

さまざまな検定

確率・統計 講義 13

村田 昇

2020年7月20日

復習

統計的仮説検定

- ある現象・母集団に対して仮定された仮説の真偽をデータに基づいて統計的に検証する方法
- 推定と大きく異なるのは、母集団の分布に対して何らかの仮説を考えると

検定における仮説

- 帰無仮説 H_0
検定統計量の分布を予想するために立てる仮説
- 対立仮説 H_1
“帰無仮説が誤っているときに起こりうるシナリオ”として想定する仮説

検定の基本的手続き

1. 仮説を立てる
2. 仮説のもとで検定統計量が従う帰無分布を調べる
3. 実際のデータから検定統計量の値を計算する
4. 計算された検定統計量の値が仮説が正しいときに十分高い確率で得られるかどうかを判断する

検定の用語

- 仮説の判定
 - 帰無仮説を棄却: 帰無仮説は誤っていると判断すること
 - 帰無仮説を受容: 帰無仮説を積極的に棄却できないこと
- 検定の誤り
 - 第一種過誤: “正しい帰無仮説を棄却する” 誤り
 - 第二種過誤: “誤った帰無仮説を受容する” 誤り
- 検定の設計
 - サイズ: “第一種過誤が起きる確率” を小さく
 - 検出力: “第二種過誤が起きない確率” を大きく

有意水準と p 値

- 有意水準

第一種過誤が起きる確率 (サイズ) として許容する上限

- p 値 (有意確率): (検定統計量 T , 棄却域 R_α)

検定統計量の値が棄却域に含まれる有意水準の最小値

$$(p \text{ 値}) = \min\{\alpha \in (0, 1) | T \text{ が } R_\alpha \text{ に含まれる}\}$$

- 有意水準と p 値の関係

p 値が有意水準未満のときに帰無仮説を棄却する

正規分布を用いた平均値の検定

- 問題

確率変数列の平均値が μ と等しいか検定せよ.

$$X_1, X_2, \dots, X_n$$

- 検定問題

$$X_i = \theta + \varepsilon_i, \quad i = 1, \dots, n \quad \varepsilon_i \sim N(0, \sigma^2)$$

を観測値の確率モデル (σ^2 は既知) とするとき

$$H_0: \theta = \mu \quad \text{vs} \quad H_1: \theta \neq \mu$$

- 検定統計量

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

は 帰無仮説が正しいとき標準正規分布に従う.

- 棄却域 (両側検定の場合)

$$R_\alpha = (-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$$

正規分布を用いた平均値の差の検定

- 問題

2 つの確率変数列の平均値が等しいか検定せよ.

$$X_1, X_2, \dots, X_n, \quad Y_1, Y_2, \dots, Y_m$$

- 検定問題

$$X_i = \theta_1 + \varepsilon_{1i}, \quad i = 1, \dots, n \quad \varepsilon_{1i} \sim N(0, \sigma^2)$$

$$Y_j = \theta_2 + \varepsilon_{2j}, \quad j = 1, \dots, m \quad \varepsilon_{2j} \sim N(0, \sigma^2)$$

