



03061
Tes.
5

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Unidad Académica de los Ciclos Profesional y de Posgrado del CCH
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas

ESTUDIO COMPARATIVO DE PRUEBAS DE BONDAD DE AJUSTE PARA LA DISTRIBUCION NORMAL MULTIVARIADA

T E S I S

Que para obtener el grado de:
Maestra en Estadística e
Investigación de Operaciones
Presenta la Actuaría
Leticia Eugenia Gracia Medrano
Valdelamar





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Indice

PROLOGO	11
I MOTIVACION	1
I.1 Importancia del problema	1
II METODOS REVISADOS	4
II.1 Criterios de Elección	4
II.2 Lista de métodos revisados	5
II.2.1 El método de Mardia	5
II.2.2 Las estadísticas de Mardia y Foster	6
II.2.3 El método de Cox y Small	9
II.2.4 El método de Radios y Angulos	11
II.2.5 Los métodos de Malkovich y Afifi	13
II.2.6 El método de Henseler, Mehrotra y Michalek	16
II.2.7 El método de Csörgő	17
III EL METODO PROPUESTO	20
III.1 La idea general del método	20
III.2 Descripción detallada del método.	24
III.3 Modificación al método de Rincón-Gallardo et al.	30
IV ESTUDIO COMPARATIVO	33
IV.1 Simulación de distribuciones multivariadas no-normales	33
IV.2 Resultados del estudio comparativo	36
CONCLUSIONES	43
HISTOGRAMAS	45
REFERENCIAS	54

PROLOGO

Los estadísticos inician generalmente su análisis proponiendo una distribución para sus observaciones. A través de los años han surgido muchos métodos para probar si la distribución propuesta es adecuada. El estudio de estos métodos se ha llamado *Bondad de Ajuste*. Este trabajo versa sobre pruebas de bondad de ajuste para la Normal Multivariada (NMV) y es en el primer capítulo donde se da una motivación del problema.

En el capítulo segundo se hace una presentación de los criterios utilizados en la elección de los métodos para el estudio comparativo de pruebas de bondad de ajuste para la NMV. Luego de una amplia revisión, se eligieron siete métodos para ser incluidos en este trabajo, estos son: el método de Mardia, el de Mardia y Foster, el de Cox y Small, el método de radios y ángulos, el método de Malkovich y Afifi, el de Henseler et al. y el de Csörgő. Fueron los primeros cinco de estos métodos los que satisficieron los criterios señalados al inicio del capítulo, quedando sólo estos cinco dentro del estudio de comparación.

En el tercer capítulo se expone un método *nuevo*, exhibiendo su sustento estadístico, se hace también una descripción detallada de su aplicación y se muestra la relación que este método guarda con el trabajo previo de Rincón et al. (1979).

Al inicio del capítulo cuarto se presentan las distribuciones alternativas utilizadas en el estudio comparativo de potencias y a continuación, se exhiben los resultados de este estudio que comparó los méritos de los distintos métodos revisados, incluyendo el método propuesto en este trabajo.

Finalmente se muestran las conclusiones de este estudio. La tesis incluye un apéndice con histogramas de las distribuciones alternativas utilizadas en el estudio comparativo.

Quiero agradecer sinceramente al Dr. Federico O'Reilly por haberme iniciado en el estudio de esta área, así como por su entusiasta y dedicada dirección de este trabajo.

Quiero dar las gracias también al M. en C. Raymundo Peralta por su ayuda en el desarrollo de programas en la VAX 11-750, a María del Carmen Quintanar por su ayuda en la edición de tablas y al Dr. Arturo Olvera por su ayuda en la impresión de esta tesis.

Agradezco al Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (IIMAS) por las facilidades brindadas para el desarrollo de este trabajo.

Capítulo I

MOTIVACION

I.1 Importancia del problema

Dentro del análisis estadístico existen métodos que suponen normalidad multivariada (NMV). Algunos de éstos no son robustos ante la violación de este supuesto, razón por la cual es deseable contar con algún método previo que permita probar la validez de este supuesto.

En el caso univariado, existen varios métodos para probar normalidad, como lo son los métodos gráficos, que son menos formales que las técnicas numéricas y entre éstos están la graficación de la función de distribución empírica, los histogramas y la graficación en papel normal (probability plot). Dentro de las técnicas numéricas se encuentran:

a) La prueba de la Ji-cuadrada; que particiona el rango de la variable X en M celdas, E_1, \dots, E_M . Si N_j es el número observado de X 's en estas celdas, entonces las N_j tienen distribución multinomial con parámetros n y $p_i = \int_{E_i} d\Phi\left(\frac{x-\mu}{\sigma}\right)$, donde Φ es la función de distribución de la normal estándar. Las p_i 's deben estimarse por un método adecuado. La Ji-cuadrada de Pearson está dada por $X^2 = \sum_{i=1}^M (N_i - np_i)^2 / np_i$, que se distribuye aproximadamente como una χ_{M-3}^2 .

b) Las pruebas basadas en las estadísticas EDF¹. La función de distribución empírica se calcula como primer paso, sirviendo como estimador de la función de distribución verdadera. Las estadísticas EDF son medidas de discrepancia entre la distribución empírica y la distribución hipotética (estimada de alguna manera). Dentro de estas se encuentran la de Lilliefors, la de Kolmogorov, la de Kuiper, la de Cramér-von Mises, la de Watson y la de Anderson-Darling.

c) Las estadísticas basadas en los momentos; entre ellas están las medidas de asimetría y kurtosis, la de D'Agostino-Pearson, la de Bowman-Shenton, etc.

d) Las estadísticas llamadas de regresión y correlación como la de Shapiro-Wilk, Shapiro-Francia y la de D'Agostino, por mencionar algunas.

En cuanto a la comparación de estas pruebas se han hecho ya varios estudios. Una lista de éstos aparece en D'Agostino y Stephens (1986, sec. 9.4) y las recomendaciones en cuanto al uso de estas estadísticas aparecen también en la misma referencia.

Para el caso multivariado es muy poco lo que se ha hecho. Por ejemplo, en el año de 1986 en el compendio más completo a esa fecha, D'Agostino y Stephens (1986), los autores presentan varios métodos pero no señalan cuál de todos conviene utilizar.

D'Agostino (1986) comenta de las pruebas para la NMV

"Es inadecuado en este momento dar recomendaciones detalladas. Se requiere bastante más investigación. Todos los procedimientos aquí incluidos anteriormente tienen mérito."

El comentario de D'Agostino es muy acertado ya que existen pocos métodos propuestos y entre éstos, pocas comparaciones se han hecho.

Aparentemente lo que uno esperaría hacer para resolver el problema de pro-

¹por sus siglas en inglés por basarse en la función empírica de distribución, Empirical Distribution Function

bar la NMV es buscar en la literatura y obtener un método, el cual debería ser claro en la idea de su desarrollo, contar con las fórmulas explícitas con las que se se construye la estadística de prueba, contar con tablas contra las cuales se compara el valor de la estadística obtenida, y finalmente que el método haya sido comparado con otros existentes para conocer algo acerca de su potencia.

En la literatura se encontraron varios métodos, pero casi ninguno cumplió con los cuatro puntos mencionados anteriormente. Esta situación motivó el hacer una revisión más amplia de los métodos existentes, y hacer una selección de éstos para poder llevar a cabo un estudio comparativo entre ellos. Asimismo, se propuso un nuevo método que se incluyó en el mencionado estudio para evaluar, a través de sus potencias a todos los métodos. Con lo anterior se pretende tener una idea más clara de como enfrentar el problema de probar normalidad en el caso multivariado.

Capítulo II

MÉTODOS REVISADOS

II.1 Criterios de Elección

Después de haber revisado varios artículos que presentan pruebas para la NMV, el siguiente paso fue seleccionar los métodos que serían estudiados y comparados con el nuevo método propuesto. Los diferentes métodos revisados presentaron distintas ventajas y desventajas. Destacan dentro de estas últimas, fórmulas complicadas, distribución de la estadística de prueba no especificada para n chico, corto alcance del método, es decir que sólo es válido para $k = 2$, o bien su difícil generalización para el caso $k > 2$, donde k es el número de variables consideradas. Así también algunos tienen falta de invarianza ya sea al orden de las componentes o al orden de las observaciones, esta última falta de invarianza resulta muy importante, ya que equivale a violar el principio de verosimilitud de Fisher. De los métodos revisados, se seleccionaron aquéllos que fueran invariantes al orden de las observaciones y que no fueran excesivamente complicados de implantarse en un código de programación.

II.2 Lista de métodos revisados

II.2.1 El método de Mardia

Mardia (1974) propone una prueba para la NMV que se basa en las nociones de asimetría y la kurtosis multivariadas. En ese artículo, define la medida de asimetría multivariada poblacional como:

$$\beta_{1,k} = E \{ [(\underline{X} - \underline{\mu})' \Sigma^{-1} (\underline{Y} - \underline{\mu})]^3 \}$$

y la contraparte muestral como:

$$b_{1,k} = \frac{1}{n^2} \sum_{i,j=1}^n [(\underline{X}_i - \bar{\underline{X}})' S^{-1} (\underline{X}_j - \bar{\underline{X}})]^3$$

Define la kurtosis multivariada poblacional como:

$$\beta_{2,k} = E \{ [(\underline{X} - \underline{\mu})' \Sigma^{-1} (\underline{X} - \underline{\mu})]^2 \}$$

y la contraparte muestral como:

$$b_{2,k} = \frac{1}{n} \sum_{i=1}^n [(\underline{X}_i - \bar{\underline{X}})' S^{-1} (\underline{X}_i - \bar{\underline{X}})]^2,$$

donde Σ es la matriz poblacional de varianzas y covarianzas, S es la matriz muestral de varianzas y covarianzas, $\underline{\mu}$ es el vector de medias poblacionales y $\bar{\underline{X}}$ es el vector muestral de medias.

Se encuentra que

$$E(b_{1,k}) = \frac{k(k+2)}{(n+1)(n+3)} [(n+1)(k+1) - 6] \quad (\text{II.1})$$

$$E(b_{2,k}) = \frac{k(k+2)(n-1)}{(n+1)} \quad (\text{II.2})$$

$$\text{Var}(b_{2,k}) = \frac{8k(k+2)(n-3)}{(n+1)^2(n+3)(n+5)} (n-k-1)(n-k+1) \quad (\text{II.3})$$

Mardia obtiene las siguientes distribuciones asintóticas bajo la hipótesis nula de NMV

$$A = \frac{nb_{1,k}}{6} \sim \chi_f^2,$$

donde $f = \frac{k(k+1)(k+2)}{6}$ (II.4)

Usando el teorema central del límite y las ecuaciones II.2 II.3 se tiene que:

$$B' = \frac{[(n+1)b_{2,k} - k(k+2)(n-1)] \sqrt{(n+3)(n+5)}}{\sqrt{8k(k+2)(n-3)(n-k-1)(n-k+1)}}$$

se distribuye asintóticamente como una $\mathcal{N}(0,1)$.

Mardia muestra también que

$$B = \frac{b_{2,k} - k(k+2)}{\sqrt{8k(k+2)/n}}$$

se distribuye asintóticamente como una $\mathcal{N}(0,1)$

Mardia y Zemroch (1975) dan los algoritmos para calcular $b_{1,k}$ y $b_{2,k}$, mismos que fueron usados para implantar este método.

La estadística, aquí propuesta, finalmente utilizada fue la resultante de usar A y B al nivel $\frac{\alpha}{2}$ y rechazar la hipótesis nula si al menos una de las dos excede el valor de tablas, (la distribución para A y B fue "calculada" por el método de Monte Carlo con 5000 simulaciones), esto es, la desigualdad de Bonferroni fue utilizada para producir una prueba conservadora.

Esta estadística será llamada de aquí en adelante como $B1B2$.

II.2.2 Las estadísticas de Mardia y Foster

Mardia y Foster (1983) presentan varias pruebas de tipo general ("ómnibus") para la NMV, basadas en $b_{1,k}$ y $b_{2,k}$, las dos medidas de asimetría y kurtosis vistas anteriormente.

Las pruebas que Mardia y Foster presentan, consisten en transformar a $b_{1,k}$ y a $b_{2,k}$ en variables normales estandarizadas y luego en construir una estadística que tome en cuenta la correlación entre estas dos variables; esto es, a través de una forma cuadrática que resulta de aproximar con una Ji-cuadrada a un vector asintóticamente normal bivariado con $b_{1,k}, b_{2,k}$. Las pruebas elegidas para el estudio de comparación son las que en el artículo aparecen con el nombre de S_N^2, S_W^2, C_N^2 y C_W^2 . Utilizan dos transformaciones para normalizar a $b_{1,k}$ que son:

$$U(b_{1,k}) = \frac{b_{1,k} - 6f/n}{6\sqrt{2f/n^2}} \quad y$$

$$W(b_{1,k}) = \frac{1}{6}\sqrt{2f} \left[6\left(\frac{4}{3}nf^2b_{1,k}\right)^{\frac{1}{2}} - 18f + 4 \right],$$

donde f está definido en II.4.

