensitive components of $\underline{t}$. Assume that

$$\text{supp}_{\underline{t}_1^m}\lambda(\underline{t};\underline{\tau})\subset\mathop{\times}_{i=1}^m[-T_i/2,T_i/2]\tag{82}$$

for all $\underline{t}_{m+1}^M\in\times_{i=m+1}^M[-T_i/2,T_i/2]$ and all $\underline{\tau}\in J$. Then C^* has the upper bound

$$\begin{aligned}C^*\leq C^{**}&\triangleq E\left[\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau})\ln\frac{\lambda(\underline{t};\underline{\tau})}{\overline{\lambda}^*(\underline{t};\underline{\tau}_2^L)}\right]\\&=E\left[\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau})\ln\lambda(\underline{t};\underline{\tau})-\int_I d\underline{t}\,\overline{\lambda}^*(\underline{t};\underline{\tau}_2^L)\ln\overline{\lambda}^*(\underline{t};\underline{\tau}_2^L)\right]\end{aligned}\tag{83}$$

where $\overline{\lambda}^*$ is the virtually uniform intensity over the maximal $\underline{t}_1^m$-support, denoted as $\times_{i=1}^m I_i'$, of $\overline{\lambda}$:

$$\overline{\lambda}^*(\underline{t};\underline{\tau}_2^L)\triangleq\begin{cases}\dfrac{1}{\left|\mathop{\times}_{i=1}^m I_i'\right|}\displaystyle\int_{\times_{i=1}^m I_i'}\lambda(\underline{t};\underline{\tau})\,d\underline{t}_1^m, & \underline{t}_1^m\in\mathop{\times}_{i=1}^m I_i'\\0, & \text{otherwise}\end{cases}.\tag{84}$$

In (84) $\times_{i=1}^{m} I_i' \triangleq \cup_{\tau_1 \in J_1} \mathrm{supp}_{t_1^m} \lambda(\underline{t}; \underline{\tau})$ and $|\times_{i=1}^{m} I_i'| \triangleq \int_{\times_{i=1}^{m}} dt_1^m$ is the volume of $\times_{i=1}^{m} I_i'$.

Proof: After some manipulations, Lemma 3 gives the following form for C^*:

$$C^* = E\left[\sup_{p_{\tau_1|\underline{\tau}_2^L}} \left\{ -\int_I \bar{\lambda}(\underline{t}; \underline{\tau}_2^L) \ln \bar{\lambda}(\underline{t}; \underline{\tau}_2^L)\, d\underline{t} + \int_I \lambda(\underline{t}; \underline{\tau}) \ln \lambda(\underline{t}; \underline{\tau})\, d\underline{t} \right\}\right]. \quad (85)$$

Since, by Property 2, the second integral in (85) is independent of τ_1, the "sup" in (85) only affects the first term. Furthermore, this first term depends on $p_{\tau_1|\underline{\tau}_2^L}$ only through the quantity $\bar{\lambda} = \int_{J_1} d\tau_1 \lambda p_{\tau_1|\underline{\tau}_2^L}$. Hence

$$C^* = E\left[\sup_{\bar{\lambda} \in \mathcal{G}^o} \left\{ -\int_I \bar{\lambda}(\underline{t}; \underline{\tau}_2^L) \ln \bar{\lambda}(\underline{t}; \underline{\tau}_2^L)\, d\underline{t} \right\}\right] + E\left[\int_I \lambda(\underline{t}; \underline{\tau}) \ln \lambda(\underline{t}; \underline{\tau})\, d\underline{t}\right] \quad (86)$$

where $\mathcal{G}^o$ is the set:

$$\mathcal{G}^o \triangleq \left\{ \tilde{\lambda}: \tilde{\lambda}(\underline{t}; \underline{\tau}_2^L) = \int_{J_1} d\tau_1 \lambda(\underline{t}; \underline{\tau}) p_{\tau_1|\underline{\tau}_2^L}(\tau_1|\underline{\tau}_2^L);\ p_{\tau_1|\underline{\tau}_2^L} = \mathrm{pdf} \right\}. \quad (87)$$

An upper bound is to be established on the first term in (86). To do this we generate a set of functions $\mathcal{G}$ such that $\mathcal{G}^o \subset \mathcal{G}$, i.e., $\mathcal{G}$ contains $\bar{\lambda}$ for arbitrary density $p_{\tau_1|\underline{\tau}_2^L}$. The upper bound will follow by the obvious inequality:

$$E\left[\sup_{\bar{\lambda} \in \mathcal{G}^o} \left\{ -\int_I d\underline{t}\, \bar{\lambda}(\underline{t}; \underline{\tau}_2^L) \ln \bar{\lambda}(\underline{t}; \underline{\tau}_2^L) \right\}\right] \le E\left[\sup_{\bar{\lambda} \in \mathcal{G}} \left\{ -\int_I d\underline{t}\, \bar{\lambda}(\underline{t}; \underline{\tau}_2^L) \ln \bar{\lambda}(\underline{t}; \underline{\tau}_2^L) \right\}\right]. \quad (88)$$

