仮説検定

確率・統計 - 第 10 講

村田 昇

前回のおさらい

- 最尤法の考え方
- 最尤推定量の性質
  - 一致性
  - 漸近正規性
- 最尤推定量による区間推定
  - 標本平均の場合
  - 母数推定の場合

統計的推測の考え方

- 観測データは確率変数の集合
- 確率変数列 $X_1, X_2, \ldots, X_n$ に対する考察が重要
- 現象の理解のためには観測された実現値より確率分布にこそ興味がある
- 一般に分析対象のデータには独立性と同分布性が同時に仮定される
- 観測データの背後の確率分布を推定
  - 分布のもつ特性量 (平均や分散など) を評価する
  - 分布そのもの (確率関数や確率密度) を決定する

点推定

- 定義

  点推定とは母数 θ を $X_1, \ldots, X_n$ の関数

  $$\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$$

  で推定することで，$\hat{\theta}$ を θ の推定量と呼ぶ.

区間推定

- 定義

  区間推定とは未知母数 θ とある値 $\alpha \in (0,1)$ に対して以下を満たす確率変数 L, U を観測データから求めることをいう.

  $$P(L \leq \theta \leq U) \geq 1-\alpha$$

  - 区間 $[L, U]$: $1-\alpha$ 信頼区間 ($100(1-\alpha)$ % とも書く)
  - L : $1-\alpha$ 下側信頼限界
  - U : $1-\alpha$ 上側信頼限界
  - $1-\alpha$: 信頼係数 ($\alpha = 0.01, 0.05, 0.1$ など)

尤度関数

- 定義

  母数 $\boldsymbol{\theta}$ に対して観測データ $X_1, X_2, \ldots, X_n$ が得られる理論上の確率

  $$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_{\theta}(X_i)$$

  を $\boldsymbol{\theta}$ の尤度と言い，$\boldsymbol{\theta}$ の関数 L を尤度関数と呼ぶ．

  - 離散分布の場合 : f は確率質量関数
  - 連続分布の場合 : f は確率密度関数

最尤法

- 最尤法

  観測データに対して "最も尤もらしい" 母数の値を $\boldsymbol{\theta}$ の推定量として採用する方法を最尤法という．

- 最尤推定量

  Θ を尤度関数の定義域として，尤度関数を最大とする $\hat{\boldsymbol{\theta}}$

  $$L(\hat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}).$$

  $$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}).$$

  を $\boldsymbol{\theta}$ の最尤推定量という．

最尤推定量の区間推定

- 定理 (最尤推定量の漸近正規性)

  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

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

  の正規分布で近似できる．

- 信頼区間の構成

  母数 θ の $1-\alpha$ 信頼区間は以下で構成される．

  $$\left[\hat{\theta} - z_{1-\alpha/2} \cdot \frac{1}{\sqrt{nI(\hat{\theta})}},\ \hat{\theta} + z_{1-\alpha/2} \cdot \frac{1}{\sqrt{nI(\hat{\theta})}}\right]$$

  - サンプル数 n が十分大きい場合に近似的に正しい

検定とは

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
- 薬の治験の例

  新しい薬の効能が古い薬よりも優れていること (仮説) を薬の治験結果 (データ) から検証したい

- 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** を考えるところ

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** が従う標本分布を調べる
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

検定における仮説

- **帰無仮説**

  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**

  “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
  - 慣習として帰無仮説を H_0, 対立仮説を H_1 で表す
  - “帰無仮説を捨てて無に帰する” ことを期待する場合が多い

帰無仮説と対立仮説

- 薬の治験の例 (単純な比較)
  - 帰無仮説 H_0 : 新しい薬も古い薬も効能は同じ
  - 対立仮説 H_1 : 新しい薬と古い薬の効能は異なる
- 薬の治験の例 (新薬がより良いことを期待)
  - 帰無仮説 H_0 : 新しい薬も古い薬も効能は同じ
  - 対立仮説 H_1 : 新しい薬の方が古い薬より効能が高い

検定の用語

- **帰無分布**

  帰無仮説が正しい場合に検定統計量が従う分布
- **棄却域**

  帰無仮説の下で統計量の取り得るべき範囲の外の領域
- 仮説検定
  - 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

