離散分布の場合のノンパラメトリックなある 2-標本問題

統計数理研究所　藤　本　熙

(1980年4月　受付)

On Some Non-parametric Two-sample Problems in the Case of Discrete Distributions

Hirosi Hudimoto
(The Institute of Statistical Mathematics)

The use of statistics of Mann-Whitney type in the case of discrete distributions will be considered for two-sample problems.

In Sections 4 and 5, it is assumed that each individual in the given population π belongs to either one of two sub-populations π_1 and π_2 and that π_1 and π_2 are mixed in proportions w_1 and w_2 of individuals of π. But, w_1 and w_2 are unknown. Let $f_i(x)$ be the probability function in π_i $(i=1, 2)$. Denoting by $L_{(\xi)}$ the ξth value of the likelihood ratio $L(x)=f_2(x)/f_1(x)$ arranged in ascending order, let $g_i(\xi)$ be $f_i(x)$ arranged in the order based on $L_{(\xi)}$. Then, $G_i(y)=\sum_{\xi\leq y} g_i(\xi)$.

Now, suppose that a random sample of the size n is obtained from π. Let $(Y_1, Y_2, \cdots, Y_n)$ be a sample from π transformed into the rank order defined above. Consider the statistic $\hat{h}=(\hat{p}_1+\hat{p}_2)/2$, where $\hat{p}_i=\frac{1}{n}\sum_{k=1}^{n}\left\{G_i(Y_k)-\frac{1}{2}g_i(Y_k)\right\}$, $i=1, 2$. Then, we have

$$\hat{w}_1=\frac{1}{2}-\left(\hat{h}-\frac{1}{2}\right)\Delta_{12}^{-1},\quad \hat{w}_2=\frac{1}{2}+\left(\hat{h}-\frac{1}{2}\right)\Delta_{12}^{-1}$$

for unbiased estimates of w_1 and w_2, where $\Delta_{12}=\frac{1}{2}\sum_{v}\{G_1(y_v)g_2(y_v)-G_2(y_v)g_1(y_v)\}$.

When our decision is to classify an individual with the observed value x obtained from π to either one of π_1 and π_2, denote the probability of deciding an individual as coming from π_1 by $\delta^{(1)}(x)$. Then, a randomized decision rule can be denoted by $\delta(x)=(\delta^{(1)}(x), 1-\delta^{(1)}(x))$. Our decision rule is that $\delta_{\hat{w}}{}^{(1)}(x)=1$ if $D_{\hat{w}}(x)\geqq 0$ or $=0$ if $D_{\hat{w}}(x)<0$, where $D_{\hat{w}}(x)=\hat{w}_1 f_1(x)-\hat{w}_2 f_2(x)$. Denote the expected risk of $\delta_{\hat{w}}(x)=(\delta_{\hat{w}}{}^{(1)}(x), 1-\delta_{\hat{w}}{}^{(1)}(x))$ with respect to $\boldsymbol{w}=(w_1, w_2)'$ by $B(\boldsymbol{w}, \delta_{\hat{w}})$. Denote the Bayes risk by $B(\boldsymbol{w})=B(\boldsymbol{w}, \delta_{\boldsymbol{w}})$. When it is assumed that the loss is 0 for correct decision and 1 for incorrect decision, $\mathrm{P}\left\{B(\boldsymbol{w}, \delta_{\hat{w}})-B(\boldsymbol{w})<2\sqrt{\frac{1}{2n\Delta_{12}{}^2}\log\frac{2}{\alpha}}\right\}>1-\alpha$ is obtained for pre-assigned $\alpha>0$.

In Section 6, we deal with a two-sample problem. Two samples $(X_1{}^{(1)}, X_2{}^{(1)}, \cdots, X_{n_1}{}^{(1)})$ and $(X_1{}^{(2)}, X_2{}^{(2)}, \cdots, X_{n_2}{}^{(2)})$ are obtained from $f_1(x)$ and $f_2(x)$, respectively. Then, an unbiased estimate $\hat{p}_{12}$ for $p_{12}=\frac{1}{2}+\frac{1}{2}\sum_{v}\{F_1(x_v)f_2(x_v)-F_2(x_v)f_1(x_v)\}$ defined in Section 3 is considered. The variance of $\hat{p}_{12}$ is

$$\sigma^2(\hat{p}_{12})=\frac{1}{n_1n_2}\left\{p_{12}-\frac{1}{4}\sum_{v}f_1(x_v)f_2(x_v)+(n_1-1)A+(n_2-1)B-(n_1+n_2-1)p_{12}{}^2\right\},$$

where $$A=\sum_{v}\left(F_1(x_v)-\frac{1}{2}f_1(x_v)\right)^2 f_2(x_v),$$

and $$B=\sum_{v}\left(1-F_2(x_v)+\frac{1}{2}f_2(x_v)\right)^2 f_1(x_v).$$

In the case that $F_1(x)=F_2(x)$ for all x, it seems that $\sigma^2(\hat{p}_{12})\leqq(n_1+n_2-1)/12n_1n_2$.

