# 一般の正規分布・多次元の確率変数

樋口さぶろお https://hig3.net

龍谷大学 先端理工学部 数理・情報科学課程

確率統計 I　L06(2023-05-15 Mon)

最終更新: Time-stamp: "2023-05-15 Mon 12:10 JST hig"

今日の目標

- 一般の正規分布の確率密度関数, グラフを書ける．母期待値，確率，分位数を求められる． 岩薩林 確率・統計 §4.5
- 多次元確率変数の同時分布から周辺分布, 母期待値, 母共分散を求められる． 岩薩林 確率・統計 §3.3

L05-Q1

## Quiz 解答: 標準正規分布の確率

累積分布関数を $F(z)$ とする. `rvz=stats.norm(loc=0,scale=1)` とする.

1. $P(Z \le 1.23) = \int_{-\infty}^{1.23} f(z;0,1^2)\,\mathrm{d}z = F(1.23) - F(-\infty) = F(1.23) - 0 =$ `rvz.cdf(1.23)` $= 0.5708$.
2. $P(-0.56 < Z \le +1.23) = \int_{-0.56}^{1.23} f(z;0,1^2)\,\mathrm{d}z = F(1.23) - F(-0.56) =$ rvz.cdf(1.23) − rvz.cdf(-0.56) $= 0.6030$.
3. $d = F^{-1}(1 - 0.025) =$ `rvz.ppf(0.975)` $= 1.960$.

L05-Q2

## Quiz 解答: 正規分布の確率

1. $\mathrm{E}[X] = 4$.
2. $\mathrm{V}[X] = 3^2$.

3. $x = 4$ を真ん中に幅 3 くらいの正規分布の確率密度関数のグラフ．標準正規分布を，横に 3 倍，縦に $1/3$ 倍，右に 4 平行移動．

# ここまで来たよ

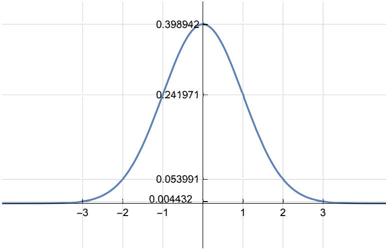
5 確率変数の標準化・標準正規分布・一般の正規分布

6 一般の正規分布・多次元の確率変数
- 一般の正規分布
- 2 次元の確率分布
- 母共分散
- 多次元の連続型確率分布

# 一般の正規分布 $\mathrm{N}(b, a^2)$

$Z \sim \mathrm{N}(0, 1^2)$ に対して，$X = aZ + b$ を考える．

$$
\begin{aligned}
\mathrm{E}[X^0] &= 1, \\
\mu = \mathrm{E}[X] &= \mathrm{E}[aZ + b] = b, \\
\sigma^2 = \mathrm{V}[X] &= \mathrm{V}[aZ + b] = a^2,
\end{aligned}
$$




$$
f_{\mathrm{Z}}(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} \rightsquigarrow f_{\mathrm{X}}(x) = \frac{1}{\sqrt{2\pi a^2}} \mathrm{e}^{-\frac{(x-b)^2}{2a^2}}
$$

## 定義 (一般の正規分布 $\mathrm{N}(\mu, \sigma^2)$)

次の確率密度関数を持つ確率変数 $X$ を，母平均値 $\mu$, 母分散 $\sigma^2$ の正規分布 $\mathrm{N}(\mu, \sigma^2)$ (normal distribution) にしたがうという．

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

# L06-Q1

## Quiz(正規分布の確率)

連続型確率変数 $X$ は, 確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi \cdot 3^2}} \mathrm{e}^{-\frac{(x-4)^2}{2\cdot 3^2}}$$

にしたがう.

1. $\mathrm{E}[X]$ を求めよう.
2. $\mathrm{V}[X]$ を求めよう.
3. $f(x)$ のグラフを, 標準正規分布の確率密度関数と重ねて描こう.

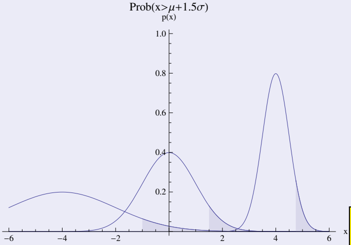
Python でも描いてみよう.

# 正規分布 $\mathrm{N}(\mu, \sigma^2)$ の確率 I

## 命題 (標準化前後の確率は同じ)

一般に, 確率は標準化 $Z = \frac{X-\mu}{\sigma}$ の前後で変わらない.

$$P(c < X \leq d) = P(\tfrac{c-\mu}{\sigma} < \tfrac{X-\mu}{\sigma} \leq \tfrac{d-\mu}{\sigma}) = P(\tfrac{c-\mu}{\sigma} < Z \leq \tfrac{d-\mu}{\sigma})$$




斜線部の面積はどれも同じ

# 標準化しても確率が同じことの別説明 (置換積分)

$X \sim \mathrm{N}(\mu, \sigma^2)$ のとき, 積分で $z = \frac{x-\mu}{\sigma}$, $\mathrm{d}z = \frac{1}{\sigma}\mathrm{d}x$ とすると,

$$
\begin{aligned}
P(c < X \leq d) &= \int_c^d \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x \\
&= \int_{\frac{c-\mu}{\sigma}}^{\frac{d-\mu}{\sigma}} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{z^2}{2}}\,\sigma \mathrm{d}z \\
&= P\left(\frac{c-\mu}{\sigma} < Z \leq \frac{d-\mu}{\sigma}\right)
\end{aligned}
$$

# L06-Q2

## Quiz(正規分布の確率)

確率変数 $X \sim \mathrm{N}(3, 2^2)$, $Z \sim \mathrm{N}(0, 1^2)$ とする.

