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　L04(2022-05-02 Mon)

最終更新: Time-stamp: "2022-05-02 Mon 07:34 JST hig"

今日の目標

- 標準化前後の確率変数の関係を説明できる
- Python の scipy.stats を使える
- 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

L03-Q1

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. $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X^1]^2 = \frac{(d-c)^2}{12}$. $\sqrt{\mathrm{V}[X]} = \frac{d-c}{\sqrt{12}} \simeq \frac{d-c}{3.5}$.

L03-Q2

Quiz 解答: 連続型一様分布の応用

1. $X \sim \mathrm{U}(18, 20)$.
2. $\mathrm{E}[X^1] = 19$ [mm].
3. $\mathrm{E}[6X^2] = 2168 \neq 6 \cdot 19^2 = 2166$ [mm^2].
4. $P(6X^2 \leq 2000) = P(X \leq (\frac{2000}{6})^{1/2}) = \int_{19}^{(2000/6)^{1/2}} \frac{1}{20-18}\,\mathrm{d}x = 5\sqrt{10/3} - 9$.
5. $\mathrm{E}[X^3] = 6878 \neq 19^3 = 6859$ [mm^3].

L03-Q3

Quiz 解答: 確率変数の変換

$\mathrm{E}[X^2] \overset{\text{V1}}{=} \mathrm{V}[X] + \mathrm{E}[X]^2 = 13.$

1. $\mathrm{E}[-X^2 + 2X - 3] \overset{\text{E2}}{=} \mathrm{E}[-X^2] + \mathrm{E}[2X] + \mathrm{E}[-3] \overset{\text{E3}}{=}$ $-\mathrm{E}[X^2] + 2\mathrm{E}[X] - 3\mathrm{E}[1] \overset{\text{E1}}{=} -13 + 2 \cdot 2 - 3 \cdot 1 = -12.$
2. $\mathrm{V}[-2X - 3] \overset{\text{V2}}{=} (-2)^2 \mathrm{V}[X] = 36.$

L03-Q4

Quiz 解答: 連続型一様分布の母期待値

$\mathrm{E}[X] = \frac{-1+1}{2} = 0, \mathrm{V}[X] = \frac{(1-(-1))^2}{12} = \frac{1}{3}$

1. $\mathrm{E}[2X + 4] = 2 \cdot 0 + 4.$
2. $\mathrm{V}[2X + 4] = 2^2 \times \frac{1}{3}.$
3. $2X + 4 > 5 \Leftrightarrow X > \frac{1}{2}.$ $P(X > \frac{1}{2}) = \int_{\frac{1}{2}}^{1} \frac{1}{2}\, \mathrm{d}x = \frac{1-\frac{1}{2}}{2} = \frac{1}{4}.$

別解 $Y = 2X + 4 \sim \mathrm{U}(2, 6)$ なので,

1. $\mathrm{E}[Y] = \frac{2+6}{2} = 4$.
2. $\mathrm{V}[Y] = \frac{(6-2)^2}{12} = \frac{1}{3}$.
3. 面積の比を考えて, $P(Y > 5) = \frac{6-5}{6-2} = \frac{1}{4}$.

L03-Q5

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, $\mathrm{V}[X] = 2^2\mathrm{V}[Z] = 2^2 \cdot 1$.

ここまで来たよ

4 連続型一様分布, 変数変換, 確率変数の標準化

4 正規分布
- 確率変数の標準化
- Python での確率変数の計算
- 標準正規分布
- 一般の正規分布

確率変数の標準化

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,
$\mathrm{E}[X] = b,$
$\mathrm{V}[Y] = a^2.$

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.
逆に $f_{\mathrm{X}}(x) \rightsquigarrow f_{\mathrm{Z}}(z) = \sigma \times f_{\mathrm{Z}}(\sigma z + \mu).$

例: 連続型一様分布の標準化

$X \sim \mathrm{U}(2,6)$,

$$f_{\mathrm{X}}(x) = \begin{cases} \frac{1}{4} & (2 \leq x \leq 6) \\ 0 & (他 \end{cases}$$

を標準化すると, $Z = \frac{X - \frac{2+6}{2}}{\frac{6-2}{\sqrt{12}}} = \frac{X-4}{2/\sqrt{3}}$

$$f_{\mathrm{Z}}(z) = \frac{2}{\sqrt{3}} \times \begin{cases} \frac{1}{4} & (-\sqrt{3} \leq z \leq \sqrt{3}) \\ 0 & (他 \end{cases}$$

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-Q1

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, $X = 2Z + 1$ を考える.

1. X の確率密度関数 $f_X(x)$ 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.

ここまで来たよ

4 連続型一様分布, 変数変換, 確率変数の標準化

4 正規分布
- 確率変数の標準化
- Python での確率変数の計算
- 標準正規分布
- 一般の正規分布

Google Colaboratory

データ分析の説明参照 `https://www-tlab.math.ryukoku.ac.jp/wiki/?Data/2021/ex02#colab`
サンプル `https://colab.research.google.com/drive/1KJB0I_DGj73r0WPtwYZYA0odLNk1DSB2`

- ファイル → ドライブにコピーを保存
  - Google Drive のフォルダ Google Colaboratory に保存される.
- または, `https://colab.research.google.com` でノートブックを新規作成, ペースト.