検定の設計指針

- 検定統計量 T に対して棄却域 R を設計
  - **サイズ** : 第一種過誤が **起きる確率**

$$(\text{サイズ}) = P(T \text{ が } R \text{ に含まれる} | \text{帰無仮説が正しい})$$

  - **検出力** : 第二種過誤が **起きない確率**

$$(\text{検出力}) = P(T \text{ が } R \text{ に含まれる} | \text{対立仮説が正しい})$$

- **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

棄却域の決め方

- 検定を構成する場合の一般的な戦略

  棄却域 R_α は以下のように設定する
  - **サイズを小さく** (有意水準 α 以下に) 抑える
  - 可能な限り **検出力を大きく** する

$$P(T \in R_\alpha | \text{帰無仮説が正しい}) \leq \alpha$$
$$P(T \in R_\alpha | \text{対立仮説が正しい}) \to \text{最大化}$$

- 対立仮説によって棄却域の形は変わりうる

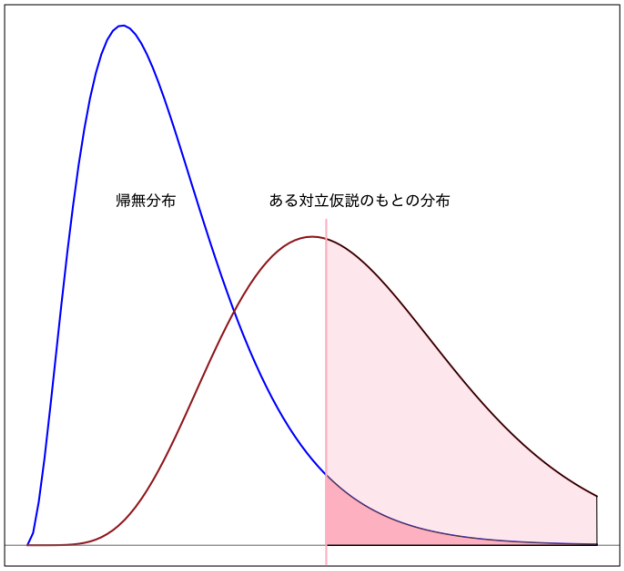


Figure 1: 有意水準が小さい場合

両側検定と片側検定

- 棄却域の形による分類
  - **右片側検定**

    ある定数 a を用いて (棄却域) = (a, ∞)
  - **左片側検定**

    ある定数 a を用いて (棄却域) = $(-\infty, a)$
  - **両側検定**

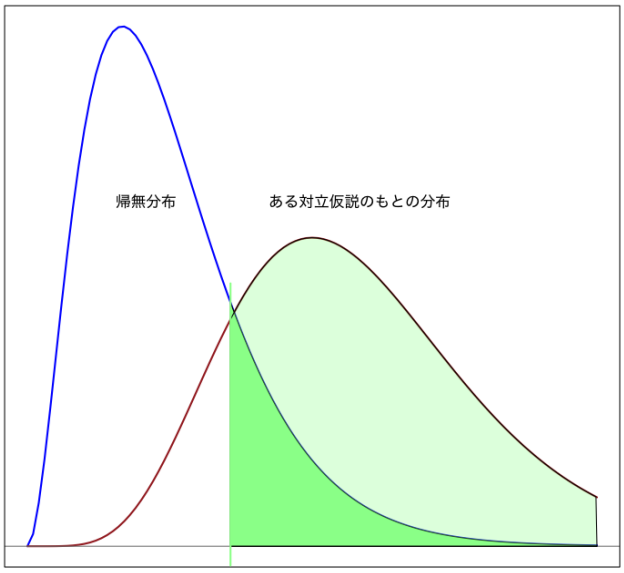


Figure 2: 有意水準が大きい場合

ある定数 $a < b$ を用いて (棄却域) $= (-\infty, a) \cup (b, \infty)$

- 平均の検定の場合
  - 対立仮説を $H_1 : \mu \neq \mu_0$ にとった場合は両側検定
  - 対立仮説を $H_1 : \mu > \mu_0$ にとった場合は右片側検定
  - 対立仮説を $H_1 : \mu < \mu_0$ にとった場合は左片側検定

p 値 (有意確率)

- 帰無分布における検定統計量の評価

  検定統計量の値が棄却域に含まれる有意水準の最小値を考える

- **p 値 (有意確率)** : (検定統計量 T, 棄却域 R_α)

