連続型確率変数の母期待値と分位数

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 L03(2023-04-24 Mon)

最終更新: Time-stamp: "2023-04-24 Mon 11:29 JST hig"

今日の目標

- 連続型確率変数の確率 岩薩林 確率・統計 §4.1 ・分位数 岩薩林 確率・統計 なし を計算できる
- 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 岩薩林 確率・統計 §4.2

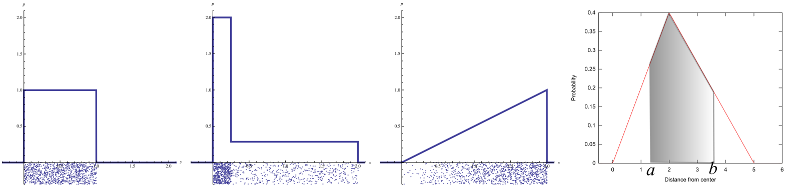
L02-Q1 $P(-\frac{1}{4} < X \leq \frac{1}{4}) = \int_{-\frac{1}{4}}^{+\frac{1}{4}} f(x)\,\mathrm{d}x = \int_{-\frac{1}{4}}^{0} 0\,\mathrm{d}x + \int_{0}^{+\frac{1}{4}} 8x\,\mathrm{d}x = \frac{1}{4}$.

ここまで来たよ

2 連続型確率変数の確率

3 連続型確率変数の母期待値と分位数
- 累積分布関数と分位数関数
- 連続型確率変数の母期待値
- 指数分布

復習: 連続型確率変数の確率密度関数



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)**)**

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

確率統計 I(2023)LL02 岩薩林 確率・統計 pp.50,51 (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)**)**

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

連続型確率変数の分位数関数

定義 (分位数関数 岩薩林 確率・統計 なし)

確率変数 X の 累積分布関数の, 区間 $[0,1]$ で定義された逆関数 $x = F^{-1}(y)$ を分位数関数, $F^{-1}(q)$ を, q-分位数 という.

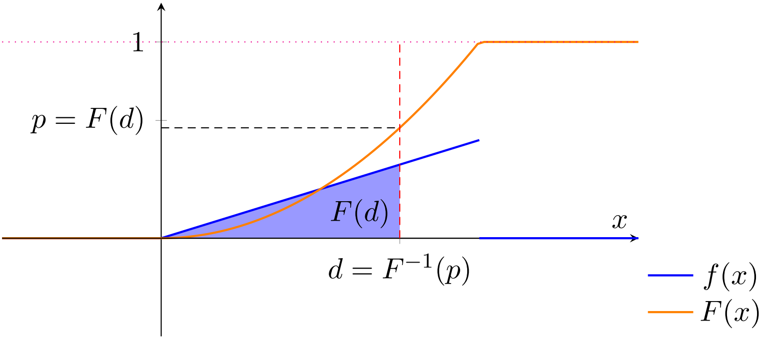
意味

q-分位数とは, 確率 $P(X \leq d)$ が q に等しくなるような境い目 $x = d$.
$F^{-1}(y)$ は, 確率 $y = q$ を与えると, そうなる境い目を返してくれる関数.

言い訳

$F(x)$ が狭義単調増加でない限り逆関数 F^{-1} は定義できない. 逆像 $F^{-1}(y)$ は集合になるので $\inf F^{-1}(y)$ とするのが正確. ただし, $F^{-1}(0)$ は極限 $\lim_{\epsilon\to+0} F^{-1}(+\epsilon)$ とする.
'確率 $P(X \leq d)$ が q に等しくなる**最小ぎりぎりの**境い目 $x = d$' のこと

```
rvx.ppf(q) # ppf=percent point function 分位数関数(の類義語)
```



例 (中央値・四分位数)

$F^{-1}(\frac{1}{2})$ は分布の (母) 中央値, $F^{-1}(\frac{1}{4}), F^{-1}(\frac{3}{4})$ は分布の (母) 四分位数.

データ分析

(標本) 中央値 岩薩林 確率・統計 p.27

L03-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 $\mathrm{E}[X^k]$ $(k = 0, 1, 2, \ldots)$
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 $P(-\frac{1}{4} < X \leq \frac{1}{4})$
7. 分布の中央値と $\frac{1}{4}$-分位数

ここまで来たよ

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 (4.9))

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

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

離散型・連続型共通の定義.

定義 (母平均値 岩薩林 確率・統計 (4.6)・母分散 岩薩林 確率・統計 (4.11))

母平均値 $\mu = \mathrm{E}[X^1]$, 母分散 $\sigma^2 = \mathrm{V}[X] = \mathrm{E}[(X-\mu)^2]$.

定義 (k 次のモーメント ($k = 0, 1, 2, 3, \cdots$) 岩薩林 確率・統計 (4.10))

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

復習: 母分散の性質 岩薩林 確率・統計 (3.8)(3.9)p.55

離散型・連続型共通の性質．公式 En, Vn.

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,

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

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

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

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

L03-Q1

岩薩林 確率・統計 例題 4.2,4.3

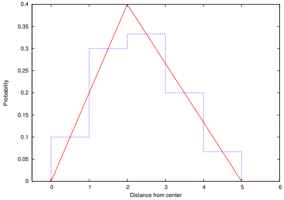
連続型確率変数の母平均値と母分散の直観的意味

$\mu = \mathrm{E}[X], \sigma^2 = \mathrm{V}[X]$

その 1: 確率密度関数のグラフから

$y=0$ と $y=f(x)$ で囲まれる形の (一様な質量密度 1 の) 板を考える.

$f(x)$ のグラフ



$\mathrm{E}[X^0]$: 板の 質量 は 1.

$\mathrm{E}[X^1]$ 板の 重心の横方向の位置 は $x=\mu$.

$\mathrm{V}[X]$ 板 の アバウトに幅 は σ,

$x=\mu$ のまわりの慣性モーメント は σ^2.

その 2: チェビシェフの不等式

母平均値から, (σ を単位に測って) 離れた値がでる確率は小さい.

その 3: 大数の弱法則

宝くじを何回も買うと, 1 回あたりの平均の賞金は,
E[賞金] に近い

ここまで来たよ

2. 連続型確率変数の確率

3. 連続型確率変数の母期待値と分位数
   - 累積分布関数と分位数関数
   - 連続型確率変数の母期待値
   - 指数分布

指数分布 岩薩林 確率・統計 p.80

定義 (指数分布)

連続型確率変数 X が 確率密度関数

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

を持つとき，X はパラメタ a,b の指数分布にしたがうといい (科目ローカル記号) $\mathrm{Exp}(b,a)$ とかく．(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，$b=0, a>0$ 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．

L03-Q2

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: 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

```
rvx = stats.expon(loc=0,scale=2) # loc=b=0, scale=a=1/lambda=2
```

分位数関数

```
rvx.ppf(q) # ppf=percent point function 分位数関数(の類義語)
```

rvx.cdf(x) の逆関数.

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