累積分布関数・分位数・指数分布・幾何分布

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(2025-04-28 Mon)

最終更新: Time-stamp: "2025-04-28 Mon 11:55 JST hig"

今日の目標

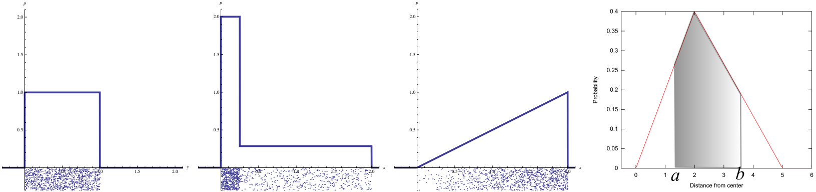
- 連続型確率変数の累積分布関数と分位点関数が計算できる 久保川 統計学入門 §5.1
- 指数分布 久保川 統計学入門 §5.3.3 の意味が説明でき計算できる
- 幾何分布 久保川 統計学入門 §4.3.4 の意味が説明でき計算できる

L02-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. $\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X^1]^2 = \frac{1}{8} - (\frac{1}{3})^2 = \frac{1}{72}$.
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. $\mathrm{Var}(2X+3) = 2^2\mathrm{Var}(X) = \frac{1}{18}$.
6. $P(-\frac{1}{4} < X \leq \frac{1}{4}) = \int_{-\frac{1}{4}}^{+\frac{1}{4}} f(x)\,\mathrm{d}x = \int_{-\frac{1}{4}}^{0} 0\,\mathrm{d}x + \int_0^{+\frac{1}{4}} 8x\,\mathrm{d}x = \frac{1}{4}$.

ここまで来たよ

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)

命題 (確率密度関数と確率 久保川 統計学入門 p.104)

$$P(c < X \leq d) = \int_c^d f(x)\,\mathrm{d}x \quad \text{(下側面積)}$$

定義 (累積分布関数 久保川 統計学入門 §4.1(p.83)§5.1(p.104))

$F(x) = P(X \le x)$ を累積分布関数（分布関数, 確率分布関数）という.

命題 (累積分布関数の性質 久保川 統計学入門 p.104)

(C1) $F(x)$ は広義単調増加関数
(C2) $F(-\infty) = 0, F(+\infty) = 1.$
(C3) $P(c < X \le d) = F(d) - F(c).$

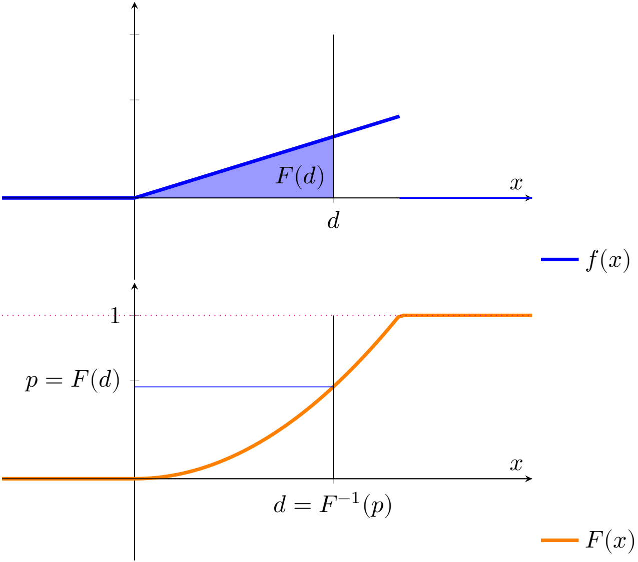
定義 (連続型確率変数の累積分布関数 久保川 統計学入門 p.104)

$$F(x) = \int_{-\infty}^{x} f(x')\, \mathrm{d}x'$$

命題 (久保川 統計学入門 公式 5.1(p.105))

F が微分可能なとき，$f(x)$ は $F(x)$ の導関数

$\boxed{\text{確率密度関数 } f}$ $\underset{\text{微分}}{\overset{\text{積分}}{\rightleftarrows}}$ $\boxed{\text{累積分布関数 } F}$ 久保川 統計学入門 例題 5.2



久保川 統計学入門 図 5.4

ここまで来たよ

2 連続型確率変数

3 累積分布関数・分位数・指数分布・幾何分布
- 累積分布関数
- 指数分布
- 分位数

指数分布 久保川 統計学入門 §5.3(p.116) 久保川 統計学入門 §5.3.3

連続型確率変数の分布の, 名前がつく有名な例. パラメタ（母数）$\lambda > 0$ を持つ family (パラメタは確率変数でない. $8x^1$ の 8 や 1 のこと).

