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(2023-04-17 Mon)

最終更新: Time-stamp: "2023-04-19 Wed 11:40 JST hig"

今日の目標

- 離散型確率変数の確率が計算できる 岩薩林 確率・統計 §3.2
- 離散型確率変数の標本抽出ができる
- 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

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

1. 母期待値 $\mathrm{E}[\mathrm{e}^X] = \frac{4}{12} \cdot \mathrm{e}^{-1} + \frac{5}{12} \cdot \mathrm{e}^{0} + \frac{3}{12} \cdot \mathrm{e}^{2}$.
2. 母平均値 $\mathrm{E}[X] = \frac{4}{12} \cdot (-1) + \frac{5}{12} \cdot 0 + \frac{3}{12} \cdot 2 = \frac{1}{6} (= \mu)$.
3. 母分散
   $\mathrm{V}[X] = \mathrm{E}[(X-\mu)^2] = \frac{4}{12} \cdot (-1 - \frac{1}{6})^2 + \frac{5}{12} \cdot (0 - \frac{1}{6})^2 + \frac{3}{12}(2 - \frac{1}{6})^2 = \frac{47}{36}$.

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

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,

$$\mathrm{E}[1] = 1 \tag{E1}$$

$$\mathrm{E}[g_1(X) + g_2(X)] = \sum_x (g_1(x) + g_2(x)) \times p(x) = \mathrm{E}[g_1(X)] + \mathrm{E}[g_2(X)]$$

(E2, 岩薩林 確率・統計 (3.5))

$$\mathrm{E}[a \cdot g(X)] = \sum_x (a \cdot g(x)) \times p(x) = a\mathrm{E}[g(X)] \tag{E3}$$

$$\mathrm{E}[aX + b] \overset{\mathrm{E2}}{=} \mathrm{E}[aX] + \mathrm{E}[b] \overset{\mathrm{E3}}{=} a\mathrm{E}[X] + b\mathrm{E}[1] \overset{\mathrm{E1}}{=} a\mathrm{E}[X] + b$$

(E4, 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)

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

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,

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

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

証明 V1 $\mu = \mathrm{E}[X]$ とする.

$\mathrm{V}[X] = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2 - 2\mu X + \mu^2] \overset{\mathrm{E2,E3}}{=}$
$\mathrm{E}[X^2] - 2\mu\mathrm{E}[X] + \mu^2\mathrm{E}[1] \overset{\mathrm{E1}}{=} \mathrm{E}[X^2] - 2\mu^2 + \mu^2.$

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

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

ここまで来たよ

1. 離散型確率変数
   - 事象と確率
   - 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 $p(x) = P(X = x)$ 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.

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) $P(a < X \leq b) = F(b) - F(a).$

(C4) $F(x) = \sum_{x' \leq x} p(x').$

小文字 $x \in \mathbb{R}$ はふつうの変数 (実数), 大文字 X は確率変数.

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)

ここまで来たよ

1. 離散型確率変数
   - 事象と確率
   - 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

```
import numpy as np # ライブラリの読込
from scipy import stats

xk = (0,   1 ,  2  , 3  ,4    , 5  ,6)    # x_1, x_2,..,x_k
pk = (0.1, 0.2, 0.3, 0.1, 0.1, 0.0, 0.2) # p_1, p_2,..,p_k
rvx = stats.rv_discrete(values=(xk, pk)) # 離散型確率変数 X を定義
sample=rvx.rvs(size=30) # サイズ30の標本抽出(毎回違った結果になる)

# 母ナントカ. 答はいつも同じ
rvx.mean()# 母平均値
rvx.var() # 母分散
rvx.std() # 母標準偏差

rvx.cdf(x)# 累積分布関数

# データ分析 でやってた標本平均値
## 標本抽出のたびに結果が変わる
sample.mean() # 標本平均値
sample.var()  # 不偏標本平均値
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
   - 事象と確率
   - 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

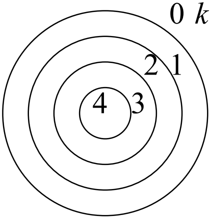
離散型確率変数の例

ジャンボ宝くじの賞金 X 円は離散型確率変数.
確率関数

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

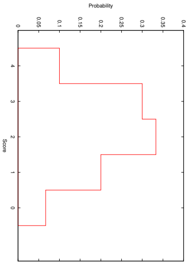
あるプレイヤーのダーツの得点確率

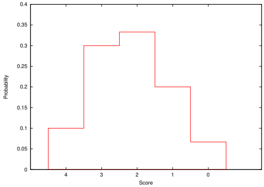
得点: 的の真ん中から順に $4, 3, 2, 1, 0$ 点



離散型確率分布

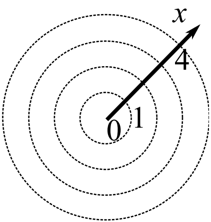
得点 s	確率関数 $f(s)$
4	0.1
3	0.3
2	0.3333
1	0.2
0	0.0667

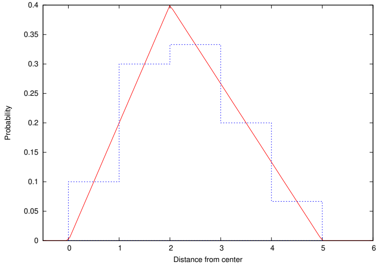




中心から x cm にあてる確率 岩薩林 確率・統計 §4.1

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.

$\leadsto$ 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	$p(x)$

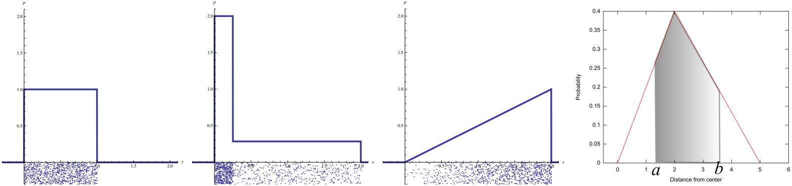
連続的

- $f(x)$ が大きいほど, その値 x がでやすい
- $0 \leq f(x)$ 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

計算科学☆演習 II

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 岩薩林 確率・統計 (4.4))

$$F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x' \qquad \text{(C4 相当)}$$

(C1),(C2) を満たす.

命題 (確率密度関数と確率 岩薩林 確率・統計 (4.1))

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)\,\mathrm{d}x \quad \text{(下側面積)} \qquad \text{(C3)}$$

連続型確率変数で, $P(x = a) =?$
ちょうど距離 $x = a$ cm となる確率は? $\rightsquigarrow \boxed{0}$.

$$1 = P(\text{全事象}) = P(-\infty < X < +\infty) = \int_{-\infty}^{+\infty} 1 \, \mathrm{d}x.$$

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))

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

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

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

定義 (k 次のモーメント ($k = 0, 1, 2, 3, \cdots$) 岩薩林 確率・統計 (4.9))

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

L02-Q1

Quiz(連続型確率変数)

次の確率密度関数 $f(x)$ を持つ確率変数 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 $\mathrm{E}[X^k]$ $(k = 0, 1, 2, \ldots)$
2. 母平均値 $\mathrm{E}[X]$
3. 母分散 $\mathrm{V}[X]$
4. 母期待値 $\mathrm{E}[(2X + 3)^2]$
5. 母分散 $\mathrm{V}[2X + 3]$
6. 確率 $P(-\frac{1}{4} < X \leq \frac{1}{4})$.

岩薩林 確率・統計 例題 4.2,4.3