1. 緒　　言

離散変量の場合について，Mann-Whitney 型統計量を使った 2-標本検定と，2 群への判別に対する経験的ベイズ決定の方法が扱われる．つまり，タイ (tie) のある場合の扱い方が問題になる．

第3節で与えられる p_{12} に対する，2-標本の場合における，不偏推定量が第6節の $\hat{p}_{12}$ である．その節では，$\hat{p}_{12}$ の分散を求めてある．

第4節の一般的な場合が，第2節に与えられるが，第4節では，混合の割合 w_1, w_2 の推定量の分散を，なるべく小さくするために，尤度比の使用を考える．第2節は [1], [2], [3] で述べた事柄の一部の要約である．

尤度比による順位に変換して考えた p_{12} の，1-標本の場合の，不偏推定量が $\hat{p}_i (i=1,2)$ のどちらかであるが，第4節では，$(X_1, X_2, \cdots, X_n)$ を複合分布 $w_1f_1(x)+w_2f_2(x)$ からのランダム標本と考えるから，$\hat{p}_1, \hat{p}_2$ から混合の割合 w_1, w_2 の推定が考えられる．

2. 一般的な議論

r 個の部分母集団 $\pi_1, \pi_2, \cdots, \pi_r$ からなる複合母集団 π を考える．つまり，π に属する個体は，r 種類の範疇に類別される．更に π_i に属する π の個体の割合を $w_i (i=1,2,\cdots,r)$ とする．ただし w_i は未知である．またその観測結果が π_i に属する個体のものであるような変量 X の分布関数を，$F_i(x)$ で示す．このとき，π から大きさ n のランダム標本 $(X_1, X_2, \cdots, X_n)$ を得て，w_i を推定することを考える．

π_i からの個体であるという条件のもとでの確率を，$\mathrm{P}(\cdot|\pi_i)$ で示すと，

(2.1) $$F_i(x)=\mathrm{P}(X\leqq x|\pi_i), \qquad i=1,2,\cdots,r$$

であって，$(X_1, X_2, \cdots, X_n)$ は，

(2.2) $$F(x)=\sum_{i=1}^{r} w_iF_i(x)$$

からのランダム標本である．

(2.3) $$\hat{p}_i=\frac{1}{n}\sum_{k=1}^{n}F_i(X_k)$$

を考え，$F_i(x)$ は連続とする．

(2.4) $$\Delta_{ij}=\int F_i(x)\,dF_j(x)-\frac{1}{2} \quad i, j=1,2,\cdots,r$$

として，

(2.5) $$\boldsymbol{w}=\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_r \end{pmatrix}, \qquad \hat{\boldsymbol{p}}=\begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \vdots \\ \hat{p}_r \end{pmatrix}$$

$$A^{(1)}=\begin{pmatrix} \frac{1}{2} & \frac{1}{2}+\Delta_{12} & \cdots & \frac{1}{2}+\Delta_{1r} \\ \frac{1}{2}+\Delta_{21} & \frac{1}{2} & \cdot & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}+\Delta_{r1} & \frac{1}{2}+\Delta_{r2} & \cdots & \frac{1}{2} \end{pmatrix} \tag{2.5}$$

とおくと，$|\boldsymbol{A}^{(1)}|\neq 0$ ならば，

$$\hat{\boldsymbol{w}}^{(1)}=(\boldsymbol{A}^{(1)})^{-1}\hat{\boldsymbol{p}} \tag{2.6}$$

は $\boldsymbol{w}$ の不偏且一致推定量である．ただし $|\boldsymbol{A}^{(1)}|$ 及び $(\boldsymbol{A}^{(1)})^{-1}$ は $\boldsymbol{A}^{(1)}$ の行列式及び逆行列である．

$$\boldsymbol{A}^{(2)}=\begin{pmatrix} 0 & \Delta_{12} & \cdots & \Delta_{1r} \\ \Delta_{21} & 0 & \cdots & \Delta_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_{r1} & \Delta_{r2} & \cdots & 0 \end{pmatrix}, \quad \boldsymbol{E}=\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \tag{2.7}$$

と，$r\times r$ 行列 $\boldsymbol{A}^{(2)}$，$\boldsymbol{E}$ を考えると，

$$\boldsymbol{A}^{(1)}=\boldsymbol{A}^{(2)}+\frac{1}{2}\boldsymbol{E}. \tag{2.8}$$

$\boldsymbol{A}^{(2)}$ は歪対称行列であるから，r が奇数であれば，$|\boldsymbol{A}^{(2)}|=0$，r が偶数ならば，$|\boldsymbol{A}^{(1)}|=|\boldsymbol{A}^{(2)}|$ である．そこで，r が偶数のときに，$|\boldsymbol{A}^{(2)}|\neq 0$ であれば，$\boldsymbol{w}$ の不偏且一致推定量として，

$$\hat{\boldsymbol{w}}^{(2)}=(\boldsymbol{A}^{(2)})^{-1}\left(\hat{\boldsymbol{p}}-\frac{1}{2}\boldsymbol{J}\right) \tag{2.9}$$

