Computer Methods for Generating Student-t Variates

E. Stadlober and U. Dieter, Graz

Abstract - Zusammenfassung

Computer Methods for Generating Student-t Variates. All previously published methods for generating t-variates are discussed and two new algorithms are developed. The first procedure is based on a modified acceptance-rejection technique using samples from the t_3-distribution. The second method requires nothing but a transformed normal deviate in the majority of cases. The two proposed algorithms apply to parameters $a > 3$ and $a \geq 1$ respectively. Comparisons are carried out in terms of generation and initialization times, expected numbers of uniform deviates and memory requirements.

AMS (MOS) subject classifactions (1980): Primary 65C10; Secondary 68J05

Key words and phrases: Random Numbers, Student-t distribution, simulation.

Computermethoden zur Erzeugung Student t-verteilter Zufallszahlen. Alle bisher publizierten Methoden zur Erzeugung t-verteilter Zufallsgrößen werden diskutiert und zwei neue Algorithmen entwikkelt. Das erste Verfahren basiert auf einer modifizierten Verwerfungsmethode, wobei Stichproben aus der t_3-Verteilung verwendet werden. Bei der zweiten Methode ist in den meisten Fällen nichts anderes als eine normalverteilte Zufallszahl zu transformieren. Die zwei vorgeschlagenen Algorithmen sind anwendbar für Parameter $a > 3$ bzw. $a \geq 1$. Vergleiche bezüglich der Ausführungs- und Initialisierungszeiten, der durchschnittlichen Anzahl gleichverteilter Zufallszahlen und des Speicherplatzbedarfs werden ebenfalls durchgeführt.

1. Introduction

Whenever theoretical derivations fail to give an exact answer to some statistical problem, useful information can still be obtained by means of simulation. On account of the importance of the Student-t distributions in many areas of statistics, such simulations frequently require random numbers which follow a t-distribution with a given number a of degrees of freedom.

If generators for normal and gamma deviates are available, one may use the following well-known fact about t-variates:

If Z is a standard normal deviate, and if Y is a standard gamma variate with parameter $a/2$, then $X = \sqrt{a}Z/\sqrt{2Y}$ is a t-distributed random variable with a degrees of freedom.

This theorem translates into the following procedure:

Algorithm TNG ($a > 0$)

1. Generate a standard normal deviate Z.
2. Generate a standard gamma variable Y with parameter $a/2$.
3. Return $X \leftarrow \sqrt{a}Z/\sqrt{2Y}$ as a sample from the t-distribution with a degrees of freedom.

However, the algorithm TNG has some disadvantages. Since the commonly used uniform pseudo-random number generators cannot produce truly independent successive samples, one should always aim at *small* expected numbers N of uniform deviates required for the production of one t-sample. But even a very efficient gamma sampling method such as GD in [3] ($N = 1.7016$ to $N = 2.6719$) combined with, say, algorithm FL5 in [2] ($N = 1.2316$) for the normal deviates in Step 1 will yield an overall N between 3 and 4.

Alternative methods utilize the explicit form of the Student-t probability density function:

$$t_a(x) = c_a\left(1+\frac{x^2}{a}\right)^{-\frac{a+1}{2}} \quad \text{for } -\infty < x < \infty,\ a \geq 1, \tag{1.1}$$

where

$$c_a = \frac{1}{\sqrt{a}\,B(a/2,1/2)} = \frac{\Gamma((a+1)/2)}{\sqrt{a\pi}\,\Gamma(a/2)}.$$

The t-family contains the Cauchy distribution for $a = 1$ and the normal distribution for $a \to \infty$ as extreme cases, for which many sampling methods are available; for the normal distribution see the survey papers of Ahrens/Dieter [1], and Kinderman/Ramage [11] where many procedures are discussed. For the Cauchy distribution see the article [17] of the first author and the recent paper of Ahrens/Dieter [4].

Furthermore, there are two special sampling methods for $a = 2$ and $a = 3$: If U denotes a $(0,1)$-uniform random variable, then

$$X \leftarrow \frac{U-1/2}{\sqrt{(U-U^2)/2}} \quad \text{samples from } t_2,$$

and values from the t_3-distribution can be generated efficiently by the ratio-of-uniforms method of Kinderman/Monahan [8]. The t_3-method is one step in the acceptance-rejection procedure (for all $a > 3$) of Best [6], and it will also be incorporated into our modified acceptance-rejection algorithm in Section 3.

A survey of published material, consisting of algorithms based on acceptance-rejection, probability mixing and ratio-of-uniforms, is given in Section 2. Section 3 deals with our suggested method based on a modified acceptance-rejection technique. Details of an exact-approximation method are explained in Section 4. Section 5 contains comparisons of the considered algorithms with respect to various performance characteristics. Some recommendations are stated which depend on the intended applications. Proofs of squeeze inequalities for algorithm TMA in Section 3 are given in the Appendix.

2. Existing Methods

2.1. Acceptance-Rejection Methods. They are based on the following refinement of J. v. Neumann's classical acceptance-rejection method [15]:

Let $f(x)$ be the density from which we want to sample and let some hat function $h(x)$ be selected for which $f(x) \le h(x)$ holds for all x. Assume that the integral of $h(x)$ is a finite number α. Hence $g(x) = h(x)/\alpha$ is a density function. The idea is to sample first of all from $g(x)$ and to accept each sample X with probability $f(X)/h(X)$. Since this quotient is often difficult to compute, it will be assumed that easy-to-calculate bounds $b(x)$ and $B(x)$ — so called squeeze-functions — are known such that

$$b(x) \le \frac{f(x)}{h(x)} \le B(x)$$

or, equivalently,

$$l(x) = b(x)\,h(x) \le f(x) \le B(x)\,h(x) = L(x)$$

holds for all x. The second formula shows that it is often sufficient to find lower and upper bounds $l(x)$ and $L(x)$ for the density $f(x)$ itself. Then sampling from $f(x)$ may be carried out as follows.

Procedure AR (Acceptance-Rejection Method)

1. Generate a random sample X from the distribution with density $g(x)$.
2. Generate a $(0,1)$-uniform deviate U.
3. If $U \le b(X)$ $(= l(X)/h(X))$ deliver X. (Quick Acceptance)
4. If $U > B(X)$ $(= L(X)/h(X))$ go to 1. (Quick Rejection)
5. If $U \le f(X)/h(X)$ deliver X, otherwise go to 1.

Note that the average number of trials is equal to α , the *efficiency* of the algorithm.

Kinderman, Monahan and Ramage explain three algorithms TAR, TIR and TIRS in [10]. They used

$$h(x) = \begin{cases} 1 & \text{for } |x| < 1 \\ 1/x^2 & \text{for } |x| \ge 1 \end{cases} \tag{2.1}$$

since

$$u_a(x) = \frac{t_a(x)}{c_a} = \left(1 + \frac{x^2}{a}\right)^{-\frac{a+1}{2}} \le h(x)$$

holds for all x. It means that in the center ($|x| < 1$) a uniform deviate can be used whereas in the tail ($|x| \ge 1$) the reciprocal of a uniform deviate samples from $h(x)$. The efficiency α_a is equal to $4c_a$, hence $1.273 = \alpha_1 \le \alpha_a \le \alpha_\infty = 1.596$. In all three algorithms the lower bound

$$l(x) = 1 - \frac{1}{2}|x| \le u_a(x) \qquad \text{for } |x| \le 2 \tag{2.2}$$

is used resulting in a quick acceptance step when $|x| \le 2$.

In TAR there is no upper bound. In TIRS the inequalities

$$u_a(x) \le 2u_1(x)\,u_a(1) = L_a(x) \le h(x) \qquad \text{for } \sqrt{2u_a(1) - 1} \le |x| \le 2, \tag{2.3}$$

are applied, whereas in TIR the squeeze is constructed more loosely from

$$u_a(x) \le 2\,u_1(x)u_\infty(1) = L(x) \le h(x) \qquad \text{for } \sqrt{2e^{-1/2} - 1} \le |x| \le 2. \tag{2.4}$$

In TIRS every change of the parameter involves computations of $2u_a(1)$ to obtain the tighter upper bound $L_a(x)$ of $u_a(x)$ in (2.3). The set-up costs can be reduced evaluating $2u_a(1)$ by means of Chebychev-economized polynomials. For a complete description of the algorithms we refer to Kinderman et al. [10].

