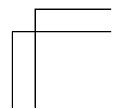
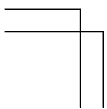
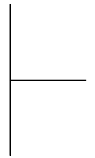
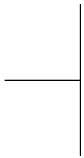


ノンパラメトリック法 , セミパラメトリック法 のアルゴリズム



第 1 章

1 標本問題

1.1 ノンパラメトリック法

$(x_1, \dots, x_n)$ を連続型の分布関数 $F_\sigma(x - \mu) = F\left(\frac{x - \mu}{\sigma}\right)$ をもつ母集団からの大きさ n の無作為標本の観測値とする. さらに, 密度関数 $f_\sigma(x) = F'_\sigma(x)$ は $f_\sigma(-x) = f_\sigma(x)$ を満たす 0 について対称な関数. すなわち $x_1, \dots, x_n$ は互いに独立で各 x_i は同一の連続型の分布 $F_\sigma(x - \mu)$ をもつ.

[1] 符号付き順位検定; 帰無仮説 $H_0 : \mu = \mu_0$ ($\sigma^2 > 0$) v.s. 対立仮説 $H_1 : \mu \neq \mu_0$ ($\sigma^2 > 0$) の有意水準 α の検定を考える. $y_i \equiv x_i - \mu_0$ ($i = 1, \dots, n$) とおき, $|y_1|, \dots, |y_n|$ を小さい方から並べたときの $|y_i|$ の順位を r_i^+ とし, y_i の符号を $\text{sign}(y_i)$ で定義する. ただし, $|y_1|, \dots, |y_n|$ の中に同じ値がある場合を考慮して,

$$r_i^+ \equiv \frac{\sum_{k=1}^n I(|y_k| \leq |y_i|) + \sum_{k=1}^n I(|y_k| < |y_i|) + 1}{2}$$

とし,

$$\text{sign}(y) = \begin{cases} 1 & (y > 0 \text{ のとき}) \\ 0 & (y = 0 \text{ のとき}) \\ -1 & (y < 0 \text{ のとき}). \end{cases}$$

とおく, ただし,

$$I(A) = \begin{cases} 1 & (\text{不等式 } A \text{ が真のとき}) \\ 0 & (\text{不等式 } A \text{ が偽のとき}). \end{cases}$$

とする. このとき,

$$t_r \equiv \sum_{i=1}^n \text{sign}(y_i) \cdot r_i^+$$

とおき, $|t_r|$ が大きいとき H_0 を棄却する.

まず, n が 20 未満のときの手順を述べる.

- 手順 1. $0 < \alpha < 1$ となるように有意水準 α を決め, t_r を計算する.
- 手順 2. シミュレーションの繰り返し数 K として 1000 を採り,
 $i = 1, 2, \dots, K-1$ について以下の手順 2.1 と手順 2.2 を繰り返す.

・ 手順 2.1. 乗算合同法の疑似一様乱数を n 個生成させ, それを

$u_1(i), \dots, u_n(i)$ とする.

・ 手順 2.2. $t_r(i) \equiv \sum_{j=1}^n \text{sign}(u_j(i) - \frac{1}{2}) \cdot r_j^+$ を計算する.

★ 手順 2.1, 2.2 をコインを使った例で説明すると, コインを n 回投げ表を 1 裏を -1 としたものを $s_1(i), s_2(i), \dots, s_n(i)$ とし, $\sum_{j=1}^n s_j(i) \cdot r_j^+$ を計算しそれを $t_r(i)$ とする.

- 手順 3. $N_u \equiv \sum_{i=1}^{K-1} I(|t_r| \leq |t_r(i)|)$, $N_\ell \equiv \sum_{i=1}^{K-1} I(|t_r| < |t_r(i)|)$ とおく,
 一様乱数 1 つを生成させ, それを v とする.

$$\frac{N_u}{K} \leq \alpha \text{ または } (\frac{N_u}{K} > \alpha \text{ かつ } \frac{N_\ell}{K} < \alpha \text{ かつ } v < \frac{K \cdot \alpha - N_\ell}{N_u - N_\ell})$$

$\Rightarrow H_0$ を棄却する.

★ $|t_r(1)|, |t_r(2)|, \dots, |t_r(K-1)|$ のうち, 大きい方から $K\alpha$ 番目の値を $|t_\alpha|$ とすれば, (4.1) より $|t_\alpha|$ は上側 100α 点の推定となり, $|t_\alpha| < |t_r|$ ならば H_0 を棄却する.

次に $n \geq 20$ のときには $z_r = \frac{t_r}{\sqrt{\sum_{i=1}^n (r_i^+)^2}}$ は帰無仮説の下で近似的に標準

正規分布に従うので, 検定方式は楽で,

$$p \text{ 値} \approx 2 \int_{|z_r|}^{\infty} \varphi(y) dy \leq \alpha \implies H_0 \text{ を棄却する,}$$

ただし, $\varphi(y)$ は標準正規分布の密度関数.

帰無仮説 H_0 v.s. 対立仮説 $H_2: \mu > \mu_0$ ($\sigma^2 > 0$) の検定に対しては t_r が大きいとき H_0 を棄却する. 有意水準 α を決め, 一様乱数 1 つを生成させそ

れを v^* とする. $N_u^* \equiv \sum_{i=1}^{K-1} I(t_r \leq t_r(i)), N_\ell^* \equiv \sum_{i=1}^{K-1} I(t_r < t_r(i))$ とおく.

$$\frac{N_u^*}{K} \leq \alpha \text{ または } \left(\frac{N_u^*}{K} > \alpha \text{ かつ } \frac{N_\ell^*}{K} < \alpha \text{ かつ } v^* < \frac{K \cdot \alpha - N_\ell^*}{N_u^* - N_\ell^*} \right)$$

$\implies H_0$ を棄却する.

