

確率分布

第8講 - いろいろな離散分布と連続分布

村田 昇

講義概要

- 一般の確率変数
- 離散分布
 - 離散一様分布・二項分布
 - Poisson 分布・幾何分布
- 連続分布
 - 一様分布・正規分布
 - ガンマ分布・ t 分布・ F 分布

一般の確率変数

確率分布

- 確率変数 X
 - 値がランダムに決定される変数
 - 任意の区間 $[a, b]$ に含まれる確率が定められている
- 確率分布 P (または単に分布)
 - 区間 $[a, b]$ ($a \leq b$) に X が含まれる確率

$$P(a \leq X \leq b)$$

- 「確率変数 X は確率分布 P に従う」という
- 現象の理解のためには実現値より確率分布にこそ興味がある

極限定理

- 大数の法則, 中心極限定理, 少数の法則など
- 緩やかな条件のもとで一般の確率変数の列についても成立する
 - 大数の強法則は平均が存在すれば成立する
(平均と分散は後に定義する)
 - 2次のモーメントをもたない (計算できない) 場合には中心極限定理は成立しない
(そもそも分散が定義できない)

確率変数列の独立性

- 定義

n 個の確率変数 X_1, X_2, \dots, X_n が **独立** であるとは, $a_i \leq b_i$ ($i = 1, \dots, n$) なる任意の実数 $a_1, b_1, \dots, a_n, b_n$ に対して

$$\begin{aligned} P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \dots, a_n \leq X_n \leq b_n) \\ = P(a_1 \leq X_1 \leq b_1)P(a_2 \leq X_2 \leq b_2) \cdots P(a_n \leq X_n \leq b_n) \end{aligned}$$

が成り立つことをいう.

確率変数列の同分布性

- 定義

n 個の確率変数 X_1, X_2, \dots, X_n が **同分布** であるとは, $a \leq b$ なる任意の実数 a, b に対して

$$P(a \leq X_1 \leq b) = P(a \leq X_2 \leq b) = \cdots = P(a \leq X_n \leq b)$$

が成り立つことをいう.

離散分布

離散分布

- 離散型確率変数
 - 取りうる値が有限個
もしくは可算無限個 (例えば整数値のみとる場合)
- 離散分布
 - 離散型確率変数の確率分布
 - $X = x$ となる確率 $P(X = x)$ で完全に決定される
 - 関数 $f(x) = P(X = x)$ を **確率質量関数 (確率関数)** と呼ぶ

離散分布の平均

- 定義

以下の右辺が定義できる (絶対収束する) とき離散型の確率変数 X の **平均** (または **期待値**) を

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} xP(X = x)$$

で定義する.

ただし \mathcal{X} は X の取りうる値 (無限個も可) の集合

- 級数 $\sum_{x \in \mathcal{X}} xP(X = x)$ が定義できない場合は X は **平均をもたない** とする

- 確率変数の変換 $\varphi(X)$ の期待値

$$\mathbb{E}[\varphi(X)] = \sum_{x \in \mathcal{X}} \varphi(x)P(X = x)$$

- p 次の **モーメント** (あるいは **積率**)

$$\mathbb{E}[X^p] = \sum_{x \in \mathcal{X}} x^p P(X = x)$$

- X が正整数 p 次のモーメントをもてば, $q \leq p$ なるすべての正整数 q 次のモーメントをもつ

離散分布の分散

- 定義

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \sum_{x \in \mathcal{X}} (x - \mathbb{E}[X])^2 P(X = x)\end{aligned}$$

- X が 2 次のモーメントをもつ ときのみ定義できる
- 分散の平方根 $\sqrt{\text{Var}(X)}$ を **標準偏差** と呼ぶ

離散一様分布

離散一様分布の確率質量関数

- 集合 $\{x_1, \dots, x_n\}$ 上の離散一様分布
相異なる実数 x_1, \dots, x_n を箱 (壺) から無作為に取り出す試行
- 確率質量関数

$$f(x) = \frac{1}{n}, \quad x \in \{x_1, \dots, x_n\}$$

離散一様分布の平均と分散

- 平均

$$\mathbb{E}[X] = \sum_{i=1}^n x_i f(x_i) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

- 分散

$$\text{Var}(X) = \sum_{i=1}^n (x_i - \mathbb{E}[X])^2 f(x_i) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

