RESEARCH SUPPORT
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

DECEMBER 1996

A FULL BAYESIAN NONPARAMETRIC ANALYSIS INVOLVING A NEUTRAL TO THE RIGHT PROCESS

WORKING PAPER #9612-41

STEPHEN WALKER
IMPERIAL COLLEGE, LONDON, UK

AND

PAUL DAMIEN
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

A Full Bayesian Nonparametric Analysis Involving a Neutral to the Right Process

Stephen Walker[1]
and
Paul Damien[2]

[1] Department of Mathematics. Imperial College. 180 Queen's Gate. London SW7 2BZ

[2] Department of Statistics and Management Science. School of Business. University of Michigan. Ann Arbor. 48109-1234, USA.

SUMMARY

Implementation of a full Bayesian nonparametric analysis involving *neutral to the right* processes (apart from the special case of the Dirichlet process) have been difficult for two reasons: first, the posterior distributions are complex and therefore only Bayes estimates (posterior expectations) have previously been presented: secondly, it is difficult to obtain an interpretation for the parameters of a *neutral to the right* process.

In this paper we extend Ferguson and Phadia (1979) by presenting a general method for specifying the prior mean and variance of a *neutral to the right* process, providing the interpretation of the parameters. Additionally, we provide the basis for a full Bayesian analysis, via simulation, from the posterior process using a hybrid of new algorithms that is applicable to a large class of *neutral to the right* process (Ferguson and Phadia only provide posterior means). The ideas are exemplified through illustrative analyses.

Keywords: Hierarchical models. Lévy process, Neutral to the right process. Beta-Stacy process, Infinitely divisible law, Latent variables. Gibbs sampler.

1 Introduction

A large section of the Bayesian nonparametric literature is concerned with the construction of prior distributions on the space of probability distributions on $(0, \infty)$. One particular application which has received a great deal of attention includes survival analysis. An important and rich class of prior distribution is the *neutral to the right* process (Doksum, 1974; Ferguson, 1974) which includes the well known Dirichlet process (Ferguson, 1973).

DEFINITION (Ferguson, 1974). A random distribution function $F(t)$ on $(0, \infty)$ is neutral to the right (from now on NTR) if for every m and $0 < t_1 < \ldots < t_m$ there exist independent random variables $V_1, \ldots, V_m$ such that

$$\mathcal{L}\Big(1 - F(t_1), \ldots, 1 - F(t_m)\Big) = \mathcal{L}\Big(V_1, V_1 V_2, \ldots, \Pi_{j=1}^m V_j\Big).$$

If F is NTR then $Z(t) = -\log(1 - F(t))$ has independent increments. The converse is also true: if $Z(t)$ is an independent increments process, nondecreasing almost surely (a.s.), right continuous a.s., $Z(0) = 0$ a.s. and $\lim_{t \to \infty} Z(t) = +\infty$ a.s. then $F(t) = 1 - \exp(-Z(t))$ is NTR. For the process $Z(.)$ there exist at most countably many fixed points of discontinuity at $t_1, t_2, \ldots$ with jumps $J_1, J_2, \ldots$ (independent) having densities $f_{t_1}, f_{t_2}, \ldots$. Then $Z_c(t) = Z(t) - \sum_{t_j \leq t} J_j$ has no fixed points of discontinuity and therefore has a Lévy representation with log-Laplace transform given by

$$-\log E\exp\Big(-\phi Z_c(t)\Big) = \int_0^\infty \Big(1 - \exp(-\phi z)\Big) dN_t(z),$$

where $N_t(.)$, a Lévy measure, satisfies: $N_t(B)$ is nondecreasing and continuous for every Borel set B; $N_t(.)$ is a measure on $(0, \infty)$ for each t and $\int_{(0,\infty)} z(1+z)^{-1} dN_t(z) < \infty$. In short we have

$$F(t) = 1 - \exp\{-Z(t)\},$$

with Z an independent increments (Lévy) process.

There has only been limited application involving NTR process priors in the statistics literature (besides the well known Dirichlet process). The reason for this is the difficulty in obtaining anything more than an expression for the posterior expectation F (Ferguson and Phadia, 1979, Section

6). Applications with a full Bayesian analysis have appeared involving the extended gamma process (Laud et al., 1996) and the beta process (Damien et al., 1996), neither of which are NTR processes.

