# Random Variable Generation Using Concavity Properties of Transformed Densities

by

Michael Evans
Department of Statistics
University of Toronto

and

Tim Swartz
Department of Mathematics and Statistics
Simon Fraser University

Technical Report No. 9606, July 5, (1996)

TECHNICAL REPORT SERIES

University of Toronto
Department of Statistics

# Random Variable Generation Using Concavity Properties of Transformed Densities

M. Evans* and T. Swartz*
U. of Toronto and Simon Fraser U.

**Abstract**

Algorithms are developed for constructing random variable generators for families of densities. The generators depend on the concavity structure of a transformation of the density The resulting algorithms are rejection algorithms and the methods of this paper are concerned with constructing good rejection algorithms for general densities.

*Keywords:* random variate generator, $T$-concavity and $T$-convexity, points of inflection, adaptive rejection sampling.

## 1 Introduction

Good algorithms for generating from univariate distributions are a necessary part of many applications where approximations to integrals or expectations are required. For a wide class of commonly used distributions there exist excellent algorithms and for many non-standard distributions there are classes of tools that can be applied to construct good algorithms; see for example, Devroye (1986). But this is not always the case. In many situations an algorithm can be constructed by sheer brute force inversion; i.e. tabulate the distribution function at many points, but this is inelegant and rarely results in a satisfactory solution. By this we mean that the time taken to generate many independent realizations can be considerable. Further, often we want an algorithm that can generate from a family of distributions and the specific distribution in the family cannot be prespecified; e.g. it may

*Partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

depend on data that varies from application to application, or the distribution may be changing dynamically as the simulation progresses, as in Gibbs sampling. In such contexts the brute force algorithm is not feasible.

In spite of the extensive amount of development in this area the authors have still encountered numerous situations where there is no obvious algorithm available beyond brute force inversion. We describe some of these in section 5. The purpose of this paper is to describe a general black-box algorithm that is capable of handling a wide variety of situations. In essence the user is required to input a minimal amount of information about densities belonging to a family of distributions and then the algorithm constructs an efficient generator. This is not a universal black-box as certain information is required to be available, or at least easily computed, and sometimes this is not the case. Basically, the information required is at most the first three derivatives, and the roots of the first and second derivatives, of some simple transformations of the density. This information is readily available for many univariate densities. A strong point of the algorithm described here is that excellent algorithms for specific distributions can be easily constructed and this does not demand deep insight into the properties of the distributions or great amounts of development time. The computer does all the work.

The algorithm we develop can be thought of as a generalization in several ways of the adaptive rejection algorithm developed in Gilks and Wild (1992). We will refer to this algorithm hereafter as the Gilks-Wild algorithm. The Gilks-Wild algorithm is a black-box algorithm for distributions with log-concave densities. For example, the $N(0,1)$, Gamma($\alpha$) for $\alpha > 1$ and Beta($a, b$) for $a, b \geq 1$ all have log-concave densities. Further log-concavity is maintained under location-scale transformations and truncation. On the other hand the Student($\lambda$) and the F($a, b$) densities are not log-concave. While there are simple good algorithms for generating from the full Student or F distributions this is not the case for truncations. The algorithm we describe here leads to new, good algorithms for the full distributions and also easily handles truncations. We note that in Gilks, Best and Tan (1993) a Markov chain algorithm is developed that combines the Gilks-Wild algorithm with the Metropolis algorithm to give an approximate generator for a general univariate density. In addition to only being approximate this algorithm also suffers from the existence of correlation between realizations. The algorithms developed here are exact and generate independent realizations.The Gilks-Wild algorithm is an adaptive rejection algorithm and this is particularly suitable in a number of applications of Gibbs sampling when the full conditional densities are log-concave. The algorithms developed here are adaptive and do not require the log-concave restriction on densities.

In section 2 we discuss $T$-concavity and specific examples of transformations $T$. In section 3 we indicate how these concepts are used to construct generators. In section 4 we consider the design of good generators and in section 5 we present examples. Conclusions are given in section 6.

