

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 L10(2022-06-13 Mon)

最終更新: Time-stamp: "2022-06-17 Fri 16:12 JST hig"

今日の目標

- チェビシェフの不等式 (母平均値・母分散の正確な意味), 大数の法則が説明できる 岩薩林 確率・統計 §4.3, §4.4
- 中心極限定理が説明できる 岩薩林 確率・統計 §4.4
- 多数の iid の和の分布を正規分布で近似できる



L09-Q1

Quiz 解答: 独立同分布にしたがう確率変数の和

- ① $E[S_{10}] = 800, V[S_{10}] = 2000.$
- ② $E[\frac{1}{10}S_{10}] = 80, V[\frac{1}{10}S_{10}] = 20.$
- ③ $E[\frac{1}{100}S_{100}] = 80, V[\frac{1}{100}S_{100}] = 2.$

L09-Q2

L09-Q3

Quiz 解答: 二項分布

- ① $X \sim B(100, \frac{2}{3})$
- ② $P(X = 40)$ を求めればよいから,
 ${}_{100}C_{40}p^{40}(1-p)^{100-40} = \frac{100!}{40!60!}(\frac{2}{3})^{40}(1-\frac{2}{3})^{60}.$
- ③ $E[X] = n \times p = \frac{200}{3}. V[X] = n \times p(1-p) = \frac{200}{9}.$

L09-Q4 L09-Q5

Quiz 解答: 二項分布または独立同分布

- ① $X \sim B(10, 0.3)$ とすると, $Y = 2X$.
- ② $E[Y] = E[2X] = 2 \times 10 \cdot 0.7 = 14$ 個.
- ③ $V[Y] = 2^2 V[X] = 2^2 \times 10 \cdot 0.7 \cdot 0.3 = 8.4$ 個².
- ④ $P(Y = 16) = P(X = 8) = \frac{10!}{2!8!} 0.7^8 0.3^2$

別解. 10 パックそれぞれに入っている卵の個数を確率変数 Y_i とすると,
 $Y = \sum_{i=1}^{10} Y_i$ で Y_i は独立同分布にしたがうので,

$$E[Y] = 10E[Y_i], V[Y] = 10V[Y_i].$$

$Y_i = 2X_i$ とおくと, X_i はパラメタ $p = 0.7$ のベルヌーイ分布にしたがう. すなわち, $X_i \sim B(1, 0.7)$.

$$E[X_i] = 0.7, V[X_i] = 0.7 \cdot (1 - 0.3) \text{ なので, } E[Y_i] = 2 \cdot 0.7,$$

$$V[Y_i] = 2^2 \cdot 0.7 \cdot (1 - 0.3) \text{ である.}$$

$$\text{よって, } E[Y] = 10E[Y_i] = 14, V[Y] = 10V[Y_i] = 8.4.$$

ここまで来たよ

10 二項分布, 独立同分布の和

10 中心極限定理と正規近似

- チェビシェフの不等式と大数の法則
- 連続型確率変数と正規分布の復習
- 中心極限定理
- 正規近似

チェビシェフの不等式

チェビシェフの不等式 Chebyshev's inequality

岩薩林 確率・統計 §4.3(4.15)

X を離散型または連続型確率変数とする. $\mu = E[X]$: 母平均値,

$\sigma^2 = V[X]$: 母分散

$a > 0$: 任意の正の実数.

このとき次が成立する.

$$P(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2}$$

どんな X にも使えて便利な不等式. これが母平均値・母標準偏差のひとつの意味づけ.

$a > 1$ と思う. 母平均値 μ からの距離が母標準偏差 $\times a$ 以上離れた値が出る確率は, $1/a^2$ 以下. 母平均値から離れるほど, 確率は小

チェビシェフの不等式の証明 (離散型)

意味のわからない $E[I_{[|X-\mu|\geq a\sigma]}(X)(X-\mu)^2]$ を定義に戻って書くと,

$$\begin{aligned}(a\sigma)^2 \times P(|X - \mu| \geq a\sigma) &= \sum_x I_{[|X-\mu|\geq a\sigma]}(x)(a\sigma)^2 p(x) \\ &\leq \sum_x I_{[|X-\mu|\geq a\sigma]}(x)(x - \mu)^2 p(x) \\ &\leq \sum_x (x - \mu)^2 p(x) \\ &= V[X] = \sigma^2\end{aligned}$$

