母比率・母平均値の区間推定

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 L11(2016-12-15 Thu)

最終更新: Time-stamp: "2016-12-16 Fri 08:31 JST hig"

今日の目標

- 標本から母平均値を区間推定できる (母分散既知, 母分散未知) 塚田確率統計 §7.3.1,7.3.2
- t 分布の確率が求められる 塚田確率統計 7.3.2
- 標本から母比率 塚田確率統計 §7.3.7 を区間推定できる

http://hig3.net

L10-Q1
Quiz 解答:二項分布と正規分布と中心極限定理

1. $X \sim \mathrm{B}(1, \frac{2}{3})$, $U \sim \mathrm{B}(18, \frac{2}{3})$.
2. $X_{,} \cdots, X_{18}$ は独立なので, $\mathrm{E}[U] = 12$, $\mathrm{V}[U] = 4$ である．$N = 18$ が大きいと考えると, 中心極限定理より, U は近似的に正規分布 $\mathrm{N}(12, 2^2)$ にしたがう．
3. 中心極限定理より, $Z = \frac{U-12}{2}$ は近似的に標準正規分布 $\mathrm{N}(0, 1^2)$ にしたがう．よって, 求める確率は, $P(X \leq 9) = P(Z < -\frac{3}{2}) = 1/2 - I(\frac{3}{2})$.

L10-Q2
Quiz 解答:中心極限定理

1. U は母平均値が 100, 分散が $\frac{100}{4}$. 近似的に正規分布 $\mathrm{N}(100, \frac{100}{4})$ に従う.
   $Z = \frac{U-100}{\sqrt{100/4}}$ は近似的に標準正規分布 $\mathrm{N}(0, 1^2)$ に従う.
   $P(U > 110) = P(Z > 2) = \frac{1}{2} - I(2)$.
2. W は母平均値が 1, 分散が $\frac{1}{400}$. 近似的に正規分布 $\mathrm{N}(1, \frac{1}{400})$ に従う.
   $Z = \frac{W-1}{\sqrt{1/400}}$ は近似的に標準正規分布 $\mathrm{N}(0, 1^2)$ に従う.
   $P(W < 1.01) = P(Z < 0.2) = \frac{1}{2} + I(0.2)$.

L10-Q3

L10-Q4

Quiz 解答:母平均値, 母分散の点推定

1. 標本平均値は, $\frac{1}{6}(117 + \cdots + 112) = 111\mathrm{g}$ なので, 母平均値は $111\mathrm{g}$ と推定できる.
2. 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.

L10-Q5

Quiz 解答:推定

これはサイズ 10 の標本.
標本平均値は

$$\frac{1}{10}[0+0+0+0+0+0+10+10+30+100]=15(\text{円}).$$

よって, 母平均値は 15 円と推定される.
不偏標本分散は

$$\frac{1}{10-1}[(0-15)^2\times 6+(10-15)^2\times 2+(30-15)^2+(100-15)^2]=983.3(\text{円}^2)$$

. よって, 母分散は 983.3 円^2 と推定される.
母標準偏差は $\sqrt{983.3}=31.4$ 円と推定される.

ここまで来たよ

点推定 対 区間推定

点推定

真の母平均値はわからないが, 標本平均値を使って,

「母平均値は A 円と推定される」

それどのくらい正確なの? 実は

区間推定

「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 (母分散既知) 塚田確率統計 §7.3.1 高校 数学 B

任意の母集団 (母平均値 μ, 母分散 σ^2) 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.

母集団が $\mathrm{N}(\mu, \sigma^2)$ のとき, $W = \overline{X}_{(n)}$ は $\mathrm{N}(\mu, \sigma^2/n)$ にしたがう. $\frac{\overline{X}_{(n)} - \mu}{\sqrt{\sigma^2/n}}$ は $\mathrm{N}(0, 1^2)$ にしたがう.

