Transformed Density Rejection with Inflection Points

Carsten Botts, Wolfgang Hörmann, Josef Leydold

Research Report Series
Report 110, July 2011

Institute for Statistics and Mathematics
http://statmath.wu.ac.at/

Transformed Density Rejection with Inflection Points

Carsten Botts · **Wolfgang Hörmann** · **Josef Leydold**

Abstract The acceptance-rejection algorithm is often used to sample from non-standard distributions. For this algorithm to be efficient, however, the user has to create a hat function that majorizes and closely matches the density of the distribution to be sampled from. There are many methods for automatically creating such hat functions, but these methods require that the user transforms the density so that she knows the exact location of the transformed density's inflection points. In this paper, we propose an acceptance-rejection algorithm which obviates this need and can thus be used to sample from a larger class of distributions.

Keywords Nonuniform random variate generation · transformed density rejection · inflection points

Mathematics Subject Classification (2000) 65C05 · 65C10

1 Introduction

Sampling random variates from non-standard distributions is a crucial part of Monte Carlo methods and stochastic simulation. Acceptance-rejection sampling is often used to generate values from non-standard distributions, but to execute this algorithm, the user has to find a multiple of some density function that majorizes f, the density of the distribution to be sampled from. This majorizing function is referred to as the hat function and is denoted as h. The user may also want to find a squeeze function, s. The squeeze function is majorized by f and is typically used to reduce the computational expense of the acceptance-rejection algorithm. Once values of h and s have been found, to generate a value of X from a distribution with density f, the following steps are necessary:

1. Generate a random variate X with density proportional to h.
2. Generate a $(0,1)$ uniform random number, U.
3. If $U\,h(X) \leq s(X)$, then return X.
4. If $U\,h(X) \leq f(X)$, then return X.
5. Otherwise, try again.

Although executing the five steps above is simple, the challenge in implementing acceptance-rejection sampling is in finding appropriate values of h and s.

Devroye (1984) proposed a method to construct hat functions when the distribution to be sampled from has a log-concave density. Gilks and Wild (1992) partition the domain of the distribution into non-overlapping intervals and use tangents and secants of the log-density to construct hat and squeeze functions, respectively. This subdivision can be refined with adaptive rejection sampling (ARS). In ARS, the hat and squeeze functions "adapt" to the density to be sampled from with every proposed value of X that is rejected. To be more specific, every time a proposed value of X is rejected, the interval in which the rejected value lies is split into two non-overlapping intervals at this point. Hat and squeeze functions are then calculated for the density within these two intervals. The area between hat and squeeze functions thus tends to 0, and the marginal generation time hardly depends on the target distribution.

C. Botts
Department of Mathematics and Statistics
Williams College
Williamstown, MA 01267, USA
E-mail: cbotts@williams.edu

W. Hörmann
Department of Industrial Engineering
Boğaziçi University
34342 Bebek-İstanbul, Turkey
E-mail: hormannw@boun.edu.tr

J. Leydold
Institute for Statistics and Mathematics
WU (Vienna University of Economics and Business)
Augasse 2–6, A-1090 Wien, Austria
E-mail: josef.leydold@wu.ac.at

Hörmann (1995) generalized this idea for the class of T-concave distributions. A density f is called T-concave if the transformed density $\tilde{f} = T \circ f$ is concave, where $T:\ (0,\infty) \to \mathbb{R}$ is a differentiable and monotonically increasing transformation. If f is T-concave, the tangent $\tilde{t}(x) = \alpha + \beta x$ to $\tilde{f}$ is greater than $\tilde{f}$ for all x in the domain of f, making the function $t(x) = T^{-1}[\tilde{t}(x)] = T^{-1}(\alpha + \beta x)$ a hat function to f. Similarly, if f is T-concave, the secant to the transformed density, $\tilde{r}$, can be used to construct the squeeze function, s, for the density in a given interval. Evans and Swartz (1998) show that the opposite applies (in that tangents are used to construct the squeeze function and secants are used to construct the hat function) when f is T-convex. Since hat and squeeze functions can be constructed for densities that are either T-concave or T-convex, hat and squeeze functions can be constructed for any density as long as the user knows exactly where the density is T-convex and T-concave. In such cases, the domain of $\tilde{f}$ should be split into intervals such that within each interval, $\tilde{f}$ is either entirely concave or entirely convex. Separate hat and squeeze functions will then be calculated within each interval, and the techniques used to calculate these hat and squeeze functions will depend, of course, on whether $\tilde{f}$ is concave or convex within the interval. Identifying these intervals, however, requires identifying the inflection points of $\tilde{f}$, and this may not be a trivial task.

Botts (2010) recently relaxed the requirement of knowing the exact position of these inflection points. He proposes a method where the domain of the distribution is subdivided into intervals where the transformed density is either concave, convex, or has exactly one inflection point. For the latter case, he introduces an additional transformation and compiles a new sampling algorithm.

In this paper we propose a sampling algorithm that works for all these intervals. Hat and squeeze functions are constructed by means of tangents and secants of the transformed density, making the resulting acceptance-rejection algorithm simple. The paper is organized as follows: In Section 2 we derive the proposed sampling method. Section 3 compiles the resulting algorithm, and in Section 4 we apply the method to some examples.

2 The Proposed Method

For the algorithm proposed in this paper, the user does not have to precisely identify the inflection points of the transformed density, $\tilde{f}$ (recall that this is required in transformed density rejection). In the proposed algorithm, the user just has to partition the domain of f into intervals in which the transformed density is either entirely concave, entirely convex, or has just one inflection point. Once these intervals have been supplied by the user, we provide a method for constructing the hat and squeeze functions. In Section 2.1 we give some conditions a density has to satisfy for the proposed algorithm to apply. In Section 2.2 details are provided on how the hat and squeeze functions are calculated for a given partition. In Section 2.3, more details are given on how the partition should be constructed, Section 2.5 discusses how the density should be transformed, and in Section 2.6 we discuss some computational issues that should be noted in the algorithm.

2.1 Conditions

The algorithm proposed in this paper can be used to sample from any distribution with a density that satisfies the following three conditions (note that these conditions are always satisfied when the given density is twice continuously differentiable and has only a finite number of inflection points).

