公開コピー誌
確率分布のノート

暗黒通信団

はじめに

確率や統計というのは哀しいジャンルである。経営部門から開発部門まで、社会人になると実は最も求められる数学分野なのだが、大学入試では大きなウェイトを占めないので、高校ではあまり真面目に勉強されない。大学にも確率論や統計学の授業はあるが、ガチの確率は最初から測度論だし、ガチの統計学は検定の話ばかりで確率分布（あるいは統計分布）について真面目に教わる機会はほとんどない。近隣分野に目を向けても、物理でいう統計力学は「統計学」とは全く別物と思っていい。情報科学は近いことを扱うにもかかわらず、用語が異なっているのでかえって混乱する。そういうわけで、確率や統計は「苦手科目」になりがちである。

本稿は高卒程度の知識を前提として、社会人に向けて書いたノートである。特に、確率に関する基礎的な用語と計算方法、各種の分布を簡潔にまとめた。ただし本稿はあくまで分布に関するノートなので、統計学の他の様々な話題（特に検定について）や情報理論の話題には触れない。また、本稿は確率の本にありがちな、日常的で卑俗な例を排除し、純粋に計算方法だけを書いている。頻度主義やベイズ主義の問題にも立ち入らない。

❶ 確率に関する基本的な知識

❶ 定義

以下では仮に連続分布を例にする。離散分布の場合は積分が $\sum$ になる。

- 確率密度関数 (probability function) と確率分布関数 (distribution function)

  変数を x とし[*1]、x が定義された全区間で正であり、積分したとき結果が 1 になるような関数を確率密度関数 $p(x)$、あるいは「確率分布」と呼ぶ。すなわち、

  $$\int_{-\infty}^{\infty} p(x)dx = 1$$

  である。この性質を満たせば、積分可能などのような関数でも確率密度関数とみなすことができる。パラメータ a, b をもつ x についての確率密度関数は $p(x; a, b)$ と書く。

  確率密度関数の x における積分を確率分布関数 $P(x)$ と呼び、以下の関係がある。

  $$\int_{-\infty}^{a} p(x)dx = P(a)$$

  確率分布関数は、変数 x が任意の実数 a 以下になる確率（$\Pr[x \leq a]$）を表す関数である。確率分布関数は単調増加関数であり、x が大きくなるにつれて 1 に近づく。

- 期待値 μ（あるいは $E(x)$）

  確率密度関数 $p(x)$ が与えられたとき、次式で定義される値 μ を期待値と呼ぶ[*2]。

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

- 分散 σ^2（あるいは $V(x)$）

  次の式で定義された値を分散 σ^2 と呼ぶ。σ^2 の平方根 σ は「標準偏差」と呼ばれる。

  $$\begin{aligned}\sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx \\ &= \int_{-\infty}^{\infty} x^2 p(x)dx - \mu^2\end{aligned}$$

- モーメント（積率）M

  期待値は $xp(x)$ の積分であり、分散は定数部分を除けば $x^2p(x)$ の積分である。ここで、$p(x)$ に掛ける x の次数を一般化し、

  $$M(k) = \int_{-\infty}^{\infty} x^k p(x)dx$$

  を考えることができる。これを原点周りの k 次のモーメントと呼ぶ。原点ではなく期待値周りのモーメントは、x^k を $(x-\mu)^k$ で置き換えると計算できる。期待値周りの 1 次のモーメントは 0 である[*3]。3 次のモーメントを歪度と呼び、4 次のモーメントを尖度と呼ぶ。

- 中心極限定理

  狭い意味での中心極限定理とは、多くの確率密度関数で、その確率密度関数に従う標本の平均と確率密度関数の期待値との差はサンプルのサイズを大きくしたとき近似的に正規分布に従う、というものである。「多くの」という表現はくせ者で、中心極限定理が成り立たない分布も存在する。例えばコーシー分布はその例である。コーシー分布は期待値が有限値ではないので、差が定義できない。

- 安定分布

  標本の平均ともとの分布の期待値との差が、もとの分布と同じになるような分布を「安定分布」と呼ぶ。より正確には、同じ確率密度関数 X に従う独立な標本 x_i について

  $$\frac{1}{a}\sum_{i=1}^{n} x_i - b$$

  がもとの X と同じ分布になるような $a > 0$ と b が存在するとき、X を安定分布 (stable distribution) という。$b = 0$ なら厳密に安定 (strictly stable) であるという。厳密に安定な分布は次の 3 つが有名である。

  ▷ 正規分布：$a = n^{1/2}$
  ▷ コーシー分布：$a = n$
  ▷ レヴィ分布：$a = n^2$

[*1] これを特に「確率変数」と呼ぶ。
[*2] 次節以降の様々な分布では、「期待値」と書く方が適切だが「平均」と書いている。これは執筆者の趣味である。
[*3] これは簡単な計算で確認できる。

