確率変数の標準化・標準正規分布・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(2023-05-08 Mon)

最終更新: Time-stamp: "2023-05-08 Mon 11:26 JST hig"

今日の目標

- 確率変数の標準化を説明できる 岩薩林 確率・統計 p.80
- 正規分布の確率密度関数, グラフを書ける．母期待値，確率，分位数を求められる． 岩薩林 確率・統計 §4.5

L04-Q1

Quiz 解答: 指数分布の母平均値・母分散・確率

1. 部分積分で $\mathrm{E}[X] = 3$.
2. 部分積分で $\mathrm{E}[X^2] = 2 \cdot 3^2$, $\mathrm{V}[X] = 3^2$.
3. $0 \le x$ で $F(x) = 1 - \mathrm{e}^{-x/3}$.
   $P(1 < X \le 3) = F(3) - F(1) = \int_1^3 \frac{1}{3}\mathrm{e}^{-x/3}\mathrm{d}x = \mathrm{e}^{-1/3} - \mathrm{e}^{-1} = 0.349$.
4. $F^{-1}(y) = -3\log(1-y)$. $F^{-1}(\frac{1}{3}) = -3\log\frac{2}{3} = 1.216$.

L04-Q2

Quiz 解答: 連続型一様分布

1. $\mathrm{E}[X^k] = \frac{1}{d-c}\int_c^d x^k \,\mathrm{d}x = \frac{1}{k+1}\frac{d^{k+1}-c^{k+1}}{d-c} = \frac{1}{k+1}(d^k + d^{k-1}c + \cdots + c^k)$.
2. $\mathrm{E}[X^1] = \frac{c+d}{2}$.
3. $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X^1]^2 = \frac{(d-c)^2}{12}$. $\sqrt{\mathrm{V}[X]} = \frac{d-c}{\sqrt{12}} \simeq \frac{d-c}{3.5}$.

L04-Q3

L04-Q4

Quiz 解答: 連続型一様分布

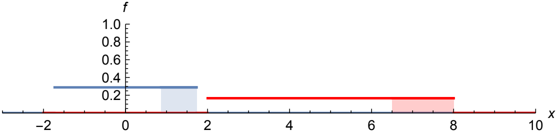
1. $X \sim \mathrm{U}(\sqrt{3}(-\sqrt{3})+5, \sqrt{3}(+\sqrt{3})+5)$, すなわち，$X \sim \mathrm{U}(2, 8)$. 関数とグラフ略.
2. $\mathrm{E}[X] = \frac{2+8}{2}$, または, $\mathrm{E}[X] = \sqrt{3}\mathrm{E}[Z] + 5 = \sqrt{3} \cdot 0 + 5$.
3. $\mathrm{V}[X] = \frac{(8-2)^2}{12} = 3$, または, $\mathrm{V}[X] = (\sqrt{3})^2 \mathrm{V}[Z] = 3 \cdot 1$.

確率変数の変数変換 $X = aZ + b$ の意味 岩薩林 確率・統計 標準化 (p.81)

$Z \sim \mathrm{U}(c, d)$ (連続型一様分布) のとき, $X = aZ + b \sim \mathrm{U}(ac + b, ad + b)$.
一般に確率変数 X の確率密度関数のグラフは, Z のものを横に a 倍, 横に b 平行移動, (縦に $1/a$ 倍) 岩薩林 確率・統計 例題 4.4

$f_X(x) = \frac{1}{a} \times f_Z(\frac{x-b}{a})$

左 $Z \sim \mathrm{U}(-\sqrt{3}, \sqrt{3})$ 右 $X = aZ + b = \sqrt{3}Z + 5 \sim \mathrm{U}(2, 8)$ の確率密度関数



$$f_X(x) = \frac{1}{a} \times f_Z(\tfrac{x-b}{a}) = \frac{1}{a} \times \begin{cases} \frac{1}{2\sqrt{3}} & (-\sqrt{3} \leq \frac{x-b}{a} \leq +\sqrt{3}) \Leftrightarrow (b - a\sqrt{3} \leq x < b + a\sqrt{3}) \\ 0 & (\text{他}) \end{cases}$$

対応する確率=塗った箇所の面積は同じ

$P(\frac{13}{2} < X \leq 8) = P(\frac{\frac{13}{2}-5}{\sqrt{3}} < Z \leq \frac{8-5}{\sqrt{3}})$.

