2014 European Control Conference (ECC)
June 24-27, 2014. Strasbourg, France

Remarks on model-based estimation of nonhomogeneous Poisson processes and applications to biological systems

Maja Skataric and Eduardo Sontag

Abstract— **This paper makes some remarks concerning model-based estimation methods for the rate of a nonhomogeneous Poisson processes that describes events arising from modeling biological phenomena in which discrete events are measured. We describe an approach based on observers and Kalman filters as well as preliminary simulation results, and compare these to other methods (not model-based) in the literature.**

I. INTRODUCTION

This work is motivated by questions arising from modeling biological phenomena in which discrete measurements such as "spikes" or "tumbles" are measured, and the objective is to estimate the underlying rate $\lambda(t)$ of a nonhomogeneous Poisson processes (NHPP) that describes these events. In such problems, often a small number of realizations is available (for example, through single-cell measurements of a number of neurons or of a population of bacteria). Current microfluidic (for bacteria) or electrode (for neurons) technology allows one to assume that internal variables (voltages, protein concentrations) in these systems behave reasonably identically, so that the only source of randomness is in a jump process driven by these internal variables. In this paper, we outline an approach based on observers and Kalman filters, providing simulation results and comparing these to other methods (not model-based) in the literature. A follow-up paper will further develop theory.

We remark that superficially analogous problems have been studied in which a diffusion (stochastic differential equation) describes the evolution of internal parameters and a nonhomogeneous Poisson process is observed, see for example [15], [2]. However, our problem is substantially different, because no randomness is assumed in the internal variables, but repeated realizations are observed.

In the statistics and signal processing literature, this problem has been studied for certain special classes of signals. There are many ways to pose mathematically the estimation problem that we are interested in. For example, one could look for a maximum likelihood (ML) estimate of parameters, if a parametric model class is assumed for $\lambda(t)$. This amounts essentially to a maximum a posteriori (MAP) estimate with a uniform distribution on parameters as prior [5]. However, there is no natural probabilistic structure in our applications to suggest the most natural prior; even a conjugate prior to the given Poisson model would not be necessarily justified. On the other hand, often the ultimate objective is control, in which case estimates of current states (and specifically of the current $\lambda(t)$) are more important than estimates of the entire history. This suggests the use of observers, or more generally Kalman filters, as done routinely in control theory, to obtain estimates that asymptotically improve. This method will be discussed in the next Section.

Numerous papers, see for example [8], [1], [7] study models of systems whose rate varies over time, and propose nonparametric techniques for estimating the cumulative intensity function of NHPP, $\Lambda(t) = \int_0^t \lambda(s)ds$, on the time interval $[0, T]$. In [7] the authors propose a piecewise-linear estimator of the cumulative intensity function between the time values $t_{(1)}, t_{(2)}, \cdots, t_{(n)}$, which represent the order statistics of the superposition of the k realizations, as

$$\tilde{\Lambda}(t) = \frac{i}{k} + \frac{t - t_{(i)}}{k(t_{(i+1)} - t_{(i)})},$$

where $\lambda(t) > 0, \forall t \in (0, T]$, and $\lambda(t)$ is continuous for almost all $t \in (0, T]$. The cumulative intensity function $\Lambda(t)$ is estimated from k (overlapping or non-overlapping) realizations on $(0, T]$, with n_i, $i = 1, 2, \cdots, k$ being the number of observations in the $i-$th realization, and $n = \sum_{i=1}^{k} n_i$ being the total number of observations. For this estimator, a strong consistency result and an asymptotically valid $100(1-\alpha)\%$ confidence interval for $\Lambda(t), \forall t \in (0, T]$ were shown, as

$$\tilde{\Lambda}(t) \to \Lambda(t),$$

with probability one as $k \to \infty, \forall t \in (0, T]$, and

$$\hat{\Lambda}(t) - z_{\alpha/2}\sqrt{\frac{\hat{\Lambda}(t)}{k}} < \Lambda(t) < \hat{\Lambda}(t) + z_{\alpha/2}\sqrt{\frac{\hat{\Lambda}(t)}{k}},$$

