

連続型確率変数

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 L02(2024-04-22 Mon)

最終更新: Time-stamp: "2024-04-22 Mon 13:14 JST hig"

今日の目標

- (Colab) 離散型確率変数の標本抽出ができる
- 岩薩林 確率・統計 §3.3 連続型確率変数とは何か説明できる
- 岩薩林 確率・統計 §3.3 連続型確率変数の母期待値母平均値母分散が計算できる



L01-Q1

Quiz 解答: 離散的な確率変数の母平均・母分散・母標準偏差

$$\textcircled{1} \text{ 母期待値 } E[e^X] = \frac{4}{12} \cdot e^{-1} + \frac{5}{12} \cdot e^0 + \frac{3}{12} \cdot e^2.$$

$$\textcircled{2} \text{ 母平均値 } E[X] = \frac{4}{12} \cdot (-1) + \frac{5}{12} \cdot 0 + \frac{3}{12} \cdot 2 = \frac{1}{6} (= \mu).$$

$$\textcircled{3} \text{ 母分散}$$

$$V[X] = E[(X - \mu)^2] = \frac{4}{12} \cdot (-1 - \frac{1}{6})^2 + \frac{5}{12} \cdot (0 - \frac{1}{6})^2 + \frac{3}{12} (2 - \frac{1}{6})^2 = \frac{47}{36}.$$

$$\text{別解 } E[X^2] = (-1)^2 \cdot \frac{4}{12} + 0^2 \cdot \frac{5}{12} + 2^2 \cdot \frac{3}{12} = \frac{4}{3}. \quad V[X] \stackrel{V1}{=} E[X^2] - \mu^2 = \frac{47}{36}.$$

計算を楽にする母期待値の性質

命題 (母期待値の性質 高校 数学 B)

X : 確率変数, $a, b \in \mathbb{R}$: 定数 のとき,

$$E[1] = 1 \quad (\text{E1})$$

$$E[g_1(X) + g_2(X)] = \sum_x (g_1(x) + g_2(x)) \times p(x) = E[g_1(X)] + E[g_2(X)] \quad (\text{E2, } \text{岩薩林 確率・統計 (3.5)})$$

$$E[a \cdot g(X)] = \sum_x (a \cdot g(x)) \times p(x) = aE[g(X)] \quad (\text{E3})$$

$$E[aX + b] \stackrel{\text{E2}}{=} E[aX] + E[b] \stackrel{\text{E3}}{=} aE[X] + bE[1] \stackrel{\text{E1}}{=} aE[X] + b \quad (\text{E4, } \text{岩薩林 確率・統計 (3.6)})$$

定義 (k 次のモーメント 岩薩林 確率・統計 p.53)

$E[X^k]$ を X の k 次のモーメントという ($k = 0, 1, 2, \dots$).

k 次のモーメントを部品として計算しておき, 他の母期待値は性質から導くのが楽.

母分散の性質 岩薩林 確率・統計 (3.8)(3.9)p.55母分散の性質 高校 数学 B 岩薩林 確率・統計 (3.9)p.55

X : 確率変数, $a, b \in \mathbb{R}$: 定数 のとき,

$$V[X] = E[X^2] - (E[X])^2 \quad (\text{V1, 分散公式 (3.8)})$$

$$V[aX + b] = a^2 V[X]. \quad (\text{V2, (3.9)})$$

証明 V1 $\mu = E[X]$ とする.

$$V[X] = E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2] \stackrel{E2, E3}{=} E[X^2] - 2\mu E[X] + \mu^2 E[1]$$

$$\stackrel{E1}{=} E[X^2] - 2\mu^2 + \mu^2.$$

証明 V2 $\mu = E[X]$ と, $E[Y] = \mu_Y = a\mu + b$ とする.

$$\begin{aligned} V[Y] &= E[(Y - \mu_Y)^2] \\ &= E[(aX + b) - (a\mu + b)]^2 \\ &= a^2 E[(X - \mu)^2] = a^2 V[X]. \end{aligned}$$

