数理解析研究所講究録
第 850 巻 1993 年 152-164

非一様乱数生成の近似正確化法

慶大理工　渋谷政昭（Masaaki SIBUYA）

グンベル分布，一般極値分布に従う乱数の，近似正確化法（exact－approximation method）による生成について述べる．また，この仕事の背景の事情について説明する．

1．近似正確化法

一様乱数 U を変換して，連続１次元分布関数 $H(\cdot)$ に従う乱数 Y を得る方法は数多く提案されている．たとえば，Devroye（1986）参照．その中で Marsaglia の近似正確化法（1984）は比較的に良い性質をもっている．この方法は次のような単純な事実に基いている．

分布関数 $H(y)$，$-\infty \le a < y < b \le \infty$，が確率密度関数 $h(y)$ をもつとする．$H(\cdot)$ の逆関数（確率点，quantile function）を近似する'簡単な'関数

$$g(u) \doteqdot H^{-1}(u), \qquad 0<u<1,$$

があれば $H(g(u)) \doteqdot u$，したがって $h(g(u))\, g'(u) \doteqdot 1$，

$0<u<1$, である.

もしも $g(\cdot)$ が単調であり, $g(0)=a$, $g(1)=b$ であるならば $H(g(u))$, $h(g(u))\,g'(u)$, $0<u<1$, はそれぞれ分布関数, 確率密度関数である. もしも $(0,1)$ 上の確率変数 X が分布関数 $H(g(\cdot))$ に従うとすると,

$$X \le v \iff g(X) \le g(v)$$

$$\Pr\{X \le v\} = H(g(v)) = \Pr\{g(X) \le g(v)\}$$

だから, $y=g(v)$ とおくと, 確率変数 $Y=g(X)$ の分布関数は

$$\Pr\{Y \le y\} = H(y), \qquad a<y<b,$$

となる. つまり,

$$X \sim F(v) := H(g(v)), \quad 0<v<1,$$

$$\iff \quad Y := g(X) \sim H(y), \qquad a<y<b,$$

であり,

$$f(v) = \frac{d}{dv}F(v) = h(g(v))\,g'(v), \qquad 0<v<1,$$

は1に近い. そこで

$$p := \inf_{0<v<1} f(v) \quad < 1$$

とおき, 確率密度関数 f を

$$f(v) = p\,\mathbb{1}[0<v<1] + (1-p)\,r(v),$$

$$r(v) := \frac{f(v)-p}{1-p}\,\mathbb{1}[0<v<1]$$

と, $(0,1)$ 一様分布と確率密度 $r(\cdot)$ の混合分布として表現できる.

そこで一様乱数 U を，分布関数 $H(\cdot)$ に従う乱数 Y に変換する，次の近似正確化法の算法が得られる．

算法 EA

(1) 一様乱数 U を生成する．

(2a) $U<p$ ならば $X:=U/p$ とおき (3) へ．

(2b) $U>p$ ならば，確率密度 $r(v)$ をもつ乱数 X を（たとえば棄却法で）生成する．

(3) $Y:=g(X)$ とする． □

この方法は p が 1 に近く $g(\cdot)$ が簡単な関数のときに有効である．

2. グンベル分布乱数の生成

区間 (a,b), $-\infty\leq a<b\leq\infty$, に打切ったグンベル分布（2重指数分布）の分布関数は

$$H(y;a,b)=(H(y)-A)/(B-A), \quad a<y<b,$$

$$H(y)=\exp(-\exp(-y)), \quad A=H(a),\ B=H(b),\quad 0<A<B<1,$$

である． $y=0$ （$H(y)=1/e$）の近くの近似を求めるために，$H(y;a,b)$ の逆関数

$$-\log(-\log(A+(B-A)u))$$

を，$(e^{-1}-A)/(B-A)$ で3次まで展開すると，

$$g_0(u)=a+\frac{b-a}{y_1-y_0}\left\{w\left(1+\frac{w^2}{6}\right)-y_0\right\},$$

$$w=e(A+(B-A)u)-1,\quad y_0=(eA-1)+(eA-1)^3/6,\quad y_1=(eB-1)+(eB-1)^3/6,$$

を得る．これを出発値として，3次式

$$g(u) = a + (b-a)(c_1 u + c_2 u^2 + c_3 u^3), \quad 0 < u < 1,$$
$$(g(0) = a, \ g(1) = b)$$

