# 二項分布からの正規乱数生成

中村　永友[1]　　土屋　高宏[2]

**要　旨**

注目する事象の生起確率 $p$ が0.5である独立試行を $n$ 回繰り返したとき，その関連する確率は二項分布 $B(n, p)$ をもとに計算できる．試行回数が十分大きいときに，この二項分布に関する何らかの確率を求める際には，古くから正規分布による近似が行われる．統計学のテキストでは必ずこの項目は触れられる事項でもある．本報告はこの近似の良さを利用して，二項分布に従う乱数から正規乱数を，高速かつ大量に生成するためのアルゴリズムを提案する．この方法はある種の近似を伴うため，数値実験を通して正規分布として判定される条件を提示する．

**キーワード**：正規乱数，一様乱数，二項分布，適合度検定．

## 1　はじめに

電子計算機が登場して以来，計算機上で数値実験研究やシミュレーション研究を行う際には，必ず任意の確率分布にしたがう疑似乱数が使われていて，同時にその生成法も数多く提案されている．さらに，大規模なシミュレーションを行う際には，確率分布にしたがう乱数が超大量に必要である．特定の確率分布にしたがう乱数を生成するには，質の良い一様乱数をいかに入手するかということがこれまでの研究の中心を占めてきた．物理乱数を個人でも入手できるようになったが，これが使えない場合には，より長周期の疑似一様乱数を用いることになる．現在はメルセンヌ＝ツイスター法といった優れた方法が，R をはじめとするかなり多くの数値計算が可能な環境で用いることができるようになったので，一様乱数に関する話題は，一段落している．

一方，一様乱数から特定の確率分布にしたがう乱数を得るためには，性質の良い確率分布では逆関数法等が利用できるが，そうでない場合には，棄却法や重点標本抽出法（Importance Sampling）等々の技法を使うことになる（四辻，2010など）．例えば，正規乱数を生成する方法は Box＝Muller 法（BM 法）が標準的であるが，中心極限定理（清水，1976）に依拠する一様乱数の複数個の和からも生成できる（12個の場合は，それらの和から 6 を引くと $N(0, 1)$ にしたがう）．しかし，超多数個を生成する際に少しでも高速に生成した方がよりよいことは容易に想像できる．超大量の連続型確率分布にしたがう乱数を得るために，目的の確率分布を近似した離散型確率分布にしたがう乱数を通して，目的の疑似乱数を生成する方法を提案した（Nakamura, 2015）．この方法は少なくとも BM 法よりも演算回数が少なく，同時に高速な方法である．離散型確率分布であるオイラリアン分布（土屋・中村，2009；Tsuchiya, 2015）を使って正規乱数を生成する方法については，Nakamura (2015)，中村・土屋 (2015, 2016, 2017*a*, 2017*b*) で報告されているが，本報告は，二項分布を使って正規乱数を生成する方法を提案する．

## 2　二進数の二項分布性

まず，必要な記号等の標記を示す：

$$\alpha：整数，\ \alpha \in [0, 2^n - 1],$$
$$\alpha_{10}：十進数表記，$$
$$\alpha_2：二進数表記．$$

このとき，二進数$\alpha_2$ は $n$ 桁の 0 と 1 の並び（bit sequence）として次のように表すことができる．

[1] 札幌学院大学 経済学部；nagatomo@sgu.ac.jp.
[2] 城西大学 理学部；takahiro@josai.ac.jp.

$$\alpha_2=(a_1a_2a_3\cdots a_k\cdots a_n)_2$$

$a_k$ ($k=1, \ldots, n$) は 0 か 1 の値をとる．一方，0 から $2^n-1$ までのすべての整数を二進数で表すと，

$$\left.\begin{aligned}
0_{10}&=(000\cdots000)_2\\
1_{10}&=(000\cdots001)_2\\
2_{10}&=(000\cdots010)_2\\
3_{10}&=(000\cdots011)_2\\
4_{10}&=(000\cdots100)_2\\
5_{10}&=(000\cdots101)_2\\
&\vdots\quad\vdots\quad\vdots\\
(2^n-2)_{10}&=(111\cdots110)_2\\
(2^n-1)_{10}&=(\underbrace{111\cdots111}_{n\text{桁}})_2
\end{aligned}\right\}2^n\text{個}\qquad(1)$$