Best [5] proposes a sampling method for all $a > 3$ with an envelope proportional to the t_3-density

$$t_a(x) \le \alpha_a t_3(x), \qquad a > 3, \tag{2.5}$$

where

$$\alpha_a = \max_x \left\{\frac{t_a(x)}{t_3(x)}\right\} = \frac{8}{9}\sqrt{3}\,c_a\,u_a(1).$$

α_a is very close to one ($1 = \alpha_3 < \alpha_a \le \alpha_\infty = 1.170$), which indicates a good fit of the suggested hat. The calculation of the ratio

$$q_a(x) = \frac{t_a(x)}{\alpha_a t_3(x)} = \frac{9}{16}\left(1 + \frac{x^2}{3}\right)^2 \left(\frac{a+1}{a+x^2}\right)^{\frac{a+1}{2}} \tag{2.6}$$

is involved, but the following inequalities help to avoid its calculation in most cases:

$$\begin{aligned} .8601 - .1163x^2 &= \left(\frac{9}{16}\right)^2 e\,(1 - .1352x^2) \\ &\le \left(\frac{9}{16}\right)^2 \left(1 + \frac{x^2}{3}\right)^4 e^{1-x^2} = q_\infty^2(x) \le q_a^2(x). \end{aligned} \tag{2.7}$$

The first inequality in (2.7) was established numerically. The well-known inequality

$$\frac{y}{1+y} < \ln(1+y) < y \qquad \text{for } y > -1$$

leads to

$$\frac{(a+1)(x^2-1)}{a+x^2} < (a+1)\ln\left(\frac{a+x^2}{a+1}\right) < x^2 - 1 \qquad \text{where } y = \frac{x^2-1}{a+1}.$$

The right-hand side yields $q_a(x) \ge q_\infty(x)$, which is the second inequality in (2.7). The left-hand side results in the upper bound

$$q_a^2(x) \le \left(\frac{9}{16}\right)^2 \left(1 + \frac{x^2}{3}\right)^4 \exp\left(\frac{(a+1)(1-x^2)}{a+x^2}\right)$$

which is applied in Step 4 of the formal statement of Best's algorithm below.

Algorithm T3T ($a > 3$; Best [6], modified)

1. Generate U, V and set $V \leftarrow V - 1/2$. If $U^2 + V^2 \le U$ set $X \leftarrow \sqrt{3}\,V/U$, otherwise go to 1. (Generation of a t_3-deviate)
2. Generate U. If $U^2 \le .8601 - .1163X^2$ deliver X. (Immediate Acceptance)
3. Set $Z \leftarrow 2\ln((\sqrt{e}/16)(3+X^2)^2/U)$. If $Z \ge X^2$ deliver X. (Quick Acceptance)
4. Set $b \leftarrow a+1$, $XA \leftarrow X^2 + a$. If $a < 15$ go to 5.
   If $(Z-1)XA < b(X^2-1)$ go to 1. (Quick Rejection)
5. If $Z - 1 \ge b\ln(XA/b)$ deliver X, otherwise go to 1.

For $a > 2$ a simple but slow method was published by Marsaglia in [13], where the density of a transformed t-deviate is majorized by a normal one. Since the resulting algorithm proved inferior even to algorithm TNG, it will not be discussed in detail.

2.2. Mixing Methods. Suppose that the density $f(x)$ can be written as

$$f(x) = \sum_{i=1}^{k} p_i f_i(x), \qquad p_i \ge 0, \quad \sum_{i=1}^{k} p_i = 1,$$

where each $f_i(x)$ is a density, for which we have some method to sample. Then sampling from $f(x)$ can be carried out in two steps: First, an integer I is selected with probability p_I; subsequently, one generates a variate X whose density is $f_I(x)$.

In their algorithms TMX and TMXS, which are valid for $a \ge 1$, Kinderman, Monahan and Ramage [10] choose

$$p_1 = 2\,c_a\,, \qquad f_1(x) = \begin{cases} \dfrac{1-|x|/2}{2} & \text{for } |x| < 2 \\ 0 & \text{otherwise;} \end{cases}$$

$$p_2 = 1 - p_1\,, \qquad f_2(x) = \frac{c_a}{p_2}\begin{cases} u_a(x) - 1 + |x|/2 & \text{for } |x| < 2 \\ u_a(x) & \text{otherwise.} \end{cases}$$

Deviates from $f_1(x)$ can be generated as $X \leftarrow 2(U_1 + U_2 - 1)$, with independent $(0,1)$-uniform variates U_i. For sampling from $f_2(x)$ one applies the acceptance-rejection method based on the numerically proved inequalities

$$u_a(x) - 1 + \frac{|x|}{2} \le \begin{cases} .13528 & \text{for } |x| \le 1.7922 \\ .2 & \text{for } 1.7922 \le |x| < 2 \end{cases}$$

and the obvious relationship $u_a(x) \le 1/x^2$ for $|x| > 2$.

It follows from

$$.6366 = \frac{2}{\pi} = 2c_1 \le p_1 \le 2c_\infty = \sqrt{\frac{2}{\pi}} = .7979,$$

that the exact value of p_1 has to be calculated only in 16.13% of all cases: with an initial $(0,1)$-uniform deviate U one samples from $f_1(x)$ if $U \le .6366$, from $f_2(x)$ if $U \ge .7979$, but only if $.6366 < U < .7979$ p_1 has to be known precisely to make the decision.

In our implementation stated in [17] the algorithm TMX includes the calculation of $2c_a$ and $2u_a(1)$, whereas the version TMXS determines these quantities only initially and after a change of the parameter a. For speed $2c_a$ and $2u_a(1)$ are approximated by economized polynomials in a.

2.3. Ratio-of-Uniforms Method. Kinderman/Monahan [8] invented a variant of the acceptance-rejection procedure which is called *ratio-of-uniforms* method.

Let $h(x) = kf(x)$, $k > 0$, $f(x)$ the desired density and let C be the two-dimensional set $C = \{(u,v) \mid 0 \le u \le \sqrt{h(v/u)}\}$. If (U,V) is uniformly distributed over C, then $X = U/V$ has the density $f(x)$.

For applying this method to the t_a-density, set $h(x) = u_a(x)$. C is contained in the rectangle $R = \{(u,v) \mid 0 \le u \le 1,\ -v_M \le v \le v_M\}$, where the constants v_M are given by $v_M = \sqrt{2a}(a+1)^{-(a+1)/4}(a-1)^{(a-1)/4}$ if $a > 1$ and $v_M = 1$ if $a = 1$. Now generate a point (U,V) uniformly in R and accept if it is inside C. The expected number of trials needed is equal to $4c_a v_M$ which increases from 1.273 at $a = 1$ to 1.369 as a tends to ∞.

The acceptance step $u^2 \le h(x) = (1 + x^2/a)^{-(a+1)/2}$ implies

$$u^{-\frac{4}{a+1}} \ge 1 + \frac{x^2}{a} \quad \text{or} \quad x^2 \le a(u^{-\frac{4}{a+1}} - 1).$$

Numerical considerations lead to the inequalities

$$5 - cu \le a(u^{-\frac{4}{a+1}} - 1) \le \frac{16}{cu} - 3, \qquad \text{where} \quad c = 4\left(1 + \frac{1}{a}\right)^{\frac{a+1}{4}}, \tag{2.8}$$

which yield fast acceptance-rejection decisions in most cases. It should be mentioned that the left hand inequality in (2.8) is sharp, whereas the right hand inequality is not optimal and only valid if $a \ge 3$.

In their first algorithm TROUO Kinderman/Monahan use solely the lower bound in (2.8). In TROU, which works only for $a \ge 3$, both bounds are utilized. A combination of these two algorithms reads as follows.