を，

$$p = \min_{0 \le u \le 1} f(u), \qquad (f(u) = h(g(u))\,g'(u))$$

を最大にするように定める．$f(0) = f(1)$ とすると，c_1, c_2, c_3 のうち自由なパラメータは1個だけであり，逐次修正により p の最大化ができる．

$A = 0.0659$，$B = 0.8659$ のときに

$$c_1 = 1.3429999, \quad c_2 = -1.6165521, \quad c_3 = 0.7254478$$

と選んで，$p = 0.8841164$ を達成できる．このときの，$f(u)$ のグラフを第1図に示す．

実用のためには，残りの区間についても計算しなければならない．これ以外の生成法については Sibuya (1984).

3．近似正確化法の拡張

第1節で述べたように算法EAの本質的な点は

$$X \sim H(g(\cdot)) \iff Y = g(X) \sim H(\cdot)$$

という関係式であり，$g \fallingdotseq H^{-1}$ と選ぶことにより，Xを一様乱数に近いものとすることができた．これを一般化すると，$H(\cdot) \fallingdotseq \Phi(\cdot)$ であり，分布関数 $\Phi(\cdot)$ に従う乱数の生成が容易であれば，関数 $g(v) \fallingdotseq v$ によって Xを Φ に近い乱

数とすることができる．

たとえば分布関数 $H_n(y)$, $-\infty \le a_n < y < b_n \le \infty$, が中心極限定理により，標準正規分布関数 $\Phi(\cdot)$ に近づくならば，$y_n = H_n^{-1}(u)$ を，$x = \Phi_n^{-1}(u)$ により

$$y_n(x) = c_{n1}x + c_{n2}x^2 + \cdots$$

と表わすことができ，$n \to \infty$ のとき $c_{n1} \to 1$, $c_{nj} \to 0$ $(j = 2, 3, \cdots)$ である．これは，いわゆる Cornish-Fisher 展開である．たとえば Barndorff-Nielsen and Cox (1989) 参照．

無限級数 $y_n(x)$ を簡単な関数 $g_n(x)$ で近似する．$g_n(-\infty) = a_n$, $g_n(\infty) = b_n$ で $g_n(\cdot)$ が単調であれば $H_n(g_n(x)) \doteqdot \Phi(x)$ は分布関数である．

$$f_n(x) := h_n(g_n(x))\, g_n'(x) \doteqdot \phi(x)$$

を

$$f_n(x) = p\phi(x) + (1-p) r_n(x), \quad -\infty < x < \infty, \quad n = n_0, n_0+1, \cdots$$

$$p = \inf_{n \ge n_0} \inf_{-\infty < x < \infty} f_n(x), \qquad r_n(x) = \frac{f_n(x) - p\phi(x)}{1-p},$$

という混合分布への分解が得られる．その算法は，次のより一般の場合について述べる．

分布関数の族 $H(\cdot;\theta)$, $\theta \in \Theta$ （Θ は R の閉区間），

$0 < H(y;\theta) < 1$, $-\infty \le a_0 < y < b_0 \le \infty$, $H(a_0+;\theta) = 0$, $H(b_0-;\theta) = 1$, の乱数 Y_0 を生成する．確率変数 X の分布関数を Φ：

$0<\Phi(x)<1$, $-\infty\le A<x<B\le\infty$, $\Phi(A+)=0$, $\Phi(B-)=1$, とする．関数 $g(\cdot;\theta)$ が単調増加で，$g(A_\theta+;\theta)=a_\theta$, $g(B_\theta-;\theta)=b_\theta$ ならば，$F(x;\theta)=F(g(x;\theta);\theta)$ は分布関数である．

$$f(x;\theta):=\frac{d}{dx}F(x;\theta),\quad \phi(x):=\frac{d}{dx}\Phi(x),$$

$$p:=\inf_{\theta\in\Theta}\ \inf_{A<x<B} f(x;\theta)/\phi(x),\quad r(x;\theta):=\frac{f(x;\theta)-p\phi(x)}{1-p},$$

