中心極限定理・母集団・標本抽出・点推定

樋口さぶろお

龍谷大学理工学部数理情報学科

確率統計☆演習 I L10(2016-12-08 Thu)

最終更新: Time-stamp: "2016-12-08 Thu 06:30 JST hig"

今日の目標

- 中心極限定理の意味が説明でき, 確率の計算に利用できる. 塚田確率統計 §5.3
- 母集団, 標本, 推定を説明できる 塚田確率統計 §6.2
- 標本から母平均値, 母分散 塚田確率統計 §7.2 を点推定できる

http://hig3.net

L09-Q1

Quiz 解答:標準正規分布の確率

標準正規分布の確率密度関数は偶関数 ($z = 0$ に関して対称) なので,

$$P(Z < -2) = \int_{-\infty}^{-2} f(z)\,\mathrm{d}z = \int_{+2}^{+\infty} f(z)\,\mathrm{d}z$$

$$\begin{aligned}\int_{+2}^{+\infty} f(z)\,\mathrm{d}z =& Q(2.00) = 0.0228.\\ \int_{+2}^{+\infty} f(z)\,\mathrm{d}z =& \int_{2}^{0} f(z)\,\mathrm{d}z + \int_{0}^{\infty} f(z)\,\mathrm{d}z\\ =& -I(2.00) + \tfrac{1}{2} = 0.0228.\end{aligned}$$

L09-Q2

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. $\mathrm{E}[Z^2] = \mathrm{V}[Z] + (\mathrm{E}[Z])^2 = 1.$
2. $P(-0.56 < Z < +1.23) = \int_{-0.56}^{1.23} f(z)\,\mathrm{d}z.$
   $\int_{-0.56}^{1.23} f(z)\,\mathrm{d}z = 1 - \int_{-\infty}^{-0.56} f(z)\,\mathrm{d}z - \int_{1.23}^{\infty} f(z)\,\mathrm{d}z =$
   $1 - Q(1.23) - Q(0.56) = 1 - 0.1093 - 0.2877 = 0.6030.$
   $\int_{-0.56}^{1.23} f(z)\,\mathrm{d}z = \int_{0}^{0.56} f(z)\,\mathrm{d}z + \int_{0}^{1.23} f(z)\,\mathrm{d}z = I(0.56) + I(1.23) =$
   $0.6030.$

Quiz 解答:正規分布の確率

定義にしたがって積分しても求まるが, 正規分布の確率密度関数と比較すると, $X \sim \mathrm{N}(4, 3^2)$ なので,

1. $\mathrm{E}[X] = 4.$
2. $\mathrm{V}[X] = 3^2.$

Quiz 解答:正規分布の確率

平方完成すると,

$$f(x) = C \cdot \mathrm{e}^{-\frac{(x-4)^2}{2 \cdot 3^2}} \mathrm{e}^{\frac{16}{18}}$$

と書ける. すなわち, $X \sim \mathrm{N}(4, 3^2)$.

1. $\mathrm{E}[X] = \int_{-\infty}^{+\infty} C \cdot \mathrm{e}^{-\frac{1}{18}x^2 + \frac{4}{9}x}\ x\ \mathrm{d}x$ だが, $X \sim \mathrm{N}(4, 3^2)$ なので, $\mathrm{E}[X] = 4.$
2. $\mathrm{V}[X] = \int_{-\infty}^{+\infty} C \cdot \mathrm{e}^{-\frac{1}{18}x^2 + \frac{4}{9}x}\ x^2\ \mathrm{d}x - \mathrm{E}[X]^2$ だが, $X \sim \mathrm{N}(4, 3^2)$ なので, $\mathrm{V}[X] = 3^2.$
3. $\mathrm{E}[1] = 1$ より, 定数 $C^{-1} = \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{(x-4)^2}{2 \cdot 3^2}} \mathrm{e}^{\frac{16}{18}}\ \mathrm{d}x$ だが, $\mathrm{E}[1] = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi 3^2}} \mathrm{e}^{-\frac{(x-4)^2}{2 \cdot 3^2}}\ \mathrm{d}x = 1$ に注意すると, $C = \frac{1}{\sqrt{2\pi 3^2}} \mathrm{e}^{-\frac{16}{18}}$ とわかる.

