

マルチメディアプログラミングによる 基本的分布の乱数発生プログラム

岡村好庸*, 岩田佳治**, 春山和男*

A software for generating the basic distributions of random numbers by multi-media programing

Yoshinobu OKAMURA, Yoshiharu IWATA and Kazuo HARUYAMA

Abstract

A software for generating the basic distributions of random numbers such as uniform distribution in an arbitrary region, normal distribution, exponential distribution and Poisson distribution is developed by using Visual Basic (version 4), which we call multi-media programing language. Those distributions are obtained by redistributing the random numbers of generatig function RND. The results can be compared with the theoretical curves and give the good agreements.

1. はじめに

プログラミング言語には一般に乱数を発生させる関数が用意されている。Visual Basic ではRND が0以上1未満の値を返す乱数発生関数である。ただし初期シード値が変わらない限り、一連のRND関数が返す乱数系列は同じになるので、使用にあたりシステムタイマーから新しいシード値を取得して乱数ジェネレータを初期化する方法が通常用いられる。これにはRandomize ステートメントを

実行するとよい¹⁾。このように得られた乱数系列は区間[0,1)の一樣疑似乱数となる。しかし自然に生じる現象や人為的現象などをシミュレーションする場合には一樣分布以外の特定分布をもつ乱数が必要とされることが多いが、このような乱数発生関数は通常用意されていない。我々はこのRND関数が返す疑似乱数系列を利用して任意区間における一樣分布、正規(ガウス)分布、指数分布、ポアソン分布をもつ乱数発生プログラムを作成し、その分布が理論値と良い一致をすることを検証した。

(1998年9月24日受理)

*宇部工業高等専門学校 電気工学科

**宇部工業高等専門学校 電気工学科, 現在 NHK

2. 基本的分布と乱数発生

任意区間の一様分布, 正規分布(ガウス分布), 指数分布, ポアソン分布などは基本的分布と呼ばれる. 確率現象や物理現象にはこれらのどれかの分布を用いて記述することが妥当と考えられるものが数多く存在する. 例えば容器内の理想気体の速度分布は正規分布(統計物理ではマクスウェルの速度分布と一般に呼ばれる)にしたがう²⁾. 待ち行列ではM/M/1(∞)モデルがよく用いられるが, これは到着分布がポアソン分布でサービス時間が指数分布の場合にあたる³⁾. またポアソン分布はまれにしか起こらない現象が一定時間内に起こる回数を記述するので, デジタルエレクトロニクスの物理的限界に関して不純物のランダムフラクチュエーションの議論でも用いられている⁴⁾. また日常生活の中で時間に関して発生する多くの確率的事象は指数分布に従うことが知られている. このような現象を解析する際に確率を用いて数学的に解析するだけでなく, 乱数を利用してシミュレーションを行うことによりさらに詳しい解析が可能となる. しかし上記分布を発生させる関数は通常用意されていないので, 0から1の範囲の一様乱数を発生させる関数RNDを利用してこれらの分布をする乱数発生関数をプログラムする必要がある. 次にその手順³⁾とプログラムにつき述べる.

2.1 任意区間の一様分布

区間 $[a, b]$ の一様分布は

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x < b \\ 0 & x < a \text{ または } x \geq b \end{cases} \quad (2.1.1)$$

と表わされる. ここで, $f(x)$ を確率密度関数と呼ぶ. 累積分布関数はこのとき

$$\begin{aligned} F(x) &= \int_0^x f(x)dx \\ &= \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases} \end{aligned} \quad (2.1.2)$$

となる. この分布の平均 μ , 標準偏差 σ は

$$\mu = \int_{-\infty}^{\infty} xf(x)dx$$

$$= \frac{b+a}{2} \quad (2.1.3)$$

$$\begin{aligned} \sigma &= \sqrt{\int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx} \\ &= \frac{b-a}{2\sqrt{3}} \end{aligned} \quad (2.1.4)$$

