ASYMPTOTICALLY OPTIMAL DESIGN POINTS FOR REJECTION ALGORITHMS

Gerhard Derflinger and Wolfgang Hörmann
Institut für Statistik und Mathematik
Wirtschaftsuniversität Wien
Augasse 2-6
A-1090, Wien, Austria
derfling@statistik.wu-wien.ac.at hormannw@boun.edu.tr

Key Words: Random variate generation, rejection method, transformed density rejection, optimal design points

Thanks: The authors thank Josef Leydold and Günter Tirler for valuable discussions and comments. This work was supported by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung*, Proj. P16767-N12.

ABSTRACT

Very fast automatic rejection algorithms were developed recently which allow to generate random variates from large classes of unimodal distributions. They require the choice of several design points which decompose the domain of the distribution into small sub-intervals. The optimal choice of these points is an important but unsolved problem. So we present an approach that allows to characterize optimal design points in the asymptotic case (when their number tends to infinity) under mild regularity conditions. We describe a short algorithm to calculate these asymptotically optimal points in practice. Numerical experiments indicate that they are very close to optimal even when only six or seven design points are calculated.

1. INTRODUCTION

Random variate generation started about 45 years ago with the design of generation algorithms for standard distributions like e.g. the normal, gamma and beta distributions. Devroye (2) developed the first algorithms for larger distribution families. The user of

such a "Black-box" (also called universal or automatic) algorithm has to provide typically the density of the desired distribution, often together with some other information like e.g. the mode of the distribution. Then the *set-up* step of the algorithm computes all necessary constants whereas the *sampling* part of the algorithm utilizes these constants to generate random variates from the desired distribution. In the last decade several fast and reliable universal algorithms were suggested (e.g. (4), (6), (1), (9), (3)). They clearly have important advantages for the user, as they provide the possibility to sample from very different distributions with a single algorithm. This facilitates the change of input distributions in simulation programs and saves the time to design and code random variate generation programs for little known or newly defined distributions.

All these automatic algorithms are based on the rejection method. The decomposition of the domain of the distribution into many sub-intervals leads to good fitting hat- and squeeze functions and thus to (very) fast algorithms. This implies that the set-up of all these algorithms contains the choice of these sub-intervals or of design points within each sub-interval. The natural question of the optimal choice of these design points was posed in all of the above references. As there is no solution to this problem published yet, this paper introduces an asymptotic theory, asymptotic in the sense that the number of design points tends to infinity, that leads to approximately optimal design points.

We start with a short description of transformed density rejection in Section 2. Section 3 develops approximations for the area below the hat that are used in Section 4 to prove our main theorem on asymptotically optimal design points. Section 5 shortly presents similar results for other variants of automatic rejection algorithms, whereas Section 6 shows that our method can be easily extended to the rejection method using piecewise constant hats and squeezes. Section 7 discusses how we can compute nearly optimal design points in practice, even under extreme conditions.

2. TRANSFORMED DENSITY REJECTION (TDR)

This method was introduced in (4) and (6), generalizations were suggested in (3), a detailed explanations of many variants are given in (8). It is based on the idea that the given

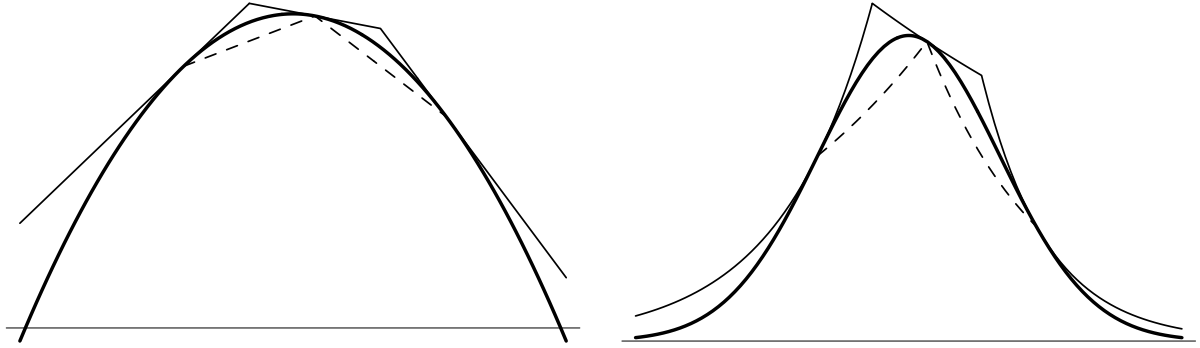

Figure 1: Hat function (thin line) and squeeze (dashed line) with three points of contact for the normal density (thick line) and logarithm as transformation, design points at -1, 0.2, and 1.4. Transformed scale (l.h.s.) and original scale (r.h.s.)

density f is transformed by a strictly monotonically increasing transformation $T\colon (0,\infty) \to \mathbf{R}$ such that $T(f(x))$ is concave. The density f is then called *T-concave*; log-concave densities are an example with $T(x) = \log(x)$.

By the concavity of $T(f(x))$ it is easy to construct a majorising function for the transformed density as the minimum of N tangents. Transforming this function back into the original scale we get a hat function $h(x)$ for the density f. By using secants between the touching points of the tangents of the transformed density we analogously can construct squeezes $s(x)$. Figure 1 illustrates the situation for the standard normal distribution, $T(x) = \log(x)$ and $N = 3$ touching points.

