Metropolis-Hastings for Lévy Random Fields

Robert L. Wolpert
Department of Statistical Science
Duke University, Durham, NC, USA

Vsn 29, 2012-03-23

1 Hastings' Ratio

We begin with a derivation and review of the basic Metropolis Hastings approach to what is now called MCMC, on an abstract measurable space. In Section (2) we move to the case of reversible-jump MCMC schemes for Lévy processes with (nearly) arbitrary Lévy measures.

Let $(\Theta, \mathcal{F}, P)$ be a probability space; we would like to construct an ergodic discrete-time Markov chain $\{\theta^t\}_{t\in\mathbb{N}}$ taking values in Θ such that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{0\le t<T} \phi\big(\theta^t\big) = \int_\Theta \phi(\theta)\, P(d\theta) \tag{1}$$

for (at least) bounded continuous functions $\phi(\cdot)$. We begin in Section (1.1) with the simplest possible case: where Θ is a finite set $\Theta = \{\theta_1, ..., \theta_n\}$ and where $\mathcal{F} = 2^\Theta$ is all possible subsets; we then consider Euclidean space in Section (1.2), and finally Lévy random fields in Section (2).

1.1 Finite Spaces

Let $P : 2^\Theta \to [0, 1]$ be a probability measure on a finite set $\Theta = \{\theta_1, ..., \theta_n\}$; our goal in this section is to construct a Markov chain on Θ whose stationary distribution is $P(d\theta)$. We begin with a specified initial distribution $P^0(d\theta)$, an auxilary transition kernel $Q : 2^\Theta \times \Theta \to [0, 1]$ (so $Q(\cdot \mid \theta)$ is a probability measure on Θ for each fixed $\theta \in \Theta$), and a $[0, 1]$-valued function $A(\theta^*, \theta)$, all to be specified later.

Our approach will be to construct a random walk θ^t on Θ by drawing $\theta^0 \sim P^0(d\theta)$ from $P^0(d\theta)$ and at time $t = 0$ and then, at each time-step t,

1. Propose a new value $\theta^* \sim Q(d\theta^* \mid \theta^t)$;
2. With probability $A(\theta^*, \theta^t)$, *accept* the proposal and set $\theta^{t+1} := \theta^*$;
3. Otherwise, *reject* the proposal and set $\theta^{t+1} := \theta^t$;
4. Increment $t \leftarrow t + 1$ and repeat.

Note that the distributions $P^0(d\theta)$ and $P(d\theta)$, the transition kernel $Q(d\theta^* \mid \theta)$, and the function $A(\theta^*, \theta)$ are all determined (respectively) by the n-vectors and $n{\times}n$ matrices

$$\begin{aligned} p_i^0 &:= P^0\big(\{\theta_i\}\big) \\ p_i &:= P\big(\{\theta_i\}\big) \\ q_{ij} &:= Q\big(\{\theta_j\} \mid \theta_i\big) \\ a_{ij} &:= A\big(\theta_j \mid \theta_i\big) \end{aligned}$$

(note conventional ordering of i, j in q and a differ from that of θ_j, θ_i in Q and A). We now turn to the selection of A.

To achieve Equation (1) we must approach equilibrium— *i.e.*, the probability distribution $P^t(d\theta)$ of θ^t must converge to $P(d\theta)$. Suppose we in fact *reach* (or even begin at) equilibrium— *i.e.*, have $\mathsf{P}[\theta^t = \theta_i] = p_i$ for each i. To maintain equilibrium with our proposed algorithm, we must have:

$$p_j = \sum_i p_i R_{ij} \qquad \text{where } R \text{ is our new chain's transition matrix,}$$

$$R_{ij} = \begin{cases} q_{ij} a_{ij} & \text{for } i \neq j \\ q_{ii} a_{ii} + \sum_k q_{ik}[1 - a_{ik}] & \text{for } i = j \end{cases}$$

Thus

$$\begin{aligned} p_j &= \sum_i p_i q_{ij} a_{ij} \quad + \sum_k p_j q_{jk}[1 - a_{jk}] \\ &= \sum_i p_i q_{ij} a_{ij} \quad + \sum_k p_j q_{jk} \quad - \sum_k p_j q_{jk} a_{jk} \\ &= \sum_i p_i q_{ij} a_{ij} \quad + \quad p_j \quad - \sum_i p_j q_{ji} a_{ji} \end{aligned}$$

and so

$$\sum_i p_i q_{ij} a_{ij} = \sum_i p_j q_{ji} a_{ji} \tag{2}$$

for each i, j. The simplest way to achieve this is to ensure that the stronger condition of "detailed balance" holds: for *every* i, j,

$$\begin{aligned} p_i q_{ij} a_{ij} &= p_j q_{ji} a_{ji}, \qquad \textit{i.e.}, \\ \frac{a_{ij}}{a_{ji}} &= \frac{p_j\, q_{ji}}{p_i\, q_{ij}}. \end{aligned} \tag{3}$$

