Timing Estimation for a Filtered Poisson Process in Gaussian Noise

Alfred O. Hero, III, Member, IEEE

Abstract — We treat the problem of estimation of time shift of an inhomogeneous causally filtered Poisson process in the presence of additive Gaussian noise. Approximate expressions for the likelihood function, the MAP estimator, and the mse estimator are obtained, which become increasingly accurate as the per-unit-time density of superimposed filter responses becomes small. The optimal MAP estimator takes the form of a cascade of linear and memoryless nonlinear components. For low signal amplitudes the MAP estimator is equivalent to maximizing the output of a linear matched filter arising in optical communications receivers. For smooth point process intensities, the performance of the MAP estimator is studied via local bias and local variance. A rate distortion type lower bound on the mse of any estimator of time delay is then derived by identification of a communications channel that accounts for the mapping from time delay to observation process. Finally, results of numerical studies of estimator performance are presented. Based on the examples considered it is concluded: 1) the small error rate of the nonlinear MAP estimator can be significantly better than the small error rate of the optimal linear estimator; 2) the rate distortion lower bound can be significantly tighter than the Poisson limited bound determined in previous studies.

I. INTRODUCTION

WE TREAT in this paper the problem of time shift estimation for inhomogeneous, causally filtered, Poisson point processes, equivalently nonstationary "shot noise," in the presence of additive Gaussian noise when the time delay parameter is imbedded in the intensity function of the point process. We call this combined pulse superposition and Gaussian noise signal a Poisson–Gaussian process. Here we consider only the simple case of constant filter gain; an extension to a random gain model is considered elsewhere [15]. Our model is a special case of a random gain model arising in applications including: nuclear particle detection systems [10], optical communications systems [16], neural spike train analysis [6], acoustic echo-localization [5], seismic signal processing [21], and analysis of underwater acoustic reverberation noise [8]. For example, in Positron emission tomography (PET) a nuclear decay event produces gamma rays that are detected and amplified through avalanche production of photons in photo-multiplier (PM) tubes [23]. Each PM tube drives an electronic circuit whose output is a combination of the superposition of single photon filter

responses, forming a filtered Poisson process, and thermal Gaussian noise. The objective of the estimator is to give an accurate estimate of the time-of-flight of the gamma rays, from which the approximate spatial position of the nuclear decay can be determined. In optical communications, an optical receiver is to decide between the presence of a logical "1" and a logical "0" based on the number of photons detected over a certain time interval that is specified by slot-synchronization information. In the case of pulse position modulation (PPM) this binary signal is coded in the time shifts of the photon packets. In other modulation formats, such as phase shift keying (PSK) and pulse amplitude modulation (PAM), lack of synchronization of the receiver to the phase/time delay of the photon intensity can significantly degrade detection performance [9].

Previous work on estimation of time shift of point process intensities has focussed on the Poisson limited regime, e.g., [11,14,15], or on the Gaussian limited regime [9,10]. For the Poisson limited regime, the point process can be observed directly and exact analytical forms of the likelihood function can be found. For this regime the major difficulty lies in the manipulation of the likelihood function into an tractable form for maximum likelihood (ML), maximum a posteriori (MAP), or minimum mean-square error (MMSE) estimation of time shift. For the Gaussian limited regime the time shift parameter is imbedded in the mean and covariance functions of the observation waveforms and an analytical functional form for the likelihood function can be found, see for example [7]. For the mixed Poisson–Gaussian regime considered in this paper, an analytical form for the process density function is more difficult to obtain even though the characteristic function has an analytical form [24]. Here we focus on an approximation to the likelihood function based on a low-density condition similar to [22, Ch. 11]; the product of the filter pulse-width and the intensity amplitude is uniformly small over time. This condition can be viewed as complementary to the condition of uniformly large intensity amplitude for which case one can assert that the filtered Poisson process is approximately Gaussian [27].

It is worth pointing out that estimation of point process intensity time shift differs from the related problem of estimation of the amplitudes and arrival times of the superimposed filter response pulses. In particular, in [5,20] the pulse arrivals are not Poisson and the number of pulses is assumed known. In [21] Poisson arrivals are considered but the point process intensity, assumed constant therein, is only incident to the pulse amplitude and pulse position estimation problem. The asymptotic likelihood function approximation developed here is obtainable as a special case of an approximation proposed in a paper [18] that appeared a number of

Manuscript received March 23, 1988; revised June 12, 1988. This work was supported in part by the National Institutes of Health (NIH) under Contract R01-CA 44622-01. This work was presented in part at the 1988 Conference on Information Sciences and Systems, Princeton, NJ. March 1988.

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

IEEE Log Number 904281.
months after this paper was submitted. However, in [18] the estimation problem is not considered.

In this paper we use the low pulse density approximation to obtain forms for the likelihood ratio used to detect the time shifted point process for known intensity time shift. The expression is then used to obtain approximations to the MAP estimator and the mse estimator of time delay. The log-likelihood ratio is seen to have the following structure: After suppression of the Gaussian noise component of the observations via classical matched filtering, an exponential transformation generates a spike train that is subsequently correlated against a time shifted version of the intensity function.

The log-likelihood ratio is maximized over the set of possible times shifts to yield the MAP estimate. It is significant that for low process amplitude, and a narrow superposition filter response, the nonlinear log-likelihood ratio reduces to a classical linear filter receiver that has been previously proposed as a suboptimal detector for optical communications systems. This is a uniformly low intensity extension of the well-known result that the linear matched filter is an approximatively optimal linear estimator/detector for uniformly high intensity. While we do not consider random gains on the generated pulses, it is shown in [15] that with random Gaussian distributed gains the resultant likelihood function approximations are quite similar to those in this paper. Specifically, the only difference is that the random gain model introduces a quadratic memoryless nonlinearity between the matched-filter and the exponential transformation in the log-likelihood structure previously described.

To evaluate the performance of the approximation to the MAP estimator, we derive expressions for the local bias and local variance under the assumption that the point process intensity is differentiable. These expressions characterize the small error behavior of the approximate MAP estimator over the full range of low to high pulse density. The local bias and local variance can be related to the asymptotic rates of decay of the actual bias and variance of a MAP estimator obtained from observing successive independent versions of the Poisson-Gaussian process through the mechanism of repeated experiments. Alternatively, the small error variance can be interpreted as an approximate Cramer-Rao lower bound for low-pulse density. To establish fundamental limits on estimation performance, a rate distortion lower bound on estimator mse is then derived. The bound is obtained by identifying a cascaded Poisson and Gaussian channel which maps random time delays to the measurement process. The capacity of the Poisson channel is upper bounded using results in [14]. On the other hand, the capacity of the Gaussian channel is upper bounded using standard results of rate distortion theory [3]. The final form of the rate distortion bound is obtained via the data processing theorem. Finally, a numerical evaluation of the expressions for small-error mse and the rate distortion lower bound is presented for the special case of Gaussian intensity and exponential superposition filter response. Based on the examples considered we conclude: 1) the small-error mse of the nonlinear MAP estimator can be significantly better than the small-error mse of the optimal linear estimator; 2) the rate distortion lower bound can be significantly tighter than the Poisson bounded limits studied in [14].

The organization of the paper is as follows. Section II introduces the main assumptions and gives a low-pulse density approximation to the likelihood ratio. Section III reviews conventional linear estimator structures and develops forms for the MAP and mse estimators. Section IV develops expressions for local bias and local variance of the MAP estimator. Section V presents lower bounds on mse. In Section VI numerical performance comparisons are presented.

II. PROBLEM STATEMENT

A few words about notation are useful. In general, bold faced variables, e.g., \( \mathbf{X} \) denote random variables, vectors, or processes. A notable exception is the random variable \( \tau \). The probability distribution of \( X \) is denoted by the generic \( \text{P}(X) \).

Observe that \( \text{P}(X) \) and \( \text{P}(\tau) \) are different functions, not the evaluation of a function \( \text{P} \) for two different values of its argument. The probability density of \( X \), with respect to some usually unspecified dominating measure \( \mu \), is denoted by \( f(X) \).

We will consider the case of a general random time delay \( \tau \). Estimators for the nonrandom case can be treated by specializing to a uniform prior on \( \tau \). Conditioned on \( \tau \), let

\[
N(t) = (N(t); t \in \{0, T\}) \text{ be an inhomogeneous Poisson process with intensity function } \lambda(t - \tau), \quad t \in \{0, T\},
\]

where

\[
\lambda(t) = \lambda(\tau) + \lambda_0.
\]

In view of (1), \( N \) corresponds to the sum of an inhomogeneous Poisson signal process, \( N_s \), with intensity \( \lambda_s(t - \tau) \), and an independent homogeneous Poisson noise process, \( N_0 \), with intensity \( \lambda_0 \). It will be assumed that \( \tau \) affects \( \lambda \) as a shift, without truncation in the sense that the set \( \{\lambda(t - \tau) > 0\} \) is contained in the observation interval \( [0, T] \) for all \( \tau \) over the support set \( \{\tau; f(\tau) > 0\} \).

