Informationの多変数解析への応用について

羽鳥, 司
東京学芸大学

https://doi.org/10.15017/12723

出版情報：統計科学研究. 1 (3), pp.15-27, 1956-09. Research Association of Statistical Sciences
バージョン：

Informationの多変数解析への応用について

羽鳥　司（東京学芸大）

§1. 序

最近 *information* を用いて統計解析を行う傾向があり，以下述べて見たいと思う．

定義1

確率密度函数 $f_1(x)$, $f_2(x)$ の間の差 $I(f_1, f_2)$ とは

(1) …… $$I(f_1, f_2) = I(1,2) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} dx$$

以下積分範囲は特にことわらないときは $(-\infty, \infty)$ とす．

又，弁別距離（*discriminative information or divergence*）とは

(2) …… $$J(f_1, f_2) = J(1,2) = \int [f_1(x) - f_2(x)] \log \frac{f_1(x)}{f_2(x)} dx$$

こゝで(2)は Mahalanobis の *distance* などを特別の場合として含んでおり，Shannon, Wiener の *information*の考えによる *measure* である．S. Kullback, R. A. Leiblerが1951 年の論文で使われ始めた．

次ぎに，これを k 変数正規母集団の場合に適用したらどうなるか．証明を述べず結果だけのべよう．

定理1.

二つの k 変数正規分布の密度函数 $f_1(x)$, $f_2(x)$ において

(3) …… $$f_i(x) = \frac{1}{\sqrt{(2\pi)^k |\sigma_{(i)}|}} e^{-\frac{1}{2}(x-\mu_{(i)})' \sigma_{(i)}^{-1} (x-\mu_{(i)})}$$

こゝで　$\mu_{(i)}$ は *mean vector*

$\sigma_{(i)}$ は *variance and covariance matrix*

$x' = (x_1, x_2, \cdots, x_k)$

$$\mu'_{(i)} = (\mu_{1(i)}\ \mu_{2(i)},\ \cdots,\ \mu_{k(i)})$$
$$\sigma_{(i)} = (\sigma_{st}) \qquad s, t = 1, \cdots, k \quad 、\sigma_{ts} = \sigma_{st}$$
$$i = 1, 2$$

とするとき

$$I(1, 2) = I(f_1, f_2)$$

(4) ……
$$= \frac{1}{2} \log \frac{|\sigma_{(2)}|}{|\sigma_{(1)}|} + \frac{1}{2}\, tr(\sigma_{(1)} - \sigma_{(2)})\, \sigma_{(2)}^{-1} + \frac{1}{2}\, \delta'\, \sigma_{(2)}^{-1}\, \delta$$

ただし　$\delta = \mu_{(1)} - \mu_{(2)}$

$$J(1, 2) = J(f_1, f_2)$$

(5) ……
$$= \frac{1}{2}\, tr\left[(\sigma_{(1)} - \sigma_{(2)})(\sigma_{(2)}^{-1} - \sigma_{(1)}^{-1})\right] + \frac{1}{2}\, \delta'(\sigma_{(1)}^{-1} + \sigma_{(2)}^{-1})\, \delta$$

or

(6) ……
$$= \frac{1}{2}\, tr\, \sigma_{(1)} \sigma_{(2)}^{-1} + \frac{1}{2}\, tr\, \sigma_{(2)} \sigma_{(1)}^{-1} - k + \frac{1}{2}\, \delta'(\sigma_{(1)}^{-1} + \sigma_{(2)}^{-1})\, \delta$$

証明.　S. Kullback : *An application of information theory to multivariate analysis.* Ann. Math. Stat. Vol. 23 (1952)参照.

当然のことであるが，次の系がでる.

系1．1

両母集団の平均ベクトル等しいとき，及び分散共分散行列等しいときは次のように簡単になる。

(7) $\mu_{(1)} = \mu_{(2)} \Rightarrow I(1, 2) = \frac{1}{2} \log \frac{|\sigma_{(2)}|}{|\sigma_{(1)}|} + \frac{1}{2}\, tr(\sigma_{(1)} - \sigma_{(2)})\, \sigma_{(2)}^{-1}$

(8) $\mu_{(1)} = \mu_{(2)} \Rightarrow J(1, 2) = \frac{1}{2}\, tr(\sigma_{(1)} - \sigma_{(2)})(\sigma_{(2)}^{-1} - \sigma_{(1)}^{-1})$

例えば一変数のときは $J(1, 2) = \frac{1}{2}(\sigma_1^2 - \sigma_2^2)\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right) = \frac{1}{2}\left(\frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2}\right) - 1$.

