2 9 20

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

TEORIA ASINTOTICA PARA PRUEBAS DE HIPOTESIS ESTADISTICAS PARA VALORES PROPIOS DE UNA MATRIZ DE COVARIANZAS EN EL CASO NORMAL

T E S I S

Que para obtener el título de:

A C T U A R I O

p r e s e n t a:

JORGE LEONCIO GUTIERREZ VALDES

México, D. F. 1989

I N D I C E

Pag.

Prefacio

Capítulo 1 Matriz de Covarianzas (Σ)

1.1 Introducción 1

1.2 Estimador de Σ 3

1.3 Propiedades de S 6

Capítulo 2 Distribución asintótica de los eigenvalores de S

2.1 Introducción 13

2.2 Distribución asintótica de S^* 14

2.3 Distribución asintótica de D (Caso $\Lambda = \lambda$ I) 16

2.4 Distribución asintótica de D (Caso General) 20

Capítulo 3 Estimadores máximo verosímiles de las raices características de Σ

3.1 Introducción 27

3.2 Estimador máximo verosímil de Σ 27

Capítulo 4 Contrastes de hipótesis

4.1 Introducción 34

4.2 Hipótesis de que q_1 eigenvalores de Σ son iguales 34

4.3 Hipótesis de que q_2 eigenvalores de Σ son iguales a λ 41

4.4 Distribución Cocientes de Verosimilitud .. 44

Capítulo 5 Intervalos de confianza..........50

Conclusiones..........53

Apéndice..........54

Bibliografía..........59

P R E F A C I O

El presente trabajo tiene por objeto analizar los principales resultados para muestras grandes concernientes a los eigenvalores de una matriz muestral de covarianzas S cuando la muestra proviene de una población normal p-variada.

Un caso importante para la aplicación de este tipo de estudios lo es el Análisis en Componentes Principales (ACP). El ACP es una técnica estadística que determina combinaciones lineales de variables aleatorias que tienen propiedades especiales en términos de varianza.

En el ACP se tienen N observaciones, sobre p variables aleatorias, que son agrupadas en una matriz de datos $X^T_{p \times N} = [X_1, \ldots, X_N]$; lo que interesa al investigador es encontrar alguna relación lineal entre las p variables que se estudian de tal manera de reducir el número de variables. Para esto se determinan combinaciones lineales de las variables originales de acuerdo a lo siguiente :

$$Y = X C$$

donde :

$$S = C D C^T$$

$$S = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})(X_i - \bar{X})^T$$

$$D = \text{Matriz diagonal con } d_i \geq d_j \text{ si } i < j$$

$$I = C C^T$$

Lo anterior es equivalente a buscar las coordenadas de los puntos $\{ X_i \}$ con respecto a ejes ortonormales que maximisen la variación de las proyecciones de los puntos $\{ X_i \}$ sobre cada eje.

En el ACP se pretende sustituir a la matriz de datos X por las primeras k $(k \leq p)$ columnas de la matriz Y, obteniendo como ganancia ortogonalidad y reducción en el número de variables.

Los criterios para determinar k (el número de variables a conservar) estan basados en el análisis específico que se esta realizando, algunos ejemplos son los siguientes :

a) Suponiendo que parte de la variabilidad del conjunto de datos esta dada por un error común en la medición de cada una de las variables, esto implicaría que el vector aleatorio del cual proceden las observaciones se puede descomponer como $X = X^{*}_{2} + e\, I$, donde e es una variable aleatoria con media 0 y varianza σ^2; en consecuencia Σ, la matriz de covarianzas de X, se puede descomponer como $\Sigma = \Psi + \sigma^2 I$, donde Ψ es de rango k $(k \leq p)$.

Para este caso se busca un subespacio de dimensión k que no contenga la variabilidad debida al error. El criterio usual es que de una grafica del $\ln(d_i)$ contra i, se toma k como el punto a partir del cual la gráfica es aproximadamente recta (ver Castell-1966).

b) En otros casos el investigador esta dispuesto a perder una cantidad no significativa de la variabilidad de las variables originales a consecuencia de la ventaja que le implica reducir el número de variables. Para esto se determina la máxima k para la cual $\sum_{i=k+1}^{p} d_i \leq \gamma$, donde γ es la cantidad de la variabilidad total que se esta dispuesto a perder.

Como vemos en los ejemplos anteriores, se aplican simples reglas empíricas, no existe un soporte probabilístico para decidir que variables se deben conservar o desechar; lo anterior es debido a la complejidad extrema de los cálculos involucrados. En este trabajo este problema se resuelve para el caso de muestras grandes de un vector aleatorio con una distribución normal p-variada.

Si bien existen algunos estudios sobre el tema, las referencias estan esparcidas y forman muchas veces parte de análisis particulares. La importancia de este trabajo reside en conjuntar, uniformizar y clarificar los resultados más importantes sobre el tema.

Con esto se pretende dar un mayor soporte al uso de las diferentes técnicas de análisis multivariado que involucran el uso de los eigenvalores en el caso de una población normal multivariada.

En el capítulo I, se establecen antecedentes y resultados básicos que se usan en capítulos posteriores.

Es necesario para la comprensión del material que aquí se presenta, conocimientos de algebra matricial y estadística matemática.

1. MATRIZ DE COVARIANZAS (Σ)

1.1 INTRODUCCION

Se dirá que un vector aleatorio

$$X^T = (x_1,\dots,x_p) \qquad x_i \in \mathbb{R}$$

tiene distribución normal multivariada, si la función de densidad de X esta dada por

$$f_X(X;\mu,\Sigma) = |2\pi\ \Sigma|^{-1/2} \exp\left\{-\frac{1}{2}\ (X-\mu)^T\ \Sigma^{-1}\ (X-\mu)\right\} \qquad \forall\ \ X \in \mathbb{R}^p$$

donde $\mu \in \mathbb{R}^p$ y Σ_{pxp} es positiva definida.

En este trabajo se consideran observaciones $X_1, X_2,\dots,X_N$ ($X_i \in \mathbb{R}^p$) de un vector aleatorio con distribución normal multivariada, en particular, para el caso en que las N observaciones sobre X son independientes. Las observaciones $X_1, X_2,\dots,X_N$ son independientes si su función de distribución, $F(X_1,..,X_N)$ puede descomponerse como el producto de las funciones de distribución marginales de cada vector :

$$F(X_1,\dots,X_N) = F(X_1)\ F(X_2)\dots F(X_N)$$

Definiremos al vector muestral de medias como

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

y a la matriz muestral de covarianzas como

$$S = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X}) (X_i - \bar{X})^T$$

Se considerará que una sucesión de funciones de distribución $\{ F_N \}$ (X_N las variables aleatorias correspondientes) converge a la distribición F, si

$$\lim_{N \to \infty} F_N = F$$

para todo intervalo acotado de continuidad de F; para este caso, nos referiremos a convergencia en distribución , y lo denotaremos como

$$X_N \overset{N \to \infty}{\sim} F$$

Las referencias a resultados que aparecen numerados, en los que se antepone una A al número, aparecen en el apéndice de este trabajo.

1.2 ESTIMADOR DE Σ

El propósito de esta sección es el de determinar los estimadores máximo verosímiles de los parámetros de la distribución $N_p(\mu,\Sigma)$ correspondiente a un conjunto de observaciones independientes $X_1,\dots,X_N$ donde $X_i \sim N_p(\mu,\Sigma)$ $i=1,\dots,N$.

Teorema 1.1 , Sea $X \sim N_p(\mu,\Sigma)$ y $X_1,\dots,X_N$ una muestra aleatoria de la distribución. Los estimadores máximo verosímiles de μ y Σ estan dados por :

$$\hat{\mu} = \frac{1}{N} \sum_{i}^{N} X_i = \bar{X}$$

$$\hat{\Sigma} = \frac{1}{N} \sum_{i}^{N} (X_i - \bar{X})(X_i-\bar{X})^T = \frac{N-1}{N} S \quad .$$

Demostración . La función de verosimilitud de los parámetros esta dada por la siguiente ecuación :

$$L(\mu,\Sigma) = |2\pi\Sigma|^{-N/2} \exp\{- \frac{1}{2} \sum_{i}^{N}(X_i-\mu)^T \Sigma^{-1} (X_i-\mu) \}$$

Con objeto de simplificar el álgebra para maximizar la función de verosimilitud, puede obtenerse una transformación monótona de L.

En este caso la función logaritmo natural resulta apropiada, obteniéndose

$$\ln(L) = -\frac{N}{2} \ln|2\pi\Sigma| - \frac{1}{2} \sum_{i}^{N} (X_i-\mu)^T \Sigma^{-1} (X_i-\mu)$$

$$= -\frac{N}{2} \ln|2\pi\Sigma| - \frac{1}{2} \operatorname{tr} \Sigma^{-1} \sum_{i=1}^{N} (X_i-\mu)(X_i-\mu)^T \qquad (1)$$

Del Lema A-1 (Apéndice) sabemos que

$$\sum_{i=1}^{N} (X_i-\mu)\ (X_i-\mu)^T = (N-1)\ S + N\ (\bar{X}-\mu)\ (\bar{X}-\mu)^T$$

y sustituyendo esta ecuación en (1), se obtiene

$$\ln (L) = - \frac{N}{2} \ln |2\pi\ \Sigma| - \frac{1}{2} \operatorname{Tr} \Sigma^{-1} (N-1)\ S - \frac{N}{2} (\bar{X}-\mu)^T \Sigma^{-1} (\bar{X}-\mu)$$

Debido a que Σ^{-1} es positiva definida, $(X-\mu)^T \Sigma^{-1}(X-\mu) \geq 0$ $\forall\ \mu$. en consecuencia,

$$\hat{\mu} = \bar{X}$$

Sólo resta maximizar el primero y segundo términos de la ecuación anterior, que son sólo funciones de Σ, para lo cual haremos uso del Teorema A-2. Haciendo $C = \Sigma^{-1}$obtenemos

$$\begin{aligned} & - \frac{N}{2} \ln |2\pi\ \Sigma| - \frac{1}{2} \operatorname{Tr} \Sigma^{-1}(N-1)\ S \\ = & - \frac{Np}{2} \ln (2\pi) + \frac{N}{2} \ln |C| - \frac{1}{2} \operatorname{Tr} C\ (N-1)S \\ = & - \frac{Np}{2} \ln (2\pi) + f(C) \end{aligned}$$

De acuerdo al Teorema A-2, f(C) es máximo en $C = \frac{N}{N-1} S^{-1}$ es decir en $\hat{\Sigma} = \frac{N}{N-1} S$, donde toma el valor

$$f(\hat{\Sigma}) = \frac{Np}{2} \ln (N) - \frac{N}{2} \ln |(N-1)S| - \frac{Np}{2}$$

de donde se sigue que

$$L(\hat{\mu},\hat{\Sigma}) = \frac{(2\pi)^{-Np/2} \exp (-Np/2)}{(\frac{N-1}{N})^{Np/2}\ |S|^{N/2}}$$

∎

El estimador máximo verosímil $\hat{\Sigma} = \frac{N-1}{N} S$ no resulta insesgado, ya que

$$E(\widehat{\Sigma}) = E\left(\frac{1}{N} \sum_{i}^{N} (X_i-\mu+\mu-\bar{X}) \quad (X_i-\mu+\mu-\bar{X})^T \right)$$

$$= \frac{1}{N} \sum_{i}^{N} E\left((X_i-\mu)(X_i-\mu)^T \right) - E\left((\bar{X}-\mu)(\bar{X}-\mu)^T \right)$$

$$= \Sigma - \frac{1}{N} \Sigma = \frac{N-1}{N} \Sigma$$

De manera natural consideramos el estimador

$$S = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X}) (X_i - \bar{X})^T$$

que resulta insesgado.

1.3 PROPIEDADES DE S

En esta sección se definen y examinan algunas propiedades de la matriz S (matriz de varianzas y covarianzas muestral).

