統計の分析と利用
（旧カリ：データ分布と予測）

推定：点推定と区間推定

堀田敬介

2011/12/9,Fri.

Contents

- 推定
  - 点推定と区間推定
    - <u>点推定 point estimation</u>
      - モーメント法 method of moments
      - 最尤法 maximum likelihood method
    - <u>区間推定 interval estimation</u>
      - 母平均の推定:母分散が既知の場合〔<u>**Z推定**</u>〕
      - 母平均の推定:母分散が未知の場合〔<u>***t*** **推定**</u>〕
      - 母分散の推定:〔<u>**χ^2推定**</u>〕
      - 母比率の推定:〔<u>**Z推定**</u>〕
  - 2つの正規母集団の比較
    - <u>母平均の差</u>の区間推定
    - <u>母分散の比</u>の区間推定

推測統計

- 推測対象による分類

推測統計学
inferential statistics

- **母数の値**
  母平均
  母分散
- **母集団の従う分布**
- **その他**

推測対象

推定：点推定，区間推定

仮説検定

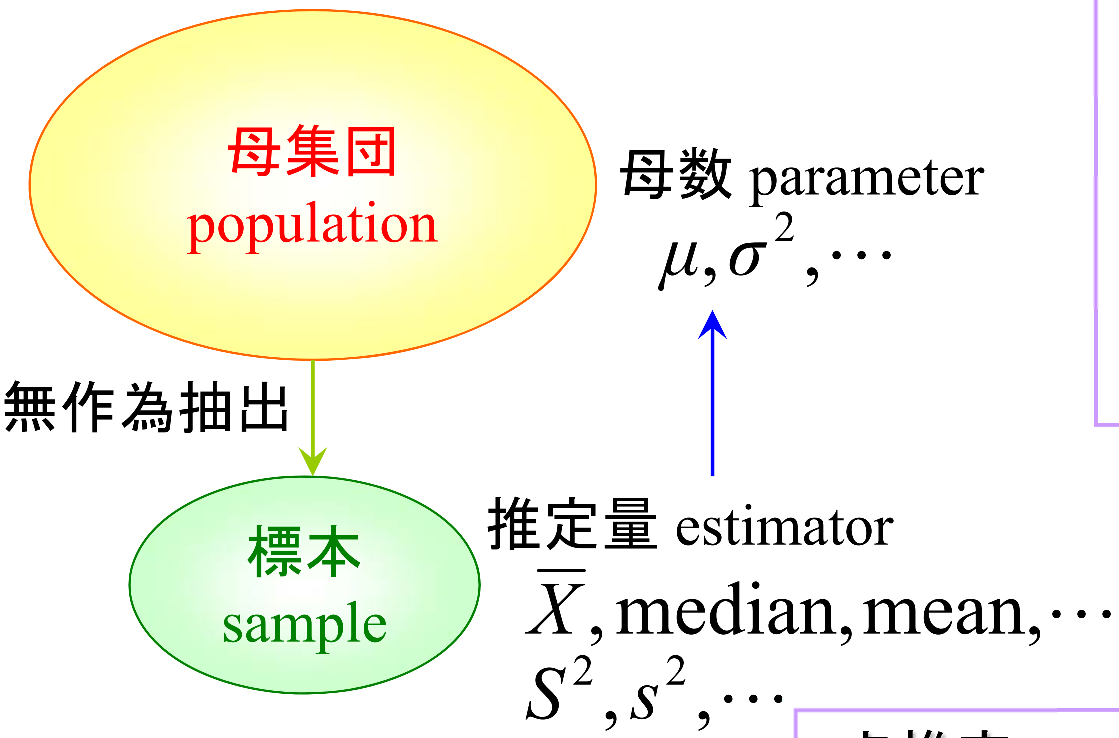
適合度検定

独立性の検定

推測方法

母数の推定

■ 母集団の推定



- パラメトリック
  母数θがわかると母集団分布がわかる場合
- ノン・パラメトリック
  母数θのみ推定したい(母集団分布に関心がない)場合

- 点推定
  母数θをある1つの値 $\hat{\theta}$ で指定する方法
- 区間推定
  母数θの値が入る確率がある値以上を保証する区間を求める方法

母数の推定:不偏推定量

標本の観測値から
計算される量

- 母数の推定量・推定値
  - 母数θを推定するために用いる統計量Wを，θの**推定量**という
  - 推定量Wの実現値をθの**推定値**という
- 不偏推定量
  - $E(W)=\theta$が成り立つとき，統計量Wをθの**不偏推定量**という
    - 例1:標本平均 $\overline{X}$

      $E(\overline{X})=\mu$ より不偏推定量
    - 例2:標本分散 S^2

      $E(S^2)=\dfrac{n-1}{n}\sigma^2$ より不偏推定量ではない
    - 例3:不偏分散 $s^2=\dfrac{1}{n-1}\left\{(X_1-\overline{X})^2+\cdots+(X_n-\overline{X})^2\right\}$

      $E(s^2)=\sigma^2$ より不偏推定量

母平均の推定：点推定〔point estimation〕

- ある店舗の36日分の週末来客数のデータ

300	356	319	213	229	244
317	306	390	287	268	257
274	231	370	275	186	327
365	272	335	167	289	352
351	299	327	405	259	376
301	337	229	244	279	243

標本 sample

母集団 population
ある店舗の
週末平均来客数
母平均 μ？

この店舗の週末の平均来客数を知りたい！

$X_1=300,\ X_2=356,\ \ldots,\ X_{36}=243\ (n=36)$

標本平均：$\bar{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

標本平均値：294

点推定

この店舗の週末平均来客数は 294 である
即ち，母平均 $\mu=294$ である

母平均の点推定

- 点推定
  - □ 積率法 method of moments
    - 積率(モーメント)を利用する方法
  - □ 最尤法 maximum likelihood method
    - 最尤原理:「現実の標本は確率最大のものが実現した」に基づく方法

Xの(原点まわりの)r次積率
$$\mu_r = E(X^r)$$
Xの期待値まわりのr次積率
$$\mu_r' = E(X-\mu)^r$$
Xのr次標準化積率
$$\alpha_r = E\{(X-\mu)/\sigma\}^r$$

$$\begin{cases} \text{母数} \quad \theta \\ \text{標本} \quad X_1,\cdots,X_n \\ \text{母集団確率分布} \quad f(x,\theta) \end{cases}$$

n個の標本の実現値(観測値)
$$x_1, x_2 \cdots, x_n$$

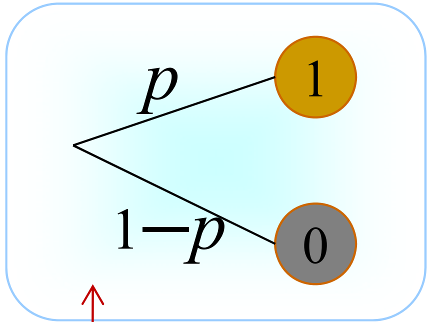
$$\longrightarrow \quad \text{尤度関数} \; L(\theta) = f(x_1,\theta)\times\cdots\times f(x_n,\theta) = \prod_{i=1}^{n} f(x_i,\theta)$$

尤度関数を母数空間Θ上で最大にするものを推定値・推定量とする
- 尤度関数を最大にするθ:最尤推定値 maximum likelihood estimate
- 母数空間Θ parameter space :母数がとりうる値の集合

※注意:最尤法は尤度関数を作る関係上，母集団分布がわからないときは使えない！

母平均の点推定

- **最尤法 maximum likelihood method**
  - 例：母集団分布が X=1,0 で 1をとる確率 p のベルヌーイ分布 $Bi(1,p)$ とする．母数 p を推定したい．