(9) $\sigma_{(1)} = \sigma_{(2)} \Rightarrow I(1, 2) = \frac{1}{2}\, \delta' \sigma\, \delta \qquad \sigma \equiv \sigma_{(1)} = \sigma_{(2)}$

(10) $\sigma_{(1)} = \sigma_{(2)} \Rightarrow J(1, 2) = \delta' \sigma\, \delta = 2I(1, 2)$

これは常数倍のちがいでMahalanobis の*distance*を表わす．例えば一変数の場合だと

$$2I(1, 2) = J(1, 2) = \frac{(\mu_1 - \mu_2)^2}{\sigma^2}$$　である。

§2. Maximum information principleについて

今 $I(1,2)\overset{/}{=}I(f_1,f_2)$ において f_2 が既知のとき *Information* の差 $I(f_1,f_2)$ を最大ならしむるように f_1 をきめる．この *principle* を *Maximum information principle* とよび，これを用いて，検定，推定等の *principle* にする．

先ず，その一般論として次にのべる．

<u>定理2</u>

$I(g^*(x),\ g_2(x|\theta))$ において

$g_2(x|\theta)$ は与えられたとし

$g^*(x)$ の平均 $\int x g^*(x)dx=a$ も分ったとき

$\Rightarrow I(g^*(x),\ g_2(x|\theta))$ を *Maximize* する g^* は

(11) …… $$g^*(x)=e^{tx}g_2(x)/M_2(t)$$

たゞし

$$M_2(t)=\int e^{tx}g_2(x)\,dx$$

<u>証明</u>

(12) …… $$I(g^*,g_2)=\int g^*(x)\log\frac{g^*(x)}{g(x)}\,dx$$

(13) ……今，條件 $\int g^*dx=1,\quad \int xg^*dx=a$

の下で求める

(14) ……これは $X=\int\left(g^*(x)\log\frac{g^*(x)}{g_2(x)}+kg^*(x)+\ell xg^*(x)\right)dx$

を *Max.* にすることと同値

故に変分法で

(15) …… $$\delta X=0=\int\delta g^*(x)\left[\log\frac{g^*(x)}{g_2(x)}+k+\ell x\right]dx$$