Consideremos como antes una población normal p-variada, y denotemos por $X^T = [X_1, X_2, \ldots, X_N]_{p\times N}$ donde $X_i \sim N_p(\mu, \Sigma)$ $i=1,\ldots,N$.

Teorema 1.2.1. Sea $X \sim N_p(\mu,\Sigma)$ y $X_1,\ldots,X_N$ una muestra aleatoria de la distribución. La matriz $(N-1)S = \sum_i^N (X_i-\bar{X})(X_i-\bar{X})^T$ tiene una distribución Wishart con parámetros $(N-1)$ y Σ.

Demostración. Sea D una matriz cuadrada de orden N, tal que sus columnas sean ortonormales, es decir,

$$d_i^T d_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Lo anterior implica que $D^TD = I$ ($\Rightarrow$ $D D^T = I$), de donde

$$X^TX = X^T D D^T X = X^T \left[\sum_i^N d_i d_i^T \right] X = \sum_i^N X^T d_i d_i^T X = \sum_i^N V_i V_i^T$$

donde $V_i = X^T d_i$

Del Teorema A-4, se obtiene que $V_i \sim N_p(M^T d_i, \Sigma)$ donde $M^T = [\mu, \mu, \ldots, \mu]$.

Consideremos el caso especial de la matriz D ortonormal, cuya primera columna sea $d_1^T = \frac{1}{\sqrt{N}}(1,1,\ldots,1)$, tenemos entonces que $D = [d_1, D_2]$, por lo tanto

$$X^TX = X^T DD^TX = X^T d_1 d_1^T X + X^T D_2 D_2^T X \qquad (2)$$

Puesto que $X^T d_1 = \frac{1}{\sqrt{N}} \sum_i^N X_i = \sqrt{N} \; \bar{X}$,

$$X^T d_1 d_1^T X = N \bar{X} \bar{X}^T$$; despejando de (2), obtenemos

$$X^T D_2 D_2^T X = X^T X - N \bar{X} \bar{X}^T = \sum_i^N (X_i-\bar{X})^T (X_i-\bar{X}) = (N-1) S$$

$$= \sum_{i=2}^N V_i V_i^T$$

Para $i > 1$ obtenemos

$$E(V_i) = E(X^T d_i) = M^T d_i = [\mu,\mu,\ldots,\mu] d_i = \mu \sqrt{N} d_1^T d_i$$
$$= 0$$

Lo anterior debido a la ortogonalidad de la matriz D. Por lo tanto $V_i \sim N_p (0,\Sigma)$ para $i > 1$.

De acuerdo a la Definición A-5, y tomando en cuenta que

$$(N-1) S = \sum_{i=2}^N V_i V_i^T$$, obtenemos

$$(N-1) S \sim W (N-1 , \Sigma)$$

■

Corolario 1.2.1 . La distribución de la matriz S, esta dada por

$$S \sim W(N-1 , \frac{1}{N-1} \Sigma)$$

Demostración. Considerando el mismo análisis del teorema anterior pero con los vectores $\frac{1}{\sqrt{N-1}} X_i$, se obtiene el resultado.

∎

Denotemos por $V^T = [V_1, \ldots, V_{N-1}]_{p \times (N-1)}$, donde $V_i \sim N(0, \frac{1}{N-1}\Sigma)$, en consecuencia por el corolario anterior, $S \sim V V^T$.

Las matrices con distribución Wishart poseen una importante propiedad que se establece en el siguiente teorema.

Teorema 1.2.2. . Sea $S \sim W(N-1, \frac{1}{N-1}\Sigma)$. Si $N-1 \geq p$, entonces S es positiva definida con probabilidad 1.

Demostración. Como $a^T S a = 0$ con $a = 0$, se verifica únicamente cuando S es una matriz de rango completo, se sigue que cuando $N-1 \geq p$

$$P (a^T S a > 0 \; \forall a \neq \bar{o}) = P (\text{Rango } S = p)$$

$$= P (\text{Rango } V = p)$$

$$= 1 - P ({}^{\text{Al menos una columna de } V \text{ es}}_{\text{dependiente de las otras}})$$

$$= 1 - P (\bigcup_{j=1}^{p} \{ \sum_{i}^{p} w_i V_{(i)} = 0 \text{ con } w_j = 1 \})$$

$$\geq 1 - \sum_{j}^{p} P (\sum_{i}^{p} w_i V_{(i)} = 0 \text{ con } w_j = 1)$$

$$= 1$$

∎

Teorema 1.2.3 . La matriz de varianzas y covarianzas de de la matriz $V_a V_a^T$, donde $V_a \sim N_p(0,\Sigma)$ esta dada por

$$\mathrm{Cov}\,(V_{ia} V_{ja}\ ,\ V_{ka} V_{la}) = \sigma_{ik}\,\sigma_{jl} + \sigma_{il}\,\sigma_{jk}$$

donde σ_{ij} es el elemento que se encuentra en el renglón i, columna j de la matriz Σ.

Demostración.

$$E\{ (V_{ia} V_{ja} - \sigma_{ij}) (V_{ka} V_{la} - \sigma_{kl}) \}$$

$$= E\{ V_{ia} V_{ja} V_{ka} V_{la} \} - \sigma_{kl}\,\sigma_{ij}$$

Recurriendo al Teorema A-3, la función característica del - vector V_a esta dada por

$$\phi(t) = \exp\left(- \tfrac{1}{2}\, t^T \Sigma\, t \right)$$

y recurriendo al siguiente resultado

$$E\{ V_{ia} V_{ja} V_{ka} V_{la} \} = \frac{1}{i^4} \left. \frac{\partial^4 \phi(t)}{\partial t_i\, \partial t_j\, \partial t_k\, \partial t_l} \right|_{t=0}$$

obtenemos las siguientes derivadas parciales:

$$\frac{\partial\,\phi(t)}{\partial t_l} = -\frac{1}{2}\left[\sum_{s=l}^{p}\sigma_{ls}t_s + \sum_{s=1}^{p}\sigma_{sl}\ t_s\ (\ 1 + \delta_{sl}\ )\right]\ \phi(t)$$

donde δ_{sl} es la delta de Kronecker

$$\frac{\partial^2\phi(t)}{\partial t_k\,\partial t_l} = -\frac{1}{2}\left[\ \sigma_{lk} + \sigma_{kl}\ \right]\phi(t)$$

$$-\frac{1}{2}\left[\sum_{s=l}^{p}\sigma_{ls}\ t_s + \sum_{s=1}^{p}\sigma_{sl}\ t_s\ (\ 1 + \delta_{sl}\ )\right]\ \frac{\partial\phi(t)}{\partial t_k}$$

$$\frac{\partial^3\phi(t)}{\partial t_j\,\partial t_k\,\partial t_l} = -\sigma_{lk}\ \frac{\partial\phi(t)}{\partial t_j} - \frac{1}{2}\left[\ \sigma_{lj} + \sigma_{jl}\ \right]\ \frac{\partial\phi(t)}{\partial t_k}$$

$$-\frac{1}{2}\left[\sum_{s=l}^{p}\sigma_{ls}\ t_s + \sum_{s=1}^{p}\sigma_{sl}\ t_s\ (1 + \delta_{sl})\right]\ \frac{\partial^2\phi(t)}{\partial t_j\,\partial t_k}$$

(Observese que $\left.\frac{\partial^3\phi(t)}{\partial t_j\,\partial t_k\,\partial t_l}\right|_{t=0} = 0$)

$$\frac{\partial^4\phi(t)}{\partial t_i\,\partial t_j\,\partial t_k\,\partial t_l} = -\sigma_{lk}\ \frac{\partial^2\phi(t)}{\partial t_i\,\partial t_j} - \sigma_{lj}\ \frac{\partial^2\phi(t)}{\partial t_i\,\partial t_k}$$

$$-\sigma_{li}\ \frac{\partial^2\phi(t)}{\partial t_j\,\partial t_k} + \frac{\partial\phi(t)}{\partial t_l}\left[\ \frac{1}{\phi(t)}\ \right]\ \frac{\partial^3\phi(t)}{\partial t_i\,\partial t_j\,\partial t_k}$$

Obtenemos entonces que

$$E(\, V_{ia}\, V_{ja}\, V_{ka}\, V_{la}\,) = \frac{1}{i^4} \left. \frac{\partial^4 \phi(t)}{\partial t_i\, \partial t_j\, \partial t_k\, \partial t_l} \right|_{t=0}$$

$$= \sigma_{lk} \left. \frac{\partial^2 \phi(t)}{i^2\, \partial t_i\, \partial t_j} \right|_{t=0} + \sigma_{lj} \left. \frac{\partial^2 \phi(t)}{i^2\, \partial t_i\, \partial t_k} \right|_{t=0}$$

$$+ \sigma_{li} \left. \frac{\partial^2 \phi(t)}{i^2\, \partial t_j\, \partial t_k} \right|_{t=0} - \frac{1}{i^4} \frac{\partial \phi(t)}{\partial t_l} \left[\frac{1}{\phi(t)} \right] \left. \frac{\partial^3 \phi(t)}{\partial t_i\, \partial t_j\, \partial t_k} \right|_{t=0}$$

$$= \sigma_{lk}\ E(\, V_{ia}\, V_{ja}\,) + \sigma_{lj}\ E(\, V_{ia}\, V_{ka}\,) + \sigma_{li}\ E(\, V_{ja}\, V_{ka}\,)$$

$$= \sigma_{lk}\, \sigma_{ij} + \sigma_{lj}\, \sigma_{ij} + \sigma_{li}\, \sigma_{jk}$$

Sustituyendo en la expresión inicial obtenemos

$$\mathrm{Cov}\, (V_{ia}\, V_{ja}\, ,\, V_{ka}\, V_{la}) = \sigma_{ik}\, \sigma_{jl} + \sigma_{il}\, \sigma_{jk}$$

Corolario 1.2.3. . La matriz de varianzas y covarianzas de la matriz S, esta dada por

$$\mathrm{Cov}(s_{ij} , s_{kl}) = \frac{\sigma_{ik}\,\sigma_{jl} + \sigma_{il}\,\sigma_{jk}}{N - 1}$$

Demostración. Al tomar en cuenta que los vectores V_i, V_j con $i \neq j$, son independientes, el resultado es inmediato.

∎

2. DISTRIBUCION ASINTOTICA DE LOS EIGENVALORES DE S

2.1. INTRODUCCION

En esta sección se presenta el desarrollo de la distribución asintótica de las raíces características de la matriz de covarianzas muestrales S, que corresponde a N observaciones de una población normal multivariada.

Debido a que Σ es una matriz simétrica, esta puede ser descompuesta en el siguiente producto

$$\Sigma = A \, \Lambda \, A^T$$

donde A es una base espectral para Σ ($A A^T = I$) y Λ es una matriz diagonal.

La matriz Λ, es la misma transformación lineal que Σ, pero respecto a la base formada por las columnas de A, que son los eigenvectores correspondientes a Σ.

El elemento i'ésimo sobre la diagonal de Λ, λ_i, es el eigenvalor correspondiente a la i'ésima columna de A.

En nuestro análisis consideraremos la base A que corresponde a la transformación asociada donde $\lambda_i \geq \lambda_j$ si y solo si $i \leq j$.

Será conveniente considerar en nuestro análisis las variables $A^T x_\alpha$, ya que las raíces características de la matriz de covarianzas no cambian.

Al considerar las variables $A^T x_\alpha$, de acuerdo al Teorema A-6, tenemos que

$$A^T x_\alpha \sim N_p \; (A^T \mu \; , \; A^T \Sigma A = \Lambda)$$

2.2 DISTRIBUCION ASINTOTICA DE S^*

En esta sección se denota por S^* al estimador usual insesgado de Λ que corresponde a N observaciones de la variable $A^T x_\alpha$

$$S^* = A^T S A$$

Los resultados de esta sección serán utilizados para el desarrollo de secciones posteriores de este capítulo que se aplican a la determinación de la distribución de los eigenvectores de S^*.