Evidently anything of the form $a_{ij} = p_j\, q_{ji}/c_{ij}$ with $c_{ij} = c_{ji} > 0$ symmetric will work, provided $c_{ij} \geq p_j\, q_{ji}$ (necessary to ensure that $a_{ij} \leq 1$ and $a_{ji} \leq 1$, as required for acceptance probabilities!). One suitable choice is $c_{ij} := p_i\, q_{ij} + p_j\, q_{ji}$; the smallest possible choice, leading to the largest possible acceptance probabilities (and so the most mobile chain $\{\theta^t\}$), is $c_{ij} := \max(p_i\, q_{ij}, p_j\, q_{ji})$, leading to

$$a_{ij} := \frac{p_j\, q_{ji}}{p_i\, q_{ij} \vee p_j\, q_{ji}} = 1 \wedge H_{ij}, \qquad H_{ij} := \frac{p_j\, q_{ji}}{p_i\, q_{ij}}. \tag{4}$$

The general idea of constructing such a Markov chain is usually attributed to Metropolis et al. (1953), in the course of designing the first hydrogen bomb, who only considered symmetric proposals

$q_{ij} = q_{ji}$ leading to a simpler acceptance probability of $a_{ij} = 1 \wedge (p_j/p_i)$. The more general form is due to Hastings (1970), who studied failure probabilities for dams and in honor of whom H is called the *Hastings ratio*. The special case in which $H \equiv 1$ (now called "Gibbs sampling") was (re)discovered in an image reconstructing context by Geman and Geman (1984) and in a more general context by Gelfand and Smith (1990) (several others had similar ideas independently—*e.g.*, Tanner and Wong (1987) and Besag et al. (1995)). Tierney (1994) offers a particularly lucid exposition of the different ways to construct such chains.

By construction, $\{\theta^t\}$ is a stationary Markov chain on Θ with initial distribution $P^0(\theta_i) = p_i^0$ and transition probability matrix R_{ij}. It is easy to show that R will be transitive, irreducible, and aperiodic if Q is on $\text{supp}(P) \equiv \{\theta \mid P(\{\theta\}) > 0\}$ (and if $P^0\big(\text{supp}(P)\big) = 1$), so by the Perron-Frobenius theorem (see, for example, Horn and Johnson 1990, chap. 8)

$$\sup_j \Big| p_j - \mathsf{P}[\theta^t = \theta_j] \Big| \le r^t$$

for some $0 < r < 1$ (namely, the second-largest eigenvalue of R). This implies geometric convergence in Equation (1).

1.2 Euclidean Spaces

A similar approach holds for state spaces $\Theta \subset \mathbb{R}^d$ with Borel sets $\mathcal{F} = \mathcal{B}(\Theta)$. Here we must specify an initial distribution $P^0(d\theta)$ on $\mathcal{F}$ and transition kernel $Q(d\theta^* \mid \theta)$ on $\mathcal{F} \times \Theta$; we begin with initial value $\theta^0 \sim P^0(d\theta)$ and accept each proposed move from θ^t to $\theta^* \sim Q(d\theta^* \mid \theta^t)$ with probability $1 \wedge H(\theta^* \mid \theta^t)$ where

$$H(\theta^* \mid \theta) := \frac{P(d\theta^*)\, Q(d\theta \mid \theta^*)}{P(d\theta)\, Q(d\theta^* \mid \theta)},$$

the Radon-Nikodym derivative of two measures on $\Theta \times \Theta$— the denominator is the joint equilibrium probability distribution of $(\theta, \theta^*) = (\theta^t, \theta^{t+1})$, while the numerator is that of $(\theta, \theta^*) = (\theta^{t+1}, \theta^t)$. When P and Q have densities with respect to a common reference measure (such as Lebesgue measure $d\theta$ on $\mathbb{R}^d$), this reduces to a ratio of densities

$$H(\theta^* \mid \theta) := \frac{P(\theta^*)\, Q(\theta \mid \theta^*)}{P(\theta)\, Q(\theta^* \mid \theta)}. \tag{5}$$

Note that $H(\theta^* \mid \theta)$ depends on P only through the ratio $P(\theta^*)/P(\theta)$; this is an important feature for Bayesian posterior statistical inference, where the posterior distribution

$$\pi(d\theta \mid \mathbf{X}) \propto \pi(d\theta)\, L(\theta \mid \mathbf{X})$$

is often given only up to an unknown proportionality constant that cancels in Equation (5).

