指数分布・一様分布と確率変数の標準化

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 L04(2023-05-01 Mon)

最終更新: Time-stamp: "2023-05-01 Mon 12:00 JST hig"

今日の目標

- 指数分布 岩薩林 確率・統計 p.80, 一様分布 岩薩林 確率・統計 p.78 を説明できる
- scipy.stat の location と scale を説明できる
- 確率変数の標準化を説明できる 岩薩林 確率・統計 p.80

L03-Q1
Quiz 解答: 連続型確率変数

1. k 次のモーメントは,

$$\mathrm{E}[X^k] = \int_{-\infty}^{+\infty} x^k \cdot f(x)\,\mathrm{d}x = \int_0^{1/2} x^k 8x\,\mathrm{d}x = \frac{2^3 \cdot 2^{-k-2}}{k+2}.$$

   $\mathrm{E}[X^0] = 1$ が確認できる.
2. $\mathrm{E}[X^1] = \frac{1}{3}$.
3. $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X^1]^2 = \frac{1}{8} - (\frac{1}{3})^2 = \frac{1}{72}$.
4. $\mathrm{E}[(2X+3)^2] = 4\mathrm{E}[X^2] + 12\mathrm{E}[X^1] + 9\mathrm{E}[X^0] = 4 \times \frac{1}{8} + 12 \times \frac{1}{3} + 9 = \frac{27}{2}$.
5. $\mathrm{V}[2X+3] = 2^2\mathrm{V}[X] = \frac{1}{18}$.
6. $P(-\frac{1}{4} < X \leq \frac{1}{4}) = \int_{-\frac{1}{4}}^{+\frac{1}{4}} f(x)\,\mathrm{d}x = \int_{-\frac{1}{4}}^{0} 0\,\mathrm{d}x + \int_0^{+\frac{1}{4}} 8x\,\mathrm{d}x = \frac{1}{4}$.

7. 累積分布関数は,

$$F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x' = \begin{cases} 0 & (x \leq 0) \\ 4x^2 & (0 < x \leq \frac{1}{2}) \\ 1 & (\frac{1}{2} < x) \end{cases}.$$

区間 $[0,1]$ で逆関数を考えると，分位点関数は

$$F^{-1}(y) = \tfrac{1}{2}\sqrt{y}.$$

中央値すなわち $\frac{1}{2}$-分位数は $F^{-1}(\frac{1}{2}) = \frac{1}{2\sqrt{2}}$, $\frac{1}{4}$-分位数は $F^{-1}(\frac{1}{4}) = \frac{1}{4}$.

別解 $\int_{-\infty}^{d} f(x)\mathrm{d}x = \frac{1}{2}$ を d について解く.

ここまで来たよ

3 連続型確率変数の母期待値と分位数

4 指数分布・一様分布と確率変数の標準化
- 指数分布
- 連続型一様分布
- 確率変数の標準化

指数分布 岩薩林 確率・統計 p.80

連続型確率変数の分布の, 名前がつくくらい有名な例

定義 (指数分布)

連続型確率変数 X が 確率密度関数

$$f(x;b,a)=\begin{cases}\frac{1}{a}\mathrm{e}^{-\frac{x-b}{a}} & (x>b)\\ 0 & (\text{他})\end{cases}$$

を持つとき, X はパラメタ $a(>0), b$ の指数分布に従うといい, 記号 $\sim$(従う) で $X\sim\mathrm{Exp}(b,a)$ と書く (Exp は科目ローカル記号) (パラメタは確率変数ではない. $8x^1$ の 8 や 1 のこと.)
世の中 岩薩林 確率・統計 p.80 では, $b=0$ の場合だけを指数分布とよぶ.

性質 $\mathrm{E}[X]=b+a$, $\mathrm{V}[X]=a^2$.
意味時間的にランダムに起きる事象，例えば,「放射性元素が崩壊する」「宝くじが当たる」のような事象の，時間間隔 x のしたがう分布.

L04-Q1

Quiz(指数分布の母平均値・母分散・確率)

パラメタ $b = 0, a = 3$ の指数分布にしたがう確率変数 $X \sim \mathrm{Exp}(0, 3)$ を考える．すなわち，X の確率密度関数は次で与えられる．

$$f(x) = \begin{cases} \frac{1}{3}\mathrm{e}^{-x/3} & (x \geq 0) \\ 0 & (\text{他}) \end{cases}$$

1. $\mathrm{E}[X]$ を求めよう．
2. $\mathrm{V}[X]$ を求めよう．
3. 累積分布関数を求めよう．確率 $P(1 < X \leq 3)$ を小数点以下第 3 位まで求めよう．
4. 分位数関数を求めよう．確率 $P(X \leq d) = \frac{1}{3}$ となるような d を小数点以下第 3 位まで求めよう．

公式: $\int x^0 \mathrm{e}^{-x} \mathrm{d}x = \mathrm{e}^{x}$, $\int x^1 \mathrm{e}^{-x} \mathrm{d}x = -(x+1)\mathrm{e}^{-x}$, $\int x^2 \mathrm{e}^{-x} \mathrm{d}x = -(x^2 + 2x + 2)\mathrm{e}^{-x}$.