Por los resultados del Teorema 1.2.1., sabemos que $(N-1)S^*$ tendrá la misma distribución que $\sum_{\alpha=1}^{N-1} V_\alpha V_\alpha^T$ donde $V_1, V_2, \ldots, V_{N-1}$ se distribuyen independientemente según $N_p(\bar{0}, \Lambda)$.

Sabemos también por aplicación del Teorema 1.2.3. que los momentos que corresponden a la matriz $V_\alpha V_\alpha^T$ estan dados por

$$\mathrm{Cov}(V_{i\alpha} V_{j\alpha}, V_{k\alpha} V_{l\alpha}) = \delta_{ik}\,\delta_{jl}\,\lambda_i\,\lambda_j + \delta_{il}\,\delta_{jk}\,\lambda_i\,\lambda_j$$

$$= \lambda_i\,\lambda_j\,(\delta_{ik}\,\delta_{jl} + \delta_{il}\,\delta_{jk})$$

Denotemos a la matriz de covarianzas de la matriz $V_\alpha V_\alpha^T$ por T, cuyas componentes estarán dadas por la expresión anterior.

Mencionamos a continuación el siguiente corolario del Teorema Multivariado del Limite Central.

Corolario 2.2.1. Sea $S^* = A^T S A$, donde

$$S = \frac{1}{N-1} \sum_{\alpha=1}^{N} (X_\alpha - \bar{X})(X_\alpha - \bar{X})^T$$

con $X_\alpha \sim N_p(\mu, \Sigma)$ y $A\,\Lambda\,A^T = \Sigma$ con $A^T A = I$. Entonces

$$U = \sqrt{N-1}\,(S^* - \Lambda) \overset{N\to\infty}{\sim} N(0, T)$$

Demostración. Notamos que $(V_\alpha V_\alpha^T)$ satisfacen las condiciones del Teorema Multivariado del Limite Central (Teorema A - 7) se concluye que :

$$\frac{1}{\sqrt{N-1}} \sum_{\alpha=1}^{N-1} (V_\alpha V_\alpha^T - \Lambda) \overset{N\to\infty}{\sim} N(\bar{0},T)$$

Ahora bien, puesto que $\sum_{\alpha=1}^{N-1} V_\alpha V_\alpha^T$ tiene la misma distribución que $(N-1)\, S^*$,

$$\frac{1}{\sqrt{N-1}} \left(\sum_{\alpha=1}^{N-1} V_\alpha V_\alpha^T - (N-1)\,\Lambda \right) \sim \frac{1}{\sqrt{N-1}} \left((N-1)\, S^* - (N-1)\,\Lambda \right)$$

$$= \sqrt{N-1} \; (S^* - \Lambda) \overset{N\to\infty}{\sim} N(\bar{0},T).$$

∎

Denotemos la descomposición espectral de la matriz S^* por

$$S^* = E\, D\, E^T$$

donde

$$d_i \geq d_j \quad si \quad i \leq j$$

$$E^T E = I$$

Nótese que los eigenvalores de la matriz S no cambian.

$$S = C\, D\, C^T \qquad \text{donde} \qquad C = A\, E$$

2.3 DISTRIBUCION ASINTOTICA DE D (CASO $\Lambda = \lambda I$)

En esta sección se determina la distribución asintótica de la matriz D ($S = C D C^T$ con $S = \frac{1}{N-1} \sum (X_i-\bar{X}) (X_i-\bar{X})^T$) para el caso en que $X_i \sim N_P (\mu , \Sigma=\lambda I)$.

El procedimiento que se sigue en esta sección sirve como base para el desarrollo del caso general que se expone en la siguiente sección.

Denotemos por $U = \sqrt{N-1} (S^* - \Lambda)$, donde S^* es el estimador insesgado de la matriz de covarianzas Λ ($\Sigma = A \Lambda A^T$) correspondiente a N observacionesde de la variable $A^T X_\alpha$.

Al despejar S^*, obtenemos

$$S^* = \Lambda + (N-1)^{-1/2} U = \lambda I + (N-1)^{-1/2} U \qquad (3)$$

Denotemos ahora por $H = \sqrt{N-1} (D - \lambda I)$, donde D es la matriz similar a S^* pero con la base dada por la matriz E ($S^* = E D E^T$, $E^T E = I$).

Al despejar D y sustituyendo en la descomposición espectral de S^*, se obtiene

$$S^* = E D E^T = \lambda I + (N-1)^{-1/2} E H E^T \qquad (4)$$

Al igualar (3) y (4) resulta que la descomposición espectral de U ($U = \sqrt{N-1} (S^* - \Lambda)$) esta dada por

$$U = E H E^T$$

Obsérvese que con el ordenamiento de los elementos de la diagonal de H($h_i \geq h_j$ si $i \leq j$)y que los elementos de la diagonal de E sean positivos, se define a E y H como una función continua de U.

Teorema 2.3.1. La función de densidad asintótica de la matriz $H = \sqrt{N-1}\ (D - \lambda I)$ esta dada por

$$g(H;\lambda,p) = \frac{\exp\left(-\frac{\lambda^2}{4} \sum h_i^2\right) \prod_{i<j} (h_i - h_j)}{2^{p(p+3)/4}\ \lambda^{p(p+1)/2} \prod_i \Gamma\left(\frac{1}{2}(p-i+1)\right)}$$

Demostración. De acuerdo a la sección 2.2.,

$$U = \sqrt{N-1}\ (S^{*} - \Lambda)$$

es una matriz simetrica que tiene una distribución asintótica normal con media 0 y una matriz diagonal de varianzas y covarianzas T, donde

$$E(U_{ii}^2) = 2\lambda^2$$

$$E(U_{ij}^2) = \lambda^2 \qquad i \neq j$$

$$\mathrm{Cov}(U_{ij}, U_{kl}) = 0$$

Nótese que los componentes de la matriz U son independientes.

Debido a que $U = U^T$, consideraremos el vector

$$u^T = [u_{11}, u_{12}, \ldots, u_{1p}, u_{22}, u_{23}, \ldots, u_{pp}]$$

que tiene $p(p+1)/2$ componentes.

Derivamos a continuación la función de densidad de u.

Debido a que

$$T = \begin{vmatrix} 2\lambda^2 \lambda^2 \ddots & & & 0 \\ & 2\lambda^2 \lambda^2 \lambda^2 & & \\ & & \ddots & \\ 0 & & & 2\lambda^2 \end{vmatrix}$$

obtenemos la forma cuadrática necesaria para obtener la densidad de u.

$$u^T T^{-1} u = \sum_i \frac{1}{2\lambda^2} u_{ii}^2 + \sum_{i<j} \frac{1}{\lambda^2} u_{ij}^2$$

$$= \frac{1}{2\lambda^2} \left[\sum_i u_{ii}^2 + \sum_{i<j} 2\, u_{ij}^2 \right] = \frac{1}{2\lambda^2} \sum_i \sum_j u_{ij}\, u_{ji}$$

$$= \frac{1}{2\lambda^2} \operatorname{Tr}(U^2)$$

Además $\quad |T| = (2\lambda^2)^p \; (\lambda^2)^{p(p-1)/2} = 2^p \; \lambda^{2p} \; \lambda^{p(p-1)}$
por lo tanto

$$|T|^{-1/2} = 2^{-p/2} \; \lambda^{-p(p+1)/2}$$

Concluimos entonces que la función de densidad de la matriz U, está dada por :

$$f_u = (2\pi)^{-p(p+1)/4} \, 2^{-p/2} \; \lambda^{-p(p+1)/2} \; \exp\left\{ \frac{1}{4\lambda^2} \operatorname{Tr}(U^2) \right\}$$

Haciendo uso del siguiente Teorema A-8, el cual se aparece a continuación, se encontrará la distribución de las raices de

$$|U - hI| = 0.$$

Si la matriz $U = U^T$, tiene una densidad de la forma $f(h_1,..,h_p)$ donde $h_1 > \dots > h_p$ son las raíces características de U, entonces la distribución conjunta de las raíces es

$$\frac{\pi^{p(p+1)/4} \; f(h_1,\dots,h_p) \prod_{i>j} (h_i - h_j)}{\prod_{i=1}^{p} \Gamma\left(\frac{1}{2} (p-i+1) \right)}$$

y la matriz de vectores característicos normalizados E se distribuye de manera independiente según la distribución condicional invariante Haar (ver Anderson T. W. 1958).

El Teorema A-8 es aplicable directamente al caso de la matriz U, al tomar en cuenta que las raíces características de la matriz U son $h_1^2, h_2^2, \dots, h_p^2$ y sustituimos la traza de U^2 por $\sum_{i=1}^{p} h_i^2$ función de densidad de U.

$$f_u = (2\pi)^{-p(p+1)/4} \; 2^{-p/2} \; \lambda^{-p(p+1)/2} \; \exp\left\{ -\frac{1}{4\lambda^2} \sum_{i=1}^{p} h_i^2 \right\}$$

Por los resultados anteriores, concluimos que las raíces características de $U = E H E^T$ (donde $H = \sqrt{N-1}(D - \lambda I)$) tienen la función de densidad

$$g(H;\lambda,p) = \frac{\exp\left(-\frac{1}{4\lambda^2} \sum h_i^2 \right) \prod_{i<j} (h_i - h_j)}{2^{p(p+3)/4} \; \lambda^{p(p+1)/2} \prod_{i=1} \Gamma\left(\frac{1}{2}(p-i+1)\right)}$$

2.4 DISTRIBUCION ASINTOTICA DE D (CASO GENERAL)

En esta sección se determina la distribución asintótica de las componentes de la matriz D ($S = C\, D\, C^T$) para el caso general en que las raíces de Σ ($\Sigma = A\,\Lambda\, A^T$) $k_1, k_2, \ldots, k_r$ tengan multiplicidades $q_1, q_2, \ldots, q_r$ respectivamente , y donde $k_1 > k_2 > \ldots > k_r > 0$.

Se hará uso del método y resultados de la sección anterior.

Haremos la siguiente partición de las matrices de interés:

$$\Lambda = \begin{pmatrix} \Lambda_1 & & & \\ & \Lambda_2 & & \\ & & \ddots & \\ & & & \Lambda_r \end{pmatrix} = \begin{pmatrix} k_1 I & & & \\ & k_2 I & & \\ & & \ddots & \\ & & & k_r I \end{pmatrix}$$

$$U = \begin{pmatrix} U_{11} & U_{12} & \cdots & U_{1r} \\ U_{21} & U_{22} & \cdots & U_{2r} \\ \vdots & & & \\ U_{r1} & U_{r2} & \cdots & U_{rr} \end{pmatrix} \qquad D = \begin{pmatrix} D_1 & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_r \end{pmatrix}$$

$$H = \begin{pmatrix} H_1 & & & \\ & H_2 & & \\ & & \ddots & \\ & & & H_r \end{pmatrix} \qquad E = \begin{pmatrix} E_{11} & E_{12} & \cdots & E_{1r} \\ E_{21} & E_{22} & \cdots & E_{2r} \\ \vdots & \vdots & & \vdots \\ E_{r1} & E_{r2} & \cdots & E_{rr} \end{pmatrix}$$

donde $H_s = \sqrt{N-1}\ (D_s - k_s I)$

Teorema 2.4.1. Sean $d_1 > d_2 > \dots > d_p$, las raices carac- terísticas de una matriz de covarianzas S, basada en una muestra de N observaciones de $N_p\ (\mu,\Sigma)$; donde $k_1>k_2>\dots>k_r$ son las raices características de Σ con multiplicidades $q_1,q_2,\dots,q_r$ respectivamente. Entonces H_s tiene la distribución dada en el Teorema 2.3.1 de la sección anterior, pero con λ reemplazada por k_s, y p por q_s.