が得られる．ここに，$\boldsymbol{J}'=(1,1,\cdots,1)$，$'$ はベクトルの転置を示す．

次に X が離散変量の場合を考える．X の確率関数を，

$$f_i(x_v)=\mathrm{P}(X=x_v\,|\,\pi_i),\quad v=1,2,\cdots,s \tag{2.10}$$

とする．このとき，(2.3) を，

$$\hat{p}_i=\frac{1}{n}\sum_{k=1}^{n}\left\{F_i(X_k)-\frac{1}{2}f_i(X_k)\right\} \tag{2.11}$$

に，(2.4) を

$$\Delta_{ij}=\frac{1}{2}\sum_{v=1}^{s}\{F_i(x_v)f_j(x_v)-F_j(x_v)f_i(x_v)\} \tag{2.12}$$

に変えれば，X が離散変量の場合にも，今迄の結果は，そのまま成立つ．

以降では，$r=2$ の場合を，X が離散変量の場合について考える．

3. $p_{12}=\frac{1}{2}+\Delta_{12}$ について

$X^{(i)}$ $(i=1,2)$ は，確率関数 $f_i(x)$ をもつ次のような離散変量とする：

$$\mathrm{P}(X^{(i)}=x_v)=f_i(x_v), \quad v=1,2,\cdots,s. \tag{3.1}$$

またその分布関数を

$$F_i(x)=\sum_{x_v\leqq x} f_i(x_v), \quad i=1,2 \tag{3.2}$$

で示す.

このとき，独立な $X^{(1)}, X^{(2)}$ に対し，$X^{(1)}<X^{(2)}$ の確率 $\mathrm{P}(X^{(1)}<X^{(2)})$ は,

$$\begin{aligned}\mathrm{P}(X^{(1)}<X^{(2)})&=\sum_{v=1}^{s}(F_1(x_v)-f_1(x_v))f_2(x_v)\\&=\sum_{v=1}^{s}(1-F_2(x_v))f_1(x_v).\end{aligned}$$

これから,

$$\mathrm{P}(X^{(1)}<X^{(2)})=\frac{1}{2}+\Delta_{12}-\frac{1}{2}\sum_{v=1}^{s}f_1(x_v)f_2(x_v), \tag{3.3}$$

$$\Delta_{12}=\frac{1}{2}\sum_{v=1}^{s}(F_1(x_v)f_2(x_v)-F_2(x_v)f_1(x_v)). \tag{3.4}$$

ところで，$F_1\equiv F_2$ のとき，$F_i(x)$ $(i=1,2)$ が連続ならば，$\mathrm{P}(X^{(1)}<X^{(2)})=\dfrac{1}{2}$ であるが，ここでは，$\sum f_1(x_v)f_2(x_v)=\sum f_i^2(x_v)>0$ であるから，$\mathrm{P}(X^{(1)}<X^{(2)})<\dfrac{1}{2}$ であって，その大きさは，$\sum f_i^2(x_v)$ の値によるので，$F_1\equiv F_2$ を帰無仮説とする 2-標本検定の場合など不便がある．しかし，Δ_{12} は,

$$-\frac{1}{2}\leqq\Delta_{12}\leqq\frac{1}{2}$$

であるから，$\mathrm{P}(X^{(1)}<X^{(2)})$ の替りに

$$p_{12}=\frac{1}{2}+\Delta_{12} \tag{3.5}$$

を使うことにすると,

$$0\leqq p_{12}\leqq 1$$

であって，$p_{12}=0$ または $p_{12}=1$ となるような極端な場合を考えるのは容易であって，p_{12} を $\mathrm{P}(X^{(1)}<X^{(2)})$ の替りに使うのが適当であることがわかる.

4. 混合の割合の推定 ($r=2$ の場合)

$(X_1, X_2, \cdots, X_n)$ を $w_1f_1(x)+w_2f_2(x)$ からのランダム標本とする．つまり，w_1, w_2 の割合で $f_1(x), f_2(x)$ からの標本が混合しているのである．$0\leqq w_i\leqq 1$, $w_1+w_2=1$ である.

ここでは，w_1, w_2 は未知であるが，$f_1(x), f_2(x)$ は既知として，w_1, w_2 の推定を考える.

まず尤度比 $L(x)=f_2(x)/f_1(x)$ を考え，この値の順序に配列した $L(x)$ の ξ-番目を $L_{(\xi)}$ とする.

$$0\leqq L_{(1)}\leqq\cdots\leqq L_{(\xi)}\leqq\cdots\leqq L_{(s)}\leqq\infty$$

である.

