

# The Ornstein-Uhlenbeck Dirichlet Process and other time-varying processes for Bayesian nonparametric inference

J.E. Griffin\*

Department of Statistics, University of Warwick, Coventry, CV4 7AL, U.K.

## Abstract

This paper introduces a new class of time-varying, measure-valued stochastic processes for Bayesian nonparametric inference. The class of priors generalizes the normalized random measure (Kingman 1975) construction for static problems. The unnormalized measure on any measurable set follows an Ornstein-Uhlenbeck process as described by Barndorff-Nielsen and Shephard (2001). Some properties of the normalized measure are investigated. A particle filter and MCMC schemes are described for inference. The methods are applied to an example in the modelling of financial data.

## 1 Introduction

Bayesian nonparametric models based on Dirichlet process mixtures have been popular for clustering and density estimation problems. Müller and Quintana (2004) give a comprehensive review of these ideas. The model assumes that the data  $y_1, \dots, y_n$  are an i.i.d. sample from a distribution with density  $f$  which can be expressed as  $f(y) = \int k(y|\theta) dG(\theta)$  where  $k(y|\theta)$  is a continuous density function, which will be referred to as a kernel. A prior,  $\Pi$ ,

---

\*Corresponding author: Jim Griffin, Department of Statistics, University of Warwick, Coventry, CV4 7AL, U.K.  
Tel.: +44-1227-82 3865; Fax: +44-24-7652 4532; Email: J.E.Griffin@warwick.ac.uk.

will be placed on  $G$  that allows a wide support for  $f$ . The model can also be represented hierarchically

$$\begin{aligned} y_i &\sim k(y_i|\theta_i) \\ \theta_i &\sim F \\ F &\sim \Pi. \end{aligned}$$

A popular choice for  $\Pi$  is the Dirichlet process (Ferguson, 1973). There has recently been some criticism of the Dirichlet process specification and alternatives have been proposed including: Normalised Inverse Gaussian (NIG) processes (Lijoi *et al* 2005), Normalised Generalized Gamma Process (Lijoi *et al* 2006), and Stick-Breaking Priors (Ishwaran and James 2001, 2003).

Suppose that we know the  $i$ -th observation,  $y_i$ , is made at time  $t_i$ . The extension of this nonparametric mixture model to these data has been an area of recent interest. A natural extension to the nonparametric model assumes that  $f(y|t)$ , the distribution of  $y$  at time  $t$ , can be expressed as  $f(y|t) = \int k(y|\theta) dG_t(\theta)$ . The prior specification is completed by defining a prior distribution for  $G_t$  over a suitable range of values of  $t$ . If the observations arrive in discrete time then a prior can be defined by the recursion

$$G_{t+1} = w_t G_t + (1 - w_t) \epsilon_t \tag{1}$$

where  $\epsilon_t$  is a realisation of a random discrete distribution and  $w_t$  is a random variable in the interval  $(0, 1)$ . Pennell and Dunson (2006) consider a time-invariant  $w_t$  and a Dirichlet process with time-invariant parameters for  $\epsilon_t$ . Griffin and Steel (2006a) consider constructing a process where  $w_1, w_2, w_3, \dots$  are i.i.d. Beta(1,  $M$ ) random variables and  $\epsilon_t$  is a distribution concentrated on a single point drawn from some distribution  $H$ . The dependence structure can be modelled by subordinating this process to a Poisson process with intensity  $\lambda$ . This construction ensures that if  $G_t$  follows a Dirichlet process then so will  $G_{t+j}$ . An alternative approach is considered by Zhu *et al* (2005) who specify the predictive distribution of  $y_{n+1}$ , observed at time  $t_{n+1}$ , which is proportional to

$$\sum_{i=1}^n \exp\{-\lambda(t_{n+1} - t_i)\} \delta_{y_i} + MH$$

where  $\delta_x$  is the Dirac measure that places mass 1 on  $x$  and  $\lambda$  and  $M$  are positive parameters, that will control the marginal process and the dependence between distributions at different times. The form generalizes the famous Pólya urn scheme representation of the Dirichlet process (Blackwell and MacQueen 1974).

This paper develops an alternative approach which guarantees stationarity of the finite distribution of the measure-valued process using the Ornstein-Uhlenbeck (OU) processes

with non-Gaussian marginal distributions developed by Barndorff-Nielsen and Shephard (2001), which will henceforth be referred to as BNS, for financial modelling. This implies that the marginal process at any time is known, which gives some intuition about the behaviour of the process and allows separate elicitation of parameters for the marginal process and the dependence structure. The construction mimics the approach of Zhu *et al* (2005) by defining an exponential decay for a process proportional to  $G_t$ . The process is defined in continuous time but in discrete time it will have a structure like (1) where  $\epsilon_t$  usually has an infinite number of atoms. Markov chain Monte Carlo and particle filtering methods are proposed for posteriors simulation which involve a truncation of the random process. Dirichlet process marginals are an interesting exception and arise from a process with an increment that has a finite number of atoms. Consequently, MCMC and particle filtering methods for this class can be defined without truncation. The paper is organized in the following way: Section 2 reviews the Normalized Random Measures (NRM) class of priors and the construction of OU processes with specified marginal distributions, Section 3 uses these ideas to construct measure valued processes whose marginal process follow specific NRMs, Section 4 describes both Markov chain Monte Carlo and particle filtering methods for inference, Section 5 applies these methods to financial time series, and Section 6 discusses the future development of these types of processes.

## 2 Background

### 2.1 Normalized random measures

Normalized random measures (NRMs) have played an important role in Bayesian nonparametrics since the Dirichlet process was introduced as a normalized Gamma process by Ferguson (1973). The idea was generalized by Kingman (1975) who takes a Poisson process on  $(0, \infty)$  say  $J_1 > J_2 > J_3 > \dots$  such that  $\sum_{i=1}^{\infty} J_i$  is finite and defines probabilities  $p_i = \frac{J_i}{\sum_{i=1}^{\infty} J_i}$ . A random probability measure can be defined by  $G = \sum_{i=1}^{\infty} \frac{J_i}{\sum_{i=1}^{\infty} J_i} \delta_{\theta_i}$  where  $\theta_1, \theta_2, \theta_3, \dots$  are independently drawn from some distribution  $H$ . This guarantees that  $E[G(B)] = H(B)$ . It will be useful to consider the unnormalized measure  $G^* = \sum_{i=1}^{\infty} J_i \delta_{\theta_i}$ . If the support of  $H$  is  $\mathcal{S}$  then  $G = \frac{G^*}{G^*(\mathcal{S})}$ . The class was extended by Regazzini *et al* (2003) to increasing additive processes in one-dimension and by James *et al* (2005) to multivariate processes. The construction of James *et al* (2005) re-expresses Kingman (1975) approach by defining a Poisson process with intensity  $\nu(J, \theta)$  on  $(0, \infty) \times \mathcal{S}$  where  $\mathcal{S}$  is the support of  $H$ . James *et al* (2005) refer to an NRM as *homogeneous* if the intensity function has the form,  $\nu(J, \theta) = W^+(J)h(\theta)$  where  $h$  is differentiable. This paper will concentrate on this homogeneous case and define this process to be  $\text{NRM}(W^+, H)$ . In this paper, the construction

will be extended to a third dimension, time, using a representation of the BNS OU process.

## 2.2 The Gamma OU process

This section describes the development of OU processes with given non-Gaussian marginal distribution and draws heavily on the material in BNS. Suppose that  $z(t)$  is a Lévy process (e.g. Sato, 1999 or Bertoin, 1996) which has dynamics governed by the Stochastic Differential Equation

$$dZ(t) = -\lambda Z(t) + d\phi(\lambda t)$$

where  $\phi(t)$  is called the background driving Lévy process (BDLP). BNS show that the Lévy density  $u$  of  $Z(t)$  is linked to Lévy density  $w$  of  $\phi(t)$  by the following relationship

$$w(x) = -u(x) - xu'(x)$$

which implies that the tail integral function  $W^+(x)$  has the form

$$W^+(x) = \int_x^\infty w(y) dy = xu(x).$$

This will be an important function for the definition of OU-type processes with specified NRM marginal processes.

### Example 1: Gamma

Suppose that  $Z(t) \sim \text{Ga}(\nu, \alpha)$  with density  $\frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} \exp\{-\alpha x\}$  then

$$W^+(x) = \nu \exp\{-\alpha x\}.$$

### Example 2: Inverse Gaussian

Suppose that  $Z(t)$  follows an Inverse Gaussian distribution which has density

$$\frac{\delta}{\sqrt{2\pi}} \exp\{\delta\gamma\} x^{-3/2} \exp\left\{-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right\}$$

then

$$W^+(x) = \frac{\delta}{\sqrt{2\pi}} x^{-1/2} \exp\left\{-\frac{1}{2}\gamma^2 x\right\}.$$

Ferguson and Klass (1972) describe how a Lévy process can be expressed as a transformation of a Poisson process. First we write the process as

$$Z(t) = \exp\{-\lambda t\}Z(0) + \int_0^t \exp\{-\lambda(t-s)\} du(\lambda s). \quad (2)$$

The stochastic integral can be expressed using the Ferguson and Klass (1972) method as

$$\int_0^t f(s) du(\lambda s) \stackrel{d}{=} \sum_{i=1}^{\infty} W^{-1}\left(\frac{a_i}{\lambda t}\right) f(\lambda \tau_i)$$

where  $a_i$  and  $\tau_i$  are two independent sequences of random variables with the  $\tau_i$  independent copies of a uniform random variable  $\tau$  on  $[0, t]$ ,  $a_1 < a_2 < \dots < a_i < \dots$  as the arrival times of a Poisson process with intensity 1 and  $W^{-1}$  is the inverse of the tail mass function  $W^+(x)$ . This representation suggest an alternative interpretation in terms of a jump sizes  $J_i = W^{-1}\left(\frac{a_i}{\lambda t}\right)$  and an arrival times  $\tau_i$ . Let  $(J, \tau)$  be a Poisson process on  $(0, \infty) \times [0, t]$  with intensity  $\lambda W^+(J)$  then

$$\int_0^t f(s) du(\lambda s) \stackrel{d}{=} \sum_{i=1}^{\infty} J_i f(\tau_i).$$

This can clearly be extended to

$$\int_{-\infty}^t f(s) du(\lambda s) \stackrel{d}{=} \sum_{i=1}^{\infty} \mathbf{I}(\tau_i < t) J_i f(\tau_i)$$

where  $(J, \tau)$  be a Poisson process on  $(0, \infty) \times (-\infty, \infty)$  with intensity  $\lambda W^+(J)$ .