Let the total number of points of \( N \) over \( [0, T] \) be denoted by \( n \), specifically

\[
n = \text{N}(T) \text{.}
\]

Let \( (t_i)^n \) be the \( n \) occurrence times associated with the points of \( N \). The joint probability distribution of \( (t_i)^n \) and \( n \) given \( \tau \) is specified by the conditional density \( f(t_i)^n \mid n, \tau \) of \( (t_i)^n \) and the probability mass function \( p(n \mid \tau) \) of \( n \) [27]:

\[
f(t_i)^n \mid n, \tau) p(n \mid \tau) = \begin{cases} e^{-n} \prod_{i=1}^{n} \Lambda(t_i - \tau), & n > 0, \\ e^{-\Lambda}, & n = 0,
\end{cases}
\]

where

\[
\Lambda \overset{\text{def}}{=} \int_{0}^{T} \Lambda(t - \tau) dt + \int_{0}^{T} \lambda(t) dt = \Lambda_s + \Lambda_0,
\]

is the energy or integrated rate of the point process which, by assumption, is functionally independent of \( \tau \).

Available for observation over the time interval \( t \in \{0, T\} \) is the sum of a filtered Poisson process and Gaussian noise

\[
X(t) = \sum_{i=1}^{n(\tau)} p(t - t_i) + W(t),
\]

where \( p(t) \) is a known continuous and square-integrable filter impulse response with 3-dB time-width \( T_p \), \( W \) is zero-mean white Gaussian noise independent of \( \tau \) and \( (t_i)^n \), with \( E[W(t)w(u)] = (N_0/2)B(t - u) \). For the purposes of mathematical analysis, the notation in (4) is to be interpreted as shorthand for the equivalent second-order integrated observation

\[
Y(t) = \int_{-\infty}^{t} \sum_{i=1}^{n(\tau)} p(u - t_i) du + W(t),
\]

where \( W(t) \) is a standard Weiner process [28]. We will assume that the filter
a deterministic signal plus a white Gaussian random process. A
form for the conditional density of the observations given
\((t_{i},)_{i=1}^{n}, n \) and \( \tau \) can then be obtained using
the Cameron–Martin formula [25]:
\[
f(X|t_{i}, n, \tau) = e^{(2/N_{x}) \gamma \lambda(t - \tau)} \frac{\mu(t - \tau)}{\mu(t)} \frac{\gamma}{\mu(t - \tau)} d\tau.
\]
(7)

The numerator and denominator of the likelihood ratio
statistic are obtained by taking the expectation of (7) over
\((t_{i},)_{i=1}^{n} \) and \( n \), denoted \( E_{n-1} \), given \( H_{1} \) and \( H_{0} \) respectively:
\[
f(X|H_{1}, \tau) = E_{n-1} \left[ f(X|t_{i}, n, \tau) \right| H_{1}, \tau],
\]
\[
f(X|H_{0}) = E_{n-1} \left[ f(X|t_{i}, n, \tau) \right| H_{0}].
\]
(8)

The two conditional density functions of (8) are not in
analytic form. In Appendix A, \( O(T_{p}N, \lambda^{+}) \) approximations to
these functions are obtained, where
\[
\lambda^{+} = \max_{T_{p}} \frac{1}{T_{p}} \int_{t}^{t+T_{p}} \lambda(t) dt \leq \max \lambda(t).
\]

Therefore, the approximation is accurate under the assumption
\( T_{p} \ll 1/\Lambda \max \lambda(t) \). This assumption is valid when
the pulse width \( T_{p} \) is sufficiently small so that the nonoverlapping
pulses dominate in the sum \( \sum \gamma \) in (7). Substitution of
expressions (A.9) and (A.10) of Appendix A into the likelihood
ratio statistic \( L(X|\tau) = f(X|H_{1}, \tau) / f(X|H_{0}) \) gives
\[
L(X|\tau) = \exp \left[ \int_{0}^{T} e^{(2/N_{x}) \gamma \lambda(t - \tau)} dt \right],
\]
(9)

where the asterisk denotes convolution and \( \gamma \) is the pulse-
to-noise ratio (PNR):
\[
\gamma = 2 \frac{\int_{0}^{T} \rho^{2}(t) dt}{N_{x} t}. \quad (10)
\]

III. ESTIMATOR STRUCTURES

In this section we will investigate the forms of the MAP
and MMSE estimators for \( \tau \), under the approximation (9) to
the likelihood ratio. In Section VI performance analysis is
presented for the MAP estimator. Let the MAP estimator
and the MMSE estimator for \( \tau \) be denoted \( \hat{\tau}_{MAP} \) and \( \hat{\tau}_{MMSE} \) respectively.

A. MAP Estimation

The MAP estimator maximizes the \textit{a posteriori} probability
\( f(\tau|X, H_{1}) \), or, equivalently, it maximizes the quantity
\( \ln(L(X|\tau)f(\tau)) \), over the set of possible \( \tau \). From (9) of
Section II, it is seen that the maximum occurs at the same
point as the maximum of the following function:

\[
\hat{\tau}_{MAP} \overset{\text{def}}{=} \arg \max_{\tau} \left\{ \int_{0}^{T} \left[ e^{(2/N_{x}) \gamma \lambda(t - \tau)} - 1 + f(\tau) \right] \right\}
\]
\[
\times \left[ \lambda(t - \tau) dt + \ln f(\tau) \right]. \quad (11)
\]
B. \textit{Mmse Estimation}

The mmse estimator is the conditional mean $E[\tau|X]$ [25] which, using (9), has the form

$$\hat{\tau}_{\text{mmse}} \overset{\text{def}}{=} \frac{E[X|\tau]f(\tau)}{E[X]} = \frac{\int \tau L(X|\tau)f(\tau) d\tau}{\int L(X|\tau)f(\tau) d\tau}. \quad (12)$$

Substitution of the right-hand side of (9) into the right-hand side of (12) gives

$$\hat{\tau}_{\text{mmse}} = \frac{\int \tau \exp\left(\int_0^\tau dt e^{t/2} \lambda_{\alpha}(\sigma^2)\right) f(\tau) d\tau}{\int \exp\left(\int_0^\tau dt e^{t/2} \lambda_{\alpha}(\sigma^2)\right) f(\tau) d\tau}.$$ \quad (13)

C. \textit{Linear Estimator Structures}

For the purposes of comparison we mention the class of linear estimator structures for time shift. These estimators result from maximization of a linear function of the observed waveform $X$:

$$\hat{\tau}_L = \arg\max\{X(\tau) \times h(\tau)\}, \quad (14)$$

where $h$ is a function of the underlying parameters. Representative examples of $h$ are: the "optical matched filter,

$$h(t) = \frac{\lambda_{\alpha}(-t)}{\lambda_{\alpha}(-t) + \lambda_\alpha + \frac{N_0}{2}}. \quad (15)$$

and robust versions of (15), [12], and the point process domain optimal filter [9]:

$$h(t) = \ln\left(1 + \frac{\lambda_{\alpha}(-t)}{\lambda_\alpha}\right). \quad (16)$$

Often the estimator $\hat{\tau}$ is implemented by detecting a zero crossing of the right-hand side of (14). Examples include the first electron and the constant fraction timing estimators implemented in scintillation counters [10]. These various filter functions $h(t)$ are substantially different both in their form and in their theoretical motivation. The optical matched filter is an optimal linear filter for signal detection in the white noise limited, voltage waveform domain. The point process domain optimal filter gives a maximum-likelihood estimator structure for ideal photon detection. The first electron timing estimator is a suboptimal filter which has been common in particle detection systems such as those implemented in PET. For the purposes of mne comparisons, we will focus on the optical matched filter (15) estimator since it is the optimal linear estimator for the observation model considered.

D. \textit{Discussion}

The approximate expression for the likelihood function $f(\tau)$ becomes exact as the pulse width $\sigma^2$ becomes small. In this case the approximate MAP and mmse estimators (11) and (13) become exact. For the remainder of the discussion, we confine our attention to the MAP estimator and the case of uniform $\tau$, $f(\tau) = \text{constant}$. For notational convenience, we let $\hat{\tau} = \hat{\tau}_{\text{MAP}}$.

A block diagram of the MAP estimator (11) is given in Fig. 3. Observe that the statistic to be maximized is a simple nonlinear function of the observations. The estimator structure can be divided into three sequential tasks, indicated in Fig. 3, accomplished by subsystems A, B, and C respectively: A) a cascade of a noise prewhitening filter $K_w^{-1}$, and a classical linear matched filter of the form $h(t) = p(-t)$ extracts a raw signal $X_A$ by smoothing the noise process $w$ and enhancing the superposition process $\Sigma_{i=1}^n p(t - \tau_i) K_w^{-1}$ is just a multiplicative factor $2/N_0$ for white noise $w$; B) the nonlinearity $\exp(-\cdot)$ has the effect of producing a spike train, $X_B$, which emulates a point process by increasing the dynamic range of $X_A$; C) a correlation of the spike train $X_B$ with $\lambda_\alpha(-\tau)$ generates the log-likelihood ratio, $\ln L(\tau)$, and a subsequent maximization over $\tau$ extracts the time delay parameter.