Define Q:

$$Q(\underline{t}_{m+1}^M; \underline{\tau}_2^L) \triangleq \int_{-T_1/2}^{T_1/2} dt_1 \cdots \int_{-Tm/2}^{Tm/2} dt_m \lambda(\underline{t}; \underline{\tau}). \quad (89)$$

Q is independent of $\underline{t}_1^m$ and, by the τ_1-insensitivity assumption, also independent of τ_1. Define the set of nonnegative functions $\mathcal{G}$:

$$\mathcal{G} \triangleq \left\{ \tilde{\lambda}: \mathrm{supp}_{t_1^m} \tilde{\lambda} \subset \times_{i=1}^{m} I_i';\ \int_{-T_1/2}^{T_1/2} dt_1 \cdots \int_{-T_m/2}^{T_m/2} dt_m \tilde{\lambda}(\underline{t}; \underline{\tau}_2^L) = Q(\underline{t}_{m+1}^M; \underline{\tau}_2^L) \right\} \quad (90)$$

where $\times_{i=1}^{m} I_i'$ is the $\underline{t}_1^m$-set defined in the lemma. The claim is that $\bar{\lambda} \in \mathcal{G}$. First observe that $\bar{\lambda}$ has a support which is contained in the union of the supports of $\lambda(\underline{t}; \underline{\tau})$ as τ_1 ranges over J_1. Specifically,

$$\mathrm{supp}_{t_1^m} \bar{\lambda} \subset \bigcup_{\tau_1 \in J_1} \mathrm{supp}_{t_1^m} \lambda(\underline{t}; \underline{\tau}) = \times_{i=1}^{m} I_i'. \quad (91)$$

Therefore, $\bar{\lambda}$ satisfies the first condition in the definition of $\mathcal{G}$ (90). As for the second condition in $\mathcal{G}$, apply the assumption (82) to (91) to obtain

$$\mathrm{supp}_{t_1^m} \bar{\lambda} \subset \times_{i=1}^{m} \left[-\frac{T_i}{2}, \frac{T_i}{2}\right], \quad (92)$$

so that $\bar{\lambda}$ is wholly supported in the observation interval I. Next consider the sequence of equalities:

$$\begin{aligned} \int_{-T_1/2}^{T_1/2} dt_1 \cdots \int_{-T_m/2}^{T_m/2} dt_m \bar{\lambda}(\underline{t}; \underline{\tau}_2^L) &= \int_{-T_1/2}^{T_1/2} dt_1 \cdots \int_{-T_m/2}^{T_m/2} dt_m E[\lambda(\underline{t}; \tau_1, \underline{\tau}_2^L)] \\ &= E\left[\int_{-T_1/2}^{T_1/2} dt_1 \cdots \int_{-T_m/2}^{T_m/2} dt_m \lambda(\underline{t}; \tau_1, \underline{\tau}_2^L)\right] \\ &= E[Q(\underline{t}_{m+1}^M; \underline{\tau}_2^L)] = Q(\underline{t}_{m+1}^M; \underline{\tau}_2^L), \end{aligned} \quad (93)$$

where the exchange in order of integration and expectation in (93) is justified by (92) and Tonelli's theorem for iterated integration of nonnegative functions [10]. Hence $\bar{\lambda}$ also satisfies the second condition in the definition of $\mathcal{G}$ (90) and is thus contained in $\mathcal{G}$ as claimed.

It is straightforward to verify that the function $\bar{\lambda} \in \mathcal{G}$, which maximizes the bound on C^* in (88) is a function that is uniform over the (finite) maximal support set $\underline{t}_1^m \in \times_{i=1}^{m} I_i' \subset \times_{i=1}^{m} [-T_i/2, T_i/2]$. This follows from the fact that the differential entropy of a support limited density function is maximized by the uniform density [14]. Application of the constraint (89) to this uniform function gives $\bar{\lambda}^*$ (84). Finally, substitution of $\bar{\lambda} = \bar{\lambda}^*$ into (88) gives the bound C^{**} in (83). This finishes the proof of Lemma 5.

Lemma 5 can be modified to handle smoothed tailed intensity functions with infinite support by replacing the support set (81) by an "ϵ-support set" containing all but a factor ϵ of the mass of the intensity function. The details are omitted.

Theorem 2: Let N be a conditionally Poisson process with almost everywhere bounded intensity function $\lambda = \{\lambda(\underline{t}; \underline{\tau}): \underline{t} \in I\}$, given the random parameter vector $\underline{\tau} = \underline{\tau} \in J$. Assume that λ is an intensity function which satisfies the energy and conditional entropy invariance properties, Properties 1 and 2, with respect to τ_1. Then a lower bound on the mse of an estimator $\hat{\tau}_1$ of τ_1 is

$$\mathrm{mse} \ge \frac{1}{2\pi e} e^{2H(\tau_1|\underline{\tau}_2^L)} \cdot \exp\left(-2E\left[\int_I d\underline{t}\, \lambda(\underline{t}; \underline{\tau}) \ln \frac{\lambda(\underline{t}; \underline{\tau})}{\bar{\lambda}^*(\underline{t}; \underline{\tau}_2^L)}\right]\right)$$

where $\bar{\lambda}^*$ is the function given in (84) of Lemma 5.