2 Poisson and Lévy Random Fields

We now turn our attention to constructing an ergodic Markov chain $\{\theta^t\}$ whose stationary distribution is absolutely continuous with respect to a Poisson random measure $\theta \sim \mathsf{Po}\big(\nu(dx)\big)$ on some measure space $\big(\mathcal{X}, \mathcal{B}, \nu(dx)\big)$, with some density function $L(\theta)$. For $\mathcal{X}$ of the form $\mathcal{X} = \mathbb{R} \times \mathcal{S}$ this will let us generate from the posterior distribution of a Lévy random field

$$\Gamma[\phi] = \int_{\mathcal{S}} \phi(\sigma)\, \Gamma(d\sigma) = \iint_{\mathcal{X}} \phi(\sigma)\, v\, \theta(dv\, d\sigma) = \sum \phi(\sigma_i) v_i$$

upon observing any data $\mathbf{Y}$ related by a measurement-error model to θ (represented through a likelihood function $L(\theta)$), such as:

Normal Regression:	$Y(t) \sim \mathsf{No}(f(t), \sigma^2),$	$f(t) := \Gamma[k(t, \cdot)]$
Gamma Regression:	$Y(t) \sim \mathsf{Ga}(f(t)\phi, \phi),$	$f(t) := \Gamma[k(t, \cdot)]$
Poisson Regression:	$Y(t) \sim \mathsf{Po}(f(t)\, dt),$	$f(t) := \Gamma[k(t, \cdot)]$
Survival:	$S(t) \sim e^{-H(t)}$	$H(t) := \Gamma[\mathbf{1}_{\{(0,t]\}}(\cdot)]$

The new wrinkle is that the space Θ of possible value of θ is more complicated than $\mathbb{R}^d$. One representation is to identify a finite integer-valued measure on $\mathcal{X}$ with the (superfluous but convenient) label J along with an ordered vector $\{x_j\}_{0 \le j < J}$ of the J (not necessarily distinct) points to which it assigns unit mass; thus

$$\Theta := \bigcup_{J=0}^{\infty} \mathcal{X}^J \ni \theta := (J; \{x_j\}_{0 \le j < J}),$$

the disjoint union of the J^{th} Cartesian power $\mathcal{X}^J$ over all integers $J \ge 0$.

2.1 Densities

Finding “densities” is more subtle here. If dx is a fixed reference measure on $\mathcal{X}$ (perhaps Lebesgue measure, if $\mathcal{X} \subset \mathbb{R}^d$), one possibility is to use

$$d\theta := \sum_{J=0}^{\infty} \mathbf{1}_{\{\Theta^J\}}(\theta)\, dx_0\, dx_2 \cdots dx_{J-1}$$

as a reference measure on Θ. The $\mathsf{Po}\big(\nu(dx)\big)$ distribution with mean measure $\nu(dx) = \nu(x)dx$ can then be represented

$$P(d\theta) = \frac{(\nu^+)^J}{J!} e^{-\nu^+} \prod_{0 \le j < J} \frac{\nu(dx_j)}{\nu^+} = \frac{e^{-\nu^+}}{J!} \prod_{0 \le j < J} \nu(dx_j)$$

where $\nu^+ \equiv \nu(\mathcal{X})$, with density function (w.r.t. $d\theta$)

$$P(\theta) = \exp\left\{ -\nu(\mathcal{X}) - \log J! + \sum_{0 \le j < J} \log \nu(x_j) \right\} \tag{6}$$

Notice that in this representation the $\{x_j\}$ are *ordered*, even though $P(d\theta)$ is symmetric; the $J!$ factor accounts for the multiple labelings the same point might have.

2.2 Transitions

To implement MCMC in a multi-dimensional space like Θ we must "jump" back and forth among the disjoint subspaces $\mathcal{X}^J$. The first implementation of such a scheme (and the name "reversible jump MCMC", or RJ-MCMC) appeared in (Green 1995), although our treatment is rather different.

We must build a transition probability kernel $Q(d\theta^* \mid \theta)$ to generate proposed moves on Θ that is transitive, irreducible, and aperiodic. Transitivity requires that we be able to reach any level $\mathcal{X}^J$ from any other $\mathcal{X}^I$; obviously it's enough to be able to increment $J \geq 0$ and decrement $J \geq 1$ by one. Incrementing entails a probability distribution $\beta(dx)$ for the "birth," which we take to have density function $\beta(x)$; movement within $\mathcal{X}^J$ can be built from any convenient Markov kernel $q(dx^* \mid x)$ on $\mathcal{B} \times \mathcal{X}$, or even from a *sub*-Markov kernel (*i.e.*, one for which $q(\mathcal{X} \mid x) \leq 1$— we just one can always extend it to be a Markov kernel on some $\tilde{\mathcal{X}} \supset \mathcal{X}$) if we regard a step "outside" of $\mathcal{X}$ as a "death." Specify a strictly probability triplet[1] $\mathbf{p} = (p_-, p_=, p_+)$ with $p_- + p_= + p_+ = 1$, a birth probability density $\beta(x)$ on $\mathcal{X}$, and a sub-Markov kernel $q(dx^* \mid x)$; with these in hand we describe transitions on Θ as follows. Beginning at $\theta = (J;\ \{x_j\}_{0 \leq j < J})$,