![Fig. 3. Block diagram of approximate MAP estimator of time shift parameter for uniform $\tau$.](image)

Inferred license and unlimited institutional license. Downloaded on 08/26/20 @ 11:47 from IEEE Xplore. Restricted access.
It is interesting to study the form of the MAP estimator (11) in the limit of small values of the argument of the exponential function: \(2/N\lambda(t) = p(t)\). In this case, using the standard approximation \(e^t = 1 + t\) and the fact that under our assumptions \(\int_0^\infty \lambda(t) dt\) is functionally independent of \(t\) over the support set of \(f(t)\), the MAP estimator reduces to the form of a linear estimation structure (14) with

\[
h(t) \overset{\text{def}}{=} \lambda(t) * p(t).
\]

(17)

For a narrow superposition filter response, \(p(t)\), and for \(A_N \ll 1\), where \(y\) is the PNR defined by (10), \(h(t)\) of (17) is equivalent to the "optical matched filter" (15).

(IV. Estimator Performance)

The relative performance of the MAP and linear estimators is characterized by local estimator bias and local estimator variance. Conditioned on a particular value of \(t\), let \(M\) repeated identical experiments be performed yielding \(M\) independent and identically distributed versions \(\{X(t)\}_i\) of the observation \(X(t); t \in [0, T]\). Let \(\hat{\tau}^{(i)}\) be an estimate of \(\tau\) formed from these \(M\) observations.

As a relevant example consider \(\lambda(t)\) to be periodic with period \(T_0\), and assume that within each period \(\lambda(t)\) is zero except over a small interval of length \(T_0/2\). This model for the signal \(\lambda(t)\) arises in optical telemetry and in optical communication experiments for synchronization. Define \(T_0/2\) the duty cycle of \(\lambda(t)\). If, conditioned on \(\tau\), \(X\) satisfies suitable asymptotic mixing conditions [2], the trajectory of \(X\) over increasingly separated nonoverlapping segments of time are asymptotically conditionally independent. For such a process, the observation of \(X\) over \(M\) successive periods is asymptotically equivalent to \(M\) identical experiments \(\{X(t); t \in ]0, T]\} as \(T \to \infty\). Furthermore, under such mixing conditions it can be directly shown that the appropriate likelihood function \(L(X|\tau)\) of the\(N\)-dimensional functions is the product of \(M\) approximately independent identically distributed (i.i.d) factors \(L(X_i|\tau)\), \(i = 1, \ldots, M\).

Let \(\xi = z_{\xi}\) be a zero of the function \(E[(d/d\xi)\lambda(X|\xi)]\), which is nearest to \(\tau\). Conditioned on \(\tau\), the local variance \(\sigma^2(\tau)\) of the ML estimator \(\hat{\tau}\) is given by

\[
\sigma^2(\tau) = \frac{1}{M} \left[ \frac{E\left[ \left( \frac{d \ln L(X|\tau)}{d \tau} \right)^2 \right]}{d^2 \ln L(X|\tau)} \right].
\]

(18) [26, Theorem 7.2.2B]. With exception of weak consistency, these technical conditions are easily verified for the function \(L(X|\xi)\) displayed in (9) using results derived in Appendix B and using the a.s. continuity of the trajectories of \(X(t)\) w.p. \(-\). Weak consistency of \(\hat{\tau}\) is more difficult and has not been verified. Local weak consistency of \(\hat{\tau}\), on the other hand, is simple to verify using a monotone approximation to \(d/d\xi L(X|\xi)\) in the neighborhood of \(\xi = z_{\xi}\) [26, Lemma 7.2.2A].

Conditioned on \(\tau\) the local variance is now naturally defined as

\[
\text{mse}(\tau) = \sigma^2(\tau) + b^2(\tau).
\]

(19)

If the function \(L(X|\tau)\) were the exact probability density function for \(X\) given \(\tau\), then it could be verified that the local bias is identically zero, and the local variance (18) is identical to the linear estimator variance (17). The above linear error analysis applies to any estimator obtained by maximizing an objective function \(L(X|\tau)\) over \(\tau\) which satisfies the additive decomposition property: \(L(X|\tau) = \sum_{\lambda = 0}^\infty L(X|\tau)\), e.g., the linear estimator structure (14).

A. MAP Estimator Local Bias

We assume that \(\tau\) is uniformly distributed over its domain. If the local bias is small, i.e., the zero \(\xi = z_{\xi}\) of \(\Psi(\xi) = E[(d/d\xi)\lambda(X|\xi)]\) is close to \(\tau\), a first-order linear approximation in the neighborhood of \(\tau = z_{\xi}\) holds:

\[
\Psi(\xi) = \Psi(\tau) - \tau(\tau - z_{\xi}).
\]

Assuming exchange of derivative and expectation is justified, this gives an analytical approximation to estimator local bias \(b(\tau) = z_{\xi} - \tau\):

\[
b(\tau) = \frac{E\left[ \left( \frac{d \ln L(X|\tau)}{d \tau} \right)^2 \right]}{E\left[ \frac{d^2 \ln L(X|\tau)}{d \tau^2} \right]}.
\]

(20)

Expressions for the numerator and denominator of (20) are calculated in Appendix B (18.7) and (18.8). Substitution of these quantities into (20) gives the following expression for the local bias of the approximate MAP estimator \(\hat{\tau}\):

\[
b(\tau) = \int_0^{T_0} \exp \left[ \int_0^{T_0} \lambda(u) du \right] \lambda(\tau) dt - \int_0^{T_0} \exp \left[ \int_0^{T_0} \lambda(u) du \right] \lambda(\tau) dt.
\]

(21)

where \(R_p(\tau) = \int_0^{T_0} \lambda(\tau) \lambda(u) du\) is the pulse autocorrelation function. Note that the local bias is functionally independent of \(\tau\), and the expression (21) is thus equivalent to the unconditional bias \(E[b(\tau)] = E[z_{\xi} - \tau]\). Observe also that the bias of \(\hat{\tau}\) depends on the particular structure of \(\lambda\) and the pulse autocorrelation function \(R_p\). Since autocorrelation functions are symmetric about the origin, it can easily be shown that the local bias is identically zero if \(\lambda\) is symmetric about any point. Hence, for an even signal inten-
sity function $\lambda(t)$, the approximate MAP estimator is locally unbiased.

For the special case of small PNR $\gamma$ (10), the local bias (21) reduces to the simple form

$$b(\tau) = \frac{\int_0^{T_1} dt_1 \int_0^{T_2} dt_2 \dot{\lambda}(t_1) \dot{\lambda}(t_2) \dot{R}_p(t_1, t_2)}{\int_0^{T_1} dt_1 \int_0^{T_2} dt_2 \dot{\lambda}(t_1) \dot{\lambda}(t_2)}$$  \hspace{1cm} (22)

where we have defined the normalized quantities:

$$\dot{\lambda}(t) = \frac{R_p(t)}{\hat{\lambda}(t)}, \quad \dot{\lambda}(t) = \frac{\hat{\lambda}(t)}{\hat{\lambda}(t)},$$  \hspace{1cm} (23)

and $\hat{\lambda}(0) = 0$ is the energy of the pulse $p(t)$. It is important to observe that to a small PNR approximation, the local bias is only dependent on the shape of the intensity function $\lambda(t)$ and the pulse autocorrelation function $R_p(t)$, i.e., bias is independent of the Poisson $\gamma$ and the rate $\lambda$.

B. MAP Estimator Local Variance

Consider the case of uniformly distributed $\tau$ again. The numerator and denominator of (18) are calculated in Appendix B, (B.11) and (B.8) respectively. Thus we have the following expression for the variance of the approximate MAP estimator conditioned on $\tau$:

$$\sigma^2(\tau) = \frac{\int_0^{T_1} dt_1 \int_0^{T_2} dt_2 \exp \left[ \int_0^{T_1} \left[ e^{\gamma \dot{R}(uipersont;\tau)} + \dot{R}_p(uipersont;\tau) \right]^2 \lambda(u) du + \int_0^{T_2} \dot{\lambda}(t_1, t_2) \right] \dot{\lambda}(t_1) \dot{\lambda}(t_2) M \int_0^{T_1} \exp \left[ \int_0^{T_2} \lambda(u) du \right] \dot{\lambda}(t_1) dt_1 \right]^2}{\int_0^{T_1} \int_0^{T_2} \dot{\lambda}(t_1) \dot{\lambda}(t_2)},$$  \hspace{1cm} (24)

where $\dot{\lambda}$ and $\dot{R}_p$ are defined in (23).

We make the following observations based on the local variance (24). Similarly to the local bias, the local variance (24) is functionally independent of $\tau$. Note that, since $R_p(uipersont;\tau) = 0$, $|uipersont;\tau| > T_1$, and $\lambda(u)$ do not jointly affect the local variance for $|uipersont;\tau| > T_1$. The same property holds for the joint intensity of $\lambda(t)$, $\lambda(u)$, and $\lambda(t)$. However, memory smears the intensity function over time as it affects its local variance. This directly affects the variance of the MAP estimator. This is different from the pure Poisson observations case, where it can be shown that the local variance and the local bias do not depend on the correlation between different time samples of $\lambda(t)$ (see also the form (27) of the Poisson limited CR bound in Section V).