❷ 確率密度関数の変数変換

x についての確率密度関数 $X(x)$ を y についての確率密度関数 $Y(y)$ に変換するには、

$$X(x)dx = Y(y)dy$$

の関係を使う[*4]。この式を変形すると、

$$Y(y)\frac{dy}{dx} = X(x)$$

となるので、別途与えられた x と y の関係式から微分係数 dy/dx を求め、代入して、$Y(y)$ を求めることができる。

❸ 確率密度関数同士の演算

2 つの独立な確率密度関数 X, Y があるとする。X, Y からサンプル x_0, y_0 をとり、何らかの演算をする。その結果の値 $z_0 = f(x_0, y_0)$ が従う確率密度関数 Z を考えたい。

X と Y の同時分布は、両者が独立であることから、

$$\iint X(x)Y(y)dxdy$$

で与えられる。ここで $z = f(x,y)$ という変数変換をする。dz と dy の関係を求めて、上記積分を、

$$\iint X(x)Y(z,x)dxdz$$

という形にして、z について解くと、

$$\int f(x)dx$$

という形になるので、その $f(x)$ が確率密度関数になる。

ここで、積分変換の仕方について書く。

$$\iint f(x,y)dxdy = f(x(u,v), y(u,v))Xdudv$$

という変換をしたい場合、

$$\begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{pmatrix} \begin{pmatrix} du \\ dv \end{pmatrix}$$

という変換を考えると、行列部分 (ヤコブ行列 J) は $(du, dv) \to (dx, dy)$ という一次変換を与える。$u-v$ 空間の直交単位ベクトルが張る面積は 1 だが、$x-y$ 空間ではこれが $\det J$ 倍される。ゆえに、$X = \det J$ とすれば積分変換が可能である。

以下に例として、加減乗除についての具体的計算を書く。

サンプル同士の加算は、すなわち平均値が従う分布を導出することである。一般的に、演算 $f(x,y)$ が $z = x+y$ の場合、X,　Y がどのような分布であっても（仮に X, Y が異なる分布であっても）X, Y が独立であるか否かに関わらず、平均 μ_z は $\mu_x + \mu_y$ である。X, Y が互いに独立であれば（それらの分布によらず）、$\sigma_z^2 = \sigma_x^2 + \sigma_y^2$ が成立する。以下では、この計算が畳み込み積分であることを示す。

$z = x+y$ とおくと、$dz = dx + dy$ で、ヤコビアンは以下の式になる。デターミナントは 1 である。

$$\begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} dx \\ dz \end{pmatrix}$$

確率分布 $p(x)$ と $p(y)$ の加算合成は、

$$\iint p(x)p(y)dydx = \iint p(x)p(z-x)dzdx$$

となり、畳み込み積分となっている。

正規分布同士の合成は、

$$\iint \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\exp\left(-\frac{(z-x-\mu)^2}{2\sigma^2}\right)dzdx$$

$$= -\frac{\sqrt{2}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{4\sqrt{\pi}\sigma}\int erf(\sqrt{2}(x-z+\mu)/2\sigma)dx$$

である。

減算の場合 $z = x-y$ とすると、

$$J = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

となり、同じく $\det J = 1$ である。確率分布 $p(x)$ と $p(y)$ の減算合成は、

$$\iint p(x)p(y)dydx = \iint p(x)p(z+x)dzdx$$

となる。

乗算の場合、$z = xy$ とすると、$dz = xdy + ydz = xdy + (z/x)dz$。

$$J = \begin{pmatrix} 1 & 0 \\ z/x^2 & -1/x \end{pmatrix}$$

となり、$\det J = -1/x$ である。確率分布 $p(x)$ と $p(y)$ の乗算合成は、

$$\iint p(x)p(y)dydx = \iint p(x)p(z/x)(-1/x)dzdx$$

となる。

除算の場合、$z = x/y$ とすると、$dz = dx/y - xdy/y^2$。$y = x/z$ より、$dz = (z/x)dx + (-z^2/x)dy$。

❷ 離散分布

この節では、よく知られている離散分布をまとめる。

- 離散一様分布 (discrete uniform distribution)
  - ▷ $p(x;n) = 1/n$
  - ▷ 平均：$\dfrac{n+1}{2}$
  - ▷ 分散：$\dfrac{n^2-1}{12}$

  その名の通り、一様な分布である。
- ラダマチャー分布 (Rademacher distribution)
  - ▷ $p(x) = -1$ or 1
  - ▷ 平均：0

