A Rejection Technique for Sampling from T-Concave Distributions

Hörmann, Wolfgang

A Rejection Technique for Sampling from T-Concave Distributions

Wolfgang Hörmann

Department of Applied Statistics and Data Processing
Wirtschaftsuniversität Wien

Preprint Series

Preprint 11
March 1994

http://statmath.wu-wien.ac.at/

Version: 1.3.1994: Accepted for publication in ACM TOMS

A Rejection Technique for sampling from T-Concave Distributions

Wolfgang Hörmann
Institut f. Statistik, WU-Wien, Augasse 2-6, A-1090 Wien, Austria
e-mail: whoer@statrix2.wu-wien.ac.at

Abstract: A rejection algorithm – called transformed density rejection – that uses a new method for constructing simple hat functions for an unimodal, bounded density f is introduced. It is based on the idea to transform f with a suitable transformation T such that $T(f(x))$ is concave. f is then called T-concave and tangents of $T(f(x))$ in the mode and in a point on the left and right side are used to construct a hat function with table-mountain shape. It is possible to give conditions for the optimal choice of these points of contact. With $T = -1/\sqrt{x}$ the method can be used to construct a universal algorithm that is applicable to a large class of unimodal distributions including the normal, beta, gamma and t-distribution.

AMS Subject Classification: 65C10, 68C25.
CR Categories and Subject Descriptors: G.3 [Probability and Statistics]: Random number generation
General Terms: Algorithms
Additional Key Words and Phrases: Rejection method, log-concave distributions, universal method

1. Introduction

In papers on random number generation the main emphasis is often laid on the speed of algorithms tailored for standard distributions. On the other hand some universal algorithms were proposed (see [6]) on which one can fall back when no standard algorithm is available. But these algorithms are very slow compared with algorithms specialized for only one distribution. Algorithms that are fast and can be used for a large class of continuous distributions (see e.g. [1], [17], [6] chapter VII) need a slow set-up step and large tables. So we aimed to design a universal method which is not too slow and needs only a short set-up. In this paper we introduce a general method, called transformed density rejection, that can be applied to a variety of unimodal continuous distributions with bounded densities which need not be log-concave. It is a generalization of a method for log-concave distributions that uses rejection from a distribution with uniform center and exponential tails and was proposed for several continuous and discrete standard distributions (see e. g. [18], [19], [14], and [6] chapter VII.2.6). In [7] a black box method for discrete log-concave distributions is based on that idea, in [9] an adaptation was suggested as an automatic method for continuous log-concave distributions, in [11] a method with uniform center and geometric tails was used to design a universal method for discrete log-concave distributions. Transformed density rejection can be used to design short algorithms for fixed distributions and – adding a conceptually simple set-up step — it results in a universal algorithm that is not much slower than most of the specialized algorithms and can be used for a variety of standard and non-standard distributions.
The paper is organized as follows. In Section 2 we give the theorems and the basic algorithm for the most general version of transformed density rejection. Section 3 contains the two most important special cases; their application to standard distributions is compared with the ratio of uniforms method. Section 4 discusses the possibility to design automatic or universal methods using transformed density rejection and gives the detailed description of the most useful algorithm of this kind. In Section 5 the computational experience with that algorithm for various distributions is compared with black-box methods and specialized algorithms given in literature.

2. Transformed density rejection

The idea of transformed density rejection is very simple: Transform the density function of the desired distribution with a suitable transformation $T(x)$ defined for $x \geq 0$. We define $h(x) = T(f(x))$ and a piecewise linear function $l(x)$ with $l(x) \geq h(x)$ for all x in the support of f (i.e. the closure of $\{x|f(x) > 0\}$). $T^{-1}(l(x))$ is then a dominating function for $f(x)$ and rejection can be used to sample from the desired distribution. In order that the choice of $l(x)$ can be automated in a simple way we restrict our attention to the case that $h(x) = T(f(x))$ is concave. (We call a function concave if its derivative is monotonically decreasing on its support as this definition admits single points where $f'(x)$ does not exist.) Among the many possibilities to choose $l(x)$ we take the simplest one and define $l(x)$ as the minimum of the three lines touching $h(x)$ in the mode m, in $x_l < m$ and in $x_r > m$ respectively. So we have

$$h(x) \leq l(x) = \min\left(h(x_l) + h'(x_l)(x - x_l),\ h(m),\ h(x_r) + h'(x_r)(x - x_r)\right)$$

In order that the method works the following four conditions are necessary:
a) $\lim_{x\to 0} T(x) = -\infty$;
b) T(x) is differentiable and $T'(x) > 0$, which implies that T^{-1} exists;
c) $\int_0^\infty T^{-1}(h(m) - x)\, dx < \infty$
d) $h(x) = T(f(x))$ is concave.
To make transformed density rejection applicable in practice we add the conditions: $F(x) = \int T^{-1}(x)\, dx$ is not too complicated and $F^{-1}(x)$ exists. Without loss of generality we assume $\lim_{x\to -\infty} F(x) = 0$.
If we want to use rejection it is necessary to compute the two intersection points b_l and b_r of the three parts of $l(x)$ and to compute the areas between x-axis and $T^{-1}(l(x))$ for the three intervals which are called v_l, v_c and v_r. Sampling from a density proportional to $T^{-1}(l(x))$ is done by inversion for the left and the right tail region, in the center $T^{-1}(l(x))$ is constant. The details are contained in the below algorithm.

