Proceedings of the 1989 Winter Simulation Conference
E.A. MacNair, K.J. Musselman, P. Heidelberger (eds.)

RATIO OF UNIFORMS AS A CONVENIENT METHOD FOR SAMPLING FROM CLASSICAL DISCRETE DISTRIBUTIONS

Ernst Stadlober
Institute of Statistics
Graz University of Technology
A–8010 Graz, Lessingstrasse 27
Austria

ABSTRACT

Many computer methods for generating variates from classical discrete distributions are available; some of them are simple and others are very fast. However, simple or convenient procedures are slow when the means μ are large. Very fast algorithms are rather involved, so that most users will not go to the trouble of implementing them. Fortunately, algorithms having the advantage of being simple and fast are obtained by applying the ratio of uniforms method to discrete distributions in a skilful way. We discuss several issues of this approach with respect to Poisson, binomial and hypergeometric distributions.

1. INTRODUCTION

Random variate generation from discrete distributions has received considerable attention in the literature. Several types of generators have been developed, see e.g. Devroye (1986) for an excellent overview. Simple inversion by sequential search and methods based upon distributional properties are convenient, but their execution times are not uniformly bounded over the whole set of parameter values. On the other hand ultra–fast algorithms are obtained by table–aided inversion (Chen and Asau, 1974, Ahrens and Kohrt, 1981) or by the alias method of Walker (1977). However, both methods are not efficient if the parameters of the distribution in hand vary all the time, because the initial set–up of new tables is costly. The acceptance rejection approach leads to uniformly fast algorithms, i.e. algorithms with compution time of order O(1), but competitive procedures are rather complex; see Ahrens and Dieter (1982), and Schmeiser and Kachitvichyanukul (1981) for the Poisson case, or Kachitvichyanukul and Schmeiser (1985, 1988) for the hypergeometric and binomial cases, respectively. The extension of the ratio of uniforms method – originally designed for continuous distributions by Kinderman and Monahan (1977) – to unimodal discrete distributions leads to algorithms combining simplicity with efficiency in the fixed and variable parameter case. The principle of the method may be described briefly as follows. The standardized histogram function $f(x)$ is covered by the *hat function* $h(x) = min(1, s^2/(x-a)^2)$ with suitably chosen location parameter a and scale parameter s. Then the following simple rejection procedure can be carried out.

Generate a random pair (U, V) uniformly distributed over the rectangle $R = (0,1) \times (-1,1)$, set $K \leftarrow \lfloor s\,V/U + a \rfloor$ and return K as a sample from $f(x)$ if $U^2 \le f(K)$ is fullfilled. Otherwise reject K and try again.

In the next section details of the general sampling method are considered. Section 3 is devoted to the construction of hat functions with suitable parameters a and s. Algorithmic aspects are discussed in Section 4. In the final Section 5 some comparisons with other algorithms are given. The development in this paper is based on material which appeared in Stadlober (1989b).

2. THE GENERAL SAMPLING METHOD

The ratio of uniforms method as suggested by Kinderman and Monahan (1977) for continuous distributions is based on the following idea.

Let $f(x)$ be a rescaled density with finite integral k, and let $C = \{(u,v) \mid 0 < u < \sqrt{f(v/u)}\}$. Then if (U,V) is uniformly distributed over C, $X = V/U$ has density $\frac{1}{k} f(x)$.

In order to create a generator, a method for sampling from C has to be established. If $f(x)$ and $x^2 f(x)$ are bounded then C can be encased in the rectangle

$$R = \{(u,v) \mid 0 < u < u_+, v_- < v < v_+\},$$

where

$$u_+ = \sup_x \sqrt{f(x)},\ v_- = \inf_x x\sqrt{f(x)},\ v_+ = \sup_x x\sqrt{f(x)}.$$

In this case acceptance rejection can be applied to obtain (U,V) uniformly distributed over C by generating a point uniformly in the rectangle R and rejecting this point if it is not in C. For applications see Kinderman and Monahan (1980), and Monahan (1987).