とすると，混合分布への分解

$$f(x;\theta)=p\phi(x)+(1-p)r(x;\theta)$$

が得られる． p を大きくするように $g(\cdot;\theta)$ を選ぶ．

分布関数 $H(\cdot;\theta)$ をもつ乱数 Y_θ, $\theta\in\Theta$, を生成する一般近似正確化法の算法は次の通りである．

算法 GEA

(1) 一様乱数 U を生成する．

(2a) $U<p$ ならば，確率密度 ϕ をもつ乱数を生成して X とする．

(2b) $U>p$ ならば，確率密度 $r(\cdot;\theta)$ をもつ乱数を生成して X とする．

(3) $Y_\theta=g(X;\theta)$ とする．

この方法は $\phi()$ の乱数の生成が容易であり，関数 $g(\cdot;\theta)$ が簡単なときに有効である．

一般に a, b, g が θ に依存し，p を θ について最小化

しなければならない．Marsaglia (1984) は例としてガンマ分布

$$H(x;\theta) = \frac{1}{\Gamma(\theta)}\int_0^x t^{\theta-1}e^{-t}dt, \quad 0<x<\infty, \quad \theta_0<\theta<\infty,$$

の正規近似を議論しているが，$g(-\infty;\theta)=0$ が満たされていないために'正確化法'となっていない．

4．一般極値分布乱数の生成

一般極値分布

$$H(y;\tau) = \begin{cases} \exp(-(1+\tau y)^{-1/\tau}), & -\infty<y<-1/\tau, \ \tau<0 \\ \quad '' & -1/\tau<y<\infty, \ \tau>0 \\ \exp(-\exp(-y)), & -\infty<y<\infty, \ \tau=0 \end{cases}$$

に従う乱数 Y_τ $(\tau\neq 0)$ を生成したい．τ が 0 に近いときに限り，グンベル分布 $H(\cdot;0)$ に従う乱数 Y_0 が得られるという条件の下で変換する．$H(\cdot;0)$ の逆関数 $z_0=-\log(-\log u)$，による，$H(\cdot;\tau)$ の逆関数 $z=(-1/\tau)\cdot(1-(-\log u)^{-\tau})$ の展開式

$$z = z_0 + \frac{1}{2}\tau z_0^2 + \frac{1}{6}\tau^2 z_0^3 + \cdots$$

を修正した近似式

$$g(x) = \left(x+\frac{\tau}{2}x^2+\frac{\tau^2}{6}x^3\right)\Big/\left(1-\frac{\tau^3}{6}x^3\,\mathbb{1}[\tau x<0]\right)$$

を用いる．

$$g(x) = x+\frac{\tau}{2}x^2+\frac{\tau^2}{6}x^3, \quad \tau x>0$$

$$g(\infty) = -1/\tau \ (\tau<0), \quad g(-\infty) = -1/\tau \ (\tau>0)$$

である．

τの範囲を狭く $(-0.14,\ 0.34)$ と限ると，$p=0.9841$ を達成することができ，そのときの $R(x;\tau)$ は第2図のようになる.

5. 確率点関数による乱数生成

分布関数 $H(\cdot)$ の逆関数である確率点 (quantile) 関数 $y=H^{-1}(u)$ が簡単に計算できるならば，一様乱数 U を $Y=H^{-1}(U)$ と変換して，分布関数 $H(\cdot)$ をもつ乱数 Y を生成できる. ところで $(0,1)$ 一様乱数 U は実際には M 個 ($M\fallingdotseq 2^{30}$) の離散値 $j/(M+1)$, $1\leq j\leq M$, の上の一様分布であり，Y も次の離散点上の一様分布に従う:

$$\xi_j := H^{-1}(j/(M+1)),\quad 1\leq j\leq M,$$

この性質を乱数の粒子性 (glanularity) と呼ぶこともある.