  $x \in \{-1, +1\}$ の二値変数を，それぞれ確率 1/2 でとる分布。
- 二項分布 (binomial distribution)
  - ▷ $p(x;n,p) = {}_nC_x p^x (1-p)^{n-x}$
  - ▷ 平均：np
  - ▷ 分散：$np(1-p)$

[*4] 確率密度関数は積分すると常に同じ値 (1) になるので、その任意の微小面積が一致する必要がある。

ある集団において、ある特性 A を持つものの割合が p であるなら、当然その特性 A を持たないものの割合は $1-p$ である。この集団から無作為に n 個の抽出（n 回のベルヌーイ試行）をしたとき、その特性 A を持つものが x 個である確率を考える。もちろん、$n \to \infty$ ならば、x は np に近づいていくはずである。

さて、抽出した n 個のうち x 個がその特性 A を持つ場合は $_nC_x$ 通りである。その各々に対して x 個が特性 A を持つ確率は p^x、残り $n-x$ 個が特性 A を持たない確率は $(1-p)^{n-x}$ であり、両者が共に起こるのはこれらの積で与えられる。よって、上記の密度関数が得られる。$n \to \infty$ で正規分布に近づく。

- ポアソン分布 (Poisson distribution)
  - ▷ $p(x;\lambda) = \dfrac{e^{-\lambda}\lambda^x}{x!}$
  - ▷ 平均 : λ
  - ▷ 分散 : λ

  単位時間中に平均で λ 回発生する事象がちょうど x 回（x は 0 を含む自然数、$x = 0, 1, 2, ...$）発生する確率。パラメータが n と λ/n である二項分布において、λ を一定に保ったまま n を無限大に近づけると、その分布は平均 λ のポアソン分布に近づく。計算機的で乱数発生させるためには、指数分布 e^{-u} に従う確率変数 a_i を順に吐き出し、足し合わせる。$\sum_{i=1}^{n}$ が最初に λ を越えたときの n がポアソン乱数になる。

- ベルヌーイ分布 (Bernoulli distribution)
  - ▷ $p(x) = 0$ or 1
  - ▷ 平均 : p
  - ▷ 分散 : $pq = p(1-p)$

  確率 p で 1 を、確率 $q = 1-p$ で 0 をとる離散確率分布。ベルヌーイ分布に従う事象を生じさせることをベルヌーイ試行という。

- 幾何分布 (geometric distribution)
  - ▷ $p(x;p) = p(1-p)^{x-1} \quad (0 < p < 1)$
  - ▷ 平均 : $1/p$
  - ▷ 分散 : $(1-p)/p^2$

  成功確率 p のベルヌーイ試行で、初めて成功するまでの試行の回数の分布である。

- 超幾何分布 (hypergeometric distribution)
  - ▷ $p(x;M,N,k) = \dfrac{\binom{M}{x}\binom{N-M}{k-x}}{\binom{N}{k}}$
  - ▷ 平均 : $\dfrac{kM}{N}$
  - ▷ 分散 : $\dfrac{nM(N-M)(N-n)}{N^2(N-1)}$
  - ・ N が大きくなると、二項分布に近づく。また N と M の比が小さく、サンプル数 k が大きいとき、ポアソン分布に近づく。
  - ・ N 個の玉が入っている容器があり，このうち M 個が赤で残りが白であるとする。この容器から k 個の玉を非復元抽出 (取り出した玉は容器に戻さない) するとき，x 個の玉が赤である確率を示す分布である。ちなみに復元抽出なら二項分布である。

- 負の二項分布 (negative binomial distribution)
  - ▷ $p(x;k,p) = \begin{pmatrix} x+k-1 \\ k-1 \end{pmatrix} p^k(1-p)^x \quad k > 0,\ 0 < p < 1,\ 0 \le x$
  - ▷ 平均 : $k(1-p)/p$
  - ▷ 分散 : $k(1-p)/p^2$

  x は整数である。k が正整数のときはパスカル分布 (Pascal distribution) という。$k = 1$ なら幾何分布。確率 p で当たりが出るベルヌーイ試行において、当たりを k 回引くまでにはずれを引く回数の分布でもある。

❸ 連続分布

- 一様分布 (uniform distribution)
  - ▷ $p(x) = 1 \quad -0.5 < x < 0.5$
  - ▷ 平均 : 0
  - ▷ 分散 : 1/12

  標準偏差 σ は $1/2\sqrt{3}$ である。もし密度関数が $-1 < x < 1$ で定義されるなら（これは規格化されてるとは言えない）、$\sigma = 1/\sqrt{3}$ となる。

- 三角分布 (triangular distribution)
  - ▷ $p(x;a,b,c) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & (a \le x \le c) \\ \dfrac{2(b-x)}{(b-a)(b-c)} & (c \le x \le b) \end{cases}, \quad a \le x \le b,\ a < c < b$
  - ▷ 平均 : $(a+b+c)/3$
  - ▷ 分散 : $\dfrac{a^2+b^2+c^2-ab-bc-ca}{18}$