It is obvious that the transformation T must have the property that the area below the hat is finite, and that generating a random variable with density proportional to the hat function by inversion must be easy (and fast). Thus we have to choose the transformations T carefully. Hörmann (6) suggests the family T_c of transformations which can be represented by

$$T_c(x) = \frac{x^c - 1}{c}. \tag{1}$$

As a linear transformation applied to T_c does not change the resulting hat, Hörmann (6)

suggested the simpler version

$$T_0(x) = \log(x) \quad \text{and} \quad T_c(x) = \text{sign}(c)\, x^c \text{ for } c \neq 0\,. \tag{2}$$

$T_0(x)$ is the limit of $T_c(x)$ for $c \to 0$, sign(c) makes T_c increasing for $c \neq 0$. Also note that, as a consequence of the mentioned independence under linear transformations, the class T_c, $c \in \mathbf{R}$, has the convenient property that multiplying the density with a constant factor does not change the constructed hat and squeeze functions; both are only multiplied with the same factor. (For details see (8).) This implies that TDR works without problems if the integral of f is bounded but not equal to one. We call such a function f a *quasi-density*. For densities with unbounded domain we must have $c \in (-1, 0]$.

For the choice of c it is important to note that the area below the hat increases when c decreases. Moreover we find that if f is T_c-concave, then f is $T_{c'}$-concave for every $c' \leq c$. Because of computational reasons, the choice of $c = -1/2$ (if possible) is suggested. Then TDR can generate random variates for a larger family than the log-concave family, all $T_{-1/2}$-concave distributions. (All distributions of this family are unimodal with subquadratic tails.) For the case that a density is not T_c-concave (3) and (8) discuss generalizations of TDR that are applicable to such densities.

TDR – like any rejection algorithm – works fastest when the area below the hat is as small as possible. This can be reached easily by using many design points N. But how should we choose the location of the design points? In the literature it is mainly suggested to use adaptive methods. This idea is quite simple and leads to a fast set-up and good results for most densities. However for the case that we want to code a single generator for a fixed distribution the idea of randomly selected design points is unelegant and the speed of the set-up is irrelevant. Thus an optimal choice of a fixed number of design points is of relevance as we can save about fifty percent of the evaluations of f compared to the adaptive selection procedures (see (8) p. 87 for results of numerical experiments). This has little influence on the speed of the algorithm for standard distributions but is important if the evaluation of the density is very expensive which is for example the case for order statistics.

Optimal design points should minimize either the total area below the hat, i.e. the expected number α of iterations of the rejection algorithm, or the total area between hat and squeeze which minimizes the expected number N_f of calls to the density f.

It is possible to find simple characterizations for the optimal design points for the cases $N = 1,\ N = 2$ (see (8)), and the case $N = 3$ when the mode of the distribution is taken as a design point (5). For $N > 3$ conditions for optimal design points are so complicated that they are not useful for practical purposes. That is the reason why we develop an asymptotical theory that allows to compute even for very small N (e.g. $N \gtrsim 5$) close to optimal design points. [1] The results of this paper are valid for general transformations T. It is only necessary to assume that T is monotonically increasing and five times differentiable on $\mathbf{R}^+$.

3. THE AREAS OF SINGLE CELLS

To obtain asymptotic formulas for the total area below the hat, and the total area between squeeze and hat we start to compute the area $a_1(p, \Delta)$ of the cell between hat and density (see Figure 2) in the interval $(p,\ p + \Delta)$ that must be a subset of the domain of the distribution. The hat is given by the minimum of the two backtransformed tangents in p and $p + \Delta$. For the hat function $h(p + t),\ 0 < t \le \Delta$, we can write

$$h(p+t) = \begin{cases} T^{-1}(T(f(p)) + T(f(p))'\, t) & \text{for } t \le \mu(p,\ p+\Delta), \\ T^{-1}(T(f(p+\Delta)) + T(f(p+\Delta)')(t-\Delta)) & \text{for } t > \mu(p,\ p+\Delta), \end{cases} \tag{3}$$

where $\mu(p,\ p + \Delta)$,

$$\mu(p,\ p+\Delta) = \frac{T(f(p)) - T(f(p+\Delta)) - p\,T(f(p))' + (p+\Delta)\,T(f(p+\Delta))'}{T(f(p+\Delta))' - T(f(p))'}, \tag{4}$$

is the abscissa of the intersection point of the tangents in p and $p + \Delta$. The area $a_1(p, \Delta)$ between hat and density is given by

$$a_1(p, \Delta) = \int_0^\Delta h(p+t) - f(p+t)\, dt. \tag{5}$$

[1] For the Cauchy (Student's t_1), Student's t_3, and the exponential distributions, all with $c = -1/2$, we have derived exact formulas for calculating the design points, and α and N_f, in the case of finite N.

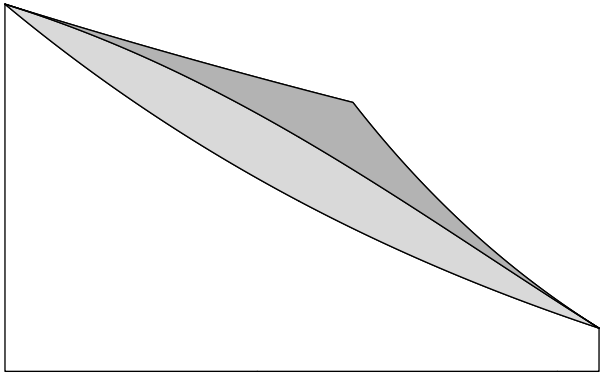

Figure 2: The cells asociated with a_1 (dark shaded), a_2 (light shaded) and a_3 (both shaded regions together)

To derive an asymptotic formula we consider for given p the case $\Delta \to 0$. Clearly we have $\lim_{\Delta\to 0} a_1(p,\Delta) = 0$ and we can also calculate the derivatives of $a_1(p,\Delta)$ with respect to Δ. This is very cumbersome and the formulas of the higher of the used derivatives become extremely extensive. But using computer algebra and the definition above it is not difficult to obtain the first three derivatives evaluated at 0. We get

$$a_1(p,0) = 0, \quad a_1'(p,0) = 0, \quad a_1''(p,0) = 0, \quad a_1'''(p,0) = -\frac{T(f(p))''}{4\,T'(f(p))}\,. \tag{6}$$

