A NOTE ON MONTE CARLO MODEL SENSITIVITIES TO DISTRIBUTIONS SAMPLED BY REJECTION

JACQUES DU TOIT

Abstract. Rejection sampling is a powerful method for sampling from many distributions. Unfortunately the method is very sensitive to the parameters of the distribution: changing them even slightly can lead to an entirely different sequence of random numbers being produced. This is especially problematic when computing model sensitivities using Monte Carlo integration. We briefly illustrate how recent research into quantiles of distributions can be adapted to deal with this problem.

1. Introduction

The rejection method is a well known and very powerful tool for drawing random numbers from various distributions. Its great strength is that it allows us to sample from virtually any distribution for which we have a density function. For many complicated distributions it is the only method we have of generating random numbers.

The downside of the rejection method is traditionally seen as being its speed: it is often not very efficient. The method requires one to throw away (or *reject*) some of the samples that have been generated, and so some of the computing effort is inevitably wasted. Much research has been devoted to reducing the proportion of samples that are rejected when sampling from various distributions.

However another problem is that rejection methods are very sensitive to the parameters of the distribution being sampled from. To illustrate, consider sampling from a Normal distribution with mean 0 and variance 1 via the inverse cumulative distribution function (CDF) method. We might obtain the following sequence of random numbers:

$$-1.2051 \quad 1.0102 \quad 0.05104 \quad -3.5096 \quad 1.7015 \quad 0.5065 \quad -0.4571 \quad \ldots$$

Re-initializing the generator to the same seed and sampling from a Normal distribution with mean 0.01 and variance 1 would then give

$$-1.1951 \quad 1.0202 \quad 0.06104 \quad -3.4996 \quad 1.7115 \quad 0.5165 \quad -0.4471 \quad \ldots,$$

in other words, a small perturbation in the distribution parameters results in a small perturbation in *each random number*. Now consider sampling from a gamma distribution with parameters $\alpha = 1$ and $\beta = 1$ by rejection. We might obtain the following sequence of random numbers:

$$2.2150 \quad 1.1105 \quad 2.0571 \quad 0.0052 \quad 0.9841 \quad 1.9015 \quad 1.4002 \quad \ldots$$

If we now re-initialize the generator to the same seed and sample from a gamma distribution with parameters $\alpha = 1.01$ and $\beta = 1$, we might obtain

$$2.2234 \quad 1.1211 \quad 2.0631 \quad 1.0549 \quad 2.9102 \quad 1.0014 \quad 0.0517 \quad \ldots.$$

Some time after the third gamma random number is generated, a sample which was previously below the rejection envelope is pushed above it and is rejected, and from that point on the two sequences will be completely different.

This becomes a problem when computing the sensitivity of a model to the parameters of the distribution. Suppose that $X_{\eta,n} = \big(x_1(\eta), x_2(\eta), \ldots, x_n(\eta)\big)$ is a sample of n random

numbers from some distribution with parameter η and let $H_{\eta,n} \equiv H(X_{\eta,n})$ denote a Monte Carlo estimator based on the sample $X_{\eta,n}$. One way of approximating $\partial H_{\eta,n}/\partial\eta$ is to compute

$$\frac{\partial H_{\eta,n}}{\partial \eta} \approx \frac{H_{\eta+\varepsilon,n} - H_{\eta,n}}{\varepsilon} = \frac{H(X_{\eta+\varepsilon,n}) - H(X_{\eta,n})}{\varepsilon}. \tag{1}$$

for some small $\varepsilon > 0$. If $x_i(\eta) \approx x_i(\eta+\varepsilon)$ for each $i = 1, 2, \ldots, n$, then the approximation (1) would be reasonably accurate, and averaging a few of these estimates could produce an overall estimate which is reasonably stable. However if $x_i(\eta)$ is completely different from $x_i(\eta+\varepsilon)$, then noise from the Monte Carlo estimation itself enters (1) and may completely drown the perturbations in the model that we wish to observe. Typically the only way to get around this is to increase n significantly so that the Monte Carlo noise is small in comparison to the model perturbations.

When the underlying random numbers are generated by rejection, computing model sensitivities to the distribution parameters can be problematic.