この事実の影響が大きいのは，分布 H の裾が重い場合である. 以下では右裾だけを扱う. まず $M\gg 1$ のとき

$$\xi_{j+1}-\xi_j = H^{-1}\left(\frac{j+1}{M+1}\right)-H^{-1}\left(\frac{j}{M+1}\right)=\frac{1}{h(\xi_j)}\frac{1}{M+1}+O\left(\frac{1}{M^2}\right)$$

であるが，右裾の重い分布では

$$1/h(y) = 1/h(H^{-1}(u))$$

が $y\to\infty$ $(u\to 1)$ のときに非常に大きくなる. ただし，これを $M+1$ で割り，相対的な距離 $(\xi_{j+1}/\xi_j)-1$ を計算すると，それほど大きくはない.

グンベル分布，ワイブル分布の右裾の行動は指数分布に近

く，

$$1/h(H^{-1}(u)) \doteqdot 1/(1-u), \quad (\xi_M/\xi_{M-1})-1 \doteqdot (\log 2)/(\log M)$$

である．

確率変数 $Y_1,\dots,Y_n$ が独立で同一分布関数 $H(\cdot)$ に従うとき，$Z=\max(Y_1,\dots,Y_n)$ の分布関数は $H^n(\cdot)$ である．n個の乱数を生成して最大値を求めるのではなく $Z=H^{-1}(U^{1/n})$ により直接に生成する．このときの粒子性はどうなるか．

$$Z=H^{-1}(U^{\frac{1}{n}})\le z \iff H^n(Z)=U\le (H(z))^n$$

であるから $\Pr\{Z\le\xi_j\}=(j/(M+1))^n$，つまり Z のとる値は Y_1 と同じであるが，今度の場合は一様分布でなく，$\xi_j,\ j\doteqdot M$，の確率が大きくなる：

$$\Pr\{Z=\xi_j\}=\left(\frac{j}{M+1}\right)^n-\left(\frac{j-1}{M+1}\right)^n=\frac{n}{M+1}\left(1-\frac{(n-1)k}{M+1}+O\left(\frac{1}{M^2}\right)\right),\quad k=M+1-j.$$

第3図で，曲線はワイブル分布（べき乗パラメータ 2.5）の分布関数 H（実線）と H^n，$n=5$（点線）である．階段状の線は $M=20$ の場合の，ξ_j 上の一様分布の分布関数と，$(j/(M+1))^n$ である．$M=1000$ 以上になれば H の方は‘目に見える’差が無くなるが n/M が 10^{-3} の位であれば H^n の上側で‘目につく’差が生じる．

乱数の粒子性を避けるのに，Mを大きくするだけが解決ではない．むしろYの分布範囲を分割する方がよい．Yの区間 $[v_{i-1}, v_i]$ 上の条件付き分布関数は

$(H(y)-V_{i-1})/(V_i-V_{i-1}), \quad v_{i-1}<y<v_i, \quad V_i=H(v_i),$

であり，これに従う乱数は $H^{-1}((1-U)V_{i-1}+UV_i)$，$0\leq V_{i-1}<V_i\leq 1$，により生成できる．Yの分布範囲を $a=v_0<v_1<\cdots<v_K=b$ と分割し，$V_{i-1}<U\leq V_i$ のときに区間 (v_{i-1}, v_i) を選び，別の一様乱数 U^*により，$H^{-1}((1-U^*)V_{i-1}+U^*V_i)$ で乱数を生成すればよい．

古典的な棄却法や本稿の近似正確化法は計算速度を早めるのが主目的で，このような分割を前提としている．棄却法では，分割が細かく，必要な定数が増加するのが欠点であるが，近似正確化法により，この点を改善できる．

Ahrens and Dieter (1988) は，近似正確化法により，指数乱数，コーシー乱数を生成し，さらにこれらを組合わせて正規乱数を生成している．コーシー乱数の生成には，確率点関数 $\mathrm{Tan}^{-1}(\pi(u-1/2))$ の有理関数近似を用いているが $(-\infty, \infty)$ 全区間の近似を行なわず，粒子性を回避していない．

