

# 色々な確率分布

- 黒木玄
- 2022-04-11~2022-04-12

このノートではJulia言語 (<https://julialang.org/>)を使用している:

- [Julia言語のインストールの仕方の一例](https://nbviewer.org/github/genkuroki/msfd28/blob/master/install.ipynb) (<https://nbviewer.org/github/genkuroki/msfd28/blob/master/install.ipynb>)

自明な誤りを見つけたら、自分で訂正して読んで欲しい。大文字と小文字の混同や書き直しが不完全な場合や符号のミスは非常によくある。

このノートの内容よりもさらに詳しいノートを自分で作ると勉強になるだろう。膨大な時間を取られることになるが、このノートの内容に関係することで飯を食っていく可能性がある人にはそのためにかけた時間は無駄にならないと思われる。

## 目次

- ▼ [1 正規分布](#)
  - [1.1 正規分布のロケーションスケール変換も正規分布](#)
  - [1.2 問題: 正規分布の平均と分散](#)
  - [1.3 問題: 正規分布に従う独立な確率変数達の和も正規分布に従う](#)
  - [1.4 問題: 正規分布における確率がほぼ95%または99%になる区間](#)
  - [1.5 問題: 対数正規分布の確率密度関数](#)
- ▼ [2 Poisson分布](#)
  - [2.1 二項分布の連続時間極限](#)
- ▼ [3 ガンマ分布](#)
  - [3.1 負の二項分布の連続時間極限](#)
  - [3.2  \$\chi^2\$ 分布](#)
- ▼ [4 ベータ分布](#)
  - [4.1 一様分布の順序統計量](#)
  - [4.2 t分布](#)
  - [4.3 F分布](#)
  - [4.4 二項分布とベータ分布の累積分布関数の関係](#)

```
In [1]: 1 ENV["LINES"], ENV["COLUMNS"] = 100, 100
2 using BenchmarkTools
3 using Distributions
4 using Printf
5 using QuadGK
6 using Random
7 Random.seed!(4649373)
8 using Roots
9 using SpecialFunctions
10 using StatsBase
11 using StatsFuns
12 using StatsPlots
13 default(fmt = :png, titlefontsize = 10, size = (400, 250))
14 using SymPy
```

## 1 正規分布

$\mu, \sigma \in \mathbb{R}, \sigma > 0$  と仮定する. 確率密度関数

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

で定義される連続分布を平均  $\mu$ , 分散  $\sigma^2$  (標準偏差  $\sigma$ ) の **正規分布** (normal distribution) と呼び,  $\text{Normal}(\mu, \sigma)$  と書くのであった.

さらに, 平均 0, 分散 1<sup>2</sup> の正規分布を **標準正規分布** (standard normal distribution) と呼び,  $\text{Normal}()$  と書くのであった. 標準正規分布の密度関数は次の形になる:

$$p(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}.$$

## 1.1 正規分布のロケーションスケール変換も正規分布

$a, b \in \mathbb{R}$ ,  $a \neq 0$  と仮定する. 一般に確率変数  $X$  について,

$$E[aX + b] = aE[X] + b, \quad \text{var}(aX + b) = a^2 \text{var}(X)$$

が成立するのであった.  $X$  が正規分布に従う確率変数の場合には  $aX + b$  も正規分布に従うことを示せる. 実際,  $X \sim \text{Normal}(\mu, \sigma)$  のとき,  $y = ax + b$  すなわち  $x = (y - b)/a$  とおくと,

$$\begin{aligned} E[f(aX + b)] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(ax + b) \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(y) \exp\left(-\frac{((y - b)/a - \mu)^2}{2\sigma^2}\right) \frac{dy}{|a|} \\ &= \frac{1}{\sqrt{2\pi(a^2\sigma^2)}} \int_{-\infty}^{\infty} f(y) \exp\left(-\frac{(y - (a\mu + b))^2}{2a^2\sigma^2}\right) dy \end{aligned}$$

なので,  $aX + b$  は平均  $a\mu + b$ , 分散  $a^2\sigma^2$  の正規分布に従う:  $aX + b \sim \text{Normal}(a\mu + b, |a|\sigma)$ .

## 1.2 問題: 正規分布の平均と分散

分布  $\text{Normal}(\mu, \sigma)$  の平均と分散がそれぞれ  $\mu, \sigma^2$  になることを示せ.

