第 17 回　2 標本問題（10.5）

村澤 康友

2024 年 11 月 26 日

今日のポイント

1. 標本を用いて 2 つの母集団を比較する問題を 2 標本問題という．
2. $\mathrm{N}\left(\mu_X,\sigma_X^2\right),\mathrm{N}\left(\mu_Y,\sigma_Y^2\right)$ から独立に抽出した無作為標本 $(X_1,\dots,X_m)$, $(Y_1,\dots,Y_n)$ の標本平均の差 $\bar{X}-\bar{Y}$，標本分散の比 s_X^2/s_Y^2 の分布を求める．
3. $\bar{X}-\bar{Y}$ の分布は $\sigma_X^2=\sigma_Y^2=\sigma^2$ なら $\left[\bar{X}-\bar{Y}-(\mu_X-\mu_Y)\right]/\sqrt{s^2(1/m+1/n)}\sim \mathrm{t}(m+n-2)$．ただし s^2 はプールした標本分散．$\sigma_X^2\neq\sigma_Y^2$ なら厳密な分布は求まらない．
4. $U\sim\chi^2(m)$ と $V\sim\chi^2(n)$ が独立のとき，$(U/m)/(V/n)\sim \mathrm{F}(m,n)$．
5. s_X^2/s_Y^2 の分布は $\left(s_X^2/s_Y^2\right)/\left(\sigma_X^2/\sigma_Y^2\right)\sim \mathrm{F}(m-1,n-1)$．

目次

1　2 標本問題（p. 204）

母集団分布 $\mathrm{N}\left(\mu_X,\sigma_X^2\right),\mathrm{N}\left(\mu_Y,\sigma_Y^2\right)$ から独立に抽出した無作為標本を $(X_1,\dots,X_m),(Y_1,\dots,Y_n)$ とする．μ_X と μ_Y の比較なら標本平均の差 $\bar{X}-\bar{Y}$，σ_X^2 と σ_Y^2 の比較なら標本分散の比 s_X^2/s_Y^2 を用いる．ただし標本分布を考慮する必要がある．

定義 1. 標本を用いて 2 つの母集団を比較する問題を ***2* 標本問題**という．

例 1. 男女別の成績の分布の比較．

2　標本平均の差

2.1　母分散が既知の場合（p. 205）

母集団分布を $\mathrm{N}\left(\mu_X,\sigma_X^2\right),\mathrm{N}\left(\mu_Y,\sigma_Y^2\right)$ とする．μ_X と μ_Y を比較したい．各母集団から独立に抽出した無作為標本を $(X_1,\dots,X_m),(Y_1,\dots,Y_n)$，標本平均を $\bar{X},\bar{Y}$ とする．

定理 1.

$$\bar{X}-\bar{Y}\sim \mathrm{N}\left(\mu_X-\mu_Y,\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}\right)$$

証明. 標本平均の分布は

$$\bar{X}\sim \mathrm{N}\left(\mu_X,\frac{\sigma_X^2}{m}\right)$$
$$\bar{Y}\sim \mathrm{N}\left(\mu_Y,\frac{\sigma_Y^2}{n}\right)$$

$\bar{X}$ と $\bar{Y}$ は独立だから

$$\bar{X}-\bar{Y}\sim \mathrm{N}\left(\mu_X-\mu_Y,\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}\right)$$

□

系 1.

$$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\sigma_X^2/m+\sigma_Y^2/n}} \sim \mathrm{N}(0,1)$$

注 1. $\bar{X}-\bar{Y}$ の累積確率は標準正規分布表から次のように求める．

$$\begin{aligned}&\Pr\left[\bar{X}-\bar{Y}\le c\right]\\ &=\Pr\left[\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\sigma_X^2/m+\sigma_Y^2/n}}\le\frac{c-(\mu_X-\mu_Y)}{\sqrt{\sigma_X^2/m+\sigma_Y^2/n}}\right]\\ &=\Phi\left(\frac{c-(\mu_X-\mu_Y)}{\sqrt{\sigma_X^2/m+\sigma_Y^2/n}}\right)\end{aligned}$$

注 2. $\mu_X=\mu_Y$ なら

$$\frac{\bar{X}-\bar{Y}}{\sqrt{\sigma_X^2/m+\sigma_Y^2/n}} \sim \mathrm{N}(0,1)$$

例 2 (p. 205)**.** 男子と女子の身長の母集団分布を，それぞれ $\mathrm{N}(172.3,30)$，$\mathrm{N}(160.2,25)$ とする．独立に抽出した男子 10 人，女子 15 人の無作為標本の標本平均をそれぞれ $\bar{X}$，$\bar{Y}$ とすると

$$\begin{aligned}\bar{X}&\sim\mathrm{N}\left(172.3,\frac{30}{10}\right)\\ \bar{Y}&\sim\mathrm{N}\left(160.2,\frac{25}{15}\right)\end{aligned}$$