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

- p 値が有意水準未満のときに帰無仮説を棄却する

検定に関する注意

- 帰無仮説の受容の意味

  帰無仮説が正しいと仮定しても矛盾は生じない

  帰無仮説の正しさを積極的に支持する結果ではない

- 第二種過誤

  一般に第二種過誤の起こる確率については何ら仮定がないため，その確率は非常に大きい可能性がある

正規分布を用いた検定

例題

- 平均値の検定

  ある工場で製造される機械の寿命は平均 μ 時間であることが望まれている．n 個の機械をランダムに選び耐久試験をしたところ，寿命はそれぞれ $X_1, X_2, \ldots, X_n$ であった．この機械の平均寿命は μ であると言えるだろうか？

- 平均値の差の検定

  A 県と B 県で生産されたの林檎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林檎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ので，それぞれの標本平均を計算したところ

  $$X_1, X_2, \ldots, X_n \to \hat{\mu}^X = \bar{X}$$
  $$Y_1, Y_2, \ldots, Y_m \to \hat{\mu}^Y = \bar{Y}$$

  となった．$\hat{\mu}^X$ と $\hat{\mu}^Y$ を比べることによって，本来の平均値が等しい

  $$\mu^X = \mu^Y ?$$

  と言えるだろうか？

- McNemar 検定

  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？

平均値の差の検定

- 問題

  2 つの確率変数列

  $$X_1, X_2, \ldots, X_n$$
  $$Y_1, Y_2, \ldots, Y_m$$

  が与えられたとき，その平均が等しいか検定せよ．

- 検定問題

  それぞれの平均を θ_1, θ_2 とすると

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

- 観測値の確率モデル

  $$X_i = \theta_1 + \varepsilon_{1i}, \quad i = 1, \ldots, n$$
  $$Y_j = \theta_2 + \varepsilon_{2j}, \quad j = 1, \ldots, m$$

- 誤差の仮定
  - $\varepsilon_{1i}, \varepsilon_{2j}$ は互いに独立に同一の分布に従う
  - $\varepsilon_{1i}, \varepsilon_{2j}$ は平均 0，分散 σ^2 (既知) である

$$\mathbb{E}[\varepsilon_{1i}] = \mathbb{E}[\varepsilon_{2j}] = 0$$
$$\mathbb{E}[\varepsilon_{1i}^2] = \mathbb{E}[\varepsilon_{2j}^2] = \sigma^2 < \infty$$

- $\varepsilon_{1i}, \varepsilon_{2j}$ は正規分布に従う

- 母数の推定量

$$\hat{\theta}_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$\hat{\theta}_2 = \bar{Y} = \frac{1}{m}\sum_{j=1}^{m} Y_i$$

- 直感的な議論

$$|\hat{\theta}_1 - \hat{\theta}_2| \text{が小さい} \Rightarrow \text{同じと思って良い}$$
$$|\hat{\theta}_1 - \hat{\theta}_2| \text{が大きい} \Rightarrow \text{どうやら違うらしい}$$

- 帰無仮説が正しい場合

  誤差 ε が平均 0，分散 σ^2 の正規分布に従うので

$$\hat{\theta}_1 - \hat{\theta}_2 = \bar{X} - \bar{Y}$$

  は

$$\text{平均} \quad \mathbb{E}[\bar{X} - \bar{Y}] = 0$$
$$\text{分散} \quad \mathrm{Var}(\bar{X} - \bar{Y}) = \left(\frac{1}{n} + \frac{1}{m}\right)\sigma^2$$

