Appl. Statist. (1979),
28, *No.* 2, *pp.* 152–157

Efficient Simulation of the von Mises Distribution

By D. J. Best and N. I. Fisher

CSIRO Division of Mathematics and Statistics, P.O. Box 52, North Ryde, Australia

[Received December 1977. Final revision November 1978]

Summary

An algorithm is given to simulate samples from the von Mises distribution. A wrapped Cauchy density is used as an envelope to give an acceptance–rejection method which is both simple to program and fast for all values of the concentration parameter of the von Mises distribution.

Keywords: SIMULATION; VON MISES DISTRIBUTION; ACCEPTANCE–REJECTION METHOD; WRAPPED CAUCHY DISTRIBUTION

1. Introduction

The von Mises distribution for points on a circle is analogous, in many respects, to the normal distribution of points on a line and appears to have been first used by von Mises (1918) to study deviations of atomic weights from integer values. Since then it has been applied, for example, to directions of surface fault lines in geology and in studies of wildlife movement. Thus the von Mises distribution has become important in the statistical theory of directional data.

Its properties, and detailed references to applications, are well documented in the book and the review paper by Mardia (1972, 1975). One particular application appearing in the literature is the interesting study by Kendall (1974) on bird navigation, in which simulated von Mises variables are used in a model for the distribution of angular errors when a bird is lost to sight over the horizon.

However, because much of the statistical theory associated with the distribution is analytically intractable recent studies, such as those by Fisher and Willcox (1978) on the distribution of test statistics for circular data, and by Collett (1978) on problems relating to outliers, have relied on generation of pseudo-random observations from the von Mises distribution.

An angular random variable Θ has the von Mises distribution VM (μ_0, κ) if its probability density function (p.d.f.) has the form

$$\phi(\theta) = \{2\pi I_0(\kappa)\}^{-1} \exp\{\kappa \cos(\theta - \mu_0)\}, \quad -\pi \leqslant \theta < \pi, \quad \kappa > 0, \quad -\pi \leqslant \mu_0 < \pi,$$

where $I_p(\kappa)$ is the modified Bessel function of the first kind and order p, $p = 0, 1, \ldots$. Without loss of generality μ_0 will be taken as zero. The problem of efficient generation of a pseudo-random observation from VM $(0, \kappa)$ is made difficult by the following characteristics of the distribution:

(i) Its distribution function does not have a closed form.

(ii) There is no simple transformation $g(\Theta) = \Theta'$ say, such that Θ' has the VM $(0, \kappa')$ distribution, $\kappa' \neq \kappa$. (Hence it is not possible to concentrate on efficient generation of a random variable from VM $(0, 1)$ and transform to VM $(0, \kappa)$.)

(iii) There are no characterizations of the von Mises distribution, as yet available, which can be usefully employed, nor are there any convenient transformations of other distributions.

The following sections will examine this problem of efficient generation of pseudo-random observations from $\mathrm{VM}(0,\kappa)$. Section 2 will consider two previously suggested methods each of which has certain deficiencies, Section 3 will suggest some alternatives, Section 4 will consider in detail the most promising of these and finally Section 5 will compare timings of a number of the methods.

2. Previous Methods

There are very few published suggestions for simulating the von Mises distribution. Two methods are discussed here.

2.1. *Mardia's Approximate Method*

Mardia (1972, pp. 66–67) suggests using pseudo-random variates from $\mathrm{WN}(0, V_0)$, the wrapped normal distribution with mean zero and variance V_0 chosen to be $-2\ln\{I_1(\kappa)/I_0(\kappa)\}$. This approximate method is good in the extreme cases as $\kappa \to 0$ and as $\kappa \to \infty$ but the situation for intermediate κ values is not so clear.

In his work on the closeness of the von Mises distribution and the wrapped normal, Stephens (1963) compared the two distribution functions for $\theta = \frac{1}{40}\pi(\frac{1}{40}\pi)\pi$ and found that the maximum absolute difference between them, denoted by $D(V;\kappa)$, was small if the wrapped normal had a variance V close to V_0. For each κ the best V was found numerically and, as this is somewhat laborious, it is more convenient just to use V_0 as Mardia suggests. Table 1 indicates that if $D(V_0;\kappa)$ values are calculated rather than $D(V;\kappa)$ values then the differences between the two distribution functions are greater than 0·001 even for $\kappa = 20$. The $D(V_0;\kappa)$ values given in Table 1 are for $\theta = \frac{1}{400}\pi(\frac{1}{400})\pi$ and hence the actual maximum difference between the two distribution functions is at least as great as that shown.