今 $f_1(x)$, $f_2(x)$ を上の順序に配列し直して，$g_1(\xi)$, $g_2(\xi)$ で示し，

$$G_i(y)=\sum_{\xi\leq y}g_i(\xi) \tag{4.1}$$

とする．更に $(X_1, X_2, \cdots, X_n)$ を $L_{(\xi)}$ にもとづく順位に変換して，$(Y_1, Y_2, \cdots, Y_n)$ とする．ここで，

$$\begin{aligned}\hat{p}_i&=\frac{1}{n}\sum_{k=1}^{n}\left\{G_i(Y_k)-\frac{1}{2}g_i(Y_k)\right\},\qquad i=1,2\\ \Delta_{12}&=\frac{1}{2}\sum_{v}\{G_1(y_v)g_2(y_v)-G_2(y_v)g_1(y_v)\}\end{aligned} \tag{4.2}$$

とすれば，(2.9) に対応する w_1, w_2 の推定量は，

$$\hat{w}_1{}^{(2)}=\frac{1}{\Delta_{12}}\left(\frac{1}{2}-\hat{p}_2\right),\qquad \hat{w}_2{}^{(2)}=\frac{1}{\Delta_{12}}\left(\hat{p}_1-\frac{1}{2}\right). \tag{4.3}$$

尤度比を使って $G_i(y)$ をつくったのは，この $\hat{w}_1{}^{(2)}$, $\hat{w}_2{}^{(2)}$ の分散をなるべく小さくするためである．$\hat{w}_1{}^{(2)}$ と $\hat{w}_2{}^{(2)}$ の和は 1 になるとはかぎらない．また，

$$\tilde{w}_1{}^{(2)}=\frac{\frac{1}{2}-\hat{p}_2}{\hat{p}_1-\hat{p}_2},\qquad \tilde{w}_2{}^{(2)}=\frac{\hat{p}_1-\frac{1}{2}}{\hat{p}_1-\hat{p}_2} \tag{4.4}$$

は，$\tilde{w}_1{}^{(2)}+\tilde{w}_2{}^{(2)}=1$ ではあるが，不偏ではない．

その和が 1 で不偏なものは，

$$\hat{h}=\frac{1}{2}(\hat{p}_1+\hat{p}_2)$$

を考えると，共通の確率関数

$$g(\xi)=w_1g_1(\xi)+w_2g_2(\xi) \tag{4.5}$$

をもつ $Y_1, Y_2, \cdots, Y_n$ の同時分布に関する $\hat{h}$ の期待値は，

$$\begin{aligned}\mathscr{E}_{gn}\{\hat{h}\}&=\frac{1}{2}+(w_2-w_1)\frac{\Delta_{12}}{2}\\ &=\frac{1}{2}+\frac{\Delta_{12}}{2}-w_1\Delta_{12}\\ &=\frac{1}{2}-\frac{\Delta_{12}}{2}+w_2\Delta_{12}\end{aligned}$$

である．従ってこれから，w_1, w_2 の不偏推定量は，

$$\begin{aligned}\hat{w}_1&=\frac{1}{2}-\left(\hat{h}-\frac{1}{2}\right)\frac{1}{\Delta_{12}},\\ \hat{w}_2&=\frac{1}{2}+\left(\hat{h}-\frac{1}{2}\right)\frac{1}{\Delta_{12}}.\end{aligned} \tag{4.6}$$

勿論 $\hat{w}_1+\hat{w}_2=1$．又尤度比 $f_2(x)/f_1(x)$ にもとづき，$G(y)$ をつくったここでは，

$$0\leq\Delta_{12}\leq\frac{1}{2} \tag{4.7}$$

であることが示される．

更にまた，$|\hat{w}_i-w_i|$ について，次の確率不等式が成立する，[1]: 予め与えた $\alpha>0$ に対して

$$(4.8)\qquad \mathrm{P}\left\{\left|\hat{h}-\frac{1}{2}-(w_2-w_2)\frac{\varDelta_{12}}{2}\right|<\sqrt{\frac{1}{2n}\log\frac{2}{\alpha}}\right\}$$

$$=\mathrm{P}\left\{|\hat{w}_i-w_i|<\sqrt{\frac{1}{2n\varDelta_{12}{}^2}\log\frac{2}{\alpha}}\right\}>1-\alpha.$$

$\hat{w}_i$ は w_i の不偏推定量であるから，$\hat{w}_i$ の値が負であることがある．しかしながら，$w_1=0$ ならば，

$$\mathrm{P}\left\{\frac{1}{2}+\frac{\varDelta_{12}}{2}-\sqrt{\frac{1}{2n}\log\frac{1}{\alpha}}\geqq\hat{h}\right\}\leqq\alpha,$$

また $w_2=0$ ならば，

$$\mathrm{P}\left\{\frac{1}{2}-\frac{\varDelta_{12}}{2}+\sqrt{\frac{1}{2n}\log\frac{1}{\alpha}}\leqq\hat{h}\right\}\leqq\alpha$$

