多標本問題

白石高章
南山大学情報理工学部

データ解析が最初に行えるように, 第 1 節でモデルの説明を行い, 引き続き第 2 節でプログラムソフトでの統計解析方法を述べている. 第 3 節以下は手法の数理的説明を行っている.

1 モデルの設定

ある要因 A があり, k 個の水準 $A_1, \ldots, A_k$ を考え, 水準 A_i における標本の観測値 $(X_{i1}, X_{i2}, \cdots, X_{in_i})$ は第 i 標本と呼ばれ, 平均 μ_i, 分散 σ^2 である同一の連続型分布関数 $F(\frac{x-\mu_i}{\sigma})$ をもつとする. さらにすべての X_{ij} は互いに独立であると仮定する. 総標本数を $n \equiv n_1 + \cdots + n_k$ とおき,

$$\nu \equiv \frac{1}{n}\sum_{i=1}^{k} n_i\mu_i, \ \ \alpha_i \equiv \mu_i - \nu$$

とおく. このとき $\sum_{i=1}^{k} n_i\alpha_i = 0$ である. α_i は要因 A の水準 A_i における**主効果**, ν を**全平均**とよぶ. ここで

$$X_{ij} = \mu_i + \epsilon_{ij} = \nu + \alpha_i + \epsilon_{ij}, \ \text{ただし} \ \sum_{i=1}^{k} n_i\alpha_i = 0$$

と書き直せ, ϵ_{ij} は誤差確率変数とよばれ独立で平均は 0 である.

(i) $F(x)$ が $N(0,1)$ であるかどうかの正規性の検定法や $F(x)$ の近い分布を探す方法を解説する.

(ii) 興味ある帰無仮説と対立仮説は,

$$H_0: \ \mu_1 = \cdots = \mu_k \ \text{vs.} \ H_1: \ \text{ある} \ i \neq i' \ \text{について} \ \mu_i \neq \mu_{i'}$$

$$\Longleftrightarrow \quad H_0: \ \alpha_1 = \cdots = \alpha_k = 0 \ \text{vs.} \ H_1: \ \text{ある} \ i \ \text{について} \ \alpha_i \neq 0$$

である. また, パラメータ $(\alpha_1, \cdots, \alpha_k)$, ν, $(\mu_1, \cdots, \mu_k)$ の点推定に興味がある.

2　解析ソフト

20 代，40 代, 60 代の健康な成人男性のコレステロール値 ($mg/100ml$) を測定したものが表 1 で得られた．年齢はコレステロール値に影響を及ぼすかを調べる．

表 1．年代とコレステロール値

番号	1	2	3	4	5	6	7	8	9	10
20 才代	135	222	252	260	269	235	235	386	252	201
40 才代	294	311	286	264	277	336	208	346	239	
60 才代	370	244	353	420	333	357	317			

この例を基に以下の節で詳しく紹介される統計手法のソフトプログラムの使用方法を紹介する．

$<$ 1 $>$ 正規性の検定法

Nfitk.exe：正規性の検定法.

[実行例:入力データをテキストファイル『A:¥input¥ksamp21.txt』から読み込む]

(1) 『ワードパッド』を起動し, 標本の個数 k, k 個の標本サイズ (個数), k 組の標本観測値の順で表 1 のデータを例に次のように入力しそれを『A:¥input¥ksamp1.txt』として保存する．

```
3
10 9 7
135 222 252 260 269 235 235 386 252 201
294 311 286 264 277 336 208 346 239
370 244 353 420 333 357 317
```

(2) Nfitk.exe を使って, 実行結果を画面に表示する．

E:¥Esoft > **Nfitk < A:¥input¥ksamp1.txt** 　エンターキー
によって実行結果が画面に表示される．

$<$ 2 $>$ データの従っている分布に近い分布を探す

Seekfk.exe：データの従っている分布に近い分布を探す.

[実行例:入力データをテキストファイル『A:¥input¥ksamp1.txt』から読み込む]

E:¥Esoft > **Seekfk < A:¥input¥ksamp1.txt** 　エンターキー
によって実行結果が画面に表示される.