A primary aim of this paper is to construct a simple and easy-to-interpret method for specifying the prior mean and variance of a random distribution function. We develop such a method for NTR processes in Section 2. A second aim is to present a new hybrid sampling based approach for making inferences on the posterior NTR process and in particular we provide an algorithm for sampling $F[a, b)$ from the posterior distribution.

2 Prior and Posterior Distributions

Prior Specifications

Central to any nonparametric Bayesian analysis is the ability to use available prior information to obtain expressions for the mean and variance of the random distribution function, $F(t)$. For the Dirichlet process it is only the mean of F, say F_0, which can be modelled arbitrarily since, given F_0, then

$$\text{var}F(t) = \frac{F_0(t)\{1 - F_0(t)\}}{c+1},$$

for $c > 0$. The parameter of the Dirichlet process is then defined as $\alpha(.) = cF_0(.)$. Ferguson and Phadia (1979) point out that for other NTR processes, the interpretation of the prior parameters of the process tends to be nebulous. (We return to this later.)

Noting from the previous section that any NTR process is defined by a Lévy process, our goal is to describe the prior mean and variance of the random distribution function in terms of the Lévy measure that characterises the process.

Consider then the expressions for $-\log\{ES(t)\}$ and $-\log\{E[S^2(t)]\}$ where $S(t) = 1 - F(t)$. Using the Lévy representation in Section 1 (with no fixed points of discontinuity) it follows that

$$\mu(t) = -\log\{ES(t)\} = \int_0^\infty (1 - e^{-z})dN_t(z)$$

and

$$\lambda(t) = -\log\{E[S^2(t)]\} = \int_0^\infty (1 - e^{-2z})dN_t(z).$$

Note that it is necessary for

$$0 < \mu(t) < \lambda(t) < 2\mu(t)$$

which derives from the condition $\{ES(t)\}^2 < E[S^2(t)] < ES(t)$.

We want to find a Lévy measure $N_t(.)$ which satisfies these two equations and in particular we consider Lévy measures of the type

$$dN_t(z) = dz\{1 - e^{-z}\}^{-1} \int_0^t \exp\Big(-z\beta(s)\Big)d\alpha(s).$$

where $\beta(.)$ is a nonnegative function and $\alpha(.)$ a measure. This type of Lévy measure characterises beta-Stacy processes which are considered in greater detail in Walker and Muliere (1995). Our motivation for working with the beta-Stacy process stems from the fact that this process encapsulates virtually all of the NTR processes mentioned in the literature. As examples, the Dirichlet process arises when $\alpha(.)$ is a finite measure and $\beta(t) = \alpha(t,\infty)$. The *simple homogeneous* process (Ferguson and Phadia, 1979) is obtained when β is constant.

LEMMA. There exists $\alpha(.)$ and $\beta(.)$ such that

$$\mu(t) = \int_0^\infty \int_0^t \exp\Big(-z\beta(s)\Big)d\alpha(s)dz$$

and

$$\lambda(t) = \int_0^\infty \int_0^t \Big\{\frac{1 - e^{-2z}}{1 - e^{-z}}\Big\}\exp\Big(-z\beta(s)\Big)d\alpha(s)dz.$$

PROOF. It is obvious that the first of these conditions is satisfied when

$$\int_0^t d\alpha(s)/\beta(s) = \mu(t),$$

that is, when

$$d\alpha(t) = \beta(t)d\mu(t).$$

The second condition becomes, using the transformation $y = 1 - \exp(-z)$,

$$\lambda(t) = \int_0^t \Big\{2 - 1/(1 + \beta(s))\Big\}d\mu(s)$$

giving

$$d\lambda(t)/d\mu(t) = 2 - 1/(1+\beta(t)),$$

concluding the proof.

In particular the result allows us to obtain an interpretation for the parameters of a simple homogeneous process, not resolved by Ferguson and Phadia. If β is constant then we have $\mu(t) = \alpha(t)/\beta$ and $\lambda(t) = c\alpha(t)/\beta$ where $c = (1+2\beta)/(1+\beta)$.

Example 1.