Proof: From Proposition 2, $\mathrm{mse} \ge \exp(2H(\tau_1|\underline{\tau}_2^L) - 2C^c)/(2\pi e)$ (37). On the other hand, combination of Lemmas 5 and 3 gives $C^c \le C^* \le C^{**}$. This establishes Theorem 2.

The following comments are useful. Observe that the bound in Theorem 2 is an exponentially decreasing function of the quantity:

$$\rho(\hat{\lambda}, \hat{\lambda}^*) \triangleq E\left[\Lambda\left(\underline{\tau}_2^L\right)\int_I d\underline{t}\, \hat{\lambda}(\underline{t}; \underline{\tau}) \ln \frac{\hat{\lambda}(\underline{t}; \underline{\tau})}{\hat{\lambda}^*\left(\underline{t}; \underline{\tau}_2^L\right)}\right] \quad (94)$$

where $\hat{\lambda} \triangleq \lambda/\Lambda(\underline{\tau}_2^L)$ and $\hat{\lambda}^* \triangleq \bar{\lambda}^*/\Lambda(\underline{\tau}_2^L)$ are normalized functions of $\underline{t}$; i.e., density functions over $\underline{t} \in I$. The integral in (94) is known as the information discrimination between the pdf's $\hat{\lambda}$ and $\hat{\lambda}^*$ [14]. The discrimination is a measure of the dissimilarity between two pdf's in the sense that it is equal to zero if and only if the two pdf's are identical, and it is positive otherwise. In view of the fact that $\hat{\lambda}^*$ (84) is constant over its $\underline{t}_1^m$-support set $\times_{i=1}^m I_i'$, Theorem 2 asserts that the closer the intensity of the conditionally Poisson observations is to a uniform function over the τ_1-sensitive components of $\underline{t}$, the more difficult it becomes to estimate τ_1 with low mse. This is consistent with the inherent difficulty in estimating a parameter for which the conditional observation density is not a sharp function of the unknown parameter of interest.

VI. Applications

For energy invariant processes, e.g., processes for which λ is shift or scale invariant with respect to τ_1, Theorem 1 can in principle be applied to develop lower bounds. In the case of shift invariance, the additional property of conditional entropy invariance holds, so that the simpler bound of Theorem 2 can be applied. In this section we specialize Theorem 2 to shift and relative shift estimation.

A. Shift Estimation

The following Theorem follows directly from Theorem 2.

Theorem 3: Let N be a conditionally Poisson process with almost everywhere bounded intensity function $\lambda = \{\lambda(\underline{t}; \underline{\tau}) \colon \underline{t} \in I\}$, given the random parameter vector $\underline{\tau} = \underline{\tau} \in J$. Assume

- $I = [-T_1/2, T_1/2] \times I_2 \times I_M$, $J = [-T_{\tau_1}/2, T_{\tau_1}/2] \times J_2 \times \cdots \times J_L$;
- $\lambda(\underline{t}; \underline{\tau}) = \lambda(t_1 - \tau_1, \underline{t}_2; 0, \underline{\tau}_2^L)$ for all $\underline{t} \in I$ and all $\underline{\tau} \in J$;
- a $T_\lambda = T_\lambda(\underline{t}_2^m, \underline{\tau}_2^L) \le T_1$ exists such that $\text{supp}_{t_1} \lambda(t_1, \underline{t}_2^m; 0, \underline{\tau}_2^L) \subset [-T_\lambda/2, T_\lambda/2]$, and $T_\lambda \le T_1 - T_{\tau_1}$.

Then a lower bound on the mse of an estimator $\hat{\tau}_1$ of τ_1 is the bound of Theorem 2 with $\bar{\lambda}^*$ given by

$$\bar{\lambda}^*\left(\underline{t}; \underline{\tau}_2^L\right) = \begin{cases} \dfrac{1}{T_{\tau_1} + T_\lambda} \displaystyle\int_{T_1/2}^{T_1/2} \lambda\left(\underline{t}; 0, \underline{\tau}_2^L\right) dt_1, & -\dfrac{T_{\tau_1} + T_\lambda}{2} \le t_1 \le \dfrac{T_{\tau_1} + T_\lambda}{2} \\ 0, & \text{otherwise} \end{cases} \quad (95)$$