であるので，予め与えた小さな値 $\alpha>0$ に対して，

$$\hat{h}\leqq\frac{1}{2}+\frac{\varDelta_{12}}{2}-\sqrt{\frac{1}{2n}\log\frac{1}{\alpha}}$$

であれば，$w_1=0$ と考えて処理し，

$$\hat{h}\geqq\frac{1}{2}-\frac{\varDelta_{12}}{2}+\sqrt{\frac{1}{2n}\log\frac{1}{\alpha}}$$

ならば，$w_2=0$ と考えて処理することが考えられる．

5. 判別に対する経験的ベイズ法則（$r=2$ の場合）

2つの部分母集団 π_1, π_2 からなる π を考える．このとき，π からの個体を，X の観測値によって，なるべく正しく，π_1 か π_2 へ判別することを考える．そこで，X の値 x から，個体を π_1 からのものと決定する確率を $\delta^{(1)}(x)$ とすると，確率化決定関数は，$\delta(x)=(\delta^{(1)}(x),\ 1-\delta^{(1)}(x))$ で示される．

今 π からの大きさ n のランダム標本により，(4.6) から得られる $\hat{w}_1, \hat{w}_2$ を使って，次のような決定法則をつくる:

$$(5.1)\qquad \delta_{\hat{w}}{}^{(1)}(x)=\begin{cases}1 & (D_{\hat{w}}(x)\geqq 0\ \text{のとき}),\\ 0 & (D_{\hat{w}}(x)<0\ \text{のとき}),\end{cases}$$

ただし，

$$(5.2)\qquad D_{\hat{w}}(x)=\hat{w}_1f_1(x)-\hat{w}_2f_2(x)$$

である．

このとき，$\delta_{\hat{w}}(x)=(\delta_{\hat{w}}{}^{(1)}(x),\ 1-\delta_{\hat{w}}{}^{(1)}(x))$ の $\boldsymbol{w}=(w_1, w_2)'$ に関する期待危険度を $B(\boldsymbol{w}, \delta_{\hat{w}})$ で示し，そのときのベイズ危険度を $B(\boldsymbol{w})=B(\boldsymbol{w}, \delta_{\boldsymbol{w}})$ で示す．ここで簡単に，正しい決定（判別）に値0を，誤った決定に値1を与えるような損失関数を与えることにすると，

$$(5.3)\qquad 0\leqq B(\boldsymbol{w}, \delta_{\hat{w}})-B(\boldsymbol{w})\leqq\sum_{i=1}^{2}|\hat{w}_i-w_i|.$$

である．それ故，(4.8) から，

$$\text{(5.4)} \qquad \mathrm{P}\left\{B(\boldsymbol{w}, \delta_{\hat{\boldsymbol{w}}}) - B(\boldsymbol{w}) < 2\sqrt{\frac{1}{2n\Delta_{12}^2}\log\frac{2}{\alpha}}\right\} > 1-\alpha$$

が予め与えた $\alpha>0$ に対していえる.

6. p_{12} の不偏推定量の分散

第 3 節で扱った p_{12} の不偏推定量 $\hat{p}_{12}$ と, その分散を考える.

$$(X_1^{(1)}, X_2^{(1)}, \cdots, X_{n_1}^{(1)}), \qquad (X_1^{(2)}, X_2^{(2)}, \cdots, X_{n_2}^{(2)})$$

は, それぞれ, $f_1(x)$ 及び $f_2(x)$ からのランダム標本とする. ただし $f_i(x)$ $(i=1, 2)$ は, (3.1) で与えられる.

$$\text{(6.1)} \qquad Z_{ij} = \begin{cases} 1 & (X_i^{(1)} < X_j^{(2)} \text{ のとき}) \\ 0 & (\text{その他のとき}) \end{cases}$$

及び

$$\text{(6.2)} \qquad z_{ij} = \begin{cases} 1 & (X_i^{(1)} = X_j^{(2)} \text{ のとき}) \\ 0 & (\text{その他のとき}) \end{cases}$$

とすると,

$$\text{(6.3)} \qquad \hat{p}_{12} = \frac{1}{n_1 n_2} U_{12} = \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{Z_{ij} + \frac{1}{2} z_{ij}\right\}$$

をつくれば, $\hat{p}_{12}$ の期待値は,

$$\mathscr{E}\hat{p}_{12} = \mathrm{P}(X^{(1)} < X^{(2)}) + \frac{1}{2}\sum_{v=1}^{s} f_1(x_v) f_2(x_v) = p_{12}.$$

$\hat{p}_{12}$ の分散を求めるために, まず, U_{12}^2 の期待値を計算する.