In addition it is possible to check that $a_1^{(4)}(p,\Delta)$ depends on the 5^{th} derivative $f^{(5)}(p+\Delta)$ of the density and lower order derivatives. Using (6) we can expand $a_1(p,\Delta)$ into a MacLaurin series:

$$a_1(p,\Delta) = -\frac{T(f(p))''}{24\,T'(f(p))}\,\Delta^3 + O(\Delta^4)\,. \tag{7}$$

From the formulas for the remainder of the Taylor series after Lagrange or after Cauchy it follows that the existence of a continuous fourth derivative of $a_1(p,\Delta)$ is sufficient for the validity of the remainder in Eqn. (7). This is guaranteed by the existence of a 5^{th} continuous derivative $f^{(5)}(x)$ of the density. In Section 6 we will see that this condition on the density is not necessary for computing close to optimal design points in practice.

To minimize the expected number N_f of evaluations of the density $f(x)$ we have to consider the area $a_3(p,\Delta)$ of the cell between hat and squeeze (see Figure 2). As $a_3(p,\Delta)$ is

the sum of $a_1(p,\Delta)$ and the area

$$a_2(p,\Delta) = \int_0^{\Delta} f(p+t) \; - \; T^{-1}(T(f(p)) + s(p,\Delta)\, t)\, dt \tag{8}$$

between density and squeeze we first consider $a_2(p,\Delta)$. In (8) $s(p,\Delta) = (T(f(p+\Delta)) - T(f(p)))/\Delta$ denotes the slope of the transformed squeeze which, for the interval $(p, p+\Delta)$, is simply a linear function connecting the points $(p, T(f(p)))$ and $(p+\Delta, T(f(p+\Delta)))$. Thus it is possible to find a MacLaurin series for the area $a_2(p,\Delta)$ as well. Again using computer algebra we get

$$a_2(p,0) = 0, \quad a_2'(p,0) = 0, \quad a_2''(p,0) = 0, \quad a_2'''(p,0) = -\frac{T(f(p))''}{2\,T'(f(p))}\,. \tag{9}$$

For $a_2^{(4)}(p,\Delta)$ the existence of the forth derivative $f^{(4)}(x)$ of the density is required. We thus obtain the MacLaurin series for the area between density and squeeze in a single cell:

$$a_2(p,\Delta) = -\frac{T(f(p))''}{12\,T'(f(p))}\,\Delta^3 \; + \; O(\Delta^4). \tag{10}$$

The area $a_3(p,\Delta)$ of the cell between hat and squeeze is the sum of (7) and (10):

$$a_3(p,\Delta) = -\frac{T(f(p))''}{8\,T'(f(p))}\,\Delta^3 \; + \; O(\Delta^4). \tag{11}$$

Eqns. (7), (10) and (11) can be unified as

$$a_j(p,\Delta) = -\frac{j\,T(f(p))''}{24\,T'(f(p))}\,\Delta^3 \; + \; O(\Delta^4), \quad j = 1,2,3. \tag{12}$$

4. ASYMPTOTIC OPTIMIZATION

Let q, r be the borders of the domain of the density $f(x)$ and let $p_0 = q,\ p_1, ..., p_{N-1}$, $p_N = r$ be the design points which decompose the domain into N intervals. We can apply the above MacLaurin series to all sub-intervals of a decomposition of the domain. For this purpose we replace p by p_i and Δ by $\Delta_i,\ \Delta_i = p_{i+1} - p_i$. Defining the abbreviation $\theta(p)$,

$$\theta(p) = -\frac{T(f(p))''}{24\,T'(f(p))}, \tag{13}$$

we can write for the MacLaurin series of the areas in a single cell

$$a_j(p_i,\ \Delta_i) = j\ \theta(p_i)\ \Delta_i^3 + O(\Delta_i^4), \quad i = 0, ..., N-1, \quad j = 1, 2, 3. \tag{14}$$

To obtain the corresponding areas A_j for the whole domain $[q, r]$ we simply sum Eqn. (14) over all i

$$A_j = \sum_{i=0}^{N-1} (j\ \theta(p_i)\Delta_i^3 + O(\Delta_i^4)), \quad j = 1, 2, 3. \tag{15}$$

Let ψ, Ψ, ω, Ω be positive constants. Then, in order to compute asymptotically optimal design points we make the following assumptions:

- $T'(x)$ is bounded according to

$$\psi \le T'(x) \le \Psi, \quad \text{for} \quad x \in [\min_z(f(z)), f(m)], \tag{16}$$

  where m is the mode of the density.

- $-T(f(x))''$ is bounded by

$$\omega \le -T(f(x))'' \le \Omega. \tag{17}$$

- The quasi-density $f(x)$ is continuously differentiable five times.

The last assumption was discussed in Section 3. It is necessary to obtain the $O(\Delta^4)$ error bound for the MacLaurin series expansion of the areas. Assumptions (16) and (17) can be easily justified. Assumption (16) guarantees that the contraction of $T(x)$, $L_T(x_1, x_2) = (T(x_2) - T(x_1))/(x_2 - x_1)$, and that of $T^{-1}(x)$, $L_{T^{-1}}(x_1, x_2)$, as well cannot have values arbitrarily close to zero. For the class T_c of transformations and a T_c-concave density f Assumption (16) is only not fulfilled if $f(x)$ can approach arbitrarily close to zero. However, as f is T_c-concave it is no problem to find an interval (p_0, p_N) for which Assumption (16) is fulfilled. (See Section 7 below for the choice of that interval.) Assumption (17) excludes densities whose transformed density $T(f(x))$ has points that almost are corner points or

regions where $T(f(x))$ is arbitrarily close to linear. The problem with a linear region for $T(f(x))$ can easily be solved by using a transform T_c with smaller c, the problem with densities with corners can only be solved by decomposing the domain of such a distribution at the corner points.