$n \geq 20$ のときには,

$$p \text{ 値} \approx \int_{z_r}^{\infty} \varphi(y) dy \leq \alpha \implies H_0 \text{ を棄却する.}$$

帰無仮説 H_0 v.s. 対立仮説 $H_3: \mu < \mu_0$ ($\sigma^2 > 0$) の有意水準 α の検定に対しては $-t_r$ が大きいとき H_0 を棄却する. 有意水準 α を決め, 一様乱数 1 つを生成させそれを v' とする.

$$N'_u \equiv \sum_{i=1}^{K-1} I(-t_r \leq -t_r(i)), \quad N'_\ell \equiv \sum_{i=1}^{K-1} I(-t_r < -t_r(i))$$

とおく.

$$\frac{N'_u}{K} \leq \alpha \text{ または } \left(\frac{N'_u}{K} > \alpha \text{ かつ } \frac{N'_\ell}{K} < \alpha \text{ かつ } v' < \frac{K \cdot \alpha - N'_\ell}{N'_u - N'_\ell} \right)$$

$\implies H_0$ を棄却する.

$n \geq 20$ のときには,

$$p \text{ 値} \approx \int_{-z_r}^{\infty} \varphi(y) dy \leq \alpha \implies H_0 \text{ を棄却する.}$$

[2] 平均の順位推定量;

$$\hat{\mu} = \text{med}\left\{\frac{x_i + x_j}{2} : 1 \leq i \leq j \leq n\right\}$$

$$= \left(\frac{n(n+1)}{2}\right)\text{個の値 } \left\{\frac{x_i+x_j}{2} : 1 \leq i \leq j \leq n\right\} \text{ の中央値}$$

を μ の順位推定量と呼ぶ。

1.2 セミパラメトリック法

$(x_1, \dots, x_n)$ を連続型の分布 $F(\frac{x-\mu}{\sigma})$ をもつ母集団からの大きさ n の無作為標本の観測値とする。さらに、密度関数 $f(x) = F'(x)$ は $f(-x) = f(x)$ を満たす 0 について対称な関数で一般性を失うことなく $\int_{-\infty}^{\infty} x^2 dF(x) = 1$ とする。すなわち $x_1, \dots, x_n$ は互いに独立で各 x_i は同一の連続型の分布 $F(\frac{x-\mu}{\sigma})$ をもつ。

[1] M 検定; 帰無仮説 $H_0 : \mu = \mu_0$ ($\sigma^2 > 0$) v.s. 対立仮説 $H_1 : \mu \neq \mu_0$ ($\sigma^2 > 0$) の有意水準 α の検定を考える。 $y_i \equiv x_i - \mu_0$ ($i = 1, \dots, n$) とおき、 y_i の符号を $\text{sign}(y_i)$ で定義する。 M 関数 $\psi(\cdot)$ を $\psi(x) = \max(\min(x, b), -b)$ で定義し、尺度母数を $\hat{\sigma}_n = \frac{1}{0.6745}(|y_1|, \dots, |y_n| \text{ の中央値})$ で推定し、

$$t_m \equiv \sum_{i=1}^n \psi\left(\frac{y_i}{\hat{\sigma}_n}\right) = \sum_{i=1}^n \text{sign}(y_i) \cdot \psi\left(\frac{|y_i|}{\hat{\sigma}_n}\right)$$

とおき、 $|t_m|$ が大きいとき H_0 を棄却する。関数 $\psi(x)$ の式の中の b の決め方は『統計科学』の 5.6 節で説明されているが今後 $b = 1.399$ とする。

まず、 n が 20 未満のときの手順を述べる。

- 手順 1. $0 < \alpha < 1$ となるように有意水準 α を決め、 t_m を計算する。
- 手順 2. シミュレーションの繰り返し数 K として 1000 を採り、
 $i = 1, 2, \dots, K-1$ について以下の手順 2.1 と手順 2.2 を繰り返す。
 - ・ 手順 2.1. 乗算合同法の疑似一様乱数を n 個生成させ、それを $u_1(i), \dots, u_n(i)$ とする。
 - ・ 手順 2.2. $t_m(i) \equiv \sum_{j=1}^n \text{sign}(u_j(i) - \frac{1}{2}) \cdot \psi\left(\frac{|u_j(i)|}{\hat{\sigma}_n}\right)$ を計算する。

- 手順 3. $N_u \equiv \sum_{i=1}^{K-1} I(|t_m| \leq |t_m(i)|)$, $N_\ell \equiv \sum_{i=1}^{K-1} I(|t_m| < |t_m(i)|)$ とおく,
一様乱数 1 つを生成させ, それを v とする.

$$\frac{N_u}{K} \leq \alpha \text{ または } (\frac{N_u}{K} > \alpha \text{ かつ } \frac{N_\ell}{K} < \alpha \text{ かつ } v < \frac{K \cdot \alpha - N_\ell}{N_u - N_\ell})$$

$\Rightarrow H_0$ を棄却する.

$$n \geq 20 \text{ のときには } z_m = \frac{t_m}{\sqrt{\sum_{i=1}^n \{\psi(\frac{y_i}{\sigma_n})\}^2}} \text{ は帰無仮説の下で近似的に標準}$$

正規分布に従うので, 検定方式は楽で,

$$p \text{ 値} \approx 2 \int_{|z_m|}^{\infty} \varphi(y) dy \leq \alpha \Rightarrow H_0 \text{ を棄却する.}$$