Algorithm TDR:

1: (Set-up) Prepare a function $f(x)$ that returns values proportional to the density function of the distribution and a function $h'(x)$.
Set $m \leftarrow$ mode of the distribution,
$i_l \leftarrow \inf\{x|f(x) > 0\}$, $i_r \leftarrow \sup\{x|f(x) > 0\}$ (i_l and i_r need not be finite).
Choose x_l in the interval (i_l, m) and x_r in the interval (m, i_r).
Set $b_l \leftarrow x_l + (h(m) - h(x_l))/h'(x_l)$, $b_r \leftarrow x_r + (h(m) - h(x_r))/h'(x_r)$,
$v_l \leftarrow (F(h(m)) - F(h'(x_l)(i_l - x_l) + h(x_l)))/h'(x_l)$, $v_c \leftarrow f(m)(b_r - b_l)$,
$v_r \leftarrow (F(h'(x_r)(i_r - x_r) + h(x_r)) - F(h(m)))/h'(x_l)$.

2: Generate a uniform random number U and set $U \leftarrow U \cdot (v_l + v_c + v_r)$.

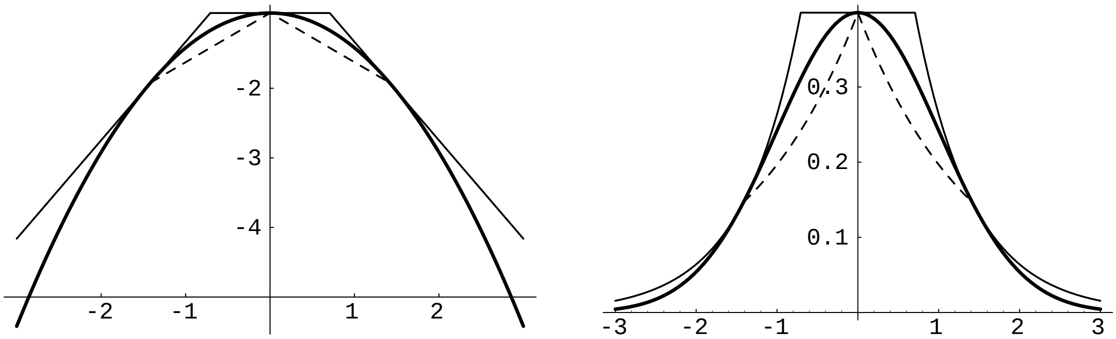
2.1: If $U \leq v_l$ set $X \leftarrow (F^{-1}(-Uh'(x_l) + F(h(m))) - h(x_l))/h'(x_l) + x_l$,
$l_x \leftarrow T^{-1}(h'(x_l)(X - x_l) + h(x_l))$.
Else if $U \leq v_l + v_c$ set $X \leftarrow ((U - v_l)/v_c)(b_r - b_l) + b_l$, $l_x \leftarrow f(m)$.
Else set $X \leftarrow F^{-1}((U - (v_l + v_c))h'(x_r) + F(h(m)))/h'(x_r) + x_r$,
$l_x \leftarrow T^{-1}(h'(x_r)(X - x_r) + h(x_r))$.

2.2: Generate a uniform random number V and set $V \leftarrow V \cdot l_x$.

2.3: If $V \leq f(X)$ return X, else go to 2.

As it was stated in the introduction similar methods with $T(x) = \log(x)$ were already suggested in literature. In this case the above conditions are obviously fulfilled for any log-concave density. Figure 1 shows in the left part $h(x)$ (thick line) and $l(x)$ (thin line) on the right hand side $f(x)$ (thick line) and $T^{-1}(l(x))$ (thin line), both for the normal distribution and $x_l = -\sqrt{2}$, $x_r = \sqrt{2}$.

Figure 1



For most of the distributions the evaluation of $f(x)$ is time consuming. Therefore it is worthwhile to use the two lines connecting the three points of contact x_l, m and x_r as simple squeezes in the interval (x_l, x_r) (shown as dashed lines in Figure 1). The validity of these squeezes follows from the fact that $h(x)$ is concave.

Algorithm TDRS: (The following two steps must be inserted in Algorithm TDR)

1.1: (inserted after step 1:)
Set $s_l \leftarrow (h(m) - h(x_l))/(m - x_l)$, $s_r \leftarrow (h(m) - h(x_r))/(m - x_r)$.