**解答例1:** すでに標準正規分布  $\text{Normal}(0, 1)$  の平均と分散がそれぞれ 0 と  $1^2$  になることは示してある.  $X \sim \text{Normal}(\mu, \sigma)$  のとき,  $Z = (X - \mu)/\sigma$  とおくと  $Z \sim \text{Normal}((\mu - \mu)/\sigma, \sigma/\sigma) = \text{Normal}(0, 1)$  となるので,  $X = \sigma Z + \mu$  の平均と分散はそれぞれ  $\sigma \cdot 0 + \mu = \mu$ ,  $\sigma^2 \cdot 1^2 = \sigma^2$  になる.

**解答終**

**解答例2:** 直接的に計算してみよう.  $X \sim \text{Normal}(\mu, \sigma)$  のとき, 積分変数を  $x = \sigma z + \mu$  と変換すると,

$$E[X] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/(2\sigma^2)} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu) e^{-z^2/2} dz = \mu.$$

最後の等号で  $\int_{-\infty}^{\infty} z e^{-z^2/2} dz = 0$  と  $\int_{-\infty}^{\infty} e^{-z^2/2} dy = \sqrt{2\pi}$  を使った. この結果を使うと,

$$\text{var}(X) = E[(X - \mu)^2] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-(x-\mu)^2/(2\sigma^2)} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 z^2 e^{-z^2/2} dz = \sigma^2.$$

最後の等号で  $\int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz = \sqrt{2\pi}$  を使った. この結果は

$$\int_{-\infty}^{\infty} e^{-\alpha z^2} dz = \sqrt{\pi} \alpha^{-1/2}$$

の両辺を  $\alpha$  で微分して  $-1$  倍して  $\alpha = 1/2$  においても得られるし, ガンマ函数に帰着する方法でも得られる.

**解答終**

## 1.3 問題: 正規分布に従う独立な確率変数達の和も正規分布に従う

$X, Y$  はともに正規分布に従う独立な確率変数達であるとき,  $X + Y$  も正規分布に従うことを示せ.  $X + Y$  の平均と分散は  $X, Y$  の平均と分散でどのように表されるか?

**解答例:**  $X \sim \text{Normal}(\mu_X, \sigma_X)$ ,  $Y \sim \text{Normal}(\mu_Y, \sigma_Y)$  でかつ  $X, Y$  は独立であると仮定する.

$X, Y$  は独立なので  $X + Y$  の平均と分散はそれぞれ  $\mu_X + \mu_Y$ ,  $\sigma_X^2 + \sigma_Y^2$  になる.

$X' = X - \mu_X$ ,  $Y' = Y - \mu_Y$  とおくと,  $X' \sim \text{Normal}(0, \sigma_X)$ ,  $Y' \sim \text{Normal}(0, \sigma_Y)$ ,  $X + Y = (X' + Y') + (\mu_X + \mu_Y)$  となるので,  $X' + Y'$  が正規分布に従うことを示せば十分である. ゆえに  $\mu_X = 0$ ,  $\mu_Y = 0$  と仮定してよいので, そのように仮定する.

$X + Y$  の確率密度函数を計算して, それが正規分布の密度函数になっていることを示せばよい.

$$E[f(X + Y)] = \frac{1}{\sqrt{(2\pi)^2 \sigma_X^2 \sigma_Y^2}} \iint_{\mathbb{R}^2} f(x + y) \exp\left(-\frac{1}{2} \left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2}\right)\right) dx dy$$

であり,  $x = x + y$  すなわち  $y = z - x$  において,  $x$  について平方完成すると,

$$\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2} = \frac{x^2}{\sigma_X^2} + \frac{(z - x)^2}{\sigma_Y^2} = \frac{\sigma_X^2 + \sigma_Y^2}{\sigma_X^2 \sigma_Y^2} \left(x - \frac{\sigma_X^2 z}{\sigma_X^2 + \sigma_Y^2}\right)^2 + \frac{z^2}{\sigma_X^2 + \sigma_Y^2}.$$

となる(この計算を自分で実行してみることに). ゆえに,

$$\begin{aligned} E[f(X+Y)] &= \frac{1}{\sqrt{(2\pi)^2 \sigma_X^2 \sigma_Y^2}} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} f(z) \exp\left(-\frac{\sigma_X^2 + \sigma_Y^2}{2\sigma_X^2 \sigma_Y^2} \left(x - \frac{\sigma_X^2 z}{\sigma_X^2 + \sigma_Y^2}\right)^2 - \frac{z^2}{2(\sigma_X^2 + \sigma_Y^2)}\right) dx \right) dz \\ &= \frac{1}{\sqrt{(2\pi)^2 \sigma_X^2 \sigma_Y^2}} \sqrt{\frac{2\pi \sigma_X^2 \sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}} \int_{\mathbb{R}} f(z) \exp\left(-\frac{z^2}{2(\sigma_X^2 + \sigma_Y^2)}\right) dz \\ &= \frac{1}{\sqrt{2\pi(\sigma_X^2 + \sigma_Y^2)}} \int_{\mathbb{R}} f(z) \exp\left(-\frac{z^2}{2(\sigma_X^2 + \sigma_Y^2)}\right) dz \end{aligned}$$

以上によって  $X+Y \sim \text{Normal}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$  を示せた.

