Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005



# System Identification with Analog and Counting Process Observations I: Hybrid Stochastic Intensity and Likelihood.

Victor Solo

*Abstract*— **In a number of application areas there is growing interest in system identification for systems whose observation processes consist of both analog and counting process signals. But so far few system identification techniques exist for these cases and likelihood functions have so far not been available. Here we introduce a new hybrid stochastic intensity and use it to construct, for the first time, an analog-counting process likelihood.**

## I. Introduction

Signal estimation and system identification for systems observed through point processes (or counting processes) have become extremely important over the years. Initial interest developed in the late 1960s and early 1970s e.g. [1],[2]. More recently there has been renewed interest from the communications networks area and especially from neuroscience [3],[4] where hundreds of cortical spike trains can now be simultaneously recorded in awake animals. Further in neuroscience now, attention is being drawn to cases where the observation process is *hybrid* also involving analog signals such as EEG recordings and local field potentials [5],[6].

For system identification to proceed in a principled way that includes all the information in the signals it is necessary to be able to construct likelihood functions. Methods based on second or third moments do not include all the information in non-Gaussian settings. But so far this likelihood construction has not been fully achieved. For a scalar point process with stochastic intensity depending on the past, the likelihood was first constructed by [7] and then in [8] extended, via a martingale framework, to general marked point processes thus including multivariate point processes as a special case. The development it must be noted is very abstract and few examples are given. A more accessible presentation of the marked point process results is in [9]. In a state space setting, when a *scalar* point process serves as an observation process for an underlying unobserved state space process so that its stochastic intensity depends on the state then [7], [10] constructed the likelihood function (actually Rubin does not require that the underlying state obey a Markov process law). The multivariate extension was obtained in [1]. But for the situation of interest here, where some observation processes may be analog and others point process and there may or may not be an underlying unobserved state, the problems of likelihood construction are open. It is the aim of this paper to sketch solutions to this class of problems.

V Solo is with Dept of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. vsolo@umich.edu

Turning to the literature on joint description of analog and point process signals there is little to report. There are firstly two papers of Willie [11],[12]. The first defines a cross- covariance function between a scalar analog process and a scalar point process and discusses estimation of it and asymptotic behaviour. The second discusses "linear" modelling in which the past of the analog process and the stochastic intensity may be allowed to depend on the past of both point process and analog signal. But no likelihood functions are constructed and as we shall see could not be constructed from the second-order quantities introduced by Willie. The other work is by [5]. This work develops cross-spectral estimation between point process and analog signals. It deals essentially with the Fourier transform of the statistics proposed by Willie (although his work is not referenced) and draws on the framework of [2]. Again likelihood based methods are not developed, although third order statistics are discussed.

We make no attempt here to provide a rigorous derivation of our results. That would require a lot more space and will be pursued elsewhere. Rather we discretise time to tiny subintervals and approximate the point process on these subintervals by a conditional Bernoulli process and thus proceed to derive results informally. This has two advantages. Firstly it leads to extremely simple derivations that are new and informative even for the known results already cited. Secondly this kind of development makes the results accessible to a mathematically less sophisticated community. The conditional Bernoulli heuristic is well known and has been mentioned briefly in connexion with deriving Rubin's likelihood [13],p72 and also used as a computational procedure [14]. But here we push the method far into unchartered territory, using it to derive new results and give new insights.

The remainder of the paper is organized a follows. In the first few sections we rederive the known results mentioned above. This considerably speeds up the development since subsequent argument will build on these results. Also it allows calibration against the known results. In section II we derive the scalar Rubin likelihood and extend this in section III to the multivariate Jacod likelihood. In section IV we then derive the Snyder-Rubin scalar partially observed state space likelihood. Then in section V we begin with the bivariate case of joint observation of a scalar analog signal and scalar point process . We introduce the new hybrid stochastic intensity which is the necessary object for likelihood construction and proceed to construct the likelihood. Extensions to the multivariate case easily follow. In section VI we extend this to the partially observed state space setting. Section VII contains conclusions. We conclude this section by describing some notation and basic definitions.

*Notation and Defintions*. In the sequel $\delta$ denotes a tiny time interval; $t$ denotes a continuous time and $k$ a discrete time so that $t = k\delta$. $N_{(t)}$=# events up to time $t$ and in discrete time $N_k = N_{(k\delta)}$. Next,$\delta N_{(t)}$ = incremental count = # events in $(t, t+\delta]$. Also $\delta N_k = \delta N_{(k\delta)}$. Continuing, $N_0^k$ = history of the counting process up to time $k = (\delta N_0 = \delta n_0, \cdots, \delta N_k = \delta n_k)$.And $N^0 = (\delta N_0 = \delta n_0)$. We write $a(\delta) = o(\delta)$ to mean $a(\delta)/\delta \rightarrow 0$ as $\delta \rightarrow 0$. Finally if $[0, T]$ is an observation interval we write $T = n\delta$.