定義 (指数分布)

連続型確率変数 X が確率密度関数

$$f(x;\lambda) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

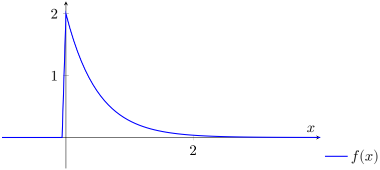
を持つとき, X はパラメタ $\lambda > 0$ の指数分布に従うといい, 記号 $\sim$(従う) で $X \sim Ex(\lambda)$ と書く ($Ex()$ は 久保川 統計学入門 ローカル記号)

性質 実は $\mathrm{E}[X^k] = \lambda^{-k}$ (多数回部分積分) 久保川 統計学入門 公式 5.19
意味時間的にランダムに起きる事象, 例えば,「ある機械が x 秒後に初めて故障する」「サッカーで x 秒後に次のゴールが起きる」のような事象の, 時間間隔 x のしたがう分布. 母平均値 $\mathrm{E}[X] = 1/\lambda$.

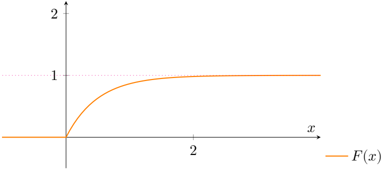
$X \sim Ex(\lambda)$, 確率密度関数 $f(x) = \lambda e^{-\lambda x}$ のとき，累積分布関数

$$F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x' = \begin{cases} \int_{-\infty}^{x} 0\ \mathrm{d}x' = 0, & (x \le 0) \\ \int_{-\infty}^{0} 0\ \mathrm{d}x' + \int_{0}^{x} \lambda e^{-\lambda x'}\mathrm{d}x' = 1 - e^{-\lambda x}. & (x > 0) \end{cases}$$

確率密度関数



累積分布関数



L03-Q1

Quiz(指数分布の累積分布関数・確率・分位数)

連続型確率変数 X はパラメタ $\lambda = 3$ の指数分布 $Ex(3)$ にしたがう．確率密度関数 $f(x)$ は次で与えられる．

$$f(x) = \begin{cases} 3\mathrm{e}^{-3x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}.$$

1. X の累積分布関数 $F(x)$ を求めよう．
2. 確率 $P(X \leq 4)$ を求めよう．
3. 確率 $P(-5 < X \leq 2)$ を求めよう．
4. 確率 $P(X \leq d) = \frac{1}{7}$ となる $d \in \mathbb{R}$ を求めよう．

指数・対数関数の値は Windows の電卓．iPhone を横向きにして関数電卓で．または, Google Colab で import numpy as np のあと，np.exp(数値), np.log(数値).

L03-Q2

Quiz(指数分布)

あるサッカーチームが，開始後 X 分で最初の得点をするとしたとき，この X はパラメタ $\lambda = 1/30$ の指数分布にしたがう（開始直後の特別なイベントは無視してゲームは時間的に一様とする．毎週の試合をつなげて考える, 本当は相手のチームによるだろうけど無視）.

1. 開始後 5 分から 10 分に最初の得点する確率を求めよう．有効数字 3 桁で答えよう．
2. 開始後 80 分たっても得点していない確率を求めよう．有効数字 3 桁で答えよう．
3. 得点と次の得点の間の時間の母平均値を求めよう．有効数字 3 桁で答えよう．
4. 1 点目がはいる確率が $2/3$ を超える時刻を求めよう．

指数・対数関数の値は Windows の電卓．iPhone を横向きにして関数電卓で．または, Google Colab で import numpy as np のあと，np.exp(数値), np.log(数値).

ここまで来たよ

2. 連続型確率変数

3. 累積分布関数・分位数・指数分布・幾何分布
   - 累積分布関数
   - 指数分布
   - 分位数

連続型確率変数の分位数関数

定義 (分位数関数 久保川 統計学入門 p.109)

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-(下側) 分位点 $F^{-1}(1-q)$ を, q-上側分位点 という.

