

Rで学ぶ確率統計

第4回 確率分布①

1

Rで学ぶ確率統計

第4回
確率分布①

講師 秋光 淳生



2

第4回 確率分布①

- 1 離散型確率分布
- 2 連続型確率分布
- 3 Rによるシミュレーション

R

3

①-1 確率変数と確率分布(1)

確率変数

確率変数

- 事象に割り当てられた変数
- 起こる事象について確率が分かっている
- 離散型と連続型がある
- 通常大文字で表す
(例) サイコロ $P(X = 1) = \frac{1}{6}$

確率分布の特性値

確率分布

- 確率変数を取る確率を表す関数
- 数式から色々な値を計算できる
- 代表的な分布
- 状況をよく説明できる

特性値

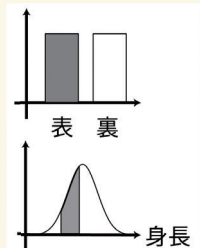
- 期待値
- 分散
- モーメント

4

①-1 確率変数と確率分布(2)

(例1) 離散型
コインを投げる
表が出たら1、裏が出たら0とする
同じ確率で出るとすると
 $P(X = 1) = \frac{1}{2}$

(例2) 連続型
日本人の平均身長を考える
・172.022...cm 精度をとことん高めて
測定することができれば値は連続
・ある範囲にある確率を考える
 $P(a \leq X \leq b) = \int_a^b f(x)dx$



Rで学ぶ確率統計

第4回 確率分布①

5

①-1 確率変数と確率分布(3)

確率分布

離散型の確率変数の場合

- 確率関数

連続型の確率変数の場合

- 確率密度関数
- (累積) 分布関数

- 確率変数が取る全ての値 $\{x_1, x_2, \dots, x_n\}$ に対して、 $P(X = x_i) = p(x_i)$ が決まっている (無限個の場合も)

$$\sum_{i=1}^n p(x_i) = 1$$

- 確率密度関数

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

- 累積分布関数

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du$$

6

①-1 確率変数と確率分布(4)

確率分布の特性値

- 期待値
- 分散
- モーメント

期待値

離散型 $E[X] = \sum_{i=1}^n x_i p(x_i) = \mu$

連続型 $E[X] = \int_{-\infty}^{\infty} x f(x) dx = \mu$

- 分散 (分布のばらつき具合)

離散型 $V[X] = \sum_{i=1}^n (x_i - \mu)^2 p(x_i) = \sigma^2$

$\sigma (> 0)$: 標準偏差

連続型 $V[X] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \sigma^2$

- モーメント

原点のまわりのモーメント

$$\mu_k = E[X^k]$$

平均のまわりのモーメント
(中心モーメント)

$$\mu_k = E[(X - \mu)^k]$$

7

①-2 二点分布 (ベルヌーイ分布 Bernoulli Distribution)

ベルヌーイ分布

- ベルヌーイ試行
2種類のうち、どちらかが一定の確率で起こる
- 確率関数

$$\begin{cases} P(X = 0) = 1 - p \\ P(X = 1) = p \end{cases}$$
- パラメータ (母数): p

- 二項分布 (①-3)

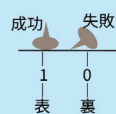
$n = 1$ の場合であるので $B(1, p)$ と書く

- 期待値・分散

$$E[X] = p$$

$$V[X] = p(1 - p)$$

(例) コイン投げ



8

①-3 二項分布 (Binomial Distribution)

二項分布 (Binomial)

- n 回の独立したベルヌーイ試行
 X : 成功回数
 $x: \{0, 1, 2, \dots, n\}$
- 確率関数
$$p(x) = {}_n C_x p^x (1 - p)^{n-x}$$
- パラメータ: n, p

- 確率変数 X が二項分布に従うことを $X \sim B(n, p)$ と書く

- 期待値・分散

$$E[X] = np$$

$$V[X] = np(1 - p)$$

- 二項係数

$${}_n C_x = \frac{n!}{(n-x)! x!}$$

Rで学ぶ確率統計

第4回 確率分布①

9

①-4 多項分布 (Multinomial Distribution)