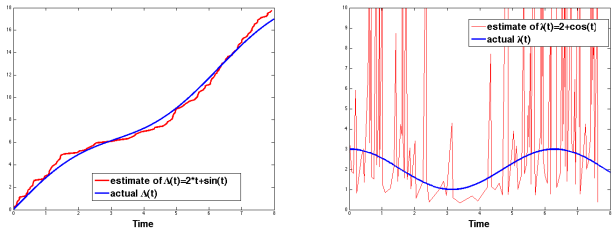
where $z_{\alpha/2}$ is the $1 - \alpha/2$ fractile of the standard normal distribution. However these methods fail to provide good estimates of the derivative of $\Lambda(t)$, shown on the right panel of Fig. 1. Our interest, recall, is on $\lambda(t)$, computed with this method as $\frac{1}{k(t_{(i+1)} - t_{(i)})}$.

Similar results can be obtained by finely sampling the time axis, and for every time point on the grid, finding the closest spike, and then computing the average wait time until a spike occurs across all realizations. Then the rate $\lambda(t)$ can be found for every time point on the grid by computing the inverse of the average wait time at that time point, shown on Fig. 2a.

Another possibility is to use what we will call "the naive method," in which one observes the average number of events r in a bin containing a time t, and estimates the rate $\lambda(t)$ as

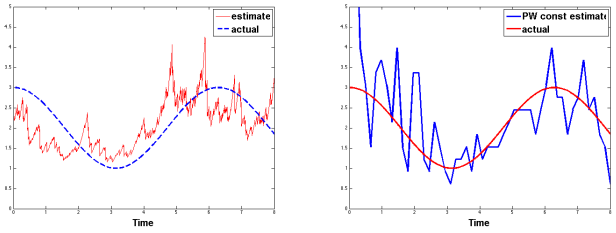
M. Skataric is with the Department of Electrical and Computer Engineering, Rutgers University, skataric@rutgers.edu

E. Sontag is with the Department of Mathematics, Rutgers University, sontag@math.rutgers.edu



(a) Cumulative Intensity function (b) Rate of a Poisson Process

Fig. 1: Piecewise linear estimator of the Cumulative Intensity function that uses k=8 realizations. mesh size= 10^{-4}, $n = 143$, i.e $n_i = \{16, 22, 23, 16, 19, 20, 15, 12\}$, $i = 1, \cdots, 8$



(a) (b)

Fig. 2: Modified piecewise-linear method, $k = 20$ realizations (a), and Naive (piecewise constant) method, $k = 20$, $M = 50$ (b)

$\frac{r}{bt}$, where b is a predetermined bin size. As an example of an application of this method, we obtain an estimate as shown in Fig. 2b. In general, finding an optimal bin size is a difficult problem (see [11]). Many other methods have been proposed. The method developed in [9] deals with the estimation of an arrival rate function $\lambda(t) = a + bt$, $t \in [0, T]$ by dividing the entire time interval into M equispaced subintervals (bins), and observing an average number of events in each. The authors propose several methods for finding the estimates $\hat{a}$ and $\hat{b}$, namely, ordinary least squares, iterative weighted least squares and maximum likelihood method, and elaborate on the effect of the time interval length T, and number of measurement subintervals M, but do not provide the optimal length of the subinterval. This method works well for slowly changing $\lambda(t)$, however it fails in cases of fast-varying signals and would not be a good starting choice, especially in cases where we have some prior information about the signal we are estimating.

II. Problem Formulation and an Observer Design

Let N_t be a counting process with rate $\lambda(t)$, i.e N_t is a Poisson random variable with rate $y(t) = \int_0^t \lambda(s)ds$. The probability of having j events until time t is

$$P_{y(t)}(N_t = j) = P(j \text{ events on } [0, t]) = \frac{y(t)^j e^{-y(t)}}{j!}.$$

Observe that, $E[N_t] = \text{var}[N_t] = y(t)$.

Fix an integer $k > 0$, which denotes the number of realizations of a counting process N_t and consider k IID N's. For a fixed t, each N_i^t, $i = 1 \ldots k$, is an IID random variable with rate $y(t)$ that gives the number of events until time t for a realization i. Let

$$N_k(t) := \frac{1}{k}(N_1^t + N_2^t + N_3^t + \cdots + N_k^t),$$

be the average number of events (across all realizations) until time t.