連続型での証明

岩薩林 確率・統計 §4.3

大数の (弱) 法則

大数の (弱) 法則アバウト版 岩薩林 確率・統計 §4.4

X_1, \dots, X_n が独立同分布にしたがい, $E[X_i] = \mu$, $V[X_i] = \sigma^2$,

$U_n = \frac{1}{n} \sum_{i=1}^n X_i$ 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, \dots, X_n が独立同分布にしたがい, $E[X_i] = \mu$,

$V[X_i] = \sigma^2, U_n = \frac{1}{n} \sum_{i=1}^n X_i$ のとき,

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

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

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

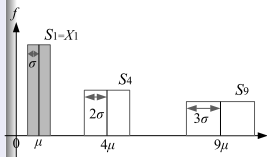
X_1, \dots, X_n : i.i.d. 母平均値 $E[X_i] = \mu$, 母分散 $V[X_i] = \sigma^2$.

和の確率変数 $S_n = X_1 + \dots + X_n$.

$$E[S_n] = \sum_{i=1}^n E[X_i] \stackrel{\text{同分布}}{=} n \times \mu.$$

$$V[S_n] \stackrel{\text{独立}}{=} \sum_{i=1}^n V[X_i] \stackrel{\text{同分布}}{=} n \times \sigma^2$$

S_n の確率密度関数はこんな感じ?



U_n の確率密度関数はこんな感じ?

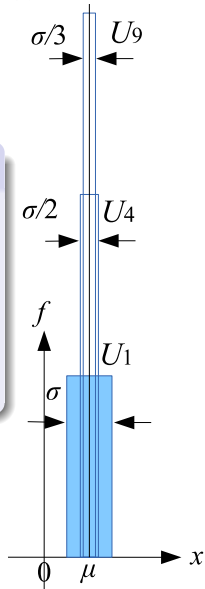
i.i.d にしたがる確率変数の和の $1/n$

X_1, \dots, X_n : i.i.d.

新しい確率変数: $U_n = \frac{1}{n} S_n = \frac{1}{n} (X_1 + \dots + X_n)$

$$E[U_n] = E\left[\frac{1}{n} S_n\right] \stackrel{\text{同分布}}{=} \frac{1}{n} \times n \times \mu.$$

$$V[U_n] = V\left[\frac{1}{n} S_n\right] \stackrel{\text{iid}}{=} \left(\frac{1}{n}\right)^2 \times n \times \sigma^2.$$



L10-Q1

Quiz(独立同分布にしたがう変数の和)

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

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

大数の弱法則の証明

$E[U_n] = \mu, V[U_n] = \sigma^2/n$ 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,

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

不等式の右辺を ϵ にしたい下心から $a = \frac{\epsilon}{\frac{\sigma}{\sqrt{n}}}$ とすると, $n \rightarrow +\infty$ で

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

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 E

ここまで来たよ

10 二項分布, 独立同分布の和

10 中心極限定理と正規近似

- チェビシエフの不等式と大数の法則
- 連続型確率変数と正規分布の復習
- 中心極限定理
- 正規近似

連続型確率変数

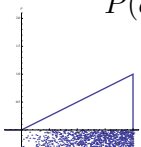
連続型確率変数

連続型確率変数 X とは、実数値をとり、確率が確率密度関数 $f(x)$ (= pdf(x)) で指定されるもの。

$$\text{確率 } P(a < X \leq b) = \int_a^b \mathbf{I}_{[a < X \leq b]}(x) f(x) dx = \int_a^b 1 \cdot f(x) dx$$

$$\text{確率 } P(-\infty < X \leq b) = \int_{-\infty}^b f(x) dx =: F(b) \text{ 累積分布関数 (=cdf}(b))$$

$$P(a < X \leq b) = F(b) - F(a)$$



累積分布関数 $F(x)$ さえ積分で計算しておけば、'すべての' 確率がそこから求められる。

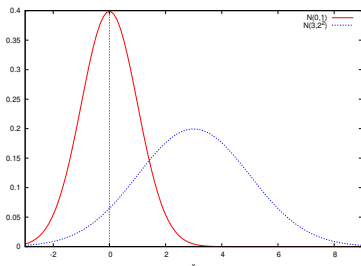
正規分布

(一般の) 正規分布 $N(\mu, \sigma^2)$

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

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

`f(x) = scipy.stats.norm(loc= μ , scale= σ).pdf(x)`



正規分布の $f(x)$ は加減乗除平方根指数関数で書けるが, 原始関数 $F(x)$ は簡単な形に書けない (ので表や Python で値を求める).