となる．各二進数の中の 1 の総数は，順に

$$0, 1, 1, 2, 1, 2, \ldots, n-1, n$$

となる．これらから $0, 1, \ldots, n$ の数字の個数を集計すると，$\binom{n}{k}$，($k=0, 1, \ldots, n$) と一致する．0 を失敗，1 を成功とすると，任意の整数の二進数表記の0−1の並びは，ベルヌーイ試行の結果を並べたものと見ることができる．(1)式は，$n$ 回の試行を行うときすべての組み合わせを並べたものとなる．つまり，任意の整数の二進数の並びは $n$ 回の試行の結果とみることができ，その中の 1 の総数は成功回数であることから，$n$ 桁の二進数は二項分布をしていることになる．

二進数が二項分布をしているということの指摘は，Ahrens & Dieter (1974) にある．これを引用する形で，Knuth (1998) の演習問題にある (3.4.1節，27番)．しかし，それらには二項分布性に関する直感的な説明があるだけで，厳密な証明はない．

以上から，整数 $\alpha$ を二進数で表したときに，以下を命題として挙げることができる．

**命題 1**　区間 $[0, 2^n-1]$ 内の任意の 1 つの整数 $\alpha$ を二進数で表記した$\alpha_2$ において，任意の桁数における 1 (または 0 ) の出現確率は1/2である．

**命題 2**　$\alpha_2$ の中の 1 の総数を $X$ とすると，$X$ は，試行回数 $n$，生起確率1/2 (平均 $n/2$，分散 $n/4$) の二項分布 $B(n, 1/2)$ にしたがう．

**系**　任意の長さ$n<\infty$のビット列において，$[0, 2^n-1]$ のすべての整数の 2 進数の 1 の数の計数値は，二項係数 $\binom{n}{k}$，($k=0, 1, \ldots, n$) と一致する．

## 3　二項乱数から正規乱数への変換

二項分布において試行回数 $n$ が十分大きく，成功確率が $p=1/2$のとき，正規分布で近似できる (正規近似)．このことを利用して二項乱数を通して，正規乱数を生成する方法を提案する．

正規乱数生成の基本的考え方は，以下の通りである．区間 $[0, 2^n-1]$ から整数値の一様乱数 $\alpha$ の二進数表示 $\alpha_2$ の 1 の総数 $x$ は二項分布 $B(n, 1/2)$ をしている．二項分布の正規近似の性質を利用すると，$x$ の確率変数 $X$ は $N(n/2, n/4)$ にしたがう．したがって，

$$\frac{X-n/2}{\sqrt{n/4}}$$

と変換すれば，これは標準正規分布にしたがうことになる．この方法によって，生成した乱数は離散的なものであるため，連続化する必要がある．そこで，一つの方法として，その時点で利用した一様乱数 $\alpha$ は，$t=\alpha/(2^n-1)$の変換で区間 $[0, 1]$ の一様乱数となる．したがって，

$$\frac{(x+t-0.5)-n/2}{\sqrt{n/4}}$$

とすることによって，連続化と正規化ができる．

しかし，ここで行ったことは，二項分布をしている乱数があり，その整数の離散値に対して±0.5の範囲で (隣のビンを超えないように)，一様分布させているだけである (図 2 (a, b)を参照)．そこで，次の命題が正しいと仮定すると，この操作によって生成した全体の乱数は大域的には正規分布にしたがうことが保証される．