ゆえに $\log\frac{g^*(x)}{g_2(x)}+k+\ell x=0$

(16) よって $$g^*(x)=e^{-k-\ell x}g_2(x)$$

又 $$1=\int g^*(x)\,dx=\int e^{-k-\ell x}\,g_2(x)$$

(17) $$=e^{-k}\int e^{-\ell x}\,g_2(x)$$

$-\ell=t$ とおくと

$$=e^{-k}\int e^{tx}\,g_2(x)$$

(18) $$=e^{-k}\,M_2(t)$$

(19) よって $g^*(x)=e^{tx}\,g_2(x)\big/M_2(t)$

(20) $$M_2(t)=\int e^{tx}\,g_2(x) \qquad (Q.E.D)$$

これで最大ならしめるものの形がわかった。

このときの最大 *information* $I^*(\theta,t)$ は

<u>系2，1</u>

$$I^*(\theta,t)\equiv I(g^*(x),\ g_2(x|\theta))$$

(21) …… $$=\int g^*(x)\log\frac{g^*(x)}{g_2(x)}\,dx=at-\log M_2(t)$$

<u>証明</u>． 明かである．

又，弁別距離 $J(g^*,g_2)$ の方については

<u>系2，2</u>．

(22) …… $$J^*(\theta,t)\equiv J(g^*,g_2)=\int(g^*-g_2)\log\frac{g^*}{g_2}\,dx$$

$$=t(a-E_2(x))$$

<u>証</u> $$\int g_2(x)\log\frac{g^*(x)}{g_2(x)}\,dx=\log M_2(t)-tE_2(x)$$

より出る．（終）

次に a,t 共に *Matrix* で支えられているときに拡張すると容易に分る．

<u>系2，3</u>．

(23) …… $$I^*(\theta,t)=tr[a,t]-\log M_2(t)$$

§3. k変数正規母集団の解析への応用

Maximum information principle を多変数正規母集団の等分散仮説の検定，等平均仮説の検定等への応用をのべる．

先ず，上記の $M_2(t)$ 即ち *moment of generating function* の一般化したものを求めよう．

補題1

$i=1,\cdots,r$ として π_i なる r 個の k変数正規母集団からそれぞれ n_i 個の *observations* をとったとき，その *sample mean vector* を $\bar{x}_i'=(\bar{x}_{i1},\cdots,\bar{x}_{ik})$．*unbiased sample variance and covariance matrix* を S_i 即ち 変動 V_i，自由度 N_i としたとき $V_i=N_i S_i$ $(i=1,\cdots,r)$

$(\bar{x}_1,\cdots,\bar{x}_r,V_1,V_2,\cdots,V_r)$ の *moment of generating function* $M_2(t)$ は

$$(24)\quad\cdots\quad M_2(t)=M_2(t_1,\cdots,t_r,T_1,\cdots,T_r)$$

$$=\prod_{i=1}^{r}\left|I-2\sigma_i T_i\right|^{-\frac{N_i}{2}}exp(t_i'\mu_i+\frac{1}{2}t_i'\frac{\sigma_i}{n_i}t_i)$$

たゞし $t_i=(t_{i1},t_{i2},\cdots,t_{ik})$

$T_i=(t_{lm}^{(i)})$ $\quad l,m=1,\cdots,k\quad$ k次の *matrix*

I は単位行列

証明

$(\bar{x}_1,\cdots,\bar{x}_r)$ は正規分布とし，$(V_1,\cdots,V_r)$ は独立に Wishart 分布をすることを考慮すれば S.S.Wilks の *Mathematical Statistics* にある低次元の場合を同様に拡張出来る．(終)

この $M_2(t)$ を用いて多変量解析の場合の(21)式にある $I^*(\theta|t)$ に相当するものを求めよう．

補題2

$\theta=\mu_1,\cdots,\mu_r,\sigma_1,\sigma_2,\cdots,\sigma_r)$

$a=(\bar{x}_1,\cdots,\bar{x}_r,V_1,V_2,\cdots,V_r)$

とするとき，前定理の系の $I^*(\theta|t)$ は次のように与えられる．

$$(25)\quad\cdots\quad I^*(\theta,t)=\sum_{i=1}^{r}t_i'\bar{x}_i-t_i'\mu_i-\frac{1}{2}t_i\frac{\sigma_i}{n_i}t_i$$

$$+\sum_{i=1}^{r} tr\ T_i V_i$$

$$+\frac{1}{2}\sum_{i=1}^{r} N_i \log|I-2\sigma_i T_i|$$

<u>証明</u>　$I^*(\theta,t)=tr(a,t)-\log M_2(t)$

こゝで $t=(t_1,\cdots\cdots,t_r,\ T_1,\cdots,\ T_r)$ であるから $tr(a,t)=\sum t_i'\bar{x}_i+\sum tr$ $(T_i V_i)$

$$\log M_2(t)=\sum t_i'\mu_i+\frac{1}{2}t_i'\frac{\sigma_i}{n_i}t_i-\sum\frac{N_i}{2}\log|I-2\sigma_i T_i|$$

からである．（終）

<u>補題3</u>

(25)式の $I^*(\theta,t)$ を $Maximize$ する $t=(t,\cdots,t_r,T_1,\cdots,T_r)$ は

(26)　$t_i=n_i\sigma_i^{-1}(\bar{x}_i-\mu_i)$

(27)　$T_i=\frac{1}{2}\sigma_i^{-1}-\frac{1}{2}S_i^{-1}$

であり，その最大値 $I^*(\theta)$ は

(28)　……　$I^*(\theta)=\sum\frac{n_i}{2}(\bar{x}_i-\mu_i)'\sigma_i^{-1}(\bar{x}_i-\mu_i)$

$$+\sum\frac{N_i}{2}\left(\log\frac{|\sigma_i|}{|S_i|}-k+tr[\sigma_i^{-1},S_i]\right)$$