  三角分布は $x = a$ から直線で上昇し、$x = c$ で最大になり、そこから $x = b$ で 0 になるように直線的に減少する。

- 正規分布 (normal distribution)
  - ▷ $p(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$
  - ▷ 平均 : μ
  - ▷ 分散 : σ^2

  あらゆる統計分布の中で、最も有名。
  - ・ $\mu = 0,\ \sigma^2 = 1$ とした場合を標準正規分布
    $$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$
    と呼ぶ。その微分は、
    $$\frac{-x}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$
    で与えられる。
  - ・ 変曲点は標準偏差 σ である[*5]。
  - ・ 二つの異なる正規分布 $N(\bar{x}_1, \sigma_1^2)$、$N(\bar{x}_2, \sigma_2^2)$ から、一つずつ要素を取り出して和をとった値は $N(\bar{x}_1 + \bar{x}_2, \sigma_1^2 + \sigma_2^2)$ の分布に従う（加法性）。
  - ・ ボックス-ミューラー (Box-Muller) 法によって一様分布から作られる。それは次のように行う。
    $(0,1]^2$ 上の一様分布に従う確率変数 x_1, x_2 から
    $y_1 = \sqrt{-2\ln x_1}\sin(2\pi x_2),\ \ y_2 = \sqrt{-2\ln x_1}\cos(2\pi x_2)$

[*5] エントロピーは $\ln(\sqrt{2\pi e\sigma^2})$ である。

によって変換する。この y_1, y_2 が標準正規分布に従う。これは、x_1 を一様乱数としたとき $(-2\ln x_1)$ が平均値 2 の指数分布に従い、$\sqrt{-2\ln x_1}$ がレイリー分布に従うことを使っている。
  - ・再生性をもつ。

- 多変量正規分布 (multivariate normal distribution)
  - ▷ $p(x;\mu,\sigma) = \dfrac{1}{(2\pi)^{k/2}|\sigma|^{1/2}} \exp\left[-\dfrac{(\mathbf{x}-\mu)^\top \sigma^{-1}(\mathbf{x}-\mu)}{2}\right]$
  - ▷ 平均 : μ (平均ベクトル)
  - ▷ 分散 : σ^2 (共分散行列)

  k は次元。共分散行列 σ^2 は正定値行列である。
- 逆正規分布 (inverse normal distribution)
  - ▷ $p(x;\mu,\lambda) = \sqrt{\dfrac{\lambda}{2\pi x^3}} \exp\left[-\dfrac{\lambda(x-\mu)^2}{2\mu^2 x}\right] \quad x>0,\ \mu>0,\ \lambda>0$
  - ▷ 平均 : μ
  - ▷ 分散 : μ^3/λ
- 対数正規分布 (log-normal distribution)
  - ▷ $p(x;\mu,\sigma) = \dfrac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\dfrac{(\ln x-\mu)^2}{2\sigma^2}\right]$
  - ▷ 平均 : $\exp\left(\mu + \dfrac{1}{2}\sigma^2\right)$
  - ▷ 分散 : $\exp(2\mu+\sigma^2)(\exp(\sigma^2)-1)$

  対数を取ると正規分布になるような分布。確率変数 a と b が対数正規分布に従うときには、ab や a/b も対数正規分布に従う。
- 指数分布（exponential distribution）
  - ▷ $p(x;\lambda) = \lambda e^{-\lambda x}$
  - ▷ 平均 : $\sigma = 1/\lambda$
  - ▷ 分散 : $\sigma = 1/\lambda^2$

  分布関数は $P(x) = 1-\exp(-x)$、その逆関数は $P^(-1)(x) = -\log(1-x)$ である。指数分布に従った乱数の発生は、自由度 2 の χ^2 分布に従う確率変数 a を 2λ で割って求める。
- カイ分布 (chi distribution)
  - ▷ $p(x;n) = \dfrac{2^{1-n/2}}{\Gamma(n/2)} \exp\left[-\dfrac{x^2}{2}\right] x^{n-1}, \quad x \geq 0,\ n>0$
  - ▷ 平均 : $\dfrac{\sqrt{2}\Gamma((n+1)/2)}{\Gamma(n/2)}$
  - ▷ 分散 : $n - 2\left(\dfrac{\Gamma((n+1)/2)}{\Gamma(n/2)}\right)^2$

  n を自由度と呼ぶ。X^2 が自由度 n のカイ二乗分布に従うとき，X は自由度 n のカイ分布に従う。$n=2$ ならレイリー分布。
- カイ二乗分布 (chi square distribution)
  - ▷ $p(x;n) = \dfrac{(1/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2}, \ \ x \geq 0,\ n>0$
  - ▷ 平均 : n
  - ▷ 分散 : $2n$

