Particle Filters for State and Parameter Estimation of Cox Process

Omiros Papaspiliopoulos[1], Miguel A. G. Belmonte[2], and Michael K. Pitt[2]

[1] Department of Economics, Universitat Pompeu Fabra,
Ramon Trias Fargas, 25-27
08005 Barcelona,
Spain
(e-mail: omiros.papaspiliopoulos@upf.edu)

[2] Department of Statistics and Department of Economics, University of Warwick,
Coventry, CV4 7AL, UK
(e-mail: m.a.g.belmonte@warwick.ac.uk, m.pitt@warwick.ac.uk)

Abstract. We consider inference for Cox processes in time where the intensity is a parametrised function of an unobserved diffusion process. We design a particle filter for the on-line estimation of the intensity function. The filter relies on a time-discretisation in order to approximate the conditional density of the data given the signal. We apply the smooth particle filter methodology of Pitt (2002) to obtain maximum likelihood estimates of unknown parameters. We find that the discretisation size has some moderate impact on the estimation of the state, when parameters are assumed known, but big impact on the estimation of parameters. Furthermore, we discover potential parameter identifiability problems. The paper is concluded with currently investigated extensions.
Keywords. smooth particle filter, sequential Monte Carlo, diffusion process, duration data

1 Introduction

Cox processes (e.g. Møller and Waagepetersen (2004)) are a very popular modelling tool for clustered point patterns. The observed data are generated by an inhomogeneous Poisson process with intensity function $\nu(\cdot)$, which is random positive function. The emphasis in on specifying computationally convenient but flexible enough models for $\nu(\cdot)$. Typically inference for the intensity and unknown parameters is carried out by Markov chain Monte Carlo algorithms. Such algorithms can be costly in computing time and difficult to design efficiently to result in reasonable Monte Carlo error in the estimates.

The focus of this paper is on point processes which evolve in time for modelling duration-type data. There is a lot of interest in such data structures, e.g. in financial econometrics for analyzing credit defaults (Schönbucher (2003)) and the inter-arrival times (so called durations) of financial transactions in a high frequency regime (Engle and Russell (1998)). In this context we require a modelling and computational framework which can manage well with long time-series, as for instance one can have hundreds of durations per day, and can be extended to model the joint evolution of several dependent point processes. Hence we choose to model the intensity as a transformation of a latent Markov process X with state space $\mathcal{X}$, achieved by an appropriate link function $\nu : \mathcal{X} \to R_+$. Due to the Markov assumption on X we can express the joint evolution of observed and latent process in a continuous-time state-space form, and resort to sequential Monte Carlo methods for the on-line estimation of the intensity. This approach has the computational advantage of managing very well with long time series. Additionally, it leads immediately to on-line predictions and it facilitates the use of certain model diagnostics which can be applied to assess model assumptions. Further we estimate unknown parameter using simulated maximum likelihood resorting to the so-called smooth particle filter methodology of Pitt (2002).

We restrict attention to Markov processes which are solutions to stochastic differential equations:

$$\mathrm{d}X_s = b(X_s)\mathrm{d}s + \sigma(X_s)\mathrm{d}B_s\,, \quad s \geq 0 \tag{1}$$

where B is a standard Brownian motion, b is called the drift and σ the diffusion coefficient. Monte Carlo methodology for such Cox processes was recently considered in Fearnhead et al. (2006). This framework naturally extends in continuous-time existing latent autoregressive models for financial duration data (Bauwens and Veredas (2004)), and it is amenable to multivariate extensions (by correlating the corresponding Brownian motions).

2 A discretised particle filter

Throughout this paper we will denote by upper case letter random variables and by lower case their realisations. Let $\{y_i\}_{i=0}^n$ denote the observed arrivals, where by convention we set $y_0 = 0$. The conditional density of each arrival given the previous and the path is

$$\nu(x_{y_i}) \exp\left\{-\int_{y_{i-1}}^{y_i} \nu(x_s)\,\mathrm{d}s\right\}, \tag{2}$$

where the second term accounts for the absence of arrivals on (y_{i-1}, y_i) and the first for the arrival at y_i. Hence, the conditional likelihood given only the value of x at the arrival times is

$$p(y_i \mid y_{i-1}, x_{y_{i-1}}, x_{y_i}) := \nu(x_{y_i}) E\left[\exp\left\{-\int_{y_{i-1}}^{y_i} \nu(X_s)\mathrm{d}s\right\}\right], \tag{3}$$

where the expectation is w.r.t. the conditional law of X given the endpoints at times y_{i-1}, y_i, i.e. w.r.t. the diffusion bridge law.