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. $P(X\geq 5)=P(Z\geq\frac{5-3}{2})=\int_1^\infty f(z)\,\mathrm{d}z$.
   $\int_1^\infty f(z)\,\mathrm{d}z=Q(1.00)=0.1587$.
   $\int_1^\infty f(z)\,\mathrm{d}z=\int_1^0 f(z)\,\mathrm{d}z+\int_0^\infty f(z)\,\mathrm{d}z=-I(1.00)+\frac{1}{2}$.
2. $Z=\frac{X-3}{2}$ とすると, Z は標準正規分布にしたがう.
   $P(1\leq X\leq 7)=P(-1\leq Z\leq 2)=\int_{-1}^2 f(z)\,\mathrm{d}z$.
   $\int_{-1}^2 f(z)\,\mathrm{d}z=1-Q(2.00)-Q(1.00)=0.8186$.
   $\int_{-1}^2 f(z)\,\mathrm{d}z=\int_0^1 f(z)\,\mathrm{d}z+\int_0^2 f(z)\,\mathrm{d}z=I(1)+I(2)=0.8186$.

L09-Q6

Quiz 解答:正規分布の確率

1. $\int_{0.5}^{0.7} f(z)\,\mathrm{d}z=Q(0.5)-Q(0.7)=I(0.7)-I(0.5)=0.3085-0.2420=0.0665$.
2. $z=\frac{X}{2}$ とすると, Z は標準正規分布に従う.
   $P(0.5\leq X\leq 0.7)=P(0.25\leq Z\leq 0.35)=\int_{0.25}^{0.35} f(z)\,\mathrm{d}z=$
   $Q(0.25)-Q(0.35)=I(0.35)-I(0.25)=0.4013-0.3632=0.0381$.
3. $Z=\frac{X-3}{2}$ とすると, Z は標準正規分布に従う.
   $P(4.0\leq X\leq 4.4)=P(0.5\leq Z\leq 0.7)=$ (1. と同様) $=0.0665$.

ここまで来たよ

1. 大数の法則・正規分布
2. 中心極限定理・母集団・標本抽出・点推定
   - 中心極限定理
   - 母集団と標本
   - 母平均値・母分散の (点) 推定

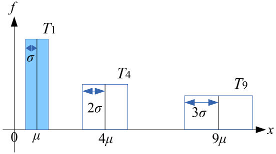
独立同分布の分布の復習 塚田確率統計 5.1

$X_1, \ldots, X_n$ が独立同分布に従うとする. $\mathrm{E}[X_i] = \mu, \mathrm{V}[X_i] = \sigma^2$.

新しい確率変数: $U_n = X_1 + \cdots + X_n$

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

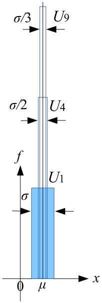
$$\mathrm{V}[U_n] = \sum_{i=1}^{n} \mathrm{V}[X_i] = n \times \sigma^2.$$



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

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

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



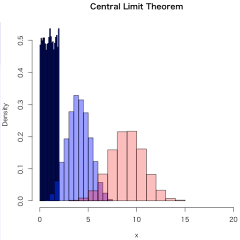
中心極限定理 塚田確率統計 5.3

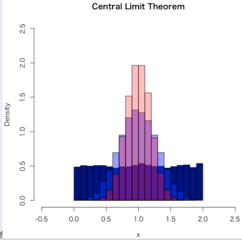
最初が一様分布でも $n \to \infty$ で U_n や W_n の確率密度関数の形は長方形から崩れていく. 分布の個性が消える! っていうか美しい形に!

