

A Nonparametric Bayesian Approach to Inverse Problems

ROBERT L. WOLPERT

Duke University, USA

wolpert@stat.duke.edu

KATJA ICKSTADT

Universität Dortmund, Germany

ickstadt@statistik.uni-dortmund.de

MARTIN B. HANSEN

Aalborg Universitet, Denmark

mbh@math.auc.dk

SUMMARY

We propose a new method for making inference about an unknown measure $\Gamma(d\lambda)$ upon observing some values of the Fredholm integral $g(\omega) = \int k(\omega, \lambda)\Gamma(d\lambda)$ of a known kernel $k(\omega, \lambda)$, using Lévy random fields as Bayesian prior distributions for modeling uncertainty about $\Gamma(d\lambda)$. Inference is based on simulation-based MCMC methods. The method is illustrated with a problem in polymer chemistry.

Keywords: GAMMA PROCESS; LÉVY PROCESS; POLYMER; RANDOM FIELD; REVERSIBLE JUMP MCMC; RHEOLOGY.

1. INTRODUCTION

Fredholm (1900) initiated the formal study of integral equations of the first kind, in which we try to impute an unknown measure $\Gamma(d\lambda)$ from finitely many observed values of the integrals

$$G(\omega_i) = \int_{\Lambda} k(\omega_i, \lambda)\Gamma(d\lambda) \quad (1)$$

of a known kernel $k(\omega, \lambda)$. The problem is difficult in part because the integral operator $K : \Gamma \mapsto G$ is *smoothing*, making the “inverse problem” $K^{-1} : G \mapsto \Gamma$ ill-posed in the sense that small changes in G may be associated with large changes in Γ .

The most common approaches to solving Equation (1) for the unknown Γ begin by approximating this infinite-dimensional continuous problem with the finite-dimensional discrete one

$$G_i = \sum_{j \in J} k_{ij} \Gamma_j. \quad (2)$$

The approximate solution of Equation (2) is available from the normal equations as $\Gamma \approx [K'K]^{-1}K'G$ (see Kirsch (1996) for a discussion of such discretization methods and of the numerical obstacles that arise in trying to solve the resulting linear systems). Typically the matrix $[K'K]$ is ill-conditioned and so Equation (2) remains ill-posed— if J is small there are no solutions, while if J is sufficiently large there are infinitely many which differ wildly. Commonly this is addressed through the inclusion of some form of “roughness penalty” (see for example the *method of regularization* of Tikhonov (1963)).

2. A NEW BAYESIAN NONPARAMETRIC APPROACH

Our approach is to treat the solution of Equation (1) as a Bayesian statistical inference problem, that of estimating the uncertain element Γ of the space $\mathcal{M}_+(\Lambda)$ of positive measures on a set Λ upon observing, perhaps with error, the quantities $G_i \approx G(\omega_i) \in \mathcal{G}$ at some finite set of points $\{\omega_i\}_{i \in I} \subset \Omega$. To complete the Bayesian model specification we must select a *prior distribution* $\pi(d\Gamma)$ on $\mathcal{M}_+(\Lambda)$, making Γ a random measure, and we must select a *measurement error model* for G_i given $g_i \equiv G(\omega_i)$, leading to a likelihood function $L(\Gamma)$.

In many applications (including ours in Section (3) below) localization arguments suggest that the uncertain positive measures $\Gamma(A)$ and $\Gamma(B)$ assigned to disjoint sets $A, B \subset \Lambda$ may be regarded as stochastically independent *a priori*. Under mild regularity conditions this leads to a Lévy-Khinchine-like representation for stochastic integrals $\Gamma[\phi] \equiv \int_{\Lambda} \phi(\lambda) \Gamma(d\lambda)$ of measurable functions $\phi : \Lambda \rightarrow \mathbb{R}$ of the form

$$\log \mathbb{E} \left[e^{i\Gamma[\phi]} \right] = \iint_{\mathbb{R}_+ \times \Lambda} \left(e^{iu\phi(\lambda)} - 1 \right) \nu(du d\lambda)$$

for some positive measure $\nu(du d\lambda)$ on $\mathbb{R}_+ \times \Lambda$ satisfying the integrability condition $\iint_{\mathbb{R}_+ \times K} (1 \wedge u) \nu(du d\lambda) < \infty$ for compact $K \subset \Lambda$. Jacod and Shiryaev (1987, Chapter II, §4c) give details about this generalization of the usual Lévy-Khinchine formula to non-stationary processes and random fields. The Inverse Lévy Measure (ILM) algorithm of Wolpert and Ickstadt (1998a, 1998b) offers an explicit construction of such random fields, predicated on the representation

$$\Gamma[\phi] = \iint_{\mathbb{R}_+ \times \Lambda} u\phi(\lambda) H(du d\lambda) = \sum_{j \in J} u_j \phi(\lambda_j)$$

of $\Gamma(d\lambda)$ in terms of a Poisson measure $H(du d\lambda)$ on $\mathbb{R}_+ \times \Lambda$ with Lévy mean measure $\mathbb{E}[H(du d\lambda)] = \nu(du d\lambda)$; here $\{u_j, \lambda_j\}_{j \in J}$ represents an instance of the (at most countable) random support of $H(du d\lambda)$.

If $\nu(\mathbb{R}_+ \times \Lambda) < \infty$ then $H(du d\lambda)$ (and hence $\Gamma(d\lambda)$) will have only finitely many points of support. If $\nu(du d\lambda)$ has a density function $\nu(u, \lambda)$ with respect to some finite reference measure $m(du d\lambda)$, then Γ will have a probability density function

$$\pi(\Gamma) = \left[\prod_{j \in J} \nu(u_j, \lambda_j) \right] e^{m(\mathbb{R}_+ \times \Lambda) - \nu(\mathbb{R}_+ \times \Lambda)}$$

with respect to the random field with Lévy measure m .

If recorded measurements $G_i \in \mathcal{G}$ may be taken to differ only by independent measurement errors from the true values

$$g_i \equiv G(\omega_i) = \int_{\Lambda} k(\omega_i, \lambda) \Gamma(d\lambda) = \sum_{j \in J} k(\omega_i, \lambda_j) u_j,$$

with probability density functions $f(G_i | g_i)$, then the likelihood function is simply $L(\Gamma) = \prod_{i \in I} f(G_i | g_i)$ and by Bayes' theorem the posterior distribution for Γ has a probability density function

$$\pi(\Gamma | \{\vec{G}_i\}_{i \in I}) \propto \left[\prod_{j \in J} \nu(u_j, \lambda_j) \right] \left[\prod_{i \in I} f(G_i | g_i) \right] e^{m(\mathbb{R}_+ \times \Lambda) - \nu(\mathbb{R}_+ \times \Lambda)}. \quad (3)$$

This posterior distribution forms the basis for statistical inference about the solution Γ of the inverse problem in Equation (1). Features of Γ that are well-determined by the data (or the prior) will show little posterior variation, while the system's ill-posedness will be expressed in wide posterior variability of features that are undetermined by the prior and data.