帰無仮説 H_0 v.s. 対立仮説 $H_2: \mu > \mu_0$ ($\sigma^2 > 0$) の検定に対しては t_m が大きいとき H_0 を棄却する. 有意水準 α を決め, 一様乱数 1 つを生成させそれを v^* とする. $N_u^* \equiv \sum_{i=1}^{K-1} I(t_m \leq t_m(i))$, $N_\ell^* \equiv \sum_{i=1}^{K-1} I(t_m < t_m(i))$ とおく.

$$\frac{N_u^*}{K} \leq \alpha \text{ または } (\frac{N_u^*}{K} > \alpha \text{ かつ } \frac{N_\ell^*}{K} < \alpha \text{ かつ } v^* < \frac{K \cdot \alpha - N_\ell^*}{N_u^* - N_\ell^*})$$

$\Rightarrow H_0$ を棄却する.

$n_1, n_2 \geq 20$ のときには,

$$p \text{ 値} \approx \int_{z_m}^{\infty} \varphi(y) dy \leq \alpha \Rightarrow H_0 \text{ を棄却する.}$$

帰無仮説 H_0 v.s. 対立仮説 $H_3: \mu < \mu_0$ ($\sigma^2 > 0$) の有意水準 α の検定に対しては $-t_m$ が大きいとき H_0 を棄却する. 有意水準 α を決め, 一様乱数 1 つを生成させそれを v' とする. $N_u' \equiv \sum_{i=1}^{K-1} I(-t_m \leq -t_m(i))$,

$$N_\ell' \equiv \sum_{i=1}^{K-1} I(-t_m < -t_m(i)) \text{ とおく.}$$

$$\frac{N_u'}{K} \leq \alpha \text{ または } (\frac{N_u'}{K} > \alpha \text{ かつ } \frac{N_\ell'}{K} < \alpha \text{ かつ } v' < \frac{K \cdot \alpha - N_\ell'}{N_u' - N_\ell'})$$

$\Rightarrow H_0$ を棄却する.

$n_1, n_2 \geq 20$ のときには,

$$p \text{ 値} \approx \int_{-z_m}^{\infty} \varphi(y) dy \leq \alpha \implies H_0 \text{ を棄却する.}$$

[2] 平均の M 推定量;

$$t_m(\theta) \equiv \sum_{i=1}^n \psi\left(\frac{x_i - \theta}{\check{\sigma}_n}\right) = 0$$

の解 $\check{\theta}$ を μ の点推定量で M 推定量と呼ばれている, ただし,

$$\check{\sigma}_n = \frac{1}{0.6745} (|x_1 - \text{med}(x)|, \dots, |x_n - \text{med}(x)| \text{ の中央値}) \text{ で}$$

$$\text{med}(x) = (x_1, \dots, x_n \text{ の中央値}).$$

$\varepsilon > 0$ 以下の誤差の範囲で $\check{\theta}$ を探すには, $\check{\theta}_\ell \equiv \min\{x_i - \check{\sigma}_n b : i = 1, \dots, n\}$,
 $\check{\theta}_u \equiv \max\{x_i + \check{\sigma}_n b : i = 1, \dots, n\}$ とおくと, $t_m(\check{\theta}_\ell) > 0$, $t_m(\check{\theta}_u) < 0$, $\check{\theta}_\ell < \check{\theta}_u$ が成り立つ,

$$\check{\theta}_x \equiv \frac{\check{\theta}_\ell + \check{\theta}_u}{2} \text{ とし}$$

$$t_m(\check{\theta}_x) > 0 \implies \text{新たに } \check{\theta}_\ell \equiv \check{\theta}_x \text{ とおき,}$$

$$t_m(\check{\theta}_x) < 0 \implies \text{新たに } \check{\theta}_u \equiv \check{\theta}_x \text{ とおく.}$$

$t_m(\check{\theta}_x) = 0$ または $\max\{|\check{\theta}_u - \check{\theta}_x|, |\check{\theta}_x - \check{\theta}_\ell|\} < 10^{-5}$ になるまで 四角の枠内を繰り返す.

第 2 章

2 標本問題

2.1 ノンパラメトリック法

$(x_1, \dots, x_{n_1})$ を平均 μ_1 の連続型の分布関数 $F_\sigma(x - \mu_1) = F(\frac{x - \mu_1}{\sigma})$ をもつ母集団からの大きさ n_1 の無作為標本の実現値, $(y_1, \dots, y_{n_2})$ を平均 μ_2 の分布関数 $F_\sigma(x - \mu_2) = F(\frac{x - \mu_2}{\sigma})$ をもつ母集団からの大きさ n_2 の無作為標本の実現値とする. すなわち, $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$ は互いに独立な確率変数の実現値で, 各 x_i は同一の分布関数 $F_\sigma(x - \mu_1)$ をもち, 各 y_j は同一の分布関数 $F_\sigma(x - \mu_2)$ をもつとする. $n \equiv n_1 + n_2$ と置く.