< 3 > 経験分布関数と正規分布関数の重ねがきグラフ

Empk.exe 及び Excel : 関数の値を求めエクセルで重ねがきグラフ.

○ Empk.exe を使って入力データを『A:¥input¥ksamp2.txt』から読み込み, ファイル『A:¥kekka¥empk.txt』に実行結果を書き込む. 続いて, A:¥kekka¥empk.txt をエクセルで開き, 経験分布関数と正規分布関数のグラフを描く手順を示す.

(1) 『ワードパッド』を起動し, 観測値の小数点以下の桁数 (整数のときは 0), 標本の個数 k, k 個の標本サイズ (個数), k 組の標本観測値の順で表 1 のデータを入力しそれを『A:¥input¥ksamp2.txt』とする. すなわち, このファイルは『A:¥input¥ksamp1.txt』の 1 行目に 0 を追加したものである.

(2) Empk.exe を使って, エクセルに読み込むテキストファイルを作る.

E:¥Esoft > **Empk < A:¥input¥ksamp2.txt > A:¥kekka¥empk.txt**
エンターキー

(3) ¥empn.txt を ¥empk.txt に替えた著書『統計科学』の 5.8 節 < 3 > の (3) から (9) と同様の操作を行う.

< 4 > 正規母集団での最良手法

Seik.exe : 正規母集団での最良手法による解析.

[実行例:キーボードで入力し画面に実行結果を表示する]

E:¥Esoft > **Seik** 　エンターキー

```
 標本の個数を入力して下さい
3        ←  入力後エンターキー
 3つの標本サイズを入力して下さい
10 9 7  ←  入力後エンターキー
 第 1 標本の観測値を 10
 個入力してください.
135 222 252 260 269 235 235 386 252 201 ←  入力後エンターキー
 第 2 標本の観測値を 9
 個入力してください.
294 311 286 264 277 336 208 346 239      ←  入力後エンターキー
 第 3 標本の観測値を 7
```

個入力してください.
```
370 244 353 420 333 357 317          ←  入力後エンターキー
```
有意水準を入力して下さい.
```
0.01 ←  入力後エンターキー
```
『実行結果が画面に表示される』

[実行例 2:入力データをテキストファイル『A:¥input¥ksamp3.txt』から読み込む]

著書『統計科学』の 1.7 節 < 3 > 実行例 2 の手順 (1) から (5) のようにして, A:¥input¥ksamp3.txt に入力データを入力し保存する.

E:¥Esoft > **Seik < A:¥input¥ksamp3.txt** 　エンターキー

によって実行結果が画面に表示される.

< 5 > ノンパラメトリック法

Nonpk.exe : 順位手法による解析.

[実行例:入力データをテキストファイル『A:¥input¥ksamp3.txt』から読み込む]

E:¥Esoft > **Nonpk < A:¥input¥ksamp3.txt** 　エンターキー

によって実行結果が画面に表示される.

< 6 > M 統計量に基づくセミパラメトリック法.

Robk.exe : セミパラメトリック法による解析.

[実行例:入力データをテキストファイル『A:¥input¥ksamp3.txt』から読み込む]

E:¥Esoft > **Robk < A:¥input¥ksamp3.txt** 　エンターキー

によって実行結果が画面に表示される.

♢ プログラム集

番号と内容	プログラム
< 1 > 正規性の検定	Nfitk.exe
< 2 > 分布の探索	Seekfk.exe
< 3 > 経験分布と正規分布関数	Empk.exe
< 4 > 正規母集団での最良手法	Seik.exe
< 5 > ノンパラメトリック法	Nonpk.exe
< 6 > セミパラメトリック法	Robk.exe