The role of the Lévy prior distribution is analogous to that of the roughness penalty in conventional regularization methods, resolving features left unspecified by the data, but with the important benefit of easy interpretability and coherence.

3. A RHEOLOGY EXAMPLE

A Newtonian fluid suspended between two horizontal plates exhibits *viscous* behavior: a tangential force applied to one of the plates leads to a velocity gradient in the fluid proportional to the force per unit area, $\tau = \eta \frac{\partial v}{\partial y}$ (here τ represents the stress, or force per unit area, and v the horizontal velocity at any height y). The proportionality constant in this linear relationship, the viscosity η , is measured in pascal-seconds (“pascal” is the SI unit for pressure or stress, equal to one newton per square meter or one kilogram per meter per second squared, so one pound per square inch (p.s.i.) is about 6.89 kPa). The viscosity of familiar fluids ranges from about 10^{-5} Pa·s for air to 10^{-3} Pa·s for water to 1 Pa·s for glycerine.

A tangential force applied to one side of a springy or *elastic* substance induces a proportional deformation, $\tau = G\gamma$, where (unitless) γ represents the relative length change induced and the proportionality constant G (measured in Pa) is called the elastic modulus.

Polymers are gooey non-Newtonian compounds whose behavior lies in between highly viscous Newtonian fluids, with viscosities in the range of 10^2 – 10^5 Pa·s, and elastic compounds. Boltzmann (1876) had the idea of modeling these *visco-elastic* compounds by introducing time-dependence to the stress $\tau(t)$, elastic modulus $G(t)$, and deformation $\gamma(t)$ (with time-derivative $\dot{\gamma}(t)$), and relating all of them by a time-dependent extension of the elasticity equation, $\tau(t) = \int_{-\infty}^t G(t-s)\dot{\gamma}(s) ds$ or, upon changing variables,

$$\tau(t) = \int_0^\infty G(\omega) \dot{\gamma}(t-\omega) d\omega. \quad (4)$$

Boltzmann took G to have a “fading memory,” i.e., to be completely monotonically decreasing and so, by Bernstein’s Theorem (see Feller (1971), §XIII.4) representable in the form

$$G(\omega) = \int_0^\infty e^{-\omega/\lambda} \Gamma(d\lambda)/\lambda \quad (5)$$

as the Laplace transform of some positive measure $\Gamma(d\lambda)$ on \mathbb{R}_+ called the *relaxation spectrum*. For infinitesimal stresses the behavior is approximately that of a viscous fluid with *zero-shear viscosity* $\nu_0 = \int G(\omega) d\omega = \Gamma(\mathbb{R}_+)$. For periodic strains $\gamma(\omega) = \gamma_0 \sin(\omega t)$ the solution to Equations (4,5) is available in closed form:

$$\tau(\omega) = \gamma_0 \left[G'(\omega) \sin(\omega t) + G''(\omega) \cos(\omega t) \right]$$

where the elastic, in-phase, energy-conserving *storage modulus* and the viscous, out-of-phase, energy-dissipating *loss modulus* are given (respectively) by

$$G'(\omega) \equiv \int_0^\infty \frac{\omega^2 \lambda}{1 + \omega^2 \lambda^2} \Gamma(d\lambda) \quad G''(\omega) \equiv \int_0^\infty \frac{\omega}{1 + \omega^2 \lambda^2} \Gamma(d\lambda). \quad (6)$$

Table 1. Experimental measurements of storage modulus $G'(\omega_i)$ and loss modulus $G''(\omega_i)$, both in Pa, at various frequencies ω_i (s^{-1}) for a polybutadiene melt at 23°C, from Berger (1988).

ω	$G'(\omega)$	$G''(\omega)$	ω	$G'(\omega)$	$G''(\omega)$
2.493×10^0	2052	34526	7.680×10^1	432105	359952
3.670×10^0	4156	50445	1.144×10^2	534678	343388
5.373×10^0	8847	73294	1.654×10^2	619214	327629
7.864×10^0	18834	105329	2.433×10^2	701325	307419
1.144×10^1	37737	149699	3.539×10^2	772708	290069
1.695×10^1	74730	206936	5.238×10^2	841878	278292
2.451×10^1	136257	266220	7.529×10^2	897344	264055
3.608×10^1	223611	313420	1.114×10^3	956262	249131
5.218×10^1	324937	345321			

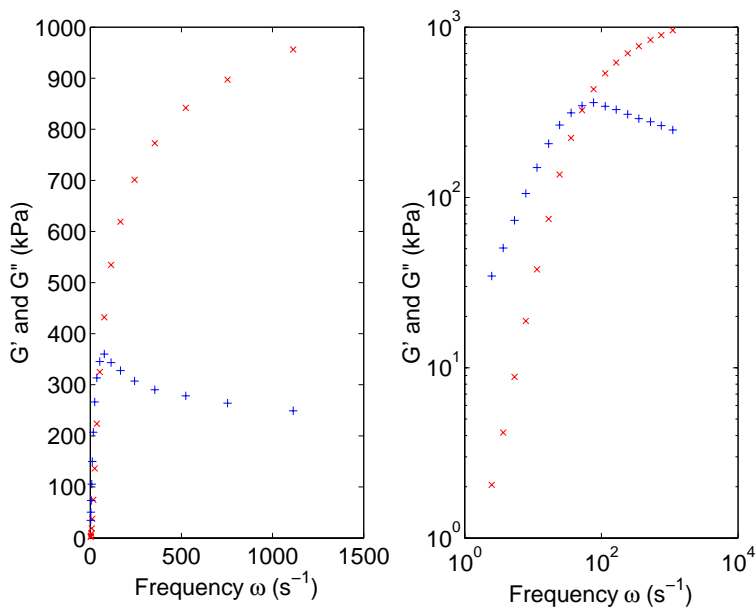


Figure 1. Measured $G'(\omega_i)$ (\times) and $G''(\omega_i)$ ($+$) from Berger (1988)

For small periodic strains both $G'(\omega)$ and $G''(\omega)$ can be measured experimentally using an oscillatory shear rheometer; our analysis below is based on measurements of Berger (1988), reproduced in Table (1) and plotted on both linear and logarithmic scale in Figure (1).

Equation (6) is a two-dimensional Fredholm inverse problem of the form of Equation (1) with $\omega \in \Omega \equiv \mathbb{R}_+$, $G(\omega) \equiv [G'(\omega), G''(\omega)] \in \mathcal{G} \equiv \mathbb{R}_+^2$, $\Lambda = \mathbb{R}_+$, $k : \Omega \times \Lambda \rightarrow \mathcal{G}$ with $k(\omega, \lambda) = [\omega^2 \lambda, \omega](1 + \omega^2 \lambda^2)^{-1}$, and $\Gamma(d\lambda)$ unknown. To proceed with the Bayesian modeling approach of Section (2) we will need to specify a measurement-error model for $\{G_i = G(\omega_i)\}_{i \in I}$ and a prior distribution for Γ .