Then

$$E[N_k(t)] = y(t),$$

$$\text{var}[N_k(t)] = \frac{1}{k^2}\text{var}(N_1^t + N_2^t + N_3^t + \cdots + N_k^t) = \frac{y(t)}{k}.$$

The random variables $N_k(t)$ are computed from the data, and we can rewrite them in the form

$$N_k(t) \ = \ y(t) + (N_k(t) - y(t)).$$

Our goal is to find an estimate of $\lambda(t)$. The schematic representation of the problem is shown on Fig. 3.

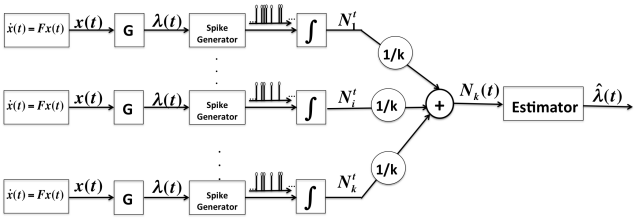


Fig. 3: The schematic representation of the problem formulation

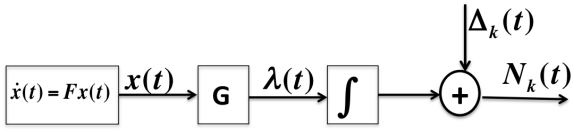


Fig. 4: Exact problem formulation in terms of key system variables. $\Delta_k(t) := N_k(t) - \int_0^t \lambda(s)ds$

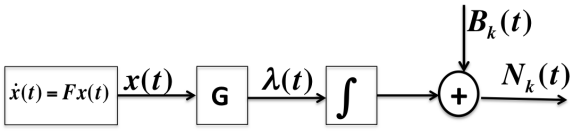


Fig. 5: Problem formulation using an approximation of $\Delta_k(t)$ with Brownian motion, in case when k is sufficiently large

Denote

$$\Delta_k(t) := N_k(t) - y(t),$$

so that

$$N_k(t) \ = \ y(t) + \Delta_k(t),$$

which is represented on Fig. 4. By the Central Limit Theorem, for large k, $\Delta_k(t) \sim N(0, \frac{y(t)}{k})$. Using the fact that N is a process with independent increments, and the fact that the N_i are IID, one can show that Δ_k is approximately a Brownian motion, $\Delta_k \approx B_k$ (Fig. 5). Purely formally, we define

$$\xi_k(t) := \frac{dB_k(t)}{dt},$$

and think of $\xi_k(t)$ as a white noise, normally distributed with zero mean and variance, $\text{var}[\xi(t)] = \frac{\text{var}[N_k(t)]}{t} = \frac{y(t)}{kt}$. (In this conference paper, we focus on algorithms. Follow-up work will discuss a proper mathematical formulation.)
Thus,

$$\begin{aligned} N_k(t) &= y(t) + \Delta_k(t) \approx y(t) + B_k(t), \\ \dot{N}_k(t) &\approx \lambda(t) + \xi_k(t). \end{aligned}$$

See Fig. 6.

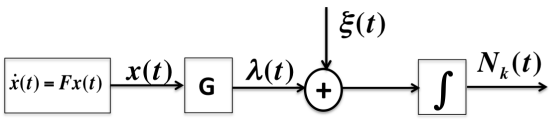


Fig. 6: Formally writing $\xi_k(t) = \frac{dB_k(t)}{dt}$, where B_k was introduced on Fig. 5.

Suppose now that we have a deterministic system,

$$\dot{x} = f(x),$$

which we will at first assume is linear

$$\dot{x} = Fx.$$

The observation that is available for estimation is a non-homogeneous Poisson random process N_t, whose rate is $\lambda(t) = h(x(t)) = Gx$. We assume that $\lambda(t)$ is always positive. This will be guaranteed if the matrix F is Metzler, $G > 0$, and $x(0)$ has positive coordinates. The objective is to obtain an estimate of the state x, or more specifically, of the rate $\lambda(t)$.

Assume next that we have access to k realizations of the process, with the same $x(0)$, so that the rate of the process, $\lambda(t)$ is the same in every realization.