を観測値の確率モデル (σ^2 は既知) とするとき

$$H_0: \theta_1 = \theta_2 \quad \text{vs} \quad H_1: \theta_1 \neq \theta_2$$

- 検定統計量

$$T = \sqrt{\frac{nm}{n+m}} \frac{\bar{X} - \bar{Y}}{\sigma}$$

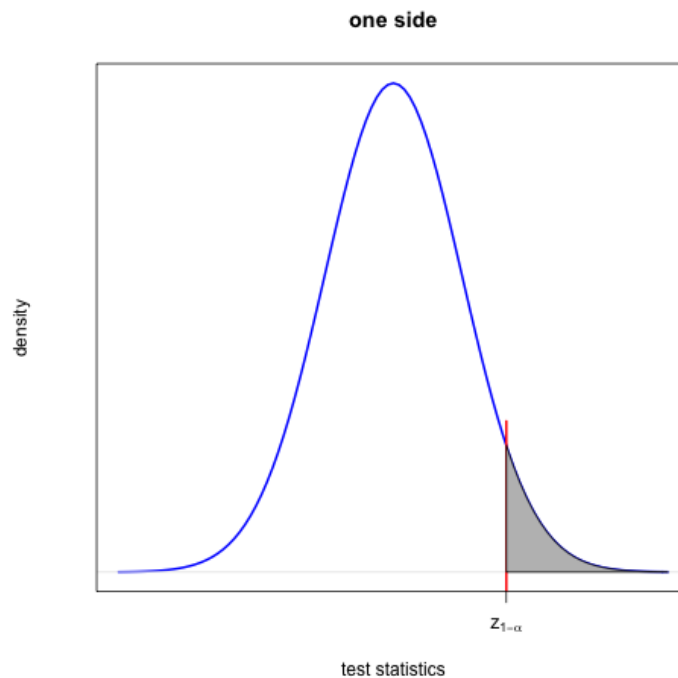
は仮説が正しいとき標準正規分布に従う.

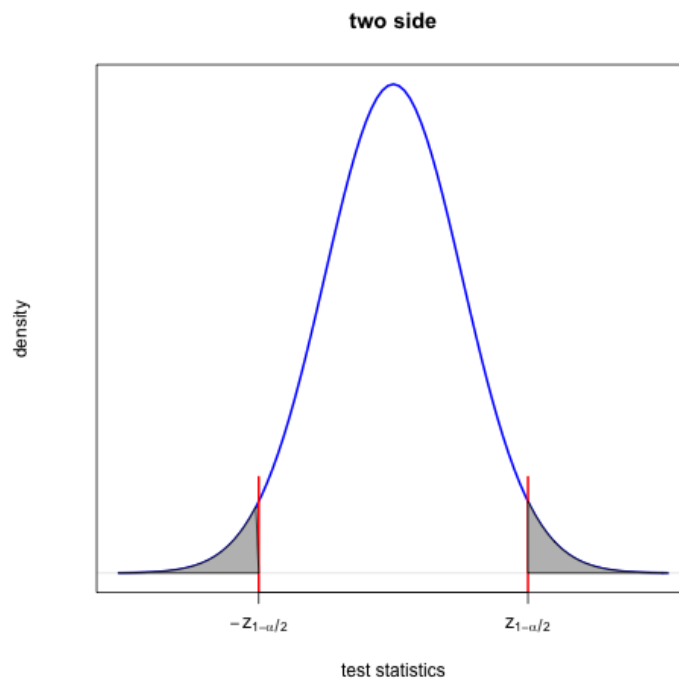
- 棄却域 (両側検定の場合)

$$R_\alpha = (-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$$

両側検定と片側検定

- 対立仮説によって棄却域の形は変わりうる
 - 例: 2つの薬の治療結果に対する仮説:
 - * 古い薬 (高価) と新しい薬 (安価) の効能が変わらない
 - * 古い薬に比べて新しい薬の効能が改善した
- 右片側検定: 棄却域がある定数 a によって (a, ∞)
- 左片側検定: 棄却域がある定数 a によって $(-\infty, a)$
 - 右片側検定と左片側検定を合わせて片側検定と呼ぶ
- 両側検定: 棄却域がある定数 $a < b$ によって $(-\infty, a) \cup (b, \infty)$





演習

- 以下の問に答えよ.

Y 県と F 県で生産されたの桃の甘さの違いを調べるために、ある八百屋で売られているそれぞれの桃一山の糖度を調べたとする. 一つ一つの甘さはばらついているが、どちらの県でも分散 $\sigma^2 = 0.6$ の正規分布に従うことが知られている. それぞれの標本平均を計算したところ

$$X_1, X_2, \dots, X_6 \rightarrow \bar{X} = 15.2$$

$$Y_1, Y_2, \dots, Y_{10} \rightarrow \bar{Y} = 14.5$$

となった. 本来の平均値が等しいと言えるだろうか?