母集団が正規でないときも, 標本サイズ n が大きい (30 くらい) なら, 中心極限定理から, 近似的に $\frac{\overline{X}_{(n)} - \mu}{\sqrt{\sigma^2/n}}$ は $\mathrm{N}(0, 1^2)$ にしたがう.

母平均値から遠く外れる, 確率 $\alpha = 0.05$ の例外的事象が起きない確率は

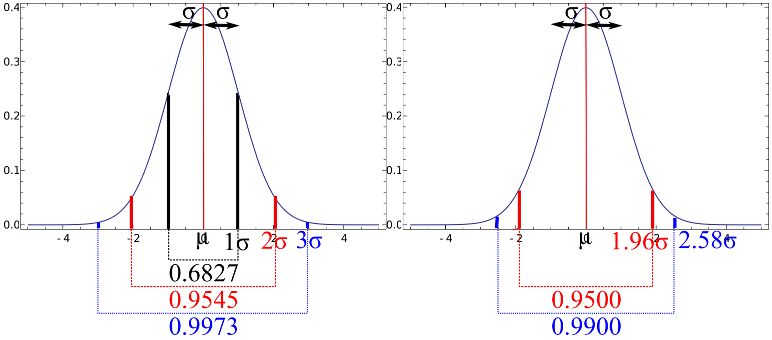
$$P\left(-1.96 < \frac{\overline{X}_{(n)} - \mu}{\sqrt{\sigma^2/n}} < +1.96\right) = 1 - 0.05.$$

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

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) のグラフに関係した面積=確率



母平均値の信頼区間 塚田確率統計 §7.3.2 高校 数学 B

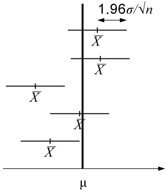
母平均値 μ の 信頼係数 $1-\alpha = 0.95$ の信頼区間 (95%信頼区間) は

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

母平均値 μ の 信頼係数 $1-\alpha = 0.99$ の信頼区間 (99%信頼区間) は

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

$\boxed{}$ 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 $[a, b]$ で答えることになってる．
σ^2 の代わりに，(不偏 $\frac{1}{n-1}$ じゃない) $S^2 = \frac{1}{n}\sum_i (X_i - \overline{X})^2$ の標本標準偏差を使っていいことになってる．

L11-Q1

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次々に製造するドーナツの重さ X_ig は, 独立同分布である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.

L11-Q2 塚田確率統計問 7.4.2

推定が正確であるとは 信頼区間が $\boxed{}$ であること.

Quiz(区間推定の性質)

標本からの母平均値の区間推定について, 正しいのはどれ?

1. 母分散が大きいほど, 信頼区間は大きくなる
2. 標本サイズが大きいほど, 信頼区間は大きくなる
3. 母平均値が大きいほど, 信頼区間は小さくなる
4. 信頼係数が大きいほど, 信頼区間は小さくなる

ここまで来たよ

母平均値の区間推定 (正規母集団, 母分散未知) 塚田確率統計 §7.3.2

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.
母集団が厳密に正規分布でなくても近似的に正しいことが多い.

t 分布

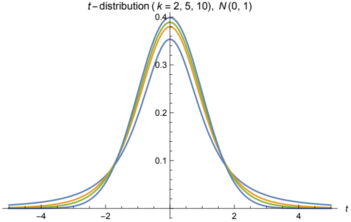
- 自由度 $k \to +\infty$ で $\mathrm{N}(0, 1^2)$ に一致する.
- 自由度 k が小さいとき, $\mathrm{N}(0, 1^2)$ より低く広い.

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

t 分布表 塚田確率統計付録 B 表 B3

上側確率 $\alpha = 0.025, 0.005$, 自由度 k に対して, $\alpha = P(T > t_k^*)$ となる t_k^* の値の表.

