数理解析研究所講究録
第498巻 1983年 1-12

乱数生成のための
プログラム・パッケージ(R-PACK)の試作

筑波大学 逆瀬川浩孝 (Hirotaka SAKASEGAWA)

0. はじめに.

統計的シミュレーションに使われる乱数生成用のプログラムパッケージとしては、次のような条件を満足していなくてはならない。すなわち、よく使われる分布に従う乱数の生成プログラムを含み、一つ一つの乱数はできるだけ正確に、かつ高速に生成されることが必要である。汎用と呼ばれる計算機には、科学計算用としてのサブプログラムライブラリが具っていて、そこには必ずと言ってよい程、乱数生成用のプログラムが入っている。これらも、標題のようなパッケージの一種であるが、それらの多くは、分布の種類も少なく、生成アルゴリズムも最新の成果が取り入れられていない。したがって上記のような条件をみたしているプログラム・パッケージはあまり見あたらない。

ここでは、いくつかの主要な確率分布に従う乱数について

、それぞれに最良と思われる生成アルゴリズムをプログラム化したパッケージ（R-PACKと呼ぶ）の試作例を報告する。

1．R-PACK の概要．

乱数生成用のプログラム・パッケージR-PACKは，各種乱数の効率よい生成アルゴリズムをFORTRAN サブプログラムの形で実現させたものである。その主要な目的はいうまでもなく統計的シミュレーションにおける良質の乱数源を提供することにあるが，それに加えて，乱数生成技術の最新の到達レベルの文書化ということも重要な目的の一つである。
R-PACKを使って生成できる乱数の従う分布は，現在のところ以下の通りである。

(1) 一様分布
(2) 標準正規分布
(3) ガンマ分布（指数分布を含む）
(4) ベータ分布
(5) 多次元正規分布
(6) 一般の離散分布
(7) ポアソン分布
(8) 二項分布
(9) ラテン方格行列

R-PACK の特徴は、正確さと、生成速度の高速性にある。一様分布以外の分布に従う乱数は、一様乱数を適当なアルゴリズムに従って変換することによって得られる。例えば、逆関数が存在する分布関数$F(x)$に従う乱数は、一様乱数$u_1, u_2, \cdots$を使って、$F^{-1}(u_1), F^{-1}(u_2), \cdots$によって得られる。一様分布に従う確率変数を$U$とした時、$F^{-1}(U)$は$F(x)$に従う確率変数となるから、この変換アルゴリズムは正確なものであるということができる。R-PACK はこのような意味で、すなわち、一様乱数として、一様分布に従う確率変数の実現値を考えた時、得られる乱数は与えられた分布に従う確率変数の実現値とみなせるという意味で、正確な生成アルゴリズムだけを採用し、近似を使う方法は取り上げなかった。近似法を使う理由は、アルゴリズムが簡単であるか、生成速度が速いかのいずれかであるが、前者についてはパッケージとしては考慮する必要がないこと、後者については正確なもので、十分に高速なものが得られること、などから、それらの理由が、正当な根拠にはならないからである。

第二の特徴である高速性は、統計的シミュレーションで推定の精度を良くするための欠かすことのできない要件である。この種のパッケージのみたすべき条件の一つとして、使用する計算機がなんであれ正常に働くという意味での汎用性が

求められるが、乱数生成のパッケージの場合は、高速性を犠牲にしてまで汎用性を追求する必要はない、というのが筆者の考えである。すなわち、計算機のハードウェア・ソフトウェアの知識を使うことによって乱数の生成速度が飛躍的に増大するならば、汎用性に固執すべきではない。実際、一様乱数の生成アルゴリズムとして需要の多い乗算合同法を汎用性を持たせてプログラムすることは、かなり余分な仕事を必要とし、使用する計算機を限定した効率よいプログラムに比べれば、生成速度は相当に落ちる。R-PACK は 高速性を実現するために、2の補数表示による 32 ビット語の計算機で動くFORTRANによってコード化されているので、他の方式の計算機で使用する場合には若干の変更が必要となる。

