母平均値・母比率の区間推定

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(2017-12-13 Wed)

最終更新: Time-stamp: "2017-12-25 Mon 17:33 JST hig"

今日の目標

- 正規母集団の標本から母平均値を区間推定できる (母分散既知 西川確率統計 §8.2, 未知 西川確率統計 §8.3)
- ベルヌーイ母集団の標本からに母比率を区間推定できる 西川確率統計 §8.4
- t 分布の確率が求められる 西川確率統計 §6.4.4

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. 表の出る回数 U は, 二項分布 $\mathrm{B}(400, \frac{1}{10})$ にしたがう. よって, $\mathrm{E}[U] = 40$, $\mathrm{V}[U] = 36$ である.
2. 各回 i の表裏について, 確率変数
$$X_i = \begin{cases} 1 & (\text{表}) \\ 0 & (\text{裏}) \end{cases}$$
を考えると, $U = X_1 + \cdots + X_{400}$ である. X_i $(i = 1, \ldots, 400)$ は独立同分布にしたがうので, $n = 400$ 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 $\mathrm{N}(40, 36)$ にしたがう.
3. $Z = \frac{U-40}{\sqrt{36}}$ 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, $P(U > 31) = P(Z > -\frac{9}{6}) = Q(-\frac{3}{2}) - Q(\infty) = (1 - Q(\frac{3}{2})) - 0 = 0.9332$.

L10-Q2

Quiz 解答:母平均値, 母分散, 母比率の点推定

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.
3. 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

10 中心極限定理・母集団・標本抽出・点推定

11 母平均値・母比率の区間推定

- 母平均値の区間推定 (正規母集団, 母分散既知)
- 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

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

西川確率統計 §8.1

「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.

「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 西川確率統計 §8.2

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

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

厳密には 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, 中心極限定理から,
$\dfrac{\overline{X}_{(n)} - \mu}{\sqrt{\sigma^2/n}}$ は $\mathrm{N}(0, 1^2)$ に近似的にしたがうと思える.

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 $Q(1.96) = 0.05/25$ だから,

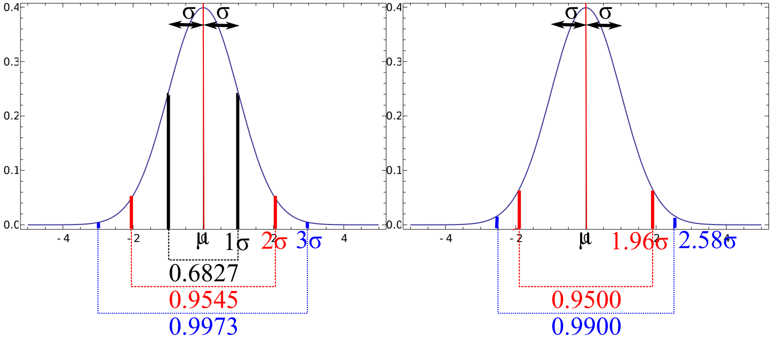
$$P\left(-1.96 < \frac{\overline{X}_{(n)} - \mu}{\sqrt{\sigma^2/n}} < +1.96\right) = 1 - 0.05.$$

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

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



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

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

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

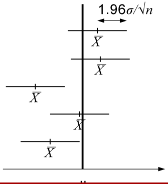
母平均値 (正規母集団, 母分散既知) の信頼区間 高校 数学 B 西川確率統計 §8.2

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

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

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-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製造するドーナツの重さ $X\mathrm{g}$ は, 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 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 $\mu = \mathrm{E}[X]$ 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-Q2

西川確率統計演習 8.1

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

10 中心極限定理・母集団・標本抽出・点推定

11 母平均値・母比率の区間推定
- 母平均値の区間推定 (正規母集団, 母分散既知)
- 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