解答例

- 観測値の確率モデル

$$X_i = \theta_1 + \varepsilon_{1i}, \quad i = 1, \dots, n \quad \varepsilon_{1i} \sim N(0, \sigma^2)$$

$$Y_j = \theta_2 + \varepsilon_{2j}, \quad j = 1, \dots, m \quad \varepsilon_{2j} \sim N(0, \sigma^2)$$

$$- n = 6, m = 10, \sigma^2 = 0.6$$

- 検定統計量

$$T = \sqrt{\frac{nm}{n+m}} \frac{\bar{X} - \bar{Y}}{\sigma}$$

- 棄却域 (例えば $\alpha = 0.05$ のとき $z_{0.975} = 1.96$)

$$R_\alpha = (-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$$

- 検定統計量の計算

$$T = \sqrt{\frac{6 \times 10}{6 + 10}} \frac{15.2 - 14.5}{\sqrt{0.6}} = 1.75$$

- “Y 県の方が F 県より甘い” という対立仮説の場合 (右片側検定)
- 棄却域 (例えば $\alpha = 0.05$ のとき $z_{0.95} = 1.64$)

$$R_\alpha = (z_{1-\alpha}, \infty)$$

正規分布にもとづいたいろいろな検定

観測値の仮定

- 確率モデル

$$X = \mu + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

- σ^2 が未知の場合を考える

平均の検定

- 問題

μ_0 を既知の定数として, 平均 μ が真の平均 μ_0 であるか否かを検定する

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0$$

考え方

- 平均と分散の推定量の性質
 - 標本平均: (正規分布に従う)

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- 不偏分散: (χ^2 分布に従う)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

- 標本平均と不偏分散は互いに独立

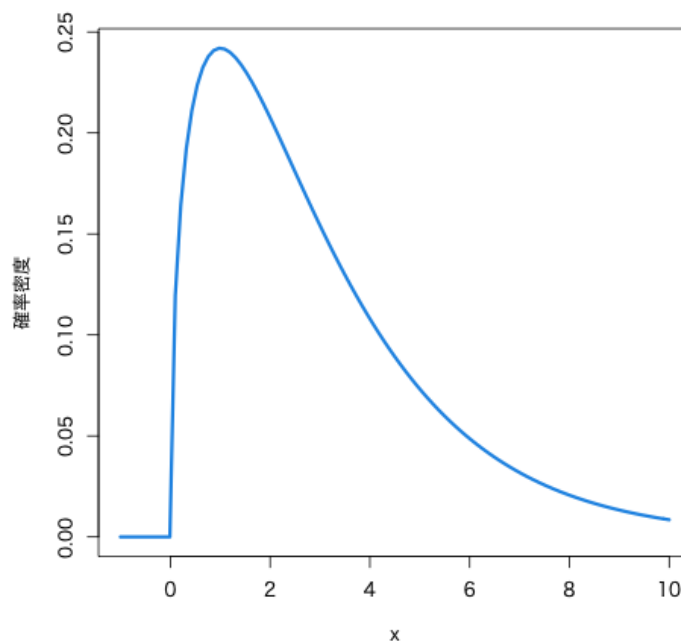


図 1: χ^2 -分布 (自由度 3)

χ^2 -分布

- 見本空間: $[0, \infty)$
- 母数: 自由度 ν
- 密度関数:

$$f(x) = \frac{1}{2^{\nu/2} \Gamma(\frac{\nu}{2})} x^{\nu/2-1} e^{-x/2}$$

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$$

– 備考: ν 個の標準正規分布の 2 乗和の分布で, 検定に利用される.

t -分布

- 見本空間: $(-\infty, \infty)$
- 母数: 自由度 ν
- 密度関数:

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)}$$

– 備考: 標準正規分布と自由度 ν の χ^2 -分布の比の分布で, 検定に利用される.