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,

- $U_n = X_1 + \cdots + X_n$, の確率分布は, の $\boxed{}$ に似る
- $W_n = \frac{1}{n}(X_1 + \cdots + X_n)$ の確率分布は, $\boxed{}$ に似る
- $Z_n = \frac{W_n - \mu}{\sigma\sqrt{n}}$ の確率分布は, $\boxed{}$ に似る





中心極限定理 (厳密バージョン)

確率変数 $X_1, X_2, \ldots, X_n$ が, 母平均値 μ, 母分散 σ^2 の独立同分布に従うとする. 正規分布じゃない. どんな分布でも可

$Z_n = \frac{\frac{1}{n}(X_1+\cdots+X_n)-\mu}{\sigma} \times \sqrt{n}$ とすると,
Z_n は, $n \to +\infty$ の極限で, $\mathrm{N}(0, 1^2)$ に従う. すなわち

$$\lim_{n\to+\infty} P(c \leq Z_n < d) = \int_c^d \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2}\, \mathrm{d}x$$

「Z_n は $\mathrm{N}(0, 1^2)$ にしたがう Z に法則収束する」

法則収束とは, 関数列がある関数に収束すること.

証明

$\mathrm{E}[Z_n] = 0, \mathrm{V}[Z_n] = 1$ はすぐわかるが…
モーメント母関数を使うと瞬殺

確率統計☆演習 II

L10-Q1

Quiz(二項分布と正規分布と中心極限定理)

表が確率 $\frac{2}{3}$, 裏が確率 $\frac{1}{3}$ でてるコインを, 1 回投げるとき, 表がでる回数を確率変数 X とする. 18 回投げるとき, 表がでる回数を確率変数 W とする.

1. W はどのような二項分布にしたがうか. $\mathrm{B}(?,?)$ の形で答えよう.
2. W は近似的にどのような正規分布にしたがうか. $\mathrm{N}(?,?)$ の形で答えよう.
3. $N = 18$ が大きいと考えるとき, 表が 9 回以下でる確率を, 正規分布表を用いて近似的に求めよう.

L10-Q2

Quiz(中心極限定理)

確率変数 $X_1, \ldots, X_{100}$ は $\mathrm{E}[X_i] = 1, \mathrm{V}[X_i] = \frac{1}{4}$ の独立同分布に従う．次の確率変数の母平均値と母分散を求めよう．また, $n = 10$ が十分に大きいと思って, 指定の確率を求めよう．

1. 確率変数 $U = X_1 + X_2 + X_3 + \cdots + X_{100}$. 確率 $P(U > 110)$.
2. 確率変数 $W = \frac{1}{100}(X_1 + X_2 + X_3 + \cdots + X_{100})$. 確率 $P(W < 1.01)$.

L10-Q3 塚田確率統計 §5.4 問 1-4

実験 (あとでいう U_1, U_4, U_9 の標本抽出)

$X_n \sim \mathrm{B}(1, \frac{2}{3})$

$$f(x) = \begin{cases} \frac{2}{3} & (x = 1) \quad \text{サイコロで} \boxed{3}\boxed{4}\boxed{5}\boxed{6} \\ \frac{1}{3} & (x = 0) \quad \text{サイコロで} \boxed{1}\boxed{2} \end{cases}$$

記入欄 $U_n = X_1 + \cdots + X_n$.

n		**1**	2	3	**4**	5	6	7	8	**9**
目	(1–6)									
X_n	(0–1)									
U_n	(0–9)	*			*					*

`https://manaba.ryukoku.ac.jp` に送信.

ここまで来たよ

1. 大数の法則・正規分布
2. 中心極限定理・母集団・標本抽出・点推定
   - 中心極限定理
   - 母集団と標本
   - 母平均値・母分散の (点) 推定

母集団と標本 (1) 有限母集団 塚田確率統計 6.2

AKB48 の身長ふたたび

- AKB48 メンバー全員 (→ 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!
  - メンバー 1 名を等確率で選んでくる, という試行を考えると, 確率変数 X の母平均値 $\mathrm{E}[X]$.
- メンバー全員分のデータがあれば定義の式使うだけ
- 握手会でメンバー 1 人ずつに質問しなければいけないとしたら?
- 握手会参加券 74 枚集めないで何とかすませたい.