B Birth step: with probability p_+, draw an index $0 \leq j \leq J$ uniformly and a new point $x^* \sim \beta(x)dx$; set

$$\theta^* = (J{+}1;\ \{x_0, \ldots, x_{j-1}, x^*, x_j, \ldots, x_{J-1}\}).$$

D Death step: with probability p_- and $J \geq 1$, "kill" a point— draw an index $0 \leq j < J$ uniformly and set

$$\theta^* = (J{-}1; \{\, x_0, \ldots, x_{j-1}, \quad x_{j+1}, \ldots, x_{J-1}\}).$$

M Movement step: with probability $p_=$ and $J \geq 1$, draw $0 \leq j < J$ uniformly and a new point $x^* \sim q(x^* \mid x_j)\, dx$. If $x^* \in \mathcal{X}$ and $\nu(x^*) > 0$, set

$$\theta^* = (J;\ \{x_0, \ldots, x_{j-1}, x^*, x_{j+1}, \ldots, x_{J-1}\});$$

if $x* \neq \mathcal{X}$ (recall the *sub*-Markov transition may have $q(\mathcal{X} \mid x_j) < 1$) or $\nu(x^*) = 0$, treat this as the death of x_j, as in step **D** above.

Altogether the transition density w.r.t. $d\theta$ is:

$$Q(\theta^* \mid \theta) = \begin{cases} \mathbf{B}: & \frac{1}{J+1}\, p_+ \beta(x^*) \\ \mathbf{D}: & \frac{1}{J}\, [p_- + p_= q^-(x_j)] \\ \mathbf{M}: & \frac{1}{J}\, p_= q(x^* \mid x_j) \end{cases} \tag{7}$$

where $q^-(x) := \big[1 - q(\mathcal{X} \mid x)\big]$.

2.3 Hastings Ratio

We now construct the Hastings ratio $H(\theta^* \mid \theta)$ of Equation (5) from the ingredients in Equations (6, 7). If $L(\theta)$ is a likelihood function (or other expression for which $L(\theta)P(d\theta)$ is proportional

[1] Actually it's possible to have $\mathbf{p} = \mathbf{p}(J)$ depend on J... convenient to arrange $p_-(0) = p_=(0) = 0$, for example.

to the intended stationary distribution for our chain— *e.g.*, $L \equiv 1$ to draw samples from the prior distribution itself), the Hastings ratio is:

$$H(\theta^* \mid \theta) = \begin{cases} \mathbf{B}: & \frac{\nu(x^*)}{J+1} \frac{L(\theta^*)}{L(\theta)} \frac{[p_- + p_= q^-(x^*)]}{p_+ \beta(x^*)} \\ \mathbf{D}: & \frac{J}{\nu(x_j)} \frac{L(\theta^*)}{L(\theta)} \frac{p_+ \beta(x_j)}{[p_- + p_= q^-(x_j)]} \\ \mathbf{M}: & \frac{\nu(x^*)}{\nu(x_j)} \frac{L(\theta^*)}{L(\theta)} \frac{q(x_j | x^*)}{q(x^* | x_j)}. \end{cases} \tag{8}$$

Note that we needn't have required that $\nu(dx)$, $\beta(dx)$, and $q(dx \mid x^*)$ all have densities with respect to some specific measure dx, but we do need $\beta(dx) \ll \nu(dx)$; to allow "death" moves from anywhere in $\mathcal{X}$, the birth distribution $\beta(dx)$ must also have full support (so $\beta(dx) \equiv \nu(dx)$). Also note that the $\mathbf{M}$ step is simply the ratio of posterior densities in the (common) case of a symmetric proposal distribution with $q(y \mid x) = q(x \mid y)$.

3 Examples

3.1 Gamma RF in $\mathbb{R}^2$

The homogeneous Gamma random field $\Gamma(ds) \sim \mathsf{Ga}(\alpha ds, \beta)$ on the unit square $\mathcal{S} = [0,1]^2$ has infinite Lévy measure

$$\nu(du\,ds) = \alpha e^{-\beta u} u^{-1} \mathbf{1}_{\{u>0\}} du\,ds$$

on $\mathbb{R} \times \mathcal{S}$. To use the methods of Section (2) we must first approximate the distribution by one with finite Lévy measure. One way is to select a small number $\epsilon > 0$ and construct a random field with Lévy measure

$$\nu_\epsilon(du\,ds) = \alpha e^{-\beta u} u^{-1} \mathbf{1}_{\{u>\epsilon\}} du\,ds$$

on $\mathbb{R} \times \mathcal{S}$, with finite mass

$$\nu_\epsilon^+ := \nu_\epsilon(\mathbb{R} \times \mathcal{S}) = \alpha \int_\epsilon^\infty e^{-\beta u} u^{-1} du = \alpha \mathrm{E}_1(\beta\epsilon),$$

where $\mathrm{E}_1(z) := \int_z^\infty x^{-1} e^{-x} dx$ denotes Gauss's exponential integral function (Abramowitz and Stegun 1964, *p.* 228). We may view ν_ϵ as a measure on $\mathcal{X} := \mathbb{R}_+ \times \mathcal{S}$ and, from a Poisson random measure $H \sim \mathsf{Po}(\nu_\epsilon(dx))$, construct an approximate Gamma RF by setting

$$\begin{aligned} \Gamma(A) = \int_A \Gamma(ds) \qquad &= \iint_{\mathbb{R}_+ \times A} u N(du\,ds) \\ \Gamma[\phi] = \int_{\mathcal{S}} \phi(s) \Gamma(ds) &= \iint_{\mathbb{R}_+ \times \mathcal{S}} \phi(s)\, u N(du\,ds) \end{aligned}$$