Table 1

Approximate maximum difference $D(V_0;\kappa)$ and the value θ (in radian measure) which gives this difference

κ	$D(V_0;\kappa)$	θ
0·1	0·0002	0·77754
0·5	0·004	0·75398
1·0	0·012	0·70686
2·0	0·016	0·58119
3·0	0·012	0·47124
4·0	0·008	0·40055
5·0	0·006	0·35343
10·0	0·003	0·24347
20·0	0·001	0·16493

In some applications it is important that the pseudo-random observations be correct in the tails. Table 2 indicates that the probability of a wrapped normal random variate with variance V_0 being greater than a given von Mises percentile may be smaller than is desirable even for κ as large as 20.

Thus, for some applications, use of Mardia's approximate method would not be appropriate even for $\kappa = 20$. This conclusion would be even stronger for cruder normal approximations. Just how big or small κ should be for the approximate method to be useful is a subjective decision depending on the application. A possible alternative is the use of an exact method.

TABLE 2

Probability that a random WN $(0, V_0)$ *variate exceeds a certain percentile of the* VM $(0, \kappa)$ *distribution, for various percentiles and various values of* κ (V_0 *defined in text*)

	Percentiles of the von Mises distribution			
κ	0·05%	0·5%	2·5%	5%
0·1	0·00050	0·00499	0·02493	0·04988
0·5	0·00046	0·00458	0·02298	0·04654
1·0	0·00032	0·00320	0·01741	0·04040
2·0	0·00005	0·00087	0·01729	0·04732
3·0	0·00001	0·00168	0·02209	0·05081
4·0	0·00003	0·00287	0·02346	0·05084
5·0	0·00010	0·00348	0·02391	0·05059
10·0	0·00030	0·00436	0·02453	0·05023
20·0	0·00040	0·00470	0·02478	0·05011

2.2. *Seigerstetter's Exact Method*

The second method is that proposed by Seigerstetter (1974). Unfortunately this method employs a very crude envelope function which is reasonable only for $\kappa < 2$. The acceptance ratio (see Abramowitz and Stegun, 1965, p. 952) for this method is $\{2\pi\phi(0)\}^{-1}$, or $e^{-\kappa} I_0(\kappa)$, which tends to zero as $\kappa \to \infty$. For $\kappa = 0{\cdot}0$ the acceptance ratio is unity, for $\kappa = 0{\cdot}5$ it is 0·645, for $\kappa = 1{\cdot}0$ it is 0·466 and for $\kappa = 10{\cdot}0$ it is 0·128. Such a decrease in acceptance ratio results in a rather slow, albeit simple, computer algorithm. Hence, although this method has the advantage of being exact when compared with the approximate wrapped normal method it has the disadvantage of being costly to use on the computer. The next section will examine other exact alternatives.

3. EXACT ALGORITHMS BASED ON THE ENVELOPE-REJECTION METHOD

A modified version of the classical acceptance–rejection method is as follows. Let $f(x)$ be the p.d.f. of a random variable X which is to be sampled. Let Y be a random variable with p.d.f. proportional to $g(x)$, an upper envelope for $f(x)$ (i.e. $g(x) \geqslant f(x)$, all x), and let U be a $U(0,1)$ random variable. If (y, u) is a realization of (Y, U), y is accepted as a realization of X if $f(y)/g(y) > u$. The distribution of the accepted values is then exactly the required distribution. However, for the method to be efficient it must (a) be easy to generate a realization, y, (b) have an acceptance ratio nearer unity than zero and (c) be easy to evaluate $f(y)/g(y)$.

For the von Mises p.d.f. $\phi(\theta)$, we have considered the following types of envelopes:

(i) *Wrapped normal p.d.f.* $\omega(\theta)$. The difficulty with using an envelope proportional to $\omega(\theta)$ is that

$$\omega(\theta) = \sum_{j=-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{(\theta+2\pi j)^2}{\sigma^2}\right) \Big/ (\sigma(2\pi)^{\frac{1}{2}})$$

has to be evaluated each time a rejection test is made. That is, point (c) above is not satisfied.