2.3.0: (inserted as the first part of step 2.3:)
If $(X < m)$
if $(X > x_l$ and $V \leq T^{-1}(h(m) - (m - x) * s_l))$ return X.
Else if $(X < x_r$ and $V \leq T^{-1}(h(m) - (m - x) * s_r))$ return X.

The question that is left open in Algorithm TDR is the choice of the points of contact x_l and x_r. For a fixed distribution with unbounded support it is quite simple to choose these points such that the area between the dominating curve and the x-axis is minimized. The below theorem contains everything necessary. (The special case $T(x) = \log(x)$ was already proved in [6] chapter VII.2.6 Theorem 2.6 .)

Theorem 1: Let $f(x) > 0, \forall x > m$ be a bounded monotone density with mode at m, or let $f(x) > 0, \forall x$ be a bounded unimodal density with mode at m. Let $T(x)$ be a transformation fulfilling the conditions a) to d) of above, $h(x)$ and $F(x)$ defined as above.
The area under the dominating curve of Algorithm TDR is minimized when x_r and x_l fulfill the condition:

$$f(x_r) = T^{-1}\left(h(m) - \frac{F(h(m))}{f(m)}\right) \qquad (*)$$

The area below the dominating curve, which is equal to the expected number of iterations, equals $f(m)(x_l - x_r)$ in the two-sided and $f(m)(x_r - m)$ in the monotone case.

Among the class of distributions which are T-concave (i.e. $T(f(x))$ is concave) for a fixed transformation T the area below the dominating curve of the optimal algorithm is bounded by $t_o = -F(T(1))/(F(-F(T(1)) + T(1)) - F(T(1)))$. (In [6] the bound $2t_o$ instead of t_o is given for the log-concave case).
Proof: Let $x_i = x_r + (h(m) - h(x_r))/h'(x_r)$ be the intersection between the center part and the right tail. Then the area below the dominating curve at the right hand side of m is

$$v_r = f(m)(x_i - m) + \int_{x_i}^{\infty} T^{-1}(h(x_r) + h'(x_r)(x - x_r))\, dx$$

which can be simplified to

$$v_r = f(m)(x_r - m) + (f(m)(h(m) - h(x_r)) - F(h(m)))/h'(x_r).$$

The derivative of this expression with respect to x_r is $\frac{-h''(x_r)}{(h'(x_r))^2}(f(m)(h(m) - h(x_r)) - F(h(m)))$; setting it equal to zero gives $(*)$. Due to the T-concavity we have $h''(x) \leq 0$ and thus it is easy to check that the derivative is nonpositive for x_r smaller than the solution of $(*)$ and nonnegative for x_r larger than that solution, which proves that we have a global maximum. As T-concavity implies continuity of the density and as the support of f is unbounded it is obvious that there is always a point that fulfills $(*)$. Substituting this point into v_r results in $f(m)(x_r - m)$ which is the area below the curve in the right sided case, in the two sided we add the left area and get $f(m)(x_r - x_l)$.
Now we proof that $f(m)(x_r - m) \leq t_o$ in the monotone case for all T-concave distributions. Without loss of generality we restrict ourselves to the class with $m = 0$ and $f(0) = 1$ as any density can be transformed into this class by relocating and rescaling it. Now we construct the T-concave function $g(x) = T^{-1}(kx + T(1))$ for $0 \leq x$. We choose k and t_o such that $\int_0^{t_o} g(x)\, dx = 1$ and $\int_0^{\infty} g(x)\, dx = t_o$. After integration we get $F(kt_o + T(1)) - F(T(1)) = k$ and $\frac{-F(T(1))}{k} = t_o$ which is solved by $k = F\,(-F(T(1) + T(1)) - F(T(1))$.
For an arbitrary f in our class T-concavity and the definition of g imply that $f(t_o) \leq g(t_o) = T^{-1}(k \cdot t_o + T(1)) = T^{-1}(-F(T(1)) + T(1)) = T^{-1}(h(0) - F(T(1)))$ which implies for the optimal choice: $x_r \leq t_o$; as $f(0) = 1$ we have $v_r \leq t_o$ which establishes the monotone case.
For the two-sided case let us denote the probability $P(X \leq m)$ with p. It is then obvious from the above that the area below the dominating curve left of m for the optimal x_l is always bounded by $t_o p$ the area right of m is bounded by $t_o(1 - p)$ which completes the proof. □

Theorem 1 gives the optimal choice of the points of contact for the case of unbounded support. In addition it implies that the expected number of iterations of Algorithm TDR is uniformly bounded over the class of all T-concave densities with arbitrary support (if there is no point fulfilling condition $(*)$ take the border of the support). This makes Algorithm TDR a good candidate for an automatic algorithm. But choosing x_l and x_r in a set-up step by solving condition $(*)$ can be very time consuming. So we give the following simple choice for x_l and x_r which guarantees the uniform boundedness as well. (In [6] p 304 the same choice for log-concave distributions is called minimax approach.)