### Example 1: Gamma (continued)

Usually the inverse tail mass function  $W^{-1}$  will not be available analytically. However, if the process has a marginal Gamma distribution with parameters  $\nu$  and  $\alpha$  then

$$W^{-1}(x) = \max\left\{0, -\frac{1}{\alpha} \log\left(\frac{x}{\nu}\right)\right\}$$

$$\int_0^t \exp\{-\lambda s\} du(\lambda s) \stackrel{d}{=} \sum_{i=1}^{\infty} \max\left\{0, -\frac{1}{\alpha} \log\left(\frac{a_i}{\lambda \nu t}\right)\right\} \exp\{-\lambda \tau_i\}$$

We are interested in Gamma distribution with shape 1 so we can re-write the expression in the following way

$$\int_0^t \exp\{-\lambda s\} du(\lambda s) \stackrel{d}{=} \sum_{i=1}^{\infty} [-\log a_i] \exp\{-\lambda \tau_i\}$$

where  $a_i$  is a Poisson process with intensity  $\lambda \nu$ .

### 3 The OUNRM processes

The Ferguson and Klass (1972) representation of the BNS OU Lévy process and James *et al* (2005)'s representation of the NRM process are both expressed through functions of Poisson processes. The OUNRM process combines these two ideas to give a Poisson process-based definition. For any NRM process, the distribution of the unnormalized measure  $G^*$  on any measurable set  $B$  has a known form. If we define a time-varying unnormalized measure  $G_t^*$ , then its marginal distribution on a measurable set  $B$  can be fixed over time using the BNS OU process. This leads to the following definition.

**Definition 1** Let  $(\tau, J, \theta)$  be a Poisson process on  $\mathbb{R} \times \mathbb{R}^+ \times \mathcal{S}$  which has intensity  $\lambda W^+(J)h$  and define

$$G_t = \sum_{i=1}^{\infty} \frac{I(\tau_i \leq t) \exp\{-\lambda(t - \tau_i)\} J_i}{\sum_{i=1}^{\infty} I(\tau_i \leq t) \exp\{-\lambda(t - \tau_i)\} J_i} \delta_{\theta_i}$$

then  $\{G_t\}_{t \in \mathbb{R}}$  follows an OUNRM process, which will be written  $\text{OUNRM}(W^+(x), H, \lambda)$

**Theorem 1** If  $\{G_t\}_{t \in \mathbb{R}}$  follows an  $\text{OUNRM}(W^+(x), H, \lambda)$  process then  $G_t$  follows an  $\text{NRM}(W^+(x), H)$  for all  $t$ .

The process can be defined in the form of equation (1) as a mixture of  $G_t$  and an innovation  $\epsilon_t$  where

$$w_t = \frac{\exp\{-\lambda\} G_{t-1}^*(\mathcal{S})}{\exp\{-\lambda\} G_{t-1}^*(\mathcal{S}) + \sum I(t-1 < \tau_i < t) \exp\{-\lambda(\tau_i - t + 1)\} J_i} \quad (3)$$

where  $G^*(t-1) = \sum_{i=1}^{\infty} I(\tau_i \leq t-1) \exp\{-\lambda(t-1 - \tau_i)\} J_i$  which by construction will have the distribution of the unnormalized measure on  $\mathcal{S}$  and

$$\epsilon_t = \frac{\sum I(t-1 < \tau_i < t) \exp\{-\lambda(\tau_i - t + 1)\} \delta_{\theta_i} J_i}{\sum I(t-1 < \tau_i < t) \exp\{-\lambda(\tau_i - t + 1)\} J_i}. \quad (4)$$

The weight  $w_t$  will, in general, be correlated with  $G_t$  and  $\epsilon_{t+1}$ . An important exception is the process with a marginal Dirichlet process, which will be developed in the following subsection. The recursion leads to the following Markov structure for  $G_{t+1}|G_t = E[w_t|G_t]G_t + E[(1-w_t)\epsilon_t|G_t]$ . The process is mean-reverting for all measurable sets. This recursion for the unknown probability distribution naturally extends to expectations of functionals of the unknown distribution. Let  $E_t[a(X)] = \int a(X) dG_t(X)$ . If this expectation exists then

$$E_{t+1}[a(X)]|E_t[a(X)] \stackrel{d}{=} E[w_t]E_t[a(X)] + (1 - E[w_t])\mu$$

where  $\mu$  is the expectation of  $a(X)$  with respect to the centring distribution  $H$ .

A useful summary of the dependence is given by the autocorrelation function of the measure on a set  $B$ , which was suggested by Mallick and Walker (1997). In general the

autocorrelation can be calculated directly from equation (3) and (4) combined with the stationarity of the process. The autocorrelation function will decay like an exponential function asymptotically but at a slower rate for small  $t$ . The effect will be observed in plots for the Dirichlet process marginal case.

### Result 1

The autocorrelation decays like  $\exp\{-\lambda t\}$  if  $E\left[\frac{1}{G_t^*(\mathcal{S})}\right]$  exists. Otherwise the autocorrelation will decay at a slower rate.

## 3.1 The OUDP process

The OUDP is an important subclass of these processes for two reasons: firstly, the Dirichlet process has been the most extensively applied example of an NRM and secondly  $G_t^*(\mathcal{S})$  is independent of  $G_t(B)$  for all  $B$ , which leads to some simplified expressions for summaries of the process. The Dirichlet process (Ferguson, 1973) is defined by normalising a Gamma process with mass parameter  $MH$  and will be written  $DP(M, H)$ . The parameter  $M$  is usually considered a precision parameter since if  $G$  follows a  $DP(M, H)$  then  $\text{Var}(G(B)) = \frac{H(B)(1-H(B))}{M+1}$ .

**Definition 2** Let  $(\tau, J, \theta)$  be a Poisson process on  $\mathbb{R} \times \mathbb{R}^+ \times \mathcal{S}$  which has intensity  $M\lambda \exp\{-J\}h$  and define

$$G_t = \sum_{i=1}^{\infty} \frac{I(\tau_i \leq t) \exp\{-\lambda(t - \tau_i)\} J_i}{\sum_{i=1}^{\infty} I(\tau_i \leq t) \exp\{-\lambda(t - \tau_i)\} J_i} \delta_{\theta_i}$$

then  $\{G_t\}_{t \in \mathbb{R}}$  follows an OUDP process which we will write  $\text{OUDP}(M, H, \lambda)$

**Corollary 1** If  $\{G_t\}_{t \in \mathbb{R}}$  follows an  $\text{OUDP}(M, H, \lambda)$  then  $G_t$  follows a Dirichlet process with mass parameter  $MH$  for all  $t$ .

It follows from the independence of  $G_t(B)$  and  $G_t^*(\mathcal{S})$  that

$$\text{Cov}(G_t, G_{t+k}) = E \left[ \prod_{j=t}^{t+k-1} w_j \right] \text{Var}(G_t).$$

The stationarity of  $G_t$  implies that if the marginal process follows a Dirichlet process then the autocorrelation is

$$\text{Corr}(G_t, G_{t+k}) = E \left[ \frac{\exp\{-\lambda k\} G_t^*(\mathcal{S})}{\exp\{-\lambda k\} G_t^*(\mathcal{S}) + \sum I(\tau_i < k) \exp\{-\lambda \tau_i\} J_i} \right].$$

The autocorrelation function for various choices of  $M$  and  $\lambda$  are shown as the solid line in Figure 1. The parameter  $\lambda$  has the largest effect on the autocorrelation and larger  $\lambda$  are

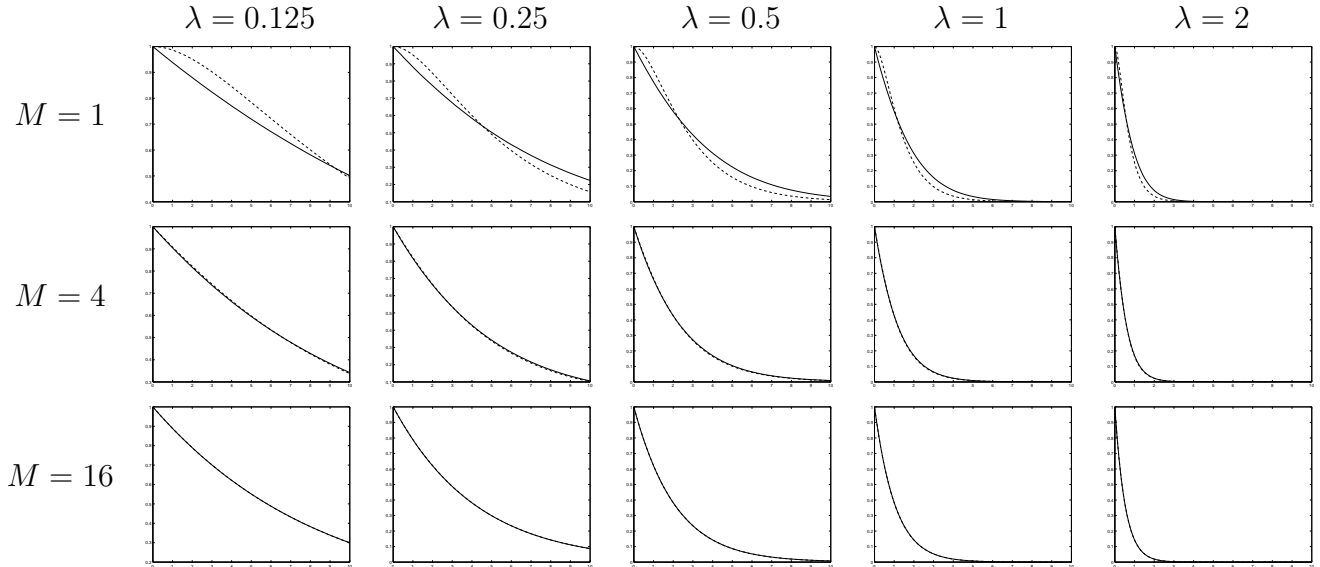


Figure 1: The autocorrelation function of the OUDP for various values of  $M$  and  $\lambda$  (solid line) with the approximation of result 2 (dashed line)

associated with more quickly decaying autocorrelation functions. For a fixed  $\lambda$ , smaller values of  $M$  lead to a more slowly decaying autocorrelation function and in fact if  $M \leq 1$  then result 1 shows that the autocorrelation will decay at a sub-exponential rate. If  $M > 1$ , then  $\text{Corr}(G_t, G_{t+k}) \rightarrow \frac{M}{M-1} \exp\{-\lambda k\}$  as  $k \rightarrow \infty$ . It is useful in these case to have an approximation of the autocorrelation function, which can be derived using the delta method,

$$\text{Corr}(G_t, G_{t+k}) \approx \exp\{-\lambda k\} \left[ 1 + \frac{\sigma^2}{\mu^2} (1 - \exp\{-\lambda k\}) \right]$$

where  $\mu = \mathbb{E}[G^*(S)]$  and  $\sigma^2 = \mathbb{V}[G^*(S)]$ .