$z = x + y, y = z - x$  とおくと密度関数中の指数関数の中身が  $z$  についての二次式になることから,  $Z = X + Y$  が正規分布に従うことは計算しなくても明らかだと考えることもできる. 上のように  $z$  以外の変数は積分して消せる. そして,  $Z = X + Y$  の期待値は  $X, Y$  の期待値の和になり,  $X, Y$  が独立という仮定からそれらは無相関になるので  $Z = X + Y$  の分散は  $X, Y$  の分散の和になることもわかる. これだけで  $X+Y \sim \text{Normal}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$  を証明でいたともみなせる.

### 解答終

上の解答中の最も面倒な部分の計算が正しいことはコンピュータで以下のように確認できる.

$a = \sigma_X^2, b = \sigma_Y^2$  とおくと,

```
In [2]: 1 @vars a b x z
        2 (a+b)/(a*b)*(x - a*z/(a+b))^2 + z^2/(a+b) - x^2/a - (z-x)^2/b > simplify
```

Out[2]: 0

以下のように素朴に計算することもできる.

```
In [3]: 1 @vars a b positive=true
        2 @vars x y z t
        3 expr = x^2/a + y^2/b
```

Out[3]:  $\frac{y^2}{b} + \frac{x^2}{a}$

```
In [4]: 1 expr = expr(y=>z-x).expand()
```

Out[4]:  $\frac{x^2}{b} - \frac{2xz}{b} + \frac{z^2}{b} + \frac{x^2}{a}$

```
In [5]: 1 A, B = 1/a+1/b, z/b
        2 expr = sympy.poly(expr(x => t + B/A), t)
```

Out[5]:  $\text{Poly}\left(\frac{a+b}{ab}t^2 + \frac{z^2}{a+b}, t, \text{domain} = \mathbb{Z}(z, a, b)\right)$

```

In [6]: 1 # 以下のセルでの説明で使う図の準備
2
3  $\mu, \sigma, c = 2, 3, 1.5$ 
4 normal = Normal( $\mu, \sigma$ )
5
6 P1 = plot(normal,  $\mu - 4\sigma, \mu + 4\sigma$ ; label="Normal( $\mu, \sigma$ )", xlabel=" $x = \sigma z + \mu$ ")
7 vline!([ $\mu$ ]; label=" $x=\mu$ ", xtick = ([ $\mu - c\sigma, \mu, \mu + c\sigma$ ], [ $\mu - c\sigma, \mu, \mu + c\sigma$ ]), ytick=false)
8 plot!(normal,  $\mu - 4\sigma, \mu - c\sigma$ ; label="", c=1, frange=0, fc=:red, fa=0.5)
9 plot!(normal,  $\mu + c\sigma, \mu + 4\sigma$ ; label="", c=1, frange=0, fc=:red, fa=0.5)
10 annotate!( $\mu - 1.3c\sigma, 0.8pdf(normal, \mu - c\sigma)$ , (" $P(X \leq \mu - c\sigma)$ ", 10, :red, :right))
11 annotate!( $\mu - 1.5c\sigma, 0.5pdf(normal, \mu - c\sigma)$ , (" $= \alpha/2$ ", 10, :red, :right))
12 annotate!( $\mu + 1.5c\sigma, 0.5pdf(normal, \mu + c\sigma)$ , (" $= \alpha/2$ ", 10, :red, :left))
13 annotate!( $\mu + 1.3c\sigma, 0.8pdf(normal, \mu + c\sigma)$ , (" $1 - P(X \geq \mu + c\sigma)$ ", 10, :red, :left))
14
15 P2 = plot(Normal(), -4, 4; label="Normal(0,1)", xlabel=" $z = (x - \mu)/\sigma$ ")
16 vline!([0]; label=" $z=0$ ", xtick = ([-c, 0, c], [-c, "0", "c"]), ytick=false)
17 plot!(Normal(), -4, -c; label="", c=1, frange=0, fc=:red, fa=0.5)
18 plot!(Normal(), c, 4; label="", c=1, frange=0, fc=:red, fa=0.5)
19 annotate!(-1.3c, 0.8pdf(Normal(), -c), (" $P(Z \leq -c)$ ", 10, :red, :right))
20 annotate!(-1.5c, 0.5pdf(Normal(), -c), (" $= \alpha/2$ ", 10, :red, :right))
21 annotate!(1.5c, 0.5pdf(Normal(), c), (" $= \alpha/2$ ", 10, :red, :left))
22 annotate!(1.3c, 0.8pdf(Normal(), c), (" $1 - P(Z \geq c)$ ", 10, :red, :left))
23
24 var"P( $\mu - c\sigma \leq X \leq \mu + c\sigma$ ) =  $1 - \alpha \Leftrightarrow 1 - P(Z \leq c) = \alpha/2$ " =
25     plot(P1, P2; size=(800, 250), bottommargin=4Plots.mm);