(ii) *Piecewise linear envelope*. With this type of envelope, the interval $(-\pi, \pi]$ is partitioned into a convenient number of sub-intervals, and the envelope is linear within each sub-interval. A difficulty here is the need to calculate constants involving the modified Bessel function for each different value of κ. In fact, the partition may need to be altered according to the value of κ. Thus, condition (a) is not satisfied.

(iii) *Polynomial envelope.* Use of the inequality $a+b\theta > \exp(\cos\theta)$ for $a = 2{\cdot}8$, $b = 1{\cdot}8/\pi$, $-\pi \leqslant \theta < 0$, leads to the envelope $\pi(\theta) = (a+b\theta)^{\kappa}/\{2\pi I_0(\kappa)\}$. Whilst this method is reasonable for small κ, the acceptance ratio, $\{2\int_{-\pi}^{0}\pi(\theta)\,d\theta\}^{-1}$, equals $\{\pi b(\kappa+1)\,I_0(\kappa)\}/(a^{\kappa+1}-1)$, which tends, albeit slowly, to zero as κ tends to infinity. In general, then, condition (b) is not satisfied.

(iv) *Cardioid p.d.f.* A possible envelope is $\psi(\theta) = (1+2\rho\cos\theta)/\{2\pi I_0(\kappa)(1+2\rho)\exp(-\kappa)\}$ for $|\rho| < \frac{1}{2}$ but as the acceptance ratio $\{\int_0^{2\pi}\psi(\theta)\,d\theta\}^{-1}$ equals $\{(1+2\rho)\,I_0(\kappa)\}\exp(-\kappa)$, which tends to zero as κ tends to infinity, condition (b) is not satisfied. Further, it is not easy to generate realizations from a cardioid distribution and hence condition (a) is not satisfied either.

(v) *Wrapped Cauchy p.d.f.* Use of this envelope will be discussed in the next section.

4. An Algorithm using the Wrapped Cauchy Distribution

Define

$$\gamma(\theta) = (2\rho/\kappa)\exp\{\kappa(1+\rho^2)/2\rho-1\}/\{(1+\rho^2-2\rho\cos\theta)(2\pi I_0(\kappa))\}, \quad -\pi \leqslant \theta < \pi,$$

$$\rho = (\tau-(2\tau)^{\frac{1}{2}})/2\kappa, \quad \tau = 1+(1+4\kappa^2)^{\frac{1}{2}}.$$

Then $\gamma(\theta)$ is proportional to the p.d.f. of a wrapped Cauchy distribution (Mardia, 1972, p. 57). In the Appendix it is shown that $\gamma(\theta)$ is an upper envelope for $\phi(\theta)$ and that the choice of ρ maximizes the acceptance ratio. The distribution function of the wrapped Cauchy distribution is $C(\theta) = (2\pi)^{-1}\cos^{-1}[\{(1+\rho^2)\cos\theta-2\rho\}\{1+\rho^2-2\rho\cos\theta\}^{-1}]$, $-\pi \leqslant \theta < \pi$, and hence pseudo-random samples from this distribution are easily obtained. The algorithm is as follows. Let u_1, u_2 and u_3 be pseudo-random observations from $U(0, 1)$. Then, if it is understood that new observations u_1, u_2 and u_3 are used each time lines 1, 2 or 4 are executed, the successive steps of the algorithm are as follows:

0. Set $r = (1+\rho^2)/(2\rho)$.
1. Set $z = \cos(\pi u_1)$, $f = (1+rz)/(r+z)$, $c = \kappa(r-f)$.
2. If $c(2-c)-u_2 > 0$ go to step 4.
3. If $\ln(c/u_2)+1-c < 0$ return to step 1.
4. Set $\theta = [\operatorname{sign}(u_3-0{\cdot}5)]\cos^{-1}(f)$.

Step 2 depends on the bound, $e^x \geqslant 1+x$, and is intended to avoid, at least some of the time, the use of the logarithm function in step 3. The acceptance ratio is

$$\left\{\int_{-\pi}^{\pi}\gamma(\theta)\,d\theta\right\}^{-1} = \{(1-\rho^2)\,I_0(\kappa)\}/[(2\rho/\kappa)\exp\{\kappa(1+\rho^2)/2\rho-1\}]$$

and hence the acceptance ratio tends to unity as κ tends to zero and, as $\kappa \to \infty$, tends to $(2\pi/e)^{-\frac{1}{2}} = 0{\cdot}6577\ldots$.