p を推定したい！

5つの標本をとったところ…

$$X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 1$$

$$\bar{X} = \frac{X_1 + \cdots + X_5}{5} = \frac{4}{5}$$

→ 尤度関数は

$$L(p) = p^4(1-p)$$

→ 尤度関数を最大にするpを求めると…

$$\frac{dL(p)}{dp} = p^3(4-5p) = 0$$

$$\rightarrow \hat{p} = \frac{4}{5}$$

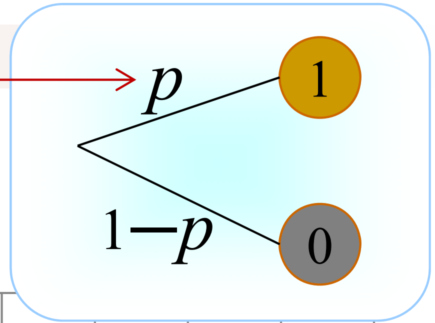
最尤推定値

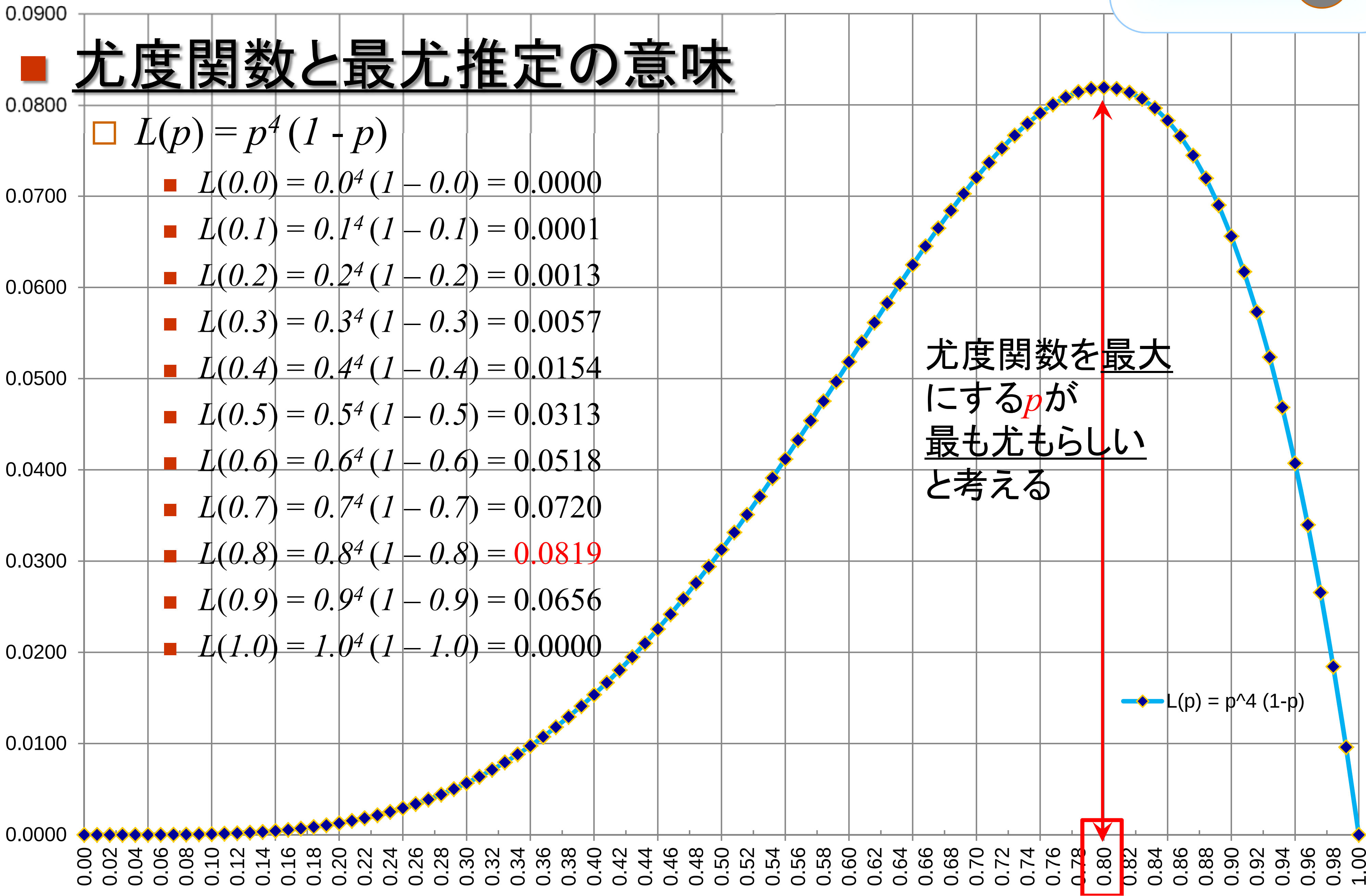
母平均の点推定

尤度関数と最尤推定の意味

- $L(p) = p^4 (1 - p)$
  - $L(0.0) = 0.0^4 (1 - 0.0) = 0.0000$
  - $L(0.1) = 0.1^4 (1 - 0.1) = 0.0001$
  - $L(0.2) = 0.2^4 (1 - 0.2) = 0.0013$
  - $L(0.3) = 0.3^4 (1 - 0.3) = 0.0057$
  - $L(0.4) = 0.4^4 (1 - 0.4) = 0.0154$
  - $L(0.5) = 0.5^4 (1 - 0.5) = 0.0313$
  - $L(0.6) = 0.6^4 (1 - 0.6) = 0.0518$
  - $L(0.7) = 0.7^4 (1 - 0.7) = 0.0720$
  - $L(0.8) = 0.8^4 (1 - 0.8) = 0.0819$
  - $L(0.9) = 0.9^4 (1 - 0.9) = 0.0656$
  - $L(1.0) = 1.0^4 (1 - 1.0) = 0.0000$

<u>p を推定したい</u>





尤度関数を<u>最大</u>
にするpが
<u>最も尤もらしい</u>
と考える

補足：母平均の点推定

点推定の基準

不偏推定量
unbiased estimator

- 不偏性
  - 推定量 $\hat{\theta}$ の期待値が，真の母数 θ の値となる性質
    - 例1：標本平均 $\overline{X}$ は母平均 μ の不偏推定量
    - 例2：標本分散 S^2 は母分散 σ^2 の不偏推定量ではない
    - 例3：不偏分散 s^2 は母分散 σ^2 の不偏推定量

$$E(\hat{\theta}) = \theta$$

一致推定量
consistent estimator

- 一致性
  - 標本数nが大きくなれば，推定量 $\hat{\theta}$ が真の母数 θ に近づく性質
    - 例1：標本平均 $\overline{X}$ は母平均 μ の一致推定量
    - 例2：標本分散 S^2 は母分散 σ^2 の一致推定量
    - 例3：不偏分散 s^2 は母分散 σ^2 の一致推定量

$$\forall \varepsilon > 0, P(|\hat{\theta}_n - \theta| > \varepsilon) \to 0$$

この2つの性質は，
推定量が最小限
満たすべき性質

モーメント法による
- 母平均の推定量 $\overline{X}$
- 母分散の推定量 S^2

補足：母平均の点推定

点推定の基準

- **漸近正規性 asymptotic normality**
  - 標本分布の漸近分布が正規分布である性質
    - 例：標本平均 $\bar{X}$ の漸近分布は，中心極限定理より，母集団分布に関係なく正規分布となる
- **有効性 efficiency**
  - 不偏性と一致性を満たす他のいかなる推定量よりも，分散が小さいという性質
    - 例：母集団分布が正規分布の場合，標本平均 $\bar{X}$ は母平均 μ の有効推定量
- **漸近有効性 asymptotic efficiency**
  - 漸近分布が正規分布となる推定量のうち，漸近分散が最小となる性質
    - 例：最尤推定量は一般に漸近有効性を持つ

漸近正規推定量
asymptotic normally estimator

有効推定量 efficient estimator
〔最小分散不偏推定量 minimum variance unbiased estimator〕

漸近的有効推定量
asymptotically efficient estimator

有効性の検証が難しいため，漸近有効性を用いる

母平均の点推定

- 例題
  - 一学年200人でテストを実施した．10人の採点をしたところで結果は以下のとおりだった．全体の平均は何点だろうか？

70	62	82	73	67	75	85	71	60	65

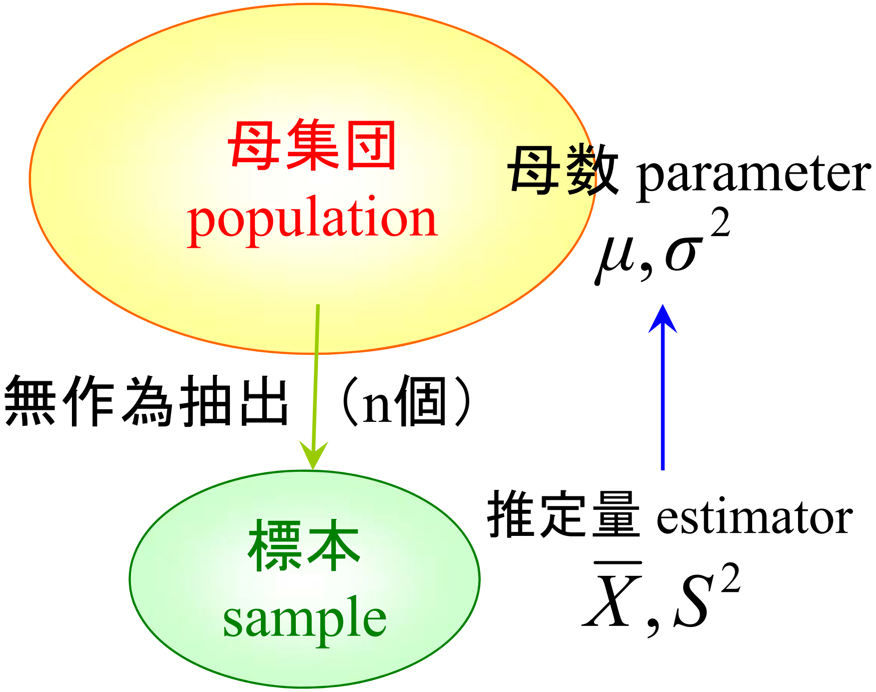
(1) 点推定で母平均を推定せよ

$$\overline{X}\text{の値}: \frac{70+62+82+73+67+75+85+71+60+65}{10} = 71.0$$

従って，母平均 $\underline{\mu = 71.0}$

母数の推定：区間推定〔interval estimation〕

■ 母平均・母分散の区間推定



- **<u>母平均μの区間推定</u>**
  - 母分散σ^2が既知の場合

→ **Z推定**（標準正規分布：$N(0,1)$）

  - 母分散σ^2が未知の場合

→ **t推定**（自由度n-1のt分布：t(n-1)）

- **<u>母分散σ^2の区間推定</u>**

→ **χ^2推定**（自由度n-1のχ^2分布：χ^2(n-1)）

母平均の区間推定

- 母平均の区間推定 …母平均の取りうる区間を推定

「母平均 μ は<u>○から△の間</u>にある」

推測の区間の幅が広ければ広いほど，
当たる可能性は高くなる

例：文教大学の男子学生の平均身長は？
「平均身長は0cm〜300cmの間にある」
「平均身長は100cm〜200cmの間にある」
「平均身長は160cm〜180cmの間にある」
「平均身長は170cm〜175cmの間にある」

⇩

推測の区間だけではなく，
推測の当たる可能性（確からしさ）も重要

「母平均 μ は<u>□%の確からしさ</u>で，
<u>○から△の間</u>にある」

信頼度（信頼係数）

信頼区間

母平均の区間推定

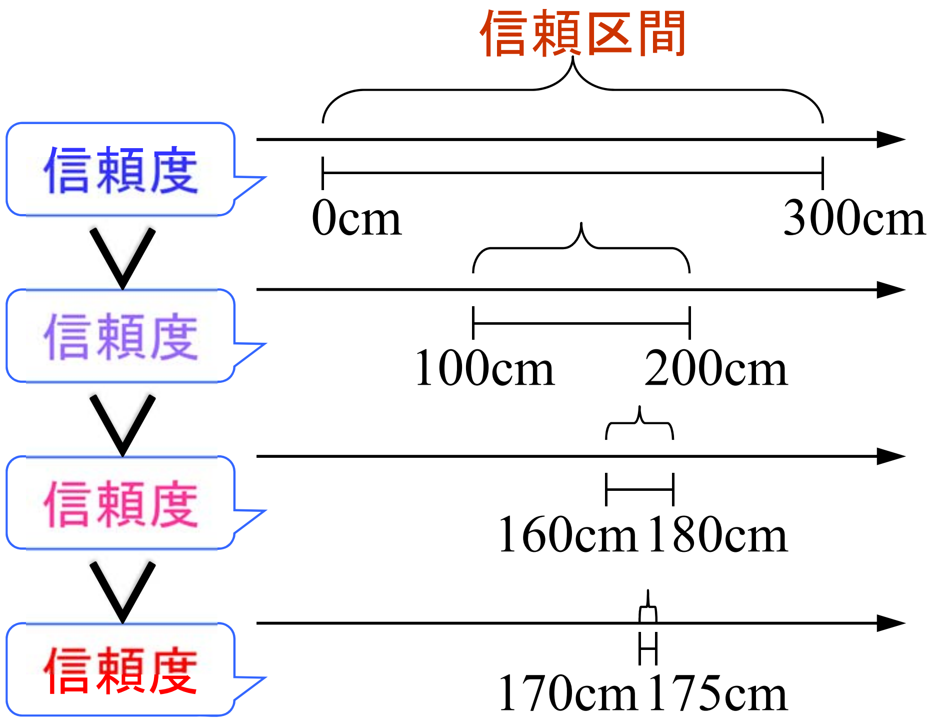
- 母平均の区間推定
  - 信頼度（信頼係数）
    - 推測した結果がどれだけ信頼できるかの目安
  - 信頼区間
    - 推測の範囲

