Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

SIMULATION FROM NON-STANDARD DISTRIBUTIONS USING ENVELOPE METHODS

Michael J. Evans

Department of Statistics
University of Toronto
Toronto, Ont M5S3G3, CANADA

Tim B. Swartz

Department of Mathematics and Statistics
Simon Fraser University
Burnaby, BC V5A1S6, CANADA

ABSTRACT

This paper considers the development of envelope methods as a tool for simulation. Envelope methods are based on the construction of simple envelopes to functions. The proposed envelopes are general, require little input from the user and are based on the concavity structure of the function or some transformation of the function. The construction of these envelopes facilitates variate generation using the adaptive rejection algorithm.

1 INTRODUCTION

Simulation is fundamental to the investigation of complex stochastic systems. To that end, various general techniques have been developed to simulate from particular distributions. For example, inversion and the ratio-of-uniforms algorithm are two long-standing strategies for simulating from continuous univariate distributions (see Fishman, 1996). Nevertheless, there arise in practice non-standard distributions for which standard simulation strategies are inapplicable. In this context it would be useful if "black-box" generators were available so that variate generation could be accomplished without extensive effort. Simulation using the envelope methods described in this paper is an attempt in this general direction.

The construction of envelopes for general random variate generation has historical precedents in the work of Marsaglia and Tsang (1984), Devroye (1986) and Zaman (1991). In Gilks and Wild (1992), a useful generator is formed by adaptively constructing upper and lower linear splines to the logarithm of log-concave functions and then applying the rejection algorithm. Hoermann (1995) introduces the idea of T-concavity which extends variate generation to densities f where $T(f)$ is concave.

In Evans and Swartz (1998a, 1998b), generalizations of Gilks and Wild (1992) and Hoermann (1995) are considered for the purpose of variate generation. For example, the notion of T-concavity is extended, T-convexity is introduced, the assumptions of unimodality and boundedness are dropped, special considerations are given to the tails of functions and higher order envelopes are constructed. Evans and Swartz (1998c) also use envelopes to approximate integrals and provide exact error bounds.

In section 2, we describe the general framework for constructing lower and upper envelopes to general functions. In section 3, we consider the application of the envelopes to variate generation based on the adaptive rejection algorithm. Various new examples based on non-standard distributions are provided to highlight the utility of the methods.

2 ENVELOPE CONSTRUCTION

Consider the construction of a lower envelope $l(x)$ and an upper envelope $u(x)$ to a function $f(x)$ defined on $\mathcal{R}$.

We begin with the specification of a transformation $T : [0, \infty) \to \mathcal{R}$ and an integer $n \geq 0$. The specification of T and n provides the ingredients for envelope construction. There are many possible transformations T that may be considered such as the logarithm transformation (Gilks and Wild, 1992) and the class of power transformations $T(f) = f^p$. This is a very general framework; we may choose a single T over $\mathcal{R}$ or we may define T piecewise. With respect to choosing the integer n, it is generally true that increasing the value of n improves the accuracy of the envelopes at the expense of simplicity. We expand on the choice of T and n as we continue.

Having defined T and n, we obtain the inflections points $x_1 < \cdots < x_m$ of $(T(f))^{(n)}$. The inflection points are often easily obtained via a symbolic package such as Maple. As demonstrated in Evans and Swartz (1998b), typically the calculation of inflection points can be entirely avoided. It follows that the function $(T(f))^{(n)}$ has constant concavity in an interval (x_l, x_r) where $r = l+1$ and $l = 1, \ldots, m-1$. If $(T(f))^{(n)}$ is concave on (x_l, x_r) then the chord

$$(T(f))^{(n)}(x_l) + \frac{(T(f))^{(n)}(x_r) - (T(f))^{(n)}(x_l)}{x_r - x_l}(x - x_l) \quad (1)$$

bounds $(T(f))^{(n)}$ from below; moreover, an upper bound on $(T(f))^{(n)}$ is the tangent

$$(T(f))^{(n)}(x_l) + (T(f))^{(n+1)}(x_l)(x - x_l). \qquad (2)$$

If $(T(f))^{(n)}$ is convex on (x_l, x_r), then the chord and tangent expressions are simply reversed.

It is the specification of T and n, and the calculation of $x_1, \ldots, x_m$ that we view as the minimal information needed for envelope construction. Because of the minimal information, we refer to the resultant algorithms as "semi-automatic" or "nearly black-box" procedures. This development allows for the construction of envelopes to $f(x)$ on intervals (x_l, x_r) as given by the following proposition.

