中心極限定理・独立同分布の和の正規近似

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(2024-06-17 Mon)

最終更新: Time-stamp: "2024-06-17 Mon 12:16 JST hig"

今日の目標

- 岩薩林 確率・統計 §4.3 チェビシェフの不等式 (母平均値・母分散の正確な意味) が説明できる
- 岩薩林 確率・統計 §4.4 大数の法則・中心極限定理が説明できる
- 独立同分布の和の確率を正規近似できる

L08-Q1
Quiz 解答: 同分布の確率母期待値母分散

1.

$x_2 \backslash x_1$	0	1
0	$1-p$	0
1	0	p

$x_1 + x_2$	確率
0	$1-p$
1	0
2	p

1. $P(X_1 = X_2 = 1) = p$.
2. $\mathrm{E}[X_1 + X_2] = \mathrm{E}[X_1] + \mathrm{E}[X_2] = 2p$.
3. $\mathrm{V}[X_1 + X_2] = \mathrm{E}[(X_1 + X_2)^2] - \mathrm{E}[X_1 + X_2]^2 = 2^2 p - (2p)^2 = 4p(1-p)$.
4. $\mathrm{Cov}[X_1, X_2] = \mathrm{E}[X_1 X_2] - \mathrm{E}[X_1]\mathrm{E}[X_2] = p - p^2 = p(1-p) = \mathrm{V}[X_1]$.

2.

$x_2 \backslash x_1$	0	1
0	$(1-p)^2$	$p(1-p)$
1	$(1-p)p$	p^2

$x_1 + x_2$	確率
0	$(1-p)^2$
1	$2p(1-p)$
2	p^2

1. $P(X_1 = X_2 = 1) = p_{X_1} x_1 \times p_{X_2} x_2 = p^2$.
2. $\mathrm{E}[X_1 + X_2] = \mathrm{E}[X_1] + \mathrm{E}[X_2] = 2p$.

3. $\mathrm{V}[X_1 + X_2] = \mathrm{V}[X_1] + \mathrm{V}[X_2] = 2p(1-p)$.
4. $\mathrm{Cov}[X_1, X_2] = \mathrm{E}[X_1 X_2] - \mathrm{E}[X_1]\mathrm{E}[X_2] = p^2 - p^2 = 0$.

L08-Q2

Quiz 解答: 離散的な確率変数の母平均値・母分散・母標準偏差・確率

1. $$\mathrm{E}[\mathrm{I}_{[X\le 5]}(X)] = \sum_{x=0}^{10} \frac{x}{55} \mathrm{I}_{[X\le 5]}(x) = \sum_{x=1}^{5} \frac{x}{55} = \frac{15}{55} = \frac{3}{11}.$$
2. $$\mathrm{E}[X] = \sum_{x=0}^{10} \frac{x}{55} \cdot x = \frac{\frac{1}{6} \cdot 10 \cdot (10+1)(2\cdot 10+1)}{55} = \frac{385}{55} = 7.$$
3. $$\mathrm{V}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \sum_{x=0}^{10} \frac{x}{55} \cdot x^2 - 7^2 = 55 - 7^2 = 6.$$

L08-Q3

Quiz 解答: 二項分布

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

L08-Q4

Quiz 解答: 二項分布または独立同分布

1. $X \sim \mathrm{B}(10, 0.3)$ とすると, $Y = 2X$.
2. $\mathrm{E}[Y] = \mathrm{E}[2X] = 2 \times 10 \cdot 0.7 = 14$ 個.
3. $\mathrm{V}[Y] = 2^2\mathrm{V}[X] = 2^2 \times 10 \cdot 0.7 \cdot 0.3 = 8.4$ 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,
$\mathrm{E}[Y] = 10\mathrm{E}[Y_i], \mathrm{V}[Y] = 10\mathrm{V}[Y_i]$.
$Y_i = 2X_i$ 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, $X_i \sim \mathrm{B}(1, 0.7)$.
$\mathrm{E}[X_i] = 0.7$, $\mathrm{V}[Y_i] = 0.7 \cdot (1 - 0.3)$ なので, $\mathrm{E}[Y_i] = 2 \cdot 0.7$,
$\mathrm{V}[Y_i] = 2^2 \cdot 0.7 \cdot (1 - 0.3)$ である.
よって, $\mathrm{E}[Y] = 10\mathrm{E}[Y_i] = 14, \mathrm{V}[Y] = 10\mathrm{V}[Y_i] = 8.4$.