解析チャート

< 1 > 正規性の検定,　< 2 > 分布の探索,　< 3 > 経験分布関数と正規分布関数の重ねかきグラフ

↓

< 1 > で正規性が棄却されず, < 2 > で正規分布が選択され, < 3 > により正規性が妥当と認められれば < 4 > を選択

< 2 > で両側指数分布またはワイブル分布, 対数正規分布, 指数分布が選択されれば < 5 > を選択

これら以外であれば < 6 > を選択

3　正規性の検定と分布の探索

第 i 標本の標本平均を, $\bar{X}_i \equiv \frac{1}{n_i}\sum_{j=1}^{n_j} X_{ij}$ とおく. $\sum_{i'=0}^{i-1} n_{i'} + 1 \le m \le \sum_{i'=0}^{i} n_{i'}$ となる整数 m と $j \equiv m - \sum_{i'=0}^{i-1} n_{i'}$ に対して

$$Z_m \equiv (X_{ij} - \bar{X}_i)/\sqrt{1 - \frac{1}{n_i}} \qquad (j = 1, \cdots, n_i) \tag{3.1}$$

で $Z_1, \cdots, Z_n$ を定義する, ただし, $n_0 \equiv 0$ とおく. このとき $E(Z_i) = 0$, $V(Z_i) = \sigma^2$ である. $\hat{G}_n(x)$ を

$$\hat{G}_n(x) \equiv \frac{1}{n}\#\{Z_i |\ Z_i \le x, 1 \le i \le n\} = \frac{1}{n}\{x \text{ 以下となる } Z_i \text{の個数}\}$$

で定義する.

観測値の従っている分布を調べる方法として,

$$D_F \equiv \sup_{-\infty<x<\infty} \left|\hat{G}_n(x) - F\left(\frac{x}{\breve{\sigma}_n}\right)\right| \tag{3.2}$$

とおく, ただし,

$$\breve{\sigma}_n \equiv \frac{1}{n\int_{-\infty}^{\infty}|x|dF(x)}\sum_{i=1}^{n}|Z_i|.$$

$\breve{\sigma}_n$ は σ の頑健推定量である.

[1] **正規性の検定**; 『帰無仮説 $H : F(x)$ は正規分布 vs. 対立仮説 $H^c : F(x)$ は正規分布ではない』の検定として

$$D_\Phi = \sup_{-\infty<x<\infty} |\hat{G}_n(x) - \Phi(\frac{x}{\breve{\sigma}_n})| \tag{3.3}$$

が大きいとき H を棄却する方法が採用できる. シミュレーションによる検定手順を以下に示す.

- 手順 1. $0 < \alpha < 1$ となるように有意水準 α を決め, D_Φ 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,
  $\ell = 1, 2, \cdots, K-1$ 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. 正規乱数を n 個生成させ, それを $\{X_{ij}(\ell) |\ j = 1, \cdots, n_i,\ i = 1, \cdots, k\}$ とし, (1) と同様に $Z_1(\ell), \cdots, Z_n(\ell)$ を定義する.
  - 手順 2.2. $Z_1, \cdots, Z_n$ の替わりに $Z_1(\ell), \cdots, Z_n(\ell)$ で (3.3) 式を計算し, それを $D_\Phi(\ell)$ とする.
- 手順 3. $N \equiv \sum_{\ell=1}^{K-1} I(D_\Phi < D_\Phi(\ell))$ とおく. $\frac{N}{K}$ は p 値となり, この値が α より小さいならば H を棄却する.

残念ながら, 1 標本の場合と同様に観測値の従っている分布が正規分布でないときでも n が非常に大きくないと正規性を棄却できないことが多い.

[2] **分布の探索**; D_F は, 観測値の従っている分布がどれくらい $F(\frac{x-\mu}{\sigma})$ に近いかのはかりとして見れる. $F(x)$ として正規分布, 混合正規分布, ロジスティック分布, 両側指数分布, 異常値を持つ混合正規分布, ワイブル分布, 対数正規分布, 指数分布を当てはめ, 統計量 D_F のもっとも小さい値を与える分布 $F(x)$ を探し, それを観測値の従っている分布に最も近い分布と見なすことができる.

4 正規母集団での最良手法

$F(x)$ を正規分布 $N(0,1)$ の分布関数とする.