If both assumptions are fulfilled $\theta(p)$, see (13), is bounded by the positive constants $\phi = \omega/(24\,\Psi)$ and $\Phi = \Omega/(24\,\psi)$,

$$\phi \leq \theta(p) \leq \Phi. \tag{18}$$

In order to compute asymptotically optimal design points we define the *partitioning function* $v(x)$, continuous on $[q, r]$ with $v(x) > 0$ and

$$\int_q^r v(x)dx = 1. \tag{19}$$

The partitioning function $v(x)$ generates the design points $p_1, p_2, ..., p_{N-1}$ according to

$$\int_{p_i}^{p_{i+1}} v(x)dx = \frac{1}{N}, \quad i = 0, ..., N-1. \tag{20}$$

We call a partitioning function $v(x)$ *regular* if there exist positive constants $m,\ M$ with $m \leq v(x) \leq M$ for $q \leq x \leq r$. Clearly, $v(x)$ is not unique. But we can prove the following main result:

THEOREM 1: Let $f(x)$ be a five times continuously differentiable quasi-density, defined on $[q, r]$, which fulfills (18). For $p_0 = q$ and $p_N = r$ fixed, and $N \to \infty$ the uniquely determined regular partitioning function $\tilde{v}(x)$, which leads to the minimal A_j for $j = 1, 2, 3$ among all regular partitioning functions is:

$$\tilde{v}(x) = C\,\theta(x)^{\frac{1}{3}}, \quad \text{where} \quad C = \frac{1}{\int_q^r \theta(x)^{\frac{1}{3}}dx}. \tag{21}$$

For the asymptotically optimal partition we obtain the area-formula

$$\lim_{N\to\infty} N^2 \tilde{A}_j = j\left(\int_q^r \theta(x)^{\frac{1}{3}}dx\right)^3, \quad j = 1, 2, 3. \tag{22}$$

Proof: Applying the mean value theorem on (20) shows that there are $\xi_i,\ i=0,...,N-1$, with

$$\Delta_i = \frac{1}{N v(\xi_i)}, \quad p_i \le \xi_i \le p_{i+1}. \tag{23}$$

Inserting this into (15) we obtain

$$A_j\ N^2 = \sum_{i=0}^{N-1} \frac{j\ \theta(p_i)}{v(\xi_i)^2}\Delta_i + N^2 \sum_{i=0}^{N-1} O(\Delta_i)^4, \quad j=1,2,3. \tag{24}$$

For the limit $N \to \infty$, $\Delta_i = O(1/N)$ for all $i = 1, 2, 3\ldots$, as v is a regular partitioning function. Together with condition (18) this also implies that $\frac{j\,\theta(x)}{v(x)^2}$ is bounded. Thus for $N \to \infty$ the first summand of Eqn. (24) can be written as an integral whereas the second summand converges to 0. We have:

$$\lim_{N\to\infty}(A_j\ N^2) = \int_q^r \frac{j\ \theta(x)}{v(x)^2} dx, \quad j=1,2,3. \tag{25}$$

Now the asymptotically optimal partitioning function $\tilde{v}(x)$ can be obtained by minimizing (25) under the constraint (19). This is a problem of the Calculus of Variations. $\tilde{v}(x)$ has to be found in such a way that

$$\int_q^r \frac{j\ \theta(x)}{v(x)^2}\ dx + \lambda \left[\int_q^r v(x)dx - 1\right] \tag{26}$$

becomes stationary for $v(x) = \tilde{v}(x)$. λ is the Lagrange multiplier. The Euler equation is

$$\frac{\partial K}{\partial \theta} - \frac{d}{dx}\frac{\partial K}{\partial \theta_x} = 0,$$

where

$$K = \frac{j\ \theta(x)}{v(x)^2} + \lambda\, v(x).$$

As K does not depend on the derivative v_x it follows that

$$\tilde{v}(x) = C\ \theta(x)^{\frac{1}{3}} = \frac{\theta(x)^{\frac{1}{3}}}{\int_q^r \theta(x)^{\frac{1}{3}}\ dx}\,. \tag{27}$$

This proves (21). The value of the constant C has been obtained from (19). As j, $j=1,2,3$, is only a constant factor of the function to be minimized $\tilde{v}(x)$ is the asymptotically optimal

partitioning function for all j. It follows that the asymptotically optimal partition is the same for A_1, A_2, A_3, (i.e. the same for minimizing α and N_f). Inserting (27) into (25) we obtain the asymptotic area-formula (22). Because of (18) $\tilde{v}(x)$ is a regular partitioning function.

5. OTHER VARIANTS OF TDR

The basic idea of TDR is that hat-functions are constructed as tangents of the transformed density in the points p_i and then transformed back into the original scale by T^{-1}. So we have

$$h_i(x) = T^{-1}(T(f(p_i)) + T(f(p_i))'(x - p_i)) \text{ for } \pi_{i-1} \leq x \leq \pi_i$$

but different variants for the choice of π_i, $i = 0, ..., N-1$, are possible. In Section 3 we have derived the series expansion of the area between hat and density for $\pi_i = \mu(p_i, p_{i+1})$, i.e. the intersection points of the corresponding tangents, see Eqn. (4). We call this the *original variant* of TDR. It is in a sense the best variant as it is the only of the variants listed here that leads always to a continuous hat and a continuous squeeze and thus to the minimal possible areas A_1 and A_3 for given design points.