## II. **Univariate Likelihood**

We suppose the point process obeys the following assumptions:

**NS** No Simultaneity.

$$P(\delta N_k > 1 | N_0^{k-1}) = o(\delta)$$

This means that in a small time interval $\delta$ only 1 or 0 events occur. This property is called orderliness in the point process literature [9].

**SI** Stochastic Intensity.

$$\begin{aligned} P(\delta N_k = 1 | N_0^{k-1}) &= \lambda_{(k\delta)}\delta + o(\delta) \\ &= \lambda_k \delta + o(\delta) \end{aligned}$$

Here $\lambda_{(t)}$ is called the stochastic (conditional) intensity and is a non-negative functional of the past history. A more formal definition of the stochastic intensity can be found in [13],[9].

In view of assumptions **NS** and **SI** we have:

**CBD** Conditional Bernoulli Description.

$$\begin{aligned} P(\delta N_k = 0 | N_0^{k-1}) &= 1 - \lambda_{(k\delta)}\delta + o(\delta) \\ &= 1 - \lambda_k \delta + o(\delta) \end{aligned}$$

With this setup we can now develop the Rubin likelihood.

Consider a fixed observation interval $0 \leq t \leq T = n\delta$. The likelihood is just the joint density of the incremental counting process random variables,

$$\begin{aligned} L_n &= P(N_0^n) \\ &= P(\delta N_0 = \delta n_0, \delta N_1 = \delta n_1, \cdots, \delta N_n = \delta n_n) \\ &= P(N_0^{n-1}, \delta N_n = \delta n_n) \\ &= P(\delta N_n = \delta n_n | N_0^{n-1}) P(N_0^{n-1}) \\ &= P(\delta N_n = \delta n_n | N_0^{n-1}) L_{n-1} \end{aligned}$$

Iterating this gives

$$L_n = \Pi_0^n P(\delta N_k = \delta n_k | N_0^{k-1})$$

Since $\delta N_k | N_0^{k-1}$ is Bernoulli (i.e. $\delta n_k = 0$ or 1) this gives

$$L_n = \Pi_0^n (\lambda_k \delta + o(\delta))^{\delta n_k} (1 - \lambda_k \delta + o(\delta))^{1 - \delta n_k}$$

To derive the likelihood formula we would like to let $\delta \rightarrow 0$. But the term $\Pi_0^n (\delta)^{\delta n_k}$ will cause a singularity problem. To resolve this one must work with a likelihood ratio. We can use any convenient reference model but the most natural is the unit rate Poisson. The resulting likelihood ratio is (dropping the $o(\delta)$ terms for simplicity)

$$\begin{aligned} LR_n &= \Pi_0^n \frac{(\lambda_k \delta)^{\delta n_k}}{\delta^{\delta n_k}} \frac{(1 - \lambda_k \delta)^{1 - \delta n_k}}{(1 - \delta)^{\delta n_k}} \\ &= \Pi_0^n \lambda_k^{\delta n_k} (1 - (\lambda_k - 1))^{1 - \delta n_k} \end{aligned}$$

We see that the singular term has been removed. Continuing we approximate further (dropping always only terms that are $o(\delta)$) to get

$$LR_n = \Pi_0^n e^{\delta n_k log \lambda_k - (\lambda_k - 1)\delta(1 - \delta n_k)}$$

In the exponent the first term is a Riemann-Stieltjes sum $\Sigma_0^n (N_{(k\delta+\delta)} - N_{(k\delta)}) log \lambda_{(k\delta)}$ which under some regularity conditions will converge in probability to the Riemann-Stieltjes integral $\int_0^T log \lambda_{(t)} dN_{(t)}$. Note that $N_{(t)}$ is of finite variation so there is no technical problem here: see e.g.[15]. Also $\lambda_{(t)}$ only depends on history up to $t-$. The second term consists of two parts. The first $\Sigma_0^n (\lambda_k - 1)\delta$ which will converge to $\int_0^T (\lambda_{(t)} - 1) dt$; the second $-\delta \Sigma_0^n (\lambda_k - 1) \delta n_k$ will be of order $\delta \int_0^T (\lambda_{(t)} - 1) dN_{(t)}$ and so vanish. We thus obtain the classical Rubin log-likelihood ratio with respect to a unit rate Poisson as,

Result I: Univariate likelihood ratio

$$lnLR_T = \int_0^T ln \lambda_{(t)} dN_{(t)} - \int_0^T (\lambda_{(t)} - 1) dt$$

Note that the derivation has made clear the appearance of a singularity and the consequent necessity for a likelihood ratio rather than a likelihood.

