LOCKHEED

80-10224

ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

JSC-16670 NASA CR-

MAY 9 1980 160656

"Made available under NASA sponsorship in the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof."

Ref: 643-7957
Contract NAS 9-15800
Job Order 73-302

TECHNICAL MEMORANDUM

A NOTE ON THE GENERATION OF POISSON-DISTRIBUTED RANDOM NUMBERS WITH LARGE MEAN

By

George R. Terrell

Approved By: T. C. Minter
T. C. Minter, Supervisor
Techniques Development Section

(E80-10224) A NOTE ON THE GENERATION OF POISSON-DISTRIBUTED RANDOM NUMBERS WITH LARGE MEAN (Lockheed Engineering and Management) 11 p HC A02/MF A01 CSCL 12A N80-27773
Unclas
G3/43 00224

April 1980 LEMSCO-14851

1830 NASA ROAD 1, Houston, Texas 77058 • Tel (713) 333-5411

A Subsidiary of Lockheed Corporation

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
JSC-16670		

4. Title and Subtitle	5. Report Date
A Note on The Generation of Poisson-Distributed Random Numbers with Large Mean	April 1980
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
George R. Terrell	LEMSCO-14851
9. Performing Organization Name and Address	10. Work Unit No.
Lockheed Engineering and Management Services, Inc. 1830 NASA Road 1 Houston, Texas 77058	63-2307-3302
	11. Contract or Grant No. NAS 9-15800
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Exact methods for computer generation of Poisson pseudo-random variables tend to be expensive for large values of the parameter. This memorandum suggests an approximate method far more accurate in those cases than the usual normal approximation.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Pseudo-random numbers, simulation, Poisson distribution, goodness of fit, discrete random variables, computer simulation	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified		

*For sale by the National Technical Information Service, Springfield, Virginia 22161

JSC Form 1424 (Rev Nov 75) NASA — JSC

CONTENTS

TABLES

Table **Page**

1. INTRODUCTION

The computer simulation of picture-element (pixel) data sets in remote sensing problems is an important method for testing proposed methods of data analysis. In such simulations the distribution of the number of pixels to be assigned to a rare signature or crop type can be modeled by a Poisson process. Unfortunately, Poisson pseudo-random generators for moderate mean rates (~50) tend to be either slow or inaccurate. This technical memorandum proposes an alternative method of generation to partially overcome these problems.

2. APPROXIMATELY POISSON PSEUDO-RANDOM VARIABLES

A Poisson-distributed random variable x with probability $e^{-\lambda}\lambda^x/x$ has mean and variance λ. Because of the importance of Poisson processes, computational procedures for generating pseudo-random Poisson-distributed numbers have received some attention in the literature (Snow 1968, Schaffer 1970). The methods described there work quite satisfactorily for λ small, but involves a sequence of calculations of duration order λ. Thus, as λ becomes large, the time required to generate a single Poisson variate may be prohibitive. The IMSL program library (IMSL 1979) which implements the above procedures replaces them with a normally distributed random variable with λ mean and variance for $\lambda>50$. As will be seen later, this approximation is not particularly good.

The criterion for a good approximation to a particular distribution will be the ∞-metric

$\max_x |F_{approx}(x)-F(x)|$ where $F(x)$ is the cummulative distribution function of the random variable to be approximated. Since satisfactory procedures exist to generate normal pseudo-random variables, we will use approximations of the following form:

$$\text{If } Z \overset{\text{dist.}}{\sim} \text{Normal } (0,1), \text{ then } x_{approx} = T(\mu+\sigma Z)$$

where T is an appropriate one-one transformation and μ and σ are constants depending on T and λ. Thus, if ϕ is the normal (0,1) cumulative distribution function,

$$F_{approx} = \phi\left[(T^{-1}(x)-\mu)/\sigma\right]$$

The Edgeworth expansion (Abramovitz and Stegun 1970) of the function $T^{-1}(x)$ (where x is Poisson distributed) shows that the highest order correction term to the normality of $T^{-1}(x)$ is the skewness $\gamma_1 = \mu_3/\sigma^3$ (where μ_3 is the third central moment) times a function of x independent of T. Thus, we wish to find a T such that $T^{-1}(x)$ has skewness as small as possible.

3. POWER TRANSFORMATIONS

The family of functions T^{-1} which we will explore will be the powers x^a, a>o, which are all defined since x Poisson ≥ 0. This choice of family is suggested by the well-known "variance-stabilizing" transformation $\sqrt{x}$ whose variance is constant to order λ^{-1}. This has been useful in analyzing Poisson-count data since after transformation the separate counts may be treated as homoscedastic.

