ANALISIS ESTADISTICO MULTIVARIANTE Y REPRESENTACION CANONICA DE FUNCIONES ESTIMABLES

TESIS PARA OPTAR AL GRADO
DE DOCTOR EN CIENCIAS, SECCION
DE MATEMATICAS, PRESENTADA POR

CARLOS M. CUADRAS AVELLANA

UNIVERSIDAD DE BARCELONA

$$\sqrt{n}_j(\bar{y}_j - m_j^*) \quad \text{es} \quad N(0,\Sigma),$$

y estocásticamente independiente de $(n-k)\hat{\Sigma}$. Según el teorema anterior, la v.a.

$$T^2 = n_j(\bar{y}_j - m_j^*)' \hat{\Sigma}^{-1}(\bar{y}_j - m_j^*)$$

es tal que

$$\frac{T^2}{(n-k)} \frac{n-k-p+1}{p}$$

sigue una distribución F de Fisher, con p y $(n-k-p+1)$ grados de libertad.

La estimación de m_j^* es $\hat{m}_j^* = \bar{y}_j$, y el conjunto

$$\|\hat{m}_j^* - m_j^*\|^2_{\hat{\Sigma}^{-1}} \leq \frac{F_\alpha}{n_j} \frac{(n-k)p}{(n-k-p+1)}$$

es un elipsoide en E^*, y la probabilidad de que m_j^*, que es desconocido, pertenezca a este elipsoide es $(1-\alpha)$, siendo F_α tal que $P(F > F_\alpha) = \alpha$

Por la proyección canónica, la imagen de este elipsiode es una esfera $R(\hat{m}^*_{cj})$, de centro $\hat{m}^*_{cj}$, y radio $R_\alpha/\sqrt{n_j}$, siendo $R_\alpha^2 = F_\alpha(n-k)p/(n-k-p+1)$. La proyección canónica de m_j^* pertenece a esta esfera con probabilidad $1-\alpha$.

c. q. d.

Queda resuelta la representación canónica de indivíduos medios en el caso en que dim M = k.

Si dim M = r < k, la estimación de cada indivíduo medio no coincide con la media muestral de la población respectiva. Además, los k indivíduos medios no forman un sistema M-linealmente independiente. La representación canónica en este caso, se expone en § 9.2.

Finalmente, puede suceder que las diferencias entre poblaciones sean debidas, en parte, a la influencia de unas variables concomitantes. Esta influencia debe ser eliminada si deseamos diferenciar las poblaciones respecto a las v.a. $\{Y_i\}$ exclusivamente. Este problema ha sido tratado en §9.3.

7 LA ESTRUCTURA DE LA COVARIABILIDAD ENTRE POBLACIONES

El modelo de efectos aleatorizados del análisis de la varianza de una variable (SCHEFFÈ,1953,.1959), es generalizado para p variables aleatorias, lo que permite representar la estructura probabilística de una infinidad de poblaciones. Se llega, por un camino distinto, a la estructura factorial de las variables canónicas o funciones discriminantes (COOLEY Y LOHNES,1971). Además, la posibilidad de aislar la covariabilidad entre poblaciones, permite efectuar un análisis factorial a las medidas de las variables sobre el conjunto de las infinitas poblaciones posibles.

7.1. UN MODELO PROBABILISTICO

Supongamos que las k poblaciones que constituyen Ω, H_1, $\ldots, H_k$ proceden en realidad de una infinidad de posibles poblaciones, cuyas diferencias obedecen a unas causas aleatorias. Planteemos un modelo basado en las siguientes hipótesis:

1) Ω es reunión de una colección no numerable de poblaciones disjuntas

$$\Omega = \bigcup_{\alpha \in \mathcal{C}} H_\alpha \ , \qquad H_\alpha \cap H_\beta = \emptyset$$

$\mathcal{C}$ es un conjunto no numerable de índices.

2) Sobre el conjunto de poblaciones

$$\mathcal{H} = \{ H_\alpha \ , \ \alpha \in \mathcal{C} \}$$

se conoce una estructura de espacio de probabilidades

$$(\mathcal{H}, \mathcal{A}_H, P_H)$$

que determina las diferencias aleatorias entre las poblaciones.

3) Cada población H_α posee una misma estructura de espacio de probabilidades isométricas a un mismo $(\Omega_o, \mathcal{A}_o, P_o)$

(7.1.1) $$(H_\alpha, \mathcal{A}_\alpha, P_\alpha) \overset{is}{\simeq} (\Omega_o, \mathcal{A}_o, P_o)$$

Es una condición de homogeneidad entre las poblaciones.

Con estas hipótesis vamos a construir un espacio de probabilidades $(\Omega, \mathcal{A}, P)$.

