中心極限定理と正規近似

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 L09(2021-06-09 Wed)

最終更新: Time-stamp: "2021-06-04 Fri 16:45 JST hig"

今日の目標

- 指数分布, カイ二乗分布の確率密度関数と母平均値と母分散を説明できる 岩薩林 確率・統計 p.80, 岩薩林 確率・統計 p.123
- 中心極限定理 岩薩林 確率・統計 定理 4.2 の内容を説明でき, 独立同分布の和を正規近似できる

L08-Q1

Quiz 解答:標準正規分布の確率

確率密度関数が偶関数であることに注意する.

1. $P(-\infty < Z < 0) = \int_{-\infty}^{0} f(z; 0, 1^2)\,\mathrm{d}z = -I(-\infty) - I(0) = I(\infty) - 0 = \frac{1}{2}$.
   $\frac{1}{2} = 0.5$.
2. $P(-0.56 < Z < +1.23) = \int_{-0.56}^{1.23} f(z; 0, 1^2)\,\mathrm{d}z = -I(-0.56) + I(1.23) = I(0.56) + I(1.23)$.
   表より, $0.2123 + 0.3907 = 0.6030$.

L08-Q2

Quiz 解答:正規分布の確率

1. $\mathrm{E}[X] = 4$.
2. $\mathrm{V}[X] = 3^2$.
3. $x = 4$ を真ん中に幅 3 くらいの正規分布の確率密度関数のグラフ.

L08-Q3

Quiz 解答:正規分布の確率

$Z = \frac{X-3}{2}$ とすると, Z は標準正規分布 $\mathrm{N}(0, 1^2)$ にしたがう.

1. $\mathrm{E}[X^2] = \mathrm{V}[X] + \mathrm{E}[X]^2 = 2^2 + 3^2$.
2. $P(X \geq 5) = P(\frac{5-3}{2} \leq Z < +\infty) = \int_1^\infty f(z)\,\mathrm{d}z = I(+\infty) - I(1) = \frac{1}{2} - I(1) = 0.1587$.
3. $P(1 \leq X \leq 7) = P(-1 \leq Z \leq 2) = \int_{-1}^2 f(z)\,\mathrm{d}z = I(2) - I(-1) = I(2) + I(1) = 0.8186$.
4. $P(3 \leq X \leq 7) = P(0 \leq Z \leq 2) = \int_0^2 f(z)\,\mathrm{d}z = I(2) - I(0) = I(2) - 0$.

ここまで来たよ

8 正規分布

9 中心極限定理と正規近似
- 指数分布と正規分布と応用
- カイ二乗分布
- 大数の法則 (復習)
- 中心極限定理
- 正規近似

指数分布

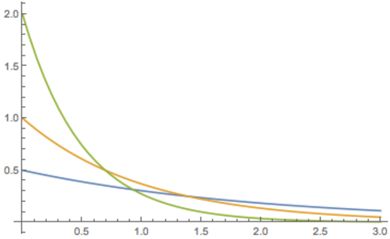
指数分布 岩薩林 確率・統計 例題 4.2(p.80)

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.

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

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

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

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

$1/\lambda$ 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 $^{-1}$.

L09-Q1

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.

正規分布の例題

L09-Q2

Quiz(正規分布の応用)

ある試験を受験したときの点数は連続型確率変数 X で，母平均値 50，母分散 10^2 の正規分布にしたがうという．点数が 60 から 65 である確率を求めよう．

ここまで来たよ

標準正規分布にしたがう独立な確率変数 $Z_1, Z_2, \ldots$ で作る確率変数

$Z, Z_i \sim \mathrm{N}(0, 1^2)$ (標準正規分布), i.i.d のとき

$X_1 = 2Z$

$X_2 = Z + 3$

$X_3 = 2Z + 3$

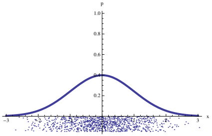
$W_k = Z_1 + Z_2 + \cdots + Z_k$

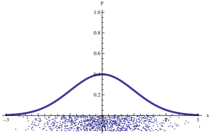
$W_1 = Z^2$
(注:$\mathrm{V}[Z] = \mathrm{E}[Z^2] - 0^2$)

$WW_2 = Z_1^2 + Z_2^2$

$\vdots$

$W_k = Z_1^2 + Z_2^2 + \cdots + Z_k^2$





カイ二乗分布 岩薩林 確率・統計 p.123

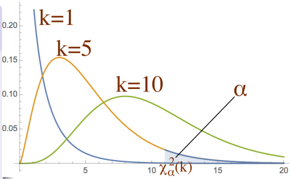
カイ二乗分布

$Z_1, \dots, Z_k \sim \mathrm{N}(0, 1^2)$, iid のとき, 確率変数 $W = Z_1^2 + \cdots + Z_k^2$ は, 自由度 k のカイ二乗分布 χ_k^2 にしたがう.