1. $X, Z$ の確率密度関数の式を書こう. グラフを重ねて描こう.
2. 母期待値 $\mathrm{E}[X^2]$ を求めよう.
3. 確率 $P(X > 5) = P(c < Z \le d)$ となるように $c$,$d$ を定めよう. また, 確率を Python や表を使って小数で求めよう.
4. 確率 $P(+1 < X \le 7) = P(c \le Z \le d)$ となるように $c$,$d$ を定めよう. また, 確率を Python や表を使って小数で求めよう.
5. 確率 $P(+3 < X \le 9) = P(c \le Z \le d)$ となるように $c$,$d$ を定めよう. また, 確率を Python や表を使って小数で求めよう.
6. 確率 $P(X \le d) = 3/4$ となる $d$ を, 確率 $P(Z \le d') = 3/4$ となるような $d'$ で表そう. また, Python や表を使って小数で求めよう.

岩薩林 確率・統計 §4.5 例題 4.10,4.11, 問題 10, 第 4 章練習問題 5

# ここまで来たよ

# 2 つの離散型確率変数の同時確率分布 高校 数学 B 岩薩林 確率・統計 p.56

例 6 枚のカードから無作為に 1 枚引く. ♡7　♡8　♡9　◇8　♠9　♣9

2 つの離散型確率変数の同時分布

$X=$ 数, $Y=0$(赤札), $1$(黒札) とすると $(x,y)$ を得る確率は 2 変数の確率関数で書ける. 同時分布, 結合分布, joint distribution という.

$$p(x,y)=\begin{cases}\frac{1}{3} & ((x,y)=(8,0)) \\ \frac{1}{6} & ((x,y)=(9,0)) \\ \frac{1}{3} & ((x,y)=(9,1)) \\ \frac{1}{6} & ((x,y)=(7,0)) \\ 0 & (\text{他})\end{cases}$$

表で書いた方がまし. ここでは,「他」は省略.

岩薩林 確率・統計 では, $p_{xy}$ とも.

$P(X=x,Y=y)$ とも. $P$(条件) 型の表記.

| $y\backslash x$ | 7 | 8 | 9 | 計 |
|---|---|---|---|---|
| 0 | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{6}$ | |
| 1 | 0 | 0 | $\frac{1}{3}$ | |
| 計 | | | | |

## 定義 (同時分布の母期待値 岩薩林 確率・統計 (3.16)p.60)

$$\mathrm{E}[g(X,Y)] \overset{\text{定義}}{=} \sum_{x=-\infty}^{+\infty} \sum_{y=-\infty}^{+\infty} g(x,y) \cdot p(x,y)$$

## 定義 (同時分布の確率)

条件 $a < X \leq b$ かつ $c < Y \leq d$ で定まる確率は,

$$P(a < X \leq b \text{ かつ } c < Y \leq d) = \sum_{a<x\leq b} \sum_{c<y\leq d} p(x,y) \tag{P2}$$

## L06-Q3

### Quiz(多変数の確率変数の期待値)

2 変数 $X, Y$ の離散型確率分布を考える．同時分布 $p(x, y)$ が下の表で与えられる．

| $y \backslash x$ | 1 | 2 | 3 |
|---|---|---|---|
| 0 | $0$ | $\frac{2}{12}$ | $\frac{1}{12}$ |
| 2 | $\frac{4}{12}$ | $0$ | $\frac{5}{12}$ |

1. 母期待値 $\mathrm{E}[2X^2 + e^Y]$ を求めよう．
2. 確率 $P(XY \geq 2)$ を求めよう．
3. 周辺分布 $p_{\mathrm{X}}(x)$, $p_{\mathrm{Y}}(y)$ を求めよう．

## 命題 (2 次元の確率分布の母期待値の性質高校 数学 B 岩薩林 確率・統計 (3.17)p.61)

$$\mathrm{E}[g_1(X,Y)+g_2(X,Y)] = \mathrm{E}[g_1(X,Y)]+\mathrm{E}[g_2(X,Y)] \quad \text{(E5)}$$

$$\text{特に}\mathrm{E}[X+Y] = \mathrm{E}[X]+\mathrm{E}[Y] \quad \text{(E6,(3.17))}$$

証明

$$\begin{aligned}\mathrm{E}[g_1(X,Y)+g_2(X,Y)] &= \sum_x\sum_y (g_1(x,y)+g_2(x,y))p(x,y)\\ &= \mathrm{E}[g_1(X,Y)]+\mathrm{E}[g_2(X,Y)].\end{aligned}$$