Algorithm TRU ($a \ge 1$; Kinderman/Monahan [9])

0. Pre-set $a' \leftarrow 0$ at compilation time.
1. [Set-up] If $a \ne a'$ set $a' \leftarrow a$, $r \leftarrow 1/a$, $p \leftarrow 1/(1+r)$, $q \leftarrow -(a+1)/4$, $c \leftarrow 4p^q$, $e \leftarrow 16/c$. If $a = 1$ set $v_M \leftarrow 1$, otherwise set $v_M \leftarrow \sqrt{2p}((1-r)p)^{(a-1)/4}$.
2. Generate U, V and set $V \leftarrow v_M(V + V - 1)$, $X \leftarrow V/U$.
3. If $cU \le 5 - X^2$ deliver X. (Immediate Acceptance)
4. If $a < 3$ go to 6.
5. If $(3 + X^2)U > e$ go to 1. (Quick Rejection)
6. If $U \le (1 + rX^2)^q$ deliver X, otherwise go to 1.

For each choice of the parameter a the algorithm needs some inital set-up time. However, for fixed parameters a it is one of the fastest procedures — see Table 5.

3. Modified Acceptance-Rejection Method

3.1. The Method. The modified acceptance-rejection method has been described by Kronmal/Peterson [12] and it was also used by Ahrens/Dieter [3] for sampling from gamma distributions:

Let $f(x)$ be the density of the target distribution, $g(x)$ be another density close to $f(x)$ and let $I = \{x|g(x) \leq f(x)\}$. Sampling from $f(x)$ may be carried out in the following way:

Procedure MA (Modified Acceptance-Rejection Method)

(I) 1. Generate a random sample X from the distribution with density $g(x)$.
2. If $X \in I$ deliver X. (Immediate Acceptance)

(Q) 3. Generate a $(0,1)$-uniform deviate U.
4. If $U \leq f(X)/g(X)$ deliver X. (Quotient Acceptance)

(D) 5. Generate a *new* sample X from the distribution with density proportional to $f(x) - g(x)$, $x \in I$, and deliver X. (Difference Acceptance)

For applying the modified acceptance-rejection method one has to find a density $g(x)$ which is close to the target density $f(x)$ such that the set I is of a simple form, e.g. $I = \{x|x \leq x_0\}$. In the case of the t_a-distribution the normal density $g(x) = \phi(x)$, would be a possible choice. However, the corresponding set $I = \{x|\phi(x) \leq t_a(x)\}$ is so small that the probability of case (I) is between 6.4% and 12%.

After some experimentation the following selection proved more appropriate: $f(x)$ is equal to a transformed t_a-density and $g(x)$ is equal to the t_3-density. More precisley,

$$f(x) = s\, t_a(sx), \qquad \text{where} \quad s = \sqrt{\frac{8}{3\pi}} + \frac{3}{a}\left(1 - \sqrt{\frac{8}{3\pi}}\right), \quad a > 3, \tag{3.1}$$

$$g(x) = t_3(x)\,. \tag{3.2}$$

The transformation constant s in (3.1) was chosen such that $f(x) \equiv g(x)$ for the lower limit $a = 3$, $f(0) = g(0)$ as a tends to the upper limit ∞, and $f(0) > g(0)$ for $3 < a < \infty$. Also, $f(x)$ and $g(x)$ intersect at some points $\pm w_a$ and the set $I_a = \{-w_a \leq x \leq w_a|g(x) \leq f(x)\}$ contains $I = \{-w_\infty \leq x \leq w_\infty\}$, since $w = w_\infty < w_a$ for all a (see [17]).

In Step 4 of MA the evaluation of the ratio $r(x) = f(x)/g(x)$ is required, but in most cases it will be avoided through the application of the squeeze inequalities

$$b(z) = 1 - \frac{1}{\beta}z \leq r(x) \leq 1.0184 - \frac{1}{\gamma}z + \frac{1}{\delta}z^2 = B(z), \quad \text{for } w \leq |x|, \tag{3.3}$$

where

$$z = x^2 - w^2, \quad \beta = 6.845 + \frac{42.8}{a-3}, \quad \gamma = 7.13 + \frac{40.9}{a-3}, \quad \delta = 201.3 + \frac{2207.3}{a-3}.$$

The proof of (3.3) will be given in the Appendix.

There are many possibilities for sampling from the density proportional to $f(x) - g(x)$ in Step 5 of MA. In algorithm TD of [16] the first author used the acceptance-rejection method with a double-exponential hat function

$$h(x) = \frac{c}{2b}\exp\left(-\frac{|x-m|}{b}\right) \geq f(x) - g(x) \qquad \text{for } x \geq 0.$$

In this paper we employ the simpler triangular envelope

$$h(x) = \frac{c}{b^2}(b - |x-m|) \geq f(x) - g(x) \qquad \text{for } 0 \leq x \leq m + b. \tag{3.4}$$

The constants c, b and m will be determined later on. With this the method can be outlined as follows.

Procedure PS (Preliminary Sampling Procedure)

(I) 1. Take a sample X from the t_3-density $g(x)$ and set $T \leftarrow sX$.
2. If $|X| \leq w$ deliver T. (Immediate Acceptance)

(S) 3. Generate U and set $Z \leftarrow X^2 - w^2$.
4. If $U \leq b(Z)$ deliver T. (Squeeze Acceptance)
5. If $U > B(Z)$ go to 7. (Squeeze Rejection)

(Q) 6. If $\ln U \leq q(X)$ $(= \ln r(X))$ deliver T. (Quotient Acceptance)

(D) 7. Generate a new sample X from the density proportional to $f(x) - g(x)$ in $(0, w_a)$ by the acceptance-rejection method with the triangular envelope $h(x)$ and deliver $T \leftarrow \pm sX$. (Difference Acceptance)

In 86% of the cases T is accepted immediately (Step 2). The probability of squeeze acceptance in Step 4 decreases from 13.6% for $a = 3.1$ to 6.1% as $a \to \infty$. In at most .7% of the time the rarest case (Q) leads to a success in Step 6. The probability that finally a sample from the difference in Step 7 is needed, increases from .2% for $a = 3.1$ to 7.4% as $a \to \infty$.

The method will be complete once the acceptance-rejection procedure in (D) has been explained in detail. First of all the constants c, b and m of the triangular hat function $h(x)$ in (3.4) have to be fixed. They are determined in such a way that the area below $h(x)$ is as small as possible. For optimal c, b and m the function $h(x)$ will touch $d(x) = f(x) - g(x)$ at three points. In Figure 1 the cases $a = 10$ and $a = 20$ are displayed.

(i) If $3 < a \leq 12.3585$, the optimal $h(x)$ touches $d(x)$ at L, R_1 and R_2, where $.70 < L < .73$, $.28 < R_1 < 1.48$ and $2.03 < R_2 = w_a < 2.17$.

(ii) If $12.3585 < a$, the optimal $h(x)$ contacts $d(x)$ at L_1, L_2 and R, where $L_1 = 0$, $.70 < L_2 < .72$ and $1.40 < R < 1.44$.

The optimal m, b and c were calculated by elementary geometric methods: In the case (i), where the triangular envelope of smallest area touches $d(x)$ at three points $L < m < R_1 < R_2 = w_a$, L and R_1 are determined by

$$d'(L) = -d'(R_1), \qquad d'(R_1) = \frac{d(R_1)}{R_1 - w_a}. \tag{3.5}$$

Thereafter m, b and c are obtained from

$$2m = w_a + L - \frac{d(L)}{d'(L)}, \qquad b = w_a - m, \qquad c = \frac{b^2 d(L)}{b + L - m}. \tag{3.6}$$

In the case (ii), $d(x)$ is covered by a triangular hat of minimum area touching at $0 = L_1 < L_2 < m < R$. Now L_2 and R fulfill the equations

$$d'(L_2) = \frac{d(L_2) - d(0)}{L_2}, \qquad d'(R) = -d'(L_2) \tag{3.7}$$

and m, b and c are calculated as

$$2m = R + \frac{d(R) - d(0)}{d'(L_2)}, \qquad b = m + \frac{d(0)}{d'(L_2)}, \qquad c = b^2 d'(L_2). \tag{3.8}$$

Second, a simple triangular squeeze function $l(x) \le d(x)$ is established which helps to avoid the laborious evaluation of $d(x)$ almost entirely. An isosceles triangle $l(x)$ is constructed such that it contacts $d(x)$ at its top at x_3 and at two other points $x_2 < x_3 < x_4$. Let $l(x)$ cross the x-axis at x_1 and x_5. Then x_2, x_3 and x_4 are determined by

$$x_2 = x_3 - \frac{d(x_3) - d(x_2)}{d'(x_2)}, \qquad x_3 = \frac{1}{2}(x_1 + x_5), \qquad x_4 = x_3 + \frac{d(x_3) - d(x_4)}{d'(x_2)}, \tag{3.9}$$

where x_1 and x_5 are obtained from

$$x_1 = x_2 - \frac{d(x_2)}{d'(x_2)}, \qquad x_5 = x_4 - \frac{d(x_4)}{d'(x_4)}. \tag{3.10}$$

The squeeze inequality with $b_l = x_3 - x_1$ finally reads

$$l(x) = \frac{d(x_3)}{b_l}(b_l - |x - x_3|) \le d(x) \qquad \text{for } x_1 \le x \le x_5 = 2x_3 - x_1. \tag{3.11}$$

This is also illustrated in Figure 1.