**命題 3**　任意の連続型確率分布 $f$ にしたがう乱数は，ある条件下では大域的には $f$ にしたがい，局所的には (任意の微小区間では) 一様分布にしたがう．

以下に変換アルゴリズムを示す．

**二項乱数から正規乱数への変換アルゴリズム**

**手順 1**　ビット数 $m$ を指定する．

**手順 2**　$r_0\in(0, 1]$ なる実数値をとる乱数を生成する．

**手順 3**　$r_1\in(0, 2^m]$ なる整数値をとる乱数を生成す

る．

**手順 4**　$r_1$ を 2 進数で表したときの 1 の個数を数え，その数を $b_1$ とする．

**手順 5**　$r_0 \leftarrow r_1/2^m$

**手順 6**　$b \leftarrow b_1 + r_0 - 0.5$

**手順 7**　$b \leftarrow (b - m/2)/\sqrt{m/4}$

**手順 8**　手順 3 に戻り，必要な個数の回数を繰り返す．

手順 4 の $b_1$ は二項分布にしたがう乱数となる．さらに手順 7 でできた $b$ が標準正規分布に従う乱数になる．また，この手順 2 と 3 で乱数を 2 回生成しているが，その理由は5.3節で説明する．

## 4　切断される分布の裾の確率

表 1 には本提案手法で正規乱数を生成したときに，分布の裾の最大値を標準化した値（$z$ 値）とその値から外側の確率の 2 倍（$p$ 値）を示す．同様な方法であるオイラリアン分布による正規乱数を生成したときの同様の値を同時に示す（Nakamura, 2015；中村・土屋, 2015）．この表からわかることは，ビット数（ビン数）を増やすほど分布の切断が平均から離れていくこと，正規分布への近似という意味ではオイラリアン分布のほうが二項分布より優れているため，裾のカバーする範囲が広いことがわかる．

正規乱数の標準的手法である BM 法は，32bit のコンピュータ（PC）で，裾の最大値（$z$ 値）は6.66で $p$ 値は $2.37\times10^{-11}$，64bit PC で，$z$ 値は 9.42 で $p$ 値は $4.54\times10^{-21}$ である．この値と比べると，提案手法は切断確率の制御が可能であることが指摘できる．また，表 1 内の数値を比較すると，二項分布の bit 数96と BM 法の64bit PC，bit 数48と32bit PC がほぼ対応

**表 1：分布の裾の切断と確率**

| bit 数 | 二項分布 | | オイラリアン分布 | |
|---|---|---|---|---|
| | $z$ 値 | $p$ 値 | $z$ 値 | $p$ 値 |
| 16 | 4.25 | $2.14\times10^{-5}$ | 6.72 | $1.80\times10^{-11}$ |
| 24 | 5.10 | $3.34\times10^{-7}$ | 8.31 | $9.27\times10^{-17}$ |
| 32 | 5.83 | $5.42\times10^{-9}$ | 9.64 | $4.99\times10^{-22}$ |
| 48 | 7.07 | $1.52\times10^{-12}$ | 11.88 | $1.56\times10^{-32}$ |
| 64 | 8.13 | $4.47\times10^{-16}$ | 13.75 | $5.13\times10^{-43}$ |
| 96 | 9.90 | $4.16\times10^{-23}$ | 16.88 | $6.02\times10^{-64}$ |
| 128 | 11.40 | $4.08\times10^{-30}$ | 19.52 | $7.45\times10^{-85}$ |
| 256 | 16.06 | $4.67\times10^{-58}$ | 27.66 | $2.18\times10^{-168}$ |
| 512 | 22.67 | $8.54\times10^{-114}$ | 39.15 | $2.64\times10^{-335}$ |

bit 数：オイラリアン分布の場合はビン数を表す．

**表 2：Box＝Muller 法との裾の最大値（$z$ 値）の比較**

| 基礎となる乱数生成法 | $z$ 値 | $p$ 値 |
|---|---|---|
| Box＝Muller 法 32bitPC | 6.66 | $2.37\times10^{-11}$ |
| 二項分布 42bit | 6.64 | $3.24\times10^{-11}$ |
| オイラリアン分布 16bin | 6.72 | $1.80\times10^{-11}$ |
| Box＝Muller 法 64bitPC | 9.42 | $4.54\times10^{-21}$ |
| 二項分布 87bit | 9.43 | $3.93\times10^{-21}$ |
| オイラリアン分布 31bin | 9.49 | $2.27\times10^{-21}$ |