$\rightsquigarrow$ 質問できたメンバー 5 人の身長 (=標本) から推定したい.
5 人を '無作為に' 選ぶ (=標本抽出する)
母集団サイズ=$\boxed{}$, 標本サイズ=$\boxed{}$, 標本の個数=$\boxed{}$.

母集団と標本 (2) 離散 or 連続型確率変数 塚田確率統計 6.2

賞金額, 個数が謎のスピードくじ (引いて賞金額を見た後で箱に戻す).
賞金額 X は離散型確率変数 $\to$ 無限母集団 (何回でもひけるから).

- 賞金の母平均値 $\mathrm{E}[X] = \sum_x f(x) \times x$ を求めたい.
- くじの中を見れば ($f(x)$ の式を知れば定義の式使うだけ)
- しかし, 中を見ることはできない.
- $+\infty$ 回くじを買わず, 何とかすませたい.

$\rightsquigarrow$ 引いた 5 枚のくじの賞金額 (=標本) から推定したい.
5 枚を '無作為に' 選ぶ (=標本抽出する).
母集団サイズ= $\boxed{}$, 標本サイズ= $\boxed{}$, 標本の個数= $\boxed{}$.

母集団・標本抽出・推定

- 母集団 population ＝考えたい集団．どんな分布，母平均値，母分散，などわかっていないことがあるが，全体を調べるわけにはいかない集団．
- 標本 sample (名詞) ＝母集団から '無作為に' とってきた一部分
- 標本抽出 する sample(動詞)＝母集団から '無作為に' とってくる ⇝ sampling (動名詞)
- 推定 する estimate(動詞) ＝標本を調べて母集団について正しそうな事実を見つける ⇝ estimation (名詞)

推定には誤差あるかも．標本の選び方ごとに答は違うし．

ここまで来たよ

1. 大数の法則・正規分布
2. 中心極限定理・母集団・標本抽出・点推定
   - 中心極限定理
   - 母集団と標本
   - 母平均値・母分散の (点) 推定

母平均値の (点) 推定 塚田確率統計 §7.2 高校 数学 B

$X_1, X_2, \ldots, X_n$ はサイズ n の標本.
各 X_i $(i = 1, \ldots, n)$ は母平均値 $\mu = \mathrm{E}[X_i]$, 母分散 $\sigma^2 = \mathrm{V}[X_i]$ の独立同分布にしたがう確率変数.
μ, σ^2 は母集団のパラメタ.

標本平均値

$$\text{標本平均値}\ \overline{X}_{(n)} = \frac{1}{n}(X_1 + \cdots + X_n) = \text{先週の}\ W_n$$

が, 母平均値 μ の 'よい' 推定値になっている.

母平均値は μ はひとつに定まっているが, 標本平均値 $\overline{X}_{(n)}$ は確率変数であり, 試行=標本抽出のたびにかわる ($\overline{X}_{(n)}$ は確率分布をもつ)

L10-Q4

Quiz(母平均値, 母分散の点推定)

フライドチキン屋さんのフライドチキンの在庫 (=母集団) から, 無作為に 6 本のチキンを取り出したところ, 重さは次のようだった.

$117\mathrm{g}, \quad 109\mathrm{g}, \quad 109\mathrm{g}, \quad 119\mathrm{g}, \quad 100\mathrm{g}, \quad 112\mathrm{g}.$

1. 重さの母平均値を点推定しよう.
2. 重さの母分散を点推定しよう.

よい推定値って?

塚田確率統計 §7.2

標本平均値 $\overline{X}_{(n)}$ は不偏性を持つ

「標本平均値 $\overline{X}_{(n)}$」の母平均値$=X_i$ の母平均値

先週の $\mathrm{E}[W_n] = \mu$

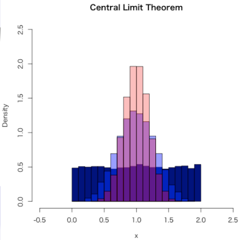
$$\forall n \qquad \mathrm{E}[\overline{X}_{(n)}] = \mu$$

標本平均値 $\overline{X}_{(n)}$ は一致性を持つ

標本サイズ n が大きくなると, $\overline{X}_{(n)}$ と母平均値 μ が離れている確率は 0 に近づく.