Consideremos los subconjuntos de Ω de la forma,

$$A \subset \Omega \ , \qquad A = \bigcup_{\alpha \in \mathcal{C}'} A_\alpha , \qquad A_\alpha \in H_\alpha , \ \mathcal{C}' \subset \mathcal{C}$$

con la condición de que:

$$H_{\mathscr{C}'} = \{ H_\alpha \ , \ \alpha \in \mathscr{C}' \}$$

sea un suceso de $\mathcal{A}_H$, y que la imagen de cada H_α, sea un mismo suceso de $\mathcal{A}_o$,

(7.1.2) $\qquad is(H_\alpha) = A_o, \forall \alpha \in \mathscr{C}', \quad A_o \in \mathcal{A}_o$

Sobre A definimos la probabilidad

(7.1.3) $\qquad P(A) = P_o(A_o) \cdot P_H(H_{\mathscr{C}'})$

Finalmente, sea $\mathcal{A}$ la σ-álgebra generada por la colección de subconjuntos A con las condiciones anteriores. La probabilidad (7.1.3) se extiende de forma única sobre el álgebra $\mathcal{A}$. Obtenemos así, el espacio $(\Omega,\mathcal{A},P)$, que vamos a relacionar con los espacios $(\Omega_o,\mathcal{A}_o,P_o)$ y $(\mathscr{H}, \mathcal{A}_H,P_H)$.

<u>Teorema 7.1.1</u>

$(\Omega,\mathcal{A},P)$ es isométrico con el producto exterior entre los espacios de probabilidades $(\Omega_o,\mathcal{A}_o,P_o)$ y $(\mathscr{H},\mathcal{A}_H,P_H)$

(7.1.4) $\qquad (\Omega,\mathcal{A},P) \simeq (\Omega_o \times \mathscr{H}, \mathcal{A}_o \otimes \mathcal{A}_H, P_o \times P_H)$

<u>Demostración</u>

La aplicación

$$\Omega \longrightarrow \Omega_o \times \mathscr{H}$$

$$\omega \longrightarrow (\omega_o, H_\alpha) \quad \text{si } \omega \in H_\alpha, \quad \omega_o = is(\omega)$$

es biyectiva.

Además, al suceso de $\mathcal{A}$ de la forma $A = \bigcup_{\alpha \in \mathscr{C}'} A_\alpha$ le podemos hacer corresponder el suceso $(A_o, H_{\mathscr{C}'})$ de $\mathcal{A}_o \times \mathcal{A}_H$. Existe pues, una correspondencia biyectiva entre los sucesos que generan $\mathcal{A}$ y los sucesos que generan $\mathcal{A}_o \otimes \mathcal{A}_H$, luego ambas σ-álgebras son isomorfas.

Como la probabilidad definidas sobre ambas es, por (7.1.3), la misma, deducimos la isometría (7.1.4).

c. q. d.

7.2. DESCOMPOSICION DE LA COVARIABILIDAD

Supongamos ahora que tenemos tres conjuntos de v.a., con distribución normal multivariante:

$(Y_1,\ldots,Y_p)$ v.a. observables sobre $(\Omega,\mathcal{A},P)$

$(\alpha_1,\ldots,\alpha_p)$ v.a. sobre $(\mathcal{H},\mathcal{A}_H,P_H)$ con distribución $N(0,\Sigma_1)$

$(\xi_1,\ldots,\xi_p)$ v.a. sobre $(\Omega_o,\mathcal{A}_o,P_o)$ con distribución $N(0,\Sigma)$

siendo ambas distribuciones no singulares.

La covariabilidad de las v.a. $\{Y_i\}$ estará influida por la covariabilidad entre los otros dos conjuntos de v.a., que por la isometría (7.1.4), pueden considerarse definidas también sobre Ω, poniendo

$$\alpha_i(\omega) = \alpha_i(H_\beta) \quad \text{si } \omega \in H_\beta, \quad i=1,\ldots,p.$$

$$\xi_i(\omega) = \xi_i(\omega_o) \quad \text{si } \omega \in H_\beta, \text{ y } \omega_o = is(\omega)$$

Admitiremos que la influencia tiene un efecto lineal,

$$Y_i = \mu_i + \alpha_i + \xi_i \qquad i=1,\ldots,p \tag{7.2.1}$$

(generalización de la parametrización 6.1.2), representando

$$\mu = (\mu_1,\ldots,\mu_p)'$$

un vector constante de medias generales.

Teorema 7.2.1

Sobre el espacio $(\Omega,\mathcal{A},P)$ se verifica:

1) Los conjuntos de v.a. $(\alpha_1,\ldots,\alpha_p)$ y $(\xi_1,\ldots,\xi_p)$ son estocásticamente independientes.

2) La distribución de las v.a. $(Y_1,\ldots,Y_p)$ es $N(\mu,\Sigma_Y)$ siendo:

$$\Sigma_Y = \Sigma + \Sigma_1 \tag{7.2.2}$$

Demostración:

La independencia de las v.a. $\{\alpha_i\}$ y $\{\xi_i\}$ es consecuencia de estar definidas sobre Ω_o y $\mathcal{H}$, respectivamente, que están asociadas a σ-álgebras independientes en $\Omega_o \times \mathcal{H}$. De aquí se deduce inmediatamente (7.2.2).

c. q. d.