  分割表の独立性を調べるカイ二乗検定で使われる。n は自由度と呼ばれる。
  - ・分布関数は、

    $$P(x) = \frac{\gamma(n/2, x/2)}{\Gamma(n/2)}$$

    となる。$\Gamma(n,z)$ は不完全ガンマ関数。
  - ・自由度 n のカイ二乗分布に従った乱数を発生させるには、分散 1 の正規分布に従う独立な n 個の確率変数を平方和する。つまり、確率変数 x が平均 0、分散 1 の正規分布に従うとき，x^2 は自由度 1 のカイ二乗分布に従うことになる。より一般に、x_i を、平均 μ_i で分散 σ_i^2 の正規分布に従う、n 個の独立なランダム変数とすると、

    $$\sum_{i=1}^{n} \left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$$

    は χ^2 分布に従う。
  - ・自由度 2 のカイ二乗分布は，$\beta = 2$ の指数分布に対応する。
  - ▷ カイ二乗分布は、自由度が $n/2$，$\lambda = 1/2$ のガンマ分布でもある。
  - ・カイ二乗分布に従う確率変数 x_1, x_2 をそれぞれの自由度 n_1, n_2 で割って比をとると F 分布に従う。
  - ・$x \sim \chi_2^2$ （自由度 2）ならば、x は期待値 2 の指数分布に従う。
  - ・カイ二乗分布に従う乱数を作るには、確率変数 x が自由度 $n/2$ のガンマ分布に従うとき、$2x$ が自由度 n のカイ二乗分布に従うことを使う。
- F 分布 (F-distribution)
  - ▷ $p(x;n_1,n_2) = \dfrac{1}{\mathrm{B}(n_1/2, n_2/2)} \left(\dfrac{n_1 x}{n_1 x + n_2}\right)^{n_1/2} \left(1 - \dfrac{n_1 x}{n_1 x + n_2}\right)^{n_2/2} x^{-1}$
  - ▷ 平均 : $\dfrac{n_2}{n_2-2}$
  - ▷ 分散 : $\dfrac{2{n_2}^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$

  B はベータ関数。分散の比を検定する F 検定で利用される。パラメータ : n_1 と n_2 は正整数で自由度と呼ばれる。
- Student の t 分布 (Student's t-distribution)
  - ▷ $p(x;n) = \dfrac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\dfrac{n}{2}\right)\left(1+\dfrac{x^2}{n}\right)^{(n+1)/2}}, \quad n>0$
  - ▷ 平均 : 0
  - ▷ 分散 : $n/(n-2)$

  単に t 分布とも呼ばれる。平均値の差の検定などに使われる。n は自由度と呼ばれる整数である。
  - ・x が平均 0，分散 1 の正規分布，v が自由度 n のカイ二乗分布に従うとき $x/\sqrt{v/n}$ は自由度 n の t 分布に従う。
  - ・正規分布を仮定した平均値の差の検定に使われるが、元の正規分布のパラメータには依らない（正規化したあとに t 分布が使われる)。

    x が、平均 μ 、分散 σ^2 の正規分布に従う独立な変数であるとする。その標本平均を $\overline{x}$ とし、分散を σ^2 とすると、

    $$t = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$$

    である。
  - ・自由度 n の t 分布に従う乱数は、正規乱数 x と自由度 n のカイ二乗乱数 v を組み合わせて、$x/\sqrt{v/n}$ の計算で作る。この乱数は、平均 0 で分散 $n/(n-2)$ である。
- ガンマ分布 (Gamma distribution)
  - ▷ $p(x;n) = x^{n-1} \dfrac{e^{-x/\theta}}{\Gamma(n)\,\theta^n}, \ x>0,\ n>0,\ \theta>0$

▷ 平均：$n\theta$
▷ 分散：$n\theta^2$

いくつかの分布を一般化した分布である。

・分布関数は、$P(x) = \dfrac{\gamma(n, x/\theta)}{\Gamma(n)}$ である。
・パラメータが n_1, θ である確率変数を x_1、n_2, θ である確率変数を x_2 を加算した分布もまたガンマ分布であり、パラメータは $n_1 + n_2$ と θ になる（再生性）。
・n が整数のときアーラン分布と呼ぶ。特に $n = 1$ なら指数分布である。n が半整数で $\theta = 2$ のとき χ^2 分布である。
・パラメータ θ の互いに独立な n 個の指数分布の和は、パラメータ n, θ のガンマ分布（アーラン分布）である。n が大きくなると正規分布に近づく。
・自由度 1 のガンマ分布は平均 1 の指数分布である。
・自由度 1/2 のガンマ分布は標準正規分布に従う確率変数 x に対して、$x^2/2$ の分布に対応する。
・ガンマ分布に従う乱数を作るには、n 個の一様乱数 $u_1, u_2 \cdots u_n$ を使って、

