中心極限定理・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 L08(2025-06-09 Mon)

最終更新: Time-stamp: "2025-06-20 Fri 16:37 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) 久保川 統計学入門 §7.3, 大数の法則 久保川 統計学入門 §7.3, 中心極限定理 久保川 統計学入門 §7.4 が説明できる
- 独立同一分布の和の確率を正規近似できる

L07-Q1 略

L07-Q2

Quiz 解答: 二項分布

1. $X \sim Bin(100, \frac{2}{3}))$
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. $\mathrm{E}[X] = n \times p = \frac{200}{3}$. $\mathrm{Var}(X) = n \times p(1-p) = \frac{200}{9}$.

L07-Q3

Quiz 解答: 二項分布または独立同一分布

1. $X \sim Bin(10, 0.3)$ とすると, $Y = 2X$.
2. $\mathrm{E}[Y] = \mathrm{E}[2X] = 2 \times 10 \cdot 0.7 = 14$ 個.
3. $\mathrm{Var}(Y) = 2^2\mathrm{Var}(X) = 2^2 \times 10 \cdot 0.7 \cdot 0.3 = 8.4$ 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 は独立同一分布にしたがうので,
$\mathrm{E}[Y] = 10\mathrm{E}[Y_i], \mathrm{Var}(Y) = 10\mathrm{Var}(Y_i)$.
$Y_i = 2X_i$ 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, $X_i \sim Ber(0.7)$.
$\mathrm{E}[X_i] = 0.7$, $\mathrm{Var}(Y_i) = 0.7 \cdot (1 - 0.3)$ なので, $\mathrm{E}[Y_i] = 2 \cdot 0.7$,
$\mathrm{Var}(Y_i) = 2^2 \cdot 0.7 \cdot (1 - 0.3)$ である.
よって, $\mathrm{E}[Y] = 10\mathrm{E}[Y_i] = 14, \mathrm{Var}(Y) = 10\mathrm{Var}(Y_i) = 8.4$.

L07-Q4

Quiz 解答: 独立同一分布にしたがう確率変数の和

1. $\mathrm{E}[S_{10}] = 800$, $\mathrm{Var}(S_{10}) = 2000$.
2. $\mathrm{E}[\frac{1}{10}S_{10}] = 80$, $\mathrm{Var}(\frac{1}{10}S_{10}) = 20$.
3. $\mathrm{E}[\frac{1}{100}S_{100}] = 80$, $\mathrm{Var}(\frac{1}{100}S_{100}) = 2$.

L08-Q5

Quiz 解答: 独立同一分布にしたがう変数の和

1. S_n は母平均値が $n\mu$, 母分散が $n\sigma^2$.
2. U_n は母平均値が μ, 母分散が $\frac{\sigma^2}{n}$.
3. $Y_n = \frac{U_n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}\sigma}(X_1 + X_2 + X_3 + \cdots + X_n - n\mu)$
   $Y_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$.

ここまで来たよ

7 二項分布・独立同一分布の和

8 中心極限定理・正規近似
- チェビシェフの不等式
- 統計モデル
- 大数の法則・中心極限定理
- 正規近似

チェビシェフの不等式 Chebyshev's inequality 久保川 統計学入門 §7.3

定理 (チェビシェフの不等式応用形は 久保川 統計学入門 公式 7.6(p.158))

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 $\sigma^2 = \mathrm{Var}(X)$, $a > 0$: 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)

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

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，

$$\begin{aligned}
(a\sigma)^2 \times P(|X-\mu| \geq a\sigma) =& (a\sigma)^2 \sum_{|x-\mu| \geq a} p(x) \\
=& \sum_{x=-\infty}^{+\infty} g(x)\cdot(a\sigma)^2\, p(x) \\
\leq& \sum_{x=-\infty}^{+\infty} g(x)\cdot(x-\mu)^2\, p(x) \\
\leq& \sum_{x=-\infty}^{+\infty} 1\cdot(x-\mu)^2\, p(x) = \mathrm{Var}(X) = \sigma^2
\end{aligned}$$

ここまで来たよ

7 二項分布・独立同一分布の和

8 中心極限定理・正規近似
- チェビシェフの不等式
- 統計モデル
- 大数の法則・中心極限定理
- 正規近似

統計モデル 久保川 統計学入門 §7.1

統計モデル 久保川 統計学入門 §7.1

$\boxed{\text{構造}} + \boxed{\text{ゆらぎ＝確率変数}} \underset{\overleftarrow{\text{推定}}}{\xrightarrow{\text{標本抽出}}} \boxed{\text{観測データ}}$

記述統計の母集団と標本，有限母集団 久保川 統計学入門 §7.1

某アイドルグループ全員 (→ 有限母集団) の身長 x_i の平均値
$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ を求めたい!
握手券を 5 枚買って質問する． 標本 $X_1, \dots, X_5$
標本平均値 $\frac{1}{5}(X_1 + \cdots + X_5)$

データ分析

推測統計の母集団と標本，無限母集団 久保川 統計学入門 §7.1

某アイドルグループは，正規分布 $N(?,?)$ の身長をもって無限に生成される (→ 無限母集団)

```
rvx=stats.norm(loc=?,scale=?) #N(?,?^2) 実は身長は正規分布
rvx.rvs(size=5)
```

「確率変数 $X \sim N(?,?)$ にしたがう試行を 5 回して，5 個の数値を得る」
$=$「5 個の確率変数 $X_1, \dots, X_n$, i.i.d. $\sim N(?,?)$ の実現値を得る」
「平均値」$\overline{X} = \frac{1}{5}(X_1 + \cdots + X_5)$ も確率変数.