Theorem 2: The choice of $x_r = m + t_o/f(m)$ with t_o as in Theorem 1 implies that the expected number of iterations in Algorithm TDR is lower or equal t_o for arbitrary monotone T-concave distributions.
With the choice $x_l = m - t_o/f(m)$, $x_r = m + t_o/f(m)$, the number of iterations of Algorithm TDR is lower or equal $2t_o$ for arbitrary T-concave distributions.
Proof: Following the proof of Theorem 1 we compute the area below the dominating curve for the monotone case. In the simplified expression of v_r it is easy to see that $v_r \leq f(m)(x_r - m)$ for the case that x_r is larger than the optimal x_r. The choice $x_r = m + t_o/f(m)$ is the rescaled and relocated version of t_o in the last part of the proof of Theorem 1. Following the arguments there it is obvious that the optimal x_r is always lower or equal $m + t_o/f(m)$ which completes the proof for the monotone case. For the two-sided case the bound is simply multiplied by two. □

Remark: For the case that x_r or x_l are not in the support of $f(x)$ leave away the left or right tail part (set $v_l \leftarrow 0$ or $v_r \leftarrow 0$ in Algorithm TDR).

Remark: It is easy to see that for the case that $p = \int_{-\infty}^{m} f(x)\,dx$ is known the choice $x_l = m - t_o p/f(m)$ and $x_r = m + t_o(1-p)/f(m)$ yields an expected number of iterations which is bounded by t_o for any T-concave distribution. This is especially useful for symmetric distributions.

For the case of bounded support the optimal choice of x_l and x_r is much more difficult and was entirely neglected in [6]. A necessary condition for optimality is given in Theorem 3.

Theorem 3: Let $f(x)$ be a T-concave monotone density on (0,a) with mode at 0. A necessary (but generally not sufficient) condition for $x = x_r$ to minimize the expected number of iterations of Algorithm TDR is:

$$f(m)(h(m)-h(x))-F(h(m))+F(h(x)+h'(x)(a-x))-h'(x)(a-x)T^{-1}(h(x)+h'(x)(a-x)) = 0$$

Proof: $area = f(m)(x_i - m) + \int_{x_i}^{a} T^{-1}(h(x_r) + h'(x_r)(x - x_r))\,dx$ is simplified and its derivative with respect to x_r set equal to 0. □

3. The choice of the Transformation

Among the possible transformations that fulfill the conditions a) to d) of the previous section we restrict our attention to the class T_c for $-1 < c \leq 0$. $T_0(x) = \log(x)$ was the transformation where transformed density rejection started from (see Section 1 and the references given there), $T_c(x) = -x^c$ for $-1 < c < 0$ is an important generalization. The two most important special cases for implementation on a computer are of course T_0 and $T_{-1/2} = -1/\sqrt{x}$ as T, T^{-1}, F and F^{-1} are so simple in these cases. Using transformation T_c the setup of Algorithm TDR constructs a dominating density with constant center and tails that behave like $x^{1/c}$ for $c < 0$ and like e^{-x} for $c = 0$. For the case that f is two times differentiable the condition for T_c-concavity is $f''(x) + (c-1)f'(x)^2/f(x) \leq 0 \quad \forall x$ in the support of f, which implies that any T_0-concave (i.e log-concave) density is T_c-concave or more generally that any T_{c_2}-concave density is T_{c_1}-concave for $c_1 < c_2$. Examples for standard distributions that are $T_{-1/2}$-concave but not log-concave include the t-family, the generalized inverse Gaussian distribution for $\lambda < 1$ and the Pearson VI distribution. Details are contained in Table 1 which gives for a variety of unimodal continuous distributions the transformation T_c which defines the smallest class of T_c-concavity the distribution falls into. Distributions which have a simple inverse cumulative distribution function (e.g. Weibull and Pareto distribution) are no problem for random variate generation and were therefore left out. (More information about the distributions contained in Table 3 is given e.g. in [12], for the last distribution see [13] or [3].)

Table 1

Name of distribution	density proportional to	parameters	T-concave for
Normal	$e^{-\frac{x^2}{2}}$		logarithm
Gamma	$x^{a-1}e^{-x}$	$a \geq 1$	logarithm
Beta	$x^{a-1}(1-x)^{b-1}$	$a, b \geq 1$	logarithm
Student's t	$\left(1+\frac{x^2}{a}\right)^{-\frac{a+1}{2}}$	$a > 0$	$c = \frac{-1}{1+a}$
Pearson VI (or beta-prime)	$\frac{x^{a-1}}{(1+x)^{a+b}}$	$a \geq 1$	$c = \frac{-1}{1+b}$
Perks	$\frac{1}{e^x+e^{-x}+a}$	$a > -2$	logarithm
Generalized inverse Gaussian	$x^{\lambda-1}e^{-(\omega/2)(x+1/x)}$	$\lambda \geq 1,\ \omega > 0$ $\lambda > 0,\ \omega \geq 0.5$	logarithm $c = -0.5$

For the class of transformations T_c we have defined above it is easy to verify that multiplying a density with a constant factor leads to the same factor in the dominating function constructed; everything else remains unchanged as long as the same points of contact are used. Therefore it is enough for Algorithm TDR to know the densities up to proportionality. To facilitate the use of Algorithm TDR Table 2 contains what we need to know about the three transformations.