多項分布

- k 種類の事象 $A_j: j = 1, 2, \dots, k$

- 各 A_j が確率 p_j で起こる

X_j : 各 A_j が起こる回数

$x_j: \{1, 2, \dots\}$

- 確率関数

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

- パラメータ: n, p_j

- $k = 2$ の場合がベルヌイ試行

- $p_1 + p_2 + \dots + p_k = 1$

- 事象 A_j が起こる回数の期待値と分散

$$E[X_j] = np_j$$

$$V[X_j] = np_j(1 - p_j)$$

$$\text{Cov}[X_i, X_j] = -np_i p_j$$

10

①-5 幾何分布 (Geometric Distribution)

幾何分布

- 独立したベルヌイ試行

X : 初めて成功するまでの試行回数

$x: \{1, 2, \dots\}$

- 確率関数

$$p(x) = p(1-p)^{x-1}$$

- パラメータ: p

- 初めて成功するまでの失敗回数

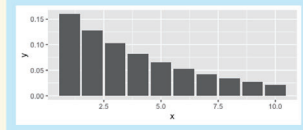
$$p(x_i) = p(1-p)^x$$

とする定義もある (Rはこちら)

期待値・分散

$$E[X] = \frac{1}{p}$$

$$V[X] = \frac{1-p}{p^2}$$



11

①-6 負の二項分布 (Negative Binomial Distribution)

負の二項分布

- ベルヌイ試行で r 回成功するまでの失敗回数

$x: \{0, 1, 2, \dots\}$

$x + r$ 回目に成功

- 確率関数

$$p(x) = {}_{x+r-1}C_{r-1} p^r (1-p)^x$$

- パラメータ: r, p

- 確率変数 X が負の二項分布に従う

- ことを $X \sim NB(r, p)$ と書く

- 期待値・分散

$$E[X] = r \frac{1-p}{p}$$

$$V[X] = r \frac{1-p}{p^2}$$

12

①-7 ポアソン分布 (Poisson Distribution)

ポアソン分布

- 一定時間に発生する回数

- 確率関数

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

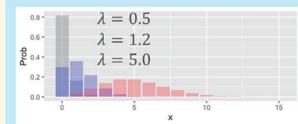
- パラメータ: λ

- 二項分布において $np = \lambda$ とした上で $n \rightarrow \infty$ としたもの

- 期待値・分散

$$E[X] = \lambda$$

$$V[X] = \lambda$$



Rで学ぶ確率統計

第4回 確率分布①

13

②-1 一様分布 (Uniform Distribution)

一様分布

- ある区間 $[a, b]$ の中で同じ確率で起こる

- 確率密度関数

$$f(x) = \frac{1}{b-a}$$

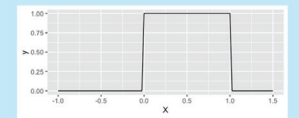
- パラメータ: a, b

- 確率変数 X が一様分布に従うことを $X \sim U(a, b)$ と書く

- 期待値・分散

$$E[X] = \frac{a+b}{2}$$

$$V[X] = \frac{(b-a)^2}{12}$$



14

②-2 指数分布 (Exponential Distribution)

指数分布

- 最初に起こるまでの時間 (ポアソン分布で最初に起こるまでの時間)

- 確率密度関数

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

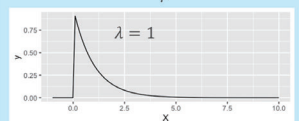
- パラメータ: λ

- 期待値・分散

$$E[X] = \frac{1}{\lambda} (= \beta), \quad V[X] = \frac{1}{\lambda^2}$$

- β を用いて書くと $x \geq 0$ は

$$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$$



15

②-3 ガンマ分布 (Gamma Distribution)

ガンマ分布

- 指数分布で α 回発生するまでの時間

- α が整数のときをアーラン分布

- 確率密度関数

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

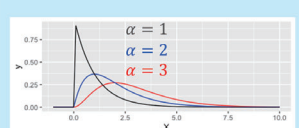
- パラメータ: α, λ

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

- α が整数のとき、 $\Gamma(\alpha+1) = \alpha!$

- 期待値・分散

$$E[X] = \frac{\alpha}{\lambda}, \quad V[X] = \frac{\alpha}{\lambda^2}$$



16

②-4 正規分布 (Normal Distribution)