Let's rewrite the full state-space model for the approximate system on Fig. 6 as:

$$\begin{aligned} \frac{dx}{dt} &= Fx, \\ \dot{N}_k(t) &= Gx + \xi, \end{aligned} \tag{1}$$

with unknown initial condition $x(0)$. Notice that if we estimate (x, N_k), we will have in particular an estimate of $\lambda(t) = Gx(t)$. In order to more compactly describe the extended system (1), we introduce the state $z_1 = \begin{bmatrix} x \\ N_k \end{bmatrix}$, with the output y being equal to the last variable $N_k(t)$. Since $N_k(t)$ will typically be very noisy, we do not want our estimator to compute numerical derivatives, so the output that we would like to feed as the input to the state estimator is $N_k(t)$, which contains colored noise.

The extended system can be written in the form:

$$\begin{aligned} \frac{dz_1}{dt} &= F_1 z_1 + G_1 \xi\,, \\ F_1 &= \begin{bmatrix} F & 0 \\ G & 0 \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}\,,\ G_1 = \begin{bmatrix} 0_n \\ 1 \end{bmatrix}. \end{aligned} \tag{2}$$

The output of this system corresponds to the last state of (2), namely

$$y_1(t) = H z_1(t)\,, \quad \text{with} \quad H = \begin{bmatrix} 0_n & 1 \end{bmatrix}. \tag{3}$$

If we would attempt to build a Kalman filter for the estimation of the state z_1, in (2), we would encounter a singular problem with zero output noise covariance matrix. To avoid this case, we apply the method developed in [3], [4], [10], and look for the derivative of y_1:

$$\dot{y}_1 = H_1 z_1(t) + v_1(t),$$

where $H_1 = HF_1$, $v_1(t) = HG_1\xi$, and $E[v_1(t)v_1^T(t)] = HG_1G_1^TH^Tm(t)$ with $\text{cov}[\xi(t)\xi^T(t)] = m(t)$. Now the estimate of $z_1(t)$ is $\hat{z}_1$ obtained as

$$\dot{\hat{z}}_1 = F_1\hat{z}_1 + L(t)(\dot{y}_1 - \dot{\hat{y}}_1) = F_1\hat{z}_1 + L(t)(\dot{y}_1 - H_1\hat{z}_1), \tag{4}$$

where $L = L(t)$ is time varying gain. Since differentiation introduces the noise in the filtering problem, introduce the following change of variables

$$g(t) = \hat{z}_1 - Ly_1. \tag{5}$$

Then,

$$\dot{g}(t) = \dot{\hat{z}}_1 - \dot{L}(t)y_1(t) - L(t)\dot{y}_1(t)$$

If we differentiate g

$$\begin{aligned} \dot{g} &= \dot{\hat{z}}_1 - L\dot{y}_1 - \dot{L}y_1 \\ \dot{g} &= (F_1 - LH_1)g(t) + [F_1L - LH_1L - \dot{L}]y_1(t), \end{aligned} \tag{6}$$

from where we can find $\hat{z}_1(t) = g(t) + L(t)y_1(t)$.
In order to find the time-varying gain $L(t)$, we now solve a routine Kalman filtering problem for (4).
Define the observation error as $e(t) = z_1(t) - \hat{z}_1(t)$, so its derivative is

$$\dot{e}(t) = \dot{z}_1(t) - \dot{\hat{z}}_1 = F_1z_1 + G_1\xi - F_1\hat{z}_1 - L\dot{y}_1 + L\dot{\hat{y}}_1,$$

Therefore,

$$\dot{e}(t) = (F_1 - LHF_1)e(t) - (LHG_1 - G_1)\xi(t) \tag{7}$$

where $\dot{\hat{y}}_1 = HF_1\hat{z}_1$. Denote the covariance of $e(t)$ by $\Sigma := E[e(t)e^T(t)]$, and $\text{Var}[e(t)] = E[(e(t) - \bar{e}(t))(e(t) - \bar{e}(t))^T]$, with $\bar{e}(t) = E[e(t)]$. Also, $\text{cov}[\xi(t)\xi^T(t)] = m(t)$, where $\xi(t)$ is the white noise, $E[\xi(t)] = 0$. The covariance matrix Σ satisfies the Riccati equation:

$$\begin{aligned} \dot{\Sigma} &= (F_1 - LHF_1)\Sigma + \Sigma(F_1 - LHF_1)^T \\ &+ (LHG_1 - G_1)(LHG_1 - G_1)^T m(t) \end{aligned} \tag{8}$$

with

$$L_{opt} = [\Sigma F_1^T H^T + G_1 G_1^T H^T m(t)](HG_1G_1^T H^T)^{-1} m(t)^{-1} \quad (9)$$

$$\dot{L}_{opt} = \dot{\Sigma} F_1^T H^T (HG_1G_1^T H^T)^{-1} m(t)^{-1} - \Sigma F_1^T H^T (HG_1G_1^T H^T)^{-1} \Big(\frac{\dot{m}(t)}{m^2(t)}\Big) \quad (10)$$