The actual autocorrelation and the approximate autocorrelation are shown in figure 1. If  $\sigma^2$  is small relative to  $\mu^2$ , which happens in the Dirichlet process as  $M \rightarrow \infty$ , the approximate autocorrelation becomes increasingly like an exponential function (in other words it inherits the dependence in the unnormalized random measure) and it has a much slower rate of decay when the ratio is large, which is illustrated by the cases with small  $M$ . A second aspect of the dependence is captured through the moments of the unknown distribution. In this case we can show the moments themselves follow the same type of process as the measures on a set  $B$ .

Figures 2 and 3 show the effect of  $\lambda$  and  $M$  on the dependence structure and variability of  $G_t((0, 0.5))$  and the mean of distribution when the Dirichlet process is centred over a uniform distribution. Large values of  $M$  leads to smaller variability and large values of  $\lambda$  lead to quicker mean reversion of the process.



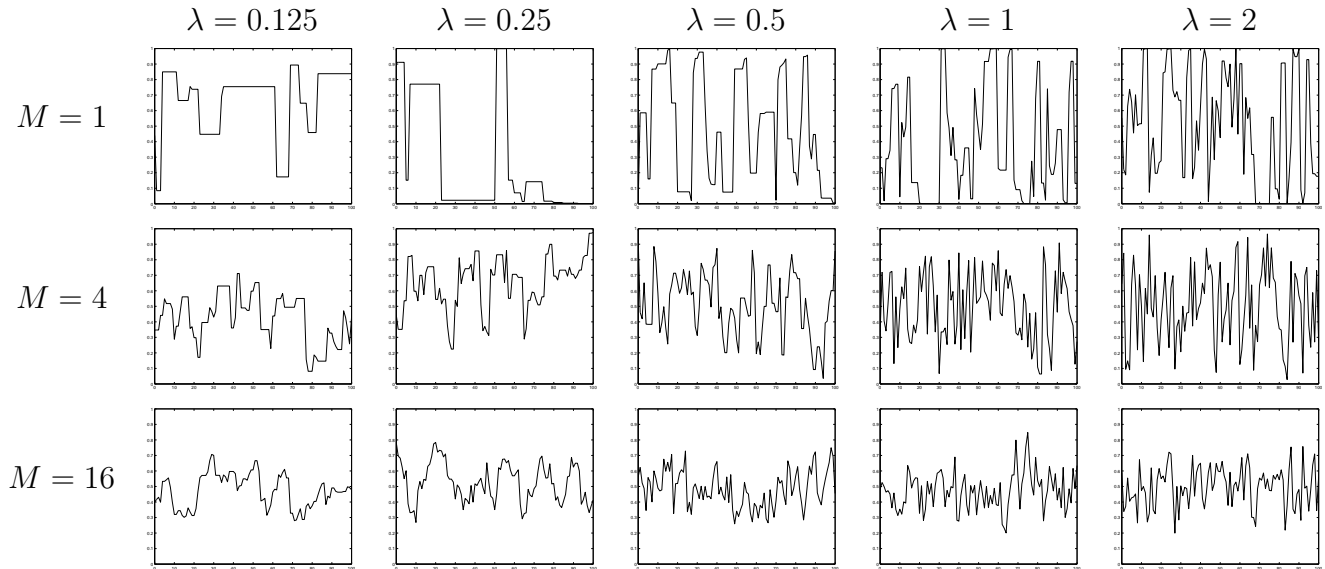


Figure 2: The dynamics of  $G_t((0, 0.5))$  of the OUDP for various values of  $M$  and  $\lambda$  when the centring distribution is uniform on  $(0, 1)$ .

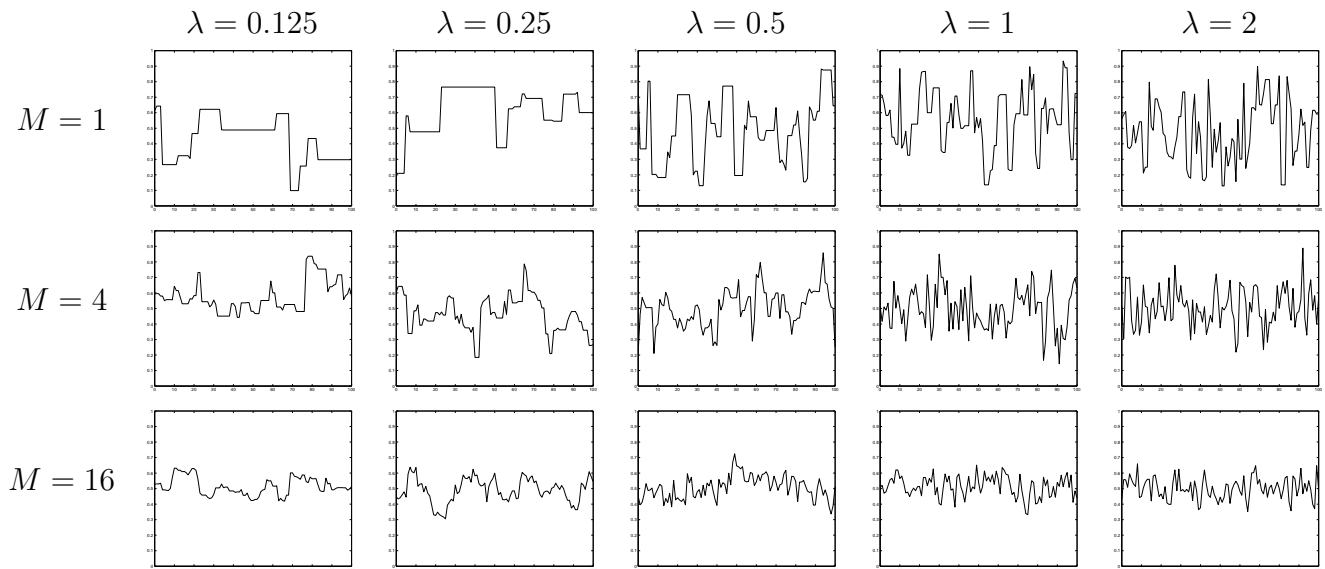


Figure 3: The dynamics of the mean of the OUDP for various values of  $M$  and  $\lambda$  when the centring distribution is uniform on  $(0, 1)$ .

### 3.2 Other marginal processes

We can define dynamic versions of other NRM that have been introduced into the literature. The Pitman-Yor process can be represented as a normalized Stable process with Lévy density

$bx^{-1-a} \exp\{-x\}$  (Pitman and Yor, 1997)

$$W^+(x) = bx^{-a} \exp\{-x\}.$$

The Pitman-Yor process can be generalized to an OUPY process.

**Definition 3** Let  $(\tau, J, \theta)$  be a Poisson process on  $\mathbb{R} \times \mathbb{R}^+ \times \mathcal{S}$  which has intensity  $b\lambda E^{-a} \exp\{-E\}h$  and define

$$G_t = \sum_{i=1}^{\infty} \frac{I(\tau_i \leq t) \exp\{-\lambda(t - \tau_i)\} J_i}{\sum_{i=1}^{\infty} I(\tau_i \leq t) \exp\{-\lambda(t - \tau_i)\} J_i} \delta_{\theta_i}$$

then  $\{G_t\}_{t \in \mathbb{R}}$  follows an OUPY process which we will write  $\text{OUPY}(a, b, H, \lambda)$

Similarly, the Normalized Inverse Gaussian process can be generalized to an OUNIG process.

**Definition 4** Let  $(\tau, J, \theta)$  be a marked Poisson process on  $\mathbb{R} \times \mathbb{R}^+ \times \mathcal{S}$  which has intensity  $\frac{\delta}{\sqrt{2\pi}} J^{-1/2} \lambda \exp\{-\frac{1}{2}\gamma^2 E\}h$  and define

$$G_t = \sum_{i=1}^{\infty} \frac{I(\tau_i \leq t) \exp\{-\lambda(t - \tau_i)\} J_i}{\sum_{i=1}^{\infty} I(\tau_i \leq t) \exp\{-\lambda(t - \tau_i)\} J_i} \delta_{\theta_i}$$

then  $\{G_t\}_{t \in \mathbb{R}}$  follows an OUNIG process which we will write  $\text{OUNIG}(\delta, \gamma, H, \lambda)$

## 4 Computational methods

This section describes computational methods for fitting a hierarchical mixture model where we observe a sample  $y_1, y_2, \dots, y_n$  at times  $0 = t_1 \leq t_2 \leq \dots \leq t_n = T$  respectively. It is assumed that

$$y_i \sim k(y_i | \theta_i, \phi)$$

$$\theta_i \sim F_{t_i}$$

$$\{F_t\} \sim \text{OUNRM}(W^+(x), H, \lambda)$$

The model can be written in an alternative way by introducing latent variable  $s_1, s_2, \dots, s_n$  that link the observations to the elements of the Poisson process representation

$$y_i \sim k(y_i | \theta_{s_i}, \phi)$$

$$p(s_i = j) = \frac{I(\tau_j < t_i) \exp\{-\lambda(t_i - \tau_j)\} J_j}{\sum I(\tau_j < t_i) \exp\{-\lambda(t_i - \tau_j)\} J_j}$$

and  $(\tau, J, \theta)$  follow the appropriate Poisson process.