$$H_s \overset{N\to\infty}{\sim} g(H_s;k_s,q_s)$$

$$g(H_s;k_s,q_s) = \frac{\exp\left(-\frac{1}{4k_s^2}\sum^{q_s} h_i\right)\prod_{i<j}(h_i - h_j)}{2^{q_s(q_s+3)/4}\ k_s^{q_s(q_s+1)/2}\ \prod \Gamma\left(\frac{1}{2}(q_s-i+1)\right)}$$

Demostración. Sabemos que $U = \sqrt{N-1}\,(S^* - \Lambda)$ y además

$$S^* = E\ D\ E^T$$

entonces :

$$U = \sqrt{N-1}\,(E\ D\ E^T - \Lambda) \qquad (5)$$

como $H = \sqrt{N-1}\,(D - \Lambda) \implies D = \Lambda + (N-1)^{-1/2} H$

Al sustituir en (5), obtenemos

$$U = \sqrt{N-1}\,(E\,(\Lambda + (N-1)^{-1/2} H)\,E^T - \Lambda)$$

Consideremos ahora la expresión de la submatriz U_{kk} de la diagonal de U.

$$U_{kk} = \sqrt{N-1}\,(E_{k1},E_{k2},\dots,E_{kr})\,(\Lambda+(N-1)^{-1/2} H)\,(E_{k1},E_{k2},\dots,E_{kr})^T - \sqrt{N-1}\,\Lambda_{kk}$$

$$= \sqrt{N-1}\,(E_{k1}\Lambda_1,E_{k2}\Lambda_2,\dots,E_{kr}\Lambda_r)\,(E_{k1},E_{k2},\dots,E_{kr})^T$$

$$+ (E_{k1}H_1,E_{k2}H_2,\dots,E_{kr}H_r)\,(E_{k1},E_{k2},\dots,E_{kr})^T - \sqrt{N-1}\,\Lambda_{kk}$$

$$= \sqrt{N-1} \sum_{s=1}^{r} E_{ks} \Lambda_s E_{ks}^T + \sum_{s=1}^{r} E_{ks} H_s E_{ks}^T - \sqrt{N-1} \Lambda_{kk}$$

$$= \sqrt{N-1} \sum_{s=1}^{r} k_s E_{ks} E_{ks}^T + \sum_{s=1}^{r} E_{ks} H_s E_{ks}^T - \sqrt{N-1} \Lambda_{kk}$$

por lo tanto

$$U_{kk} = \sqrt{N-1} \sum_{s=1}^{r} k_s E_{ks} E_{ks}^T + E_{kk} H_k E_{kk}^T + \sum_{s\neq k}^{r} E_{ks} H_s E_{ks}^T - \sqrt{N-1} \Lambda_{kk}$$

En este caso H_k y E_{kk} son funciones de U, y dependen de N. A continuación mostramos que si una secuencia de matrices simétricas $U_{(N)}$ converge a U, entonces las correspondientes secuencias de $H_{k(N)}$, y $E_{kk(N)}$ convergen a la solución de

$$U_{kk} = E_{kk} H_k E_{kk}^T \qquad I = E_{kk} E_{kk}^T$$

Probaremos el caso particular de $H_{(N)}$ y $E_{11(N)}$.

Consideremos $d_{1(N)} > \ldots > d_{q1(N)}$, las q_1 raices caracte - rísticas mas grandes de $S^*_{(N)}$ ($S^*_{(N)} = \Lambda + {}_{(n-1)}{}^{-1/2} U_{(N)}$), entonces las q_1 raices más grandes de

$$| S^*_{(N)} - (k_1 + (N-1)^{-1/2} h) I |$$

$$= | \Lambda + (N-1)^{-1/2} U_{(N)} - (k_1 + (N-1)^{-1/2} h) I | = 0 \qquad (5)$$

son $h_{1(N)} = \sqrt{N-1} \, (d_{1(N)} - k_1) > \ldots . > h_{q1(N)} = \sqrt{N-1} \, (d_{q1(N)} - k_1)$

Ahora bien, denotemos por

$$\Lambda = \begin{pmatrix} k_1 I & 0 \\ 0 & \Lambda^* \end{pmatrix} \qquad U_{(N)} = \begin{pmatrix} U_{11(N)} & U^{*T}_{(N)} \\ U^*_{(N)} & U^{**}_{(N)} \end{pmatrix}$$

y sustituyamos en (6).

$$\begin{vmatrix} (N-1)^{-1/2} (U_{11(N)} - h\ I) & (N-1)^{-1/2} U^{*T}_{(N)} \\ (N-1)^{-1/2} U^*_{(N)} & \Lambda^* - k_1 I + (N-1)^{-1/2}(U^{**}_{(N)} - h\ I) \end{vmatrix} = 0$$

Al factorizar $(N-1)^{-1/2}$ de las primeras q_1 columnas, obtenemos el siguiente determinante

$$\begin{vmatrix} U_{11(N)} - h\ I & (N-1)^{-1/2} U^{*T}_{(N)} \\ U^*_{(N)} & \Lambda^* - k_1 I + (N-1)^{-1/2} (U^{**}_{(N)} - hI) \end{vmatrix}$$

$$= \begin{vmatrix} U_{11(N)} - hI & 0 \\ U^*_{(N)} & \Lambda^* - k_1 I \end{vmatrix} + \begin{vmatrix} U_{11(N)} - hI & (N-1)^{-1/2} U^{*T}_{(N)} \\ U^*_{(N)} & (N-1)^{-1/2}(U^{**}_{(N)} - hI) \end{vmatrix}$$

La última expresión debida al hecho de que

$$\text{Det } [A_1, A_2+A_3] = \text{Det } [A_1, A_2] + \text{Det } [A_1, A_3]$$

Los coeficientes del polinomio asociado al segundo determinante tienden a 0 conforme N tiende a ∞.

La matriz del primer determinante puede ser expresada como

$$\begin{bmatrix} U_{11(N)} - hI & 0 \\ U^{*}_{(N)} & \Lambda^{*} - k_1 I \end{bmatrix} = \begin{bmatrix} U_{11(N)} - hI & 0 \\ U^{*}_{(N)} & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \Lambda^{*} - k_1 I \end{bmatrix}$$

y en consecuencia obtenemos conforme n tiende a ∞, el polinomio de grado q_1 siguiente.

$$| U_{11} - hI | \; | \Lambda^{*} - k_1 I | = 0$$

Por lo tanto q_1 raíces del polinomio original tienden a las q_1 raíces de $| U_{11} - hI | = 0$, porque las raices son funciones continuas de los coeficientes.

Como para $i > q_1$, $d_i \longrightarrow k_s$ $(s > 1)$, esto implica que $h_i \longrightarrow 0$, por lo tanto las q_1 raices mas grandes del polinomio original tienden a la solución de $| U_{11} - hI | = 0$.

De manera similar se demuestra que $H_{k(N)}$ tiende a la solución de

$$U_{kk} = E_{kk} H_k E_{kk}^{T} \quad \text{para} \quad k > 1.$$

Con ayuda del siguiente teorema determinamos que la distribución - límite de E_{kk} y H_k es la única distribución tal que E_{kk} es orto- normal y la distribución de $E_{kk} H_k E_{kk}^{T}$ es la distribución de U_{kk}.

Teorema

Sea $F_N(U)$ la distribución de la matriz $U_{(N)}$, $G_N(V)$ la distribución inducida de $V_{(N)} = f_N(U_N)$, y supongamos que dada una sucesión de matrices $\{ u_{(N)} \}$

$$\lim_{N \to \infty} u_{(N)} = u$$

$$\lim v_{(n)} = v$$

entonces si $U_{(N)} \overset{N \to \infty}{\sim} F(U)$, se tiene que

$$V_{(N)} \overset{N \to \infty}{\sim} G(V)$$

donde $G(V)$ es la distribución de $V=f(U)$.

Por lo tanto H_s tiene la distribución dada en el Teorema 2.3.1., pero con λ reemplazada por k_s y p por q_s.

∎

Corolario 2.4.1. Para el caso en que todos los eigenvalores poblacionales son positivos y distintos, se tienen los siguientes resultados asintóticos :

$$E(d_i) = \lambda_i$$

$$\mathrm{Cov}(d_i , d_j) = \begin{cases} 2 \lambda_i^2 / (N-1) & i = j \\ 0 & i \neq j \end{cases}$$

$$h_i = (N-1)^{1/2} (d_i - \lambda_i) \overset{N \to \infty}{\sim} N (0,2 \lambda_i^2)$$

Demostración. De acuerdo al teorema anterior, cuando alguna de las raíces poblacionales es simple, tenemos :

$$g(h_i;\lambda_i,1) = \frac{\exp\{-\frac{1}{4\lambda_i^2} h_i^2\}}{2\lambda_i \; \Gamma(\frac{1}{2})}$$

$$= \frac{\exp\{-\frac{1}{2}(\frac{h_i}{\sqrt{2}\lambda_i})^2\}}{2 \; \lambda_i \; \sqrt{\pi}} = \frac{\exp\{-\frac{1}{2}(\frac{h_i}{\sqrt{2}\lambda_i})^2\}}{\sqrt{2\pi} \; \sqrt{2} \; \lambda_i}$$

es decir

$$h_i = \sqrt{N-1} \; (d_i - \lambda_i) \overset{N \to \infty}{\sim} N(0, 2\lambda_i^2)$$

∎

3. ESTIMADORES MAXIMO VEROSIMILES DE LAS RAICES CARACTERISTICAS DE Σ

3.1. INTRODUCCION

En esta sección se determinan los estimadores máximo verosímiles de las raíces características de una matriz de covarianzas Σ a partir de N observaciones de una población normal p-variada con parámetros μ y Σ.

Se ha usado el método de máxima verosimilitud debido a que ha sido muy útil en varios problemas de estimación y pruebas de hipótesis, además de que usualmente los estimadores derivados por este método ó funciones de ellos tienen algunas propiedades óptimas (consistentes, convergencia asintótica normal, etc).

Obsérvese que la distribución de las raíces características D_s, está completamente especificada por k_s, es por esto que el primer problema estadístico es cómo estimar este parámetro en base a una muestra de observaciones.

3.2. ESTIMADORES MAXIMO VEROSIMILES DE Σ

Deduciremos a continuación estimadores máximo verosímiles (M.V.) para el caso de una matriz de covarianzas Σ, con raíces $k_1 > k_2 > .. > k_r$ de multiplicidades $q_1, q_2, ..., q_r$ respectivamente ($\sum_{i=1}^{r} q_i = p$).

Teorema 3.2.1., Sea $X_1, X_2, ..., X_N$, N observaciones independientes de una N_p (μ, Σ) y sean $k_1 > k_2 > ... > k_r$ las raíces características de la matriz Σ, entonces el estimador máximo verosímil de k_s esta dado por

$$\hat{k}_s = \frac{1}{q_s} \left(\frac{N-1}{N} \right) \sum_{i \in L_s} d_i$$

donde L_s es el conjunto de enteros

$$q_1 + ... + q_{s-1} + 1 , ..., q_1 + ... + q_s$$

Demostración. La función de verosimilitud correspondiente es:

$$L(\mu,\Sigma) = -\frac{N}{2}\ln|2\pi\,\Sigma| - \frac{1}{2}\,\mathrm{tr}\,\Sigma^{-1}(N-1)\,S - \frac{N}{2}\,(\bar{X}-\mu)^T\,\Sigma^{-1}(\bar{X}-\mu) \quad (7)$$

Debido a que Σ^{-1} es positiva definida (Rango Completo)

$$(\bar{X}-\mu)^T\,\Sigma^{-1}\,(\bar{X}-\mu) > 0 \qquad \forall \quad \mu \neq \bar{X}$$

en consecuencia $\hat{\mu} = \bar{X}$.

Por lo tanto, sólo resta maximizar el 1er y 2do término de (7), que son sólo funciones de Σ.