Berger's choice of sampling frequencies ω_i and the general form of both $G'(\omega)$ and $G''(\omega)$ in Figure (1) suggest that the logarithmic transformation simplifies the relation and may stabilize the variance, leading to our choice of a bivariate lognormal measurement error model

$$\log G'_i \sim N(\log g'_i, \sigma^2) \quad \log G''_i \sim N(\log g''_i, \sigma^2)$$

or, more succinctly, $\vec{G}_i \sim \text{LN}(\vec{g}_i, \sigma^2 I_2)$ for some fixed $\sigma^2 > 0$ with $\vec{G}_i \equiv [G'_i, G''_i]$. In the

absence of repeated measurements which might let us validate the log-normal model and help select a value for σ^2 , we based our choices on an exploratory study of the residuals $\{(\vec{G}_i - \hat{G}_i)\}_{i \in I}$ from the best fit $\{\hat{G}_i\}_{i \in I}$ of Equation (6) to Berger's data.

For a prior distribution we have chosen the Gamma random field with Lévy measure

$$\nu(du d\lambda) = \alpha u^{-1} \lambda^{-1} e^{-\beta u} du d\lambda$$

with uniform shape measure (on a logarithmic scale) on the interval $\Lambda \equiv (\lambda_-, \lambda^+)$ with $\lambda_- = e^{-7}$, $\lambda^+ = e^{-1}$, a convenient approximation to the interval $1/\omega_{\max} < \lambda < 1/\omega_{\min}$ that the localization principle argument of Davies and Anderssen (1997) suggests is the widest range on which we can hope to learn from the data about $\Gamma(d\lambda)$. The localization principle is also the basis for our choice of an independent-increment prior distribution.

With our choice of prior distribution the zero-shear viscosity $\int_0^\infty G(t) dt = \Gamma(\mathbb{R}_+)$ has a $\text{Ga}(\alpha \log(\lambda^+/\lambda_-), \beta)$ distribution. Values for the parameters α, β were chosen to ensure that the mean and variance of $\Gamma(\mathbb{R}_+)$ would be approximately $6\alpha/\beta \approx 13,500$ and $6\alpha/\beta^2 \approx 50^2$, as suggested by the empirical evidence of Berger's observations (the zero-shear viscosity is also the slope $dG''/d\omega$ at $\omega = 0$). See Wolpert, Hansen and Ickstadt (to appear) for a wider range of prior distributions and measurement error models and see Anderssen and Hansen (to appear) for a more specific rheological treatment and discussion of the sampling localization theorem in the context of relaxation spectral analysis.

In summary, the complete Bayesian model specification is:

$$\begin{aligned} \text{Data:} \quad \vec{G}_i &\sim \text{LN}(\vec{g}_i, \sigma^2 I_2) \text{ in } \mathcal{G} = \mathbb{R}_+^2 \\ \text{Model:} \quad \vec{g}_i &= \left[\int_0^\infty \frac{\omega_i^2 \lambda}{1 + \omega_i^2 \lambda^2} \Gamma(d\lambda), \int_0^\infty \frac{\omega_i}{1 + \omega_i^2 \lambda^2} \Gamma(d\lambda) \right] \\ \text{Prior:} \quad \Gamma(d\lambda) &\sim \text{Levy}(\nu(du, d\lambda)), \quad u \in \mathbb{R}_+, \lambda \in \Lambda \equiv (\lambda_-, \lambda^+) \\ &= \text{Ga}(\alpha \lambda^{-1} d\lambda, \beta) \end{aligned}$$

and our goal is to estimate $\Gamma(d\lambda)$ upon observing $\vec{G}_i \equiv [G'(\omega_i), G''(\omega_i)] \approx \vec{g}_i$ for several frequencies ω_i .

4. COMPUTATIONS

Our choice of the Gamma Lévy prior features an infinite Lévy measure $\nu(\mathbb{R}_+ \times \Lambda) = \infty$ and, therefore, almost surely there are infinitely many terms in the representation

$$\vec{g}_i = \sum_{j \in J} [\omega_i^2 \lambda_j, \omega_i] (1 + \omega_i^2 \lambda_j^2)^{-1} u_j.$$

For any $\epsilon > 0$ the number $M_\epsilon = |J_\epsilon|$ of points with $u_j > \epsilon$ (indexed by $J_\epsilon = \{j \in J : u_j > \epsilon\}$) is a random variable whose prior distribution is Poisson with mean

$$\mathbb{E}[M_\epsilon] = \int_{\lambda_-}^{\lambda^+} \int_\epsilon^\infty \alpha u^{-1} \lambda^{-1} e^{-\beta u} du d\lambda = \alpha E_1(\beta \epsilon) \log \frac{\lambda^+}{\lambda_-} < \infty,$$

where $E_1(x) \equiv \int_x^\infty t^{-1} e^{-t} dt$ is the exponential integral function (Abramowitz and Stegun (1964), §5.1). The expected total mass $\sum_{j \in J_\epsilon} \{u_j : u_j \leq \epsilon\}$ of all points (u_j, λ_j) with $u_j \leq \epsilon$ is only

$$\mathbb{E}\left[H\left((0, \epsilon] \times \Lambda\right)\right] = \int_{\lambda_-}^{\lambda^+} \int_0^\epsilon \alpha \lambda^{-1} e^{-\beta u} du d\lambda = \alpha \beta^{-1} \log \frac{\lambda^+}{\lambda_-} (1 - e^{-\beta \epsilon}),$$

a fraction $(1 - e^{-\beta\epsilon})$ of the total prior expected mass $E[\Gamma(\mathbb{R}_+)] = \alpha\beta^{-1} \log \frac{\lambda^+}{\lambda_-}$. In our implementation we select ϵ small enough that this represents 0.5% of the total mass, and include only the mass points $u_j \in U_\epsilon \equiv (\epsilon, \infty)$. The posterior distribution of M_ϵ is not Poisson, of course.