## 2 $T$-Concavity

The Gilks-Wild algorithm is based on the log-concavity of a density $f$; i.e. the function $\ln \circ f$ is concave. In fact there is no reason to restrict just to the logarithm transformation and there is no reason to restrict to concavity. With appropriate restrictions on transformations $T : (0, \infty) \to R$, similar algorithms can be constructed.

Accordingly we say $f : D \to R$ is *$T$-concave* , where $D$ is a convex subset of $R$, if $T \circ f$ is concave. If $T \circ f$ is smooth then $f$ is $T$-concave if and only if $(T \circ f)'' = (T'' \circ f)(f')^2 + (T' \circ f)f'' \leq 0$. Further we say that $f$ is *$T$-convex* if $-(T \circ f\ )$ is concave. We call a convex subset $C$ of $D$ a *domain of $T$-concavity* of $f$ if $T \circ f$ is concave or convex there. We will restrict our discussion hereafter to functions $f$ with a finite partition $\{D_1, \ldots, D_m\}$ of $D$ by domains of $T$-concavity. For such an $f$ there is a coarsest such partition, which we call the *$T$-partition* of $f$, and note that this can be constructed by finding the inflection points of $T \circ f$. The $T$-partition together with the concavity of $T \circ f$ on each partition element will be referred to as the *$T$-concavity structure* of $f$. It is clear that $T$-concavity or $T$-convexity is preserved under location-scale transformations and truncations.

We recall some elementary facts about inflection points for smooth functions $g$ defined on an open interval. First $x$ is a point of inflection for $g$ if and only if $g''(x) = 0$ and $g'''(x) \neq 0$. Further, if $x$ is a point of inflection of $g$ and $g'''(x) > 0$ then $g$ changes from concave to convex as we proceed from left to right through $x$. Similarly $g$ changes from convex to concave if $g'''(x) < 0$. The inflection points of $T \circ f$ , perhaps for several $T$ transformations, is typically the information needed to construct a generator using the methods of this paper. For many distributions and transformations this information is readily available.

For the transformations $T$ considered here the $T$-concavity structure of a function does not change under positive multiples of $f$. A sufficient condition to ensure this is that $T'$ be homogeneous of degree $\mu \in R$. For $T'$ homogeneous of degree $\mu$ implies that $T''$ is homogeneous of degree $\mu - 1$ and then the above expression for $(T \circ f)''$ shows that the sign of this quantity does not change under positive multiples of $f$. In such a case our algorithm does not require that the density be normalized. This is practically significant, as often determining a norming constant can be a substantial computation. Further, it allows for great convenience in our development as we will ignore norming constants. As such, when referring to a density $f$, hereafter, we will only require that it be nonnegative and integrable.

As might be imagined an arbitrary $T$ does not suffice for the construction of good, or even feasible, generators. For convenience we list what seem to be necessary characteristics. The necessity of these will become apparent when we present the algorithm.

1. $T : (0, \infty) \to R$ is smooth, monotone and $T'$ is homogeneous of some degree,
2. $T$ and its derivatives and $T^{-1}$ are easy to compute,
3. the anti-derivative of $T^{-1}\ (\alpha + \beta x)$ is easy to compute for $x \in D$ and is integrable on $D$; i.e. $T^{-1}\ (\alpha + \beta x)$ is a density on $D$
4. it is easy to generate from the distribution with density $T^{-1}\ (\alpha + \beta x)$ via inversion.

We now present some examples of transformations that satisfy items 1-4.