## III. **Multivariate Likelihood**

To keep the discussion brief we develop the bivariate case. But in fact it already exhibits all the essential issues so that the full multivariate result will follow easily. We consider then two counting processes $N_{(t)}, M_{(t)}$ with corresponding discrete time counts $N_k = N_{(k\delta)}, M_k = M_{(k\delta)}$ and so on as before. We also denote the joint history as $\mathcal{H}_0^k = (N_0^k, M_0^k)$. We now introduce the following assumptions.

**NS** No-simultaneity.

Given any past trajectory only 0 or 1 events (of either type) can occur in the next small time interval.

$$P(\delta N_k + dMk > 1 | \mathcal{H}_0^k) = o(\delta)$$

This of course implies marginal no-simultaneity.

**SI** Joint Stochastic Intensities.

$$\begin{aligned} P(\delta N_k = 1 | \mathcal{H}_0^{k-1}) &= \lambda_{(k\delta)}^{NJ} \delta + o(\delta) \\ &= \lambda_k^{NJ} \delta + o(\delta) \\ P(\delta M_k = 1 | \mathcal{H}_0^{k-1}) &= \lambda_{(k\delta)}^{MJ} \delta + o(\delta) \\ &= \lambda_k^{MJ} \delta + o(\delta) \end{aligned}$$

It is important to keep note of the fact that these two stochastic intensities depend on the joint history and so will differ from the marginal stochastic intensity previously introduced.

As before assumptions **NS** ,**SI** yield:
**CBD** Conditional Multi-Bernoulli Description.
Firstly we have the semi-marginal relations,

$$\begin{aligned} P(\delta N_k = 0|\mathcal{H}_0^{k-1}) &= 1 - \lambda_k^{NJ}\delta + o(\delta) \\ P(\delta M_k = 0|\mathcal{H}_0^{k-1}) &= 1 - \lambda_k^{MJ}\delta + o(\delta) \end{aligned}$$

But we need to consider bivariate conditional probabilities and this is most usefully pursued in the context of constructing the likelihood. Proceeding much as before we find

$$\begin{aligned} L_n &= P(\mathcal{H}_0^n) \\ &= \Pi_0^n P(\delta N_k = \delta n_k, \delta M_k = \delta m_k|\mathcal{H}_0^{k-1}) \end{aligned}$$

Because of the **NS** condition we have only to calculate four probabilities,

$$\begin{aligned} &P(\delta N_k = 1, \delta M_k = 1|\mathcal{H}_0^{k-1}) \\ &P(\delta N_k = 0, \delta M_k = 0|\mathcal{H}_0^{k-1}) \\ &P(\delta N_k = 1, \delta M_k = 0|\mathcal{H}_0^{k-1}) \\ &P(\delta N_k = 0, \delta M_k = 1|\mathcal{H}_0^{k-1}) \end{aligned}$$

Now the joint no-simultaneity ensures the first probability is $o(\delta)$. So we need only specify the other three. And now remarkably this can be done in terms of semi-marginal quantities. We have

$$\begin{aligned} & P(\delta N_k = 1|\mathcal{H}_0^{k-1}) = P(\delta N_k = 1, \delta M_k = 0|\mathcal{H}_0^{k-1}) \\ + \quad & P(\delta N_k = 1, \delta M_k = 1|\mathcal{H}_0^{k-1}) \\ \Rightarrow \quad & P(\delta N_k = 1, \delta M_k = 0|\mathcal{H}_0^{k-1}) = \lambda_k^{NJ}\delta + o(\delta) \end{aligned}$$

Similarly

$$P(\delta N_k = 0, \delta M_k = 1|\mathcal{H}_0^{k-1}) = \lambda_k^{MJ}\delta + o(\delta)$$

Finally by subtraction we can conclude

$$P(\delta N_k = 0, \delta M_k = 0|\mathcal{H}_0^{k-1}) = 1 - \lambda_k^{NJ}\delta - \lambda_k^{MJ}\delta + o(\delta)$$

But now we are are able to conclude the following remarkable result:
Result II :**CI** Conditional Independence.
$\delta M_k, \delta N_k$ are conditionally independent given the history $\mathcal{H}_0^{k-1}$ i.e.

$$\begin{aligned} & P(\delta N_k = \delta n_k, \delta M_k = \delta m_k|\mathcal{H}_0^{k-1}) \\ = \quad & (P(\delta N_k = \delta n_k|\mathcal{H}_0^{k-1}) + o(\delta)) \\ \times \quad & (P(\delta M_k = \delta m_k|\mathcal{H}_0^{k-1}) + o(\delta)) \end{aligned}$$