scipy.stats での確率変数の扱い

loc(ation)=分布の位置, scale=分布の幅

```
from scipy import stat
# 確率密度関数
stats.確率分布名(loc=位置, scale=幅).pdf(x)
# 累積分布関数
stats.確率分布名(loc=位置, scale=幅).cdf(x)
# サンプル
stats.確率分布名(loc=位置, scale=幅).rvs(size=サンプルサイズ
確率分布にしたがうランダムな数を サンプルサイズ個 取得する.
# 母平均値
stats.確率分布名(loc=位置, scale=幅).mean()
# 母分散
stats.確率分布名(loc=位置, scale=幅).var()
# 母標準偏差
stats.確率分布名(loc=位置, scale=幅).std()
# 母期待値
stats.確率分布名(loc=位置, scale=幅).expect(関数)
```

確率分布名=uniform, expon, norm, $\cdots$.

scipy.stats での分布の名前

https://docs.scipy.org/doc/scipy/reference/stats.html

分布	一様分布	指数分布	正規分布
X	$\mathrm{U}(c,d)$	$\mathrm{Exp}(b,a)$	$\mathrm{N}(b,a^2)$
scipy.stats.	uniform()	expon()	norm()
上の引数	loc=c,scale=$d-c$	loc=$b-a$,scale=a	loc=b,scale=a
母平均値	$\frac{c+d}{2}$	b	b
母分散	$\frac{(d-c)^2}{12}$	a^2	a^2

標準 Z	$\mathrm{U}(-\sqrt{3},\sqrt{3})$	$\mathrm{Exp}(1,1)$	$\mathrm{N}(0,1^2)$
母平均値	0	0	0
母分散	1^2	1^2	1^2

$Z=\frac{X-\mu}{\sigma}$, $X=\sigma Z+\mu$.

$\mathrm{Exp}(b,a)$ はこの授業ローカル記号.

ここまで来たよ

4 連続型一様分布, 変数変換, 確率変数の標準化

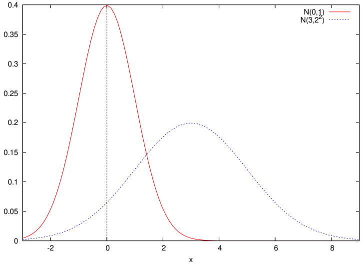
4 正規分布
- 確率変数の標準化
- Python での確率変数の計算
- 標準正規分布
- 一般の正規分布

一般の正規分布 normal distribution 高校 数学 B 岩薩林 確率・統計 §4.5(4.23)

(一般の=標準と限らない) 正規分布 $\mathrm{N}(b, a^2)$ の確率密度関数

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

b, a^2: パラメタ ($a > 0$).



`scipy.stats.norm(loc=`b`,scale=`a`)`

Excel で `=norm.dist()`

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

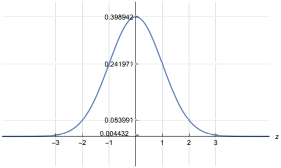
難しいので, (実は標準化された) $b=0, a=1$ の場合を考える.

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

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

`scipy.stats.norm(loc=0,scale=1)`

Excel で`=norm.s.dist(z,FALSE)`



標準正規分布の母期待値

$Z \sim \mathrm{N}(0, 1^2)$

k 次のモーメント (k: 自然数)

$$\mathrm{E}[Z^{2k-1}] = 0, \quad \text{奇関数}$$

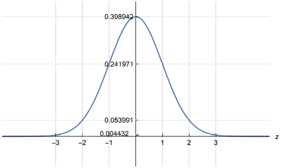
$$\mathrm{E}[Z^{2k}] = (2k-1)!!$$

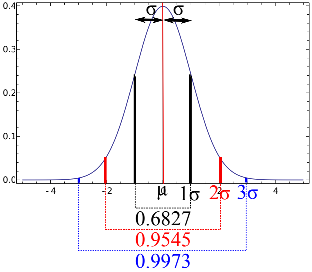
$$= 1 \cdot 3 \cdot 5 \cdots \cdot (2k-1) \quad \text{置換積分}$$

全確率$\mathrm{E}[Z^0] = 1$,　岩薩林 確率・統計 (4.19)　微積分 II

母平均値$\mathrm{E}[Z] = 0$,　岩薩林 確率・統計 (4.20)　奇関数

母分散$\mathrm{V}[Z] = 1$　岩薩林 確率・統計 (4.21)





つまり, *Z は標準化済み*

標準では $\mu = 0, \sigma = 1$

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.
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.

L04-Q2

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 $P(-0.56 < Z < +1.23)$ を Python または表で求めよう.
2. 確率 $P(Z < 1.23)$ を Python または表で小数として求めよう.

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

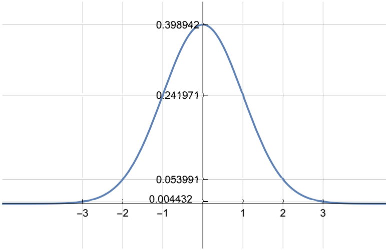
4 連続型一様分布, 変数変換, 確率変数の標準化

4 正規分布
- 確率変数の標準化
- Python での確率変数の計算
- 標準正規分布
- 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.

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



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

$X \sim \mathrm{N}(b, a^2)$. 標準化すると $Z \sim \mathrm{N}(0, 1^2)$.

(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,

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

L04-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.

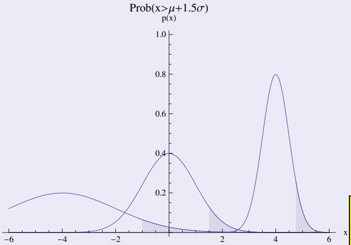
Python でも描いてみよう.

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

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.

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



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,

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

L04-Q4

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. 確率密度関数の式を書こう. そのグラフを描こう.
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.

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