$$\frac{-1}{\theta}\log(u_1 u_2 \cdots u_n)$$

という計算をすると、平均 θ、自由度 n のガンマ乱数が得られる。

- ディリクレ分布 (Dirichlet distribution)
  ▷ $p(\mathbf{x}, \alpha) = \dfrac{1}{B(\alpha)}\prod_{i=1}^{k} {x_i}^{\alpha_i - 1}, \quad x_i \geq 0, \quad \sum_{i=1}^{k} x_i = 1, \ \ \alpha_i \geq 0$

別名多変量ベータ分布。$A = \sum_i^k \alpha_i$、B はベータ関数である。同時に発生することのない k 個の事象がそれぞれ $\alpha_i - 1$ 回発生したときに、各事象の起こる確率が x_i である確率を意味する。$\mathbf{x} = (x_1, \ldots, x_k)$ は実数ベクトル、$\alpha = (\alpha_1, \ldots, \alpha_k)$ はパラメータベクトルと呼ばれる。

- コーシー分布 (Cauchy distribution)
  ▷ $p(x; x_0, \gamma) = \dfrac{1}{\pi}\left[\dfrac{\gamma}{(x - x_0)^2 + \gamma^2}\right]$

別名ローレンツ分布、ブライト・ウィグナー分布ともいう。x_0 は分布の最頻値、γ は半値半幅に対応する。平均と分散は存在しない。

・$x_0 = 0$、$\gamma = 1$ のときを標準コーシー分布 $p(x) = \dfrac{1}{\pi(1 + x^2)}$ と呼ぶ。その分布関数は、

$$P(x) = \frac{1}{\pi}\arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}$$

である。
・コーシー分布は、$x = 0$ 付近で発散するため、平均値や分散が定義できない。最頻値と中央値は常に定義され、それらはいずれも x_0 である。
・x をコーシー分布に従う確率変数とすると、コーシー分布の特性関数は以下のように与えられる。

$$\phi_x(t) = \exp(i\, x_0\, t - \gamma\, |t|)$$

・x_1 と x_2 を標準正規分布に従う独立な確率変数とすると、それらの比 x_1/x_2 は標準コーシー分布に従う。
・$x_1, \cdots, x_n$ を標準コーシー分布に従う独立な確率変数とすると、それらの平均は標準コーシー分布に従う。
・自由度 1 の t 分布は、標準コーシー分布である。
・正規分布から独立にサンプルを二つとってきて割った商はコーシー分布に従う。
・$x_0 = 0$, $\gamma = 1$ のときは自由度 1 の t 分布である。
・確率変数が Cauchy 分布に従うとき, その逆数も Cauchy 分布に従う。

- ラプラス分布 (Laplace distribution)
  ▷ $p(x; \mu, \phi) = \dfrac{1}{2\phi}\exp(-\dfrac{|x - \mu|}{\phi}), \quad -\infty < \mu < \infty, \ \phi > 0$
  ▷ 平均：μ
  ▷ 分散：$2\phi^2$

二重指数分布とも呼ばれる。

- レイリー分布 (Rayleigh distribution)
  ▷ $p(x; \theta) = \dfrac{x}{\theta^2}\exp\left(-\dfrac{x^2}{2\theta^2}\right), \quad 0 \leq x, \ \theta > 0$
  ▷ 平均：$\theta\sqrt{\dfrac{\pi}{2}}$
  ▷ 分散：$\left(2 - \dfrac{\pi}{2}\right)\theta^2$
- ロジスティック分布 (logistic distribution)
  ▷ $p(x) = \dfrac{\exp(-\frac{x - \mu}{\sigma})}{\sigma(1 + \exp(-\frac{x - \mu}{\sigma}))^2}, \quad -\infty < x < \infty, \ \mu > 0, \ \sigma > 0$
  ▷ 平均：μ
  ▷ 分散：$\pi^2\sigma^2/3$

確率分布関数 $P(x) = 1/(1 + \exp(-\dfrac{x - \mu}{\sigma}))$ はシグモイド関数の形をしている。

- 第一種パレート分布 (Pareto distribution)
  ▷ $p(x) = \dfrac{\frac{a}{b}}{\left(\frac{x}{b}\right)^{a+1}}, \quad a > 0, \ b > 0, \ x \geq b$
  ▷ 平均：$\dfrac{ab}{b - 1}$
  ▷ 分散：$\dfrac{a^2 b}{(b - 1)^2 (b - 2)}$