Proposition 1: Consider the interval (x_l, x_r). Suppose that T is an increasing invertible function with $n+3$ derivatives and that $(T(f))^{(n)}$ is concave. Then for every $x \in (x_l, x_r)$, we have that $l(x) \le f(x) \le u(x)$ where

$$\begin{aligned} l(x) &= T^{-1}\left[\sum_{k=0}^{n} \frac{(T(f))^{(k)}(x_l)}{k!}(x - x_l)^k \right. \\ &+ \left. \frac{(T(f))^{(n)}(x_r) - (T(f))^{(n)}(x_l)}{x_r - x_l} \frac{(x - x_l)^{n+1}}{(n+1)!}\right] \end{aligned}$$

and

$$u(x) = T^{-1}\left[\sum_{k=0}^{n+1} \frac{(T(f))^{(k)}(x_l)}{k!}(x - x_l)^k\right].$$

Proof: The proof generalizes Lemma 2 in Evans and Swartz (1998a) which proceeds by taking the anti-derivatives of (1) and (2) n times and then inverting via T^{-1}.

If $(T(f))^{(n)}$ is convex on (x_l, x_r) then the expressions for $l(x)$ and $u(x)$ in Proposition 1 are reversed. The expressions are again reversed if T is a decreasing function.

Now there are various way that the approximating envelopes $l(x)$ and $u(x)$ can be improved. First, the envelopes can generally be made tighter by choosing $n > 0$ which results in higher order envelopes. When $n = 0$, the approximations to $T(f)$ are linear. Second, it is a simple matter to improve the envelopes on any interval (x_l, x_r) by compounding. By introducing a new point x^* where $x_l < x^* < x_r$, the concavity structure on (x_l, x^*) and (x^*, x_r) is the same as on (x_l, x_r). Therefore improved lower and upper envelopes can be defined on each of the subintervals.

The above development is all that is needed to construct envelopes for truncated functions $f(x)$. For suppose that $f(x)$ has a left truncation point t_1 and a right truncation point t_2. Then we simply set $x_0 = t_1$ and $x_{m+1} = t_2$ and note that the intervals (x_0, x_1) and (x_m, x_{m+1}) have constant concavity. Truncated functions are common in Bayesian statistics where f is a density function and x is a parameter subject to order restrictions.

In problems where we have tails, special consideration is given to the construction of the upper envelope $u(x)$. As a first priority, we should choose x_0 and x_{m+1} somewhat extreme so as to limit the impact of tail calculations for variate generation. Without loss of generality, suppose that we are interested in constructing an upper envelope for the right tail. We should choose a smooth and invertible function T such that $T^{-1}(\alpha + \beta x)$ is integrable on (x_{m+1}, ∞). Now suppose that f is T-concave on (x_{m+1}, ∞) when T is increasing or T-convex on (x_{m+1}, ∞) when T is decreasing. Then for $x \in (x_{m+1}, \infty)$ we have that

$$T(f(x)) \le T(f(x_{m+1})) + T'(f(x_{m+1}))f'(x_{m+1})(x - x_{m+1})$$

when f is T-concave and

$$T(f(x)) \ge T(f(x_{m+1})) + T'(f(x_{m+1}))f'(x_{m+1})(x - x_{m+1})$$

when f is T-convex. Then taking the T-inverse of both of these inequalities we obtain

$$\begin{aligned} u(x) &= T^{-1}\{T(f(x_{m+1})) \\ &+ T'(f(x_{m+1}))f'(x_{m+1})(x - x_{m+1})\} \end{aligned}$$

which serves as an upper bound for $f(x)$ when $x \in (x_{m+1}, \infty)$.

Example 1: The Exponential Distribution

With such a tractable distribution as the exponential, there is certainly no need to develop new algorithms to generate exponential variates. However, our methods are easily and effectively implemented, and more importantly, the results provide building blocks for more complex problems.

Therefore consider the function $f(x) = e^{-x}$ and choose T equal to the identity map. Note that $\left(e^{-x}\right)^{(n)} = (-1)^{(n)} e^{-x}$ and so the $n-th$ derivative is concave when n is odd and convex when n is even. Therefore, on the interval (a, b), using the $n-th$ derivative, we obtain the upper envelope for e^{-x} which is given by

$$u(x) = \begin{cases} e^{-a}\sum_{k=0}^{n+1} \frac{(-1)^k}{k!}(x-a)^k & n \ odd \\ e^{-a}\sum_{k=0}^{n} \frac{(-1)^k}{k!}(x-a)^k + \frac{(-1)^n}{(n+1)!}\frac{e^{-b}-e^{-a}}{b-a}(x-a)^{n+1} & n \ even \end{cases}$$

and the lower envelope for e^{-x} is obtained by reversing the expressions for n odd and n even.