$\bar{X}$ と $\bar{Y}$ は独立だから

$$\bar{X}-\bar{Y}\sim\mathrm{N}\left(12.1,\frac{14}{3}\right)$$

標準化すると

$$\frac{\bar{X}-\bar{Y}-12.1}{\sqrt{14/3}}\sim\mathrm{N}(0,1)$$

2.2 母分散が未知で等しい場合（p. 205）

$\sigma_X^2=\sigma_Y^2=\sigma^2$ とすると

$$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\sigma^2(1/m+1/n)}} \sim \mathrm{N}(0,1)$$

定義 2. $(X_1,\dots,X_m)$ と $(Y_1,\dots,Y_n)$ を**プールした標本分散**は

$$s^2:=\frac{1}{m+n-2}\left[\sum_{i=1}^m\left(X_i-\bar{X}\right)^2+\sum_{j=1}^n\left(Y_j-\bar{Y}\right)^2\right]$$

補題 1. $X\sim\chi^2(m)$ と $Y\sim\chi^2(n)$ が独立なら

$$X+Y\sim\chi^2(m+n)$$

証明. $Z_1,\dots,Z_{m+n}\sim\mathrm{N}(0,1)$ を独立とすると，χ^2 分布の定義より

$$\begin{aligned}X&:=Z_1^2+\cdots+Z_m^2\\ Y&:=Z_{m+1}^2+\cdots+Z_{m+n}^2\end{aligned}$$

したがって

$$X+Y=Z_1^2+\cdots+Z_{m+n}^2$$

χ^2 分布の定義より $X+Y\sim\chi^2(m+n)$. □

定理 2.

$$\frac{(m+n-2)s^2}{\sigma^2}\sim\chi^2(m+n-2)$$

証明. 標本分散を s_X^2,s_Y^2 とすると

$$\begin{aligned}\frac{(m-1)s_X^2}{\sigma_X^2}&\sim\chi^2(m-1)\\ \frac{(n-1)s_Y^2}{\sigma_Y^2}&\sim\chi^2(n-1)\end{aligned}$$

ただし $\sigma_X^2=\sigma_Y^2=\sigma^2$．両者は独立なので，前補題より

$$\begin{aligned}&\frac{(m+n-2)s^2}{\sigma^2}\\ &=\frac{\sum_{i=1}^m\left(X_i-\bar{X}\right)^2}{\sigma^2}+\frac{\sum_{j=1}^n\left(Y_j-\bar{Y}\right)^2}{\sigma^2}\\ &=\frac{(m-1)s_X^2}{\sigma^2}+\frac{(n-1)s_Y^2}{\sigma^2}\\ &\sim\chi^2(m+n-2)\end{aligned}$$

□

定理 3.

$$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{s^2(1/m+1/n)}}\sim\mathrm{t}(m+n-2)$$

証明. 式変形すると

$$\begin{aligned}&\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{s^2(1/m+1/n)}}\\ &=\frac{\left[\bar{X}-\bar{Y}-(\mu_X-\mu_Y)\right]/\sqrt{\sigma^2(1/m+1/n)}}{\sqrt{s^2/\sigma^2}}\\ &=\frac{\left[\bar{X}-\bar{Y}-(\mu_X-\mu_Y)\right]/\sqrt{\sigma^2(1/m+1/n)}}{\sqrt{[(m+n-2)s^2/\sigma^2]/(m+n-2)}}\end{aligned}$$

ここで

$$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\sigma^2(1/m+1/n)}} \sim \mathrm{N}(0,1)$$
$$\frac{(m+n-2)s^2}{\sigma^2} \sim \chi^2(m+n-2)$$

分子と分母の独立性も証明できる（省略）. □

注 3. $\bar{X}-\bar{Y}$ の累積確率は t 分布表から次のように求める.

$$\begin{aligned}
&\Pr\left[\bar{X}-\bar{Y} \le c\right] \\
&= \Pr\left[\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{s^2(1/m+1/n)}} \le \frac{c-(\mu_X-\mu_Y)}{\sqrt{s^2(1/m+1/n)}}\right] \\
&= \Pr\left[\mathrm{t}(m+n-2) \le \frac{c-(\mu_X-\mu_Y)}{\sqrt{s^2(1/m+1/n)}}\right]
\end{aligned}$$

注 4. $\mu_X = \mu_Y$ なら

$$\frac{\bar{X}-\bar{Y}}{\sqrt{s^2(1/m+1/n)}} \sim \mathrm{t}(m+n-2)$$

2.3　母分散が未知で異なる場合（p. 206）

$\sigma_X^2 \ne \sigma_Y^2$ だと σ_X^2, σ_Y^2 に分布が依存しない統計量を作れない. ただし大数の法則と中心極限定理より

$$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{s_X^2/m+s_Y^2/n}} \overset{a}{\sim} \mathrm{N}(0,1)$$

もっとよい近似もある（ウェルチの近似）.