2. アルゴリズム

各乱数の生成アルゴリズムについては、一般論も含めて、文献[1] にあるので、ここでは概略のみにとどめる。

2.1 一様乱数

簡便な方法として乗算合同法、長周期を得る方法として最大周期列法を採用している。乗算合同法は

(1) $x_{n+1} \equiv \lambda x_n \pmod{P}$ （ただし $P = 2^{32}$）

の形である。$\lambda \equiv \pm 5 \pmod 8$ の時、この数列の周期は 2^{30}

である。最大周期列法は

(2) $x_n = x_{n-p} \oplus x_{n-q}$ （但し $p=521,\ q=32$）

の形である。ここで $a \oplus b$ は、a, b を二進展開して、各ビットごとに2を法として足し算を行った結果を表し、x_n は31ビットの0ではない整数とする。初期値として $x_{-p+1}, x_{-p+2}, \cdots, x_0$ のための $31 \times p$ ビット乱数が必要となるが、これには、(1)式を使って最初の x_n の最上位ビットだけを取り出して p 個のビットを作り、(2)式を2進一桁の数列に適用して、それらの p ビットを初期値として残りのビットを生成している。

2.2 一様分布以外の分布に従う乱数

この場合は、層別棄却法、あるいは合成棄却法と呼ばれる方法で一様乱数を変換することが多い。目的の乱数の従う分布の密度関数を $f(x)$、$f(x)$ と x 軸とによって囲まれる領域を A とする。A の中の点をランダムに生成し、その x 座標の値を並べると、それらは $f(x)$ に従った乱数になっている。直接 A の点をランダムに生成することは一般にむづかしいので次のような工夫を考える。$\forall x$ で $f(x) \leq g(x)$ となるような関数 $g(x)$ をとり（$g(x)$ は連続でなくても、有界でなくてもよくその積分が存在するだけでよい）、$g(x)$ と x 軸とに囲まれる領域を D、D の一つの層別を $D_1, D_2, \ldots$ （$\bigcup_j D_j = D,\ D_i \cap D_j = \phi,\ i \neq j$）

とする。$D_1, D_2, \ldots$ としては、それらの中のランダムな点が容易に生成できるようなものを選ぶものとする。$D_1, D_2 \ldots$ から、それらの面積に応じてランダムに点を生成し、それらの点のうちAに含まれないものを棄て、Aに含まれている点だけを集めると、それらはAの中でランダムに散らばっているので、これらの点から$f(x)$に従う乱数が生成できる。これが層別棄却法の考え方で、正規分布、ガンマ分布、ベータ分布に従う乱数は、この方法によって生成することができる。ある点PがAに含まれている$(P \in A)$か否かを調べるためには$f(x)$の値を計算する必要があり、これは一般にはかなり時間がかゝるのが普通である。そこで、A又はA^cに完全に含まれる領域$D_1', D_2', \ldots$で、ある点がそこに含まれるか否かを簡単に調べることができるようなものがもしあれば、$P \in A$を調べる前に$P \in D_1'$, $P \in D_2', \ldots$という一連のテストによって、$P \in A$か否かをある程度判定することができることになり、棄却法の時間を短縮することができる。これをしぼり出し法と呼ぶ。層別棄却法はしぼり出し法を併用することによって乱数生成の速度を大幅に改善できる場合が多い。

3．パッケージ使用上の注意

3.1. 汎用性がない最大の原因は、乱数生成の高速性を実現

するために、各乱数の基本となる一様乱数を、32 ビット語の計算機の特徴を使った乗算合同法によって生成していることにある。したがって、この方式ではない計算機で使用する時は、その部分だけ修正すれば、殆んどの場合正常に動く筈である。例えば、2の補数表示であるが、一語のビット数が違っている場合には、式(1) で得られる1からP-1までの整数乱数を(0,1) 区間の実数乱数になおす為の定数を変えるだけでよく、プログラム上は DATA 文をさしかえるだけである。

3.2　式(1) による一様乱数は周期が 2^{30} しかないことから多次元における過疎性という好ましくない性質がある。しかし、この性質は、パラメータλとして、λを2進表示した時の1または0の個数が極端に少なくないようなものを選べば、殆んど実害を及ぼさないということが経験的に確かめられているので、R-PACKではこの方法を採用している。とはいえ、どんな場合にでも最適となるようなパラメータは存在しないので、一つのパラメータによる乱数列を用いたシミュレーション結果だけから結論を導くのは楽観的すぎるかもしれない。重要な計算では、複数個のパラメータを使って別々の乱数列を作り出し、それぞれの乱数列を用いた独立なシミュレーションを行って結果に偏りがないかどうかを検討した方

