Robustez Asintótica de la Estadística de Hotelling

Mario Antonio Gneri[1] - *Emanuel Pimentel Barbosa*

Abstract

Es sabido que, bajo la hipótesis de normalidad, la distribución asintótica de la estadística T^2 de Hotelling es uma chi-quadrado. En este trabajo damos una demostración de que la hipótesis de normalidad puede ser suprimida, permaneciendo el resultado válido bajo condiciones bastante generales (finitud de momentos de orden 2 y matriz de covariancia definida positiva). Esta afirmación más general es también mencionada en la literatura, pero nos parece necesaria una demostración rigurosa de la misma.

Palabras llave: estadística T^2 de Hotelling, robustez asintótica.

1 Introducción

1.1 La estadística T^2 de Hotelling bajo la hipótesis de normalidad

Sea X un vector aleatorio p-dimensional con vector de medias μ y matriz de covariancia Σ. Consideremos una muestra de n observaciones independientes e idénticamente distribuídas $X_1, X_2, ..., X_n$ del vector X con sus respectivos vector de medias $\bar{X}(n)$ y matriz de covariancias muestrales $S(n)$, donde:

$$\bar{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{y} \quad S(n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}(n))(X_i - \bar{X}(n))'.$$

Hotelling (1931) definió la estadística T^2 mediante la fórmula:

$$T^2 = n(\bar{X}(n) - \mu)'(S(n))^{-1}(\bar{X}(n) - \mu).$$

[1] Corresponding author. E-mail adress: gneri@ime.unicamp.br

En los libros de análisis multivariado (por ejemplo, Anderson (1971) o Seber (1984)) se prueba que si los vectores $X_1, X_2, ..., X_n$ tienen distribución normal multivariada $\text{Normal}_p(\mu, \Sigma)$ la estadística T^2 tiene distribución $\frac{(n-1)p}{(n-p)}$ $F_{p,n-p}$ y - consecuentemente - cuando n tiende a infinito la distribución de T^2 converge a una chi-cuadrado con p grados de libertad.

La robustez asintótica de T^2 también es mencionada en la literatura y hasta alguna demostración es esbozada, pero sin rigor (como por ejemplo en Johnson and Wichern (1998), p. 187).

Mostraremos aquí que, substituyendo la normal por cualquier distribución continua con momentos de orden 2 finitos y matriz de covariancia positiva definida, la distribución asintótica de T^2 es también una chi-cuadrado con p grados de libertad.

La forma usual de probar que la estadística t de Student es asintóticamente robusta se basa en el Teorema Central del Límite y en el conocido "Teorema de Slutzky". Aquí seguiremos la misma línea para mostrar la robustez asintótica de T^2.

1.2 Notación, normas y convergencia

1. $\mathbb{N}$ and $\mathbb{R}$ denotan los conjuntos de números naturales y reales, respectivamente.

2. Si $m \in \mathbb{N}$, $k \in \mathbb{N}$ y $A \in \mathbb{R}^{m \times k}$, A' denota la transpuesta de A.

3. La norma de A para $A = [a_{ij}] \in \mathbb{R}^{m \times k}$ se define por:

$$\| A \| = \text{máximo}\{| a_{ij} |, i = 1, ..., m, j = 1, ..., k\}.$$

Es fácil verificar que si $A \in \mathbb{R}^{m\times k}$ and $B \in \mathbb{R}^{k\times t}$, entonces

$$\| AB \| \leq k \| A \| \| B \| \tag{1}$$

para m, k y t naturales.

4. Como es usual, diremos que la sucesión A_n in $\mathbb{R}^{m\times k}$ converge a $A \in \mathbb{R}^{m\times k}$ si y sólo si $\| A_n - A \|$ converge a cero. Recordamos que todas las métricas inducidas por normas son equivalentes en $\mathbb{R}^{m\times k}$, particularmente la convergencia en la norma $\| \, \|$ es equivalente a la convergencia coordenada a coordenada.

5. Si Z es un vector aleatorio p-dimensional y P una medida de probabilidad, PZ^{-1} denotará la distribución inducida por Z en $\mathbb{R}^p$.