ある程度充分な数の標本（n個）を収集し，信頼度を保ちながら，なるべく狭い信頼区間を推定したい！

信頼区間の幅が広い
⇒　推測が当たる可能性高い
⇒　信頼度が高い

信頼区間の幅が狭い
⇒　推測が当たる可能性低い
⇒　信頼度が低い

例：文教大学の男子学生の平均身長は？



母平均の区間推定

中心極限定理より　　　　標準化

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

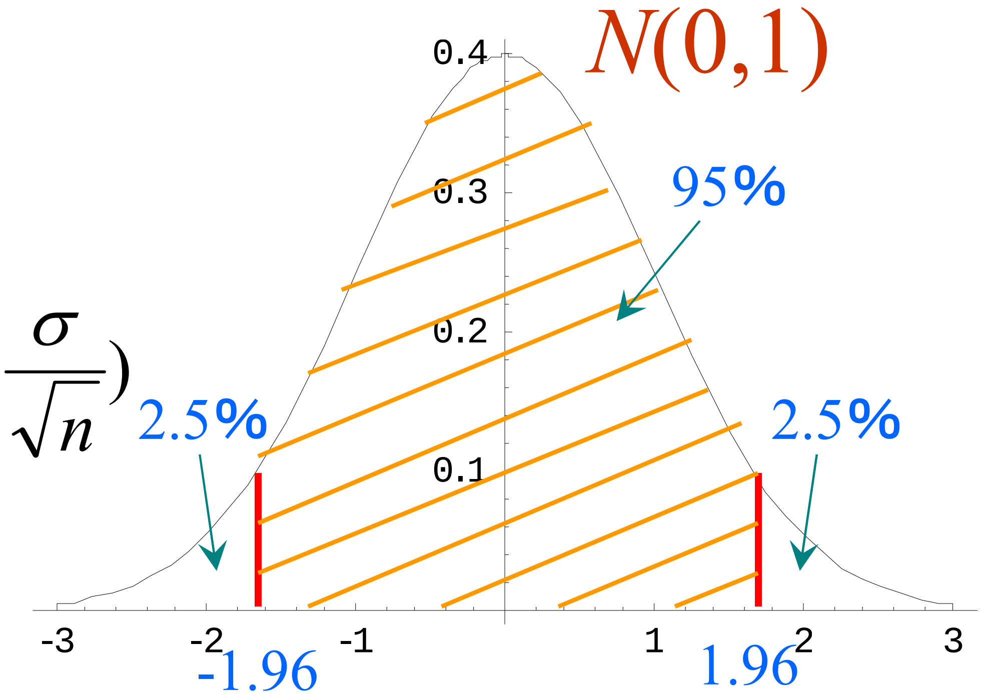
母平均の区間推定

- 標準正規分布 $N(0,1)$ に従う確率変数 Z を使う
- 標準正規分布 $N(0,1)$ に従う確率変数 Z が -1.96以上1.96以下の値をとる確率は0.95である

$$0.95 = P(-1.96 \le Z \le 1.96)$$

$$= P\left(-1.96 \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le 1.96\right)$$

$$= P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

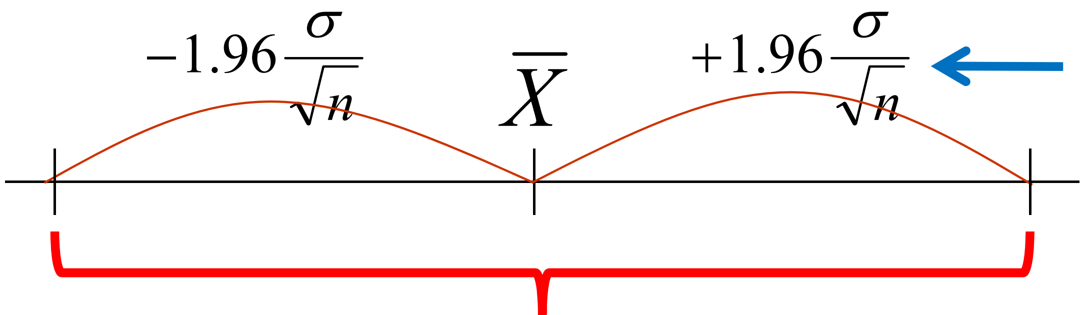


母平均の区間推定

- 母平均の区間推定(母分散が既知の場合)
  - 母平均μは信頼度95%で以下信頼区間にあると推定

$$\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

母分散σ^2がわかれば計算可能



【95%信頼区間】

μはこの区間のどこかにいる(注:どこかはわからない)

標本数(n)が分母にある，即ち，
nが大きければ，区間幅は狭くなり，
nが小さければ，区間幅は広くなる．
つまり，たくさん標本をとってくれば，同じ信頼度で区間幅を狭くできる！

注:母集団が有限(N)の場合

$$\sigma \to \sigma\sqrt{\frac{N-n}{N-1}}$$

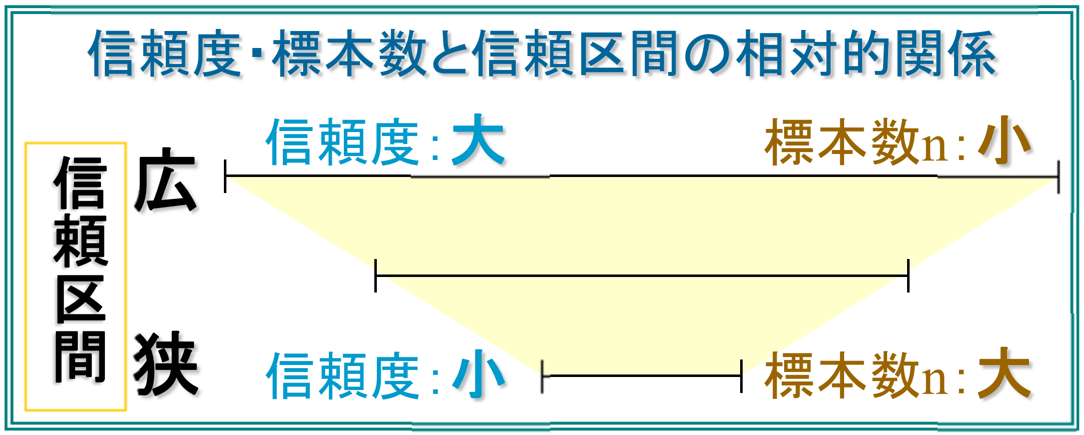
母平均の区間推定

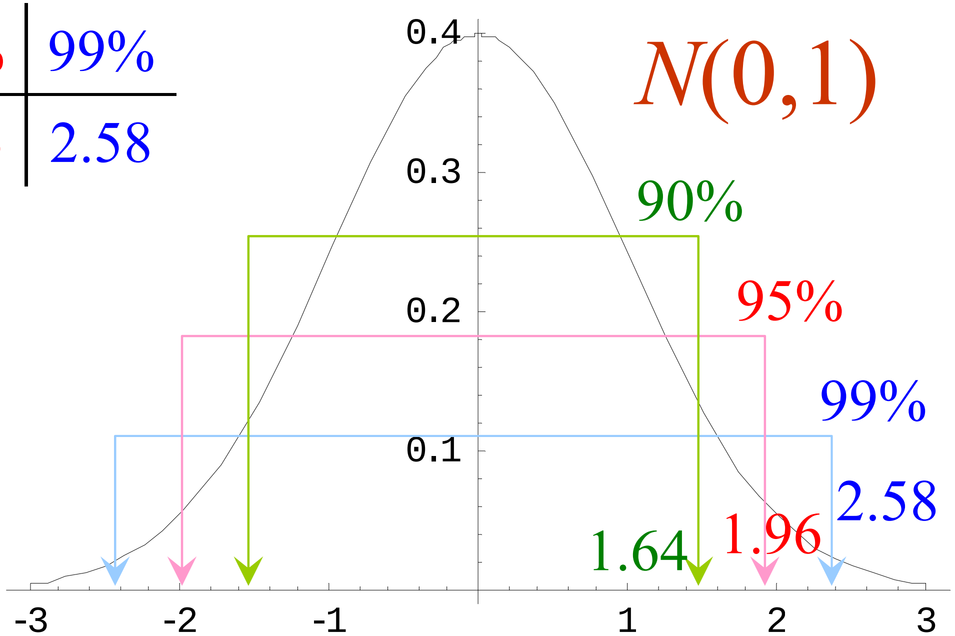
- Z推定（母分散が既知の場合）
  - 母平均μは100(1-α)%の信頼度で以下信頼区間の間にある