がより安全である。そのために、このパッケージでは、乱数の初期値（x_0、補遺のプログラムではKR）同様、パラメータ（λ、同じくLD）の値も、使用者が与えることができるようになっている。LDの値λ_0が0でない時に限って、パラメータの値λ_{old}を次の式に従って新しくしている。

$$\lambda_{new} = 8\left(|\lambda_0| + \left[\frac{\lambda_{old}}{8}\right] \pmod{2^{23}}\right) + 11111005$$

λが新しくなると同時にLDの値は0におきかわるので、次に使用者がLDを別の値におきかえない限りは、このパラメータに従って一様乱数が生成される。なお、パラメータは、使用する乱数の個数が多く、例えば周期の十分の一（約1億）を越えるというような場合を除いては、計算の途中で最初にセットした値を変えるべきでない。

3.3　一様乱数は、普通の規模のシミュレーションでは乗算合同法で十分であるが、これは、サブルーチンとして使わずin-lineとして使用すべきである。FUJITSU M-200/OS-IVでは、一つの乱数を得るのにサブルーチンの場合は約5μ秒、in-lineの場合は約1μ秒であった。これは、シミュレーションの精度（推定量の分散比）が、前者に比べて後者が約5倍になることを意味している。

実際の使用に際してのサブルーチンの呼び出し方については補遺を参照のこと。

4. おわりに。

ここで報告された乱数生成のためのプログラムパッケージR-PACKはまだ完全なものではなく、計画の第一段階を終えたにすぎない。確率分布の種類も少なく、文書化もまだこれからである。今後は、これらの点を徐々に改良しつつ、冒頭に書かれたような目的にそったパッケージの完成を目指したい。

