'Unobserved' Monte Carlo method for identification of partially observed nonlinear state space systems, Part II: Counting Process Observations.

Victor Solo
School of Electrical Engineering and Telecommunications
University of New South Wales
Sydney NSW 2052
Australia
vsolo@syscon.ee.unsw.edu.au

Abstract

We present a simple simulation method for generating an approximate likelihood function for fitting partially observed (with counting process observations) nonlinear stochastic differential equations. We also discuss use of the method to generate approximate maximum likelihood estimators. We also mention methods based on density evolution equations.

1 Introduction

Parameter estimation and signal estimation for systems with counting process observations are becoming increasingly important. Applications include optical communications [1],[2]; auditory physiology [1] and communication networks [2]. And an exciting problem of recent origin which with this author is involved is neural encoding [3],[4]. In these applications the doubly stochastic Poisson process is the most common model having an underlying state of interest generated from a nonlinear state space equation (a diffusion) and nonhomogeneous Poisson observations with rate function depending on the state.

The problem of system identification for these kinds of problems has remained unresolved because the likelihood function cannot be computed in closed form. Previous approaches have relied on approximate filters generated by Taylor series methods [1],[4]. Related work [5] has used Markov Chain Monte Carlo methods on a particular problem with point process observations.

As described in part I [6] of this work for analog observations, we have develped a simulation approach suited to this problem but much simpler than Markov Chain Monte Carlo . Here we extend the method to counting process observations. We also discuss the use of the unnormalised nonlinear state estimation filter (UNF) [2] to construct the likelihood function in simple cases. [7] has discussed the use of the UNF in the context of neural encoding but not to do statistical inference, nor are simulation methods discussed. Although the use of the UNF for system identification has been discussed for analog observations in [6],[8] its use with counting process observations seems to be new. Note that the development here is self contained, not relying on [6], however the ideas behind the Part I and Part II are much the same.

2 Poisson Observations of a Diffusion

To describe our simulation method we setup a discrete diffusion

$$x_{i+1} - x_i = f(i, x_i) + \sigma(i, x_i)n_i \tag{2.1}$$

where n_i is a Gaussian white noise of unit covariance. We think of (2.1) as a discretisation of the continuous time diffusion or nonlinear state $x(\tau)$. So with δ as a small sampling interval, $(x_i, f_i = f(i, x_i), \sigma_i = \sigma(i, x_i))$ approximates

$$(x(i\delta), f_c(i\delta, x_i)\delta, \sigma_c(i\delta, x_i)\sqrt{\delta})$$

where $f_c, Q_c = \sigma_c\sigma_c^T$ are the infinitesimal drift and covariance (with $X_\delta(\tau) = X(\tau + \delta) - X(\tau)$)

$$\begin{aligned} E(X_\delta(\tau)|X(\tau)) &= f_c(\tau, X(\tau)\delta + o(\delta) \\ E(X_\delta(\tau)X_\delta(\tau)^T|X(\tau)) &= Q_c(\tau, X(\tau)\delta + o(\delta) \end{aligned} \tag{2.2}$$

The observation process is a Poisson process N_t(= # spikes in $0, t$) related to the state through its rate function

$$dN_t = \lambda(t, x(t))dt + dM_t \tag{2.3}$$

i.e. conditional on the whole state trajectory N_t is an inhomogeneous Poisson process with rate function

$\lambda(t, x(t))$. And M_t is then a martingale. For the simulation method we consider discrete observations

$$\begin{aligned} y_i &= N_i^\delta = \lambda(i\delta, x_i)\delta + m_i && (2.4) \\ &= \#\text{events in } i\delta, i\delta + \delta && (2.5) \end{aligned}$$

We collect the observations into data and state sequences

$$N_{1,n} = (N_1^\delta, \cdots, N_n^\delta)^T; X_{1,n} = (X_1, \cdots, X_n)^T$$

and suppose $f_i, \lambda_i = \lambda(i\delta, x_i), Q_i = \sigma_i\sigma_i^T$ are parameterised by a vector of parameters θ.