[3] F **検定**; $\bar{X}_{i\cdot} \equiv \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$, $\bar{X}_{\cdot\cdot} \equiv \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}$ とし

$$\begin{aligned} R_0^2 &\equiv \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i\cdot})^2 \\ R_1^2 &\equiv \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{\cdot\cdot})^2 \end{aligned}$$

をそれぞれ, **級内変動**, **総変動**といい,

$$R_1^2 - R_0^2 = \sum_{i=1}^{k} n_i(\bar{X}_{i\cdot}-\bar{X}_{\cdot\cdot})^2 \tag{4.1}$$

の関係がある. この (4.1) を**級間変動**という. これらを使って帰無仮説 H_0 vs. 対立仮説 H_1 の検定統計量は

$$T_S \equiv \frac{(R_1^2-R_0^2)/(k-1)}{R_0^2/(n-k)}$$

と表され, T_S が大きいとき H_0 を棄却する. H_0 の下で, T_S は自由度 $(k-1, n-k)$ の F 分布に従う. また, $n-k$ が十分大きければ,

$$Z_S \equiv (k-1)T_S$$

は自由度 $k-1$ のカイ二乗分布に従う. 故に, 自由度 $(k-1, n-k)$ の F 分布の密度関数を $f_{k-1,n-k}(y)$, 自由度 $k-1$ の χ^2 分布の密度関数を $g_{k-1}(y)$ とすると,

a. $n-k \leq 50$ のとき,

$$p\,値 = \int_{T_S}^{\infty} f_{k-1,n-k}(y)dy \leq \alpha \Longrightarrow H_0\ を棄却する.$$

b. $n-k > 50$ のとき,

$$p\,値 \approx \int_{Z_S}^{\infty} g_{k-1}(y)dy \leq \alpha \Longrightarrow H_0\ を棄却する.$$

[4] **点推定量**; $\tilde{\alpha}_i = \bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}$ $(i = 1, \cdots, k)$, $\tilde{\nu} = \bar{X}_{\cdot\cdot}$, $\tilde{\mu}_i = \bar{X}_{i\cdot}$ $(i = 1, \cdots, k)$.

5　ノンパラメトリック法

$F(x)$ を未知の分布関数であってもよいとする.

[5] **クラスカル・ワリスの順位検定**; n 個すべての観測値 $X_{11},\cdots,X_{kn_k}$ を小さい方から並べたときの X_{ij} の順位を R_{ij} とする. ただし, $X_1,\cdots,X_{kn_k}$ の中に同じ値がある場合を考慮して,

$$R_{ij} \equiv \frac{\sum_{\ell=1}^{k}\sum_{m=1}^{n_i} I(X_{\ell m} \le X_{ij}) + \sum_{\ell=1}^{k}\sum_{m=1}^{n_i} I(X_{\ell m} < X_{ij}) + 1}{2}$$

とおく. このとき,

$$T_R \equiv \sum_{i=1}^{k} n_i(\bar{R}_{i\cdot} - \frac{n+1}{2})^2$$

とおき, T_R が大きいとき H_0 を棄却する. ただし, $\bar{R}_{i\cdot} \equiv \sum_{j=1}^{n_i} R_{ij}/n_i$ とする.

まず, ある i について n_i が 30 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,

$$\boldsymbol{R}(0) \equiv (R_1(0),\cdots,R_n(0)) \equiv (R_{11},\cdots,R_{1n_1},R_{21},\cdots,R_{kn_k})$$

とおき, 統計量

$$T_R \equiv \sum_{i=1}^{k} n_i\{\bar{R}_{i\cdot}(0) - \frac{n+1}{2}\}^2$$

を計算する, ただし, $\bar{R}_{i\cdot}(0) \equiv \sum_{j=L_{i-1}+1}^{L_i} R_j(0)/n_i$, $L_0 \equiv 0$,
$L_i \equiv \sum_{j=1}^{i} n_j$ とする.