Y otras dos para transformar a $b_{2,k}$:

$$U(b_{2,k}) = \frac{b_{2,k} - k(k+2)(n-1)/(n+1)}{\sqrt{8k(k+2)/n}} \quad y$$

$$W(b_{2,k}) = 3\sqrt{\frac{f_1}{2}} \left[1 - \frac{2}{9f_1} - \left\{ \frac{(1 - \frac{2}{f_1})}{1 + x\sqrt{\frac{2}{f_1-4}}} \right\}^{\frac{1}{2}} \right]$$

donde

$$x = \frac{b_{2,k} - E[b_{2,k}]}{\sqrt{\text{Var}[b_{2,k}]}}$$

y

$$f_1 = 6 + \sqrt{\frac{8k(k+2)}{(k+8)^2}} \sqrt{n}$$

$$\left[\sqrt{\frac{k(k+2)}{2}} \frac{\sqrt{n}}{(k+8)} + \sqrt{1 + \frac{k(k+2)n}{2(k+8)^2}} \right] \quad (\text{II.5})$$

Como se dijo anteriormente, estas transformaciones normalizan a $b_{1,k}$ y a $b_{2,k}$. Las estadísticas que no toman en cuenta la correlación entre las variables transformadas en la construcción de la forma cuadrática son:

$$S_N^2 = U^2(b_{1,k}) + U^2(b_{2,k}),$$

$$S_W^2 = W^2(b_{1,k}) + W^2(b_{2,k}).$$

Y las estadísticas que sí toman en cuenta la correlación entre las variables transformadas son:

$$C_N^2 = \underline{b}'V^{-1}\underline{b},$$

donde

$$\underline{b}' = (b_{1,k} - 6f/n, b_{2,k} - k(k+2)(n-1)/(n+1))$$

y

$$V = \begin{bmatrix} 72fn^{-2} & 12k(8k^2 - 13k + 23)n^{-2} \\ 12k(8k^2 - 13k + 23)n^{-2} & 8k(k+2)n^{-1} \end{bmatrix}.$$

Y

$$C_W^2 = (W(b_{1,k}), W(b_{2,k}))W^{-1}(W(b_{1,k}), W(b_{2,k})),$$

donde W es una matriz simétrica de 2×2 con unos en la diagonal y la siguiente expresión fuera de la diagonal:

$$\text{Cov}(W(b_{1,k}), W(b_{2,k})) = 3\sqrt{\frac{f_1}{2}} \frac{1}{\sqrt{72f_1}} \left\{ -\frac{40}{9} \left(1 - \frac{2}{f_1}\right) \frac{1}{f_1 - 4} + \frac{n}{3\sigma} \left(1 - \frac{2}{f_1}\right)^{\frac{1}{2}} \right. \\ \left. \sqrt{2/(f_1 - 4)} \text{Cov}(b_{1,k}, b_{2,k}) \right\}$$

donde f_1 está definido en II.5 y $\sigma^2 = \text{Var}(b_{2,k})$ y la $\text{Cov}(b_{1,k}, b_{2,k})$ está dada por la siguiente expresión:

$$\text{Cov}(b_{1,k}, b_{2,k}) = \frac{12k(8k^2 - 13k + 23)}{n^2}.$$

En el caso de C_N^2 , se observó que la matriz V es positiva definida para $k = 2$ sólo cuando $n \geq 26$; de aquí que la distribución asintótica que presentan los autores sólo puede utilizarse como aproximación para $n \geq 26$; pero más aún, la estadística C_N^2 sólo puede evaluarse si $n \geq 26$.

Mardia y Foster señalan que debido a la normalidad aproximada de las variables transformadas, las estadísticas anteriores tienen una distribución próxima a una χ_2^2 bajo el supuesto de NMV. Sin embargo, se hizo un estudio de simulación para $k = 2$ y para $n = 40$ con 5000 repeticiones y las distribuciones de estas estadísticas no resultaron muy cercanas a la distribución asintótica señalada. En el estudio de comparación las pruebas se llevaron a cabo utilizando la distribución obtenida vía Monte Carlo. La notación utilizada de aquí en adelante para estas estadísticas es la siguiente:

$$SU = S_N^2, SW = S_W^2, CU = C_N^2, y CW = C_W^2$$

II.2.3 El método de Cox y Small

Cox y Small en (1978) presentan una estadística de prueba para el caso $k = 2$ basada en la linealidad de una componente en la otra, ya que el hecho de que las esperanzas condicionales sean lineales es una caracterización de la normalidad.

Llamando X_1 a la primer componente de las observaciones, y X_2 a la segunda, hacen las regresiones siguientes:

$$X_1 = a + bX_2 + cX_2^2 + e_1 \quad y \quad (II.6)$$

$$X_2 = a' + b'X_1 + c'X_1^2 + e_2 \quad (II.7)$$

y calculan las estadísticas t de student para probar que el coeficiente que multiplica a los términos cuadráticos es diferente de cero, siendo Q_{12} para el modelo II.6 y Q_{21} para el modelo II.7. Cabe aclarar que así como se postula una parte no lineal como cuadrática, podría haberse escogido cualquier otra función no lineal; esto aparenta ser una debilidad del procedimiento.

La distribución conjunta de (Q_{12}, Q_{21}) es complicada, aunque las marginales sean relativamente simples. Bajo la hipótesis nula (Q_{12}, Q_{21}) se distribuye asintóticamente como una normal bivariada con medias 0, varianzas 1 y con $Corr(Q_{12}, Q_{21}) = \rho(2 - 3\rho^2)$ donde $\rho = Corr(X_1, X_2)$. De lo anterior, la estadística propuesta por ellos es:

$$[Q_{2,1}, Q_{1,2}] \begin{bmatrix} 1 & r_{1,2}(2 - 3r_{1,2}^2) \\ r_{1,2}(2 - 3r_{1,2}^2) & 1 \end{bmatrix}^{-1} \begin{bmatrix} Q_{2,1} \\ Q_{1,2} \end{bmatrix},$$

donde $r_{1,2}$ es el coeficiente de correlación muestral de X_1 y X_2 , ya que ρ es desconocido. La forma cuadrática se distribuye asintóticamente como una χ_2^2 .

Como se hizo para las estadísticas de las secciones anteriores, se calculó su distribución por Monte Carlo para muestras tamaño $n = 40$, con 5000 simulaciones.

En ese artículo los autores no señalan qué cola de la distribución hay que utilizar para rechazar la hipótesis nula, y se encontró al hacer una simulación que en ocasiones hay que usar la cola izquierda para rechazar y en otras la cola derecha; al no haber una recomendación de los autores respecto a esto, se decidió considerar tres pruebas, usando las siguientes opciones: rechazar para valores grandes (cola derecha), rechazar para valores pequeños (cola izquierda) y rechazar para valores pequeños o grandes (dos colas). A las pruebas se les hará referencia de aquí en adelante como CSD, CSI y CSDI respectivamente.

Los autores muestran cómo podría generalizarse esta idea para $k > 2$ y lo que obtienen es una serie de estadísticas semejantes a las Q 's anteriores; el número de éstas crece considerablemente con k . Los autores sugieren hacer gráficas de estas estadísticas ordenadas contra los valores esperados de las estadísticas de orden de la normal, (desviaciones de una línea recta indicarán alejamiento de la NMV). Lo anterior supone que se cuenta con tamaños de muestra suficientemente grandes como para considerar que las estadísticas tienen una distribución normal. Esta generalización es conceptualmente sencilla, sin embargo la apre-

ciación de las gráficas no deja de ser un un método subjetivo que hace difícil asignarle una potencia a una prueba hecha de esta manera. Estos procedimientos no resultan ser invariantes a transformaciones lineales no singulares.

Cox y Small sugieren en el mismo artículo, un método que es invariante a transformaciones lineales no singulares. El procedimiento también esta basado en la idea de la linealidad de una componente en la otra. Este enfoque trata de encontrar dos pares de variables tales que sean combinaciones lineales de las variables originales y que una tenga la máxima curvatura en la regresión de la otra. La curvatura alcanzada es la estadística de prueba. Llevar a cabo este procedimiento resulta bastante complicado, requiere de maximizaciones y aún el encontrar los puntos iniciales resulta complicado. Por los motivos dados anteriormente sólo CSD, CSI y CSDI aparecen en el estudio de comparación realizado.

II.2.4 El método de Radios y Angulos

En el artículo de Andrews et. al. (1973), se presentan varios métodos para detectar fallas al supuesto de NMV. Entre ellos está el método de descomposición en radios y ángulos, para el caso bivariado. Este consiste en calcular los residuales escalados

$$\underline{Z}_i = S^{-\frac{1}{2}}(\underline{X}_i - \underline{\bar{X}}), i = 1, \dots, n,$$

donde \underline{X}_i denota el vector de observaciones, $\underline{\bar{X}}$ y S son el vector muestral de medias y la matriz muestral de covarianzas respectivamente. Bajo el supuesto de NMV los cuadrados de las normas de los \underline{Z}_i ,

$$R_i = \underline{Z}_i' \underline{Z}_i = (\underline{X}_i - \underline{\bar{X}})' S^{-1} (\underline{X}_i - \underline{\bar{X}}),$$

se distribuyen aproximadamente como χ^2 con 2 grados de libertad . También en el caso bivariado el ángulo θ_i que \underline{Z}_i forma con el eje de las abscisas se distribuye

aproximadamente como $U(0, 2\pi)$. Los residuales escalados y los ángulos para n grande resultan ser casi independientes.

Los procedimientos propuestos son técnicas gráficas, sin embargo se hizo una variación para poder introducir este procedimiento al estudio de comparación. Este procedimiento consiste en probar que los R_i provengan de una χ^2_2 y que los $\theta_i/2\pi$ provengan de una $U(0, 1)$, en forma simultánea. Para hacer esto se utilizó la desigualdad de Bonferroni y la estadística A^2 de Anderson-Darling evaluada para los R_i y para los θ_i por separado.

La función de distribución de una Ji-cuadrada con 2 grados de libertad es

$$F(x) = 1 - e^{-\frac{1}{2}x}, \quad x > 0$$

entonces a cada R_i^2 se le mapea con la función anterior y se obtienen las correspondientes U_i , es decir

$$U_i = 1 - e^{-\frac{1}{2}R_i^2} \quad \text{para } i = 1, \dots, n.$$

Después de hecho lo anterior se ordenan las U_i y se calcula la estadística de Anderson-Darling.

El caso para las θ_i es más sencillo ya que

$$F(x) = x, \quad x \in [0, 1]$$

entonces sólo se ordenan las θ_i y se calcula la estadística de Anderson-Darling.

Debido a que las distribuciones de las θ_i y los R_i son asintóticas, no se pueden usar las tablas dadas por D'Agostino y Stephens (1986) (caso 0) para los dos valores resultantes de A^2 . Se procedió a calcular las distribuciones de éstas por Monte Carlo, únicamente para $n = 40$ y para $k = 2$.

A la prueba que sigue este procedimiento se le llamará de aquí en adelante como *RAN*.

II.2.5 Los métodos de Malkovich y Afifi

Las dos primeras pruebas que presentan Malkovich y Afifi (1973) están basadas en generalizaciones de las medidas univariadas de asimetría y kurtosis .

Definen a la distribución de \underline{X} con asimetría multivariada si

$$\beta_1(\underline{C}) = \frac{[E\{(\underline{C}'\underline{X} - \underline{C}'E(\underline{X}))^3\}]^2}{[Var(\underline{C}'\underline{X})]^3} > 0,$$

para algún vector \underline{C} . Y definen la asimetría poblacional multivariada como

$$\beta_1^* = \max_{\underline{C}} \beta_1(\underline{C}) > 0.$$

La distribución de \underline{X} tiene kurtosis multivariada si

$$[\beta_2(\underline{C})]^2 = \left[\frac{E\{(\underline{C}'\underline{X} - \underline{C}'E(\underline{X}))^4\}}{\{Var(\underline{C}'\underline{X})\}^2} \right]^2 > 9,$$

para algún vector \underline{C} .

Y en forma similar definen

$$(\beta_2^*)^2 = \max_{\underline{C}} [\beta_2(\underline{C}) - 3]^2$$

como la medida poblacional de kurtosis multivariada.

La hipótesis de no asimetría para $\underline{C}'\underline{X}$ es aceptada si $b_1(\underline{C}) \leq K_{b_1}$, donde K_{b_1} es una constante y $b_1(\underline{C})$ está dado por

$$b_1(\underline{C}) = n \frac{[\sum_{j=1}^n (Z_j - \bar{Z})^3]^2}{[\sum_{j=1}^n (Z_j - \bar{Z})^2]^3}$$

con $Z_j = \underline{C}'\underline{X}_j$.

Usando el principio de Roy, la hipótesis de que la distribución multivariada no sea asimétrica se acepta si

$$b_1^* = \max_{\underline{C}} b_1(\underline{C}) \leq K_{b_1}.$$

Similarmente la hipótesis de que la distribución multivariada no tiene kurtosis es aceptada si

$$(b_2^*)^2 = \max_{\underline{C}} [b_2(\underline{C}) - K]^2 \leq K_{b_2},$$

donde K y K_{b_2} son constantes y $b_2(\underline{C})$ está dado por

$$b_2 = \frac{n \sum (Z_j - \bar{Z})^4}{[\sum (Z_j - \bar{Z})^2]^2}$$

con $Z_j = \underline{C}' \underline{X}_j$.

Estas medidas resultan invariantes ante transformaciones de la forma

$$\underline{Y}_j = T \underline{X}_j + \underline{M}, \quad j = 1, \dots, n,$$

donde T es no singular y \underline{M} es cualquier vector de constantes. Entonces, sin pérdida de generalidad restringen a los vectores \underline{C} a aquellos que satisfacen $\underline{C}' \underline{C} = 1$. Para obtener estas estadísticas utilizan el método de multiplicadores de Lagrange.