Due to the presence of integrands of the form $e^{\gamma \dot{R}_p(uipersont;\tau)}$ in the numerator and denominator of (24), the numerical computation of the local approximation is impracticable for the case of large PNR-rate product $\gamma A$. For the case of small PNR-rate product, $\gamma A \ll 1$, it is shown in Appendix B that the expression (24) for local variance reduces to the expression:

$$\sigma^2(\tau) = \frac{1}{\gamma A \int_0^{T_1} dt_1 \int_0^{T_2} dt_2 \dot{\lambda}(t_1) \dot{\lambda}(t_2)}$$  \hspace{1cm} (25)

Under the assumed low $\gamma A$ conditions, it can easily be shown that the expression (25) is equivalent to the local variance for the linear matched filter estimator (17). For low PNR-rate product, a rate of decay of the variance on the order of $A^2$ as a function of the energy $\Lambda$ of the Poisson process is indicated by (25). The low PNR expression for local variance (25) is monotonically decreasing in the PNR $\gamma$. It is also decreasing in the second derivative of the normalized intensity function $\dot{\lambda}(t)$ for those values of $t$ such that $\dot{\lambda}(t)$ is large, $|t - t| \leq T_1$. On the basis of these observations, the estimator can be expected to have low variance for the cases where: 1) the process $\dot{\lambda}(t)$ is easy to estimate by virtue of high PNR; 2) the intensity function $\lambda(t)$ is highly resolved in the sense that $\lambda(t)$ is high and $\lambda(t)$ is "sharply peaked," i.e., $\lambda(t) \gg 0$, in $t$-regions of high intensity.

V. LOWER BOUNDS ON MSE

Here we derive lower bounds on estimator error for the estimation problem outlined in Section II. Two bounds will be presented: the Cramer-Rao (CR) lower bound, derived under the optimistic Poisson limited (high PNR) regime, and a rate distortion bound. Both of these bounds specify a lower limit on achievable mse for the general Gaussian-Poisson regime and, unlike the MAP approximation (9), are applicable to arbitrary pulse density conditions.

A. CR Bound

The CR bound on $\tau$ is given by the inequality

$$\text{mse} \geq \frac{1}{E \left[ \frac{d}{d\tau} \ln L(X(t)) f(\tau) \right]},$$  \hspace{1cm} (26)

where $L(X(t))$ is the likelihood ratio for the hypotheses (6). For the case of pure Poisson observations, $X(t)$, the CR bound is derived in [14]:

$$\text{mse} \geq \frac{1}{\lambda(t) dt} + E \left[ \frac{\dot{\lambda}(t)}{dt} \right],$$  \hspace{1cm} (27)

where it has been assumed that derivatives in (27) exist. Since the addition of Gaussian noise to the observations only introduces additional $\gamma$-uncoupled nuisance parameters to the estimation problem, the Poisson limited CR bound (27) is less than or equal to the exact but uncomputable CR bound valid for the observation model of Section II [4]. Unfortunately, the right-hand side of (27) fails to be a useful bound for a large class of intensity functions, e.g., biexponential, rectangular, or any intensity for which the ratio of the squared first derivative and the intensity magnitude is not absolutely integrable.
Fig. 4. Effects of mapping random timing parameter $\tau$ into observation space as decomposed into cascades of two transformations: 1) channel $C_1$ accounts for generation of point process through Poisson mechanism, this is point process channel; 2) channel $C_2$ accounts for finite bandwidth and additive noise effects introduced via detection process, this is continuous observation channel.

**B. Rate-Distortion Type Lower Bound**

For the case of perfect Poisson observations $X = \{t_i\} \rightarrow$, a rate-distortion type lower bound was derived in [14]. The bound in [14] was derived from a special case of Shannon's inequality

$$ H(\tau) + \frac{1}{2} \ln (2\pi e \text{mse}) \leq R(\text{mse}) \leq C. \quad (28) $$

where $C$ is the capacity of a channel taking the source symbols $\tau$ to destination symbols $X$, $R$ is the rate-distortion function associated with $\text{mse}$ distortion measure, and $H(\tau)$ is the entropy of the p.d.f. of $\tau$. Shannon's inequality (28) gives a lower bound on $\text{mse}$ in terms of $C$:

$$ \text{mse} \geq \frac{1}{2\pi e} e^{2H(\tau)+C2e}. \quad (29) $$

In the present situation, the channel $C$ can be represented as the cascade of two separate channels $C_1$ and $C_2$, where $C_1$ takes the source symbols $\tau$ into the occurrence times $(t_i)_{i=1}^n$, and $C_2$ takes the occurrence times $(t_i)_{i=1}^n$ into the observations $X$ (see Fig. 4). Hence, $C_1$ is an intensity modulated Poisson point process channel with associated intensity $\lambda(t-\tau)$, while $C_2$ is a Gaussian white noise channel with impulse response $p(-t)$ and Poisson input statistics. An upper bound on the overall capacity $C$ of this cascaded channel is given by the "data processing theorem" [3] as

$$ C \leq \min\{C_1, C_2\}. \quad (30) $$

Furthermore, an explicit upper bound, $C^*$, on $C$ was given in [14]:

$$ C_1 \leq C^*_1 = A \int_0^T \lambda(t) \ln \frac{\lambda(t)}{\hat{\lambda}(t)} \, dt. \quad (30) $$

In (30) the integral quantity is the nonnegative "information divergence" between $\lambda$ and the uniform normalized intensity $1/T$ [3]. This can be interpreted as a (asymmetric) distance measure between the two intensities.

An upper bound on $C_2$ is obtained by recalling the decomposition formula for the capacity of a channel with input $Z$ and output $X$, which is the sum of a "signal" $S = g(Z)$ plus an independent additive noise $w$: $X = g(Z) + w$. Specifically:

$$ C_2 = \sup_{P(Z)} \{ H(X) - H(X|Z) \} $$

$$ = \sup_{P(Z)} \{ H(X) - H(w|Z) \} $$

$$ = \sup_{P(Z)} \{ H(X) - H(w) \}. \quad (31) $$

where $H(X)$ and $H(X|Z)$ are the entropy of $X$ and the conditional entropy of $X$ given the input $Z$ respectively, $P(X|Z)$ is a conditional distribution, and $H(w)$ is the entropy of the noise. In (31) we have used the independence of $Z$ and $w$ to equate the conditional entropy $H(w|Z)$ to the noise entropy $H(w)$, which is independent of $P(X|Z)$. On the other hand, it is well known that, for a fixed output autocovariance function $K_X = K_x + K_w$, the entropy $H(X)$ is maximized for Gaussian $X$ [3]. Hence the channel capacity $C_2$ is upper bounded by the capacity of a Gaussian channel.

Here $Z$ can be identified with the point process sequence $(z_i)_{i=1}^n$, and $S(t) = g(Z) = \sum_{i=1}^n p(t-t_i)$, where $n = N(t)$. The autocovariance $K_z(z_1, z_2)$, at times $z_1$ and $z_2$, is computed in Appendix C (C.4):

$$ K_z(z_1, z_2) = \int_0^{\min(z_1, z_2)} p(z_1-u)p(z_2-u)\lambda(u) \, du $$

$$ + \int_0^{\min(z_1, z_2)} \int_0^{\min(z_2, z_1)} p(z_1-u_1)p(z_2-u_2) \text{cov}[\lambda(u_1-\tau), \lambda(u_2-\tau)], \quad (32) $$

where $\lambda(u) = E[\lambda(u-\tau)]$; and the expectation in "cov" is over the random variable $\tau$. For the special case that $K_z(z_1, z_2)$ depends only on $z_1 - z_2$, i.e., $S$ is covariance stationary, $C_2$ is upper bounded by the capacity of the stationary Gaussian channel with capacity $C_2^*$ [3]:

$$ C_2 \leq C_2^* = \frac{1}{4\pi} \int_{-\infty}^{\infty} \ln \left(1+\frac{2G_0(\omega)}{N_0}\right) d\omega, \quad (33) $$

where $G_0$ is the power spectral density of the signal $S$, i.e., the Fourier transform of the covariance $K_s$. For the general case of nonstationary $K_s$, a more complicated channel capacity formula involving the eigenvalues of the Karhunen-Love expansion associated with $K_s$ must be used [3].

Combination of (30) and (33) gives, from (29), the rate distortion lower bound:

$$ \text{mse} \geq B_{\text{dist}} = \frac{1}{2\pi e} e^{2H(\tau)+2\text{min}(\tau, C^2)}. \quad (34) $$

The rate distortion lower bound (34) has some interesting features. Assume that $\tau$ is uniformly distributed over $[0, T]$ and that $p(t)$ is a causal impulse response. In this case (see Appendix C) $\lambda(t) = \Lambda(t)$, $t \in [0, T]$ and, as $T \rightarrow \infty$,

$$ G_s(\omega) = \overline{p^2 G_0(\omega)} \quad (35) $$

where

$$ \hat{G}_s(\omega) = \left[ \hat{P}(\omega)^2 + \left| \hat{S}(\omega) - \hat{\lambda}(\omega) \right|^2 \right] $$

$$ = \left| \hat{P}(\omega)^2 + \left| \Lambda + \lambda(\omega) - \lambda(t) \right|^2 \right| \quad (36) $$

and $\text{sinc}(x) = \sin(x)/x$. In (36), $\Lambda(\omega) = \mathcal{F}(\lambda(t))$ is the Fourier transform of $\lambda$, hence $\Lambda(0) = \Lambda$, the integrated rate of the point process. Also the pulse energy normalized Fourier transform of the superposition filter $p(t)$ has been defined: $\hat{P}(\omega) = \mathcal{F}(p(t))/\sqrt{\int p^2(t) dt}$. Note that the minimum value $G_s(\omega) = \overline{p^2 G_0(\omega)}$ is attained for the case that $\lambda(t) = \lambda(t)$. Using (35) in the bound (34) we obtain
the final form:

\[ B_{\text{dB}} = \begin{cases} \frac{1}{2\pi T e^{2\lambda r}} \exp \left( -2\lambda \int_0^T \tilde{\lambda}(t) \ln \left( \frac{1}{\gamma_T} \right) dt \right), & \gamma > \gamma_o, \\ \frac{1}{2\pi T e^{2\lambda r}} \exp \left( -\frac{1}{2\pi T} \int_{-\infty}^{\infty} \ln \left( 1 + \gamma \hat{G}(\omega) \right) d\omega \right), & \gamma \leq \gamma_0, \end{cases} \]