3.1.1 Birth Steps

With a birth distribution $\beta(dx)$ for $x = (u, s)$ assigning independent exponentially distributed magnitudes $u \sim \epsilon + \mathsf{Ex}(\lambda)$ and uniformly distributed locations $s \sim \mathsf{Un}(\mathcal{S})$, the Lévy measure ν and birth distribution β are mutually absolutely continuous, with Lebesgue density functions

$$\nu(x) = \alpha e^{-\beta u} u^{-1} \mathbf{1}_{\{[\epsilon,\infty) \times \mathcal{S}\}}(x) \qquad \beta(x) = \lambda e^{-\lambda(u-\epsilon)} \mathbf{1}_{\{[\epsilon,\infty) \times \mathcal{S}\}}(x)$$

3.1.2 Movement Steps and Hastings Ratios

Any symmetric Markov random walk on $[\epsilon, \infty) \times \mathcal{S}$ (for example, one taking independent Gaussian steps in each of the three dimensions, with reflecting boundary conditions at $u \geq \epsilon$ and at $0 \leq s_i \leq 1$) yields $q^-(x) \equiv 0$ and a symmetric $q(x^* \mid x) = q(x \mid x^*)$; with unit likelihood $L(x) \equiv 1$ this leads to

$$H(\theta^* \mid \theta) = \begin{cases} \mathbf{B}: & \frac{\alpha\ p_-}{\lambda\ \exp(\lambda\epsilon)\ p_+} \frac{\exp((\lambda-\beta)u^*)}{(J+1)\ u^*} \\ \mathbf{D}: & \frac{\lambda\ \exp(\lambda\epsilon)\ p_+}{\alpha\ p_-} \frac{J\ u_j}{\exp((\lambda-\beta)u_j)} \\ \mathbf{M}: & \exp\big(\beta(u_j - u^*)\big)(u_j/u^*) \end{cases}$$

Conversely, independent normal random walks $s_i^* \mid s_i \sim \mathsf{No}(s_i, \sigma_s^2)$ for locations (which might step outside $\mathcal{S}$, leading to a "death") and log-normal $u^* \mid u \sim \mathsf{LN}(\log u, \sigma_u^2)$ for magnitudes (which might step below $u < \epsilon$, again leaving the domain), renders a subMarkov transition with

$$\begin{aligned} q^-(x) &= 1 - \Phi\Big(\log(u/\epsilon)/\sigma_u\Big) \times \left[\Phi\Big(\frac{1-s_1}{\sigma_s}\Big) - \Phi\Big(\frac{s_1}{\sigma_s}\Big)\right] \\ &\qquad\qquad\qquad\qquad\qquad \times \left[\Phi\Big(\frac{1-s_2}{\sigma_s}\Big) - \Phi\Big(\frac{s_2}{\sigma_s}\Big)\right] \\ q(x^* \mid x) &= \frac{1}{\sigma_u u^*}\varphi\Big(\frac{1}{\sigma_u}\log(\frac{u^*}{u})\Big) \times \frac{1}{\sigma_s}\varphi\Big(\frac{s_1^* - s_1^*}{\sigma_s}\Big) \times \frac{1}{\sigma_s}\varphi\Big(\frac{s_2^* - s_2^*}{\sigma_s}\Big) \end{aligned}$$

where $x = (u, s)$ and $x^* = (u^*, s^*)$; here $\varphi(z)$ and $\Phi(z)$ denote the pdf and CDF of the standard $\mathsf{No}(0, 1)$ distribution, respectively. Note our transition kernel $q(x^* \mid x)$ is subMarkov; we treat random walk steps that lead $x^* \notin \mathcal{S}$ or $u^* < \epsilon$ as the death of a point at x. These (along with $L(x) \equiv 1$) let us calculate the Hastings ratio $H(\theta^* \mid \theta)$ of Equation (8), all that's needed to generate a Markov chain $\{\theta^t\}$ and hence $\{\Gamma^t\}$ from the intended distribution:

$$H(\theta^* \mid \theta) = \begin{cases} \mathbf{B}: & \frac{\alpha\ [p_- + p_= q(x^*)]}{\lambda\ \exp(\lambda\epsilon)\ p_+} \frac{\exp((\lambda-\beta)u^*)}{(J+1)\ u^*} \\ \mathbf{D}: & \frac{\lambda\ \exp(\lambda\epsilon)\ p_+}{\alpha\ [p_- + p_= q(x_j)]} \frac{J\ u_j}{\exp((\lambda-\beta)u_j)} \\ \mathbf{M}: & \exp\big(\beta(u_j - u^*)\big) \end{cases}$$