The following description of the method (Stadlober, 1989a) allows us to use ratio of uniforms in the discrete case. By taking (U,V) uniformly distributed over the *standardized* rectangle

$$R_s = \{(u,v) \mid 0 < u < 1, -1 < v < 1\}$$

and by transforming

$$(U,V) \quad \text{to} \quad (X,Y) = (a + s\,V/U,\ U^2)$$

we obtain the transformed *table mountain*

$$T(R_s) = \{(x,y) \mid -\infty < x < \infty\ ,\ 0 < y < h(x)\},$$

where

$$h(x) = \begin{cases} 1 & , \quad a - s \le x \le a + s \\ \frac{s^2}{(x-a)^2} & , \quad \text{elsewhere} \end{cases}.$$

$T(R_s)$ should cover the domain

$$T(C) = \{(x,y) \mid -\infty < x < \infty\ ,\ 0 < y < f(x)\}$$

such that $h(x)$ may serve as hat function of $f(x)$.

In this way *ratio of uniforms with rectangles can be interpreted as acceptance rejection with table mountain hats* $h(x)$ (see Figure 1 on the right column). Obviously, the random pair (X,Y) is uniformly distributed over $T(R_s)$ which has area $4s$. The marginal density of X is simply $g(x) = \frac{1}{4s}h(x)$ and Y is uniformly distributed over $(0, h(x))$ for every fixed x. Since we are restricting $0 < u < 1$ and hence $h(x) \le 1$ we must standardize the histogram function

$$f_*(x) = p_j,\ j \le x < j + 1 \text{ for all mass points } j,$$

to

$$f(x) = f_*(x)/p_m,\ \text{where } p_m = \max_j p_j.$$

Now $f(x)$ can never exceed $h(x)$ at the top. But the validity of $f(x) \le h(x)$ on the slopes depends on the free parameters a and s of $h(x)$. Ideally the hat parameters a and s should be determined such that the average number of trials - called also efficiency of the method - is as small as possible. Note that the best choices of a and s usually need to be computed numerically. Suboptimal values of a and s for the Poisson, binomial and hypergeometric distributions are considered in the next section. We conclude that ratio of uniforms methods for unimodal discrete distributions can be stated in the following standard format.

$$\alpha_s = \frac{\int h(x)dx}{\int f(x)dx} = 4s\,p_m$$

Ratio of uniforms for discrete distributions

A. Generate X with density $g(x)$ and set $K \leftarrow \lfloor X \rfloor$.
(Generate (U,V) uniformly over $(0,1) \times (0,1)$ and set $X \leftarrow a + s(2V-1)/U$, $K \leftarrow \lfloor X \rfloor$).

B. Take Y uniformly over $(0, 4sg(X))$.
(Set $Y \leftarrow U^2$).

C. If $Y \le f(K)$ return K as a sample from $\{p_j\}$.
Otherwise goto A.

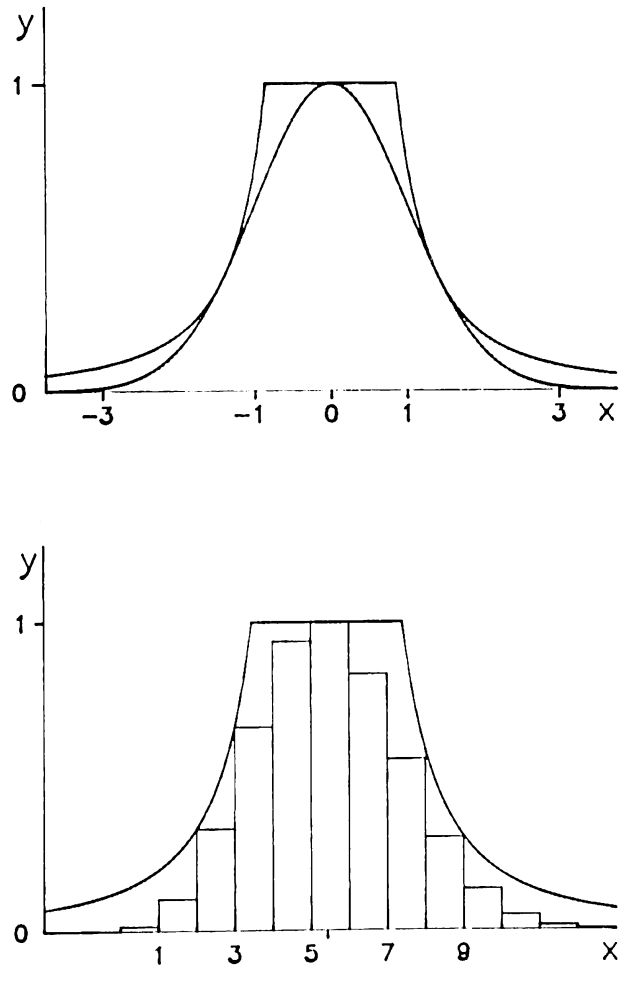

Figure 1: Optimal table mountains for standardized normal and binomial $(20, \frac{1}{4})$-distributions.

