中心極限定理・正規近似・区間推定

樋口さぶろお　https://hig3.net

龍谷大学理工学部数理情報学科

確率統計☆演習 I　L11(2019-12-16 Mon)

最終更新: Time-stamp: "2019-12-16 Mon 18:09 JST hig"

今日の目標

- 中心極限定理の内容を説明できる
- 正規近似できる
- 母平均値, 母期待値, 母比率を区間推定できる

L10-Q1

Quiz 解答:母平均値, 母分散, 母比率の点推定　在庫のフライドチキンの重さを X とすると,

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

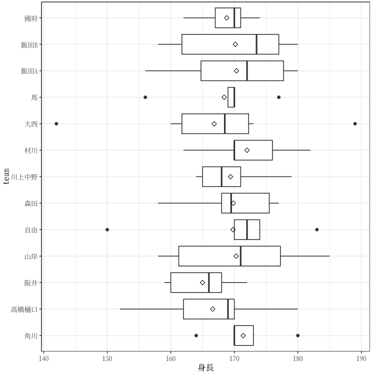
ここまで来たよ

10. 大数の法則/母集団・標本抽出・推定
    - 母比率とその (点) 推定
11. 中心極限定理・正規近似・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, 母分散既知)
    - 母比率の区間推定

標本抽出 → 点推定の結果

母集団=回答者全体, 母集団サイズ=75
身長の母平均値=169.3cm, 身長の母分散=72.9cm^2, 滋賀県高校の母比率=0.31.
標本平均値 $\overline{X}$, 不偏標本分散 S^2, 標本比率 p, 標本サイズ N

team	$\overline{X}$(cm)	S^2(cm^2)	p	N
國府	168.80	20.700	0.20000	5
飯田 B	170.17	95.367	0.16667	6
飯田 A	170.33	90.667	0.33333	6
馬	168.40	58.300	0.40000	5
大西	166.83	240.567	0.50000	6
村川	172.00	49.600	0.50000	6
川上中野	169.40	36.300	0.20000	5
森田	169.83	50.167	0.33333	6
自由	169.80	147.200	0.20000	5
山岸	170.30	92.900	0.40000	10
阪井	165.00	30.000	0.20000	5
高橋樋口	166.60	107.800	0.20000	5
角川	171.40	33.800	0.20000	5



比率=ratio 岩薩林 確率・統計 p.107

確率変数 $Y \sim \mathrm{B}(1, p)$ ベルヌーイ分布, を考える.
こういう Y は, いろんな母集団を, 条件 $f(X) =$「X は…である」の成立不成立で 2 つに類別して作れる. カテゴリ変数

- $X \sim$ ある分布, $Y = \mathrm{I}_{[\cdots である]}(X)$, たとえば $X > 10$ なら $Y = 1$.
- 母集団=日本国民, 国民 X の血液型が A であるなら $Y = 1$.

母比率 岩薩林 確率・統計 p.107

$\mathrm{B}(1, p)$ の p. または母集団で条件 $f(x)$ から $\mathrm{B}(1, p)$ を作ったとき, '母集団の「…である」ものの母比率', ともいう.
有限母集団なら,

$$\text{母集団の「…である」母比率 } p = \frac{\text{「…である」データ } x \text{ の個数}}{\text{母集団サイズ}} = \mathrm{E}[Y]$$

母比率の (点) 推定 岩薩林 確率・統計 p.115

標本比率 岩薩林 確率・統計 p.115

標本のデータ n 個中 k 個が「…である] とき,

$$\text{標本比率 } \hat{p} = \frac{k}{n}$$

が「…」の母比率 p のよい推定値になっている.

母比率 p の推定=母平均値 $\mathrm{E}[Y]$ の推定

サイズ n の標本中 k 個が「…である」とき,

$$\text{母平均値 } \mathrm{E}[Y] \text{ の推定値 } = \text{標本平均値} \overline{Y}$$

$$= \frac{1}{n}[\underbrace{1 + \cdots + 1}_{k} + \underbrace{0 + \cdots + 0}_{n-k}] = \frac{k}{n} = \hat{p}.$$

岩薩林 確率・統計問題 6(p.116)

ここまで来たよ

10 大数の法則/母集団・標本抽出・推定
- 母比率とその (点) 推定

11 中心極限定理・正規近似・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, 母分散既知)
- 母比率の区間推定

(復習) 独立同分布 (i.i.d) にしたがう確率変数の和

岩薩林 確率・統計例題 4.6(p.84)

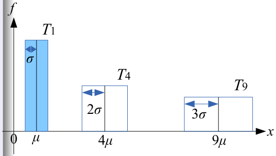
i.i.d にしたがう確率変数 X_i の和

母平均値 $\mathrm{E}[X_i]=\mu$, 母分散 $\mathrm{V}[X_i]=\sigma^2$.
和の確率変数 $T_n = X_1+\cdots+X_n$.