A nice variant is the *midpoint method* (see Figure 3). Here the intersection points are replaced by the midpoints of the sub-intervals:

$$\pi_i = \frac{p_i + p_{i+1}}{2}. \tag{28}$$

Although the hat is not continuous at π_i this method has interesting properties:

- Its asymptotic behaviour is identical with that of the original variant. Even the 4^{th} derivatives of the $a_j(p, \Delta)$, see Eqn. (12), are identical. Therefore the computed design points are the same as for the original variant of TDR.
- As no intersection points of tangents are to be calculated the set-up is faster.
- Applied to finite N there is only a very slight increase in α or N_f, respectively.

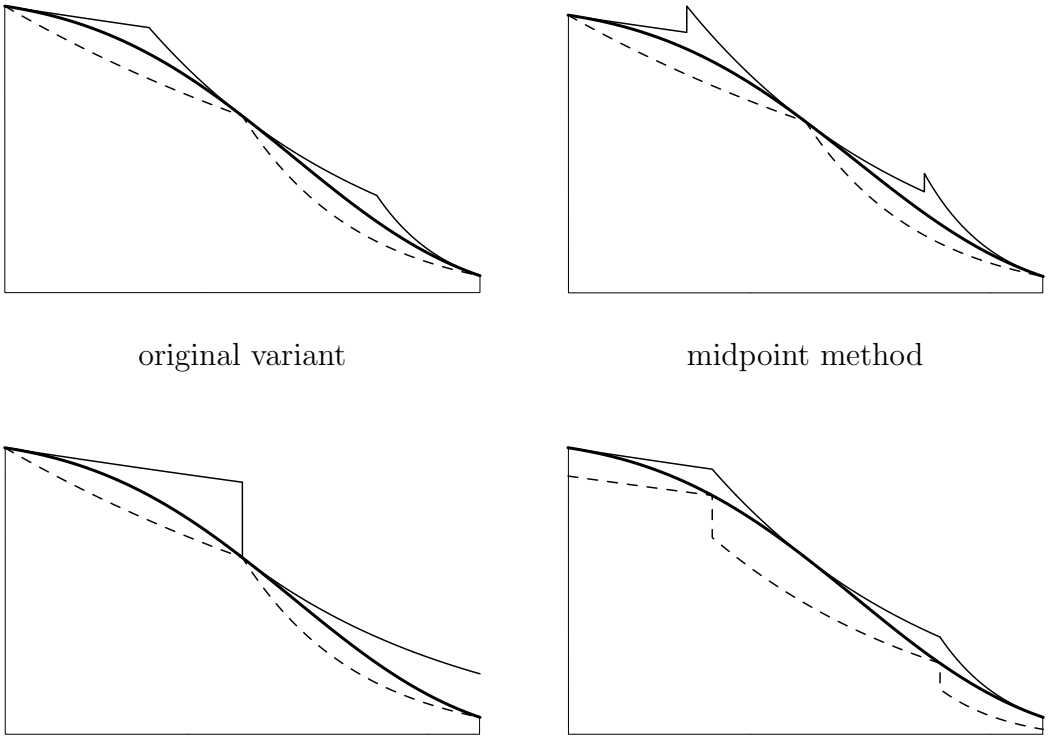

Figure 3: Hat (thin line) and squeeze (dashed line) of the different variants

Slightly different is the situation for the variant of TDR suggested by Evans and Swartz (3). To obtain a simpler sampling algorithm they suggest to take $\pi_i = p_i$ and $\pi_{i+1} = p_{i+1}$. This hat is not only not continuous, the fit is worse as we have on average a longer distance to the design point used for constructing the hat. Setting $\mu(p_i, p_{i+1})$, cf. (4) and (28), equal to p_{i+1}, then proceeding according to Section 3, with p_i as point of contact, one can derive that the area between hat and density for a single cell has the following MacLaurin series

$$a_1(p, \Delta) = -\frac{T(f(p))''}{6\,T'(f(p))}\,\Delta^3 \;+\; O(\Delta^4)$$

and thus is four times larger than the corresponding area of the original variant of TDR. This does not change the asymptotically optimal choice of the design points; only the estimated area between hat and density (a_1) is four times larger than the one given in the theorem. This implies that this version requires twice the number of design points to obtain the same area A_1. This does not mean that the algorithm in (3) is slower as that choice of the hat function leads to a simplification of the sampling.

A different simple and very fast implementation of TDR is possible for *proportional squeezes* as suggested in (10). There the hat is constructed as explained above with π_i as the intersection points of the tangents. The squeeze is no longer defined for the sub-intervals (p_i, p_{i+1}) but for the same interval (π_{i-1}, π_i) as the hat-function; and it is defined such that it is proportional to the hat in that interval. For this variant the hat is exactly the same as for standard TDR but the squeeze is slightly changed. Using the same techniques as in Section 3 it is possible to calculate the MacLaurin series for a_2, the area between density and squeeze, which is the same as before. So our theorem is also fully valid for this variant of TDR. This is especially important in practice, as the heuristic methods for finding good (but not optimal) design points like for example adaptive rejection sampling (see (4)) do not work for this variant.