2. Quantile Mechanics and Recycling

Recent research (see [1]) into computing quantiles of various distributions may prove useful in tackling this problem. The idea is very simple. Suppose we have a distribution with cumulative distribution function (CDF) F that we wish to sample from. The so-called inverse-CDF method would set

$$X = F^{-1}(U) \tag{2}$$

where U is a random variable with a $U(0,1)$ distribution, so that the random variable X now has CDF given by F. However we can also set

$$X = F^{-1}\big(G(G^{-1}(U))\big) \tag{3}$$

for any smooth monotonic function G. If we set $J = F^{-1} \circ G$, then (2) becomes

$$X = J\big(G^{-1}(U)\big). \tag{4}$$

This procedure is termed *quantile mechanics* in [1], and the random variable $G^{-1}(U)$ is said to be *recycled* into the target distribution F.

If G is chosen judiciously, then $G^{-1}(U)$ is cheap to compute and J is well approximated by rational functions over almost all of its domain. As an example, consider the inverse error function `erfinv`. Typically `erfinv` is approximated by splitting the domain into three different sub-domains, with a different rational approximation in each sub-domain. On any SIMD-based (single instruction multiple data) architecture such as GPUs or SSE/AVX units on modern CPUs, this will lead to problems when `erfinv` is used to transform a sample of $U(0,1)$ random numbers into a sample of Normal random numbers. On NVIDIA GPUs, for example, it will lead to *warp divergence* and it was this divergence which led the authors of [1] to re-examine the approximation of `erfinv`. They show that by choosing G so that J is approximated well by a single rational function over virtually all of its domain, the proportion of divergent warps can be reduced.

It is possible to adapt the ideas in [1] to the present problem of computing sensitivities to distributions sampled via rejection. Suppose we have a sample $X_{\eta,n} = \big(x_1(\eta), x_2(\eta), \ldots, x_n(\eta)\big)$ of random numbers from a distribution with parameter $\eta > 0$ that are sampled by rejection (or indeed by any sampling method). Let F_η denote the CDF of the distribution and define $U_{\eta,n} = \big(u_1(\eta), u_2(\eta), \ldots, u_n(\eta)\big)$ by setting

$$u_i(\eta) := F_\eta\big(x_i(\eta)\big) \tag{5}$$

for each $i = 1, \ldots, n$ so that $U_{\eta,n}$ is an independent and identically distributed sequence of $U(0,1)$ random numbers.

Suppose that we now wish to perturb η to $\eta + \varepsilon$ for some small $\varepsilon > 0$ in order to compute sensitivities. Ideally then we would want to generate a *dependent* sample $X_{\eta+\varepsilon,n} = \big(x_1(\eta+\varepsilon), x_2(\eta+\varepsilon), \ldots, x_n(\eta+\varepsilon)\big)$ from our current sample $X_{\eta,n}$ by setting

$$x_i(\eta+\varepsilon) = F^{-1}_{\eta+\varepsilon}\big(u_i(\eta)\big) = F^{-1}_{\eta+\varepsilon}\big(F_\eta\big(x_i(\eta)\big)\big) := J\big(x_i(\eta)\big) \tag{6}$$

for all $i = 1, \ldots, n$ where

$$J(x) := F^{-1}_{\eta+\varepsilon} \circ F_\eta(x). \tag{7}$$

In other words, we are *recycling* $X_{\eta,n}$ into $X_{\eta+\varepsilon,n}$. If ε is small, then F_η and $F_{\eta+\varepsilon}$ should be close together so that the function J should be well approximated by a single rational function. This should allow an efficient method for transforming one sample of random numbers into a dependent sample from a slightly perturbed distribution.

3. Closing Remarks

NAG is interested to hear from individuals, academics and practitioners about whether or not this approach is one that could be used in practice for commercial or research work. This methodology is one we are interested in exploring. In particular, we are interested in collaborating with anyone who has a use case for this approach, so that we can better evaluate client needs and ensure the resulting routines are fit for purpose.

References

[1] Steinbrecher, G. and Shaw, W. T (2008). Quantile Mechanics. European Journal of Applied Mathematics, Vol 19, Issue 2.

Numerical Algorithms Group Ltd
E-mail address: jacques@nag.co.uk