The space Θ of configurations we model may be represented as the countable union of Cartesian powers

$$\Theta = \cup_{M=0}^{\infty} (\mathbb{R}_+ \times \Lambda)^M,$$

where the measure $\Gamma_\theta(d\lambda) \in \mathcal{M}_+(\Lambda)$ associated with index $\theta \in \Theta$ is

$$\Gamma_\theta(d\lambda) = \sum_{j=1}^M u_j \delta_{\lambda_j}(d\lambda),$$

the sum of M point masses of magnitudes $u_j \in \mathbb{R}_+$ at points $\lambda_j \in \Lambda$. From Equation (3) we can compute the probability density function of the posterior distribution of $\theta \in \Theta$ (with respect to the Poisson random measure on Θ with rate $m(du d\lambda)$) upon observing the $N \equiv |I|$ vectors $\vec{G}_i = [G'_i, G''_i]$:

$$\begin{aligned} \log \pi(\theta | \{\vec{G}_i\}_{i \in I}) &= c + \sum_{j \in J} \log \nu(u_j, \lambda_j) + \sum_{i \in I} \log f(\vec{G}_i | g_i) \\ &= c + M \log \alpha - \sum_{j \in J} \log u_j \lambda_j - \beta \sum_{j \in J} u_j \\ &\quad - N \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i \in I} \left[\log^2 \frac{G'_i}{g'_i} + \log^2 \frac{G''_i}{g''_i} \right] \end{aligned}$$

where $c = m(U_\epsilon \times \Lambda) - \nu(U_\epsilon \times \Lambda)$ does not depend on θ .

To implement the Metropolis-Hastings version of the Markov Chain Monte Carlo method (see, e.g., Tierney (1994)) we must select an irreducible transition probability distribution $Q(d\theta^* | \theta)$ on Θ . Our choice reflects our intention to *model* uncertainty about $\Gamma(d\lambda)$ using the Gamma random field with its infinite Lévy measure, even though our *implementation* permits us to simulate only the finite number M_ϵ Gamma masses of magnitude $u_j > \epsilon$.

The heuristic behind our proposal distribution is to imagine infinitely many particles at locations $(u_j, \lambda_j) \in \mathbb{R}_+ \times \Lambda$ all undergoing simultaneous ergodic diffusion with the posterior as a stationary distribution. At any given time only finitely many points M_ϵ will lie above the line $u_j > \epsilon$; now and then one of these will diffuse below that line, causing M_ϵ to fall by one, while now and then one of the infinitely many points below the line will rise above it, increasing M_ϵ by one. Sampled at discrete times this would be a random walk similar to that we propose below, with three types of steps—those with M_ϵ unchanged ($\Delta M_\epsilon = 0$) and those where M_ϵ increases or decreases by one ($\Delta M_\epsilon = \pm 1$). For fixed logarithmic step size δ (we use $\delta = 0.25$), re-entry probability $0 < p < 1$ (we use $p = 0.01$), and re-entry distribution with density $f(u, \lambda)$ on $(0, \infty) \times \Lambda$ (see below), the move proposals are:

$$\begin{aligned} \Delta M_\epsilon = 0 & \text{ With probability } 1-p, \text{ choose } j \text{ uniformly from the integers } 1:M \text{ and propose} \\ & \text{ the lognormal step } u_* = u_j \exp(\delta Z_1), \lambda_* = \lambda_j \exp(\delta Z_2) \text{ with } Z_1, Z_2 \sim \\ & \text{N}(0, 1). \text{ Reflect at the boundaries if necessary to ensure that } \lambda_* \in \Lambda = \\ & (\lambda_-, \lambda^+). \text{ If } u_* > \epsilon, \text{ then the proposed new } M_\epsilon \text{ remains unchanged. Other-} \\ & \text{wise,} \end{aligned}$$

- $\Delta M_\epsilon = -1$ If $u_* \leq \epsilon$ above, remove j from J and decrease M_ϵ by one; the resulting proposal is to delete the single point (u_j, λ_j) from the ensemble θ .
- $\Delta M_\epsilon = +1$ With probability p , increment M_ϵ by one and introduce a new index M to J and draw a new mass point $(u_M, \lambda_M) \sim f(u_M, \lambda_M)$ from the re-entry distribution.

For our re-entry distribution we draw λ from the uniform distribution on a logarithmic scale on Λ and, independently, draw u from the exponential distribution with mean $\mu = E[H]/E[M_\epsilon]$, conditioned to satisfy $u > \epsilon$, giving $f(u, \lambda) = (\lambda \mu \log \frac{\lambda^+}{\lambda^-})^{-1} \exp(-\frac{\epsilon-u}{\mu})$ on $U_\epsilon \times \Lambda$.

The conditional p.d.f. $Q(\theta_* | \theta)$ of the proposal transition probability distribution $Q(d\theta_* | \theta)$ (again, with respect to the $\text{Po}(m(du d\lambda))$ Poisson random measure) is easily calculated from this prescription. Finally the MCMC algorithm proceeds as follows:

0. Initialize $t = 0$, $M^{(0)} \sim \text{Po}(E[M_\epsilon])$, $J = \{1, \dots, M^{(0)}\}$, $\{(u_j, \lambda_j)\}_{j \in J} \sim f(u, \lambda)$, and set $\theta^{(0)} = \{(u_j, \lambda_j)\}_{j \in J}$.
1. Find a proposed new point $\theta_* \sim Q(d\theta_* | \theta^{(t)})$ and compute the Metropolis-Hastings log acceptance probability

$$\zeta^{(t+1)} = \log \pi(\theta_* | \{\vec{G}_i\}) + \log Q(\theta^{(t)} | \theta_*) - \log \pi(\theta^{(t)} | \{\vec{G}_i\}) - \log Q(\theta_* | \theta^{(t)})$$

2. Generate a standard exponential random variable $Z \sim \text{Ex}(1)$ and set

$$\theta^{(t+1)} = \begin{cases} \theta_* & \text{if } Z + \zeta^{(t+1)} \geq 0 \\ \theta^{(t)} & \text{if } Z + \zeta^{(t+1)} < 0 \end{cases}.$$

Adjust $M^{(t)}$ and J if necessary. Increment $t \leftarrow t + 1$.

3. Periodically (e.g. at 100 evenly-spaced times following “burn-in”) store $\theta^{(t)}$.
4. If $t < TMAX$, repeat steps 1–4.

The parameters δ , μ , etc. are adjusted in trial runs to ensure that the rate of accepting proposed moves is approximately 20-50%. On contemporary small computers (2GHz dual-processor Unix workstations), our MatLab implementation can complete approximately two million steps per hour.

5. RESULTS

Figure (2) shows a representation of the posterior distribution of the model’s predictions $G(\omega) = [G'(\omega), G''(\omega)]$, with the measurements (\times for $G'(\omega_i)$, $+$ for $G''(\omega_i)$); the figure shows the 25%, 50% and 75% percentile bands along with 100 MCMC iterations (equally spaced from among one million). These curves lie so close together that it is difficult to distinguish them in the plot, showing that there is little posterior uncertainty about $G(\omega)$.

Figure (3) shows the prior (dotted line) and posterior means of the spectral density $\Gamma(d\lambda)/d\lambda$, and Figure (4) the prior (again, dotted) and posterior cumulative spectral distribution $H[(0, \lambda)]$ with the 25%, 50% and 75% posterior percentiles and 100 MCMC iterations to illuminate the distribution. Evidently the spectral density is unimodal (or, if not, has no strong second mode), centered at about $\lambda \approx 0.020s$, with half of its mass in the interval $[0.015, 0.027]$ and 90% in the interval $[0.004, 0.086]$.