し，オイラリアン分布の bin 数32と BM 法の64bit PC，bin 数16と32bit PC がほぼ対応している．より厳密に比較したものを表 2 に示す．

## 5　数値実験

### 5.1　実験 1 ：利用可能なビット数とデータ数

提案したアルゴリズムによって，どのような乱数が生成されるのかを検証する．

bit 数を$2^2$～$2^9$（4～512），データ数を$10^1$～$10^5$ として，それぞれ適当な間隔の組み合わせで疑似乱数を生成し，そのデータに対して適合度検定（Anderson Darling 検定）を行い，$p$ 値を計算した．これを1000回繰り返して等高線を描いたのが図 1 である．

この結果の図より，左上の方向に $p$ 値が0.05より小さい領域が存在する．この条件下では正規乱数と見なせないということである．それ以外の領域は有意水準5%では棄却できない，すなわち正規乱数と見なせる条件であることがわかる．

同様の適合度検定を何種類か行っているが，どの検

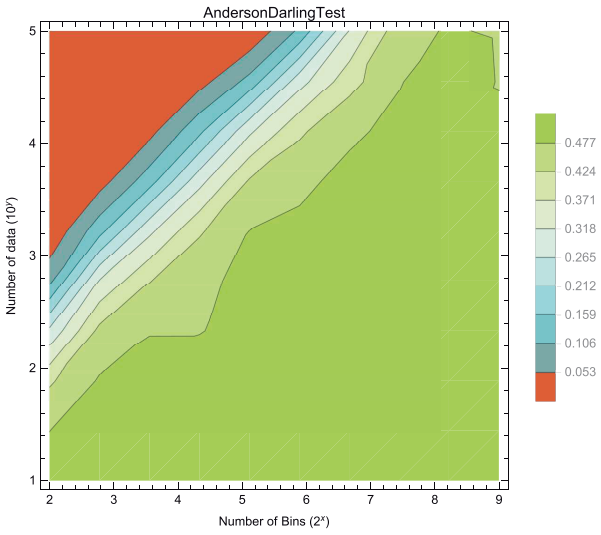



**図 1：ビット数とデータ数による仮説検定の $p$ 値**
縦軸：データ数（$10^y$），横軸：ビット数（$2^x$）．

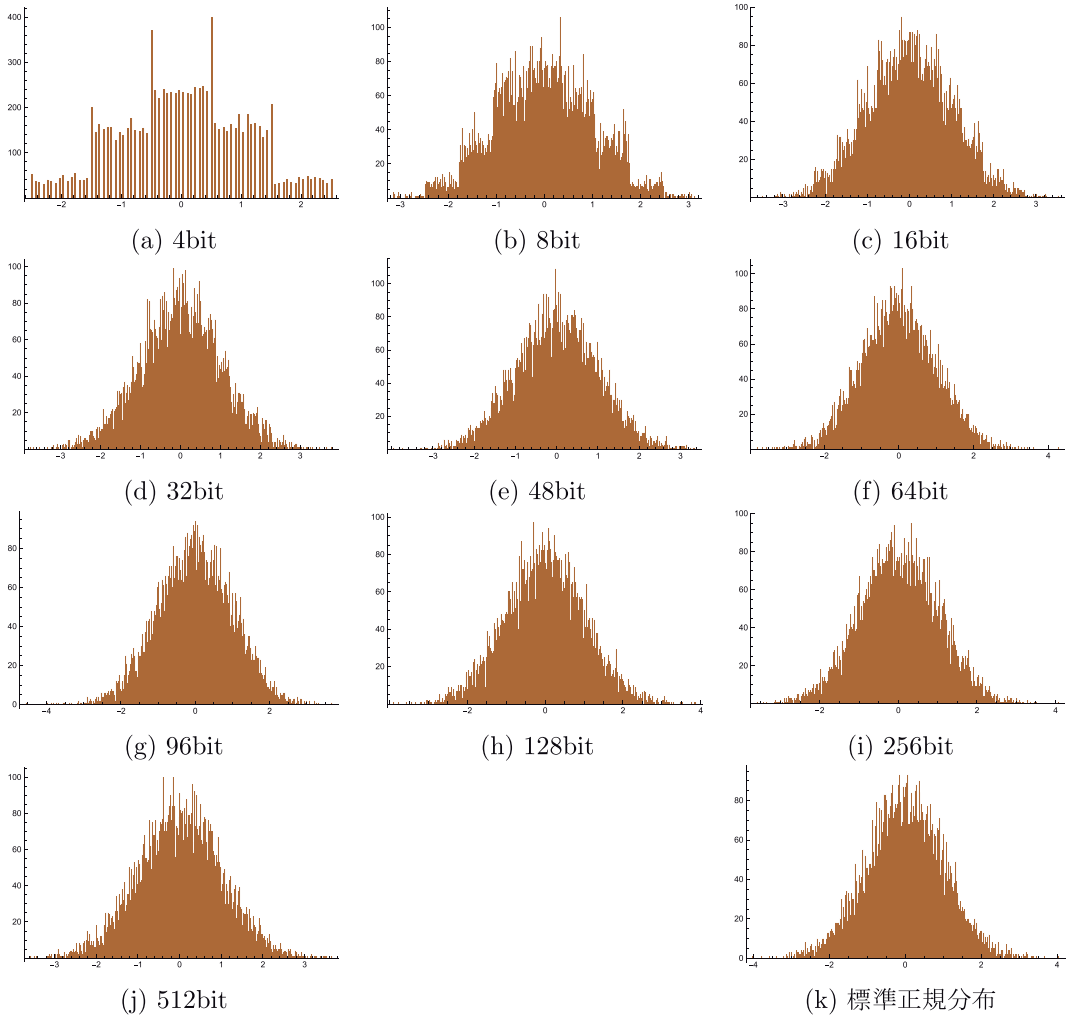