6．シミュレーションの設計

計算機が高速となり，$M=2^m$ ($m\doteqdot 30$) 個の乱数を使用するシミュレーションが可能となっている．現在'良い'とされている合同法（Park and Miller (1988) および Ferrenberg, et al. (1992) 参照）の乱数1周期を消費して

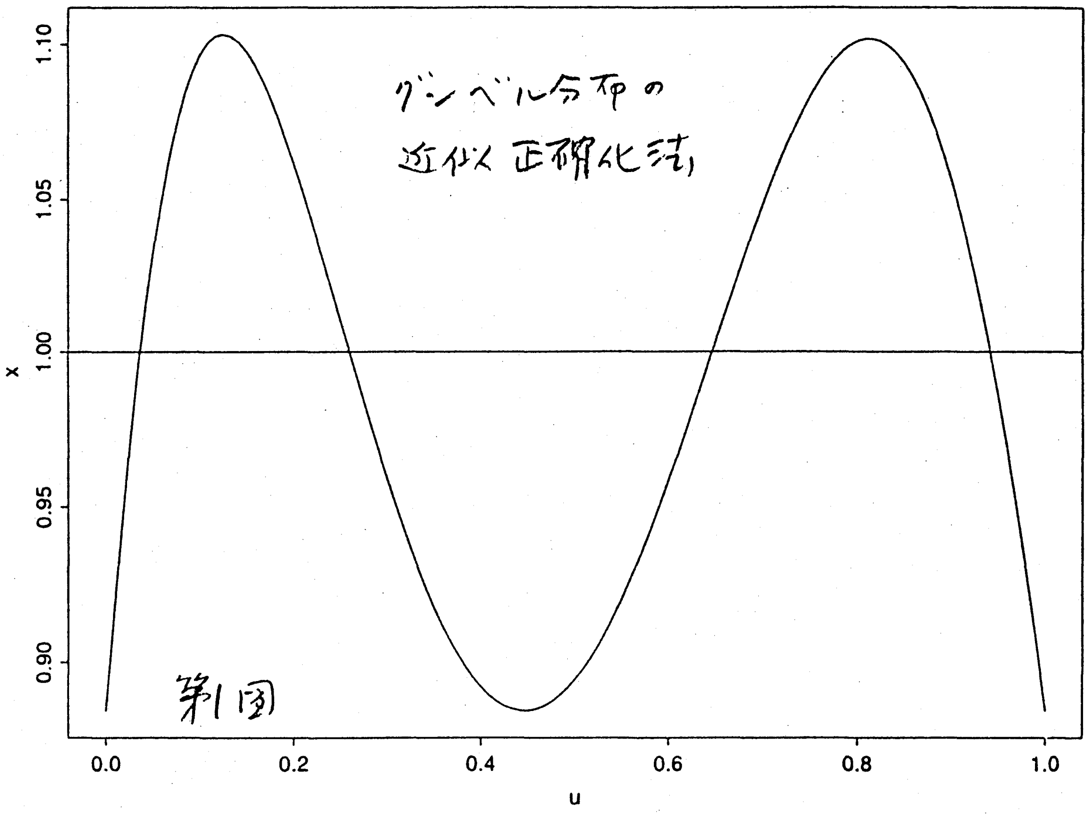
しまう．

シミュレーション反復回数を増してサンプリング誤差を減らすためには，第1に周期のより長い乱数を使用しなければならない．幸にその方向の成果が得られている．Fushimi (1990) の乱数は使用経験が積まれており，今特集の最近の研究成果も期待される．(L'Ecuyer (1990) の総合報告が，新しい方向について詳しいが，その後の成果も目ざましい．)一様乱数そのものを倍長とすることは必要なさそうである．