3　標本分散の比

3.1　F 分布（p. 207）

定義 3. $U \sim \chi^2(m)$ と $V \sim \chi^2(n)$ が独立のとき, $(U/m)/(V/n)$ の分布を**自由度** (m,n) **の** F **分布**という.

注 5. $\mathrm{F}(m,n)$ と書く.

注 6. 累積確率は F 分布表を参照.

注 7. $X \sim \mathrm{F}(m,n)$ なら $1/X \sim \mathrm{F}(n,m)$.

注 8. $t \sim \mathrm{t}(n)$ なら $t^2 \sim \mathrm{F}(1,n)$.

例 3. F 分布の pdf の例は図 1 の通り.

3.2　母平均が既知の場合

母集団分布を $\mathrm{N}\left(\mu_X, \sigma_X^2\right), \mathrm{N}\left(\mu_Y, \sigma_Y^2\right)$ とする. σ_X^2 と σ_Y^2 を比較したい. 各母集団から独立に抽出した無作為標本を $(X_1,\dots,X_m),(Y_1,\dots,Y_n)$, 標本分散を $\hat{\sigma}_X^2, \hat{\sigma}_Y^2$ とする.

定理 4.
$$\frac{\hat{\sigma}_X^2/\hat{\sigma}_Y^2}{\sigma_X^2/\sigma_Y^2} \sim \mathrm{F}(m,n)$$

証明. 標本分散の分布は

$$\frac{m\hat{\sigma}_X^2}{\sigma_X^2} \sim \chi^2(m)$$
$$\frac{n\hat{\sigma}_Y^2}{\sigma_Y^2} \sim \chi^2(n)$$

両者は独立なので

$$\begin{aligned}
\frac{\hat{\sigma}_X^2/\hat{\sigma}_Y^2}{\sigma_X^2/\sigma_Y^2} &= \frac{\hat{\sigma}_X^2/\sigma_X^2}{\hat{\sigma}_Y^2/\sigma_Y^2} \\
&= \frac{\left(m\hat{\sigma}_X^2/\sigma_X^2\right)/m}{\left(n\hat{\sigma}_Y^2/\sigma_Y^2\right)/n} \\
&\sim \mathrm{F}(m,n)
\end{aligned}$$

□

3.3　母平均が未知の場合（p. 208）

標本分散を s_X^2, s_Y^2 とする.

定理 5.

$$\frac{s_X^2/s_Y^2}{\sigma_X^2/\sigma_Y^2} \sim \mathrm{F}(m-1,n-1)$$

証明. 標本分散の分布は

$$\frac{(m-1)s_X^2}{\sigma_X^2} \sim \chi^2(m-1)$$
$$\frac{(n-1)s_Y^2}{\sigma_Y^2} \sim \chi^2(n-1)$$

両者は独立なので

$$\begin{aligned}
\frac{s_X^2/s_Y^2}{\sigma_X^2/\sigma_Y^2} &= \frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} \\
&= \frac{\left[(m-1)s_X^2/\sigma_X^2\right]/(m-1)}{\left[(n-1)s_Y^2/\sigma_Y^2\right]/(n-1)} \\
&\sim \mathrm{F}(m-1,n-1)
\end{aligned}$$