Condition 1 Density f and thus $\tilde{f}$ are continuous. Notice that this is always the case for a concave or convex function in some open interval.

Condition 2 $\tilde{f}$ is continuously differentiable except in a finite number of points where all one-sided derivatives exist. In abuse of language we set $\tilde{f}'(x) = \infty$ if $\tilde{f}$ has a vertical tangent like in $\sqrt[3]{x}$ at $x = 0$.

Condition 3 $\tilde{f}$ is twice continuously differentiable except in a finite number of points. These points must be inflection points of $\tilde{f}$, that is, $\tilde{f}''$ must change sign at such points. This excludes transformed densities with cusp-like structure like in $|x| - x^2$ at $x = 0$. We also assume that there is only a finite number of points where $\tilde{f}'' = 0$.

2.2 Constructing the hat and squeeze functions

We begin by assuming that the user has partitioned the domain of f, into mutually exclusive intervals. Within each interval, $\tilde{f}$ is either entirely concave, entirely convex, or contains just one inflection point. In the cases where $\tilde{f}$ is entirely concave, tangents and secants to $\tilde{f}$ are used to construct the hat and squeeze functions, respectively, and when $\tilde{f}$ is convex, the opposite applies. In cases where $\tilde{f}$ contains one inflection point, then we rely on the result in Theorem 1 to construct a hat function.

Theorem 1 *Let $[b_l, b_r]$ be a closed interval where $\tilde{f}$ has (at most) one inflection point $y \in (b_l, b_r)$. Then (at least) one of the following cases holds. (Note that the tangents of $\tilde{f}$ at b_l and b_r are denoted by $\tilde{t}_l$ and $\tilde{t}_r$, respectively, and the secant between these two points by $\tilde{r}$.)*

(Ia) $\tilde f$ *is concave near* b_l *and* $\tilde t_r(x) \le \tilde f(x) \le \tilde t_l(x)$,
(Ib) $\tilde f$ *is convex near* b_l *and* $\tilde t_l(x) \le \tilde f(x) \le \tilde t_r(x)$,

(IIa) $\tilde f$ *is concave near* b_l *and* $\tilde r(x) \le \tilde f(x) \le \tilde t_l(x)$,
(IIb) $\tilde f$ *is convex near* b_l *and* $\tilde r(x) \le \tilde f(x) \le \tilde t_r(x)$,

(IIIa) $\tilde f$ *is concave near* b_l *and* $\tilde t_r(x) \le \tilde f(x) \le \tilde r(x)$,
(IIIb) $\tilde f$ *is convex near* b_l *and* $\tilde t_l(x) \le \tilde f(x) \le \tilde r(x)$,

(IVa) $\tilde f$ *is concave on* $[b_l, b_r]$ *and* $\tilde r(x) \le \tilde f(x) \le \tilde t_m(x)$,
(IVb) $\tilde f$ *is convex on* $[b_l, b_r]$ *and* $\tilde t_m(x) \le \tilde f(x) \le \tilde r(x)$,

where $\tilde t_m$ *is the tangent of* $\tilde f$ *in* b_l *or* b_r *wherever* $\tilde f$ *is larger.*

Cases (Ia), (IIa), and (IIIa) are illustrated in Figure 1. Also note that Case (IVa) is a special case of (IIa), and Case (IVb) is a special case of (IIIb).

Proof Recall that all the functions $\tilde t_l$, $\tilde t_r$, and $\tilde r$ are linear, and that $\tilde r(b_l) = \tilde t_l(b_l) = \tilde f(b_l)$ and $\tilde r(b_r) = \tilde t_r(b_r) = \tilde f(b_r)$. Assume first that there is an inflection point $y \in (b_l, b_r)$ such that $\tilde f$ is concave on $[b_l, y]$ and thus convex on $[y, b_r]$. Let $\tilde r_{b_l y}$ be the secant between $\tilde f(b_l)$ and $\tilde f(y)$. Then by concavity of $\tilde f$ we have $\tilde t_l(x) \ge \tilde f(x) \ge \tilde r_{b_l y}(x)$ for all $x \in [b_l, y]$. Analogously, we have $\tilde t_r(x) \le \tilde f(x) \le \tilde r_{y b_r}(x)$ for all $x \in [y, b_r]$.

If $\tilde t_l(b_r) \ge \tilde f(b_r)$ and $\tilde t_r(b_l) \le \tilde f(b_l)$, then $\tilde t_l(x) \ge \tilde t_r(x)$ and we find by continuity of $\tilde f$ that $\tilde t_l(y) \ge \tilde r_{b_l y}(y) = \tilde r_{y b_r}(y) = \tilde f(y)$ and $\tilde t_l(b_r) \ge \tilde r_{y b_r}(b_r) = \tilde f(b_r)$. Consequently, $\tilde t_l(x) \ge \tilde r_{y b_r}(x) \ge \tilde f(x)$ for all $x \in [y, b_r]$ and hence $\tilde t_l$ is an upper bound for $\tilde f$ on $[b_l, b_r]$. Analogously, $\tilde t_r$ is a lower bound for $\tilde f$. Thus we have case (Ia).

If $\tilde t_l(b_r) \ge \tilde f(b_r)$ and $\tilde t_r(b_l) \ge \tilde f(b_l)$, then again $\tilde t_l$ is an upper bound for $\tilde f$. However, $\tilde t_l$ and $\tilde t_r$ intersect. We now have $\tilde r(x) \le \tilde t_r(x) \le \tilde f(x)$ for all $x \in [y, b_r]$ (by construction of $\tilde r$ and $\tilde t_r$). In particular this inequality holds in y and hence it also implies that $\tilde r(x) \le \tilde r_{y b_r}(x) \le \tilde f(x)$ for all $x \in [b_l, y]$. Thus we have case (IIa).

If $\tilde t_l(b_r) \le \tilde f(b_r)$ and $\tilde t_r(b_l) \le \tilde f(b_l)$, then we analogously get case (IIIa).