第二種や第三種も存在する。X が指数分布 $\exp(-y)$ に従うとき、 $\alpha\exp(X/\beta)$ がパレート分布に従う。

- レヴィ分布 (Lvy distribution)
  ▷ $p(x; \mu, \sigma) = \sqrt{\dfrac{\sigma}{2\pi}}\exp\Big[-\dfrac{\sigma}{2(x - \mu)}\Big](x - \mu)^{-3/2}, \quad \mu > 0, \ \ \sigma, \ \mu < x$

有名な安定分布の一つ。期待値と分散は共に無限大に発散している。

- ミンコフスキー分布 (Minkowski distribution)
  ▷ $p(x; \mu, \sigma^2, q) = \dfrac{q}{2(2\sigma^2)^{1/q}\Gamma(1/q)}\exp\left(-\dfrac{|x - \mu|^q}{2\sigma^2}\right)$

$q = 2$ なら正規分布になるらしい。

- ポアソン分布 (Poisson distribution)
  ▷ $p(x; \lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}, \quad x \geq 0, \ \lambda > 0$
  ▷ 平均：λ
  ▷ 分散：λ

x は整数で、λ は実数。歪度は $1/\sqrt{\lambda}$, 尖度は $3 + 1/\lambda$ である。

- ワイブル分布 (Weibull distribution)
  ▷ $p(x; \sigma, \gamma) = \dfrac{\gamma}{\sigma}\Big(\dfrac{x}{\sigma}\Big)^{\gamma - 1}\exp\Big[-\Big(\dfrac{x}{\sigma}\Big)^{\gamma}\Big], \quad \sigma > 0, \ \gamma > 0, \ x \geq 0$

▷ 平均：$\sigma\Gamma\left(\frac{\gamma+1}{\gamma}\right)$

▷ 分散：$\sigma^2\Gamma\left(\frac{\gamma+2}{\gamma}\right)-\sigma^2\left[\Gamma\left(\frac{\gamma+1}{\gamma}\right)\right]^2$

分布関数は $P(x;\sigma,\gamma)=1-\exp\left[-\left(\frac{x}{\sigma}\right)^\gamma\right]$ である。

- ウィシャート分布 (Wishart distribution)
  - ▷ $p(\mathbf{X};\mathbf{\Sigma},n,p)=\frac{1}{Z(\mathbf{\Sigma},n)}|\mathbf{X}|^{\frac{n-p-1}{2}}\exp\left[-\frac{1}{2}\mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{X})\right]$, $Z(\mathbf{\Sigma},n,p)=|\mathbf{\Sigma}|^{\frac{n}{2}}2^{\frac{np}{2}}\Gamma_p(n/2),\ n\geq p$
  - ▷ 平均：$n\mathbf{\Sigma}$

  $\Gamma_p(x)$ は多変量ガンマ関数、$\mathbf{\Sigma}$ は $p\times p$ の対称行列、n は自由度とよばれる自然数である。
  - ・多変量分布 $N(0,\mathbf{\Sigma})$ に従う n 個の p 次元ベクトル $\mathbf{x}_i$ について，$\mathbf{X}=\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top$ が従う分布である。
  - ・ウィシャート分布はカイ二乗分布の多次元化バージョンであり、$p=1$ ならカイ二乗分布そのものである。
- フォンミーゼス分布 (von Mises distribution)
  - ▷ $p(x;\mu,\sigma)=\frac{\exp[\sigma\cos(x-\mu)]}{2\pi I_0(\sigma)},\quad -\pi\leq\mu\leq\pi,\quad \sigma>0,\quad -\pi\leq x\leq\pi$

  円上の正規分布。μ で最も大きくなり，$\mu\pm\pi$ で最小になる。循環正規分布 ともいう。多次元化すると次のミセスフィッシャー分布になる。$I_j(x)$ は第 1 種変形 Bessel 関数。再生性は持たない。
- ミーゼスフィッシャー分布 (von Mises-Fisher distribution)
  - ▷ $p(\mathbf{x};\mu,\sigma)=\frac{\sigma^{d/2-1}}{(2\pi)^{d/2}I_{d/2-1}(\sigma)}\exp(\sigma\mu^\top\mathbf{x}),\quad \sigma>0$
  - ▷ 最頻値：μ

  意味としては超球上の正規分布。μ は d 次元ベクトルで、$\|\mu\|=1$。I_j は第 1 種変形 Bessel 関数。
- ベキ級数分布 (power series distribution)
  - ▷ $p(x;a_x,\theta)=\frac{a_x\theta^x}{\sum_{y=0}^{\infty}a_y\theta^y},\quad a_x>0,\quad x=0,1,\ldots$

