Modeling and Computation with Lévy RFs

Robert L. Wolpert

February 18, 2016

1 Motivation

A common flexible way to construct stationary time series (discrete-time stochastic processes) is to begin with an iid sequence $\{\zeta_i : i \in \mathbb{Z}\}$ from an arbitrary distribution and a suitable set $\{b_i\}$ of coefficients and set

$$f(i) := \sum b_{i-j}\zeta_j$$

for $i \in \mathbb{Z}$. These "moving average" or MA processes may be used as prior distributions for uncertain functions on (any subset of) $\mathbb{Z}$; their means, covariances, ch.f.s, *etc.* are all easily computed from those of the $\{\zeta_i\}$.

A natural extension of this idea to *continuous* time would be to set

$$f(t) := \int_{\mathbb{R}} b(t-s)\zeta(ds)$$

for a suitable *function* b and *random measure* $\zeta(ds)$. The analogue of "iid" for the $\{\zeta_i\}$ above is that $\zeta(ds)$ should assign independent random variables to disjoint sets, whose distributions are translation-invariant.

The requirements that

$$\zeta(A) \perp\!\!\!\perp \zeta(B) \qquad \zeta(t+B) \sim \zeta(B)$$

for disjoint A, B and $t \in \mathbb{R}$ imply that $\zeta_t \equiv \zeta\big((0,t]\big)$ is an SII process. This can be extended immediately to non-stationarity (replacing $b(t-s)$ with $k(t,s)$, not necessarily a function only of $(t-s)$), and beyond $\mathbb{R}$ for both t and (separately) for s.

2 Eg 1: Nonlinear Regression

Let $\mathcal{X}$ be a compact interval in $\mathbb{R}^1$ or rectangle in $\mathbb{R}^2$ (or any other Polish space) and consider the problem of estimating some unknown function $f : \mathcal{X} \to \mathbb{R}$, from noisy measurements

$$Y_i \overset{\text{ind}}{\sim} \mathsf{No}\big(f(x_i),\ 1\big), \quad i \in I$$

at specified design points $\{x_i\}_{i\in I}$ with known variance (say, one)[1]. One way to proceed is to model the uncertain function $f(x)$ with a "LARK" model (Wolpert et al., 2011) for the prior. In that

[1] In fact, any measurement error model will do— all we need is a family of distributions $g(dy \mid \theta)$ for Y, completely determined by some parameter θ, all with density functions ("likelihoods", for us) $g(y \mid \theta)$. In our nonlinear regression context, the parameter θ is taken to be an uncertain function $f(x_i)$ of the explanatory variable vector x.

approach one chooses a family of basis or kernel functions such as

$$k(x,\omega) = e^{-\lambda(x-\chi)}\mathbf{1}_{\{\chi<x\}}$$

on $\mathcal{X}\times\Omega$ for some convenient space Ω (here $\omega=(\chi,\lambda)\in\Omega = X\times\mathbb{R}_+$, for locations $\chi\in\mathcal{X}$ and decay rates $\lambda\ge 0$) and represents $f(x)$ in the form

$$f(x) = \sum \upsilon_j k(x,\omega_j) \tag{1a}$$

as a linear combination of the kernel functions with random coefficients υ_j and parameters ω_j (typically determining the location and shape)— or, a little more generally, the form

$$= \int_\Omega k(x,\omega)\,\zeta(d\omega) \tag{1b}$$

for a random measure $\zeta(d\omega)$ on some Polish space Ω. Choose a family $\{k(\cdot,\omega)\}$ to make (1a) efficient.

The original motivation (mimicing MA time series— see Section (1)) led to ID random measures $\zeta(d\omega)$ on $\mathbb{R}$; even in the present more general setting (where $\mathcal{X}$ and Ω are arbitrary "lccb" spaces, not necessarily $\mathbb{R}$, and where $k(x,\omega)$ need not be of convolution form), it's convenient to consider measures $\zeta(d\omega)$ that assign independent ID random variables to disjoint sets in Ω. One way to achieve that is to select a Lévy measure $\nu(d\upsilon\,d\omega)$ on $\mathbb{R}\times\Omega$ that satisfies

$$\int_{\mathbb{R}\times K} \big(1\wedge|\upsilon|\big)\,\nu(d\upsilon\,d\omega) < \infty \tag{2}$$

for each compact $K\subset\Omega$, construct a Poisson random measure $H\sim\mathsf{Po}\big(\nu(d\upsilon\,d\omega)\big)$ on $\mathbb{R}\times\Omega$, and set

$$\begin{aligned} f(x) &= \int_\Omega k(x,\omega)\,\zeta(d\omega) \\ &= \int_{\mathbb{R}\times\Omega} k(x,\omega)\,\upsilon\,H(d\upsilon\,d\omega). \end{aligned} \tag{3}$$

This will of course have ch.f.

$$\mathsf{E}\exp\{i\omega f(x)\} = \exp\left\{\int_{\mathbb{R}\times\Omega}\big(e^{i\omega k(x,\omega)\,\upsilon}-1\big)\,\nu(d\upsilon\,d\omega)\right\} \tag{4}$$

from which means, variances, *etc.* are available; also, this shows that the distribution of $f(x)$ itself is also ID. For example, we can arrange for $\zeta(A)\sim\mathsf{Ga}(\alpha(A),\beta)$ for some σ-finite measure $\alpha(d\omega)$ on Ω, some $\beta>0$, and all $A\subset\Omega$ of finite α-measure by taking

$$\nu(d\upsilon\,d\omega) = \alpha(d\omega)e^{-\beta\upsilon}\upsilon^{-1}\mathbf{1}_{\{\upsilon>0\}}d\upsilon.$$

If $k(x,\omega)=b\mathbf{1}_B(x,\omega)$ is constant on some set B in $\mathbb{R}\times\Omega$ and zero elsewhere then $f(x)$ will have a Gamma distribution too (figure out the parameters), but in general it would be a linear combination of Gammas with different rate parameters.