For the remaining case where $\tilde t_l(b_r) < \tilde f(b_r)$ and $\tilde t_r(b_l) > \tilde f(b_l)$ we have $\tilde r(x) \le \tilde f(x) \le \tilde r(x)$ for all $x \in [b_l, b_r]$. Thus $\tilde f$ is a linear function and we find $\tilde t_l(x) = \tilde f(x) = \tilde t_r(x)$, a contradiction.

If $\tilde f$ is convex on $[b_l, y]$, then cases (Ib), (IIb), and (IIIb) follows completely analogously.

If there is no inflection point, then $\tilde f$ is concave or convex in $[b_l, b_r]$, i.e., (IVa) and (IVb), respectively. Then we may use any tangent with point of contact in $[b_l, b_r]$. The particular choice of tangents deserve some explanation. When $\tilde f$ is concave and strictly monotone on some interval we always use its maximum as construction point for the tangent as this ensures a valid hat function for f, see Sect. 2.3 below. □

With the result in Theorem 2, it is simple to diagnose which case in Theorem 1 holds.

Theorem 2 *Let* $[b_l, b_r]$ *be a closed interval where* $\tilde f$ *has at most one point* y *which is an inflection point or where* $\tilde f''(y) = 0$. *Let* $R = \frac{\tilde f(x_r) - \tilde f(x_l)}{x_r - x_l}$ *be the slope of the secant of* $\tilde f$. *Then the cases from Theorem 1 occur when the following conditions are satisfied:*

(Ia) if and only if $\tilde f'(b_l) \ge R$ *and* $\tilde f'(b_r) \ge R$.
(Ib) if and only if $\tilde f'(b_l) \le R$ *and* $\tilde f'(b_r) \le R$.

(IIa) when $\tilde f''(b_l) < 0$ *and* $\tilde f''(b_r) > 0$ *and* $\tilde f'(b_l) \ge R \ge \tilde f'(b_r)$.
(IIb) when $\tilde f''(b_l) > 0$ *and* $\tilde f''(b_r) < 0$ *and* $\tilde f'(b_l) \ge R \ge \tilde f'(b_r)$.

(IIIa) when $\tilde f''(b_l) < 0$ *and* $\tilde f''(b_r) > 0$ *and* $\tilde f'(b_l) \le R \le \tilde f'(b_r)$.
(IIIb) when $\tilde f''(b_l) > 0$ *and* $\tilde f''(b_r) < 0$ *and* $\tilde f'(b_l) \le R \le \tilde f'(b_r)$.

(IVa) if and only if $\tilde f''(b_l) \le 0$ *and* $\tilde f''(b_r) \le 0$.
(IVb) if and only if $\tilde f''(b_l) \ge 0$ *and* $\tilde f''(b_r) \ge 0$.

Proof Recall that $\tilde r(b_l) = \tilde t_l(b_l) = \tilde f(b_l)$ and $\tilde r(b_r) = \tilde t_r(b_r) = \tilde f(b_r)$. Moreover, notice that $\tilde t_l(b_r) \ge \tilde f(b_r)$ if and only if $\tilde f'(b_l) \ge R$. Similarly, $\tilde t_r(b_l) \le \tilde f(b_l)$ if and only if $\tilde f'(b_r) \ge R$. Thus using our considerations from the proof of Theorem 1 the statements for cases (Ia) and (Ib) immediately follow.

If we have $\tilde f''(b_l) < 0$ and $\tilde f''(b_r) > 0$, then $\tilde f$ is concave near b_l and there is exactly one inflection point $y \in (b_l, b_r)$. Moreover, $\tilde f'(b_l) \ge R \ge \tilde f'(b_r)$ implies $\tilde t_l(b_r) \ge \tilde f(b_r)$ and $\tilde t_l(b_r) \ge \tilde f(b_r)$ and hence we have case (IIa) by the arguments from the proof of Theorem 1. The statements for cases (IIb), (IIIa) and (IIIb) follow analogously.

At last (IVa) implies $\tilde f''(b_l) \le 0$ and $\tilde f''(b_r) \le 0$. On the other hand if these two inequalities hold then at least one most be strict because there is at most one point y with $\tilde f''(y) = 0$. Moreover, $\tilde f''(y) = 0$ cannot change sign and thus $\tilde f''(x) \ge 0$ for all $x \in [b_l, b_r]$, i.e., $\tilde f$ is concave and we have case (IVa). Similarly, the statement for case (IVb) follows. □

2.3 Partitioning the domain of f

In this subsection, we give more details on how the domain of f should be partitioned. The proposed algorithm initially requires that the user splits the domain of f into non-overlapping intervals and that if the domain of f is unbounded, then in the left and/or right-most intervals, $\tilde f = T \circ f$ should be concave and strictly monotone. We formally state this in the following condition.

Condition 4 Let $b_0 < b_1 < \ldots < b_{n-1} < b_n$ be the breaking points of a partition of the domain of density f. Then the following must hold:

- In each bounded interval $[b_i, b_{i+1}]$ of the partition, there is at most one inflection point of the transformed density $\tilde f$, or one point where $\tilde f''$ vanishes.

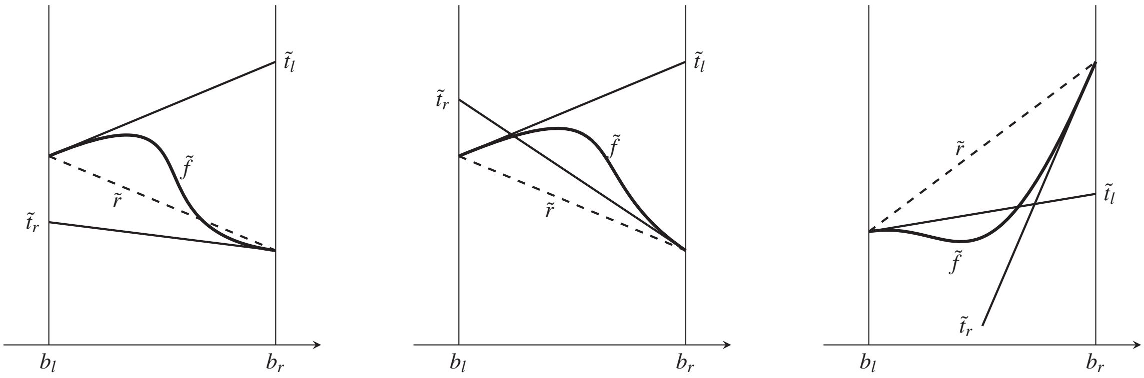

Fig. 1 Cases (Ia), (IIa), and (IIIa) from Theorem 1

– In each unbounded interval $(-\infty, b_1]$ or $[b_{n-1}, \infty)$, $\tilde{f}$ must be concave and strictly monotone.