If it is required that $ES(t) = \exp\{-at\}$, $a > 0$, and $E[S^2(t)] = \exp\{-b(t)\}$, $b(0) = 0$, b nondecreasing and $b(t) \to \infty$ as $t \to \infty$, then

$$\beta(t) = 1/(2 - a^{-1}db/dt) - 1$$

and

$$d\alpha(t) = a\beta(t)dt.$$

Note that here we need the condition $0 < a < db/dt < 2a$ which corresponds to the necessary condition $0 < \mu(t) < \lambda(t) < 2\mu(t)$. (Note also that the two conditions are in general impossible for the Dirichlet process which has the extra constraint of $\beta(t) = \alpha(t,\infty)$).

Example 2.

Suppose we wish to center, up to and including second moments, the nonparametric model on the parametric Bayesian model given by $S(t) = \exp(-at)$ with $a \sim$ gamma(p,q). Then we would have $\mu(t) = p\log(1+t/q)$ and $\lambda(t) = p\log(1+2t/q)$ giving

$$\beta(t) = q/(2t) \text{ and } d\alpha(t) = pqdt/\{2t(q+t)\}.$$

Firstly, since $\beta(t) \neq \alpha(t,\infty)$, the Dirichlet process is not applicable. Secondly, one could use different gamma(p,q) densities along the time axis. Thirdly, some other parametric model could be chosen resulting, of course, in different forms for $\alpha(t)$ and $\beta(t)$. All these allow for greater modelling flexibility.

Example 3.

For a two-parameter Weibull 'centering' we have $ES(t) = \exp\{-(at)^b\}$ leading to $\mu(t) = (at)^b$. Since we require $\mu < \lambda < 2\mu$ it seems appropriate to take $\lambda(t) = c(at)^b$ for some $c \in (1,2)$. It is now easy to see that $\beta(t) = \beta = (c-1)/(2-c)$ and $d\alpha(t) = \beta a^b b t^{b-1} dt$. Here we have a simple homogeneous process. We could consider uncertainty in c via a prior distribtuion on $(1,2)$ which would result in a mixture of simple homogeneous processes.

Thus, our method allows one to specify the mean and variance for F, which is not, in general, possible for the Dirichlet process, or other well known NTR processes. Ferguson and Phadia (1979, Remarks 1 and 2) discuss the problems associated in interpreting the mean and variance of $F(t)$. They rely on such ideas as *prior sample size* and *strength of belief* when considering uncertainty in the centering of F, and conclude they are not satisfactory. In fact, as a consequence of the Lemma, their assertion that $c = \alpha(0,\infty)$ represents a prior sample size, which tends to become noninformative as $c \to 0$ is questionable. For if $\beta(t) = \alpha(t,\infty) = cF_0(t,\infty)$ then $d\mu(t) = dF_0(t)/F_0(t,\infty)$ and as $c \to 0$, it is clear from the Lemma, that $\lambda(t) \to \mu(t)$ implying that $E[S^2(t)] \to ES(t)$. (See, also, Sethuraman and Tiwari (1982) for further insight on this point.)

We next provide details of the posterior distribution of a NTR process.

Posterior Distributions

THEOREM 1 (Ferguson, 1974). If F is NTR and $X_1,\ldots,X_n$ is a sample from F, including the possibility of right censored samples (where X_i represents the censoring time if applicable), then the posterior distribution of F given $X_1,\ldots,X_n$ is NTR.

The prior distribution for $Z(.)$ is characterised by

$$M = \left\{t_1, t_2, \ldots\right\}, \quad \left\{f_{t_1}, f_{t_2}, \ldots\right\},$$

the set of fixed points of discontinuity with corresponding densities for the jumps, and $N_t(.)$, the Levy measure for the part of the process without fixed points of discontinuity. The posterior distribution is now given for a single observation X. The case for n observations can then be obtained by repeated application. In the following, we assume the Lévy measure to be of the type $dN_t(z) = dz \int_{(0,t]} K(z,s)ds$ (the beta-Stacy process with parameters $\alpha(.)$ and $\beta(.)$ arises when $K(z,s)ds = (1-e^{-z})^{-1}\exp\{-z\beta(s)\}d\alpha(s)$).

THEOREM 2(Ferguson, 1974; Ferguson and Phadia, 1979). Let F be NTR and let X be a random sample from F.

(i) Given $X > x$ the posterior parameters (denoted by a $*$) are $M^* = M$,

$$f^*_{t_j}(z) = \begin{cases} c.e^{-z} f_{t_j}(z) & \text{if } t_j \leq x \\ f_{t_j}(z) & \text{if } t_j > x. \end{cases}$$

and $K^*(z,s) = \exp\{-zI(x \geq s)\}K(z,s)$ (here c is the normalising constant).