<u>証明</u>

(29)　$\frac{\partial}{\partial t_i}I^*(\theta,t)=\bar{x}_i-\mu_i-\frac{\sigma_i}{n_i}t_i=0$

$$\frac{\partial}{\partial T_i}I^*(\theta,t)=0$$

(30)　すなわち　$tr(dT_i)V_i-N_i tr(I-2\sigma_i T_i)^{-1}\sigma_i(dT_i)=0$

これを計算すると

$$tr(I-2\sigma_i T_i)^{-1}\sigma_i=\frac{V_i}{N_i}=S_i$$

よって $T_i=\frac{1}{2}(\sigma_i^{-1}-S_i^{-1})$,　(26),（27)を得た。

これを(25)式に代入すると

$t_i'(\bar{x}_i-\mu_i)=n_i(\bar{x}_i-\mu_i)'\sigma_i^{-1}$　（σ_i は $symmetric$ だから）

$$\frac{1}{2}\, t_i' \frac{\sigma_i}{n}\, t_i = \frac{1}{2}\left[n_i (\bar{x}_i - \mu_i)' \sigma_i^{-1} \right] \frac{\sigma_i}{n_i} \left[n_i \sigma_i^{-1} (\bar{x}_i - \mu_i) \right]$$

$$= \frac{1}{2}\, n_i (\bar{x}_i - \mu_i)' \sigma_i^{-1} (\bar{x}_i - \mu_i)$$

$$tr\, T_i V_i = tr\left[\left(\frac{1}{2} \sigma_i^{-1} - \frac{1}{2} S_i^{-1} \right) N_i S_i \right]$$

$$= tr\left[\frac{1}{2} N_i \left(\sigma_i^{-1} S_i - I \right) \right]$$

$$= \frac{1}{2} N_i \left(tr\, \sigma_i^{-1} S_i - k \right)$$

$$I - 2\sigma_i T_i = I - 2\sigma_i \left(\frac{1}{2} \sigma_i^{-1} - \frac{1}{2} S_i^{-1} \right)$$

$$= I - I + \sigma_i S_i^{-1} = \sigma_i S_i^{-1}$$

$$\log |I - 2\sigma_i T_i| = \log |\sigma_i S_i^{-1}| = \log \frac{|\sigma_i|}{|S_i|}$$

これらより (28)式を得ることが出来る．（終）

いよいよ等平均仮説検定の問題にうつろう。

定理 3

$$H_2 : \mu_1 = \mu_2 = \cdots = \mu_r \; (= \mu)$$

$$\sigma_1, \sigma_2, \cdots \quad \sigma_r$$

と仮説をたてたとき

$I^*(\mu, \sigma_1, \sigma_2, \cdots, \sigma_r)$ の Mim. を与えるものは

$$(31) \cdots \quad \begin{cases} \hat{\mu} = \left(\sum_{i=1}^{r} n_i S_i^{-1} \right)^{-2} \left(\Sigma n_i S_i^{-1} \bar{x}_i \right) \\ \hat{\sigma}_i = S_i \end{cases}$$

かつ，その値 $\hat{I}$ は

$$(32) \cdots \quad \hat{I} \equiv I^*(\hat{\mu}, \hat{\sigma}, \cdots, \hat{\sigma}_r)$$

$$= \frac{1}{2} \Sigma n_i (\bar{x}_i - \bar{x})' S_i^{-1} (\bar{x}_i - \bar{x})$$

たゞし $\bar{x} = \left(\Sigma n_i S_i^{-1} \right)^{-1} \left(\Sigma n_i S_i^{-1} \bar{x}_i \right)$ とす

証明 $\quad I^*(\mu, \sigma_1, \cdots, \sigma_r)$

(33) …… $= \sum \frac{n_i}{2}(\bar{x}_i - \mu)' \sigma_i^{-1}(x_i - \mu) + \sum \frac{N_i}{2}\left(\log \frac{|\sigma_i|}{|S_i|} - k + t_r \sigma_i^{-1} S_i\right)$

今 σ_i について minimize するものは容易に分るように $\hat{\sigma}_i = S_i$

又 μ につき $0 = \delta I^* = \sum n_i S_i^{-1}(\bar{x}_i - \mu) = \sum n_i S_i^{-1} \bar{x}_i - \mu \sum n_i S_i^{-1}$

ゆえに $\mu = (\sum n_i S_i^{-1})^{-1} (\sum n_i S_i^{-1} \bar{x}_i)$

$\hat{I}$ は代入すれば求められる. (終)

次は，等分散仮説の場合である.