Some remarks regarding this condition include

Remark 1 By Conditions 1–3 we always can find a partition that satisfies Condition 4. The monotonicity condition for concave $\tilde{f}$ on an unbounded interval can be easily checked by $\tilde{f}'(b_r) > 0$ and $\tilde{f}'(b_l) < 0$, respectively. These assumptions allow us to apply Theorem 2 for distinguishing between the cases from Theorem 1.

Remark 2 The second part of Condition 4 ($\tilde{f}$ must be concave on unbounded intervals) is not crucial for the development of the theory. For $\tilde{f}(x) = |x| - \log|x|$ we can construct a hat function on $[b, \infty)$ using the linear function $\tilde{h}(x) = \tilde{f}(b) + \beta(x - b)$ where $\beta = \lim_{t\to\infty} \tilde{f}'(t)$. However, for practical reasons we excluded this case.

Once the initial partition of f's domain has been entered, it will be used to partition the domain into even more intervals. The more intervals the domain of f is divided into, the closer the hat and squeeze functions become, and the closer

$$\rho = \frac{\text{area below hat}}{\text{area below squeeze}} \tag{1}$$

gets to 1. As ρ approaches 1, the expected number of iterations until a point is accepted also tends to 1, making the algorithm more efficient.

Gilks and Wild (1992) accomplish this for log-concave distributions using adaptive rejection sampling (ARS). In ARS, the hat function for a log-concave density is initially constructed by considering two points in the domain of f (each point on either side of a mode, if a mode exists), calculating the tangents to the log-density at these two points, and then "connecting" these tangents at their point of intersection. The point where these two tangents intersect divides the domain of f into two intervals. The domain of f is divided into more intervals when a candidate value of X is rejected. When this happens, the tangent to the log-density is calculated at this point, a new hat function is created by connecting this tangent line to the tangent lines adjacent to it, and the points where this tangent intersects the adjacent tangent lines are the boundaries of the new interval created. This procedure is repeated for every candidate value of X rejected.

ARS is a very powerful tool when only a few (or just one) random variate has to be drawn from a particular distribution. It does not clearly distinguish between the setup and generation part, however. ARS therefore introduces additional complexity into the resulting algorithm when many variates are drawn from the same distribution. Leydold et al (2002) thus proposed *derandomized adaptive rejection sampling* (DARS). In DARS, after the initial decomposition has been established, one splits all intervals where the area between hat and squeeze is above some threshold value which is just the average over all intervals of the current decomposition. This procedure is repeated until the ratio ρ is as small as requested. The splitting points can be computed by the "arc-mean" of the boundaries of interval (b_{i-1}, b_i):

$$p_{\text{arc}} = \tan\left(\tfrac{1}{2}(\arctan(b_{i-1}) + \arctan(b_i))\right) \tag{2}$$

where $\arctan(\pm\infty)$ is set to $\pm\pi/2$, see also Hörmann et al (2004, Sect. 4.4.6).

2.4 Starting Intervals

The algorithm proposed in this paper requires that the user initially splits the domain of f into three types of intervals: those intervals in which $\tilde{f}$ is entirely concave, those intervals in which $\tilde{f}$ is entirely convex, and those intervals in which $\tilde{f}$ contains one inflection point. These intervals can be obtained just by examining a plot of $\tilde{f}''$. In those intervals where $\tilde{f}''$ is entirely negative, $\tilde{f}$ is concave; in those intervals where $\tilde{f}''$ is entirely positive, $\tilde{f}$ is convex, and in those intervals where there exists some x such that $\tilde{f}''(x) = 0$, $\tilde{f}$ contains an inflection point. This technique of identifying the initial intervals assumes, of course, that the user has correctly identified the number of $\tilde{f}$'s inflection points, i.e., that the user sees how many times $\tilde{f}''$ is 0. If we were to do away with this assumption, we would have to employ a method similar to that of Botts (2010), in which the user ascertains (after the sample has been collected) that the correct number of inflection points have been identified.

2.5 Appropriate Transformations

We restrict our interest to the family of T_c transformations proposed by Hörmann (1995), see Table 1. With this family of transformations, generating candidate values of X with density proportional to h can be done by means of the inversion method. To generate values of X using the inversion method, all one needs is the anti-derivative of T_c^{-1}, F_T, and its inverse, F_T^{-1}. These are easy to compute for this family of transformations and are given for all values of c in Table 1. With this family of transformations, it is also simple to decrease the number of inflection points in the transformed density, $\tilde{f}$. Hörmann et al (2004) observe that when a density is T_{c_o}-concave in some interval (b_l, b_r), then it is also T_c-concave for every $c < c_o$. Decreasing c may thus decrease the number of inflection points in $\tilde{f}$. In the following paragraph, more details are given on how the value of c should be selected.

For densities with unbounded domain, $c > -1$ is required, since otherwise the hat function has an unbounded integral. For unbounded densities, c must be sufficiently small to get rid of poles in the transformed scale, but c should still be greater than -1 to get an integrable hat function. It is often preferred to set $c = -1/2$, as this is in some sense equivalent to the ratio-of-uniforms method (see Leydold 2000) and thus leads to very fast marginal generation times. Setting $c = 0$ (setting $T = \log$) may also be preferred, since working with log-densities makes computing less susceptible to numerical overflow or underflow.