指数関数, 対数関数の値の求め方: Windows の電卓．iPhone を横向きにして関数電卓で．または, Google Colab で np.exp(数値), np.log(数値).

Python で連続型確率変数

- expon: 指数分布 (固有名詞)
- loc(ation): 分布の (重心とは限らないけど) 位置を表す．指数分布の b.
- scale: 分布の (標準偏差とは限らないけど) 幅を表す．指数分布の a.

指数分布

```
import scipy.stats
rvx = stats.expon(loc=0,scale=a) # Exp(0,a) loc=b=0, scale=a=1
# 連続だけにある rvx.cdf(x)の逆関数
rvx.ppf(q) # ppf=percent point function 分位数関数(の類義語)

# 離散にも実はあった
rvx.moment(4) # 4次のモーメント
rvx.expect(lambda x: x**2 + 3* x +1) # E[X^2 + 3X + 1]
```

Google Colab 連続型確率変数-指数分布.ipynb hig3Moodle

指数分布の拡大縮小

確率変数 $X \sim \mathrm{Exp}(0, a)$, $Y = 2X$ とする.

1. Y の確率密度関数 $f_{\mathrm{Y}}(y)$ を求め, グラフを描こう. $f_{\mathrm{X}}(x)$ との関係は?
2. $\mathrm{E}[Y], \mathrm{V}[Y]$ を 2 つの方法で求めよう.

$$\mathrm{E}[2X] = \int_0^\infty 2x \frac{1}{a} \mathrm{e}^{-x/a} \, \mathrm{d}x \overset{y=2x}{=} \int_0^\infty y \frac{1}{a} \mathrm{e}^{-y/(2a)} \frac{1}{2} \, \mathrm{d}y = \mathrm{E}[Y]$$

ここまで来たよ

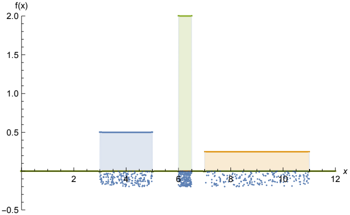
連続型一様分布 岩薩林 確率・統計 例題 4.1(p.78)

連続型一様分布は, 連続型確率変数の分布の, 名前がつくくらい有名な例.

連続型一様分布 $\mathrm{U}(c,d)$

確率変数 X の確率密度関数が次で与えられるとき, X は区間 $[c,d)$ の連続型一様分布 $\mathrm{U}(c,d)$ に従うという.

$$f(x) = \begin{cases} \frac{1}{d-c} & (c \le x \le d) \\ 0 & (\text{他}) \end{cases}$$



$\mathrm{U}(4,6), \mathrm{U}(6,6.5), \mathrm{U}(7,11).$

連続型一様分布 岩薩林 確率・統計 例題 4.1(p.78)

連続型一様分布

```
import scipy.stats
rvx = stats.uniform(loc=c,scale=d-c) # U(c,d)
```

Google Colab 連続型確率変数-連続型一様分布.ipynb hig3Moodle

L04-Q2

Quiz(連続型一様分布)

連続型確率変数 X が連続型一様分布 $\mathrm{U}(c,d)$ にしたがう.

1. モーメント $\mathrm{E}[X^k]$ を求めよう.
2. 母平均値 $\mathrm{E}[X]$ を求めよう.
3. 母標準偏差 $\sqrt{\mathrm{V}[X]}$ を求めよう.

$\mathrm{U}(c,d)$ に対するこの結果は, 公式のように記憶して使おう.
母平均値・母標準偏差の意味とマッチしてる?

L04-Q3

Quiz(連続型一様分布の応用)

あるおんぼろキューブアイス製造マシンから立方体の氷が, 一辺 X が 18mm 以上 20mm 以下の範囲で, 同じ確からしさで決まってランダムに出てくる.

1. X のしたがう分布を答えよう.
2. キューブアイスの一辺の長さの母期待値を求めよう.
3. キューブアイスの表面積の母期待値を求めよう.
4. キューブアイスの表面積が $2000[\mathrm{mm}^2]$ 以下である確率を求めよう.
5. キューブアイスの体積の母期待値を求めよう.

ここまで来たよ

3 連続型確率変数の母期待値と分位数

4 指数分布・一様分布と確率変数の標準化
- 指数分布
- 連続型一様分布
- 確率変数の標準化

確率変数の標準化

定義 (確率変数の標準化 岩薩林 確率・統計 例題 4.4(p.80))

任意の確率変数 X に対して, $\mu = \mathrm{E}[X], \sigma^2 = \mathrm{V}[X], \sigma > 0$ とする.
確率変数 $Z = \frac{X-\mu}{\sigma}$ を「X を標準化した確率変数」という.