3 'Unobserved' Monte Carlo

The discussion in this section parallels that in [6]. For the model (2.1),(2.4) we can write down formulae for the joint densities $P(X_{1,n}), P(N_{1,n}, X_{1,n})$ and then the likelihood is given by (with $dX_{1,n}$ meaning $dx_1 \cdots dx_n$)

$$L_n = P(N_{1,n}) = \int P(N_{1,n}|X_{1,n})P(X_{1,n})dX_{1,n} \quad (3.1)$$

But the integration here is intractable and this leads to our simple simulation technique. Generate independent draws of the *unobserved* state trajectory $X_{1,n}^{(r)}, r = 1, \cdots, m$ and estimate the likelihood by a Monte Carlo average.

$$\hat{L}_n = \frac{1}{m}\Sigma_1^m P(N_{1,n}|X_{1,n}^{(r)}) \quad (3.2)$$

Generating the draws is easy. It just means simulating (2.1) for n time steps and requires only generating a Gaussian white noise. The approach is direct and has none of the difficulties associated with Markov Chain Monte Carlo .

3.1 Simulated Gradient

To generate a gradient for maximizing the likelihood we find (with $\nabla_\theta = \frac{d}{d\theta}$) on differentiating

$$\begin{aligned} \nabla_\theta L_n &= \nabla_\theta P(N_{1,n}) \\ &= \int P(N_{1,n}|X_{1,n})P(X_{1,n})SdX_{1,n} \\ S &= S(N_{1,n}|X_{1,n}) + S(X_{1,n}) \\ S(N_{1,n}|X_{1,n}) &= \nabla_\theta lnP(N_{1,n}|X_{1,n}) && (3.3) \\ S(X_{1,n}) &= \nabla_\theta lnP(X_{1,n}) && (3.4) \end{aligned}$$

and the Monte-Carlo estimate of the likelihood gradient is then

$$\hat{\nabla}_\theta L_n = \frac{1}{m}\Sigma_1^m[S(N_{1,n}|X_{1,n}^{(r)}) + S(X_{1,n}^{(r)})]P(N_{1,n}|X_{1,n}^{(r)}) \quad (3.5)$$

To get the gradient of the log-likelihood, just divide (3.5) by (3.2).

3.2 Simulated EM algorithm

For an EM algorithm [9] the criterion function $\int lnP_1P_0(X_{1,n}|N_{1,n})dX_{1,n}, lnP_1 = lnP_1(N_{1,n}, X_{1,n})$ (where subscripts 0,1 correspond to parameter values θ_0, θ_1) can be replaced by

$$M(\theta_1, \theta_0) = \int (lnP_1)P_0(N_{1,n}|X_{1,n})P_0(X_{1,n})dX_{1,n}$$

and this leads to the Monte-Carlo estimate

$$\begin{aligned} \hat{M}(\theta_1, \theta_0) &= \frac{1}{m}\Sigma_1^m Z^{(r)}P_0(N_{1,n}|X_{1,n}^{(r)}) && (3.6) \\ Z^{(r)} &= [lnP_1(N_{1,n}|X_{1,n}^{(r)}) + lnP_1(X_{1,n}^{(r)})] \end{aligned}$$

We need to elaborate the SLIK method further to show how to compute $P(N_{1,n}|X_{1,n})$ in an efficient way; how to compute gradients with respect to θ in an efficient way Resolution of these issues follows on the result in the next subsection.

4 Conditional Likelihood Ratio

Introduce the reference or null model or $*$ model (the reason for this will become apparent)

$$N_i^\delta = \delta + m_i$$

which describes a Poisson process with unit rate. The corresponding likelihood function is

$$P_*(N_{1,n}) = \Pi_1^n(\delta)^{N_i^\delta}e^{-\delta}/N_i^\delta!$$

Then we have the decomposition

$$\begin{aligned} P(N_{1,n}|X_{1,n}) &= \Lambda_n P_*(N_{1,n}) \\ \Lambda_n &= e^{H_n} \\ H_n &= \Sigma_1^n N_i^\delta log\lambda_i - (\lambda_i - 1)\delta \end{aligned}$$

Note that the reference model for N_i^δ does not depend on unknown parameters.