$$\text{(6.4)} \qquad \mathscr{E}\, U_{12}^2 = \mathscr{E}\left\{\sum_i^{n_1}\sum_j^{n_2}\left(Z_{ij} + \frac{1}{2} z_{ij}\right)\right\}^2$$

$$= \sum_i^{n_1}\sum_j^{n_2}\mathscr{E}\left(Z_{ij} + \frac{1}{2} z_{ij}\right)^2 + \sum_{i\neq\lambda}^{n_1(n_1-1)}\sum_j^{n_2}\mathscr{E}\left(Z_{ij} + \frac{1}{2} z_{ij}\right)\left(Z_{\lambda j} + \frac{1}{2} z_{\lambda j}\right)$$

$$+ \sum_i^{n_1}\sum_{j\neq\mu}^{n_2(n_2-1)}\mathscr{E}\left(Z_{ij} + \frac{1}{2} z_{ij}\right)\left(Z_{i\mu} + \frac{1}{2} z_{i\mu}\right)$$

$$+ \sum_{i\neq\lambda}^{n_1(n_1-1)}\sum_{j\neq\mu}^{n_2(n_2-1)}\mathscr{E}\left(Z_{ij} + \frac{1}{2} z_{ij}\right)\left(Z_{\lambda\mu} + \frac{1}{2} z_{\lambda\mu}\right).$$

$$\mathscr{E}\left(Z_{ij} + \frac{1}{2} z_{ij}\right)^2 = \mathscr{E}\left(Z_{ij}^2 + Z_{ij} z_{ij} + \frac{1}{4} z_{ij}^2\right)$$

$$= \mathrm{P}(X_i^{(1)} < X_j^{(2)}) + \frac{1}{4}\mathrm{P}(X_i^{(1)} = X_j^{(2)})$$

$$= p_{12} - \frac{1}{4}\sum_v f_1(x_v) f_2(x_v).$$

$$\mathscr{E}\left(Z_{ij}+\frac{1}{2}z_{ij}\right)\left(Z_{\lambda j}+\frac{1}{2}z_{\lambda j}\right)$$

$$=\mathscr{E}\left(Z_{ij}Z_{\lambda j}+\frac{1}{2}z_{ij}Z_{\lambda j}+\frac{1}{2}Z_{ij}z_{\lambda j}+\frac{1}{4}z_{ij}z_{\lambda j}\right)$$

$$=\mathrm{P}(X_i^{(1)}<X_j^{(2)},\ X_\lambda^{(1)}<X_j^{(2)})+\mathrm{P}(X_i^{(1)}<X_j^{(2)},\ X_\lambda^{(1)}=X_j^{(2)})$$

$$+\frac{1}{4}\mathrm{P}(X_i^{(1)}=X_j^{(2)},\ X_\lambda^{(1)}=X_j^{(2)})$$

$$=\sum_v(F_1(x_v)-f_1(x_v))^2f_2(x_v)+\sum_v(F_1(x_v)-f_1(x_v))f_1(x_v)f_2(x_v)$$

$$+\frac{1}{4}\sum_v f_1^2(x_v)f_2(x_v)$$

$$=\sum_v\left(F_1(x_v)-\frac{1}{2}f_1(x_v)\right)^2f_2(x_v).$$

$$\mathscr{E}\left(Z_{ij}+\frac{1}{2}z_{ij}\right)\left(Z_{i\mu}+\frac{1}{2}z_{i\mu}\right)$$

$$=\sum_v\left(1-F_2(x_v)+\frac{1}{2}f_2(x_v)\right)^2f_1(x_v).$$

それ故，

$$A=\sum_v\left(F_1(x_v)-\frac{1}{2}f_1(x_v)\right)^2f_2(x_v),$$

$$B=\sum_v\left(1-F_2(x_v)+\frac{1}{2}f_2(x_v)\right)^2f_1(x_v)$$

とおくと，

$$\mathscr{E}U_{12}^2=n_1n_2\left(p_{12}-\frac{1}{4}\sum_v f_1(x_v)f_2(x_v)\right)$$
$$+n_1(n_1-1)n_2A+n_1n_2(n_2-1)B$$
$$+n_1(n_1-1)n_2(n_2-1)p_{12}^2.$$

よって，

$$\mathscr{E}U_{12}^2-(\mathscr{E}U_{12})^2$$

$$=n_1n_2\Big\{p_{12}-\frac{1}{4}\sum_v f_1(x_v)f_2(x_v)+(n_1-1)A+(n_2-1)B$$

$$-(n_1+n_2-1)p_{12}^2\Big\}.$$

かくして，$\hat{p}_{12}=\dfrac{U_{12}}{n_1n_2}$ の分散は，

$$\sigma^2(\hat{p}_{12})=\frac{1}{n_1n_2}\Big\{p_{12}-\frac{1}{4}\sum_v f_1(x_v)f_2(x_v)+(n_1-1)A \tag{6.5}$$