Student の t 検定

- 検定統計量

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{s}$$

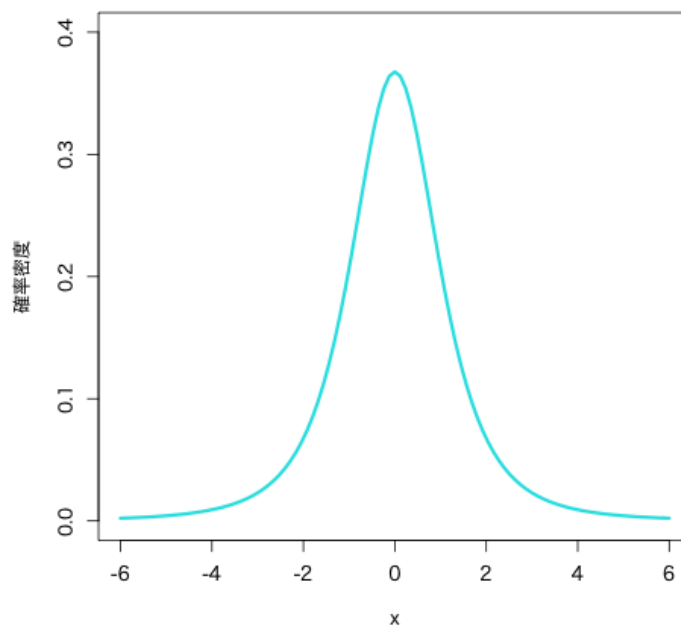


図 2: t -分布 (自由度 3)

- 帰無分布は自由度 $n-1$ の t -分布
- $t_{1-\alpha/2}(n-1)$:
自由度 $n-1$ の t 分布の $1-\alpha/2$ 分位点
- 棄却域:

$$R_\alpha = (-\infty, -t_{1-\alpha/2}(n-1)) \cup (t_{1-\alpha/2}(n-1), \infty)$$

平均の差の検定

- 問題
2 種類のデータの平均が等しいか否かを検定する
$$H_0: \mu_1 = \mu_2 \quad \text{vs} \quad H_1: \mu_1 \neq \mu_2$$
- Behrens-Fisher 問題
– 正確かつ適切な検定を導出することは実は難しい

考え方

- X_1, \dots, X_m および Y_1, \dots, Y_n の不偏分散:

$$s_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2, \quad s_2^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

- 標本平均と不偏分散の性質:
 - $\bar{X} - \bar{Y}, s_1^2, s_2^2$ は互いに独立
 - $(m-1)s_1^2/\sigma_1^2$ は自由度 $m-1$ の χ^2 分布に従う
 - $(n-1)s_2^2/\sigma_2^2$ は自由度 $n-1$ の χ^2 分布に従う

Welch の t 検定

- 検定統計量

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{s_1^2/m + s_2^2/n}}$$

- 帰無分布は自由度 $\hat{\nu}$ の t -分布 (Welch の近似)

$$\hat{\nu} = \frac{(s_1^2/m + s_2^2/n)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}}$$

- $t_{1-\alpha/2}(\hat{\nu})$:
自由度 $\hat{\nu}$ の t 分布の $1-\alpha/2$ 分位点
- 棄却域:

$$R_\alpha = (-\infty, -t_{1-\alpha/2}(\hat{\nu})) \cup (t_{1-\alpha/2}(\hat{\nu}), \infty)$$

分散の検定

- 問題
 σ_0^2 を既知の定数として, 分散 σ^2 が σ_0^2 であるか否かを検定する
 $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$

考え方

- 分散の推定量の性質:
 - 不偏分散: (χ^2 分布に従う)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

χ^2 検定

- 検定統計量:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

- 帰無分布は自由度 $n-1$ の χ^2 分布
- $\chi_{\alpha/2}^2(n-1), \chi_{1-\alpha/2}^2(n-1)$:
自由度 $n-1$ の χ^2 分布の $\alpha/2, 1-\alpha/2$ 分位点
- 棄却域:

$$R_\alpha = (0, \chi_{\alpha/2}^2(n-1)) \cup (\chi_{1-\alpha/2}^2(n-1), \infty)$$

分散の比の検定

- 問題

2 種類のデータの分散が等しいか否かを検定する

$$H_0: \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1: \sigma_1^2 \neq \sigma_2^2$$

考え方

- X_1, \dots, X_m および Y_1, \dots, Y_n の不偏分散:

$$s_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2, \quad s_2^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