Otra prueba que sugieren Malkovich y Afifi está basada en la generalización de la estadística univariada de Shapiro-Wilk. La hipótesis de NMV es aceptada si

$$\min_{\underline{C}} W(\underline{C}) \geq K_w,$$

donde

$$W(\underline{C}) = \frac{[\sum a_j (Z_{(j)} - \bar{Z})]^2}{\sum (Z_j - \bar{Z})^2} \quad \text{con } Z_j = \underline{C}' \underline{X}_j$$

y K_w es una constante. Calcular el mínimo anterior es difícil y en vez de eso, consideran que $W(\underline{C})$ tiene una cota inferior cuando \underline{C} satisface las condiciones

siguientes:

$$\underline{C}'(\underline{X}_1 - \underline{X}) = \frac{n-1}{na_1}, \quad \underline{C}'(\underline{X}_j - \underline{X}) = \frac{-1}{na_1} \text{ para } j = 2, \dots, n,$$

y encuentran que no hay \underline{C} que cumpla esas condiciones, por lo que toman una aproximación a \underline{C} , que minimiza

$$\left[\underline{C}'(\underline{X}_1 - \bar{\underline{X}}) - \frac{n-1}{na_1} \right]^2 + \sum_{j=2}^n \left[\underline{C}'(\underline{X}_j - \bar{\underline{X}}) + \frac{1}{na_1} \right]^2. \quad (\text{II.8})$$

El vector \underline{C} resultante es

$$\underline{C} = \frac{1}{a_1} A^{-1} (\underline{X}_1 - \bar{\underline{X}}),$$

donde A está dado por

$$A = \sum_{j=1}^n (\underline{X}_j - \bar{\underline{X}})(\underline{X}_j - \bar{\underline{X}})'$$

Como el vector \underline{X}_1 en II.8 pudo ser cualquier $\underline{X}_j, j = 1, \dots, n$, escogen \underline{X}_1 de manera que el denominador de $W(\underline{C})$ sea maximizado. Los pasos para calcular esta estadística son los siguientes:

i. Encontrar \underline{X}_m , la observación para la cual

$$(\underline{X}_m - \bar{\underline{X}})' A^{-1} (\underline{X}_m - \bar{\underline{X}}) = \max_{1 \leq j \leq n} \{ (\underline{X}_j - \bar{\underline{X}})' A^{-1} (\underline{X}_j - \bar{\underline{X}}) \}.$$

ii. Ordenar las variables

$$U_j = (\underline{X}_m - \bar{\underline{X}})' A^{-1} (\underline{X}_j - \bar{\underline{X}}), \quad j = 1, \dots, n,$$

y denotarlas como $U_{(1)}, \dots, U_{(n)}$.

iii. Calcular la estadística como

$$W^* = \frac{[\sum a_j U_{(j)}]^2}{(\underline{X}_m - \bar{\underline{X}})' A^{-1} (\underline{X}_m - \bar{\underline{X}})}$$

donde las a_j son las constantes tabuladas por Shapiro-Wilk (1965).

Los autores del artículo señalan que las a_j no dependen del número k de componentes. Valores de W^* pequeños indican la no normalidad multivariada. La distribución exacta de las estadísticas bajo la hipótesis nula no se conoce.

Debido a la facilidad con que puede programarse W^* , esta estadística quedó incluida dentro del estudio de comparación, de nuevo calculando su distribución por Monte Carlo para $k = 2$ y $n = 40$ con 5000 simulaciones; mientras que b_1^* y b_2^* no se incluyeron, debido a que su programación es complicada.

II.2.6 El método de Henseler, Mehrotra y Michalek

El procedimiento que estos autores proponen (Henseler et al. 1977) reduce el problema de probar la hipótesis *compuesta multivariada* de normalidad en probar k hipótesis *simples* de normalidad *univariada*. En el artículo demuestran un teorema que permite reducir el problema de probar que una distribución es una $\mathcal{N}(\underline{\mu}, \Sigma)$ en probar que cierta transformación de los datos proviene de una $\mathcal{N}(0, \Sigma)$, con esto se pierde un vector de observaciones al usar este resultado para eliminar $\underline{\mu}$. En seguida presentan con detalle como llevar la prueba de NMV para el caso $k = 2$. La prueba usa el teorema demostrado y caracterizaciones de la distribución normal. Demuestran que para probar que la distribución es una $\mathcal{N}(0, \Sigma)$, es equivalente el probar que la segunda componente es una $\mathcal{N}(0, \sigma_{22})$ y que una nueva transformación de los datos que allí proponen, proviene de una $\mathcal{N}(0, \sigma^2)$. Con esta transformación se pierde otra observación. Para probar esas hipótesis univariadas de normalidad llevan a cabo otras transformaciones que reducen el problema a probar uniformidad de una sola muestra. En este proceso se pierden otras dos observaciones. En total en todo el proceso se pierden cinco observaciones; el mismo número de parámetros eliminados ($\mu_1, \mu_2, \sigma_{11}, \sigma_{22}, \sigma_{12}$). Finalmente presentan como se haría esta prueba para el caso $k > 2$ que se resumiría en lo siguiente:

- i. Eliminar el parámetro de localización.

- ii. Probar normalidad univariada de la última componente.
- iii. Probar la normalidad de la distribución condicional de las primeras $k - 1$ componentes dada la última.

Lo anterior reduce el problema a uno equivalente pero con tamaño de muestra $n - 2$ para una normal $(k - 1)$ -variada con media 0. El procedimiento anterior se repite hasta llegar a probar la normalidad bivariada. Después de estas reducciones se llega a probar la normalidad univariada de k muestras independientes de tamaños $n - 1, n - 2, \dots, n - k$. En total se pierden $k(k + 1)/2 + k$ observaciones, que es el mismo número de parámetros eliminados.

Los autores no presentan estudio alguno acerca de la potencia de esta prueba.

Esta prueba tiene un gran inconveniente, que es la no invarianza al orden de las observaciones, esta característica hace que el resultado obtenido por dos investigadores utilizando la misma muestra pero en distinto orden pueda ser diferente, por esta razón esta prueba se ha dejado fuera del estudio de comparación, ya que como se ha mencionado, este tipo de invarianza viola el principio de verosimilitud de Fisher.

II.2.7 El método de Csörgö

El enfoque que da Csörgö está basado en el comportamiento asintótico de la función característica multivariada estudentizada empírica y en una extensión multivariada del enfoque de Murota y Takeuchi (1981) para probar la normalidad univariada.

La función característica estudentizada empírica está definida como:

$$C_n(S_n^{-1/2}\underline{t}) = n^{-1} \sum_{j=1}^n \exp(i\langle S_n^{-1/2}\underline{t}, \underline{X}_{(j)} \rangle)$$

donde S_n es la matriz de covarianza muestral y $\langle \underline{t}, \underline{v} \rangle = \sum_{i=1}^k t_i v_i$ es el producto

interno de \underline{t} y \underline{v} .

El módulo $|C_n(S_n^{-1/2}\underline{t})|^2$ es invariante ante transformaciones lineales no singulares, por lo que se supone $E(\underline{X}) = 0$ y $\Sigma = I$.

En el artículo se demuestra que si $\underline{X} = (X_1, \dots, X_k)$ se distribuye como $\mathcal{N}(\underline{\mu}, \Sigma)$ entonces el proceso estocástico

$$Z_n(\underline{t}) = n^{1/2} \left\{ |C_n(S_n^{-1/2}\underline{t})|^2 - e^{-(\underline{t}, \underline{t})} \right\}$$

converge a un determinado proceso Gaussiano $Z(\underline{t})$. Como estadística de prueba propone la funcional siguiente:

$$M_n^{(k)}(T) = \sup_{\underline{t} \in [-T, T]^k} |Z_n(\underline{t})|$$

donde T es un número positivo. Encuentran que bajo la hipótesis de NMV

$$\lim_{n \rightarrow \infty} P\{M_n^{(k)}(T) \geq y\} = P\left[\sup_{\underline{t} \in [-T, T]^k} |Z(\underline{t})| \geq y \right].$$

Esta distribución no se conoce, y sólo se da una cota superior para los desvíos. Presenta una tabla con los percentiles asintóticos de la cota superior de $M_n^{(k)}$ para $k = 1, \dots, 6$. y las fórmulas para calcular $M_n^{(1)}$ y $M_n^{(2)}$.

Para darse cuenta de qué tan cercana está la cota superior de $M_n^{(1)}$, Csörgő hace un estudio de simulación y muestra los percentiles empíricos de $M_n^{(1)}$ para $\alpha = 0.1, 0.05, 0.01$ calculados para muestras de tamaño 50 y 100. Al comparar éstos con los valores de la cota superior, se ve que no está tan cerca de $M_n^{(1)}$. El autor da una cota superior más cercana a los percentiles obtenidos usando la desigualdad de Fernique y la de Borell, así como de la mediana empírica obtenida en el estudio de simulación.

Para los casos con $k > 2$ el autor no presenta ningún estudio de simulación, pero podría esperarse que el alejamiento de la cota superior para la probabilidad

deseada y la de valores grandes de $M_n^{(k)}$ sea semejante o mayor; en tal caso, para llevar a cabo la prueba sería conveniente como en las subsecciones anteriores calcular los percentiles empíricos de $M_n^{(k)}$ ($k > 2$). Debido a la complejidad que esto implica se decidió no incluir esta prueba en el estudio de comparación.

Capítulo III

EL METODO PROPUESTO

III.1 La idea general del método

La idea del método propuesto consiste en hacer transformaciones que conviertan a los vectores $\underline{X}'_i = (X_{i1}, X_{i2}, \dots, X_{ik})$, $i = 1, \dots, n$, en cantidades univariadas $U_{11}, U_{12}, \dots, U_{n1}, \dots, U_{nk}$ que bajo la hipótesis nula se distribuyen "uniformemente" en el intervalo unitario. Una vez hecho esto, el problema se convierte en probar la uniformidad de una muestra univariada de tamaño nk .

Antes de entrar de lleno a los detalles de cómo se llevaría a cabo la prueba de la hipótesis nula compuesta:

H_0 : F es una distribución normal *multivariada* $\mathcal{N}(\underline{\mu}, \Sigma)$ con $\underline{\mu}$ y Σ ambas *desconocidas*,

se tratará de ilustrar la idea del método para un caso mucho más sencillo; es decir, cuando la hipótesis nula es la siguiente:

H_0 : F es una normal *univariada* $\mathcal{N}(\mu, \sigma^2)$ con μ y σ^2 *conocidas*.

En el caso univariado, para construir la estadística de prueba, se tienen n

variables aleatorias X_1, \dots, X_n . Es ampliamente conocido que las variables

$$U_i = \Phi \left(\frac{X_i - \mu}{\sigma} \right)$$

son v.a.i.i.d $U(0,1)$ bajo el supuesto de H_0 , donde Φ es la función de distribución de la normal $\mathcal{N}(0,1)$;

La prueba consiste entonces en probar que las U_i provienen de una $U(0,1)$. Existen varias estadísticas en la literatura para probar esto, entre ellas se encuentran las EDF como son la Cramér-von Mises, Watson, Kolmogorov-Smirnov, la A^2 de Anderson-Darling, etc., sin embargo esta última resulta ser más potente que las otras según se menciona en D'Agostino y Stephens (1986, pág. 110). La estadística A^2 mide la distancia, utilizando cierta métrica, entre la distribución empírica $F_n(x)$ y la distribución que se está probando.

La estadística A^2 para probar una distribución conocida F , continua, con base en una muestra tamaño n es:

$$A_n^2 = n \int_{\mathbb{R}} \frac{[F_n(x) - F(x)]^2}{F(x)(1 - F(x))} dF(x),$$

y una fórmula alternativa para su cómputo es:

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\log(U_{(i)}) + \log(1 - U_{(n-i+1)})],$$

donde las $U_{(i)}$ son las transformadas $U_i = F(X_i)$ ordenadas. La distribución de A_n^2 es conocida y aparece en las tablas (caso 0) de D'Agostino y Stephens (1986, pág. 105).

Para el caso aún univariado, en que μ y σ son desconocidas, en vez de calcular

$$U_i = \Phi \left(\frac{X_i - \mu}{\sigma} \right),$$

se calcula

$$\hat{U}_i = \Phi \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}} \right)$$

donde $\hat{\mu}$ y $\hat{\sigma}$ son estimadores de μ y σ , usualmente los máximo verosímiles. Es importante señalar aquí que las \hat{U}_i ya no son independientes; esto afecta a la distribución de A_n^2 que aparece en las tablas (caso 3) de D'Agostino y Stephens (1986, pág. 127).

Ahora, en el caso en que la hipótesis nula se establece como:

$H_0: F$ es una $\mathcal{N}(\underline{\mu}, \Sigma)$ k -variada con $\underline{\mu}$ y Σ conocidas,

se requiere de un paso adicional por tratarse del caso multivariado. Este paso utiliza la transformación T de Rosenblatt (1952), que mapea un vector con una distribución absolutamente continua k -variada F conocida, en un vector con distribución uniforme en el hipercubo k -dimensional.

A continuación se describe esa transformación T .

Sea $\underline{X} = (X_1, \dots, X_k)$ un vector aleatorio con distribución $F(x_1, \dots, x_k)$ y sea $\underline{z} = (z_1, \dots, z_k) = T(\underline{x}) = T(x_1, \dots, x_k)$, dada por

$$z_1 = P[X_1 \leq x_1] = F_1(x_1)$$

$$z_2 = P[X_2 \leq x_2 \mid X_1 = x_1] = F_2(x_2 \mid x_1)$$

⋮

$$z_k = P[X_k \leq x_k \mid X_1 = x_1, \dots, X_{k-1} = x_{k-1}] = F_k(x_k \mid x_1, \dots, x_{k-1}).$$

Para el caso que se está tratando F corresponde a una $\mathcal{N}(\underline{\mu}, \Sigma)$, y T queda dada por

$$F_1(x_1) = \Phi\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}}\right)$$

$$F_2(x_2 \mid x_1) = \Phi\left(\frac{x_2 - \mu_2 + (\Sigma_{21}^{(2)}/\Sigma_{22}^{(2)})(x_1 - \mu_1)}{\sqrt{|\Sigma^{(2)}| / \Sigma_{22}^{(2)}}}\right)$$

⋮

$$F_k(x_k | x_1, \dots, x_{k-1}) = \Phi \left(\frac{x_k - \mu_k + \sum_{j=1}^{k-1} (\Sigma_{kj}^{(k)} / \Sigma_{kk}^{(k)}) (x_j - \mu_j)}{\sqrt{|\Sigma^{(k)}| / \Sigma_{kk}^{(k)}}} \right)$$

donde $\Sigma^{(r)} = [\sigma_{ij}]$, $i, j = 1, \dots, r \leq k$, y $\Sigma_{ij}^{(r)}$ es el cofactor de σ_{ij} en $\Sigma^{(r)}$.

Aplicando esta transformación T a cada vector \underline{X}_i , $i = 1, \dots, n$; esto es, calculando $T(\underline{X}_1), \dots, T(\underline{X}_n)$, obtenemos $n \times k$ cantidades univariadas aleatorias $U_{11}, \dots, U_{1n}, \dots, U_{n1}, \dots, U_{nk}$, que bajo la hipótesis nula de NMV se distribuyen uniformemente en el intervalo unitario. Construyendo finalmente con estas U_{ij} , $i = 1, \dots, n$ y $j = 1, \dots, k$, la estadística A^2 .

