OPTIMAL FILTERING FOR PARTIALLY OBSERVED POINT PROCESSES USING TRANS-DIMENSIONAL SEQUENTIAL MONTE CARLO

Arnaud Doucet†, Luis Montesano‡ and Ajay Jasra★

†Departments of Computer Science and Statistics, University of British Columbia, CA.
‡Department of Computer Science and Systems Engineering, University of Zaragoza, Spain.
★Department of Mathematics, Imperial College, UK.
Email: arnaud@stat.ubc.ca, montesano@unizar.es, ajay.jasra@imperial.ac.uk

ABSTRACT

Continuous-time marked point processes appear in many areas of science and engineering including queuing theory, seismology, neuroscience and finance. In numerous applications, these point processes are unobserved but actually drive an observation process. Here, we are interested in optimal sequential Bayesian estimation of such partially observed point processes. This class of filtering problems is non-standard as there is typically no underlying Markov structure and the likelihood function relating the observations to the point process has a complex form. Hence, except in very specific cases it is impossible to solve them in closed-form. We develop an original trans-dimensional Sequential Monte Carlo method to address this class of problems. An application to partially observed queues is presented.

1. INTRODUCTION

Point processes have found many applications in many areas of science including queuing theory [2], nuclear science [1] [17], seismology [5] and finance [16]. When the point process is directly observed, there has been much work on developing efficient statistical methods to perform inference. However, in many realistic applications the point process of interest is unobserved. Estimation of such partially observed point processes is much more complicated. In particular, in this case "the number of unknowns" (the number of points) "is something you don't know" [13]. In a Bayesian framework, analysis of such problems has become possible thanks to the introduction of trans-dimensional Markov chain Monte Carlo (MCMC) methods [13]; e.g. [1]. Unfortunately, these iterative batch methods are not adequate when massive datasets and/or real-time constraints are present.

We present here an original simulation-based approach for online Bayesian inference in partially observed point processes. Our methodology relies on Sequential Monte Carlo (SMC) methods in which the sequence of posterior distributions of interest is approximated by a cloud of random samples termed particles which evolve over time using sampling and resampling mechanisms. SMC is now routinely used to solve optimal filtering problems in non-linear non-Gaussian state-space models [10]. However, we emphasize that the filtering problems addressed here are of a completely different nature and cannot be solved using standard SMC methods. In this article, we develop an efficient trans-dimensional SMC method to achieve this. Simple trans-dimensional algorithms have appeared recently either implicitly [3] [6] [15] or explicitly [4] [18] in the literature. Here we make use of the fact that trans-dimensional SMC methods are actually a subset of SMC methods developed recently for static inference. Following the general framework detailed in [7] [8], we propose a principled way to design efficient sampling strategies within the trans-dimensional SMC framework. Our methodology allows the user to design and combine complex moves in an optimal way.

We demonstrate our methodology on optimal filtering for queues. The problem consists of reconstructing the state of a queue (number of customers waiting, number of dropouts etc.) based uniquely on departure data. Under restrictive assumptions on the interarrival and service times distributions, this problem admits a closed-form solution whose complexity grows cubically with time [12]. However as soon as these assumptions are relaxed, there is no longer an analytic solution [14]. Our methodology allows us to address realistic queuing models including general arrival and service times distributions, random dropouts and queues of finite capacity.

The paper is organized as follows: In Section II we formalize the optimal filtering problem. In Section III we describe the SMC sampler methodology and its applications to the trans-dimensional case. In Section IV we describe the queuing application.

2. OPTIMAL FILTERING FOR PARTIALLY OBSERVED POINT PROCESSES

2.1. Point Process Models

We consider the following marked point process

$$M(d\tau, du) = \sum_{k\geq 1} \delta_{\tau_k, U_k}(d\tau, du)$$

where $\{\tau_k\}_{k\geq 1}$ with $\tau_k > \tau_{k-1} \geq 0$ are the random occurrence times and $\{U_k\}_{k\geq 1}$ the associated random marks with U_k taking values in $\mathcal{U}$.