(37)

where \( \gamma_o \) is a PNR threshold determined by the condition \( C_1^* = C_2^* \) in (34). Specifically, \( \gamma_o \) is the solution of the equation:

\[ \Lambda \int_0^T \tilde{\lambda}(t) \ln \left( \frac{1}{\gamma_T} \right) dt = \frac{1}{4\pi} \int_{-\infty}^{\infty} \ln \left( 1 + \gamma_o \hat{G}(\omega) \right) d\omega, \]

(38)

when the solution exists. An important case where the solution of (38) does not exist is when the superposition filter \( \rho(t) \) approaches a delta function, i.e., \( \hat{P}(\omega) \equiv \text{constant} \). In this case, since the right-hand side of (38) diverges, \( C_1^* < C_2^* \), and hence \( \min(C_1^*, C_2^*) = C_1^* \), for all \( \gamma > 0 \). Therefore, \( \gamma_o = 0 \)

for a zero width pulse \( \rho(t) \) and nonzero PNR, and the lowerbound (37) is identical to the rate distortion lower bound of [14] for the pure Poisson observation \( X = \{ x_n \}_{n=1}^\infty \).

The lower bound (37) separates the performance into two PNR regimes: the Poisson limited regime (\( \gamma > \gamma_o \)), and the Gaussian limited regime (\( \gamma \leq \gamma_o \)). In the high PNR Poisson limited regime, we have a PNR independent bound that is independent of the superposition filter \( \rho(t) \) and depends principally on the information divergence between \( \lambda \) and the worst case uniform intensity \( 1/T \) over \([0, T]\). The closer \( \lambda \) is to the uninformative uniform intensity, the poorer becomes the estimator mse. This Poisson limited bound decays to zero at an exponential rate as the point process rate \( \Lambda \) increases. This rate is controlled by the information divergence, which is the magnitude difference between the entropy of the normalized intensity \( \lambda \) and the normalized maximum entropy uniform intensity. On the other hand, in the low PNR Gaussian limited regime, we have a bound that depends on the pulse shape through its magnitude Fourier transform, and depends on the information function through the magnitude squared difference between the Fourier transforms of \( \lambda \) and the uniform intensity \( \Lambda \). Observe that, unlike the Poisson limited case, the decrease of the rate distortion bound for \( \gamma < \gamma_o \) is subexponential in \( \Lambda \) for large \( \Lambda \). In both the Poisson and the Gaussian limited regimes the effect of prior information on \( \tau \) is manifested through the quantity: \( \exp(2H(\tau))/2\pi e \), which is the \textit{a priori} "entropy power" of \( \tau \) [5]. The entropy power of \( \tau \) is the variance of a Gaussian random variable with equal entropy as \( \tau \). Thus the bound (37) provides an indication of the \textit{a posteriori} variance reduction achievable by making use of the observations \( X \).

VI. NUMERICAL COMPARISONS

In this section the mse performance of the MAP estimator (11), the linear "optimal matched filter" (15), and the lower bounds (27) and (37) are investigated numerically.

Using the expressions (21) and (24) with the identification \( \ln(L(X|\theta) = X(t) - h(t)) \), we have the following forms for local bias and local variance of a linear estimator structure of the form (14):

\[ b(\tau) = \frac{-\int_0^T h(-t) \tilde{\lambda}(t) \delta(t) dt}{\int_0^T h(-t) \tilde{\lambda}(t) \delta(t) dt}, \]

(39)

\[ \sigma^2(\tau) = \frac{1}{M} \int_0^T dt \int_0^T dt' h(-t) h(-t') K_1(t, t') + \frac{N_0}{2} \delta(t_1 - t_2), \]

(40)

where \( K_1 \) is the covariance:

\[ K_1(t_1, t_2) = \int_0^T \rho(t - z_1) \rho(t - z_2) \lambda(t) dt. \]

A numerical study was performed for the case that the signal intensity function \( \lambda \) is a Gaussian pulse with one sided standard width \( T_s \) and \( M = 1 \). The superposition filter response \( \rho(t) \) is a Gaussian pulse with one sided exponential with time constant \( T_a \). Due to the symmetry of \( \lambda \), both the approximate MAP estimator (11) and the optimal matched filter (15) are locally unbiased. Fig. 5 shows the local mse of the MAP estimator (24) and the local mse of the optimal matched filter (40) as functions of Poisson rate \( \lambda = \lambda_e + \lambda_o \) for the following parameters: Poisson signal-to-noise energy ratio \( \rho = \lambda_e/\sigma_e = 50 \); PNR \( \gamma = 0 \) dB; intensity time-width to filter time-width ratio \( T_a/T_s = 333 \); a priori interval-width to intensity time-width ratio \( T_a/T_s = 100 \). Observe that for low Poisson rate, \( \lambda \ll 5 \) dB, the matched filter and the MAP estimator have similar local mse performance. For higher \( \lambda \), however, the MAP estimator has uniformly smaller local mse than the linear matched filter. The improvement of the MAP local mse over the matched filter local mse can be over 10 dB for high \( \lambda \). While this dramatic improvement may not occur in the global mse, which takes large errors into account, Fig. 5 is suggestive of performance gains. Also shown for comparison are the CR bound (27) and the rate distortion bound (37). For \( \lambda \) below a rate threshold of 10 dB, the rate distortion lower bound is a tighter bound. This threshold specifies the lower boundary of the \( \lambda \) region over which
Fig. 5. Local mse approximations for matched filter (mf), MAP (map), rate distortion lower bound (rdlb), and Cramer–Rao lower bound (crib), as functions of rate \( \lambda \) of Poisson process. Poisson SNR, denoted \( \rho \), is 50 dB and PNR = \( \gamma \) is 0 dB. mse axis has been normalized so that 0.0 corresponds to the mse of uniform random variable over a priori interval [0, 100].

Fig. 6. Local mse approximations and lower bounds as function of rate \( \lambda \). Poisson SNR, denoted \( \rho \), is 50 dB and PNR = \( \gamma \) is –3 dB. mse axis is normalized as in Fig. 5.
small errors are theoretically attainable. A decrease of $A$ below this threshold implies a sudden and precipitous increase in the achievable mse. In Fig. 6 the PNR has been decreased to $-3$ dB below the PNR of Fig. 5. The effect of the decreased PNR on the lower bound is a raising of the low $A$ threshold that decreases the range of $A$, over which small mse is theoretically achievable. The local mse and the CR bound as functions of the Poisson signal-to-noise ratio $\rho$ for a fixed total rate of $\lambda = 20$ dB and PNR = 0 dB are shown in Fig. 7. The remaining parameters are identical to the ones used to generate Fig. 5. Note that the local mse of the MAP estimator has similar appearance to the CR bound, both curves displaying abrupt thresholds at approximately the same value of SNR. As before, the local mse performance of the MAP estimator is uniformly lower than that of the matched filter estimator. For these numerical evaluations, the PNR-rate product $\gamma A$ had to be held sufficiently low to avoid numerical overflow in the numerator and denominator of the MAP local mse expression (24). For the values of $\gamma A$ studied, it was observed that in the exponent of the rate distortion bound (34), we have $C_j^* > C_j^+$; that is, the bound was active only for the Gaussian limited regime. At higher values of PNR the Poisson limited rate distortion bound would normally become active. Additional numerical studies of the rate distortion lower bound are presented in [13].

VII. CONCLUSION

Approximations to the likelihood function, the MAP estimator, and the mmse estimator have been given for the time-shifted intensity function of a causally filtered Poisson process observed in additive Gaussian noise. The approximation becomes more accurate as the per-unit-time density of superimposed filter impulse responses becomes small. The MAP estimator has a simple nonlinear structure as a function of the observations. First, it attempts to enhance the filtered Poisson process via classical matched filtering. Then, a memoryless nonlinear transformation produces a spike train that emulates the underlying point process. Finally, the spike train is correlated against shifted versions of the intensity function. The maximizing shift provides a time delay estimate. For smooth intensities, a local analysis of the bias and mse of the approximate likelihood ratio statistic indicated the conditions under which good estimation and detection performance can be obtained: 1) a high pulse-to-noise power ratio; 2) a high Poisson signal-to-Poisson noise ratio; 3) a large second derivative of the intensity function over regions where the amplitude of the intensity is large.