Our aim is to recursively calculate the filtering densities, that is the posterior densities $\pi_i(x_{y_i}) := p(x_{y_i} \mid y_{0:i})$, where $y_{0:i} = (y_0, \ldots, y_i)$, where π_0 is a prior density chosen to reflect our prior belief for X_0 (typically chosen to have large support). Note that the normalising constant of this density is $p(y_i \mid y_{0:i})$. Analytic calculation of the π_i's is typically impossible. The standard Monte Carlo solution is to approximate $\pi_i(x_{y_i})$ by a discrete distribution whose support is a set of N particles, $\{x_{y_i}^{(j)}\}_{j=1}^{N}$, with associated probability weight $\{w_i^{(j)}\}_{j=1}^{N}$. Such approximation is propagated to the next time point via Bayes' theorem to yield

$$\tilde{\pi}_{i+1}(x_{y_{i+1}}) \propto \sum_{j=1}^{N} w_i^{(j)} p(y_{i+1}|y_i, x_{y_i}^{(j)}, x_{y_{i+1}}) p_{y_{i+1}-y_i}(x_{y_{i+1}}|x_{y_i}^{(j)}), \tag{4}$$

where $p_t(z \mid x)$ denotes the transition density of the diffusion. A particle filter consists of the sequential application of these approximations, together with an importance sampling mechanism for sampling from (4). To avoid explosion of the variance, one needs to impose interaction among particles by occasionally re-sampling them with probabilities proportional to their weights. This would replicate particles with high weight and eliminate others with negligible weight. We use the stratified re-sampling algorithm of Carpenter et. al (1999). The error of the approximation induced by the particle filter is of Monte Carlo type and in standard cases it is reduced as $1/\sqrt{N}$.

Here we consider sampling from (4) by propagating each particle $x_{y_i}^{(j)}$ to the next time point according to the system transition density, $p_{y_{i+1}-y_i}(\cdot|x_{y_i}^{(j)})$, in which case the weight of each new particle, $x_{y_{i+1}}^{(j)}$, is $w_i^{(j)} p(y_{i+1}|y_i, x_{y_i}^{(j)}, x_{y_{i+1}})$. This scheme works well when the conditional likelihood is not very peaked relative to the transition density (which is the case in our problem).

There are two main challenges for applying the particle filter in the family of Cox processes we consider. First, simulating from the system transition density can be difficult since the latter is typically intractable. Second, the conditional likelihood involved in the weights is intractable, since it requires analytic computation of the expectation in (3), which for most processes X will be impossible. A solution for both of these issues, when b and σ in (1) satisfy certain conditions was provided in Fearnhead et. al (2008). Here we take a more basic approach and augment the observed data with additional observations in-between arrival times. We define a parameter $\delta > 0$, which is the maximum distance in time between two observations and discretise time accordingly. Hence we obtain data at times $0 = t_0 < t_1 < \ldots < t_M = y_n$, which is a superset of $\{y_i\}_{i=0}^{n}$, and construct a data set $\{z_m\}_{m=1}^{M}$, where $z_m = 1$ if $t_m = y_i$ for some i, and 0 otherwise. Then we define

$$P[Z_m = 0 \mid z_{m-1}, x_{t_{m-1}}, x_{t_m}] = \exp\{-\nu(x_{t_{m-1}})(t_m - t_{m-1})\}\,. \tag{5}$$

In this setting the conditional likelihood depends only the previous state $x_{t_{m-1}}$. However, this is only one of the possible discretisations of the system. Hence, in the new filter between time points t_m and t_{m+1} we propagate the particles according to the system dynamics and multiply their existing weights by the conditional likelihood in (5). Simulation from the transition of the diffusion process can be achieved by the Exact Algorithm (see Fearnhead et. al (2008)) or alternatively it can be done approximately by some discretization scheme. For fixed N, we recover the exact particle filter approximation at the arrival times as $\delta \to 0$. Convergence to the filtering distributions as $\delta \to 0$ and $N \to \infty$ is a subtle matter, see Del Moral et. al (2001).