となり, 平均と標準偏差が与えられると一様分布の上下限が

$$a = \mu - \sqrt{3}\sigma \quad (2.1.5)$$

$$b = \mu + \sqrt{3}\sigma \quad (2.1.6)$$

と求められるが本プログラムでは直接に一様分布の上下限を入力する. さて, 0と1の間の乱数発生関数RNDを任意の区間 $[a, b]$ の一様分布に変換するには $F(x)$ にRNDを代入して

$$\text{RND} = \frac{x-a}{b-a} \quad (2.1.7)$$

すなわち

$$x = (b-a)\text{RND} + a \quad (2.1.8)$$

とすればよい.

図1はこの分布発生のための入力画面である.

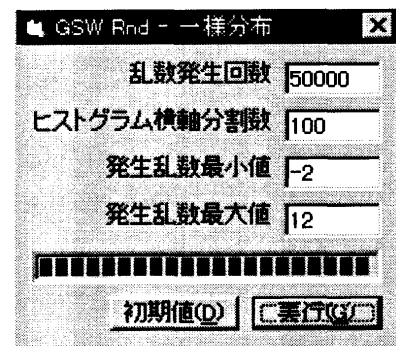


図1

この例では $a = -2, b = 12$ の区間で一様乱数を50000回発生させ, その分布を図示するためのヒストグラムの分割数を100とするように設定している. 図2は実行後の乱数値の分布を表わす. 1%の横線は理論値である.

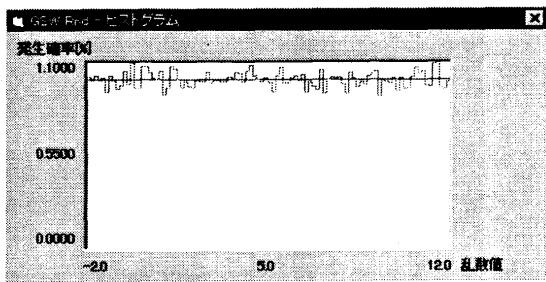


図2

バラツキはあるがほぼ一様な分布が発生している。バラツキは乱数発生回数をふやすことにより減少する。

2.2 正規分布 (ガウス分布)

正規分布の確率密度関数は

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (2.2.1)$$

$$-\infty < x < \infty$$

である。 μ は x の平均値であり、 σ は x の標準偏差である。正規分布の乱数を作成するために n 個の独立した $[0,1)$ の一様分布を考え、その i 番目の分布の確率変数、平均、標準偏差をそれぞれ r_i, μ_i^*, σ_i^* とする。 n 個の独立した $[0,1)$ の一様分布は明らかに同じ平均と標準偏差をもつ。これらは、(2.1.3)、(2.1.4) 式を用いて

$$\mu_i^* = \frac{1}{2} \quad (2.2.2)$$

$$\sigma_i^* = \sqrt{\frac{1}{12}} \quad (2.2.3)$$

と得られる。次にこれら確率変数の和をとり、新たに $\sum_{i=1}^n r_i$ を確率変数とする分布を考える。これは n を大きくすると中心極限定理により、平均 $\mu^* = \sum_{i=1}^n \mu_i^*$ 、分散 (標準偏差の2乗) $(\sigma^*)^2 = \sum_{i=1}^n (\sigma_i^*)^2$ の正規分布に近づく。したがって n 個の独立した $[0,1)$ の一様分布を用いると、 $\sum_{i=1}^n r_i$ の平均、分散は

$$\mu^* = \sum_{i=1}^n \mu_i^* = \frac{n}{2} \quad (2.2.4)$$

$$(\sigma^*)^2 = \sum_{i=1}^n (\sigma_i^*)^2 = \frac{n}{12} \quad (2.2.5)$$

となる。確率変数 z をつぎのように導入する。

$$z = \frac{\sum_{i=1}^n r_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \quad (2.2.6)$$