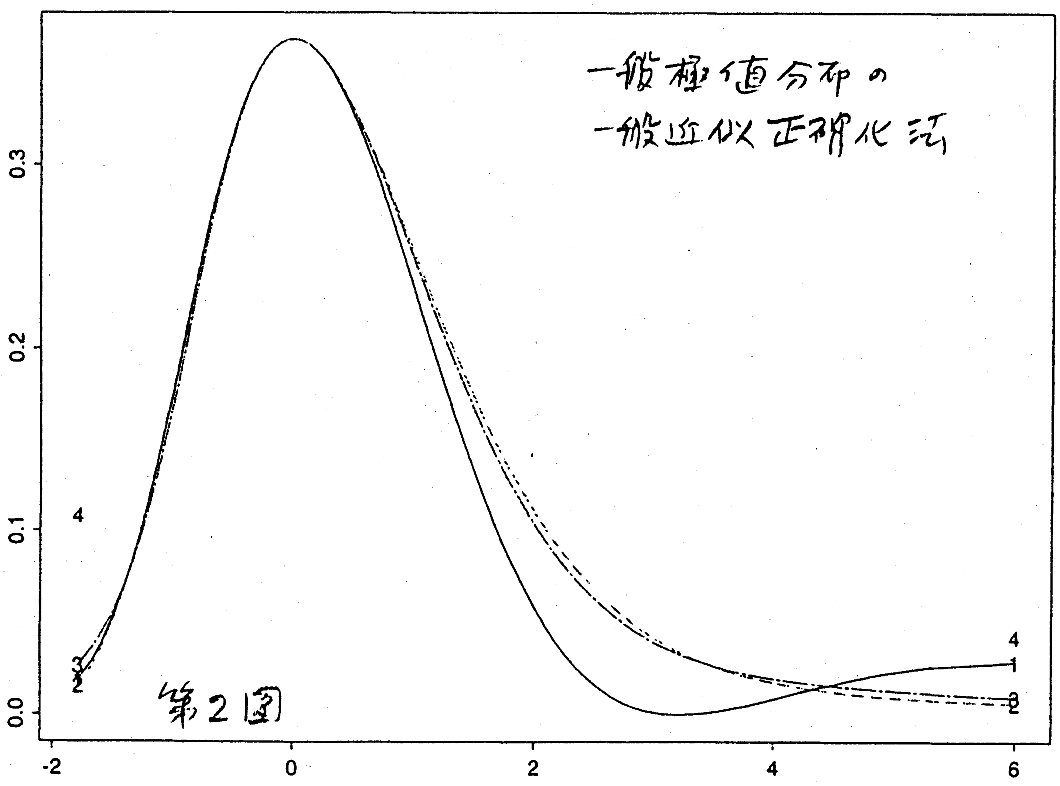
複雑な計算を行って，その結果が思わしくないとき，乱数生成法に不安が残ることは避けたい．