定理4

$H_2 : \sigma_1 = \sigma_2 = \cdots = \sigma_r \ (= \sigma), \ \mu_1, \cdots, \mu_r$

$I^*(\mu_1, \mu_2, \cdots, \mu_r, \sigma)$ を Min. ならしめるものは

(35) …… $\begin{cases} \hat{\mu}_i = \bar{x}_i \\ \hat{\sigma}_i = \frac{1}{\sum N_i}\left(\sum_{i=1}^{r} N_i S_i\right) = \frac{1}{N}\left(\sum N_i S_i\right) = S \end{cases}$

($\sum N_i S_i \equiv NS$ とすれば)

又その値 $\hat{I}$ は

(36) …… $\hat{I} = I^*(\hat{\mu}_1, \cdots, \hat{\mu}_r, \hat{\sigma})$

$$= \frac{1}{2} \sum N_i \log \frac{|S|}{|S_i|}$$

証明

$I^*(\mu_1, \cdots, \mu_r, \sigma)$ は

$$= \sum \frac{n_i}{2}(\bar{x}_i - \mu_i)' \sigma^{-1}(\bar{x}_i - \mu_i) + \sum \frac{N_i}{2}\left(\log \frac{|\sigma|}{|S_i|} - k + t_r \sigma^{-1} S_i\right)$$

δ をとると μ_i につき

(37) …… $0 = \delta I^* = n_i \sigma^{-1}(\bar{x}_i - \mu_i) = 0$ ゆえに $\mu_i = \bar{x}_i$

σ について

(38) …… $0 = \delta I^* = \sum -\frac{n_i}{2}(\bar{x}_i - \mu_i)'(\sigma^{-1})(\delta\sigma)(\sigma^{-1})(\bar{x}_i - \mu_i)$

$$+ \sum \frac{N_i}{2} t_r \sigma^{-1}(\delta\sigma) - \sum \frac{N_i}{2} t_r S_i \sigma^{-1}(\delta\sigma)\sigma^{-1}$$

ゆえに

$$\sum N_i \sigma = \sum N_i S_i, \quad N\sigma = \sum N_i S_i, \quad \text{ゆえに} \quad \sigma = \frac{1}{N} \sum N_i S_i$$

次に $I^*(\mu_1, \cdots, \mu_r, \sigma)$ に $\hat{\mu}_1 \cdots, \hat{\mu}_r, \hat{\sigma}$ を入れると

$$I^*(\hat{\mu}_1,-\hat{\mu}_r,\hat{\sigma})=\frac{1}{2}\sum N_i\left\{\log\frac{|S|}{|S_i|}-k+t_r(S^{-1}S_i)\right\}$$

$$=\frac{1}{2}\sum N_i\log\frac{|S|}{|S_i|}-\frac{1}{2}k\sum N_i+\frac{1}{2}t_r S^{-1}\sum N_i S_i$$

$$=\frac{1}{2}\sum N_i\log\frac{|S|}{|S_i|}-\frac{k}{2}N+\frac{1}{2}t_r S^{-1}NS$$

$$=\frac{1}{2}\sum N_i\log\frac{|S|}{|S_i|}-\frac{k}{2}N+\frac{1}{2}Nk$$

$$=\frac{1}{2}\sum N_i\log\frac{|S|}{|S_i|}$$ （証終）

系４，１

定理４の I^* に対する J^* の $Min.$ の値は

(39) …… $$\hat{J}=J^*(\hat{\mu}_1,\cdots,\hat{\mu}_r,\hat{\sigma})$$

$$=\sum_{i>j}\frac{N_iN_j}{2N}\left(t_r S_i S_j^{-1}+t_r S_j S_i^{-1}-2k\right)$$