2 Compacidad relativa de uma familia de distribuciones

Sea $\mathcal{F}$ el espacio de las medidas finitas definidas en la σ-álgebra de los conjuntos borelianos de um espacio métrico separable y completo (Ω, d). Denotemos por $\mathcal{T}$ la topología de la convergencia en distribición en $\mathcal{F}$.

Prohorov (1956) demostró que existe una métrica π en $\mathcal{F}$ que induce $\mathcal{T}$ y que $(\mathcal{F}, \pi)$ es separable y completo. Prohorov también caracterizó los conjuntos relativamente compactos de $(\mathcal{F}, \mathcal{T})$. En el caso de una familia $\mathcal{P}$ de medidas de probabilidad, Prohorov provó que $\mathcal{P}$ es relativamente compacto si y sólo si la familia es "tight", o sea, si dado $\epsilon > 0$ existe un compacto $K = K(\epsilon, \mathcal{P})$ en (Ω, d) tal que

$$P(K) > 1 - \epsilon \quad \text{para toda } P \text{ en } \mathcal{P}.$$

Utilizaremos esta propiedad en la demostración de la Proposición 2 de 4.1.

3 Convergencia en probabilidad en $\mathbb{R}^{m\times k}$

Probaremos aquí la equivalencia entre convergencia en probabilidad coordenada a coordenada y convergencia en probabilidad en el sentido de la norma $\| \ \|$.

Lema 1 *Sea* $\{A_n, n \in \mathbb{N}\}$ *una sucesión en* $\mathbb{R}^{m\times k}$, $A \in \mathbb{R}^{m\times k}$. *Entonces:* $A_n \to A$ *en probabilidad coordenada a coordenada si y sólo si* $\| A_n - A \| \to 0$ *en probabilidad.*

Demostración: i) En primer lugar, dada la invariancia por translaciones de las distancias inducidas por normas, basta probar que:

$$A_n \to 0 \text{ en probabilidad coordenada a coordenada si y solamente si } \| A_n \| \to 0 \text{ en probabilidad.}$$

ii) supongamos que $A_n \to 0$ en probabilidad coordenada a coordenada, esto significa que

$$|\ (A_n)_{ij}\ | \to 0 \text{ en probabilidad, para } i = 1, 2, ..., n \text{ y } j = 1, 2, .., n$$

sean $\epsilon > 0$ y $\delta > 0$, debemos probar que existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces

$$Probabilidad\,(\text{máximo}\{|\ (A_n)_{ij}\ |,\ i = 1, ..., m, j = 1, ..., k\} \geq \epsilon) \leq \delta.$$

En virtud de la hipótesis de convergencia coordenada a coordenada, dados $\epsilon > 0$ y $\delta > 0$ existen naturales $n_{ij},\ i = 1, ..., m, j = 1, ..., k$ tales que si $n \geq n_{ij}$ entonces

$$Probabilidad\,(\{|\ (A_n)_{ij}\ | \geq \epsilon\}) \leq \frac{\delta}{m \times k},$$

sea $n_0 = \text{máximo}\{n_{ij},\ i = 1, ..., m, j = 1, ..., k\}$, si $n \geq n_0$

$$Probabilidad\,(\{|\ (A_n)_{ij}\ | < \epsilon\}) \geq 1 - \frac{\delta}{m \times k}\ ,\ i = 1, ..., m, j = 1, ..., k\ ,$$

entonces

$$Probabilidad\,(\text{máximo}\{\mid (A_n)_{ij} \mid,\ i=1,...,m, j=1,...,k\} < \epsilon) =$$

$$Probabilidad\left(\bigcap_{i=1,...,m,j=1,...,k} \mid (A_n)_{ij} \mid < \epsilon\right) \geq 1-\delta.$$

iii) Finalmente, supongamos que $\parallel A_n \parallel \rightarrow\ 0$, o sea,

$$\text{máximo}\{\mid (A_n)_{ij} \mid, i=1,...,m, j=1,...,k\} \rightarrow 0 \quad \text{en probabilidad,}$$