ここまで来たよ

7 二項分布・独立同一分布の和

8 中心極限定理・正規近似
- チェビシェフの不等式
- 統計モデル
- 大数の法則・中心極限定理
- 正規近似

独立同一分布にしたがう確率変数の和の性質

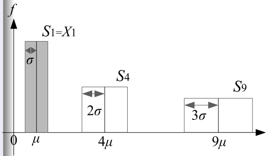
復習 確率統計 I(2025)L07 i.i.d にしたがう確率変数の和

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

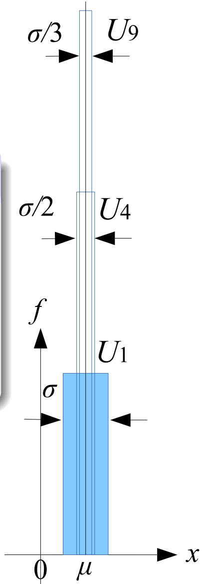
$$\mathrm{E}[S_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] \overset{\text{同一分布}}{=} n \times \mu.$$

$$\mathrm{Var}(S_n) \overset{\text{独立}}{=} \sum_{i=1}^{n} \mathrm{Var}(X_i) \overset{\text{同一分布}}{=} n \times \sigma^2$$

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

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

大数の (弱) 法則

$X_1, \dots, X_n$ が独立同一分布にしたがい (n 回のくじの賞金), $\mathrm{E}[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$ とする.

確率変数 $\overline{X} = U_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ を考える (n 回の賞金の相加平均) と

$\mathrm{E}[U_n] = \mu$, $\mathrm{Var}(U_n) = \sigma^2/n$.

定理 (大数の (弱) 法則アバウト版 久保川 統計学入門 公式 7.8)

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 久保川 統計学入門 公式 7.8)

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

「U_n は μ に確率収束する」

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,

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

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」

定理 (大数の強法則)

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

「U_n は μ に概収束 する」(確率収束とは極限の位置が違う)
U_n の「近づく」(!) 先が μ でない確率はゼロである
確率にまつわる「収束」にはいろいろある…

$$\begin{tikzcd}
\text{概収束} \arrow[d] & \text{平均収束} \arrow[dl] \\
\text{確率収束} \arrow[d] & \\
\text{法則収束} &
\end{tikzcd}$$

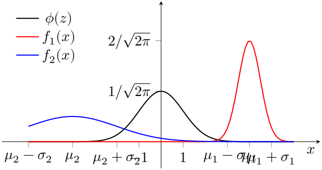
復習: 正規分布 久保川 統計学入門 §5.3.2

定義 ((一般の) 正規分布 $N(\mu, \sigma^2)$ 久保川 統計学入門 §5.3.2)

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,

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

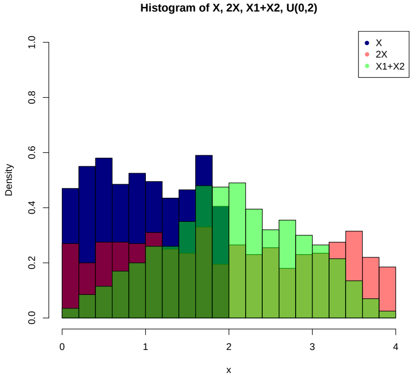
$f(x)$ =scipy.stats.norm(loc=μ,scale=σ).pdf(x)



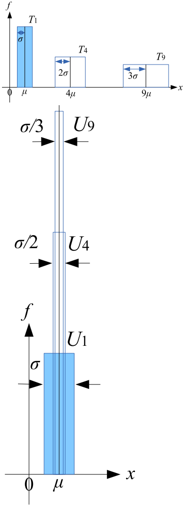
対称軸は $x=\mu$, $f(\mu)=\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}$, σ 離れると 0.6 倍.

$Z=\frac{X-\mu}{\sigma} \sim N(0,1^2)$ 標準正規分布

一様分布に従う確率変数の定数倍と和



$X_1, X_2 \sim U(0,2)$, 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？

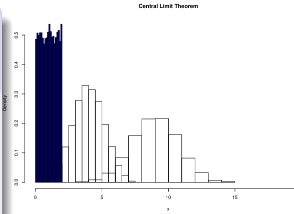
中心極限定理 久保川 統計学入門 §7.4

定理 (中心極限定理 (いいかげんバージョン) 久保川 統計学入門 公式 7.11)

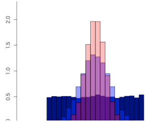
$X_1, \dots, X_n$: 独立同一分布にしたがう． $\mathrm{E}[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$. $n \to +\infty$ で,

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

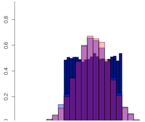
S



U



Z



ここまで来たよ

7 二項分布・独立同一分布の和

8 中心極限定理・正規近似
- チェビシェフの不等式
- 統計モデル
- 大数の法則・中心極限定理
- 正規近似

独立同一分布にしたがう確率変数の和の正規近似 I

正規近似

久保川 統計学入門 p.162, 例題 7.14

L08-Q1

Quiz(独立同一分布と中心極限定理)

確率変数 X_i $(i = 1, \dots, 400)$ は独立同一分布にしたがい, $\mathrm{E}[X_i] = \frac{1}{10}$, $\mathrm{Var}(X_i) = \frac{9}{100}$ 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.

L08-Q2

Quiz(独立同一分布と中心極限定理)

確率変数 X_i $(i = 1, \ldots, 400)$ は独立同一分布にしたがい, $X_i \sim U(3, 5)$ 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

久保川 統計学入門 基本問題問 1(p.166)

L08-Q3

Quiz(ベルヌーイ分布の独立同一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.

二項分布の正規近似高校 数学 B 久保川 統計学入門 公式 7.13

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