(ii) Given $X = x \in M$ the posterior parameters are $M^* = M$,

$$f^*_{t_j}(z) = \begin{cases} c.e^{-z} f_{t_j}(z) & \text{if } t_j < x \\ c.(1-e^{-z}) f_{t_j}(z) & \text{if } t_j = x \\ f_{t_j}(z) & \text{if } t_j > x \end{cases}$$

and, again, $K^*(z,s) = \exp\{-zI(x \geq s)\}K(z,s)$.

(iii) Given $X = x \notin M$ the posterior parameters are $M^* = M \cup \{x\}$, with $f_x(z) = c.(1-e^{-z})K(z,x)$,

$$f^*_{t_j}(z) = \begin{cases} c.e^{-z} f_{t_j}(z) & \text{if } t_j < x \\ f_{t_j}(z) & \text{if } t_j > x \end{cases}$$

and, again, $K^*(z,s) = \exp\{-zI(x \geq s)\}K(z,s)$.

With this general characterisation of a posterior NTR process, a procedure for sampling the posterior process is developed in the next section.

3 Simulating the Posterior Process

We are now concerned with simulating from the posterior distribution of $F[a,b)$. Since $F(t) = 1 - \exp\{-Z(t)\}$ we have

$$F[a,b) = \exp\Big(-Z(0,a)\Big)\Big\{1 - \exp\Big(-Z[a,b)\Big)\Big\}$$

and note that $Z(0,a)$ and $Z[a,b)$ are independent. Therefore we only need to consider sampling random variables of the type $Z[a,b)$. This will involve sampling from densities $\{f_{t_j}^* : t_j \in [a,b)\}$, corresponding to the jumps in the process, and secondly, the continuous component, the random variable $Z_c^*[a,b)$.

The type of densities corresponding to points in M^*, the jumps, are defined, up to a constant of proportionality, by

$$f^*(z) \propto \Big(1-\exp(-z)\Big)^{\lambda} \exp\Big(-\mu z\Big) f(z), \quad \lambda, \mu \geq 0 \text{ (integers)}$$

and

$$f^*(z) \propto \Big(1-\exp(-z)\Big)^{\lambda} K(z,x), \quad \lambda > 0 \text{ (integer)}$$

where the subscripts t_j have been omitted.

Simulating the jump component

The second of these two types of densities for the random variable corresponding to the jump in the process will most likely be a member of the SD family of densities (Damien et al., 1995) and algorithms for sampling from such densities is given in Walker (1995) and Damien and Walker (1996).

However, the first of these two types can be done by introducing latent variables, whose densities are of the form described in Damien and Walker (1996), and setting up the conditional distributions required to implement a Gibbs sampler (Smith and Roberts, 1993).

Define the joint density of z, $u = (u_1, ..., u_\lambda)$ and v by

$$f(z,u,v) \propto e^{-uv} I(v > z) \Big\{\Pi_{l=1}^{\lambda} e^{-u_l} I(u_l < z)\Big\} f(z).$$

Clearly the marginal density of z is as required. The full conditional densities for implementing Gibbs sampling are given by

$$f(u_l | u_{-l}, v, z) \propto e^{-u_l} I(u_l < z), \quad l = 1, ..., \lambda,$$

$$f(v | u, z) \propto e^{-uv} I(v > z)$$

and

$$f(z | u, v) \propto f(z) I\Big(z \in (\max_l\{u_l\}, v)\Big).$$

So, provided it is possible to sample from $f(z)$, this algorithm is easy to implement. In the next section, we show that, for all the well-known NTR processes, simulating from $f^*(z)$ is trivial.

Simulating the continuous component

The random variable $Z = Z_c^*[a,b)$ corresponding to the continuous component is *infinitely divisible* (id) with log-Laplace transform given by

$$-\log E\exp(-\phi Z) = \int_0^\infty \Big(1-\exp(-\phi z)\Big) \int_a^b \exp\Big(-z\sum_{i=1}^n I(X_i \geq s)\Big) K(z,s) ds dz.$$

There are two algorithms for simulating an id random variable with a given Laplace transform: Damien et al. (1995), which relies on the characterisation of an id random variable as the limit of sequences of sums of Poisson types, and Walker (1996), which uses the fact

$$\mathcal{L}Z = \mathcal{L}\Big(\int_0^\infty z dP(z)\Big),$$

where $P(.)$ is a Poisson process with intensity measure $dz \int_{[a,b)} K^*(z,s)ds$. The Walker (1996) method is described in some detail in the Appendix since it is an improvement to the Damien et al. algorithm, and hence was used to analyse the example data in the next section.