証明

$$\hat{J}=\sum\frac{N_i}{2}\left(t_r S_i S^{-1}+t_r S S_i^{-1}\right)-kN$$

$$=t_r\left(\sum\frac{N_i}{2}S_i\right)S^{-1}+\sum\frac{N_i}{2}t_r SS_i^{-1}-kN$$

$$=t_r\left[\frac{1}{2}(NS)S^{-1}\right]+\sum\frac{N_i}{2}t_r SS_i^{-1}-kN$$

$$=\sum\frac{N_i}{2}t_r SS_i^{-1}-\frac{k}{2}N$$

$$=\sum\frac{N_i}{2}t_r\left(\sum_j\frac{1}{N}N_jS_j\right)S_i^{-1}-\frac{k}{2}N$$

$$=\frac{1}{2N}\sum_{i,j}N_iN_j\left(t_r S_j S_i^{-1}\right)-\frac{k}{2}N$$

$$=\sum_{i>j}\frac{N_iN_j}{2N}\left\{t_r S_j S_i^{-1}+t_r S_i S_j^{-1}-2k\right\}$$ （終）

この定理を用いてつぎに π_1, π_2 二つの正規母集団の判別函数を求めて見よう.

定理５　π_1,π_2 の判定函数 $y=\alpha'x=\sum_{i=1}^{k}\alpha_i x_i$ は S_1, S_2 を夫々 π_1, π_2 からの N_1, N_2 個 *sample* の *unbiased variance and covariance* としたとき $S_1\alpha=\lambda S_2\alpha$ を満足する α である。たゞし λ は $\det|S_1-\lambda S_2|=0$ の最大根又は最小根である.

証明

定理４において　$r=2$ の場合を考えると

(40)　…　$\hat{J}=\frac{N_1N_2}{2N}(tr\,S_1S_2^{-1}+tr\,S_2S_1^{-1}-2k)$

こゝでは y についての $\hat{J}$ 即ち $\hat{J}(y)$ を考えるわけである．yは正規分布をするから $E_1(y)=\alpha E_1(x)$，　$E_2(y)=\alpha E_2(x)$，　$\sigma_1^2(y)=\alpha'\sigma_1\alpha$，　$\sigma_2^2(y)=\alpha'\sigma_2\alpha$
又　$\hat{\sigma}_1=S_1$　$\hat{\sigma}_2=S_2$ を用い，一次元のことに注意すれば

(41)　…　$\hat{J}'(y)=\frac{N_1N_2}{2N}\left(\frac{\alpha'S_1\alpha}{\alpha'S_2\alpha}+\frac{\alpha'S_2\alpha}{\alpha'S_1\alpha}-2\right)$

を得る．これを最大ならしむるには $S_1\alpha=\lambda\cdot S_2\alpha$ なる α をとれば良い．
そこで $|S_1-\lambda S_2|=0$ の決定方程式の根を求めるとよい．それを $\lambda_1,\cdots,\lambda_k$ その一つを λ_i とすれば

$$\hat{J}'(y,\lambda_i)=\frac{N_1N_2}{2N}\left(\lambda_i+\frac{1}{\lambda_i}-2\right)$$

よって，もし $\lambda_1\lambda_k<1$ なら $\lambda_i=\lambda_1$ をとればよいし，$\lambda_1\lambda_k>1$ なら $\lambda_i=\lambda_k$ 即ち最小根をとればよい．　（証終）

<u>系５，１</u>　　判別函数に求めるときに出た $\hat{J},\hat{J}'(y,\lambda_i)$ について

(42)　…　$\hat{J}=\sum_{i=1}^{k}\hat{J}'(y,\lambda_i)$

<u>証明</u>　　$tr\,S_1S_2^{-1}=\sum_{i=1}^{k}\lambda_i$，　$tr\,S_2S_1^{-1}=\sum_{i=1}^{k}\frac{1}{\lambda_i}$

なることを注意すればよい．

§４．次に線型回帰の問題を扱って見よう．

<u>補題４</u>

(43)　…　$Z_{(i)}=Y_{(i)}-BX_{(i)}$　　　$i=1,\cdots,n$

こゝで　$Z_{(i)}$ は k_2 変数正規分布（平均＝０，*covariance matrix* は皆 σ）とする．

$$\left.\begin{aligned}X'_{(i)}&=(x_{1i},\cdots\cdots,x_{k_1i})\\ Y'_{(i)}&=(y_{1i},\cdots\cdots,y_{k_2i})\qquad k_1\geq k_2\end{aligned}\right\},\quad B=(b_{rs})_{\substack{1\leq r\leq k_2\\1\leq s\leq k_1}}$$