Or, we can arrange for $\zeta(A) \sim \mathsf{St}_\alpha\big(\alpha, \beta, \gamma(A), \delta(A)\big)$ for some constants $0 < \alpha < 1$ and $-1 \le \beta \le 1$ and σ-finite measures $\gamma(d\omega)$ and $\delta(d\omega)$ by taking

$$\nu(dv\, d\omega) = c_\alpha(1 + \beta \operatorname{sgn} v)\alpha|v|^{-1-\alpha} dv\, \gamma(d\omega)$$

for $c_\alpha = \frac{1}{\pi}\Gamma(\alpha)\sin\frac{\pi\alpha}{2}$, in which case $f(x)$ will have an α-stable distribution too (find the parameters; note that if $\delta \neq 0$ we must add the non-random offset $\int_\Omega k(x,\omega)\delta(d\omega)$ to Eqn (1b)).

SO, think of f as the sum Eqn (1a), with a random number of terms each of which has a random coefficient v_j and attribute (location, shape, etc.) ω_j.

2.1 Generating f from the Prior Distribution

If ν is finite then Eqn (3) already shows just how to draw random functions f from this distribution. For infinite ν, both approximation methods we've discussed in class will work— either draw the v_j in monotonically decreasing order of absolute value using the ILM algorithm, or fix a "cut-off" $\epsilon > 0$ and approximate f by f^ϵ based on the finitely-many (say, J_ϵ) mass points of H on $B_\epsilon^c \times \Omega$. The martingale methods we've used for computing bounds on the approximation errors carry through.

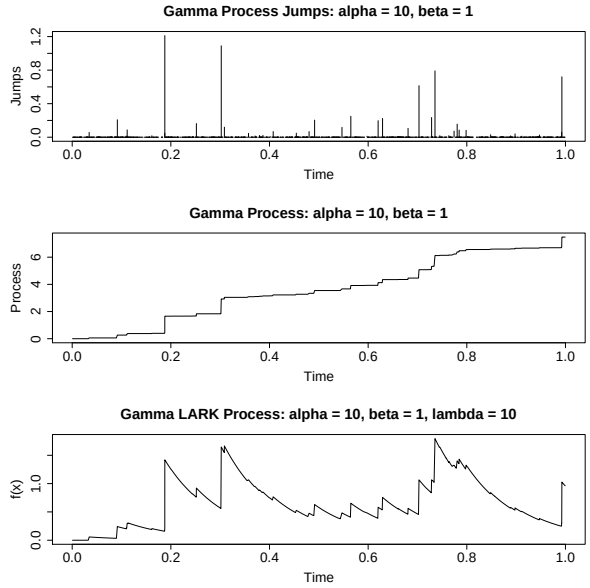

Figure 1: Top: sample draw from $\zeta(dt) \sim \mathsf{Ga}(\alpha\, dt, \beta)$ random measure on unit interval $[0,1]$, with $\alpha = 10$, $\beta = 1$. Middle: SII Gamma process $X_t = \zeta(0,t]$. Bottom: Stationary LARK process $f(t) = \int_0^t e^{-\lambda(t-s)}\zeta(ds)$, with $\lambda = 10$.

More interesting is:

2.2 Posterior Distributions

Consider a data-set $\{(x_i, Y_i)\}_{i\in I}$ with the iid Normal error model described above. If we fix a space Ω, a kernel $k(x,\omega)$, and a Lévy measure $\nu(dv\, d\omega)$, we have a complete Bayesian model for f. How can we find the *posterior* distribution?

Usually in Bayesian analysis we begin with some observation vector $\mathbf{Y}$ (we've got one!) and a parametric family of probability distributions $\{f(y \mid \theta)\}$ for $\mathbf{Y}$ (still okay— the iid normal model), indexed by a parameter vector θ from some set Θ of possible values (uh oh, what's θ here?). Then we need a prior probability distribution $\pi(d\theta)$ on Θ (all this ID distribution stuff must have something to do with that, right?), and must do some integration (or simulation) to evaluate (or draw samples from) the posterior distribution.

Nonparametric Bayesian analysis is more involved because the space Θ is big and sometimes unwieldy, and the business of building priors on big sets Θ is more challenging.

What is Θ here? Somehow it's the space of "all possible regression functions f", but we'll need to be much more explicit and specific.

For the ϵ approximations to the LARK models above, each possible f^ϵ is determined by:

- J_ϵ, the number of terms to include in the sum Eqn (1a);
- $\{\upsilon_j\}_{j\le J_\epsilon}$, the magnitudes of the terms;
- $\{\omega_j\}_{j\le J_\epsilon}$, the attributes of the terms.

The set of all possible values "θ" of this form is:

$$\Theta = \cup_{J=0}^{\infty}\big[\mathbb{R}\times\Omega\big]^J,$$

the union over all possible values of J (including zero) of all the configurations with exactly J terms.

Now we need to specify the **prior density** on Θ, evaluate the **likelihood function** on Θ, and implement an MCMC scheme to draw replicated samples from the *posterior* distribution. For that, take a look at the class notes "mh.pdf", *Metropolis-Hastings for Lévy Random Fields*.

References

Wolpert, R. L., Clyde, M. A., and Tu, C. (2011), "Stochastic Expansions using Continuous Dictionaries: Lévy Adaptive Regression Kernels," *Annals of Statistics*, 39, 1916–1962, doi: 10.1214/11-AOS889.

Last edited: February 18, 2016