3. CONSTRUCTION OF TABLE MOUNTAIN HATS

The performance of ratio of uniforms algorithms for specific discrete distributions is governed by the choice of the hat parameters a and s. Appropriate values of a and s should be easy to calculate and they should guarantee a good fit of the table mountains over a broad range of the distribution parameters.

3.1 Poisson and binomial cases

The first ratio of uniforms procedures for a special discrete distribution are due to Ahrens and Dieter (1989), who developed some algorithmic versions for Poisson distributions with means $\mu \geq 1$. Similar algorithms for binomial distributions with means $\mu = np \geq 1$ ($p \leq \frac{1}{2}$ without loss of generality) were proposed in Stadlober (1989a). In both cases the location parameter a was fixed at

$$a = \mu + \frac{1}{2},$$

which is of course only suboptimal for skewed Poisson histograms, but it is clearly the best idea for symmetric binomial or hypergeometric distributions. After some experimentation the scale parameter of the hat $h(x)$ was chosen as

$$\hat{s} = \sqrt{\frac{2}{e}\left(\sigma^2 + \frac{1}{2}\right)} + \frac{3}{2} - \sqrt{\frac{3}{e}}, \qquad (1)$$

where σ denotes the standard deviation of the corresponding distribution. The validity of $\hat{s}$ was established numerically. Note that for fixed $a = \mu + \frac{1}{2}$, $\hat{s}$ is still not optimal (except for the Poisson(1) distribution and the limiting normal case), but it is possible to replace the convenient choices $\hat{s}$ by the true optima s^*. The following simple rule for computing values of s^* in the binomial case has been derived in Stadlober (1989b, Lemma 4.2). We recall that the Poisson distribution is contained as limiting case $n \to \infty$, $p \to 0$, μ fixed.

Lemma 1.

The quotient $q_k = \frac{f(k)}{h(k)}$ attains its maximal value 1 at $k^* = \lfloor z \rfloor$ or $k^* = \lceil z \rceil$, where $z = a - \sqrt{2a(1-p)}$.

The best possible value s^* is then simply determined as

$$s^* = (a - k^*)\sqrt{f(k^*)}. \qquad (2)$$

In the third line of Table 1 on the right column some efficiences $\alpha_{s^*} = 4s^*p_m$ are displayed, whereas the slightly larger values $\alpha_{\hat{s}}$ resulting from the approximations $\hat{s}$ (1) are given in the bottom line of the table.

For comparison purposes, the *true optima* of a and s were calculated by numerical search methods within the range $1 \leq \mu \leq 10000$. The first two lines of Table 1 below contain some differences $a - \mu$ and the best possible efficiencies α_{s^*}. It can be observed that the optimal values of a are always between $\mu + .24$ and $\mu + \frac{1}{2}$, indicating the usefulness of the suggestion $a = \mu + \frac{1}{2}$.

Table 1: Efficiencies in the binomial case resulting from different choices of hat parameters

n	20	100	1000	2000	Poisson
	.261	*.246*	*.244*	*.243*	*.243*
	1.856	1.834	1.829	1.829	1.829
$\mu = 1$	2.207	2.207	2.207	2.207	2.207
	2.237	2.213	2.208	2.207	2.207
	.500	*.332*	*.289*	*.287*	*.285*
	1.598	1.536	1.524	1.523	1.522
$\mu = 10$	1.598	1.595	1.599	1.599	1.599
	1.735	1.632	1.617	1.617	1.616
		.500	*.336*	*.337*	*.324*
		1.468	1.438	1.437	1.437
	$\mu = 50$	1.468	1.463	1.462	1.462
		1.522	1.478	1.476	1.475
			.500	*.419*	*.325*
			1.400	1.394	1.390
		$\mu = 500$	1.400	1.398	1.398
			1.415	1.406	1.401
				.500	*.329*
				1.390	1.384
			$\mu = 1000$	1.390	1.389
				1.401	1.392

The four entries in each block are:
$a - \mu$ *for optimal* a.
α_{s^*} = best possible efficiency
($\alpha_{s^*} \to 4/\sqrt{\pi e} = 1.3687\ldots$,
as $n, \mu \to \infty, p$ fixed).
α_{s^*} = best efficiency for $a = \mu + \frac{1}{2}$.
$\alpha_{\hat{s}}$ = efficiency resulting from $\hat{s}$ in (1).

