

Sampling Shortcuts in Two Gibbs Samplers for Spatial Models

Levi John Wolf

May 13, 2016

In reviewing and improving implementations of the spatial hierarchical models, I noticed errors in the statistical assumptions of the inversion and M-H samplers used in Dong & Harris (2014) and Lacombe & McIntyre (2016), respectively. These issues only affect sampling of the spatial parameters, but will, in general, affect the performance of the entire sampler. I hope the following discussion illuminates these issues so they are avoided in future.

0.1 Dong & Harris Gridded Gibbs

Dong & Harris (2014) use a gridded inversion sampling method they ascribe to the Bayesian spatial probit sampler in Smith & LeSage (2004). In short, the implementation error is unique to Dong & Harris’s implementation of the sampler, and affects samples of both spatial parameters.

The sampler must evaluate a density for the lower-level spatial parameter, ρ , with the following non-standard form:

$$\rho|\beta, \Theta, \sigma_e^2, \sigma_u^2, \lambda \sim |A| \exp \left\{ -(2\sigma_e^2)^{-1} (Ay - X\beta - \Delta\Theta)' (Ay - X\beta - \Delta\Theta) \right\}$$

In this notation, yy is the response variable, A is the standard Laplacian matrix, $I - \rho\mathbf{W}$, X contains all lower- & upper-level exogenous covariates, Δ is the matrix that categorizes all lower-level observations into upper-level units, and Θ is the vector of spatially correlated random effects. Let us call the probability density function shown above \mathcal{D} , and let its cumulative distribution be Φ .

To sample $\rho \sim \mathcal{D}$ the log determinant, $\ln(|A|)$, is first evaluated over a grid of possible ρ values. This allows precomputation of the log determinant, possibly improving performance. Then, the kernel of the density shown above is evaluated at grid points. With this vector of probabilities, Φ is approximated using a simple cumulative sum. The next draw of ρ , called ρ_{new} , conditional on the rest of the sampler state is:

$$\rho_{new} | (\dots) = \hat{\Phi}^{-1}(u) \quad u \sim Unif(0, 1)$$

In theory, this generates consistent estimates of ρ when the grid values approximate all of \mathcal{D} well. Smith & LeSage (2004) correctly suggest evaluating \mathcal{D} over

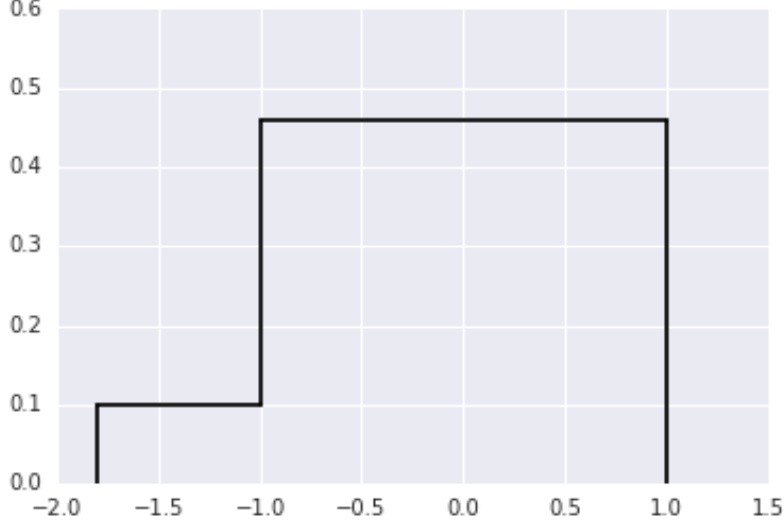


Figure 1: Arbitrary probability density function for ρ_{new} with some amount of probability residing outside of the inversion sampling range

a grid with boundaries determined by the minimal and maximal eigenvalues, $[e_{min}^{-1}, e_{max}^{-1}]$, and such an approximation should result in consistent estimates of parameters. Instead of this, Dong & Harris (2014) evaluate the grid over a domain of $[-1, 1]$. Intuitively, this is reasonable, because the largest possible ρ value for a row-standardized weights matrix is 1. In addition, ρ values less than -1 imply strange behavior, and ρ is usually constrained to be above -1 because of this.

But, in the case where $e_{min}^{-1} < -1$, this choice of bounds will introduce positive bias into the estimates. This is relatively easy to show. Let the maximal lower value of ρ according to the distribution above be l and the upper bound be r . Then, the total probability Dong & Harris (2014) *do not* sample from is:

$$\Phi(l) + (1 - \Phi(r))$$

A simple example is shown in Figure 1, with a vastly simplified \mathcal{D} . The corresponding Φ are shown in 2. The two are related through the truncated probability:

$$\Phi^*(x) = \frac{\Phi(x)}{1 - (\Phi(l) + (1 - \Phi(r)))}$$

Thus, for a given u draw in an inversion sampling step, $\Phi^*(x)$ will be larger than $\Phi(x)$. This is also shown in 2 by the difference in the new parameter values for a given draw of u .