  $a_x=\begin{pmatrix}n\\x\end{pmatrix},\theta=p/(1-p)$ とすると二項分布で､$a_x=1/x!$, $\theta=\lambda$ とすると Poisson 分布になる。
- ベキ関数分布 (power function distribution)
  - ▷ $p(x;\beta,\gamma)=\left(\frac{\gamma}{\beta}\right)\left(\frac{x}{\beta}\right)^{\gamma-1},\quad 0\leq x\leq\beta$
  - ▷ 平均：$\frac{\beta\gamma}{\gamma+1}$,
  - ▷ 分散：$\frac{\beta^2\gamma}{(\gamma+2)(\gamma+1)^2}$ $a=\gamma,\ b=1$ のときは、$\beta=1$ のベキ関数分布に相当する。
- ベータ分布 (beta distribution)
  - ▷ $p(x;a,b)=\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$, x は0〜1 の実数, $B(a,b)$ はベータ関数
  - ▷ 平均：$\frac{a}{a+b}$,
  - ▷ 分散：$\frac{ab}{(a+b)^2(a+b+1)}$

  ディリクレ分布はこれを多変量にしたものらしい。
- アーラン分布 (Erlang distribution)
  - ▷ $p(x;n,\theta)=\frac{1}{\theta^n(n-1)!}x^{n-1}\exp\left(-\frac{x}{\theta}\right),\quad n>0,\ \theta>0$
  - ▷ 平均：$n\theta$
  - ▷ 分散：$n\theta^2$

  $n=1$ なら指数分布に対応する。歪度は $2/\sqrt{n}$ で、尖度は $3+6/n$ である。
- ガンベル分布 (Gumbel distribution)
  - ▷ $p(x;\alpha,\beta)=\frac{1}{\beta}\exp\left(-\exp\left(\frac{x-\alpha}{\beta}\right)+\frac{x-\alpha}{\beta}\right)$
  - ▷ 平均：$\alpha-\gamma\beta$ 　γ はオイラー定数。
  - ▷ 分散：$\frac{1}{6}\pi^2\beta^2$

  サンプルの最大値の分布などを扱う順序統計などで利用されるらしい。
- 正準分布 (Canonical distribution)
  - ▷ $p(E)=\frac{1}{Z}\exp(-E/kT)$

  物理が起源。熱平衡状態にある温度 T の系がエネルギー E をとる確率のこと。Z は分配関数と呼ばれる規格化定数。
- 対数分布 (logarithmic distribution)
  - ▷ $p(n;p)=-\frac{p^n}{n\ \ln(1-p)}\quad n\geq 1, 0<p<1$
  - ▷ 平均：$\frac{-p}{(1-p)\ln(1-p)}$,
  - ▷ 分散：$\frac{-p\big(p+\ln(1-p)\big)}{\big((1-p)\ln(1-p)\big)^2}$

  対数級数分布ともいう。
- 指数ベキ分布 (exponential power distribution)
  - ▷ $p(x;\mu,p)=\frac{1}{2\sigma p^{1/p}\Gamma(1+1/p)}\exp\left[-\frac{|x-\mu|^p}{p\sigma^p}\right]$
  - ▷ 平均：μ
  - ▷ 分散：σ^2

  パラメータ μ は任意の実数，σ と p は正の実数。$p=2$ で正規分布，$p=1$ でラプラス分布に相当。
- 指数分布 (exponential distribution)
  - ▷ $p(x;\beta)=\frac{1}{\beta}\exp\left(-\frac{x}{\beta}\right)$
  - ▷ 平均：β
  - ▷ 分散：β^2

  ガンマ分布で $k=1$ の場合に相当する。

❹ 参考文献、参考ページなど

本稿は以下の書籍、ページを参考に作られている。

- 統計分布ハンドブック（朝倉書店；ISBN 978-4254121544）：この種の話題のバイブル。高いけど。
- 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；ISBN 978-4903814353）：小さい字でびっしり。ジュンク堂書店で暗黒団フェアをやってもらったときに見つけた。リアル書店って大事だよ。
- 朱鷺の杜 wiki：http://ibisforest.org/index.php 機械学習の人たちの牙城。
- Wikipedia：毎度おなじみ。

確率分布のノート

2012 年 8 月 12 日 初版発行
2013 年 2 月 19 日 PDF 版公開

著　者　シンキロウ（しんきろう）
発行者　星野 香奈（ほしのかな）
発行所　同人集合　暗黒通信団（http://www.mikaka.org/~kana/）
〒277-8691 千葉県柏局私書箱 54 号 D 係
頒　価　0 円 / ISBN 978-4-87310-172-9 (C0041)

内容ミス等の指摘は遠慮なくお寄せください。随時訂正します。

©Copyright 2012-2013 暗黒通信団　　Printed in Japan