$$+(n_2-1)B-(n_1+n_2-1)p_{12}^2\Big\}.$$

$F_i(x)$ $(i=1,2)$ が連続のときには，

$$A_1=\int_{-\infty}^{\infty}F_1^2(x)\,dF_1(x)=\frac{1}{3},$$

$$B_1=\int(1-F_1(x))^2\,dF_2(x)=\frac{1}{3}$$

であるから，すべての x に対して，$F_1(x)=F_2(x)$ であれば，

$$\sigma^2(\hat{p}_{12})=\frac{n_1+n_2-1}{12n_1n_2}$$

であるが，ここで考えている離散変量の場合には，実際に数値例で当ってみると，

$$A_1=\sum_v\Big(F_1(x_v)-\frac{1}{2}f_1(x_v)\Big)f_1(x_v)\leqq\frac{1}{3}$$

$$B_1=\sum_v\Big(1-F_1(x_v)+\frac{1}{2}f_1(x_v)\Big)f_1(x_v)\leqq\frac{1}{3}$$

となる．そこで

$$\sigma^2(\hat{p}_{12})\leqq\frac{n_1+n_2-1}{12n_1n_2}\tag{6.6}$$

であるようにみえる．

順序統計量を扱うときに，タイを避けるために，通常連続な分布関数 $F(x)$ を考える．連続な $F(x)$ からの，大きさ n のランダム標本による経験分布関数を $\hat{F}_n(x)$ とするとき，$F(x)$ と $\hat{F}_n(x)$ の差や，差の絶対値についての確率的な評価は，離散分布の場合には，幾分控え目なものになるので，そのまゝ使っても差し支えない．このことはよく知られていることであるが，これと同じようなことが，ここでの場合にもあるのではないかと思われる．

7. 適　用　例

第6節で述べた事柄は，この節で扱おうとしている適用例の為に必要になったのである．

データは，日本大学歯学部小児歯科学教室の赤坂守人氏の得たものであって，幼児の歯の齲蝕の有様が，約4ケ月毎に7回，経時的に調べられてある．

第1図，前歯群の0型とあるのは，齲蝕のない幼児であって，86名ある．また I 型とあるのは，4回目から7回目の調査の間で，左上 ($\underline{AB|}$) 前歯に 1〜2 面齲蝕の生じている幼児であって，52名ある．つまり，第1図の6つの図では，横軸に調査回数が，縦軸に齲蝕の生じた歯の面の数が示してある．したがって，II, III, IV, V 型それぞれ図示のような齲蝕の傾向をもつ幼児であって，39名，18名，22名，19名がそれぞれの頻度である．

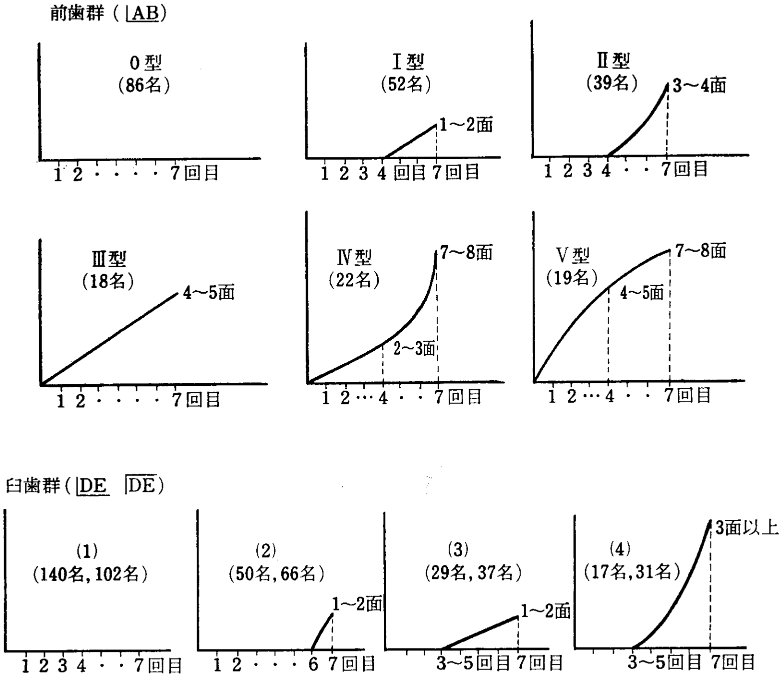
又臼歯群 ($\underline{DE|}$, $\overline{DE|}$) については，(1) とあるのが，齲蝕のない幼児，(2), (3), (4) は，それぞれ図示のような齲蝕の傾向をもつ幼児を示している．

ところで，ここで分析しようとしているのは，前歯に生じた齲蝕の臼歯への関連である．

前歯では，0, I, II, III, IV, V の順序で，臼歯では，(1), (2), (3), (4) の順序で齲蝕が進行しているとみられ，前歯は臼歯より早く生える傾向があるから，まず，0, I〜V の隣合った2つをとり，(1)〜(4) を齲蝕の順位と見做して，$\hat{p}_{\alpha\beta}$ $(\alpha,\beta=0,\mathrm{I},\cdots,\mathrm{V})$ を計算してみる．