## 4.1 MCMC sampler

The model can be fitted using the Gibbs sampler. Methods for fitting the OU model in stochastic volatility models have been described by Roberts *et al* (2004) and Griffin and Steel (2006b) for the Gamma marginal distribution. Extensions to Generalized Inverse Gaussian marginal distributions are described in Gander and Stephens (2006). A major problem with these models are the dependence of the timing of the Lévy process on  $\lambda$  which also controls the rate of decay of the jumps. This is a particular problem when updating the mean of the underlying Poisson process conditional on the number of jumps and mixing can usually be improved by jointly updating the Poisson process of jumps with the parameters controlling the mean number of jumps. In the OUDP process, for example, there is usually a high correlation between  $M$  and  $\lambda$  and it is useful to reparameterize to  $\zeta = M\lambda$  and  $\lambda$ . The OUNRM process can be separated into two parts: the initial distribution  $G_0$  with weight  $\gamma$  and the subsequent jumps. Indicator variables  $r_1, r_2, \dots, r_n$  are introduced that indicate whether an observation is drawn from  $G_0$  or from the subsequent jumps. A second set of indicators  $s_1, \dots, s_n$  link an observation to a jump (if  $r_i = 1$ ) or the distinct values of  $G_0$  (if  $r_i = 0$ ). Let  $K$  be the number of distinct elements of  $G_0$  which have had observations allocated to them and let  $L$  be the number of jumps between 0 and  $T$ . Let  $\theta_i$  be the distinct values of  $G_0$ . Let  $J_i$  be the values of the jumps and let  $\phi_i$  be the value of the distinct value for that jump. I will concentrate on the OUDP case although the methods can be simply extended to other OUNRM processes. For example, updating parameters connected to  $G_0$  could be implemented using the methodology of James *et al* (2005). In general, the number of subsequent jumps will be almost surely infinite in any region and a method for truncation is described in Gander and Stephens (2006). Finally, the exposition is helped by defining the quantity

$$D_i = \exp\{-\lambda t_i\} \gamma_k + \sum_{m=1}^L \mathbf{I}(\tau_m < t_i) \exp\{-\lambda(t_i - \tau_m)\} J_m$$

### 4.1.1 Updating $r$ and $s$

The full conditional distribution of the latent variables  $(s_i, r_i)$  is a discrete distribution with the following probabilities

$$p(r_i = 0, s_i = j) \propto k(y_i | \theta_j) \exp\{-\lambda t_i\} \gamma \frac{n_j}{M + n^*}, \quad 1 \leq j \leq K$$

$$p(r_i = 1, s_i = j) \propto \mathbf{I}(\tau_j < t_i) k(y_i | \psi_j) \exp\{-\lambda(t_i - \tau_m)\} J_m, \quad 1 \leq j \leq L$$

where  $n_j = \#\{i | r_i = 0, s_i = j\}$  and  $n^* = \sum_{j=1}^L n_j$ .

### 4.1.2 Updating $G_0$ and $\gamma$

The full conditional distribution of  $\gamma$  is proportional to

$$p(\gamma) \prod_{i=1}^n \frac{(\exp\{-\lambda t_i\}\gamma)^{1-r_i} (\exp\{-\lambda(t_i - \tau_{s_i})\}J_{s_i})^{r_i}}{\exp\{-\lambda t_i\}\gamma + \sum_{\{j|\tau_j < t_i\}} \exp\{-\lambda(t_i - \tau_j)\}J_j}.$$

This can be simulated using a slice sampler (Damien *et al* (1999), Neal 2003) by introducing auxiliary variables  $u_1, \dots, u_n$  which are uniformly distributed and the density of the joint distribution can be expressed as

$$p(\gamma)p(u_1, \dots, u_n) \prod_{i=1}^n \prod_{i=1}^n \mathbf{I} \left( u_i < \frac{(\exp\{-\lambda t_i\}\gamma)^{1-r_i} (\exp\{-\lambda(t_i - \tau_{s_i})\}J_{s_i})^{r_i}}{\exp\{-\lambda t_i\}\gamma + \sum_{\{j|\tau_j < t_i\}} \exp\{-\lambda(t_i - \tau_j)\}J_j} \right).$$

The marginal distribution of  $\gamma$  is the full conditional distribution above. Then the full conditional distribution of  $u_i$  is a uniform distributed on the region

$$\left( 0, \frac{(\exp\{-\lambda t_i\}\gamma)^{1-r_i} (\exp\{-\lambda(t_i - \tau_{s_i})\}J_{s_i})^{r_i}}{\exp\{-\lambda t_i\}\gamma + \sum_{\{j|\tau_j < t_i\}} \exp\{-\lambda(t_i - \tau_j)\}J_j} \right).$$

The full conditional distribution of  $\gamma$  is now the prior  $p(\gamma)$  truncated to the region  $A < \gamma < B$  where

$$A = \max_{i \in \mathcal{A}} \left\{ \frac{u_i}{1 - u_i} \exp\{\lambda t_i\} (D_i - \exp\{-\lambda t_i\}\gamma) \right\}$$

$$B = \min_{i \in \mathcal{B}} \left\{ \exp\{\lambda t_i\} \left[ \frac{\phi_i}{u_i} - (D_i - \phi_i) \right] \right\}.$$

where  $\mathcal{A} = \{i | r_i = 0\}$ ,  $\mathcal{B} = \{1, \dots, n\} - \mathcal{A}$ ,  $\phi_i = \exp\{-\lambda(t_i - \tau_{s_i})\}J_{s_i}$ . If we want a marginal Dirichlet process then  $p(\gamma_l) \sim \text{Ga}(M, 1)$ .

### 4.1.3 Updating the jumps

The number and size of the jumps can be updated using a hybrid Metropolis-Hastings sampler with four possible moves: Add, Delete, Change and Move. At each iteration a move is chosen at random with the constraint that probability of choosing an Add and a Delete are equal. The Add move proposes to increase the number of jumps by one by drawing  $J_{L+1}$  from an exponential(1) distribution,  $\tau_{L+1}$  from a uniform distribution on  $(0, T)$  and  $\psi_{L+1}$  from  $H$ . The Metropolis-Hastings acceptance ratio is

$$\min \left\{ 1, \prod_{i=1}^n \frac{D_i}{D_i + \mathbf{I}(\tau_{L+1} < t_i) \exp\{-\lambda(t_i - \tau_{L+1})\}J_{L+1}} \frac{T\lambda}{L+1} \right\}.$$

The Delete move proposes to decrease the number of jumps by selecting one jump at random to be removed. Let this jump be the  $k$ -th one. If number of observations allocated to that

jump is non-zero then the acceptance probability is 0. Otherwise, the acceptance probability is

$$\min \left\{ 1, \frac{D_i}{D_i - \mathbf{I}(\tau_k < t_i) \exp\{-\lambda(t_i - \tau_k)\} J_k} \frac{L}{T\lambda} \right\}.$$

The other two moves uses a slice sampler to update the size of jump (Change) and the timing of the jump (Move). In both cases, the  $k$ -th jump is chosen at random. Latent variable  $u_1, \dots, u_n$  are introduced drawing from a uniform distribution on the interval

$$\begin{aligned} & \left( 0, \frac{\exp\{-\lambda t_i\} \gamma}{\exp\{-\lambda t_i\} \gamma + \sum_{\{j|\tau_j < t_i\}} \exp\{-\lambda(t_i - \tau_j)\} J_j} \right) \text{ if } r_i = 0 \\ & \left( 0, \frac{\exp\{-\lambda(t_i - \tau_{s_i})\} J_{s_i}}{\exp\{-\lambda t_i\} \gamma + \sum_{\{j|\tau_j < t_i\}} \exp\{-\lambda(t_i - \tau_j)\} J_j} \right) \text{ if } r_i = 1. \end{aligned}$$

If we are performing a Change move then the full conditional distribution of  $J_k$  is the distribution of  $J_k$  truncated to the region  $(A, B)$  where

$$\begin{aligned} A &= \max_{i \in \mathcal{A}} \left\{ \exp\{\lambda_P(t_i - \tau_k)\} \frac{u_i}{1 - u_i} (D_i - \exp\{-\lambda(t_i - \tau_k)\} J_k) \right\} \\ B &= \min_{i \in \mathcal{B}} \left\{ \exp\{\lambda(t_i - \tau_k)\} \left[ \frac{\phi_i}{u_i} - (D_i - \phi_i) \right] \right\} \end{aligned}$$

where

$$\begin{aligned} \mathcal{A} &= \{i | r_i = 1, s_i = k\} \\ \mathcal{B} &= \{i | t_i > \tau_k\} \setminus \mathcal{A} \\ \phi_i &= \exp\{-\lambda(t_i - \tau_{s_i})\} J_{s_i} \end{aligned}$$

If we are performing a Move update then the full conditional distribution of  $\tau_k$  has the following form. Let  $r_{max} = \min\{T, \max\{t_i | s_i = k\}\}$ . Let  $t_1^* < \dots < t_h^*$  be an ordered version of all the values of  $t$  less than  $r_{max}$ ,  $t_0^* = 0$  and  $t_h^* = r_{max}$  then  $\tau_k$  is drawn from a uniform distribution on the region  $(A, B_0) \cap (0, t_1) \cup \bigcup_{i=1}^{h-1} (A, B_i) \cap (t_i, t_{i+1}) \cup (A, B_h)(t_h, r_{max})$  where

$$\begin{aligned} A &= \max_{j \in \mathcal{A}} \left\{ t_j + \frac{1}{\lambda} [\log u_j - \log(1 - u_j) - \log J_k + \log (D_j - \exp\{-\lambda(t_i - \tau_k)\} J_k)] \right\} \\ B_i &= \max_{j \in \mathcal{B}_i} \left\{ t_j + \frac{1}{\lambda} [\log((1 + u_j)\phi_j - u_j D_j) - \log J_k - \log u_j] \right\} \end{aligned}$$

where

$$\begin{aligned} \phi_i &= \exp\{-\lambda(t_i - \tau_{s_i})\} J_{s_i} \\ \mathcal{A} &= \{r_i = 1, s_i = k\} \\ \mathcal{B}_i &= \{j | t_j > t_i^*\} \setminus \mathcal{A} \end{aligned}$$

#### 4.1.4 Updating $\theta$ and $\psi$

The full conditional distribution of  $\theta_i$  is proportional to

$$h(\theta_i) \propto \prod_{\{j|s_j=i \text{ and } r_j=0\}} k(y_j|\theta_i, \phi)$$

and the full conditional distribution of  $\psi_i$  is proportional to

$$h(\psi_i) \propto \prod_{\{j|s_j=i \text{ and } r_j=1\}} k(y_j|\psi_i, \phi).$$

These full conditional distribution will arise in Dirichlet process or NRM mixture models and can be sampled using methods for the corresponding static model.

