

2 標本 t 検定と Hotelling の T^2 検定

樋口さぶろお <https://hig3.net>

龍谷大学大学院 先端理工学研究科 数理・情報科学コース

計算科学特論 L14(2026-07-20 Mon)

最終更新: Time-stamp: "2026-08-08 Sat 09:09 JST hig"

今日の目標

- (対応のない) 2 群の母平均値の差の t 検定を実行できる
- (対応のない) 2 群の母平均値の差ベクトルの Hotelling の T^2 検定を実行できる



L13-Q1

TA Prob and Sol: 共分散行列の変換

$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix}$ を m 個の確率変数のベクトルとする.

$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_p \end{bmatrix}$ を, 適当な次元の定数の行列 R により $\mathbf{Y} = R\mathbf{X}$ で定める.

X_i の共分散 $\text{Cov}[X_i, X_j]$ を (i, j) 成分とする行列を分散-共分散行列 Σ_{XX} といい, $\text{Cov}[\mathbf{X}, \mathbf{X}]$ と略記する.

- Σ は, 行列 $(\mathbf{X} - \boldsymbol{\mu})^t(\mathbf{X} - \boldsymbol{\mu})$ の各成分の母期待値からなるものであることを示そう.
- $\text{Cov}[\mathbf{Y}, \mathbf{Y}] = R\text{Cov}[\mathbf{X}, \mathbf{X}]^t R$ を示そう.

略解

$\boldsymbol{\mu} = E[\mathbf{X}]$ とし, その第 i 成分を $\mu_i = E[X_i]$ とする.

- 行列

$$(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t$$

の (i, j) 成分は

$$(X_i - \mu_i)(X_j - \mu_j)$$

である．したがって，各成分の母期待値からなる行列の (i, j) 成分は

$$E[(X_i - \mu_i)(X_j - \mu_j)] = \text{Cov}[X_i, X_j]$$

となる．これは Σ_{XX} の (i, j) 成分なので，

$$\Sigma_{XX} = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t].$$

- $\boldsymbol{\nu} = \mathbb{E}[\mathbf{Y}]$ とすると, R は定数行列なので,

$$\boldsymbol{\nu} = \mathbb{E}[R\mathbf{X}] = R\boldsymbol{\mu}.$$

よって,

$$\begin{aligned} \text{Cov}[\mathbf{Y}, \mathbf{Y}] &= \mathbb{E}[(\mathbf{Y} - \boldsymbol{\nu})(\mathbf{Y} - \boldsymbol{\nu})^{\mathsf{t}}] \\ &= \mathbb{E}[R(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathsf{t}}R] \\ &= R\mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{\mathsf{t}}]R \\ &= R\text{Cov}[\mathbf{X}, \mathbf{X}]R. \end{aligned}$$

L13-Q2

Quiz(因子分析の共分散行列)

因子分析のモデル

$$\mathbf{X} = \boldsymbol{\mu}_0 + \Lambda \mathbf{f} + \boldsymbol{\epsilon}.$$

において,

$$\begin{aligned} \text{Cov}[\mathbf{X}, \mathbf{X}] &= \text{Cov}[\Lambda \mathbf{f}, \Lambda \mathbf{f}] + \text{Cov}[\boldsymbol{\epsilon}, \boldsymbol{\epsilon}] \\ &= \Lambda \text{Cov}[\mathbf{f}, \mathbf{f}] \Lambda^t + \text{diag}(\sigma_1^2, \dots, \sigma_p^2) = \Lambda^t \Lambda + \text{diag}(\sigma_1^2, \dots, \sigma_p^2) \end{aligned}$$

を導出しよう.

ここまで来たよ

13 因子分析 Factor Analysis

14 2 標本 t 検定と Hotelling の T^2 検定

- t 分布
- 2 標本 t 検定 (1 変数)
- Hotelling の T^2 検定

カイ二乗分布

岩薩林 p.123

定義 (カイ二乗分布)

$Z_1, \dots, Z_k$ が独立同一分布 $Z_i \sim N(0, 1^2)$ のとき, 確率変数

$$W = Z_1^2 + \dots + Z_k^2$$

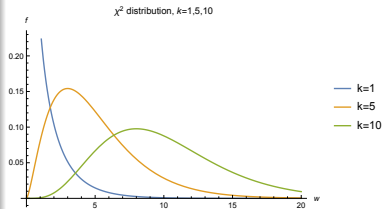
は自由度 k のカイ二乗分布 χ_k^2 に従う.