Proof:

We have

$$P(N_{1,n}|X_{1,n}) = P(N_n^\delta|N_{1,n-1}, X_{1,n})P(N_{1,n-1}|X_{1,n})$$

But given $X_{1,n-1}$ then the density of $N_{1,n-1}$ is fully specified and so independent of x_n. Also given x_n the density of N_n^δ is fully specified so

$$P(N_{1,n}|X_{1,n}) = p(N_n^\delta|x_n)P(N_{1,n-1}|X_{1,n-1})$$

and iterating this gives

$$P(N_{1,n}|X_{1,n}) = \Pi_1^n p(N_i^\delta|x_i)$$

Now invoking the inhomogeneous Poisson assumption gives

$$= \Pi_1^n e^{-\lambda_i\delta}(\lambda_i\delta)^{N_i^\delta}/N_i^\delta!$$

Now this likelihood has a singularity problem as $\delta \to 0$. It is to fix that problem, that we introduce the reference model. We can write then

$$\begin{aligned} P(N_{1,n}|X_{1,n}) &= \Pi_1^n \frac{e^{-\lambda_i \delta}(\lambda_i \delta)^{N_i^\delta}}{e^{-\delta}(\delta)^{N_i^\delta}} \frac{e^{-\delta}(\delta)^{N_i^\delta}}{N_i^\delta !} \\ &= \Pi_1^n e^{-(\lambda_i - 1)\delta} e^{N_i^\delta log \lambda_i} P_*(N_{1,n}) \end{aligned}$$

and this gives the result.

4.1 Simulated Gradient Revisited

Applying the LRL to (3.3) gives

$$\begin{aligned} S(N_{1,n}|X_{1,n}^{(r)}) &= \nabla_\theta H_n^{(r)} \\ &= \Sigma_1^n N_i^\delta \frac{dlog\lambda_i^{(r)}}{d\theta} - \frac{d\lambda_i^{(r)}}{d\theta}\delta \end{aligned}$$

And from this using (3.2),(3.5) we get the simulated gradient log-likelihood

$$\hat{\nabla}_\theta ln L_n = \frac{\frac{1}{m}\Sigma_{r=1}^m (\nabla_\theta H_n^{(r)} + \nabla_\theta ln P(X_{1,n}^{(r)}))\Lambda_n^{(r)}}{\frac{1}{m}\Sigma_{r=1}^m \Lambda_n^{(r)}} \tag{4.1}$$

The formula simplifies in particular cases. To get $\nabla_\theta ln P(X_{1,n})$ we write via (2.1)

$$\begin{aligned} P(X_{1,n}) &= \Pi_1^n p(x_i|x_{i-1}) \\ \Rightarrow \nabla_\theta ln P(X_{1,n}) &= \Sigma_1^n \nabla_\theta ln p(x_i|x_{i-1}) \end{aligned} \tag{4.2}$$

where by $p(x_1|x_0)$ we mean $p(x_1)$.

4.2 Simulated EM Revisited

Here we find from the LRL

$$\begin{aligned} ln P_1(N_{1,n}|X_{1,n}^{(r)}) &= H_n^{(r)(1)} \\ &= \Sigma_1^n (N_i^\delta ln \lambda_i^{(r)(1)} - (\lambda_i^{(r)(1)} - 1)\delta) \end{aligned}$$

where the superscript (1) denotes evaluation with $\theta = \theta_1$. Applying the LRL, the simulated M function can be replaced by

$$\hat{M}(\theta_1, \theta_0) = \frac{1}{m}\Sigma(H_n^{(r)(1)} + ln P_1(X_{1,n}^{(r)}))\Lambda_n^{(r)(0)}$$

5 Nonlinear Filtering

Here we give a new heuristic derivation of the UNF and show how to use it to compute the likelihood function. The heuristic derivation will be useful in making the UNF more accessible. As discussed in [10] the UNF obeys a forwards Kolmogorov equation (FKE) and can be derived in two ways. One route is via a data modified forward Chapman Kolmogorov equation (CKE) and the other via Ito's lemma. Both methods use a test function technique. A very nice and rigorous Ito type derivation can be obtained from the development in [11]. The original derivation in [2] is less transparent; while the method in [7] is adhoc. We follow the CKE route.