3.2 Hypergeometric case

The hypergeometric distribution causes more problems for the designer of random variate generators, mainly because of its three shape parameters N, M and n. Let a be fixed at $a = \mu + \frac{1}{2}$ as before, where $\mu = n\frac{M}{N} \geq 1$. Here the discussion is reduced to the cases $n \leq \frac{N}{2}$, $M \leq \frac{N}{2}$, which is not restrictive. First of all the optimum scale parameter s^* was determined numerically for various parameter combinations (N, M, n) and compared with the approximation $\hat{s}$ as defined in (1). In

all cases we confirmed that $\hat{s}$ was larger than the minimum value s^*. There is no reason to doubt the validity of $\hat{s} \geq s^*$ for hypergeometric distributions not explicitely investigated. However, it appeared that an efficient algorithm based on the better values s^* could be developed, since the following lemma (see Stadlober,1989b, Lemma 5.2) allows us to evaluate s^* directly without any numerical optimization.

Lemma 2.
The ratio $q_k = \frac{f(k)}{h(k)}$ *is maximal at one of the two points* $k^* = \lfloor z \rfloor$ *or* $k^* = \lceil z \rceil$, *where* $z = a - \sqrt{2a(1-\frac{M}{N})(1-\frac{n}{N})}$.
s^* *is obtained as in (2).*

4. ALGORITHMIC ASPECTS

In this section we consider some implementation issues and we explain specific properties of ratio of uniforms algorithms. The detailed statement of binomial generator BRUE^t at the end of the section is meant as an illustration of a concrete application. Important characteristics of ratio procedures are.

(a) Our Poisson, binomial and hypergeometric generators work for all means $\mu \geq 1$, under the appropriate restriction to symmetric or right skewed binomial ($p \leq \frac{1}{2}$) and hypergeometric ($n \leq \frac{N}{2}, M \leq \frac{N}{2}$) distributions, but simple inversion by sequential search from the bottom is faster if μ is smaller than some breakpoint μ^*.

(b) The evaluation of the probabilities p_j involves the computation of one (Poisson) or more (binomial, hypergeometric) logarithms of factorials $\ln k!$. The proposed methods need no table, if a routine for $\ln \Gamma(k+1) = \ln k!$ is available. In order to maintain e.g. 9 decimal digits precision one could store the values $\ln 0!, \ln 1!, \ldots, \ln 9!$ and use for $k \geq 10$ the Stirling approximation

$$\ln k! \approx \ln \sqrt{2\pi} + (k+\frac{1}{2})\ln k - k + \frac{1}{12k} - \frac{1}{360k^3},$$

whose *relative error* ϵ is smaller than 5.2×10^{-10}. We emphasize that for moderate modes m the generators can be speeded up even further by using an idea of Schmeiser and Kachitvichyanukul (1981). They suggest to compute the values of $f(k) = p_k/p_m$ via the well known recursive relationships for the probabilities p_k, starting at the mode m where $f(m) = 1$.

(c) However, the fastest algorithms are obtained when a table of $\ln k!$ is stored beforehand. In these implementations no external functions are needed as long as the distribution parameters remain the same (see binomial generator BRUE^t below).

(d) The computation times are bounded since the expected number of trials α_{s^*} per variate can never exceed the *upper bound*

$$\frac{6}{e} = 2.207276647\ldots,$$

which is attained for the Poisson (1) case. On the other hand α_{s^*} is always *larger than*

$$\frac{4}{\sqrt{\pi e}} = 1.368793121\ldots,$$

the efficiency of the limiting normal case.

Below we describe binomial generator BRUE^t, which needs stored table values $\gamma_k = \ln k!$. Note that the restriction $p \leq \frac{1}{2}$ is dropped.

Binomial generator BRUE^t
($n \min(p, 1-p) \geq 1$, stored table $\gamma_k = \ln k!$)
Constants $\ln 2 = .693147181\ldots$ and B, where B depends on the accuracy of the computer;
e.g $B = 5$ for 9 decimal digits precision.

