中心極限定理, 母平均値母比率の区間推定

樋口さぶろお　http://hig3.net

龍谷大学理工学部数理情報学科

確率統計☆演習 I　L11(2018-12-12 Wed)

最終更新: Time-stamp: "2018-12-12 Wed 11:38 JST hig"

今日の目標

- 中心極限定理の内容を説明できる　前園確率統計 §3.4
- 母平均値, 母比率を区間推定できる　前園確率統計 §5.2,§5.3,§5.4

L10-Q1

Quiz 解答:母平均値, 母分散, 母比率の点推定　在庫の重さを X とすると,

1. 標本平均値 $\overline{X} = \frac{1}{6}(117 + \cdots + 112) = 111\mathrm{g}$ なので, 母平均値は $111\mathrm{g}$ と推定できる.
2. 標本期待値 $\overline{X^2} = \frac{1}{6}(117^2 \cdots + 112^2) = 37078/3\mathrm{g}^2$ なので, 母期待値は $37078/3\mathrm{g}^2$ と推定できる.
3. 不偏標本分散は, $\frac{1}{6-1}[(117-111)^2 + \cdots + (112-111)^2] = 46\mathrm{g}^2$ なので, 母分散は $46\mathrm{g}^2$ と推定できる.
4. 標本比率は, $\frac{1}{6}[1+0+0+1+0+1] = 0.5$ なので, 母比率は 0.5 と推定できる.

ここまで来たよ

独立同分布の性質 (復習)

前園確率統計 §3.2 定理 3.4

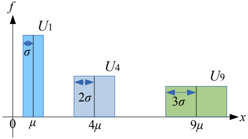
$X_1, \ldots, X_n$: 独立同分布. 母平均値 $\mathrm{E}[X_i] = \mu$, 母分散 $\mathrm{V}[X_i] = \sigma^2$.

和の確率変数 $U_n = X_1 + \cdots + X_n$

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

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

U_n の確率密度関数はこんな感じ?

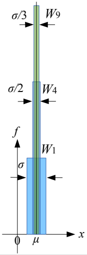


確率変数: $W_n = \frac{1}{n} U_n = \frac{1}{n}(X_1 + \cdots + X_n)$

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

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

W_n の確率密度関数はこんな感じ?

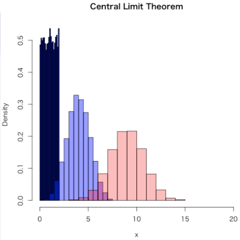


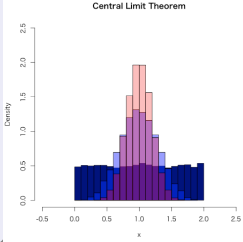
中心極限定理 前園確率統計 §3.4

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

- $U_n = X_1 + \cdots + X_n$, 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) 前園確率統計定理 3.6

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

$Z_n = \dfrac{\frac{1}{n}(X_1+\cdots+X_n)-\mu}{\sigma} \times \sqrt{n}$ 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

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

「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()L

二項分布の正規近似 高校 数学 B I

L11-Q1

Quiz(二項分布と正規分布と中心極限定理)

表が確率 $\frac{1}{10}$, 裏が確率 $\frac{9}{10}$ でるコインを, 400 回投げるとき, 表がでる回数を確率変数 U とする.

1. U 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. U 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. 表が 31 回より多くでる確率を, 標準正規分布の上側確率 $Q(z)$ 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.

ここまで来たよ

10 略解:母集団・標本・標本抽出と推定

11 中心極限定理, 母平均値母比率の区間推定

- 中心極限定理と正規近似
- 母平均値の区間推定 (正規母集団, 母分散未知)
- 母比率 (ベルヌーイ分布の p) の区間推定

点推定 対 区間推定

点推定 前園確率統計 §5.1

真の母平均値はわからないが, 標本平均値を使って,

「母平均値を A 円と推定する」

それどのくらい正確なの? 実は

区間推定 前園確率統計 §5.2

「母平均値が, B 円以上 C 円以下である '確率' は $1-\alpha=0.95$」

ここで '確率' というのは不誠実.