The idea is to develop an evolution equation based on the joint density $p(x_n, N_{1,n})$. We have

$$\begin{aligned} p(x_{n+1}, N_{1,n+1}) &= p(x_{n+1}, N_{n+1}^\delta, N_{1,n}) \\ &= p(N_{n+1}^\delta|x_{n+1}, N_{1,n}) p(x_{n+1}, N_{1,n}) \end{aligned}$$

and using the definition of the observation process gives

$$\begin{aligned} &= p(N_{n+1}^\delta|x_{n+1}) p(x_{n+1}, N_{1,n}) \\ &= p(N_{n+1}^\delta|x_{n+1}) \int p(x_{n+1}, x_n, N_{1,n}) dx_n \\ &= p(N_{n+1}^\delta|x_{n+1}) \int p(x_{n+1}|x_n) p(x_n, N_{1,n}) dx_n \end{aligned}$$

Now apply one step of the LRL to get

$$\begin{aligned} &= p_*(N_{n+1}^\delta) l_{n+1} \int p(x_{n+1}|x_n) p(x_n, N_{1,n}) dx_n \\ l_{n+1} &= e^{N_{n+1}^\delta ln \lambda_{n+1} - (\lambda_{n+1} - 1)\delta} \end{aligned}$$

Finally we introduce the state dependent likelihood ratio

$$\rho(x_{n+1}, N_{1,n+1}) = \frac{p(x_{n+1}, N_{1,n+1})}{P_*(N_{1,n+1})} \tag{5.1}$$

to find that

$$\rho(x_{n+1}, N_{1,n+1}) = l_{n+1} \int p(x_{n+1}|x_n) \rho(x_n, N_{1,n}) dx_n \tag{5.2}$$

which is the data adjusted forwards CKE, the main result of this section. Without the data factor l_{n+1} it is a fully observed forwards CKE.

The important observation to be made now is that the likelihood ratio can be recovered by integrating out the state, thus

$$\frac{P(N_{1,n+1})}{P_*(N_{1,n+1})} = \int \rho(x_{n+1}, N_{1,n+1}) dx_{n+1} \tag{5.3}$$

It is straightforward to derive the counting process version of Bucy type representation theorem [12] which as in [10] follows from Bayes rule and the LRL lemma. We have

$$\begin{aligned} p(x_n|N_{1,n}) &= \frac{p(N_{1,n}|x_n)\pi(x_n)}{P(N_{1,n})} \\ &= \frac{p(N_{1,n}|x_n)\pi(x_n)}{\int p(N_{1,n}|x_n)\pi(x_n) dx_n} \end{aligned}$$

However we now use the LRL to find

$$\begin{aligned} p(N_{1,n}|x_n) &= \int P(N_{1,n}|X_{1,n}) P(X_{1,n}|x_n) dX_{1,n-1} \\ &= P_*(N_{1,n}) \int \Lambda_n P(X_{1,n}|x_n) dX_{1,n-1} \\ &= P_*(N_{1,n}) E_*(\Lambda_n|x_n) \end{aligned} \tag{5.4}$$

and so we get Bucy's representation theorem for the conditional density.

$$p(x_n|N_{1,n}) = \frac{E_*(\Lambda_n|x_n)\pi(x_n)}{E_*(\Lambda_n)}$$

We also note from (5.1) and (5.4) that

$$\rho(x_n, N_{1,n}) = E_*(\Lambda_n|x_n)\pi(x_n) \qquad (5.5)$$

And the continuous time version of this result is used by [11], together with Ito's lemma to develop the UNF. Our route is different proceeding instead from (5.2).