In summary we have:
1. Specify the prior mean and variance of a neutral to the right process, $F(t)$ via the mean and variance of its corresponding Lévy process using, say, a hierarchical parametric model.
2. The posterior process is given by Theorem 2. This process is comprised of a jump component and a continuous component. The jump components are random variables that can be simulated via latent variable substitutions, and, if necessary, a Gibbs sampler.
3. The continuous component for any NTR process, always, has an infinitely divisible distribution. Simulating from such a distribution is described in the Appendix.

4 Illustrative Analysis

Walker and Muliere (1995) developed the beta-Stacy process and showed that it is a generalisation of the Dirichlet process. Unlike the Dirichlet process, the beta-Stacy process is conjugate to arbitrarily right censored data which commonly arise in survival analysis. Moreover. Walker and Muliere prove that if the prior process is Dirichlet. then the posterior process is beta-Stacy if the data include right censored observations. Here these results are used to model exact and right-censored data.

The Kaplan-Meier (1958) data was chosen because Ferguson and Phadia (1979) examined these data previously. thus offering a basis for comparison. The data consist of exact observations at 0.8. 3.1. 5.4 and 9.2 months and right censored data at 1.0. 2.7. 7.0 and 12.1 months.

Example 1

We follow Ferguson and Phadia and take $\beta(s) = \exp(-0.1s)$ and $d\alpha(s) = 0.1\exp(-0.1s)ds$. The prior therefore is a Dirichlet process but. with the inclusion of censored observations within the data set. the posterior process is not Dirichlet but a beta-Stacy process. We take $M = \emptyset$ a priori: i.e.. there are no jumps in the prior process. Then, given the data. the posterior parameters are given by

$$M^* = \{X_i : \delta_i = 1\}.$$

where $\delta_i = 1$ indicates that X_i is an exact observation.

$$f_x^*(z) \propto \Big(1 - \exp(-z)\Big)^{N\{x\}-1} \exp\Big(- z\{\beta(x) + Y(x) - N\{x\}\}\Big).$$

where $x \in M^*$. $N\{x\} = \sum_{X_i = x} \delta_i$ and $Y(x) = \sum_i I(X_i \geq x)$.

A striking consequence of the beta-Stacy process model in this first illustration is to note that if Z. the jump random variable has density $f_x^*(.)$ then $Z =_d -\log(1 - Y)$ where

$$Y \sim \text{beta}\Big(N\{x\}, \beta(x) + Y(x) - N\{x\}\Big).$$

however. if $N\{x\} = 1$ then Z has an exponential density with mean value $1/(\beta(x) + Y(x) - 1)$. From a simulation perspective, this obviates the use of both the latent variable steps required for simulating the jump random variable. and a Gibbs sampler.

The continuous component of the posterior process has Lévy measure given by

$$K^*(z, s)ds = \Big(1 - \exp(-z)\Big)^{-1} \exp\Big(- z\{\beta(s) + Y(s)\}\Big) d\alpha(s).$$

so note therefore that the posterior process is also a beta-Stacy process. We obtain samples from the posterior distribution of $F(0,1)$. This will involve sampling $Z_\cdot^*(0,0.8)$, $Z_\cdot^*[0.8,1)$, which is done using the algorithm of Walker (1996), and sampling $J_{0.8}$ from the density $f_{0.8}^*(\cdot)$, the exponential density with mean $1/\{\exp(-0.08)+7\}$. A required sample from the posterior distribution of $F(0,1)$ is then given by $1-\exp\{-Z_\cdot^*(0,0.8)-Z_\cdot^*[0.8,1)-J_{0.8}\}$.

We collected 1000 samples from the posterior and the resulting histogram representation with kernel density estimate is given in Figure 1. The mean value is given by 0.12 which is the (exact) value obtained by Ferguson and Phadia.

Example 2

We also reanalyse the data set using the 'base' hierarchical centering model given in Section 2. Here, with $p=q=1$, in an attempt to be noninformative with the gamma parametric prior, we have $\beta(t)=1/(2t)$ and $d\alpha(t)=dt/\{2t(1+t)\}$. Again, we collected 1000 samples from the posterior and the resulting histogram representation with kernel density estimate is given in Figure 2.