```

#### 1.4 問題: 正規分布における確率がほぼ95%または99%になる区間

$X \sim \text{Normal}(\mu, \sigma)$  であると仮定する. このとき,  $X$  が区間  $[\mu - c\sigma, \mu + c\sigma]$  に含まれる確率

$$P(\mu - c\sigma \leq X \leq \mu + c\sigma)$$

が 95% になる  $c$  と 99% になる  $c$  を小数点以下第2桁目まで求めよ.

**解答例:** 標準正規分布の場合の計算に帰着することを考えよう.

$Z = (X - \mu)/\sigma$  とおくと  $Z$  は標準正規分布に従うのであった:  $Z \sim \text{Normal}(0, 1)$ . このとき,

$$\mu - c\sigma \leq X \leq \mu + c\sigma \iff -c \leq Z = \frac{X - \mu}{\sigma} \leq c$$

なので, 標準正規分布に従う確率変数  $Z$  について  $P(-c \leq Z \leq c)$  が 95% になる  $c$  と 99% になる  $c$  を求めればよい. 正規分布は左右対称なので,  $P(-c \leq Z \leq c) = 1 - \alpha$  となることと,  $1 - P(Z \leq c) = \alpha/2$  すなわち

$$P(Z \leq c) = 1 - \alpha/2$$

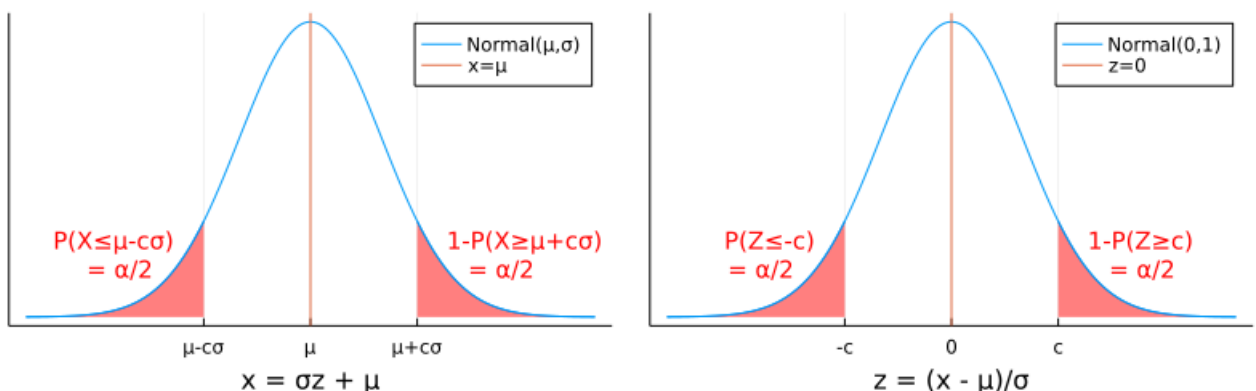
となることは同値である. 下の図を見よ.

```

In [7]: 1 var"P( $\mu - c\sigma \leq X \leq \mu + c\sigma$ ) =  $1 - \alpha \Leftrightarrow 1 - P(Z \leq c) = \alpha/2$ "

```

Out[7]:



$F(z) = P(Z \leq z)$  は標準正規分布の累積分布関数である. 標準正規分布の累積分布関数はコンピュータでの基本特殊関数ライブラリに含まれている誤差関数