# 周辺分布

## 定義 (確率変数の周辺分布)

同時分布 $p(x,y)$ に対して,
$X$ の周辺分布 $p_{\mathrm{X}}(x)$, $Y$ の周辺分布 $p_{\mathrm{Y}}(y)$ は,

$$p_{\mathrm{X}}(x) = p(x,\bullet) = \sum_y p(x,y), \quad p_{\mathrm{Y}}(y) = p(\bullet,y) = \sum_x p(x,y)$$

要するに**一方を無視した分布. 小計.**

岩薩林 確率・統計 例題 3.4(p.57)

| $y \backslash x$ | 7 | 8 | 9 | 計 |
|---|---|---|---|---|
| 0 | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{6}$ | |
| 1 | 0 | 0 | $\frac{1}{3}$ | |
| 計 | | | | |

### 命題 ($X$ だけ, $Y$ だけの関数の母期待値)

$x$ だけ, $y$ だけの関数の母期待値は,
下の左辺= $\boxed{\text{同時分布}}$ で計算しても
下の右辺= $\boxed{\text{周辺分布}}$ で計算しても
同じ結果.

$$\mathrm{E}[g(X)] = \sum_x \sum_y g(x) \cdot p(x,y) = \sum_x g(x) \sum_y p(x,y) = \sum_x g(x) \cdot p_{\mathrm{X}}(x)$$

$$\mathrm{E}[g(Y)] = \sum_y \sum_x g(y) \cdot p(x,y) = \sum_y g(y) \sum_x p(x,y) = \sum_y g(y) \cdot p_{\mathrm{Y}}(y)$$

# L06-Q4

## Quiz

さっきの問で

1. 周辺分布を求めよう.
2. 周辺分布から $\mathrm{E}[2X^2 + \mathrm{e}^Y]$ を求めよう.
3. 母分散 $\mathrm{V}[X]$ を求めよう.

# ここまで来たよ

5. 確率変数の標準化・標準正規分布・一般の正規分布

6. 一般の正規分布・多次元の確率変数
   - 一般の正規分布
   - 2 次元の確率分布
   - 母共分散
   - 多次元の連続型確率分布

# 母共分散 高校 数学 B

**定義 (母共分散 covariance 岩薩林 確率・統計 (3.19)p.61,(3.20)p.62)**

$X, Y$ が確率変数で，$\mu_X = \mathrm{E}[X], \mu_Y = \mathrm{E}[Y]$ とおいたとき，

$$\text{母共分散}\mathrm{Cov}[X,Y] \overset{\text{定義}}{=} \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] \quad (3.19)$$

$$= \boxed{\text{岩薩林 確率・統計 (3.20)}} \cdots = \mathrm{E}[XY] - \mathrm{E}[X]\times\mathrm{E}[Y]. \quad (\mathrm{C1},(3.20))$$

L06-Q5

**母共分散**

さっきの問で

1. 母期待値 $\mathrm{E}[XY]$ を求めよう．
2. 母共分散 $\mathrm{Cov}[X,Y]$ を求めよう．

## 2 次元の確率分布の母期待値のよくある間違い

$$\mathrm{E}[g(X,Y)] \neq g(\mathrm{E}[X],\mathrm{E}[Y])$$

だって, $\sin(\log_x y) \neq \log_{\sin x}(\sin y)$ じゃん.
特に $\mathrm{E}[X \times Y] \neq \mathrm{E}[X] \times \mathrm{E}[Y]$. $\times$ を, $+$ と区別しよう.

## 2 次元の確率分布の母分散のよくある間違い

$$\mathrm{V}[X+Y] \neq \mathrm{V}[X] + \mathrm{V}[Y]$$

母分散を, 母期待値と区別しよう.

# L06-Q6

## 母分散

さっきの問で母分散 $\mathrm{V}[X+Y]$ を求めよう.

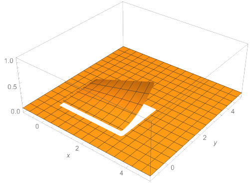
# ここまで来たよ

5. 確率変数の標準化・標準正規分布・一般の正規分布

6. 一般の正規分布・多次元の確率変数
   - 一般の正規分布
   - 2 次元の確率分布
   - 母共分散
   - 多次元の連続型確率分布

# 多次元の連続型確率分布 岩薩林 確率・統計 §4.6

2 次元の同時確率密度関数 $f(x,y)$.
$f(x,y) \geq 0$, 大きいほど，その $(x,y)$ が「出やすい」




## 定義 (連続型次元確率変数の母期待値 岩薩林 確率・統計 SS3.3,4.6)

離散型 $$\mathrm{E}[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot p(x,y) \tag{3.16}$$

連続型 $$\mathrm{E}[g(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) \cdot f(x,y)\,\mathrm{d}x\mathrm{d}y \tag{4.32}$$

確率 $P(a < X \leq b \text{かつ} c < Y \leq d) = \int_a^b (\int_c^d f(x,y)\mathrm{d}y)\mathrm{d}x$. 平面の領域上にある部分の体積.