**図２：ビット長の違いによるデータ生成**

二項乱数を連続化して正規乱数近似を行った結果．ここで示すビット（bit）とは二項分布における試行回数に相当する．$n=10000$.

定でも同様の結果であった．

### 5.2　実験２：ビットの長さの違いによる確認

ビット長の違いによるデータの生成の具合を比較する．

４ビットから512ビットと設定して，どのような疑似乱数が生成されているのかを検証する．その結果が図２である．何れもデータ数は10000である．比較のために図２(k)には標準正規分布の乱数を示す．見た目には64ビット程度以上では，正規乱数と見分けがつかないことが確認できる．このことは，図１において縦軸4(⇒$10^4$)と横軸6(⇒$2^6=64$)の交点の $p$ 値は0.45程度であることからも，正規分布と見なしてもよいことが確認できる．

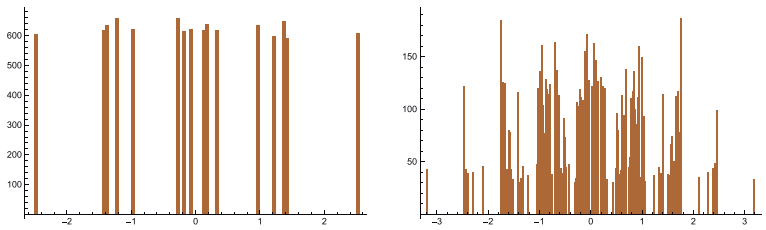
### 5.3　１つの乱数の２度利用

提案したアルゴリズムは，１つの疑似乱数を得るために，直前で使った１つの疑似一様乱数と，新たにもう１つ必要である．その役割は，後者の疑似乱数は二項乱数を得るためのもので，前者は連続化のためである．１つの疑似一様乱数でこの両者を行ったときの結果を図３に示す．これを見てわかるように，両方の役割をさせると何らかの関連性があるために，図２(a)，(b)のようにうまく連続化ができないことがわかる．