  の正規分布に従う．

- 検定統計量の構成と帰無分布の導出

  分散が 1 になるようにを正規化した統計量

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

  は仮説が正しいとき

$$\text{平均} \quad \mathbb{E}[T] = 0$$
$$\text{分散} \quad \mathrm{Var}(T) = 1$$

  の正規分布に従う．

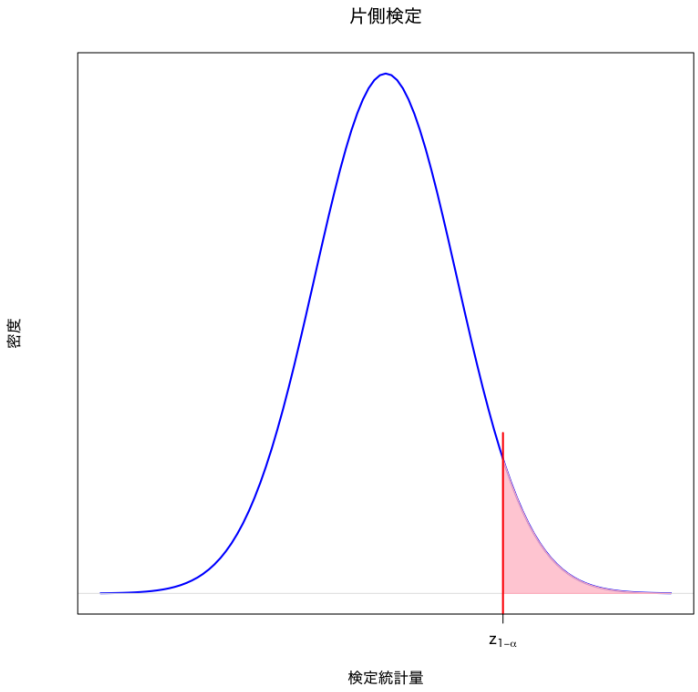
- 有意水準と分位点

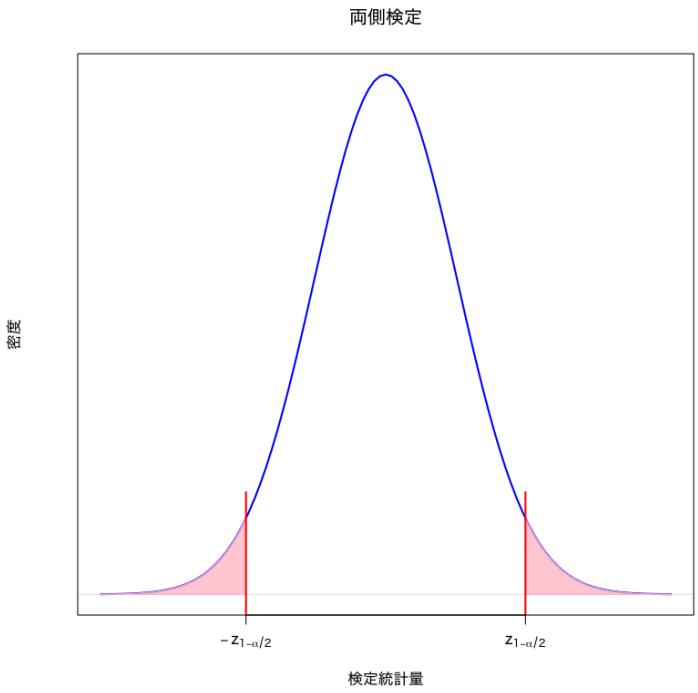
  正規分布の確率密度関数から $|T| > z_{1-\alpha/2}$ という事象がおこる確率が α となる $z_{1-\alpha/2}$ ($1-\alpha/2$ 分位点)

$$P\left(|T| > z_{1-\alpha/2}\right) = \alpha$$

  を求める．

  - 仮説が正しいのなら $|T| > z_{1-\alpha/2}$ という事象は確率 α 程度でしか起こらない





- 棄却域の構成

  統計量 T が以下の領域に入ったら帰無仮説を棄却する.

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

  - $|T| > z_{1-\alpha/2}$ が起こったのなら，“滅多に起こらないことが起こった” ので仮説を正しいと考えることは疑わしいと考える
  - “仮説を有意水準 α で **棄却** する” ($\Leftrightarrow$ “仮説を **受容** する”)

演習

練習問題

- 平均値の検定の問題を考えてみよ．

  加法的雑音が分散 σ^2 の正規分布に従う確率変数列

  $$X_1, X_2, \ldots, X_n$$

  が与えられたとき，平均値が μ と等しいか検定せよ．

練習問題

- 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

  $$X_1, X_2, \ldots, X_6 \to \bar{X} = 15.2$$
  $$Y_1, Y_2, \ldots, Y_{10} \to \bar{Y} = 14.5$$

  となった．本来の平均値が等しいと言えるだろうか？

今回のまとめ

- 統計的仮説検定
  - 検定の手続き
  - 帰無仮説と対立仮説
  - 有意水準，棄却域，p 値
  - 過誤と検出力
  - 両側検定と片側検定
- 正規分布を用いた検定
  - 平均値の差の検定
  - 平均値の検定