In the computation of the optimal gain $L(t)$ from the Riccati equation we use $m(t)$, which we will estimate from the sample variance, which, in turn is estimated from the data. However in the expression (6), the derivative of $m(t)$ will appear through the term $\dot{L}(t)$, but this numerical derivative does not affect the estimate directly.

In the next Section, we first consider simple Luenberger observers, by picking any constant stabilizing L, and show that results also look satisfactory. For the Luenberger observer design we observe the system defined by (2) and (3), and ignore the noise, ξ. The results for the full Kalman filter design will be provided in a follow-up paper.

III. Examples

In this section we will demonstrate the advantage of model based methods in cases where we have some prior information about the system or about the inputs acting upon the system. Consider an output $\lambda(t)$ generated by the following system with a known frequency ω_0 :

$$\begin{aligned} \dot{x}_0 &= 0\,, & x_0(0) &= 2.5, \\ \dot{x}_1(t) &= x_2\,, & x_1(0) &= 0.15, \\ \dot{x}_2(t) &= -\omega_0^2 x_1\,, & x_2(0) &= 0.25, \\ \lambda(t) &= x_0(t) + x_2(t), \end{aligned}$$

where the initial conditions are such that $\lambda(t) > 0$ $(\forall t)$, and $\omega_0 = 10$. If we use the prior information of the frequency, an oscillatory estimator for this system can be represented as

$$\begin{aligned} &\dot{x} = \begin{bmatrix} \dot{x}_0 & \dot{x}_1 & \dot{x}_2 \end{bmatrix}^T = Fx\,, \quad F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -w_0^2 & 0 \end{bmatrix}, \\ &\lambda(t) = Gx\,, \quad G = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}, \\ &\dot{N} = Gx, \end{aligned} \quad (11)$$

where we introduced another state, N, to represent the counting process. The extended system is now

$$\dot{z}_1 = \begin{bmatrix} \dot{x}_0 & \dot{x}_1 & \dot{x}_2 & \dot{N} \end{bmatrix}^T = F_1 z_1\,,$$

$$F_1 = \begin{bmatrix} F & 0 \\ G & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -w_0^2 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix},$$

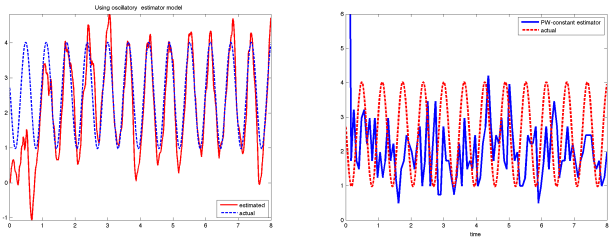
with the output

$$\dot{y}_1 = Hz_1\,, \quad H = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}.$$

Thus, we can build an observer for the system as

$$\frac{d\hat{z}_1}{dt} = F_1 \hat{z}_1 + L(y_1 - H\hat{z}_1). \quad (12)$$

We picked the observer gains L so that the observer poles are arbitrarily chosen as $-10, -9, -8$ and -7. The simulation results and comparison with the piecewise constant estimator are shown on Fig. 7. We have challenged the "naive method" by increasing and decreasing the bin size, b, while keeping the same number of realizations k as in our model-based approach, and the simulation results do not indicate any improvement of the naive, piecewise constant method, in comparison to the observer based method.