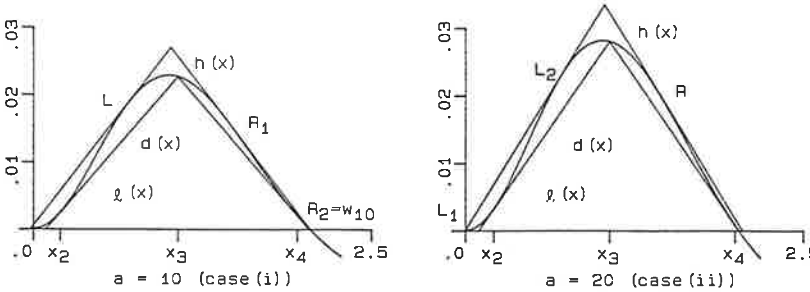

Figure 1: Squeezes $l(x)$, difference functions $d(x)$ and envelopes $h(x)$.

For the final algorithm the exact values of x_3, b_l, $c_l = \frac{\pi\sqrt{3}}{2}\, d(x_3)/b_l$ in (3.11) and the true optimal m, b and $c_h = \frac{\pi\sqrt{3}}{2}\, c$ in (3.3) are not used. Instead we worked out convenient approximations which are stated in Step 6 of algorithm TMA. Note that we equated x_3 with m. It was verified by extensive numerical checks that the inequality $l(x) < d(x) < h(x)$ remains correct.

In the left part of Table 1 the intersections w_a ($d(w_a) = 0$) and the exact values of x_3, b_l and c_l are listed. The right-hand side of Table 1 contains the true optima m, b, c_h and the best possible efficiencies $\alpha = 2c/P(D)$. α is the expected number of trials until a sample from $h(x)$ is accepted as a sample from the difference $d(x)$. In the last column the slightly larger $\hat{\alpha}$ resulting from the approximations of m, b and c_h are given.

Table 1. Intersections w_a, Optimum Squeeze and Hat Parameters.

a	w_a	x_3	b_l	c_l	m	b	c_h	α	$\hat{\alpha}$
3.1	2.1619	1.0842	1.0176	.0026	.9650	1.1969	.0036	1.1651	1.2402
3.5	2.1438	1.0822	1.0132	.0116	.9725	1.1713	.0160	1.1525	1.1643
4	2.1261	1.0800	1.0081	.0207	.9799	1.1461	.0279	1.1405	1.1493
5	2.1007	1.0764	.9998	.0340	.9906	1.1101	.0443	1.1243	1.1366
7	2.0712	1.0711	.9881	.0503	1.0032	1.0680	.0628	1.1073	1.1259
10	2.0486	1.0661	.9775	.0633	1.0129	1.0357	.0765	1.0958	1.1194
20	2.0218	1.0587	.9625	.0794	1.0222	1.0249	.0934	1.0971	1.1131
50	2.0055	1.0533	.9520	.0898	1.0275	1.0286	.1042	1.1034	1.1099
100	2.0000	1.0513	.9483	.0933	1.0293	1.0298	.1078	1.1056	1.1089
200	1.9972	1.0503	.9463	.0951	1.0302	1.0305	.1096	1.1068	1.1085
500	1.9956	1.0496	.9451	.0962	1.0307	1.0308	.1107	1.1075	1.1082
1000	1.9950	1.0494	.9447	.0966	1.0309	1.0310	.1111	1.1077	1.1081
∞	1.9945	1.0492	.9443	.0969	1.0311	1.0311	.1114	1.1079	1.1080

3.2. The Algorithm. With the following expositions the formal statement of algorithm TMA below should be understandable.

The switches a', a'' and a''' ensure that the quantities s (Step 1), β, γ, δ (Step 4), r, ss, $q(0)$, and the simple choices of m, b, c_h, $b_l = m - x_1$ and c_l (Step 6) are recalculated only if this is required by a shift of the parameter a. Steps 2 and 3 express case (I).

The squeeze functions $b(z)$ and $B(z)$ of (3.3) are used in Step 5, which is denoted as case (S). Squeeze acceptance in Step 5.1 is indicated if $U \le 1 - Z/\beta$; that is, if $\beta(1 - U) \ge Z$. Otherwise the squeeze rejection test $U > 1.0184 - Z/\gamma + Z^2/\delta$, which is equivalent to $\gamma\,\delta\,(1.0184 - U) < Z(\delta - \gamma Z)$, has to be carried out in Step 5.2.

$q(0)$ in Step 7 is evaluated as an economized expression $q(0) = \sum q_k a^{-k}$. In Table 2 a set of coefficients q_k for 9 decimal digits accuracy is listed. If $Y = s^2 X^2/a > .5$, $Q = q(X)$ in Step 7 is calculated according to its definition. Whenever $Y \le .5$ the term $\frac{1}{2}(a+1)\ln(1+Y)$ is substituted by the polynomial $\frac{1}{2}(1+1/a)\, s^2 X^2 \sum a_k Y^k$ with coefficients a_k also specified in Table 2. Thus loss of precision for small Y is avoided.

Step 8 contains case (Q). In Step 9 case (D) is entered: the new triangular deviate X can be rejected immediately if $X \leq 0$ (Step 9.1). Preliminary acceptance is indicated in Step 9.2 if $U \leq l(X)/h(X)$; using (3.4), (3.11) and $X - m = bW$ (Step 9), acceptance occurs whenever $c_h U\,(1-|W|) \leq b\,c_l\,(b_l - |X-m|)$. Otherwise a new $Q(X)$ is calculated in Step 9.3. X is rejected in Step 9.4, if $Uh(X) > g(X)\big(f(X)/g(X) - 1\big)$; with (3.1), (3.2), (3.4) and $X - m$ as above the condition for rejection reads $c_h U(1-|W|)(1+X^2/3)^2 > b(e^Q - 1)$. $e^Q - 1$ is evaluated as $\sum e_k Q^k$ ($Q < .18$) with coefficients e_k from Table 2 below.

Table 2. Coefficients of Approximating Polyomials.

q_1	.006205644	a_0	.999999997	e_1	1.000000001
q_2	−.032820259	a_1	−.499999318	e_2	.499999848
q_3	.047246697	a_2	.333303669	e_3	.166673413
q_4	−.000456172	a_3	−.249502214	e_4	.041562461
q_5	−.056949440	a_4	.195812062	e_5	.008985695
q_6	.042166133	a_5	−.146816128	ϵ = truncation error	
q_7	.044134067	a_6	.086787712		
q_8	−.064037334	a_7	−.027465302		
$\lvert\epsilon\rvert < 1.1 \times 10^{-9}$		$\lvert\epsilon\rvert < 2.6 \times 10^{-9}$		$\lvert\epsilon\rvert < 5.6 \times 10^{-10}$	

Algorithm TMA ($a > 3$; Stadlober/Dieter)

0. Pre-set $a' \leftarrow 0$, $a'' \leftarrow 0$, $a''' \leftarrow 0$ at compilation time.
1. [Set-up for (I)] If $a \neq a'$ set $a' \leftarrow a$, $s \leftarrow .921317732 + .236046804/a$.
2. [t_3-deviate] Generate U, V and set $V \leftarrow V - 1/2$. If $U^2 + V^2 > U$ go to 2, otherwise set $X \leftarrow \sqrt{3}\,V/U$.
3. [Case (I)] Set $T \leftarrow sX$. If $|X| \leq w = 1.994464166$ deliver T.
4. [Set-up for (S)] If $a \neq a''$ set $a'' \leftarrow a$, $\beta \leftarrow 6.845 + 42.8/(a-3)$, $\gamma \leftarrow 7.13 + 40.9/(a-3)$, $\delta \leftarrow 201.3 + 2207.3/(a-3)$.
5. [Case (S)] Generate U and set $Z \leftarrow X^2 - w^2$.