Aplicando propiedades del logaritmo y determinante, obtenemos

$$-\frac{N}{2}\ln|2\pi\,\Sigma| - \frac{1}{2}\,\mathrm{Tr}\,\Sigma^{-1}(N-1)S$$

$$= -\frac{Np}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\,\mathrm{Tr}\,\Sigma^{-1}(N-1)S$$

Haremos uso de la descomposición espectral de Σ y S, como sigue :

$$-\frac{N}{2}\ln|\Sigma| - \frac{1}{2}\,\mathrm{Tr}\,\Sigma^{-1}(N-1)S$$

$$= -\frac{N}{2}\ln|A\Lambda A^T| - \frac{1}{2}(N-1)\,\mathrm{Tr}\,A\,\Lambda^{-1}A^T\,C\,D\,C^T$$

$$= -\frac{N}{2}\ln|\Lambda| - \frac{1}{2}(N-1)\,\mathrm{Tr}\,\Lambda^{-1}\,P^T\,D\,P \quad (8)$$

donde : $\quad P = C^TA \qquad S = C\,D\,C^T \qquad C^T\,C = I$

Obsérvese que siendo Λ^{-1} y D matrices diagonales,

$$\Lambda^{-1}\ P^T = \begin{pmatrix} 1/\lambda_1 & & & \\ & 1/\lambda_2 & & \\ & & \ddots & \\ & & & 1/\lambda_p \end{pmatrix} \begin{pmatrix} p_{11} & p_{21} & \dots & p_{p1} \\ p_{12} & p_{22} & \dots & p_{p2} \\ \vdots & \vdots & & \vdots \\ p_{1p} & p_{2p} & \dots & p_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} p_{11}/\lambda_1 & p_{21}/\lambda_1 & \dots & p_{p1}/\lambda_1 \\ p_{12}/\lambda_2 & p_{22}/\lambda_2 & \dots & p_{p2}/\lambda_2 \\ \vdots & \vdots & & \vdots \\ p_{1p}/\lambda_p & p_{2p}/\lambda_p & \dots & p_{pp}/\lambda_p \end{pmatrix}$$

$$D\ P = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_p \end{pmatrix} \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1p} \\ p_{21} & p_{22} & \dots & p_{2p} \\ \vdots & \vdots & & \vdots \\ p_{p1} & p_{p2} & \dots & p_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} d_1\ p_{11} & d_1\ p_{12} & \dots & d_1\ p_{1p} \\ d_2\ p_{21} & d_2\ p_{22} & \dots & d_2\ p_{2p} \\ \vdots & \vdots & & \vdots \\ d_p\ p_{p1} & d_p\ p_{p2} & \dots & d_p\ p_{pp} \end{pmatrix}$$

En consecuencia, de la matriz $\Lambda^{-1}\ P^T\ D\ \ P$, el término - i'ésimo de la diagonal viene dado por

$$\sum_{s=1}^{p} p_{si}^2\ \frac{d_s}{\lambda_i}$$

Por lo tanto el 2do. término de (8) Tr $\Lambda^{-1}\ P^T\ D\ \ P$, puede ser escrito como :

$$\sum_{i=1}^{p}\ \sum_{s=1}^{p} p_{si}^2\ \frac{d_s}{\lambda_i} = \sum_{s=1}^{p}\ d_s \sum_{i=1}^{p}\ \frac{p_{si}^2}{\lambda_i}$$

$$= \sum_{s=1}^{p} d_s \left\{ \frac{1}{k_1}\sum_{j=1}^{q_1} p_{sj}^2 + \frac{1}{k_2}\sum_{j=q_1+1}^{q_2} p_{sj}^2 + \ldots + \frac{1}{k_r}\sum_{j=q_{r-1}+1}^{q_r} p_{sj}^2 \right\}$$

$$= \sum_{s=1}^{p} d_s \sum_{i=1}^{r} \frac{q_{si}}{k_i} \qquad (9)$$

donde $q_{si} = \sum_{j\in L_i} p_{sj}^2$ y L_i es el conjunto de enteros

$$q_1 + \ldots + q_{i-1} + 1, \ldots , q_1 + \ldots + q_i.$$

Debido a que $P = C^T A$ es ortonormal ($P P^T = C^T A A^T C = C^T I C = I$),

$$\sum_{i=1}^{r} q_{si} = \sum_{j=1}^{p} p_{sj}^2 = 1 \qquad (10)$$

$$\sum_{s=1}^{p} q_{si} = \sum_{s=1}^{p} \sum_{j\in L_i} p_{sj}^2 = \sum_{j\in L_i} \sum_{s=1}^{p} p_{sj}^2 = \sum_{j\in L_i} 1 = q_i \qquad (11)$$

De acuerdo a lo anterior, sólo resta minimizar (9), sujeto a las condiciones (10) y (11), el cual es claramente un problema de programación lineal.

La solución al problema anterior, esta dada por la siguiente solución básica factible (degenerada) :

$$q_{is} = \begin{cases} 1 & \text{si} \quad i \in L_s \\ 0 & \text{si} \quad i \notin L_s \end{cases} \qquad (12)$$

siendo el mínimo de la función $\sum_{s=1}^{r} \sum_{i\in L_s} \frac{d_i}{k_s}$.

Se puede obtener la solución dada por aplicación del método simplex de dos fases al siguiente problema :

$$\text{Min} \quad z = \sum_{i=1}^{p} \sum_{s=1}^{r} \frac{d_i}{K_s} \, q_{is}$$

$$\text{s.a.} \qquad \sum_{s=1}^{r} q_{is} = 1 \qquad i = 1,..,p$$

$$\sum_{i=1}^{p} q_{is} = q_s \qquad s = 1,..,r$$

$$q_{is} \geq 0 \qquad \forall\ i = 1,..,p \ \wedge \ s=1,..,r$$

Sin embargo el carácter general del problema y la necesidad del uso de variables artificiales, complican la aplicación del método.

Probaremos a continuación que dada una solución factible distinta de (12), el valor de la función objetivo puede ser disminuido.

Consideremos una solución factible $\{q_{ik}\}$ y denotemos por

$$l = \min \{ k \ / \ q_{ik} < 1 \quad i \in L_k \}$$

$$j = \min \{ i \ / \ i \in L_l \quad q_{il} < 1 \}$$

En consecuencia para $k < l$

$$q_{ik} = \begin{cases} 0 & i \notin L_k \\ 1 & i \in L_k \end{cases}$$

ya que $\sum_i q_{ik} = q_k$.

Además para $i < j$ e $i \in L_k$, $q_{ik} = 0$ ya que $\sum_k q_{ik} = 1$.

Lo anterior implica que existen índices k e l, tales que :

$$q_{jk} > 0 \qquad \text{con} \quad k > l$$

$$q_{il} > 0 \qquad \text{con} \quad i > q_1 + \ldots + q_l$$

Sea $\Delta = \min (q_{jk} , q_{il})$.

Una solución factible, se obtiene reemplazando q_{jl}, q_{jk}, q_{il}, q_{ik} por $q_{jl}+\Delta$, $q_{jk}-\Delta$, $q_{il}-\Delta$, $q_{ik}+\Delta$.

Denotemos por $\hat{z}$ el valor de la función objetivo, evaluada en la nueva solución factible, entonces :

$$z - \hat{z} = \Delta (d_j - d_i) \left(\frac{1}{\lambda_l} - \frac{1}{\lambda_k} \right) < 0$$

ya que $k > l$ ($\Longrightarrow \lambda_l > \lambda_k$).

Por lo tanto, la expresión final de (8) a maximizar (una vez maximizada con respecto a P) es :

$$- \frac{1}{2} N \ln |\Lambda| - \frac{1}{2} (N-1) \sum_{s=1}^{r} \sum_{i \in L_s} \frac{d_i}{k_s} \qquad (13)$$

Ahora bien como

$$\ln |\Lambda| = \ln \prod_{i=1}^{p} \lambda_i = \ln \prod_{s=1}^{r} k_s^{q_s} = \sum_{s=1}^{r} q_s \ln k_s$$

sustituyendo en la expresión (13), tenemos

$$- \frac{1}{2} N \sum_{s=1}^{r} q_s \ln k_s - \frac{1}{2} (N-1) \sum_{s=1}^{r} \sum_{i \in L_s} \frac{d_i}{k_s}$$

Al derivar la última expresión con respecto a k_s e igualando a 0, tenemos :

$$- \frac{1}{2} N \frac{q_s}{k_s} + \frac{1}{2} (N-1) \frac{1}{k_s^2} \sum_{i \in L_s} d_i = 0$$

$$\Longrightarrow \quad \hat{k}_{s_{mv}} = \frac{1}{q_s} \left(\frac{N-1}{N} \right) \sum_{\in L_s} d_i \qquad (14)$$

∎

Corolario 3.2.1. . Si la descomposición espectral de la matriz de covarianzas ($\Sigma = A \Lambda A^T$) se supone que es única,esto es, si las raices λ_i son todas diferentes, entonces el estimador máximo verosímil de Λ esta dado por :

$$\hat{\Lambda} = \left(1 - \frac{1}{N} \right) D$$

y el estimador máximo verosímil de A por

$$\hat{A} = C$$

donde $S = C \ D \ C^T$.

Demostración.

∎

4. CONTRASTES DE HIPOTESIS

4.1 INTRODUCCION

En este capítulo presentamos el desarrollo del contraste de algunas hipótesis relativas a los eigenvalores de una matriz de varianzas y covarianzas que corresponde a una población normal multivariada.

Para cada uno de los contrastes de hipótesis que se presentan se da en la Sección 4.4 la distribución límite del - cociente de verosimilitudes.

Se ha empleado el método de cociente de verosimilitudes - debido a que ha sido muy usado en varios problemas de estimación y contrastes de hipótesis con relación a la distribución normal multivariada.

Como antes, suponemos que $X_1,\ldots,X_N$ es una muestra aleatoria de $N_p(\upsilon,\Sigma)$, donde la descomposición espectral de Σ está dada por $\Sigma = A \Lambda A^T$ con $A^T A = I$.

Las raíces características de la matriz Σ, son $k_1>k_2>\ldots>k_r$ con multiplicidades $q_1,\ldots,q_r$ respectivamente ($\sum_{i=1}^{r} q_i = p$).

Para S (la matriz de varianzas y covarianzas muestral), la descomposición espectral esta dada por :

$$S = C \; D \; C^T \quad \text{con} \quad C^T C = I$$

4.2. HIPOTESIS DE QUE q_s EIGENVALORES DE Σ SON IGUALES.

El siguiente resultado nos da la razón de verosimilitudes para la hipótesis de que q_s eigenvalores son iguales, $q_1+q_2+\ldots+q_{s-1}$ son menores que los q_s, y $q_{s+1}+\ldots+q_p$ son mayores.

Teorema 4.2.1. . El criterio de la razón de verosimilitudes para probar la hipótesis

$$H: \quad \lambda_{q_1+\dots+q_{s-1}+1} = \dots = \lambda_{q_1+\dots+q_s}$$

esta dada por

$$\lambda^*_{(s)} = \left[\prod_{j \in L_s} d_j \;/\; (q_s^{-1} \sum_{j \in L_s} d_j)^{q_s} \right]^{N/2}$$

Demostración. El criterio de la razón de verosimilitudes es

$$\lambda^*_{(s)} = \frac{\underset{\mu,\Sigma \in \omega}{\mathrm{Max}} \; L(\mu,\Sigma)}{\underset{\mu,\Sigma \in \Omega}{\mathrm{Max}} \; L(\mu,\Sigma)}$$

donde Ω es el espacio de μ y Σ (positiva definida); y ω es la región de este espacio donde H se cumple.