6. PIECEWISE CONSTANT HATS AND SQUEEZES

Theorem 1 is based on the fact that the lowest order term of the series expansion of the area between hat and squeeze is $k = 3$ for TDR. It is not difficult to formulate and prove

Theorem 1 for general k,

$$a_j(p,\ \Delta) = \gamma_{kj}\ \theta(p)\ \Delta^k + O(\Delta^{k+1}), \quad j = 1, 2, 3, \quad \gamma_{k1} = 1,\ \gamma_{k2} = k-1,\ \gamma_{k3} = k, \tag{29}$$

cf. eqn.(14). For TDR we need the case $k = 3$ with $\gamma_{3j} = j$. The case $k = 2$ is also of great interest. Using $k = 2,\ \gamma_{21} = \gamma_{22} = 1, \gamma_{23} = 2$, we can compute asymptotically optimal design points for the rejection method for piecewise constant hats and squeezes that is discussed for example in (1). It can be easily verified that, analogously to (12), the MacLaurin series for $a_1,\ a_2,\ , a_3$, are

$$a_j(p, \Delta) = \frac{\gamma_{2j}}{2}|f'(p)|\Delta^2 + O(\Delta^3), \qquad j = 1, 2, 3, \qquad \gamma_{2j} = 1, 1, 2.$$

Thus, in analogy to (13), the coefficient in the leading term of the MacLaurin series (29) is given by

$$\theta(p) = \frac{1}{2}\ |f'(p)|.$$

The details of the further development are in close analogy to the TDR case. We therefore only state the main result:

For a three times differentiable density f, $p_0 = q$ and $p_N = r$ fixed, we consider a rejection method that uses $\max_{p_i \le x \le p_{i+1}} f(x)$ as hat and $\min_{p_i \le x \le p_{i+1}} f(x)$ as squeeze in the N subintervals with $i = 0, 1, \ldots, N-1$. For $N \to \infty$ the uniquely determined regular partitioning function $\tilde{v}(x)$, which leads to the minimal A_j for $j = 1, 2, 3$ among all regular partitioning functions is:

$$\tilde{v}(x) = C\ \theta(x)^{\frac{1}{2}}, \quad \text{where} \quad C = \frac{1}{\int_q^r \theta(x)^{\frac{1}{2}} dx}. \tag{30}$$

For the asymptotically optimal partition we obtain the area-formulas.

$$\lim_{N\to\infty} N\tilde{A}_j = \frac{\gamma_{2j}}{2}\left(\int_q^r \theta(x)^{\frac{1}{2}} dx\right)^2, \qquad j = 1, 2, 3, \qquad \gamma_{2j} = 1, 1, 2. \tag{31}$$

Eqns. (30) and (31) are the analogues of (21) and (22).

7. CALCULATING THE POINTS

We return to TDR: How can we use our theorem to obtain nearly optimal design points in practice? In the case of an infinite domain we cannot use $p_0 = q$ and $p_N = r$ as they must remain finite. And even in the case of bounded domain it happens very often that the optimal choice implies $p_0 > q$ and/or $p_N < r$. So if we use our asymptotic theorem to find approximately optimal design points we have to vary p_0 and p_N. Outside of the interval $[p_0, p_N]$ there are no squeezes or, as one could say, squeezes which formally are equal to zero. For given N, we therefore have to minimize, at least approximately, either, if the rejection constant α is the objective, the asymptotic approximation $\tilde{H}(p_0, p_N)$ of the area below the hat,

$$\tilde{H}(p_0, p_N) = \int_{x \le p_0} h(x)dx + \int_{p_0}^{p_N} f(x)dx + \frac{1}{N^2}\left(\int_{p_0}^{p_N} \theta(x)^{\frac{1}{3}}dx\right)^3 + \int_{x \ge p_N} h(x)dx \qquad (32)$$

or, if the expected number N_f of calls to $f(x)$ is the criterion, the asymptotic approximation $\tilde{S}(p_0, p_N)$ of the area between hat and squeeze,

$$\tilde{S}(p_0, p_N) = \int_{x \le p_0} h(x)dx + \frac{3}{N^2}\left(\int_{p_0}^{p_N} \theta(x)^{\frac{1}{3}}dx\right)^3 + \int_{x \ge p_N} h(x)dx. \qquad (33)$$

On the right hand side of Eqn. (32) the area below the hat over the interval $[p_0, p_N]$ is decomposed into the area below the density and the area between hat and density, for which the asymptotic approximation (22), $j=1$, is used, which is an excellent approximation even for small N. With a symbolic computation package like e.g. Mathematica or Maple it is not difficult to evaluate $\tilde{H}(p_0, p_N)$ or $\tilde{S}(p_0, p_N)$ for given values of p_0, p_N. So the selection of p_0 and p_N is reduced to a minimization problem with two variables that can be easily solved using a numerical search procedure available in symbolic computation packages as well. The other design points can be then be calcuated easily using Theorem 1.

7.1 The Algorithm Using Numerical Approximations

For code in a general purpose programming language it is helpful to observe that it is not necessary to calculate the design points with great accuracy as the objective function (the total area between hat and squeeze) is very flat close to the minimum. It is therefore possible

to use rough numerical approximations for the derivatives necessary to evaluate $\theta(x)$; and it is enough to compute these approximations on a rough one-dimensional grid. In erecting this grid one starts from a point m which is the mode or a point near to the mode. Depending on the behaviour of the quasi-density the stepsize cannot be kept constant in most cases. This will be discussed below. What we need at the beginning are scale units. Taking into account the possibility of skew or even very skew distributions the left scale unit u_l, used at the left hand side of the mode, and the right scale unit u_r must have the freedom to be different. Experience has shown that the distances of the right and left design points to the mode in the three-point case (5) with $c = -1/2$ are suitable scale units in all cases. These distances u_l, u_r fulfill the equations

$$f(m + u_l) = f(m + u_r) = \phi\, f(m), \quad \phi = 1/4, \quad u_l < 0,\ u_r > 0.$$

In order to accelerate the algorithm we do not solve these equations exactly. For the factor ϕ it is enough to be in the interval $0.15 < \phi < 0.4$. (Note that the geometric mean of the borders is approximately 1/4.) For partly correcting these crude approximations u_l and u_r are refined according to

$$u_l \leftarrow u_l \sqrt{4\, f(m + u_l)/f(m)}, \quad u_r \leftarrow u_r \sqrt{4\, f(m + u_r)/f(m)}.$$

As initial stepsizes the values $k\, u_l$ and $k\, u_r$ are used with k between 0.02 and 0.15. (For the normal distribution, e.g., $k = 0.15$ leads to about 40 grid points.)