意味

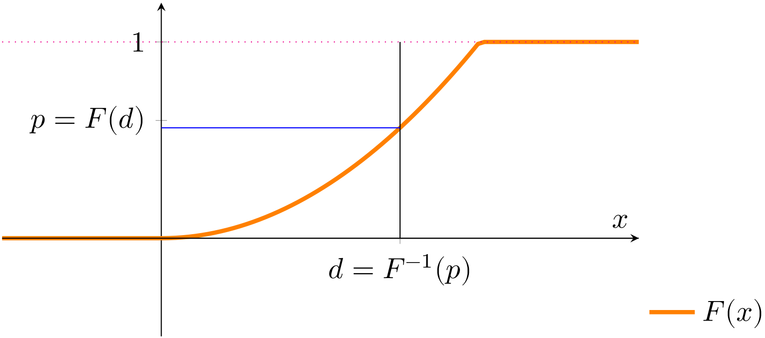
'q-下側分位点とは，確率 $P(X \leq d)$ が q に等しくなる**最小ぎりぎりの**境い目 $x = d$'

$F^{-1}(y)$ 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.

'q-上側分位点とは，確率 $P(X \geq d)$ が q に等しくなる**最大ぎりぎりの**境い目 $x = d$'

言い訳

定義域 $[0,1]$ の端は含められない場合もある． $F(x)$ が狭義単調増加でない限り逆関数 F^{-1} は定義できない．これらのややこしいケースには深入りしない．離散型にも深入りできない．



例 (中央値・四分位点)

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

データ分析

(標本) 中央値，四分位点 久保川 統計学入門 §1.2.3

scipy による連続型確率変数

https://colab.research.google.com

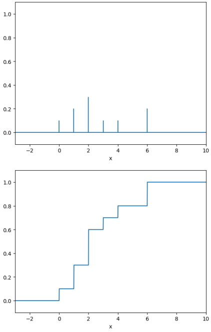
```
import numpy as np # ライブラリの読込
from scipy import stats

rvx = stats.expon(loc=0,scale=0.2) # 指数分布, 専用の名前がある.
                                   # scale=1/λ
# 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.moment(k) k 次のモーメント
rvx.pdf(x) 確率密度関数 probability density function
rvx.cdf(x) 累積分布関数 cumurative density function
rvx.ppf(p) 分位数関数 percent point function

# データ分析 でやってた標本ナントカ 毎回違った結果
sample=rvx.rvs(size=30) # サイズ30の標本抽出
sample.mean() # 標本平均値
sample.var()  # 不偏標本平均値
sample.std()  # 標本標準偏差
```

離散型確率変数の累積分布関数

連続型と同様の定義 $F(x) = P(X \leq x)$

命題 (離散型確率変数の累積分布関数 久保川 統計学入門 p.83)

累積分布関数の累積分布関数は，確率関数 $p(x)$ で，$F(x) = \sum_{x' \leq x} p(x')$ と書ける．



久保川 統計学入門 図 4.1

幾何分布 久保川 統計学入門 §4.3.4(p.97)

定義 (幾何分布 久保川 統計学入門 §4.3.4(p.97))

整数値をとる離散型確率変数 X が，パラメタ $0 < p < 1$ として，確率関数

$$p(x) = \begin{cases} 0 & (x < 0) \\ (1-p)^x p & (0 \le x) \end{cases}$$

を持つとき，X はパラメタ p の幾何分布 $Geo(p)$ にしたがうという．

表の出る確率が p のコインを繰り返し投げ続ける中で，初めて表が出るまでに要した裏の回数 X のしたがう分布．

L03-Q3

Quiz(離散的な確率変数の累積分布関数)

整数に値をとる離散型確率変数 X が次の確率関数を持つ．この X のしたがう分布をパラメタ p の幾何分布 $Geo(p)$ という．

$$p(x) = \begin{cases} (1-p)^x p & (x = 0, 1, 2, \dots) \\ 0 & (\text{他}) \end{cases}$$

1. 累積分布関数 $F(x)$ を，整数 x に対して求めよう．