3 VARIATE GENERATION

The rejection algorithm is a general method for variate generation from a density that is proportional to f. Suppose then that the function $f(x)$ is bounded above by an upper envelope $Mg(x)$ where $M > 0$ is a known constant and $g(x)$ is a standard density function from which we can sample. Then generate $X \sim g$ and generate $w \sim \text{Uniform}(0, 1)$. If $w < f(x)/Mg(x)$, then retain X and note that X arises from a distribution with density proportional to f. It is easy to show that X is accepted with probability $\int f(x)\,dx/M$. Therefore, a successful implementation of rejection sampling involves obtaining g and M such that M is not too large and g is a density from which we can efficiently generate variates.

Using the piecewise envelopes $u(x)$ and $l(x)$ from section 2, implementation of rejection sampling is straightforward. We simply set $Mg(x) = u(x)$ where the constant M is determined by integrating $u(x)$ over the support of $f(x)$. The density $g(x)$ is viewed as a mixture with mixture intervals determined by $x_0, \ldots, x_{m+1}$.

To sample from $g(x)$, we must first sample a mixture component. This is done using the aliasing method (see Devroye (1986)) and only requires the generation of a single uniform variate. Once the mixture component has been determined, our second step in the rejection algorithm involves the generation of a variate from the particular mixture. Suppose then that we are sampling from the mixture on (x_l, x_r). A similar development can be given for the tails. The distribution function for this mixture is given by

$$U(x) = \frac{\int_{x_l}^{x} u(z)\,dz}{\int_{x_l}^{x_r} u(z)\,dz}. \tag{3}$$

Now, we need to be able to generate variates from $U(x)$ in (3). A direct method is to generate $p \sim U(0, 1)$ and then solve $U(X) = p$ for X. Note that because U is strictly increasing on (x_l, x_r), there is a unique X and this is the unique root of $U(x) - p$ lying in the interval. Note that if T is the identity map, then U is a polynomial whose coefficients are easily calculated. In this case, the root can be obtained using polynomial root-finding algorithms although we have found that the secant method is usually more efficient. For certain other choices (eg. $T = \ln$ and $n = 0$), $U(x)$ reduces to convenient forms that are also easily invertible.

The lower envelope $l(x) \geq 0$ which bounds f from below can also be used in the rejection algorithm. If f is computationally expensive to evaluate, then the squeezing condition $w < l(x)/Mg(x)$ is checked prior to checking $w < f(x)/Mg(x)$. If the squeezing condition holds, there is no need to check the expensive condition $w < f(x)/Mg(x)$ as we know that it also holds.

As described above, the final step in the rejection algorithm involves checking whether we retain the variate X. This step leads to an adaptive algorithm. For if we reject the variate X, then we replace the interval (x_l, x_r) with the 2 subintervals (x_l, X) and (X, x_r) and update the envelopes. The expensive part of adaptive rejection sampling is the setup time for aliasing. Accordingly, it makes sense to stop adapting when the upper and lower envelopes provide accurate approximations to f. As discussed in Evans and Swartz (1998a), a good rule is to stop adapting in the interval (x_l, x_r) when the ratio

$$\frac{\int_{x_l}^{x_r} l(z)\,dz}{\int_{x_l}^{x_r} u(z)\,dz}$$

is sufficiently close to 1.

The above algorithm leaves open the question of what is a suitable choice of the order of the approximation n? A natural criterion to assess this is efficiency of computation; namely which choice of n leads to the fewest rejection steps or, perhaps more importantly, the fastest computation time. However, as is shown in some of the examples, the choice $n = 0$ frequently leads to a perfectly satisfactory algorithm. The real virtue of the higher order polynomial envelopes is that such envelopes can often be computed very easily. For example, suppose that f can be factored as $f(x) = g(x)h(x)$ where $g \geq 0, h \geq 0$ and we have linear (i.e. $n = 0$) envelopes $l_g \leq g \leq u_g$ and $l_h \leq h \leq u_h$ for these functions. We then have quadratic envelopes $l_g l_h \leq f \leq u_g u_h$ for f. Application of techniques similar to this can often allow us to entirely avoid the computation of derivatives of f and also the need to calculate inflection points of such derivatives.