Por aplicación del Teorema 1.1.1 obtenemos que

$$\underset{\mu,\Sigma \in \Omega}{\mathrm{Max}} \; L(\mu,\Sigma) = \frac{(2\pi)^{-Np/2} \; \exp(-Np/2)}{(\frac{N-1}{N})^{Np/2} \quad |S|^{N/2}}$$

Por aplicación del Teorema 3.2 con

$q_i = \begin{cases} \geq 1 & i = s \\ 1 & i \neq s \end{cases}$ sabemos que L es máxima en ω en los puntos

$$\hat{\mu} = \bar{X}$$

$$\hat{k}_s = \frac{N-1}{N} \frac{1}{q_s} \sum_{j \in L_s} d_j$$

$$\hat{\lambda}_i = \frac{N-1}{N} d_i \qquad i \in L_s$$

donde toma el valor

$$\underset{\mu,\Sigma \in \omega}{\text{Max}} \; L(\mu,\Sigma) = \frac{(2\pi)^{-Np/2} \exp(-Np/2)}{\prod_{i \in L_s}^{p} \left(\frac{N-1}{N} d_i \right)^{N/2} \hat{k}_s^{Nq_s/2}}$$

por lo tanto

$$\lambda^{*}_{(s)} = \frac{\text{Max}_{W} L(\mu,\Sigma)}{\text{Max}_{\Omega} L(\mu,\Sigma)} = \frac{\dfrac{(2\pi)^{-Np/2} \exp(-Np/2)}{\prod \left(\frac{N-1}{N} d_i \right)^{N/2} \hat{k}_s^{Nq_s/2}}}{\dfrac{(2\pi)^{-Np/2} \exp(-Np/2)}{\left(\frac{N-1}{N} \right)^{Np/2} |S|^{N/2}}}$$

$$= \left[\frac{\left(\frac{N-1}{N} \right)^{p} |S|}{\prod_{i \in L_s}^{p} \left(\frac{N-1}{N} d_i \right) \hat{k}_s^{q_s}} \right]^{N/2}$$

Observese que $|S| = \prod_{i=1}^{p} d_i$, sustituyendo obtenemos

$$\lambda^*_{(s)} = \left\{ \frac{\left(\frac{N-1}{N}\right)^{q_s} \prod_{i \in L_s} d_i}{\hat{k}_s^{q_s}} \right\}^{N/2}$$

∎

Un caso especial del teorema anterior ocurre cuando la hipótesis nula es igualdad de todas las raices poblacionales, lo cual es equivalente a decir que todas las variables son independientes y tienen la misma varianza.

Corolario 4.2.1. , Hipótesis de Esfericidad

Cuando la hipótesis a considerar es $\lambda_1 = \lambda_p$ (igualdad de todas las raices poblacionales), el criterio anterior resulta ser :

$$\left[\left(\frac{p}{Tr(S)} \right)^{p} |S| \right]^{N/2}$$

Demostración. ∎

Teorema 4.2.2. , El criterio de la razón de verosimilitudes para la hipótesis que consiste de r conjuntos de igualdades de la forma

$$\lambda_{q_1+..+q_{s_{i-1}}+1} = \lambda_{q_1+..+q_{s_i}} \qquad i = 1,..,r$$

esta dado por

$$\lambda^{**}_{(r)} = \prod_{i=1}^{r} \lambda^*_{(s_i)}$$

Demostración. Por los mismos argumentos del Teorema 4.1 , obtenemos

$$\lambda^{**}_{(r)} = \frac{\mathrm{Max}_{\mu,\Sigma\in\omega} L(\mu,\Sigma)}{\mathrm{Max}_{\mu,\Sigma\in\Omega} L(\mu,\Sigma)} = \frac{\dfrac{\prod_{j\in\{L_{si}\}} \left(\frac{N-1}{N} d_j\right)^{-N/2} \left(\prod_{i=1}^{r} \hat{k}_{si}^{\,q_{si}}\right)^{-N/2}}{(2\pi)^{Np/2} \quad \exp\,(pN/2)}}{\dfrac{|S|^{-N/2} \quad \left(\frac{N-1}{N}\right)^{-N/2}}{(2\pi)^{Np/2} \quad \exp\,(Np/2)}}$$

$$= \frac{\prod_{j\in\{L_{si}\}}^{p} \left(\frac{N-1}{N} d_j\right)^{-N/2} \left(\prod_{i=1}^{r} \hat{k}_{si}^{\,q_{si}}\right)^{-N/2}}{\left(\prod_{j=1}^{p} d_j\right)^{-N/2} \quad \left(\frac{N-1}{N}\right)^{-Np/2}}$$

$$= \left[\frac{\prod_{j\in\{L_{si}\}} d_j \qquad \left(\frac{N-1}{N}\right)^{\sum_{i=1}^{r} q_{si}}}{\prod_{i=1}^{r} \hat{k}_{si}^{\,q_{si}}} \right]^{N/2}$$

$$= \prod_{i=1}^{r} \lambda^{*}_{(si)}$$

∎

Con frecuencia, lo que se busca es decidir el número de componentes principales necesarias para representar las variables iniciales. Un criterio puede ser el de seleccionar las k componentes principales que tengan una parte "importante" de la variabilidad total, de tal manera de no considerar las componentes principales cuya variación sea cercana a cero. Sin embargo debe notarse que cualquier eigenvalor distinto de cero es significativamente distinto de cero, ya que si $\lambda_p = 0$, entonces con probabilidad 1, $d_p=0$. Es por esto que la hipótesis $\lambda_k = 0$, no tiene sentido.

La siguiente hipótesis nula

$$H : \quad \lambda_p = \lambda_{p-1} = \ldots = \lambda_{k+1}$$

que prueba si la variación en el subespacio indicado es la misma, es útil debido a que las primeras k componentes principales podrían estar midiendo alguna variación substancial en X, y como las últimas p-k componentes principales tienen igual variación, se puede considerar que sólo miden "ruido" matemáticamente esto es equivalente a decir que $\Sigma = \Psi + v\, I$, donde ψ es positiva se - midefinida de rango k.

Se realizan con frecuencia, una secuencia de pruebas de hipótesis de este tipo empezando con k = 0, e incrementando k hasta que la hipótesis nula es aceptada, de esta manera, se puede decidir cuantas componentes principales son distinguibles del "ruido" y que vale la pena retener.

Corolario 4.2.1. . La razón de verosimilitudes para probar la hipótesis nula

$$H : \lambda_p = \lambda_{k+1}$$

esta dada por

$$\lambda^{***}_{(k+1)} = \left\{ \prod_{s=k+1}^{p} d_s \Big/ \left(\sum_{s=k+1}^{p} d_s / (p-k) \right)^{p-k} \right\}^{N/2}$$

Un caso especial del corolario anterior, esta dado por el siguiente lema.

Lema 4.2.1. . La razón de verosimilitudes para probar la hipótesis nula

$$H : \quad \Sigma = v^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

esta dada por $\lambda^{***}_{(2)}$.

Demostración. Al sumar los p-1 últimos renglones (de la matriz $\Sigma - xI$) al primero, obtenemos

$$\Sigma - xI = \begin{bmatrix} 1+(p-1)\rho-X & 1+(p-1)\rho-X & \cdots\cdots & 1+(p-1)\rho-X \\ \rho & 1-X & \cdots\cdots & \rho \\ \vdots & \vdots & & \vdots \\ \rho & \rho & \cdots\cdots & 1-X \end{bmatrix}$$

Restando a cada renglón, el renglón inmediato anterior (a partir del segundo renglón) obtenemos :

$$v^2 \operatorname{Det} (\Sigma - xI)$$

$$= (1-(p-1)\rho-X/v^2)\,(1-\rho-X/v^2)^{p-2} \begin{vmatrix} 1 & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 \\ 0 & 1 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot \\ \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 1 & -1 \\ \rho & \rho & \rho & \rho & \cdot & \cdot & \cdot & \rho & 1-X/v^2 \end{vmatrix}$$

Multiplicando el primer renglón por ρ y restandoselo al último, obtenemos :

$$\mathrm{Det}\ (\ \Sigma - xI\ ) = 0 \qquad \Longrightarrow$$

$$v^2(1+(p-1)\rho) = \lambda_1 > \lambda_2 = \lambda_3 = \ldots\ldots = \lambda_p = v^2\ (1-\rho)$$

∎

4.3 HIPOTESIS DE QUE q_s EIGENVALORES DE Σ SON IGUALES A λ

Teorema 4.3.1. . El criterio de la razón de verosimilitudes para probar la hipótesis

$$H \quad : \quad \lambda_{q_1+\ldots+q_{s-1}+1} = \lambda = \lambda_{q_1+\ldots+q_s}$$

esta dada por

$$\lambda^*_{(s,\lambda)} = \left[\frac{\prod_{i\in L_s} (\frac{N-1}{N}\ d_i)}{\lambda^{q_s}} \right]^{N/2} \exp \left\{ - \frac{N}{2} \left[\frac{N-1}{N} \sum \frac{d_i}{\lambda} - q_s \right] \right\}$$

Demostración. Del teorema 3.2 , resulta

$$\max_{\mu\ \Sigma\ \in\ \omega} L(\mu,\Sigma) = (2\pi)^{-Np/2}\ \lambda^{-Nq_s/2} \prod_{i\in L_s} (\frac{N-1}{N}\ d_i)^{-N/2}$$

$$\exp\ \{ - \frac{N}{2}(p-q_s) - \frac{1}{2}(N-1) \sum_{i\in L_s} \frac{d_i}{\lambda}\ \}$$

y del teorema 1.1, se obtiene

$$\underset{\mu\,\Sigma\,\in\,\Omega}{\mathrm{Max}}\ L(\mu,\Sigma) = \frac{(2\pi)^{-Np/2}\quad \exp(-Np/2)}{\prod_{i=1}^{p}\left(\frac{N-1}{N}\,d_i\right)^{N/2}}$$

Por lo tanto

$$= \frac{\dfrac{(2\pi)^{-Np/2}\ \exp\left\{-\frac{N}{2}(p-q_s)-\frac{N-1}{2}\sum\frac{d_i}{\lambda}\right\}}{\lambda^{Nq_s/2}\ \prod_{i\in L_s}\left(\frac{N-1}{N}\,d_i\right)^{N/2}}}{\dfrac{(2\pi)^{-Np/2}\ \exp\left\{-Np/2\right\}}{\prod_{i=1}^{p}\left(\frac{N-1}{N}\,d_i\right)^{N/2}}}$$

$$= \frac{\prod_{i\in L_s}\left(\frac{N-1}{N}\,d_i\right)^{N/2}\quad \exp\left\{Nq_s/2\right\}}{\lambda^{Nq_s/2}}\quad \exp\left\{-\frac{1}{2}(N-1)\sum\frac{d_i}{\lambda}\right\}$$

$$= \left[\frac{\prod_{i\in L_s}\left(\frac{N-1}{N}\,d_i\right)}{\lambda^{q_s}}\right]^{N/2}\quad \exp\left\{-\frac{N}{2}\left[\frac{N-1}{N}\sum\frac{d_i}{\lambda}-q_s\right]\right\}$$

En el caso de raíces de multiplicidad 1, los mismos resultados obtenidos en el capítulo 3 ($d_k \sim N(\lambda_k, 2\lambda_k^2/(N-1))$, pueden ser usados para construir pruebas de hipótesis.

Por ejemplo, para probar la hipótesis $H: \lambda_k = \lambda$, el estadístico de prueba adecuado sería

$$\frac{d_k - \lambda}{\sqrt{2/(N-1)}\ \lambda} \sim N(0,1)$$

rechazándose la hipótesis si

$$\left| \frac{d_k - \lambda}{\frac{2}{N-1}\ \lambda} \right| \geq Z_{\alpha/2}$$

donde $Z_{\alpha/2}$ es el percentil que acumula $\alpha/2$ % de la distribución normal a partir de 0.

4.4. DISTRIBUCION COCIENTES DE VEROSIMILITUD

En esta sección se presenta el desarrollo de la distribu - ción de los estadísticos de prueba mencionados en las secciones an teriores de este capítulo.