Table 2

$T(x)$	$R^+ \to R:\ \log(x)$	$R^+ \to R^-:\ -x^c$	$-x^{-1/2}$
$T^{-1}(x)$	$R \to R^+:\ e^x$	$R^- \to R^+:\ (-x)^{1/c}$	x^{-2}
$F(x)$	$R \to R^+:\ e^x$	$R^- \to R^+:\ \frac{-(-x)^{1+1/c}}{1+1/c}$	$-1/x$
$F^{-1}(x)$	$R^+ \to R:\ \log(x)$	$R^+ \to R^-:\ (x(1+1/c))^{\frac{c}{1+c}}$	$-1/x$
$(*)$	$f(x_r) = f(m)/e$	$f(x_r) = f(m)\left(\frac{1}{c+1}\right)^{\frac{1}{c}}$	$f(x_r) = f(m)/4$
t_o	$\frac{e}{e-1} = 1.582$	$\frac{1}{1-\left(\frac{1}{1+c}\right)^{1+\frac{1}{c}}}$	2

Line $(*)$ of Table 2 corresponds to Theorem 1 and tells us how to place the points of contact to minimize the area below the dominating curve. For this choice t_o is the upper bound of the expected number of iterations necessary in Algorithm TDR for an arbitrary T-concave distribution. It is obvious that for a fixed distribution the area below the dominating curve with optimal points of contact is lowest for the transformation with the largest c possible. Therefore T_0 gives the best fit for log-concave distributions. As an example Table 3 gives for some standard distributions and the transformations T_0 and $T_{-1/2}$ the optimal points of contact and the expected number of iterations α.

Table 3

distribution		m	T	x_l	x_r	α	tl_{opt}	tr_{opt}
normal		0	log	$-\sqrt{2}$	$\sqrt{2}$	1.1284	0.5642	0.5642
			$T_{-1/2}$	$-\sqrt{\log(16)}$	$\sqrt{\log(16)}$	1.3286	0.6643	0.6643
gamma	$a = 2$	1	log	0.1586	3.1462	1.0881		
				0.3162	3.1462	1.0779	0.2516	0.7859
			$T_{-1/2}$	0.1018	3.6926	1.3066		
				0.3243	3.6926	1.2816	0.2486	0.9906
gamma	$a = 20$	19	log	13.483	25.848	1.1264		
				13.508	25.848	1.1264	0.5004	0.6240
			$T_{-1/2}$	12.635	27.210	1.3065		
				13.221	27.210	1.3010	0.5266	0.7481
beta	$a = 2$, $b = 3$	1/3	log	0.0619	0.7260	1.1392		
				0.1159	0.6760	1.1163	0.3866	0.6092
			$T_{-1/2}$	0.0402	0.7824	1.2324		
				0.1187	0.6717	1.1460	0.3815	0.6015
t	$a = 1$	0	$T_{-1/2}$	$-\sqrt{3}$	$\sqrt{3}$	1.1027	0.5513	0.5513
t	$a = 10$	0	$T_{-1/2}$	−1.6931	1.6931	1.3176	0.6588	0.6588
			$T_{-1/11}$	−1.4491	1.4491	1.1278	0.5639	0.5639

The results of Table 3 show the good fit of the dominating curve for the log-concave distributions when log is used as transformation but for $T_{-1/2}$ the results are not so bad as well. Thus Algorithm TDR with $T_{-1/2}$ can be faster than the logarithmic version even for log-concave distributions as T^{-1}, F and F^{-1} are more easy to compute for $T_{-1/2}$. But the main advantage of $T_{-1/2}$ is the fact that it is applicable for a wider range of distributions. For the case of bounded support we computed the suboptimal points according to $(*)$ (upper line) and the optimal points according to Theorem 3 (lower line). The difference between optimal and suboptimal solution is larger for $T_{-1/2}$ than for the logarithm. Interesting is the fact that in the case of bounded support the optimal point of contact is in most cases nearer to the mode for $T_{-1/2}$ than for the logarithm, in the unbounded case it is always the other way round.