確率変数の標準化

定義 (標準化された確率変数 岩薩林 確率・統計 例題 4.4(p.80)**)**

確率変数 Z が, $\mathrm{E}[Z]=0, \mathrm{V}[Z]=1^2$ であるとき, Z は標準化された確率変数という.

任意の確率変数 X は, 1 次式 $Z=\frac{X-b}{a}$ で標準化された確率変数に変換できる.

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 に対して,
$\mathrm{E}[aZ+b]=b$, $\mathrm{V}[aZ+b]=a^2$.

標準一様分布

任意の連続型一様分布 $\mathrm{U}(c, d)$ を標準化すると, $\mathrm{U}(-\sqrt{3}, +\sqrt{3})$ になる.

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.

確率変数の標準化とは

確率密度関数 $f_X(x)$ を横方向に拡大縮小して '幅' 1 にする（副産物として縦にも拡大縮小する），平行移動して平均値 '重心' の位置を 0 にすること

ここまで来たよ

4 指数分布・一様分布と確率変数の標準化

5 確率変数の標準化・標準正規分布・一般の正規分布
- 標準正規分布
- 一般の正規分布

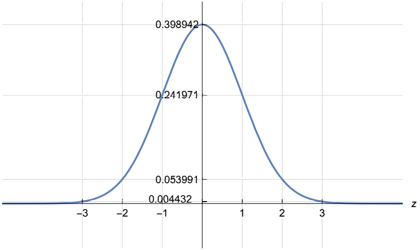
標準正規分布 standard normal distribution $\mathrm{N}(0,1^2)$ の性質

定義 (標準正規分布 $\mathrm{N}(0,1^2)$ 岩薩林 確率・統計 (4.17))

次の確率密度関数 を持つ確率変数 Z を，標準正規分布 $\mathrm{N}(0,1^2)$ にしたがうという.

$$f(z;0,1^2)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}.$$

Python で `scipy.stats.norm(loc=0,scale=1)`,
Excel で`=norm.s.dist(z,FALSE)`



標準正規分布 $Z \sim \mathrm{N}(0, 1^2)$ の母期待値

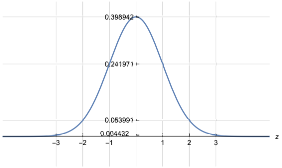
k 次のモーメント (k: 自然数)

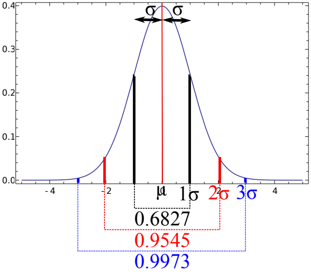
$$
\begin{aligned}
\mathrm{E}[Z^{2k-1}] &= 0, \quad \text{奇関数} \\
\mathrm{E}[Z^{2k}] &= (2k-1)!! \\
&= 1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2k-1) \quad \text{部分積分} \\
\text{全確率}\mathrm{E}[Z^0] &= 1, \\
\text{母平均値}\mathrm{E}[Z] &= 0, \\
\text{母分散}\mathrm{V}[Z] &= 1
\end{aligned}
$$

岩薩林 確率・統計 (4.19) 微積分 II

岩薩林 確率・統計 (4.20)

岩薩林 確率・統計 (4.21)





たしかに, Z は標準化された分布. Google Colab Python による正規分布 hig3Moodle

標準正規分布の確率と $I(z)$ の数表

$Z \sim \mathrm{N}(0, 1^2)$ のとき,

$$P(c < Z \leq d) = \int_c^d f(z'; 0, 1^2)\,\mathrm{d}z' = \int_c^d \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z'^2}{2}}\,\mathrm{d}z'$$ 岩薩林 確率・統計 (4.22)

不定積分は, 簡単な関数では書けないが, 累積分布関数を使って,

$$P(c < Z \leq d) = F(d) - F(c)$$

とし, $F(z)$ を `scipy.stats.norm().cdf(z)` や数表 岩薩林 確率・統計 付表 1(p.227) から求める. 分位数関数 $F^{-1}(q)$ も Python や数表 岩薩林 確率・統計 付表 1(p.227) に頼る.