はじめに，左上 ($\underline{DE|}$) の臼歯についてみると，

$$\hat{p}_{0\,\mathrm{I}}=0.58,\quad \hat{p}_{\mathrm{I\,II}}=0.65,\quad \hat{p}_{\mathrm{II\,III}}=0.47,\quad \hat{p}_{\mathrm{III\,IV}}=0.64,\quad \hat{p}_{\mathrm{IV\,V}}=0.58$$



第1図 齲蝕の時間的経緯パターン

第1表 前歯群と臼歯群の同時頻度分布

臼歯群 頻度 前歯群		⌊DE (1)	⌊DE (2)	⌊DE (3)	⌊DE (4)	⌈DE (1)	⌈DE (2)	⌈DE (3)	⌈DE (4)	計
⌊AB	0	(0.83) 71	(0.14) 12	(0.02) 2	(0.01) 1	(0.64) 55	(0.25) 21	(0.09) 8	(0.02) 2	86
	I	(0.67) 35	(0.25) 13	(0.04) 2	(0.04) 2	(0.44) 23	(0.29) 15	(0.15) 8	(0.12) 6	52
	II	(0.41) 16	(0.31) 12	(0.20) 8	(0.08) 3	(0.39) 15	(0.28) 11	(0.13) 5	(0.20) 8	39
	III	(0.50) 9	(0.22) 4	(0.17) 3	(0.11) 2	(0.17) 3	(0.28) 5	(0.28) 5	(0.28) 5	18
	IV	(0.27) 6	(0.23) 5	(0.32) 7	(0.18) 4	(0.18) 4	(0.36) 8	(0.32) 7	(0.14) 3	22
	V	(0.16) 3	(0.21) 4	(0.37) 7	(0.26) 5	(0.11) 2	(0.31) 6	(0.21) 4	(0.37) 7	19
		140	50	29	17	102	66	37	31	236

であって，0〜V の隣合った順位の (α, β) については，$p_{\alpha\beta}=0.5$ は棄却できそうにない．試みに，(6.6) の右辺 $(n_1+n_2-1)/12n_1n_2$ を求めてみると，

(0, I) の場合： $0.0026 \fallingdotseq (0.050)^2$,

(I, II) の場合： $0.0037 \fallingdotseq (0.061)^2$,

(II, III) の場合: $0.0066 \fallingdotseq (0.082)^2$,
(III, IV) の場合: $0.0082 \fallingdotseq (0.091)^2$,
(IV, V) の場合: $0.0080 \fallingdotseq (0.089)^2$.

このように, 0〜V の隣合うものでは, 傾向にずれがみえないとみて, III, IV, V を合併して, 0 と合併したもの $\underbrace{\mathrm{III\,V}}$ との間でみると,

$$\hat{p}_{0\,\underbrace{\mathrm{III\,V}}} = 0.79$$

であって, $(n_1+n_2-1)/12n_1n_2=144/12\times86\times59=0.0024\fallingdotseq(0.049)^2$. これから, 前歯の齲蝕が臼歯の齲蝕に幾許か関連をもつものと考えられる.

この傾向は, 左下 ($\lceil$DE) の臼歯についてみても同様であって, 念のために計算してみると,

$$\hat{p}_{0\,\mathrm{I}} = 0.62,\quad \hat{p}_{\mathrm{I\,II}} = 0.55,\quad \hat{p}_{\mathrm{II\,III}} = 0.63,\quad \hat{p}_{\mathrm{III\,IV}} = 0.43,\quad \hat{p}_{\mathrm{IV\,V}} = 0.62.$$

又, 0 と III, IV, V を合併したものについては,

$$\hat{p}_{0\,\underbrace{\mathrm{III\,V}}} = 0.80.$$

参 考 文 献

[1] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Ass.*, **53**, 13–30.
[2] Hudimoto, H. (1968). On the empirical Bayes procedure (1), *Ann. Inst. Statist. Math.*, **20**, 169–185.
[3] Hudimoto, H. (1976). On the empirical Bayes approach to classification problems, *Essays in Probability and Statistics*, The Committee for Publishing the Birthday Volume in Honor of Prof. Ogawa.
[4] 藤本 熙, 松原 望 (1976). 決定の数理, 筑摩書房.
[5] Hudimoto, H. (1979). On the empirical Bayes approach to classification in the case of discrete multivariate distribution having only finite mass points, *Ann. Inst. Staist. Math.*, **31**, A, 1–7.
[6] Hudimoto, H. (1980). Empirical Bayes two-way decision in the case of discrete distributions, *Recent Developments in Statistical Inference and Data Analysis* (ed. K. Matusita), North-Holland Publishing Company, Amsterdam · New York · Oxford.
[7] Klotz, J. (1966). The Wilcoxon, ties and the computer, *J. Amer. Statist. Ass.*, **61**, 772–787.
[8] Putter, J. (1955). The treatment of ties in some nonparametric tests, *Ann. Math. Statist.*, **26**, 368–386.