Figure (5) displays the posterior distribution of M_ϵ , the number of mass points for the spectral measure of magnitude $u > \epsilon$, with the posterior mean $E[M_\epsilon | \{\vec{G}_i\}_{i \in I}] = 303.9$ indicated with a vertical line (the prior distribution was $M_\epsilon \sim \text{Po}(118.0)$).

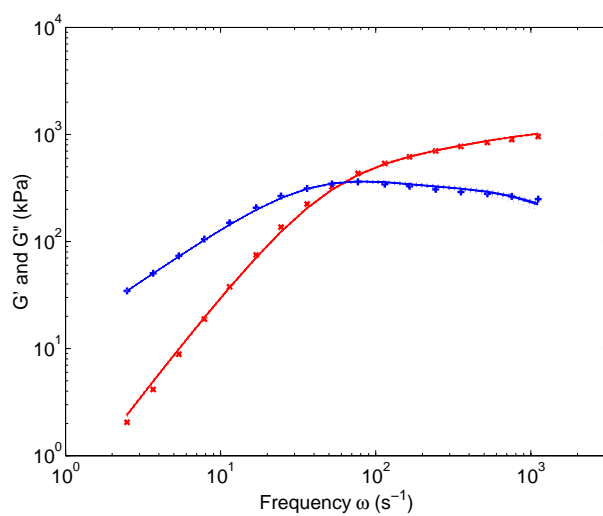


Figure 2. Posterior distributions for $G'(\omega_i)$ (\times) and $G''(\omega_i)$ ($+$).

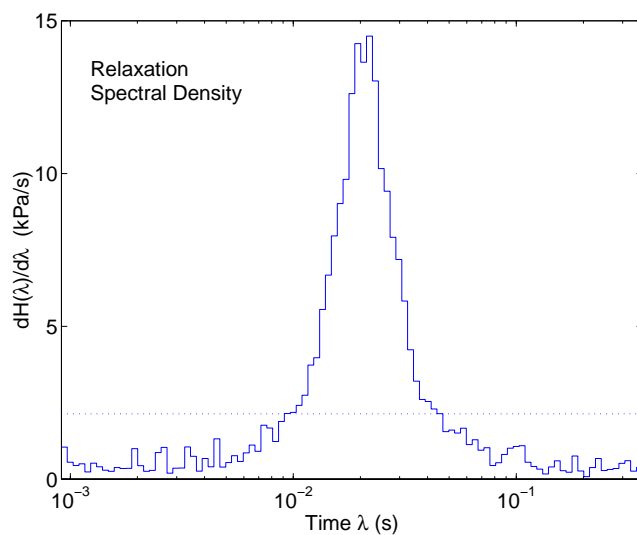


Figure 3. Prior and Posterior Spectral Density

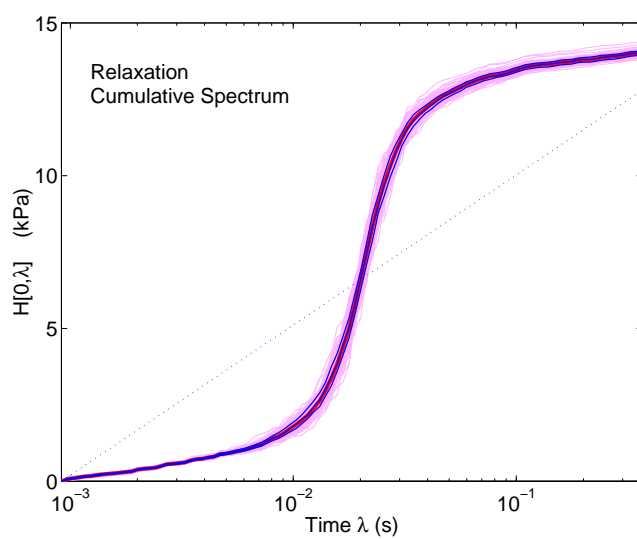


Figure 4. Prior and Posterior Cumulative Spectrum

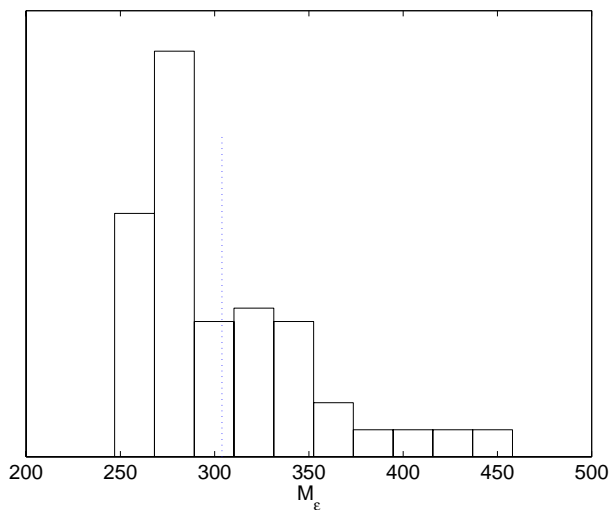


Figure 5. Posterior Distribution of M_ϵ

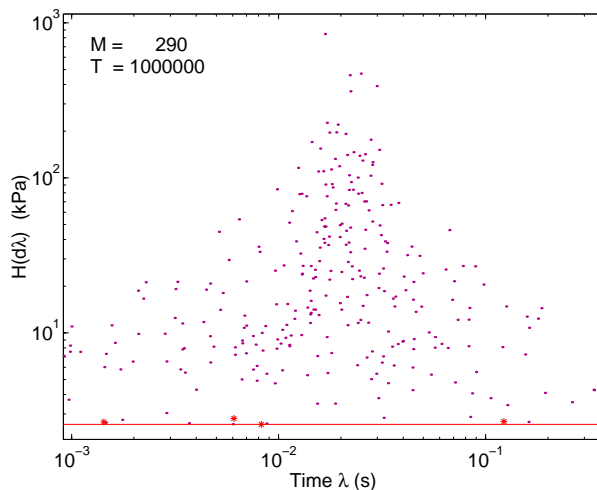


Figure 6. A single configuration $\theta^{(t)}$ from the simulation

Figure (6) shows a single state $\theta^{(t)}$ from the simulation (the final one, with $t = 1,000,000$). This step features $M_\epsilon = 290$ mass points (u_j, λ_j) (dots represent recently-moved points, stars represent recently-added points from the re-entry distribution). Following the 100,000-step burn-in phase in this simulation run, M_ϵ ranged from a minimum of 247 to a maximum of 458, with a mean of 303.9. Magnitudes u_j are represented on a logarithmic scale, so a very large fraction of the mass is represented by the largest few points; the 30 points above $u > 100$ hold 50% of the total mass, and the 186 points above $u > 10$ hold about 95%. Here $\epsilon = 2.57$ was chosen to ensure that 99.5% of the prior mass exceeded ϵ .