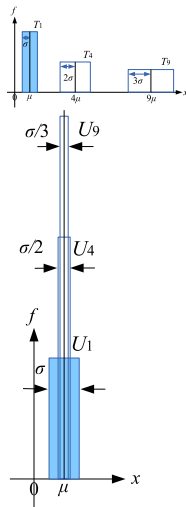
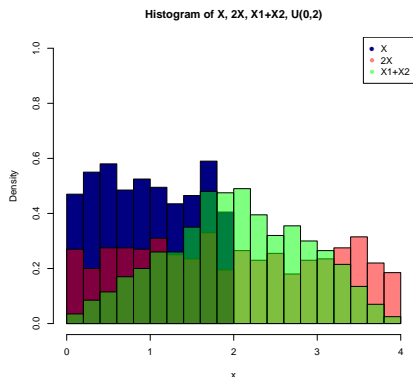
ここまで来たよ

10 二項分布, 独立同分布の和

10 中心極限定理と正規近似

- チェビシェフの不等式と大数の法則
- 連続型確率変数と正規分布の復習
- 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



$X_1, X_2 \sim U(0, 2)$, i.i.d.

$2 \times X_1 \sim$

$X_1 + X_2 \sim$

https://colab.research.google.com/drive/1GsHKZsvfo5_w1t1QuXjgeY5BgEb3FfR8

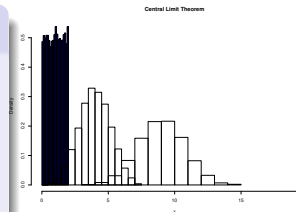
中心極限定理 岩薩林 確率・統計 定理 4.2(p.87)

中心極限定理 (いいかげんバージョン)

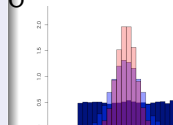
X_1, \dots, X_n が母平均値 μ , 母分散 σ^2 の独立同分布に従うとき, $n \rightarrow +\infty$ で

- $S_n = X_1 + \dots + X_n$ の確率分布は,
 正規分布 $N(n\mu, n\sigma^2)$ に似る
- $U_n = \frac{1}{n}(X_1 + \dots + X_n)$ の確率分布は,
 正規分布 $N(\mu, \sigma^2/n)$ に似る
- 標準化した $Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ の確率分布は,
 標準正規分布 $N(0, 1^2)$ に似る

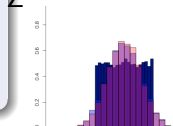
S



U



Z



ここまで来たよ

10 二項分布, 独立同分布の和

10 中心極限定理と正規近似

- チェビシェフの不等式と大数の法則
- 連続型確率変数と正規分布の復習
- 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

L10-Q2

Quiz(独立同分布と中心極限定理)

確率変数 X_i ($i = 1, \dots, 400$) は独立同分布にしたがいで、 $E[X_i] = \frac{1}{10}$, $V[X_i] = \frac{9}{100}$ である.

$S = X_1 + \dots + X_{400}$ とする.

S の分布を正規分布で近似し、 $P(S > 31)$ の確率を、標準正規分布の累積分布関数 $F_Z(z)$ (または $Q(z)$ または $I(z)$) を用いて表し、さらに `scipy.stats.norm.cdf()` (または正規分布表) を用いて小数値として近似的に求めよう.

独立同分布にしたがう確率変数の和の正規近似 I

L10-Q3

Quiz(独立同分布と中心極限定理)

確率変数 X_i ($i = 1, \dots, 400$) は独立同分布にしたがいで、 $X_i \sim U(3, 5)$ である。 $S = X_1 + \dots + X_{400}$ とする。

S の分布を正規分布で近似し、 $P(S \leq 1500)$ の確率を、標準正規分布の累積分布関数 $F_Z(z)$ (または $Q(z)$ または $I(z)$) を用いて表し、さらに `scipy.stats.norm.cdf()` (または正規分布表) 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

L10-Q4

Quiz(二項分布と正規分布と中心極限定理)

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

標準正規分布の累積分布関数 $F_Z(z)$ (または $Q(z)$ または $I(z)$) を用いて表し, さらに `scipy.stats.norm.cdf()` (または正規分布表) を用いて小数値として近似的に求めよう.

岩薩林 確率・統計 第 5 章練習問題 1(3)

二項分布の正規近似 高校 数学 B

二項分布 $B(n, p)$ は, 正規分布 $N(np, p(1 - p))$ で近似できる.