参考文献

[1] 逆瀬川浩孝「モンテカルロシミュレーション」応用統計学10巻1号(1981), 3-21。

補遺. R-PACKのリストの一部

```
C  PROGRAM F1.   UNIFORM RANDOM NUMBER GENERATOR
C  ==========    =======
C   METHOD : MULTIPLICATIVE CONGRUENTIAL METHOD
C             ( VALID ONLY FOR 32-BIT-WORD MACHINES
C              WITH 2'S COMPLEMENT (E.G. FUJITSU M-200, ETC.) ).
C   COMMENT : HAD BETTER USE THE ROUTINE IN IN-LINE MODE.
C        BEFORE ANY EXECUTABLE STATEMENTS, R32, LMD AND KR ARE SET AS
C             DATA R32/.23283064E-9/, LMD/39894229/, KR/111111111/.
C        EACH TIME A UNIFORM RANDOM NUMBER IS NEEDED, THE FOLLOWING
C        STATEMENT IS PUT:
C             KR = LMD*KR.
C        < KR > IS SEEN AS AN INTEGER-VALUED RANDOM NUMBER
C        BETWEEN -2**31 AND 2**31.
C        THEN < KR*R32+0.5 > IS A REAL-VALUED RANDOM NUMBER ON (0,1).
C
C     FUNCTION  UNIFRN ( KR, LD )
C               ======
C          KR ... SEED ( ANY ODD NUMBER OF LESS THAN 10 FIGURES )
C          LD ... RANDOM FACTOR OF MULTIPLIER ( IF NOT ZERO )
C                 ( ANY NUMBER OF 7 FIGURES )
```

```
C   PROGRAM F2.   UNIFORM RANDOM NUMBER GENERATOR
C   ==========    =======
C    METHOD : MAXIMUM SEQUENCE METHOD.
C    REFERENCE : FUSHIMI,M. AND TEZUKA,S. (1982),
C                       APPLIED STATISTICS (JAPAN) 10.3.
C
       FUNCTION   UNIFRM ( J )
C                 ======
C           J : POINTER OF THE TABLE
C              ( J MUST BE EQUAL TO -1 FOR THE FIRST TIME )
C          MSEQTB : NAME OF COMMON BLOCK OF 521 WORDS
C              ( MUST BE PREPARED BY USER AS FOLLOWS :
C                   COMMON /MSEQTB/ MTAB(521) ).
C
C
C   PROGRAM F3.   NORMAL RANDOM NUMBER GENERATOR
C   ==========    ======
C    METHOD : STRATIFIED REJECTION METHOD.
C    REFERENCE : SAKASEGAWA,H. (1978),
C                       ANNALS INST. STATIST. MATH. A30.2.
C
       FUNCTION   SNORRN ( KR, LD )
C                 ======
C          KR,LD : PARAMETERS OF BUILT-IN UNIFORM RANDOM NUMBER GENERATOR
C                  ( MULTIPLICATIVE CONGRUENTIAL METHOD )
C            KR ... SEED ( ANY ODD NUMBER OF LESS THAN 10 FIGURES )
C            LD ... RANDOM FACTOR OF MULTIPLIER ( IF NOT ZERO )
C                  ( ANY NUMBER OF 7 FIGURES )
C
C
C   PROGRAM F4.   GAMMA RANDOM NUMBER GENERATOR
C   ==========    =====
C     CASE 1) A>.35.
C      METHOD : WILSON-HILFERTY'S APPROXIMATION WITH (X/A)**(1/3).
C      REFERENCE : MARSAGLIA,G. (1977), COMPUTER MATH. APPLIC. 3.
C     CASE 2) A<.35.
C      METHOD : STRATIFIED REJECTION METHOD.
C      REFERENCE : AHRENS,J.H. AND DIETER,U. (1974), COMPUTING 12.
C     COMMENT : SUPERIOR TO THE PROGRAM F5
C               WHEN A PARAMETER CHANGES EVERY TIME.
C
       FUNCTION   GAMARM ( A, KR, LD )
C                 ======
C          A      : SHAPE PARAMETER OF GAMMA DISTRIBUTION
C              ( F(X) = X**(A-1) / GM(A) * EXP(-X) )
C          KR,LD : SAME AS THE PROGRAM F3
C
C
C   PROGRAM F5.   GAMMA RANDOM NUMBER GENERATOR
C   ==========    =====
C     CASE 1) A>.55.
C      METHOD : WILSON-HILFERTY'S APPROXIMATION WITH X**(1/3).
C      REFERENCE : NIKI,N. (1979), MANUSCRIPT.
C     CASE 2) A<.55. SAME AS THE PROGRAM F4.
C     COMMENT : SUPERIOR TO THE PROGRAM F4
C                 IN CASE OF A CONSTANT PARAMETER.
C
```

```
      FUNCTION   GAMARN ( A, KR, LD )
                 ======
C
C         A      : SHAPE PARAMETER OF GAMMA DISTRIBUTION
C             ( F(X) = X**(A-1) / GM(A) * EXP(-X) )
C         KR,LD : SAME AS THE PROGRAM F3
C
C
C   PROGRAM F6.  EXPONENTIAL RANDOM NUMBER GENERATOR
C   ==========   ===========
C      METHOD : INVERSE FUNCTION METHOD
C    COMMENT : HAD BETTER USE THE ROUTINE IN IN-LINE MODE.
C
      FUNCTION   EXPORM ( KR, LD )
C                ======