To derive the UNF we need to move to discretised continuous time and to a more careful notation for describing densities. Thus in place of $p(x_{n+1})$ we write

$$\begin{aligned} p(n+1, x) &= Prob(X_{n+1} \text{ is near } x) \\ &= lim_{||h||\to 0} P(x \leq X_{n+1} \leq x+h)\frac{1}{||h||} \end{aligned}$$

or in continuous time

$$p(\tau, x) = Prob(X(\tau) \text{ is near } x)$$

This 'equation' should be interpreted to mean that the likelihood ratio $p(\tau,x)/p(\tau,x^{'})$ gives the likelihood that $X(\tau)$ is near x rather than near $x^{'}$.

The natural continuous time notation for $\rho(x_n, N_{1,n})$ will be $\rho(\tau, x, N_{0,\tau})$ where $N_{0,\tau} = N(t) : 0 \leq t \leq \tau$ but we compact this to $\rho(\tau, x)$ allowing τ to do double duty. To keep expressions compact we use $p_{\tau+\delta,x|\tau,u}$ for the Markov transition density rather than $p(\tau+\delta, x|\tau, u)$. Note that this notation anyway differs from the more standard Markov notation $p(\tau, u : \tau+\delta, x)$.

Now the data adjusted forward CKE (5.2) can be written

$$\begin{aligned} \rho(\tau+\delta, x) &= l_{\tau,\delta}(N_\delta|x)\int p_{\tau+\delta,x|\tau,u}\rho(\tau,u)du \\ &= [l_{\tau,\delta}(N_\delta|x) - 1]\int p_{\tau+\delta,x|\tau,u}\rho(\tau,u)du \\ &+ \int p_{\tau+\delta,x|\tau,u}\rho(\tau,u)du \qquad (5.6) \end{aligned}$$

where (note the switch to continuous time quantities)

$$\begin{aligned} l_{\tau,\delta}(N_\delta|x) &= e^{N_\delta log\lambda(\tau,x) - (\lambda(\tau,x)-1)\delta} \\ N_\delta &= N(\tau+\delta) - N(\tau) \end{aligned}$$

Looking carefully at the two terms in (5.6) we see that the second term uses data up to time τ and brings it forward to time $\tau+\delta$ through the Markov transition density. The first term does this also but also inputs the new data supplied in the interval $(\tau, \tau+\delta)$. With a fully observed state the first term vanishes.

Let $\psi(x)$ be an arbitary test function and consider the normalised version of

$$\frac{1}{\delta}E(\psi(X(\tau+\delta))|N_{0,\tau+\delta}) - E(\psi(X(\tau))|N_{0,\tau})$$

namely

$$\begin{aligned} \mathcal{E}_\delta &= \frac{1}{\delta}[\int \rho(\tau+\delta,x)\psi(x)dx - \int \rho(\tau,x)\psi(x)dx] \\ \to \mathcal{E} &= \int \frac{\partial\rho}{\partial\tau}\psi dx, \text{ as } \delta \to 0 \end{aligned}$$

On the other hand using the forward CKE in the first part of $\mathcal{E}_\delta$ we get $\mathcal{E}_\delta = B_\delta + C_\delta$ with

$$\begin{aligned} B_\delta &= \int \frac{(l_{\tau,\delta}(N_\delta|x) - 1)}{\delta}[\int p_{\tau+\delta,x|\tau,u}\rho(\tau,u)]\psi(x)dxdu \\ C_\delta &= \frac{1}{\delta}[\int p_{\tau+\delta,x|\tau,u}\rho(\tau,u)du\psi(x)dx \\ &- \int \rho(\tau,x)\psi(x)dx] \end{aligned}$$

Note that C_δ is the term that would appear in the fully observed case. Changing dummy variables gives

$$\begin{aligned} C_\delta &= \frac{1}{\delta}[\int p_{\tau+\delta,u|\tau,x}\rho(\tau,x)dx\psi(u)du \\ &- \int \rho(\tau,x)\psi(x)dx] \end{aligned}$$

which is the unnormalised version of iterated conditional expectation i.e.

$$E(\psi(X(\tau+\delta))|N_{0,\tau+\delta}) = E(E(\psi(X(\tau+\delta))|X(\tau))|N_{0,\tau+\delta})$$