$k\backslash$ 上側確率$\alpha/2$	0.025	0.005
1	12.71	63.66
2	4.303	9.925
3	3.182	5.841
4	2.776	4.604
5	2.571	4.032
6	2.447	3.707
7	2.365	3.499
8	2.306	3.355
9	2.262	3.250
10	2.228	3.169
11	2.201	3.106
12	2.179	3.055
13	2.160	3.012
14	2.145	2.977
15	2.131	2.947
16	2.120	2.921
17	2.110	2.898
18	2.101	2.878
19	2.093	2.861
20	2.086	2.845
30	2.042	2.750
40	2.021	2.704
50	2.009	2.678
60	2.000	2.660
80	1.990	2.639
100	1.984	2.626
$+\infty$	**1.960**	**2.576**



母平均値の信頼区間 (母分散未知)

母集団が (母分散未知の) 正規分布にしたがうとき, 母平均値 μ の信頼係数 $1-\alpha$ の信頼区間は

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

ただし, s^2:不偏標本分散, n: サンプルサイズ, t^*_{n-1}: 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次々に製造するドーナツの重さ Xg は, 独立な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

塚田確率統計問 7.4.3

母平均値の区間推定 (母分散未知, 大標本) 高校 数学 B

自由度 $n-1$ が大きいとき, t 分布のかわりに $\mathrm{N}(0,1^2)$ を使っても大した誤差じゃない. また, 母集団が正規分布でなくても, 中心極限定理から, 近い結果になることが多い.

物理実験

L11-Q5

Quiz(母平均値の区間推定 (母分散未知, 大標本))

あるドーナツ製造マシンが次々に製造するドーナツの重さ X_ig は, 独立同分布にしたがう確率変数である.

製造された 400 個のドーナツの重さを測定したところ, 次のようだった.

$51\mathrm{g}, 52\mathrm{g}, 47\mathrm{g}, \ldots, 50\mathrm{g}$.

ここから標本平均値, 不偏標本分散を計算したところ, $m = 51\mathrm{g}, s^2 = 4\mathrm{g}^2$ だった.

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.

ここまで来たよ

1. 中心極限定理・母集団・標本抽出・点推定
2. 母比率・母平均値の区間推定
   - 母平均値の区間推定 (母分散既知の場合:標準正規分布)
   - 母平均値の区間推定 (正規母集団, 母分散未知)
   - 母比率または二項分布の母平均値の区間推定

母比率の信頼区間 塚田確率統計 §7.3.6 高校 数学 B

$X \sim \mathrm{B}(1,p)$ とするとき,

$$f(x) = \begin{cases} p & (x=1,\cdots\text{である}) \\ 1-p & (x=0,\cdots\text{である以外}) \end{cases}$$

復習: $\mathrm{E}[X]=p, \mathrm{V}[X]=p(1-p)$.
n 回試行を繰りかえしたとき,

$$\text{母平均値の推定値} = \text{標本平均値} = \frac{1}{n}\sum_i X_i = \frac{Y}{n}$$

Y は $X=1$ の回数, 表が出た回数, A 型が出た回数.
$Y \sim \mathrm{B}(n,p)$. n が大きいとき近似的に $Y \sim \mathrm{N}(np, np(1-p))$.
二項分布では, 特に p を母比率, $\hat{p} = \frac{y}{n}$ を標本比率と言う.

$\hat{p}(1-\hat{p})$ が母分散そのものと思って, 母平均値の区間推定 (分散未知, 大標本) を使うと,

母比率の区間推定

母比率の信頼係数 $1-\alpha = 0.95$ の信頼区間は, サンプルサイズ n と標本比率 $\hat{p} = Y/n$ により,

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

母比率の信頼係数 $1-\alpha = 0.99$ の信頼区間は

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

L11-Q6

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のうち A 候補に投票した人の母比率, すなわち投票者全体での A 候補の得票率を考える.

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.

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.

L11-Q7　塚田確率統計問 7.4.4

連絡

- 予習問題は, $\boxed{\text{次々回の授業直前}}$ 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.
- ごめんなさい Manaba のレポートを改めて公開してます.
- 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)
- 次回は母平均値の統計的仮説検定 塚田確率統計 §8.1 塚田確率統計 §8.2

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