The process $\{\tau_k\}_{k\geq 1}$ admits the following joint probability density

$$f_{\tau_{1:k}}(\tau_{1:k}) = f_{\tau_1}(\tau_1) \prod_{j=2}^{n} f_{\tau_j|\tau_{1:j-1}}(\tau_j | \tau_{1:j-1}).$$

For example, if the arrival process $\{\tau_k\}_{k\geq 1}$ is a Poisson process of intensity λ then

$$f_{\tau_j|\tau_{1:j-1}}(\tau_j | \tau_{1:j-1}) = \lambda \exp(-\lambda(\tau_j - \tau_{j-1})).$$

For the sake of simplicity we will assume that the marks $\{U_k\}_{k\geq 1}$ are independent of $\{\tau_k\}_{k\geq 1}$ and independent and identically distributed according to a density $f_U(\cdot)$. However, it is possible to consider more complex models if necessary.

2.2. Likelihood Function

The process $M(d\tau, du)$ is unobserved and we only have access at time t to a (possibly random) number of observations denoted Y_t. We assume here that the likelihood function satisfies

$$p\left(Y_t \middle| \{\tau_k, u_k\}_{k\geq 1}\right) = p\left(Y_t \middle| \tau_{1:k_t}, u_{1:k_t}\right) \tag{1}$$

where K_t is an integer-valued stopping time satisfying $K_{t'} \geq K_t$ (almost surely) for $t' > t$. Broadly speaking a stopping time is a random variable determined entirely by Y_t and $\{\tau_k, U_k\}_{k\geq 1}$. The stopping time definition depends on the problem of interest.

Example. Noisy Shot Noise Process. Consider the discrete time observations $\{Y(t_n)\}_{n\geq 1}$

$$Y(t_n) = \int uh(t_n - \tau) M(d\tau, du) + b(t_n)$$

where $b(t_n) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ and $h(t) = 0$ for $t < 0$. In this case we have $Y_t = \{Y(t_n) : t_n \leq t\}$ and

$$K_t = \arg\max_k \{k : \tau_k \leq \arg\max\{t_n : t_n \leq t\}\}.$$

Example. Randomly Delayed Process. Consider the observation process

$$Y(dt) = \sum_{k\geq 1} \delta_{\tau_k + U_k}(dt)$$

where $\{U_k\}_{k\geq 1}$ are positive random variables then

$$K_t = \arg\max_k \{k : \tau_k + U_k \leq t\}.$$

2.3. Optimal Filtering

We are interested in the optimal estimation of $M(d\tau, du)$ given the observations available at any time t. Hence the posterior distribution of interest is

$$p(\{\tau_k, u_k\}_{k\geq 1} | Y_t) \propto p(Y_t | \{\tau_k, u_k\}_{k\geq 1}) p(\{\tau_k, u_k\}_{k\geq 1}).$$

Clearly it is impossible to estimate this posterior distribution as it includes an infinite number of variables. Because of (1), it is more sensible to restrict ourselves to the estimation of $p\left(\tau_{1:k_t}, u_{1:k_t} \middle| Y_t\right)$. The variable K_t being an integer-valued stopping time, the distribution $p\left(\tau_{1:k_t}, u_{1:k_t} \middle| Y_t\right)$ is defined on a space of the form $\biguplus_k \mathbb{D}_k \times \mathcal{U}^k$ where $\mathbb{D}_k = \{\tau_{1:k} : 0 \leq \tau_1 < \tau_k < \cdots < \tau_k\}$. Typically, this distribution does not admit a closed-form. If t was fixed, it would be possible to approximate this distribution using trans-dimensional MCMC [13]. However, we are here interested in estimating a sequence of posterior distributions $\left\{p\left(\tau_{1:k_{t_n}}, u_{1:k_{t_n}} \middle| Y_{t_n}\right)\right\}$ where $t_{n-1} < t_n$.