命題 (χ_k^2 の確率密度関数)

確率密度関数 $f_k(w)$ は

$$f_k(w) = \begin{cases} C_k w^{\frac{k}{2}-1} e^{-\frac{1}{2}w} & (w \geq 0), \\ 0 & \text{それ以外} \end{cases}$$

で与えられる.



不偏標本分散のしたがう分布

正規分布 $N(\mu, \sigma^2)$ にしたがう母集団のサイズ n の標本

- 標本平均値の分布 $N(\mu, \sigma^2/n)$ をうまく平行移動, 拡大縮小すると標準正規分布 $N(0, 1^2)$
- 不偏標本分散の分布をうまく拡大縮小すると **カイ二乗分布** χ_k^2

命題 (不偏標本分散のしたがう分布 岩薩林 定理 5.6)

正規分布 $N(\mu, \sigma^2)$ からサイズ n の標本 $X_1, X_2, \dots, X_n$, i.i.d. $\sim N(\mu, \sigma^2)$ を取り出すとき, 不偏標本分散

$$S^2 = \frac{1}{n-1} [(X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2]$$

から定めた

$$W = (n-1) \times \frac{V^2}{\sigma^2}$$

は, **自由度** $k = n-1$ の **カイ二乗分布** χ_{n-1}^2 にしたがう.

実は W は標本平均値 $\bar{X}_n$ と独立

t 分布

定義 (t 分布)

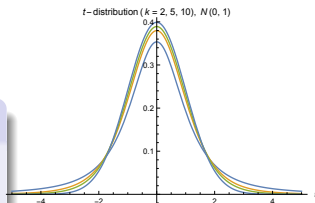
$Z \sim N(0, 1^2)$, $W \sim \chi_k^2$, Z と W が独立,
 のとき連続型確率変数 $T = \frac{Z}{\sqrt{W/k}}$ のしたがう分布を自由度 k の (ス
 チューデントの, またはゴセットの)t 分布 t_k という.

自由度 k が小さいとき, $N(0, 1^2)$ より
 低く広い.

自由度 $k \rightarrow +\infty$ で $N(0, 1^2)$ に一致
 する.

自由度 k の t 分布の確率密度関数

$$f_k(x) = A_k \cdot \left(1 + \frac{1}{k}x^2\right)^{-\frac{k+1}{2}}.$$



F 分布

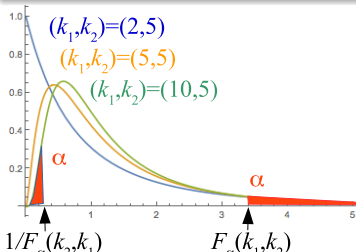
定義 (F 分布)

$W_{k_1} \sim \chi_{k_1}^2, W_{k_2} \sim \chi_{k_2}^2$ 独立な確率変数.

このとき、分子分母の母期待値を 1 に調節した比の確率変数

$$F = \frac{W_{k_1}/k_1}{W_{k_2}/k_2}$$

のしたがう分布を自由度 (k_1, k_2) の F 分布といい, $F \sim F(k_1, k_2)$ とかく.



Python では

```
from scipy import stats
rvf=stats.f(dfn=k1,dfd=k2)
rvf.ppf(p)
df=degrees of freedom, n=numerator 分子, d=denominator 分母
```

t 分布と F 分布の関係

命題 (t 分布と F 分布の関係)

確率変数 $t \sim t_k$ のとき, $t^2 \sim F(1, k)$.

$t = Z / \sqrt{W_k/k}$ と書ける.

$t^2 = \frac{Z^2/1}{W_k/k}$ より, $t^2 \sim F(1, k)$.

なお, 分子だけ見たら, $Z \sim N(0, 1^2)$ のとき $Z^2 \sim \chi_1^2$.

$t^2 = f$ とすると,

$P(|t| > t_c) = P(f = t^2 > t_c^2)$.