Σ es la matriz de covarianzas entre las variables $\{Y_i\}$ dentro de cada población. Σ_1 es la matriz de covarianzas entre las variables $\{\alpha_i\}$, y refleja la covariabilidad entre las infinitas poblaciones. La interpretación de Σ_1 es distinta de Σ_E (§ 6.3), porque esta segunda matriz introduce una nueva covariabilidad sobre las v.a. $\{Y_i\}$, que explica las diferencias entre las poblaciones.

Indicaremos:

$$\sigma_{ij} = (Y_i, Y_j) = \mathrm{cov}(\xi_i, \xi_j) \quad \text{cov. respecto a } \Sigma$$

$$\sigma^1_{ij} = (Y_i, Y_j)_1 = \mathrm{cov}(\alpha_i, \alpha_j) \quad \text{cov. respecto a } \Sigma_1$$

Con la matriz Σ_1 queda aislada la covariabilidad que existe entre las poblaciones, dependiente de la districión de las v.a. $\{\alpha_i\}$. La estructura de esta covariabilidad la podemos estudiar en dos sentidos:

1) Por la influencia que tiene en las v.a. $\{Y_i\}$, lo que nos dará la aportación de cada variable a la discriminación de las poblaciones. El análisis se realiza entonces sobre $(\Omega, \mathcal{A}, P)$.

2) Considerando la "población" de indivíduos medios de cada población, es decir, $(\mathcal{H}, \mathcal{A}_H, P_H)$, y analizando la estructura factorial de las v.a. $\{\alpha_i\}$, a través de Σ_1. Obtendríamos los factores que son causa de la diferenciación entre poblaciones.

7.3. ESTIMACION DE LAS MATRICES Σ Y Σ_1

Supongamos que tenemos una muestra de k poblaciones de $\mathcal{H}$, siendo $k > p$, y en cada una de ellas, n valores muestrales de una v.a. Y_j,

$$(7.3.1) \qquad (y_{ji1},\ldots,y_{jin}) \qquad \begin{matrix} j=1,\ldots,p \\ i=1,\ldots,k \end{matrix}$$

Los k.n valores muestrales son v.a. independientes sobre el espacio muestral

$$(\mathcal{H}\times\Omega_0\times \overset{(n}{\ldots}\times\Omega_0)\times\overset{(k}{\ldots}\times(\mathcal{H}\times\Omega_0\times\overset{(n}{\ldots}\times\Omega_0)$$

siendo

$$Y_{jir}=\mu_j+\alpha_j+\xi_{jr}$$

Definamos en el espacio $\mathcal{H}\times\Omega_0\times\overset{(n}{\ldots}\times\Omega_0$ las v.a.

$$h_{j.}=\mu_j+\alpha_j+\frac{1}{n}\sum_r \xi_{jr} \qquad j=1,\ldots,p$$

Se verifica:

$$\mathrm{cov}(h_{j.},h_{l.}) = \sigma^1_{jl}+\frac{1}{n}\sigma_{jl}$$

Indicando:

$$\bar{y}_{ji.}=\frac{1}{n}\sum_r y_{jir}, \qquad \bar{y}_{j..}=\frac{1}{k}\sum_i y_{ji.}$$

tendremos entonces, que

$$\bar{y}_{j1.},\bar{y}_{j2.},\ldots,\bar{y}_{jk.}$$

es una muestra, de tamaño k, de la v.a. $h_{j.}$, de donde, el estimador:

(7.3.2) $$S^1_{jl}= \frac{1}{k-1} \sum_i (\bar{y}_{ji.}-\bar{\bar{y}}_{j..})(\bar{y}_{li.}-\bar{\bar{y}}_{l..})$$

nos dará una estimación insesgada de $\sigma^1_{jl}+\frac{1}{n}\sigma_{jl}$.

Por otra parte, para una población H_i determinada, $(y_{ji1},\dots,y_{jin})$ es una muestra de la v.a. sobre Ω_o,

$$Y_{j/H_i} = \mu_j + \alpha_j(H_i) + \varepsilon_j$$

y teniendo en cuenta que $\alpha_j(H_i)$ será un valor constante,

(7.3.3) $$S_{jl}= \frac{1}{k(n-1)} \sum_i \sum_r (y_{jir}-\bar{y}_{ji.})(y_{lir}-\bar{y}_{li.})$$

será una estimación insesgada de σ_{jl}

<u>Teorema 7.3.1</u>

Los elementos de la matriz $\hat{\Sigma}_1 = (\hat{\sigma}^1_{jl})$, siendo

$$\hat{\sigma}^1_{jl} = \frac{1}{k-1} \sum_i (\bar{y}_{ji.}-\bar{\bar{y}}_{j..})(\bar{y}_{li.}-\bar{\bar{y}}_{l..}) -$$

$$- \frac{1}{n.k(n-1)} \sum_i \sum_r (y_{jir}-\bar{y}_{ji.})(y_{lir}-\bar{y}_{li.})$$

son estimadores insesgados de Σ_1.