It may be preferable on some computers to avoid the use of the cosine function (as suggested, for example, in Marsaglia and Bray, 1964) by the replacement of step 1 with the following two steps:

(1a) Set $v = u_1-0{\cdot}5$, $w = u_4-0{\cdot}5$, $d = v^2$ and $e = w^2$.

(1b) If $d+e > 0{\cdot}25$ go to 1a otherwise set $t = d/e$, $z = (1-t)/(1+t)$, $f = (1+rz)/(r+z)$ and $c = \kappa(r-f)$.

Here u_4 is understood to be a new pseudo-random observation from $U(0,1)$ each time step 1a is executed. If this modification is made then the value of u_2 used in steps 2 and 3 can be taken to be $4(d+e)$.

5. Timings

Table 3 compares timings of the wrapped Cauchy method with those of the Seigerstetter and polynomial envelope methods. The timings were made on a CDC 7600 machine using a FTN 4·6 compiler and were based on a DO loop generating pseudo-random samples of size 10,000 by function call. Constants required by each routine were preset and saved between calls. Pseudo-random $U(0,1)$ samples and pseudo-random normal samples (for Mardia's approximate method) were obtained as in Cheng (1977). It is noted that it may be difficult to extrapolate these timings to other machines although broad conclusions should remain unchanged. Each of the three methods compared in Table 3 is easy to program as is Mardia's approximate wrapped normal method if it is assumed that V_0 can be calculated using the asymptotic expansion given as equation (3. 4.50) by Mardia, 1972, p. 63).

Table 3

Central processing time (μs) required on a CDC 7600 to generate one pseudo-random von Mises variate

	0·1	0·5	1·0	2·0	5·0	10·0	100·0
Seigerstetter	15·4	20·6	27·2	39·1	63·0	91·0	—
Polynomial	30·4	35·9	40·9	46·8	44·8	41·4	196·9
Best–Fisher	13·6	14·5	16·1	17·6	19·0	19·8	20·3

The wrapped Cauchy method is clearly superior to either of the other two exact methods in terms of timings and on the CDC 7600 was as fast as Cheng's (1977) algorithm GB for simulating random gamma samples. The approximate wrapped normal time for comparison with the exact method times given in Table 3 is 10·7 μs (for any κ). This suggests, bearing in mind previous qualifications, that the user may wish to use the approximate method for $\kappa > \kappa_0$ where κ_0 would probably be about 10·0 although this would vary with the application. The exact wrapped Cauchy method could, however, be used for all κ without increasing computer time greatly.

Appendix

The aim of this Appendix is to show that $\gamma(\theta)$ is the best upper envelope (for $\phi(\theta)$) which is proportional to the wrapped Cauchy p.d.f. Following the envelope-rejection procedure described, for example, in Cheng (1977) it is first necessary to find the global maximum of the function, $R(\theta)$, proportional to the ratio of the von Mises and wrapped Cauchy p.d.f.'s.

Thus take $R(\theta) = (1+\rho^2-2\rho\cos\theta)\exp(\kappa\cos\theta)$ where $-\pi \leqslant \theta < \pi$ and ρ is the constant of the wrapped Cauchy distribution function, $C(\theta)$. An optimal value of ρ will be determined below. Hence $R'(\theta) = 2\rho\sin\theta\exp(\kappa\cos\theta)-(1+\rho^2-2\rho\cos\theta)\kappa\sin\theta\exp(\kappa\cos\theta)$ and thus $R'(\theta)=0$ when $\sin\theta=0$ or when $\cos\theta=(1+\rho^2-2\rho/\kappa)/2\rho$. By examining $R''(\theta)$, we find that $R(\theta)$ has a local maximum value $R_1 \equiv (1-\rho)^2\exp(\kappa)$ at $\sin\theta = 0$ if

$$2\rho/(1-\rho)^2 < \kappa, \tag{1}$$

and a local maximum value $R_2 \equiv (2\rho/\kappa)\exp\{\kappa(1+\rho^2)/2\rho-1\}$ at $\cos\theta=(1+\rho^2-2\rho/\kappa)/2\rho$, if

$$2\rho/(1+\rho)^2 < \kappa < 2\rho/(1-\rho)^2. \tag{2}$$

Let ρ_0 and ρ_1 denote the roots of $2\rho/(1-\rho)^2=\kappa$ and $2\rho/(1+\rho)^2=\kappa$, respectively, which lie in $(0,1)$. Then (2) implies that $\rho_0<\rho<\min(\rho_1,1)$ where $\rho_1<1$ only if $\kappa<0{\cdot}5$. Notice that (for fixed κ) equations (1) and (2) define non-intersecting ranges of ρ.