確率変数にまつわる確率

定義 (確率関数 岩薩林 確率・統計 (3.1))

x の関数 $f(x) = p(x) = P(X = x)$ を離散型確率変数 X の確率関数という。

定義 (累積分布関数 岩薩林 確率・統計 (3.2))

x の関数 $F(x) = P(X \leq x)$ を確率変数 X の累積分布関数という。

$x \in \mathbb{Z}$ に限ると, 数列 $F_x = F(x)$ の階差数列が $f_x = f(x)$.

命題 (累積分布関数の性質 岩薩林 確率・統計 pp.50,51)

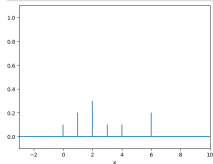
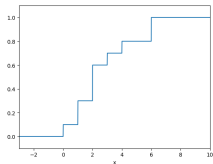
- (C1) $F(x)$ は広義単調増加関数
- (C2) $F(-\infty) = 0, F(+\infty) = 1$.
- (C3) $P(c < X \leq d) = F(d) - F(c)$.
- (C4) $F(x) = \sum_{x' \leq x} p(x')$.

累積分布関数

確率変数 X の確率関数を次とする。

$$p(x) = \begin{cases} 0.1 & (x=0) \\ 0.2 & (x=1) \\ 0.3 & (x=2) \\ 0.1 & (x=3) \\ 0.1 & (x=4) \\ 0 & (x=5) \\ 0.2 & (x=6) \\ 0 & (\text{他}) \end{cases} =$$

p_k	x_k
0.1	($x=0$)
0.2	($x=1$)
0.3	($x=2$)
0.1	($x=3$)
0.1	($x=4$)
0	($x=5$)
0.2	($x=6$)

 $p(x)$  $F(x)$

L01-Q2

Quiz(離散的な確率変数の母平均・母分散・母標準偏差)

整数に値をとる離散型確率変数 X は次の確率分布に従う.

$$f(x) = \begin{cases} \frac{4}{12} & (x = -1) \\ \frac{5}{12} & (x = 0) \\ \frac{3}{12} & (x = 2) \\ 0 & (\text{他}) \end{cases}$$

- ① 母期待値 $E[e^X]$ を求めよう.
- ② X の母平均値を求めよう.
- ③ X の母分散を求めよう.
- ④ X の母標準偏差を求めよう.
- ⑤ 事象 $X \leq 1$ の確率を求めよう.

ここまで来たよ

- 1 離散型確率変数
 - 標本抽出と Python の scipy による確率変数の扱い
- 2 連続型確率変数
 - 確率密度関数

確率変数の標本抽出

岩薩林 確率・統計 §5 を先取り

定義 (標本, 標本抽出)

確率変数 X の値を得る試行を n 回行って得られる x の長さ n の列を,
サイズ (size) n の標本 (sample) という.
標本を得る操作を標本抽出 (sampling) という.

標本は '毎回違った結果' になる.

データ分析 では標本を扱っていた.

scipy による確率変数の標本抽出

<https://colab.research.google.com>

```
1 import numpy as np # ライブラリの読込
2 from scipy import stats
3
4 xk = (0, 1, 2, 3, 4, 5, 6) #  $x_1, x_2, \dots, x_k$ 
5 pk = (0.1, 0.2, 0.3, 0.1, 0.1, 0.0, 0.2) #  $p_1, p_2, \dots, p_k$ 
6 rvx = stats.rv_discrete(values=(xk, pk)) # 離散型確率変数  $X$  を定義
7 sample=rvx.rvs(size=30) # サイズ30の標本抽出(毎回違った結果になる)
8
9 # 母ノントカ. 答はいつも同じ
10 rvx.mean() # 母平均値
11 rvx.var() # 母分散
12 rvx.std() # 母標準偏差
13
14 rvx.cdf(x) # 累積分布関数
15
16 # データ分析 でやってた標本平均値
17 ## 標本抽出のたびに結果が変わる
18 sample.mean() # 標本平均値
19 sample.var() # 不偏標本平均値
20 sample.std() # 不偏標本標準偏差
```