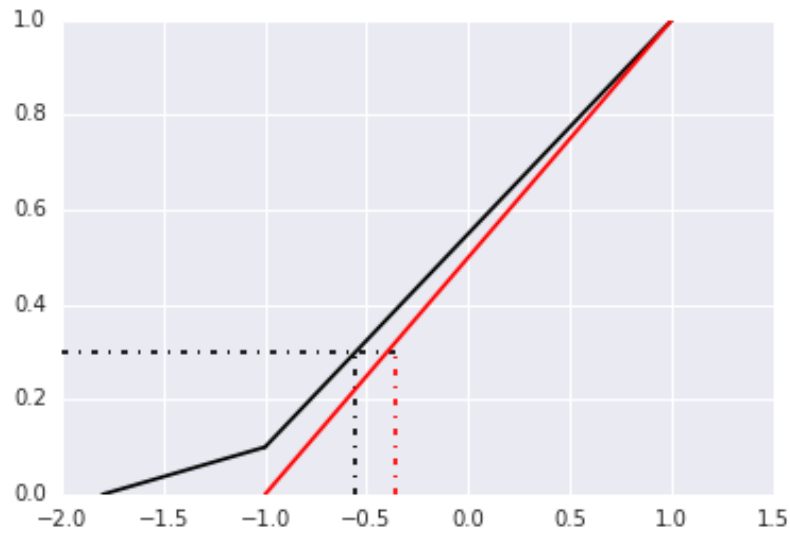


Figure 2: True cumulative distribution function Φ (black) and truncated distribution function Φ^* (red). At almost every possible inversion sample, a draw from the red truncated cumulative density function will dominate the black true cumulative density function. Parameters estimated using Φ^*-1 will exhibit positive bias.

0.2 Lacombe & McIntyre Metropolis

Lacombe & McIntyre’s provided code exhibits a similar domain-based issue in their Metropolis-Hastings implementation. The simplest Metropolis-Hastings deployment samples ρ_{new} from \mathcal{D} by adding some random step δ to ρ . The random shock is often a normally-distributed innovation with variance adapted according to the rejection rate. Then, this new value is accepted with the acceptance probability:

$$A = \min \left\{ 1, \frac{\mathcal{D}(\rho_{new})}{\mathcal{D}(\rho)} \frac{f(\rho|\rho_{new})}{f(\rho_{new}|\rho)} \right\}$$

for a symmetric proposal distribution in continuous space, the farthest right term, sometimes called the “proposal scaling factor,” cancels out, since the probability of moving from ρ to ρ_{new} is the same as the probability of moving from ρ_{new} to ρ . Then, to finish the draw, another uniform random test value like the u discussed above is drawn. The proposed ρ_{new} is accepted if $A > u$ and ρ_{new} is assigned to ρ , or is rejected if $A < u$ and the sampler proceeds.

The M-H deployment in Lacombe & McIntyre’s code adds a seemingly reasonable innovation: if $\rho + \delta$ is outside $[-1,1]$, simply draw another δ in the same step. If ρ_{new} is not in a valid part of the domain, it will be rejected anyway. By re-drawing δ until ρ_{new} is valid, we will accept more ρ , and reduce the waste of drawing and rejecting many invalid ρ_{new} .

Unfortunately, this means the scaling factor is not one, since $f(\rho|\rho_{new}) \neq f(\rho_{new}|\rho)$. This becomes clear when you consider what happens to the set of valid transitions when ρ is very close to the boundary. If ρ_{new} is *away* from the boundary, its truncation will be different, resulting in a different probability of moving from ρ_{new} to ρ (i.e. $f(\rho_{new}|\rho)$). That is, the probability of moving from ρ to ρ_{new} will be higher than that of ρ_{new} to ρ , since the truncation is much more aggressive at ρ than ρ_{new} .

This may or may not induce consistent bias in the parameter estimates, but can cause unexpected behavior in the sampling routine. While this statistical incorrectness is in pursuit of sampling efficiency, it is unlikely that the efficiency gained is really *worth* it, since the gains to efficiency (and statistical incorrectness) only occur when the sampler takes many draws close to the boundary. In addition, since Lacombe & McIntyre also assume spatial parameters below -1 are invalid *prima facie*, it is likely that this compounds with the same issue noted above.