Teorema 4.4.1. . Considerese el estadístico de prueba correspondiente al Teorema 4.2.1.

$$\lambda^*_{(s)} = \left[\prod_{j\in L_s} d_j \ / \ (q_s^{-1} \sum_{j\in L_s} d_j)^{q_s} \right]^{N/2}$$

La distribución limite de $-2 \ln \lambda^*_{(s)}$ esta dada por

$$-2 \ln \lambda^*_{(s)} \overset{N\to\infty}{\sim} \chi^2_{\frac{1}{2}q_s(q_s+1)-1}$$

Demostración. El logaritmo de $\lambda^*_{(s)}$, multiplicado por -2 es asintóticamente equivalente a :

$$-2 \ln \lambda^*_{(s)} = -N \left[\ln \prod_{j\in L_s} d_j - q_s \ln (\sum_{j\in L_s} d_j / q_s) \right]$$

Recordando que $h_i = \sqrt{N-1}\,(d_i-k_s) \Longrightarrow d_i = (N-1)^{-1/2}h_i+k_s$, al sustituir en la expresión anterior $(n = N-1)$:

$$= N \left\{ - \sum \ln (k_s+n^{-1/2}h_i) + q_s \ln (\sum_{j\in L_s} \frac{(k_s + n^{-1/2}h_j)}{q_s}) \right\}$$

$$= N \left\{ - \sum \ln (1 + \frac{n^{-1/2}h_j}{k_s}) + q_s \ln (1 + \frac{\sum h_i}{q_s k_s \sqrt{n}}) \right\}$$

Debido a que $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots$, tenemos que

$$= N \left\{ - \Sigma \left[\frac{h_i}{k_s\sqrt{n}} - \frac{h_i^2}{2k_s^2 n} + \ldots \right] + q_s \left[\frac{\sum h_i}{q_s k_s \sqrt{n}} - \frac{(\sum h_i)^2}{2q_s^2\, k_s^2\, n} + \ldots \right] \right\}$$

$$= N \left\{ - \left[- \frac{\sum h_i}{k_s}\sqrt{n} + \frac{\sum h_i^2}{2k_s^2\, n} - \ldots \right] + \left[\frac{\sum h_i}{k_s}\sqrt{n} - \frac{(\sum h_i)^2}{2q_s\, k_s^2\, n} + \ldots \right] \right\}$$

$$= N \left\{ \left[\frac{\sum h_i^2}{2k_s^2\, n} - \ldots \right] - \frac{(\sum h_i)^2}{2q_s\, k_s^2\, n} + \ldots \right\}$$

La distribución asintótica de esta cantidad, es la distribución asintótica de :

$$\frac{1}{2\, k_s^2} \left\{ \Sigma\, h_j^2 - \frac{1}{q_s} (\sum h_j)^2 \right\} = \frac{1}{2\, k_s^2} \left\{ tr\ U_{ss}^2 - \frac{1}{q_s} (tr\ U_{ss})^2 \right\}$$

$$= \frac{1}{2\, k_s^2} \left\{ tr\ (U_{ss}\, U_{ss}^T) - \frac{1}{q_s} (tr\ U_{ss})^2 \right\}$$

$$= \frac{1}{2\, k_s^2} \left\{ \Sigma\, u_{ii}^2 + 2 \sum_{i<j} u_{ij}^2 - \frac{1}{q_s} (\Sigma\, u_{ii})^2 \right\}$$

En el límite, las componentes de la matriz U_{ss}, tienen una distribución normal con media 0 y componentes independientes con varianzas (vease Capítulo 3) :

$$V(\, u_{ii}) = 2\, k_s^2 \qquad\qquad V(\, u_{ij}) = k_s^2$$

Por lo tanto :

$$\frac{1}{k_s^2} \sum_{i<j} u_{ij}^2 \sim \chi^2_{q_s(q_s-1)/2}$$

$$\frac{1}{2k_s^2} \left[\sum u_{ii}^2 - \frac{(\sum u_{ii})^2}{q_s} \right] = \frac{\sum (u_{ii} - \bar{u})}{2\ k_s^2} \sim \chi^2_{q_s - 1}$$

Lo anterior implica que

$$\frac{1}{2k_s^2} \left[\sum u_{ii}^2 + 2 \sum_{i<j} u_{ij} - \frac{1}{q_s} (\sum u_{ii})^2 \right]$$

tiene una distribución asintótica χ^2 con $\frac{1}{2}$ $q_s(q_s+1) - 1$ grados de libertad.

∎

Corolario 4.4.1. . La distribución límite de menos dos veces el logaritmo del estadistico de prueba correspondiente al teorema 4.2.2. esta dada por :

$$-2 \ln \lambda^{**}_{(r)} = -2 \ln \prod_{i=1}^{r} \lambda^{*}_{(s_i)} \overset{N \to \infty}{\sim} \chi^2 \quad \text{con}$$

$\frac{1}{2} (\sum_{i=1}^{r} q_{s_i}^2 + \sum_{i=1}^{r} q_{s_i}) - r$ grados de libertad.

Demostración. Debido a que

$$- 2 \ln \prod_{i=1}^{r} \lambda^*_{(s_i)} = \sum_{i=1}^{r} - 2 \ln \lambda^*_{(s_i)}$$

y puesto que se trata de una suma de variables aleatorias independientes donde el i'ésimo término tiene una distribución $\aleph^2$ con $\frac{1}{2} q_{s_i}(q_{s_i}+1)-1$ grados de libertad, el resultado es inmediato.

∎

Teorema 4.4.2. . Considérese el estadístico de prueba correspondiente al teorema 4.3.1..

$$\lambda^*_{(s,\lambda)} = \left[\frac{\prod_{i \in L_s} (\frac{N-1}{N} d_i)}{\lambda^{q_s}}\right]^{N/2} \exp \left\{- \frac{N}{2} \left[\frac{N-1}{N} \sum \frac{d_i}{\lambda} - q_s\right] \right\}$$

La distribución límite de $- 2 \ln \lambda^*_{(s,\lambda)}$ está dada por :

$$- 2 \ln \lambda^*_{(s,\lambda)} \overset{N \to \infty}{\sim} \aleph^2_{\frac{1}{2} q_s(q_s+1)}$$

Demostración.

$$- 2 \ln \lambda^{*}_{(s,\lambda)} = - N \ln \left[\frac{\Pi(\frac{N-1}{N} d_i)}{\lambda^{q_s}} \right] + N \left[\frac{N-1}{N} \sum \frac{d_i}{\lambda} - q_s \right]$$

$$= - N \left[\sum \ln (\frac{N-1}{N} d_i) - q_s \ln \lambda \right] + N \left[\frac{N-1}{N} \sum \frac{d_i}{\lambda} - q_s \right]$$

$$= - N \left[\sum \ln d_i + \sum \ln \frac{N-1}{N} - q_s \ln \lambda \right] + (N-1) \sum \frac{d_i}{\lambda} - N q_s$$

$$= -N\left[\sum \ln(n^{-1/2}h_i+\lambda)+\sum \ln\frac{N-1}{N} - q_s \ln\lambda\right]+(N-1) \sum(n^{-1/2}h_i+\lambda)-N q_s$$

$$= - N \left[\sum \ln (1 + \frac{h_i}{\sqrt{n}\,\lambda}) + q_s \ln \frac{N-1}{N} \right] + (N-1) \sum \frac{h_i}{\sqrt{n}\,\lambda} - q_s$$

$$= - N \left[\sum (\frac{h_i}{\sqrt{n}\,\lambda} - \frac{h_i^2}{2n\,\lambda^2} + \dots) + q_s \ln \frac{N-1}{N} \right] + (N-1) \frac{\sum h_i}{\sqrt{n}\,\lambda} - q_s$$

$$= - N \left[\frac{\sum h_i}{\lambda\sqrt{n}} - \frac{\sum h_i^2}{2n\lambda^2} + \dots \right] + N \frac{\sum h_i}{\lambda\sqrt{n}} - \frac{\sum h_i}{\lambda\sqrt{n}} - q_s \left(\ln(\frac{N-1}{N})^N + 1\right)$$

$$= N \left(\frac{\sum h_i^2}{2n\lambda^2} - \dots \right) - \frac{\sum h_i}{\lambda\sqrt{n}} - q_s \left(\ln(\frac{N-1}{N})^N + 1\right)$$

La distribución asintótica de esta cantidad es la distribución asintótica de :

$$\frac{\sum h_i^2}{2\lambda^2} = \frac{1}{2\lambda^2} \operatorname{tr} U_{ss}^2 = \frac{1}{2\lambda^2} \operatorname{tr} (U_{ss} U_{ss}^T)$$

$$= \frac{1}{2\lambda^2} \left\{ \sum u_{ii}^2 + 2 \sum_{i<j} u_{ij}^2 \right\}$$

donde u_{ij} es un elemento de la matriz $U = \sqrt{N-1}\ (S^{*} - \Lambda)$ que - como se vió en el corolario 2.2.1.

$$\frac{u_{ii}}{\sqrt{2}\lambda} \overset{n \to \infty}{\sim} N (0, 2 \lambda^{2})$$

$$\frac{u_{ij}}{\lambda} \overset{n \to \infty}{\sim} N (0, \lambda^{2}) \qquad i \neq j$$

y donde cada componente de la matriz U es independiente.

De acuerdo a lo anterior

$$\frac{1}{2\lambda^{2}} \left\{ \sum u_{ii}^{2} + 2 \sum_{i<j} u_{ij}^{2} \right\} \overset{n \to \infty}{\sim} \chi^{2}_{\frac{1}{2}q_{s}(q_{s}+1)}$$

∎

5. INTERVALOS DE CONFIANZA

Se considera ahora el problema de determinar intervalos de confianza para las raíces poblacionales.

Supongamos que las últimas q_r raíces son iguales.

Sea $\bar{d} = \sum_{i \in L_r} \frac{d_i}{q_r}$, tenemos entonces que

$$\sqrt{n} \, (\bar{d} - k_r) = \frac{1}{q_r} \sum_{i \in L_r} h_i = \mathrm{tr}\, U_{rr} / q_r = \frac{1}{q_r} \sum_{p-q_r+1}^{p} u_{ii}$$

tiene una distribución asintótica normal con media cero y varianza $2 \, k_r^2 / q_r$, en consecuencia

$$\frac{(\frac{1}{2} n \, q_r)^{1/2} (d - \bar{k}_r)}{k_r} \quad \underset{\sim}{N} {}^{\to \infty} \quad N (0 , 1)$$

Entonces con probabilidad $1 - \alpha$ ($n \to \infty$), sucede que

$$\left[n \, q_r (\bar{d} - k_r)^2 \right] / \, 2 \, k_r^2 \leq t^2_{1-\alpha/2}$$

lo cual proporciona el siguiente intervalo de confianza asintótico para k_r

$$\bar{d} / (1 + (2/nq_r)^{1/2} \, t_{1-\alpha/2}) \leq k_r \leq d / (\bar{1} + (2/nq_r)^{1/2} \, t_{1-\alpha/2})$$

siempre y cuando, n, q_r y α sean tales que

$$1 - (2/nq_r)^{1/2}\, t_{1-\alpha/2} \ > \ 0 \ .$$

Un aspecto importante en la construcción de intervalos de confianza esta en el de poder juzgar las magnitudes de los eigenvalores para determinar cuales pueden ser despreciados (considerados como ceros), de tal manera que se puedan ignorar las correspondientes componentes principales.

Si se supone que los últimos q_r eigenvalores son iguales, se puede obtener una cota superior asintótica de confianza.

Con probabilidad $1 - \alpha / 2 \quad (n \to \infty)$

$$\sqrt{q_r\, n/2}\ \ (\bar{d} - k_r) / k_r \ \ \geq \ - t_{1-\alpha}$$

lo cual implica que

$$k_r \ \leq \ \bar{d} / (1 - (2/nq_r)^{1/2}\ t_{1-\alpha})$$

De esta manera se puede decidir estudiar sólo las primeras $p - q_r$ componentes principales si la cota que se obtiene es "suficientemente" pequeña.