4 Posterior ILM Sampling

As an alternative to the RJ-MCMC ϵ-truncation approach of Section (2), we can use the Inverse Lévy Measure algorithm of Wolpert and Ickstadt (1998a,b) in which a fixed number J of mass points are generated. The classic ILM approach begins by writing a Lévy measure ν on $\mathcal{X} = \mathbb{R}_+ \times \mathcal{S}$ in semidirect product form

$$\begin{aligned} \nu(dx) &= \nu_u(dr)\, \nu_s(ds \mid u) \\ \nu^+(r) &= \nu_u\big((u, \infty)\big) \\ \nu^{\leftarrow}(t) &= \inf\big\{r > 0 :\ \nu^+(r) \leq t\big\}\,. \end{aligned}$$

Now fix $J \in \mathbb{N}$ and draw the first J event times $0 < \tau_1 < \tau_2 < \cdots < \tau_J$ of a unit-rate Poisson process. Set

$$\begin{aligned} r_j &= \nu^{\leftarrow}(\tau_j) \\ s_j &\sim \nu_s(ds \mid r_j) \\ \Gamma(ds) &= \sum_{j=1}^{J} r_j \delta_{s_j}(ds). \end{aligned}$$

This sum with $J = \infty$ would have exactly the target Lévy distribution; since the $\{r_j\}$ are drawn in decreasing order, with finite $J < \infty$, it includes the J largest mass points and for that reason can be more efficient than some other approximate methods. For *posterior* sampling, a Metropolis-Hastings approach will be required— but, this time, with a fixed number J of mass points and so without need for reversible jumps.

When both $\nu_u(dr) = \nu_u(r)\,dr$ and $\nu_s(ds \mid u) = \nu_s(s \mid u)\,ds$ have density functions (wrt arbitrary reference measures dr and ds on $\mathbb{R}_+$ and $\mathcal{S}$, respectively), the prior pdf is available by change of variables from that of the $\{\tau_j = \nu^+(r_j)\}$:

$$\begin{aligned} &\tau_1, \ldots, \tau_J \sim e^{-\tau_J} \mathbf{1}_{\{0<\tau_1<\cdots<\tau_J\}} d\tau_1 \cdots d\tau_J \qquad \Rightarrow \\ &r_1, \ldots, r_J \sim \exp\big(-\nu^+(r_J)\big) |\nu^{+\prime}(r_1) \cdots \nu^{+\prime}(r_J)|\, \mathbf{1}_{\{0<r_J<\cdots<r_1\}} dr_1 \cdots dr_J \\ &s_1, \ldots, s_J \sim \nu_s(s_1 \mid r_1) \cdots \nu_s(s_J \mid r_J)\, ds_1 \cdots ds_J. \end{aligned}$$

With this and the likelihood ratio in hand, a M-H scheme can be constructed with only conventional moves of the $\{r_j\}$ (preserving order) and the $\{s_j\}$. Block moves (in which the entire vector $\vec{u}$ is replaced with another of the form $\nu^{\leftarrow}(\vec{\tau})$) are a good choice in some problems. The Hastings ratio for a move $\theta \to \theta^*$ for $\theta = (\vec{u}, \vec{s})$ is

$$H(\theta^* \mid \theta) = \frac{L(\theta^*)}{L(\theta)} \left\{ e^{\nu^+(r_J) - \nu^+(r_J^*)} \prod_{j=1}^{J} \frac{\nu^{+\prime}(r_j^*)\, \nu_s(s_j^* \mid r_j^*)}{\nu^{+\prime}(r_j)\, \nu_s(s_j \mid r_j)} \right\} \frac{Q(\theta \mid \theta^*)}{Q(\theta^* \mid \theta)}.$$

4.1 Explicit Example: ILM for Gamma Random Fields

Again we consider the Gamma $\mathsf{Ga}(\alpha ds,\ \beta)$ random field on $\mathcal{S} = [0,1]^2$. This time we use the ILM algorithm, with reflecting symmetric Gaussian random walk steps in $s_j \in \mathcal{S}$ and log-scale Gaussian random walk steps in $r_j \in \mathbb{R}_+$. For this example $\nu^+(r) = \alpha \mathrm{E}_1(\beta r)$ with derivative $\nu^{+\prime}(r) = -\alpha r^{-1} e^{-\beta r}$, while $Q(\theta \mid \theta^*)/Q(\theta^* \mid \theta) = \prod (r_j^*/r_j)$, so

$$H(\theta^* \mid \theta) = \frac{L(\theta^*)}{L(\theta)} \exp\Big(\alpha[\mathrm{E}_1(\beta r_J) - \mathrm{E}_1(\beta r_J^*)] + \beta \sum (r_j - r_j^*) \Big).$$

Random walk steps are allowed to change the ordering of the $\{r_j\}$; just sort after the proposed move, to ensure that $r_J^* = \min \left\{ r_j^* \right\}$. Since $\mathrm{E}_1(z) \approx -\log z - \gamma_e$ for small z,

$$h(\theta^* \mid \theta) \equiv \log H(\theta^* \mid \theta) \approx [\ell(\theta) - \ell(\theta^*)] + \alpha \log(r_J^*/r_J) + \beta \sum (r_j - r_j^*)$$

where $\ell(\theta) = -\log L(\theta)$. A good starting point is $\vec{u} = \{r_j\}$, $r_j = \exp(-\gamma_e - j/\alpha)/\beta$ (why?).