ここまで来たよ

- 1 離散型確率変数
 - 標本抽出と Python の `scipy` による確率変数の扱い
- 2 連続型確率変数
 - 確率密度関数

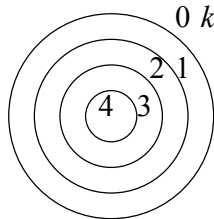
離散型確率変数の例

ジャンボ宝くじの賞金 X 円は離散型確率変数.
確率関数

$$p(x) = \begin{cases} \frac{889}{1000} & (x = 0) \\ \frac{1}{10} & (x = 300) \\ \frac{1}{100} & (x = 3000) \\ \frac{1}{1000} & (x = 10000) \\ 0 & (\text{他}) \end{cases}$$

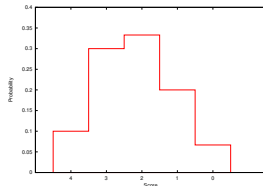
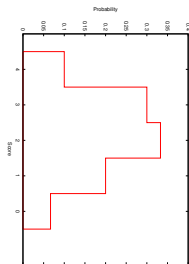
あるプレイヤーのダーツの得点確率

得点: 的の真ん中から順に 4, 3, 2, 1, 0 点



離散型確率分布

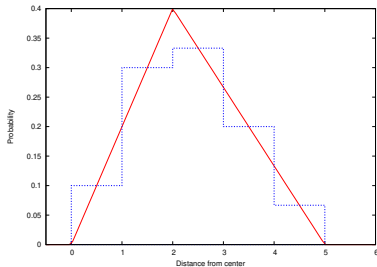
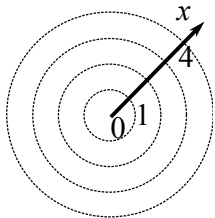
得点 s	確率関数 $f(s)$
4	0.1
3	0.3
2	0.3333
1	0.2
0	0.0667



中心から x cm にあてる確率

岩薩林 確率・統計 §4.1

的の真ん中からの距離 x cm, 得点 $s = 4 - x$ 点 (実数).



$x = 0.5$ cm と 0.9 cm への当たりやすさは違う. $x = 1.0$ cm を境に急に変わるわけじゃない. これを表現したい.

↪ 点数の出やすさは x のある関数 $p(x)$ で表される!

連続型確率変数 連続型確率分布

連続型 確率密度関数 $f(x)$ (x は実数)

離散型 確率関数 $p(x)$ (x は整数またはとびとびの値)

連続型確率変数

定義 (連続型確率変数)

連続型確率変数 X とは, 実数値をとり, 確率が確率密度関数 $f(x)$ で指定されるもの.

離散的

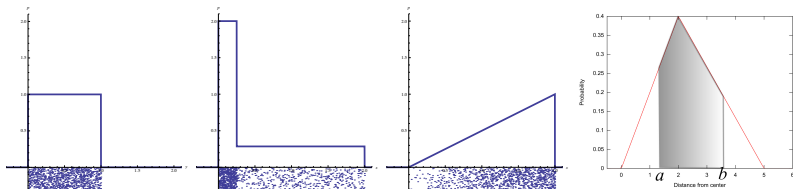
得点 x	確率 $p(x)$
0	0.1
1	0.3
\vdots	
x	$p(x)$

連続的

- $f(x)$ が大きいほど, その値 x がやすい
- $0 \leq f(x)$ である. $f(x) \leq 1$ とは限らない.