$\mathrm{E}[Z] = \mathrm{E}[\frac{X-\mu}{\sigma}] = \boxed{0}$,
$\mathrm{V}[Z] = \mathrm{V}[\frac{X-\mu}{\sigma}] = \boxed{1}$
逆に, 標準化された Z から $X = aZ + b$ を考えると,
$\mathrm{E}[X] = b$,
$\mathrm{V}[X] = a^2$.

Z の確率密度関数 $f_{\mathrm{Z}}(z) \rightsquigarrow X$ の確率密度関数 $f_{\mathrm{X}}(x) = \frac{1}{a} \times f_{\mathrm{Z}}(\frac{x-b}{a})$.
b だけ平行移動, a だけ拡大.
$\frac{1}{a}\times$ は面積を 1 に保つもの.

グラフの平行移動と拡大縮小

$y=f(x)$ のグラフを, y 軸を中心に横 (x 軸方向) に a 倍すると $y=f(\frac{x}{a})$.
さらに, 横 (x 軸方向) に b 平行移動すると $y=f(\frac{x-b}{a})$.
$y=x, y=x^2$ を例に考えてね.
$z=\frac{x-b}{a}, x=az+b$ とおいてる感じです.

連続型一様分布の標準化

$X \sim \mathrm{U}(2,8)$,

$$f_{\mathrm{X}}(x) = \begin{cases} \frac{1}{6} & (2 \leq x \leq 8) \\ 0 & (\text{他} \end{cases}$$

を標準化しよう.

$\mathrm{E}[X] = 5, \mathrm{V}[X] = (8-2)^2/12 = 3$ より，$Z = \frac{X-5}{\sqrt{3}}$

$$f_{\mathrm{Z}}(z) = \begin{cases} \frac{\sqrt{3}}{6} & (-\sqrt{3} \leq z \leq \sqrt{3}) \\ 0 & (\text{他} \end{cases}$$

標準連続型一様分布 $Z \sim \mathrm{U}(-\sqrt{3}, +\sqrt{3})$.

標準化の逆: 標準からもとのやつ I

L04-Q4

Quiz(連続型一様分布)

連続型確率変数 $Z \sim \mathrm{U}(-\sqrt{3}, \sqrt{3})$ に対して, $X = \sqrt{3}Z + 5$ を考える.

1. X の確率密度関数 $f_{\mathrm{X}}(x)$ とそのグラフを答えよう.
2. $\mathrm{E}[X]$ を求めよう.
3. $\mathrm{V}[X]$ を求めよう.

それぞれ 2 つの方法で.

復習

$X = aZ + b$ の母平均値と母分散

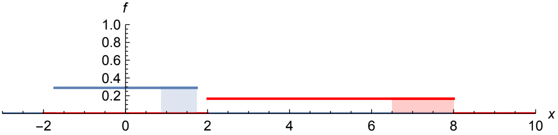
$$
\begin{aligned}
\mathrm{E}[X] =& \mathrm{E}[aZ + b] = a\mathrm{E}[Z] + b \\
\mathrm{V}[X] =& \mathrm{V}[aZ + b] = a^2\mathrm{V}[Z]
\end{aligned}
$$

$X = aZ + b$ の意味 岩薩林 確率・統計 標準化 (p.81)

Z が 連続型一様分布 $\mathrm{U}(c, d)$ にしたがうとき,
$X = aZ + b$ は 連続型一様分布 $\mathrm{U}(ac + b, ad + b)$ にしたがう.
一般に X の確率密度関数のグラフは, Z のものを横に a 倍して b 平行移動, (縦に $1/a$ 倍)
$f_\mathrm{X}(x) = \frac{1}{a} \times f_\mathrm{Z}(\frac{x-b}{a})$

岩薩林 確率・統計 例題 4.4

確率変数 $X = \sqrt{3}Z + 5$ の確率密度関数



左から $Z \sim \mathrm{U}(-\sqrt{3}, \sqrt{3})$, $X = aZ + b = \sqrt{3}Z + 5 \sim \mathrm{U}(2, 8)$.

岩薩林 確率・統計 例題 4.4

$X = aZ + b$ のとき

$$f_\mathrm{X}(x) = \frac{1}{a} \times f_\mathrm{Z}(\tfrac{x-b}{a}) = \frac{1}{a} \times \begin{cases} \frac{1}{2\sqrt{3}} & (-\sqrt{3} \le \frac{x-b}{a} \le +\sqrt{3}) \Leftrightarrow (b - a\sqrt{3} \le x < b + a\sqrt{3}) \\ 0 & (\text{他}) \end{cases}$$