During the construction of the grid we evaluate the function $\theta(x)$, see eqn.(13), at every grid point κ_i. For this purpose we have included a self-written routine for symbolic differentiation. Alternatively, estimates of $T(f(\kappa_i))''$, obtained as second derivatives of a quadratic interpolation of $T(f(x))$ based on $x = \kappa_{i-1}, \kappa_i, \kappa_{i+1}$, can be used. (Our numerical experiments with many distributions indicate that it is sufficient to work with an average $T(f(x))''$ in every interval. The final results when using this crude numerical approximation and symbolic differentation were always very close.) For $\tilde{v}(x)$, see (27), we construct the approximation $\hat{v}(x)$ consisting of the linear interpolations between every pair $\tilde{v}(\kappa_i)$, $\tilde{v}(\kappa_{i+1})$. In (20) we use

$\hat{v}(x)$ instead of $v(x)$ and solve for $p_i,\ i = 1, N-1$. As $\hat{v}(x)$ is a piecewise linear function this is quite easy.

We consider the functions $\tilde{H}$ and $\tilde{S}$ of (32), (33) being dependent on the current leftmost grid point κ_{i_l} and the rightmost gridpoint κ_{i_r}, and calculate approximations $\hat{H}(\kappa_{i_l}, \kappa_{i_r})$ and $\hat{S}(\kappa_{i_l}, \kappa_{i_r})$ for $\tilde{H}(\kappa_{i_l}, \kappa_{i_r})$ and $\tilde{S}(\kappa_{i_l}, \kappa_{i_r})$. The integral over $\theta(x)^{\frac{1}{3}}$ in (32) and (33) is approximated by the following generalized trapezoidal rule also applicable to not equidistant grid points:

$$\int_{\kappa_{i_l}}^{\kappa_{i_r}} \theta(x)^{\frac{1}{3}} dx \ \approx\ \frac{1}{2} \sum_{i=i_l}^{i_r-1} (\theta(\kappa_i)^{\frac{1}{3}} + \theta(\kappa_{i+1})^{\frac{1}{3}})(\kappa_{i+1} - \kappa_i). \tag{34}$$

After every extension of the grid by one point the value of the integral is updated (one trapezoidal area more). Synchronously to the integral (34) the loss function

$$\hat{H}(\kappa_{i_l}, \kappa_{i_r}) = \int_{x \le \kappa_{i_l}} h(x)dx + \int_{\kappa_{i_l}}^{\kappa_{i_r}} f(x)dx + \frac{1}{N^2} \left(\int_{\kappa_{i_l}}^{\kappa_{i_r}} \theta(x)^{\frac{1}{3}} dx \right)^3 + \int_{x \ge \kappa_{i_r}} h(x)dx \tag{35}$$

or, respectively,

$$\hat{S}(\kappa_{i_l}, \kappa_{i_r}) = \int_{x \le \kappa_{i_l}} h(x)dx + \frac{3}{N^2} \left(\int_{\kappa_{i_l}}^{\kappa_{i_r}} \theta(x)^{\frac{1}{3}} dx \right)^3 + \int_{x \ge \kappa_{i_r}} h(x)dx \tag{36}$$

is evaluated. Compared with (36) the integral over f in (35) causes nearly no additional costs. As the values of the density at the grid points are to be calculated anyway, only one additional multiplication and some additions are necessary per interval. If the function values of three successive points at one side are not monotone anymore an approximation of p_0 or p_N, respectively, is obtained by minimzing the quadratic interpolation based on these three points. (For N not too small the left and right side do not influence each other.)

Starting from m every step is made on that side of the mode where the decrease in the loss function is greater. If after a certain number - we use $\lfloor 20/k \rfloor$ - of steps at one side the minimum is not yet met then for each further step the new stepsize will be obtained by multiplying the previous one by a number slightly greater than one (1.01 in our programs). There are cases (e.g. the Cauchy distribution, $c = -1/2,\ N > 3$, minimizing N_f) where at

the minimum solution p_0 and/or p_N are infinite (asymptotes instead of tangents). Therefore, in such cases the procedure has to be terminated at a certain stage. We stop it if $\kappa_{i_l} < m + 1000\, u_l$ or $\kappa_{i_r} > m + 1000\, u_r$. Surely, also in this case one will obtain an excellent approximation. In the case of a bounded domain it may happen that a newly calculated grid point falls out of the domain. If $f(q) > 0$ and $f'(q) < \infty$ then the procedure will be stopped. Otherwise, after rejecting the point outside, every new grid point is obtained by bisecting the line segment from q to the former grid point, analogously for the right bound r. The derivatives $f'(q)$ or $f'(r)$ are considered to be infinite if for an x near to q or near to r

$$f'(x) \;>\; 1000\,\frac{f(u_l)}{-u_l} \quad \text{or} \quad -f'(x) \;>\; 1000\,\frac{f(u_r)}{u_r},$$

holds. (Here $f(u_l)$, $f(u_r)$ are used as scale units in the $y-$direction.)

Note that the costs of the main part of our asymptotic point finding method, namely the calculation of p_0 and p_N, are nearly independent of N. Only the range of integration according to (35) or (36) is slightly larger for greater N. After this main part the design points $p_2, \ldots, p_{N-1}$ are obtained by the evaluation of the inverse of a piecewise linear function at $N-2$ arguments. The corresponding algorithm is very short, the required time, approximately proportional to N, is negligible.