「母平均値の信頼係数 $1-\alpha=0.95$ の信頼区間は B 円以上 C 円以下」

というのが正しい言葉遣い. 以下でその意味と B, C の求め方.

母平均値の区間推定 (正規母集団，母分散既知) 高校 数学 B 前園確率統計 p.80

$\mathrm{N}(\mu, \sigma^2)$ にしたがう母集団 (正規母集団) の，サイズ n の標本を何回も取り出して，毎回，標本平均値 $\overline{X}_{(n)}$ を計算する．実は，

$W_n = \overline{X}_{(n)} \sim \mathrm{N}(\mu, \sigma^2/n)$. $\dfrac{\overline{X}_{(n)} - \mu}{\sqrt{\sigma^2/n}} \sim \mathrm{N}(0, 1^2)$.

厳密には正規分布の再生性 前園確率統計定理 2.2 確率統計☆演習 II()L で．$n \to +\infty$ で正しいことは中心極限定理からわかる．正規母集団でないときも，標本サイズ n が大きい (30 くらい) なら，近似的に成立することが多い．

標本平均値が母平均値から大きく外れない確率は大きい (ここでは $1 - \alpha = 1 - 0.05$) という式を書くと… 表から $Q(1.96) = 0.05/2$ だから，

$$P\left(-1.96 < \frac{\overline{X}_{(n)} - \mu}{\sqrt{\sigma^2/n}} < +1.96\right) = 1 - 0.05.$$

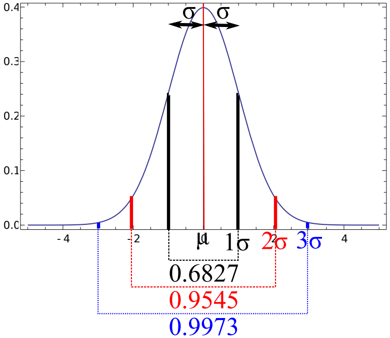
$$P\left(\mu - 1.96 \times \sqrt{\sigma^2/n} < \overline{X}_{(n)} < \mu + 1.96 \times \sqrt{\sigma^2/n}\right) = 1 - 0.05.$$

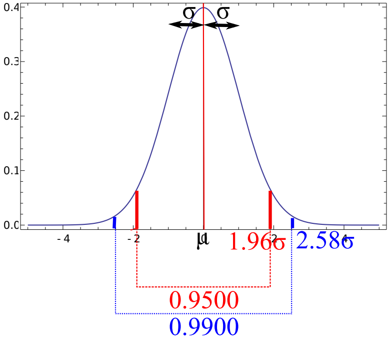
μ について不等式を解くと，

$$P\left(\overline{X}_{(n)} - 1.96 \times \sqrt{\sigma^2/n} < \mu < \overline{X}_{(n)} + 1.96 \times \sqrt{\sigma^2/n}\right) = 1 - 0.05.$$

標準正規分布 (ガウス分布) の確率

前園確率統計付表 1, 付表 2





切りがいい $0.05/2, 0.01/2$ を $Q(z)$ の表から探すと,

$$P(-1.96 < Z < 1.96) = 0.95,$$

$$P(-2.58 < Z < 2.58) = 0.99$$

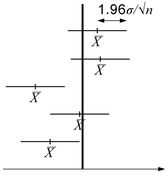
https://www.geogebra.org/classic#probability

母平均値 (正規母集団, 母分散既知) の信頼区間 高校 数学 B 前園確率統計 p.80

$\mathrm{N}(\mu, \sigma^2)$ にしたがう母集団の, σ^2 がわかっているとき, サイズ n の標本から区間推定すると,
母平均値 μ の 信頼係数 $1-\alpha = 0.95$ の信頼区間 (95%信頼区間), $1-\alpha = 0.99$ の信頼区間 (99%信頼区間) は,

$$\overline{X}_{(n)} - 1.96 \times \sqrt{\sigma^2/n} < \mu < \overline{X}_{(n)} + 1.96 \times \sqrt{\sigma^2/n},$$
$$\overline{X}_{(n)} - 2.58 \times \sqrt{\sigma^2/n} < \mu < \overline{X}_{(n)} + 2.58 \times \sqrt{\sigma^2/n}$$

$\boxed{\quad}$ とき, 信頼区間が μ を含む確率は 0.95 or 0.99.