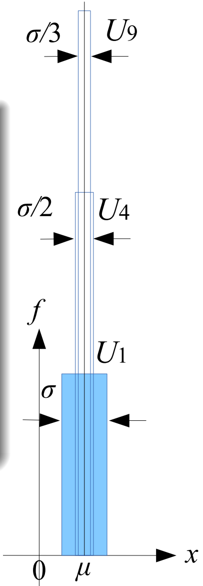
$$\mathrm{E}[T_n]=\sum_{i=1}^{n}\mathrm{E}[X_i]=n\times\mu.$$

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

T_n の確率密度関数はこんな感じ?



U_n の確率関数は
こんな感じ?



i.i.d にしたがう確率変数 X_i の和の $1/n$

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

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

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

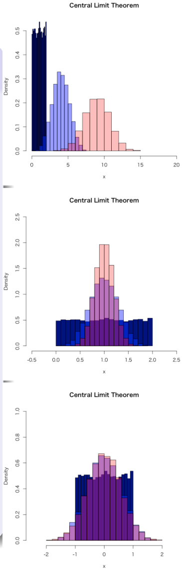
大数の法則と話があってる! 実際, これを使って大数の法則が証明できる

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)

$X_1, \ldots, X_n$ 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

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



独立同分布にしたがう確率変数の和の正規近似 I

L11-Q1

Quiz(二項分布と正規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.
$T = X_1 + \cdots + X_{400}$ とする.
$P(T > 31)$ の確率を, 標準正規分布の $Q(z)$ または $I(z)$ 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.

二項分布の正規近似 高校 数学 B I

L11-Q2

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

1. T 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. T はどのような独立同分布の和と考えられるか. T 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 $Q(z)$ または $I(z)$ 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.

岩薩林 確率・統計第 5 章練習問題 1(3)

ここまで来たよ

10 大数の法則/母集団・標本抽出・推定
- 母比率とその (点) 推定

11 中心極限定理・正規近似・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, 母分散既知)
- 母比率の区間推定

点推定 対 区間推定

点推定 岩薩林 確率・統計 §6.1

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 岩薩林 確率・統計 §6.1

「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. 信頼係数=信頼度を使って,

「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 岩薩林 確率・統計 p.144

$\mathrm{N}(\mu, \sigma^2)$ 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,

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

> $n \to +\infty$ 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 0 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

$P(-1.96 < Z < +1.96) = 2I(1.96) = 1 - 0.05$ だから,

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

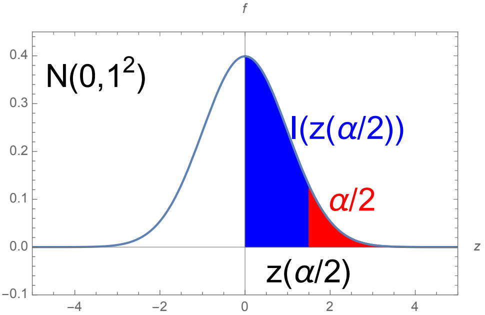
$$P\left(\mu - 1.96 \times \sqrt{\sigma^2/n} < \overline{X}_{(n)} < \mu + 1.96 \times \sqrt{\sigma^2/n}\right) = 1 - 0.05.$$

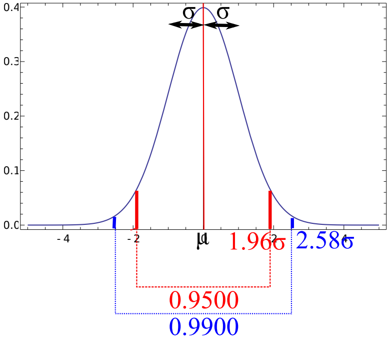
μ について不等式を解くと,

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

標準正規分布 (ガウス分布) の確率

岩薩林 確率・統計付表 1





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

$$P(-1.96 < Z < 1.96) = 2I(1.96) = 0.95,$$

$$P(-2.58 < Z < 2.58) = 2I(2.58) = 0.99$$

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 岩薩林 確率・統計定理 6.1(p.18)

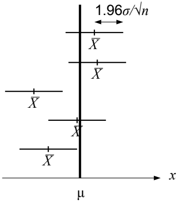
$\mathrm{N}(\mu, \sigma^2)$ にしたがう母集団の, σ^2 がわかっているとき, サイズ n の標本から区間推定すると, 母平均値 μ の 信頼係数 $1-\alpha$ の信頼区間 ($(1-\alpha) \times 100\%$信頼区間) は, $\overline{X}_{(n)}$ を標本平均値として,

$$\overline{X}_{(n)} - z(\tfrac{\alpha}{2}) \times \sqrt{\sigma^2/n} < \mu < \overline{X}_{(n)} + z(\tfrac{\alpha}{2}) \times \sqrt{\sigma^2/n}.$$

α 推定がはずれる確率. $z(\frac{0.05}{2}) = 1.96, z(\frac{0.01}{2}) = 2.58.$

とき, 信頼区間が μ を含む確率は, 信頼係数 $1-\alpha$.