## 2.1 Logarithm transformation

If we take $T = \ln$ then $T$ is smooth and increasing, $T'$ is homogeneous of degree -1, $T^{-1}(x) = \exp(x)$ and $\int T^{-1}\ (\alpha + \beta x)\, dx = \frac{1}{\beta} \exp(\alpha + \beta x)$. Therefore $T^{-1}\ (\alpha + \beta x)$ is a density on $(a, b)$ whenever $a, b \in R$, a density on $(-\infty, b)$ when $\beta > 0$ and is a density on $(a, \infty)$ when $\beta < 0$. Further the inverse cdf of any of these distributions is easily obtained using the log function so that generating using inversion is easy.

## 2.2 Power transformations

We define $T_p$ for $p \neq 0$ by $T_p(f) = f^p$. Then $T_p$ is smooth, increasing when $p > 0$, decreasing when $p < 0$ and $T'$ is homogeneous of degree $p - 1$. Further $T_p^{-1}(x) = x^{1/p}$ for $x > 0$ and, provided that $\alpha$ and $\beta$ are chosen so that $\alpha + \beta x \geq 0$ on $D$, then

$$\int T_p^{-1}(\alpha + \beta x)\, dx = \begin{cases} \frac{1}{\beta} \frac{p}{p+1} (\alpha + \beta x)^{\frac{p+1}{p}} & if \quad p \neq 0, -1 \\ \frac{1}{\beta} \ln(\alpha + \beta x) & if \quad p = -1 \end{cases}.$$

From this we see that, provided that $\alpha+\beta x \geq 0$ on the interval in question, $T_p^{-1}(\alpha+\beta x)$ is a density on $(a,b)$ for every $p \neq 0$ and is a density on $(-\infty,b)$ or $(a,\infty)$ whenever $p \in (-1,0)$ . In all of these cases the inverse cdf is easily obtained and so it easy to generate from these distributions via inversion. As we will see the requirement that $\alpha+\beta x \geq 0$ on an interval, is typically easy to satisfy as part of the algorithm.

We note that another family of transformations given by $T_p^*(f) = (f^p - 1)/p$ for $p \neq 0$ and $T_0^*(f) = \ln(f)$ includes the log and power transformations in a continuous family. There seems to be no apparent advantage to this family, however, and it ignores the fundamental difference in the restriction placed on $\alpha+\beta x$ between the log and power transformations. Also while $T_p \circ f$ and $T_p^* \circ f$ have the same concavity structure they lead to different generators.

# 3 The Algorithm

We restrict ourselves initially to the situation where $D = [a,b]$ is a bounded interval and suppose that we have chosen $T$, satisfying 1-4, so that $f$ is $T$-concave or $T$-convex on $D$ and $T \circ f$ is smooth. We note, however, that in general we can use different $T$ tranformations on different parts of the support of $f$. Further we suppose that we have chosen points $a \leq x_1 < \ldots < x_m \leq b$. In section 4 we will discuss how to choose these points.

Now let $t_i(x) = (T \circ f)(x_i) + (T \circ f)'(x_i)(x - x_i)$ be the equation of the tangent line to $T \circ f$ at $x_i$. Let $z_i \in (x_i, x_{i+1})$ be a point satisfying $t_i(z_i) = t_{i+1}(z_i)$ for $i = 1, \ldots, m-1$ and put $z_0 = a$, $z_m = b$. Note that the $T$-concavity or $T$-convexity ensures that $z_i$ exists and if $(T \circ f)'(x_i) \neq (T \circ f)'(x_{i+1})$ then it is unique. If $(T \circ f)'(x_i) = (T \circ f)'(x_{i+1})$ then $T \circ f = t_i$ on $(x_i, x_{i+1})$. In this case we will see that there is no benefit to having both $x_i$ and $x_{i+1}$ in the partition and so we can delete one. Henceforth we will assume that this has been done and then

$$z_i = \frac{[(T \circ f)'(x_i)x_i - (T \circ f)'(x_{i+1})x_{i+1}] - [(T \circ f)(x_i) - (T \circ f)(x_{i+1})]}{[(T \circ f)'(x_i) - (T \circ f)'(x_{i+1})]}.$$

Further let $c_i(x) = (T \circ f)(z_i) + [(T \circ f)(z_i) - (T \circ f)(z_{i-1})]\,(x - z_i)/(z_i - z_{i-1})$ be the equation of the secant from $(z_{i-1}, (T \circ f)(z_{i-1}))$ to $(z_i, (T \circ f)(z_i))$ for $i = 1, \ldots, m$.