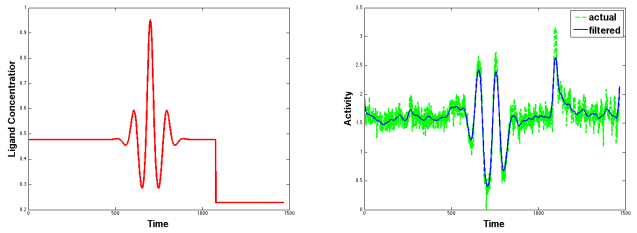


(a) Observer Method, $k = 50$ (b) Naive method, k=50 M=100

Fig. 7: Comparison of estimates obtained using piecewise constant estimator and a model-based oscillatory observer

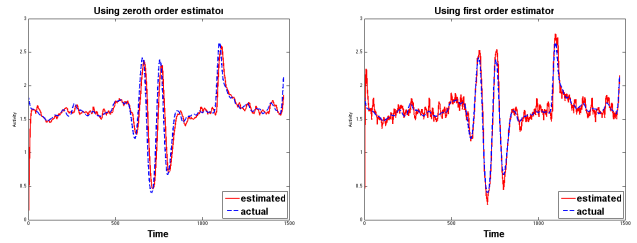
A. A biological example

This work was motivated by current work we are pursuing with experimental collaborators in the design of microfluidics devices that will allow the same inputs to be fed to a population of chemotactic bacteria, and microscope-based observations of tumbling events will be used for estimation of the tumbling rate (a function of chemotactic protein concentrations). Since this data is not available yet, we use here experimental data from the paper [6], which measured the actual rates through FRET (Fluorescence Resonance Energy Transfer) techniques for a particular strain of *E. coli* bacteria. Since FRET measurements are very noisy, we first low-pass filtered this data in order to simulate $\lambda(t)$ and generate artificial events. We see numerically that a simple observer-based method recovers the FRET measurement with roughly the same amount of noise. (Of course, if FRET data is available, there would be no need for our observers. The goal, however, is to study similar questions for other bacterial species for which FRET measurements, which require extensive genetic modifications, are not available.) Note that since we did not have a priori information about the nature of the output, we have tested the zeroth order observer, which is essentially an observer designed to estimate constant signals, and also a first order or linear observer which would be optimal for linear or piecewise linear signals. The results are depicted on Figures 8, 9, and 10. As expected the linear observer shows satisfactory results.



(a) (b)

Fig. 8: (a) Input (ligand concentration) and (b) Measured output and a filtered output used for the estimation process



(a) (b)

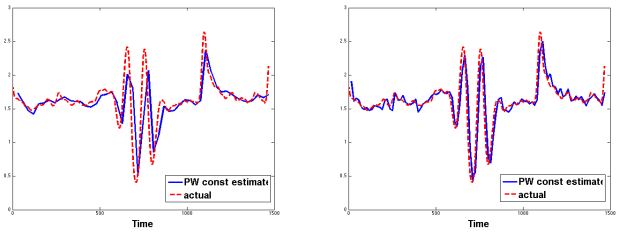
Fig. 9: Estimation using the observer method of zeroth order (i.e. constant estimator) (a), and first order (i.e. linear estimator) (b). $k = 50$

To further motivate our work, we will use the developed approach to test for the scale-invariance property, namely scale invariance of the complete output trajectory with respect to a rescaling of input magnitudes [13], [6], for various bacterial species and classes of inputs. To illustrate, we will use the "gold standard" model for *E. coli* chemotactic pathway from [12], to generate the artificial events, as before, and test for scale invariance using the following two experiments: in the first experiment we preadapt the system to a constant ligand concentration $L_1 = 200$, and use the preadapted values as the initial conditions when this system is presented with an input $L_2 = 200 + 100\sin(5t)$; in the second experiment the inputs are scaled by a factor $p = 2$. The values are selected according to the results presented in [12], so that the system should exhibit the scale invariance property. The oscillatory estimator model presented above, can help in detecting the scale invariance property from the "spike" data. The results and the comparison with the naive method are shown on Figures 11 and 12.