離散一様分布に従う乱数

- 歪みのないサイコロを 1 回投げたときに出る目の分布は、集合 $\{1, \dots, 6\}$ 上の離散一様分布に従う
 - 乱数の発生は関数 `sample()` が利用できる
(`replace=TRUE`; 復元抽出)
- ```
a <- 1:6 # サンプリング対象の集合をベクトルとして定義
sample(a, size=20, replace=TRUE) # 離散一様分布 (20 個)
```

## 二項分布

### 二項分布の確率質量関数

- 試行回数  $n$ , 成功確率  $p$  の二項分布  
確率  $p$  で成功する試行を  $n$  回行った総成功回数
- 確率質量関数

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$$

## 二項分布の平均

- 平均

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x=0}^n x f(x) = \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \\ &= np \sum_{x=0}^{n-1} \binom{n-1}{x} p^x (1-p)^{n-1-x} \\ &= np\end{aligned}$$

## 二項分布の分散

- 2 次モーメント

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{x=0}^n x^2 f(x) = \sum_{x=0}^n x(x-1) \binom{n}{x} p^x (1-p)^{n-x} + \mathbb{E}[X] \\ &= n(n-1)p^2 \sum_{x=0}^{n-2} \binom{n-2}{x} p^x (1-p)^{n-2-x} + np \\ &= n(n-1)p^2 + np = (np)^2 + np(1-p)\end{aligned}$$

- 分散

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = np(1-p)$$

## 二項分布に従う乱数

- 表が出る確率が  $p$  のコインを  $n$  回投げたときに表が出る回数は、試行回数  $n$ 、成功確率  $p$  の二項分布に従う
- 試行回数 1 の二項分布を **Bernoulli 分布** と呼ぶ
- 乱数の発生は関数 `rbinom()` を用いる

```
rbinom(10, size=1, prob=0.2) # Bernoulli 分布 (10 個)
rbinom(20, size=5, prob=0.6) # 二項分布 (20 個)
```

```
[1] 0 0 0 0 1 0 0 0 1 0
[1] 2 2 4 2 4 2 2 4 5 3 2 5 2 3 2 3 5 2 3
```

- 確率質量関数は関数 `dbinom()` で計算できる

## 実習

### 数値実験の雛形 (離散分布)

- 人工的に生成した乱数と理論曲線 (確率質量関数) との比較を行う
- 数値実験の雛形は以下のようになる

```
mc <- 10000 # 実験回数を指定
my_random <- function(){...} # 乱数生成のプログラム
my_dist <- function(){...} # 必要であれば理論曲線を定義
#' ただし多くの場合 d"乱数名"で定義されているのでそれらを用いればよい
my_data <- replicate(mc, my_random()) # Monte-Carlo 実験
my_table <- table(my_data)/mc # 同じ値の頻度から確率を計算
tibble(count = as.numeric(names(my_table)),
 observation = as.numeric(my_table),
 theory = dbinom(値, size = n, prob = p)) |>
pivot_longer(!count, values_to = "probability") |>
ggplot(aes(x = count, y = probability, fill = name)) +
geom_bar(stat = "identity", width = 0.8, position = "dodge")
```

## 練習問題

- 二項分布に関して以下を考察せよ.
  - 平均と分散の計算を確認しなさい.
  - $X_1, \dots, X_n$  を成功確率  $p$  の Bernoulli 分布に従う独立同分布な確率変数列とする. このとき  $\sum_{i=1}^n X_i$  の分布は, 試行回数  $n$ , 成功確率  $p$  の二項分布に従う. これをグラフに描画して確認しなさい.

## Poisson 分布

### Poisson 分布の確率質量関数

- 強度  $\lambda$  の Poisson 分布  $Po(\lambda)$  ( $\lambda > 0$ )  
少数の法則に現れた減多に起こらない事象の起こる回数の分布
- 確率質量関数

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, \dots$$

### Poisson 分布の平均

- 平均

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x=0}^{\infty} x f(x) = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \\ &= \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \lambda \end{aligned}$$

### Poisson 分布の分散

- 2 次モーメント

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{x=0}^{\infty} x^2 f(x) = \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x}{x!} e^{-\lambda} + \mathbb{E}[X] \\ &= e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!} + \lambda = \lambda^2 e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} + \lambda \\ &= \lambda^2 + \lambda \end{aligned}$$