<u>Demostración:</u>

Hemos visto que:

$$E(S^1_{jl}) = \sigma^1_{jl} + \frac{1}{n}\sigma_{jl} \quad E(S_{jl}) = \sigma_{jl}$$

Tomando entonces,

$$\hat{\sigma}^1_{jl} = S^1_{jl} - \frac{1}{n} S_{jl},$$

$$E(\hat{\sigma}^1_{jl}) = E(S^1_{jl}) - \frac{1}{n}E(S_{jl}) = \sigma^1_{jl} + \frac{1}{n}\sigma_{jl} - \frac{1}{n}\sigma_{jl} =$$

$$= \sigma^1_{jl} \qquad \text{c. q. d.}$$

Finalmente, utilizando las notaciones §6.2, salvo que ahora es:

$$n_1 = n_2 = \ldots = n_k = n, \qquad \sum_i n_i = k.n,$$

se comprueba que:

$$(7.4.3) \qquad \hat{\Sigma}_1 = \frac{1}{n}\left(\frac{1}{k-1}(R_1 - R_0) - \frac{1}{k(n-1)} R_0\right)$$

$$\hat{\Sigma} = \frac{1}{k(n-1)} R_0$$

7.4. COMPONENTES CANONICAS

La descomposición que admiten, según el modelo propuesto, las v.a. observables $\{Y_i\}$,

$$(7.4.1) \qquad Y_i = \mu + \alpha_i + \xi_i$$

sugiere considerar a las v.a. $\{\alpha_i\}$ como factores que influyen es las primeras. Para analizar la relación entre las v.a. $\{Y_i\}$ y las v.a. $\{\alpha_i\}$ que miden las diferencias entre poblaciones, podemos adaptar a esta situación el <u>método factorial canónico</u> de RAO (1954), que consiste en hallar las variables

$$Y = \sum_{h=1}^{p} a_h Y_h, \qquad \alpha = \sum_{h=1}^{p} b_h \alpha_h$$

tales que la correlación canónica (al cuadrado),

$$\upsilon^2 = (\mathrm{corr}(Y, \alpha))^2$$

sea máxima, lo que exige la resolución de la ecuación matricial

$$\det(\Sigma_1 - \nu^2 \Sigma_T) = 0$$

De la relación $\Sigma_T = \Sigma_1 + \Sigma$, y por el lema 7.5.1, esta ecuación equivale a

(7.4.2) $$\det(\Sigma_1 - \lambda\Sigma) = 0$$

Entonces, según el teorema fundamental (capítulo V) se puede interpretar $Y = \sum_h a_h Y_h$ como aquella v.a. tal que la relación de módulos al cuadrado respecto a Σ_1 y a Σ es máxima.

Lamaremos <u>componentes canónicas</u> a las v.a. $\{V_i\}$ tales que

$$\Sigma_1 V_i = \lambda_i \Sigma V_i \qquad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$$

$$\mathrm{var}(V_i)_\Sigma = 1 \qquad i=1,\ldots,p$$

Analizando los coeficientes de

$$V_i = v_{i1} Y_1 + \cdots + v_{ip} Y_p \qquad i=1,\ldots,p$$

podremos detectar las contribuciones de las v.a. $\{Y_j\}$ en la componente $\{V_i\}$. Esta influencia se estudia más fácilmente a través de la matriz de estructura factorial

$$R = V' \Sigma D \qquad R = (r_{ij})$$

siendo $r_{ij} = \mathrm{corr}(V_i, Y_i)$

$$D = \begin{pmatrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_p \end{pmatrix} \qquad \sigma_i = \sqrt{\mathrm{var}(Y_i)}$$

En particular, las correlaciones

$$(r_{11}, r_{12}, \ldots, r_{1p})$$

de las v.a. $\{Y_i\}$ en la primera componente canónica, nos informan de la influencia de cada variablee en la causa principal de la covariabilidad entre poblaciones.

Desde luego, estimaremos Σ_1 y Σ mediante las matrices $\hat{\Sigma}_1$ y $\hat{\Sigma}$ obtenidas en el apartado anterior. Los vectores propios de $\hat{\Sigma}_1$ respecto de $\hat{\Sigma}$ nos darán las componentes canónicas estimadas.

7.5. RELACION ENTRE COMPONENTES Y VARIABLES CANONICAS

La no existencia de covariabilidad entre poblaciones, equivale a $\Sigma_1 = 0$. Dada una muestra de $k \geq p$ poblaciones, el test estadístico para contrastar la anulación de Σ_1 se puede resolver comparando los individuos medios como en § 6.2. Si son iguales, podemos aceptar $\Sigma_1 = 0$, con el correspondiente nivel de significación. No hay entonces, componentes canónicas.