6. DISCUSSION

The nonparametric Bayesian approach to inference in inverse problems allows us to model explicitly our prior beliefs or expert understanding about features of the solution $\Gamma(d\lambda)$, and to represent honestly and coherently whatever uncertainty remains about these features following the observation of our data. The prior distribution serves the same role as the roughness penalty in the regularization approach, but with the added benefits of interpretability and coherence.

The specific choice of Lévy prior distributions leads both to (usually welcome) prior independence of the spectral measure of disjoint sets and to tractable computational problems.

Note that the *posterior* distribution of $\Gamma(d\lambda)$, while still discrete, is *not* Lévy—the point process $H(du d\lambda)$ in the representation $\Gamma(d\lambda) = \int_{\mathbb{R}_+} u H(du, d\lambda)$ assigns independent random variables $H(A_k) \sim \text{Po}(\nu(A_k))$ to disjoint sets $A_k \subset \mathbb{R}_+ \times \Lambda$ but the likelihood function induces dependence among the $\{H(A_k)\}$.

Our nonparametric Bayesian approach offers a number of advantages over earlier methods, including

Flexibility : Different choices for the Lévy measure $\nu(du, d\lambda)$ will lead to “smooth” or “bumpy” measures, allowing the analysis to reflect any expert opinion about features of $\Gamma(d\lambda)$ such as smoothness, uni- or multi-modality, zero-shear viscosity $\Gamma(\mathbb{R}_+)$, etc.

Tractability : The MCMC approach described here works equally well for any of these choices of Lévy measure (and measurement-error model). Posterior distributions of any quantity $\Gamma[\phi] = \int \phi(\lambda) \Gamma(d\lambda)$, including interval measures $\Gamma(A)$ and zero-shear viscosity $\Gamma(\mathbb{R}_+)$, are easily computed.

Parsimony : Some choices of $\nu(du, d\lambda)$ will almost-surely have finite numbers M of mass points, and can even have $E[M]$ as small as two or three, leading to strikingly parsimonious representations of $\Gamma(d\lambda)$ as a sum of a small number of point masses, similar to the representations of Anderssen and Davies (2001).

In our work with the Gamma prior we found no apparent sensitivity to the choice of the cut-off $\epsilon > 0$. We settled on a value small enough that our truncated approximation includes 99.5% of the prior expected mass, but in a sensitivity analysis we varied ϵ over a wide range.

We also explored a similar algorithm modeling a fixed number M of mass points (u_j, λ_j) diffusing over $U_\epsilon \times \Lambda$ with *reflecting* boundary conditions at $u = \epsilon$, rather than the *free* boundary conditions of the present implementation. Model fit and posterior distributions of $\Gamma(A)$ for intervals A were very similar to those found in the present study, for a wide range of values of M .

We continue to explore the prior elicitation issues that arise in this modeling approach, studying a range of different Lévy measures and seeing how they affect posterior inference. We are also exploring inference for other inverse problems; some of this work will be described in Wolpert, Hansen and Ickstadt (to appear).

7. ACKNOWLEDGMENTS

The authors acknowledge gratefully the support of US NSF grant DMS-9626829, of US EPA grant R828686-01-0, of EC TMR grant ERB-FMRX-CT96-0096, and of the Centre for Mathematical Physics and Stochastics (MaPhySto), funded by the Danish National Research Foundation. We would like to thank Robert S. Anderssen for suggesting the problem and Jesper Møller and the Department of Mathematical Sciences of Aalborg University for facilitating our collaboration.

REFERENCES

- Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables*. Washington, D.C.: National Bureau of Standards.
- Anderssen, R. S. and Davies, A. R. (2001). Simple moving-average formulae for the direct recovery of the relaxation spectrum. *J. Rheol.* **45**, 1–27.
- Anderssen, R. S. and Hansen, M. B. Bayesian relaxation spectrum recovery. (to appear).
- Berger, L. (1988). *Untersuchung zum Rheologischen Verhalten von Polybutadienen mit Bimodaler Molmassenverteilung*. Ph.D. Thesis, Eidgenössische Technische Hochschule (ETH), Zurich.
- Boltzmann, L. (1876). Zur Theorie der elastischen Nachwirkung. *Ann. Phys. Chem.* **7**, 624–654.

- Davies, A. R. and Anderssen, R. S. (1997). Sampling localization in determining the relaxation spectrum. *J Non-Newtonian Fluid Mech.* **73**, 163–179.
- Feller, W. (1971). *An Introduction to Probability Theory and its Application* **2**, (2nd ed). Chichester: Wiley.
- Fredholm, Erik I. (1900). Sur une nouvelle méthode pour la résolution du problème de Dirichlet. *Œuvres complètes: publiées sous les auspices de la Kungliga svenska vetenskapsakademien par l'Institut Mittag-Leffler*, 61–68, reprinted 1955. Malmö: Litos reprotryck.
- Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes*. Berlin: Springer.
- Kirsch, A. (1996). *An Introduction to the Mathematical Theory of Inverse Problems*. New York: Springer.
- Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.* **22**, 1701–1762.
- Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Doklady* **4**, 1035–1038.
- Wolpert, R. L., Hansen, M. B. and Ickstadt, K. Inverse problems and a Lévy process solution. (to appear).
- Wolpert, R. L. and Ickstadt, K. (1998a). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–267.
- Wolpert, R. L. and Ickstadt, K. (1998b). Simulation of Lévy random fields. *Practical Nonparametric and Semi-parametric Bayesian Statistics* (D. Dey, P. Müller and D. Sinha, eds.), New York: Springer, 227–242.

DISCUSSION

SUBHASHIS GHOSAL (*North Carolina State University, U. S. A.*)

First I would like to congratulate the authors for an excellent paper. It once again demonstrates the usefulness of statistical, in particular Bayesian, methods in hard problems in numerical analysis. Statistical, and in particular Bayesian, methods have a lot to contribute in numerical analysis. Bayesian approaches to numerical analysis problems, in fact, goes back to Poincaré. He considered a Bayes procedure for interpolation polynomials. His approach consisted of putting a prior distribution on the coefficients of the power series expansions of the interpolated functions. Despite the enormous potential of Bayesian approaches for numerical analysis problems, it is somewhat surprising that the methods did not receive a lot of attention. Notable exceptions are Bakhvalov (1959) and Diaconis (1988). The former author considered numerical quadrature problems and put a prior that is supported on finitely many functions. Diaconis (1988) formulated a Bayesian approach for a general numerical problem and considered several examples. In particular, he showed that for a quadrature problem, if the prior on the functions is taken to be a Brownian motion, then the trapezoid rule is the Bayes solution. The paper of Ickstadt, Hansen and Wolpert renews the hope of a revival of Bayesian methods for numerical analysis.