Debe de tomarse en cuenta que el supuesto de igualdad de raices conlleva una subestimación de la varianza y una reducción del intervalo de confianza cuando las raíces no son iguales. Esto debido a que

$$\sqrt{n}\left[\bar{d}-\left(\frac{\sum\lambda_i}{q_r}\right)\right]=\frac{1}{q_r}\sum h_i \quad \overset{N\to\infty}{\sim} \quad N\left(0,\ 2\,\frac{\sum\lambda_i^2}{q_r}\right)$$

Un procedimiento más conservador es el de reemplazar la varianza $2\,\frac{\sum\lambda_i^2}{q_r^2}$ por su estimador consistente $2\,\frac{\sum d_i^2}{q_r^2}$.

C O N C L U S I O N E S

Como se indicó al inicio de este trabajo, uno de los principales problemas en cualquier análisis de datos es el de determinar que tanto de la variación en los datos es "sistemática" y cuanta es aleatoria o "ruido". En este trabajo este problema se atacó dedeterminando la distribución asintótica para cocientes de verosimilitud correspondientes a hipótesis estadísticas sobre los eigenvalores de una matriz de covarianzas muestral cuando la muestra es grande y proviene de un vector aleatorio con una distribución normal p-variada.

Se consideraron además hipótesis estadísticas concernientes a probar que tanto de la variación esta representada por un número específico de componentes principales.

Para todas estas pruebas de hipótesis, resulta que la distribución asintótica del cociente de verosimilitudes tiene una distribución chi-cuadrada.

Se determinó también la distribución asintótica de los de los eigenvalores de la matriz de covarianzas muestral, que resultan ser las varianzas correspondientes a las combinaciones lineales resultamtes de un ACP. Para el caso de raíces de multiplicidad 1, se obtuvo que la distribución limite era una normal.

Con estos resultados se permite un uso más formal de técnicas estadísticas (ACP, Análisis Factorial Clásico, etc.) que involucren el uso de los eigenvalores en sus resultados.

La aportación principal de este trabajo es la de permitir un análisis satisfactorio, secuencial y consistente del tema, representando el trabajo material de consulta importante.

A P E N D I C E

En esta sección se presentan algunos resultados de algebra y probabilidad que fueron usados en este trabajo.

LEMA A-1. Sean $X_1, X_2, \ldots, X_N$, N vectores donde $X_i \in R^p$ y sea $\bar{X} = \sum_{i=1}^{N} X_i / N$, entonces $\forall\ \mu \in R^p$,

$$\sum_{\alpha=1}^{N} (X_\alpha - \mu)(X_\alpha - \mu)^T = \sum_{\alpha=1}^{N} (X_\alpha - \bar{X})(X_\alpha - \bar{X})^T + N(\bar{X} - \mu)(\bar{X} - \mu)^T$$

Demostración.

$$\sum_{\alpha=1}^{N} (X_\alpha - \mu)(X_\alpha - \mu)^T = \sum_{\alpha=1}^{N} \Big[(X_\alpha - \bar{X}) + (\bar{X} - \mu)\Big] \Big[(X_\alpha - \bar{X}) + (\bar{X} - \mu)\Big]^T$$

$$= \sum_{\alpha=1}^{N} \Big[(X_\alpha - \bar{X})(X_\alpha - \bar{X})^T + (X_\alpha - \bar{X})(\bar{X} - \mu)^T$$

$$+ (\bar{X} - \mu)(X_\alpha - \bar{X})^T + (X_\alpha - \mu)(\bar{X} - \mu)^T \Big]$$

$$= \sum_{\alpha=1}^{N} (X_\alpha - \bar{X})(X_\alpha - \bar{X}) + \left[\sum_{\alpha=1}^{N} (X_\alpha - \bar{X})\right] (\bar{X} - \mu)^T$$

$$+ (\bar{X} - \mu) \sum_{\alpha=1}^{N} (X_\alpha - \bar{X})^T + N(\bar{X} - \mu)(\bar{X} - \mu)^T$$

Debido a que $\sum (X_\alpha - \bar{X}) = 0$, el resultado es inmediato.

$\blacksquare$

Teorema A - 2. Sea

$$f(C) = \frac{N}{2} \ln|C| - \frac{1}{2} \operatorname{Tr} CA$$

donde C y A son matrices positivas definidas. Entones el máximo de f(C) esta en $C = N A^{-1}$ donde toma el valor :

$$f(N A^{-1}) = \frac{Np}{2} \ln N - \frac{N}{2} \ln|A| - \frac{Np}{2}$$

Demostración.

Al derivar f(C) con respecto a C_{kl}, se obtiene :

$$\frac{\partial f}{\partial C_{kl}} \begin{cases} \frac{1}{2} N \frac{\operatorname{cof}(C_{kk})}{|C|} - \frac{1}{2} a_{kk} & \text{si } k = l \\ N \frac{\operatorname{cof}(C_{kl})}{|C|} - a_{kl} & \text{si } k \neq l \end{cases}$$

Al igualar a cero, obtenemos :

$$N C^{-1} = A \quad \text{es decir}$$

$$C = N A^{-1} \quad \text{donde f toma el valor}$$

$$f(N A^{-1}) = \frac{N}{2} \ln (N^{p} |A^{-1}|) - \frac{N}{2} \operatorname{Tr} A^{-1} A$$

$$= \frac{Np}{2} \ln N - \frac{N}{2} \ln |A| - \frac{Np}{2}$$

■

Teorema A - 3. Sea X ~ $N_p(\mu,\Sigma)$. La función característica de X esta dada por

$$\phi_X(t) = \exp\left(it^T\mu + \frac{1}{2}\, t^T \Sigma\, t \right)$$

Demostración.

La función característica de X puede obtenerse mediante el uso de la transformación $Y = \Sigma^{-1/2}(X-\mu)$, donde $E(Y)=0$ y $V(Y)=I$. La función característica de Y puede escribirse como

$$\phi_Y(t) = E(\exp(it^TY)) = \prod_{i=1}^{p} \phi_{Y_i}(t_i)$$

por se $y_1,\ldots,y_p$ variables independientes. El valor $\phi_{Y_i}(t_i)$ corresponde a

$$E(\exp(it_iy_i)) = \int_{-\infty}^{\infty} \exp(it_iy_i)\,(2\pi)^{-1/2}\exp(-\tfrac{1}{2}\,y_i^2)\,dy$$

$$= (2\pi)^{-1/2}\exp(-\tfrac{1}{2}\,t_i^2)\int_{-\infty}^{\infty}\exp(-\tfrac{1}{2}\,(y_i-it_i)^2)\,dy_i$$

$$= \exp(-\tfrac{1}{2}\,t_i^2)$$

por lo que

$$\phi_X(t) = \prod_{i=1}^{p}\phi_{Y_i}(t_i) = \exp\{-\tfrac{1}{2}\,t^T t\}$$

de donde se obtiene que

$$\phi_X(t) = E\{\exp(it^TX)\} = E\{\exp(it^T[\Sigma^{1/2}Y+\mu])\}$$

$$= \exp(it^T\mu)\; E\{\exp(it^T\Sigma^{1/2}Y)\}$$

$$= \exp\{it^T\mu - \tfrac{1}{2}\,(t^T\Sigma^{1/2})(t^T\Sigma^{1/2})^T\}$$

$$= \exp\{it^T\mu - \tfrac{1}{2}\,t^T\Sigma\, t\}$$

∎

Teorema A - 4. Sea d_s un vector de $N \times 1$, y denotemos por

$$X^T = [\, X_1, X_2, \ldots, X_N \,]_{p \times N}$$

donde $X_i \sim N_p(\mu, \Sigma)$. Entonces $X^T d_s$ tiene una distribución normal con parámetros $M^T d_s$ y $d_s^T d_s \Sigma$.

Demostración.

La función característica de $X^T d_s$ corresponde a

$$\phi_{X^T d_s}(t) = E\{\exp(i t^T X^T d_s)\} = E\{\exp(i \sum_{i=1}^{N} t^T X_i d_{si})\}$$

Utilizando la independencia de las variables $X_1, \ldots, X_N$, y el Teorema A-3, el valor de $\phi_{X^T d_s}(t)$ puede escribirse como

$$\phi_{X^T d_s}(t) = \prod_{i=1}^{N} \phi_{X_i}(d_{si}\, t) = \prod_{i=1}^{N} \exp\{ i\, d_{si}\, t^T \mu - \frac{1}{2} d_{si}^2\, t^T \Sigma\, t \}$$

$$= \exp\{ i\, t^T \sum_{i=1}^{N} d_{si}\, \mu - \frac{1}{2} \sum_{i=1}^{N} d_{si}\ \ t^T \Sigma\, t \}$$

$$= \exp\{ i\, t^T M^T d_s - \frac{1}{2} t^T d_s^T d_s \Sigma\, t \}$$

de donde se deduce que $X^T d_s \sim N_p(M^T d_s, d_s^T d_s \Sigma)$.

∎

Definición A - 5. Se dice que la matriz $M_{p\times p}$ tiene distribución Wishart de dimensión p, con matriz de escala Σ y N grados de libertad, si puede ser escrita como $M = \sum_{i=1}^{N} X_i^T X_i$ donde X_i se distribuye de acuedo a una $N_p(\bar{0},\Sigma)$.

Teorema A - 6. Sea $X \sim N_p(\mu,\Sigma)$ y $Y = A X$, donde $A_{p\times p}$ es una matriz de constantes. La distribución de Y es normal multivariada de media $A\mu$ y matriz de covarianzas $A \Sigma A^T$.

Demostración. La función de densidad de Y puede obtenerse evaluando la función de densidad de X en la transformación inversa y multiplicando por el Jacobiano de la transformación. La función de densidad de Y esta dada por

$$f_Y(Y) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(A^{-1}Y-\mu)^T\Sigma^{-1}(A^{-1}Y-\mu)\right\} |A^{-1}|$$

$$= (2\pi)^{-p/2} |A \Sigma A^T|^{-1/2} \exp\left\{-\frac{1}{2}(Y-A\mu)'(A\Sigma A^T)^{-1}(Y-A\mu)\right\}$$

que corresponde a la densidad de una variable normal de dimensión p, de media $A\mu$ y matriz de covarianzas $A \Sigma A^T$.

∎

Teorema A - 7. (Teorema del Límite Central) Sean $V_1, V_2, \ldots$ una sucesión de vectores aleatorios independientese identicamente distribuidos, provenientes de una población con media μ y matriz de covarianzas T , entonces $\frac{1}{\sqrt{N}} \sum_{\alpha=1}^{N} (V_\alpha - \mu)$ converge en distribución a una $N_p(\bar{0},\Sigma)$ cuando el valor de N tiende a infinito.

Demostración. Ver Anderson (1958 pág. 74).

B I B L I O G R A F I A

- Anderson T. W. (1958) An Introduction to Multivariate Statis - tical Analysis. Wiley, New York.

- Bartlett M. S. (1947) Multivariate Analysis. Journal of the - Royal Statistical Society. B,9, 176-197.

- Castell R. B. (1966) The Scree Test for the Number of Factors. Multivariate Behavioral Research 1, 245-276.

- Cramer H. (1946) Mathematical Methods of Statistics. Princeton University Press.

- Dietrich Morgenstern. (1944) Einführung in die Wahrscheinlich-keitrechnung und Mathematische Statistik. Springer Verlag.

- Feller W. . Introducción a la Teoría de Probabilidades y sus - aplicaciones. Vol 2, Segunda edición. Editorial Limusa.

- Hoel Paul, Port, Sidney C. Introduction to probability theory. Houghton Mifflin Company. International Dolfin Edition, Boston.

- Hogg R. V. & Craig A. T. (1978) Introduction to Mathematical Statistics. Macmillan Publishing Company, Ney York.

- Johnson N.L. & Kotz S. (1972) Distributions in Statistics : Continuous Multivariate Distributions. Wiley, New York.

- Mardia K. V., Kent J. T. & Bibby J. M. (1979) Multivariate Analysis. Academic Press, London.

- Mendoza Blanco Rodolfo. La distribución Normal Multivariada y sus relaciones con otras distribuciones. Tésis para Título de Actuario (1987).