高校 数学 B では, $z(\frac{0.05}{2}) = 1.96$ 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-Q3

Quiz(母平均値の区間推定 (母分散既知))

あるドーナツ製造マシンが i 番目に製造するドーナツの重さ X_ig は, 独立で, 同じ正規分布にしたがう確率変数である. あらかじめ行った調査により, X_i の母分散は $\sigma^2 = 9\mathrm{g}^2$ であることがわかっている.
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 $\mu = \mathrm{E}[X_i]$ を, 信頼係数 $1 - \alpha = 0.95$ で区間推定しよう.
2. 母平均値 $\mu = \mathrm{E}[X_i]$ を, 信頼係数 $1 - \alpha = 0.99$ で区間推定しよう.

岩薩林 確率・統計例題 6.4(p. 145) | 岩薩林 確率・統計問題 3(p.146) | 岩薩林 確率・統計第 6 章練習問題 1

ここまで来たよ

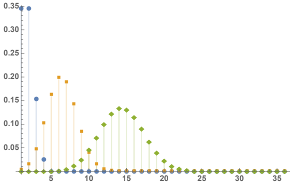
10 大数の法則/母集団・標本抽出・推定
- 母比率とその (点) 推定

11 中心極限定理・正規近似・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, 母分散既知)
- 母比率の区間推定

母比率の信頼区間高校 数学 B

- 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 $Y \sim \mathrm{N}(np, np(1-p))$.
$\frac{Y}{n} \sim \mathrm{N}(p, \frac{1}{n}p(1-p))$.



$p = 0.8, n = 4, 20, 40$.
信頼係数 $1-\alpha$.

$$P\left(p - z(\tfrac{\alpha}{2}) \times \sqrt{\frac{1}{n}p(1-p)} < \hat{p} < p + z(\tfrac{\alpha}{2}) \times \sqrt{\frac{1}{n}p(1-p)}\right) = 1-\alpha$$

$\sqrt{}$ の中で $p = \hat{p}$ とする近似で, p について解く.

母比率の信頼区間 (母分散未知) 岩薩林 確率・統計 §7.3

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 $1-\alpha$ の信頼区間 ($(1-\alpha) \times 100\%$信頼区間) は,

$$\hat{p} - z(\tfrac{\alpha}{2}) \times \sqrt{\tfrac{1}{n}\hat{p}(1-\hat{p})} < p < \hat{p} + z(\tfrac{\alpha}{2}) \times \sqrt{\tfrac{1}{n}\hat{p}(1-\hat{p})}.$$

$z(\frac{0.05}{2}) = 1.96, z(\frac{0.01}{2}) = 2.58.$

L11-Q4

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.

岩薩林 確率・統計例題 7.6(p.170) | 岩薩林 確率・統計問題 7(p.171) | 岩薩林 確率・統計第 7 章練習問題 2(2)

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

- 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
  - 2019-12-19 木 12:50-13:20
  - 図書館地下ナレッジコモンズ ナレッジスクエア
- 2020-01-20 月 1 テスト B1 実施．45 分, PC 受験の予定．持込不可．教科書の p. xii–xv のコピーを配布．
- 2020-01-27 月 1(定期試験期間) テスト A2,B2 実施．A1,B1 と同等の評価内容．
- 2020-02-15 土 学力認定試験 2 年生も受験可能 `https://wiki.math.ryukoku.ac.jp` (全学認証)
- 成績の 70 ピーナッツ分$=\min(\max(A1, A2), \max(B1, B2))$.
- 予習復習問題を, 期限後も (再/初) 受験できるようにしてます．ピーナッツにはカウントしないけど, テスト準備に活用してね．
- 2019-12-23 は統計的仮説検定．§§6.1,7.1,7.3 読んできてね．
- 2019-12-23 の Trial 予告 $\boxed{}$
- 2019-12-16 までの (短い作文の) レポート課題 `https://manaba.ryukoku.ac.jp`
- オフィスアワー木 6(1-539) 金昼 (1-542), Math ラウンジ (1-536/538)

学籍番号での LINE 公式アカウント登録

龍大 manaba course