Now define the *upper envelope function* by

$$u(x) = \begin{cases} t_i(x) \quad \text{if } z_{i-1} \le x \le z_i, & \begin{matrix} T \circ f \text{ concave}, \\ T \text{ increasing} \end{matrix} \quad \text{or} \quad \begin{matrix} T \circ f \text{ convex}, \\ T \text{ decreasing} \end{matrix} \\ c_i(x) \quad \text{if } z_{i-1} \le x \le z_i, & \begin{matrix} T \circ f \text{ concave}, \\ T \text{ decreasing} \end{matrix} \quad \text{or} \quad \begin{matrix} T \circ f \text{ convex}, \\ T \text{ increasing} \end{matrix} \end{cases}$$

and the *lower envelope function by*

$$l(x) = \begin{cases} c_i(x) \quad \text{if } z_{i-1} \le x \le z_i, & \begin{matrix} T \circ f \text{ concave}, \\ T \text{ increasing} \end{matrix} \quad \text{or} \quad \begin{matrix} T \circ f \text{ convex}, \\ T \text{ decreasing} \end{matrix} \\ t_i(x) \quad \text{if } z_{i-1} \le x \le z_i, & \begin{matrix} T \circ f \text{ concave}, \\ T \text{ decreasing} \end{matrix} \quad \text{or} \quad \begin{matrix} T \circ f \text{ convex}, \\ T \text{ increasing} \end{matrix} \end{cases}.$$

We then have that $T^{-1}(l(x)) \le f(x) \le T^{-1}(u(x))$ for every $x \in (a,b)$ and on $(z_{i-1}, z_i)$, $T^{-1}(u(x)) = T^{-1}(\alpha + \beta x)$ for some $\alpha, \beta$. Define the mixture density $g(x) = T^{-1}(u(x))/\int_a^b T^{-1}(u(z))\,dz = \sum_{i=1}^m p_i g_i(x)$ where $p_i = d_i/(d_1 + \cdots + d_m)$, $d_i = \int_{z_{i-1}}^{z_i} T^{-1}(u(x))\,dx$ and $g_i(x) = T^{-1}(u(x))/d_i$ on $[z_{i-1}, z_i]$ and is equal to 0 otherwise. We can generate from $g$ since it is easy to calculate the $p_i$ and easy to generate from component $g_i$ using inversion. We use the aliasing algorithm; see Devroye (1986), to generate from the discrete distribution $(p_1, \ldots, p_m)$. Thus generating from $g$ only requires the generation of 2 uniforms. Then the rejection sampling algorithm for $f$ proceeds by (i) generating $X \sim g$, (ii) generating $V \sim U(0,1)$, (iii) if $f(X) \ge VT^{-1}(u(X))$ then return $X$ else go to (i). In contexts where the computation of $f(X)$ is expensive we can add a squeezer step, between (ii) and (iii), by first testing $T^{-1}(l(X)) \le VT^{-1}(u(X))$ and returning $X$ if this holds, otherwise carrying out step (iii). In the adaptive version of this algorithm the point $X$ is added to $\{x_1, \ldots, x_m\}$ and a new $l$ and $u$ computed whenever $f(X) < VT^{-1}(u(X))$.

We recall here the requirement that $\alpha + \beta x \ge 0$ whenever $T$ is a power transformation. Consideration of the above shows that this restriction will automatically be satisfied piecewise by $u$ and $l$ whenever we require that $x \in \{x_1, \cdots, x_m\}$ if $(T \circ f)'(x) = 0$.