更に　$X'=(X_{(1)},\cdots\cdots,X_{(n)})$
　　　$Y'=(Y_{(1)},\cdots\cdots,Y_{(n)})$．

$\Longrightarrow B$ の *unbiased estimate* $\hat{B}$ は

(44) …… $\hat{B}=(Y'X)(X'X)^{-1}$, 又 σ の *estimate* $\hat{\sigma}$ は $(n-k_1)\hat{\sigma}=(Y'Y)-(Y'X)(X'X)^{-1}(X'Y)$

<u>証明</u> 以下 *Matrix* の行数，列数を明らかにするために p 行 q 列の *matrix* A を $A(p,q)$ で表わす．

$Z(n,k_2)=Y(n,k_2)-X(n,k_1)B'(k_1,k_2)$ から

$B(k_2,k_1)X'(k_1,n)=Y'(k_2,n)-Z'(k_2,n)$ 　両辺に $X(n,k)$ を乗ずる

$B(k_2,k_1)X'(k_1,n)X(n,k)=[Y'(k_2,n)-Z'(k_2,n)]X(n,k)$

$X'X$ は (k_1,k_1) だから *inverse* 存在するから

$B(k_2,k_1)=Y'(k_2,n)X(n,k_1)(X'X)^{-1}-Z'(k_2,n)X(n,k_1)(X'X)^{-1}$

B の *unbiased estimate* は $E(Z)=0$ だから

$$\hat{B}(k_2,k_1)=Y'(k_2,n)X(n,k_1)(X'X)^{-1}$$

即ち $\hat{B}=(Y'X)(X'X)^{-1}$

次に σ については

$(n-k_1)\hat{\sigma}=\hat{Z}'\hat{Z}$ より 今の $\hat{B}$ を代入すれば

$$=Y'Y-(Y'X)(X'X)^{-1}(X'Y)$$ (終)

<u>定理 6</u>

$Z_{(i)}=Y_{(i)}-BX_{(i)}$ において

(45) …… $\begin{cases} H_1: E(Y_{(i)})=BX_{(i)} & i=1,\cdots,n \\ H_2: E(Y_{(i)})=0 & i.e \quad B=0 \end{cases}$

$\Rightarrow J(1,2)=2I(1,2)$

(46) …… $=(n-k_1)\left(\frac{r_1^2}{1-r_1^2}+\cdots\cdots+\frac{r_{k_2}^2}{1-r_{k_2}^2}\right)$

たゞし $r_1,\cdots\cdots,r_{k_2}$ は *canonical correlation coefficient* とする．

<u>証明</u>

$J(1,2)=\delta'\sigma^{-1}\delta$ 　(σ は共通だから)

$$=(X'_{(1)}B',\cdots\cdots,X'_{(n)}B')\begin{pmatrix}\sigma^{-1} & & \\ & \sigma^{-1} & \\ & & \ddots \\ & & & \sigma^{-1}\end{pmatrix}\begin{pmatrix}BX_{(1)}\\ \vdots \\ BX_{(n)}\end{pmatrix}$$

$$=\Sigma X'_{(i)}B'\sigma^{-1}BX_{(i)}$$

こゝで $X'_{(i)}B'$ は $(1,k_2)$ 型の *matrix*

$\sigma^{-1}BX_{(i)}$ は $(k_2,1)$ 型の *matrix*

よって　$Q(1,k_2)P(k_2,1)=tr\ PQ$　の事実により上は Σ の中は

$tr\ \sigma^{-1}BX_{(i)}\cdot X'_{(i)}B'$　にひとし

故に　上式は

$$=\Sigma\, tr\ \sigma^{-1}BX_{(i)}\cdot X'_{(i)}B'$$
$$=tr\ \sigma^{-1}\Sigma\, B\, X_{(i)}X'_{(i)}B'$$
$$=tr\ \sigma^{-1}B(\Sigma X_{(i)}X'_{(i)})B'$$
$$=tr\ \sigma^{-1}BX'XB'$$

こゝで σ, B の上の *lemma* の *estimate* を代入し

$$\hat{J}(1,2)=tr\ \hat{\sigma}^{-1}\hat{B}XX'\hat{B}'$$ を計算すると

$$=(n-k_1)\,tr\,[YY'-(Y'X)(X'X)^{-1}(X'Y)]^{-1}$$
$$X(Y'X)(X'X)^{-1}(X'Y)$$

こゝで $X'X$, $X'Y$, $Y'X$, $Y'Y$ を夫々 nS_{11}, nS_{12}, nS_{21}, nS_{22} で表わすと

$$\hat{J}(1,2)=(n-k_1)\,tr\ S_{22,1}^{-1}\ S_{21}S_{11}^{-1}S_{12}$$

こゝで　$S_{22.1}=S_{22}-S_{21}S_{11}^{-1}S_{12}$　とす.