0. *[Set-up of constants. Necessary only if the value of n or p changes]*
Set $t \leftarrow \min(p, 1-p)$, $q \leftarrow 1-t$, $a \leftarrow nt + \frac{1}{2}$,
$c \leftarrow \ln \frac{t}{q}$, $d \leftarrow \sqrt{2aq}$, $m \leftarrow \lfloor (n+1)t \rfloor$,
$g \leftarrow \gamma_m + \gamma_{n-m}$, $b \leftarrow \min(n, \lfloor a + Bd \rfloor)$,
$k \leftarrow \lfloor a - d \rfloor$, $x \leftarrow (a-k-1)/(a-k)$.
If $(n-k)tx^2 > (k+1)q$ set $k \leftarrow k+1$.
Set $h \leftarrow (a-k)\exp(\frac{1}{2}((k-m)c + g - \gamma_k - \gamma_{n-k}) + \ln 2)$.

1. *[Generation of candidate variates]*
Generate independent $(0,1)$-uniforms U, V.
Set $K \leftarrow \lfloor a + h(V - \frac{1}{2})/U \rfloor$.

2. *[Quick rejection checks]*
If $K < 0$ or $K > b$ goto 1.
Otherwise set $T \leftarrow (K-m)c + g - \gamma_K - \gamma_{n-K}$.

3. *[Rejection tests]*

3.1 If $U(4-U) - 3 \leq T$ goto 4. (Fast acceptance)

3.2 If $U(U-T) \geq 1$ goto 1. (Fast rejection)

3.3 If $2\ln U > T$ goto 1.

4. *[Check if $p > \frac{1}{2}$]*
If $p > 1/2$ set $K \leftarrow n - K$.
Return K.

Remarks. The constant b in Step 0 is a safety bound. Thereby the If-statement in Step 2 passes only variates K with significant probabilities p_K. For the calculation of $h = 2s^*$ (Step 0) Lemma 1 is applied as follows. Take $k = \lfloor a - \sqrt{2aq} \rfloor$ if the ratio of the quotients

$$\frac{q_{k+1}}{q_k} = \left(\frac{a-k-1}{a-k}\right)^2 \left(\frac{n-k}{k+1}\right) \frac{t}{q}$$

is *not greater* than 1, otherwise increase $k = k+1$. Then evaluate $h = 2s^* = 2(a-k)\sqrt{f(k)}$. The quantity $T = \ln f(K)$ (Step 2) is compared with $2\ln U$ in Step 3.3. But the calculation of $\ln U$ can be avoided most of the time by the squeeze tests in Steps 3.1 and 3.2, which are based on the inequalities $u - \frac{1}{u} \le 2\ln u \le -3 + 4u - u^2$.

5. COMPARISON OF ALGORITHMS

Ratio of uniforms procedures and other state-of-the-art algorithms were implemented in Fortran and compared on a Univac 1100/81 computer. Uniform random numbers were generated by the multiplicative congruential generator URAND (factor = 5308871541, modulus = 2^{35}), coded in Assembler. The comparisons are thoroughly documented in Stadlober (1989b). We mention that the discussion is restricted to procedures which remain efficient if the parameters of the distribution vary all the time. This excludes the guide-table method of Chen and Asau (1974) and the alias method of Walker (1977). In this section we resume the most substantial results of our empirical study.

5.1 Poisson distribution

The involved but uniformly fast methods PD (modified acceptance rejection with discrete normal distributions of Ahrens and Dieter, 1982) and PTPE (triangle-parallelogram-exponential rejection of Schmeiser and Kachitvichyanukul, 1981) seem to be the favorites for $\mu \ge 10$ if only *speed* is important. They should be complemented by table-aided inversion in case of $\mu < 10$. We mention that the ultimate choice between PD and PTPE depends on the availability of a fast normal generator on which PD is based.

If *simplicity* and *speed* are the dominant considerations PRUE^t (ratio of uniforms with $s^*(2)$ and table $\ln k!$) could be the first choice when $\mu \ge 5$, whereas simple inversion PIN should be substituted for $\mu < 5$.

5.2 Binomial distribution

Kachitvichyanukul and Schmeiser (1988) report that simple inversion BIN dominates all its competitors for $\nu = n\min(p, 1-p) \le 10$, and that their own rejection method BTPE (valid for $\nu \ge 10$) is most efficient for larger ν. In our experiments algorithm BRUE (ratio of uniforms with $s^*(2)$, external function with Stirling approximation for $\ln k!$) appeared to be faster than BIN for $\nu > 7$ and also faster than BTPE for $\nu < 30$. Algorithm BRUE is also much simpler than BTPE (376 words versus 597 words of compiled code), but BTPE has the advantage of very low set-up costs, which could be important if only a few variates are needed for a fixed combination of parameters n and p. Thus for applications in which *speed* is the main concern, a combination of BIN ($\nu \le 10$) and BTPE ($\nu > 10$) could be the method of choice. However, users preferring to work with *simple* and *reasonably fast* methods would rather decide in favor of the combined algorithm BIN/BRUE ($\nu \le 7/\nu > 7$).