言語	小	大	読み
英語	x	X	エクス
ギリシャ語	χ	X	カイ

χ_k^2 の確率密度関数

$$f_k(y) = \begin{cases} C_k \times y^{\frac{k}{2}-1} \mathrm{e}^{-\frac{1}{2}y} & (y \geq 0) \\ 0 & (\text{他}) \end{cases}$$



$\mathrm{E}[W_k] = \mathrm{E}[Z_1^2 + \cdots + Z_k^2] = k, \mathrm{V}[W_k] = \text{積分} = 2k, \mathrm{E}[(W_k)^\ell] =$ 簡単じゃない.

ここまで来たよ

8. 正規分布

9. 中心極限定理と正規近似
   - 指数分布と正規分布と応用
   - カイ二乗分布
   - 大数の法則 (復習)
   - 中心極限定理
   - 正規近似

独立同分布にしたがう確率変数の和の性質

岩薩林 確率・統計 例題 4.6

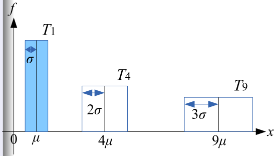
i.i.d にしたがう確率変数の和

$X_1, \ldots, X_n$: i.i.d. 母平均値 $\mathrm{E}[X_i] = \mu$, 母分散 $\mathrm{V}[X_i] = \sigma^2$.
和の確率変数 $T_n = X_1 + \cdots + X_n$.

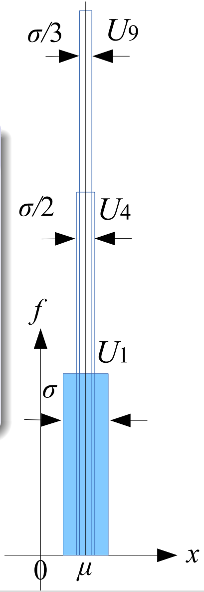
$$\mathrm{E}[T_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] = n \times \mu.$$

$$\mathrm{V}[T_n] = \sum_{i=1}^{n} \mathrm{V}[X_i] = n \times \sigma^2$$

T_n の確率関数はこんな感じ?



U_n の確率関数は
こんな感じ?



i.i.d にしたがう確率変数の和の $1/n$

$X_1, \ldots, X_n$: i.i.d.
新しい確率変数: $U_n = \frac{1}{n}T_n = \frac{1}{n}(X_1 + \cdots + X_n)$

$$\mathrm{E}[U_n] = \mathrm{E}\left[\tfrac{1}{n}T_n\right] = \frac{1}{n} \times n \times \mu.$$

$$\mathrm{V}[U_n] = \mathrm{V}\left[\tfrac{1}{n}T_n\right] = \left(\frac{1}{n}\right)^2 \times n \times \sigma^2.$$

大数の (弱) 法則アバウト版 岩薩林 確率・統計 §4.4

$X_1, \ldots, X_n$ が独立同分布にしたがい, $\mathrm{E}[X_i] = \mu$, $\mathrm{V}[X_i] = \sigma^2$, $U_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ のとき, n が十分大きいとき U_n は ‘ほぼ μ に等しい’ (U_n が μ から外れる確率はゼロに近づく)

大数の (弱) 法則 岩薩林 確率・統計 §4.4

$X_1, \ldots, X_n$ が独立同分布にしたがい, $\mathrm{E}[X_i] = \mu$, $\mathrm{V}[X_i] = \sigma^2$,$U_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ のとき,

$$\forall\epsilon > 0 \ \lim_{n\to+\infty} P(|U_n - \mu| \geq \epsilon) = 0.$$

つまり n 大で U_n は $\mathrm{E}[U_n] = \mu$ に「必ず近い」(確率収束).

大数の弱法則の証明

$\mathrm{E}[U_n] = \mu, \mathrm{V}[U_n] = \sigma^2/n$ に注意して, U_n に対するチェビシェフの不等式を書くと,

$$P(|U_n - \mu| \geq a \times \tfrac{\sigma}{\sqrt{n}}) \leq \tfrac{1}{a^2}$$

$a = \frac{\epsilon}{\frac{\sigma}{\sqrt{n}}}$ とすると, $n \to +\infty$ で

$$P(|U_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2 n} \to 0$$

これが母平均値・母期待値の直観的意味. 要するに,

何回も宝くじを買って賞金を平均すると, 必ず $\mathrm{E}[\text{賞金}]$ に近い

L09-Q3

Quiz(独立同分布にしたがう変数の和)

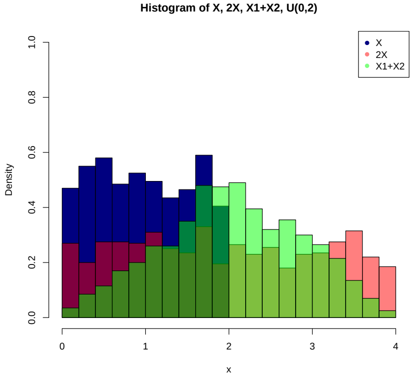
確率変数 $X_1, \ldots, X_n$ は $\mathrm{E}[X_i] = \mu, \mathrm{V}[X_i] = \sigma^2$ の独立同分布に従う.