As the ratio of uniforms method can be interpreted as table-mountain rejection (see for example [8]) and we use a dominating density with table-mountain shape as well it seems in place here to compare transformed density rejection and ratio of uniforms. The standard ratio of uniforms method first suggested in [15] must be compared with Algorithm TDR with $T_{-1/2}$ as both methods use table-mountains with tails like $1/x^2$. It is easy to see that the expected number of iterations α for TDR with optimal points of contact is for a fixed distribution lower than for the ratio of uniforms method as this method is restricted to table mountains where the area below the center part equals the area below the tails. For the normal distribution the difference is small (α=1.3286 compared with α=1.3688) but for the Cauchy distribution it is remarkable (1.1027 to 1.2732). Similar considerations can be made for generalizations of the ratio of uniforms method suggested in [20] (also discussed in [8] and "rediscovered" in [21]) when compared with TDR together with our family of transformations T_c: TDR can select the dominating density in a wider class of table-mountains and the optimal choice thus leads to a lower α.

4. A universal algorithm

Now we want to use the results from the previous sections to construct a universal or automatic algorithm that is applicable to all T-concave distributions with given mode and density. Due to the wider range of possible applications and the simplicity of the required functions we restrict ourselves to the case of $T_{-1/2}(x) = -1/\sqrt{x}$, of course the same could be done for the logarithm or T_c with arbitrary c in almost the same way. The main idea of our algorithm follows of course Algorithm TDR (for T, T^{-1}, F and F^{-1} we take the last column of Table 2). The only thing that is left open in Algorithm TDR is the choice of x_l and x_r. One possibility would be to use the result of Theorem 2, a second one to choose the points of contact by solving the condition $(*)$ of Table 2 numerically which of course results in a better fit of the dominating distribution but is not optimal for the case of bounded support. Depending on the application it can be more important to minimize the execution time for a fixed distribution with fixed parameters or to keep the setup as short as possible. As there are very fast table methods with relatively long setup available ([1], [17], [6] chapter VII) we will use the approach based on Theorem 2 to obtain an algorithm with moderate setup and good speed for fixed parameters. On the opposite extreme are the black box methods of [6] which need almost no setup but are quite slow.
The remark after Theorem 2 shows that for the case that the density has mass on both sides of the mode the choice $x_l = m - t_o/f(m)$, $x_r = m + t_o/f(m)$ ($t_o = 2$) is much too far away from the mode. Table 3 gives the optimal values t_o for x_l in the column tl_{opt}, those for x_r in the column tr_{opt}. So we suggest to start with a constant $t_o = t_1$ smaller than one and to compute the area below the dominating curve. If it is larger than 4 we take t_o=2 according to Theorem 2. The choice of t_1 can vary according to the distributions we are mainly interested in. We suggest the optimal value of the normal distribution $t_1 = 0.664$ which is good for all symmetric or nearly symmetric distributions and not so bad for asymmetric distributions as well. We tested it for the gamma, beta and t-distribution for many different parameters and the expected number of iterations α was always below 1.6, for most of the nearly symmetric distributions it was close to 1.32.
For the setup of Algorithm TDR it is necessary to compute the derivative of $h(x)$ in x_l and x_r. It can be inconvenient or slow to code the derivative of h but it is not necessary. Instead of the tangent of h in the point x_l we can take the line through the point $(x_l, h(x_l + \Delta))$ with the ascent $(h(x_l + \Delta) - h(x_l))/\Delta$. Due to the T-concavity of h it is always greater than or equal to h for arbitrary $\Delta > 0$. For x_r we can do the same with $-\Delta$. There are of course numerical difficulties if Δ is chosen too small. The details of a variant that guarantees that not more than 5 digits are lost are contained in the below algorithm step 1.3 and 1.4. It is based on the fact that $h'(x_l)$ is smaller than or equal to the ascent of the line connecting x_l and the mode (called al). (The exponent of the constant 10^{-5} gives the maximal number of (decimal) digits that can be lost due to cancellation. It should be changed if floating point numbers with less than 10

digits precision are used.)
One detail we have not explained yet refers to the case when x_l and/or x_r lie outside the support of the distribution and one or both tail-parts are therefore omitted. As the algorithm is slowed down considerably due to the missing squeeze we decided to define the missing point just for the squeeze with the distance between the point and m is 60 percent of the distance between m and the border of the support.
Now we are ready to give the details of a universal algorithm for $T_{-1/2}$-concave distributions with given mode and computable density $f(x)$.

Algorithm UTDR

1: [Set-up]

1.0: Set $m \leftarrow$ mode of the distribution, $f_m \leftarrow f(m)$, $h_m \leftarrow -1./\sqrt{f_m}$,
Set $i_l \leftarrow \inf\{x|f(x) > 0\}$, $i_r \leftarrow \sup\{x|f(x) > 0\}$,
if $(i_l = -\infty)$ set $t_l \leftarrow 0$ else set $t_l \leftarrow 1$,
if $(i_r = \infty)$ set $t_r \leftarrow 0$ else set $t_r \leftarrow 1$.
Set $c \leftarrow 0.664$.