- 分散

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda$$

## Poisson 分布に従う乱数

- 発生確率が低い事象が十分長い期間で起こる回数の分布は Poisson 分布で近似できる (少数の法則)
  - 一定時間に放射性物質から放射される粒子の数
  - 一定時間に発生する交通事故の件数
- 乱数の発生は関数 `rpois()` を用いる

```
rpois(15, lambda=1) # 強度 1 の Poisson 分布 (15 個)
rpois(15, lambda=10) # 強度 10 の Poisson 分布 (15 個)
```

```
[1] 3 3 2 1 1 2 0 2 1 0 1 0 1 0 2
[1] 10 14 11 12 9 9 12 8 7 8 15 12 8 17 10
```

- 確率質量関数は関数 `dpois()` で計算できる

## 実習

### 練習問題

- Poisson 分布に関して以下を考察せよ.
  - 平均と分散の計算を確認しなさい.
  - $X, Y$  を独立な 2 つの確率変数とし, それぞれ強度  $\lambda_1, \lambda_2$  の Poisson 分布に従うとする. このとき, 和  $X + Y$  の分布は強度  $\lambda_1 + \lambda_2$  の Poisson 分布に従う. これをグラフに描画して確認しなさい.

## 幾何分布

#' @exercise 幾何分布で用いた例

### 幾何分布の確率質量関数

- 成功確率  $p$  の幾何分布 ( $0 < p \leq 1$ )  
成功するまでの回数の分布
- 確率質量関数

$$f(x) = p(1-p)^x, \quad x = 0, 1, \dots$$

### 幾何分布の平均

- 平均

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x=0}^{\infty} x f(x) = \sum_{x=0}^{\infty} x p (1-p)^x = p(1-p) \sum_{x=1}^{\infty} x (1-p)^{x-1} \\ &= -p(1-p) \frac{d}{dp} \sum_{x=1}^{\infty} (1-p)^x = -p(1-p) \frac{d}{dp} \frac{1-p}{p} \\ &= -p(1-p) \cdot \left(-\frac{1}{p^2}\right) = \frac{1-p}{p} \end{aligned}$$

## 幾何分布の分散

- 2 次モーメント

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{x=0}^{\infty} x^2 f(x) = \sum_{x=0}^{\infty} x(x-1)p(1-p)^x + \mathbb{E}[X] \\ &= 2\left(\frac{1-p}{p}\right)^2 + \frac{1-p}{p}\end{aligned}$$

- 分散

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1-p}{p^2}$$

## 幾何分布に従う乱数

- 表が出る確率が  $p$  のコインを投げ続けて、初めて表が出るまでに表の出た裏の回数は、成功確率  $p$  の幾何分布に従う
- 乱数の発生は関数 `rgeom()` を用いる

```
rgeom(15, prob=0.1) # 成功確率 0.1 の幾何分布 (15 個)
```

```
[1] 11 11 2 17 1 11 8 7 3 4 4 8 3 26 1
```

- 確率質量関数は関数 `dgeom()` で計算できる

## 連続分布

### 連続分布

- 連続型確率変数
  - 取りうる値が任意の実数またはある範囲の実数
  - 取りうる値のパターンが数多いため実数で近似
- 連続分布
  - 連続型の確率変数の確率分布
  - 非負の値をとる実数上の関数  $f$  の積分で表される

$$P(a \leq X \leq b) = \int_a^b f(x)dx \quad (a < b)$$

- 関数  $f$  を **確率密度関数** (または **密度**) と呼ぶ

### 連続分布の平均

- 連続型の確率変数  $X$  の平均  
積分  $\int_{-\infty}^{\infty} xf(x)dx$  が絶対収束するとき  $X$  の **平均** (または **期待値**) を

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx$$

で定義する

- 右辺の積分が絶対収束しない場合は  $X$  は **平均をもたない** とする
- 確率変数の変換  $\varphi(X)$  の期待値

$$\mathbb{E}[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) f(x) dx$$

- $p$  次の **モーメント** (あるいは **積率**)

$$\mathbb{E}[X^p] = \int_{-\infty}^{\infty} x^p f(x) dx$$

- $X$  が正整数  $p$  次のモーメントをもてば、 $q \leq p$  なるすべての正整数  $q$  次のモーメントをもつ

## 連続分布の分散

- 分散

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f(x) dx \end{aligned}$$

- $X$  が **2 次のモーメントをもつ** ときのみ定義できる
- 分散の平方根  $\sqrt{\text{Var}(X)}$  を **標準偏差** と呼ぶ

## 一様分布

### 一様分布の確率密度関数

- 区間  $(a, b)$  上の一様分布  $U(a, b)$   
区間  $(a, b)$  のランダムな値の選択を表す分布
- 確率密度関数

$$f(x) = \begin{cases} \frac{1}{b-a} & (a < x < b) \\ 0 & (\text{上記以外のとき}) \end{cases}$$