母平均値の区間推定 (正規母集団, 母分散未知) 西川確率統計 §8.3

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にしたがわなくても近似的に正しいことが多い．

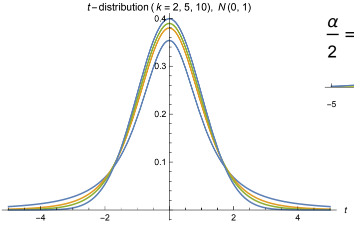
t 分布 西川確率統計 §6.4.4

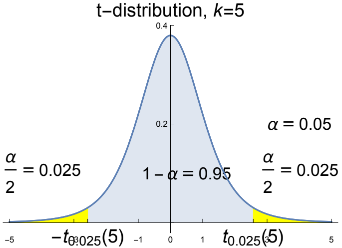
- 自由度 $k \to +\infty$ で $\mathrm{N}(0, 1^2)$ に一致する．
- 自由度 k が小さいとき，$\mathrm{N}(0, 1^2)$ より低く広い．

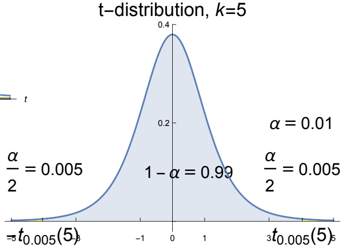
$$\text{確率密度関数 } f_k(x) = A_k \cdot \left(1 + \frac{1}{k}x^2\right)^{-\frac{k+1}{2}}.$$

t 分布表 西川確率統計付録表 B2

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 $t_\alpha(k)$ の値の表.
$t_{0.025}(k) \to 1.960, t_{0.005}(k) \to 2.576$ $(k \to +\infty)$.







母平均値の信頼区間 (母分散未知) 西川確率統計 8.3

(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

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

ただし, s^2:不偏標本分散, n: サンプルサイズ, $t_{\alpha/2}(n-1)$: 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.
$51\text{g}, 52\text{g}, 47\text{g}, 50\text{g}.$

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

西川確率統計例題 8.2, 問題 8.2(p.166), 演習 8.2(p.171)

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製造するドーナツの重さ Xg を確率変数と考える.
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

母比率の信頼区間 高校 数学 B 西川確率統計 §8.4

- 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.

確率変数があったとき, 条件「…である」を満たすかどうかによって, 満たす $x=1$ それ以外 $x=0$ を対応させる.
このとき, $X \sim \mathrm{B}(1,p)$ (ベルヌーイ分布)
復習: $\mathrm{E}[X]=p, \mathrm{V}[X]=p(1-p)$.
このようなとき, p を母比率という. 二項分布にも使う言葉.
n 回試行を繰りかえしたとき,

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

Y は $X=1$ の回数. $\hat{p}=\frac{y}{n}$ を標本比率と言う.
なお, $Y \sim \mathrm{B}(n,p)$.

$X \sim \mathrm{B}(n, p)$ に対して, 近似値 $\hat{p}(1-\hat{p})$ が X の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,

母比率の信頼区間 (母分散未知)

西川確率統計 8.4

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-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.

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

西川確率統計例題 8.4,8.5, 問題 8.4,8.5, 演習 8.4(p.170)

分類

- 正規母集団, 母分散既知 (あまりない)
  - 大標本でない 正規分布で 1.96,2.58
  - 大標本 正規分布で 1.96,2.58
- 正規母集団, 母分散未知 (よくあること)
  - 大標本でない t 分布
  - 大標本 t 分布で $n \to +\infty$ して正規分布で 1.96,2.58
- ベルヌーイ母集団
  - 大標本でない (やらなかった) 二項分布としてやるしかない.
  - 大標本 正規分布で 1.96,2.58

正規分布でなくても, 標本サイズが十分大きくなくても, 使ってしまうことが世の中ではけっこうあるらしい⋯ が, 数理卒の人は適切かどうか判断できるように.

連絡

- 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月 3.5(1-539) 金 4(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 西川確率統計 §7