このとき、 z は平均0、標準偏差1の正規分布をする。したがって平均 μ 、標準偏差 σ が与えられているときの確率変数 x は z と

$$z = (x - \mu) / \sigma \quad (2.2.7)$$

のように関係付けられ

$$x = z\sigma + \mu$$

$$= \frac{\sum_{i=1}^n r_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \sigma + \mu \quad (2.2.8)$$

となる。すなわち、 n を決めて $RND(= r_i)$ の和を計算して、上式より平均 μ 、標準偏差 σ の正規分布をもつ乱数 x を計算すればよい。

図3はこの分布発生のための入力画面である。

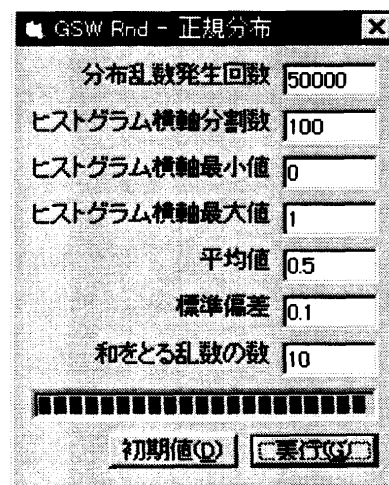


図3

この例では一様乱数10個の和をとり1個の乱数として50000回発生させ、その分布を図示するためのヒストグラムの分割数を100とする。平均値0.5、標準偏差0.1なので分布図は0と1の間でプロットするように設定している。縦軸のスケールはデータから自動的に設定される。図4は実行後の乱数値の分布と理論曲線を表わす。理論曲線は(2.2.1)式を分割間隔でスケールして描かれている。

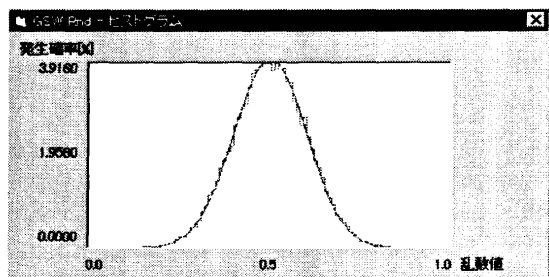


図 4

一様分布と同様バラツキは乱数発生回数をふやすことにより減少する。

2.3 指数分布

指数分布の確率密度関数 $f(x)$ は

$$f(x) = \alpha e^{-\alpha x} \quad (\alpha > 0, x \geq 0) \quad (2.3.1)$$

である。したがって、累積分布関数 $F(x)$ は

$$F(x) = \int_0^x f(x) dx = 1 - e^{-\alpha x} \quad (2.3.2)$$

となる。平均 μ , 標準偏差 σ はそれぞれ

$$\mu = \int_0^{\infty} x f(x) dx = \frac{1}{\alpha} \quad (2.3.3)$$

$$\sigma = \sqrt{\int_0^{\infty} (x - \mu)^2 f(x) dx} = \frac{1}{\alpha} \quad (2.3.4)$$

指数分布にしたがう乱数を作成するには、 $0 \leq F(x) \leq 1$ から $0 \leq 1 - F(x) \leq 1$ となるので、区間 $[0,1)$ の一様乱数 RND を $1 - F(x)$ に等しいとおく。この結果