### 一様分布の平均

- 平均

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{b-a} \int_a^b x dx \\ &= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \end{aligned}$$



## 一様分布の分散

- 2 次モーメント

$$\begin{aligned}\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{1}{b-a} \int_a^b x^2 dx \\ &= \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}\end{aligned}$$

- 分散

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{(b-a)^2}{12}$$

## 一様分布に従う乱数

- 乱数の発生は関数 `runif()` を用いる

```
runif(8) # 区間 (0,1) 上の一様乱数 (8 個)
runif(8,min=-1,max=1) # 区間 (-1,1) 上の一様乱数 (8 個)
```

```
[1] 0.50732772 0.90798226 0.73639084 0.73641601 0.05024919 0.49685352 0.69249992
[8] 0.38717180
[1] -0.341971243 0.008502312 0.210793295 -0.880408118 -0.896585018 0.086961693
[7] -0.132189428 -0.911392471
```

- 確率密度関数は関数 `dunif()` で計算できる

## 正規分布

### 正規分布の確率密度関数

- 平均  $\mu$ , 分散  $\sigma^2$  の正規分布  $N(\mu, \sigma^2)$   
( $\mu$  は実数,  $\sigma$  は正の実数)
- 確率密度関数

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

- 平均 0, 分散 1 の正規分布を **標準正規分布** と呼ぶ
- **Gauss 分布** と呼ぶこともある

### 正規分布の特徴付け

- 物理実験等の観測誤差の分布はしばしば正規分布でモデル化される
- 真の平均を標本平均で推定した際の推定誤差の確率分布は, サンプル数が大きくなるに従って正規分布に近づいていく (**中心極限定理**)
- 中心極限定理の特殊な形 (de Moivre-Laplace の定理)
  - $X_1, \dots, X_n$  を成功確率  $p$  の Bernoulli 分布に従う独立同分布な確率変数列とする
  - $Y = \sum_{i=1}^n X_i$  は試行回数  $n$ , 成功確率  $p$  の二項分布に従う

- Bernoulli 分布は平均  $p$ , 分散  $p(1-p)$  である
- 中心極限定理により

$$\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\frac{1}{n} \sum_{i=1}^n X_i - p)}{\sqrt{p(1-p)}} = \frac{Y - np}{\sqrt{np(1-p)}}$$

の分布は標準正規分布で近似できる

- $p$  が非常に小さい場合, 特に  $np$  がそれほど大きくならない程度に  $p$  が小さい場合は,  $(Y - np)/\sqrt{np(1-p)}$  の分布の正規近似よりも,  $Y$  の分布は強度  $np$  の Poisson 分布による近似の方が精度がよい (少数の法則)

## 正規分布に従う乱数

- 乱数の発生は関数 `rnorm()` を用いる

```
rnorm(8) # 標準正規乱数 (8 個)
rnorm(8, mean=1, sd=2) # 平均 1 分散 4=2^2 の正規乱数
```

```
[1] -0.4439001 1.4403117 -0.7214356 0.6895706 2.1746420 1.6652812 -0.2016271
[8] -0.7689908
[1] 2.0004849 2.8423186 3.0009585 1.0171984 0.4080415 2.6919563 0.5315107
[8] -0.6322620
```

- 確率密度関数は関数 `dnorm()` で計算できる

## 実習

### 数値実験の雛形 (連続)

- 人工的に生成した乱数と理論曲線 (確率密度) との比較を行う
- 数値実験の雛形は以下のようになる

```
mc <- 10000 # 実験回数を指定
my_random <- function(){...} # 乱数生成のプログラム
my_dist <- function(){...} # 必要であれば理論曲線を定義
'ただし多くの場合 d"乱数名"で定義されているのでそれらを用いればよい
my_data <- replicate(mc, my_random()) # Monte-Carlo 実験
tibble(x = my_data) |>
 ggplot(aes(x = x)) +
 geom_histogram(aes(y = after_stat(density))) + # 乱数の分布 (密度) の表示
 geom_function(fun = my_dist) # 理論曲線 (確率密度)
```

## 練習問題

- 正規分布に関して以下を考察せよ.
  - $U_1, U_2$  を  $(0, 1)$  上の一様分布に従う独立な確率変数とする. このとき

$$X_1 = \sqrt{-2 \log(U_1)} \cos(2\pi U_2),$$

$$X_2 = \sqrt{-2 \log(U_1)} \sin(2\pi U_2)$$

とおくと,  $X_1, X_2$  は独立かつともに標準正規分布に従う. これをグラフに描画して確認せよ. (この変換を Box-Muller 変換と呼ぶ)