B. A simple nonlinear observer model of an E. Coli chemotactic pathway

As we have seen in the previous examples, the naive method does not take advantage of the known input signal. Consider a system given by

$$\begin{aligned} \dot{x} &= f(x,u), \\ y &= h(x,u), \end{aligned} \tag{13}$$



(a) $M = 50$ (b) $M = 100$

Fig. 10: Naive piece-wise constant estimator with subinterval width M

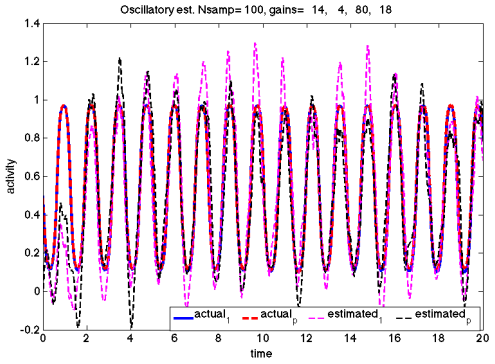


Fig. 11: Detection of scale invariance property using the oscillatory estimator. k=100 realizations.

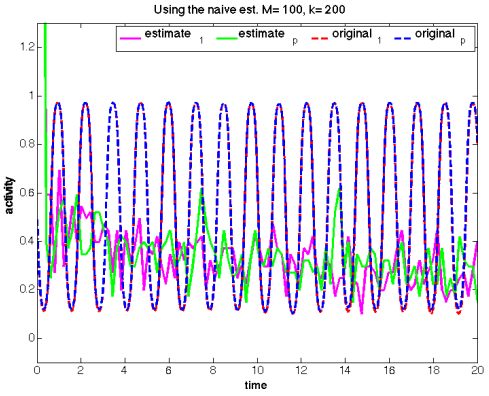


Fig. 12: Detection of scale invariance property using the naive estimator. The estimator is not able to correctly reconstruct the underlying output, regarless of the number of realizations.

where $y(t) \in \mathbb{R}^1$ corresponds to the measured output ("activity") in a simple but realistic model of the *E. coli* chemotactic pathway, see for example [6], [12], [14], [13]. Here, $x(t) \in \mathbb{R}^n$ is the unknown internal state, and $u(t)$ is a known input signal. As before, we introduce the extended $(n+1)-$dimensional system that models the counting pro-

cess by adding an integrator:

$$\begin{aligned} \dot{x} &= f(x,u), \\ \dot{z} &= h(x,u). \end{aligned} \tag{14}$$

Specifically, we take a simplified version of the model from [12], by picking the Hill coefficients in the model equal to 1, and all coefficients set to unity. Thus, we consider the following model:

$$\begin{aligned} \dot{x} &= \frac{1}{2} - \frac{1}{1+\frac{u}{x}} \\ \dot{z} &= \frac{1}{1+\frac{u}{x}}, \end{aligned} \tag{15}$$

with output z. An observer for this $n-$ dimensional system can be obtained as follows:

$$\begin{aligned} \dot{\hat{x}} &= \frac{1}{2} - \frac{1}{1+\frac{u}{\hat{x}}} \\ \dot{\hat{z}} &= \frac{1}{1+\frac{u}{\hat{x}}} - L(\hat{z}-z). \end{aligned} \tag{16}$$

Define the errors

$$\begin{aligned} e_1(t) &= \hat{x}(t) - x(t), \\ e_2(t) &= \hat{z}(t) - z(t). \end{aligned} \tag{17}$$

Then,

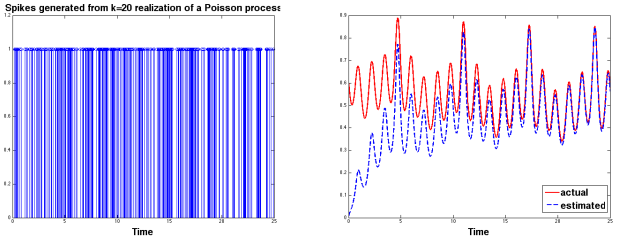
$$\begin{aligned} \dot{e}_1(t) &= \dot{\hat{x}} - \dot{x}(t) = \frac{u(x-\hat{x})}{(x+u)(\hat{x}+u)} = \frac{-ue_1}{(x+u)(\hat{x}+u)}. \\ \dot{e}_2(t) &= \dot{\hat{z}} - \dot{z}(t) = -\Big(\frac{1}{1+\frac{u}{x}} - \frac{1}{1+\frac{u}{\hat{x}}}\Big) - L(\hat{z}-z) \\ &= \frac{e_1 u}{(x+u)(\hat{x}+u)} - Le_2. \end{aligned} \tag{18}$$

It can be shown that if $0 < \alpha \leq u(t) \leq \beta \ \forall t$, then both $\hat{x}(t)$ and $x(t)$ are bounded, and $e_1(t) \to 0$ and $e_2(t) \to 0$ as $t \to \infty$.