7.2 Computational Experience

Asymptotic theory is interesting from a mathematical point of view. To demonstrate that it is also practically useful we include a small empirical study for $N = 31$ and $N = 9$. Table 1 shows the resulting α and N_f for the Normal and Gamma(3/2) distributions, the Makeham distribution, see e.g. Evans and Swartz (3), with parameters $a = 0.01$, $b = 0.02$, $c = e$, the distribution of the 29^{th} order statistic of a sample of 97 independent normal variates, the distribution of the 69^{th} order statistic of a sample of 97 independent Cauchy variates, the hyperbolic distribution with density proportional to $\exp(-\sqrt{1+x^2})$ and the exponential power distribution with density proportional to $\exp(-x^4)$, all with transformation $T(x) = -x^{-1/2}$. We used our Fortran implementation of the asymptotic point finding algorithm described in Section 6 and compared the results with the optimal results that we obtained

using a relatively sophisticated gradient method written for this purpose. (This method, however, is too slow for practical use.) The results clearly show that the asymptotic method works excellent both for $N = 31$ and for $N = 9$ for the first six distributions. We obtained similar results for all of the many other standard distributions we have tried.

The asymptotic results are not as close to optimal (but still very good) for the exponential power distribution with density proportional to $\exp(-x^4)$. This is not too astonishing as due to $T(f(0))'' = 0$ this density is not fulfilling (18) and thus our asymptotic theorem is not applicable. Note that the fact that $T'(f(0)) = \infty$ for the Gamma distribution is no problem for our algorithm as the asymptotic theorem is only applied to the interval $(\kappa_{i_l}, \kappa_{i_r})$ that never contains zero for this distribution.

Also the condition of five time differentiable densities seems to be no problem in practice. Most densities of interest can be differentiated arbritarily often anyway. We have also constructed examples of densities being only differentiable twice in up to 10^5 points per unit interval. The calculation of the (nearly) optimal design points of these densities with our algorithm can still be carried out without problems. We were not able to construct examples where our algorithm had problems with two times differentiable densities.

We also observed in our experiments that we get practically the same results when minimizing α and when minimizing N_f.

8. CONCLUSIONS

We have proven an asymptotic theorem for the choice of optimal design points in rejection algorithms (asymptotic in the sense that the number of design points tends to infinity). This theorem solves an open optimization problem for the design of automatic random variate generation algorithms. We have also demonstrated that the asymptotic result is valid for all variants of transformed density rejection and leads to excellent results for many standard distributions, even for small N

Table 1: The asymptotic approximations for α and N_f, compared with the optimal values calculated by a gradient method and with the simple heuristic of equiangular points (8). For the asymptotic algorithm $c = -1/2$, the stepsize $k = 1/12$ and symbolic differentiation were used.

Method	N	Distribution						
		Normal $N(0,1)$	Gamma $t = 3/2$	Makeham $a = 0.01$, $b = 0.02$, $c = e$	29^{th} order statistic of 97 normal variates	69^{th} order statistic of 97 Cauchy variates	Hyper-bolic $e^{-\sqrt{1+x^2}}$	Expon-ential power e^{-x^4}
α asymp.	9	1.033978	1.019890	1.018040	1.033986	1.034037	1.035766	1.023752
α equiang.	9	1.065618	1.044879	1.047892	1.065477	1.062273	1.041767	1.054558
α optimal	9	1.033955	1.019870	1.018028	1.033963	1.034012	1.035740	1.023396
N_f asymp.	9	0.091348	0.061229	0.056335	0.091377	0.091792	0.096985	0.071487
N_f equiang.	9	0.177451	0.163139	0.254090	0.177018	0.161508	0.112061	0.148874
N_f optimal	9	0.091340	0.061186	0.056334	0.091369	0.091790	0.096984	0.070753
α asymp.	31	1.002946	1.001916	1.001519	1.002947	1.002970	1.003163	1.002158
α equiang.	31	1.006800	1.006694	1.003837	1.006783	1.006226	1.004402	1.005369
α optimal	31	1.002946	1.001914	1.001518	1.002947	1.002970	1.003163	1.002144
N_f asymp.	31	0.008598	0.005815	0.004617	0.008601	0.008678	0.009250	0.006508
N_f equiang.	31	0.019944	0.029094	0.055419	0.019900	0.018329	0.013025	0.015895
N_f optimal	31	0.008597	0.005809	0.004616	0.008601	0.008677	0.009250	0.006478
T_c-concave for $c \leq$		0	0	-0.488898	0	-0.033333	0	0

BIBLIOGRAPHY

(1) Ahrens, J. H. A one-table method for sampling from continuous and discrete distributions. Computing **1995**, *54*, 127–146.

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

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

(4) Gilks, W. R.; Wild, P. Adaptive rejection sampling for Gibbs sampling. Applied Statistics, **1992**, *41*, 337–348.

(5) Hörmann, W. A note on the quality of random variates generated by the ratio of uniforms method. ACM Transactions on Modeling and Computer Simulation, **1994**, *4*, 96–106.

(6) Hörmann, W. A rejection technique for sampling from T-concave distributions. ACM Transactions on Mathematical Software **1995**, *21*, 182–193.

(7) Hörmann, W.; Derflinger, G. Rejection-inversion to generate variates from monotone discrete distributions. ACM Transactions on Modeling and Computer Simulation, **1996**, *6*, 169–184.

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

(9) Leydold, J. Automatic sampling with the ratio-of-uniforms method. ACM Transactions on Mathematical Software, **2000**, *26*, 78–98.

(10) Leydold, J; Janka, E.; Hörmann, W. Variants of transformed density rejection and correlation induction, *Monte Carlo and Quasi-Monte Carlo Methods 2000 (Hong Kong)* (Fang, K.-T.; Hickernell, F. J.; Niederreiter, H. eds.), **2002**, Springer-Verlag, pp. 345–356.