[1] ウィルコクソンの順位検定; 帰無仮説 $H_0 : \mu_1 = \mu_2$ v.s. 対立仮説 $H_1 : \mu_1 \neq \mu_2$ の有意水準 α の検定を考える. $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$ を小さい方から並べたときの x_i の順位を r_i , y_j の順位を r_{n_1+j} とする. ただし, $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$ の中に同じ値がある場合を考慮して,

$$r_i \equiv \frac{\sum_{k=1}^{n_1} I(x_k \leq x_i) + \sum_{k=1}^{n_2} I(y_k \leq x_i)}{2} + \frac{\sum_{k=1}^{n_1} I(x_k < x_i) + \sum_{k=1}^{n_2} I(y_k < x_i) + 1}{2},$$

$$r_{n_1+j} \equiv \frac{\sum_{k=1}^{n_1} I(x_k \leq y_j) + \sum_{k=1}^{n_2} I(y_k \leq y_j)}{2} + \frac{\sum_{k=1}^{n_1} I(x_k < y_j) + \sum_{k=1}^{n_2} I(y_k < y_j) + 1}{2}$$

とする。このとき、 $t_r \equiv \frac{1}{n_1} \sum_{i=1}^{n_1} (r_i - \frac{n+1}{2})$ とおき、 $|t_r|$ が大きいとき H_0 を棄却する。

まず、 n_1 , または n_2 が 20 未満のときの手順を述べる。

- 手順 1. $0 < \alpha < 1$ となるように有意水準 α を決め、

$$r(0) \equiv (r_1(0), \dots, r_n(0)) \equiv (r_1, \dots, r_n)$$

とおき、統計量

$$t_r \equiv \frac{1}{n_1} \sum_{i=1}^{n_1} (r_i(0) - \frac{n+1}{2})$$

を計算する。

★ 手順 1 をカードを使った例で説明すると、表に $r_1, r_2, \dots, r_n$ の番号の書かれたカードを n 枚作り、最初の n_1 枚までの番号をすべて加えた値を sum とし $\frac{1}{n_1}(sum - \frac{n_1(n+1)}{2})$ を t_r とする。

- 手順 2. シミュレーションの繰り返し数 K として 1000 を採り、
 $i = 1, 2, \dots, K-1$ について以下の手順 2.1 から手順 2.4 までを繰り返す。
 - ・ 手順 2.1. 乗算合同法の疑似一様乱数を n_2 個生成させ、それを $u_1(i), \dots, u_{n_2}(i)$ とする。
 - ・ 手順 2.2. $j = 1, \dots, n_2$ について以下の手順 2.2.1 と手順 2.2.2 を繰り返す。
 - 手順 2.2.1. $\ell \equiv [(n-j+1) \cdot u_j(i)] + 1$ とおき、 $\ell = n-j+2$ ならば新たに $\ell \equiv 1$ とおく、ただし、 $[x]$ は x を越えない最大の整数とする。
 - 手順 2.2.2. n 次元ベクトル $r(i-1) \equiv (r_1(i-1), \dots, r_n(i-1))$ の $n-j+1$ 成分 $r_{n-j+1}(i-1)$ と ℓ 成分 $r_\ell(i-1)$ とを入れ替え

たものを新たに, $r(i-1)$ とおく.

- ・手順 2.3. 手順 2.2 で成分入れ替えによって最終的に決定された $r(i-1)$ を $r(i)$ とおく.

- ・手順 2.4. $r(i) \equiv (r_1(i), \dots, r_n(i))$ より

$$t_r(i) \equiv \frac{1}{n_1} \sum_{j=1}^{n_1} (r_j(i) - \frac{n+1}{2})$$

を計算する.

★ カードを裏にして, 十分繰った後に n_1 枚のカードを抜き, それらの番号をすべて加えた値を $sum(1)$ とし $\frac{1}{n_1}(sum(1) - \frac{n_1(n+1)}{2})$ を $t_r(1)$ とする. 以下同様の操作により, $t_r(2), \dots, t_r(K-1)$ を求める.

- 手順 3. $N_u \equiv \sum_{i=1}^{K-1} I(|t_r| \leq |t_r(i)|)$, $N_\ell \equiv \sum_{i=1}^{K-1} I(|t_r| < |t_r(i)|)$ とおく.

一様乱数 1 つを生成させ, それを v とする.

$$\frac{N_u}{K} \leq \alpha \text{ または } (\frac{N_u}{K} > \alpha \text{ かつ } \frac{N_\ell}{K} < \alpha \text{ かつ } v < \frac{K \cdot \alpha - N_\ell}{N_u - N_\ell})$$

$\Rightarrow H_0$ を棄却する.

★ $|t_r(1)|, |t_r(2)|, \dots, |t_r(K-1)|$ のうち, 大きい方から $K\alpha$ 番目の値を $|t_\alpha|$ とすれば, (4.1) より $|t_\alpha|$ は上側 100α 点の推定となり, $|t_\alpha| < |t_r|$ ならば H_0 を棄却する.

次に $n_1, n_2 \geq 20$ のときには

$$z_r \equiv \sqrt{\frac{n(n-1)}{n_1 n_2}} \frac{n_1 t_r}{\sqrt{\sum_{j=1}^n (r_j - \frac{n+1}{2})^2}}$$

は帰無仮説の下で近似的に標準正規分布に従うので, 検定方式は楽で,

$$p \text{ 値} \approx 2 \int_{|z_r|}^{\infty} \varphi(y) dy \leq \alpha \Rightarrow H_0 \text{ を棄却する.}$$

帰無仮説 $H_0: \mu_1 = \mu_2$ v.s. 対立仮説 $H_2: \mu_1 > \mu_2$

の有意水準 α の検定に対しては t_r が大きいとき H_0 を棄却する. 有意水準 α を決め, 一様乱数 1 つを生成させそれを v^* とする. $N_u^* \equiv \sum_{i=1}^{K-1} I(t_r \leq t_r(i))$,

$N_\ell^* \equiv \sum_{i=1}^{K-1} I(t_r < t_r(i))$ とおく.

$$\frac{N_u^*}{K} \leq \alpha \text{ または } \left(\frac{N_u^*}{K} > \alpha \text{ かつ } \frac{N_\ell^*}{K} < \alpha \text{ かつ } v^* < \frac{K \cdot \alpha - N_\ell^*}{N_u^* - N_\ell^*}\right)$$

$\Rightarrow H_0$ を棄却する.