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2) du$$

を使えば

$$F(z) = \frac{1 + \operatorname{erf}(z/\sqrt{2})}{2}.$$

と書けるのであった。ゆえに誤差関数の逆関数  $\operatorname{erfinv}(y)$  (この関数のコンピュータでの基本特殊関数ライブラリに含まれている)を使えば、標準正規分布の累積分布関数  $F(z)$  の逆関数(分位点関数)は

$$Q(p) = F^{-1}(p) = \sqrt{2} \operatorname{erfinv}(2p - 1)$$

と書ける。これを使えば  $P(Z \leq c) = 1 - \alpha/2$  となる  $c$  を

$$c = Q(1 - \alpha/2) = \sqrt{2} \operatorname{erfinv}(1 - \alpha)$$

と求めることができる。

- $1 - \alpha = 95\%$  のとき,  $c = \sqrt{2} \operatorname{erfinv}(1 - \alpha) \approx 1.96$
- $1 - \alpha = 99\%$  のとき,  $c = \sqrt{2} \operatorname{erfinv}(1 - \alpha) \approx 2.58$

**解答終**

[Julia言語 \(https://julialang.org/\)](https://julialang.org/)では以下のように計算できる。(ただし, using SpecialFunctions, Distributions が必要.)

```
In [8]: 1 √2 * erfinv(0.95), quantile(Normal(), 0.975)
```

```
Out[8]: (1.9599639845400576, 1.9599639845400576)
```

```
In [9]: 1 √2 * erfinv(0.99), quantile(Normal(), 0.995)
```

```
Out[9]: (2.5758293035489053, 2.5758293035489053)
```

[WolframAlpha \(https://www.wolframalpha.com/\)](https://www.wolframalpha.com/)では以下のように計算できる。

- $\sqrt{2} \operatorname{erfinv}(0.95)$  (<https://www.wolframalpha.com/input?i=%E2%88%9A2+erfinv%280.95%29>),  $\operatorname{quantile}(\operatorname{NormalDistribution}(0,1), 0.975)$  (<https://www.wolframalpha.com/input?i=quantile%28NormalDistribution%280%2C1%29%2C+0.975%29>).
- $\sqrt{2} \operatorname{erfinv}(0.99)$  (<https://www.wolframalpha.com/input?i=%E2%88%9A2+erfinv%280.99%29>),  $\operatorname{quantile}(\operatorname{NormalDistribution}(0,1), 0.995)$  (<https://www.wolframalpha.com/input?i=quantile%28NormalDistribution%280%2C1%29%2C+0.995%29>).

## 1.5 問題: 対数正規分布の確率密度関数

$X \sim \operatorname{Normal}(\mu, \sigma)$  のときの  $Y = e^X$  が従う分布を対数正規分布と呼び,

$$\operatorname{LogNormal}(\mu, \sigma)$$

と表す。対数正規分布の確率密度関数を求めよ。

**解答例:**  $x = \log y$  と積分変数を変換すると,  $dx = dy/y$  なので,

$$E[f(Y)] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(e^x) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(y) \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right) \frac{dy}{y}.$$

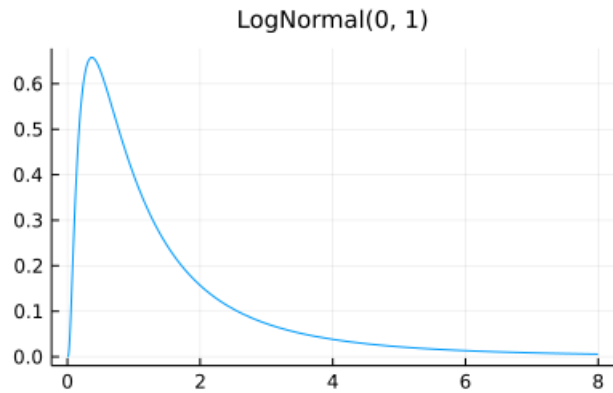
ゆえに対数正規分布の確率密度関数は次の形になる:

$$p(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right).$$

**解答終**

```
In [10]: 1 plot(LogNormal(0, 1), 0, 8; label="", title="LogNormal(0, 1)")
```

Out[10]:



## 2 Poisson分布

### 2.1 二項分布の連続時間極限

## 3 ガンマ分布

### 3.1 負の二項分布の連続時間極限

### 3.2 $\chi^2$ 分布

## 4 ベータ分布

### 4.1 一様分布の順序統計量

### 4.2 t分布

### 4.3 F分布

### 4.4 二項分布とベータ分布の累積分布関数の関係

In [ ]:

1