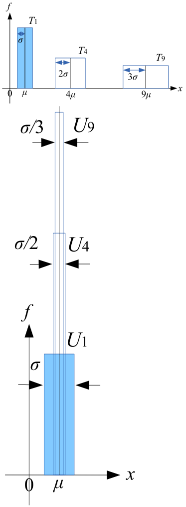
1. 確率変数 $A = X_1 + X_2 + X_3 + \cdots + X_n$ の母平均値と母分散を求めよう.
2. 確率変数 $B = \frac{1}{n}(X_1 + X_2 + X_3 + \cdots + X_n)$ の母平均値と母分散を求めよう.
3. 確率変数 A を標準化して, C を, 母平均値 0, 母分散 1^2 になるように定義しよう.

ここまで来たよ

8. 正規分布
9. 中心極限定理と正規近似
   - 指数分布と正規分布と応用
   - カイ二乗分布
   - 大数の法則 (復習)
   - 中心極限定理
   - 正規近似

一様分布に従う確率変数の定数倍と和





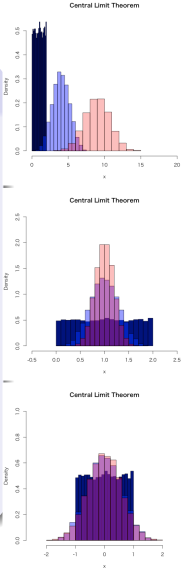
$X_1, X_2 \sim \mathrm{U}(0,2)$, i.i.d.
$2 \times X_1 \sim$
$X_1 + X_2 \sim$

中心極限定理 岩薩林 確率・統計 定理 4.2(p.87)

中心極限定理 (いいかげんバージョン)

$X_1, \ldots, X_n$ が母平均値 μ, 母分散 σ^2 の独立同分布に従うとき, $n \to +\infty$ で

- $T_n = X_1 + \cdots + X_n$, の確率分布は, 正規分布 $\boxed{\mathrm{N}(n\mu, n\sigma^2)}$ に似る
- $U_n = \frac{1}{n}(X_1 + \cdots + X_n)$ の確率分布は, 正規分布 $\boxed{\mathrm{N}(\mu, \sigma^2/n)}$ に似る
- $Z_n = \frac{U_n - \mu}{\sigma/\sqrt{n}}$ の確率分布は, $\boxed{\text{標準正規分布 } \mathrm{N}(0, 1^2)}$ に似る



独立同分布にしたがう確率変数の和

L09-Q4

Quiz(独立同分布と中心極限定理)

独立同分布の個数 n が十分大きいとき, 中心極限定理を利用した, 正規分布での近似を考える.

1. $X_1, \ldots, X_n$ を, パラメタ λ の指数分布にしたがう独立な確率変数とする. $T = X_1 + \cdots + X_n$ と T/n は, 近似的にどのような正規分布にしたがうか.
2. $X_1, \ldots, X_n$ を, ベルヌーイ分布 $\mathrm{B}(1, p)$ にしたがう独立な確率変数とする. $T = X_1 + \cdots + X_n$ と T/n は, 近似的にどのような正規分布にしたがうか.
3. $X_1, \ldots, X_n$ を, 正規分布 $\mathrm{N}(\mu, \sigma^2)$ にしたがう独立な確率変数とする. $T = X_1 + \cdots + X_n$ と T/n は, 近似的にどのような正規分布にしたがうか.
4. $Y_1, \ldots, Y_n$ を, 自由度 $k = 1$ のカイ二乗分布 $\chi(k)$ にしたがう独立な確率変数とする. $T = Y_1 + \cdots + Y_n$ と T/n は, 近似的にどのような正規分布にしたがうか.

ここまで来たよ

8 正規分布

9 中心極限定理と正規近似
- 指数分布と正規分布と応用
- カイ二乗分布
- 大数の法則 (復習)
- 中心極限定理
- 正規近似

独立同分布にしたがう確率変数の和の正規近似 I

L09-Q5

Quiz(独立同分布と中心極限定理)

確率変数 X_i $(i = 1, \ldots, 400)$ は独立同分布にしたがい, $\mathrm{E}[X_i] = \frac{1}{10}$, $\mathrm{V}[X_i] = \frac{9}{100}$ である.
$T = X_1 + \cdots + X_{400}$ とする.
$P(T > 31)$ の確率を, 標準正規分布の $Q(z)$ または $I(z)$ を用いて表し, さらに正規分布表を用いて小数値として近似的に求めよう.

二項分布の正規近似高校 数学 B I

L09-Q6

Quiz(二項分布と正規分布と中心極限定理)

表が $\frac{4}{5}$, 裏が $\frac{1}{5}$ の確率ででる超いびつなコインを, 100 回投げる. 表が 73 回以上 79 回以下でる確率を求めたい.

1. 100 が十分大きいと考えて中心極限定理を利用し, 求める確率を $I(a) = \int_0^a \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z'^2}{2}} \mathrm{d}z'$ の 1 次式で書こう. ただし, $0 < a < +\infty$ であるようにしよう.
2. 求める確率を, 上の状況のもとで, 正規分布表を用いて小数で表そう.

岩薩林 確率・統計 第 5 章練習問題 1(3)