$n_1, n_2 \geq 20$ のときには,

$$p \text{ 値} \approx \int_{z_r}^{\infty} \varphi(y) dy \leq \alpha \Rightarrow H_0 \text{ を棄却する.}$$

[2] ホッジス・レーマン順位推定量;

$$\begin{aligned} \hat{\delta} &= \text{med}\{x_i - y_j : i = 1, \dots, n_1, j = 1, \dots, n_2\} \\ &= (n_1 n_2 \text{個の値 } \{x_i - y_j : i = 1, \dots, n_1, j = 1, \dots, n_2\} \text{ の中央値}) \end{aligned}$$

を $\mu_1 - \mu_2$ の順位推定量と呼ぶ.

アルゴリズムは $i = 1, \dots, n_1, j = 1, \dots, n_2$ に対して, $\ell \equiv (i-1) \cdot n_2 + j$ とおき, さらに, $w(\ell) \equiv x_i - y_j$ とおく. $n_1 \cdot n_2$ 個の $w(\ell)$ について 例 4.4 の高速標本中央値のプログラムをサブルーチンにしたものをつかって $\hat{\delta}$ を求める.

2.2 セミパラメトリック法

$(x_1, \dots, x_{n_1})$ を平均 μ_1 の連続型の分布関数 $F(\frac{x - \mu_1}{\sigma})$ をもつ母集団からの大きさ n_1 の無作為標本の実現値, $(y_1, \dots, y_{n_2})$ を平均 μ_2 の分布関数 $F(\frac{x - \mu_2}{\sigma})$ をもつ母集団からの大きさ n_2 の無作為標本の実現値とする. $n \equiv n_1 + n_2$ と置く.

[6] M 検定; 帰無仮説 $H_0 : \mu_1 = \mu_2$ ($\sigma^2 > 0$) v.s. 対立仮説 $H_1 : \mu_1 \neq \mu_2$ ($\sigma^2 > 0$) の有意水準 α の検定を考える. 関数 $\psi(\cdot)$ を $\psi(x) = \max(\min(x, b), -b)$

で定義し, $\tilde{\mu} = \frac{1}{n}(n_1\bar{x} + n_2\bar{y})$ とおき, 尺度母数を

$$\hat{\sigma}_n = \frac{\sqrt{\pi}}{\sqrt{2} \cdot n} \left(\sum_{i=1}^{n_1} |x_i - \tilde{\mu}| + \sum_{j=1}^{n_2} |y_j - \tilde{\mu}| \right) \text{ で推定し,}$$

$$t_m \equiv \frac{1}{n_1} \sum_{i=1}^{n_1} \psi\left(\frac{x_i - \tilde{\mu}}{\hat{\sigma}_n}\right) - \frac{1}{n_2} \sum_{j=1}^{n_2} \psi\left(\frac{y_j - \tilde{\mu}}{\hat{\sigma}_n}\right)$$

で定義し, $|t_m|$ が大きいとき H_0 を棄却する. 関数 $\psi(x)$ の式の中の b の決め方は『統計科学』の 5.6 節で説明されているが今後 $b = 1.399$ とする.

まず, n_1 または n_2 が 20 未満のときの手順を述べる.

- 手順 1. $0 < \alpha < 1$ となるように有意水準 α を決め,

$$\begin{aligned} w(0) &\equiv (w_1(0), \dots, w_n(0)) \\ &\equiv \left(\psi\left(\frac{x_1 - \tilde{\mu}}{\hat{\sigma}_n}\right), \dots, \psi\left(\frac{x_{n_1} - \tilde{\mu}}{\hat{\sigma}_n}\right), \psi\left(\frac{y_1 - \tilde{\mu}}{\hat{\sigma}_n}\right), \dots, \psi\left(\frac{y_{n_2} - \tilde{\mu}}{\hat{\sigma}_n}\right) \right) \end{aligned}$$

とおき, 統計量

$$t_m \equiv \frac{1}{n_1} \sum_{i=1}^{n_1} w_i(0) - \frac{1}{n_2} \sum_{j=1}^{n_2} w_{n_1+j}(0)$$

を計算する.

★ 手順 1 をカードを使った例で説明すると, 表に $w_1(0), \dots, w_n(0)$ の数値の書かれたカードを n 枚作り, 最初の n_1 枚までの番号をすべて加えた値を $sum1$ とし, 後の n_2 枚までの番号をすべて加えた値を $sum2$ とする. $\frac{sum1}{n_1} - \frac{sum2}{n_2}$ を t_m とする.

- 手順 2. シミュレーションの繰り返し数 K として 1000 を採り,
 - $i = 1, 2, \dots, K-1$ について以下の手順 2.1 から手順 2.4 までを繰り返す.
 - ・ 手順 2.1. 乗算合同法による $U(0, 1]$ 上の疑似一様乱数を n_2 個生成させ, それを $u_1(i), \dots, u_{n_2}(i)$ とする.
 - ・ 手順 2.2. $j = 1, \dots, n_2$ について以下の手順 2.2.1 と手順 2.2.2 を繰り返す.
 - 手順 2.2.1. $\ell \equiv [(n-j+1) \cdot u_j(i)] + 1$ とおき, $\ell = n-j+2$ ならば新たに $\ell \equiv 1$ とおく,

ただし, $[x]$ は x を越えない最大の整数とする.