For the system defined by (15) we have applied an input signal given by $u(t) = 2+\sin(5t)+0.5\sin(t)-0.2\cos(3t-20)$ and generated $k = 20$ realizations of the process, shown on left panel of Figure 13. Then we applied the observer-based method described above to estimate the output of the process. We assumed that the model has an initial state $x_0 = 3$ and that the observer initial state was picked to be 0. The results indicate that the observer performs extremely well.

IV. Conclusion

We proposed a model-based approach to designing observers for jump processes driven by a deterministic system, and numerically compared their performance to that of various other techniques. Our approach is shown to perform well in identifying rates produced by several classes of linear systems as well as a model of *E. coli* chemotaxis. Theoretical analysis and comparisons of these algorithms are the subject of current work.



(a) (b)

Fig. 13: Spikes (events) used as an input to the observer (a), and comparison between the true output and an estimate obtained by using a nonlinear observer (b). $k = 20$ realizations were used. L was picked to be 1.

V. Acknowledgment

We are grateful to Triet Pham for bringing reference [11] to our attention, and for useful discussions. This work was supported in part by NIH grant 1R01GM100473.

References

[1] B. L. Arkin and L. M. Leemis. Nonparametric estimation of the cumulative intensity function for a nonhomogeneous poisson process from overlapping realizations.

[2] R. K. Boel and V. E. Benes. Recursive nonlinear estimation of a diffusion acting as the rate of an observed poisson process. *IEEE Transactions on Information Theory*, 26(5):561–575, 1980.

[3] A. E. Bryson, Jr. and L. J. Henrikson. Estimation using sampled data containing sequentially correlated noise. *Journal of Spacecraft and Rockets*, 5:662–665, June 1968.

[4] A. E. Bryson, Jr. and D. Johansen. Linear filtering for time-varying systems using measurements containing colored noise. *IEEE Trans. Automat. Contr.*, 10:4–10, January 1965.

[5] S. Koyama, U. Eden, E. Brown, and R. Kass. Bayesian decoding of neural spike trains. *Annals of the Institute of Statistical Mathematics*, 62(1):37–59, 2010.

[6] M. D. Lazova, T. Ahmed, D. Bellomo, R. Stocker, and T. S. Shimizu. Response rescaling in bacterial chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*, 108:13870–5, 2011 Aug 16 2011.

[7] L. M. Leemis. Nonparametric estimation of the cumulative intensity function for a nonhomogeneous poisson process. *Manage. Sci.*, 37(7):886–900, July 1991.

[8] L. M. Leemis. Nonparametric estimation and variate generation for a nonhomogeneous poisson process from event count data. *IIE Transactions*, 36:1155–1160, 2004.

[9] W. A. Massey, G. A. Parker, and W. Whitt. Estimating the parameters of a nonhomogeneous poisson process with linear rate. *Telecommunication Systems*, 5(2):361–388, 1996.

[10] P. J. Moylan. A note on Kalman-Bucy filters with zero measurement noise. *IEEE Trans. Automat. Contr.*, 19(3):263–264, June 1974.

[11] H. Shimazaki and S. Shinomoto. A method for selecting the bin size of a time histogram. *Neural Comput.*, 19(6), June 2007.

[12] T. S. Shimizu, Y. Tu, and H. C. Berg. A modular gradient-sensing network for chemotaxis in Escherichia coli revealed by responses to time-varying stimuli. *Mol. Syst. Biol.*, 6:382, 2010.

[13] O. Shoval, L. Goentoro, Y. Hart, A. Mayo, E. Sontag, and U. Alon. Fold-change detection and scalar symmetry of sensory input fields. *Proceedings of the National Academy of Sciences*, 107(36):15995–16000, 2010.

[14] M. Skataric and E. D. Sontag. A characterization of scale invariant responses in enzymatic networks. *PLoS Comput Biol*, 8(11):e1002748, 11 2012.

[15] J. H. Van Schuppen. Filtering, prediction and smoothing for counting process observations, a martingale approach. *SIAM Journal of Applied Math.*, 32(3):552–570, May 1977.