This corresponds to our case a = 1/2. To estimate the skewness of x^a, x Poisson, we write

$$E(x^a) = \lambda^a E\left(\left(\frac{x}{\lambda}\right)^a\right)$$

$$= \lambda^a E\left(\left(1 + \frac{x-\lambda}{\lambda}\right)^a\right)$$

and using the binomial theorem

$$= \lambda^a E\left(1 + a\frac{x-\lambda}{\lambda} + \frac{a(a-1)}{2}\left(\frac{x-\lambda}{\lambda}\right)^2 + \frac{a(a-1)(a-2)}{6}\left(\frac{x-\lambda}{\lambda}\right)^3\right.$$

$$\left. + \frac{a(a-1)(a-2)(a-3)}{24}\left(\frac{x-\lambda}{\lambda}\right)^4 + \ldots\right)$$

then using standard results about the central moments of a Poisson variable

$$= \lambda^a + \frac{a(a-1)}{2}\lambda^{a-1} + \frac{a(a-1)(a-2)(3a-5)}{24}\lambda^{a-2} + \ldots$$

Manipulation of this expansion gives us

$$\mu_{x^a} = \lambda^a + \frac{a(a-1)}{2}\lambda^{a-1} + \ldots$$

$$\sigma_{x^a} = a\lambda^{a-1/2} + 3/4a(a-1)^2\lambda^{a-3/2} + \ldots$$

$$\gamma_{1x^a} = a(3a-2)\lambda^{-1/2} + \ldots$$

Thus, by choosing a = 2/3, x^a is asymptotically unskewed as λ becomes large.

4. A GENERATION ALGORITHM

The random number generation scheme suggested by this result is as follows:

Generate a normal (0,1) pseudo-random variate Z. Multiply it by

$$\sigma = 2/3\lambda^{1/6} + 1/18\lambda^{-5/6} \text{ and add } \mu = \lambda^{2/3} - 1/9\lambda^{-1/3}$$

then raise the result to the 3/2-power. The Poisson variate x is taken to be the non-negative integer nearest to this value.

In table 1, the ∞-metric error for this approximate scheme is tabulated for various values of λ, along with the error for the usual untransformed ($a = 1$) normal approximation. It is apparent that the 3/2-scheme is enormously more accurate; for λ as small as 2 it is already better than the usual procedure at $\lambda = 50$. Also, as λ increases the 3/2-scheme becomes better as λ^{-1}, whereas the old scheme becomes better as $\lambda^{-1/2}$. Prichard (1980) has implemented this scheme on a programmable calculator with very little difficulty.

The disadvantages of the scheme include the extra computation involved in calculating μ and σ, which may be negligible if a number of variates are to be calculated, and the overhead involved in taking a 3/2 power for each variate. The higher accuracy may well compensate for this. In mixed exact-and-asymptotic for small and large λ schemes, speed may actually be enhanced because the asymptotic approximation can be tolerated for much smaller λ's.

TABLE 1.—COMPARISON OF MAXIMUM ERRORS IN TWO APPROXIMATIONS TO THE POISSON DISTRIBUTION

	Max Error In	
λ	Normal Approx.	$(\text{Normal})^{3/2}$ Approx.
2	.044	.0077
5	.0290	.00293
10	.0208	.00144
15	.0170	.00095
20	.0148	.000695
25	.0132	.000556
30	.0121	.000457
40	.0105	.000341
50	.000938	.000271
75	.00766	.000178
100	.00664	.000132
150	.00495	.000072

5. EXTENSIONS

As a further application of the results of this memorandum, we may construct a goodness-of-fit test for categories that are believed to have independent Poisson processes as generators. Let O_i, $i = 1, \ldots, n$ be the observed counts and E_i be the expected counts (λ's). Then

$$\sum_{i=1}^{n} \frac{\left(O_i^{2/3} - \mu_{2/3}(E_i)\right)^2}{\sigma_{2/3}^2(E_i)} \sim \chi_n^2$$

where the approximation is much closer for small E_i's than is the Pearson χ^2. Linear constraints, however, are no longer linear so this test warrants further investigation.

6. REFERENCES

1. Abramovitz, M.; and Stegun, I. A.: Handbook of Mathematical Functions. Dover p. 935, 1970.

2. Prichard, H. M.: Personal communication, 1980.

3. Schaffer, H. E.: Algorithm 369, Generator of Random Numbers Satisfying The Poisson Distribution. Communications of the ACM 13 (1), p. 49, 1970.

4. Snow, R. H.: Algorithm 342, Generator of Random Numbers Satisfying the Poisson Distribution. Communications of the ACM, 11 (12), p. 819, 1968.