Comparative analysis of even faster algorithmic versions, supported by a stored table of values $\ln k!$, demonstrates that our generator BRUE^t (see Section 4) is most efficient for $\nu \le 100$. For larger values of ν the table-supplied version BTPE^t is a little bit faster. Both methods have nearly the same set-up costs, but BRUE^t is simpler. Consequently, a combined procedure BIN/BRUE^t with cut-off point at $\nu^* = 5$ can be recommended, whenever *speed* and *simplicity* are essential, provided that one is prepared to store a table of $\ln k!$.

5.3 Hypergeometric distribution

For the hypergeometric case only one competing uniformly fast algorithm is known: Algorithm H2PE (uniform-exponential rejection of Kachitvichyanukul and Schmeiser, 1985), which was developed for $\nu = m - \max(0, n-N+M) \ge 10$, where $m = \lfloor (n+1)(M+1)/(N+2) \rfloor$ is the mode of the distribution. H2PE is sligthly faster than our algorithm HRUE (defined for $\nu \ge 1$) in the fixed parameter case, but its initialization of constants is about twice slower than that of HRUE. Additionally H2PE occupies more space (867 words versus 507 words). Therefore the combination HIN/HRUE ($\nu \le 3/\nu > 3$), supported with a double precision function for $\ln k!$, would be a good choice for a *fast* and *compact* sampling routine.

Probably the *fastest* and *simplest* hypergeometric sampling method could be offered by the combined generator HIN^t/HRUE^t ($\nu \le 3/\nu > 3$) at the cost of a long double precision table for values of $\ln k!$.

REFERENCES

Ahrens, J.H. and Dieter, U. (1982). Computer generation of Poisson deviates from modified normal distributions. *ACM Transactions on Mathematical Software 8*, 163-179.

Ahrens, J.H. and Dieter, U. (1989). A convenient sampling method with bounded computation times for Poisson distributions. *American Journal of Mathematical and Management Sciences* (to appear).

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

Chen, H.C. and Asau, Y. (1974). On generating random variates from an empirical distribution. *AIIE Transactions 6*, 163-166.

Devroye, L. (1986). *Non-uniform random variate generation.* Springer, New York.

Kachitvichyanukul V. and Schmeiser, B.W. (1985). Computer generation of hypergeometric random variates. *Journal of Statistical Computation and Simulation 22*, 127–145.

Kachitvichyanukul, V. and Schmeiser, B.W. (1988). Binomial random variate generation. *Communications of the ACM 31*, 216–222.

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

Kinderman, A.J. and Monahan, J.F. (1980). New methods for generating Student's t and gamma variables. *Computing 25*, 369–377.

Monahan, J.F. (1987). An algorithm for generating chi random variables. *ACM Transactions on Mathematical Software 13*, 168–172.

Schmeiser, B.W. and Kachitvichyanukul (1981). Poisson random variate generation. Research Memorandum 81–4. School of Industrial Engineering, Purdue University, West Lafayette, Indiana.

Stadlober, E. (1989a). Binomial random variate generation: A method based on ratio of uniforms. *American Journal of Mathematical Management Sciences* (to appear).

Stadlober, E. (1989b). Sampling from Poisson, binomial and hypergeometric distributions: Ratio of uniforms as a simple and fast alternative. Mathematisch–Statistische Sektion 303. Forschungsgesellschaft Joanneum, Graz, Austria.

Walker, A.J. (1977). An efficient method for generating discrete random variables with general distributions. *ACM Transactions on Mathematical Software 3*, 253–256.

AUTHOR'S BIOGRAPHY

ERNST STADLOBER is an assistent professor at the Institute of Statistics, Graz University of Technology, Austria. He received his diploma degree (Dipl.-Ing.) in Technical Mathematics and his Dr. techn. in Statistics from Graz University of Technology in 1975 and 1983 respectively. Recently he was appointed to Univ.-Dozent in Probability Theory and Statistics. His current research interests are random variate generation, biostatistics and graphical methods in statistics. He is member of the International Association For Statistical Computing and The Biometric Society.

Ernst Stadlober
Institute of Statistics
Graz University of Technology
A–8010 Graz, Lessingstraße 27, Austria
(0043)316–70616478
statistik@rech.tu–graz.ada.at