Regresando por un momento al caso univariado normal con parámetros desconocidos, se dijo que la $\hat{U}_i = \Phi\left(\frac{X_i - \hat{\mu}}{\hat{\sigma}}\right)$ eran utilizadas habiendo estimado los parámetros y que la distribución de A_n^2 estaba identificada.

En cierto modo $\Phi\left(\frac{\cdot - \hat{\mu}}{\hat{\sigma}}\right)$ es un "estimador" de la función de distribución de X_i . Existe sin embargo otro método que consiste en estimar directamente la distribución de X_i con el estimador Rao-Blackwell que se denotará aquí por \tilde{F} ; así pueden construirse $\tilde{U}_i = \tilde{F}(X_i)$ $i = 1, \dots, n$, y éstas sí resultan ser uniformes pero no son independientes.

El uso de \tilde{F} o de $\Phi\left(\frac{\cdot - \hat{\mu}}{\hat{\sigma}}\right)$, esto es, el basar las estadísticas de prueba en las \tilde{U}_i o las \hat{U}_i es asintóticamente equivalente como lo muestra Moore (1973). Con el objeto de no sobrecargar notacionalmente esta descripción, y dado que más adelante se detalla la \tilde{F} para el caso normal multivariado, no se exhibe la expresión de \tilde{F} para $k = 1$.

Para el caso de interés donde $\underline{\mu}$ y Σ son desconocidas, el método propuesto estima la función de distribución $\mathcal{N}(\underline{\mu}, \Sigma)$ por medio del estimador Rao-Blackwell y el procedimiento es el mismo que el mencionado anteriormente para parámetros conocidos, esto es con la transformación de Rosenblatt; pero aplicada al estimador Rao-Blackwell multivariado correspondiente. Al estimar la función de distribución, ya no se cuenta con la independencia entre $(\tilde{U}_{i1}, \dots, \tilde{U}_{ik})$

y $(\tilde{U}_{11}, \dots, \tilde{U}_{1k})$ (aunque \tilde{U}_{ir} y \tilde{U}_{is} si son independientes para $r \neq s$).

III.2 Descripción detallada del método.

A continuación se presentan en forma detallada los pasos a seguir con el método propuesto.

i. Estimar la función de distribución $\mathcal{N}(\underline{\mu}, \Sigma)$.

La función utilizada para estimar a la normal es la Rao-Blackwell, obtenida por Ghurye y Olkin (1969) (Sec. 3.4) y es la siguiente:

$$\tilde{F}(\underline{z}) = P[\underline{X} \leq \underline{z} | U_n] = \text{invt}^{(k)}(S^{-1}, \frac{n-1}{n}, n-k-1, \bar{\underline{X}})$$

donde

$$\begin{aligned} \bar{\underline{X}} &= \frac{1}{n} \sum_{i=1}^n \underline{X}_i = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_k \end{bmatrix}, \\ S_n &= S = \sum_{i=1}^n (\underline{X}_i - \bar{\underline{X}})(\underline{X}_i - \bar{\underline{X}})' = [s_{ij}], \\ U_n &= (\bar{\underline{X}}, S), \end{aligned}$$

En esta sección S no está dividida por n .

Un vector \underline{Z} se distribuye como $\text{invt}^{(k)}(P, Q, \nu, \underline{\eta})$ si su función de densidad es la siguiente:

$$f(\underline{z}) = \begin{cases} \frac{\Gamma(\frac{\nu+k}{2}) |P|^{\frac{1}{2}} Q^{-\frac{\nu+k-2}{2}} (Q - (\underline{z} - \underline{\eta})' P (\underline{z} - \underline{\eta}))^{-\frac{\nu-2}{2}}}{\pi^{\frac{k}{2}} \Gamma(\frac{\nu}{2})} & \text{si } Q - (\underline{z} - \underline{\eta})' P (\underline{z} - \underline{\eta}) > 0 \text{ y} \\ 0 & \text{en otro caso} \end{cases}$$

siendo P una matriz de precisión, Q un escalar, ν los grados de libertad y $\underline{\eta}$ un parámetro de localización. Esta notación se tomó de Press (1982).

ii. Aplicar la transformación de Rosenblatt a \tilde{F} .

Para encontrar las distribuciones marginales y condicionales de \tilde{F} y aplicar así la transformación de Rosenblatt se encontraron los resultados siguientes:

Si \underline{Z} se distribuye como $invt^{(k)}(P, Q, \nu, \eta)$ y particionando

$$\underline{Z} = \begin{bmatrix} \underline{Z}_1 \\ \underline{Z}_2 \end{bmatrix}, P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \text{ y } \underline{\eta} = \begin{bmatrix} \underline{\eta}_1 \\ \underline{\eta}_2 \end{bmatrix},$$

donde \underline{Z}_1 y $\underline{\eta}_1$ son vectores de dimensión $k_1 \times 1$ y \underline{Z}_2 y $\underline{\eta}_2$ son vectores de dimensión $k_2 \times 1$ con $k_1 + k_2 = k$, entonces la distribución *marginal* de \underline{Z}_1 es una $invt^{(k_1)}(P_{11.2}, Q, \nu + k_2, \underline{\eta}_1)$ y la distribución *condicional* de \underline{Z}_2 dado \underline{Z}_1 es una $invt^{(k_2)}(P_{22}, Q - (\underline{Z}_1 - \underline{\eta}_1)' P_{11.2} (\underline{Z}_1 - \underline{\eta}_1), \underline{\eta}_2 - P_{22}^{-1} P_{21} (\underline{Z}_1 - \underline{\eta}_1))$ donde $P_{11.2} = P_{11} - P_{12} P_{22}^{-1} P_{21}$.

Utilizando lo anterior se tiene que la distribución *marginal* \tilde{F}^1 , de la primera componente, es una

$$invt^{(1)}(s_{11}, \frac{n-1}{n}, n-2, \bar{x}_1),$$

y la distribución *condicional* $\tilde{F}^{2|1}$, de la segunda componente dada la primera, es una

$$invt^{(1)}\left(\left(s_{22} - \frac{s_{12}^2}{s_{11}}\right)^{-1}, \frac{n-1}{n} - \frac{(x_1 - \bar{x}_1)}{s_{11}}, n-3, \bar{x}_2 + \frac{s_{12}}{s_{11}}(x_1 - \bar{x}_1)\right).$$

En general se tiene que la distribución *condicional* $\tilde{F}^{j|1, \dots, (j-1)}$, de la j -ésima componente dadas las primeras $(j-1)$ componentes, es una

$$invt^{(1)}(P_j, Q_j, \nu_j, \underline{\eta}_j)$$

donde

$$\begin{aligned}
 P_j &= (s_{jj} - \xi_j' S_{j-1}^{-1} \xi_j)^{-1}, \\
 Q_j &= \frac{n-1}{n} (x_1 - \bar{x}_1, \dots, x_{j-1} - \bar{x}_{j-1}) S_{j-1}^{-1} \begin{bmatrix} x_1 - \bar{x}_1 \\ \vdots \\ x_{j-1} - \bar{x}_{j-1} \end{bmatrix}, \\
 \nu_j &= n - j - 1, \\
 \eta_j &= \bar{x}_j + \xi_j' S_{j-1}^{-1} \begin{bmatrix} x_1 - \bar{x}_1 \\ \vdots \\ x_{j-1} - \bar{x}_{j-1} \end{bmatrix} \text{ y} \\
 S_j &= \begin{bmatrix} S_{j-1} & \xi_j \\ \xi_j' & s_{jj} \end{bmatrix},
 \end{aligned}$$

para $j = 2, \dots, k$, y donde S_j es la matriz de covarianza muestral hasta la j -ésima componente (sin estar dividida por n).

iii. Obtener las \tilde{U}_{ij}

Este paso consiste en aplicar $\tilde{F}^{j|1, \dots, (j-1)}$ a las X_{ij} y determinar las \tilde{U}_{ij} como:

$$\tilde{U}_{ij} = \tilde{F}^{j|1, \dots, (j-1)}(X_{ij}), \text{ para } i = 1, \dots, n \text{ y } j = 1, \dots, k.$$

Con el propósito de hacer más operativo el método, en vez de calcular directamente la función $\text{invt}^{(1)}(P_j, Q_j, \nu_j, \eta_j)$, cosa que es difícil, se procedió a hacer uso de dos transformaciones.

La primera transformación mapea a una variable $X \sim \text{invt}^{(1)}(P, Q, \nu, \eta)$ en una variable $Z \sim \text{dirt}^{(1)}(P, Q, \nu, \eta)$ que en notación, también tomada de Press (1982), está dada por:

$$\frac{Q^{\frac{1}{2}}(X - \eta)}{\sqrt{Q - (X - \eta)'P(X - \eta)}} + \eta = Z.$$

Un vector \underline{Z} se distribuye como $dir^{(k)}(P, Q, \nu, \eta)$ si su función de densidad es la siguiente:

$$f(\underline{z}) = \frac{\Gamma(\frac{\nu+k}{2}) |P|^{-\frac{k}{2}} Q^{\frac{\nu}{2}} (Q + (\underline{z} - \eta)'P(\underline{z} - \eta))^{-\frac{\nu+k}{2}}}{\pi^{\frac{k}{2}} \Gamma(\frac{\nu}{2})}$$

La segunda transformación mapea a una $Z \sim dir^{(1)}(P, Q, \nu, \eta)$ en una variable $t \sim t_\nu$ (t de Student con ν grados de libertad, como aparece en cualquier libro de texto) dada por las siguiente expresión:

$$\frac{(Z - \eta)\nu^{\frac{1}{2}} |P|^{-\frac{1}{2}}}{Q^{\frac{1}{2}}} = t.$$

Finalmente, se tiene que las \tilde{U}_{ij} son de la forma:

$$G_{n-j-1} \left(\frac{(n-j-1)^{\frac{1}{2}} [X_{ij} - \bar{X}_j - \xi_j' S_{j-1}^{-1} (X_{i(j-1)} - \bar{X}_{(j-1)})]}{(s_{jj} - \xi_j' S_{j-1}^{-1} \xi_j)^{\frac{1}{2}} (\frac{n-1}{n} - (X_{i(j-1)} - \bar{X}_{(j-1)})' S_{j-1}^{-1} (X_{i(j-1)} - \bar{X}_{(j-1)} - M)^{\frac{1}{2}})} \right) \quad (\text{III.1})$$

para $j = 1, \dots, k$, $i = 1, \dots, n$, y donde

$$\underline{X}_{i(j-1)} = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{i(j-1)} \end{bmatrix},$$

$$\bar{X}_{(j-1)} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_{(j-1)} \end{bmatrix},$$

$$M = \frac{[X_{ij} - \bar{X}_j - \xi_j' S_{j-1}^{-1} (X_{i(j-1)} - \bar{X}_{(j-1)})]^2}{s_{jj} - \xi_j' S_{j-1}^{-1} \xi_j}$$

y G_ν es la función de distribución de una t de Student con ν grados de libertad.

iv. Calcular \bar{A}_{nk}^2

Se ordenan las \tilde{U}_{ij} como $\tilde{U}_{(1)}, \tilde{U}_{(2)}, \dots, \tilde{U}_{(nk)}$ y con ellas se calcula \bar{A}_{nk}^2 como

$$\bar{A}_{nk}^2 = -nk - \frac{1}{nk} \sum_{i=1}^{nk} (2i-1) \left[\ln(\tilde{U}_{(i)}) + \ln(1 - \tilde{U}_{(nk+1-i)}) \right]$$

Una vez hechos estos cuatro pasos, se tiene ya la estadística de prueba. La distribución de \bar{A}_{nk}^2 se desconoce. Para poder llevar a cabo la prueba fue necesario simular la distribución de \bar{A}_{nk}^2 bajo la hipótesis nula. Se encontró la distribución de \bar{A}_{nk}^2 para $k = 2$ y $k = 3$ mediante un estudio de simulación con 10,000 muestras de tamaño n , para $n = 15, 20, 30, 40$ y 50 ; éstas aparecen en las tablas 1 y 2.

Como observación que se considera muy importante, nótese que la distribución tanto para $k = 2$ como para $k = 3$ converge empíricamente a la distribución asintótica de A^2 cuando $k = 1$. Se postula este resultado surgido por la simulación de Monte Carlo, para su demostración formal se requiere de herramientas matemáticas fuera del alcance de este trabajo.

Con los valores de las tablas 1 y 2 es posible llevar a cabo la prueba rechazando la hipótesis nula cuando el valor de \bar{A}_{nk}^2 excede el valor que aparece en tablas.

TABLA 1. Distribución de \bar{A}_{nk}^2 .

k = 2.

p	Tamaño de muestra n					
	15	20	30	40	50	∞
.50	.336	.334	.336	.335	.337	.341
.75	.456	.460	.463	.466	.462	.470
.85	.547	.543	.552	.557	.559	.561
.90	.620	.611	.620	.625	.628	.631
.95	.738	.725	.747	.742	.748	.752
.975	.847	.847	.879	.873	.876	.873
.990	1.002	1.001	1.024	1.042	1.031	1.035
.995	1.132	1.136	1.147	1.155	1.159	1.159
.999	1.386	1.370	1.386	1.412	1.415	-

Porcentiles para varios valores de p y diferentes tamaños de muestra.

TABLA 2. Distribución de \bar{A}_{nk}^2 .

k = 3.

p	Tamaño de muestra n.					
	15	20	30	40	50	∞
.50	.337	.339	.336	.336	.338	.341
.75	.466	.469	.465	.461	.467	.479
.85	.549	.558	.556	.548	.558	.561
.90	.619	.626	.630	.622	.627	.631
.95	.747	.744	.751	.732	.749	.752
.975	.867	.859	.888	.855	.862	.873
.990	1.041	1.018	1.026	1.001	1.001	1.035
.995	1.192	1.143	1.116	1.113	1.123	1.159
.999	1.467	1.415	1.292	1.421	1.398	-

Porcentiles para varios valores de p y diferentes tamaños de muestra

III.3 Modificación al método de Rincón-Gallardo et al.