2.6 Computational Issues

We begin by writing the hat function for $\tilde{f}$ as

$$\tilde{h} = \alpha + \beta(x - x_0),$$

where x_0 is always the boundary point of the interval (b_l, b_r) where $\tilde{f}$ obtains its maximum. When the tangent serves as the hat function to $\tilde{f}$, $\alpha = \tilde{f}(x_0)$ and $\beta = \tilde{f}'(x_0)$, and when the secant serves as the hat function to $\tilde{f}$, $\alpha = \tilde{f}(x_0)$ and $\beta = \left(\tilde{f}(b_r) - \tilde{f}(b_l)\right) / (b_r - b_l)$. The area below the hat then becomes

$$\begin{aligned} A_h &= \int_{b_l}^{b_r} h(x)\,dx = \int_{b_l}^{b_r} T^{-1}(\alpha + \beta(x - x_0))dx \\ &= \frac{1}{\beta}\left[F_T(\alpha + \beta(b_r - x_0)) - F_T(\alpha + \beta(b_l - x_0))\right], \end{aligned} \tag{3}$$

and the (non-normalized) CDF, $H(x)$, of the density proportional to h becomes

$$\begin{aligned} H(x) &= \int_a^x T^{-1}(\alpha + \beta(t - x_0))dt \\ &= \frac{1}{\beta}\left[F_T(\alpha + \beta(x - x_0)) - F_T(\alpha + \beta(a - x_0))\right]. \end{aligned}$$

Notice that $H(b_r) = A_h$. Thus the inverse $H^{-1}(u)$ for $u \in [0, A_h]$ is then given by

$$H^{-1}(u) = x_0 + \frac{1}{\beta}\left[F_T^{-1}(\beta U + F_T(\alpha + \beta(a - x_0))) - \alpha\right].$$

In our extensive numerical experiments, we observed problems when x_0 is very close to a local maximum or another point where $\tilde{f}'(x_0) \approx 0$. To be more specific, we observed the situation where cancellation errors resulted in only two remaining digits for α and β and consequently in an invalid "hat" function. The resulting random sample then showed large defects.

For $c = 0$ we can avoid these problem in the following way. Let

$$z = \begin{cases} \beta(b_r - b_l), & \text{if } x_0 = b_l, \\ -\beta(b_r - b_l), & \text{if } x_0 = b_r. \end{cases}$$

Then a straightforward computation gives

$$\begin{aligned} A_h &= \frac{1}{\beta}\left(e^{\alpha + \beta(b_r - x_0)} - e^{\alpha + \beta(b_l - x_0)}\right) \\ &= e^{\alpha}(b_r - b_l)\frac{1}{z}(e^z - 1) \\ &\approx f(x_0)(b_r - b_l)\left(1 + \frac{z}{2} + \frac{z^2}{6}\right), \end{aligned} \tag{4}$$

where the last approximation follows from Taylor's theorem and is accurate up to machine precision (which is $2^{-52} \approx 2.2 \cdot 10^{-16}$) for $|z| < 10^{-6}$. Table 2 lists transformations and approximations for other values of parameter c. And for the two most important values of c, Table 3 lists the inverse functions $H^{-1}(u)$ and their approximations.

When $c \neq 0$ there might be a problem when using tangents to construct hat or squeeze functions. From Table 1 we can see that $\tilde{t}$ cannot be transformed back into a valid hat or squeeze function when $\tilde{t}$ vanishes inside the corresponding interval, that is, when there is a point $x \in [b_l, b_r]$ where $\tilde{t}(x) = 0$. (By construction this cannot happen for secants.) If such a problem arises and prevents the construction of

Table 1 The family T_c of transformations. F_t denotes the antiderivative of T_c^{-1}

c	T_c	$T_c(x)$	$T_c^{-1}(x)$	$F_T(x)$	$F_T^{-1}(x)$
>0	$(0,\infty)\to(0,\infty)$	x^c	$x^{1/c}$	$\frac{c}{c+1}x^{(c+1)/c}$	$\left(\frac{c+1}{c}x\right)^{c/(c+1)}$
0	$(0,\infty)\to\mathbb{R}$	$\log(x)$	e^x	e^x	$\log(x)$
<0	$(0,\infty)\to(-\infty,0)$	$-x^c$	$(-x)^{1/c}$	$-\frac{c}{c+1}(-x)^{(c+1)/c}$	$-\left(-\frac{c+1}{c}x\right)^{c/(c+1)}$
$-1/2$	$(0,\infty)\to(-\infty,0)$	$-1/\sqrt{x}$	$1/x^2$	$-1/x$	$-1/x$
-1	$(0,\infty)\to(-\infty,0)$	$-1/x$	$-1/x$	$-\log(-x)$	$-\exp(-x)$

Table 2 Area below hat function and its approximation for small values of $\tilde{f}'(x_0)$. ($\sigma = 1$ for $x_0 = b_l$ and $\sigma = -1$ for $x_0 = b_r$)

c	A_h	Approximation	z
1	$f(x_0)(b_r-b_l)\frac{1}{2}(2+z)$		$\sigma\frac{1}{\alpha}(b_r-b_l)$
0	$f(x_0)(b_r-b_l)\frac{1}{z}(e^z-1)$	$f(x_0)(b_r-b_l)\left(1+\frac{z}{2}+\frac{z^2}{6}\right)$	$\sigma(b_r-b_l)$
$-1/2$	$f(x_0)(b_r-b_l)\frac{1}{1+z}$	$f(x_0)(b_r-b_l)\left(1-z+z^2\right)$	$\sigma\frac{1}{\alpha}(b_r-b_l)$
-1	$f(x_0)(b_r-b_l)\frac{1}{z}\log(1+z)$	$f(x_0)(b_r-b_l)\left(1-\frac{1}{2}z+\frac{1}{3}z^2\right)$	$\sigma\frac{1}{\alpha}(b_r-b_l)$
otherwise	$f(x_0)(b_r-b_l)\frac{c}{c+1}\cdot\frac{1}{z}\left[(1+z)^{(c+1)/c}-1\right]$		$\sigma\frac{1}{\alpha}(b_r-b_l)$