5.1 If $\beta\,(1-U) \geq Z$ deliver T. (Squeeze Acceptance)

5.2 If $\gamma\,\delta\,(1.0184 - U) < Z(\delta - \gamma Z)$ go to 9. (Squeeze Rejection)

6. [Set-up for (Q) and (D)] If $a \neq a'''$ set $a''' \leftarrow a$, $r \leftarrow 1/a$, $ss \leftarrow s^2$ and calculate q_0, m, b, c_h, b_l and c_l as follows:
   $q_0 \leftarrow \sum q_k r^k$ (instead of $\ln s + \frac{1}{2}\ln\frac{3}{4a} + \ln\Gamma(\frac{a+1}{2}) - \ln\Gamma(\frac{a}{2})$.)
   $m \leftarrow 1.03109 - r(.15268 + .24891r)$, if $a \leq 12.4$ set $b \leftarrow .95938 + .76577r$,
   else set $b \leftarrow 1.03109 - .09338r$. $c_h \leftarrow .11146 - .33355r$,
   $b_l \leftarrow m - .1094 + .0691r$, $c_l \leftarrow .099 - .305r$.
7. [Calculation of Q] Set $TT \leftarrow ssX^2$, $Y \leftarrow rTT$ and calculate $Q = q(X)$ as follows:
   If $Y \leq .5$: $Q \leftarrow q_0 + 2\ln(1+X^2/3) - \frac{1}{2}(1+r)\,TT \sum a_k Y^k$;
   If $Y > .5$: $Q \leftarrow q_0 + 2\ln(1+X^2/3) - \frac{1}{2}(1+a)\,\ln(1+Y)$.
8. [Case (Q)] If $\ln U \leq Q$ deliver T.
9. [Case (D)] Generate U, V and set $W \leftarrow U + V - 1$, $X \leftarrow bW + m$.

9.1 If $X \leq 0$ go to 9. Otherwise generate U and set $H \leftarrow c_h\,U\,(1-|W|)$.

9.2 If $H \leq b\,c_l\,(b_l - |X-m|)$ go to 9.5. (Preliminary Acceptance)

9.3 Set $TT \leftarrow ssX^2$, $Y \leftarrow r\,TT$ and calculate Q as in Step 7.

9.4 If $H(1+X^2/3)^2 > b(e^Q - 1)$ go to 9.
($e^Q - 1$ is calculated as $\sum e_k Q^k$)

9.5 Deliver $T \leftarrow s\,X\,sign(V-U)$.

4. The Exact Approximation Method

4.1. The Method. The basis of this method is the following well-known fact about the density of transformed random variables.

Let $f(x)$ be a density and let $h(y)$ be a differentiable function, with $h'(y) > 0$ for all y. If Y is a random variable with density $g(y) = f(h(y))h'(y)$, then $X = h(Y)$ has density $f(x)$.

For a given density $f(x)$ one has to find a suitable function $h(y)$ such that sampling from $g(y)$ can easily be accomplished. In the case of $f(x) = t_a(x)$, $a \geq 1$, Marsaglia [14] suggests

$$h(y) = \sqrt{a}\sinh\Big(\frac{y}{\sqrt{a}}\Big), \tag{4.1}$$

such that the density

$$g(y) = t_a\big(h(y)\big)h'(y) = c_a \cosh^{-a}\Big(\frac{y}{\sqrt{a}}\Big) \tag{4.2}$$

is close to the standard normal density $\phi(y)$. Because of $\cosh^{-a}(\frac{y}{\sqrt{a}}) \geq e^{-y^2/2}$, $g(y)$ can be represented as a mixture

$$g(y) = p\,\phi(y) + (1-p)\,r(y), \qquad \text{where} \quad p = \sqrt{2\pi}c_a. \tag{4.3}$$

The residual density

$$r(y) = \frac{g(y) - p\,\phi(y)}{1-p} \tag{4.4}$$

tends to the χ_5-density:

$$\lim_{a\to\infty} r(y) = \chi_5(y) = \frac{y^4}{3}\phi(y). \tag{4.5}$$

Extensive numerical calculations show that

$$r(y) \geq (1 - .56/a)\,\chi_5(y) \qquad \text{for all} \quad a \geq 1. \tag{4.6}$$

Hence $r(y)$ itself can be expressed as the mixture

$$r(y) = (1 - .56/a)\,\chi_5(y) + (.56/a)\,s(y). \tag{4.7}$$

Combination of (4.3) and (4.7) yields the three-term mixture

$$g(y) = p_1\phi(y) + p_2\chi_5(y) + p_3 s(y), \tag{4.8}$$

where

$$p_1 = p = \sqrt{2\pi}\, c_a = \sqrt{\frac{2}{a}}\, \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} \approx 1 - \frac{1}{4\,a} + \frac{1}{32\,a^2},$$

$$p_2 = (1 - \frac{.56}{a})(1 - p_1) \approx \frac{1}{4\,a} - \frac{1.37}{8a^2},$$

$$p_3 = 1 - p_1 - p_2 \approx \frac{.14}{a^2}.$$

The approximations are derived from Stirling's formula; in Table 3 the probabilities p_1, p_2, p_3 are listed for various parameters a.

Deviates Y from $\chi_5(y)$ can be generated as the signed square root of the sum of a χ_4^2 and a χ_1^2 deviate, that is $Y \leftarrow \pm\sqrt{-2\ln(UV) + Z^2}$, with U, V (0,1)-uniforms and Z standard normal. Since Marsaglia did not work out a method for sampling from $s(y)$ we had to develop a generation procedure for variables with density

$$s(y) = \frac{a}{.56}\left(r(y) - \left(1 - \frac{.56}{a}\right)\chi_5(y)\right),$$

$$= \frac{a}{.56}\frac{1}{\sqrt{2\pi}}\left(c_r\left(\cosh^{-a}\left(\frac{y}{\sqrt{a}}\right) - e^{-y^2/2}\right) - c_h y^4 e^{-y^2/2}\right)$$

where

$$c_r = \frac{p_1}{1 - p_1} \qquad \text{and} \qquad c_h = \frac{1}{3}\left(1 - \frac{.56}{a}\right).$$

We employ the acceptance-rejection method with a double-exponential envelope

$$h_s(y) = \frac{c}{2b}\exp\left(-\frac{|y - m|}{b}\right) \geq s(y) \quad \text{for } y \geq 0. \tag{4.10}$$

The optimal cover functions $h_s(y)$ touch $s(y)$ at two or three points (see Figure 2):

(i) If $1 \leq a < 3.0681$, the optimal $h_s(y)$ contacts $s(y)$ at L_1 and R, where $1.22 < L_1 < 1.50$ and $3.73 < R < 4.04$. The local minimum (for $a > 2.2$) of the difference $h_s(y) - s(y)$ at L_2 ($2.43 < L_2 < 2.73$) is positive.

(ii) If $3.0681 \leq a$, $h_s(y)$ touches $s(y)$ at L_1, L_2 and R, where $1.0 < L_1 < 1.23$, $2.42 < L_2 < 2.75$, $3.69 < R < 3.74$.