Example 2: Generating from Truncated Exponentials

As mentioned in Example 1, there is certainly no need to develop new algorithms for generating from the exponential distribution. Therefore this example serves as somewhat of a worse case scenario for the envelope methodology described in this paper. We consider the generation of 10^6 variates from the standard exponential distribution truncated on the interval $(1.0, 5.0)$. We first use the IMSL procedure DRNUN to generate a uniform variate u and then obtain the required variate $v = -\log[e^{-1} - (e^{-1} - e^{-5})u]$ via inversion. This requires 10 seconds of computation. Note that this is one of the few practical examples where inversion leads to an analytic formula. We compare this to the envelope methodology where a degree $n = 0$ expansion is used in the expression given in Example 1. The generation of 10^6 variates requires 70 seconds of computation.

Example 3: Generating from Truncated Student Distributions

A naive approach to this problem involves generating a variate x from the full Student distribution and retaining the variate if it lies within the required bounds. Evans and Swartz (1998a) develop a new and simple algorithm for generating from the Student family and truncations of Students. The approach is based on the recognition that the Student(α) distribution is T-convex everywhere when $T(f) = f^{-1/(\lambda+1)}$.

As an example, we consider generating from a Student(.5) distribution truncated to (-1,2). Generating 10^6 values using our methods requires 38 seconds of CPU time. For comparison purposes, generating 10^6 truncated variates using the naive approach with IMSL routines requires 76 seconds of CPU time. The advantage of our algorithm is even more dramatic with shorter truncation intervals.

We mention that these sorts of truncated distributions are common in Bayesian analyses with order restrictions. In another context, Evans and Swartz (1996) generate from truncated F distributions to implement stratified multivariate Student importance sampling.

Example 4: Generating from Truncated Polynomial Distributions

Chan (2000) considered the estimation of pupping probabilities of Grey Seals captured over a period of years. Based on independent binomial models, latent data and certain order constraints on the pupping probabilities, the analysis requires simulation from distributions whose densities are proportional to truncated polynomials.

For example, suppose that we need to generate the variate p_x (e.g. the probability that a female Grey Seal gives birth at age x) where the density of p_x is proportional to a non-negative q-th degree polynomial $f(p_x)$ truncated between a and b. The simple but inefficient approach in Chan (2000) begins with a root-finding algorithm to obtain the value $\hat{p}$ which maximizes f on (a, b). The rejection algorithm is then used where u is generated according to the Uniform$(0, 1)$ distribution, p_x is generated according to the Uniform(a, b) distribution and p_x is retained if $u \leq f(p_x)/f(\hat{p})$.

Envelope methods can be used in this application by choosing T equal to the identity map and choosing $n = 0$. The critical points are easily obtained using a root-finding algorithm for polynomials. The adaptive aspect of the algorithm quickly provides good envelopes to the function f. Alternatively, choosing T equal to the identity map and choosing $n = q - 1$ gives essentially an inversion method as the upper envelope $u(p_x)$ is equal to $f(p_x)$.

Example 5: Rational Normal Generators

We consider densities (possibly truncated) that are products of normal densities and positive rational functions (i.e. the quotient of 2 polynomials). We note that this is a huge family of distributions accomodating a wide range of shapes. Such distributions may have Bayesian applications in prior elicitation and importance sampling.

Evans and Swartz (1998b) develop algorithms for variate generation from these distributions as well as rational-beta distributions. One approach is based on a direct implementation of Proposition 1 where T is chosen as the identity map and derivatives are obtained via the Faà di Bruno formula. As an example, consider the product of the Normal$(0, 1)$ density with the rational function $(x^2 + 4x + 4.01)(x^2 - 4x + 4.01)/(x^2 + 1)$. Using a generator with order $n = 2$, the acceptance rate is a highly respectable .74.

Example 6: Gibbs Sampling from CIHM

Albert and Chib (1997) consider the Bayesian analysis of conditionally independent hierarchical models (CIHM). These are a broad and useful class of models which may be analysed using Markov chain Monte Carlo methods. In one such example, $y_i \sim$ Poisson(η_i), $\theta_i = \ln(\eta_i) \sim$ Normal$(x_i^T \beta, \tau^2)$ and (β, τ^2) are independent with $\beta \sim$ Normal(β_0, B_0^{-1}) and $\tau^2 \sim$ Inverse Gamma(a, b). For this problem, the conditional distribution of θ_i is proportional to

$$\exp\{y_i \theta_i - e^{\theta_i}\} \phi \left(\frac{\theta_i - x_i^T \beta}{\tau} \right) \tag{4}$$

where ϕ is the density function of the standard normal distribution.