- 手順 2. シミュレーションの繰り返し数 M として 1000 を採り,
$\imath = 1,2,\cdots,M-1$ 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 $n-n_1$ 個生成させ, それを
$u_1(\imath),\cdots,u_{n-n_1}(\imath)$ とする.
  - 手順 2.2. $j = 1,\cdots,n-n_1$ 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. $\ell \equiv [(n-j+1)\cdot u_j(\imath)] + 1$ 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 $\boldsymbol{R}(\imath-1) \equiv (R_1(\imath-1),\cdots,R_n(\imath-1))$ の $n-j+1$ 成分 $R_{n-j+1}(\imath-1)$ と ℓ 成分 $R_\ell(\imath-1)$ とを入れ替えたものを新たに, $\boldsymbol{R}(\imath-1)$ 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 $\boldsymbol{R}(\imath-1)$ を $\boldsymbol{R}(\imath)$ とおく.

· 手順 2.4. $\boldsymbol{R}(\imath) \equiv (R_1(\imath), \cdots, R_n(\imath))$ より

$$T_R(\imath) \equiv \sum_{i=1}^{k} n_i \{\bar{R}_{i\cdot}(\imath) - \frac{n+1}{2}\}^2$$

を計算する, ただし, $\bar{R}_{i\cdot}(\imath)$ は $\bar{R}_{i\cdot}(0)$ と類似の定義.

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

一様乱数 1 つを生成させ, それを v とする.

$\dfrac{N_u}{M} \leq \alpha$ または $(\dfrac{N_u}{M} > \alpha$ かつ $\dfrac{N_\ell}{M} < \alpha$ かつ $v < \dfrac{M \cdot \alpha - N_\ell}{N_u - N_\ell})$

$\Longrightarrow H_0$ を棄却する.

次に $n_1, \cdots, n_k \geq 30$ のときには

$$Z_R \equiv \frac{(n-1) \cdot T_R}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (R_{ij} - \frac{n+1}{2})^2}$$

は帰無仮説の下で近似的に自由度 $k-1$ のカイ二乗分布に従うので, 検定方式は楽で,

$$p \text{ 値} \approx \int_{Z_R}^{\infty} g_{k-1}(y)dy \leq \alpha \Longrightarrow H_0 \text{ を棄却する.}$$

[6] **順位推定量**; $1 \leq i \neq i' \leq k$ に対して

$$\begin{aligned} \hat{\delta}_{ii'} &= \mathrm{med}\{X_{ij} - X_{i'j'} | j = 1, \cdots, n_i, j' = 1, \cdots, n_{i'}\} \\ &= (n_i \cdot n_{i'} \text{個の値 } \{X_{ij} - X_{i'j'} | j = 1, \cdots, n_i, j' = 1, \cdots, n_{i'}\} \text{ の中央値}) \end{aligned}$$

は著書『統計科学』の 6.4 節 [7] より $\mu_i - \mu_{i'}$ のホッジス・レーマン順位推定量である. また, $\alpha_i = \frac{1}{n} \sum_{i'=1}^{k} n_{i'}(\mu_i - \mu_{i'})$ より, α_i の順位推定量は

$$\hat{\alpha}_i = \frac{1}{n} \sum_{i'=1}^{k} n_{i'} \hat{\delta}_{ii'} \ (i = 1, \cdots, k)$$

となる, ただし $\hat{\delta}_{ii} = 0$ とする. $1 \leq m \leq n$ なる整数 m に対して, $m = \sum_{p=0}^{i-1} n_p + j$ かつ $1 \leq j \leq n_i$ を満たす i, j が存在する, ただし, $n_0 \equiv 0$ とおく. この m, i, j に対して $Y_m \equiv X_{ij} - \hat{\alpha}_i$ とおき, ν の順位推定量は

$$\begin{aligned} \hat{\nu} &= \mathrm{med}\{\frac{Y_i + Y_j}{2} | 1 \leq i \leq j \leq n\} \\ &= (\frac{n(n+1)}{2} \text{個の値 } \{\frac{Y_i + Y_j}{2} | 1 \leq i \leq j \leq n\} \text{ の中央値}) \end{aligned}$$

μ_i の順位推定量は

$$\hat{\mu}_i = \hat{\nu} + \hat{\alpha}_i \quad (i = 1, \cdots, k).$$

$\hat{\alpha}_i$ は密度関数 $f(x) \equiv F'(x)$ に対称性を必要としないが, $\hat{\nu}$, $\hat{\mu}_i$ には $f(x)$ の対称性を必要とする.