Proof: First it is shown that t_1 is the only τ_1-sensitive component of $\underline{t}$. To this end consider the simple change of variable $t_1' = t_1 - \tau_1$ in the integral on the right of the following:

$$\int_{-T_1/2}^{T_1/2} dt_1\, \lambda(\underline{t}; \underline{\tau})$$
$$= \int_{-T_1/2}^{T_1/2} dt_1\, \lambda\left(t_1 - \tau_1, \underline{t}_2^M; 0, \underline{\tau}_2^L\right)$$
$$= \left\{\int_{-T_1/2}^{T_1/2} + \int_{-(T_1+\tau_1)/2}^{-T_1/2} - \int_{(T_1-\tau_1)/2}^{T_1/2}\right\} dt_1\, \lambda\left(t_1, \underline{t}_2^M; 0, \underline{\tau}_2^L\right)$$
$$= \int_{-T_1/2}^{T_1/2} dt_1\, \lambda\left(t_1, \underline{t}_2^M; 0, \underline{\tau}_2^L\right) \quad (96)$$

where in going from line two to line three of (96) we have used the assumption $T_\lambda \le T_1 - T_{\tau_1}$, which ensures that, of the three integrals inside the braces, only the first is nonzero. Recalling Definition 1, since the integral (96) is functionally independent of τ_1, $\underline{t}_2^M$ can be identified as the maximal set of τ_1 insensitive components. Since λ is supported on a t_1-interval, the maximal t_1-support, $I_1' = \bigcup_{\tau_1 \in J_1} \text{supp}_{t_1} \lambda$, of λ, (84) of Lemma 5, referenced in Theorem 2, is the interval: $[-(T_{\tau_1} + T_\lambda)/2, (T_{\tau_1} + T_\lambda)/2]$.

It remains to show that Properties 1 and 2 are satisfied. By (96), $\int_I \lambda = \int_{I_M} \cdots \int_{I_1} \lambda$ must also be independent of τ_1 so that Property 1 holds. Property 2 follows from the same change of variable argument as used in (96), $t_1' = t_1 - \tau_1$, applied to the innermost integral:

$$\int_I d\underline{t}\, \lambda(\underline{t}; \underline{\tau}) \ln \lambda(\underline{t}; \underline{\tau})$$
$$= \int_{I_M} dt_M \cdots \int_{-T_1/2}^{T_1/2} dt_1\, \lambda(\underline{t}; \underline{\tau}) \ln \lambda(\underline{t}; \underline{\tau}). \quad (97)$$

Hence the assumptions of Theorem 2 hold and the proof is complete.

B. Time Delay Estimation

A special case of Theorem 3 of interest is the time delay estimation problem for which λ is a one-dimensional pulse as a function of t_1. To simplify the notation, we assume that the sum of the length of the prior uncertainty interval on τ_1 and the pulsewidth is equal to the length of the observation interval: $T_{\tau_1} + T_\lambda = T_1$.

We first consider the case for which there are no nuisance parameters:

- $\lambda(\underline{t}; \underline{\tau}) = \lambda_1(t_1 - \tau_1)$;
- τ_1 is a random variable with support contained in $[-T_{\tau_1}/2, T_{\tau_1}/2]$;
- $\lambda_1(t)$ is a known function with t-width (pulsewidth) T_λ and $T_{\tau_1} + T_\lambda = T_1$.

For this case Theorem 3 gives the result:

$$\mathrm{mse} \geq \frac{1}{2\pi e} e^{2H(\tau_1|\tau_2^L)} \cdot \exp\left(2\Lambda_1 \ln \frac{\Lambda_1}{T_1} - 2\int_{-T_1/2}^{T_1/2} \lambda_1(t) \ln \lambda_1(t)\, dt\right)$$
$$= \frac{1}{2\pi e} e^{2H(\tau_1|\tau_2^L)} \cdot \exp\left(2\Lambda_1\left[\ln \frac{1}{T_1} - \int_{-T_1/2}^{T_1/2} \hat{\lambda}_1(t) \ln \hat{\lambda}_1(t)\, dt\right]\right), \quad (98)$$

where $\Lambda_1 \triangleq \int_{I_1} \lambda_1(t_1)\, dt_1$ is the energy of the pulse and the normalized intensity (density) $\hat{\lambda}_1$ has been defined:

$$\hat{\lambda}_1(t) \triangleq \frac{\lambda_1(t)}{\Lambda_1}. \quad (99)$$

The bound (98) has several notable characteristics. The leading exponential term is independent of the observation statistics. This term is determined by the *a priori* uncertainty in the time delay τ_1. The exponential factor in brackets is a channel capacity C^{**} that is always negative. Hence the mse lower bound is an exponentially decreasing function of the total energy of the pulse Λ_1. The exponential rate of decay is determined by the information discrimination between the conditionally Poisson inhomogeneous process N and an equal energy homogeneous Poisson process. This is, equivalently, the difference between the differential entropy of the normalized intensity $\hat{\lambda}$ and the entropy of a uniform density over the observation interval. Thus the lower bound on the error in estimating τ_1 decreases very rapidly as either the pulse energy is increased or as the intensity function becomes more concentrated over a small subset of the observation interval.