Table 3 Inverse CDF of "hat distribution" and its approximation for small values of $\tilde{f}'(x_0)$

c	$H^{-1}(u)$	Approximation	z
0	$b_l+\frac{u}{\exp(\alpha+\beta(b_l-x_0))}\cdot\frac{1}{z}\log(1+z)$	$b_l+\frac{u}{\exp(\alpha+\beta(b_l-x_0))}\cdot\left(1-\frac{z}{2}+\frac{z^2}{3}\right)$	$\frac{\beta u}{\exp(\alpha+\beta(b_l-x_0))}$
$-1/2$	$b_l+u(\alpha+\beta(b_l-x_0))^2\cdot\frac{1}{1-z}$	$b_l+u(\alpha+\beta(b_l-x_0))^2\cdot(1+z+z^2)$	$\beta u(\alpha+\beta(b_l-x_0))$

a valid squeeze function in (b_l,b_r), the algorithm will proceed without a squeeze function in (b_l,b_r). If such a problem arises and prevents the construction of a valid hat function, however, the interval (b_l,b_r) will be split into smaller intervals (using the "arc-mean" method described in Section 2.4) until a valid hat function is obtained.

3 The Algorithm

Now we can compile an algorithm that is based on Theorems 1 and 2. Table 4 presents Algorithm `Tinflex-log` that implements case of $T = \log$. Notice that Step 14 can be executed in constant time (i.e., independent of the number of intervals) by means of the alias method or the guide table method (see, e.g., Hörmann et al 2004, Sect. 3).

It is obvious that this algorithm can easily be generalized for arbitrary transformations T_c by using the formulæ from Tables 2 and 3. However, for $c < 0$ one must check whether a tangent results in a valid (bounded) hat function. Otherwise, we have to split the corresponding interval. This can be easily implemented by setting the area in such intervals to $A_{h,i} = \infty$.

In Condition 4 we have demanded that $\tilde{f}$ must be concave and strictly monotone in each unbounded interval $(-\infty,b_1]$ or $[b_{n-1},\infty)$ of the given starting partition. In practice it is only necessary that intervals $(-\infty,b_r]$ and $[b_l,\infty)$ with this property exist and that there is at most one inflection point of $\tilde{f}$ in the given starting intervals for $(-\infty,b_1]$ or $[b_{n-1},\infty)$. If $[b_{n-1},\infty)$ contains an inflection point, then it is not possible to construct a hat function according to our rules. Thus $A_{h,n-1}$ is set to ∞. Then the logic of DARS will split that interval in the next cycle. This splitting is repeated until a point satisfying Condition 4 is found.

Remark 3 Notice that Algorithm `Tinflex-log` as well as its generalization works for any multiple of a density f. Thus there is no necessity to compute a normalization constant.

Remark 4 It is obvious that we can replace DARS (Steps 7–12) by ARS. For this purpose we have to add appropriate steps in the generation part after a candidate point X has been rejected.

We have coded a proof-of-concept implementation of Algorithm `Tinflex-log` using the **R** programming language for statistical computing (R Development Core Team 2010).

Table 4 Algorithm `Tinflex-log`

Input: Log-density $\tilde{f}$ with domain (b_l, b_r) and its derivatives $\tilde{f}'$ and $\tilde{f}''$ that satisfy Conditions 1–3;
partition $b_l = b_0 < b_1 < \ldots < b_{n-1} < b_n = b_r$ that satisfies Condition 4;
maximal accepted value for $\rho_{\max}$.
Output: Random variate X with density f.

▷ Setup: Initial intervals
for $i = 0, \ldots n$ **do**
  Compute $\tilde{f}(b_i)$, $\tilde{f}'(b_i)$, and $\tilde{f}''(b_i)$.
for all intervals $[b_i, b_{i+1}]$ **do**
  Determine type of interval using Theorem 2.
  Determine x_0 and compute intercepts α and slopes β of hat $\tilde{h}_i$ and squeeze $\tilde{s}_i$ using Theorem 1.
  Compute area $A_{h,i}$ below hat and area $A_{s,i}$ below squeeze using formula from Table 2.
▷ Setup: Derandomized adaptive rejection sampling
repeat
  $A_h \leftarrow \sum A_{h,i}$ and $A_s \leftarrow \sum A_{s,i}$.
  $\bar{A} \leftarrow (A_h - A_s)/(\#\text{ intervals})$.
  for all intervals with $(A_{h,i} - A_{s,i}) > \bar{A}$ **do**
    Split interval using "arc-mean" (2) and compute hat, squeeze and areas for the two new intervals.
until $A_h/A_s \leq \rho_{\max}$
▷ Generation
loop
  Generate J with probability vector proportional to $(A_{h,1}, A_{h,2}, \ldots)$.
  Generate X with density prop. to h_J using formula from Table 3.
  Generate $U \sim U(0,1)$.
  if $U\,h(X) \leq s(X)$ **then** ▷ evaluate squeeze
    return X.
  if $U\,h(X) \leq \exp(\tilde{f}(X))$ **then** ▷ evaluate density
    return X.

We also provide a ready-to-use version of the most general algorithm with **R** package `Tinflex` which is available at the CRAN. An advantage of the proposed algorithm is that the intervals can be treated independently from each other, i.e., virtually we have a mixture of distribution on mutually disjoint domains. Thus it allows (mostly) arbitrary values of c which may differ on different intervals of the starting partition. Moreover, points that violate Condition 3 may be used as partition points. Thus the routine can handle densities with cusps (although these have been excluded by Condition 3 for our theoretical considerations). Consequently it also can handle (to some extend) densities with poles.

Both implementations, `Tinflex-log.R` and **R** package `Tinflex`, are also available as online supplement to this paper and demonstrate the usefulness of this method.

4 Examples

In this section we demonstrate in three examples the usefulness of the proposed method.