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

標準正規分布 $N(0,1)$ で $Z_{\alpha/2}$ のとる確率によって定まる

信頼度	90%	95%	99%
$Z_{\alpha/2}$	1.64	1.96	2.58





母平均の区間推定

- 例題
  - 一学年200人でテストを実施した．10人の採点をしたところで結果は以下のとおりだった．全体の平均は何点だろうか？

70	62	82	73	67	75	85	71	60	65

(1) 母分散が59のとき，信頼度95%で区間推定せよ

信頼度95% $\rightarrow$ $\alpha=0.05$ $\rightarrow$ $Z_{\alpha/2}=1.96$

$$\rightarrow 71.0-1.96\frac{\sqrt{59}}{\sqrt{10}} \leq \mu \leq 71.0+1.96\frac{\sqrt{59}}{\sqrt{10}}$$

$$\leftrightarrow 66.239 \leq \mu \leq 75.761$$

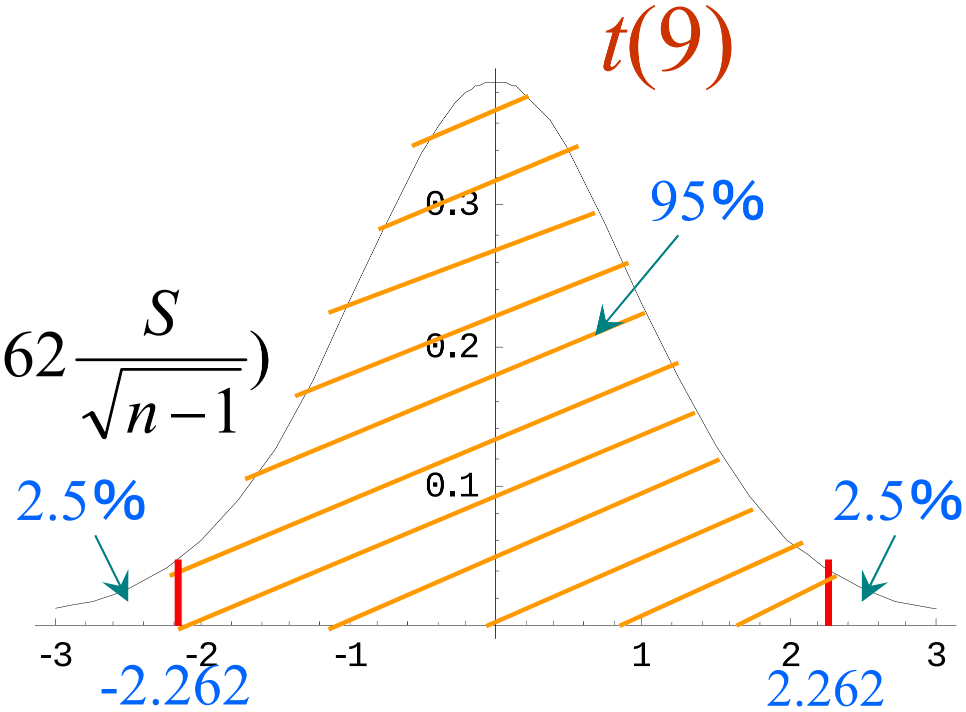
母平均の区間推定

自由度n-1の*t*分布に従う

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n-1}} \sim t(n-1)$$

- 母平均の区間推定（母分散が未知の場合）
  - 自由度n-1の*t*分布 *t*(n-1) に従う確率変数 *T* を使う
  - 標本数10のとき，自由度9の*t*分布 *t*(9)に従う確率変数 *T* が −2.262以上2.262以下の値をとる確率は0.95である

$$\begin{aligned}
0.95 &= P(-2.262 \le T \le 2.262) \\
&= P\left(-2.262 \le \frac{\overline{X} - \mu}{S/\sqrt{n-1}} \le 2.262\right) \\
&= P\left(\overline{X} - 2.262\frac{S}{\sqrt{n-1}} \le \mu \le \overline{X} + 2.262\frac{S}{\sqrt{n-1}}\right)
\end{aligned}$$



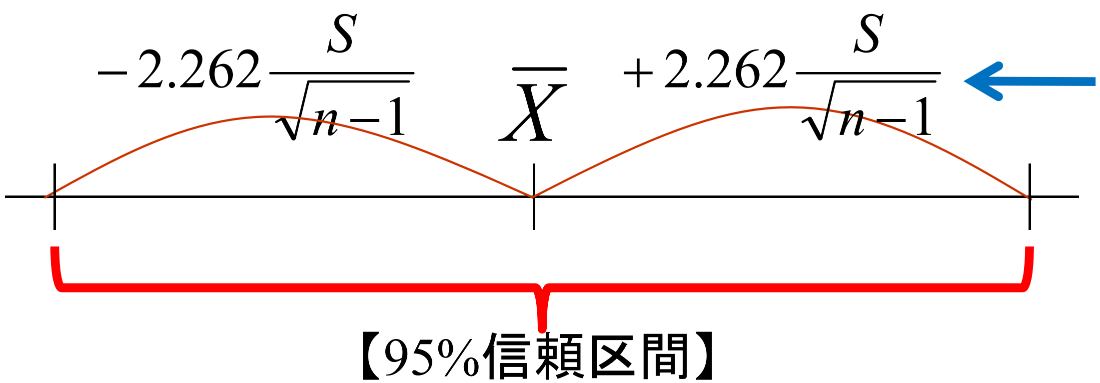
母平均の区間推定

- 母平均の区間推定（母分散が未知の場合）
  - 母平均μは信頼度95%で以下信頼区間にあると推定

（自由度9の場合）

$$\overline{X} - 2.262\frac{S}{\sqrt{n-1}} \le \mu \le \overline{X} + 2.262\frac{S}{\sqrt{n-1}}$$

標本分散S^2から計算可能



標本数(n)が分母にある，即ち，
nが大きければ，区間幅は狭くなり，
nが小さければ，区間幅は広くなる．
つまり，たくさん標本をとってくれば，同じ信頼度で区間幅を狭くできる！

μはこの区間のどこかにいる（注：どこかはわからない）

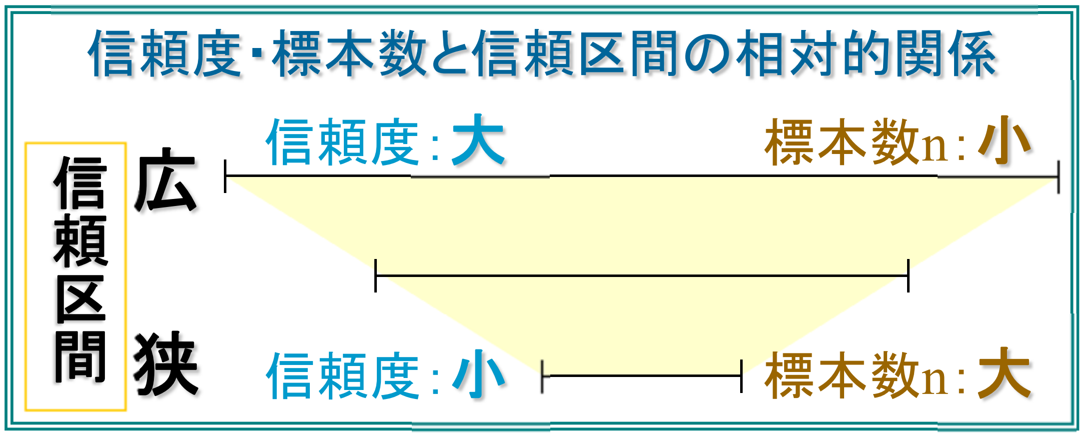
母平均の区間推定

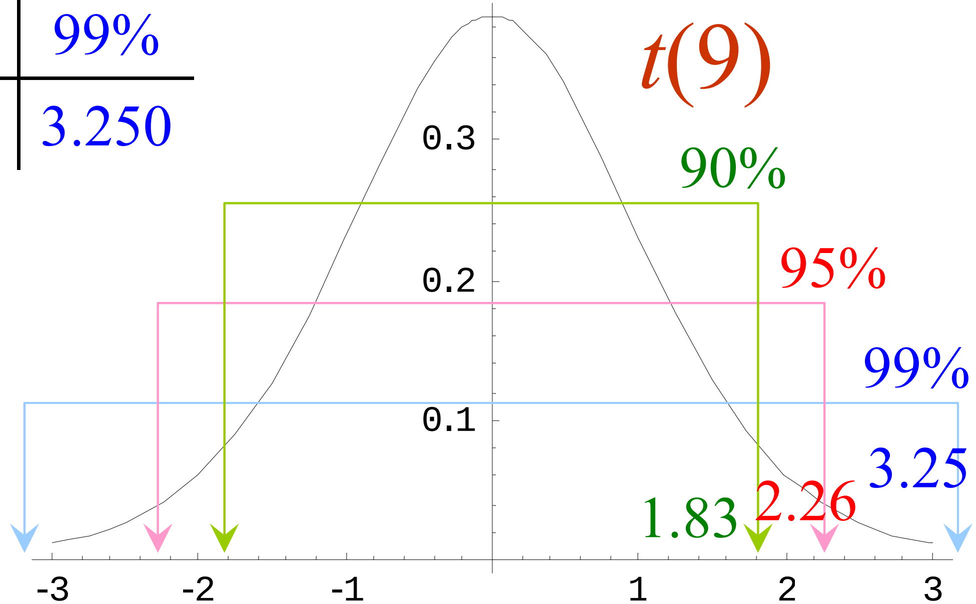
- 母平均の区間推定(母分散が未知の場合: t推定)
  - 母平均μは100(1-α)%の信頼度で以下信頼区間の間にある

$$\bar{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n-1}} \le \mu \le \bar{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n-1}}$$

自由度 n-1 の t 分布で $t_{\alpha/2}$ のとる確率によって定まる

信頼度	90%	95%	99%
$t_{\alpha/2}(9)$	1.833	2.262	3.250





母平均の区間推定

- 例題
  - 一学年200人でテストを実施した．10人の採点をしたところで結果は以下のとおりだった．全体の平均は何点だろうか？

70	62	82	73	67	75	85	71	60	65

(1) 母分散が未知のとき，信頼度90%で区間推定せよ

信頼度90% $\rightarrow$ $\alpha=0.10$ $\rightarrow$ $t_{\alpha/2}(9)=1.833$

$$\rightarrow\ 71.0-1.833\frac{\sqrt{59.20}}{\sqrt{10-1}}\le\mu\le 71.0+1.833\frac{\sqrt{59.20}}{\sqrt{10-1}}$$

$$\leftrightarrow\ 66.299\le\mu\le 75.701$$

母数(母平均)の推定:区間推定

- 演習
  - 正規母集団から標本

$$9,\quad 7,\quad 12,\quad 8,\quad 9$$

を得た.

(1) 母平均μを点推定せよ.

(2) 母分散$\sigma^2=4$の時，信頼度95%で母平均μを区間推定せよ.