A rate-distortion type lower bound on the mse of any time estimator was derived. This bound is nontrivial for some important cases where the CR bound gives trivial results. The rate distortion bound indicates the importance of several factors on the inherent estimability of the time delay. First, for high pulse-to-noise ratio, the mse bound is Poisson noise limited and it decreases exponentially as a function of Poisson rate, where the rate of decay is the "information divergence" between the intensity function and a uniform intensity over time. The higher the information divergence, i.e., the more the intensity function differs from a uniform intensity, the better the potential mse performance. Second, for low pulse-to-noise power ratio, the mse bound decreases subexponentially as a function of Poisson rate, where the mean-squared difference between the intensity and a uniform intensity over time, and the shape of the superposition filter impulse response govern the rate of decay. The more broadband the filter transfer function, the better is the potential mse performance. Third, there is a pulse-to-noise...
ratio (PNR) threshold that specifies a PNR boundary between Poisson limited and Gaussian limited use performance.

A numerical evaluation of the local mse approximations and the lower bounds indicates that the rate distortion bound is tighter than the CR bound for the larger error regime, where the opposite is true for the small error regime.

A comparison between the local mse of the approximate MAP estimator and the linear "optimal matched filter" indicates that the MAP estimator can have better local error performance than the optimal linear filter. A large error sensitivity analysis should be performed to more completely characterize the advantages of the approximate MAP structure for detection and estimation of time shift.

Acknowledgment

The author gratefully acknowledges helpful discussions with W. L. Rogers and N. H. Clinthorne of the Division of Nuclear Medicine, University of Michigan, Ann Arbor, MI.

Appendix A

Derivation of Approximate Likelihood Function

We recall the following fact about a Poisson process with intensity \( \lambda \) [27]. Given the total number of points \( n \) over \([0,T]\), the (unordered) occurrence times \( t_1, \ldots, t_n \) are independent identically distributed random variables over \([0,T]\) with marginal probability density functions \( f(t_i) = \lambda e^{-\lambda t} dt \). The following identity will be needed [27]:

\[
E_{t_1,\ldots,t_n} \left[ \prod_{i=1}^{n} Q(t_i) \right] = e^{\int_{0}^{T} \lambda(t) \left( Q(t) - 1 \right) dt},
\]

where \( Q(t) \) is an arbitrary (integrable) function. If \( Q(t) \) is independent of \( t \), (A.1) specializes to the more useful formula:

\[
E_{t_1,\ldots,t_n} [O^n] = e^{\int_{0}^{T} \lambda(t) (Q(t) - 1) dt}.
\]

Another identity that will be useful is

\[
E_{t_1,\ldots,t_n} \left[ \sum_{i=1}^{n} p(t_i - t) \right] = \int_{0}^{T} \lambda(u) [1 - \lambda(u)] du.
\]

In this Appendix we approximate the numerator and denominator in (8) of the likelihood ratio \( L(X|\tau) = f(X|H_1,\tau) / f(X|H_0) \) are derived. In the following I(A) denotes the indicator function of the set \( A \), \( R_p = \int_{0}^{T} p(t) dt \) is the pulse auto-correlation function, and \( \lambda^2 = R_p(0) \) is the energy of the pulse \( p(t) \). We start from (7) by expanding the double sum (\( \Sigma \)) and, using by definition of \( T_p \), \( R_p(t) = 0 \) for \( |t| > T_p \):

\[
f(X|H_0,\tau) = \sum_{t_1,\ldots,t_n} p(t_1 - t) \ldots p(t_n - t) f(X|H_0,\tau) dt_1 \ldots dt_n
\]

\[
= \sum_{t_1,\ldots,t_n} e^{\int_{0}^{T} \lambda(t) (Q(t) - 1) dt} \ldots e^{\int_{0}^{T} \lambda(t) (Q(t) - 1) dt} \prod_{i=1}^{n} \left[ 1 - \int_{0}^{T} f(X|H_0,\tau) dt \right]
\]

\[
= \sum_{t_1,\ldots,t_n} e^{\int_{0}^{T} \lambda(t) (Q(t) - 1) dt} \ldots e^{\int_{0}^{T} \lambda(t) (Q(t) - 1) dt} \prod_{i=1}^{n} \left[ 1 - \int_{0}^{T} f(X|H_0,\tau) dt \right]
\]

\[
= \sum_{t_1,\ldots,t_n} e^{\int_{0}^{T} \lambda(t) (Q(t) - 1) dt} \ldots e^{\int_{0}^{T} \lambda(t) (Q(t) - 1) dt} \prod_{i=1}^{n} \left[ 1 - \int_{0}^{T} f(X|H_0,\tau) dt \right]
\]

\[
= \sum_{t_1,\ldots,t_n} e^{\int_{0}^{T} \lambda(t) (Q(t) - 1) dt} \ldots e^{\int_{0}^{T} \lambda(t) (Q(t) - 1) dt} \prod_{i=1}^{n} \left[ 1 - \int_{0}^{T} f(X|H_0,\tau) dt \right]
\]

where

\[
G(X, (t_i)_{i=1}^{n}) \triangleq \left( e^{\int_{0}^{T} \lambda(t) (Q(t) - 1) dt} \right) \prod_{i=1}^{n} \left[ 1 - \int_{0}^{T} f(X|H_0,\tau) dt \right].
\]

The random variable \( \prod_{i=1}^{n} I[|t_i - t| > T_p] \) is equivalent to the random variable \( G(X, (t_i)_{i=1}^{n}) \), where \( z_k \) is the increment from the \( k \)th largest occurrence time to the \((k+1)\)st largest occurrence time. Observe the following chain of inequalities:

\[
E \left[ 1 - \prod_{i=1}^{n} I[|t_i - t| > T_p] \right] \leq E \left[ \sum_{i=1}^{n} E \left[ I[z_i > T_p] \right] \right]
\]

\[
= E \left[ \sum_{i=1}^{n} \max_{t} P(N(t, t + T_p) > 0) \right]
\]

\[
\leq E \left[ \sum_{i=1}^{n} \max_{t} P(N(t, t + T_p) > 0) \right]
\]

\[
= E \left[ \sum_{i=1}^{n} \left[ 1 - \int_{0}^{T} \lambda u (Q(t) - 1) dt \right] \right]
\]

\[
= \int_{0}^{T} \lambda u (Q(t) - 1) dt,
\]

(A.6)

where \( \lambda(t, \tau) \) is the max \( \int_{0}^{T} \lambda(u - \tau) du \) and \( \lim_{\tau \to \infty} \lambda(u) = 0 \). Defining

\[
\lambda^* \triangleq \max_{t} \int_{0}^{T} \lambda(u - t) du = \max_{t} \int_{0}^{T} \lambda(u) du,
\]

(A.4)
after applying the Schwarz inequality and (A.6) to the second term on the right-hand side of (A.4) it is seen that:

\[
E_{t_i-1}^2 \left[ G(X(t_i),x) \right] [1 - \prod_{j=1}^{n} \left[ \exp \left( 2 \sum_{j=1}^{n} \left( \int_0^T X(t) \rho(t - t_j) dt \right) \right) \right]_\tau] \\
\leq E_{t_i-1} \left[ G(X(t_i),x) \right] [1 - \prod_{j=1}^{n} \left[ \exp \left( \frac{1}{2} \sum_{j=1}^{n} \left( \int_0^T X(t) \rho(t - t_j) dt \right) \right) \right]_\tau] \\
= O(\Lambda T^\tau). 
\]  
(A.7)

Hence, using (A.7) in (A.4):

\[
f(X|H_\tau) = E_{t_i-1} \left[ \exp \left( 2 \sum_{j=1}^{n} \left( \int_0^T X(t) \rho(t - t_j) dt \right) \right) \right]_\tau] \\
+ O(\Lambda T^\tau). 
\]  
(A.8)

Thus to order \(O(\Lambda T^\tau)\) we have, using the identity (A.1),

\[
f(X|H_\tau) = E_{t_i-1} \left[ \prod_{j=1}^{n} \exp \left( \int_0^T \left( e^{2/N_x} T_X(t)(u - t_j) du - \frac{1}{2} u^2 \right) \right) \right]_\tau] \\
= \exp \left( \int_0^T \left( e^{2/N_x} T_X(t)(u - t_j) du - \frac{1}{2} u^2 \right) \right) \lambda(t - \tau) dt \\
= \exp \left( \int_0^T \left( e^{2/N_x} T_X(t)(u - t_j) du - \frac{1}{2} u^2 \right) \right) \lambda(t - \tau) dt. 
\]  
(A.9)

To obtain an order \(O(\Lambda T^\tau)\) approximation to \(f(X|H_\tau)\), simply replace \(\lambda(t - \tau)\) in (A.9) by the constant Poisson intensity \(\lambda_\tau\):

\[
f(X|H_\tau) = \exp \left( \int_0^T \left( e^{2/N_x} T_X(t)(u - t_j) du - \frac{1}{2} u^2 \right) \lambda_\tau dt \right). 
\]  
(A.10)

Recalling the relation (1), \(\lambda(t - \tau) = \lambda(t - \tau) + \lambda_\tau\), the ratio of (A.9) and (A.10) reduces to the expression (9).