It is instructive to consider representative extreme cases of the rate distortion bound (98) as a function of the pulse shape $\hat{\lambda}$. It is easily seen that the bound is minimized by minimizing the differential entropy of $\hat{\lambda}$. The differential entropy can be made to approach $-\infty$ by letting $\hat{\lambda}$ approach a delta function, corresponding to the case of perfect observation of τ_1, in which case the error bound is zero. On the other hand, the maximum of the bound is achieved for $\hat{\lambda}$ equal to a uniform density over the observation interval, corresponding to useless observations, in which case the error bound is only a function of the prior distribution on τ_1: the "entropy power" $\exp(H(\tau_1))/(2\pi e)$. In particular, for Gaussian τ_1 the entropy power is simply the *a priori* variance of τ_1 which is achievable by using the trivial estimator $\hat{\tau}_1 = E[\tau_1]$.

Finally, a comparison between the lower bound (98) and the approximate lower bound in [3] (30) can be made. Let λ be a narrow rectangular intensity of the form (29), $T_\lambda \ll T_1 = T$, and let τ_1 be uniformly distributed over $[-(T-T_\lambda)/2,(T-T_\lambda)/2] \approx [-T/2, T/2]$. Then (98) reduces to the form:

$$\mathrm{mse} \geq \frac{T^2}{2\pi e}\left(\frac{T_\lambda}{T}\right)^{2\Lambda} \quad (100)$$

where subscripts on Λ_1 and T_1 found in (98) have been suppressed. The lower bound (100) and the approximate lower bound (30) have quite different behavior over the majority of values of Λ. In particular as $\Lambda \to 0$, the bound (100) goes to the *a priori* entropy power of τ_1 while (30) goes to a much smaller quantity which is a fraction of the pulsewidth T_λ. On the other hand, as $\Lambda \to \infty$, (100) decreases exponentially while (30) has only quadratic decrease. Hence, generally speaking, the bound (100) is tighter in the small Λ region while (30) is tighter in the large Λ region.

We next consider the case of time delay estimation in the presence of uncertainty in pulsewidth $\tau_2 T_\lambda$ and uncertainty in pulse amplitude τ_3. Specifically, we assume the following:

- $\lambda(\underline{t};\underline{\tau}) = (\tau_3/\tau_2)\lambda_1((t_1-\tau_1)/\tau_2)$, $t_1 \in I = [-T_1/2, T_1/2]$;
- τ_1, τ_2, τ_3 are independent random variables with supports contained in the sets $[-T_{\tau_1}/2, T_{\tau_1}/2]$, $[0,1]$ and $[0,\infty)$, respectively;
- $\lambda_1(t)$ is a known function with t-width T_λ and $T_{\tau_1} + T_\lambda = T_1$.

Theorem 3 and a straightforward computation show that under the above assumptions the mse of $\hat{\tau}_1$ is bounded by

$$\mathrm{mse} \geq \frac{1}{2\pi e} e^{2H(\tau_1)} \exp\left(2E[\tau_3]\left[\Lambda_1 \ln \frac{\Lambda_1}{T_1} - \int_{-T_1/2}^{T_1/2} \lambda_1(t) \cdot \ln \lambda_1(t)\, dt + E[\ln \tau_2]\right]\right) \quad (101)$$

where $\Lambda_1 \triangleq \int_{-T_1/2}^{T_1/2} \lambda_1(t)\, dt$.

Observe that the mse bound (101) is similar to (98) as a function of λ and T_1. Since $\tau_3 \in [0,1]$, the mse bound decreases exponentially in the mean pulse amplitude. It is significant that the bound is not directly dependent on the variance of the amplitude scale factor τ_3. On the other hand, it is to be noted that the mean and variance of τ_3 are not independent parameters due to the positivity of τ_3. The role of uncertainty in the pulsewidth τ_2 is to increase the mse bound as a function of the mean logarithmic value of pulsewidth uncertainty. Hence a Taylor expansion shows that, unlike the pulse amplitude uncertainty, the variance of the pulsewidth, is directly implicated in the mse bound.

C. Relative Shift Estimation

When an estimate of the relative shift τ_1 between two separately observed conditionally independent Poisson processes is to be estimated, Theorem 2 is applicable. For relative shift estimation, the absolute time delays of the individual point processes are not of direct interest. Thus it is appropriate for this problem to model the midpoint τ_2 between the arrival times as an unknown nuisance parameter.