(3) 母分散$\sigma^2=4$の時，信頼度99%で母平均μを区間推定せよ.

(4) 母分散が未知の時，信頼度90%で母平均μを区間推定せよ.

(5) 母分散が未知の時，信頼度95%で母平均μを区間推定せよ.

母平均の区間推定(まとめ)

- 母平均μの区間推定
  - □母分散が既知のとき

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

〔信頼率 1-α〕

$\Rightarrow$ ***Z*推定**

母分散σ^2の値が既知のときに，**標準正規分布*N*(0,1)**の性質を利用して母平均μの信頼区間を求める

  - □母分散が未知のとき

$$\bar{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n-1}} \le \mu \le \bar{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n-1}}$$

〔信頼率 1-α〕

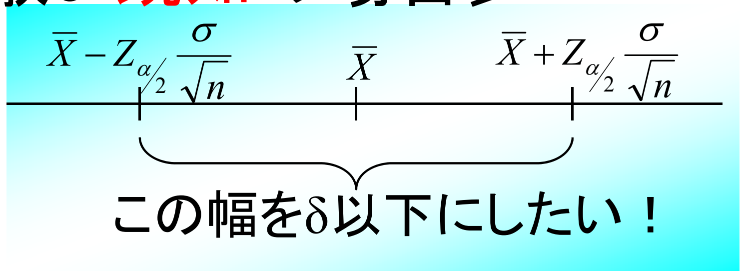
$\Rightarrow$ ***t* 推定**

母分散σ^2の値が未知のときに，**標本分散S^2**を用い，**自由度n-1の *t* 分布**の性質を利用して母平均μの信頼区間を求める

参考：母平均区間推定の標本数設計法

- 母平均μの信頼区間（信頼率1-α）〔母分散σ²既知の場合〕

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$



  - 信頼区間をδ以下に抑えるために必要な標本数の設計

$$2Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \delta \quad \Leftrightarrow \quad n \ge \frac{4Z_{\alpha/2}^2\sigma^2}{\delta^2}$$

よって，標本数 n を
この数以上にすればよい．

  - 例題：全国男子大学生の平均身長を区間推定したい．95%信頼区間を2cm以下にするには，何人の学生を調査すればよいか？　ただし，母分散はσ^2=49とする．

$$n \ge \frac{4\cdot(1.96)^2\cdot 49}{2^2} = 188.2384$$

よって，n=189人を調べれば充分

参考：母平均区間推定の標本数設計法

- 母平均μの信頼区間（信頼率1-α）〔母分散σ^2未知の場合〕

$$\bar{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n-1}} \le \mu \le \bar{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n-1}}$$

$\bar{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n-1}}$　$\bar{X}$　$\bar{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n-1}}$

この幅をδ以下にしたい！

  - 信頼区間をδ以下に抑えるために必要な標本数の設計

区間幅 $2t_{\alpha/2}(n-1)\frac{S}{\sqrt{n-1}}$ をδ以下にすればよいが，確率変数Sが含まれているので，区間幅の期待値をδ以下に抑える.

$$2t_{\alpha/2}(n-1)\frac{E(S)}{\sqrt{n-1}} \le \delta$$

$E(S)$は未知母数σに依存するので，何らかの情報からσを想定し，標本数 n を設定することになる.

$$n \ge 1 + \frac{4 \cdot t_{\alpha/2}(n-1)^2 \cdot \left(E(S)\right)^2}{\delta^2}$$

$E(S^2) = \dfrac{n-1}{n}\sigma^2$ だが　$\left(E(S^2) = \dfrac{N}{N-1} \cdot \dfrac{n-1}{n}\sigma^2 \right)$ 有限母集団の場合

$E(S) \ne \sqrt{\dfrac{n-1}{n}}\sigma$ であることに注意

母数(母分散)の推定:区間推定

$$\chi^2 = \frac{nS^2}{\sigma^2} \sim \chi^2(n-1)$$

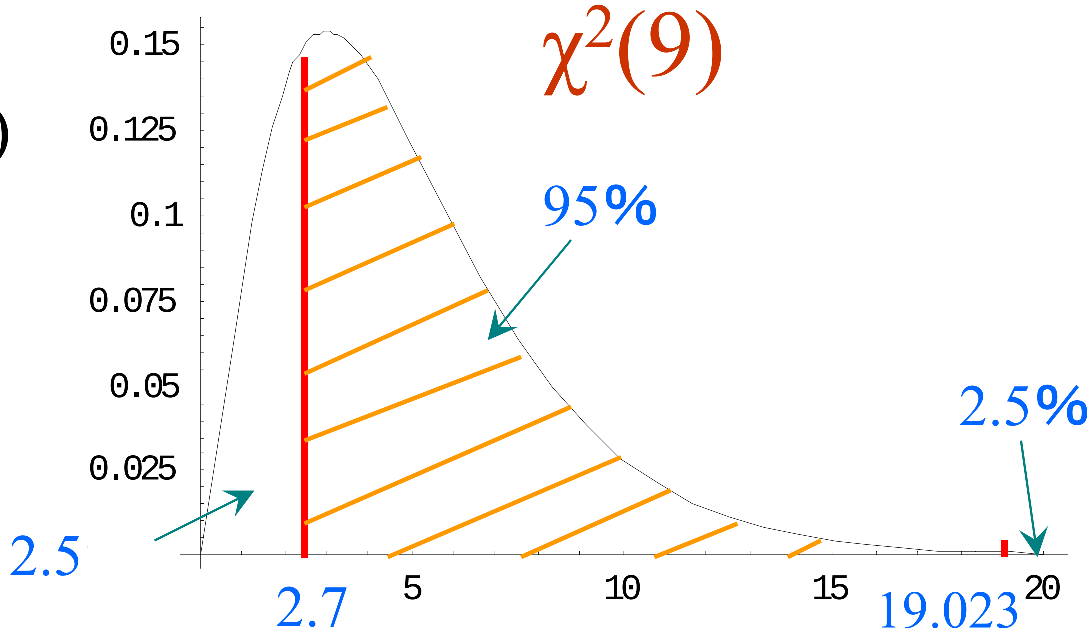
■ 母分散の区間推定

- □ 自由度n-1のχ^2分布に従う確率変数χ^2を使う
- □ 自由度9のχ^2分布に従う確率変数χ^2が−2.700以上19.023以下の値をとる確率は0.95

⇩

$$\begin{aligned} 0.95 &= P(2.700 \le \chi^2 \le 19.023) \\ &= P(2.700 \le \frac{nS^2}{\sigma^2} \le 19.023) \\ &= P(\frac{nS^2}{19.023} \le \sigma^2 \le \frac{nS^2}{2.700}) \end{aligned}$$

注:χ^2分布は左右対称ではないので,左右各々の裾の面積が0.025となる点を考える必要がある.



母数(母分散)の推定:区間推定

- 母分散の区間推定
  - 母分散σ^2は95%の信頼度で以下の信頼区間の間にあると推測できる!(自由度9の時)

$$2.700 \le \frac{nS^2}{\sigma^2} \le 19.023$$

$$\Leftrightarrow \frac{nS^2}{19.023} \le \sigma^2 \le \frac{nS^2}{2.700}$$

標本分散 S^2から計算できる

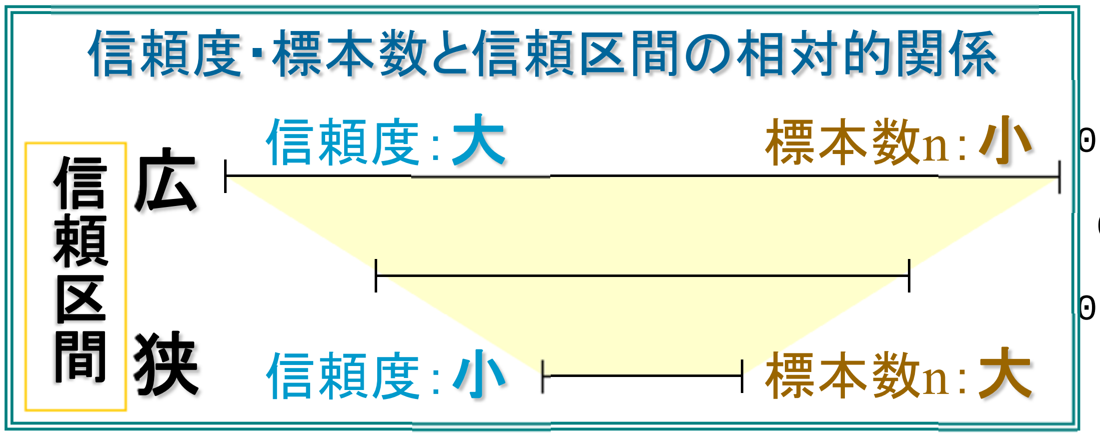
母数(母分散)の推定:区間推定

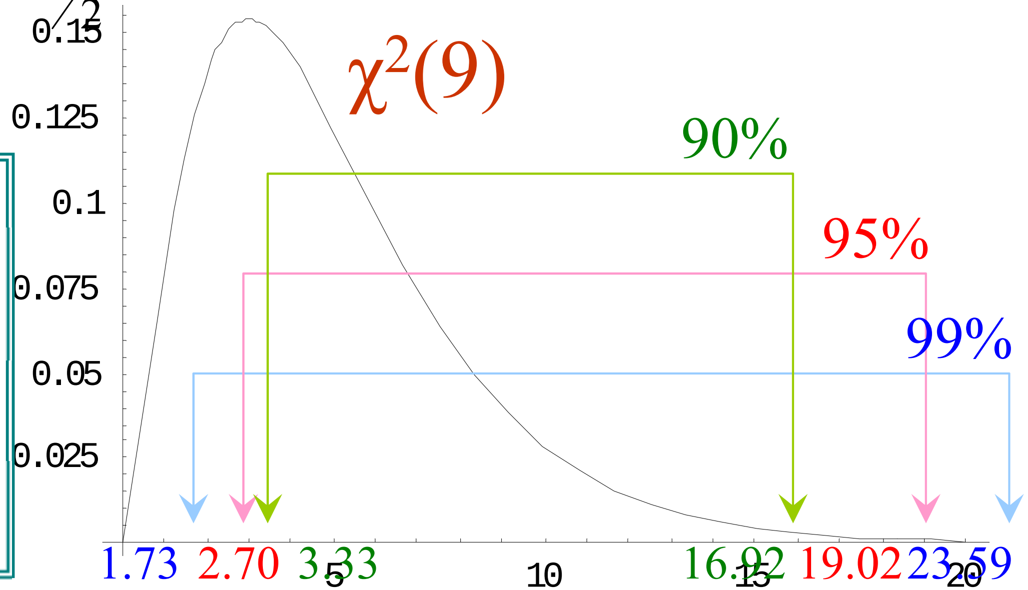
■ 母分散の区間推定(χ^2 推定)