1.1: $c \leftarrow c/f_m$, $x_l \leftarrow m - c$, $x_r \leftarrow m + c$.

1.2: If $(t_l = 1$ and $x_l < i_l)$
set $b_l \leftarrow i_l$, $v_l \leftarrow 0$,
if $(i_l < m)$ set $x_l \leftarrow m + (i_l - m) * 0.6$ and
$s_l \leftarrow (h_m + 1./\sqrt{f(x_l)})/(m - x_l)$.
else set $\tilde{y}_l \leftarrow -1/\sqrt{f(x_l)}$, $s_l \leftarrow (h_m - \tilde{y}_l)/(m - x_l)$,
$\Delta \leftarrow \max(|x_l|, -\tilde{y}_l/s_l) \cdot 10^{-5}$
$y_l \leftarrow -1/\sqrt{f(x_l + \Delta)}$, $a_l \leftarrow (y_l - \tilde{y}_l)/\Delta$.
Set $b_l \leftarrow x_l + (h_m - y_l)/a_l$, $d_l \leftarrow y_l - a_l * x_l$,
$v_l \leftarrow -1/(a_l * h_m)$, $c_l \leftarrow v_l$,
if $(t_l = 1)$ set $v_l \leftarrow v_l + 1/(a_l * (a_l * i_l + d_l))$.

1.3: If $(t_r = 1$ and $x_r > i_r)$
set $b_r \leftarrow i_r$,$v_r \leftarrow 0$,
if $(i_r > m)$ set $x_r \leftarrow m + (i_r - m) * 0.6$ and
$s_r \leftarrow (h_m + 1./\sqrt{f(x_r)})/(m - x_r)$.
else set $\tilde{y}_r \leftarrow -1/\sqrt{f(x_r)}$, $s_r \leftarrow (h_m - \tilde{y}_r)/(m - x_r)$,
$\Delta \leftarrow \max(|x_r|, \tilde{y}_r/s_r) \cdot 10^{-5}$
$y_r \leftarrow -1/\sqrt{f(x_r - \Delta)}$, $a_r \leftarrow (\tilde{y}_r - y_r)/\Delta$.
Set $b_r \leftarrow x_r + (h_m - y_r)/a_r$, $d_r \leftarrow y_r - a_r * x_r$,
$v_r \leftarrow 1/(a_r * h_m)$, $c_r \leftarrow v_r$,
if $(t_r = 1)$ set $v_r \leftarrow v_r - 1/(a_r(a_r * i_r + d_r))$.

1.4: Set $v_c \leftarrow (b_r - b_l) * f_m$, $v_{lc} \leftarrow v_l + v_c$, $v_t \leftarrow v_{lc} + v_r$.
If $(v_t \geq 4)$ set $c \leftarrow 2$ and go to step 1.1.

2: Generate a uniform random number U and set $U \leftarrow U * v_t$.

2.1: If $(U \leq v_l)$ set $X \leftarrow -d_l/a_l + 1/(a_l^2 * (U - c_l))$, $l_x \leftarrow (a_l * (U - c_l))^2$.
Else if $(U \leq v_{lc})$ set $X \leftarrow (U - v_l) * (b_r - b_l)/v_c + b_l$, $l_x \leftarrow f_m$.
Else set $X \leftarrow -d_r/a_r - 1/(a_r^2 * (U - v_{lc} - c_r))$, $l_x \leftarrow (a_r * (U - v_{lc} - c_r))^2$.

2.2: Generate a uniform random number V and set $V \leftarrow V * l_x$.

2.3: If $(X < m)$
    if $(X \geq x_l$ and $V * (h_m - (m - X) * s_l)^2 \leq 1)$ return X.
Else if $(X \leq x_r$ and $V * (h_m - (m - X) * s_r)^2 \leq 1)$ return X.
If $(V \leq f(X))$ return X,
else go to step 2.

5. Computational experience

We coded Algorithm UTDR in C and tested it for the normal, gamma, beta and t distribution on our DEC-station 5000/240. First we compared it with an implementation of Algorithm TDRS with T=log. Although the expected number of iterations is lower for the logarithm (see Table 3) UTDR was about ten percent faster for every distribution. Among the universal methods suggested in literature no one can be applied to such a large family of distributions if only the mode and the density are known. So we compared the execution times of UTDR with the black box algorithm for log-concave densities as explained in [6] p. 292 which is quite simple and short and needs almost no setup. But on the other hand the expected number of iterations α is four for that algorithm (compared with about 1.33 for UTDR) and it is about four times slower than UTDR for a fixed distribution. For the case that the distribution (or the parameters of the distribution) change after every call the algorithm of Devroye is slightly faster than UTDR due to the slow setup of UTDR which takes about six times longer than the generation of one random variate. If two samples of the same distribution are needed UTDR is already faster than Devroye's algorithm.
If we compare UTDR with specialized algorithms for the four distributions the fastest and most complicated generators coded in a high level language (see eg. [10] for the normal, [2] and [19], for the gamma, [18], and [22] for the beta and [16] for the t-distribution) are about three times faster for the normal distribution and two times faster for the gamma, beta and t-distribution. Short and simple methods for a single distribution (e.g. [4] for the gamma and [5] for the beta distribution) have about the same speed as Algorithm UTDR if we compare the fixed parameter case. For the case that the parameters vary after every call UTDR is not competitive in terms of speed. Contrary to Devroye's algorithm UTDR works well when the density is only known up to a constant factor that is not too far away from one (for example between 0.5 and 2). This fact is of importance when we need samples from truncated standard distributions. For example to generate a standard normal deviate truncated to the interval (-0.5,2) it is not necessary to change the code of Algorithm UTDR or the subprogram that delivers $f(x)$, only the borders of the support must be changed.