## ガンマ分布

### ガンマ分布の確率密度関数

- パラメータ  $\nu, \alpha$  のガンマ分布  
 $\Gamma(\nu, \alpha)$  または  $G(\alpha, \nu)$  ( $\nu > 0, \alpha > 0$ )
- 確率密度関数

$$f(x) = \begin{cases} \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

ただしガンマ関数  $\Gamma(\nu)$  は以下で定義される：

$$\Gamma(\nu) = \int_0^\infty x^{\nu-1} e^{-x} dx$$

### ガンマ分布の性質

- $\nu, \alpha$  はそれぞれ **形状パラメータ**，**レート** と呼ばれることがある
- 平均は  $\nu/\alpha$ ，分散は  $\nu/\alpha^2$  で与えられる
- ガンマ分布は応用上重要な確率分布を含む
  - 指数分布
  - $\chi^2$  分布

### ガンマ分布に従う乱数

- 乱数の発生は `rgamma()` を用いる

```
rgamma(8, shape=3, rate=1) # ガンマ分布 (8 個)
rgamma(8, shape=1, rate=3) # 異なるパラメタのガンマ分布 (8 個)
```

```
[1] 2.984398 1.686583 2.993885 3.650005 2.595698 6.104866 2.838832 5.719568
[1] 0.009950937 0.025301012 0.450376679 0.051349526 0.110350411 0.020112835
[7] 0.673363996 0.042058443
```

- 確率密度関数は `dgamma()` で計算できる

### 指数分布の確率密度関数

- パラメータ  $\lambda$  の指数分布  $\text{Exp}(\lambda)$  ( $\lambda > 0$ )

$$f(x) = \Gamma(1, \lambda) = \begin{cases} \lambda e^{-\lambda x} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

- $\lambda$  は **レート** と呼ばれることがある
- 平均，分散はそれぞれ  $\lambda^{-1}, \lambda^{-2}$  で与えられる

### 指数分布に従う乱数

- 乱数の発生は関数 `rexp()` を用いる

```
rexp(8) # レート 1 の指数分布 (8 個)
rexp(8, rate=0.5) # レート 0.5 の指数分布 (8 個)
```

```
[1] 0.3381164 0.3488348 1.0111424 0.3026318 2.5416492 0.1196099 0.5038373 1.2680349
[1] 0.695594249 4.104401875 3.630450237 0.910081916 2.048596868 0.004995573
[7] 0.387011799 0.250865612
```

## $\chi^2$ 分布の確率密度関数

- 自由度  $k$  の  $\chi^2$  分布  $\chi^2(k)$  ( $k > 0$ )

$$f(x) = \Gamma(k/2, 1/2) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

## $\chi^2$ 分布の特徴付け

- $\chi^2$  は **カイ二乗** と読む
- 平均, 分散はそれぞれ  $k, 2k$  で与えられる
- 標準正規分布に従う  $k$  個の独立な確率変数の二乗和は自由度  $k$  の  $\chi^2$  分布に従う

## $\chi^2$ 分布に従う乱数

- 乱数の発生は関数 `rchisq()` を用いる

```
rchisq(8, df=1) # 自由度 1 のカイ二乗分布 (8 個)
rchisq(8, df=4) # 自由度 4 のカイ二乗分布 (8 個)
```

```
[1] 0.07370854 0.20502638 3.59183176 0.30870410 1.05413412 2.11534768 1.74071856
[8] 0.55553581
[1] 2.132675 4.051093 2.504893 5.211969 2.870498 4.387340 7.612965 8.992941
```

- 確率密度関数は関数 `dchisq()` で計算できる

## 実習

### 練習問題

- $\chi^2$  分布に関して以下を考察せよ.
  - 標準正規分布に従う  $k$  個の独立な確率変数の二乗和は自由度  $k$  の  $\chi^2$  分布に従うことを確認しなさい.