## 6　おわりに

近年の正規乱数に関するサーベイとしては，Thomas, *et al.*（2007）や四辻（2010）に種々の生成法が紹介されていて，その長所短所などが記述されている．

Marsaglia による矩形-楔形-裾分割法（the rectangle-



(a) 4bit　　(b) 8bit

**図 3：ビット長の違いによるデータ生成**

二項乱数生成時の1つの乱数を連続化でも共用した場合の結果．$n=10000$.

wedge-tail method）という正規乱数の生成法がある（Knuth, 1998）．この方法（アルゴリズム）は，一見，本提案と類似しているように見えるが，分布の主たる部分は一様乱数で生成し，重要な部分は棄却法に依拠する方法である．

本提案手法による正規乱数生成速度に関しては，別の機会に議論する．

## 参考文献


[1] Ahrens J. H. and Dieter U.（1974）. Computer methods for sampling from Gamma, Beta, Poison and Binomial distribution Computing **12**, 223-246.

[2] Knuth D. E.（1998）. *The Art of Computer Programming*, Volume2, *Seminumerical Algorithms*, Third edition, Addison Wesley.

[3] Nakamura, N.（2015）. Pseudo-Normal Random Number Generation via the Eulerian Numbers, Josai Mathematical Monographs 8, 85-95.

[4] 中村永友・土屋高宏（2015）. 離散型確率分布を通した連続型確率分布にしたがう乱数の生成, 日本計算機統計学会　第29回シンポジウム　論文集.

[5] 中村永友・土屋高宏（2016）. 疑似乱数における局所一様性に関する統計的性質, 日本計算機統計学会　第30回シンポジウム　論文集.

[6] 中村永友・土屋高宏（2017*a*）. 正規分布の裾の確率評価と乱数生成, 日本計算機統計学会　第31回シンポジウム　論文集, 和歌山県立医科大学.

[7] 中村永友・土屋高宏（2017*b*）. 正規分布の裾の確率評価と乱数生成, 札幌学院大学　総合研究所紀要（情報科学）, Vol.4, 1-7, 2017.3.

[8] 清水良一（1976）. 中心極限定理, 教育出版.

[9] 土屋高宏・中村永友（2009）. 変形バケットソートに現れる離散型確率分布と Eulerian 数, 統計数理, Vol.57, No.1, 159-178.

[10] Tsuchiya, T.（2015）. Eulerian distribution with a missing number, Josai Mathematical Monographs 8, 73-83.

[11] Thomas, D. B, Luk, W., Leong, P. H. W. and Villasenor, J.D.（2007）. Gaussian random number generators, ACM Comuting Surveys, Vol. 39 (4), Article No.11, DOI: 10.1145/1287620.1287622.

[12] 四辻哲章（2010）. 計算機シミュレーションのための確率分布乱数生成法, プレアデス出版.

The Proceedings of the Research Institute of Sapporo Gakuin University Vol.5, 1–6 (2018)

# Normal Random Number Generation via Binomial Distribution

Nagatomo NAKAMURA[1]
and
Takahiro TSUCHIYA[2]

**Abstract**

When an independent trial of occurrence probability $p=1/2$ is repeated for $n$ times, its associated probability can be calculated from the binomial distribution. If the number of trials is sufficiently large, the probability can be approximated by a normal distribution. An algorithm for generating normal random numbers from random numbers according to binomial distribution depending on this property is proposed. Since this method involves an approximation, the conditions can be regarded as a normal distribution via numerical experiments are exhibited.

**Keywords:** Normal Random Numbers, Uniform Random Numbers, Binomial Distribution.

[1] Department of Economics, Sapporo Gakuiun University; nagatomo@sgu.ac.jp.
[2] Department of Mathematics, Josai University; takahiro@josai.ac.jp.