*Proof.*
We just have to treat the four cases. Firstly joint no-simultaneity ensures $P(\delta N_k = 1, \delta M_k = 1|\mathcal{H}_0^{k-1}) = o(\delta)$. While the product on the right side is

$$\begin{aligned} & P(\delta N_k = 1|\mathcal{H}_0^{k-1})P(\delta M_k = 1|\mathcal{H}_0^{k-1}) \\ = \quad & (\lambda_k^{NJ}\delta + o(\delta))(\lambda_k^{MJ}\delta + o(\delta)) = o(\delta) \end{aligned}$$

as required. Secondly

$$P(\delta N_k = 1, \delta M_k = 0|\mathcal{H}_0^{k-1}) = \lambda_k^{NJ}\delta + o(\delta)$$

While

$$\begin{aligned} & P(\delta N_k = 1|\mathcal{H}_0^{k-1})P(\delta M_k = 0|\mathcal{H}_0^{k-1}) \\ = \quad & (\lambda_k^{NJ}\delta + o(\delta))(1 - \lambda_k^{MJ}\delta + o(\delta)) \\ = \quad & \lambda_k^{NJ}\delta + o(\delta) \end{aligned}$$

and the result follows. The result follows similarly for the third case $\delta n_k, \delta m_k = 0, 1$. Finally

$$\begin{aligned} & P(\delta N_k = 0, \delta M_k = 0|\mathcal{H}_0^{k-1}) \\ = \quad & 1 - \lambda_k^{NJ}\delta - \lambda_k^{MJ}\delta + o(\delta) \end{aligned}$$

While

$$\begin{aligned} & P(\delta N_k = 0|\mathcal{H}_0^{k-1})P(\delta M_k = 0|\mathcal{H}_0^{k-1}) \\ = \quad & (1 - \lambda_k^{NJ}\delta + o(\delta))(1 - \lambda_k^{MJ}\delta + o(\delta)) \\ = \quad & 1 - (\lambda_k^{NJ}\delta + \lambda_k^{NJ}\delta) + o(\delta) \end{aligned}$$

and the result follows again.

With conditional independence established we can return to the likelihood to find,

$$\begin{aligned} L_n &= \Pi_0^n P(\delta N_k = \delta n_k|\mathcal{H}_0^{k-1}) \\ &\times \quad \Pi_0^n P(\delta M_k = \delta m_k|\mathcal{H}_0^{k-1}) \end{aligned}$$

which is a product of 'univariate' likelihoods. Normalizing with a product of independent unit rate Poissons and taking limits as before we find:
Result III : Bivariate Likelihood Ratio,

$$\begin{aligned} lnLR_T &= \int_0^T ln\lambda_{(t)}^{NJ}\delta N_{(t)} - \int_0^T (\lambda_{(t)}^{NJ} - 1)dt \\ &+ \quad \int_0^T ln\lambda_{(t)}^{MJ}\delta M_{(t)} - \int_0^T (\lambda_{(t)}^{MJ} - 1)dt \end{aligned}$$

which can be also deduced from Jacod's results.

We now see something not evident from previous derivations. The simple *additive structure* is a consequence of *conditional independence* which itself is induced by the joint *no-simultaneity* condition.

It is important not to be misled by the additivity/ conditional independence. Each 'univariate' likelihood does depend on the joint history through the stochastic intensities. So one is most emphatically not just adding up the marginal log-likelihood ratios.

It is also immediately apparent that under the joint no-simultaneity condition the results extend to the general multivariate case. For $m$ point processes there are $2^m$ required conditional probabilities to form the iterated likelihood (since there are $2^m$ strings of 0s and 1s). But the joint no-simultaneity ensures all but $m+1$ of these strings has probability $o(\delta)$. And these $m+1$ joint conditional probabilities can be determined from $m$ semi-marginal probabilities plus the fact that they must sum to 1. This will induce conditional independence again and so lead to the additive log likelihood ratio.

## IV. **State Space Likelihood**

Here we suppose the stochastic intensity depends on an underlying unobserved state (Snyder) as well as the past of the counting process (Rubin). We take the state to be an analog stochastic process $x_{(t)}$ for simplicity but the point process case can easily be treated. The sampled signal is $x_k = x_{(k\delta)}$. Since $\delta N_k$ looks ahead we match the history $X_1^k = (X_1 = x_1, \cdots, X_k = x_k)$ with $N_0^{k-1}$.