There are several assumptions associated with the above development. First we assumed that there exists a $T$ such that $f$ is $T$-concave or $T$-convex on $D$. This is clearly not necessarily the case but this problem is easily dealt with when $D = [a,b]$ by using the $T$-partition of $f$ and constructing $u$ amd $l$ piecewise on each element of the partition. More serious are the assumptions of bounded support and of no singularities at the end-points. As we will see in the examples, both of these problems can be dealt with in very general families of densities by making a judicious choice of a $T$ transformation for a tail interval or an interval with a singularity as an end-point. For example, it turns out that when a tail is not log-concave then there is often a power transformation $T_p$ such that the tail is $T_p$-convex and $p \in (-1, 0)$ so that $T_p$ is decreasing. In particular, infinite intervals are easily handled, in the sense that we can construct a rejection sampler $g$ on the whole interval, whenever $f$ is $T$-concave on the interval with $T$ increasing or whenever $f$ is $T$-convex on the interval with $T$ decreasing. In these cases $u$ is defined exactly as in the case of a bounded interval while $l$ must be modified so that the squeezer $T^{-1}(l(x))$ takes the value 0 on infinite intervals; e.g. in the log-concave case $l(x)$ takes the value $-\infty$ on such intervals.

## 4 Selecting the Points

Given a specific density $f$ it is natural to ask which transformation $T$ should be used, say from amongst those described in section 2. It turns out, however, that a single $T$ is sometimes not sufficient as we will require different transformations for the tails and the central region; e.g. see section 5. In the situation where a single transformation suffices then we would like to choose that $T$ and the points $\{x_1, \ldots, x_m\}$ which maximizes the probability of acceptance; namely

$$P(f(X) \geq VT^{-1}(u(X))) = \frac{\int f(x)\,dx}{\int T^{-1}(u(x))\,dx}.$$

This is not a tractable problem, however, even in very simple contexts. One thing we can say, based on the developments in section 3, is that choosing $T$ so that $T \circ f$ is approximately linear seems appropriate. Accordingly, we will suppose that $T$ has been chosen for a particular interval and consider the choice of the points $\{x_1, \ldots, x_m\}$ in this interval.

While optimal selection of the points may be a reasonable approach for distributions that are used very frequently, in general the following seems like an effective way to proceed as it demands minimal input and design of the algorithm. We start with some intial set $\{x_1, \ldots, x_m\}$ containing at least all the criticial and inflection points of $T \circ f$ and typically it pays to include more than these. For example, if the largest $x$ value is a critical point and the distribution has an infinite right tail then we must include one more point in the right tail else $g$ will not be integrable. A similar consideration arises if the smallest $x$ value is a critical point and the distribution has an infinite left tail. We then let the algorithm run adaptively; i.e. every time an $X$

function of $p \in (-1, -\frac{2}{b+2})$ and goes to $\infty$ as $p \to -\frac{2}{b+2}$ and goes to

$$c_l = \frac{b}{a}\frac{a-2}{2+b}\left(1 - \sqrt{\left(-\frac{2(a+b)}{(a-2)\,b}\right)}\right)$$

as $p \to -1$. Simple manipulations show that $c_l > 0$ for all $a \in (0,2)$, $b > 0$. So if $c > c_l$ we can find $p$ so that the largest root of $pk(x)$ equals $c$. This $p$ is given by

$$p(c) = -2\frac{(a^2b + 2a^2)\,c^2 + (-2a^2b + 4ba)\,c - b^2a + 2b^2}{((2a + ba)\,c - ba + 2b)^2}$$

and note that $p(c) \to -\frac{2}{b+2}$ as $c \to \infty$ and $p(c_l) = -1$. Thus $c = dc_l$ for any $d > 0$ is appropriate.

# 6 Conclusions

We have presented a general black-box algorithm for the construction of a random variable generator based on the concavity properties of simple transformations of densities. This generalizes the Gilks-Wild algorithm in several ways. Provided there is minimal information available about the density then the construction of excellent rejection algorithms is easy and automatic. A number of examples have demonstrated the utility of this approach in contexts where finding good generators has proven difficult. Further these examples demonstrate that it is possible to use these methods for families of distributions with little added complexity.