4.2 Explicit Example: ILM for α-Stable Random Fields

For $0 < \alpha < 1$, $\beta \in [-1, 1]$, $\gamma \in \mathbb{R}_+$, and $\delta = 0$, a random measure $\zeta(ds) \sim \mathsf{St_A}(\alpha, \beta, \gamma ds, 0)$ can be constructed on (say) the unit interval $[0, 1]$ by the ILM algorithm as

$$\zeta(ds) = \sum_{j<\infty} r_j \sigma_j \delta_{s_j}(ds)$$

for $\tau_0 = 0$ and

$$\begin{aligned} r_j &= (\tau_j/\gamma c_\alpha)^{-1/\alpha}, & [\tau_j - \tau_{j-1}] &\overset{\text{iid}}{\sim} \mathsf{Ex}(1) \\ \sigma_j &= (2\zeta_j - 1), & \zeta_j &\overset{\text{iid}}{\sim} \mathsf{Bi}\big(1, (1+\beta)/2\big) \\ s_j &\overset{\text{iid}}{\sim} \mathsf{Un}(\mathcal{S}) \end{aligned}$$

where $c_\alpha = \frac{2}{\pi}\Gamma(\alpha)\sin\frac{\pi\alpha}{2}$, or may be approximated by the first J terms of that sum. Here $\mathcal{S} = (0, 1) \times \{\pm 1\}$ with elements (s_j, σ_j). For *posterior* sampling as in Section (4), note $\nu_u^{+\prime}(r) = \gamma c_\alpha \alpha r^{-\alpha-1}$. For independent symmetric random walk (say, Gaussian) steps in r_j on a log scale, and srw steps in s_j (say, Gaussian w/ reflecting bc), and $\sigma_j = \pm 1$, the log Hastings ratio becomes

$$h(\theta^* \mid \theta) = \ell(\theta) - \ell(\theta^*) \;\; + \gamma c_\alpha (r_J^{-\alpha} - r_J^{*\,-\alpha}) + \alpha \sum_{j=1}^{J} \log(r_j/r_j^*) + \log\left(\tfrac{1+\beta}{1-\beta}\right) \sum_{j=1}^{J} (\sigma_j^* - \sigma_j)/2,$$

with $r_j = (\gamma c_\alpha/j)^{1/\alpha}$ and $\sigma_j = (2\zeta_j - 1)$, $\zeta_j \sim \mathsf{Bi}(1, (1+\beta)/2)$ a good starting point.

5 Dirichlet Random Fields

For any finite partition $\mathcal{S} = \cup \Lambda_j$ of a finite measure space $\big(\mathcal{S}, \mathcal{F}, \alpha(ds)\big)$ with $\alpha^+ \equiv \alpha(\mathcal{S}) < \infty$, the Dirichlet random field $\mathcal{D} \sim \mathsf{Di}\big(\alpha(ds)\big)$ assigns random variables $p_j = \mathcal{D}(\Lambda_j)$ whose joint distribution is Dirichlet $\vec{p} \sim \mathsf{Di}(\vec{\alpha})$ with parameter vector $\vec{\alpha} = \{\alpha_j\}$, $\alpha_j = \alpha(\Lambda_j)$. Dirichlet RFs are frequently used to model uncertain probability distributions, because they're easy to interpret (the mean and variance are $\mathsf{E}\mathcal{D}(A) = \alpha(A)/\alpha(\mathcal{X})$ and $\mathsf{V}\mathcal{D}(A) = \alpha(A)\alpha(A^c)/\alpha(\mathcal{X})^2(1 + \alpha(\mathcal{X}))$, so α/α^+ is the "prior mean" and α^+ quantifies prior precision) and trivial to compute with (they're conjugate for observations $X_j \sim \mathcal{D}$).

A Dirichlet random field can be constructed by normalizing the Gamma RF of Sec. (3.1) or (4.1):

$$\mathcal{D}(A) = \Gamma(A)/\Gamma(\mathcal{S})$$

for $A \subset \mathcal{S}$, with $\Gamma(ds) \sim \mathsf{Ga}\big(\alpha(ds), \beta\big)$ for any constant $\beta > 0$ (say, one)— the constant cancels when we normalize.