注: つねに $F(-\infty) = 0, F(+\infty) = 1$. 標準正規分布は確率密度関数が偶関数なので $F(z) + F(-z) = 1$. 特に $F(0) = 1/2$.

岩薩林 確率・統計 付表 1 には $I(z) = F(z) - \frac{1}{2}, z > 0$ の表が載っている. 高校 数学 B

$z < 0$ については, $I(-z) = I(z)$ を経由して求める.

上側確率 $Q(z) = \frac{1}{2} - I(z) = 1 - F(z)$ の表を載せてる教科書も多い.

L05-Q1

Quiz(標準正規分布の確率)

Z は標準正規分布 $\mathrm{N}(0, 1^2)$ に従う連続型確率変数である.

1. 確率 $P(Z \leq 1.23)$ 累積分布関数 $F(z)$ で表そう．Python または表で小数として求めよう.
2. 確率 $P(-0.56 < Z \leq +1.23)$ を累積分布関数 $F(z)$ で表そう. Python または表で求めよう.
3. 確率 $P(Z > d) = 0.025$ となる d を累積分布関数 $F(z)$ の逆関数で表そう．Python または表で求めよう.

岩薩林 確率・統計 例題 4.9(p.91), 問題 8(p.92), 問題 10(p.96)

ここまで来たよ

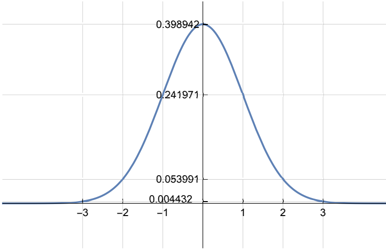
4 指数分布・一様分布と確率変数の標準化

5 確率変数の標準化・標準正規分布・一般の正規分布
- 標準正規分布
- 一般の正規分布

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.

$$
\begin{aligned}
\mathrm{E}[X^0] &= 1, \\
\mu = \mathrm{E}[X] &= \mathrm{E}[aZ + b] = b, \\
\sigma^2 = \mathrm{V}[X] &= \mathrm{V}[aZ + b] = a^2,
\end{aligned}
$$



$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} \rightsquigarrow f_X(x) = \frac{1}{\sqrt{2\pi a^2}} \mathrm{e}^{-\frac{(x-b)^2}{2a^2}}$$

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

次の確率密度関数を持つ確率変数 X を，母平均値 μ, 母分散 σ^2 の正規分布 $\mathrm{N}(\mu, \sigma^2)$ (normal distribution) にしたがうという.

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

L05-Q2

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.

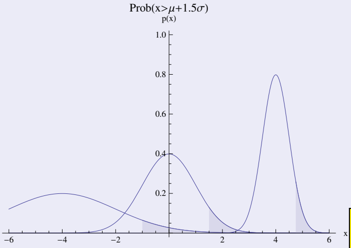
Python でも描いてみよう.

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

命題 (標準化前後の確率は同じ)

一般に, 確率は標準化 $Z = \frac{Z-b}{a}$ の前後で変わらない.

$$P(c < X \leq d) = P\left(\frac{c-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{d-\mu}{\sigma}\right) = P\left(\frac{c-\mu}{\sigma} < Z \leq \frac{d-\mu}{\sigma}\right)$$



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,

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

L05-Q3

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 $P(X > 5) = P(c < Z \leq d)$ となるように c,d を定めよう. また, 確率を Python や表を使って小数で求めよう.
4. 確率 $P(+1 < X \leq 7) = P(c \leq Z \leq d)$ となるように c,d を定めよう. また, 確率を Python や表を使って小数で求めよう.
5. 確率 $P(+3 < X \leq 9) = P(c \leq Z \leq d)$ となるように c,d を定めよう. また, 確率を Python や表を使って小数で求めよう.
6. 確率 $P(Z \leq d) = 3/4$ となる d を, Python や表を使って小数で求めよう. 確率 $P(X \leq d') = 3/4$ となるような d' を d で表そう.

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