Theorem 4: Let $[N_1, N_2]$ be a vector of two conditionally independent Poisson processes with almost everywhere bounded intensity functions: $\lambda_1 = \{\lambda_1(\underline{t};\underline{\tau}): t \in I\}$, $\lambda_2 =$

$\{\lambda_2(\underline{t};\underline{\tau}): \underline{t}\in I\}$, given the random parameter vector $\underline{\tau}=\underline{\tau}\in J$. Assume

- $I=[-T_1/2, T_1/2]\times I_2\times\cdots\times I_M$, $J=[-T_{\tau_1}/2, T_{\tau_1}/2]\times[-T_{\tau_2}/2, T_{\tau_2}/2]\times J_3\times\cdots\times J_L$;
- $\lambda_1(\underline{t};\underline{\tau})=\lambda_1(t_1-\tau_2+(\tau_1/2), \underline{t}_2^m, 0, 0, \underline{\tau}_3^L)$ for all $\underline{t}\in I$ and all $\underline{\tau}\in J$;
- $\lambda_2(\underline{t};\underline{\tau})=\lambda_2(t_1-\tau_2-(\tau_1/2), \underline{t}_2^m; 0, 0, \underline{\tau}_3^L)$ for all $\underline{t}\in I$ and all $\underline{\tau}\in J$;
- the λ_i have finite t_1-widths: $T_{\lambda_i}(t_2,\tau)\le T_1$ such that: $\mathrm{supp}_{t_1}\lambda_i\subset I_1$.

Then a lower bound on mse of an estimator $\hat{\tau}_1$ of τ_1 is

$$\mathrm{mse}\ge\frac{1}{2\pi e}e^{2H(\tau_1|\underline{\tau}_2^L)}\exp\left(-2\sum_{i=1,2}E\left[\int_I d\underline{t}\,\lambda_i(\underline{t};\tau_1,0,\underline{\tau}_3^L)\cdot\ln\frac{\lambda_i(\underline{t};\tau_1,0,\underline{\tau}_3^L)}{\bar{\lambda}_i^*(\underline{t};0,\underline{\tau}_3^L)}\right]\right)\tag{102}$$

where the $\bar{\lambda}_i^*$ are given by

$$\bar{\lambda}_i^*(\underline{t};\underline{\tau}_2^L)\triangleq\begin{cases}\dfrac{1}{T_{\lambda_1}+T_{\lambda_2}+T_{\tau_1}}\displaystyle\sum_{j=1,2}\int_{-T_1/2}^{T_1/2}\lambda_j(\underline{t};0,0,\underline{\tau}_3^L)\,dt_1, & t_1\in\left[-\dfrac{T_{\lambda_i}+T_{\tau_1}/2}{2},\dfrac{T_{\lambda_i}+T_{\tau_1}/2}{2}\right]\\ 0, & \text{otherwise}\end{cases}\tag{103}$$

Proof: Since the t_1-supports of the λ_i are contained in the finite interval $I_1=[-T_1/2, T_1/2]$, the trajectories of N_1 and N_2 can be concatenated to form the equivalent scalar process $N(\underline{t})$, $\underline{t}\in[-T_1,T_1]\times I_2\times\cdots\times I_M$, with intensity

$$\lambda(\underline{t};\underline{\tau})=\lambda_1\left(t_1+\frac{T_1}{2},\underline{t}_2^M;\underline{\tau}\right)+\lambda_2\left(t_1-\frac{T_1}{2},\underline{t}_2^M;\underline{\tau}\right).\tag{104}$$

The rest of the proof is analogous to the proof of Theorem 3. It is easily verified that t_1 is the τ_1-sensitive component of t and that Property 1 is satisfied. Furthermore, N satisfies Property 2 since, by conditional independence, $H(N_1,N_2|\tau)=H(N_1|\tau)+H(N_2|\tau)$; and N_1 and N_2 individually satisfy Property 2 with respect to τ_1. Next we identify the maximal t_1-support of $\bar{\lambda}$, namely:

$$\begin{aligned}\mathop{\times}_{i=1}^{m} I_1'=I_1'&=\bigcup_{\tau_1\in J_1}\mathrm{supp}_{t_1}\lambda\\&=\bigcup_{\tau_1\in J_1}\left[-\frac{T_{\lambda_1}}{2}+\frac{T_1}{2}-\tau_2-\frac{\tau_1}{2},\frac{T_{\lambda_1}}{2}+\frac{T_1}{2}-\tau_2-\frac{\tau_1}{2}\right]\\&\quad\cdot\bigcup\left[-\frac{T_{\lambda_2}}{2}-\frac{T_1}{2}-\tau_2+\frac{\tau_1}{2},\frac{T_{\lambda_2}}{2}-\frac{T_1}{2}-\tau_2+\frac{\tau_1}{2}\right]\\&=\left[-\frac{T_{\lambda_1}}{2}+\frac{T_1}{2}-\tau_2-\frac{T_{\tau_1}}{4},\frac{T_{\lambda_1}}{2}+\frac{T_1}{2}-\tau_2+\frac{T_{\tau_1}}{4}\right]\\&\quad\cdot\bigcup\left[-\frac{T_{\lambda_2}}{2}-\frac{T_1}{2}-\tau_2-\frac{T_{\tau_1}}{4},\frac{T_{\lambda_2}}{2}-\frac{T_1}{2}-\tau_2+\frac{T_{\tau_1}}{4}\right].\end{aligned}\tag{105}$$