C         KR,LD : SAME AS THE PROGRAM F3
C
C
C   PROGRAM F7.  EXPONENTIAL RANDOM NUMBER GENERATOR
C   ==========   ===========
C    METHOD : RUN LENGTH TEST METHOD.
C    REFERENCE : FORSYTHE,G. (1972), MATH. COMP. 26.120.
C
      FUNCTION   EXPORN ( KR, LD )
C                ======
C         KR,LD : SAME AS THE PROGRAM F3
C
C
C   PROGRAM F8.  BETA RANDOM NUMBER GENERATOR
C   ==========   ====
C    METHOD :
C     CASE 1) A=1 AND B=1. MULTIPLICATIVE CONGRUENTIAL METHOD.
C     CASE 2) A=1 AND B<>1 OR A<>1 AND B=1. INVERSE FUNCTION METHOD.
C     CASE 3) OTHERWISE. STRATIFIED REJECTION METHOD.
C    REFERENCE : SAKASEGAWA,H. (1983),
C                    ANNALS INSTIT. STATIST. MATH. B35.1.
C
      FUNCTION   BETARN ( A, B, KR, LD )
C                ======
C         A,B    : SHAPE PARAMETERS OF BETA DISTRIBUTION
C                  ( F(X) = X**(A-1) * (1-X)**(B-1) / BE(A,B) )
C         KR,LD : SAME AS THE PROGRAM F3
C
C
C   PROGRAM F9.  MULTIDIMENSIONAL NORMAL RANDOM VECTOR GENERATOR
C   ==========   ======================
C    METHOD : MARGINAL DECOMPOSITION METHOD.
C    REFERENCE : HURST,R.L. AND KNOP,R.E. (1972), COMM. ACM 15.5.
C
      SUBROUTINE   MNORRN ( R, N, V, L, IE, KR, LD )
C                  ======
C         R(N)    : GENERATED RANDOM VECTOR  ( 1 < N < 50 )
C         V(L,N)  : VARIANCE-COVARIANCE MATRIX ( CONTENTS - DISABLED )
C         IE      : SINGULARITY CHECK  (0:INITIAL CALL, -1:SINGULAR
C                                        1:POSITIVE DEFINITE)
C         SNORRN(KR,LD) : FUNCTION SUBROUTINE SUPPLYING
C                              STANDARD NORMAL RANDOM NUMBER
C         KR,LD   : SAME AS THE PROGRAM F3
C
C
```

```
C   PROGRAM F10.   POISSON RANDOM NUMBER GENERATOR
C   ===========    =======
C    METHOD : MODIFIED TABLE LOOK-UP METHOD.
C    REFERENCE : FISHMAN,G.S. (1976), COMPUTING 17.
C
      FUNCTION   NPOSRN ( A, KR, LD )
C                ======
C           A      : PARAMETER OF THE DISTRIBUTION
C               ( P(N) = A**N * EXP(-A) / N! )
C           KR,LD : SAME AS THE PROGRAM F3
C
C
C   PROGRAM F11.   BINOMIAL RANDOM NUMBER GENERATOR
C   ===========    ========
C    METHOD : SIMPLE TABLE LOOK-UP METHOD.
C
      FUNCTION   NBIMRN ( N, P, KR, LD )
C                ======
C           N,P    : PARAMETERS OF THE DISTRIBUTION
C                    ( P(K) = C(N,K) * P**K * (1-P)**(N-K) )
C           KR,LD : SAME AS THE PROGRAM F3
C
C
C... PROGRAM F11 ... BINOMIAL RANDOM NUMBER GENERATOR
C   ===========    ========
C    METHOD : SIMULATION OF THE BERNOULLI TRIAL.
C
      FUNCTION   NBINRN ( N, P, KR, LD )
C                ======
C           N,P    : PARAMETERS OF THE DISTRIBUTION
C                    ( P(K) = C(N,K) * P**K * (1-P)**(N-K) )
C           KR,LD : SAME AS THE PROGRAM F3
C
C
C   PROGRAM F13.   ANY DISCRETE RANDOM NUMBER GENERATOR
C   ===========        ========
C    METHOD : ALIAS METHOD.
C    REFERENCE : WALKER,A.J. (1977), ACM TRANS. MATH. SOFTWARE 3.3.
C
      FUNCTION   NDISRN ( N, PP, NN, QQ, MM, KR, LD )
C                ======
C           PP(N),NN(N)  : TABLE OF PROBABILITIES
C                          ( PR( K=NN(J) ) = PP(J) )
C           QQ(N),MM(N)  : WORKING AREA FOR THRESHOLD VALUES AND ALIASES
C           KR,LD : SAME AS THE PROGRAM F3
C
C
C   PROGRAM F14.   RANDOM LATIN SQUARE GENERATOR
C   ===========             ===========
C    METHOD : DIRECT ASSIGNMENT.
C
      SUBROUTINE LATINS ( N, NE, ND, L, NCON, KR, LD )
C                ======
C           N(ND,*)  : MATRIX WHERE RANDOM LATIN SQUARE WILL BE PUT
C           NE       : DIMENSION OF A MATRIX
C           L(ND,*)  : WORK AREA
C           NCON(*)  : WORK AREA FOR TRIAL NUMBER
C           KR,LD    : SAME AS THE PROGRAM F3
```