大数の (弱) 法則

$$\forall \epsilon > 0 \qquad \lim_{n \to +\infty} P(|\overline{X}_{(n)} - \mu| > \epsilon|) \to 0$$



母分散の (点) 推定 塚田確率統計 §7.2 高校 数学 B

(不偏) 標本分散

$$\text{(不偏) 標本分散 } s^2 = \frac{1}{n-1}[(X_1-\overline{X})^2+\cdots+(X_n-\overline{X})^2]$$
$$= \frac{n}{n-1}\left[\frac{1}{n}\sum_i X_i^2 - \left(\overline{X}\right)^2\right]$$

が, 母分散の ‘よい’ 推定値になっている.
ここで, $\overline{X}$ は母平均値でなく, 上の標本平均値 ($\overline{X}_{(n)}$ の略記).

$n-1$ の理由 こうするとちょうど不偏: $\mathrm{E}[s^2]=\sigma^2$.
直観的理由 $\overline{X}$ は X_i の重心だから, $(X_i-\overline{X})^2$ は $(X_i-\mu)^2$ より小さくなりがち ($\frac{n-1}{n}$ 倍) なので修正.

おぼえ方 (不偏) 標本分散は $n=1$ のとき, $\boxed{}$

$\mathrm{E}[s^2] = \sigma^2$ を $n = 2$ のときに確認

$$
\begin{aligned}
\text{左辺} =& \frac{1}{2-1}\mathrm{E}[(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2] \\
=& \mathrm{E}[X_1^2 + X_2^2 - 2(X_1 + X_2)\overline{X} + 2\overline{X}^2] \\
=& \mathrm{E}[X_1^2 + X_2^2 - 2\overline{X}^2] \\
=& \mathrm{E}[X_1^2] + \mathrm{E}[X_2^2] - 2\mathrm{E}[\overline{X}^2]
\end{aligned}
$$

ここで,

$\sigma^2 = \mathrm{V}[X_1] = \mathrm{E}[X_1^2] - (\mathrm{E}[X_1])^2 = \mathrm{E}[X_1^2] - \mu^2$,

$\frac{\sigma^2}{n} = \mathrm{V}[\overline{X}] = \mathrm{E}[\overline{X}^2] - (\mathrm{E}[\overline{X}])^2 = \mathrm{E}[\overline{X}^2] - \mu^2$, より,

$$
\begin{aligned}
\text{左辺} =& (\mu^2 + \sigma^2) + (\mu^2 + \sigma^2) - 2(\mu^2 + \tfrac{\sigma^2}{2}) \\
=& \sigma^2 \\
=& \text{右辺}
\end{aligned}
$$

母標準偏差の (点) 推定

$$\text{母標準偏差の (点) 推定値} = \sqrt{\text{母分散の (点) 推定値 } s^2}$$

L10-Q5

Quiz(推定)

賞金が, ある確率分布に従う確率変数 X であるスピードくじを 10 回ひいたところ, 賞金は,
0 円, 0 円, 0 円, 0 円, 0 円, 0 円, 10 円, 10 円, 30 円, 100 円
だった. 確率変数 X の母平均値と母分散と母標準偏差を推定しよう.

連絡

- 予習問題は, 次々回の授業直前 を締切 (そこまでの最高点を記録) とします. でも, Trial までにやったほうが効率いいと思う. 前からそうだけど, 予習問題が満点だと, Trial の満点の 1/3 まで保証されます.
- 配布資料は 1-503 向かいの引出, `http://hig3.net` で再配布.
- 加減乗除と平方根 (ルート) の使える電卓持ってきてね. 関数電卓でなくてもいいです. 携帯電話の機能・アプリでもかまいません.
- 樋口オフィスアワー木 6 金昼 (1-502), Math ラウンジ月-木昼 (1-614)
- 次回はカイ二乗分布 塚田確率統計 §4.9 t 分布 塚田確率統計 §4.10 母平均値の区間推定 塚田確率統計 7.3.2

`https://manaba.ryukoku.ac.jp`