□ 母分散σ^2が$100(1-\alpha)$%の信頼度で以下信頼区間の間

$$\frac{nS^2}{\chi^2_{\alpha/2}(n-1)} \le \sigma^2 \le \frac{nS^2}{\chi^2_{1-\alpha/2}(n-1)}$$

自由度n-1のχ^2分布で $\chi^2_{1-\alpha/2}, \chi^2_{\alpha/2}$ のとる確率によって定まる





母数(母分散)の推定:区間推定

■ 例題

- □ 一学年200人でテストを実施した. 10人の採点をしたところで結果は以下のとおりだった. 全体の平均は何点だろうか？

70	62	82	73	67	75	85	71	60	65

(1) 信頼度95%で区間推定せよ

信頼度95% $\rightarrow$ $\alpha=0.05$ $\rightarrow$ $\chi^2_{1-\alpha/2}(9)=2.70039, \chi^2_{\alpha/2}(9)=19.0228$

$$\frac{nS^2}{\chi^2_{\alpha/2}(n-1)} \le \sigma^2 \le \frac{nS^2}{\chi^2_{1-\alpha/2}(n-1)} \rightarrow \frac{10\cdot 59.2}{19.0228} \le \sigma^2 \le \frac{10\cdot 59.2}{2.70039}$$

$$\leftrightarrow 31.12055 \le \sigma^2 \le 219.2276$$

$$\rightarrow 5.57858 \le \sigma \le 14.80634$$

母数(母分散)の推定:区間推定

- 演習 (出展:「確率・統計の仕組みがわかる本」技評p.367)
  - 養鶏場における卵の重さのばらつきを調べたい. 無作為に16個の卵を抽出したときの重さは下表のとおりとなった.

46	52	54	46	51	47	52	44
50	53	48	51	48	49	54	47

(1) 信頼度90%で母分散σ^2を区間推定せよ.

(2) 信頼度95%で母分散σ^2を区間推定せよ.

(3) 信頼度99%で母分散σ^2を区間推定せよ.

母数の推定：区間推定

- 演習 (参考：「統計学入門」東大出版会 p.231)
  - 東京都の2005年11月1日～10日までの最高気温，最低気温は下表のとおりであった．正規母集団を仮定する．

日にち	11/1	11/2	11/3	11/4	11/5	11/6	11/7	11/8	11/9	11/10
最高気温(℃)	17	19	19	21	21	16	24	22	19	18
最低気温(℃)	10	10	12	12	13	13	13	12	10	10

（データ：「Yahoo!天気情報」より）

(1) 最高気温について，信頼度99%で母平均μの信頼区間を求めよ．

(2) 最高気温について，信頼度95%で母分散σ^2の信頼区間を求めよ．

(3) 最低気温について，信頼度95%で母平均μの信頼区間を求めよ．

(3) 最低気温について，信頼度90%で母分散σ^2の信頼区間を求めよ．

母数（母分散）の推定：区間推定（まとめ）

- 母分散の区間推定
  - χ^2推定

自由度n-1のχ^2分布の性質を利用して母分散σ^2の信頼区間を求める

$$\frac{nS^2}{\chi^2_{\alpha/2}(n-1)} \leq \sigma^2 \leq \frac{nS^2}{\chi^2_{1-\alpha/2}(n-1)}$$

〔信頼率 1-α〕

母数（母比率）の推定：区間推定

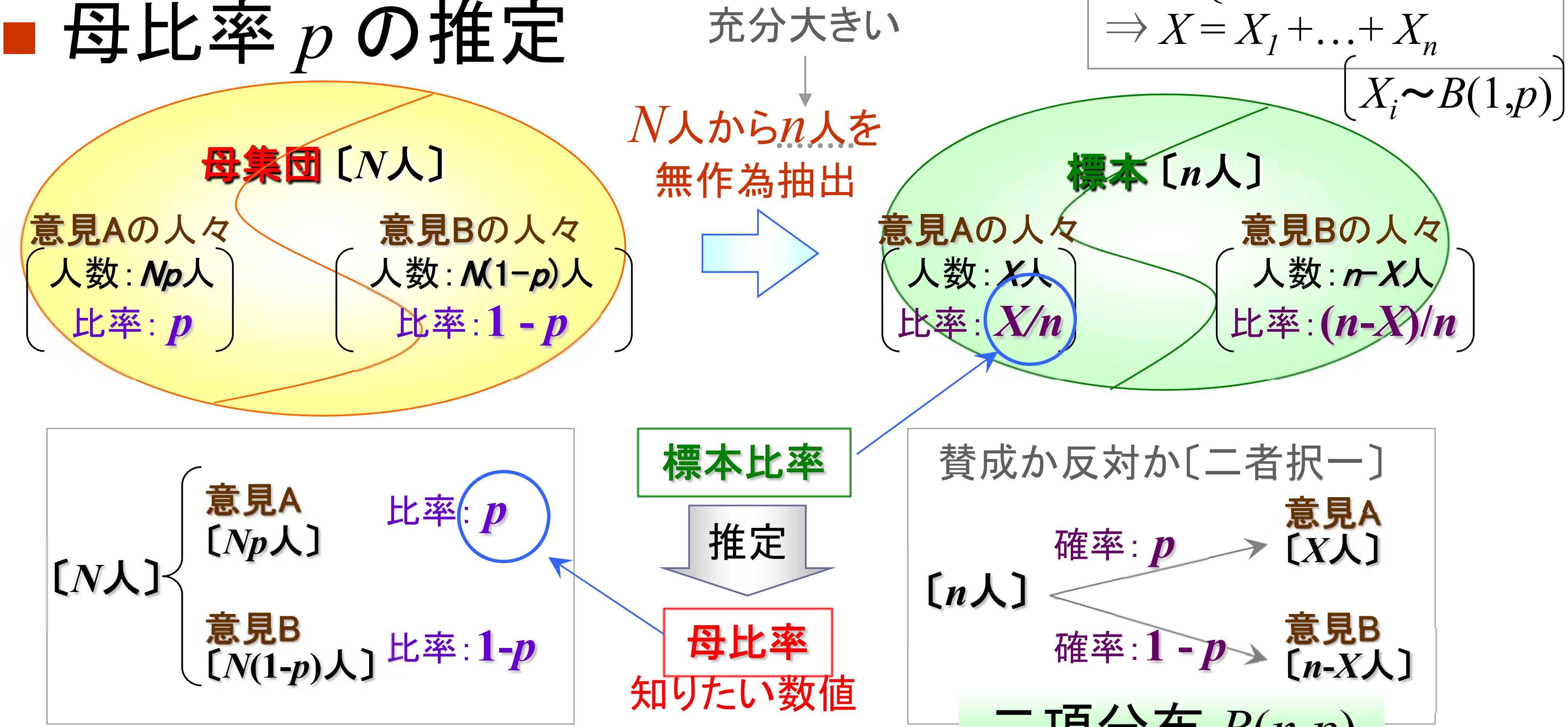
第i番目の人について

$$X_i = \begin{cases} 1 & \text{（意見Aである）} \\ 0 & \text{（意見Bである）} \end{cases}$$

$$\Rightarrow X = X_1 + \ldots + X_n$$

$X_i \sim B(1,p)$

■ 母比率 p の推定



二項分布 $B(n,p)$

（Xは二項分布$B(n,p)$に従う）

$$X \sim B(n,p) \quad \begin{cases} E(X) = np \\ V(X) = np(1-p) \end{cases}$$

中心極限定理

$$X \sim N(\, np,\, np(1\text{-}p)\,)$$

（X は正規分布$N(np,\, np(1\text{-}p))$に従う）

母数(母比率)の推定：区間推定

第i番目の人について

$$X_i=\begin{cases}1 & (\text{意見Aである})\\ 0 & (\text{意見Bである})\end{cases}$$

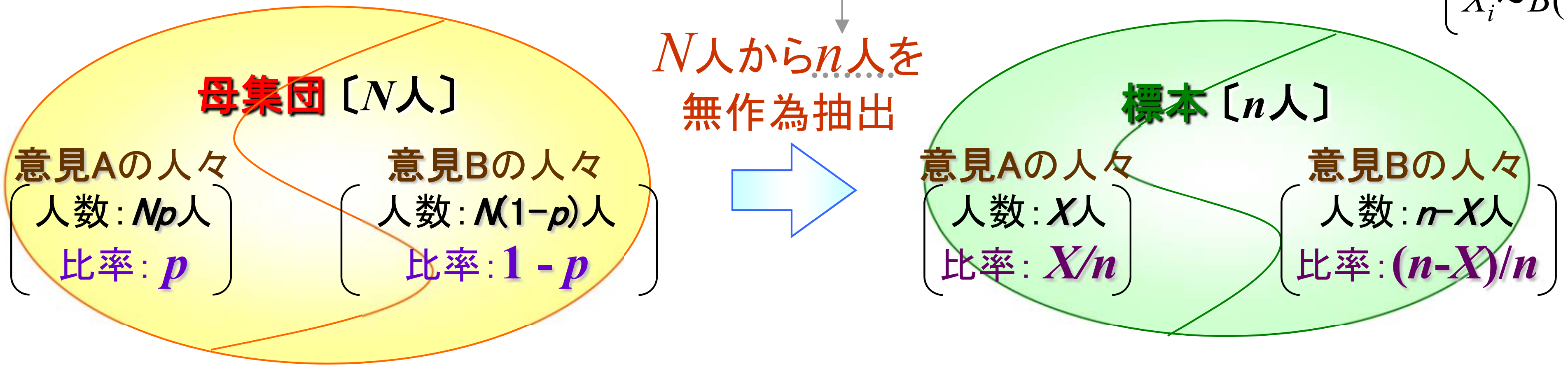
$\Rightarrow X=X_1+\ldots+X_n$

$\left(X_i \sim B(1,p)\right)$

■ 母比率 p の推定

充分大きい

N人からn人を無作為抽出



- X は 二項分布 $B(n,p)$ に従う　　$X\sim B(n,p)$　$\begin{cases}E(X)=np\\ V(X)=np(1-p)\end{cases}$