The calculation of the optimal b, c and m follows from Dieter [7] and Ahrens/Dieter [3]. In the case (i) b is determined by

$$b = \frac{1}{2}(R - L_1), \qquad L_1 < m < R, \tag{4.11}$$

and L_1 and R fulfill the equations

$$s'(L_1) = \frac{1}{2}s(L_1), \qquad s'(R) = -\frac{1}{b}s(R). \tag{4.12}$$

Equations (4.11) and (4.12) are sufficient to determine L_1, R and b. Thereafter c and m are obtained from

$$c = 2\,b\,e\sqrt{s(L_1)s(R)}, \qquad m = \frac{1}{2}\left(L_1 + R + b\ln\left(\frac{s(R)}{s(L_1)}\right)\right). \tag{4.13}$$

In the case (ii), L_1 and L_2 are determined by the left half and R by the right half of equation (4.12). b is now obtained as

$$b = \frac{L_2 - L_1}{\ln\left(\frac{s(L_2)}{s(L_1)}\right)}. \tag{4.14}$$

After the simultaneous calculation of L_1, L_2, R and b, m is worked out as before and c is calculated from

$$c = 2\,b\,s(L_1)\exp\left(\frac{m - L_1}{b}\right). \tag{4.15}$$

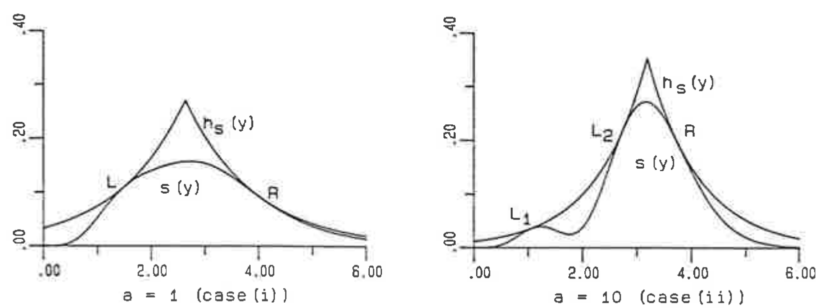

Figure 2: Densities $s(y)$ and optimal envelopes $h_s(y)$.

The optimal parameters b, $c_b = .56\sqrt{2\pi}\,c/(2b)$ and m in (4.10), listed in Table 3, lead to the best possible efficiencies $\alpha = 2c$. As in Section 3 we use suitable approximations of b, c_b and m, given in Step 6 of algorithm TEA below. These choices result in the efficiencies $\hat{\alpha}$, which are printed in the last column of Table 3. In the left-hand side of Table 3 the mixing probabilities p_1, p_2 and p_3 are included.

4.2 The Algorithm. In the formal statement of algorithm TEA below the triggers a' and a'' prevent unnecessary recalculations of r, s and p (Step 1) and of c_r, c_h, m, b and c_b (Step 6). $p = \sqrt{2\pi}\,c_a$ is evaluated as $\sum c_k a^{-k}$ with coefficients c_k from Table 4. In Step 2 the standard normal deviate Y is generated with probability $p_1 = p$ and the transformation $X = h(Y)$ in Step 3 is carried out as follows:

If $T = Y^2/a > 4$, then X is calculated according to (4.1); for $T < 4$ replace the term $(Y/\sqrt{T})\sinh\sqrt{T}$ by the economized expression $\sum a_k T^k$ with coefficients a_k also listed in Table 4. In this way we avoid the use of slow system routines for sinh(y). With

Table 3. Mixing Probabilities and Parameters for $h_s(y)$.

a	p_1	p_2	p_3	m	b	c_b	α	$\hat{\alpha}$
1	.797884	.088931	.113185	2.6399	1.2674	.3821	1.3802	1.3803
2	.886227	.081917	.031856	3.0070	1.2532	.3796	1.3556	1.3724
3	.921318	.063995	.014687	3.1760	1.2525	.3717	1.3265	1.3277
4	.939986	.051612	.008402	3.1958	1.1364	.4062	1.3153	1.3939
5	.951533	.043039	.005428	3.1991	1.0669	.4329	1.3161	1.4178
10	.975350	.023270	.001380	3.1942	.9415	.4970	1.3333	1.4362
20	.987583	.012069	.000348	3.1851	.8848	.5350	1.3490	1.4187
50	.995013	.004931	.000056	3.1771	.8525	.5599	1.3603	1.4062
100	.997503	.002483	.000014	3.1740	.8421	.5686	1.3643	1.4010
200	.998751	.001246	.000003	3.1724	.8369	.5730	1.3664	1.3983
500	.999500	.000499	.000001	3.1713	.8338	.5756	1.3676	1.3966
1000	.999750	.000250	.000000	3.1710	.8327	.5765	1.3681	1.3960

probability p_2 a deviate Y from $\chi_5(y)$ is generated and $X = h(Y)$ is accepted in Step 5. In Step 7 the acceptance-rejection procedure for sampling from $s(y)$ begins with probability p_3. The double-exponential deviate Y is rejected immediately in Step 8 if $Y < 0$. In Step 11 rejection is indicated whenever $|U| > s(Y)/h_s(Y)$, i.e. if $|U|h_s(Y) > s(Y)$. Using $V = \exp(-|Y-m|/b)$, (4.9), (4.10) and GS in Step 10, rejection of Y occurs if $c_b\,|U|\,V/a > GS$. If Y is accepted, the new $X = h(sign(V-.5)Y)$ is delivered in Step 12. The term $G = \cosh^{-a}(Y/\sqrt{a})$ of GS (Step 10) is substituted by the economized expression $\exp(-Y^2\sum b_k T^k)(1/T)\ln\cosh\sqrt{T} = \sum b_k T^k$, whenever $T = Y^2/a \le 1$. The coefficients b_k are specified in Table 4. (The direct calculation of G for large a yielded intolerable errors in the test function of Step 11.)

Algorithm TEA ($a \ge 1$; Marsaglia [14], Stadlober)

0. Preset $a' \leftarrow 0$, $a'' \leftarrow 0$ (at compilation time).
1. [Set-up for $\phi(y)$, $\chi_5(y)$] If $a \ne a'$ set $a' \leftarrow a$, $r \leftarrow 1/a$, $s \leftarrow \sqrt{a}$, $p \leftarrow \sum c_k r^k$ (instead of $\sqrt{\frac{2}{a}}\,\Gamma\left(\frac{a+1}{2}\right)/\Gamma\left(\frac{a}{2}\right)$.)
2. [$\phi(y)$] Generate U. If $U > p$ go to 4.
   Otherwise generate a standard normal deviate Y.
3. Set $T \leftarrow rY^2$ and calculate X as follows: If $T \le 4$ deliver $X \leftarrow Y\sum a_k T^k$;
   If $T > 4$ deliver $X \leftarrow s\sinh(Y/s)$.
4. If $U > p + (1-.56r)(1-p)$ go to 6.
5. [$\chi_5(y)$] Generate U, V and a standard normal deviate Z.
   Set $Y \leftarrow sign(Z)\sqrt{Z^2 - 2\ln(UV)}$, $T \leftarrow rY^2$ and calculate X as in Step 3.
6. [Set-up for $s(y)$] If $a'' \ne a$ set $a'' \leftarrow a$ and
   calculate c_r, c_h, m, b and c_b as follows:
   $c_r \leftarrow p/(1-p)$, $c_h \leftarrow (1-.56r)/3$.
   If $a < 3$, set $m \leftarrow 3.374 - .734r$, $b \leftarrow 1.239 + .0284r$, $c_b \leftarrow .3801 + .0021a$.
   If $a \ge 3$, set $m \leftarrow 3.176$, $b \leftarrow .83 + 1.2675r$, $c_b \leftarrow .59 - .654r$.
7. [$s(y)$] Generate U, V and set $E \leftarrow -\ln V$ (standard-exponential deviate),
   $U \leftarrow U + U - 1$ and $Y \leftarrow m + E\,b\,sign(U)$.
8. If $Y \le 0$ go to 7.
9. Set $T \leftarrow rY^2$ and calculate G as follows:
   If $T \le 1$: $G \leftarrow \exp(-Y^2\sum b_k T^k)$;
   If $T > 1$: $G \leftarrow \cosh^{-a}(Y/s)$.
10. Set $GS \leftarrow c_r G - \exp(-Y^2/2)(c_r + c_h Y^4)$.
11. If $r\,c_b\,|U|\,V > GS$ go to 7.
12. Set $Y \leftarrow sign(V-.5)\,Y$ and calculate X as in Step 3.