4.1 A Simple Example

We start with a simple example. Assume that we have to draw a sample from a distribution with density proportional to

$$f(x) = \exp\left(-|x|^{\alpha} + s|x|^{\beta} + \varepsilon x^2\right), \quad \alpha > \beta \geq 2,\ s, \varepsilon > 0\,.$$

The second derivative of the log-density is then

$$\tilde{f}''(x) = -\alpha(\alpha-1)|x|^{\alpha-2} + s\beta(\beta-1)|x|^{\beta-2} + 2\varepsilon\,.$$

Observe that $\tilde{f}''(0) = 2\varepsilon > 0$ and $\lim_{x\to\infty} \tilde{f}'' = -\infty$. Moreover, $\tilde{f}''(x) \leq 0$ implies that

$$\alpha(\alpha-1)|x|^{\alpha-2} \geq s\beta(\beta-1)|x|^{\beta-2} + 2\varepsilon\,.$$

Consequently, if $x > 0$ and $\tilde{f}''(x) \leq 0$ then we find for the third derivative at x,

$$\begin{aligned}\tilde{f}'''(x) &= -\frac{\alpha-2}{x}\left[\alpha(\alpha-1)|x|^{\alpha-2} + \frac{s\beta(\beta-1)(\beta-2)}{\beta-1}|x|^{\beta-2}\right] \\ &\leq -\frac{\alpha-2}{x}\left[s\beta(\beta-1)\left(1-\frac{\beta-2}{\alpha-2}\right)|x|^{\beta-2} + 2\varepsilon\right] \\ &< 0\end{aligned}$$

Hence if x_0 is the smallest root of $\tilde{f}''$ with $x_0 > 0$, then $\tilde{f}''$ is strictly decreasing in $[x_0, \infty)$. Consequently x_0 is the unique inflection point in $[0, \infty)$ and f is log-concave in the tail. By symmetry the same holds for $x \in (-\infty, 0]$. Hence we can run Algorithm `Tinflex-log` with starting partition points $\{-\infty, 0, \infty\}$ for drawing a sample from this distribution.

4.2 Exponential Power Distribution

For the second example we consider the family of exponential power distributions (EP) with density proportional to

$$f_{EP}(x) = \exp(-|x|^{\alpha})\,, \quad \alpha > 0\,.$$

It is a generalization of the normal distribution where $\alpha = 2$. For $\alpha \geq 1$ the density is log-concave. For $\alpha < 1$, however, it is strongly log-convex near the mode $x = 0$.

In the framework of transformed density rejection, the notion of the *local concavity* of a density f at a point x is a very convenient tool to analyze a given distribution (Hörmann et al 2004, Sect. 4.3). It is defined as the maximal value for c

such that the transformed density $\tilde{f}(x) = (T_c \circ f)(x)$ is concave near x. For a twice differentiable density f it is given by

$$\text{lc}_f(x) = 1 - \frac{f''(x)\, f(x)}{f'(x)^2} \; .$$

A density f is T_c-concave at x if and only if $\text{lc}_f(x) \geq c$. Moreover, $\tilde{f} = T_c \circ f$ has an inflection point in x if and only if $\text{lc}_f(x) = c$.

For the exponential power distribution we find

$$\text{lc}_{EP}(x,\alpha) = \frac{\alpha-1}{\alpha}\, |x|^{-\alpha} \; .$$

Since the distribution is symmetric, it is enough to consider positive values of x. Recall that we consider the case where $\alpha < 1$. When x approaches 0 then $\text{lc}_{EP}(x)$ diverges to $-\infty$, and for increasing x it is strictly monotonically increasing with limit 0. Therefore we cannot apply transformation $T(x) = \log(x)$, i.e., $c = 0$. However, for any fixed value of $c < 0$ the transformed density has exactly one inflection point in domain $(0,\infty)$, and it is T_c-concave for sufficiently large x. We therefore can apply the generalized algorithm with $c = -1/2$ for the truncated exponential power distribution on $[0,\infty)$ with starting partition points $\{0,\infty\}$. Note that when executing our algorithm to generate from an EP distribution with $\alpha < 1$, the first derivative of the log-density must return the right derivative at $x = 0$. We analogously generate from the truncated exponential power distribution on the domain $(-\infty,0]$, and in this case the first derivative of the log-density must return the left derivative. Drawing a random sample from the exponential power distribution using the proposed algorithm is thus simple; we merely combine the samples generated for the two truncated distributions.

A simpler approach, however, would be to implement the first derivative of the log-density such that it returns 0 for the mode at $x = 0$. Then for a narrow interval $[0,\varepsilon]$ the transformed density is classified as case (Ia) by the rules of Theorem 1. The algorithm then creates a valid constant hat using the pseudo-tangent at $x = 0$, and a valid squeeze using the tangent at $x = \varepsilon$. This interval has to be chosen such that there is no (real) inflection points of $\tilde{f}$. A suitable choice is $\varepsilon = (1-\alpha)/2$. We also have to take care that $\tilde{f}_{EP}$ forms a cusp when $\alpha < 1$. We thus must use $x = 0$ as partition point. We can therefore apply the generalized algorithm with $c = -1/2$ for the exponential power distribution using $c = -1/2$ and starting partition points $\{-\infty, -(1-\alpha)/2, 0, (1-\alpha)/2, \infty\}$.

We ran our code for values of α between 0.015 and 0.99. In all cases routine `Tinflex` had no problem to construct a hat function with $\rho_{\max} = 1.1$. The resulting ratios $\rho = A_h/A_s$ were always less than 1.1. The observed numbers of intervals were 15 for $\alpha = 0.99$, increased to 88 for $\alpha = 0.1$ and was almost 1000 for $\alpha = 0.015$. (For $\alpha < 0.015$ the routine did not work due to numeric overflow and other limitations of floating point arithmetic.)

4.3 Generalized Gaussian Distribution

As our third example we consider the generalized inverse Gaussian Distribution (GIG). For the purpose of random variate generation, we only need to consider the two parameter family with density proportional to

$$f_{GIG}(x;\lambda,\omega) = x^{\lambda-1} \exp\left(-\frac{\omega}{2}\left(x + \frac{1}{x}\right)\right) \,, \quad x > 0$$

and λ and ω positive (see Devroye 1986). Its local concavity is given by

$$\text{lc}_{GIG}(x,\lambda,\omega) = \frac{4x(\omega + (\lambda-1)x)}{(\omega + 2(\lambda-1)x - \omega x^2)^2} \; .$$

The GIG distribution (as defined above) is log-concave for $\lambda \geq 1$ and $T_{-1/2}$-concave for $\omega \geq 0.5$. We thus only consider the cases where $\lambda < 1$ and $\omega < 0.5$.