3 Parameter estimation

Let θ be a vector which contains all the unknown parameters which appear in the functions ν, b, σ. It is easy to check that the log-likelihood for a given θ can be estimated by

$$\ell(\theta; N, \delta) \approx \sum_{m=1}^{M} \log \left(\frac{1}{N} \sum_{j=1}^{N} w_m^{(j)} \right) ;$$

(the dependence in δ is implicitly in the model approximation). We are interested in exploring features of the likelihood surface such as the maximum, profile likelihoods, level sets etc. Hence we want a simulation scheme which yields a continuous (if not differentiable) map $\theta \to \ell(\theta; N, \delta)$, for given N and δ. Simulation of X as a smooth transformation of θ and random variables independent of θ is generally feasible for diffusions, and particularly easy when approximate Gaussian schemes are used for its dynamics. However, setting up the re-sampling scheme to have similar properties is hard, since small perturbations of the weights can cause big changes in the evolution of the filter and in the likelihood. The smooth particle filter of Pitt (2002) provides a novel solution to this problem which applies when X is a one-dimensional process. We make use of this technology to obtain maximum likelihood estimates and profile likelihoods.

To illustrate our results we design a small simulation experiment, when X is an Ornstein-Uhlenbeck process with drift $b(x) = -\rho(x - \mu)$, and constant diffusion coefficient σ, and $\nu(x) = e^x$; thus, $\theta = (\mu, \rho, \sigma^2)$. Exact simulation from the process transition density is simple in this case. Figure 1 shows state estimation when keeping parameters fixed at their true values. The smaller δ stabilizes the variance of the weights (we are re-sampling at every step of the algorithm). It is interesting that re-sampling more often (as δ gets smaller) does not deteriorate the filter, since it is done over increasingly smaller time horizons and there is data information at the imputed times (no arrival). With the smooth particle filter we estimate the parameters. Our results below correspond to the same parameter values as above but time series of 2000 arrivals (and 1000 particles). The effect of the discretisation is serious: for $\delta = 20$ and $\delta = 0.5$ we estimate $(-0.91, 0.44, 0.16)$ and $(-1.03, 0.31, 0.15)$

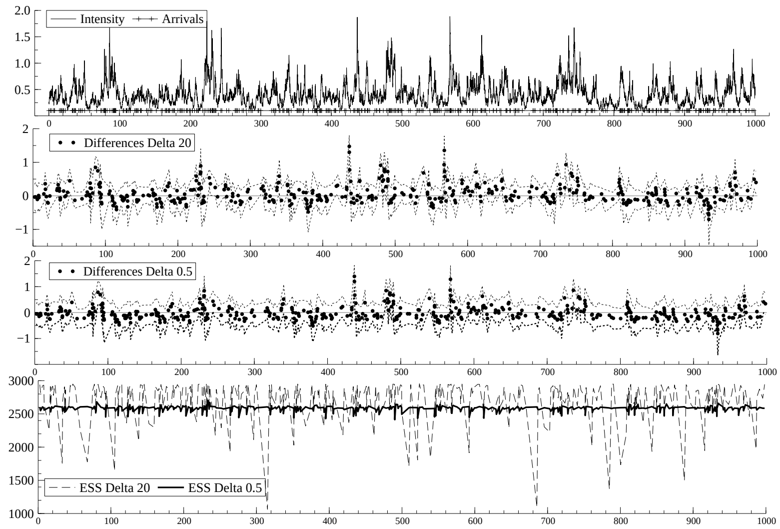

Fig. 1. Top: Intensity $\exp(x_s)$ plotted on a fine time-grid and simulated arrivals (denoted by "+"), where $\mu = -1, \rho = 0.3, \sigma^2 = 0.15$, $n = 455$, $y_n = 998.2$ and $N = 3000$. 2nd-3rd rows: difference between intensity and posterior means together with 95% symmetric confidence intervals. 4th row: Effective Sample Size approximated by $(\sum_{j=1}^{N} w_i^{(j)})^2 / \sum_{j=1}^{N} (w_i^{(j)})^2$.

respectively. Figure 2 (top) shows the profile likelihood for σ^2 under the two different discretisation levels. We have found that identification of both ρ and σ^2 is difficult unless we have very long time series. Figure 2 plots the profile likelihood of σ^2 and $\gamma := \sigma^2/2\rho$ for two different simulated datasets. It is clear that whereas the latter is quite well estimated the information about the former is weak and varies a lot with the dataset. However, experiments with datasets of 10,000 arrivals show convergence of the MLE to the true values.