The authors consider certain inverse problems and offer Bayesian solutions. If variables are measured with errors, statistical methods considering the presence of random errors must be employed. Since an inverse problem is ill-conditioned, small changes in the observation lead to large changes in the solution. Only a small rounding error in the observations can also make a big difference in the solution. Therefore, even if the observations are deterministic functions of the parameters, Bayes solutions may offer the desired stability of the solution. Bayesian procedures are expected to be much more stable than many common frequentist methods such as the least square estimate in ill-conditioned problems. For instance, the ridge regression estimator, which is an approximate Bayes procedure, gives much more stable solutions than those given by the ordinary least squares.

In the continuous analogue of the regression problem considered by the authors, a nonparametric Bayesian approach is natural. The authors argue that a random field prior based on a Lévy process is natural for this problem.

It seems that the main question in this context is whether the Bayes procedure leads to a stable solution. We believe it does, since a Bayesian method has an inherent smoothing property. A formal verification would probably consist of the computation of the norm of the linear map given by the derivative, at a given vector of observations, of the map relating the observations to the Bayes estimate. This computation may turn out to be too difficult. In that case, at least the stability property should be checked by numerically computing the variations in the Bayes estimate for small perturbations of the observations.

The authors argue that a Lévy process is a natural candidate for the prior distribution on the space of the function to be found. We partly agree with the authors. The independence structure is certainly helpful for writing down the joint posterior density of the weights and the location of the support points up to a normalizing constant. Then it is relatively easy to apply the Metropolis-Hastings algorithm to calculate the posterior distribution. However, independence has a price to be paid in terms of the roughness of the sample paths of the underlying function.

Perhaps it is not unreasonable to think that a priori the sample paths of these functions are smooth. In particular, this means that the random measure $\Gamma(ds)$ considered by the authors is absolutely continuous and the density is in fact a smooth function. In that case, one should use a prior for $\Gamma(ds)$ that sits on the absolutely continuous smooth densities. If the same symbol $\Gamma(s)$ stands for the density of $\Gamma(ds)$, then $\Gamma(s)$ should be a random process that takes values in $(0, \infty)$ and has smooth paths. Then a possible alternative to the Lévy process prior could be a Gaussian process prior for $\log \Gamma(s)$, where the covariance kernel for the Gaussian process is chosen in such a way that the sample paths have the desired smoothness properties. Another attractive alternative is to consider a mixture prior of the form $\int \psi_k(s, z) dU(z)$, where $\psi_k(s, z)$ is a probability kernel, k is a smoothing parameter and $U(z)$ is an increasing process on $(0, \infty)$ without any smoothness restriction. Note that the total mass of $\Gamma(\cdot)$ on $(0, \infty)$ is equal to $U(\infty)$. Now a Lévy process prior can be used for $U(z)$. Using the notion of a Feller approximation, Petrone and Veronese (2002) argued that the gamma kernel is a natural kernel when the domain is $(0, \infty)$ as in this case. We agree that the kernel can be combined with the likelihood at the computation stage. However, in this case, one is not interested in the posterior distribution of $U(\cdot)$ but in that of $\int \psi_k(\cdot, z) dU(z)$, the smoothed out $U(\cdot)$. Smoothing is likely to have a useful consequence in consistency and rates of convergence. The situation is similar to the difference between the Dirichlet process and Dirichlet mixtures. Without the smoothing, one can only hope consistency in the weak-star topology or a Kolmogorov-Smirnov type norm, while with smoothing one would expect consistency in the stronger L_1 or L_2 distances. For instance, in the case of estimation of a probability measure that has a density, the Dirichlet prior only gives rise to consistency in the weak or Kolmogorov-Smirnov norm while Dirichlet mixtures are consistent under the variation norm (Ghosal, Ghosh and Ramamoorthi, 1999). The effect of smoothing is even more apparent in rates of convergence. For normal mixtures, one gets a very fast rate $\log n/\sqrt{n}$ of convergence (Ghosal and van der Vaart, 2001). On the other hand, the estimation of a mixing distribution is known to be a hard problem with only a logarithmic rate of convergence.

The Inverse Lévy Measure Algorithm comes handy in the computations. As it is common with most non-parametric Bayesian procedures, computations can be done only with the help of Markov chain Monte Carlo methods. The authors use the Metropolis algorithm by first reducing the infinitely many jump points of the Lévy process to finitely many jumps by ignoring all the remaining jump points. Perhaps this is a natural way to resolve the problem. However, one needs to be cautious about the number of terms to be taken. If one takes only a fixed number of terms independent of the sample size, then for large sample sizes, the approximate posterior will necessarily be very different from the actual posterior in the natural scale of accuracy for

the sample size, since the former posterior is inconsistent (because the corresponding prior does not have the full support) while the latter is expected to be consistent — see the next paragraph. The small jumps ruled out in the prior stage may become important with increasing sample size. Therefore, it seems natural to ask the question how many support points are to be taken in the computation for a given sample size to guarantee a given level of accuracy. It seems that one needs to take enough terms so that the remaining part has a prior probability that is exponentially small with the increasing sample size.

It is important to know asymptotic properties such as consistency and rates of convergence of the Bayes solution. Consistency questions are important not only to objective Bayesians and frequentists, but also to subjective Bayesians since consistency is equivalent of merging of opinion; see Diaconis and Freedman (1986) for a discussion of this topic and a striking example of inconsistency. There seem to be two parameters — σ measuring the (lack of) precision and n , the number of observations taken, both of which may control the asymptotics. Letting $\sigma \rightarrow 0$ is the same as taking repeated measurements. Nevertheless, to estimate the whole function accurately, the number of points where observations are taken must increase and fill up the whole domain. Since a Lévy process can be chosen to have full weak-star support, consistency is expected to hold under the weak-star topology. In order to have this, the Lévy measure $\nu(\cdot)$ must necessarily have infinite total mass. To have a stronger form of consistency or a rate of convergence, one needs to use a prior that has smooth sample paths like the Gaussian or the mixture prior.

Finally, we wonder whether the log normal error distribution used in the computation is the most appropriate one. The log normal errors seem to have arisen because the measurements are positive and one switches to the logarithmic scale. Since a log normal variable can be incredibly large, and it is the actual variable rather than its logarithm that is measured, the log normal model for measurement errors may not well represent the uncertainty in the present context.

REPLY TO THE DISCUSSION

First we would like to thank Professor Ghosal for his thorough discussion and encouraging remarks about this work, and for the additional perspective he offers us on the interplay between Bayesian statistics and numerical analysis.

Ghosal begins with a rousing endorsement of the application of Bayesian methods and principles to problems in numerical analysis, for which we thank him and with which we heartily concur. A large number of factors contribute to uncertainty in applied mathematical problems such as those we consider— measurement error is the principal one we consider here, but model simplifications or misspecification, minor recording and transcription errors and, as Ghosal observes, even rounding can affect inference in ill-conditioned problems. The Bayesian approach is ideal for reflecting all the sources of uncertainty that affect inference in applied mathematics problems.