We compare this with the posterior distribution of the Bayes model given by $F(0,1)=1-\exp(-a)$ with $a\sim\text{gamma}(1+4,1+41.3)$. 1000 samples from the posterior were collected and the resulting histogram representation with kernel density estimate is given in Figure 3. The distributions are fundamentally different even though mean values are identical at 0.11.

5 Conclusions

Neutral to the right processes form a wide class of prior distributions for the space of distribution functions. An intuitive way for specifying prior expectations of a random distribution function via hierarchical parametric models was developed. From a practical perspective, this method enables the statistician to try different prior models for the mean and variance of the Lévy measure that characterises all neutral to the right processes. The form for the posterior process was stated using a general and well-defined Lévy measure. A full Bayesian solution was then obtained by simulating the posterior process using a hybrid of sampling methods, and which was exemplified with exact and censored survival data.

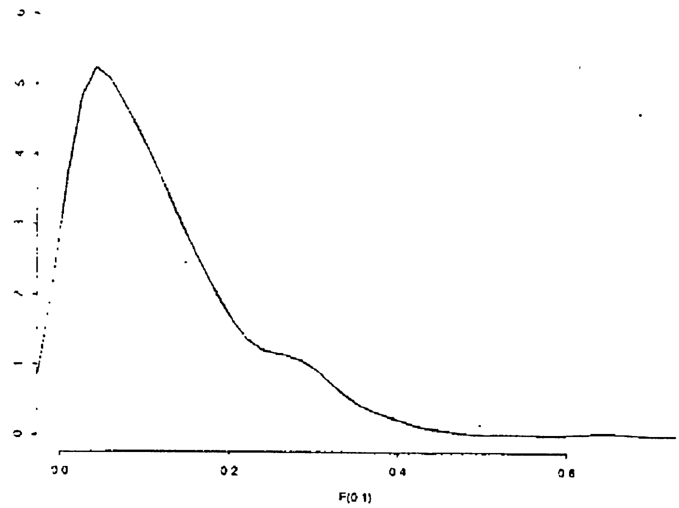

Figure 1: Histogram representation with kernel density estimate of posterior density of F(0.1).

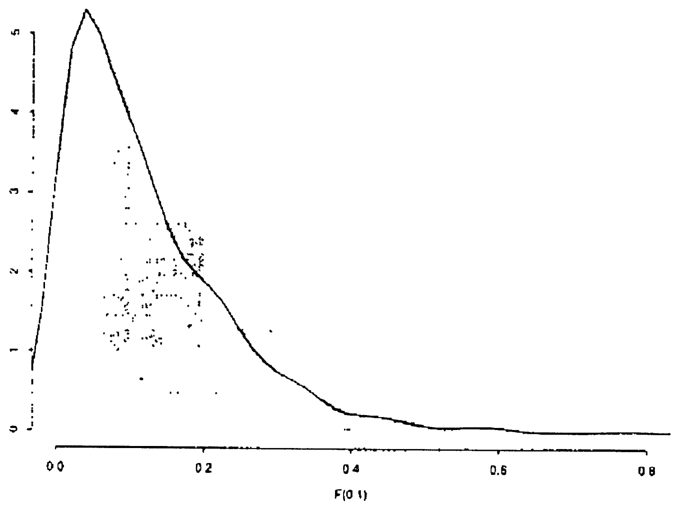

Figure 2: Histogram representation with kernel density estimate of posterior density of F(0.1).

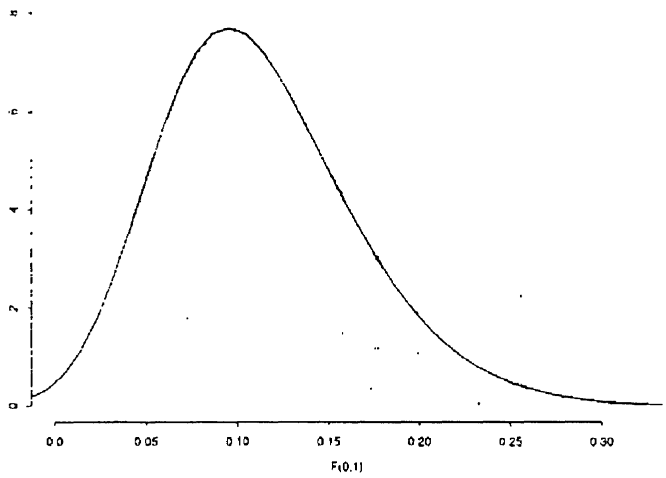

Figure 3: Histogram representation with kernel density estimate of posterior density of F(0.1).