When $\lambda, \omega < 1$ it is no problem to see that for $x \leq \omega/(1-\lambda)$, $\text{lc}_{GIG} \geq 0$ and hence f_{GIG} is log-concave. The first derivative of the local concavity is given by

$$\text{lc}'_{GIG}(x\lambda,\omega) = -4\,\omega \frac{2(\lambda-1)x^3 + 3\omega x^2 + \omega}{(\omega x^2 - 2(\lambda-1)x - \omega)^3} \; .$$

It numerator is a multiple of the cubic polynomial

$$q(x) = 2(\lambda-1)x^3 + 3\omega x^2 + \omega \; .$$

For $x > \omega/(1-\lambda)$ the local concavity thus has a single local minimum and converges to 0 when x tends to ∞. Notice that the numerator $q(x)$ has exactly one real root. Consequently, for $c = -1/2$ the transformed density has exactly two inflection points if the minimum value of lc_{GIG} is less than -0.5. Thus we have to add at least one point in the region where $\text{lc}_{GIG} < -0.5$ to the set of starting partition points as this guarantees that there is at most one inflection point in each starting interval.

In practice it is enough to check that at least for one starting point the second derivative of the transformed density is negative. It is also possible to find the approximate location of the minimum of lc_{GIG} by finding the (approximate) location of the real root r_0 of $q(x)$ using numeric search starting at $x = \omega/(1-\lambda)$. The mode of the distribution,

$$m = \left(\lambda - 1 + \sqrt{(\lambda-1)^2 + \omega^2}\right)/\omega \, ;$$

may also be used as a convenient partition point. We can therefore apply the generalized algorithm for the GIG distribution using $c = -1/2$ and starting partition points $\{0, m, r_0, \infty\}$. Of course it is important that the log-density $\log \circ f_{GIG}$ and its derivative are carefully implemented for $x = 0$.

We ran our code for values of λ between 0.01 and 0.9 and for values of ω in the set $\{10^{-15}, 10^{-14}, \ldots, 10^{-2}, 0.1, 0.2, 0.3, 0.4, 0.5\}$ In all cases routine `Tinflex` had no problem to construct a hat function for $c = -1/2$ with $\rho_{\max} = 1.1$. The resulting ratios $\rho = A_h/A_s$ were always less than 1.1. The observed numbers of intervals were 10–13 for $\omega \geq 0.1$ and increased to 100–120 for $\omega = 10^{-15}$.

Notice that for parameters $\lambda = 0.4$ and $\omega = 10^{-7}$ the most popular generation method for GIG variables by Dagpunar (1989) which is based on the ratio-of-uniforms method with a bounding rectangle is extremely slow due to its huge rejection constant of about 8500. For values of ω less than 10^{-8} all implementations of this generator known to the authors fail (or produces an invalid sample).

The above approach has the drawback that we need the approximate location r_0 of the real root of $q(x)$. An alternative approach avoids this step by using two different transformation T_c. Notice that the cubic $q(x) - \omega = 2(\lambda-1)x^3 + 3\omega x^2$ has a root at $x_0 = \frac{3}{2}\frac{\omega}{1-\lambda}$. It is not difficult to show that $q(x)$ is concave for $x \geq x_0$. We can therefore use the root $r_1 = \frac{3}{2}\frac{\omega}{1-\lambda} + \frac{2}{9}\frac{1-\lambda}{\omega}$ of the tangent $q(x_0) + q'(x_0)(x - x_0)$ as an upper bound for the location of the minimum of $\mathrm{lc}_{GIG}(x)$. We can therefore apply the generalized algorithm with $c = 0$ on the interval $[0, r_1]$ and with $c = -1/2$ on the interval $[r_1, \infty)$. Again routine `Tinflex` had no problem to construct a hat function with $\rho_{\max} = 1.1$. However, as we have more intervals where the transformed density is log-convex we need (up to 60%) more intervals.

5 Conclusions

The algorithm presented in this paper is a user-friendly adaptive acceptance-rejection algorithm. It is user-friendly in the sense that hat and squeeze functions of f are constructed without the user having to know the exact location of the inflection points of the transformed density. The only input required from the user is the transformation T and the intervals in the domain of f where the transformed density is either entirely concave, entirely convex, or contains only one inflection point. Areas of future research include how to optimally select these initial intervals for a given density, and how to generalize this algorithm so that it can be applied to densities that do not necessarily satisfy conditions 1–3.

References

Botts C (2010) A modified adaptive accept-reject algorithm for univariate densities with bounded support. Journal of Statistical Computation and Simulation, to appear.

Dagpunar J (1989) An easily implemented generalised inverse gaussian generator. Communications in Statistics, Simulation and Computing 18:703–710

Devroye L (1984) A simple algorithm for generating random variates with a log-concave density. Computing 33(3–4):247–257

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

Evans M, Swartz T (1998) Random variable generation using concavity properties of transformed densities. Journal of Computational and Graphical Statistics 7(4):514–528

Gilks WR, Wild P (1992) Adaptive rejection sampling for Gibbs sampling. Applied Statistics 41(2):337–348

Hörmann W (1995) A rejection technique for sampling from T-concave distributions. ACM Trans Math Software 21(2):182–193

Hörmann W, Leydold J, Derflinger G (2004) Automatic Nonuniform Random Variate Generation. Springer-Verlag, Berlin Heidelberg

Leydold J (2000) Automatic sampling with the ratio-of-uniforms method. ACM Trans Math Software 26(1):78–98, DOI http://doi.acm.org/10.1145/347837.347863

Leydold J, Janka E, Hörmann W (2002) Variants of transformed density rejection and correlation induction. In: Fang KT, Hickernell FJ, Niederreiter H (eds) Monte Carlo and Quasi-Monte Carlo Methods 2000, Springer-Verlag, Heidelberg, pp 345–356

R Development Core Team (2010) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org, ISBN 3-900051-07-0