Hence I_1' of (105) is a disjoint union of intervals of lengths $T_{\lambda_1}+T_{\tau_1}/2$ and $T_{\lambda_2}+T_{\tau_1}/2$. Application of Theorem 2 gives a bound with the exponential factor:

$$C^{**}=E\left[\int_I d\underline{t}\,\bar{\lambda}^*(\underline{t};\underline{\tau}_2^L)\ln\bar{\lambda}^*(\underline{t};\underline{\tau}_2^L)\right]-E\left[\int_I d\underline{t}\,\lambda(\underline{t};\underline{\tau})\ln\lambda(\underline{t};\underline{\tau})\right]\tag{106}$$

where, following (84), $\bar{\lambda}^*$ is the function defined below, having a t_1-support set I_1':

$$\begin{aligned}\bar{\lambda}^*(\underline{t};\underline{\tau}_2^L)&=\frac{1}{|I_1'|}\int_{I_1'}\lambda(\underline{t};0,\underline{\tau}_2^L)\,dt_1\\&=\frac{1}{|I_1'|}\sum_{i=1,2}\int_{-(T_{\lambda_i}/2)-((-1)^iT_1/2)-\tau_2-(T_{\tau_1}/4)}^{(T_{\lambda_i}/2)-((-1)^iT_1/2)-\tau_2-(T_{\tau_1}/4)}\cdot\lambda_i\left(t_1-\frac{(-1)^iT_1}{2}-\tau_2,\underline{t}_2^M;0,0,\underline{\tau}_3^L\right)dt_1\\&=\frac{1}{T_{\lambda_1}+T_{\lambda_2}+T_{\tau_1}}\sum_{i=1,2}\int_{-(T_{\lambda_i}/2)-(T_{\tau_1}/4)}^{(T_{\lambda_i}/2)-(T_{\tau_1}/4)}\cdot\lambda_i(t_1,\underline{t}_2^M;0,0,\underline{\tau}_3^L)\,dt_1.\end{aligned}\tag{107}$$

Substitution of (107) into (106) gives the following:

$$E\left[\int_{\times_{i=2}^M I_i}d\underline{t}_2^M\int_{-T_1}^{T_1}dt_1\,\lambda(\underline{t};\underline{\tau})\ln\frac{\lambda(\underline{t};\underline{\tau})}{\bar{\lambda}^*(\underline{t};\underline{\tau}_2^L)}\right].\tag{108}$$

Observe that λ (104) and $\bar{\lambda}^*$ (107) are functions of τ_2 only through a shift $t_1-\tau_2$. Therefore, a change of variable $t_1'=t_1-\tau_2$ in (108) renders the integrand independent of τ_2. The functions λ and $\bar{\lambda}^*$ are each the sum of two functions with disjoint support sets. Hence the integral (108) decomposes into the sum of two integrals indicated in the exponent of (102). This completes the proof of Theorem 4.

D. Relative Time Delay Estimation

We specialize Theorem 4 to the case where N_1 and N_2 are one-dimensional conditionally independent Poisson streams which undergo random relative and absolute time shifts. Specifically, assume N_1 and N_2 are the components of a vector point process observed over the time interval $I=[-T_1/2, T_1/2]$ with intensities λ_1 and λ_2 such that

- $\lambda_1(\underline{t};\underline{\tau})=\tilde{\lambda}_1(t_1-\tau_2-(\tau_1/2))$;
- $\lambda_2(\underline{t};\underline{\tau})=\tilde{\lambda}_2(t_1-\tau_2+(\tau_1/2))$;
- τ_1,τ_2 are independent with supports $[-T_{\tau_1}/2, T_{\tau_1}/2]$, $[-T_{\tau_2}/2, T_{\tau_2}/2]$, respectively;
- the $\tilde{\lambda}_i(t)$ are known functions with t-widths T_{λ_i} and $T_{\tau_2}+(T_{\tau_1}/2)\le\max\{T_1-T_{\lambda_1}, T_1-T_{\lambda_2}\}$.

The last assumption guarantees that the support of the λ_i will be contained in the observation interval $I=[-T_1/2, T_1/2]$ for all τ_1 and τ_2. The maximal t_1-support for $\bar{\lambda}_i$ is identical to the general case in Theorem 3, and, after some manipulation, the following bound on the mse of $\hat{\tau}_1$ is obtained:

$$\text{mse} \geq \frac{e^{2H(\tau_1)}}{2\pi e} \exp\left(2(\Lambda_1+\Lambda_2)\ln\frac{\Lambda_1+\Lambda_2}{T_{\lambda_1}+T_{\lambda_2}+T_{\tau_1}} - 2\sum_{i=1,2}\int_{-T_1/2}^{T_1/2} dt_1\, \lambda_i(t_1)\ln\lambda_i(t_1)\right) \tag{109}$$

where

$$\Lambda_1 \triangleq \int_{-T_1/2}^{T_1/2} \lambda_1(t)\, dt$$

$$\Lambda_2 \triangleq \int_{-T_1/2}^{T_1/2} \lambda_2(t)\, dt.$$

Observe that, under the above assumptions on τ_2, the mse bound (109) is functionally independent of the absolute delays of N_1 and N_2 since τ_2 does not appear in the bound.

