指数分布・一般の正規分布

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　L05(2022-05-09 Mon)

最終更新: Time-stamp: "2022-05-09 Mon 11:24 JST hig"

今日の目標

- Python の scipy.stats を使える
- 指数分布を例に, 累積分布関数の意味を説明できる
- 一般の正規分布の確率密度関数, グラフ, 母期待値が求められる

L04-Q1

Quiz 解答: 連続型一様分布

1. $X \sim \mathrm{U}(1-2\sqrt{3}, 1+2\sqrt{3})$.
2. $\mathrm{E}[X] = \frac{(1-2\sqrt{3})+(1+2\sqrt{3})}{2}$, または, $\mathrm{E}[X] = 2\mathrm{E}[Z]+1 = 0+1$.
3. $\mathrm{V}[X] = \frac{((1+2\sqrt{3})-(1-2\sqrt{3}))^2}{12}$, または, $\mathrm{V}[X] = 2^2\mathrm{V}[Z] = 2^2 \cdot 1$.

L04-Q2

Quiz 解答: 標準正規分布の確率

累積分布関数を $F(z)$ とする. また, $I(z) = P(0 \le Z \le z)$
岩薩林 確率・統計 付表 1 とする.

1. $P(-0.56 < Z < +1.23) = \int_{-0.56}^{1.23} f(z; 0, 1^2)\,\mathrm{d}z =$
   $F(1.23) - F(-0.56) =$ stats.norm(loc=0,scale=1).cdf(1.23) $-$
   stats.norm(loc=0,scale=1).cdf(-0.56) $= 0.6030$.
   また,
   $P(-0.56 < Z < +1.23) = I(1.23) - I(-0.56) = I(0.56) + I(1.23)$.
2. $P(Z < 1.23) = \int_{-\infty}^{1.23} f(z; 0, 1^2)\,\mathrm{d}z = F(1.23) - F(-\infty) =$
   $F(1.23) - 0 = 0.5708$.
   また, $P(-\infty < Z < 1.23) = I(1.23) - I(-\infty) = I(1.23) - (-\frac{1}{2})$.

ここまで来たよ

5 正規分布

5 指数分布・一般の正規分布
- 累積分布関数の復習
- 指数分布
- 一般の正規分布

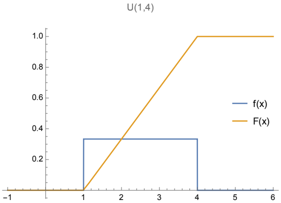
累積分布関数 岩薩林 確率・統計 pp.50,77

一般の離散型, 連続型確率変数 X に対して,

定義 (累積) 分布関数

$$\text{離散型} F(x) = \sum_{x' \le x} p(x') = P(X \le x), \tag{3.2}$$

$$\text{連続型} F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x' = P(X \le x) \tag{4.4}$$



逆に言うと $F'(x) = f(x)$.
大文字 X は確率変数 (別に書いた確率密度にしたがう) で, 値を代入できないし, 左辺に現れない. 偏微分 $f_{\mathrm{xy}}(x, y)$ みたいな記号?

確率密度関数 $f(x)(\geq 0)$
$P(c < X \leq d) = \int_c^d f(x')\mathrm{d}x'$
scipy.stats. 分布名.pdf(x)

累積分布関数 $F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x'$. $F'(X) = f(x)$.
$P(c < X \leq d) = F(d) - F(c)$. $P(X \leq d) = F(d)$.
$F(x)$ は広義単調増加. $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$.
scipy.stats. 分布名.cdf(x)

分位点関数 (quantile function) $F^{-1}(p)$. $p = F(x)$ の逆関数.
$F^{-1}(p) = (P(X \leq d) = p \text{となる} d)$.
定義域 $[0, 1]$, 値域 $[-\infty, +\infty]$. (ただし, $f(x) = 0$ となるところでは複数の原像があるのでその下限とする.
scipy.stats. 分布名.ppf(p), ppf=percent point function