#### 4.1.5 Updating $\lambda$

The parameters  $\lambda$  can be updated from its full conditional distribution using a Metropolis-Hastings random walk proposal where a new value  $\lambda'$  is chosen from the transition kernel  $q(\lambda, \lambda')$  which has the acceptance probability

$$\frac{p(\lambda')p(M')q(\lambda', \lambda)}{p(\lambda)p(M)q(\lambda, \lambda')} \frac{\varphi(\lambda')}{\varphi(\lambda)} \frac{\lambda}{\lambda'} \left(\frac{M'}{M}\right)^K \prod_{i=1}^K \frac{M+i-1}{M'+i-1}$$

where

$$\varphi(\lambda) = \prod_{i=1}^n \frac{(\exp\{-\lambda t_i\}\gamma)^{1-r_i} (\exp\{-\lambda(t_i - \tau_{s_i})\})^{r_i}}{\exp\{-\lambda t_i\}\gamma + \sum_{\{j|\tau_j < t_i\}} \exp\{-\lambda(t_i - \tau_j)\} J_j}$$

and  $M' = \frac{\zeta}{\lambda'}$ .

#### 4.1.6 Updating $\zeta$

Using a Gibbs update for  $\zeta$  can lead to slow mixing. Roberts *et al* (2004) propose a data augmentation method and Griffin and Steel (2006b) suggest using a method called “dependent thinning” to jointly update  $\zeta$  and  $\{(\tau_i, J_i, \theta_i)\}_{i=1}^K$ . In both methods, a new value  $\zeta'$  is proposed from the transition kernel  $q(\zeta, \zeta')$ , such as a normal distribution centred on  $\zeta$ . Under the Roberts *et al* (2004) scheme if  $\zeta' > \zeta$ , a random number of jumps  $K' - K$  is drawn from a Poisson distribution with mean  $T(\zeta' - \zeta)$  and

$$J'_i \sim \text{Ex}(1), \tau'_i \sim \text{U}(0, T), \theta'_i \sim H, \quad K < i \leq K'$$

and if  $\zeta' > \zeta$  jumps are deleted if  $U_i > \frac{\zeta'}{\zeta}$  where  $U_1, U_2, \dots, U_K$  are i.i.d. and  $U_i$  are uniform random variates. Under the “dependent thinning” scheme of Griffin and Steel (2006b), if

$\zeta' > \zeta$ , then a random number of jumps  $K' - K$  is drawn from a Poisson distribution with mean  $T(\zeta' - \zeta)$  and

$$J'_i = -\log \left( \frac{\zeta}{\zeta'} + \left( 1 - \frac{\zeta}{\zeta'} \right) u'_i \right), \tau'_i \sim \text{U}(0, T), \theta'_i \sim H, \quad i > K$$

where  $u'_i$  is uniform random variate and

$$J'_i = J_i + \log \zeta' - \log \zeta, \tau'_i = \tau_i, \theta'_i = \theta_i, \quad i \leq K.$$

If  $\zeta' < \zeta$  then if  $J_i < \log \zeta - \log \zeta'$  the jump is deleted and otherwise

$$J'_i = J_i - \log \zeta + \log \zeta', \tau'_i = \tau_i, \theta'_i = \theta_i.$$

In all cases the proposed values  $\lambda', (\tau', J', \theta')$  are accepted according to the Metropolis-Hastings acceptance probability

$$\min \left\{ 1, \frac{p(M') \prod_{i=1}^n p(s_i, r_i | \tau', J', \lambda') q(\lambda', \lambda)}{p(M) \prod_{i=1}^n p(s_i, r_i | \tau, J, \lambda) q(\lambda, \lambda')} \right\}$$

where  $M = \frac{\zeta}{\lambda}$  and  $M' = \frac{\zeta'}{\lambda}$ . I have found very little difference in the performance of the sampling schemes based on these two updates in applications.

#### 4.1.7 Updating other parameters

Any hyperparameters of  $h(\theta)$  giving a parametric function  $h(\theta|\phi)$  can be updated from the full conditional distribution

$$p(\phi) \prod_{i=1}^K h(\theta_i|\phi) \prod_{i=1}^L h(\psi_i|\phi).$$

In static versions of the models described in this paper this full conditional distribution will arise and can be sampled using the same methods as in computational methods for those models.

## 4.2 Particle Filters

Particle filters are a computational methods for fitting the generic state space model for observation  $y_i$  with states  $\theta_i$  described by

$$y_i | \theta_i \sim f(y_i | \theta_i)$$

$$\theta_i | \theta_{i-1} \sim g(\theta_i | \theta_{i-1})$$

for distributions  $f(y_i|\cdot)$  and  $g(\theta_i|\cdot)$  by finding a sample approximation to the posterior distribution of the states  $p(\theta_i|y_1, \dots, y_i)$  for all values of  $i$ . If we have a sample  $\theta_{i-1}^{(1)}, \theta_{i-1}^{(2)}, \dots, \theta_{i-1}^{(N)}$  from  $p(\theta_i|y_1, \dots, y_i)$  then we generate  $\tilde{\theta}_i^{(j)}$  from  $g(\cdot|\theta_{i-1}^{(j)})$ , which is called the propagation step. A sample  $\theta_i^{(1)}, \theta_i^{(2)}, \dots, \theta_i^{(N)}$  can be generated by drawing  $N$  values from a discrete distribution with atoms  $\tilde{\theta}_i^{(j)}$  which have probabilities proportional to  $f(y_i|\tilde{\theta}_i^{(j)})$ . Doucet *et al* (2001) provide a description of the method and several advances. If we fix  $\lambda$  and  $M$  then the OUDP-based hierarchical model fits into this framework where the states are  $\theta_i$  and  $F_i$ . The method can be extended to models with parameters that are static. These methods have also been used in static problems such as the Dirichlet process mixture model when all observations are made at the same time (see *e.g.* Liu *et al* (1999) and Fearnhead (2003)). Recently, Caron *et al* (2006) have developed methods for state space models where the noise in the observation equation and the system equation is assumed to arise from an unknown distribution given a Dirichlet process prior, which is time-invariant. In this paper, we propose an algorithm based on the work of Fearnhead (2003) and Caron *et al* (2006). The main difference between the idea presented here is the inclusion of the unknown distribution  $G_t$  as a state variable. In previous work the unknown distribution is assumed fixed which can lead to some problems with the convergence and consequently Monte Carlo errors of estimated distribution (*e.g.* Caron *et al* (2006)). The distribution is now time-varying and enjoys the benefits of rejuvenation for estimation of states and the unknown distribution at the propagation step. Fearnhead (2003) suggests integrating across the unknown distribution to produce the classic Polya urn scheme of Blackwell and MacQueen (1973) which fits into a Rao-Blackwellised scheme. This idea is not available for the prior defined in this paper and I will use a truncated representation of the full probability measure. I present two versions of the algorithm: a version which works with all models and a second version that relies on the conjugacy of the kernel and the centring distribution. The marginalisation used in the second method produces a Rao-Blackwellised scheme that will usually lead to much better performance than the first algorithm. The initial distribution  $G_0$  can be simulated using the stick-breaking construction of the Dirichlet process or by a truncated version of the NRM more generally.  $N$  atoms are chosen for the particle filter and for each atom a truncated Dirichlet process is generated with  $K$  distinct elements for each  $i = 1, 2, \dots, N$  where

$$w_j^{(i)} = V_j^{(i)} \prod_{k < j} (1 - V_k^{(i)})$$

where  $V_1^{(i)}, V_2^{(i)}, V_3^{(i)}, \dots, V_{K-1}^{(i)} \sim \text{Be}(1, M)$ ,  $V_K^{(i)} = 1$ ,  $\theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_K^{(i)} \sim H$  and simulate  $\gamma^{(i)}$  from a Gamma distribution with shape parameter  $M$  and scale parameter 1. Finally, a counter of the number of jumps is initialized for each atom:  $\eta_0^{(1)} = K, \eta_0^{(2)} = K, \dots, \eta_0^{(N)} = K$  and set  $J_j^{(i)} = \gamma^{(i)} w_j^{(i)}$ .



### 4.2.1 Simple Particle Filtering Algorithm

Let  $(J^{(i)}, \theta^{(i)}, n^{(i)})$  be the  $i$ -th atom of the particle filter approximation at time  $t$ . The algorithm works by first simulating an updated atom  $(\tilde{J}^{(i)}, \tilde{\theta}^{(i)}, \tilde{\eta}^{(i)})$  using the prior updating rules then uses an importance-resampling step to produce a sample from the posterior distribution for the first  $t + 1$  observations. The first step works by proposing

$$\tilde{\eta}^{(i)} - \eta^{(i)} \sim \text{Pn}(M\lambda(t_{t+1} - t_t),$$

$$\tilde{J}_j^{(i)} = \begin{cases} \exp\{-\lambda(t_{t+1} - t_t)J_j^{(i)}\} & j \leq \eta^{(i)} \\ \exp\{-\lambda\tau_j^{(i)}z_j^{(i)}\} & \eta^{(i)} < j < \tilde{\eta}^{(i)} \end{cases} .$$

where  $\tau_j^{(i)} \sim U(0, (t_{t+1} - t_t))$  and  $z_j^{(i)} \sim Ex(1)$  and

$$\tilde{\theta}_j^{(i)} = \begin{cases} \theta_j^{(i)} & j \leq \eta^{(i)} \\ \psi_j^{(i)} & \eta^{(i)} < j < \tilde{\eta}^{(i)} \end{cases} .$$

where  $\psi_j^{(i)} \sim H$ . The importance resampling step works in the following way, each atom is assigned a weight  $p_1, \dots, p_N$  according to

$$p_k = \frac{\sum_{k=1}^{\tilde{\eta}^{(i)}} J_k^{(i)} k(y_{t+1} | \theta_k^{(i)})}{\sum_{k=1}^{\tilde{\eta}^{(i)}} J_k^{(i)}}, \quad 1 \leq k \leq \tilde{\eta}^{(i)}$$

A new sample of atoms is formed by sampling the  $i$ -th atom according to a weight proportional to  $\psi_i$ . A more efficient method of re-weighting the sample is described by Carpenter *et al* (1999) for the examples.