中心極限定理

- X は 正規分布 $N(np,\ np(1\text{-}p))$ に従う　　$X\sim N(np,\ np(1\text{-}p))$

- X/n は 正規分布 $N(p, p(1\text{-}p)/n)$ に従う　　$X/n\sim N(p, p(1\text{-}p)/n)$　$\begin{cases}E(\frac{X}{n})=\frac{1}{n}E(X)=p\\ V(\frac{X}{n})=\frac{1}{n^2}V(X)=\frac{p(1-p)}{n}\end{cases}$

$P:=X/n$

- Z は 正規分布 $N(0,1)$ に従う　　$Z=\dfrac{P-p}{\sqrt{p(1-p)/n}}\sim N(0,1)$　$\begin{cases}E(Z)=0\\ V(Z)=1\end{cases}$

標準化：平均を引いて標準偏差で割る

母数(母比率)の推定:区間推定

■ 母比率 p の推定

□ 母比率 p の信頼度 100(1-α)% の信頼区間

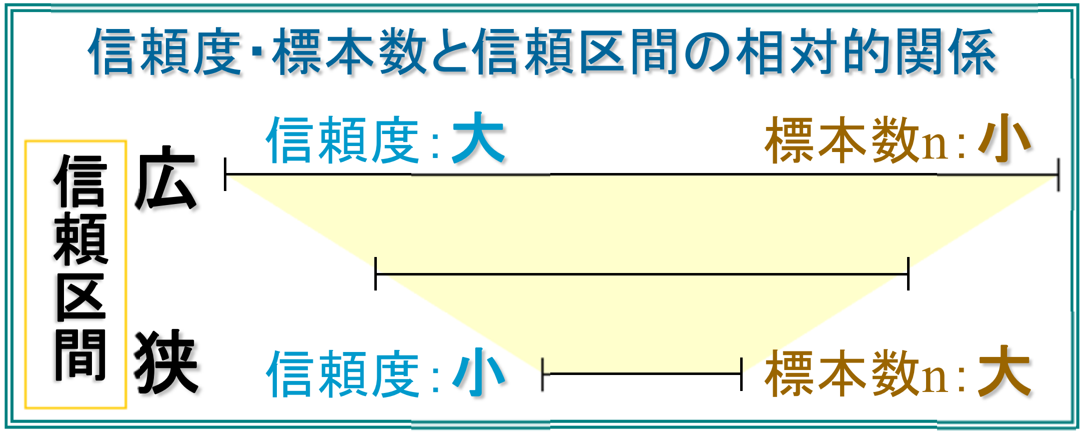
$$P - Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}} \le p \le P + Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$$

注:標本数 n が充分大きいときの信頼区間.
nが小さいときは,修正式が提案されている.

標準正規分布 $N(0,1)$ で $Z_{\alpha/2}$ のとる確率によって定まる

式中のPは標本比率で,$P:=X/n$ である

信頼度	90%	95%	99%
$Z_{\alpha/2}$	1.64	1.96	2.58



注:点推定の場合

母比率は X/n と推定

母数(母比率)の推定:区間推定

- 例題(出展:「図解雑学 統計解析」ナツメ社 p.170)
  - ある新聞社による内閣支持率調査では3000人の対象者のうち1674人が現行内閣を指示すると回答した. この国の内閣支持率はどのぐらいだろうか？　信頼度95%で母比率 p の区間推定をしよう.
    - 標本比率:(標本平均) $P=\dfrac{X}{n}=\dfrac{1674}{3000}=0.558$
    - 信頼度95%とはα=0.05, 即ち, $Z_{0.05/2}$=1.96
    - 信頼区間:

$$P-Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}\le p\le P+Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$$

$$\to 0.558-1.96\sqrt{\frac{0.558(1-0.558)}{3000}}\le p\le 0.558+1.96\sqrt{\frac{0.558(1-0.558)}{3000}}$$

$$\leftrightarrow 0.540229\le p\le 0.575771$$

    - 故に, 内閣支持率は, 信頼度95%で 54.0%～57.6%の間にある.

母数(母比率)の推定:区間推定

- 演習(出展:「確率・統計の仕組みがわかる本」技評p.375)
  - ある薬を常用している妊婦は女の子を産む確率が高いらしい．該当者のうち200人を調査したところ，赤ちゃんの124人が女の子だった．この薬を常用している妊婦が女の子を産む比率はどの程度か？

(1) 信頼度90%で母比率 p の区間推定をせよ

(2) 信頼度95%で母比率 p の区間推定をせよ

(3) 信頼度99%で母比率 p の区間推定をせよ

母数(母比率)の推定：区間推定(まとめ)

- 母比率の区間推定
  - Z推定

標準正規分布$N(0,1)$の性質を利用して母平均μの信頼区間を求める

$$P - Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}} \le p \le P + Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$$

〔信頼率 1-α〕

式中のPは標本比率で，$P:=X/n$ である

（Xは 二項分布 $B(n,p)$ に従うが，中心極限定理から
Zは 正規分布 $N(0,1)$ に従う）

参考：母比率推定に必要な標本数

- 適切な標本数
  - 母比率推定における信頼区間の幅と上限

$$2Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}} \le \frac{Z_{\alpha/2}}{\sqrt{n}} \qquad \left[\because P(1-P) = -(P-1/2)^2 + 1/4\right]$$

  - 信頼区間の幅をβ%以内にしたい場合の標本数

$$\frac{Z_{\alpha/2}}{\sqrt{n}} \le \beta \quad \Leftrightarrow \quad n \ge \frac{Z_{\alpha/2}^2}{\beta^2}$$

- 例題：信頼度95%の信頼区間の幅を5%以内にしたい場合

$$n \ge \frac{Z_{\alpha/2}^2}{\beta^2} = \frac{1.96^2}{0.05^2} = 1536.64$$

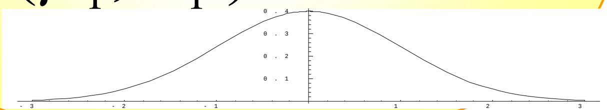
より，標本数は1537あれば充分．

二つの正規母集団の推定

- **母平均の差**の区間推定
  - 母分散が等しいとき（$\sigma_1^2=\sigma_2^2$）
  - 母分散が等しくないとき（$\sigma_1^2\neq\sigma_2^2$）
- **母分散の比**の区間推定

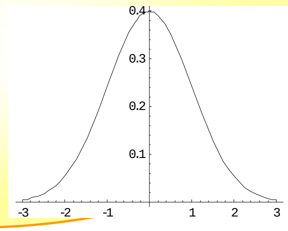
母集団1

$N(\mu_1,\sigma_1^2)$



母集団2

$N(\mu_2,\sigma_2^2)$



m個 無作為抽出

$X_1,X_2\cdots,X_m$

n個 無作為抽出

$Y_1,Y_2\cdots,Y_n$

標本1

標本平均：$\overline{X}\sim N(\mu_1,\frac{\sigma_1^2}{m})$

標本分散：$S_X{}^2$

標本2

標本平均：$\overline{Y}\sim N(\mu_2,\frac{\sigma_2^2}{n})$

標本分散：$S_Y{}^2$

二つの正規母集団の平均値の差の推定

- 母平均の差の区間推定（$\sigma_1^2 = \sigma_2^2$ のとき）
  - 母集団分布が $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ である2つの正規母集団から，個別に2つの標本 $X_1, \ldots, X_m$ と $Y_1, \ldots, Y_n$ を抽出したときの，母平均の差 $\theta = \mu_1 - \mu_2$ の信頼度100(1-α)%の信頼区間

$$\bar{X} - \bar{Y} - t_{\alpha/2}(m+n-2)\sqrt{\frac{mS_X^2 + nS_Y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}$$

自由度 $m+n-2$ の t 分布

$$\leq \theta \leq$$

$$\bar{X} - \bar{Y} + t_{\alpha/2}(m+n-2)\sqrt{\frac{mS_X^2 + nS_Y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}$$

注：ただし，先に $\sigma_1^2 = \sigma_2^2$ の検定を行う必要がある．

二つの正規母集団の平均値の差の推定

- 母平均の差の区間推定($\sigma_1^2 \neq \sigma_2^2$のとき)
  - 母集団分布が$N(\mu_1,\sigma_1^2),N(\mu_2,\sigma_2^2)$である2つの正規母集団から，個別に2つの標本 $X_1,\ldots,X_m$ と $Y_1,\ldots,Y_n$ を抽出したときの，母平均の差 $\theta=\mu_1-\mu_2$ の信頼度100(1-α)%の信頼区間

$$\bar{X}-\bar{Y}-t_{\alpha/2}(\nu)\sqrt{\frac{S_X^2}{\nu_1}+\frac{S_Y^2}{\nu_2}} \le \theta \le \bar{X}-\bar{Y}+t_{\alpha/2}(\nu)\sqrt{\frac{S_X^2}{\nu_1}+\frac{S_Y^2}{\nu_2}}$$

自由度νの t 分布

ただし，νは $\left(\frac{S_X^2}{\nu_1}+\frac{S_Y^2}{\nu_2}\right)^2 \Big/ \left(\frac{S_X^4}{\nu_1^3}+\frac{S_Y^4}{\nu_2^3}\right)$ に一番近い整数であり，

$\nu_1=m-1,\nu_2=n-1$

二つの正規母集団の平均値の差の推定

- **例題**（出展:「統計学入門」東大出版会 p.231）
  - 20匹のラットを10匹ずつ2群に分け，一方は普通の食餌，他方は血中の赤血球数を減らすと考えられる薬を混入した食餌を与えた．その結果，各群のラットの血液1mm³中の赤血球数が下表のようになった．この薬の効果を測定したい．

投薬群（100万個）	7.97	7.66	7.59	8.44	8.05	8.08	8.35	7.77	7.98	8.15
対照群（100万個）	8.06	8.27	8.45	8.05	8.51	8.14	8.09	8.15	8.16	8.42