ここまで来たよ

13 因子分析 Factor Analysis

14 2 標本 t 検定と Hotelling の T^2 検定

- t 分布
- 2 標本 t 検定 (1 変数)
- Hotelling の T^2 検定

正規分布にしたがう確率変数の和と差

命題 (正規分布の再生性)

確率変数 $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ のとき, 和の確率変数 $Z = X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

さらに, $aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$ であるから, 差の確率変数

$$X - Y \sim N(\mu_X - \mu_Y, (+1)^2\sigma_X^2 + (-1)^2\sigma_Y^2).$$

以下, 等分散 $\sigma = \sigma_X = \sigma_Y$ と仮定する.

標本平均値の差

確率変数 X_i ($i = 1, 2, \dots, n_X$) が独立同一分布 $N(\mu_X, \sigma^2)$

確率変数 Y_i ($i = 1, 2, \dots, n_Y$) が独立同一分布 $N(\mu_Y, \sigma^2)$

にしたがうとする (X, Y のサイズ n_X, n_Y の 2 標本). 標本平均値は, 再生性より,

$$\bar{X} = \frac{1}{n_X} [X_1 + \dots + X_{n_X}] \sim N(\mu_X, \frac{\sigma^2}{n_X}).$$

$$\bar{Y} = \frac{1}{n_Y} [Y_1 + \dots + Y_{n_Y}] \sim N(\mu_Y, \frac{\sigma^2}{n_Y}).$$

$$\text{よって, } \bar{X} - \bar{Y} \sim N(\mu_X - \mu_Y, \sigma^2 \cdot (\frac{1}{n_X} + \frac{1}{n_Y})).$$

注: X_i, Y_i が正規分布とかぎらない独立同一分布にしたがうときも, n_X, m が大きければ, 中心極限定理から近似的にこのことが成り立つ.

これを使って $\mu_X - \mu_Y$ について区間推定や検定ができる!

しか～し, 母分散 σ^2 が最初からわかっていることはなかなかない.
母分散を推定値で代用する.

共通の母分散の推定値

定義 (2 標本の合併した (プールした) 不偏標本分散)

標本 X_i ($i = 1, 2, \dots, n_X$), 標本 Y_i ($i = 1, 2, \dots, n_Y$) に対して,

$$\begin{aligned} S^2 &= \frac{1}{n_X + n_Y - 2} \left[(X_1 - \bar{X})^2 + \dots + (X_{n_X} - \bar{X})^2 \right. \\ &\quad \left. + (Y_1 - \bar{Y})^2 + \dots + (Y_{n_Y} - \bar{Y})^2 \right] \\ &= \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{(n_X - 1) + (n_Y - 1)} \end{aligned}$$

をプールした不偏標本分散という.

共通の母分散の推定値として使える.

2 標本の不偏標本分散 S_X^2 , S_Y^2 を自由度で重み付けして平均したもの.

命題 (標本平均値の差のしたがう分布)

$X_i \sim N(\mu_X, \sigma^2)$, $Y_j \sim N(\mu_Y, \sigma^2)$, $i = 1, \dots, n_X$, $j = 1, \dots, n_Y$ のとき,

2 標本 t 統計量

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{S^2 \left(\frac{1}{n_X} + \frac{1}{n_Y} \right)}}$$

は自由度 $n_X + n_Y - 2$ の t 分布に従う.

また, その 2 乗は

$$T^2 = \frac{((\bar{X} - \bar{Y}) - (\mu_X - \mu_Y))^2}{S^2 \left(\frac{1}{n_X} + \frac{1}{n_Y} \right)} \sim F(1, n_X + n_Y - 2).$$

1 次元 t 検定における分母の標本サイズ n が, その調和平均 $\left(\frac{1}{\frac{1}{n_X} + \frac{1}{n_Y}} \right)$ に置き換えられたと考えられる.

これを利用して, 差の区間推定や差の検定を行うことができる.