Acknowledgements

The authors are grateful to Adrian Smith for helpful discussions and comments on an earlier draft of the paper.

References

Damien,P.,Laud,P.W. and Smith,A.F.M. (1995). Random variate generation approximating infinitely divisible distributions with application to Bayesian inference. *Journal of the Royal Statistical Society Series B* **57**, 547-564.

Damien,P.,Laud,P.W. and Smith,A.F.M. (1996). Implementation of Bayesian nonparametric inference based on beta processes. *Scandinavian Journal of Statistics* **23**, 27-36.

Damien,P. and Walker,S.G. (1996). Sampling probability densities via uniform random variables and a Gibbs sampler. *Submitted for publication.*

Doksum,K.A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *The Annals of Probability* **2**, 183-201.

Ferguson,T.S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209-230.

Ferguson,T.S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics* **2**, 615-629.

Ferguson,T.S. and Phadia,E.G. (1979). Prior distributions on spaces of probability measures. *The Annals of Statistics* **2**, 615-629.

Kaplan,E.L. and Meier,P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457-481.

Laud,P.W.,Smith,A.F.M. and Damien,P. (1996). Monte Carlo methods approximating a posterior hazard rate process. *Statistics and Computing* **6**, 77-84.

Sethuraman,J. and Tiwari,R. (1982). Convergence of Dirichlet measures and the interpretation of their parameter. In *Proceedings of the Third Purdue Symposium on Statistical Decision Theory and Related Topics*, Gupta,S.S. and Berger,J. (eds.), pp.305-315. Academic Press, NY.

Smith,A.F.M. and Roberts,G.O. (1993). Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B* **55**, 3-23.

Tierney,L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics* **22**, 1701-1762.

Walker,S.G. (1995). Generating random variates from D-distributions via substitution sampling. *Statistics and Computing* **5**, 311-315.

Walker,S.G. (1996). Random variate generation from an infinitely divisible distribution via Gibbs sampling. *Submitted for publication.*

Walker,S.G. and Muliere,P. (1995). Beta-Stacy processes and a generalisation of the Polya-urn scheme. Under revision for *The Annals of Statistics*.

Appendix

Here we consider sampling an infinitely divisible variable Z with log-Laplace transform given by

$$\log E\exp(-\phi Z) = \int_0^\infty \left(e^{-z\phi} - 1\right) dN_\Delta(z),$$

where

$$dN_\Delta(z) = \frac{dz}{1 - e^{-z}} \int_\Delta \exp\Big(-z\beta(s)\Big) d\alpha(s).$$

Recall $\mathcal{L}Z = \mathcal{L}(\int_{(0,\infty)} z dP(z))$ where $P(.)$ is a Poisson process with intensity measure $dN_\Delta(.)$. For a fixed $\epsilon > 0$ let $Z = X_\epsilon + Y_\epsilon$ (X_ϵ, Y_ϵ independent) where X_ϵ is the sum of the jumps (times) of $P(.)$ in (ϵ, ∞) and Y_ϵ is the sum of the jumps (times) in $(0, \epsilon]$ and has log-Laplace transform given by $\int_{(0,\epsilon]}\{\exp(-\phi z) - 1\} dN_\Delta(z)$. The aim is to approximate a random draw Z by sampling an X_ϵ. The error variable Y_ϵ has mean $EY_\epsilon = \int_{(0,\epsilon]} z dN_\Delta(z)$ and variance $\text{var}Y_\epsilon = \int_{(0,\epsilon]} z^2 dN_\Delta(z)$. Now ϵ can be taken as small as one wishes and it is not difficult to show, for very small ϵ, that $EY_\epsilon < \epsilon$ and $\text{var}Y_\epsilon < \epsilon^2/2$.