薬の効果（平均の差）を信頼度95%で区間推定をする．母分散は等しいと仮定．

$$\begin{cases} \overline{X} = 8.004, \overline{Y} = 0.230 \rightarrow \overline{X} - \overline{Y} = -0.226 \\ S_X^2 = 0.0685, S_Y^2 = 0.0264 \end{cases}$$

信頼度95% $\rightarrow$ α=0.05 $\rightarrow$ $t_{0.025}(18)=2.101$

$$\overline{X} - \overline{Y} - t_{\alpha/2}(m+n-2)\sqrt{\frac{mS_X^2 + nS_Y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)} \le \mu_1 - \mu_2 \le \overline{X} - \overline{Y} + t_{\alpha/2}(m+n-2)\sqrt{\frac{mS_X^2 + nS_Y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}$$

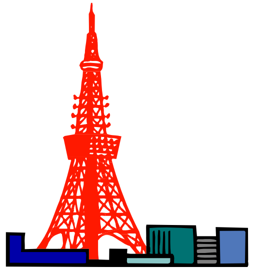
$$\rightarrow -0.226 \pm 2.101\sqrt{\frac{10\cdot 0.0685 + 10\cdot 0.0685}{10+10-2}\left(\frac{1}{10}+\frac{1}{10}\right)}$$

$$\rightarrow -0.226 - 0.216 \le \mu_1 - \mu_2 \le -0.226 + 0.216$$

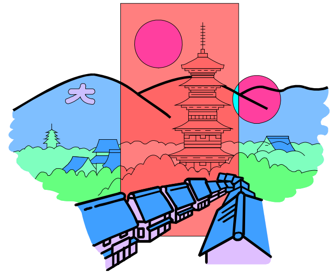
$$\leftrightarrow -0.0101 \le \mu_1 - \mu_2 \le -0.442$$

二つの正規母集団の平均値の差の推定

- 例題（参考：「統計学入門」東大出版会 p.228）
  - 京都は東京より暑いか？



v.s.



日付	8/1	8/2	8/3	8/4	8/5	8/6	8/7	8/8	8/9	8/10
東京（℃）	32	31	32	35	35	34	33	32	32	30
京都（℃）	35	35	35	36	36	33	35	36	35	35

各観測値が対として対応

（2005年8月1日～10日の東京と京都の最高気温：「Yahoo!天気情報」より）

対標本paired sample の場合は，2標本 t 統計量ではなく，差で1標本推定を行う．

東京-京都	-3	-4	-3	-1	-1	1	-2	-4	-3	-5

t 推定

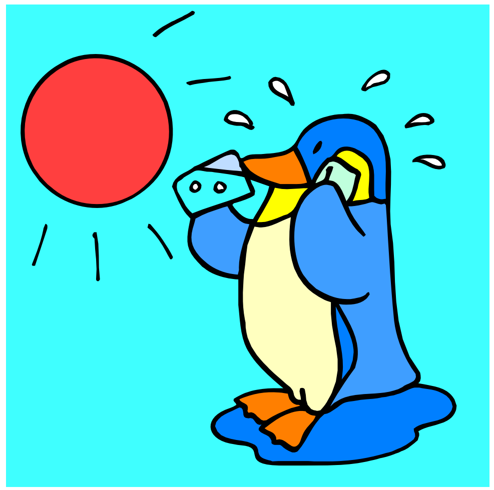
信頼度95%で母平均μの区間推定をすると…

信頼度95% $\rightarrow$ $\alpha=0.05$ $\rightarrow$ $t_{0.025}(9)=2.262$

$$\rightarrow -2.5-2.262\cdot\frac{1.688}{\sqrt{10-1}} \le \mu \le -2.5+2.262\cdot\frac{1.688}{\sqrt{10-1}}$$

$$\leftrightarrow -3.77 \le \mu \le -1.23$$

京都の方が東京より，信頼度95%で夏の最高気温平均が1.2～3.8℃の間にある．　〔→ 平均値の差の検定〕



二つの正規母集団の分散値の比の推定

母分散の比の区間推定

- 母集団分布が $N(\mu_1,\sigma_1^2), N(\mu_2,\sigma_2^2)$ である2つの正規母集団から，個別に2つの標本 $X_1,\ldots,X_m$ と $Y_1,\ldots,Y_n$ を抽出したときの，母分散の比 $\theta=\sigma_2^2/\sigma_1^2$ の信頼度100(1-α)%の信頼区間

$$\frac{1}{F_{\alpha/2}(n-1,m-1)}\frac{\frac{n}{n-1}S_2^2}{\frac{m}{m-1}S_1^2}\le\frac{\sigma_2^2}{\sigma_1^2}\le F_{\alpha/2}(m-1,n-1)\frac{\frac{n}{n-1}S_2^2}{\frac{m}{m-1}S_1^2}$$

自由度 (n-1, m-1) の F 分布

自由度 (m-1, n-1) の F 分布

$$\left(F_{1-\alpha/2}(m-1,n-1)=\frac{1}{F_{\alpha/2}(n-1,m-1)}\right)$$

母分散の不偏推定値: $\hat{\sigma}^2:=\frac{n}{n-1}S^2$

$$F:=\frac{\hat{\sigma}_1^2/\sigma_1^2}{\hat{\sigma}_2^2/\sigma_2^2}=\frac{\frac{m}{m-1}S_1^2\Big/\sigma_1^2}{\frac{n}{n-1}S_2^2\Big/\sigma_2^2}=\frac{\frac{\chi_1^2}{m-1}}{\frac{\chi_2^2}{n-1}}\sim F(m-1,n-1)$$

$$\left(\because\right)\chi^2=\frac{nS^2}{\sigma^2}$$

$$P(\mathrm{O}\le F\le\mathrm{O})=1-\alpha$$

$$\leftrightarrow F_{1-\alpha/2}(m-1,n-1)\le F\le F_{\alpha/2}(m-1,n-1)$$

$$\leftrightarrow F_{1-\alpha/2}(m-1,n-1)\le\frac{\frac{m}{m-1}S_1^2\Big/\sigma_1^2}{\frac{n}{n-1}S_2^2\Big/\sigma_2^2}\le F_{\alpha/2}(m-1,n-1)$$

$$\leftrightarrow F_{1-\alpha/2}(m-1,n-1)\frac{\frac{n}{n-1}S_2^2}{\frac{m}{m-1}S_1^2}\le\frac{\sigma_2^2}{\sigma_1^2}\le F_{\alpha/2}(m-1,n-1)\frac{\frac{n}{n-1}S_2^2}{\frac{m}{m-1}S_1^2}$$

二つの正規母集団の分散値の比の推定

- **例題**(出展:「なるほど統計学」海鳴社 p.101)
  - 某町工場では，技能オリンピック出場者を決める所である．Alpha君，Bravoさんの2人のうち，どちらかを派遣したいので，最近の2人の仕事ぶりから技能を評価する．旋盤工工員である彼らが行った30㎜のパイプ加工の品質検査をした結果以下の通りであった．どちらが優れているのだろうか？

工員	標本数	平均値(㎜)	標準偏差(㎜)
Alpha	4	30	2
Bravo	10	30	3

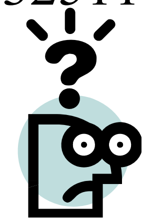
注:腕のいい旋盤工は，実際にはこんなにずれないそうです．

信頼度90%で各々の標準偏差を区間推定すると…

$$\frac{nS^2}{\chi^2_{0.05}(3)} \le \sigma_1^2 \le \frac{nS^2}{\chi^2_{0.95}(3)} \to \frac{4\cdot 4}{7.81473} \le \sigma_1^2 \le \frac{4\cdot 4}{0.351846} \to (1.43)^2 \le \sigma_1^2 \le (6.74)^2$$

$$\frac{nS^2}{\chi^2_{0.05}(9)} \le \sigma_2^2 \le \frac{nS^2}{\chi^2_{0.95}(9)} \to \frac{10\cdot 9}{16.919} \le \sigma_2^2 \le \frac{10\cdot 9}{3.32511} \to (2.31)^2 \le \sigma_2^2 \le (5.20)^2$$

いっそのこと分散比を区間推定しよう！



結局どっちが優秀なの？

二つの正規母集団の分散値の比の推定

- **例題**（出展：「なるほど統計学」海鳴社 p.101)

信頼度90%で分散比を区間推定すると…

$$\frac{1}{F_{\alpha/2}(n-1,m-1)}\frac{\frac{n}{n-1}S_2^2}{\frac{m}{m-1}S_1^2} \le \frac{\sigma_2^2}{\sigma_1^2} \le F_{\alpha/2}(m-1,n-1)\frac{\frac{n}{n-1}S_2^2}{\frac{m}{m-1}S_1^2}$$

$$\rightarrow \frac{1}{F_{0.05}(9,3)}\frac{\frac{10}{10-1}9}{\frac{4}{4-1}4} \le \frac{\sigma_2^2}{\sigma_1^2} \le F_{0.05}(3,9)\frac{\frac{10}{10-1}9}{\frac{4}{4-1}4}$$

$$\leftrightarrow (0.46)^2 \le \frac{\sigma_2^2}{\sigma_1^2} \le (2.69)^2$$

$$\begin{cases} 1 \le \dfrac{\sigma_2}{\sigma_1} & :\text{Aの方がバラツキが小さい} \rightarrow \text{Aの方が優秀} \\ \dfrac{\sigma_2}{\sigma_1} \le 1 & :\text{Bの方がバラツキが小さい} \rightarrow \text{Bの方が優秀} \end{cases}$$

結局このデータだけからは，どちらが優秀かはわからない．

優劣を判定するには，もう少しデータを集める必要があるよ．

参考文献

- 東大教養学部統計学教室編 「統計学入門」東大出版会(1991)
- 東大教養学部統計学教室編 「自然科学の統計学」東大出版会(1992)
- 鈴木達三・高橋宏一「標本抽出の計画と方法」放送大学 (1991)
- 永田靖「サンプルサイズの決め方」朝倉書店 (2003)
- 永田靖「統計的方法のしくみ」日科技連 (1996)
- 村上雅人「なるほど統計学」海鳴社 (2002)
- 久保応助・藤沢偉作「全問精解 確率・統計演習」聖文社 (1983)
- 丹慶勝市「図解雑学 統計解析」ナツメ社 (2003)
- 高橋信[著]・トレンドプロ[マンガ]「マンガでわかる統計学」オーム社(2004)