ここまで来たよ

5 正規分布

5 指数分布・一般の正規分布
- 累積分布関数の復習
- 指数分布
- 一般の正規分布

指数分布

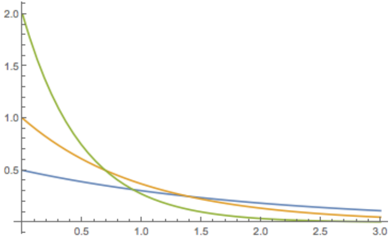
指数分布 岩薩林 確率・統計 例題 4.2(p.80)

連続型確率変数 X でつぎの確率密度関数をもつものを「パラメタ $\lambda(>0)$ の指数分布」にしたがうという.

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

意味 回数が時間の長さに比例して (単位時間に平均 λ 回), ランダムに (一定間隔でなく) 起きるできごとの, 時間間隔の x の分布.

例 全国で震度 n 以上の地震の起きる時間間隔, 機械の故障する時間間隔 (=寿命), サッカーで点が入る時間間隔



$\lambda=0.5,1,2.$

指数分布のモーメントと母期待値

$$\mathrm{E}[X^k] = \text{部分積分} = \qquad \lambda^{-k}k!$$

$$\mathrm{E}[X] = \boxed{1/\lambda}, \mathrm{V}[X] = \boxed{1/\lambda^2}$$

Python の scipy.stats.expon(loc=0,scale=$1/\lambda$)
$1/\lambda$ に 1 回起きる, 単位時間に λ 回起きる.
λ の次元 (単位) は 1/時間, 秒 $^{-1}$.

L05-Q1

Quiz(指数分布の確率)

連続型確率変数 X はパラメタ $\lambda = 2$ の指数分布にしたがう.

1. 確率 $P(X \leq 3)$ を小数で求めよう.
2. 累積分布関数 $F(x)$ を求めよう.
3. 確率 $P(3 < X \leq 4)$ を小数で求めよう.
4. 確率 $P(X \geq 4)$ を小数で求めよう.
5. 確率 $P(X \leq a) = 3/4$ となる a を小数で求めよう.

Windows の電卓や, iPhone を横向きにした関数電卓が役立つかも.

L05-Q2

Quiz(指数分布)

あるサッカーチームは, 1 ゲーム 90 分で平均 4.5 点得点できる (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. このチームの, 得点と得点の時間間隔のしたがう分布を答えよう.
2. 得点と得点の時間間隔の母平均値を求めよう.
3. 得点と得点の時間間隔が 5 分未満である確率を求めよう.
4. 得点と得点の時間間隔が 15 分以上 25 分未満になる確率を求めよう.

ここまで来たよ

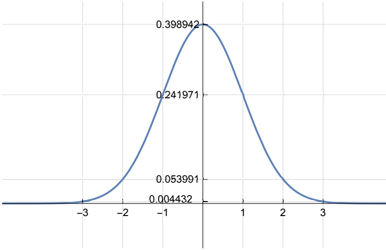
5 正規分布

5 指数分布・一般の正規分布
- 累積分布関数の復習
- 指数分布
- 一般の正規分布

ふたたび一般の正規分布 $\mathrm{N}(b, a^2)$

$Z \sim \mathrm{N}(0, 1^2)$ に対して, $X = aZ + b$ を考える.

$$\mathrm{E}[X^0] = 1, \quad \text{微積分 II}$$
$$\mu = \mathrm{E}[X] = \mathrm{E}[aZ + b] = b,$$
$$\sigma^2 = \mathrm{V}[X] = \mathrm{V}[aZ + b] = a^2, \quad \text{微積分 II}$$



$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \rightsquigarrow f_X(x) = \frac{1}{\sqrt{2\pi a^2}} e^{-\frac{(x-b)^2}{2a^2}}$$
$$Z \sim \mathrm{N}(0, 1^2) \rightsquigarrow X \sim \mathrm{N}(b, a^2).$$

(一般の) 正規分布 $\mathrm{N}(\mu, \sigma^2)$

母平均値 μ, 母分散 σ^2 の正規分布 $\mathrm{N}(\mu, \sigma^2)$ の確率密度関数は,

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

L05-Q3

Quiz(正規分布の確率)

連続型確率変数 X は, 確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi \cdot 3^2}} \mathrm{e}^{-\frac{(x-4)^2}{2\cdot 3^2}}$$

にしたがう.