$$\text{RND} = 1 - F(x) = e^{-\alpha x} \quad (2.3.5)$$

すなわち

$$x = -\frac{1}{\alpha} \ln(\text{RND}) = -\mu \ln(\text{RND}) \quad (2.3.6)$$

となる。

図5はこの分布発生のための入力画面である。

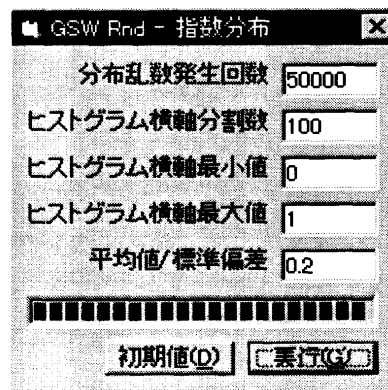


図 5

指数分布では平均値と標準偏差が同じ値となるので、ここでは0.2としている。乱数を50000回発生させ、その分布を図示するためのヒストグラムの分割数を100とする。分布図は0と1の間でプロットするように設定している。縦軸のスケールはデータから自動的に設定される。図6は実行後の乱数値の分布と理論曲線を表す。理論曲線は(2.3.1)式を分割間隔でスケールして描かれている。

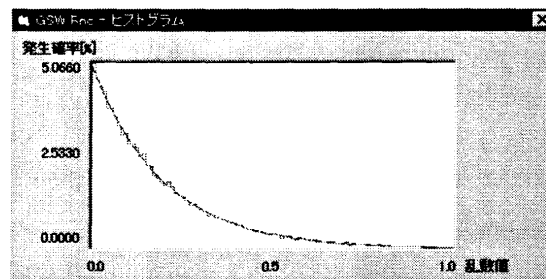


図 6

一様分布と同様バラツキは乱数発生回数をふやすことにより減少する。

2.4 ポアソン分布

単位時間内に事象が n 回発生する確率 $P(n)$ が

$$P(n) = \frac{e^{-\alpha} \alpha^n}{n!} \quad (\alpha > 0, n = 0, 1, 2, \dots) \quad (2.4.1)$$

で表されるとき、これをポアソン分布という。ここで事象が0回、1回、2回、...と発生する確率は α の関数である。平均 μ , 標準偏差 σ はそれぞれ

$$\mu = \sum_{n=0}^{\infty} n P(n) = \alpha \quad (2.4.2)$$

$$\sigma = \sqrt{\sum_{n=0}^{\infty} (n - \mu)^2 P(n)} = \sqrt{\alpha} \quad (2.4.3)$$

となる。ポアソン分布に従う乱数を発生させるには指数分布とポアソン分布の関係を用いる。単位時間内に n 人が到着する確率がポアソン分布 $P(n)$ に従うとする。このとき単位時間内に平均 α 人が到着するので、 T 時間当たりの平均到着人数は αT 人となり T 時間内に到着する人数が n 人である確率は

$$P_T(n) = \frac{e^{-\alpha T} (\alpha T)^n}{n!} \quad (2.4.4)$$

となる。従って、 T 時間内に一人も到着しない確率 $P_T(0)$ は

$$P_T(0) = e^{-\alpha T} \quad (2.4.5)$$

次に、到着間隔の確率密度関数 $f(t)$ を導入する。即ち次の人が T 時間内に到着する確率 P_{arrive} 、到着しない確率 $P_{notarrive}$ は確率密度関数を用いて次のように表される。

$$P_{arrive} = \int_0^T f(t) dt \quad (2.4.6)$$

$$P_{notarrive} = 1 - \int_0^T f(t) dt \quad (2.4.7)$$

T 時間内に一人も到着しない確率 $P_T(0)$ は、ある人が到着後次の人が T 時間内に到着しない確率 $P_{notarrive}$ に等しいので、 $P_T(0) = P_{notarrive}$ 。これより、

$$f(t) = \alpha e^{-\alpha t} \quad (2.4.8)$$

が得られる。すなわち、到着分布がポアソン分布に従うとき、相続く到着間隔分布は指数分布に従う。したがって、単位時間当たりに平均 α 人到着するポアソン分布は、平均到着間隔 $1/\alpha$ の指数分布に従うので、その指数分布乱数 t_1, t_2, t_3, \dots を次々に発生させて次の関係を満たす n を求めればよい。