**APPENDIX B**

**DERIVATION OF MOMENTS OF LOG-LIKELIHOOD DERIVATIVES**

In this Appendix we calculate the numerator and denominator quantities on the right-hand sides of (20) and (19) using the approximate form for the log-likelihood given in (9):

\[
\ln L(X) = \left[ t_0 \left( e^{2/N_x} T_X(t)(u - t_j) du - \frac{1}{2} u^2 \right) \right] \lambda(u - \tau) du - \lambda(t - \tau) dt - \Lambda. 
\]  
(B.1)

For notational simplicity, we will denote \(T_i = T\) and \(X_i = X\).

Take the first derivative with respect to \(\tau\) of the log-likelihood function approximation (B.1) and take the expectation

\[
E \left[ \frac{d \ln L(X|\tau)}{d \tau} \right]_\tau \right] \\
= -e^{-\lambda(t - \tau)} \int_0^T \left( e^{2/N_x} T_X(t)(u - t_j) du \right) \lambda(u - \tau) du, 
\]  
(B.2)

where we have used the fact \(\lambda_\tau = \lambda\). Consider the expectation within the \(\int\) on the right-hand side of (B.2). We have by definition, \(X = \sum_{i=1}^{n} p(t - t_i) + w\), and hence:

\[
E \left[ e^{2/N_x} T_X(t)(u - t_j) du \right] \\
= E \left[ e^{2/N_x} T_X(t)(u - t_j) du \right] \\
= E \left[ e^{2/N_x} T_X(t)(u - t_j) du \right] \\
= e^{2/N_x} T_X(t)(u - t_j) du. 
\]  
(B.3)

Now since, by assumption, the Gaussian noise \(w\) is independent of the Poisson occurrence times \(t_i\) and \(\tau\), we have

\[
E \left[ e^{2/N_x} T_X(t)(u - t_j) du \right]_\tau \right] \\
= e^{2/N_x} T_X(t)(u - t_j) du. 
\]  
(B.4)

Now the exponent of the argument of "\(E\)" in (B.4) is a zero-mean Gaussian random variable with variance \(\sigma^2 = \frac{\sigma^2}{2} = \frac{\sigma^2}{2}N_o\). Using the well-known form of the characteristic function of such a Gaussian random variable, \(\exp(\sigma^2/2\lambda)\), we have from (B.4)

\[
E \left[ e^{2/N_x} T_X(t)(u - t_j) du \right]_\tau \right] \\
= e^{2/N_x} T_X(t)(u - t_j) du. 
\]  
(B.5)

Substitution of (B.5) into (B.3) gives:

\[
E \left[ e^{2/N_x} T_X(t)(u - t_j) du \right] \\
= e^{2/N_x} T_X(t)(u - t_j) du. 
\]  
(B.6)

where in the fourth line of (B.6) the identity (A.1) has been used. Substitution of (B.6) into (B.2) gives the following:

\[
E \left[ \frac{d \ln L(X|\tau)}{d \tau} \right]_\tau \right] \\
= -e^{-\lambda(t - \tau)} \int_0^T \left( e^{2/N_x} T_X(t)(u - t_j) du \right) \lambda(u - \tau) du \\
= -e^{-\lambda(t - \tau)} \int_0^T \left( e^{2/N_x} T_X(t)(u - t_j) du \right) \lambda(u - \tau) du. 
\]  
(B.7)

The calculation of

\[
E \left[ \frac{d^2 \ln L(X|\tau)}{d\tau^2} \right]_\tau \right] 
\]
is similar to the calculation of
\[ E \left[ \frac{d \ln \mathcal{L}(X|\tau)}{d\tau} \right] \tau; \]
the details are omitted. The result is
\[ E \left[ \frac{2 d \ln \mathcal{L}(X|\tau)}{d\tau} \right] \tau \]
\[ = -e^{-\lambda} \int_0^T \exp \left( \int_0^t \frac{2d\mathcal{L}}{d\tau} \right) \lambda(u) \, du \, \tau \, dt. \]  \hspace{1cm} (B.8)

Next the numerator of (19) is derived. From (B.2):
\[ E \left[ \frac{d \ln \mathcal{L}(X|\tau)}{d\tau} \right] \tau \]
\[ = e^{-2\lambda} \int_0^T dt \int_0^T dt_2 \]
\[ \cdot \left[ e^{2\lambda} \mathcal{L}(X|\tau_1) \mathcal{L}(X|\tau_2) \right] \tau \]
\[ \cdot \lambda'(z_1 - \tau) \lambda'(z_2 - \tau). \] \hspace{1cm} (B.9)

Finally, it is shown that the local mse approximation (24) reduces to the expression (25) of Section IV-B under the small PNR assumption. For convenience we repeat the expression (24):
\[ E \left[ \left( \hat{\lambda}(\tau) - \lambda(\tau) \right)^2 \right] = \int_0^{T_2} dt \int_0^{T_2} dt_2 \exp \left( \int_0^{T_2} \left( e^{\lambda \hat{\lambda}(u, \tau_1) + \hat{\lambda}(u, \tau_2) - 1} \lambda(u) \, du + \gamma \hat{\lambda}_p(t_1 - t_2) \right) \hat{\lambda}(u) \, du + \frac{1}{N_0} \gamma \hat{\lambda}_p(t_1 - t_2) \right) \hat{\lambda}(u) \, du \right)^2 \]
\[ M \int_0^{T_2} \exp \left( \int_0^{T_2} \left( e^{\lambda \hat{\lambda}(u, \tau_1) + \hat{\lambda}(u, \tau_2) - 1} \lambda(u) \, du + \gamma \hat{\lambda}_p(t_1 - t_2) \right) \hat{\lambda}(u) \, du + \frac{1}{N_0} \gamma \hat{\lambda}_p(t_1 - t_2) \right) \hat{\lambda}(u) \, du \right)^2 \]  \hspace{1cm} (B.12)

Consider the expectation on the right-hand side of (B.9). In a manner analogous to the derivation of
\[ E \left[ \frac{d \ln \mathcal{L}(X|\tau)}{d\tau} \right] \tau \]

find
\[ E \left[ \frac{2 d \ln \mathcal{L}(X|\tau)}{d\tau} \right] \tau \]
\[ = E \left[ \frac{2 d \ln \mathcal{L}(X|\tau)}{d\tau} \right] \tau \]
\[ = \int_0^T dt \int_0^T dt_2 \exp \left( \int_0^T \left( e^{\lambda \hat{\lambda}(u, \tau_1) + \hat{\lambda}(u, \tau_2) - 1} \lambda(u) \, du + \gamma \hat{\lambda}_p(t_1 - t_2) \right) \hat{\lambda}(u) \, du + \frac{1}{N_0} \gamma \hat{\lambda}_p(t_1 - t_2) \right) \hat{\lambda}(u) \, du \right)^2 \]  \hspace{1cm} (B.13)

where \( \gamma \) is the PNR defined in (10). Make the following low
PNR substitutions in the numerator and denominator of
(B.12):
\[ \exp \left( \int_0^T \left( e^{\lambda \hat{\lambda}(u, \tau_1) + \hat{\lambda}(u, \tau_2) - 1} \lambda(u) \, du + \gamma \hat{\lambda}_p(t_1 - t_2) \right) \hat{\lambda}(u) \, du \right) \]
\[ = 1 + \gamma \Lambda \int_0^T \hat{\lambda}_p(u - t_1) + \hat{\lambda}_p(u - t_2) \hat{\lambda}(u) \, du \]  \hspace{1cm} (B.14)

and:
\[ \exp \left( \int_0^T \left( e^{\lambda \hat{\lambda}(u, \tau_1) + \hat{\lambda}(u, \tau_2) - 1} \lambda(u) \, du + \gamma \hat{\lambda}_p(t_1 - t_2) \right) \hat{\lambda}(u) \, du \right) \]
\[ = 1 + \gamma \Lambda \int_0^T \hat{\lambda}_p(t_1 - t_2) \hat{\lambda}(u) \, du \]  \hspace{1cm} (B.15)

Since the integrals of \( \lambda' \) and \( \lambda' \) are identically zero the substitution of the approximations (B.13) and (B.15) into (B.12) gives:
\[ E \left[ \left( \hat{\lambda}(\tau) - \lambda(\tau) \right)^2 \right] \]
\[ = \frac{1}{\gamma \Lambda} \int_0^T dt \int_0^T dt_2 \hat{\lambda}(t_1) \hat{\lambda}(t_1 - t_2) \hat{\lambda}(t_2) \]  \hspace{1cm} (B.16)
Next note that, due to the assumption $T = T_a \gg T_b$, and the differentiability of $\lambda$:

$$0 = \frac{d^2}{dt^2} \int dt_1 \int dt_2 \lambda(t_1) R_p(t_1 - t_2) \lambda(t_2)$$

$$= \frac{d^2}{dt^2} \int dt_1 \int dt_2 \lambda(t_1 - \tau) R_p(t_1 - t_2) \lambda(t_2 - \tau)$$

$$- \int dt_1 \int dt_2 \lambda(t_1 - \tau) R_p(t_1 - t_2) \lambda(t_2 - \tau)$$

$$+ \int dt_1 \int dt_2 \lambda(t_1 - \tau) R_p(t_1 - t_2) \lambda'(t_2 - \tau)$$

$$= \int dt_1 \int dt_2 \lambda'(t_1) R_p(t_1 - t_2) \lambda(t_2)$$

$$+ \int dt_1 \int dt_2 \lambda(t_1) R_p(t_1 - t_2) \lambda'(t_2).$$  \hspace{1cm} (B.17)

Finally, using the identity (B.17) in (B.18), we obtain

$$E[(\bar{\tau} - \tau)^2]$$

$$= \frac{1}{\gamma^2 M} \int_0^{T_a} \int_0^{T_a} \tilde{\lambda}(t_1) \tilde{R}_p(t_1 - t_2) \tilde{\lambda}(t_2) \ dt_1 \ dt_2$$

$$= \frac{1}{\gamma^2 M} \int_0^{T_a} \int_0^{T_a} \tilde{\lambda}(t_1) \tilde{R}_p(t_1 - t_2) \tilde{\lambda}(t_2) \ dt_1 \ dt_2.$$  \hspace{1cm} (B.18)

**APPENDIX C**

**Computation of Covariance of Filtered Poisson Process**

Here the autocovariance of the superposition $S(t) = \sum_{n=1}^{N(t)} p(t - t_n)$ is derived.