3. TRANS-DIMENSIONAL SEQUENTIAL MONTE CARLO

Standard SMC methods are algorithms designed to sample from a sequence of distributions $\{\pi_n\}_{n\geq 1}$ where π_n is defined on E_n and $E_n = E_{n-1} \times F_n$; i.e. the spaces are of "increasing dimension". Clearly these methods are inadequate in the point process case where this condition is not satisfied. Indeed in the point process case all distributions are typically defined on the same space $E = \biguplus_{k=1}^{\infty} \mathbb{D}_k \times \mathcal{U}^k$ but have positive masses only on a subset S_n of E with $S_{n-1} \subseteq S_n$. For example in the shot noise process example described earlier the posterior distribution $p\left(\{\tau_k, u_k\}_{1\leq k\leq k_{t_n}} \middle| Y_{t_n}\right)$ is defined on E but its support is restricted to $S_n = \biguplus_{k=1}^{\infty} \mathbb{D}_{k,t_n} \times \mathcal{U}^k$ where

$$\mathbb{D}_{k,t_n} = \{\tau_{1:k} : 0 \leq \tau_1 < \tau_k < \cdots < \tau_k \leq t_n\}.$$

3.1. SMC Samplers

We use interchangeably measures and densities in this section. SMC samplers are a generalization of SMC methods introduced in [7], [8]. Essentially if we consider two successive target distributions $\pi_{n-1}(x_{n-1})$ and $\pi_n(x_n)$ defined respectively on two arbitrary (measurable) spaces E_{n-1} and E_n then SMC samplers methods allow us to define valid moves between these spaces. We introduce a generalized Markov kernel $T_n : E_{n-1} \to \mathcal{B}(E_n)$ (where $\mathcal{B}(E_n)$ is a sigma-algebra). Assume $X_{n-1} \sim \pi_{n-1}$ and $X_n | X_{n-1} \sim T_n(X_{n-1}, \cdot)$ then the marginal distribution of X_n is given by

$$\mu_n(dx_n) = \int_{E_{n-1}} \pi_{n-1}(dx_{n-1}) T_n(x_{n-1}, dx_n).$$

When $E_n = E_{n-1} \times F_n$ the Markov kernel T_n is usually built by introducing $q_n : E_{n-1} \to \mathcal{B}(F_n)$ and setting

$$T_n(x_{n-1}, dx_n) = q_n(x_{n-1}, dv_n) \delta_{(x_{n-1}, v_n)}(dx_n).$$

In this case, it is possible to compute $\mu_n(dx_n)$ pointwise (up to a normalizing constant) and then reweight the resulting particles with respect to the target distribution of interest using

$$w_n(x_n) = \frac{\pi_n(x_n)}{\mu_n(x_n)} = \frac{\pi_n((x_{n-1}, v_n))}{\pi_{n-1}(x_{n-1}) q_n(x_{n-1}, v_n)}. \tag{2}$$

However in the general case it is impossible to compute $\mu_n(x_n)$ pointwise and hence to compute (2). In [7], [8], it is shown how to weight these particles consistently with respect to $\pi_n(x_n)$ without having to compute $\mu_n(x_n)$. To this aim we introduce another Markov kernel $L_{n-1} : E_n \to \mathcal{B}(E_{n-1})$ and we define the new incremental weight

$$w_n(x_{n-1}, x_n) = \frac{\pi_n(x_n) L_{n-1}(x_n, x_{n-1})}{\pi_{n-1}(x_{n-1}) T_n(x_{n-1}, x_n)}; \tag{3}$$

i.e. we are performing importance sampling with a new artificial target distribution $\pi_n(x_n) L_{n-1}(x_n, x_{n-1})$ which admits $\pi_n(x_n)$ as a marginal by construction.