En la sección anterior se describió un método que se basa en transformar a los n vectores k -variados en nk variables aleatorias que se distribuyen "uniformemente" en el intervalo $(0,1)$. Esta idea había sido tratada en el artículo de Rincón-Gallardo et al. (1979), donde los autores presentan transformaciones que llevan a los n vectores k -variados a $k(n - k - 1)$ variables aleatorias independientes idénticamente distribuidas uniformemente en el intervalo $(0,1)$. Estas $k(n - k - 1)$ variables aleatorias están dadas por

$$U_{ij} = G_{i-k+j-2} \{ Z_{ij} [(i - k + j - 2) / (1 + Z_{1i}^2 + \dots + Z_{(j-1)i}^2)]^{1/2} \}$$

donde

$$\underline{Z}_i = A_i (\underline{X}_i - \bar{\underline{X}}_i) / \{ [(i - 1) / i] - (\underline{X}_i - \bar{\underline{X}}_i)' S_i^{-1} (\underline{X}_i - \bar{\underline{X}}_i) \}^{1/2} \quad (\text{III.2})$$

para $i = k + 2, \dots, n$ y $j = 1, \dots, n$ y donde

$$\begin{aligned} A_i' A_i &= S_i^{-1}, \\ \underline{X}_i &= 1/i \sum_{l=1}^i \underline{X}_l, \\ S_i &= \sum_{l=1}^i \underline{X}_l \underline{X}_l' - i \bar{\underline{X}}_i \bar{\underline{X}}_i' \end{aligned}$$

ya que G_ν es la función de distribución de una t de Student con ν grados de libertad.

Estas transformaciones resultan de considerar la distribución condicional de $\underline{X}_{k+2}, \underline{X}_{k+3}, \dots, \underline{X}_n$ dada la estadística suficiente minimal $T_n = (\bar{\underline{X}}_n, S_n)$ para después utilizar en forma secuencial la distribución condicional de \underline{X}_n dada T_n

luego la distribución condicional de \underline{X}_{n-1} dada T_n y \underline{X}_n , así hasta la distribución condicional de \underline{X}_{k+2} dadas T_n y $\underline{X}_n, \underline{X}_{n-1}, \dots, \underline{X}_{k+3}$.

En el trabajo de Rincón-Gallardo et al. se demuestra que la distribución condicional de \underline{X}_i dadas T_n y $\underline{X}_n, \dots, \underline{X}_{i+1}$ ($i \geq k+2$), es exactamente la misma que la distribución condicional de \underline{X}_i dada T_i .

Utilizando una transformación que requiere de exhibir una factorización A_i de S_i , se utilizó en Rincón-Gallardo et al. la versión de la t multivariada de Dickey (1967). Con esta versión y la aplicación de la transformación de Rosenblatt a la condicional de \underline{X}_i dada T_i es que se obtuvieron los resultados presentados en Rincón-Gallardo et al.

Con el método presentado por Rincón-Gallardo et al. se pierden por así decirlo $k+1$ vectores de observaciones; éstos son de alguna manera "el precio" por obtener independencia entre las U 's. Sin embargo este procedimiento tiene una desventaja: no resulta ser invariante al orden de los vectores de observaciones, es decir, viola el principio de verosimilitud de Fisher.

En un intento por evitar esa falla en el procedimiento de Rincón-Gallardo et al., se hizo la modificación siguiente: en vez de utilizar S_i y \bar{X}_i en III.2 se decidió utilizar S_n y \bar{X}_n para transformar a \underline{X} en una nueva \underline{Z} . Esto equivale a transformar cada \underline{X}_i ($i = 1, \dots, n$) con la condicional de \underline{X}_i dada T_n , no dada T_i y luego usar la versión de Dickey antes de aplicar las transformaciones de Rosenblatt. Al introducir esta idea, es decir al utilizar toda la información en cada una de las transformaciones (logrando con esto que el procedimiento fuera invariante al orden de las observaciones), se perdió la independencia entre las U 's. Se pensó que esa dependencia no sería importante al aumentar el tamaño de muestra n . Con esto en mente, se calcularon las U 's y con ellas se utilizó la estadística A^2 de Anderson-Darling para verificar su "uniformidad". Como la distribución de A^2 no se conocía, se simuló ésta para $k = 2$ y para distintos tamaños de $n = 20, 30, 40$ y 50 , y se encontró que la distribución no

convergía. Ante este resultado, que no se esperaba, se evaluó el procedimiento para $k = 1$; encontrándose que en este caso la distribución de A^2 convergía a la dada por D'Agostino y Stephens (1986)(caso 0, pág. 105), sin embargo cuando se hizo para $k = 3$ se encontró que la no-convergencia era aún más grave. No se sabe la razón exacta de esta falta de convergencia, pero se pensó que pudiera ser ocasionada por la introducción de un elemento arbitrario en la construcción de la estadística, como lo es la elección de la "raíz" A_n de la matriz S_n . No se profundizó más en el tema, procediendo a desarrollar otro método que no utilizara elementos arbitrarios, llegándose al método desarrollado en la sección anterior, en el cuál el uso de la distribución t -invertida, excluye la arbitrariedad al no requerir A_n .

Capítulo IV

ESTUDIO COMPARATIVO

IV.1 Simulación de distribuciones multivariadas no-normales

¿Qué se entiende por una distribución no-normal en el caso multivariado? Esta pregunta se enfrentó cuando se decidió un análisis comparativo de potencias.

Es muy difícil exhibir distribuciones no-normales en general. Pueden hacerse marginales no-normales independientes; pueden usarse para ello las típicas Ji-cuadradas, uniformes, "colas pesadas", etc. Ante esto se decidió utilizar un simulador que se centra en las asimetrías y kurtosis de las distribuciones marginales.

El método utilizado para la generación de distribuciones alternativas fue el presentado por Vale y Maurelli (1983), que está basado en el procedimiento de Fleishman para generar distribuciones univariadas no-normales con media, varianza, asimetría y kurtosis especificadas de antemano. Este procedimiento define una variable Y que es combinación lineal de las tres primeras potencias de una variable normal estándar X :

$$Y = a + bX + cX^2 + dX^3 \quad (\text{IV.1})$$

Las constantes a, b, c y d se determinan de manera que Y tenga los momentos especificados. Para lograr esto Fleishman expresa los primeros cuatro momentos de Y en términos de los primeros cuatro momentos de X que son constantes conocidas; y encuentra que a, b, c y d son las soluciones del siguiente sistema no lineal:

$$\begin{aligned} b^2 + 6bd + 2c^2 + 15d^2 &= 1 \\ 2c(b^2 + 24bd + 105d^2 + 2) &= \gamma_1 \\ 24[bd + c^2(1 + b^2 + 28bd) + d^2(12 + 48bd + 141c^2 + 225d^2)] &= \gamma_2 \\ a &= -c \end{aligned} \quad (\text{IV.2})$$

donde γ_1 y γ_2 son la asimetría y kurtosis respectivamente, tomando ambos valores de cero para el caso de una distribución normal (a γ_2 se le restó 3 para ello). Si $\gamma_2 < 0$ tendrá una distribución más aplanada que la normal (platicúrtica). Si $\gamma_2 > 0$ tendrá una distribución más apuntada que la normal (leptocúrtica).

El procedimiento para el caso multivariado es una combinación del procedimiento de Fleishman y el procedimiento de descomposición de la matriz de correlación poblacional utilizado para generar normales multivariadas.

Se generan primero números aleatorios que provienen de una distribución normal multivariada con correlaciones especificadas y después a cada componente se le transforma en forma univariada con el método mencionado. Los dos procedimientos interactúan de tal manera que las variables no-normales tienen

correlaciones diferentes a las de las variables normales originales. La matriz de correlación intermedia se determina a través de la matriz de correlación final deseada y de las transformaciones no lineales previamente descritas.

El procedimiento se describe a continuación para el caso bivariado:

- i. Resolver el sistema no-lineal IV.2 para los valores deseados de γ_1 y γ_2 de cada marginal, para hallar así a_i, b_i, c_i y d_i $i = 1, 2$.
- ii. Se definen dos variables X_1 y X_2 con distribución normal estándar y correlación ρ_{X_1, X_2} que se determinará más adelante. Sean $\underline{X}'_i = (1, X_i, X_i^2, X_i^3)$ y $\underline{W}'_i = (a_i, b_i, c_i, d_i)$ con $i = 1, 2$.
- iii. La variable Y_i es el resultado del producto de estos vectores

$$Y_i = \underline{W}'_i \underline{X}_i$$

- iv. Si ρ_{Y_1, Y_2} es la correlación final deseada entre las variables no-normales Y_1 y Y_2 , que corresponden a las variables normales X_1 y X_2 . Se tiene que

$$\rho_{Y_1, Y_2} = E(Y_1 Y_2) = E(\underline{W}'_1 \underline{X}_1 \underline{W}'_2 \underline{X}_2) = \underline{W}'_1 R \underline{W}_2$$

donde

$$R = E(\underline{X}_1 \underline{X}'_2) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & \rho_{X_1, X_2} & 0 & 3\rho_{X_1, X_2} \\ 1 & 0 & 2\rho_{X_1, X_2}^2 + 1 & 0 \\ 0 & 3\rho_{X_1, X_2} & 0 & 6\rho_{X_1, X_2}^3 + 9\rho_{X_1, X_2} \end{bmatrix}$$

Haciendo el álgebra correspondiente y usando relaciones entre a y c se encuentra la siguiente expresión:

$$\rho_{Y_1, Y_2} = \rho_{X_1, X_2} (b_1 b_2 + 3b_1 d_2 + 3b_2 d_1 + 9d_1 d_2) + \rho_{X_1, X_2}^2 (2c_1 c_2) + \rho_{X_1, X_2}^3 (6d_1 d_2).$$

Al resolver este polinomio para ρ_{X_1, X_2} se obtiene la correlación intermedia entre X_1 y X_2 requerida para obtener la correlación deseada.

Existen problemas numéricos de cierta importancia en la solución del sistema IV.2. Para ilustrar esto, se comenta que el uso de una norma (para el error en la búsqueda de la solución) de 10^{-26} llevó en algunos casos a soluciones incorrectas. Por ello en todos los casos se verificó que las soluciones fueran muy precisas.

Este procedimiento se programó en FORTRAN en una VAX 11-750 haciendo uso de las subrutinas GGNSM y ZRPOLY del paquete IMSL.

La comparación de potencias se hizo únicamente para el caso bivariado. Las distribuciones alternativas se dividieron en tres grupos:

Grupo 1 Tiene ambas marginales no-normales y la correlación entre ellas es cero.

Grupo 2 Tiene una marginal normal y la otra no-normal, la correlación entre ellas es cero.

Grupo 3 Es una combinación de los dos grupos anteriores pero en este grupo la correlación entre las marginales se fijó en 0.7 .

En el apéndice 1 pueden verse las gráficas de las distintas marginales con las que se trabajó en el estudio comparativo, a través de histogramas generados con 5000 observaciones de cada distribución no-normal.

IV.2 Resultados del estudio comparativo

El estudio se hizo probando cada una de las estadísticas frente a 1000 muestras de tamaño $n = 40$ de las distintas distribuciones no normales (por ejemplo, en la Tabla 3 en el primer renglón, para el caso de una distribución con ambas marginales no normales, con asimetría $\gamma_1 = 0.0$ y kurtosis $\gamma = 0.5$, con la estadística \tilde{A}^2 la hipótesis de normalidad se rechazó para un $\alpha = 0.10$, en 181 muestras de las 1000 y para un $\alpha = 0.05$ en 104).

De los tres grupos, fue en el grupo 2 fue donde se encontraron las menores potencias para las estadísticas consideradas. Las estadísticas en su conjunto no mostraron un patrón definido frente a la presencia o ausencia de correlación. Por ejemplo, mientras que la estadística *RAN* aumentó su potencia al incluir correlación, otras como \bar{A}^2 la disminuyeron.

En seguida se presenta un análisis para cada uno de los grupos considerados.

En la Tabla 3, donde aparecen los resultados para el primer grupo de alternativas, puede verse que las estadísticas *SW* y *CW* son sesgadas para todas las alternativas, excepto para la alternativa con $(\gamma_1, \gamma_2) = (0.0, -1.0)$. Mientras que para esa misma alternativa aparecen con sesgo *B1B2*, *SU*, *CSDI*, y *W**. El resto de las estadísticas no muestra sesgo; éstas son \bar{A}^2 , *CU*, *CSI*, *CSD* y *RAN*. Al observar las potencias de *CSI* y *CSD*, éstas resultan ser pequeñas, cercanas al α . \bar{A}^2 aparece como la mejor opción para los casos $(\gamma_1, \gamma_2) = (0.5, 0.0)$ y $(\gamma_1, \gamma_2) = (0.5, 0.5)$, es decir los casos que presentan asimetría; mientras que para los casos simétricos $(\gamma_1, \gamma_2) = (0.0, 0.5)$ y $(\gamma_1, \gamma_2) = (0.0, 1.0)$, *CU* es la mejor opción. Para el caso simétrico $(\gamma_1, \gamma_2) = (0.0, -1.0)$ la mejor opción entre todas es *SW*, que como se dijo anteriormente es sesgada. La siguiente mejor opción es \bar{A}^2 . Para este grupo *RAN* no muestra potencias altas excepto para el caso $(\gamma_1, \gamma_2) = (0.0, -1.0)$.

En la Tabla 4, correspondiente al grupo 2, otra vez las estadísticas *SW* y *CW* presentan sesgos en la mayoría de las alternativas. Otras como *CSI*, *CSDI*, *CSD*, *W**, y *RAN* presentan sesgos sólo en algunos de los casos. En esta tabla resalta claramente *CU* como la mejor opción, seguida por *SU*. \bar{A}^2 presenta su mejor comportamiento frente a la distribución que tiene la mayor asimetría en la segunda componente ($\gamma_1 = 0.8$). Para la alternativa con kurtosis ($\gamma_2 = -1.0$) también se presenta como la mejor opción. Es en la alternativa con kurtosis ($\gamma_2 = 1.0$) donde *B1B2* presenta su mejor comportamiento. En la tabla anterior le ocurre algo similar.