Since, conditioned on $\tau$, $(t_n)_{n=1}^{N(t)}$ is Poisson, application of the identity (A.3) gives

$$E \left[ \sum_{i=1}^{n} p(t - t_i) \right] = E \left[ \sum_{i=1}^{n} p(t - t_i) \right]$$

$$E \left[ \int_0^T p(t - u) \tilde{\lambda}(u) \ du \right] = \int_0^T p(t - u) \tilde{\lambda}(u) \ du.$$  \hspace{1cm} (C.1)

where $\tilde{\lambda}(t) = E[\tilde{\lambda}(u - \tau)]$ and $n = N(t)$. Next consider the autocovariance function $K_s(z_1, z_2)$:

$$E \left[ \sum_{i=1}^{N(z_1)} p(z_1 - t_i) \sum_{j=1}^{N(z_2)} p(z_2 - t_j) \right]$$

$$= \int_0^{T_a} p(z_1 - u) \tilde{\lambda}(u) \ du \int_0^{T_a} p(z_2 - u) \tilde{\lambda}(u) \ du$$

$$= E \left[ \sum_{i=1}^{N(z_1)} p(z_1 - t_i) \sum_{j=1}^{N(z_2)} p(z_2 - t_j) \right]$$

$$- \int_0^{T_a} p(z_1 - u) \tilde{\lambda}(u) \ du \int_0^{T_a} p(z_2 - u) \tilde{\lambda}(u) \ du.$$  \hspace{1cm} (C.2)

The inner expectation of the first term on the right-hand side of (C.2) decomposes into a sum of two terms due to the independent increment property of $(t_i)_{i=1}^{N(t)}$ [27]:

$$E \left[ \sum_{i=1}^{N(z_1)} p(z_1 - t_i) \sum_{j=1}^{N(z_2)} p(z_2 - t_j) \right]$$

$$= E \left[ \sum_{i=1}^{N(z_1)} p(z_1 - t_i) p(z_2 - t_j) \right]$$

$$+ \int_0^{T_a} p(z_1 - u) \tilde{\lambda}(u) \ du$$

$$\cdot \int_0^{T_a} p(z_2 - u) \tilde{\lambda}(u) \ du$$

$$= \int_0^{T_a} \int_0^{T_a} p(z_1 - u) \tilde{\lambda}(u) \ du$$

$$\cdot \int_0^{T_a} p(z_2 - u) \tilde{\lambda}(u) \ du.$$  \hspace{1cm} (C.3)

Finally, taking the expectation of (C.3) with respect to $\tau$ gives the covariance function:

$$K_s(z_1, z_2) = \text{cov} \left( \sum_{i=1}^{N(z_1)} p(z_1 - t_i), \sum_{i=1}^{N(z_2)} p(z_2 - t_i) \right)$$

$$= \int_0^{T_a} \int_0^{T_a} p(z_1 - u) \tilde{\lambda}(u) \ du$$

$$\cdot \int_0^{T_a} p(z_2 - u) \tilde{\lambda}(u) \ du$$

$$+ \int_0^{T_a} \int_0^{T_a} p(z_1 - u) \tilde{\lambda}(u) \ du$$

$$\cdot \int_0^{T_a} p(z_2 - u) \tilde{\lambda}(u) \ du$$

$$= \int_0^{T_a} \int_0^{T_a} p(z_1 - u) \tilde{\lambda}(u) \ du$$

$$\cdot \int_0^{T_a} p(z_2 - u) \tilde{\lambda}(u) \ du.$$  \hspace{1cm} (C.4)

where

$$\text{cov} \left( \lambda(u_1 - \tau), \lambda(u_2 - \tau) \right)$$

$$= E \left[ \lambda(u_1 - \tau) \lambda(u_2 - \tau) \right]$$

$$- E \left[ \lambda(u_1 - \tau) \right] E \left[ \lambda(u_2 - \tau) \right].$$

Next it is shown that the covariance $K_s(z_1, z_2)$ of (C.4) depends only upon the time difference $z_1 - z_2$, under the following conditions: 1) $T \to \infty$; 2) uniformly distributed $\tau$ over $[0, T]$; 3) causal impulse response $p(t)$. Under these conditions we have the following results:

$$\tilde{\lambda}(t) = E[\tilde{\lambda}(t - \tau)]$$

$$- \frac{1}{T} \int_0^T \tilde{\lambda}(t - \tau) \ d\tau$$

$$= \frac{1}{T} \int_0^T \tilde{\lambda}(t - \tau) + \lambda_0 \ d\tau$$

$$= \frac{1}{T} \int_0^T \tilde{\lambda}(t - \tau) + \lambda_0 \ d\tau$$

$$= \frac{1}{T} \int_0^T \tilde{\lambda}(t) \ d\tau$$

$$= \frac{\Lambda}{T}, \quad 0 \leq t \leq T.$$  \hspace{1cm} (C.5)
and
\[
\text{cov} \left[ \lambda(u_1 - \tau), \lambda(u_2 - \tau) \right] = E \left[ (\lambda(u_1 - \tau) - \bar{\lambda}(u_1)) (\lambda(u_2 - \tau) - \bar{\lambda}(u_2)) \right] = \frac{1}{T} \int_0^T \left[ \lambda(u_1 - \tau) - \bar{\lambda}(u_1) \right] \left[ \lambda(u_2 - \tau) - \bar{\lambda}(u_2) \right] \, du.
\]

Therefore, from the causality assumption,
\[
K_s(z_1, z_2) = \text{cov} \left( \sum_{i=1}^{N_s(z_1)} p(z_1 - t_i), \sum_{j=1}^{N_s(z_2)} p(z_2 - t_j) \right)
= \int_0^T \rho(z_1 - u) p(z_2 - u) \lambda(u) \, du
+ \int_0^T d\lambda(u) \int_0^T d\lambda(u) p(z_1 - u_1) p(z_2 - u_2)
\cdot \text{cov} \left[ \lambda(u_1 - \tau), \lambda(u_2 - \tau) \right].
\]

Using (C.6) in (C.4) and taking a two-dimensional Fourier transform over the arguments \(z_1\) and \(z_2\) gives
\[
\mathcal{F} \left( K_s(z_1, z_2) \right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{F} \left( K_s(z_1, z_2) \right) e^{-j\omega_1 z_1 + j\omega_2 z_2}
= P(\omega) P(\nu) \left[ \Lambda(\omega) + \Lambda(\omega) \sin \left( [\omega + \nu] T / 2 \right) \right]
+ \Lambda(\omega) \Lambda(\omega) \sin \left( [\omega + \nu] T / 2 \right) - \Lambda^2(\omega) \sin \left( \omega T / 2 \right) \sin \left( \nu T / 2 \right),
\]
where \(\Lambda(\omega)\) and \(P(\omega)\) are the Fourier transforms of \(\Lambda(t)\) and \(p(t)\), and \(\sin(x)=\sin(x)/x\). As \(T \to \infty\), \(\sin([\omega + \nu] T / 2) = 0\), unless \(\omega = -\nu\), and \(\sin(\omega T / 2) \sin(\nu T / 2) = \sin(\omega T / 2)\) if \(\omega = \nu\) and zero otherwise. Therefore, in the limit, over its nonzero region of definition (the diagonal \(\omega = -\nu\)), the two-dimensional Fourier transform (C.8) reduces to the one-dimensional Fourier transform in the frequency variable \(\omega\). Hence, to an \(O(1/T)\) approximation, the covariance function \(K_s(z_1, z_2)\) has a one-dimensional Fourier transform over the difference \(z_1 - z_2\):
\[
G_s(\omega) = \left| P(\omega) \right|^2 \left[ \Lambda(\omega) + \Lambda(\omega) \sin(\omega T / 2) \right]^2 + \Lambda(\omega) \Lambda(\omega) \sin(\omega T / 2) \sin(\nu T / 2)
\]
and hence \(K_s\) is a function of \(z_1 - z_2\). The final form for the Fourier transform of \(K_s\) obtained from (C.10) by definition of the energy normalized Fourier transform of \(p(t)\), denoted \(\hat{P}(\omega)\), and recognition of the D.C. value \(\Lambda(0)\) as the energy \(A\) of the point process. The result is
\[
G_s(\omega) = \left| \hat{P}(\omega) \right|^2 \left[ \Lambda + \Lambda(\omega) - \Lambda \sin(\omega T / 2) \right]^2.
\]

**References**