Clearly the performance of this method is going to be highly dependent on the choice of L_{n-1}. In [7], [8] it is established that the optimal choice for L_{n-1} (with respect to the variance of the weights) is given by

$$L_{n-1}^{\text{opt}}(x_{n-1}, x_n) = \frac{\pi_{n-1}(x_n) T_n(x_n, x_{n-1})}{\mu_n(x_{n-1})}. \tag{4}$$

This result is actually intuitive as in this case (3) is equal to (2). Clearly we cannot use L_{n-1}^{opt} if in practice $\mu_n(x_n)$ cannot be evaluated pointwise but it suggests that to obtain good performance we should select L_{n-1} as an approximation of L_{n-1}^{opt}. The key point is that even if $L_{n-1} \neq L_{n-1}^{\text{opt}}$ then the algorithm remains theoretically valid. The price to pay for using some L_{n-1} different to L_{n-1}^{opt} is an increase in the variance of the weights.

When T_n is an MCMC kernel of invariant distribution π_n, an approximation of L_{n-1}^{opt} which will prove very useful in applications is

$$L_{n-1}(x_{n-1}, x_n) = \frac{\pi_n(x_n) T_n(x_n, x_{n-1})}{\pi_n(x_{n-1})} \tag{5}$$

and in this case

$$w_n(x_{n-1}, x_n) = \frac{\pi_n(x_{n-1})}{\pi_{n-1}(x_{n-1})}.$$

In practice, we are typically interested in using not one single move but a combination of p moves described by $\{T_{n,m}\}$ and selected with probability $\{\alpha_{n,m}(x_{n-1})\}$ where $m = 1, ..., p$. The total kernel is given by

$$T_n(x_{n-1}, x_n) = \sum_{m=1}^{p} \alpha_{n,m}(x_{n-1}) T_{n,m}(x_{n-1}, x_n).$$

In this case, the reverse artificial kernel should be selected as follows

$$L_{n-1}(x_{n-1}, x_n) = \sum_{m=1}^{p} \beta_{n-1,m}(x_{n-1}) L_{n-1,m}(x_{n-1}, x_n)$$

where $\{L_{n,m}\}$ are Markov kernels and $\{\beta_{n,m}(x_{n-1})\}$ are probabilities. The importance weight can be computed using (3) but this is prohibitive if p is large and in this case it can be shown that the following expression is also valid

$$w_n(x_{n-1}, x_n, m_n) = \frac{\pi_n(x_n) \beta_{n-1,m_n}(x_n) L_{n-1,m_n}(x_n, x_{n-1})}{\pi_{n-1}(x_{n-1}) \alpha_{n,m_n}(x_{n-1}) T_{n,m_n}(x_{n-1}, x_n)}$$

when at time n the move $m_n \in \{1, ..., p\}$ has been selected. In this case, the expressions for the optimal kernels $\{L_{n-1,m_n}\}$ and probabilities $\{\beta_{n-1,m_n}\}$ minimizing the variance of $w_n(x_{n-1}, x_n, m_n)$ can also be established.

3.2. Trans-dimensional SMC Samplers

We now consider here the trans-dimensional problem appearing for point processes; i.e. all distributions $p\left(\tau_{1:k_{t_n}}, u_{1:k_{t_n}} \middle| Y_{t_n}\right)$ are defined on $E = \biguplus_{k=1}^{\infty} \mathbb{D}_k \times \mathcal{U}^k$ but their supports S_n are growing over time; i.e. $S_{n-1} \subseteq S_n$. We propose a combination of various generic moves and for ease of presentation we assume here that $\alpha_{n,m_n}(x_{n-1}) = \alpha_{n,m_n}$ so that the optimal kernel is equal to (4). To simplify notation we write $\vartheta_{1:k_{t_n}} = \left(\tau_{1:k_{t_n}}, u_{1:k_{t_n}}\right)$. We emphasize that the applicability of these moves depends on the problem under study.

Update move. This move consists of using

$$T_{n,update}\left(\vartheta_{1:k_{t_{n-1}}}, d\vartheta'_{1:k_{t_n}}\right) = \delta_{\vartheta_{1:k_{t_{n-1}}}}\left(d\vartheta'_{1:k_{t_n}}\right).$$

This move cannot be used alone if $S_{n-1} \subset S_n$ as the support of the resulting importance distribution would not include the support of the target.