Vamos a relacionar ahora las componentes canónicas (rang $\Sigma_1 = p$), con las variables canónicas definidas en § 6.5, en el caso $k \geq p$ (rang $\Sigma_E = p$).

Lema 7.5.1

Sean A y B dos matrices cuadradas, B no singular, y α, β, γ, δ cuatro números reales no nulos. Si V es vector propio de A respecto B, de valor propio λ, entonces V es también vector propio de $(\alpha A + \beta B)$ respecto a $(\gamma A + \delta B)$, de valor propio $(\alpha\lambda + \beta)/(\gamma\lambda + \delta)$.

En efecto: Se verifica:

$$A\,V = \lambda B\,V$$

de donde

$$(\alpha A+\beta B)V = \alpha A\ V+\beta B\ V = (\alpha\lambda+\beta)B\ V$$

$$(\gamma A+\delta B)V = \gamma A\ V+\delta B\ V = (\gamma\lambda+\delta)B\ V$$

Luego:

$$(\alpha A+\beta B)V = \left(\frac{\alpha\lambda+\beta}{\gamma\lambda+\delta}\right)(\gamma A+\delta B)V$$

c. q. d.

Teorema 7.5.1

Sea $k \geqslant p$, y $n_1=n_2=n\ldots=n_k=n$ el tamaño muestral (común) de cada población. Las componentes canónicas estimadas coinciden, en dirección, con las variables canónicas.

Demostración:

En las mismas notaciones de §6.2, se verifica la siguiente identidad:

$$\sum_{t=1}^{k}\sum_{h=1}^{n}(y_{jth}-\bar{y}_{j..})(y_{ith}-\bar{y}_{i..}) =$$

$$= \sum_{t=1}^{k} n(\bar{y}_{it.}-\bar{y}_{i..})(\bar{y}_{jt.}-\bar{y}_{j..})$$

$$+\sum_{t=1}^{k}\sum_{h=1}^{n}(y_{jth}-\bar{y}_{jt.})(y_{ith}-\bar{y}_{it.})$$

es decir,

$$R_1(i,j) = n\hat{\sigma}_{\varepsilon ij}+R_o(i,j)$$

de donde se deduce:

$$\hat{\Sigma}_\varepsilon = \frac{1}{n}(R_1-R_o) \qquad \hat{\Sigma} = \frac{1}{k(n-1)}R_o$$

Sea ahora V un vector propio de R_1 respecto a R_o. Por el lema anterior, V es vector propio de $\hat{\Sigma}_\varepsilon$ respecto a $\hat{\Sigma}$, y por (7.4.3), lo es también de $\hat{\Sigma}_1$ respecto a $\hat{\Sigma}$.

c. q. d.

Con todo esto concluimos y justificamos plenamente que las variables canónicas y sus correlaciones, con las variables observables $\{Y_i\}$, nos proporcionan la estructura externa de la covariabilidad entre poblaciones.

7.6. ANALISIS FACTORIAL SOBRE POBLACIONES

En § 7.4 y § 7.5 se ha analizado la estructura de Σ_1 respecto a Σ, en relación al espacio $(\Omega, \mathcal{A}, P)$. Pero, en un camino distinto, podemos realizar otro tipo de análisis trabajando exclusivamente en la matriz Σ_1.

Ya hemos visto que con Σ_1 quedaba aislada la covariabilidad de las v.a. $\{\alpha_i\}$, definidas sobre el espacio $(\mathcal{H}, \mathcal{A}_H, P_H)$. Podemos, por tanto, realizar un análisis factorial a las v.a. $\{\alpha_i\}$, a través de Σ_1 (aquí prescindimos de Σ), que nos permitirá detectar unos <u>factores comunes</u> $F_1, \ldots, F_m$, y unos <u>factores únicos</u> $U_1, \ldots, U_p$, de modo que:

$$\alpha_i = a_{i1}F_1 + \ldots + a_{im}F_m + d_iU_i \qquad (i=1,\ldots,p)$$

La matriz del modelo factorial $A = (a_{ij})$ y la matriz diagonal $D = (d_i)$ deberán verificar la identidad fundamental

$$(7.6.1) \qquad \Sigma_1 = A\,A' + D^2 \qquad \text{(Teorema de Thurstone)}$$

(véase HARMAN, 1968; HORTS, 1965; TORRENS-IBERN, 1972; CUADRAS, 1970).

La solución factorial puede obtenerse por el método del factor principal (HOTELLING, 1933), si es que se desean los factores cuya contribución a la variabilidad es máxima. Si la solución debe ajustarse a la estructura simple (THURSTONE, 1947), podrá realizarse entonces una solución del tipo VARIMAX (KAISER, 1958).

En la práctica, partiremos de la estimación $\hat{\Sigma}_1$ (§7.8) Considerada como matriz aleatoria, al proceder de un muestreo con distribución normal multivariante, sigue la distribución de Wishart.