$$\sum_{i=1}^n t_i \leq 1 < \sum_{i=1}^{n+1} t_i \quad (2.4.9)$$

ここで n が単位時間に到着する人数であり、ポアソン分布となる。 i 番目の一様乱数 $(\text{RND})_i$ を用いて、平均到着間隔 $1/\alpha$ の指数分布乱数 t_i は

$$t_i = -\frac{1}{\alpha} \ln(\text{RND})_i \quad (2.4.10)$$

となるので

$$\sum_{i=1}^n \ln(\text{RND})_i \geq -\alpha > \sum_{i=1}^{n+1} \ln(\text{RND})_i \quad (2.4.11)$$

$$(\text{RND})_1 \cdots (\text{RND})_n \geq e^{-\alpha} > (\text{RND})_1 \cdots (\text{RND})_{n+1} \quad (2.4.12)$$

すなわち、ポアソン分布の平均 $\mu = \alpha$ を入力して、一様乱数を発生させ、その積が $e^{-\alpha}$ より小さくなるときの (乱数発生回数-1) をポアソン分布の確率変数 n とする。

図7はこの分布発生のための入力画面である。



図7

ポアソン分布では平均値と標準偏差の2乗が同じ値となり、ここでは15としている。整数乱数を10000回発生させ、その分布を縦棒で図示する。縦軸と横軸の最大値は発生乱数から自動的に設定されている。図8は実行後の乱数値の分布と理論曲線(2.4.1)式を表わす。

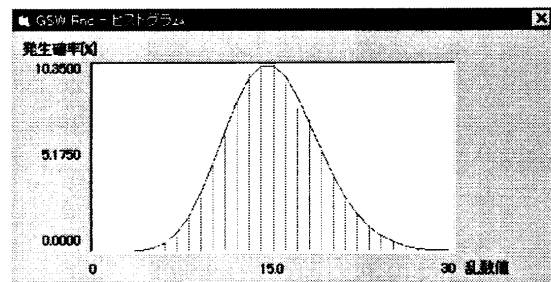


図8

縦棒の分布は理論曲線によく一致している。

3. まとめ

任意区間における一様分布、正規(ガウス)分布、指数分布、ポアソン分布の乱数を発生するプログ

ラムを Visual Basic V.4 を用いて作成し、その実行結果を理論値と比較した。分布を特徴付けるパラメータとして、原則的に平均値及び標準偏差を選んだ。図 2,4,6,8 で示されるように十分満足のいく分布が得られている。理論値からのバラツキは分布乱数の発生回数を増やすことで小さくできる。但し乱数系列には周期があるので、原則として周期以内の発生回数を用いる必要があると思われる。これらのプログラムを関数化することにより様々なシミュレーションに応用することが可能となる。

最後にプログラム作成にあたって留意した点を述べる。1. 理論値のグラフ表示により発生乱数分布の検証を容易にした。2. ヒストグラムの縦軸最大値は自動的に設定されるようにした。3. ヒストグラムの表示領域を指定することにより、見たい範囲を調整できるようにした。4. 実行経過が分かるようにプロセスバーを導入した。5. その他処理速度低下を防ぐ工夫をした。これらの詳細はソースプログラムとして文献5)に収録してある。

謝辞

本研究にあたり、基本的分布を含む様々な確率分布につき詳しい調査をしていただいた津田智寛氏に感謝いたします。

参考文献

- 1) Microsoft Visual Basic ランゲージリファレンス (マイクロソフト, 1995) pp.732-733 p.686.
- 2) 藤田重次: 統計熱物理学 (裳華房, 1989) pp.19-29.
- 3) 木村春彦, 大藪多可志: 情報科学概論 (共立出版) pp135-160.
- 4) R. W. Keyes: Proceedings of the IEEE 63 (1975) 740.
- 5) 岩田佳治, 津田智寛: 確率分布発生プログラムの開発とランダムウォークス (宇部高専電気工学科卒業研究論文, 1997年度).