### 4.2.2 Rao-Blackwellised Particle Filtering Algorithm

The first algorithm can lead to fast degeneracy of the sample, which leads to an inaccurate Monte Carlo approximations. A more accurate particle filter arises if we can integrate the locations  $\theta$  from the sampler. This will often be possible if we choose a conjugate Dirichlet process if we can calculate the predictive distribution  $m(y^* | \mathcal{A}) = \frac{\int k(y^* | \theta) \prod_{i \in \mathcal{A}} k(y_i | \theta) p(\theta) d\theta}{\int \prod_{i \in \mathcal{A}} k(y_i | \theta) p(\theta) d\theta}$  where  $\mathcal{A}$  is a subset of  $1, 2, \dots, n$ . The sampler then resembles the approach taken by Fearnhead (2003) and Caron *et al* (2006) which is described as a Rao-Blackwellisation of the algorithm. A further step must be added to make this approach work. We include allocation variables, as in the MCMC sampler,  $s_1^{(i)}, \dots, s_t^{(i)}$  for each particle. The atom now is formed by  $(J^{(i)}, s^{(i)}, n^{(i)})$  where  $s^{(i)}$  is the allocation variables. Often the predictive distribution is a function of sufficient statistics of the data and the computational burden can be eased in these situations (see Fearnhead (2003) for details). At the  $t + 1$ -th time point we propose  $(\tilde{J}^{(i)}, \tilde{s}^{(i)}, \tilde{\eta}^{(i)})$  where  $\tilde{\eta}^{(i)}$  and  $\tilde{J}^{(i)}$  are generated as in the Simple Particle Filtering Algorithm.

The new value  $\tilde{s}^{(i)}$  is proposed as follows:  $\tilde{s}_k^{(i)} = s_k^{(i)}, k \leq t$  and we  $\tilde{s}_{t+1}^{(i)}$  with the following probabilities

$$p(\tilde{s}_{t+1}^{(i)} = k) \propto \tilde{J}_k^{(i)} m(y_{t+1} | \mathcal{Y}_k^{(i)}), \quad 1 \leq k \leq \tilde{\eta}^{(i)}$$

where  $\mathcal{Y}_k^{(i)} = \{y_j | s_j^{(i)} = k, \quad 1 \leq j \leq t\}$ . The weights for the atoms now have the form

$$p_i = \sum_{k=1}^{\tilde{\eta}^{(i)}} \tilde{J}_k^{(i)} m(y_{t+1} | \mathcal{Y}_k^{(i)}), \quad 1 \leq i \leq N$$

where the re-weighting can be performed in the same way as the Simple Particle Filtering Algorithm.

## 5 Example: Volatility Estimation

In financial time series, the estimation of volatility of the returns of an asset are an object of interest. The return is defined by  $r_t = \log P_t - \log P_{t-1}$  where  $P_t$  is a time series of observed asset prices. Models for this type of data usually contain certain stylized features (see *e.g.* Taylor 1986) which include: a small mean, heavier than normal tails and a time-varying variance (or volatility, which is its square root). Most work has considered modelling the last of these characteristics. Tsay (2005) gives an excellent introduction to the modelling of financial time series. There are two main models: GARCH models and stochastic volatility models. The former assumes that the variance is some deterministic function of lagged squared returns and volatilities whereas the latter assumes that the variance follows a stochastic process. In this paper, I will consider fitting a semi-parametric model which is closer in spirit to the latter approach. We assume that the volatility is drawn from a time-varying nonparametric distribution which is unobserved and so will induce dependence between the volatilities. I consider fitting the following model

$$r_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 \sim G_t$$

$$G_t \sim \text{OUDP}(M, \text{IG}(a, b), \lambda).$$

The model is constructed in a similar way to a discrete stochastic volatility model (*e.g.* Shephard, 2005). The marginal distribution of  $r_t$  will be a scale mixture of normal distributions which will lead to heavier tails than a normal distribution. The parameters of the centring distributions are estimated by assuming no dependence in the returns and dividing each estimate by three to give a centring distribution with a wider support. In this model  $\sigma_t^2$  is now drawn from the conditional distribution of the volatility represented by  $G_t$ . The prior distributions for  $M$  is chosen to be  $\text{Ex}(1)$ . The parameter  $\lambda$  controls the temporal dependence of the processes and the prior distribution is chosen to be  $\text{Ex}(10)$ .

## 5.1 Simulated Data

The model is initially fitted to data generated from a discrete time stochastic volatility model where the log volatility follows an autoregressive process

$$r_t \sim N(0, \sigma_t^2)$$

$$\log \sigma_t^2 = a \log \sigma_{t-1}^2 + \epsilon_t$$

where  $\epsilon_t \sim N(0, b)$ . The values chosen were  $a = 0.8$  and  $b = \sqrt{1 - 0.8^2}$  which leads to a

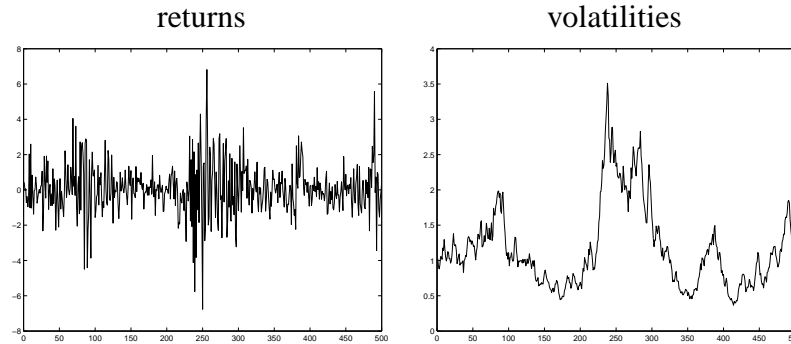


Figure 4: The simulated data and underlying volatilities

marginal log-normal distribution for the volatility. The time series of returns and volatilities are illustrated in figure 4 showing several periods of high volatility. In particular, there is a short-lived period of extreme volatility in the middle of the series.

Table 1 shows the posterior inference on the parameter of the model. The posterior distribution of  $M$  places mass on relatively small values of  $M$  indicating that the conditional distribution is far from the marginal distribution. This is not surprising given the difference between the marginal distribution and the conditional distribution. The parameter  $\lambda$  is small which implies that the autocorrelation function decays slowly: reaching 0.5 after 14 days and 0.05 after 42 days. The value of  $M\lambda$  indicates that the process will have a relatively small number of jumps. A new jump will appear on average every 11 days.

$M$	1.01 (0.53, 1.93)
$\lambda$	0.087 (0.050, 0.166)
$\exp\{-\lambda\}$	0.92 (0.85, 0.95)
$M\lambda$	0.09 (0.04, 0.16)

Table 1: Summaries of the posterior distribution showing the posterior median followed by the 95% highest posterior probability region for several parameters when the model is fitted

Smoothed and filtered estimates of the volatilities for these data are shown in figure 5. Both estimates are able to capture the features of the changing volatility and clearly show the periods of high volatility. The period of highest volatility in the middle of the series seems to be a little shrunk. The filtered estimates are calculated using the particle filtering algorithm 2 conditional on the posterior median estimates of  $\lambda$  and  $M$ . They behave in a similar way to the smoother estimates but show some predictable difference. The median estimate is similar to the smoothed estimate and gives a good indication of the underlying changes in the volatility.

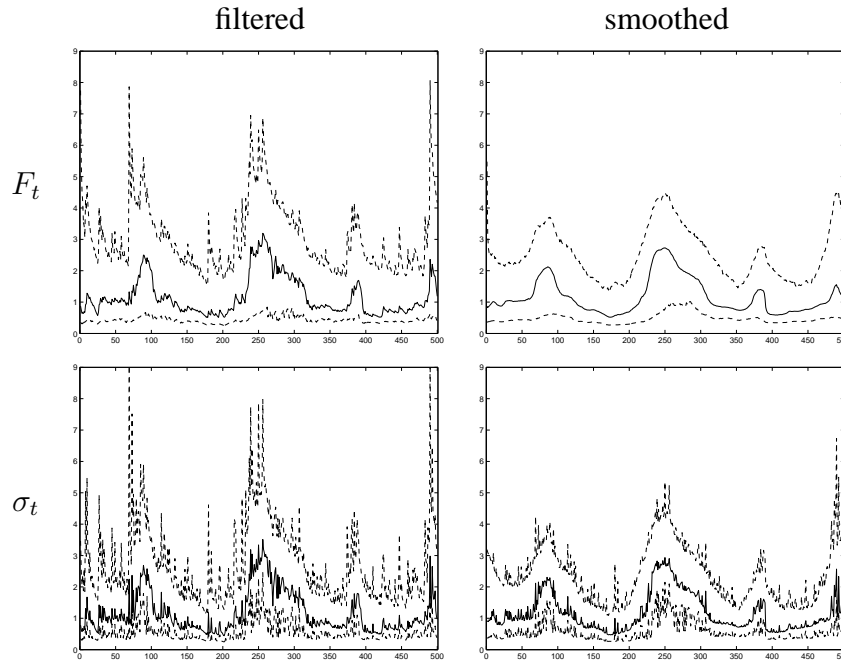


Figure 5: The filtered and smoothed estimates of the volatilities with the OUDP model for the simulated data. The results are represented by the median (solid line) and the 2.5 and 97.5 percentile (dashed lines)

## 5.2 Brazilian stock exchange data

The Ibovespa index of the Brazilian stock exchange between 14/7/92 and 16/5/96 are plotted in figure 6. The graph shows a period where the spread of returns is fairly constant upto time 600 followed by a brief period of higher volatility followed by a longer period of low volatility. Fitting the OUDP model to the data yields a posterior distribution of  $M$  that is similar to the simulated data example. However the parameter controlling dependence  $\lambda$  is estimated to be smaller indicating substantial correlation at longer times. The autocorrelation

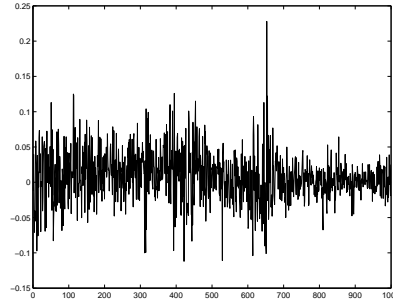


Figure 6: A time series graph of daily returns from the Ibovespa index for the Brazilian stock exchange

function equals 0.5 after 23 days and equals 0.05 after 70 days. The posterior median estimate of  $M\lambda$  is also smaller with a new jump appearing every 18 days on average.