In order to choose the best value of ρ each of the above maxima will now be examined to determine the value of ρ (in terms of κ) which minimizes the reciprocal of the acceptance ratio. For $\sin\theta = 0$ the reciprocal of the acceptance ratio, $A_1(\rho)$ say, is

$$A_1(\rho) = \frac{R_1}{2\pi I_0(\kappa)} \frac{2\pi}{1-\rho^2}\left\{\frac{1}{2\pi}\int_0^{2\pi}\frac{(1-\rho^2)\,d\theta}{1+\rho^2-2\rho\cos\theta}\right\} = \frac{\exp(\kappa)}{I_0(\kappa)}\frac{1-\rho}{1+\rho}$$

and this attains its minimum value at $\rho = \rho_0 = (\kappa+1-(1+2\kappa)^{\frac{1}{2}})/\kappa$ (cf. (1)). For

$$\cos\theta = (1+\rho^2-2\rho/\kappa)/2\rho,$$

$A_2(\rho)$, the reciprocal of the acceptance ratio is

$$A_2(\rho) = \frac{2\rho}{\kappa}\frac{1}{I_0(\kappa)}\frac{1}{(1-\rho^2)}\exp\{\kappa(1+\rho^2)/2\rho-1\}.$$

Thus,

$$\frac{\exp(\kappa)\,I_0(\kappa)}{2\kappa}A_2'(\rho) = \frac{1}{\rho(1-\rho^2)^2}(2\rho+2\kappa\rho^2+2\rho^3-\kappa-\kappa\rho^4)\exp\{\kappa(1+\rho^2)/2\rho\}.$$

The root of $-\kappa\rho^4+2\rho^3+2\kappa\rho^2+2\rho-\kappa = 0$ which lies in the interval $(0, 1)$ can be shown to be $\rho_* = \{\tau-(2\tau)^{\frac{1}{2}}\}/2\kappa$, $\tau = (4\kappa^2+1)^{\frac{1}{2}}+1$, and it can be shown that $\rho_0 < \rho_* < \rho_1$. Further, as $A_2''(\rho_*) > 0$ it follows that the stationary value at $\rho = \rho_*$ is a minimum. Finally, we have to decide between the possible choices $\rho = \rho_0$ and $\rho = \rho_*$. It is readily seen that $A_2(\rho_0) = A_1(\rho_0)$. Hence, $A_2(\rho_*) < A_1(\rho_0)$, that is, the reciprocal of the acceptance ratio in the algorithm is minimized by choosing $\rho = \rho_*$, and $\gamma(\theta) = R_2/\{(1+\rho^2-2\rho\cos\theta)(2\pi I_0(\kappa))\}$ is the appropriate envelope for this ρ.

References

Abramowitz, M. and Stegun, I. G. (1965). *Handbook of Mathematical Functions.* New York: Dover Publications.

Cheng, R. C. H. (1977). The generation of gamma variables with non-integral shape parameter. *Appl. Statist.*, **26**, 71–75.

Collett, D. (1978). Some problems related to outliers on the line and on the circle. Ph.D. Thesis, University of Hull.

Fisher, N. I. and Willcox, M. E. (1978). A useful decomposition of the resultant length for samples from von Mises–Fisher distributions. *Commun. Statist.*, B7(3), 257–267.

Kendall, D. G. (1974). Pole-seeking Brownian motion and bird navigation. *J. R. Statist. Soc.* B, **36**, 365–417.

Mardia, K. V. (1972). *Statistics of Directional Data.* London: Academic Press.

—— (1975). Statistics of directional data. *J. R. Statist. Soc.* B, **37**, 349–393.

Marsaglia, G. and Bray, T. A. (1964). A convenient method for generating normal variables. *SIAM Rev.*, **6**, 260–264.

Seigerstetter, J. (1974). Discussion of Kendall, D. G. (1974). Pole-seeking Brownian motion and bird navigation. *J. R. Statist. Soc.* B, **36**, 411–412.

Stephens, M. A. (1963). Random walk on a circle. *Biometrika*, **50**, 385–390.

Von Mises, R. (1918). Über die "Ganzzahligkeit" der Atomgewicht und verwandte Fragen. *Physikalische Z.*, **19**, 490–500.