4 Extensions

The results of our paper suggest that working with integrated quantities is desirable to avoid the bias in the parameter estimation. Hence we wish to extend the framework of Fearnhead et. al (2008) in the direction of parameter estimation. We wish to study the effect that approximation of the dynamics of the SDE has on the results. The asymptotic properties of the MLE in this class of models is also of interest. A class of model diagnostics can be applied on the output of the particle filter. Further, it is easy to consider heavier-tailed observation equations by addition of unstructured random effects. The framework lends itself to modelling dependent point processes, where state estimation should not pose any problem but parameter estimation will require careful thought.

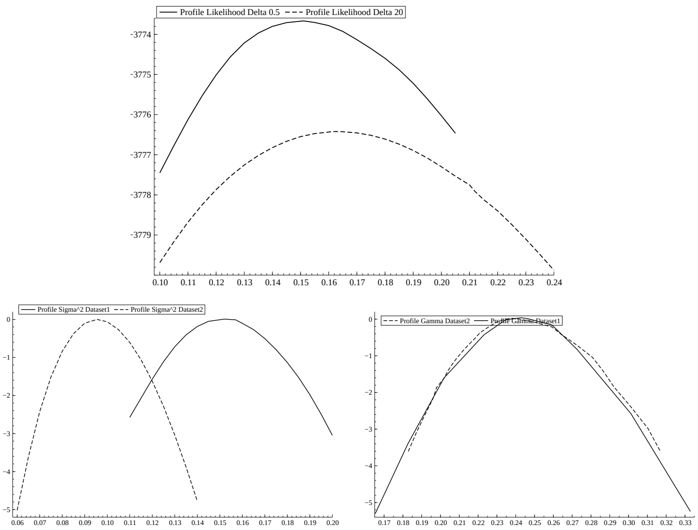

Fig. 2. Top: Profile log-likelihoods for σ^2 under two discretisation levels. Bottom: Profile log-likelihoods for σ^2 and $\gamma = \sigma^2/2\rho$, for two different datasets, with $\delta = 0.5$ in both cases.

References

Bauwens, L. and Veredas, D. (2004): The Stochastic Conditional Duration Model: A Latent Variable Model For The Analysis Of Financial Durations, *Journal of Econometrics*, *119-2*, 381–412.

Carpenter, J., Clifford, P. and Fearnhead, P. (1999): An improved particle filter for non-linear problems. *IEE proceedings-Radar, Sonar and Navigation*, **146**, 2–7.

Del Moral, P., Jacod, J. and Protter, P. (2001): The Monte-Carlo method for filtering with discrete-time observations. *Probab. Theory Related Fields*, *120*, 346–368.

Engle, R.F. and Russell, J.R. (1998): Autoregressive Conditional Duration: A New Model For Irregularly Spaced Transaction Data, *Econometrica*, *66*, 1127–1162.

Fearnhead, P., Papaspiliopoulos, O. and Roberts, G.O. (2008): Particle filters for partially observed diffusions. *J. Roy. Statist. Soc. Ser. B* to appear, available in `http://www.maths.lancs.ac.uk/~fearnhea/RWPF.pdf`

Møller, J. and Waagepetersen, R.P. (2004): Statistical inference and simulation for spatial point processes. Monographs on Statistics and Applied Probability, 100. Chapman & Hall/CRC, Boca Raton, FL,. xvi+300 pp. ISBN: 1-5

Pitt, M.K. (2002): Smooth Particle Filters For Likelihood Evaluation And Maximisation. Working paper, available in `http://www2.warwick.ac.uk/fac/soc/economics/staff/faculty/pitt/`

Schönbucher, P.J. (2003) *Credit Derivatives Pricing Models: Models, Pricing and Implementation.* John Wiley & Sons, Chichester.