We agree with Ghosal that our Bayesian approach “leads to a stable solution” (we believe he is referring to the relative insensitivity of inference to small variations in the observations), but we would not characterize stability as “the main question.” Earlier regularization methods also overcome the natural ill-conditioning of this and similar inverse problems by penalizing roughness, trading off a small degree of model fit for the comfort of a smooth (or *regular*) solution. What is new in the present approach is the *interpretability* of this trade-off (we express our anticipation of regularity or smoothness through a prior distribution with an easy probabilistic interpretation) and the opportunity to express coherently all the uncertainty in model predictions and inference that arises from measurement errors and other sources, an advantage inherent in our Bayesian formulation. The stability computation he suggests appears

to be quite problem-specific, depending on details of the measurement error model used in specific applications; our goal here is to present a widely applicable methodology useful in many commonly studied inverse problems, with widely varying measurement error models.

The frequentist question of asymptotic consistency (as the number n of observations increases) is an interesting *mathematical* issue that would be fun to study further. We do not know precise conditions on the prior distribution (i.e., on the Lévy measure), the likelihood (i.e., on the measurement error model), and the specific problem (i.e., on the kernel $k(t, s)$) that would guarantee consistency. That problem is less interesting *statistically*, however, since there is no opportunity here to acquire additional observations (the number $n = 17$ of data points is fixed). What *is* interesting in practice is how accurately and honestly we can represent the uncertainty about quantities of interest in applications, such as the elastic modulus $G(\omega) = \int_{\Lambda} \exp(-\omega/\lambda) \Gamma(d\lambda)/\lambda$ and integrals of the uncertain relaxation spectral measure $\Gamma(d\lambda)$ over intervals. We feel that the method succeeds in representing the evidence contained in the data and the prior about these issues of interest in this application.

Consistency, as Ghosal observes, can be expected to require an infinite Lévy measure $\nu(du d\lambda)$, hence a random field with infinitely many mass points, while any digital implementation must necessarily be finite. In some sense our approach actually models infinitely many mass points, though, explicitly keeping track of the locations and magnitudes of *all* of the $M_{\epsilon} < \infty$ points of macroscopic mass $u_j > \epsilon$ for some small $\epsilon > 0$ (these are the largest mass points), and also modeling the total mass $\sum \{u_j : u_j < \epsilon\}$ (but not the precise locations or individual masses) of the infinitely-many points with mass below ϵ . When a Metropolis-Hastings move drops a mass point below the ϵ -boundary the number M_{ϵ} decreases by one and the total mass of small points experiences a corresponding increase; the re-entry distribution can be interpreted as the occasional elevation of a small-mass point above the ϵ -boundary, entering the explicitly modeled configuration and thus increasing M_{ϵ} by one.

Our choice of the gamma prior distribution (with infinite Lévy measure) and very small ϵ (ensuring that 99.5% of the prior Lévy mass will lie above the ϵ -boundary) lead to prior and posterior means of about $E[M_{\epsilon}] \approx 118$ and $E[M_{\epsilon} | \text{data}] \approx 304$, respectively. For perspective, the optimal fit of the data with finitely many points M shows virtually no improvement beyond $M = 8$ or so (more precisely, there is negligible drop in deviance at the MLE for the $2M$ -dimensional models with M masses $\{u_j\}$ and locations $\{\lambda_j\}$ for $M > 8$). The data alone offer no compelling reason to use more than a small handful of points—our implementation includes many more points than are needed.

Ghosal takes issue with our choice of a lognormal measurement error model. He suggests that such a model offers the possibility of huge positive deviations that do not appear in our dataset, and asks if the data themselves were not measured (with possible measurement error), rather than their logarithms, making a linear-scale measurement error model appear to be more appropriate than the log-scale one we used. His concern appears to be well founded and we will explore that issue further.

Finally, Ghosal suggests some alternatives to our approach, based on smoothed (or mixtures of) Lévy processes or on exponentiating Gaussian processes. Of course there are many ways to model uncertainty about the relaxation measure $\Gamma(ds)$ in our application or, more generally, about the measures arising in Fredholm integral equations of the first kind; we have offered and illustrated one, and Ghosal suggests a few more of varying degrees of tractability. Part of his motivation appears to be a concern with the discrete nature of the underlying measure $\Gamma(ds)$ arising in our Lévy formulation, a consequence of our insistence on independent increments. The independent increment requirement was not made for our convenience, it followed from the localization result of Davies and Anderssen (1997). We know of no reason to expect $\Gamma(ds)$

to have a density $\Gamma(s)$, let alone a continuous one; the objects of apparent interest are not Γ itself but rather the elastic modulus $G(\omega) = \int_{\Lambda} \exp(-\omega/\lambda) \Gamma(d\lambda)/\lambda$ and the partial viscosity $\eta_{(a,b)} \equiv \int_{(a,b)} \Gamma(d\lambda)/\lambda$ associated with intervals $(a, b) \subset \Lambda$.

For investigators or problems where a continuous density function is required, it may be useful to note that the Fredholm equation $G_{\epsilon}(\omega) = \int_{\Lambda} k(\omega, \lambda) \Gamma_{\epsilon}(\lambda) d\lambda$ for a smoothed density $\Gamma_{\epsilon}(\lambda) \equiv \int \gamma_{\epsilon}(\lambda - t) \Gamma(dt)$ can be rewritten as $G_{\epsilon}(\omega) = \int_{\Lambda} k_{\epsilon}(\omega, \lambda) \Gamma(d\lambda)$ with a discrete Lévy measure $\Gamma(d\lambda)$ and a smoothed kernel $k_{\epsilon}(\omega, \lambda) \equiv \int k(\omega, t) \gamma_{\epsilon}(\lambda - t) dt$. Thus, the present formulation and implementation already includes the mixture-of-Lévy model, simply by convolving the kernel $k(\omega, \lambda)$ with a smoother $\gamma_{\epsilon}(\cdot)$.

We would like to thank the audience, organizers, and our discussant again for a lively session exploring new opportunities for Bayesian statisticians to make contributions in areas of applied mathematics where the role of uncertainty and the need for statistics have not yet fully been recognized. There are many more opportunities awaiting us.

ADDITIONAL REFERENCES IN THE DISCUSSION

- Bakhvalov, N. S. (1959). Approximate computation of multiple integrals. *Vestnik Moskov. Univ. Ser. Mat. Mekh. Astr. Fiz. Khim.* **4**, 3–18.
- Diaconis, P. (1988). Bayesian numerical analysis. *Statistical Decision Theory and Related Topics IV* **1** (S. S. Gupta and J. O. Berger, eds.). Berlin: Springer, 163–175.
- Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates (with discussion). *Ann. Statist.* **14**, 1–67.
- Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27**, 143–158.
- Ghosal, S. and van der Vaart, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29**, 1264–1280.
- Petrone, S. and Veronese, P. (2002). Non parametric mixture priors based on an exponential random scheme. *J. Italian Statist. Soc.* (to appear).