Optimal updating moves. In light of new observations, the general framework proposed earlier allows us to update say the last L points

$$\begin{aligned} &T_{n,optupdate}\left(\vartheta_{1:k_{t_{n-1}}} d\vartheta'_{1:k_{t_n}}\right) \\ &= \delta_{\vartheta_{1:k_{t_{n-1}}-L}}\left(d\vartheta'_{1:k_{t_n}-L}\right) q_n\left(\vartheta_{1:k_{t_{n-1}}-L}, d\vartheta'_{k_{t_n}-L+1:k_{t_n}}\right). \end{aligned}$$

The optimal importance distribution q_n minimizing the conditional variance of the importance weights is given by

$$p\left(\vartheta'_{k_{t_n}-L+1:k_{t_n}} \middle| Y_{t_n}, \vartheta_{1:k_{t_{n-1}}-L}\right).$$

We cannot typically sample from it but we can approximate it by say $\widehat{p}$. In this case, it is also possible to approximate (4) and the weight (3) is given by

$$\begin{aligned} &\frac{p\left(\vartheta_{1:k_{t_n}-L}, \vartheta'_{k_{t_{n-1}}-L+1:k_{t_n}} \middle| Y_{t_n}, \vartheta_{1:k_{t_{n-1}}-L}\right)}{p\left(\vartheta_{1:k_{t_{n-1}}} \middle| Y_{t_{n-1}}\right)} \\ &\times \frac{\widehat{p}\left(\vartheta_{k_{t_{n-1}}-L+1:k_{t_n}-1} \middle| Y_{t_{n-1}}, \vartheta_{1:k_{t_{n-1}}-L}\right)}{\widehat{p}\left(\vartheta'_{k_{t_n}-L+1:k_{t_n}} \middle| Y_{t_n}, \vartheta_{1:k_{t_{n-1}}-L}\right)}. \end{aligned}$$

This strategy is very useful when the discrepancy between successive distributions is high [11].

Local moves. In this case we perturb a parameter by some increments on a discrete grid; say if we want to modify $\vartheta_{k_{t_{n-1}}}$

$$\begin{aligned} &T_{n,local}\left(\vartheta_{1:k_{t_{n-1}}}, d\vartheta'_{1:k_{t_n}}\right) = \\ &\delta_{\vartheta_{1:k_{t_{n-1}}}}\left(d\vartheta'_{1:k_{t_{n-1}}}\right) \cdot \tfrac{1}{M} \sum_{i=1}^{M} \delta_{\vartheta_{k_{t_{n-1}}}+\gamma_i}\left(d\vartheta'_{k_{t_n}}\right). \end{aligned}$$

In this case it is possible to compute (2) [hence (4)].

Birth move. In a birth move, we propose to add one or several points to the current state. For ease of presentation, we suppose here that we just add some points at the end of the current path when $k_{t_{n-1}} < k_{t_n}$. We have

$$\begin{aligned} &T_{n,birth}\left(\vartheta_{1:k_{t_{n-1}}}, d\vartheta'_{1:k_{t_n}}\right) = \\ &\delta_{\vartheta_{1:k_{t_{n-1}}}}\left(d\vartheta'_{1:k_{t_{n-1}}}\right) q_n\left(\vartheta_{1:k_{t_{n-1}}}, d\vartheta'_{k_{t_{n-1}}+1:k_{t_n}}\right). \end{aligned}$$

In this case it is also possible to compute (2) [hence (4)]. The proposal q_n minimizing the conditional variance of the weight is given by

$$p\left(\vartheta'_{k_{t_{n-1}}+1:k_{t_n}} \middle| Y_{t_n}, \vartheta_{1:k_{t_{n-1}}}\right).$$

Death move. Assume one wants to remove one point; say the last one we added for ease of presentation then

$$T_{n,death}\left(\vartheta_{1:k_{t_{n-1}}}, d\vartheta'_{1:k_{t_n}}\right) = \delta_{\vartheta_{1:k_{t_{n-1}}-1}}\left(d\vartheta'_{1:k_{t_n}}\right).$$

In this case, it is usually impossible to compute (2) because

$$\mu_n\left(\vartheta'_{1:k_{t_n}}\right) = \int p\left(\vartheta'_{1:k_{t_{n-1}}} \middle| Y_{t_{n-1}}\right) d\vartheta'_{k_{t_{n-1}}}$$

does not admit a closed-form expression. However dependent on applications it might be possible to come up with sensible approximations of (4).