Consider the special case of identical intensities and maximal uncertainty in the relative time shift. This is appropriate for time-of-flight PET and PPM optical communications applications when the observation interval $[-T_1/2, T_1/2]$ is known to contain all of the process energy. For this case we take in (109) $\lambda_1 = \lambda_2$ and $T_{\lambda_1} + T_{\tau_1}/2 = T_1$ to obtain the bound

$$\text{mse} \geq \frac{1}{2\pi e} e^{2H(\tau_1)} \cdot \exp\left(4\Lambda_1\left[\ln\frac{1}{T_1} - \int_{-T_1/2}^{T_1/2} dt_1\, \hat{\lambda}_1(t_1)\ln\hat{\lambda}_i(t_1)\right]\right) \tag{110}$$

where $\hat{\lambda}_1 = \lambda_1/\Lambda_1$ is the normalized intensity. Comparison between (110) and (98) indicates that the mse for relative delay estimation is equal to the mse for (absolute) delay estimation times a factor $\exp(-2\rho(\lambda, \bar{\lambda}^*))$ where ρ is the discrimination (94) between λ and a uniform intensity $\bar{\lambda}^*$ over $[-T_1/2, T_1/2]$. Hence for $\rho \gg 1$ the mse bound for relative time delay is much better than the mse bound for single stream time delay.

VII. Conclusion

We derived a general lower bound on the mse of estimators of a random parameter for M-dimensional conditionally Poisson point processes using rate-distortion theory. For the lower bound we assumed that the integral of the conditional intensity function over the observation interval, or "energy of the process," is functionally independent of the parameter of interest. To derive the bound a maximum entropy property associated with inhomogeneous Poisson processes was applied to upper bound the capacity of a channel with nuisance parameter side information. The general lower bound depends implicitly on an optimizing conditional source probability density function for which necessary and sufficient conditions were given. Under the additional assumption that the conditional entropy of the Poisson process is independent of the parameter of interest an explicit mse bound was derived. The energy and conditional entropy invariance properties hold for the problems of shift and relative shift parameter estimation in the presence of nuisance parameters. The bound was evaluated for forms of the intensity function which are appropriate for time delay and relative time delay estimation in conditionally Poisson streams. The form of the lower bound indicated two important limiting factors of mse performance: 1) the energy of the point process; and 2) the information discrimination between the observed inhomogeneous conditionally Poisson process and an equal energy homogeneous Poisson process over the observation time interval.

References

[1] F. Gatti and V. Svelto, "Review of theories and experiments of resolving time with scintillation counters," *Nuclear Instrum. Methods*, vol. 43, pp. 248–268, 1966.

[2] R. M. Gagliardi and S. Karp, *Optical Communications.* New York: Wiley, 1976.

[3] I. Bar-David, "Minimum-mean-square-error estimation of photon pulse delay," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 326–330, May 1975.

[4] N. H. Clinthorne, W. L. Rogers, A. O. Hero, G. D. Hutchins, and K. F. Koral, "Improved coincidence timing through the application of estimation theory," presented at the Society of Nuclear Medicine 34th Annual Meeting, Toronto, ON, Canada, June 1987.

[5] I. Bar-David, "Information in the time of arrival of a photon packet: capacity of PPM channels," *J. Opt. Soc. Amer.*, vol. 63, no. 2, pp. 166–170, Feb. 1973.

[6] P. N. Misra and H. W. Sorenson, "Parameter estimation in Poisson processes," *IEE Trans. Inform. Theory*, vol. IT-21, pp. 87–90, Jan. 1975.

[7] L. P. Seidman, "Performance limitations and error calculations for parameter estimation," *Proc. IEEE*, vol. 58, pp. 644–652, 1970.

[8] R. B. Washburn and D. Teneketzis, "Rate distortion lower bound for a special class of nonlinear estimation problems," in *Proc. 24th IEEE Conf. Decision and Control*, 1985.

[9] D. L. Snyder, *Random Point Processes.* New York: Wiley, 1975.

[10] P. Billingsley, *Probability and Measure.* New York: Wiley, 1979.

[11] I. A. Ibragimov and R. Z. Has'minsky, *Statistical Estimation: Asymptotic Theory.* New York: Springer-Verlag, 1981.

[12] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, vol. I. New York: Wiley, 1969.

[13] T. Berger, *Rate Distortion Theory.* New York: Prentice-Hall, 1971.

[14] R. Blahut, *Applications of Information Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[15] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, "Some classes of global Cramer–Rao bounds," *Ann. Statist.*, vol. 15, no. 4, pp. 1421–1438, 1987.

[16] D. G. Luenberger, *Optimization by Vector Space Methods.* New York: Wiley, 1969.