Let $\lambda_\epsilon = N_\Delta(\epsilon, \infty) < \infty$ and take $\nu \sim \text{Poisson}(\lambda_\epsilon)$. Here ν is the (random) number of jumps in (ϵ, ∞) and therefore the jumps $\tau_1, \ldots, \tau_\nu$ are taken iid from $G_\epsilon(s) = \lambda_\epsilon^{-1} N_\Delta(\epsilon, s]$, for $s \in (\epsilon, \infty)$. Then X_ϵ is given by $X_\epsilon = \sum_{k=1}^{\nu} \tau_k$. To implement the algorithm therefore it is required to (a) simulate from $G_\epsilon(.)$ and (b) calculate λ_ϵ. For the beta-Stacy process the density corresponding to G_ϵ is given, up to a constant of proportionality, by

$$g_\epsilon(z) \propto \Big\{1 - \exp(-z)\Big\}^{-1} I\Big(z > \epsilon\Big) \int_\Delta \exp\Big(-z\beta(s)\Big) d\alpha(s).$$

(a). To sample from g_ϵ define the joint density function of z, s and u by

$$f(z, s, u) \propto I\Big(u < \{1 - \exp(-z)\}^{-1}\Big) \exp\Big(-z\beta(s)\Big) \alpha(s) I\Big(z > \epsilon, s \in \Delta\Big).$$

where u is a random variable defined on the interval $(0, \{1 - \exp(-\epsilon)\}^{-1})$, s is a random variable defined on $(0, \infty)$ and $\alpha(s)ds = d\alpha(s)$. Clearly the marginal distribution for z is given by g_ϵ.

We now describe a hybrid Gibbs/Metropolis sampling algorithm (Tierney, 1994) for obtaining random variates $\{z^{(l)}\}$ from g_ϵ. The algorithm involves simulating a Markov chain $\{z^{(l)}, s^{(l)}, u^{(l)}\}$ such that as $l \to \infty$ then $z^{(l)}$ can be taken as a random draw from $g_\epsilon(z)$. The algorithm is given, with starting values $\{z^{(1)}, s^{(1)}, u^{(1)}\}$, by

(i) take $z^{(l+1)}$ from the exponential distribution with mean value $1/\beta(s^{(l)})$ restricted to the interval $(\epsilon, -\log\{1 - 1/u^{(l)}\})$ if $u^{(l)} > 1$ and the interval (ϵ, ∞) if $u^{(l)} \leq 1$;

(ii) take $u^{(l+1)}$ from the uniform distribution on the interval

$$\left(0, \{1 - \exp(-z^{(l+1)})\}^{-1}\right)$$

and

(iii) (for example) take s^* from $\alpha(.)/\alpha(\Delta)$ on Δ and ξ from the uniform distibution on the interval $(0, 1)$. If

$$\xi < \min\left\{1, \exp\left(- z^{(l+1)}[\beta(s^*) - \beta(s^{(l)})]\right)\right\}$$

then $s^{(l+1)} = s^*$ else $s^{(l+1)} = s^{(l)}$.

(b). To calculate λ_ϵ define the joint density function of z and s by

$$f(z, s) \propto \exp\left(- z\beta(s)\right)\alpha(s)I\left(z > \epsilon, s \in \Delta\right)$$

and let the right-hand side of this expression be $h(z, s)$. Then

$$\lambda_\epsilon = \int_0^\infty \int_0^\infty \{1 - \exp(-z)\}^{-1} h(z, s) dz ds$$

so

$$\lambda_\epsilon = E_f\left[\{1 - \exp(-z)\}^{-1}\right] \times \int_0^\infty \int_0^\infty h(z, s) dz ds.$$

Now $E_f[\{1 - \exp(-z)\}^{-1}]$ can be obtained from $L^{-1} \sum_{l=1}^{L} \{1 - \exp(-z^{(l)})\}^{-1}$ where $\{z^{(l)} : l = 1, 2, \ldots\}$ are obtained via a Gibbs sampling algorithm which involves sampling from $f(z|s)$, which is the exponential distribution with mean value $1/\beta(s)$ restricted to the interval (ϵ, ∞), and sampling from $f(s|z)$, which is done as in (iii) from (a). Finally

$$\int_0^\infty \int_0^\infty h(z,s)dzds = \int_\Delta \exp\Big(-\epsilon\beta(s)\Big)d\alpha(s)/\beta(s),$$

so Monte Carlo methods, if an analytic expression is not available, can be used to calculate this term (it should also be adequate to take this to be $\int_\Delta d\alpha(s)/\beta(s) - \epsilon\alpha(\Delta)$).