*Additional Notation and Definitions*. We use the notation $X_k \sim x$ to mean $x \leq X_k \leq x + h$ with $0 < h << 1$. And then $\tilde{X}_1^k = (X_1 \sim x_1, \cdots, X_k \sim x_k)$. Finally by $P(A|X_1^k)$ we mean

$$\lim_{h \to 0} P(A|\tilde{X}_1^k) = \lim_{h \to 0} \frac{P(A, \tilde{X}_1^k)\frac{1}{h^{k+1}}}{P(\tilde{X}_1^k)\frac{1}{h^{k+1}}}$$

The assumptions now become:

**NS** No simultaneity.

$$P(\delta N_k > 1|N_0^{k-1}, X_1^k) = o(\delta)$$

**SDSI** State Dependent Stochastic Intensity.

$$\begin{aligned} P(\delta N_k = 1|N_0^{k-1}, X_1^k) &= P(\delta N_k = 1|N_0^{k-1}, X_k = x_k) \\ &= \lambda_{(k\delta, x_{(k\delta)})}\delta + o(\delta) \\ &= \lambda_{k,x_k}\delta + o(\delta) \end{aligned}$$

This leads as usual to the:

**CBD** Conditional Binomial Description.

$$\begin{aligned} P(\delta N_k = 0|N_0^{k-1}, X_1^k) &= P(\delta N_k = 0|N_0^{k-1}, X_k = x_k) \\ &= 1 - \lambda_{(k\delta, x_{(k\delta)})}\delta + o(\delta) \\ &= 1 - \lambda_{k,x_k}\delta + o(\delta) \end{aligned}$$

As before the likelihood ratio (with respect to a unit rate Poisson) will be

$$LR_n = \Pi_0^n \frac{P(\delta N_k = \delta n_k|N_0^{k-1})}{\delta^{\delta n_k}(1-\delta)^{1-\delta n_k}}$$

To evaluate this we write

$$\begin{aligned} &\frac{P(\delta N_k = \delta n_k|N_0^{k-1})}{\delta^{\delta n_k}(1-\delta)^{1-\delta n_k}} \\ = &\int \frac{P(\delta N_k = \delta n_k|N_0^{k-1}, X_k = x_k)}{\delta^{\delta n_k}(1-\delta)^{1-\delta n_k}} p(x_k|N_0^{k-1})dx_k \end{aligned}$$

And appealing to the **CBD** property we get (on dropping $o(\delta)$ terms)

$$\int \left(\frac{\lambda_{k,x_k}\delta}{\delta}\right)^{\delta n_k} \left(\frac{1-\lambda_{k,x_k}\delta}{1-\delta}\right)^{1-\delta n_k} p(x_k|N_0^{k-1})dx_k$$

We evaluate this as follows.

When $\delta n_k = 1$ we get

$$\int \lambda_{(k\delta, x_{(k\delta)})} p(x_k|N_0^{k-1})dx_k = \hat{\lambda}_k = \hat{\lambda}_{(k\delta)}$$

which defines $\hat{\lambda}_{(t)}$. When $\delta n_k = 0$ we get

$$\int \left(\frac{1-\lambda_{k,x_k}\delta}{1-\delta}\right) p(x_k|N_0^{k-1})dx_k = \frac{1-\hat{\lambda}_k\delta}{1-\delta}$$

Putting these together gives

$$\frac{P(\delta N_k = \delta n_k|N_0^{k-1})}{\delta^{\delta n_k}(1-\delta)^{1-\delta n_k}} = \hat{\lambda}_k^{\delta n_k}\left(\frac{1-\hat{\lambda}_k\delta}{1-\delta}\right)^{1-\delta n_k}$$

Now iterating and taking limits as before we get:

Result IV: Point Process-State Space log-likelihood ratio,

$$lnLR_T = \int_0^T ln\hat{\lambda}_{(t)}dN_{(t)} - \int_0^T (\hat{\lambda}_{(t)} - 1)dt$$

which is the Rubin/Snyder formula.

This new argument shows clearly the origin of $\hat{\lambda}_{(t)}$; how it is that the formula has the same structure as previously and why the stochastic intensity can be allowed to depend on the past as well as on the latest value of the state. Generation of $\hat{\lambda}_{(t)}$ requires a conditional density which could be generated by modern particle filtering methods. Details will be pursued elsewhere.

## V. **Analog and Point Process: Hybrid Likelihood**

Now finally we are ready to treat the hybrid case. We begin for simplicity with the bivariate case of a jointly observed scalar analog signal $y_{(t)}$ and a point process $N_{(t)}$. We extend previous notation in the natural way to cover $y_{(t)}$. In particular we introduce the joint history $\mathcal{H}_{N,Y}^k = (N_0^k, Y_1^k)$. It is not immediately clear how to define a stochastic intensity to cover this case and the utility of our definition will become clear below. We assume:

**NS** No simultaneity

$$P(\delta N_k > 1|\mathcal{H}_{N,Y}^{k-1}, Y_k = y) = o(\delta)$$

**HSI** Hybrid Stochastic Intensity

$$\begin{aligned} &P(\delta N_k = 1|\mathcal{H}_{N,Y}^{k-1}, Y_k = y) \\ = &\lambda_{(k\delta, y)}\delta + o(\delta) \\ = &\lambda_{k,y}\delta + o(\delta) \end{aligned}$$

As usual **NS,HSI** deliver:

**CBD** Conditional Bernoulli Description.

$$P(\delta N_k = 0|\mathcal{H}_{N,Y}^{k-1}, Y_k = y) = 1 - \lambda_{k,y}\delta + o(\delta)$$

There are two associated quantities of importance.