La estimación estadística $\hat{A}$ de la matriz A del modelo factorial (7.6.1) puede obtenerse entonces, por el método de la máxima verosimilitud (LAWLEY Y MAXWELL, 1963; JÖRESKOG, 1967).

Finalmente, de la solución ortogonal (factores incorrelacionados), puede realizarse una rotación oblícua (factores correlacionados), por ejemplo, según el criterio PROMAX (HENDRICSON Y WHITE, 1964), para acercarnos más a la estructura simple de Thurstone.

Cualquiera que sea el método utilizado de factorización, que dependerá de las condiciones experimentales del problema (véase CUADRAS, 1972), la obtención de A, a partir de Σ_1, nos detectará los factores que influyen en las v.a. $\{\alpha_i\}$ en el espacio $(\mathcal{H}, \mathcal{A}_H, P_H)$. Puesto que las variables directamente observables son $(Y_1, \ldots, Y_p)$, puede interpretarse como la factorización de los valores medios de $\{Y_i\}$ sobre el conjunto de las infinitas poblaciones.

La factorización de Σ_1 puede aplicarse a la situación general siguiente: supongamos que disponemos de una "población" de grupos, y que deseamos estudiar la estructura factorial de unas v.a. $\{Y_i\}$, a partir de su variabilidad y covariabilidad sobre el indivíduo medio representante de cada grupo, en la práctica desconocido. Si la matriz Σ es común en cada grupo, y se dispone de una muestra del mismo tamaño para cada grupo, podremos estimar Σ_1 y realizar el análisis factorial de las v.a. $\{Y_i\}$ sobre el conjunto de individuos medios.

8 ANALISIS CANONICO DE FUNCIONES ESTIMABLES

Se generaliza el análisis canónico de poblaciones a un sistema de funciones estimables linealmente independientes.

Adaptando a funciones estimables la distancia de Mahalanobis, se demuestra en relación a esta distancia las propiedades de la proyección canónica propuesta. Además, basándonos en algunos teoremas de análisis multivariante (ANDERSON,1958), se construye, especificado un coeficiente de confianza, una región confidencial exacta para cada función estimable.

Ni en la reciente publicación de RAO (1972) sobre los últimos avances del análisis multivariante, ni en la bibliografía consultada hemos encontrado una generalización, para funciones estimables, del análisis canónico de poblaciones. No obstante, hacemos constar que DEMPSTER (1969), analiza una representación de poblaciones tomando, como ejes, distintos pares de funciones discriminantes, con lo cual realiza la diferenciación según conceptos diversos. SAUPE (1965), y COOLEY y LOHNES (1971), con el nombre de análisis factorial discriminante, analizan la estructura factorial de sistemas de funciones discriminantes. Tales técnicas, aplicadas sólo a diseños particulares, no pueden considerarse una generalización del análisis canónico en el sentido de esta memoria.

Otras referencias consultadas: ANDREWS(1971), GOODMAN(1972), JAMES (1954), KENDALL(1957), MAHALANOBIS(1936), MATUSITA(1966), RAO(1952,1965).

8.1. LA DISTANCIA DE MAHALANOBIS

La métrica en E induce en E* una métrica, que define el producto escalar de matriz asociada Σ^{-1} en la base dual $\{Y_i^*\}$ (§ 2.2).

La distancia (al cuadrado) entre dos indivíduos medios será:

$$(8.1.1) \qquad D^2(m_i^*, m_j^*) = \| m_i^* - m_j^* \|^2_{\Sigma^{-1}} = (\xi_i - \xi_j)' \Sigma^{-1} (\xi_i - \xi_j)$$

siendo ξ_i el vector de componentes de m_i^* en E^*.

Se conoce (8.1.1) como la <u>distancia de Mahalanobis</u> entre las poblaciones H_i y H_j.

Esta distancia se puede restringir a un subespacio $F \subset E$, dim F = d.

Sea $(V_1, \ldots, V_d)$ una base ortonormal de F. Indiquemos por V la matriz cuyos vectores columna son las componentes de los vectores $\{V_i\}$ en la base $\{Y_i\}$ de E. Tendremos:

$$E^* = F^* \oplus (F^*)^\perp \quad \text{siendo} \quad F^* = \{\omega^* | \omega^*(V) = 0, \forall\, V \in F\}$$

$$\omega^* = \omega^*_1 + \omega^*_2 \qquad \omega^*_1 \in F^*, \qquad \omega^*_2 \in (F^*)^\perp$$

La restricción del producto escalar a (F^*), ortocomplementario de F^*, será:

$$(\omega^*, \omega'^*)_{(F^*)} = (\omega^*_1, \omega'^*_1) = w'(V \cdot V')\, \bar{w}$$

y la distancia retringida a (F^*) será:

$$(8.1.2) \qquad D^2_{(F^*)}(m_i^*, m_j^*) = \| m_i^* - m_j^* \|^2_{(F^*)} = (\xi_i - \xi_j) V V' (\xi_i - \xi_j)$$

En particular, cuando F es el espacio canónico EC_d, generado por las d primeras v.a. canónicas, (8.1.2) nos dará la distancia entre las proyecciones canónicas de los individuos medios en el dual canónico EC^*_d.