手順 2.2.2. n 次元ベクトル $w(i-1) \equiv (w_1(i-1), \dots, w_n(i-1))$

の $n-j+1$ 成分 $w_{n-j+1}(i-1)$ と ℓ 成分 $w_\ell(i-1)$

とを入れ替えたものを新たに, $w(i-1)$ とおく.

・手順 2.3. 手順 2.2 で成分入れ替えによって最終的に決定された

$w(i-1)$ を $w(i)$ とおく.

・手順 2.4. $w(i) \equiv (w_1(i), \dots, w_n(i))$ より

$$t_m(i) \equiv \frac{1}{n_1} \sum_{j=1}^{n_1} w_j(i) - \frac{1}{n_2} \sum_{j=n_1+1}^n w_j(i)$$

を計算する.

★ カードを裏にして, 十分繰った後に n_1 枚のカードを抜き, それらの値をすべて加えた値を $sum1(1)$, 残りの n_2 枚のカードの値の和を $sum2(1)$ とし $\frac{sum1(1)}{n_1} - \frac{sum2(1)}{n_2}$ を $t_m(1)$ とする. 以下同様の操作により, $t_m(2), \dots, t_m(K-1)$ を求める.

● 手順 3. $N_u \equiv \sum_{i=1}^{K-1} I(|t_m| \leq |t_m(i)|)$, $N_\ell \equiv \sum_{i=1}^{K-1} I(|t_m| < |t_m(i)|)$ とおく.

一様乱数 1 つを生成させ, それを v とする.

$$\frac{N_u}{K} \leq \alpha \text{ または } \left(\frac{N_u}{K} > \alpha \text{ かつ } \frac{N_\ell}{K} < \alpha \text{ かつ } v < \frac{K \cdot \alpha - N_\ell}{N_u - N_\ell} \right)$$

$\Rightarrow H_0$ を棄却する.

★ $|t_m(1)|, |t_m(2)|, \dots, |t_m(K-1)|$ のうち, 大きい方から $K\alpha$ 番目の値を $|t_\alpha|$ とすれば, (4.1) より $|t_\alpha|$ は上側 100α 点の推定となり, $|t_\alpha| < |t_m|$ ならば H_0 を棄却する.

$$\text{次に, } n_1, n_2 \geq 20 \text{ のときには } \bar{\psi}(\mathbf{x}, \mathbf{y}) \equiv \frac{1}{n} \left\{ \sum_{i=1}^{n_1} \psi\left(\frac{x_i - \tilde{\mu}}{\hat{\sigma}_n}\right) + \sum_{j=1}^{n_2} \psi\left(\frac{y_j - \tilde{\mu}}{\hat{\sigma}_n}\right) \right\}$$

とおけば,

$$z_m \equiv \frac{\sqrt{n_1 n_2 (n-1)} \cdot t_m}{\sqrt{n \left[\sum_{i=1}^{n_1} \left\{ \psi \left(\frac{x_i - \tilde{\mu}}{\check{\sigma}_n} \right) - \bar{\psi}(\mathbf{x}, \mathbf{y}) \right\}^2 + \sum_{j=1}^{n_2} \left\{ \psi \left(\frac{y_j - \tilde{\mu}}{\check{\sigma}_n} \right) - \bar{\psi}(\mathbf{x}, \mathbf{y}) \right\}^2 \right]}}$$

は帰無仮説の下で近似的に標準正規分布に従うので, 検定方式は,

$$p \text{ 値} \approx 2 \int_{|z_m|}^{\infty} \varphi(y) dy \leq \alpha \implies H_0 \text{ を棄却する.}$$

帰無仮説 $H_0 : \mu_1 = \mu_2$ v.s. 対立仮説 $H_2 : \mu_1 > \mu_2$

の有意水準 α の検定に対しては t_m が大きいとき H_0 を棄却する. 有意水準 α を決め, 一様乱数 1 つを生成させそれを v^* とする. $N_u^* \equiv \sum_{i=1}^{K-1} I(t_m \leq t_m(i))$,

$$N_\ell^* \equiv \sum_{i=1}^{K-1} I(t_m < t_m(i)) \text{ とおく.}$$

$$\frac{N_u^*}{K} \leq \alpha \text{ または } \left(\frac{N_u^*}{K} > \alpha \text{ かつ } \frac{N_\ell^*}{K} < \alpha \text{ かつ } v^* < \frac{K \times \alpha - N_\ell^*}{N_u^* - N_\ell^*} \right) \\ \implies H_0 \text{ を棄却する.}$$

$n_1, n_2 \geq 20$ のときには t_m は帰無仮説の下で近似的に標準正規分布に従うので, 検定方式は,

$$p \text{ 値} \approx \int_{z_m}^{\infty} \varphi(y) dy \leq \alpha \implies H_0 \text{ を棄却する.}$$

[7] M 推定量;

$$\check{\sigma}_n = \frac{\sqrt{\pi}}{\sqrt{2} \cdot n} \left(\sum_{i=1}^{n_1} |x_i - \bar{x}| + \sum_{j=1}^{n_2} |y_j - \bar{y}| \right) \text{ とし}$$

$$t_m(\theta) \equiv \frac{1}{n_1} \sum_{i=1}^{n_1} \psi \left(\frac{x_i - \tilde{\mu} - \theta}{\check{\sigma}_n} \right) - \frac{1}{n_2} \sum_{j=1}^{n_2} \psi \left(\frac{y_j - \tilde{\mu} + \left(\frac{n_1}{n_2} \right) \theta}{\check{\sigma}_n} \right)$$