Split and Merge moves. Similarly to trans-dimensional MCMC, it is possible to develop split and merge moves. However contrary to MCMC we emphasize that it is not necessary to design reversible moves. Moreover although one might be tempted to mimic MCMC proposals this will usually yield unbounded importance weights. It is recommended that the user always follows (4).

Trans-dimensional MCMC moves. It is always possible to use a kernel $T_{n,MCMC}$ of invariant distribution $p\left(\vartheta'_{1:k_{t_n}} \middle| Y_{t_n}\right)$; that is a trans-dimensional MCMC kernel [13]. However, it has to be applied after other moves have been used and the particles resampled so that they are approximately distributed according to $p\left(\vartheta'_{1:k_{t_n}} \middle| Y_{t_n}\right)$. If it is not the case, the problem is that it is very difficult to approximate (4) and (5) is not valid in this case.

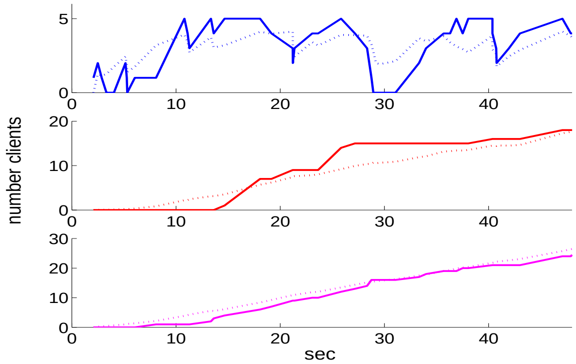

Fig. 1. The figure shows the number of customers in the queue (top), number of customers not accepted due to full buffer (middle) and number of customers that dropped the queue (bottom). Solid lines are the true number and dotted ones are the conditional expectations according to the estimated distributions. We use $N = 500$ particles, $\lambda = 2$, $\mu = 1$, $\omega = 0.2$ and $C = 5$. The results obtained are similar for a higher number of particles.

4. APPLICATION TO PARTIALLY OBSERVED QUEUES

We consider the following problem. Some customers are arriving to a queue at random times $\{\tau_k\}_{k\geq 1}$. We assume that the interarrival times are independent and identically distributed according $f_{\Delta\tau}(\cdot)$, that is

$$f_{\tau_j|\tau_{1:j-1}}(\tau_j|\,\tau_{1:j-1}) = f_{\Delta\tau}(\tau_j - \tau_{j-1}).$$

Customers are served on a first in first out basis and the service times $\{S_k\}_{k\geq 1}$ are independent and identically distributed according to $f_S(\cdot)$. However, the queue has a finite capacity of C clients. Finally, if the k^{th} customer is in the queue but has been waiting to be served longer than U_k then it drops out of the queue. Here $\{U_k\}_{k\geq 1}$ are independent and identically distributed according to $f_U(\cdot)$

We only have access to the departure times of customers who have been served; that is the observation process is also a point process

$$Y(dt) = \sum_{k\geq 1} \delta_{T_k}(dt)$$

where the departures times $\{T_k\}_{k\geq 1}$ are a deterministic function of $\{\tau_k\}_{k\geq 1}$, $\{S_k\}_{k\geq 1}$ and $\{U_k\}_{k\geq 1}$. The stopping time K_t corresponds to the index of the customer associated with the last departure observed before or at time t.