Albert and Chib (1997) recommend the Gibbs sampling algorithm with an imbedded Metropolis step to sample from the non-standard conditional distributions for θ_i, $i = 1, \ldots, n$. Rather than using the Metropolis step, one can sample from (4) directly using the envelope methods in this paper. For example, perhaps the simplest generator involves choosing $T = \ln$. In this case, the conditional density is T-concave and choosing $n = 0$ gives a slight variation of the Gilks and Wild (1992) algorithm. Note that β and τ in (4) are changing in each iteration of Gibbs sampling; therefore it does not make sense to spend too much time finding efficient generators for specific values of β and τ.

Example 7: Generation in FOA Models

Swartz (2000) considered empirical Bayes models with latent variables to analyze bidding behaviour in final offer arbitration (FOA). One step in the posterior analysis requires simulation from non-standard distributions with densities proportional to

$$f(y) = \frac{1}{y^2} \exp\{a + b/y + c/y^2\}$$

where $d < y < e$. Swartz (2000) implements the simulation using the rejection algorithm based on the rejection density $g(y) = de(e - d)^{-1}/y^2$ defined on $d < y < e$. This also requires the simple maximization of $\exp\{a + b/y + c/y^2\}$ on $d < y < e$. The approach can be extremely inefficient for certain values of a, b, c, d and e.

Alternatively, an envelope approach can be used in this application by first obtaining envelopes $l(y)$ and $u(y)$ to $\exp\{a + b/y + c/y^2\}$ based on the exponential expansion discussed in Example 1. Simple envelopes to $f(y)$ are then given by $l(y)/y^2$ and $u(y)/y^2$. Note that such envelopes provide tractable expressions for (3).

ACKNOWLEDGMENTS

Both authors were partially supported by grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Albert, J.H., and S. Chib. 1997. Bayesian tests and model diagnostics in conditionally independent hierarchical models. *J. Amer. Statist. Assoc.* 92: 916-925.

Chan, E. 2000. Pupping probabilities II - the Bayesian approach. Unpublished MSc project, Department of Mathematics and Statistics, Simon Fraser University, Burnaby BC, Canada.

Devroye, L. 1986. *Non-uniform random variable generation*. Springer-Verlag.

Evans, M., and T.B. Swartz. 1996. Bayesian integration using multivariate Student importance sampling. *Computing Science and Statistics* 27: 456-461.

Evans, M., and T.B. Swartz. 1998a. Random variable generation using concavity properties of transformed densities. *J. Comp. Graph. Stat.* 7: 514-528.

Evans, M., and T.B. Swartz. 1998b. Higher order envelope random variate generators. Manuscript.

Evans, M., and T.B. Swartz. 1998c. An algorithm for the approximation of integrals with exact error bounds. Manuscript.

Fishman, G.S. 1996. *Monte Carlo: concepts, algorithms and applications*. Springer-Verlag.

Gilks, W.R., and P. Wild. 1992. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 41: 337-348.

Hoermann, W. 1995. A rejection technique for sampling from T-concave distributions. *ACM Trans. Math. Softw.* 21 (2): 182-193.

Marsaglia, G., and W.W. Tsang. 1984. A fast easily implemented method for sampling from decreasing or symmetric unimodal density functions. *SIAM J. Sci. Stat. Comp.* 5 (2): 349-359.

Swartz, T.B. 2000. Empirical Bayes modelling and computations in final offer arbitration. Manuscript.

Zaman, A. 1991. Generating random numbers from a unimodal density by cutting corners. Unpublished manuscript.

AUTHOR BIOGRAPHIES

MICHAEL J. EVANS is a Professor in the Department of Statistics at the University of Toronto. He received a B.Sc. degree in mathematics from the University of Western Ontario; and he received M.Sc. and Ph.D. degrees in statistics from the University of Toronto. He is a Fellow of the American Statistical Association and is an associate editor for the Journal of the American Statistical Association and the Canadian Journal of Statistics. His email and web addresses are `<mevans@utstat.utoronto.ca>` and `<www.utstat.utoronto.ca/mikevans/>`.

TIM B. SWARTZ is an Associate Professor in the Department of Mathematics and Statistics at Simon Fraser University. He received a B.Math. degree in statistics with computer science from the University of Waterloo; and he received M.Sc. and Ph.D. degrees in statistics from the University of Toronto. His email and web addresses are `<tim@stat.sfu.ca>` and `<www.math.sfu.ca/ tim/>`.