$M$	1.05 (0.54, 1.82)
$\lambda$	0.052 (0.029, 0.098)
$\exp\{-\lambda\}$	0.95 (0.91, 0.97)
$M\lambda$	0.055 (0.037, 0.079)

Table 2: A summary of the posterior distribution showing the posterior median followed by the 95% highest posterior probability region for  $M$  for the Ibovespa data

Figure 7 shows the smoothed and filtered results, which were calculated in the same ways as with the simulated data. These graphs show the Fmain features of the volatility. An initial constant period of volatility is followed by a period where the distribution of the volatility becomes more widely spread. The results indicate that the right-hand tail of the distribution increases a lot (to capture the few large returns) which contrasts with the median value of volatility that changes a much less pronounced way. The final period of lower volatility is captured.

Figure 8 shows the smoothed and filtered posterior distribution of volatility at several selected time points. It shows that the distribution of volatility at several different time points which indicates how the various periods of volatility are captured and show a substantial variation in the fitted shapes.

## 6 Discussion

This paper has introduced a method for constructing stochastic processes with particular marginal processes which generalize the Normalized Random Measures priors for Bayesian

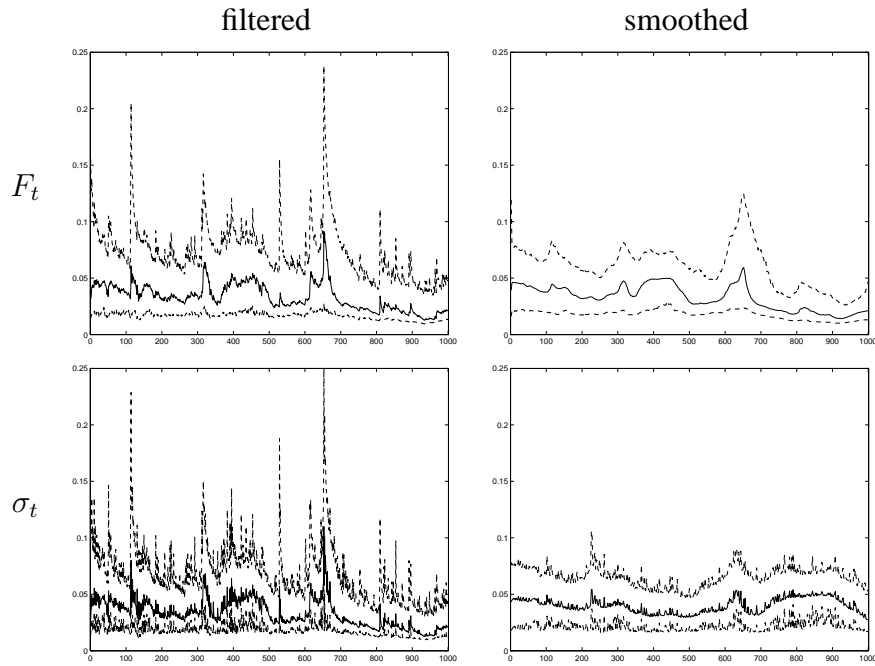


Figure 7: The filtered and smoothed estimates of the volatilities with the OUDP model for the Ibovespa data. The results are represented by the median (solid line) and the 2.5 and 97.5 percentile (dashed lines)

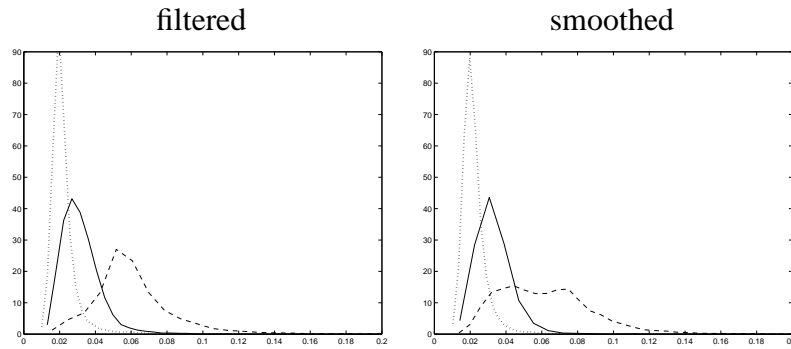


Figure 8: The filtered and smoothed estimates of the distribution with the OUDP model for the Ibovespa data at times: 200 (solid line), 652 (dashed line) and 994 (dashed line).

nonparametric inference to time-varying problems. The autocorrelation function of the unnormalized process has an exponential form which is inherited by the normalized process in many situations. The special case of the OUDP process which has a marginal Dirichlet process is considered. Inferential methods using Markov chain Monte Carlo methods and particle filtering are developed. The simple form of the process suggest a number of possible extensions to allow more general time-dependencies between distributions. One possible ap-

proach uses mixed OU processes (Barndorff-Nielsen, 2000) where the unnormalized process is defined by

$$G_t^* = \int G_t^*(\lambda)\mu(d\lambda)$$

where  $G_t^*(\lambda)$  follows stochastic differential equation in (2) for a fixed value of  $\lambda$  and  $\mu$  is a distribution. Inferential approach using MCMC methods have been considered by both Robert *et al* (2004) and Griffin and Steel (2006b) when  $G^*(t)$  as a continuous-time volatility process and  $\mu$  is a discrete distribution with a finite number of atoms, which could be extended using some of the MCMC methods developed in this paper. Alternative generalisations are explored by Gander and Stephens (2005) who apply the ideas of Wolpert and Taqqu (2006) to stochastic volatility modelling. In particular they present alternative methods for defining a long-memory process. An interesting question for future research is “how well can the autocorrelation function be estimated from the data?”

The current work has concentrated on specifying known forms for the process of  $G_t$ . An alternative approach chooses a form for the background driving Lévy process. The properties of the marginal process can then be derived. This approach can lead to simpler sampling schemes in the stochastic volatility context. In particular James (2006) shows that the infinite dimensional process  $G^*(t)$  can be integrated from the process for particular choices. The use of these models to define a time-varying normalized random measure seems an interesting alternative approach to the ideas presented in this paper. The current work has focused on Dirichlet process marginals since these processes have dominated the Bayesian nonparametric literature. There has been recent work looking at other processes in applications. In particular Teh (2006) considers the use of a Pitman-Yor process for language modelling. The methods developed in this paper allow time-varying versions of such models and an interesting direction for future work would be the exploration of non-Dirichlet process time-varying models.

## References

- Aldous, D. (1983): “Exchangeability and related topics,” in *L’École d’été de probabilités de Saint-Flour, XIII-1983*, Springer-Verlage, pgs 1-198.par
- Barndorff-Nielsen, O. E. (2000): “Superposition of Ornstein-Uhlenbeck type processes,” *Theory of Probability and its Applications*, 45, 175-194.
- Barndorff-Nielsen, O. E. and Shephard, N. (2001): “Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics (with discussion),” *Journal of the Royal Statistical Society B*, 63, 167-241.

- Bertoin, J. (1996): "Lévy Processes," Cambridge University Press: Cambridge.
- Blackwell, D. and J. B. MacQueen (1973): "Ferguson distributions via Pólya urn schemes," *Annals of Statistics*, 1, 353-355.
- Caron, F., M. Davy, A. Doucet, E. Duflos and P. Vanheeghe (2006): "Bayesian Inference for Linear Dynamic Models with Dirichlet Process Mixtures," *Technical Report*.
- Carpenter, J., P. Clifford and P. Fearnhead (1999): "An improved particle filter for non-linear problems," *IEE proceedings - Radar, Sonar and Navigation*, 146, 2-7.
- Damien, P., J. Wakefield, and S. G. Walker (1999): "Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables," *Journal of the Royal Statistical Society*, B, 61, 331-344.
- Doucet, A., N. de Freitas and N. Gordon (2001): "Sequential Monte Carlo Methods in practice," Springer-Verlag.
- Escobar, M. D. and M. West (1995): "Bayesian Inference for Density Estimation," *Journal of the American Statistical Association*, 90, 577-588.
- Fearnhead, P. (2003): "Particle filters for mixture models with an unknown number of components," *Statistics and Computing*, 14, 11-21.
- Ferguson, T. S. (1973): "A Bayesian analysis of some nonparametric problems," *Annals of Statistics*, 1, 209-230.
- Ferguson, T. S. and Klass, M. J. (1972): "A representation of independent increment processes without Gaussian components," *Annals of Mathematical Statistics*, 43, 1634-43.
- Gander, M. P. S. and D. A. Stephens (2005): "Inference for Stochastic Volatility Models Driven by Lévy Processes," *Technical Report*, Imperial College London.
- Gander, M. P. S. and D. A. Stephens (2006): "Stochastic Volatility Modelling with General Marginal Distributions: Inference, Prediction and Model Selection For Option Pricing," *Journal of Statistical Planning and Inference*, to appear.
- Griffin, J. E. and M. F. J. Steel (2006a): "Order-based Dependent Dirichlet Processes," *Journal of the American Statistical Association*, 101, 179-194.
- Griffin, J. E. and M. F. J. Steel (2006b): "Inference with non-Gaussian OrnsteinUhlenbeck processes for stochastic volatility," *Journal of Econometrics*, 134, 605-644.
- Ishwaran, H. and L. F. James (2001): "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of the American Statistical Association*, 96, 161-173.
- Ishwaran, H. and James, L.F. (2003): "Some further developments for stick-breaking priors: finite and infinite clustering and classification," *Sankhya Series A*, 65, 577-592.