El siguiente teorema confirma que el poder descriminante de las v.a. canónicas es óptimo para diferenciar los individuos medios en un subespacio de dimensión d nv = mín (p,k-1). (RAO, 1952).

<u>Teorema 8.1.1</u>

Entre todos los subespacios de E, de dimensión d, el espacio canónico EC_d, maximiza la suma de distancias restringidas (a.c.) entre las poblaciones, ponderadas por sus respectivas probabilidades,

$$\sum_{i,j} D^2_{EC^*}(m_i^*,m_j^*)p_i p_j \geq \sum_{i,j} D^2_{F^*}(m_i^*,m_j^*)p_i p_j$$

$$\forall\ F\subset E \quad \dim F=d$$

<u>Demostración:</u>

Sea $F\subset E$ y $(V_1,\ldots,V_d)$ una base de F. De (8.1.2) tenemos:

$$\sum_{i,j} D^2_{F^*}(m_i^*,m_j^*)p_i p_j = \sum_{i,j} (\xi_i-\xi_j)V.V'(\xi_i-\xi_j)p_i p_j$$

Se puede demostrar sin dificultad la siguiente identidad:

$$\sum_{i,j}(\xi_i-\xi_j)VV'(\xi_i-\xi_j)p_i p_j = 2\sum_i (\xi_i-\bar\xi)VV'(\xi_i-\bar\xi)p_i$$

siendo $\bar\xi$ las componentes de $m^*=\sum_i p_i m_i^*$ en E^*.

Por otra parte se tiene,

$$\sum_i (\xi_i-\bar\xi)(\xi_i-\bar\xi)'p_i = \Sigma_E \quad \text{de donde:}$$

$$\sum_i (\xi_i-\bar\xi)VV'(\xi_i-\bar\xi)p_i = \|V_1\|^2_{\Sigma_E} + \ldots + \|V_d\|^2_{\Sigma_E}$$

Según el teorema fundamental, para las d primeras v.a. canónicas, con norma 1 respecto a Σ, se alcanza la norma máxima respecto a Σ_ε,

$$\| V_1 \|^2_{\Sigma_\varepsilon} = \lambda_1 = \text{máx}, \ldots, \| V_d \|^2_{\Sigma_\varepsilon} = \lambda_d = \text{máx},$$

siendo $\lambda_1, \ldots, \lambda_d$ los d primeros valores propios de Σ_ε respecto a Σ.

c. q. d.

8.2. COVARIABILIDAD ENTRE FUNCIONES ESTIMABLES

Sean $\psi^*_1, \ldots, \psi^*_s$ un sistema M-linealmente independiente (§ 3.2) de s funciones estimables. Una condición necesaria de independencia es haber rechazado, con un nivel de significación α, la hipótesis nula

$$H_o \qquad \psi^*_1 = \ldots = \psi^*_s$$

Supongamos dim M = $r \leq k$, y sea

$$\psi_i = \hat{d}_{i1} m_1 + \ldots + \hat{d}_{ik} m_k \qquad i=1,\ldots,s$$

la expresión Δ-óptima de cada ψ_i, para un cierto $m \in \mathfrak{m}$. Las funciones estimables $(\psi_1, \ldots, \psi_s)$, pueden considerarse v.a. aleatorias del espacio $\mathfrak{F}_M$ (§ 6.3), y entonces,

$$\| \psi_i \|^2_M = \hat{d}^2_{i1} p_1 + \ldots + \hat{d}^2_{ik} p_k \qquad i=1,\ldots,k$$

Con el fin de introducir una probabilidad para ψ_i, que generalice la probabilidad P_M en $\mathcal{A}_M$, definamos primero

$$P(\psi_i) = \hat{d}^2_{i1} p_1 + \ldots + \hat{d}^2_{ik} p_k$$

y a continuación la probabilidad

$$(8.2.1) \qquad P_\psi(\psi_i) = P(\psi_i) \Big/ \sum_{h=1}^{s} P(\psi_h) \qquad i=1,\ldots,s.$$

Tomando $\Omega_\psi = \{\psi_1,\dots,\psi_s\}$, $\mathcal{A}_\psi = \mathcal{P}(\Omega_\psi)$, obtenemos el espacio de probabilidades $(\Omega_\psi, \mathcal{A}_\psi, P_\psi)$. La probabilidad sobre cada ψ_i es proporcional a su módulo (al cuadrado) en $\mathcal{F}_M$.