It is easy to see that random variable generation from many of the most common distributions encountered in practice can be handled by these techniques. For example, the Beta$(a, b)$ distribution is not log-concave whenever $a$ or $b$ is less than 1. But it is clear from our development of a generator for the F distribution that these cases can be handled in exactly the same way. Similar considerations apply to the Gamma$(a)$ distribution whenever $a < 1$.

The authors are currently developing general software to allow these methods to be applied to virtually any distribution for which the information detailed in the paper is available. Additional work can also be done on using other classes of transformations and on the selection of the initial points.

## References

Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer-Verlag. New York.

Evans M. and Swartz, T. (1994) Distribution theory and inference for polynomial-normal density functions. Commun. Statist. Theor. Meth. 23(4), 1123-1148.

Evans, M. and Swartz, T. (1995) Some techniques associated with multivariate Student importance sampling. To appear in the Proceedings of the 1995 Interface.

Gilks, W. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. Applied Statistics, 41, 337-348.

Gilks, W., Best, N.G., and Tan, K.K.C. (1993) Adaptive rejection Metropolis sampling for Gibbs sampling. Manuscript.

Kinderman, A.J. and Ramage, J.G. (1976). Computer generation of normal random variables. JASA, Vol. 71, No. 356, 893-896.

Scollnik, D. (1996) Simulating random variates from Makeham's distribution. To appear in the Transactions of the Society of Actuaries.

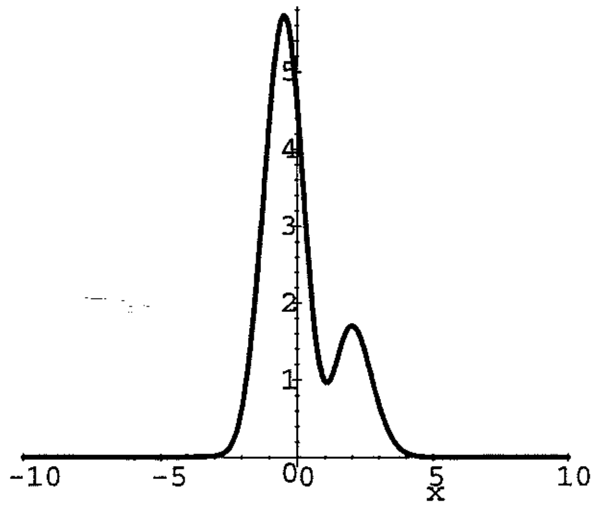


Figure 1. $f(x) = (x - z_1)(x - z_1^*)(x - z_2)(x - z_2^*)\phi(x)$ in Example 5.2.

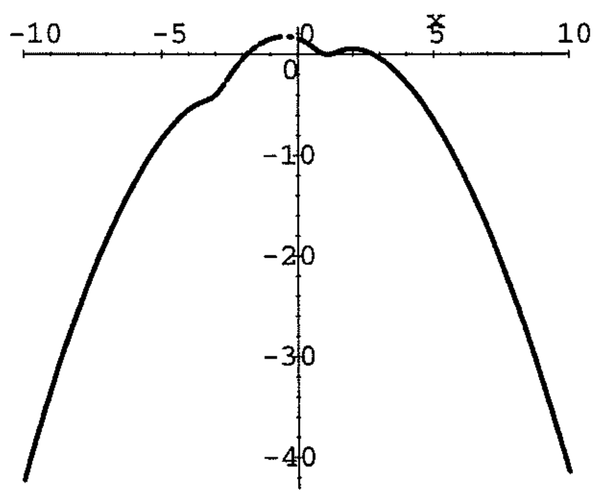


Figure 2. $\ln(f(x))$ in Example 5.2.