6. Conclusions

Transformed density rejection is a simple method that can be applied to a variety of continuous distributions. It uses a dominating density with the shape of a table mountain and is more flexible than ratio of uniforms and its generalizations. Thus it is easy to find optimal dominating distributions with a low expected number of iterations. Transformed density rejection is especially well suited to design universal algorithms for a very large class of bounded unimodal densities. The execution times for these universal algorithms are uniformly bounded over the whole class and comparable with algorithms designed for a specific distribution. Due to the important advantages of universal algorithms – one program of moderate length coded and debugged only once can do more than a collection of programs – we are convinced that the suggested Algorithm UTDR could replace the specialized algorithms for most applications thus gaining flexibility without loosing much speed.

References

[1] Ahrens, J.H. and Kohrt, K.D. Computer methods for sampling from largely arbitrary statistical distributions. Computing 26, (1981), 19-31.

[2] Ahrens, J.H. and Dieter U. Generating gamma variates by a modified rejection technique. Communications of the ACM 25, 1 (Jan. 1982), 47-54.

[3] Atkinson, A.C. The simulation of generalized inverse gaussian and hyperbolic random variables. SIAM Journal of Scientific and Statistical Computing, (1982), 502-515.

[4] Cheng, R.C.H. The generation of gamma variables with non-integral shape parameter. Applied Statistics 26, (1977), 71-75.

[5] Cheng, R.C.H. Generating beta variates with nonintegral shape parameters. Communications of the ACM 21, 4 (April 1978), 317-322.

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

[7] Devroye, L., A simple generator for discrete log-concave distributions. Computing 39, (1987), 87-91.

[8] Dieter, U. Mathematical aspects of various methods for sapling from classical distributions, in: Proceedings of the 1989 Winter Simulation Conference, eds. E.A. MacNair et al., (1989), 477-483.

[9] Gilks, W. R. and Wild, P. Adaptive rejection sampling for gibbs sampling. Applied Statistics 41, (1992), 337-348.

[10] Hörmann, W. and Derflinger, G. The ACR method for generating normal random variables. OR Spektrum 12, (1990), 181-185.

[11] Hörmann, W. A universal generator for discrete log-concave distributions. Computing 52, (1994), 89–96.

[12] Johnson, N. L. and Kotz, S. Continuous univariate distributions - 1. John Wiley, New York, (1970).

[13] Jorgensen, B. Statistical Properties of the Generalized Inverse Gaussian Distribution, Lectur Notes in Statistics 9, Springer, Berlin, (1982).

[14] Kachitvichyanukul, V. an Schmeiser, B. W. Computer generation of hypergeometric random variates. Journal of Statistical Computation and Simulation 22, (1985), 127-145.

[15] Kinderman, A. J. and Monahan, F. J. Computer generation of random variables using the ratio of uniform deviates. ACM Transactions on Mathematical Software 3, 3 (Sept. 1977), 257-260.

[16] A. J. Kinderman and J. F. Monahan New methods for generating student's t and gamma variables. Computing 25, (1980), 369-377.

[17] Marsaglia, G. and Tsang, W.W. A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions. SIAM Journal of Sientific and Statistical Computing 5, (1984), 349-359.

[18] Schmeiser, B.W. and Babu A.J.G. Beta variate generation via exponential majorizing functions. Operations Research 28, (1980), 917-926.

[19] Schmeiser, B.W. and Lal R. Squeeze methods for generating gamma variates. Journal of the American Statistical Association 75, (1980), 679-682.

[20] Stefanescu, S. and Vaduva, I. On computer generation of random vectors by transformations of uniformly distributed vectors. Computing 39, (1987), 141-153.

[21] Wakefield, J. C., Gelfand, A. E. and Smith, A. F. M. Efficient generation of random variates via the ratio-of-uniforms method. Statistics and Computing 1, (1991), 129-133.

[22] Zechner, H. and Stadlober, E. Generating beta variates via patchwork rejection. Computing 49, (1992).