ここまで来たよ

チェビシェフの不等式 Chebyshev's inequality 岩薩林 確率・統計 §4.3

定理 (チェビシェフの不等式 岩薩林 確率・統計 (4.15))

X を離散型または連続型確率変数とする.
母平均値を $\mu = \mathrm{E}[X]$, 母分散 $\sigma^2 = \mathrm{V}[X]$, $a > 0$: 任意の正の実数とする.
このとき次が成立する.

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

中心とか幅とか言ってた母平均値・母標準偏差の厳密な意味づけ.

$a > 1$ 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, 確率は小さい. 母標準偏差 σ は離れ方の基準 (分布の幅).

チェビシェフの不等式の証明 (離散型)

岩薩林 確率・統計 §4.3(連続型での証明)

$$\text{「離れた}\ x\text{」特徴関数}\ g(x) = \begin{cases} 1 & (|x-\mu| \geq a\sigma) \\ 0 & (|x-\mu| < a\sigma) \end{cases}$$

とおく．意味を考えず $\mathrm{E}[g(X)\cdot(X-\mu)^2]$ を定義に戻って書くと,

$$(a\sigma)^2 \times P(|X-\mu| \geq a\sigma) = (a\sigma)^2 \sum_{x<\mu-a\sigma \ \text{or}\ \mu+a\sigma<x} p(x)$$

$$= \sum_{x=-\infty}^{+\infty} g(x)\cdot(a\sigma)^2\, p(x)$$

$$\leq \sum_{x=-\infty}^{+\infty} g(x)\cdot(x-\mu)^2\, p(x)$$

$$\leq \sum_{x=-\infty}^{+\infty} 1\cdot(x-\mu)^2\, p(x) = \mathrm{V}[X] = \sigma^2$$

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

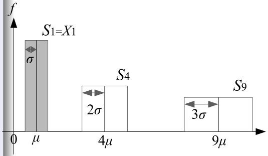
i.i.d にしたがう確率変数の和

$X_1, \ldots, X_n$: 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 $S_n = X_1 + \cdots + X_n$.

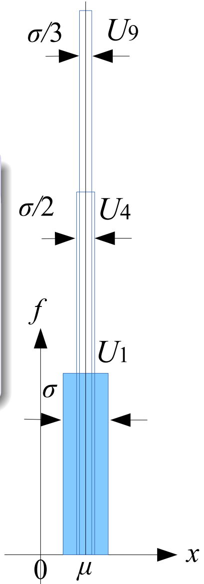
$$\mathrm{E}[S_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] \overset{\text{同分布}}{=} n \times \mu.$$

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

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: $U_n = \frac{1}{n} S_n = \frac{1}{n}(X_1 + \cdots + X_n)$

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

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, $\mathrm{E}[X_i] = \mu$, $\mathrm{V}[X_i] = \sigma^2$ とする.
確率変数 $U_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ を考える (n 回の賞金の相加平均) と $\mathrm{E}[U_n] = \mu$, $\mathrm{V}[U_n] = \sigma^2/n$.