Los resultados del grupo 3 aparecen en la Tabla 5. Ahí es fácil ver que las estadísticas $SW, CW, CSI, CSDI$ y CSD presentan sesgos. Salta a la vista el buen comportamiento de RAN en la presencia de correlación, que en ningún caso es superada por otra de las estadísticas. Sin tomar en cuenta a RAN las estadísticas $\bar{A}^2, B1B2, SU, CU$ y W^* resultan ser la mejor opción al menos frente a alguna alternativa, aunque las potencias que presentan son considerablemente menores a las que presenta RAN . \bar{A}^2 muestra mejor comportamiento frente a las alternativas con ambas marginales no simétricas. W^* resulta ser la mejor (después de RAN) para las alternativas que tienen componentes con $(\gamma_1, \gamma_2) = (0.5, 0.5)$; mientras que SU , y CU se comportan bien frente a las alternativas con $(\gamma_1, \gamma_2) = (0.0, 0.5)$.

Cabe mencionar que la potencia de \bar{A}^2 fue mayor en los casos en que la alternativa tenía la primer componente no normal y la segunda normal. Esto puede verse en los renglones 3 y 4 de la Tabla 4 y en los renglones 6 y 7 de la Tabla 5.

Como en casi todas las comparaciones de pruebas generales (ómnibus) y al igual que en estudios anteriores, se muestra que no existe una mejor opción. A pesar de lo limitado que es este estudio exploratorio, sí puede considerarse un hallazgo de cierta importancia, el que algunas pruebas mostraron ser sesgadas; unas ante varias alternativas otras ante una sola. En ese último caso cae la estadística RAN que se distinguió por otro lado por tener una muy alta potencia en las alternativas reportadas en la Tabla 5 (con variables con correlación bastante alta).

Las estadísticas RAN (con la salvedad del caso en que mostró un sesgo), SU (también con un curioso sesgo para el caso $\alpha = 0.05$ en una de las alternativas de la tabla 3), CU (que no mostró sesgo) y \bar{A}^2 (que no mostró sesgo en ninguno de los casos) se recomiendan como pruebas generales.

La generalización a más de dos dimensiones de todas estas pruebas se ve

sumamente sencilla para \bar{A}^2 , SU y CU , no siendo así para RAN para la cual se desconoce cómo podría generalizarse.

Otro aspecto que puede resultar de interés para la recomendación de pruebas generales es el que la distribución asintótica además de que debe estar disponible, sea alcanzable con un buen grado de aproximación para un tamaño moderado de muestra. En este último sentido \bar{A}^2 mostró satisfacer este requisito en los casos $k = 2$ y $k = 3$; sin embargo, SU y CU para el caso $k = 2$ mostraron que la aproximación al utilizar su distribución asintótica fue un tanto cruda (recuérdese que CU en el caso $k = 2$ inclusive dejó de estar definida computacionalmente si $n \leq 26$).

TABLA 3. Potencia

Variables no correlacionadas con la misma distribución marginal no normal

Los números en paréntesis corresponden a $\alpha = 0.05$ y los otros a $\alpha = 0.10$

Alternativa			Estadísticas										
	γ_1	γ_2	\bar{A}^2	B1B2	SU	SW	CU	CW	CSI	CSDI	CSD	W^*	RAN
X_1	0.0	0.5	181	348	376	44	388	65	107	91	88	178	139
X_2	0.0	0.5	(104)	(225)	(272)	(22)	(282)	(30)	(52)	(47)	(39)	(106)	(84)
X_1	0.5	0.0	459	320	359	39	430	17	105	104	105	264	119
X_2	0.5	0.0	(319)	(195)	(246)	(27)	(345)	(4)	(51)	(52)	(53)	(152)	(56)
X_1	0.0	-1.0	493	24	314	507	240	246	90	86	99	43	401
X_2	0.0	-1.0	(333)	(8)	(24)	(409)	(142)	(160)	(32)	(42)	(54)	(13)	(278)
X_1	0.5	0.5	390	233	266	66	244	58	110	101	96	294	132
X_2	0.5	0.5	(286)	(160)	(196)	(46)	(172)	(24)	(50)	(52)	(45)	(195)	(70)
X_1	0.0	1.0	246	457	474	83	474	84	103	91	80	267	178
X_2	0.0	1.0	(154)	(223)	(375)	(51)	(379)	(50)	(51)	(51)	(40)	(188)	(103)

γ_1, γ_2 medidas de asimetría y kurtosis

TABLA 4. Potencia

VARIABLES NO CORRELACIONADAS, UNA MARGINAL NORMAL Y OTRA NO NORMAL

Los números en paréntesis corresponden a $\alpha = 0.05$ y otros a $\alpha = 0.10$

Alternativa			Estadísticas										
	γ_1	γ_2	\tilde{A}^2	B1B2	SU	SW	CU	CW	CSI	CSDI	CSD	W^*	RAN
X_1	0.0	0.0	167	110	169	145	238	59	87	86	87	59	161
X_2	0.0	-1.0	(84)	(62)	(81)	(104)	(177)	(28)	(42)	(52)	(44)	(30)	(88)
X_1	0.0	-1.0	172	101	170	156	227	57	94	108	99	71	174
X_2	0.0	0.0	(94)	(53)	(67)	(104)	(153)	(22)	(51)	(53)	(57)	(26)	(85)
X_1	0.0	0.0	177	269	308	53	360	44	98	104	90	176	118
X_2	0.5	0.0	(98)	(164)	(210)	(34)	(266)	(14)	(59)	(58)	(45)	(101)	(47)
X_1	0.5	0.0	197	247	276	41	346	22	85	89	96	184	82
X_2	0.0	0.0	(115)	(141)	(185)	(23)	(258)	(7)	(44)	(36)	(45)	(97)	(38)
X_1	0.0	0.0	167	156	183	71	177	83	111	104	118	196	138
X_2	0.5	0.5	(110)	(99)	(124)	(36)	(112)	(27)	(49)	(56)	(55)	(123)	(75)
X_1	0.0	0.0	157	334	363	49	385	74	93	92	102	163	138
X_2	0.0	1.0	(94)	(234)	(263)	(28)	(283)	(32)	(40)	(60)	(52)	(108)	(88)
X_1	0.0	0.0	568	332	379	32	519	14	91	71	77	350	255
X_2	0.8	0.0	(445)	(196)	(239)	(21)	(408)	(10)	(37)	(29)	(34)	(267)	(143)

γ_1, γ_2 medidas de asimetría y kurtosis

TABLA 5. Potencia

Variables correlacionadas ($\rho = 0.7$)

Los números en paréntesis corresponden a $\alpha = 0.05$ y otros a $\alpha = 0.10$

Alternativa			Estadísticas										
	γ_1	γ_2	\bar{A}^2	B1B2	SU	SW	CU	CW	CSI	CSDI	CSD	W^4	RAN
X_1	0.0	0.5	144	195	219	89	214	129	75	89	115	199	637
X_2	0.0	0.5	(82)	(112)	(140)	(41)	(114)	(55)	(30)	(47)	(59)	(136)	(492)
X_1	0.5	0.0	340	189	227	51	212	22	92	102	111	290	485
X_2	0.5	0.0	(231)	(98)	(139)	(33)	(131)	(11)	(41)	(56)	(61)	(173)	(306)
X_1	0.5	0.5	281	259	296	68	243	57	81	125	141	325	519
X_2	0.5	0.5	(201)	(165)	(211)	(36)	(185)	(23)	(40)	(69)	(85)	(221)	(356)
X_1	0.0	0.0	109	146	168	88	164	116	87	83	89	137	650
X_2	0.0	0.5	(50)	(85)	(93)	(46)	(89)	(44)	(41)	(55)	(42)	(87)	(513)
X_1	0.0	0.0	148	156	185	80	167	104	53	149	209	176	453
X_2	0.5	0.0	(84)	(88)	(98)	(41)	(95)	(53)	(21)	(89)	(128)	(94)	(307)
X_1	0.0	0.0	129	208	241	83	198	79	59	150	202	226	497
X_2	0.5	0.5	(73)	(132)	(165)	(51)	(134)	(32)	(28)	(86)	(122)	(151)	(328)
X_1	0.5	0.5	179	180	208	48	174	48	58	132	188	240	621
X_2	0.0	0.0	(97)	(107)	(134)	(24)	(111)	(20)	(24)	(75)	(108)	(148)	(462)

γ_1, γ_2 medidas de asimetría y kurtosis

Apéndice A

CONCLUSIONES

En este trabajo se estudió un método nuevo que de manera razonable mostró empíricamente que \tilde{A}^2 , su estadística de prueba, converge en distribución (bajo H_0) cuando $n \rightarrow \infty$ para $k = 2$ y $k = 3$; cosa que no ocurrió con la modificación hecha al método de Rincón et al. (1979). La explicación de esta falta de convergencia puede estar en la dependencia más fuerte entre las variables construidas.

El nuevo método es susceptible de ser utilizado para cualquier k . En la construcción de la estadística de prueba, al aumentar la dimensión se utiliza la fórmula III.1, que de manera sencilla puede adaptarse para cualquier valor de k , sin requerirse de la inversión de matrices, ya que utiliza fórmulas conocidas en la literatura para evaluarse secuencialmente.

Cabe mencionar que el método propuesto no es invariante frente al orden de las componentes, lo que implica que éstas deben ordenarse de alguna manera. Los resultados del estudio sugieren que para obtener mayor potencia se deben colocar como primeras componentes aquéllas que se consideren más no normales.

En el estudio de potencias realizado, la estadística \tilde{A}^2 se comportó bien en los casos donde las alternativas presentaban asimetría en una o en ambas

marginales, cuando éstas son no correlacionadas; también tuvo buen comportamiento en las alternativas en que una o ambas marginales (no correlacionadas) presentaban kurtosis igual a -1.0 . Cuando las alternativas presentaban una alta correlación el comportamiento de \bar{A}^2 no fue tan bueno.

En general el estudio mostró que las estadísticas CU , SU y \bar{A}^2 son las mejores opciones cuando las variables son no correlacionadas, mientras que RAN fue la mejor cuando las variables están altamente correlacionadas.

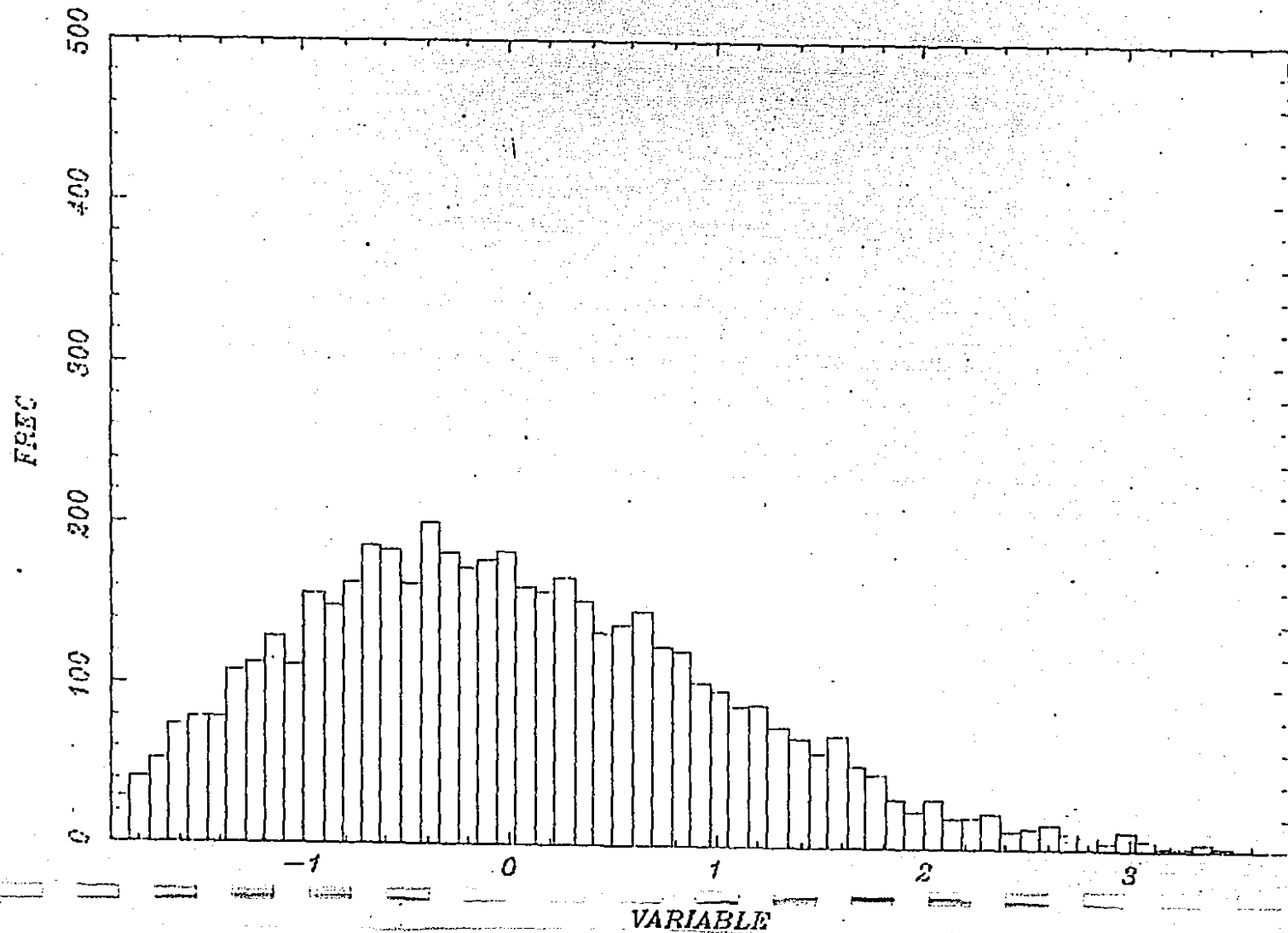
De las estadísticas propuestas hay algunas que no convergen rápidamente, y la distribución asintótica no resulta ser aproximación razonable. Otros métodos resultaron sesgados, lo cual en cierta manera los elimina para futuras consideraciones. La estadística RAN debiera estudiarse más para hacer una generalización para los casos con $k > 2$.