1. $\mathrm{E}[X]$ を求めよう.
2. $\mathrm{V}[X]$ を求めよう.
3. $f(x)$ のグラフを, 標準正規分布の確率密度関数と重ねて描こう.

Python でも描いてみよう.

L05-Q4

Quiz(正規分布の確率)

連続型確率変数 X は, 確率密度関数

$$f(x) = C \times \mathrm{e}^{-8x^2+48x}$$

を持つ. ここで C は定数である.

1. $\mathrm{E}[X]$ を求めよう.
2. $\mathrm{V}[X]$ を求めよう.
3. $f(x)$ のグラフを, 標準正規分布の確率密度関数と重ねて描こう.

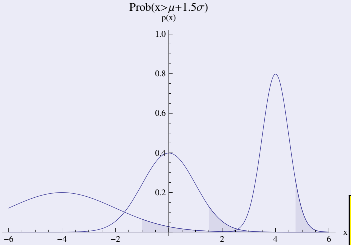
Hint. 平方完成

正規分布 $\mathrm{N}(\mu, \sigma^2)$ の確率 I

標準化前後の確率

一般に, 確率は標準化の前後で変わらない.
$X \sim \mathrm{N}(\mu, \sigma^2)$ の確率は, 標準化 $Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0, 1^2)$ して求めることができる.

$$P(c < X < d) = P(\tfrac{c-\mu}{\sigma} < \tfrac{X-\mu}{\sigma} < \tfrac{d-\mu}{\sigma}) = P(\tfrac{c-\mu}{\sigma} < Z < \tfrac{d-\mu}{\sigma})$$



斜線部の面積はどれも同じ

標準化しても確率が同じことの別説明 (置換積分)

$X \sim \mathrm{N}(\mu, \sigma^2)$ のとき, 積分で $z = \frac{x-\mu}{\sigma}$, $\mathrm{d}z = \frac{1}{\sigma}\mathrm{d}x$ とすると,

$$
\begin{aligned}
P(c < X < d) &= \int_c^d \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\, \mathrm{d}x \\
&= \int_{\frac{c-\mu}{\sigma}}^{\frac{d-\mu}{\sigma}} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{z^2}{2}}\, \sigma \mathrm{d}z \\
&= P\left(\frac{c-\mu}{\sigma} < Z < \frac{d-\mu}{\sigma}\right)
\end{aligned}
$$

L05-Q5

Quiz(正規分布の確率)

確率変数 $X \sim \mathrm{N}(3, 2^2)$, $Z \sim \mathrm{N}(0, 1^2)$ とする.

1. X, Z の確率密度関数の式を書こう. グラフを重ねて描こう.
2. 母期待値 $\mathrm{E}[X^2]$ を求めよう.
3. 確率 $P(X \geq 5)$ と等しくなるように $P(c \leq Z \leq d)$ の c,d を求めよう. また, 確率を Python や表を使って小数で求めよう.
4. 確率 $P(+1 \leq X \leq 7)$ と等しくなるように $P(c \leq Z \leq d)$ の c,d を求めよう. また, 確率を Python や表を使って小数で求めよう.
5. 確率 $P(+3 \leq X \leq 7)$ と等しくなるように $P(c \leq Z \leq d)$ の c,d を求めよう. また, 確率を Python や表を使って小数で求めよう.
6. 確率 $P(X \leq a) = 3/4$ となる a を Python や表を使って小数で求めよう.

岩薩林 確率・統計 §4.5 例題 4.10,4.11, 問題 10, 第 4 章練習問題 5