正規分布

- ガウス分布ともいう

- 確率密度関数

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

- パラメータ: μ, σ

- 二項分布で $n \rightarrow \infty$ とする

- 互いに独立に同一の分布に従う十分大きな数の平均値

- 期待値 μ 分散 σ^2

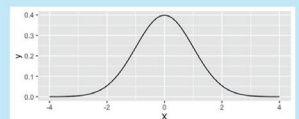
- $\mu = 0, \sigma = 1$ のときを特に

標準正規分布という

$$-\sigma \leq x - \mu \leq \sigma: 68.3\%$$

$$-2\sigma \leq x - \mu \leq 2\sigma: 95.4\%$$

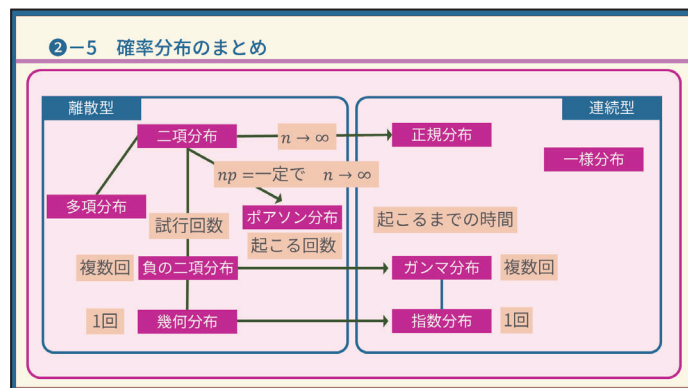
$$-3\sigma \leq x - \mu \leq 3\sigma: 99.7\%$$



Rで学ぶ確率統計

第4回 確率分布①

17



18

③-1 Rにおける分布関数(1)

Rにおける関数

- dXXXX: 確率密度関数 (確率関数)
- pXXXX: 累積分布関数
- qXXXX: 逆分布関数
- 累積分布関数が確率 p になる x
- rXXXX: 乱数
- 発生させる個数
- XXXX: 分布の名前

離散型

- 二項分布: binom
(例) dbinom(x, n, p)
- 多項分布: multinom
(例) dmultinom($c(x1, x2), n, c(p1, p2)$)
- 負の二項分布: nbinom
- 幾何分布: geom
- ポアソン分布: pois

19

③-2 Rにおける分布関数(2)

Rにおける関数

- dXXXX: 確率密度関数 (確率関数)
- pXXXX: 累積分布関数
- qXXXX: 逆分布関数
- 累積分布関数が確率 p になる x
- rXXXX: 乱数
- 発生させる個数
- XXXX: 分布の名前

連続型

- 一様分布: unif
(例) dunif($x, \text{min}, \text{max}$)
- 正規分布: norm
(例) dnorm($x, \text{mean}, \text{sd}$)
- 指数分布: exp
(例) dexp(x, rate)
- ガンマ分布: gamma
(例) dgamma($x, \text{shape}, \text{rate}$)

20

③-3 Rによるシミュレーション

離散型

```
> f1 <- tibble(x=0:10, y=dgeom(seq(0,10,1),0.2))
> ggplot(f1)+geom_bar(aes(x=x,y=y),stat="identity")
```

連続型

```
> ggplot(data=data.frame(x=c(-1,1.5)), aes(x=x)) +
  stat_function(fun=dexp, args=c(rate=1) )

> ggplot( data.frame(x= rnorm(1000) ) +
  geom_histogram(aes(x=x,y=..density..) ) +
  geom_function(fun = dnorm, colour = "red")
```

• seq(a,b,c)
a から b まで c 刻み
(例)
> seq(0,10,3)
[1] 0 3 6 9

• c(-1,1.5) で定義域
• stat_function内
(geom_function)
fun で分布
args でパラメーター
を指定する