BUT— the Dirichlet has several unfortunate features that limit its utility. One is its discreteness ($\mathcal{D}$ is a discrete distribution with probability one, so even if α has a density it is certain that observations $\{X_n\} \sim \mathcal{D}$ will feature ties), and another is the constancy of its precision α^+, which precludes assigning "vaguer" prior distributions in some parts of $\mathcal{S}$ than in others. The discreteness can be overcome by taking kernel mixtures $\int k(x, s)\, \mathcal{D}(ds)$, at the expense of losing the computational triviality, while the uniform precision can be overcome by replacing the constant β by a function $\beta(s)$ above; the same computational approach described in Section (3.1) with Lebesgue measure replaced by $\alpha(ds)$ and the constant β by a function $\beta(s)$, leading to

$$\nu(dr\, ds) = \alpha(ds) e^{-\beta(s) r}\; r^{-1} \mathbf{1}_{\{r>0\}} dr$$

will suffice to generate prior and posterior distributions for a generalization of $\mathcal{D}(ds)$.

References

Abramowitz, M. and Stegun, I. A., eds. (1964), *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, *Applied Mathematics Series*, volume 55, Washington, D.C.: National Bureau of Standards.

Besag, J., Green, P. J., Higdon, D., and Mengersen, K. (1995), "Bayesian computation and stochastic systems (with discussion)," *Statistical Science*, 10, 3–66.

Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398–409.

Geman, S. and Geman, D. (1984), "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Green, P. J. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711–732.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97–109.

Horn, R. A. and Johnson, C. R. (1990), *Matrix Analysis*, Cambridge, UK: Cambridge University Press.

Metropolis, N. C., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, 21, 1087–1092.

Tanner, M. A. and Wong, W. H. (1987), "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, 82, 528–550.

Tierney, L. (1994), "Markov chains for exploring posterior distributions (with discussion)," *Annals of Statistics*, 22, 1701–1762.

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," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. D. K. Dey, P. Müller, and D. Sinha, New York, NY: Springer-Verlag, *Lecture Notes in Statistics*, volume 133, pp. 227–242.

Appendix: Inference for Poisson Random Measures

Let $\nu_\theta(dx)$ be a family of *finite* nonnegative Borel measures on a complete separable metric space ("Polish" space) $\mathcal{X}$, indexed by $\theta \in \Theta$. In this section we consider the problem of finding a likelihood function for θ, upon observing a Poisson random field $N(dx) \sim \mathsf{Po}\big(\nu(dx)\big)$. Begin with the assumption that some single σ-finite Borel reference measure $\mu(dx)$ dominates $\nu_\theta(dx)$ for *each* $\theta \in \Theta$, and that a regular conditional probability density function exists so that

$$\nu_\theta(dx) = \nu(x,\theta)\,\mu(dx)$$

for a Borel measurable function $\nu : \mathcal{X} \times \Theta \to \mathbb{R}_+$.

For any partition $\mathcal{X} = \cup\Lambda_j$ into disjoint Borel sets with $\bar{\Lambda}_j$ compact, each $\lambda_j(\theta) \equiv \nu_\theta(\Lambda_j)$ and $\mu_j \equiv \mu(\Lambda_j)$ is finite. The random variables $N_j \equiv N(\Lambda_j)$ are independent, each Poisson distributed with mean $\nu_j(\theta)$, so the likelihood $L(\theta)$ upon observing all the $\{N_j\}$ would be any nonnegative multiple of

$$\begin{aligned} L(\theta) &= \prod_j \left\{ \frac{\nu_j(\theta)^{N_j}}{N_j!}\; e^{-\nu_j(\theta)} \right\} \\ &\propto \left\{ \prod_j \left(\frac{\nu_j(\theta)}{\mu_j} \right)^{N_j} \right\} e^{-\sum \nu_j(\theta)} \end{aligned}$$

Enumerate the (random and countable) support $\{x_n\}$ of $N(dx)$, and let j_n be the index of the partition element Λ_{j_n} containing x_n. Then

$$L(\theta) = \left\{ \prod_n \left(\frac{\nu_{j_n}(\theta)}{\mu_{j_n}} \right)^{N_{j_n}} \right\} e^{-\nu_\theta(\mathcal{X})}$$

Now take successive refinements of the partition $\{\Lambda_j\}$ with $\mathrm{diam}(\Lambda_j) \to 0$. Since every Polish space is Radon, it follows that $\nu_{j_n}(\theta)/\mu_{j_n} = \nu_\theta(\Lambda_{j_n})/\mu(\Lambda_{j_n})$ converges to $\nu(x_n,\theta)$, so

$$\to e^{-\nu_\theta(\mathcal{X})} \prod_n \nu(x_n,\theta).$$

Note that our requirement that each $\nu^+(\theta) \equiv \nu_\theta(\mathcal{X}) < \infty$ was necessary for this to be well-defined. Also the formula remains correct even if, for some θ, ν_θ (and hence μ) has atoms; in that case some of the $\{x_n\}$ may coincide. Both Bayesian and sampling-based inference about θ now depend on the data only through the negative log likelihood function,

$$\ell(\theta) = -\log L(\theta) = \nu^+(\theta) - \sum \log \nu(x_n,\theta).$$