Conditional Density

$$q(k\delta, y) = \lim_{h \to 0} \frac{1}{h} P(Y_k \sim y|\mathcal{H}_{N,Y}^{k-1})$$

Induced Stochastic Intensity

$$\begin{aligned} &P(\delta N_k = 1|\mathcal{H}_{N,Y}^{k-1}) \\ = &\int P(\delta N_k = 1|\mathcal{H}_{N,Y}^{k-1}, Y_k = y)q(k\delta, y)dy \\ = &\lambda_{(k\delta)}\delta + o(\delta) \\ \lambda_{(t)} = &\int \lambda_{(t,y)} q_{(t,y)}dy \end{aligned}$$

We are now ready to develop the new hybrid likelihood expression. We have

$$\begin{aligned} L_n &= \lim_{h\to 0} P(N_0^n, \tilde{Y}_0^n)/h^{n+1} \\ &= \lim_{h\to 0} \frac{L_{n-1}}{h} P(\delta N_n = \delta n_n, Y_n \sim y_n | N_0^{n-1}, \tilde{Y}_0^{n-1}) \\ &= \lim_{h\to 0} P(\delta N_n = \delta n_n | Y_n \sim y_n, \mathcal{H}_{N,Y}^{n-1}) \\ &\times \frac{1}{h} P(Y_n \sim y_n | \mathcal{H}_{N,Y}^{n-1}) L_{n-1} \\ &= P(\delta N_n = \delta n_n | Y_n \sim y_n, \mathcal{H}_{N,Y}^{n-1}) q_{(n\delta, y_{(n\delta)})} L_{n-1} \end{aligned}$$

We see now how our definition of the HSI fits with this factorization of the likelihood. Iterating yields a product of an analog and a point process (discrete valued) component.

$$\begin{aligned} L_n &= L_n^a L_n^d \\ L_n^a &= \Pi_0^n q_{(n\delta, y_{(n\delta)})} \\ L_n^d &= \Pi_0^n P(\delta N_k = \delta n_k | Y_k = y_k, \mathcal{H}_{N,Y}^{k-1}) \end{aligned}$$

The digital likelihood can be treated exactly as before and after normalization by the usual unit rate Poisson we are led to the digital component of the new log-likelihood formula we have sought,

Result Va: Digital Component of Hybrid Likelihood Ratio

$$lnLR_T^d = \int_0^T ln\lambda_{(t,y_{(t)})} dN_{(t)} - \int_0^T (\lambda_{(t,y_{(t)})} - 1) dt$$

Turning to the analog component we are more or less on familiar ground since there is much literature on these types of likelihood. Except that is for the fact that we need to allow dependence on the past of the point process. Rather than attempt a general specification we give a simple, but useful example to indicate the kind of result possible. We assume conditional distributions are Gaussian so we need only specify conditional first and second moments, thus,

$$\begin{aligned} E(y_{k+1} - y_k | Y_k = y, \mathcal{H}_{N,Y}^{k-1}) &= -\mu_{(k\delta, y_{(k\delta)})}\delta + o(\delta) \\ &= -\mu_{k,y_k}\delta + o(\delta) \\ var(y_{k+1} - y_k | Y_k = y, \mathcal{H}_{N,Y}^{k-1}) &= \sigma^2\delta + o(\delta) \end{aligned}$$

So $\mu_{(t,y_{(t)})}$ is a functional of the joint past and is the analog analogue(!) of the HSI. The condtional density is then

$$q_{(k\delta, y_{(k\delta)})} = -\frac{1}{2}\frac{(y_{k+1} - y_k - \mu_{k,y_k}\delta)^2}{\sigma^2\delta} - \frac{1}{2} ln\sigma^2\delta$$

As a reference ($q^0_{(k\delta, y_{(k\delta)})}$) we take the same model but with $\mu_{k,y_k} = 0$. The associated likelihood ratio is $\Pi_0^n q_{(k\delta, y_{(k\delta)})}/q^0_{(k\delta, y_{(k\delta)})}$ and taking logs, cancelling out common terms and taking limits leads to

$$lnLR_T^a = \frac{1}{\sigma^2}\int_0^T \mu_{(t,y_{(t)})} dt - \frac{1}{2\sigma^2}\int_0^T \mu^2_{(t,y_{(t)})} dt$$

This type of analog likelihood ratio is well known. Putting these together we have:

Result Vb: Hybrid Likelihood Ratio,

$$lnLR_T = lnLR_T^d + lnLR_T^a$$

The multivariate version of these results will follow much as before due to the conditional independence property. Thus $lnLR_T^d$ will be given by an additive formula.