## $t$ 分布

### $t$ 分布の確率密度関数

- 自由度  $\nu$  の (Student の)  $t$  分布  $t(\nu)$  ( $\nu > 0$ )
- 確率密度関数

$$f(x) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$$

- Student** は  $t$  分布を導入した統計学者 Gosset のペンネーム

## $t$ 分布の平均と分散

- 平均

$$\mathbb{E}[X] = 0 \quad (\text{ただし } \nu > 1)$$

- 分散

$$\text{Var}(X) = \frac{\nu}{\nu - 2} \quad (\text{ただし } \nu > 2)$$

## $t$ 分布の特徴付け

- $Z$  を標準正規分布に従う確率変数,  $Y$  を自由度  $k$  の  $\chi^2$  分布に従う確率変数とし,  $Z, Y$  は独立であるとする. このとき確率変数

$$\frac{Z}{\sqrt{Y/k}}$$

は自由度  $k$  の  $t$  分布に従う

## $t$ 分布に従う乱数

- 乱数の発生は関数 `rt()` を用いる

```
rt(8, df=1) # 自由度 1 の t 分布 (8 個)
rt(8, df=4) # 自由度 4 の t 分布 (8 個)
```

```
[1] -0.7688987 -0.6447764 -2.0438017 5.4066950 -0.3202864 -0.7711316 -1.4888384
[8] -4.2119137
[1] -0.8268906 -2.0866210 1.6772290 -0.3509583 -1.8560248 -0.1643798 1.2005594
[8] -0.2846538
```

- 確率密度関数は関数 `dt()` で計算できる

## 実習

### 練習問題

- $t$  分布に関して以下を考察せよ.
  - $Z$  を標準正規分布に従う確率変数,  $Y$  を自由度  $k$  の  $\chi^2$  分布に従う確率変数とし,  $Z, Y$  は独立であるとする. このとき確率変数

$$\frac{Z}{\sqrt{Y/k}}$$

は自由度  $k$  の  $t$  分布に従うことを確認しなさい.

## F 分布

### F 分布の確率密度関数

- 自由度  $\nu_1, \nu_2$  の F 分布  $F(\nu_1, \nu_2)$  ( $\nu_1 > 0, \nu_2 > 0$ )
- 確率密度関数

$$f(x) = \begin{cases} \frac{(\nu_1/\nu_2)^{\nu_1/2}}{B(\nu_1/2, \nu_2/2)} \frac{x^{\nu_1/2-1}}{(1+\nu_1 x/\nu_2)^{(\nu_1+\nu_2)/2}} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

### F 分布の平均と分散

- 平均

$$\mathbb{E}[X] = \frac{\nu_2}{\nu_2 - 2} \quad (\text{ただし } \nu_2 > 2)$$

- 分散

$$\text{Var}(X) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \quad (\text{ただし } \nu_2 > 4)$$

### F 分布の特徴付け

- $Y_1, Y_2$  をそれぞれ自由度  $k_1, k_2$  の  $\chi^2$  分布に従う独立な確率変数とする。このとき、確率変数

$$\frac{Y_1/k_1}{Y_2/k_2}$$

は自由度  $k_1, k_2$  の F 分布に従う

### F 分布に従う乱数

- 乱数の発生は関数 `rf()` を用いる

```
rf(10, df1=4, df2=7) # 自由度 4, 7 の F 分布 (10 個)
rf(10, df1=7, df2=12) # 自由度 7, 12 の F 分布 (10 個)
```

```
[1] 0.4476848 3.0426270 0.6468350 2.1124276 1.7329053 1.7308992 1.9159838 4.0724373
[9] 8.1904018 0.3056619
[1] 0.6543048 6.7186621 1.3987511 1.9454726 1.0256459 0.5493811 0.7942002 0.9367187
[9] 0.6611155 0.7070424
```

- 確率密度関数は関数 `df()` で計算できる

## 実習

### 練習問題

- F 分布に関して以下を考察せよ。
  - $Y_1, Y_2$  をそれぞれ自由度  $k_1, k_2$  の  $\chi^2$  分布に従う独立な確率変数とする。このとき、確率変数

$$\frac{Y_1/k_1}{Y_2/k_2}$$

は自由度  $k_1, k_2$  の F 分布に従うことを確認しなさい。

## 参考

- 練習用に興味のあるデータを収集しておくが良い
  - データ収集・整理の方法の復習になる
  - 統計量の計算やモデルのあてはめの練習になる
- 参考となるサイト
  - e-Stat (総務省) <https://www.e-stat.go.jp>
  - 気象庁 <http://www.jma.go.jp/jma/index.html>
  - World Bank <https://data.worldbank.org>
  - Yahoo! finance <https://finance.yahoo.com>

## 次回の予定

- 基本的な記述統計量
- “モーメント”に基づく記述統計量
  - 歪度と尖度
  - 相関と共分散
- “順序”に基づく統計量
- “頻度”に基づく統計量