- James, L. F., A. Lijoi and I. Prünster (2005): “Bayesian inference via classes of normalized random measures,” *Technical Report*.
- James, L. F. (2006): “Laws and likelihoods for Ornstein-Uhlenbeck Gamma and other BNS OU Stochastic Volatility models with extensions,” *Technical Report*
- Kingman, J. F. C. (1975): “Random discrete distributions,” *Journal of the Royal Statistical Society B*, 37, 1-22.
- Lijoi, A., R. H. Mena, and I. Prünster (2005): “Hierarchical mixture modelling with normalized inverse-Gaussian priors,” *Journal of the American Statistical Association*, 100, 1278-1291./par
- Lijoi, A., R. H. Mena, and I. Prünster (2006): “Controlling the reinforcement in Bayesian nonparametric mixture models,” *Preprint*
- Liu, J. S. (1996): “Nonparametric hierarchical Bayes via sequential imputations,” *Annals of Statistics*, 24, 911-930.
- MacEachern, S. N., Clyde M. and Liu J. S. (1999): “Sequential importance sampling for nonparametric Bayes models: The next generation,” *Canadian Journal of Statistics*, 27, 251-267.
- Mallick, B. K. and S. G. Walker (1997): “Combining information from several experiments with nonparametric priors,” *Biometrika*, 84, 697-706.
- Mena. R. H. and Walker, S. G. : “Stationary Autoregressive Models via a Bayesian Nonparametric Approach,” *Journal of Time Series Analysis*, 26, 789-805.
- Müller, P. and Quintana, F. (2004): “Nonparametric Bayesian Data Analysis,” *Statistical Science*, 19, 95-110.
- Neal, R. M. (2003): “Slice sampling,” *Annals of Statistics*, 31, 705-767.
- Papaspiliopoulos, O. and G. O. Roberts (2005): “Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models,” *Technical Report*, University of Lancaster.
- Pennell, M.L. and Dunson, D.B. (2005): “Bayesian semiparametric dynamic frailty models for multiple event time data,” *Biometrics*, to appear.
- Pitman, J. and Yor, M. (1997): “The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator,” *Annals of Probability*, 25, 855-900.
- Regazzini, E., A. Lijoi and I. Prünster (2003): “Distributional results for means of normalized random measures with independent increments.” *The Annals of Statistics*, 31, 560-585.

- Roberts, G. O., O. Papaspiliopoulos and P. Dellaportas (2004): “Bayesian inference for Non-Gaussian Ornstein-Uhlenbeck Stochastic Volatility models,” *Journal of the Royal Statistical Society B*, 66, 369-393.
- Sato, K. (1999): “Lévy Processes and Infinitely Divisible Distributions,” CUP : Cambridge.
- Shephard, N. (2005): “Stochastic Volatility: Selected Reading,” OUP : Oxford.
- Taylor, S. J. (1986): “Modelling Financial Time Series” John Wiley and Sons : Chichester.
- Teh, Y. W. (2006): “A Hierarchical Bayesian Language Model based on Pitman-Yor process,” *Technical Report*.
- Tsay, R. (2005): “Analysis of Financial Time Series,” Wiley.
- Wolpert, R. L. and M. Taqqu (2006): “Fractional Ornstein-Uhlenbeck Lévy Processes and the Telecom Process: Upstairs and Downstairs,” *Signal Processing*, to appear.
- Zhu, X., Ghahramani, Z., and Lafferty, J. (2005): “Time-Sensitive Dirichlet Process Mixture Models,” Carnegie Mellon University Technical Report CMU-CALD-05-104.

## A Proofs

### A.1 Proof of theorem 1

We need to show that  $G_t^* = \sum_{i=1}^{\infty} I(\tau_i \leq t) \exp\{-\lambda(t - \tau_i)\} J_i \delta_{\theta_i}$  follows the appropriate unnormalized Lévy process. Consider a measurable set  $B$  then  $G_t^*(B) = \sum_{i=1}^{\infty} I(\tau_i \leq t) \exp\{-\lambda(t - \tau_i)\} J_i \delta_{\theta_i}(B)$  which is an independent thinning of  $G^*(\mathcal{S}) = \sum_{i=1}^{\infty} I(\tau_i \leq t) \exp\{-\lambda(t - \tau_i)\} J_i \delta_{\theta_i}(\mathcal{S})$  with thinning probability  $H(B)$ . Therefore it has the same distribution as  $\sum_{i=1}^{\infty} I(\tau_i^* \leq t) \exp\{-\lambda(t - \tau_i^*)\} J_i^*$  where  $(\tau^*, J^*)$  is a Poisson process  $\mathbb{R}^+ \times \mathbb{R}^+$  which has intensity  $W^+(J)H(B)\lambda$  which by the properties of the BNS construction has the marginal distribution  $W^+(J)H(B)$ , which is the marginal distribution of the unnormalized measure on  $B$  under the  $\text{NRM}(W^+, H)$ .

### A.2 Proof of result 1

$$\begin{aligned} \text{Cov}(G_t, G_{t+k}) &= \text{Cov} \left( \frac{G_t^*(B)}{G_t^*(B) + G_t^*(\mathcal{S} \setminus B)}, \frac{\exp\{-\lambda t\} G_t^*(B) + \sum_{i=1}^k \exp\{-\lambda t_i\} J_i \delta_{\theta_i}(B)}{\exp\{-\lambda t\} G_t^*(\mathcal{S}) + \exp\{-\lambda t\} G_t^*(\mathcal{S} \setminus B) + \sum_{i=1}^k \exp\{-\lambda t_i\} J_i} \right) \\ \text{Cov}(G_t, G_{t+k}) &= \text{Cov} \left( G_t(B), \frac{\exp\{-\lambda t\} G_t^*(\mathcal{S}) G_t(B) + H(B) \sum_{i=1}^k \exp\{-\lambda t_i\} J_i}{\exp\{-\lambda t\} G_t^*(\mathcal{S}) + \sum_{i=1}^k \exp\{-\lambda t_i\} J_i} \right) \\ \text{E}(G_t G_{t+k}) &= \text{E} \left( \frac{\exp\{-\lambda t\} G_t^*(\mathcal{S}) G_t(B)^2 + G_t(B) H(B) \sum_{i=1}^k \exp\{-\lambda t_i\} J_i}{\exp\{-\lambda t\} G_t^*(\mathcal{S}) + \sum_{i=1}^k \exp\{-\lambda t_i\} J_i} \right) \end{aligned}$$

$$\text{Cov}(G_t, G_{t+k}) = \mathbb{E} \left( \frac{\exp\{-\lambda t\} G_t^*(\mathcal{S}) G_t(B)^2 + G_t(B) H(B) \sum_{i=1}^k \exp\{-\lambda t_i\} J_i}{\exp\{-\lambda t\} G_t^*(\mathcal{S}) + \sum_{i=1}^k \exp\{-\lambda t_i\} J_i} \right) - H(B)^2$$

$$\text{Cov}(G_t, G_{t+k}) = \mathbb{E} \left( \frac{G_t^*(\mathcal{S}) G_t(B)^2 + G_t(B) H(B) \sum_{i=1}^k \exp\{-\lambda(t_i - t)\} J_i}{\exp\{-\lambda t\} G_t^*(\mathcal{S}) + \sum_{i=1}^k \exp\{-\lambda t_i\} J_i} \right) - H(B)^2$$

Asymptotically

$$\exp\{\lambda t\} \text{Cov}(G_t, G_{t+k}) = \mathbb{E} \left( \frac{G_t^*(\mathcal{S}) G_t(B)^2 + G_t(B) H(B) \exp\{\lambda t\} \sum_{i=1}^k \exp\{-\lambda t_i\} J_i}{\sum_{i=1}^k \exp\{-\lambda t_i\} J_i} \right) - H(B)^2$$

$$\exp\{\lambda t\} \text{Cov}(G_t, G_{t+k}) = \mathbb{E} \left( \frac{G_t^*(\mathcal{S}) G_t(B)^2}{\sum_{i=1}^k \exp\{-\lambda t_i\} J_i} \right) + \exp\{\lambda t\} H(B)^2 \exp\{\lambda t\} - H(B)^2$$

$$\exp\{\lambda t\} \text{Cov}(G_t, G_{t+k}) = \mathbb{E} \left( \frac{G_t^*(\mathcal{S}) G_t(B)^2}{\sum_{i=1}^k \exp\{-\lambda t_i\} J_i} \right)$$

The autocorrelation is asymptotically

$$\mathbb{E} \left( \frac{G_t^*(\mathcal{S}) G_t(B)^2}{\sum_{i=1}^k \exp\{-\lambda t_i\} J_i} \right) \exp\{-\lambda t\}$$

### A.3 Derivation of autocorrelation for OUDP

The general result

$$g(X, Y) = \frac{X}{Y}$$

$$\mathbb{E}[Z] = \frac{\mathbb{E}[X]}{\mathbb{E}[Y]} + \sigma_Y^2 \frac{\mathbb{E}[X]}{\mathbb{E}[Y]^3} - \sigma_{XY} \frac{1}{\mathbb{E}[Y]^2}$$

then  $\mathbb{E}[X] = \exp\{-\lambda\Delta\} \mu$  and  $\mathbb{E}[Y] = \mu$  by stationarity where  $\mu = \mathbb{E}[G(\mathcal{X})]$ .  $\text{Cov}(X, Y) = \text{Cov}(X, X) + \text{Cov}(X, Y - X) = \mathbf{V}(X) = \exp\{-2\lambda\Delta\} \sigma^2$  where  $\sigma^2 = \mathbf{V}[G^*(\mathcal{S})]$ .

$$\mathbb{E}[Z] = \exp\{-\lambda\Delta\} \left[ 1 + \frac{\sigma^2}{\mu^2} (1 - \exp\{-\lambda\Delta\}) \right]$$

for the Dirichlet process  $\frac{\sigma^2}{\mu^2} = \frac{1}{M}$