命題 (母平均値の差の区間推定)

$X_i \sim N(\mu_X, \sigma^2), Y_i \sim N(\mu_Y, \sigma^2), i = 1, \dots, n_X, j = 1, \dots, n_Y$ のとき,
母平均値の差 $\mu_X - \mu_Y$ の信頼係数 $1 - \alpha$ の信頼区間は

$$\begin{aligned} & \bar{X} - \bar{Y} - t_{n_X+n_Y-2, \frac{\alpha}{2}} \sqrt{S^2 \left(\frac{1}{n_X} + \frac{1}{n_Y} \right)} \\ & < \mu_X - \mu_Y < \bar{X} - \bar{Y} + t_{n_X+n_Y-2, \frac{\alpha}{2}} \sqrt{S^2 \left(\frac{1}{n_X} + \frac{1}{n_Y} \right)} \end{aligned}$$

L14-Q1

Quiz(母平均値の差の区間推定 (母分散未知))

ドーナツ製造マシン 1 号, 2 号が製造するドーナツの重さ X, Y g は, 未知の母平均値 μ_1, μ_2 の独立同分布にしたがう確率変数である. 母分散も未知だが, 1 号と 2 号で等しいことがわかっている.

1 号, 2 号で製造したドーナツの重さは次のようだった.

1 号: 51g, 52g, 47g, 50g.

2 号: 55g, 56g, 51g, 52g, 56g, 54g.

- ① $\bar{X}, \bar{Y}$, 合併した不偏標本分散 S^2 を求めよう.
- ② 母平均値の差 $\mu_1 - \mu_2$ を, 信頼係数 $1 - \alpha = 0.99$ で区間推定しよう.

(両側)2 標本 t 検定 I

L14-Q2

TA Prob and Sol: 両側 2 標本 t 検定

ドーナツ製造マシン 1 号, 2 号が製造するドーナツの重さ X_i, Y_j g は, 未知の母平均値 μ_1, μ_2 , 未知の共通の母分散の独立な正規分布にしたがう確率変数である.

- 1 号の標本は, サイズ 4, 標本平均値 50g, 不偏標本分散 $\frac{14}{3}g^2$,
- 2 号の標本は, サイズ 6, 標本平均値 54g, 不偏標本分散 $\frac{22}{5}g^2$ だった.

2 つのドーナツ製造マシンの製造するドーナツの重さの母平均値に差があるか知りたい. 帰無仮説を H_0 : 差はない $\mu_1 - \mu_2 = 0$, として, 有意水準 $\alpha = 0.01$ で両側 2 標本 t 検定をしよう.

略解

(両側)2 標本 t 検定 II

- ① 有意水準 $\alpha = 0.01$ で,
- ② 母平均値の差の両側 2 標本 t 検定を行う
- ③ 帰無仮説 H_0 を, 「…ドーナツの重さの母平均値は等しい:
 $\mu_1 - \mu_2 = 0$ 」とする. すなわち, 対立仮説 H_1 を, $\mu_1 \neq \mu_2$ とする.
- ④ サイズ m, n の標本の, 標本平均値を $\bar{X}, \bar{Y}$, プールした不偏標本分散を S^2 とすると, 量 $T = \frac{\bar{X} - \bar{Y}}{\sqrt{S^2 \cdot (\frac{1}{n} + \frac{1}{m})}}$ は, 帰無仮説のもとで自由度 $n + m - 2$ の t 分布に従う. この量を検定統計量として用いる.
- ⑤ この標本に対して $T = \frac{-4}{\sqrt{\frac{1}{4+6-2}((4-1)\frac{14}{3} + (6-1)\frac{22}{5}) \cdot (\frac{1}{4} + \frac{1}{6})}} = -2.9218$.
- ⑥ t 分布表より, p 値 $P(|T| > 2.9218)$ は $\alpha = 0.01$ よりも大きい ($0.01 = P(|T| > 3.355)$ だから. あるいは, $2.9218 < t_{0.005}(8)$ だから といっても同じこと). よって帰無仮説は棄却されない.
 母平均値が異なると有意水準 0.01 では結論できない.

(両側)2 標本 t 検定 III

対応のある 2 標本 t 検定

$n = n_x = n_y$ 人の学生別に, 月曜日と水曜日の睡眠時間の睡眠時間のデータがあるとき,
それぞれの睡眠時間が正規分布に従うなら, 同じ学生の月曜日と水曜日の差も正規分布にしたがう. 1 次元の t 検定をすればよい.