密度関数 $f(x)$ に対称性を仮定できない場合は, $\tilde{\nu} = \bar{X}_{\cdot\cdot}$ を ν の推定量とする. μ_i の推定量は

$$\hat{\mu}_i = \tilde{\nu} + \hat{\alpha}_i \quad (i = 1, \cdots, k).$$

6 頑健な手法

前節と同様に $F(x)$ を未知の分布関数であってもよいとする.

[7] M 検定; 帰無仮説 H_0 と 対立仮説 H_1 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, 4 節 [4] の $\tilde{\nu}$ を使って, 尺度母数を $\hat{\sigma}_n = \frac{\sqrt{\pi}}{\sqrt{2}n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} |X_{ij} - \tilde{\nu}|$ で推定し, $\bar{\psi}_{i\cdot}(\boldsymbol{X}) = \frac{1}{n_i} \sum_{j=1}^{n_i} \psi(\frac{X_{ij} - \tilde{\nu}}{\hat{\sigma}_n})$, $\bar{\psi}_{\cdot\cdot}(\boldsymbol{X}) = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \psi(\frac{X_{ij} - \tilde{\nu}}{\hat{\sigma}_n})$ とおき

$$T_M \equiv \sum_{i=1}^{k} n_i \{\bar{\psi}_{i\cdot}(\boldsymbol{X}) - \bar{\psi}_{\cdot\cdot}(\boldsymbol{X})\}^2$$

が検定統計量である. T_M が大きいとき H_0 を棄却する.

まず, ある i について n_i が 30 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,

$$\begin{aligned} \boldsymbol{W}(0) &\equiv (W_1(0), \cdots, W_n(0)) \\ &\equiv (\psi(\frac{X_{11} - \tilde{\nu}}{\hat{\sigma}_n}), \cdots, \psi(\frac{X_{1n_1} - \tilde{\nu}}{\hat{\sigma}_n}), \psi(\frac{X_{21} - \tilde{\nu}}{\hat{\sigma}_n}), \cdots, \psi(\frac{X_{kn_k} - \tilde{\nu}}{\hat{\sigma}_n})) \end{aligned}$$

  とおき, 統計量

$$T_M \equiv \sum_{i=1}^{k} n_i \{\bar{W}_{i\cdot}(0) - \bar{W}_{\cdot\cdot}(0)\}^2$$

  を計算する, ただし, $\bar{W}_{i\cdot}(0) \equiv \sum_{j=L_{i-1}+1}^{L_i} W_j(0)/n_i$, $\bar{W}_{\cdot\cdot}(0) \equiv \sum_{j=1}^{n} W_j(0)/n$, $L_0 \equiv 0$, $L_i \equiv \sum_{j=1}^{i} n_j$ とする.

- 手順 2. シミュレーションの繰り返し数 M として 1000 を採り,
  $\imath = 1, 2, \cdots, M-1$ 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 $n - n_1$ 個生成させ, それを $u_1(\imath), \cdots, u_{n-n_1}(\imath)$ とする.
  - 手順 2.2. $j = 1, \cdots, n - n_1$ 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. $\ell \equiv [(n - j + 1) \cdot u_j(\imath)] + 1$ 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 $\boldsymbol{W}(\imath - 1) \equiv (W_1(\imath - 1), \cdots, W_n(\imath - 1))$

の $n-j+1$ 成分 $W_{n-j+1}(\imath-1)$ と ℓ 成分 $W_\ell(\imath-1)$ とを入れ替えたものを新たに, $\boldsymbol{W}(\imath-1)$ 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 $\boldsymbol{W}(\imath-1)$ を $\boldsymbol{W}(\imath)$ とおく.