□

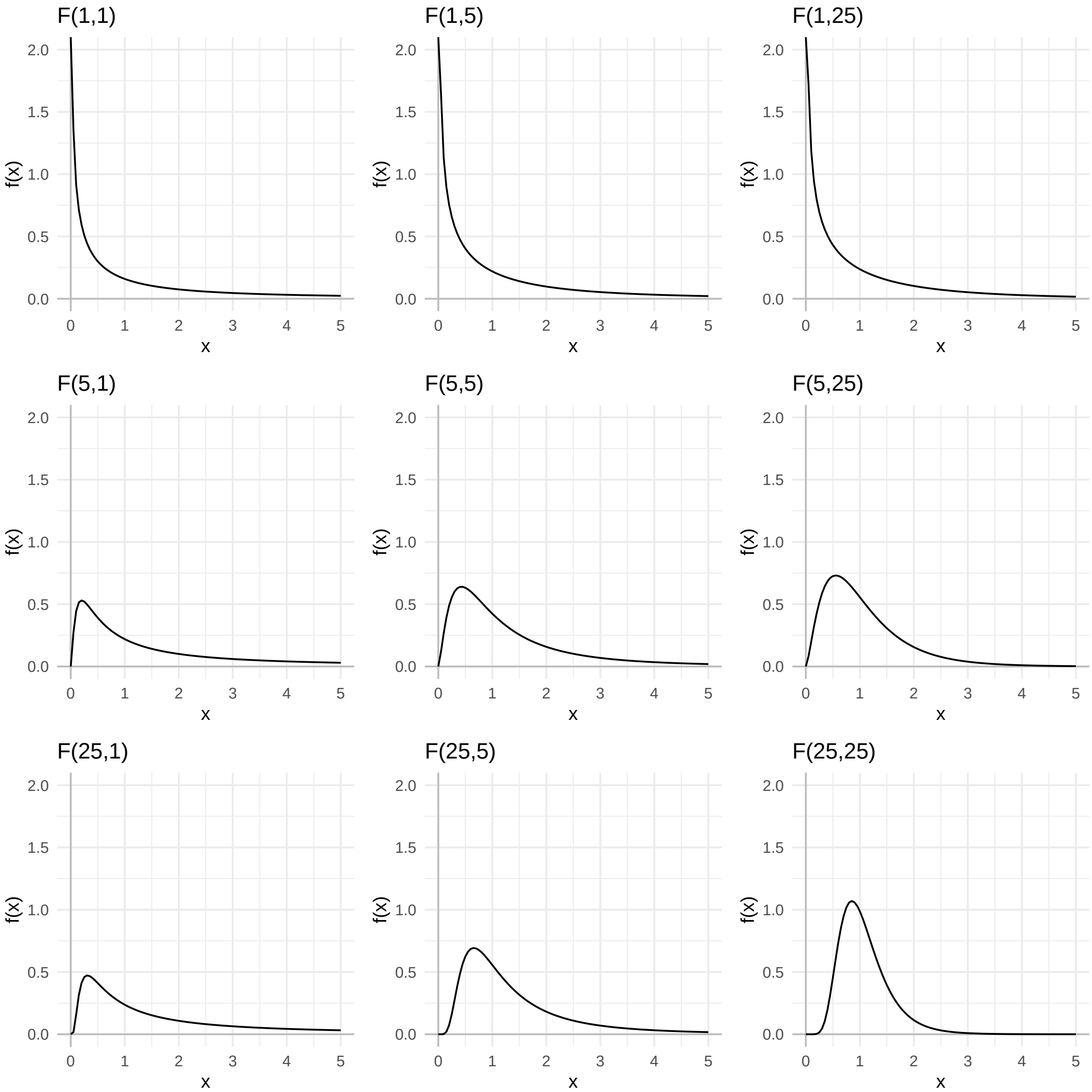

図 1　F 分布の pdf の例

注 9. s_X^2/s_Y^2 の累積確率は F 分布表から次のように求める.

$$\begin{aligned}\Pr\left[\frac{s_X^2}{s_Y^2} \le c\right] &= \Pr\left[\frac{s_X^2/s_Y^2}{\sigma_X^2/\sigma_Y^2} \le \frac{c}{\sigma_X^2/\sigma_Y^2}\right] \\ &= \Pr\left[\mathrm{F}(m-1, n-1) \le \frac{c}{\sigma_X^2/\sigma_Y^2}\right]\end{aligned}$$

注 10. $\sigma_X^2 = \sigma_Y^2$ なら

$$\frac{s_X^2}{s_Y^2} \sim \mathrm{F}(m-1, n-1)$$

例 4 (p. 209). $\mathrm{N}\left(\mu_X, \sigma^2\right), \mathrm{N}\left(\mu_Y, \sigma^2\right)$ から独立に抽出した大きさ $10, 15$ の無作為標本の標本分散をそれぞれ s_X^2, s_Y^2 とする（母分散は等しい）. $s_X^2/s_Y^2 > 3$ の確率は

$$\begin{aligned}\Pr\left[\frac{s_X^2}{s_Y^2} > 3\right] &= \Pr[\mathrm{F}(9, 14) > 3] \\ &\approx .03\end{aligned}$$

4　今日のキーワード

2 標本問題，プールした標本分散，χ^2 分布の再生性，標本平均の差の分布（母分散が既知・未知で等しい・未知で異なる），F 分布，標本分散の比の分布（母平均が既知・未知）

5　次回までの準備

提出　宿題 5

復習　教科書第 10 章 5 節，復習テスト 17

予習　教科書第 11 章 1–2, 4 節