Continuing

$$C_\delta = \frac{1}{\delta}[\int p_{\tau+\delta,u|\tau,x}[\psi(u) - \psi(x)]du\rho(\tau,x)dx$$

Now apply a Taylor series to second order giving

$$\begin{aligned} &= \frac{1}{\delta}\int p_{\tau+\delta,u|\tau,x}[(u-x)\psi^{'} \\ &+ \frac{1}{2}trace((u-x)(u-x)^T\psi^{''})]\rho(\tau,x)dx + o(\delta) \end{aligned}$$

where $\psi^{'} = \frac{\partial\psi}{\partial x}, \psi^{''} = \frac{\partial^2\psi}{\partial x\partial x^T}$. Continuing

$$= \int [-f_c^T\psi^{'} + \frac{1}{2}trace(Q_c\psi^{''})]\rho(\tau,x)dx + o(\delta)$$

where we have recognised the infinitesimal drift and covariance in (2.2). Now integrate by parts and assume boundary terms to vanish to find $C_\delta = C + o(\delta)$

$$\begin{aligned} C &= \int \psi(x)L^+\rho dx \\ L^+\rho &= -\frac{\partial}{\partial x}(f_c\rho) + \frac{1}{2}trace(\frac{\partial^2}{\partial x\partial x^T}(Q_c\rho)) \end{aligned}$$

In the absence of the B_δ term we have $\mathcal{E} = C$ and since ψ is arbitary we deduce the fully observed FKE $\frac{\partial \rho}{\partial \tau} = L^+ \rho$.

Now looking to B_δ we observe that as $\delta \to 0$ N_δ "becomes binomial" taking one of two values ; 0 (no spike) or (a spike). At a time τ away from a spike time (so that $N_\delta = 0$) we find

$$l_{\tau,\delta}(N_\delta|x) - 1 = e^{-(\lambda(\tau,x)-1)\delta} - 1 = -(\lambda(\tau,x)-1)\delta + o(\delta)$$

Also as $\delta \to 0$ the transition probability density becomes a Dirac delta so

$$\begin{aligned} & \int \frac{l_{\tau,\delta}(N_\delta|x) - 1}{\delta} p_{\tau+\delta,x|\tau,u} \psi(x) dx \\ & \to \quad -(\lambda(\tau,u) - 1)\psi(u) \\ \Rightarrow B_\delta & \to \quad B = -\int (\lambda(\tau,u) - 1)\psi(u)\rho(\tau,u) du \end{aligned}$$

Since ψ is arbitary, equating terms ($\mathcal{E} = B + C$) gives

$$\frac{\partial \rho}{\partial \tau} = L^+ \rho(\tau,x) - (\lambda(\tau,x) - 1)\rho(\tau,x) \qquad (5.7)$$

On the other hand at a time τ near a spike

$$\begin{aligned} l_{\tau,\delta}(N_\delta|x) - 1 &= \lambda(\tau,x) e^{-(\lambda(\tau,x)-1)\delta} - 1 \\ &= \lambda(\tau,x)(1 - (\lambda(\tau,x) - 1)\delta + o(\delta)) - 1 \\ &= \lambda(\tau,x) - 1 + O(\delta) \qquad (5.8) \end{aligned}$$

which blows up when we divide by $\delta \to 0$. This means $\rho(\tau, x)$ takes a jump at a spike time. To see how to handle this properly we simply return to (5.6) and recall our description of the two terms to see that on letting $\delta \to 0$ and using (5.8) we get

$$\rho(\tau+, x) - \rho(\tau-, x) = (\lambda(\tau,x) - 1)\rho(\tau-, x) \qquad (5.9)$$

We can write (5.7),(5.9) jointly as an Ito equation

$$d_\tau \rho(\tau,x) = L^+ \rho(\tau,x) d\tau + (dN - d\tau)(\lambda(\tau,x) - 1)\rho(\tau,x)$$

which is the desired UNF. Note that (5.9) was derived by [7] by a different method. Naturally rigorous justification of the limiting arguments used here is required.