· 手順 2.4. $\boldsymbol{W}(\imath)\equiv(W_1(\imath),\cdots,W_n(\imath))$ より

$$T_M(\imath)\equiv\sum_{i=1}^k n_i\{\bar{W}_{i\cdot}(\imath)-\bar{W}_{\cdot\cdot}(\imath)\}^2=\sum_{i=1}^k n_i\{\bar{W}_{i\cdot}(\imath)-\bar{W}_{\cdot\cdot}(0)\}^2$$

を計算する, ただし, $\bar{W}_{i\cdot}(\imath)$ は $\bar{W}_{i\cdot}(0)$ と類似の定義.

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

次に $n_1,\cdots,n_k\ge 30$ のときには

$$Z_M\equiv\frac{(n-1)\cdot T_M}{\sum_{i=1}^k\sum_{j=1}^{n_i}\{\psi(\frac{X_{ij}-\tilde{\nu}}{\hat{\sigma}_n})-\bar{\psi}_{\cdot\cdot}(\boldsymbol{X})\}^2}$$

は帰無仮説の下で近似的に自由度 $k-1$ のカイ二乗分布に従うので, 検定方式は楽で,

$$p\text{ 値}\approx\int_{Z_R}^{\infty}g_{k-1}(y)dy\le\alpha\Longrightarrow H_0\text{ を棄却する.}$$

[8] **M 推定量**; $i\ne i'$ に対して $\tilde{\nu}_{ii'}\equiv\dfrac{n_i\bar{X}_{i\cdot}+n_{i'}\bar{X}_{i'\cdot}}{n_i+n_{i'}}$, $\breve{\sigma}_n\equiv\dfrac{\sqrt{\pi}}{\sqrt{2}n}\displaystyle\sum_{i=1}^k\sum_{j=1}^{n_i}|X_{ij}-\bar{X}_{i\cdot}|$ とし

$$T_{ii'}(\theta)\equiv\frac{1}{n_i}\sum_{j=1}^{n_i}\psi(\frac{X_{ij}-\tilde{\nu}_{ii'}-\theta}{\breve{\sigma}_n})-\frac{1}{n_{i'}}\sum_{j=1}^{n_{i'}}\psi(\frac{X_{i'j}-\tilde{\nu}_{ii'}+(\frac{n_i}{n_{i'}})\cdot\theta}{\breve{\sigma}_n})$$

とおく. $T_{ii'}(\theta)=0$ の解を $\breve{\theta}_{ii'}$ とすれば, 著書『統計科学』の 6.5 節 [10] より $\breve{\delta}_{ii'}=(1+\frac{n_i}{n_{i'}})\breve{\theta}_{ii'}$ は $\mu_i-\mu_{i'}$ の M 推定量である. また, $\alpha_i=\frac{1}{n}\sum_{i'=1}^k n_{i'}(\mu_i-\mu_{i'})$ より, α_i の M 推定量は

$$\breve{\alpha}_i=\frac{1}{n}\sum_{i'=1}^k n_{i'}\breve{\delta}_{ii'}\quad(i=1,\cdots,k)$$

となる, ただし $\breve{\delta}_{ii}=0$ とする. i,j に対して $Y_{ij}\equiv X_{ij}-\breve{\alpha}_i$ とおき, M 統計量を $T_M(\theta)=\sum_{i=1}^k\sum_{j=1}^{n_i}\psi(\frac{Y_{ij}-\theta}{\breve{\sigma}_n})$ とする. $T_M(\theta)=0$ の θ の解 $\breve{\nu}=\hat{\theta}$ が ν の M 推定量である. μ_i の M 推定量は

$$\breve{\mu}_i=\breve{\nu}+\breve{\alpha}_i\quad(i=1,\cdots,k).$$

$\breve{\alpha}_i$ は密度関数 $f(x) \equiv F'(x)$ に対称性を必要としないが, $\breve{\nu}$, $\breve{\mu}_i$ には $f(x)$ の対称性を必要とする.

密度関数 $f(x)$ に対称性を仮定できない場合は, $\tilde{\nu} = \bar{X}_{\cdot\cdot}$ を ν の推定量とする. このとき, μ_i の推定量は

$$\breve{\mu}_i = \tilde{\nu} + \breve{\alpha}_i \quad (i = 1, \cdots, k).$$