- 分散の推定量の性質:
 - s_1^2, s_2^2 は互いに独立
 - $(m-1)s_1^2/\sigma_1^2$ 自由度 $m-1$ の χ^2 分布に従う
 - $(n-1)s_2^2/\sigma_2^2$ は自由度 $n-1$ の χ^2 分布に従う

F -分布

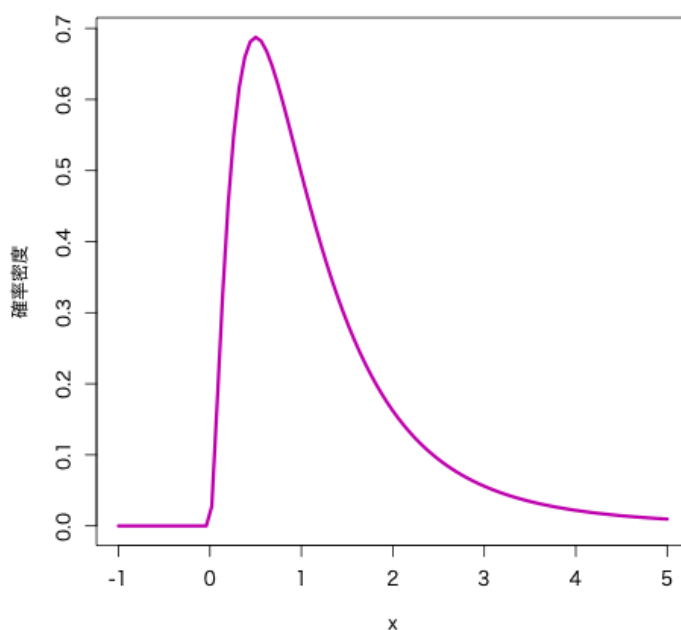


図 3: F -分布 (自由度 5, 10)

- 見本空間: $[0, \infty)$
- 母数: 自由度 ν_1, ν_2
- 密度関数:

$$f(x) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2}-1}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right) \left(1 + \frac{\nu_1}{\nu_2}x\right)^{\frac{\nu_1+\nu_2}{2}}}$$

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$$

- 備考: ν_1, ν_2 個の標準正規分布の 2 乗和の比の分布で, 検定に利用される.

F 検定

- 検定統計量:

$$F = \frac{s_1^2}{s_2^2}$$

- 帰無分布は自由度 $m-1, n-1$ の F -分布
- $F_{\alpha/2}(m-1, n-1), F_{1-\alpha/2}(m-1, n-1)$:
自由度 $m-1, n-1$ の F 分布の $\alpha/2, 1-\alpha/2$ 分位点
- 棄却域:

$$R_\alpha = (0, F_{\alpha/2}(m-1, n-1)) \cup (F_{1-\alpha/2}(m-1, n-1), \infty)$$

漸近正規性にもとづく検定

推定量の漸近正規性

- 漸近正規性 (データ数が多いときの性質)
 - 多くの推定量 $\hat{\theta}$ の分布は正規分布で近似できる
 - モーメントに基づく記述統計量は漸近正規性をもつ
 - 最尤推定量は広い範囲の確率分布に対して漸近正規性をもつ
 - いずれも中心極限定理にもとづく
- 信頼区間と同様に正規分布を用いて検定を考えることができる

最尤推定量の漸近正規性

- 定理
 - $f(x) > 0$ が連続で 2 階微分可能ならば $\sqrt{n}(\hat{\theta}^* - \theta_0)$ は $n \rightarrow \infty$ で正規分布 $N(0, I(\theta_0)^{-1})$ に近づく.
 - 観測データが十分多ければ, 最尤推定量の誤差 (分散) は Cramer-Rao 下界に一致する
 - Fisher 情報量 (f は確率質量関数または確率密度関数)

$$\begin{aligned} I(\theta_0) &= \mathbb{E}_{\theta_0} \left[-\frac{\partial^2}{\partial \theta^2} \log f(X, \theta_0) \right] \\ &= \mathbb{E}_{\theta_0} \left[\left(\frac{\partial}{\partial \theta} \log f(X, \theta_0) \right)^2 \right] \end{aligned}$$

最尤推定量の検定

- 問題
 - θ_0 を既知の定数として, 母数 θ が真の値 θ_0 であるか否かを検定する

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$$

- 上記は両側検定
- 片側検定も同様に考えることができる

考え方

- 最尤の推定量の性質

観測データ数 n が十分大きいとき, 1 次元母数 θ を含む連続分布の最尤推定量 $\hat{\theta}$ は

$$\mathbb{E}[\hat{\theta}] = \theta_0, \quad \text{Var}(\hat{\theta}) = \frac{1}{nI(\theta_0)}$$

の正規分布で近似できる.