Numerical solution in low dimensional problems can be accomplished by discretisation. A natural reliable method for this parabolic partial differential equation is the Crank-Nicholson scheme [13]. Details will be discussed elsewhere.

6 Appendix - BKE

For completeness and to facilitate easy comparison with our development, we include this standard derivation of the backwards Kolmogorov equation [14]. We start with a backward CKE

$$p_{\tau,x|\tau-\delta,u} = \int p_{\tau,x|t,z} p_{t,z|t-\delta,u} dz$$

So that

$$\begin{aligned} & p_{\tau,x|\tau-\delta,u} - p_{\tau,x|t,u} \\ = & \int [p_{\tau,x|t,z} - p_{\tau,x|t,u}] p_{t,z|t-\delta,u} dz \end{aligned}$$

Now expand in a Taylor series to second order

$$\begin{aligned} &= \int [(z-u)\frac{\partial p}{\partial u} + \frac{1}{2}(z-u)(z-u)^T \frac{\partial^2 p}{\partial u \partial u^T}] p_{t,z|t-\delta,u} dz \\ &= \delta f_c(t-\delta,u)\frac{\partial p}{\partial u} + \frac{1}{2} trace(Q_c(t-\delta,u)\frac{\partial^2 p}{\partial u \partial u^T}\delta + o(\delta) \end{aligned}$$

where we have recognised the infinitesimal drift and covariance. Divide by δ and let $\delta \to 0$ to get the result.

7 Conclusion

In this paper we have presented a new simulation method for approximate maximum likelihood identification of nonlinear pertially observed state space systems with counting process observations. The method promises considerable advantages over alternative Markov Chain Monte Carlo methods. We have developed an EM algorithm and a gradient descent procedure. We have additionally pointed out, that in low dimensional problems, solution of the unnormalised density filtering equation can be used to compute the likelihood and presented a new heuristic derivation of the update for continuous discrete filtering.

References

[1] D.L. Snyder. *Random Point Processes*. J. Wiley, New York, 1975.

[2] R. K. Boel and V. E. Benes. Recursive nonlinear estimation of a diffusion acting as the rate of an observed Poisson process. *IEEE Trans. Inf. Thy.*, 26:561–575, 1980.

[3] F. Rieke, D. Warland, R. de Ruyter van Stvenink, and W. Bialek. *Spikes: Exploring the neural code*. MIT Press, Boston, 1997.

[4] E. N. Brown, L. M. Frank, D. Tang, M. C. Quirk, and M. A. Wilson. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Jl Neuroscience*, 18:7411–7425, 1998.

[5] K. S. Chan and J. Ledolter. Monte Carlo EM estimation for time series models involving counts. *Jl. Amer. Stat. Assoc*, 90:242–252, 1995.

[6] V Solo. 'Unobserved' Monte Carlo method for system identification of partially observed nonlinear state space systems, Part I: Analog observations. Technical report, School of Elec. Eng., Univ. New South Wales, February 2000.

[7] N. T. Twum-Danso. Estimation, information and neural signals. Technical Report Unpublished PhD Thesis, Harvard University, Division of Applied Sciences, 1997.

[8] C. D. Charalambos and A. Logothetis. Maximum likelihood parameter estimation from incomplete data via the sensitivity equations: The continuous-time case. *IEEE. Trans. Autom . Contr.*, 2000.

[9] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, 39:p.1–38, 1977.

[10] V. Solo. 'Unobserved' Monte Carlo method for identiifcation of partially observed nonlinear state space systems, part I: Analog observations. Technical report, School of Elec. Eng., Univ. New South Wales, February 2000.

[11] W. P. Malcolm, M. R. James, and R. J. Elliott. Risk sensitive filtering with Poisson process observations. *Jl Appl Math Opt*, to appear, 1999.

[12] T. P. McGarty. *Stochastic Systems and State Estimation.* J Wiley, New York, 1974.

[13] G. D. Smith. *Numerical Solution of Partial Differential Equations.* Oxford University Press, Oxford, UK, 1974.

[14] A. T. Bharucha Reid. *Elements of the theory of Markov Processes.* McGraw Hill, New York, 1960.