ここまで来たよ

13 因子分析 Factor Analysis

14 2 標本 t 検定と Hotelling の T^2 検定

- t 分布
- 2 標本 t 検定 (1 変数)
- Hotelling の T^2 検定

2 標本の母平均値ベクトルの差

X 組, Y 組の学生の, 英語数学の成績ベクトル $\mathbf{X}_i, \mathbf{Y}_i$ を想像して.

確率変数 $\mathbf{X}_i$ ($i = 1, 2, \dots, n_X$) が独立同一分布 $N(\boldsymbol{\mu}_X, \Sigma)$

確率変数 $\mathbf{Y}_i$ ($i = 1, 2, \dots, n_Y$) が独立同一分布 $N(\boldsymbol{\mu}_Y, \Sigma)$

にしたがうとする. Σ は共通の共分散行列.

X, Y 組の平均成績ベクトル $\boldsymbol{\mu}_X, \boldsymbol{\mu}_Y$ は異なるか?

英語の成績, 数学の成績はそれぞれ正規分布に従うので, 英語, 数学, 別々に 2 標本 t 検定すればよい. しかし, 多重検定による有意水準の調整が必要になる. ベクトルとして扱えないか?

Hotelling の T^2 統計量

$\mathbf{X}_i, \mathbf{Y}_j$ が p 次元正規分布に従い、標本サイズが n_X, n_Y のとき、

定義 (Hotelling の T^2 統計量)

$$T^2 = {}^t(\bar{\mathbf{X}} - \bar{\mathbf{Y}}) \left(\frac{1}{\frac{1}{n_X} + \frac{1}{n_Y}} \right) (S^2)^{-1} (\bar{\mathbf{X}} - \bar{\mathbf{Y}})$$

プールされた不偏標本共分散行列は、

$$S^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}.$$

T^2 は標本平均値間のマハラノビス距離の 2 乗に $\frac{1}{\frac{1}{n_X} + \frac{1}{n_Y}}$ を掛けたものである。

マハラノビス距離: 線形判別分析 (LDA) で現れた。

1 変数では t 分布にしたがう t の 2 乗になる。

Wishart 分布 — カイ二乗分布の多変数拡張

多次元確率変数 $\mathbf{X}_i \in \mathbb{R}^p$, $\mathbf{X}_i \sim N(\mathbf{0}, \Sigma)$ $i = 1, \dots, n$ とする.

定義 (Wishart 分布)

$p \times p$ 行列 $\sum_{i=1}^n \mathbf{X}_i^t \mathbf{X}_i$ のしたがう分布を, 母数 Σ, p, n の Wishart 分布 $W(\Sigma, p, n)$ という.

T2 統計量と F 分布の間の関係

$$\frac{n_X + n_Y - p - 1}{p(n_X + n_Y - 2)} T^2 \sim F(p, n_X + n_Y - p - 1)$$

$p = 1$ とすると, t 分布とカイ二乗分布の関係に相当する.

L14-Q3

Quiz(両側 2 標本 t2 検定)

各 30 人の 2 組で、 $p = 2$ 科目の試験を行った。

X, Y 組の 2 科目の成績の標本平均値ベクトルはそれぞれ $\begin{pmatrix} 65 \\ 70 \end{pmatrix}, \begin{pmatrix} 72 \\ 74 \end{pmatrix}$

だった。プールされた不偏共分散行列は $S = \begin{pmatrix} 100 & 60 \\ 60 & 90 \end{pmatrix}$ だった。

2 組の成績は、同じ共分散行列の正規分布にしたがうと仮定して、有意水準 $\alpha = 0.05$ で、成績の母平均値ベクトルに差があるか検定しよう。

L14-Q4

Quiz(両側 2 標本 t2 検定)

各 40 人の 2 組で、 $p = 2$ 科目の試験を行った。

X, Y 組の 2 科目の成績の標本平均値ベクトルはそれぞれ $\begin{pmatrix} 60 \\ 65 \end{pmatrix}, \begin{pmatrix} 66 \\ 73 \end{pmatrix}$

だった。プールされた不偏共分散行列は $S = \begin{pmatrix} 100 & 40 \\ 40 & 80 \end{pmatrix}$ だった。

2 組の成績は、同じ共分散行列の正規分布にしたがうと仮定して、有意水準 $\alpha = 0.05$ で、成績の母平均値ベクトルに差があるか検定しよう。