$det\,|\ell\ S_{22.1}-S_{21}S_{11}^{-1}S_{12}|=0$　の根を　$\ell_1,\ell_2,\cdots,\ell_{k_2}$　とすれば

(47)　…　$\hat{J}(1,2)=(n-k_1)\,(\ell_1+\ell_2+\cdots+\ell_{k_2})$

$det\,|r^2S_{22}-S_{21}S_{11}^{-1}S_{12}|=0$　の根 r と前の根 ℓ の関係は $\ell=\frac{r^2}{1-r^2}$

ゆえに

$$\hat{J}(1,2)=(n-k_1)\left(\frac{r_1^2}{1-r_1^2}+\cdots\cdots+\frac{r_{k_2}^2}{1-r_{k_2}^2}\right)$$

よって $\hat{J}(1,2)$ が *canomical correlation coefficient* を用いて表わされた.

（証終）

<u>系６．１</u>　　定理６の場合の判別函数　$\alpha_1Y_{i1}+\alpha_2Y_{i2}+\cdots+\alpha_{k_2}Y_{k_2}$　は

(48)　…　$det\,|BX'XB-\ell\sigma|=0$　の根に対応するものである.

<u>証明</u>　は前の定理５のときと同様である.

<u>**結び**</u>　*information* の考えで実験計画法，線型計画法にも応用する理論が作れると思われる.

――――――　○　――――――

<u>註１</u>.　S. Kullback は定理４に相当するところにおいて

$$I(1,2)=\Sigma\frac{N_i}{2}\left(log\frac{|\sigma|}{|\sigma_i|}+tr\ \sigma_i\sigma^{-1}\right)-\frac{k}{2}N$$

と出しているが $\sigma_i = \sigma$ を入れると $I(1,2)=0$ になるから不充分と思う。

$\hat{\sigma}_i = S_i$, $\hat{\sigma} = S$ を代入して $\hat{I}(1,2) = \Sigma \frac{N_i}{2}\left(\log\frac{|S|}{|S_i|} + tr\, S_i S^{-1}\right) - \frac{k}{2}N$

$= \Sigma \frac{N_i}{2} \log\frac{|S|}{|S_i|}$ と出している。

註2

定理3の $2\hat{I}$ は *asymptotically* に χ^2 分布自由度 $(r-1)k$ とし

定理4の $2\hat{I}$ は $(r-1)k(k+1)/2$ の自由度の χ^2 分布をする

定理5の $\hat{J}$ については χ^2 分布の予想がたてられている

定理6の $\hat{J}$ については $k_1 \cdot k_2$ の自由度の χ^2 分布をする

ことが S. Kullback の論文 1956 年の最後の章に出ている。

参　考　文　献

[1] H. Chernoff, "*A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations.*" AMS, 23 (1952).

[2] 〃 "*Large sample theory: Parametric case*" AMS, 27 (1956).

[3] R. A. Fisher "*Contributions to Mathematical Statistics*" 1950.

[4] H. Hotelling, "*A generalized T test and measure of multivariate dispersion.*" Second Berkeley Symposium, Univ. of California Press 1951, pp. 23-41.

[5] S. Kullback, "*An application of information theory of multivariate analysis.*" A.M.S., 25 (1954).

[6] 〃 "*Certain inequalities in information theory and the Cramér-Rao inequality.*" A.M.S., 25 (1954).

[7] 〃 "*An application of information theory to multivariate analysis II.*" A.M.S., 27 (1956).

[8] S. Kullback *and* R. A. Leibler, "*On information and sufficiency.*" A.M.S., 22 (1951).

[9] S. S. Wilks, "*Mathematical Statistics.*" 1943.