Procediendo como en §6.3, sea ahora $\mathcal{F}_\psi$ el espacio vectorial de dimensión s, generado por $(\psi_1,\dots,\psi_s)$, considerados como átomos de $\mathcal{A}_\psi$. En $\mathcal{F}_\psi$ introducimos un producto escalar de matriz asociada

$$M_\psi = \begin{pmatrix} P_{\psi_1} & & \\ & \ddots & \\ & & P_{\psi_s} \end{pmatrix}$$

Sea entonces la aplicación:

$$f_\psi : E \longrightarrow \mathcal{F}_\psi$$

$$f_\psi(Y) = \sum_{i=1}^{s} ((\psi_i^* - \bar{\psi}^*)(Y))\psi_i$$

siendo $\bar{\psi}^* = \sum_{i=1}^{s} P_{\psi_i} \psi_i^*$

El producto escalar en $\mathcal{F}_\psi$, de la imagen de dos v.a. $Y, Y' \in E$ coincidirá con la covarianza en Ω_ψ,

$$f_\psi(Y) . f_\psi(Y') = \operatorname{cov}_\psi(Y,Y') =$$

$$= \sum_{h=1}^{s} ((\psi_h^* - \bar{\psi}^*)(Y))((\psi_h^* - \bar{\psi}^*)(Y'))P_{\psi_h}$$

Poniendo entonces

$$(8.2.2) \qquad Y \circ Y' = \operatorname{cov}_\psi(Y,Y') \qquad \forall\, Y, Y' \in E$$

habremos definido un nuevo producto escalar en E, que reflejará la covariabilidad entre las s funciones estimables.

Si $\psi^*_i = b_{i1}Y^*_1 + \ldots + b_{ip}Y^*_p$ $\quad i=1,\ldots,p$

entonces $\bar{\psi}^* = \bar{b}_1 Y^*_1 + \ldots + \bar{b}_p Y^*_p$

siendo $\bar{b}_i = \sum_{h=1}^{s} b_{hi} p_{\psi_h}$

Indicando

$$B_\psi = \begin{pmatrix} b_{11}-\bar{b}_1, \ldots\ldots, b_{p1}-\bar{b}_p \\ \ldots\ldots\ldots \\ b_{s1}-\bar{b}_1, \ldots\ldots, b_{sp}-\bar{b}_p \end{pmatrix}$$

la matriz asociada Σ_ψ, del producto escalar (8.2.2) en la base $\{Y_i\}$ será:

(8.2.3) $\quad \Sigma_\psi = B'_\psi M_\psi B_\psi$

Naturalmente, si la hipótesis H_o resulta ser cierta, entonces

$$\psi^*_1 = \ldots = \psi^*_s = \bar{\psi}^*$$

y la f_ψ aplica E en el cero de $\mathcal{F}_\psi$.

Por el contrario, si tales funciones estimables son M-linealmente independientes, el rango de B_ψ por columnas no puede superar a p, y por filas no puede superar a (s-1). Queda demostrado entonces:

Teorema 8.2.1

1) Si la hipótesis nula H_o es cierta, entonces $\Sigma_\psi = 0$.

2) Si las funciones lineales estimables $\psi^*_1, \ldots, \psi^*_s$ son M-linealmente independientes, entonces,

$$\text{rang}\ (\Sigma_\psi) = \text{mín}\ (p, s-1)$$

8.3. ESTIMACION DE Σ_ψ

Sea $\psi_i^* = \hat{d}_{i1} m_1^* + \dots + \hat{d}_{ik} m_k^*$ $\quad (i=1,\dots,s)$

la expresión A-óptima de cada ψ_i^*.

La estimación insesgada de ψ_i^*, con dispersión **mínima**, es

$$\hat{\psi}_i^* = \hat{b}_{i1} Y_1^* + \dots + \hat{b}_{ip} Y_p^* \qquad (i=1,\dots,s)$$

siendo

$$b_{ij} = \psi_i^*(Y_j) = \sum_{h=1}^{k} \hat{d}_{ih} \bar{y}_{jh}.$$

La estimación de $\bar{\psi}^*$ será:

$$\hat{\bar{\psi}}^* = \hat{\bar{b}}_1 \; Y_1^* + \dots + \hat{\bar{b}}_p \; Y_p^*$$

siendo

$$\hat{\bar{b}}_j = \sum_{i=1}^{s} \left(\sum_{h=1}^{k} \hat{d}_{ih} \bar{y}_{jh} \right) \hat{P}_{\psi_i} \qquad (j=1,\dots,p)$$

en donde $\hat{P}_{\psi_i}$, estimación de P_{ψ_i}, se calcula de (8.2.1) sustituyendo cada p_h por n_h/n $\quad (h=1,\dots,k)$.

Indicando **entonces**,

$$\hat{M}_\psi = \begin{pmatrix} \hat{P}_{\psi_1} & & \\ & \ddots & \\ & & \hat{P}_{\psi_k} \end{pmatrix}$$

y $\hat{B}_\psi$ la matriz $\hat{B}_\psi = (b_{ij} - \bar{b}_j)$, una buena estimación de Σ_ψ, según (8.2.3) será:

$$(8.3.1) \qquad \hat{\Sigma}_\psi = \hat{B}_\psi' \; \hat{M}_\psi \; \hat{B}_\psi$$