定理 (大数の (弱) 法則アバウト版 岩薩林 確率・統計 定理 4.1(p.84))

n が十分大きいとき確率変数 U_n は 'μ に近い値ばかり. ' (U_n が μ から外れる確率はゼロに近づく)

弱法則の正確な表現

定理 (大数の (弱) 法則 岩薩林 確率・統計 定理 4.1(p.84))

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

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

U_n に対するチェビシェフの不等式を書くと,

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

P の中の不等式の右辺を ϵ にしたい下心から $a = \frac{\epsilon}{\frac{\sigma}{\sqrt{n}}}$ とすると,

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

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の相加平均をとると, 必ず E[賞金] に近い

確率にまつわるいろいろな「収束」

大数の強法則
U_n の「近づく」(!) 先が μ でない確率はゼロである

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

確率にまつわる「収束」にはいろいろある…

- 概収束
- 確率収束
- 平均収束
- 法則収束
- ⋮

ここまで来たよ

8 独立同分布の和・二項分布

9 中心極限定理・独立同分布の和の正規近似
- チェビシェフの不等式・大数の法則
- 中心極限定理
- 正規近似

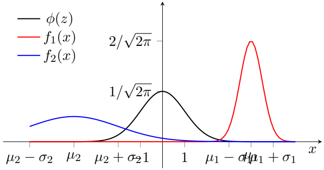
復習: 正規分布 岩薩林 確率・統計 §4.5

定理 ((一般の) 正規分布 $\mathrm{N}(\mu, \sigma^2)$ 岩薩林 確率・統計 (4.23))

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,

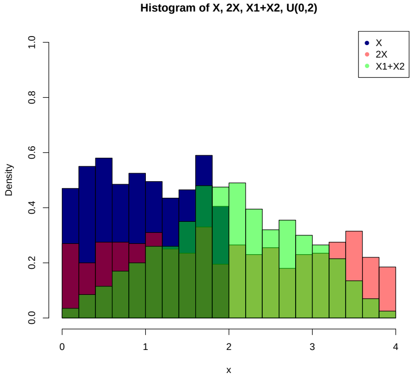
$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

$f(x) =$scipy.stats.norm(loc$=\mu$,scale$=\sigma$).pdf(x)

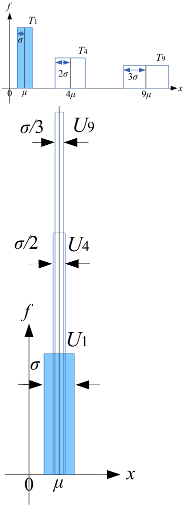


対称軸は $x = \mu, f(\mu) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}$,
$\sqrt{2}\sigma$ 離れると $1/e$ 倍.
$Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0, 1^2)$ 標準正規分布

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.
$2 \times X_1 \sim$
$X_1 + X_2 \sim$



和の確率密度関数は，実際はどんな形なんだろう？

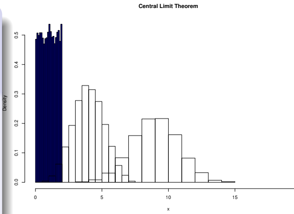
中心極限定理 岩薩林 確率・統計 §4.4

定理 (中心極限定理 (いいかげんバージョン) 岩薩林 確率・統計 定理 4.2(p.87))

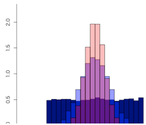
$X_1, \ldots, X_n$: 独立同分布にしたがう．$\mathrm{E}[X_i] = \mu$, $\mathrm{V}[X_i] = \sigma^2$. $n \to +\infty$ で,

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

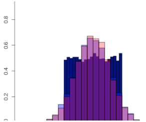
S



U



Z



ここまで来たよ

8 独立同分布の和・二項分布

9 中心極限定理・独立同分布の和の正規近似
- チェビシェフの不等式・大数の法則
- 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-Q1

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.
$S = X_1 + \cdots + X_{400}$ 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 $\Phi(z)$ 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

L09-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，$X_i \sim \mathrm{U}(3, 5)$ である．$S = X_1 + \cdots + X_{400}$ 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 $\Phi(z)$ 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

L09-Q3

Quiz(ベルヌーイ分布の独立同分布の和と中心極限定理)

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 回より多く 79 回以下でる確率を求めたい.
標準正規分布の累積分布関数 $\Phi(z)$ 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 $\mathrm{B}(n,p)$ は, 正規分布 $\mathrm{N}(np,np(1-p))$ で近似できる.