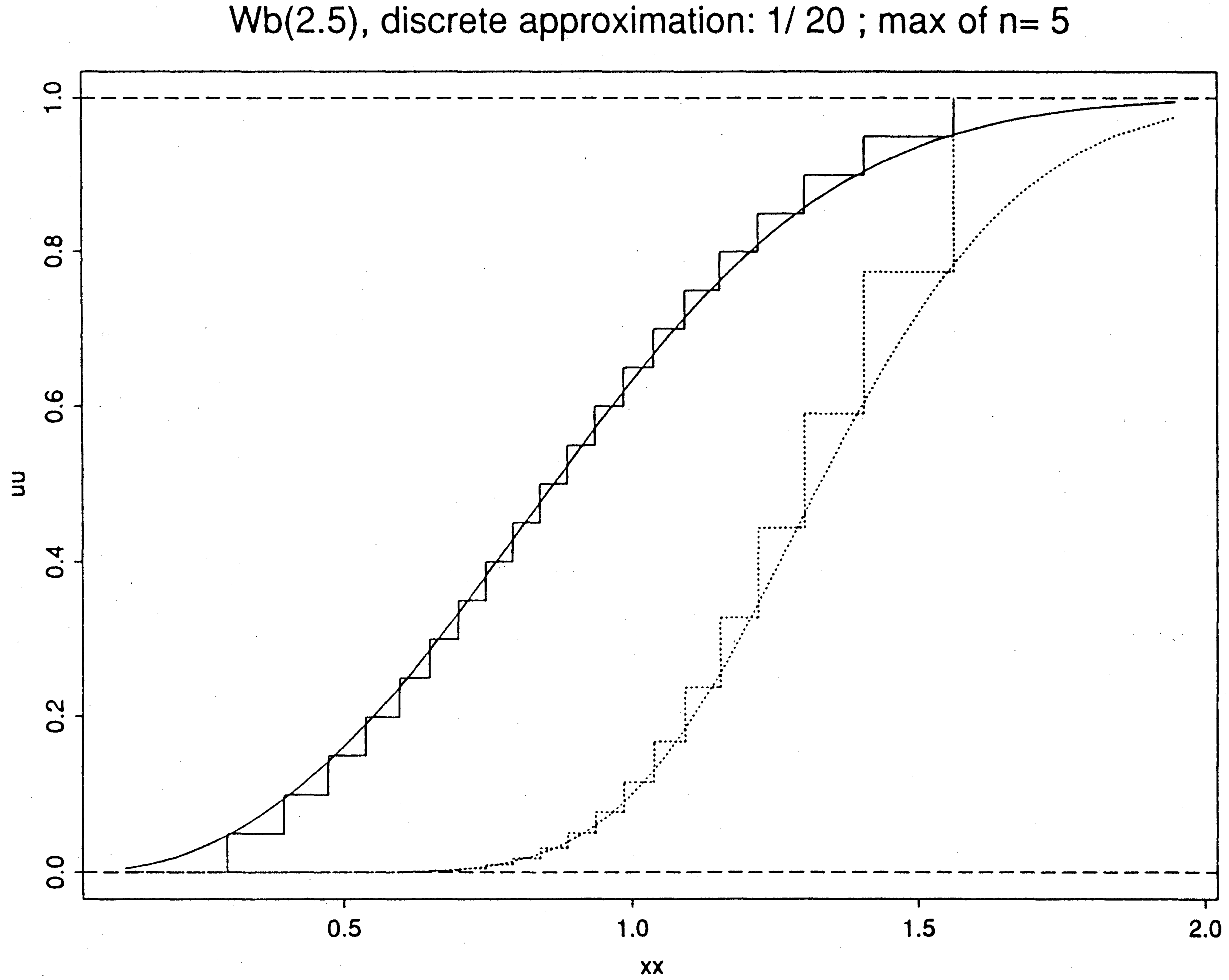
1. Ahrens, J.H. and Dieter, U. (1988) Efficient table-free sampling methods for the exponential, Cauchy, and normal distributions, *Commun. ACM* **31**, 1330-1337.
2. Barndorff-Nielsen, O.E. and Cox, D.R. (1989) *Asymptotic Techniques for Use in Statistics* , Chapman and Hall, London.
3. Devroye, L. (1986) *Non-Uniform Random Variate Generation*, Springer, New York.
4. L'Ecuyer, P. (1990) Random numbers for simulation, *Commun. ACM* **33-10**, 85–97.
5. Ferrenberg, A.M., Landau, D.P. and Wong, Y.J. (1992) Monte Carlo simulations: Hidden errors from 'good 'random number generators, *Physical Review Letters* **69-23**, 3382–3384.
6. Fushimi, M. (1990) Random number generation with the recursion $X_t = X_{t-3p} \oplus X_{t-3q}$, *Jour. Comp. Appl. Math.* **31**, 105–118.
7. Lewis, P.A., Goodman, A.S. and Miller, J.M. (1969) A pseudo-random number generator for the System/360, *IBM Syst. J.* **8**, 136–146.
8. Marsaglia, G. (1984) The exact-approximation method for generating random variables in a computer, *J. Amer. Statist. Assn.* **79**, 218–221.
9. Park, S.K. and Miller, K.W. (1988) Random number generators: good ones are hard to find, *Commun. ACM* **31**, 1192–1201. (See also, Technical Correspondenses, *Commun. ACM* **32**, 1020–1024.) (西村恕彦訳，bit 1993年4月 19–27 5月 12–20．)
10. Sibuya, M. (1984) Doubly exponential random number generators, J. Tiago de Oliveira (ed.), *Statistical Extremes and Applications*, 395–402, Reidel, Dordrecht.

h(g(u))g'(u)



r(x). tau= -0.1,-0.05,0.1,0.2 ; p= 0.995





第3図　ワイブル乱数の粒子性