高校 数学 B では, $1-\alpha = 0.95 \to 1.96$ の場合のみ.
$a < \mu < b$ でなく, 閉区間の記号 $[a, b]$ で.
真の母分散 σ^2 の代わりに, (不偏 $\frac{1}{n-1}$ じゃない) $S^2 = \frac{1}{n}\sum_i (X_i - \overline{X})^2$ の標本分散を使っていい

L11-Q2

母平均値の区間推定 (正規母集団, 母分散未知) 前園確率統計 p.82

μ はわからないのに σ^2 がわかってるケースはあまりない. ふつうはどちらもわからない.
σ^2 のかわりに不偏標本分散 s^2 (それ自身確率変数) を使っちゃいたい.
母集団が正規分布のときは, 使っちゃた量 $T = \frac{\overline{X}_{(n)} - \mu}{\sqrt{s^2/n}}$ が, 正規分布 $\mathrm{N}(0, 1^2)$ からちょっとずれた 自由度 $n-1$ の Student の t 分布にしたがうことが知られている.

$\boxed{\text{母集団が厳密に正規分布にしたがわなくても近似的に正しいことが多い.}}$

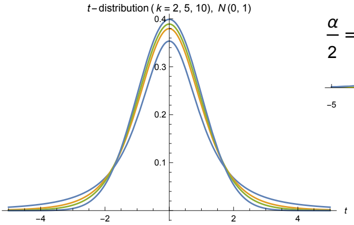
t 分布 前園確率統計 p.38

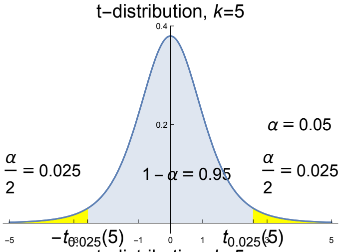
- 自由度 $k \to +\infty$ で $\mathrm{N}(0, 1^2)$ に一致する.
- 自由度 k が小さいとき, $\mathrm{N}(0, 1^2)$ より低く広い.

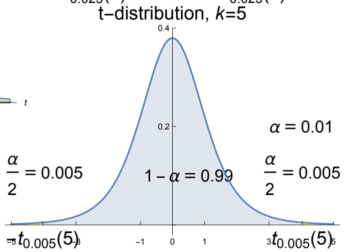
$$\text{確率密度関数 } f_k(x) = A_k \cdot \left(1 + \frac{1}{k} x^2\right)^{-\frac{k+1}{2}}.$$

t 分布表 前園確率統計付表 3

上側確率 $\alpha = 0.025, 0.005$, 自由度 k に対して, $\alpha = P(T > t(k;\alpha))$ となる $t(k;\alpha)$ の値の表.
https://www.geogebra.org/classic#probability
$t(k;0.025) \to 1.960, t(k;0.005) \to 2.576 \ (k \to +\infty)$.







母平均値の信頼区間 (母分散未知)

前園確率統計 p.82

(母分散未知の) 正規分布 $\mathrm{N}(\mu, ?^2)$ にしたがう母集団から, サイズ n の標本を得たとき, 母平均値 μ の 信頼係数 $1-\alpha$ の信頼区間は

$$\overline{X}_{(n)} - t(n-1;\alpha/2) \times \sqrt{s^2/n} < \mu < \overline{X}_{(n)} + t(n-1;\alpha/2 \times \sqrt{s^2/n}.$$

ただし, s^2:不偏標本分散, n: サンプルサイズ, $t(n-1;\alpha/2)$: 自由度 $n-1$ の t 分布の上側確率が $\alpha/2$ となる点 (表から求める).

L11-Q3

Quiz(母平均値の区間推定 (母分散未知))

あるドーナツ製造マシンが製造するドーナツの重さ Xg は, 正規分布にしたがう確率変数である
製造された 4 個のドーナツの重さを測定したところ, 次のようだった.
$51\mathrm{g}, 52\mathrm{g}, 47\mathrm{g}, 50\mathrm{g}.$

1. 母平均値 $\mu = \mathrm{E}[X]$ を, 信頼係数 $1-\alpha = 0.95$ で区間推定しよう.
2. 母平均値 $\mu = \mathrm{E}[X]$ を, 信頼係数 $1-\alpha = 0.99$ で区間推定しよう.