Otra posibilidad en futuros estudios podría ser la consideración de la suma de k estadísticas \bar{A}^2 (una para cada componente) cuya distribución límite se conjetura que es una mezcla de χ^2 's. Desde luego para otros estudios debiera utilizarse un conjunto más amplio de alternativas como podrían ser las mezclas de distribuciones normales, o el uso de distribuciones t , χ^2 , Cauchy, etc., para conocer mejor el comportamiento de las estadísticas consideradas en este trabajo.

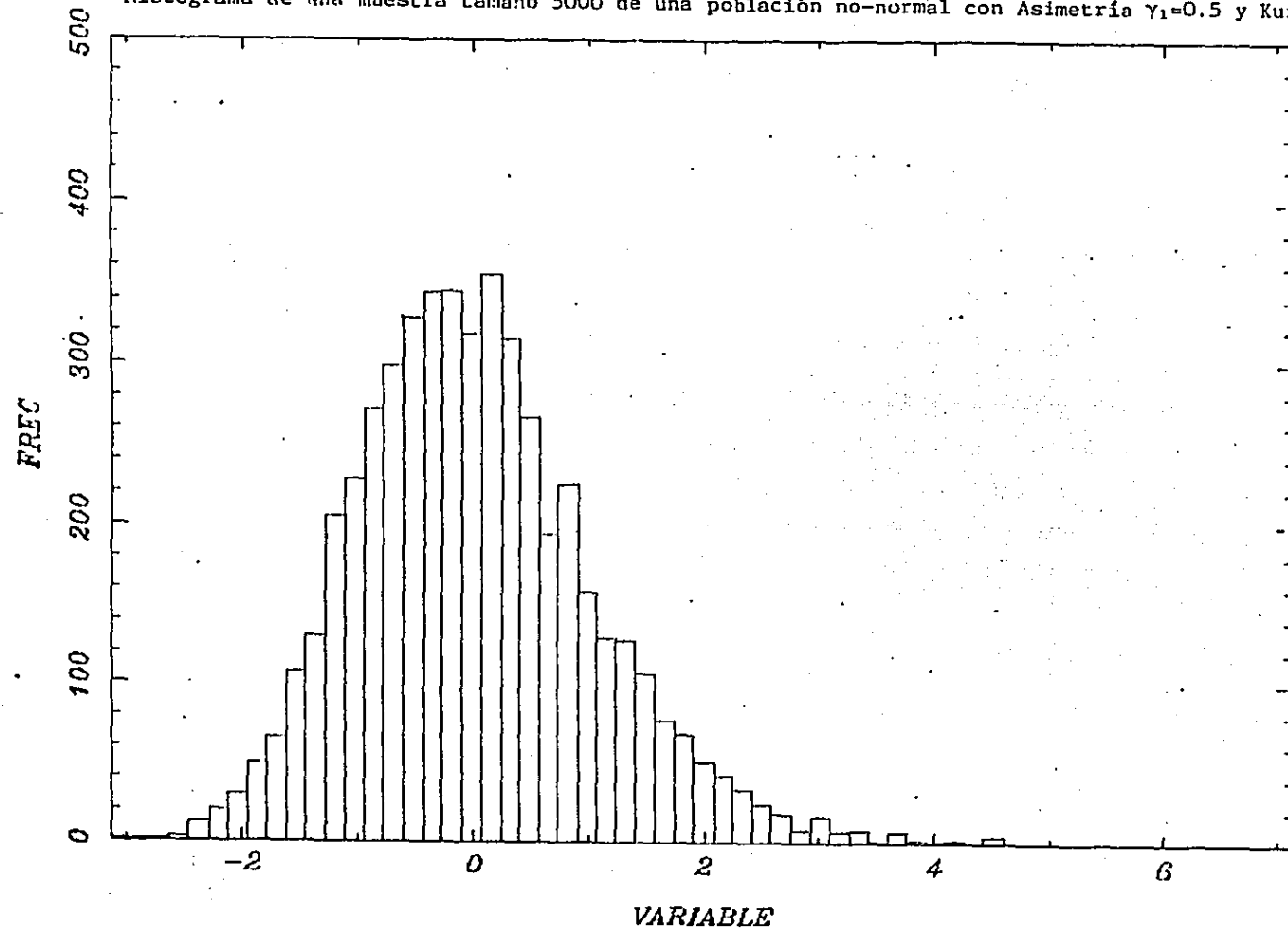
Apéndice

HISTOGRAMAS

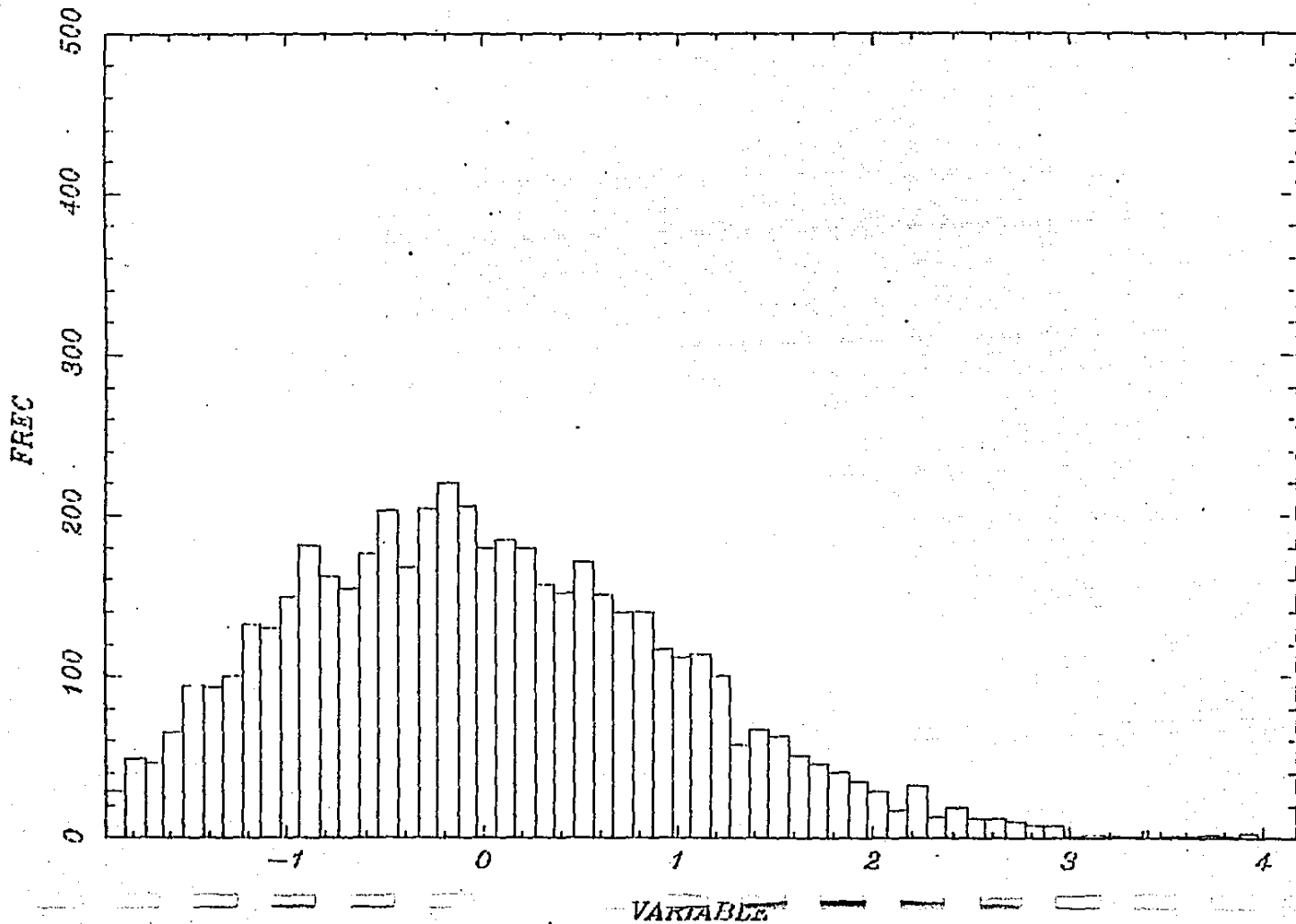
Histograma de una muestra tamaño 5000 de una población no-normal con Asimetría $\gamma_1=0.5$ y Kurtosis $\gamma_2=0.0$



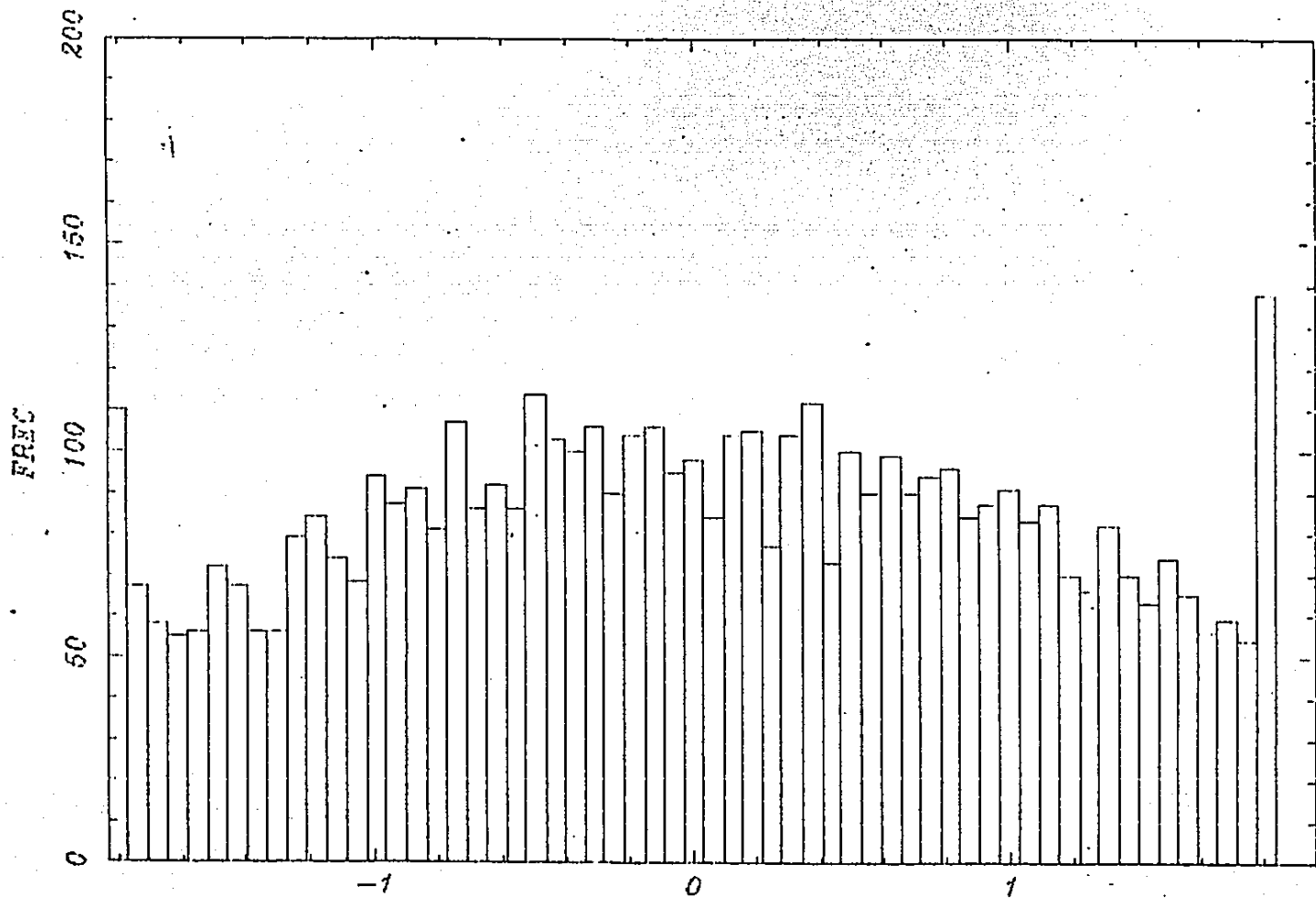
Histograma de una muestra tamaño 5000 de una población no-normal con Asimetría $\gamma_1=0.5$ y Kurtosis $\gamma_2=0.5$



Histograma de una muestra tamaño 5000 de una población no-normal con Asimetría $\gamma_1=0.5$ y Kurtosis $\gamma_2=0.0$