とおく. $t_m(\theta) = 0$ の解を $\check{\theta}$ とし, $\check{\delta} = \left(1 + \frac{n_1}{n_2} \right) \check{\theta}$ を $\mu_1 - \mu_2$ の点推定量とする. $\varepsilon > 0$ 以下の誤差の範囲で $\check{\theta}$ を探すには, $\check{\theta}_\ell \equiv \min\{x_i - \tilde{\mu} - \check{\sigma}_n b : i = 1, \dots, n_1\}$, $\check{\theta}_u \equiv \max\{x_i - \tilde{\mu} + \check{\sigma}_n b : i = 1, \dots, n_1\}$ とおくと, $t_m(\check{\theta}_\ell) > 0$,

$t_m(\check{\theta}_u) < 0, \check{\theta}_\ell < \check{\theta}_u$ が成り立つ,

$$\begin{aligned} \check{\theta}_x &\equiv \frac{\check{\theta}_\ell + \check{\theta}_u}{2} \text{ とし} \\ t_m(\check{\theta}_x) &> 0 \implies \text{新たに } \check{\theta}_\ell \equiv \check{\theta}_x \text{ とおき,} \\ t_m(\check{\theta}_x) &< 0 \implies \text{新たに } \check{\theta}_u \equiv \check{\theta}_x \text{ とおく.} \end{aligned}$$

$t_m(\check{\theta}_x) = 0$ または $\max\{|\check{\theta}_u - \check{\theta}_x|, |\check{\theta}_x - \check{\theta}_\ell|\} < 10^{-5}$ になるまで 四角の枠内を繰り返す.

第 3 章

独立性の検定

3.1 スピアマンの順位検定

$(x_1, y_1), \dots, (x_n, y_n)$ を 2 次元の連続型分布関数 $F_{XY}(x, y)$ をもつ母集団からの大きさ n の無作為標本の観測値とする. $x_1, \dots, x_n$ を小さい方から並べたときの x_i の順位を r_i , $y_1, \dots, y_n$ を小さい方から並べたときの y_i の順位を q_i とする. すなわち,

$$r_i \equiv \frac{\sum_{k=1}^n I(x_k \leq x_i) + \sum_{k=1}^n I(x_k < x_i) + 1}{2},$$

$$q_i \equiv \frac{\sum_{k=1}^n I(y_k \leq y_i) + \sum_{k=1}^n I(y_k < y_i) + 1}{2}$$

とする. このとき, $t_r \equiv \sum_{i=1}^n r_i \cdot (q_i - \frac{n+1}{2})$ とおき, $|t_r|$ が大きいとき帰無仮説 $H_0: x_i$ と y_i は互いに独立を棄却する.

まず, n が 20 未満のときの手順を述べる.

- 手順 1. $0 < \alpha < 1$ となるように有意水準 α を決め,

$$r(0) \equiv (r_1(0), \dots, r_n(0)) \equiv (r_1, \dots, r_n)$$

とおき, 統計量

$$t_r \equiv \sum_{j=1}^n r_j(0) \cdot (q_j - \frac{n+1}{2})$$

を計算する.

- 手順 2. シミュレーションの繰り返し数 K として 1000 を採り,
 $i = 1, 2, \dots, K-1$ について以下の手順 2.1 から手順 2.4 までを繰り返す.
 - 手順 2.1. 乗算合同法の疑似一様乱数を $n-1$ 個生成させ, それを
 $u_1(i), \dots, u_{n-1}(i)$ とする.
 - 手順 2.2. $j = 1, \dots, n-1$ について以下の手順 2.2.1 と手順 2.2.2 を繰り返す.
 - 手順 2.2.1. $\ell \equiv [(n-j+1) \cdot u_j(i)] + 1$ とおき, $\ell = n-j+2$ ならば新たに $\ell \equiv 1$ とおく, ただし, $[x]$ は x を越えない最大の整数とする.
 - 手順 2.2.2. n 次元ベクトル $r(i-1) \equiv (r_1(i-1), \dots, r_{n-1}(i-1))$ の $n-j+1$ 成分 $r_{n-j+1}(i-1)$ と ℓ 成分 $r_\ell(i-1)$ とを入れ替えたものを新たに, $r(i-1)$ とおく.
 - 手順 2.3. 手順 2.2 で成分入れ替えによって最終的に決定された $r(i-1)$ を $r(i)$ とおく.
 - 手順 2.4. $r(i) \equiv (r_1(i), \dots, r_n(i))$ より

$$t_r(i) \equiv \sum_{j=1}^n r_j(i) \cdot (q_j - \frac{n+1}{2})$$

を計算する.

- 手順 3. $N_u \equiv \sum_{i=1}^{K-1} I(|t_r| \leq |t_r(i)|)$, $N_\ell \equiv \sum_{i=1}^{K-1} I(|t_r| < |t_r(i)|)$ とおく,
 一様乱数 1 つを生成させ, それを v とする.
 $\frac{N_u}{K} \leq \alpha$ または $(\frac{N_u}{K} > \alpha \text{ かつ } \frac{N_\ell}{K} < \alpha \text{ かつ } v < \frac{K \cdot \alpha - N_\ell}{N_u - N_\ell})$
 $\implies H_0$ を棄却する.