## VI. State Space Hybrid Likelihood with Analog and Point Process Observations

As usual we assemble an expanded set of definitions and assumptions.

**NS** No simultaneity

$$P(\delta N_k > 1 | \mathcal{H}_{N,Y}^{k-1}, X_1^k, Y_k = y) = o(\delta)$$

**SDHSI** State Dependent Hybrid Stochastic Intensity

$$\begin{aligned} & P(\delta N_k = 1 | \mathcal{H}_{N,Y}^{k-1}, X_1^k, Y_k = y) \\ =\ & P(\delta N_k = 1 | \mathcal{H}_{N,Y}^{k-1}, X_k = x_k, Y_k = y) \\ =\ & \lambda_{(k\delta, x_{(k\delta)}, y)}\delta + o(\delta) \\ =\ & \lambda_{k,x_k,y}\delta + o(\delta) \end{aligned}$$

As usual **NS,SDHSI** deliver:

**CBD** Conditional Bernoulli Description.

$$\begin{aligned} & P(\delta N_k = 0 | \mathcal{H}_{N,Y}^{k-1}, X_1^k, Y_k = y) \\ & P(\delta N_k = 0 | \mathcal{H}_{N,Y}^{k-1}, X_k = x_k, Y_k = y) \\ =\ & 1 - \lambda_{k,x_k,y}\delta + o(\delta) \end{aligned}$$

There are two associated quantities of importance.

Conditional Density

$$q(k\delta, y) = \lim_{h\to 0} \frac{1}{h} P(Y_k \sim y | \mathcal{H}_{N,Y}^{k-1})$$

**SDCD** State dependent conditional density

$$\begin{aligned} & \lim_{h\to 0} P(Y_k \sim y | \mathcal{H}_{N,Y}^{k-1}, X_k = x_k) \\ =\ & P(Y_k \sim y | X_k = x_k) = p(y|x_k) = q_{k\delta, x_{k\delta}, y} \end{aligned}$$

Now we continue by repeating the argument of the last section to get $L_n = L_n^a L_n^d$ with $L_n^a, L_n^d$ given as before. We proceed to evaluate $L_n^d$ first normalising by a unit rate Poisson. We have (along the lines of the state space section earlier)

$$\begin{aligned} & \frac{P(\delta N_k = \delta n_k | Y_k = y_k, \mathcal{H}_{N,Y}^{k-1})}{\delta^{\delta n_k}(1-\delta)^{1-\delta n_k}} \\ =\ & \int \frac{P(\delta N_k = \delta n_k | \mathcal{H}_{N,Y}^{k-1}, Y_k = y_k, X_k = x_k)}{\delta^{\delta n_k}(1-\delta)^{1-\delta n_k}} \\ \times\ & p(x_k | \mathcal{H}_{N,Y}^{k-1}, Y_k = y_k) dx_k \end{aligned}$$

And appealing to SDCD and the **CBD** property we get (on dropping $o(\delta)$ terms)

$$\begin{aligned} & \int \left(\frac{\lambda_{k,x_k,y_k}\delta}{\delta}\right)^{\delta n_k} \left(\frac{1 - \lambda_{k,x_k,y_k}\delta}{1-\delta}\right)^{1-\delta n_k} \\ \times\ & \frac{p(y_k | X_k = x_k, \mathcal{H}_{N,Y}^{k-1}) p(x_k | \mathcal{H}_{N,Y}^{k-1})}{p(y_k | \mathcal{H}_{N,Y}^{k-1})} dx_k \\ =\ & \int (\lambda_{k,x_k,y_k})^{\delta n_k} \left(\frac{1 - \lambda_{k,x_k,y_k}\delta}{1-\delta}\right)^{1-\delta n_k} \\ \times\ & \frac{p(y_k|x_k) p(x_k | \mathcal{H}_{N,Y}^{k-1})}{q_{(k\delta, y_{(k\delta)})}} dx_k \end{aligned}$$

We evaluate this, much as before, as follows.
When $\delta n_k = 1$ we get

$$\int \lambda_{k,x_k,y_k} p(y_k|x_k)p(x_k|\mathcal{H}_{N,Y}^{k-1})\frac{dx_k}{q_{(k\delta,y_{(k\delta)})}} = \hat{\lambda}_{k,y_k}$$

which defines $\hat{\lambda}_{(t,y_{(t)})}$. When $\delta n_k = 0$ we get

$$\int \left(\frac{1-\lambda_{k,x_k,y_k}\delta}{1-\delta}\right) p(y_k|x_k)p(x_k|\mathcal{H}_{N,Y}^{k-1})\frac{dx_k}{q_{(k\delta,y_{(k\delta)})}}$$
$$= \frac{1-\hat{\lambda}_{k,y_k}\delta}{(1-\delta)}$$

where we have used

$$q_{(k\delta,y_{(k\delta)})} = \int p(y_k|x_k)p(x_k|\mathcal{H}_{N,Y}^{k-1})dx_k$$