Given a realization of the observations our aim is to estimate at any time the posterior distributions $\{p(\tau_{1:k_{t_n}}, u_{1:k_{t_n}}|\,Y_{t_n})\}$. This problem is clearly a trans-dimensional problem as the nth observed departure at time t_n does not necessarily correspond to the nth customer who came to the queue because of the finite capacity of the queue and the random dropouts. Based on these distributions, it is possible to estimate the number of customers who have dropped out of the queue, the number of customers currently in the queue etc.

Our filtering algorithm relies on the trans-dimensional methodology and combines birth moves, update moves and trans-dimensional MCMC moves. Figure 1 presents the results obtained in the case where $f_{\Delta\tau}$, f_S and $f_U(\cdot)$ are exponential distributions of parameters λ, μ and ω. Note that each path has a different stopping time. In order to compare the results with the real data, we use the distributions of the arrival, departure and waiting times to simulate the paths forward until time t_n.

Finally, we note that the trans-dimensional SMC approach allows us to compute the (marginal) likelihood of the observations. Hence this procedure could be extended to obtain the maximum likelihood estimates of the parameters of the inter-arrival, service and dropout distributions.

5. REFERENCES

[1] C. Andrieu, E. Barat and A. Doucet, “Bayesian deconvolution of noisy filtered point processes”, *IEEE Trans. Signal Processing*, vol. 49, no. 1, 134-146, 2001.

[2] P. Bremaud, *Point Processes and Queues*, New York: Springer-Verlag, 1980.

[3] Y. Chen, J. Xie and J.S. Liu, “Stopping-time resampling for sequential Monte Carlo”, *J.R. Statist. Soc.* B, 67, 199-217, 2005.

[4] M. Davy, P. Del Moral, and A. Doucet, “Méthodes Monte Carlo séquentielles pour l’analyse spectrale Bayésienne”, 19ème Colloque GRETSI, 2003.

[5] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, New York: Springer-Verlag, 1988.

[6] P. Del Moral, *Feynman-Kac Formulae. Genealogical and Interacting Particle Approximations*, New York: Springer-Verlag, 2004.

[7] P. Del Moral, A. Doucet and A. Jasra, “Sequential Monte Carlo samplers”, Technical report TR-CUED-F-INFENF 343, First version December 2002, Revised June 2004 and October 2005.

[8] P. Del Moral, A. Doucet and A. Jasra, “Sequential Monte Carlo methods for Bayesian computation” (with discussion), in *Bayesian Statistics 8*, Oxford University Press, to appear 2006.

[9] A. Doucet A., S.J. Godsill and C. Andrieu, “On sequential Monte Carlo sampling methods for Bayesian filtering”, *Statist. Comput.*, vol. 10, 197-208, 2000.

[10] A. Doucet, J.F.G. de Freitas and N.J. Gordon (eds.), *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001.

[11] A. Doucet, M. Briers and S. Sénécal, “Efficient sampling strategies for sequential Monte Carlo”, *J. Comp. Graph. Stat.*, to appear 2006.

[12] P. Fearnhead, “Filtering recursion for calculating likelihoods for queues based on inter-departure data”, *Statist. Comput.*, vol. 14, pp. 261-266, 2004.

[13] P.J. Green, “Trans-dimensional Markov chain Monte Carlo”, in *Highly Structured Stochastic Systems*, Oxford University Press, 2003.

[14] K. Heggland and A. Frigessi, “Estimating functions in indirect inference”, *J.R. Statist. Soc.* B, 66, 447-462, 2004.

[15] S. Maskell, “Joint tracking of manoeuvering targets and classification of their manoeuvrability”, *J. App. Sig. Proc.*, 15, 2339-2350, 2004.

[16] G.O. Roberts, O. Papaspiliopoulos and P. Dellaportas, “Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility models”, *J. R. Statist. Soc.* B, 66, 2, 369-393, 2004.

[17] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*, Springer-Verlag, 1991.

[18] J. Vermaak, S.J. Godsill and A. Doucet, “Radial basis function regression using trans-dimensional sequential Monte Carlo”, *Proc. IEEE Workshop. Statistical Signal Processing*, 2003.