Table 4. Coefficients of Approximating Polyomials.

c_0	1.000000002	a_0	1.000000000	b_0	.500000000
c_1	−.250000601	a_1	.166666673	b_1	−.083333300
c_2	.031280359	a_2	.008333315	b_2	.022221494
c_3	.038452052	a_3	.000198432	b_3	−.006739976
c_4	−.003839730	a_4	.000002747	b_4	.002161791
c_5	−.088380021	a_5	.000000027	b_5	−.000680136
c_6	.162459233			b_6	.000176723
c_7	−.160425061			b_7	−.000025766
c_8	.096731805				
c_9	−.033534049				
c_{10}	.005140572				
$\lvert\epsilon\rvert < 1.9 \times 10^{-9}$		$\lvert\epsilon\rvert < 3.5 \times 10^{-10}$		$\lvert\epsilon\rvert < 2.6 \times 10^{-10}$	

5. Computational Experience and Comparison

All algorithms were written as Fortran functions and tested on a UNIVAC 1100/81 computer (Compiler FTN 10R1A). Uniform deviates were obtained by the multiplicative-congruential generator URAND (factor = 5308871541, modulus = 2^{35}), coded in Assembler. For TNG and TEA an assembler version of the normal generator FL5 due to Ahrens/Dieter [2] was used. The gamma deviates in TNG were generated by a fast Fortran subprogram, given in Ahrens/Kohrt/Dieter [5]. The algorithms are compared with regard to (i) generation times, (ii) set-up costs (initialization times), (iii) words of compiled code (including constants) and (iv) expected number of uniforms N_a required per deviate. The observed computation times in Table 5 are averages of 5 replications of 10,000 variates for each choice of a.

It should be noted that the classical method TNG is considerably slower than all the other algorithms compared in Table 5, even with the fastest known subprograms for the normal and gamma distributions. In the group of the acceptance-rejection and mixing algorithms the fastest method is our modification of Best's T3T [6], followed by TMXS, TMX, TIRS, TIR and TAR, all of Kinderman et al. [10]. But T3T works only for $a > 3$ and its consumption of uniform deviates is rather high.

Table 5. Times in microseconds/deviate.

a	TNG	TAR	TIR	TIRS	T3T	TMX	TMXS	TRU	TD	TMA	TEA
1	291	106	98	93	—	97	87	110	—	—	133
2	211	116	107	104	—	104	99	97	—	—	101
3	201	121	110	108	91[1]	107	103	80	79[1]	65[1]	93
4	194	122	112	111	94	109	105	81	83	68	90
5	191	124	113	112	95	109	106	81	86	70	88
7	187	125	114	114	97	110	107	82	89	71	86
10	184	126	115	115	99	111	108	82	91	72	84
15	182	127	117	117	97	111	109	83	92	72	83
30	178	128	117	118	96	112	110	83	93	73	82
50	176	129	117	118	96	113	110	83	93	73	82
100	175	129	117	118	95	113	110	83	94	73	82
1000	174	131	119	120	95	114	111	84	95	73	82
Set up costs	43[2]	—	—	70	—	—	32	202	12	10	68
Words	59	105	223	267	14	274	311	158	362	439	420
N_1	3.90	2.55	2.55	2.55	3.57[3]	3.27	2.63	2.55	2.69[3]	2.69[3]	2.59
N_∞	2.96	3.19	3.19	3.19	4.15	3.85	3.05	2.74	2.90	2.93	2.23

[1] Read $a = 3.1$, [2] For $a \geq 2$, [3] Read $N_{3.1}$.

In terms of generation time the above methods are outperformed (for $a > 2$) by the combined ratio-of-uniforms algorithm TRU of Kinderman/Monahan [9]. On the other hand, this procedure is significantly slower than our modified acceptance-rejection algorithm TMA (valid only for $a > 3$). Actually TMA is the fastest method for fixed and frequently changing parameters. If the parameters vary a lot small set-up costs for parameter-dependent constants are important. The expected numbers of uniforms required are small for TRU and TMA. The advantage of TRU is its compactness, whereas TMA is rather complicated.

Our exact-approximation procedure TEA is comparable in speed with TRU for $a \geq 10$. However, this result and the numbers of uniforms N_a depend on our choice of the normal generator.

Finally we can give the following recommendations: If speed and flexibility (fast set-up) are the dominant considerations then a combination of TIR ($a \leq 3$) and TMA ($a > 3$) can be advocated. If speed and compactness are important, TRU could be the first choice.

Acknowledgement. The authors wish to express their appreciation and thanks to *Joe Ahrens* for his constructive and valuable comments and suggestions, which led to an improved exposition of the paper.

6. Appendix

The speed of algorithm TMA in Section 3 was strongly improved by squeeze functions which depended on the inequalities

$$1 - \frac{1}{\beta} z \leq r(x) \leq 1.0184 - \frac{1}{\gamma} z + \frac{1}{\delta} z^2 \qquad \text{for } w \leq |x| \tag{6.1}$$

where

$$z = x^2 - w^2, \quad \beta = 6.845 + \frac{42.8}{a-3}, \quad \gamma = 7.13 + \frac{40.9}{a-3}, \quad \delta = 201.3 + \frac{2207.3}{a-3} \tag{6.2}$$

and

$$r(x) = s \, \frac{c_a}{c_3} \frac{u_a(sx)}{u_3(x)}, \qquad u_a(x) = \left(1 + \frac{x^2}{a}\right)^{-\frac{a+1}{2}} \tag{6.3}$$

is the acceptance-rejection quotient of the two t-distributions. Expressing $r(x)$ as a function of z leads to

$$r(x) = R_a q(z) \tag{6.4}$$

where

$$q(z) = \left(1 + \frac{z}{as^{-2} + w^2}\right)^{-\frac{a+1}{2}} \left(1 + \frac{z}{w^2 + 3}\right)^2 \tag{6.5}$$

and

$$R_a = s \, \frac{c_a}{c_3} \left(1 + \frac{w^2 s^2}{a}\right)^{-\frac{a+1}{2}} \left(1 + \frac{w^2}{3}\right)^2. \tag{6.6}$$

Inequality (6.1) will be proved in the form

$$1 + Cz \leq q(z) \leq 1 + Dz + Ez^2 \qquad \text{for } z \geq 0 \tag{6.7}$$

where

$$C = q^{(1)}(z_C), \qquad D = q^{(1)}(0), \qquad E = \frac{q^{(1)}(z_E) - q^{(1)}(0)}{2z_E}.$$

Multiplication of (6.7) by R_a yields (6.1) since $1 \leq R_a \leq 1.0184$ holds for all a.

Applying the Taylor expansion

$$q(z) = 1 + q^{(1)}(0)z + \frac{1}{2} q^{(2)}(0)z^2 + \frac{1}{6} q^{(3)}(0)z^3 + \frac{1}{24} q^{(4)}(0)z^4 + \ldots$$

transforms the left-hand side of (6.7) into

$$C \leq q^{(1)}(0) + \frac{1}{2} q^{(2)}(0)z + \frac{1}{6} q^{(3)}(0)z^2 + \ldots$$

or, considering only quadratic terms in z,

$$C \le q^{(1)}(0) - \frac{3\left(q^{(2)}(0)\right)^2}{8q^{(3)}(0)} + \frac{1}{6}q^{(3)}(0)\left(z + \frac{3q^{(2)}(0)}{2q^{(3)}(0)}\right)^2. \qquad (6.8)$$

This means that

$$q^{(3)}(0) \ge 0 \qquad \text{for all } a \ge 3 \qquad (6.9)$$

has to be fulfilled.

Applying the Taylor expansion of $q(z)$ to the right-hand side of (6.7) yields

$$q^{(1)}(0) + \frac{1}{2}q^{(2)}(0)z + \frac{1}{6}q^{(3)}(0)z^2 + \frac{1}{24}q^{(4)}(0)z^3 + \cdots \le D + Ez. \qquad (6.10)$$

Since $q^{(3)}(0) \ge 0$ was already assumed, this can be true only if $D = q^{(1)}(0)$. Thus (6.10) transforms into

$$\frac{1}{2}q^{(2)}(0) - \frac{\left(q^{(3)}(0)\right)^2}{6q^{(4)}(0)} + \frac{1}{24}q^{(4)}(0)\left(z + 2\frac{q^{(3)}(0)}{q^{(4)}(0)}\right)^2 \le E. \qquad (6.11)$$

Therefore

$$q^{(4)}(0) \le 0 \qquad \text{for all } a > 3$$

has to be satisfied.