Putting these together gives

$$\frac{P(\delta N_k = \delta n_k|Y_k = y_k, \mathcal{H}_{N,Y}^{k-1})}{\delta^{\delta n_k}(1-\delta)^{1-\delta n_k}}$$
$$= \hat{\lambda}_{k,y_k}^{\delta n_k}\left(\frac{1-\hat{\lambda}_{k,y_k}\delta}{1-\delta}\right)^{1-\delta n_k}$$

Now iterating we find

$$L_n^d = \Pi_0^n \hat{\lambda}_{k,y_k}^{\delta n_k}\left(\frac{1-\hat{\lambda}_{k,y_k}\delta}{1-\delta}\right)^{1-\delta n_k}$$

and taking limits as usual we get:
Result VI: Point Process-State Space log-likelihood ratio,

$$lnLR_T^d = \int_0^T ln\hat{\lambda}_{(t,y_{(t)})}dN_{(t)} - \int_0^T (\hat{\lambda}_{(t,y_{(t)})} - 1)dt$$

We can obtain an expression for $lnLR_T^a$ but it will be model dependent and lack of space precludes details. Again implementation of these expressions will require particle filtering.

## VII. Conclusions

In this paper we have sketched a derivation of a number of old and new point process likelihood ratio formulae by a very simple argument. Though our derivations lack rigour the rigorous derivation of the old results requires considerable martingale machinery whereas we have done it in a few lines in a self-contained way. Further our derivation throws new light on why the formulae have the structure they do. Thus additivity in multivariate point process formulae is a direct consequence of a conditional independence property induced by the no-simultaneity assumption. Also the origin of the simple structure of state space likelihood ratios emerges clearly in the derivation. Most important however are our new results for likelihood ratios relating to systems observed with both analog and counting process observations (results Va,VI).

## References


[1] D.L. Snyder, *Random Point Processes*, J. Wiley, New York, 1975.
[2] D R Brillinger, "The spectral analysis of stationary interval functions", in *Proc 6th Berkeley Symposium on Mathematical Statistics and Probaility*, L LeCam, J Neyman, and E L Scott, Eds., 1972, pp. 483–513.
[3] F. Rieke, D. Warland, R. de Ruyter van Stvenink, and W. Bialek, *Spikes: Exploring the neural code*, MIT Press, Boston, 1997.
[4] Dayan P and Abbott LF, *Theoretical Neuroscience*, MIT Press, Cambridge MA, 2001.
[5] D M Halliday, J R Rosenberg, A M Amjad, P Breeze, B A Conway, and S F Farmer, "A framework for the analysis of mixed time series/point process data - Theory and application to the study of physiological tremor ,single motor unit discharges and electromyograms", *Prog Biophys Molec Biol*, vol. 64, pp. 237–278, 1995.
[6] Pesaran B, Pezaris J S, Sahani M, Mitra P P, and Andersen R A, "Temporal structure in neuronal activity during working memory in macaque parietal cortex", *Nature Neuroscience*, vol. 5, pp. 805–811, 2002.
[7] I Rubin, "Regular point processes and their detection", *IEEE Trans. Inf. Thy.*, vol. 18, pp. 547–557, 1972.
[8] J Jacod, "Multivariate point processes: Predictiable projection, Radon Nikodym derivatives, representation of martingales", *Z Wahrsch verw Geb*, vol. 31, pp. 235–253, 1975.
[9] D J Daley and D Vere-Jones, *An introduction to the Theory of Point Processes, Volume I (2nd. ed.)*, Springer-Verlag, New York, 2003.
[10] D L Snyder, "Filtering and detection for doubly stochastic Poisson processes", *IEEE Trans. Inf. Thy.*, vol. 18, pp. 91–102, 1972.
[11] J S Willie, "Covariation of a time series and a point process", *Jl. Appl. Prob.*, vol. 19, pp. 609–618, 1982.
[12] J S Willie, "Measuring the association of a time series and a point process", *Jl. Appl. Prob.*, vol. 19, pp. 597–608, 1982.
[13] Karr A, *Point Processes and their Statistical Inference, second edition*, Marcel Dekker, New York, 1991.
[14] M Berman, "Approximate point process likelihood with GLIM", *Applied Statistics*, vol. 41, pp. 31–38, 1992.
[15] F Klebaner, *Introduction to Stochastic Calculus with Applications*, Imperial College Press, London UK, 1998.