Z 検定 (正規分布による検定)

- 検定統計量

$$Z = \sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0)$$

- 帰無分布は標準正規分布
- $z_{1-\alpha/2}$:
標準正規分布の $1-\alpha/2$ 分位点
- 棄却域:

$$R_\alpha = (-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$$

例題 (McNemar 検定)

- 問題

A 社と B 社の開発した 2 つの文字認識機械がある. n 個の文字に対してその性能を調べたところ

	1	2	3	...	n	
A 社	○	○	×	...	○	98.1%
B 社	×	○	○	...	○	98.0%

のような正答率を示した. このとき A 社の機械は B 社より優れていると言えるだろうか?

- 試験した結果によって判断は異なるべき
($n=10,000$)

	A が正解	A が誤り
B が正解	9800	0
B が誤り	10	190

	A が正解	A が誤り
B が正解	9610	190
B が誤り	200	0

演習

- 以下の問に答えよ

A 社と B 社の開発した 2 つの文字認識機械がある. 10,000 文字に対してその性能を調べたところ

	A が正解	A が誤り
B が正解	9500	180
B が誤り	220	100

のような正答率を示した. このとき A 社の機械は B 社より優れていると言えるだろうか?

解答例

- 以下のように考える.
 - 両者が正解: 簡単な文字 (比較しても意味がない)
 - 両者が不正解: 難しい文字 (比較しても意味がない)
 - どちらかが正解: 認識すべき文字 (比較すべき)
- 以下の確率変数を考える.

$$X = \begin{cases} 1, & \text{A 社のみ正解} \\ 0, & \text{B 社のみ正解} \end{cases}$$

- 確率質量関数

A 社のみ正解の確率を θ とし, 確率変数 X の確率質量関数を以下で定義する.

$$f(x) = \theta^x (1 - \theta)^{1-x}$$

- Fisher 情報量

$$I(\theta) = \mathbb{E} \left[-\frac{\partial^2}{\partial \theta^2} \log f(X; \theta) \right] = \frac{1}{\theta(1 - \theta)}$$

- 対数尤度

$$\ell(\theta) = n\bar{X} \log \theta + n(1 - \bar{X}) \log(1 - \theta)$$

- 最尤推定量

$$\begin{aligned} \ell'(\hat{\theta}) &= \frac{n\bar{X}}{\hat{\theta}} - \frac{n(1 - \bar{X})}{1 - \hat{\theta}} = 0 \\ \hat{\theta} &= \bar{X} \end{aligned}$$

- 以下の検定を考える.
 - 帰無仮説: A 社と B 社の性能に違いはない ($\theta_0 = 0.5$)
 - 対立仮説: A 社の方が B 社より性能が良い

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

- 検定統計量

$$Z = \sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0)$$

- 帰無分布は標準正規分布

- 有意水準と棄却域 (両側・片側) に応じて以下の分位点を用いれば良い

$\alpha = 0.10$	$z_{0.95} = 1.64$	$z_{0.90} = 1.28$
$\alpha = 0.05$	$z_{0.975} = 1.96$	$z_{0.95} = 1.64$
$\alpha = 0.01$	$z_{0.995} = 2.58$	$z_{0.99} = 2.33$

- 検定統計量の計算

題意より

$$Z = \sqrt{400/0.5^2}(220/400 - 0.5)$$

まとめ

- 分散未知の正規分布を用いた検定
 - 平均値の検定
 - 平均値の差の検定
 - 分散の検定
 - 分散の比の検定
- 最尤推定量を用いた検定
 - 漸近正規性
 - 正規分布を用いた検定に帰着