The conditions (6.9) and (6.12) are checked numerically. For this consider the identity

$$q^{(1)}(z) = \sum_{n=1}^{\infty} \frac{1}{(n-1)!}q_n z^{n-1} = \frac{q^{(1)}(z)}{q(z)}\, q(z)$$

$$= \left(-\frac{a+1}{2(as^{-2} + w^2 + z)} + \frac{2}{w^2 + 3 + z}\right)\sum_{n=1}^{\infty} \frac{1}{n!}q_n z^n.$$

A multiplication by $2(as^{-2} + w^2 + z)(w^2 + 3 + z)$ leads for $n \ge 2$ to the recursion

$$q^{(n)}(0) = -\left(\frac{2n + a - 1}{2(as - 2 + w^2)} + \frac{n-3}{w^2 + 3}\right)q^{(n-1)}(0) - \frac{(n-1)(2n + a - 7)}{2(as^{-2} + w^2)(w^2 + 3)}q^{(n-2)}(0),$$

Keeping in mind that

$$q^{(0)}(0) = q(0) = 1, \qquad q^{(1)}(0) = -\frac{a+1}{2(as^{-2} + w^2)} + \frac{2}{w^2 + 3},$$

the calculation of $q^{(3)}(0)$ and $q^{(4)}(0)$ for $a > 3$ is no problem at all, and the results are that

$$q^{(1)}(0) < 0, \qquad q^{(2)}(0) < 0, \qquad q^{(3)}(0) > 0, \qquad q^{(4)}(0) < 0 \qquad \text{for } a > 3.$$

The best values of C and D follow now easily from (6.7):

$$C = \min_z\left(\frac{q(z) - 1}{z}\right) \qquad (6.13)$$

$$E = \max_z\left(\frac{q(z) - 1 - q^{(1)}(0)z}{z^2}\right) \qquad (6.14)$$

The values z_C and z_E of z for which the optima are attained are solutions of the following equations

$$q(z_C) - 1 - z_C q^{(1)}(z_C) = 0 \qquad \text{for (6.13).} \qquad (6.15)$$

$$2(q(z_E) - 1) - z_E q^{(1)}(z_E) - z_E q^{(1)}(0) = 0 \qquad \text{for (6.14).} \qquad (6.16)$$

Finally, C and E are determined to

$$C = \frac{q(z_C) - 1}{z_C} = q^{(1)}(z_C). \qquad (6.17)$$

$$E = \frac{q(z - E) - 1 - q^{(1)}(0)z_E}{z_E^2} = \frac{q^{(1)}(z_E) - q^{(1)}(0)}{2z_E}. \qquad (6.18)$$

To prove that z_C and z_E are in fact optima of the corresponding functions, the second derivatives in (6.13) and (6.14) have to be determined. After some simplification this leads to the conditions

$$q^{(2)}(z_C) > 0, \qquad z_E q^{(2)}(z_E) - q^{(1)}(z_E) + q^{(1)}(0) < 0. \qquad (6.19)$$

For the numerical calculation of second derivatives of $q(z)$ one may apply the logarithmic derivatives of $q(z)$

$$\frac{q^{(1)}(z)}{q(z)} = -\frac{a+1}{2(as^{-2} + w^2 + z)} + \frac{2}{w^2 + 3 + z}$$

which leads to

$$\frac{q^{(2)}(z)}{q(z)} = \left(\frac{q^{(1)}(z)}{q(z)}\right)^2 + \frac{a+1}{2(as^{-2} + w^2 + z)^2} - \frac{2}{(w^2 + 3 + z)^2}.$$

For checking (6.19), $q^{(1)}(z)$ and $q^{(2)}(z)$ can be calculated numerically from these two formulas.

The numerical calculation of z_C and z_E is carried out as follows. First of all, (6.15) and (6.16) are combined into

$$(1 + F)(q(z) - 1) - zq^{(1)}(z) - Fzq^{(1)}(0) = 0 \qquad (6.20)$$

where

$$F = \begin{cases} 0, & \text{for } z = z_C; \\ 1, & \text{for } z = z_E. \end{cases}$$

(6.20) is solved by Newton iteration for z. The solutions z_C and z_E determine C and E:

$$C_F = \frac{q(z) - 1 - Fzq^{(1)}(0)}{z^{1+F}} = \frac{q^{(1)}(z) - Fq^{(1)}(0)}{(1+F)z^F} \tag{6.21}$$

where

$$C_F = \begin{cases} C, & \text{for } F = 0; \\ E, & \text{for } F = 1. \end{cases}$$

Numerical calculations show that

$$C_F = C_F^0 + \frac{1}{a+\epsilon} C_F^1$$

where ϵ is nearly constant. Setting $A \leftarrow 1/a$, C_F^1 is obtained by differentiation of C_F with respect to A.

References

[1] Ahrens, J. H. and Dieter, U., *Computer methods for sampling from the exponential and normal distributions*, Comm. ACM **15** (1972), 873–882.

[2] Ahrens, J. H. and Dieter, U., *Extensions of Forsythe's method for random sampling from the normal distribution*, Math. Comput. **27** (1973), 927–937.

[3] Ahrens, J. H. and Dieter, U., *Generating gamma variates by a modified rejection technique*, Comm. ACM **25** (1982), 47–54.

[4] Ahrens, J. H. and Dieter, U., *Efficient table-free sampling methods for the exponential, Cauchy, and normal distributions*, Comm. ACM **31** (1988), 1330–1337.

[5] Ahrens, J. H., Kohrt, K. D. and Dieter, U.:, *ALGORITHM 599. Sampling from gamma and Poisson distributions*, ACM Trans. Math. Software **9** (1983), 255–257.

[6] Best, D. J., *A simple algorithm for the computer generation of random samples from a Student's t or symmetric beta distribution*, COMPSTAT 1978, 341–347.

[7] Dieter, U., *Optimal acceptance-rejection envelopes for sampling from various distributions*, in "ICOSCO-I, 30 March – 2 April 1987, Çeşme, Izmir, Turkey (Dudewicz, E.J., ed.)," pp. 505–530.

[8] Kinderman, A. J. and Monahan, J. F., *Computer generation of random variables using the ratio of uniform deviates*, ACM Trans. Math. Software **3**, (1977), 257-260.

[9] Kinderman, A. J. and Monahan, J. F., *New methods for generating Student's t and gamma variables*, Computing **23** (1980), 369–377.

[10] Kinderman, A. J., Monahan, J. F. and Ramage, J. G., *Computer methods for sampling from Student's t distribution*, Math. Comput. **31** (1977), 1009–1018.

[11] Kinderman, A. J. and Ramage, J. G., *Computer generation of normal random variables*, J.A.S.A. **71** (1976), 893–896.

[12] Kronmal, R. A. and Peterson, A. V.:, *A variant of the acceptance-rejection method for computer generation of random variables*, J.A.S.A. **76** (1981), 446–451.

[13] Marsaglia, G., *Generating random variables with a t-distribution*, Math. Comput. **34** (1980), 235–236.

[14] Marsaglia, G., *The exact-approximation method for generating random variables in a computer*, Preliminary version 1979, J.A.S.A. **79** (1984), 218–221.

[15] Neumann, J. von, *Various techniques used in connection with random digits*, in "Collected Works 5," Pergamon Press, New York, 1963, pp. 768–770.

[16] Stadlober, E., *Generating Student's t variates by a modified rejection method*, in "Probability and Statistical Inference (Grossmann, W. et al., eds.)," Reidel Publ., Dordrecht, 1982, pp. 349–360.

[17] Stadlober, E., *Die Erzeugung von Zufallszahlen aus der t-Verteilung*, Math. Statist. Sektion, Forschungszentrum Graz, Bericht Nr. **206**, 1983.