L11-Q4

前園確率統計例題 5.3

ここまで来たよ

10 略解:母集団・標本・標本抽出と推定

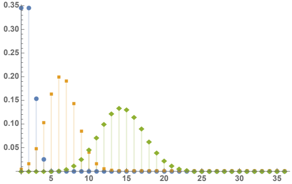
11 中心極限定理, 母平均値母比率の区間推定
- 中心極限定理と正規近似
- 母平均値の区間推定 (正規母集団, 母分散未知)
- 母比率 (ベルヌーイ分布の p) の区間推定

母比率の信頼区間 高校 数学 B 前園確率統計 §5.4

- 候補者 A の得票率は何% ? n 人に質問しただけで推定したい.
- 出荷する製品の何% が不良品? n 個だけ抜き出して調査したい.
- このコインの表が出る確率は? n 回投げるだけで推定したい.

$Y \sim \mathrm{B}(n, p)$. n が大きいとき近似的に $Y \sim \mathrm{N}(np, np(1-p))$.
$\frac{Y}{n} \sim \mathrm{N}(p, \frac{1}{n}p(1-p))$.

$p = 0.8, n = 4, 20, 40.$



$$P\left(p - 1.96 \times \sqrt{\frac{1}{n}p(1-p)} < \hat{p} < p + 1.96 \times \sqrt{\frac{1}{n}p(1-p)}\right) = 0.95$$

逆に解いて ($\sqrt{}$ の中で $p = \hat{p}$ とする近似をする.

母比率の信頼区間 (母分散未知) 前園確率統計 §5.4

X のサイズ n の標本で, 標本比率 $\hat{p} = y/n$ のとき, 母比率の信頼係数 $1 - \alpha = 0.95$, 信頼係数 $1 - \alpha = 0.99$ の信頼区間は

$$\hat{p} - 1.96 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})} < p < \hat{p} + 1.96 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})},$$

$$\hat{p} - 2.58 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})} < p < \hat{p} + 2.58 \times \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})}.$$

L11-Q5

Quiz(母比率の区間推定)

選挙で出口調査をしたところ, 50 人中 35 人が A 候補に投票したと答えた. 母集団を投票した人全体とする. そのうち A 候補に投票した人の母比率 (得票率) を考える.

1. A 候補の得票率を, (点) 推定しよう
2. A 候補の得票率を, 信頼係数 $1-\alpha=0.95$ で区間推定しよう.
3. A 候補の得票率を, 信頼係数 $1-\alpha=0.99$ で区間推定しよう.

前園確率統計例題 5.5

注: 下限, 上限が 0,1 を越えるときは, 0,1 に直してしまっていい.

連絡

Moodle
https://learn.math.ryukoku.ac.jp/

GeoGebra 確率電卓
https://www.geogebra.org/classic#probability

Moodle モバイルアプリ
https://download.moodle.org/mobile

起動後, URL https://learn.math.ryukoku.ac.jp/moodle を登録

https://learn.math.ryukoku.ac.jp → 今日のところ → 教室内 標本抽出-区間推定
https://learn.math.ryukoku.ac.jp/moodle/mod/questionnaire/view.php?id=1547

- 図書館ミニ講義「確率を学ぶ〜年末ジャンボ宝くじが当たる確率は !? 〜」by 樋口
  - 2018-12-20 木 12:45-13:15
  - 生協コンビニ地下スチューデントコモンズ (瀬田) ミーティングスペース
- 学期途中の振り返りのレポート. –2018-12-12 水. https://manaba.ryukoku.ac.jp
- 予習復習問題を, 期限後も (再/初) 受験できます. 点数にはカウントしないけど, プチテスト準備に活用してね.
- Learn Math Moodle の予習復習問題は来週期限のものがあります. プチテストに備えてね.
- 教科書の統計的検定 前園確率統計 §6.1, 前園確率統計 §6.6 読んできてね.