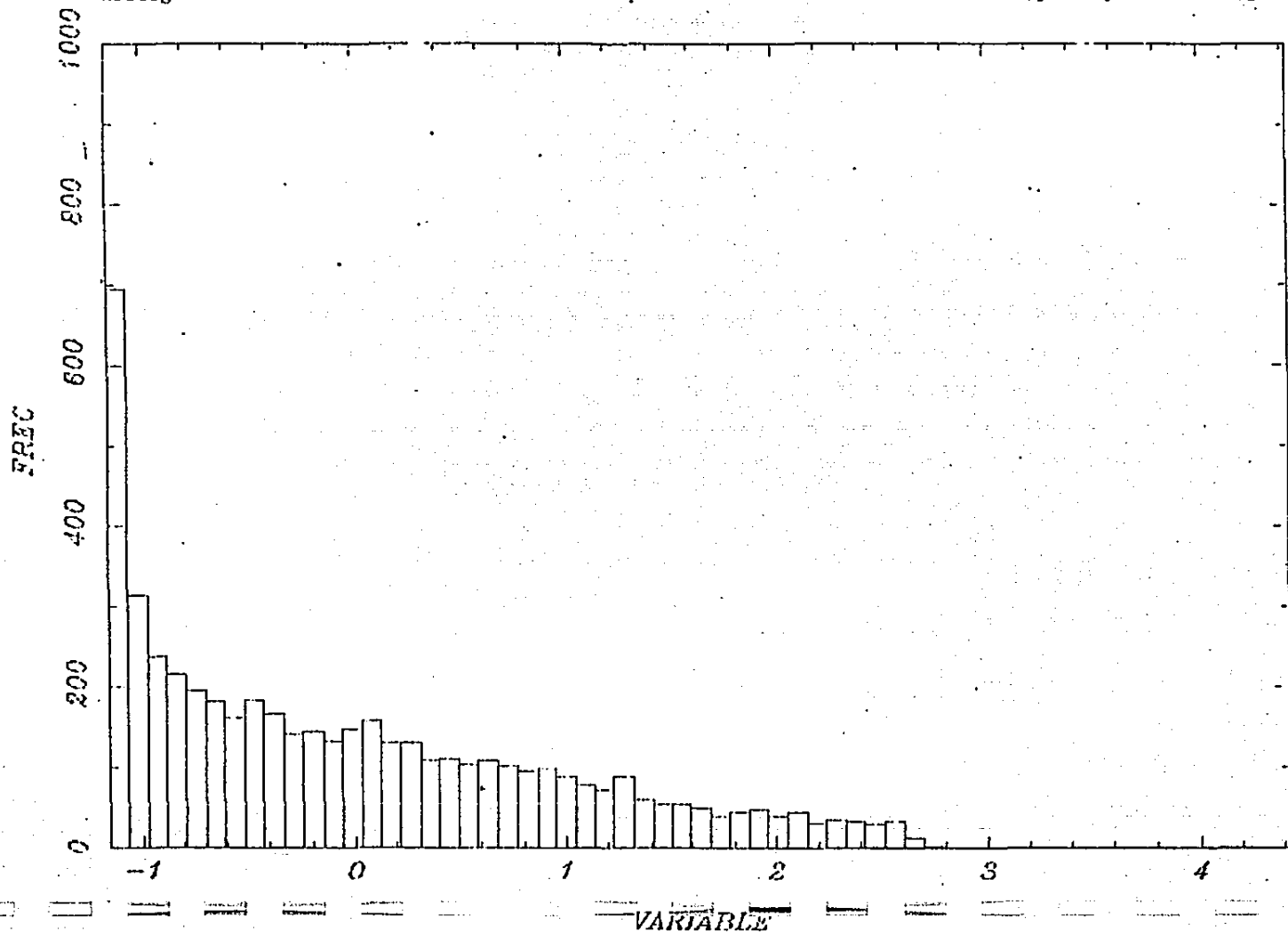


Histograma de una muestra tamaño 5000 de una población no-normal con Asimetría $\gamma_1=0$ y Kurtosis $\gamma_2=-1.0$

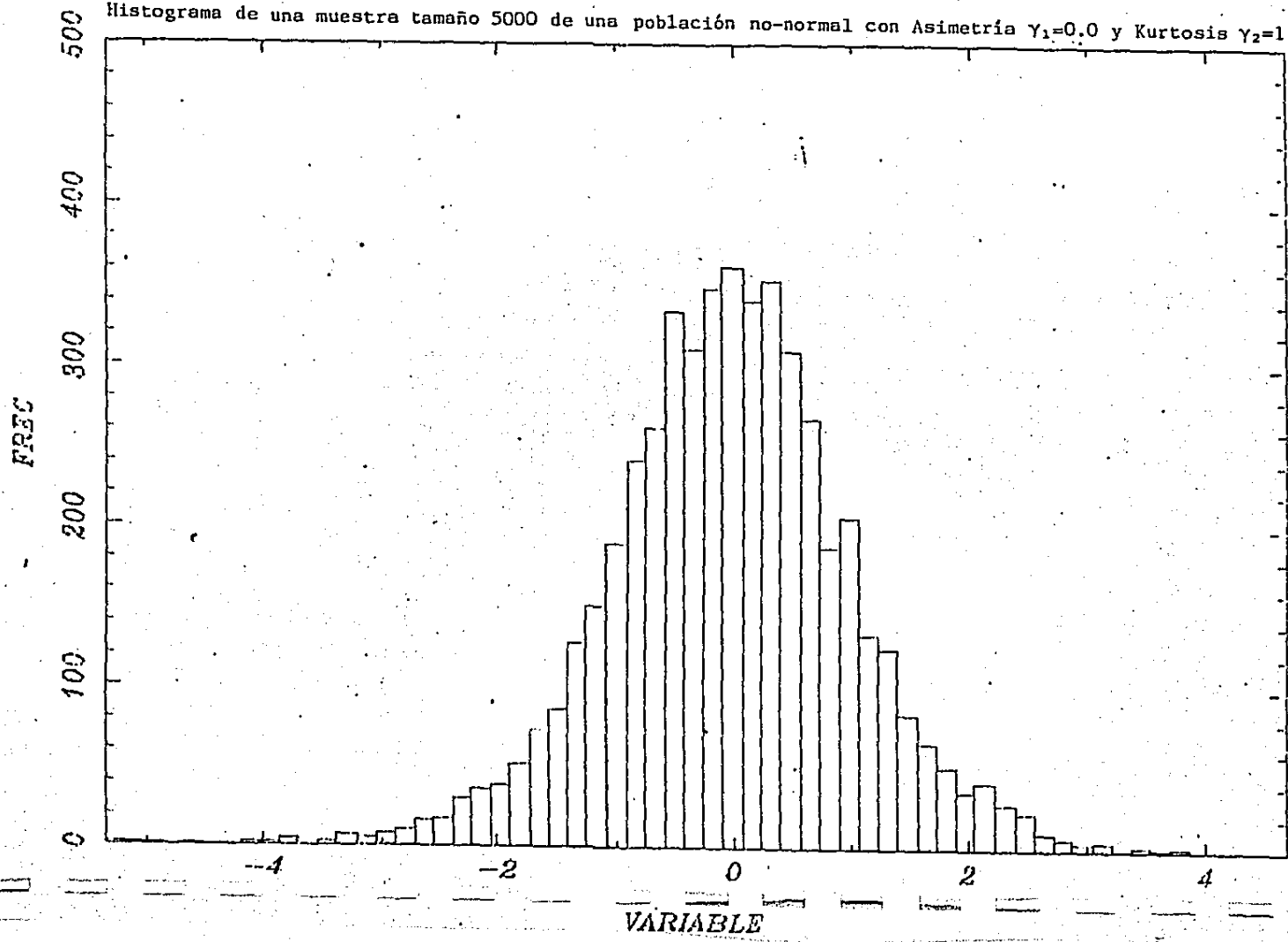


ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

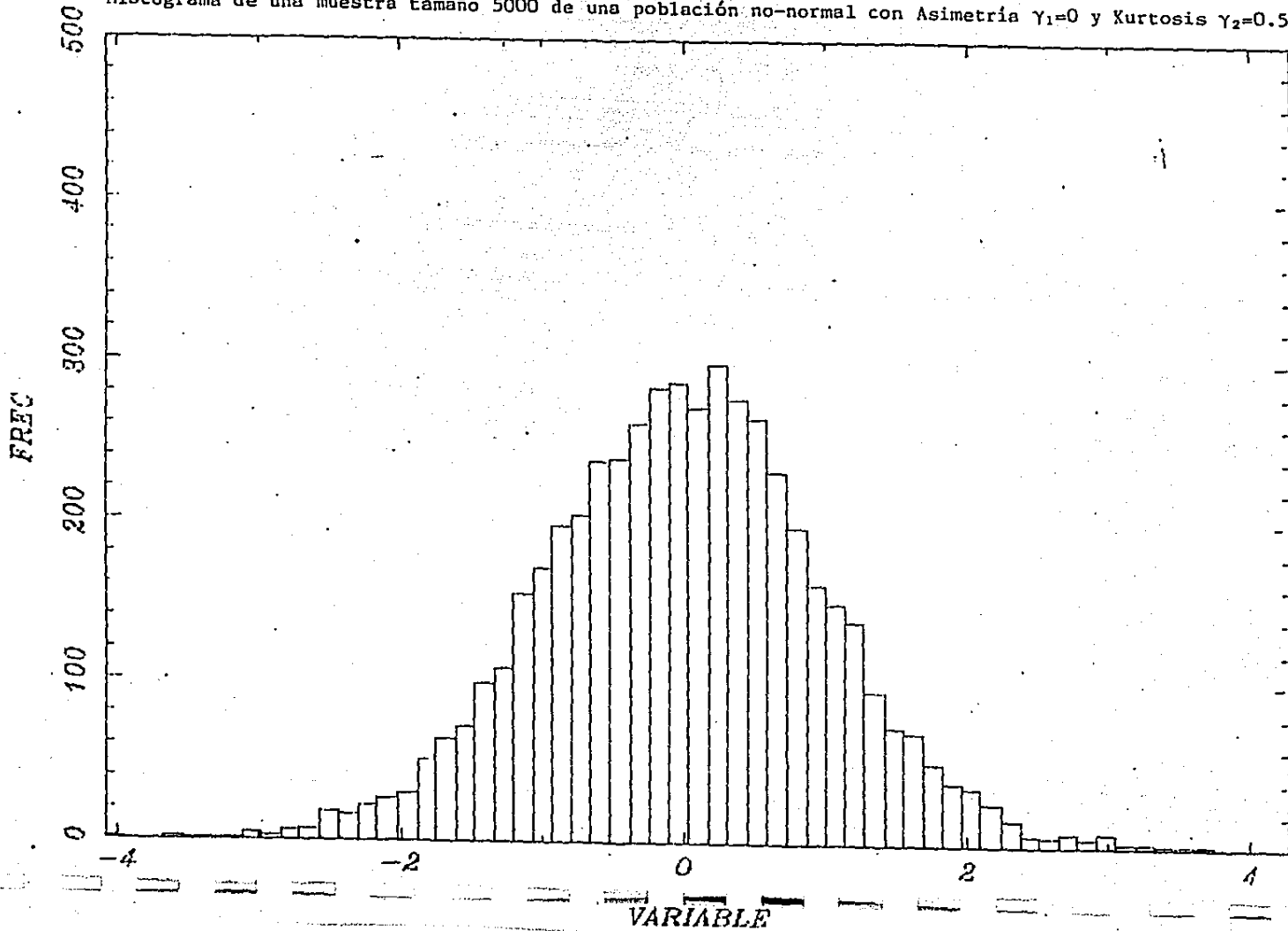
Histograma de una muestra tamaño 5000 de una población no-normal con Asimetría $\gamma_1=0.8$ y Kurtosis $\gamma_2=0.0$



Histograma de una muestra tamaño 5000 de una población no-normal con Asimetría $\gamma_1=0.0$ y Kurtosis $\gamma_2=1.0$



Histograma de una muestra tamaño 5000 de una población no-normal con Asimetría $\gamma_1=0$ y Kurtosis $\gamma_2=0.5$



REFERENCIAS

ANDREWS, D. F., GNANADESIKAN, R. Y WARNNER, J. L. (1972) "Methods for assessing multivariate normality" En: P.R. Krishnaiah ed., Multivariate analysis III, Academic press, Nueva York, 95-116.

COX, D.R. Y SMALL, N. J. H. (1978) "Testing multivariate normality" *Biometrika* 65, 263-272.

CSÖRGÖ, S. (1986) "Testing for normality in arbitrary dimension." *Ann. Math. Statist.* 14,708-723.

DICKEY, J. M. (1967) "Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution." *Ann. Math. Statist.* 38,511-518.

D'AGOSTINO, R. B. Y M. A. STEPHENS (Eds.) (1986) *Goodness of fit techniques*. Marcel Decker. Nueva York.

GHURYE, S.G. Y I. OLKIN (1969) "Unbiased estimation of some multivariate probability densities and related functions", *Ann. Math. Statist.* 40,1261-1271.

HENSELER, G.L., MEHROTA, K.G. Y MICHAELEK, J. E. (1977) "A goodness of fit test for multivariate normality." *Comm. Statist. A Theory Methods*. **6**, 33-41.

MALKOVICH, J.F. Y AFIFI, A.A. (1973) "On tests for multivariate normality." *J. Am. Statist. Ass.* **68**, 176-179.

MARDIA, K.V. (1974) "Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies." *Sankhya* **36**, (B), 115-128.

MARDIA, K.V. Y FOSTER, K. (1983) "Omnibus test of multinormality based on skewness and kurtosis." *Comm. in Statist.* **12**, 207-221.

MARDIA, K. V. Y P. J. ZEMROCH (1975) "Measures of multivariate skewness and kurtosis." *J. Roy. Statist. Soc. C* **24**, 262-265.

MOORE, D. S. (1973) "A note on Srinivasan's goodness-of-fit test." *Biometrika*. **60**, 209-211.

MUROTA, K Y TAKEUCHI, K (1981) "The studentized empirical characteristic function and its application to test the shape of distribution." *Biometrika* **68**, 55-65.

O'REILLY F. J. Y C. P. QUESENBERRY (1973) "The conditional probability integral transformation and applications to obtain composite chi-square goodness of fit techniques tests." *Am. Statist.* **1**, 74-83.

PRESS, S. J. (1982) *Applied Multivariate Analysis*. Malabar Krieger Publishing Co.

RINCÓN-GALLARDO, S. QUESENBERRY, C. P. Y F. J. O'REILLY (1979) "Conditional probability integral transformations and goodness-of-fit tests for multivariate normal distributions." *Ann. Statist.* **7**, 1052-1057.

ROSENBLATT, M. (1952) "Remarks on a Multivariate Transformation", *Ann. Math. Statist.* 28, 470-473.

SHAPIRO, S.S. Y WILK, M. B. (1965) "An analysis of variance test for normality." *Biometrika* 52, 591-611.

VALE, C. D. Y V. A. MAURELLI (1983) "Simulating multivariate non-normal distributions." *Psychometrika* 48, 465-471.