$n \geq 20$ のときには

$$z_r \equiv \frac{\sqrt{n-1} \cdot t_r}{\sqrt{\sum_{i=1}^n (r_i - \frac{n+1}{2})^2 \cdot \sum_{i=1}^n (q_i - \frac{n+1}{2})^2}}$$

は帰無仮説の下で近似的に標準正規分布に従うので、検定方式は楽で、

$$p \text{ 値} \approx 2 \int_{|z_r|}^{\infty} \varphi(y) dy \leq \alpha \implies H_0 \text{ を棄却する.}$$

3.2 M 検定

$(x_1, y_1), \dots, (x_n, y_n)$ を 2 次元の連続型分布関数 $F_{XY}(x, y)$ をもつ母集団からの大きさ n の無作為標本の観測値とする. M 関数 $\psi(\cdot)$ を $\psi(x) = \max(\min(x, b), -b)$ で定義し, x, y の尺度母数をそれぞれ

$$\hat{\sigma}_x = \frac{\sqrt{\pi}}{\sqrt{2}n} \sum_{i=1}^n |x_i - \bar{x}_n|, \hat{\sigma}_y = \frac{\sqrt{\pi}}{\sqrt{2}n} \sum_{i=1}^n |y_i - \bar{y}_n| \text{ で推定し, } \bar{\psi}_x \equiv \frac{1}{n} \sum_{i=1}^n \psi\left(\frac{x_i - \bar{x}_n}{\hat{\sigma}_x}\right),$$

$$\bar{\psi}_y \equiv \frac{1}{n} \sum_{i=1}^n \psi\left(\frac{y_i - \bar{y}_n}{\hat{\sigma}_y}\right) \text{ とおき,}$$

$$t_m \equiv \sum_{i=1}^n \left\{ \psi\left(\frac{x_i - \bar{x}_n}{\hat{\sigma}_x}\right) - \bar{\psi}_x \right\} \cdot \left\{ \psi\left(\frac{y_i - \bar{y}_n}{\hat{\sigma}_y}\right) - \bar{\psi}_y \right\}$$

とおき, $|t_m|$ が大きいとき帰無仮説 H_0 : 『 x_i と y_i は互いに独立』を棄却する.

まず, n が 20 未満のときの手順を述べる.

- 手順 1. $0 < \alpha < 1$ となるように有意水準 α を決め,

$$\begin{aligned} w(0) &\equiv (w_1(0), \dots, w_n(0)) \\ &\equiv \left(\psi\left(\frac{x_1 - \bar{x}_n}{\hat{\sigma}_n}\right), \dots, \psi\left(\frac{x_n - \bar{x}_n}{\hat{\sigma}_n}\right) \right) \end{aligned}$$

とおき, 統計量

$$t_m \equiv \sum_{j=1}^n w_j(0) \cdot \left\{ \psi\left(\frac{y_j - \bar{y}_n}{\hat{\sigma}_y}\right) - \bar{\psi}_y \right\}$$

を計算する.

- 手順 2. シミュレーションの繰り返し数 K として 1000 を採り,
 $i = 1, 2, \dots, K-1$ について以下の手順 2.1 から手順 2.4 までを繰り返す.
 - ・ 手順 2.1. 乗算合同法による $U(0, 1]$ 上の疑似一様乱数を n_2 個生成させ,
 それを $u_1(i), \dots, u_{n-1}(i)$ とする.
 - ・ 手順 2.2. $j = 1, \dots, n-1$ について以下の手順 2.2.1 と手順 2.2.2 を繰り返す.
 - 手順 2.2.1. $\ell \equiv [(n-j+1) \cdot u_j(i)] + 1$ とおき, $\ell = n-j+2$ ならば新たに $\ell \equiv 1$ とおく,
 - 手順 2.2.2. n 次元ベクトル $w(i-1) \equiv (w_1(i-1), \dots, w_n(i-1))$ の $n-j+1$ 成分 $w_{n-j+1}(i-1)$ と ℓ 成分 $w_\ell(i-1)$ とを入れ替えたものを新たに, $w(i-1)$ とおく.
 - ・ 手順 2.3. 手順 2.2 で成分入れ替えによって最終的に決定された $w(i-1)$ を $w(i)$ とおく.
 - ・ 手順 2.4. $w(i) \equiv (w_1(i), \dots, w_n(i))$ より

$$t_m(i) \equiv \sum_{j=1}^{n_1} w_j(i) \cdot \left\{ \psi\left(\frac{y_j - \bar{y}_n}{\hat{\sigma}_y}\right) - \bar{\psi}_y \right\}$$

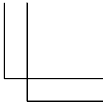
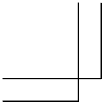
を計算する.

- 手順 3. $N_u \equiv \sum_{i=1}^{K-1} I(|t_m| \leq |t_m(i)|)$, $N_\ell \equiv \sum_{i=1}^{K-1} I(|t_m| < |t_m(i)|)$ とおく,
 一様乱数 1 つを生成させ, それを v とする.
 $\frac{N_u}{K} \leq \alpha$ または $(\frac{N_u}{K} > \alpha$ かつ $\frac{N_\ell}{K} < \alpha$ かつ $v < \frac{K \cdot \alpha - N_\ell}{N_u - N_\ell})$
 $\Rightarrow H_0$ を棄却する.

$n \geq 20$ のときには

$$z_m \equiv \frac{\sqrt{n-1} \cdot t_m}{\sqrt{\sum_{i=1}^n \left\{ \psi\left(\frac{x_i - \bar{x}_n}{\hat{\sigma}_x}\right) - \bar{\psi}_x \right\}^2 \cdot \sum_{i=1}^n \left\{ \psi\left(\frac{y_i - \bar{y}_n}{\hat{\sigma}_y}\right) - \bar{\psi}_y \right\}^2}}$$

は帰無仮説の下で近似的に標準正規分布に従うので, 検定方式は楽で,



$p \text{ 値} \approx 2 \int_{|z_m|}^{\infty} \varphi(y) dy \leq \alpha \implies H_0 \text{ を棄却する.}$

[著書と発展のページにリンク](#)