物理・工学系では $p(x)$ と書いたら確率密度関数 $f(x)$ を意味することも

確率密度関数の例



横軸下の細かい点が、標本 (縦方向の位置はランダムで意味なし)

定義 (累積分布関数 岩籙林 確率・統計 (4.4))

$$F(x) = \int_{-\infty}^x f(x') dx' \quad (\text{C4 相当})$$

$F(x)$ が微分可能なら $f(x)$ は $F(x)$ の導関数

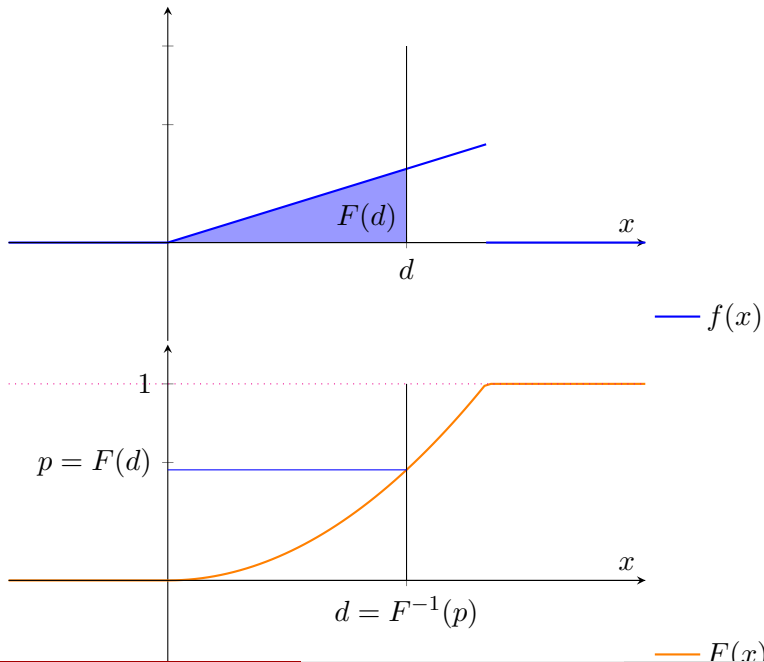
命題 (確率密度関数と確率 岩籙林 確率・統計 (4.1))

$$P(c < X \leq d) = F(d) - F(c) = \int_c^d f(x) dx \quad (\text{下側面積}) \quad (\text{C3})$$

連続型確率変数で, $P(x = a) = ?$

ちょうど距離 $x = c$ cm となる確率は? \rightsquigarrow **0**.

$$1 = P(\text{全事象}) = P(-\infty < X < +\infty) = \int_{-\infty}^{+\infty} 1 \cdot f(x) dx = E[1].$$



連続型確率変数の母期待値

岩薩林 確率・統計 §4.2

定義 (母期待値)

岩薩林 確率・統計 (4.8)

$$\text{離散型確率変数} \quad E[g(X)] = \sum_x g(x) \cdot p(x)$$

$$\text{連続型確率変数} \quad E[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx$$

- 離散型と同じ定義: 母平均値 $\mu = E[X]$, 母分散 $V[X] = E[(X - \mu)^2]$
- 離散型と同じ公式 E_n, V_n が成立

定義 (k 次のモーメント ($k = 0, 1, 2, 3, \dots$))

岩薩林 確率・統計 (4.9)

$$E[X^k] = \int_{-\infty}^{+\infty} x^k \cdot f(x) dx$$

L02-Q1

Quiz(連続型確率変数)

次の確率密度関数 $f(x)$ を持つ連続型確率変数 X を考える. 次を求めよう.

$$f(x) = \begin{cases} 8x & (0 \leq x < \frac{1}{2}) \\ 0 & (\text{それ以外}) \end{cases}$$

- ① 母期待値 $E[X^k]$ ($k = 0, 1, 2, \dots$)
- ② 母平均値 $E[X]$
- ③ 母分散 $V[X]$
- ④ 母期待値 $E[(2X + 3)^2]$
- ⑤ 母分散 $V[2X + 3]$

岩薩林 確率・統計 例題 4.2,4.3