ahora debemos probar que $\mid (A_n)_{uv} \mid \rightarrow 0$ for $u=1,...,m, v=1,...,k$; sean $\epsilon > 0$ y $\delta > 0$, por hipotesis existe un natural n_0 tal que si $n \geq n_0$ entonces

$$Probabilidad\,(\text{máximo}\{\mid (A_n)_{ij} \mid,\ i=1,...,m, j=1,...,k\} \geq \epsilon) \leq \delta$$

y esto implica que para $n \geq n_0$

$$Probabilidad\,(\{\mid (A_n)_{uv} \mid\} \geq \epsilon\}) \leq \delta$$

para cada par (u,v) con $u=1$, ..., m, $v=1$,..., k •

4 Distribución Asintótica de T^2

4.1 Una consecuencia del Teorema de Slustzky

La siguiente proposición es una consecuencia del resultado conocido como Teorema de Slutzky, que afirma lo siguiente:

si U_n, U, V son variables aleatorias, c es una constante, $U_n \rightarrow U$ en distribución y $V_n \rightarrow c$ en probabilidad, entonces $U_n + V_n \rightarrow U + c$ en distribución.

Proposición 2 *Sean $\{Y_n, n \in \mathbb{N}\}$ e Y vectores aleatorios p-dimensionales, $\{A_n, n \in \mathbb{N}\}$ un subconjunto aleatorio de $\mathbb{R}^{p \times p}$ y A un elemento fijo de $\mathbb{R}^{p \times p}$. Si $Y_n \rightarrow Y$ en distribución y $A_n \rightarrow A$ en probabilidad, entonces*

$$(Y_n)' A_n\ Y_n \rightarrow Y' A\ Y \quad \textit{en distribución.}$$

Demostración: i) Caso $A = 0$: debemos probar aquí que $(Y_n)' A_n Y_n \to 0$ en distribución, que es lo mismo que $(Y_n)' A_n \ Y_n \to 0$ en probabilidad (porque 0 es una constante); equivalentemente, debemos probar que dados $\epsilon > 0$ y $\delta > 0$, existe un natural n_0 tal que

$$Probabilidad\,(\{\| \ (Y_n)' A_n \, Y_n \ \| \leq \epsilon\}) = Probabilidad\,(\{| \ (Y_n)' A_n \, Y_n \ | \leq \epsilon\}) \geq 1 - \delta$$

para todo $n \geq n_0$.

Por hipótesis, la sucesión $(Y_n, n \in \mathbb{N})$ converge en distribución, entonces la familia $\{P(Y_n)^{-1}, n \in N\}$ es un conjunto relativamente compacto en el espacio de las medidas de probabilidad definidas en los borelianos $\mathbb{R}^p$ y por lo tanto "tight"; esto implica que dado $\delta > 0$, existe um subconjunto compacto K en $\mathbb{R}^p$ tal que

$$(P(Y_n)^{-1})(K) \geq 1 - \frac{\delta}{2} \quad \text{para todo} \quad n \in \mathbb{N}.$$

Dado que los conjuntos compactos de los espacios métricos son acotados, existe un número real $M > 0$ tal que el conjunto K está contenido en $[-M, M]^p$ y entonces

$$(P(Y_n)^{-1})([-M, M]^p) \geq 1 - \frac{\delta}{2} \quad \text{para todo} \quad n \in \mathbb{N}$$

lo que equivale a decir

$$Probabilidad\,(\{\| \ Y_n \ \| \leq M\}) \geq 1 - \frac{\delta}{2} \quad \text{para todo} \quad n \in \mathbb{N}; \tag{2}$$

dado que $A_n \to 0$ en probabilidad, dados $\epsilon > 0$, $M > 0$ y p, existe un natural n_0 tal que

$$Probabilidad\left(\left\{\| \ A_n \ \| \leq \frac{\epsilon}{M^2 p^2}\right\}\right) \geq 1 - \frac{\delta}{2} \quad \text{for all} \quad n \geq n_0. \tag{3}$$

Finalmente, en vista de las desigualdades (1), (2) y (3), podemos decir que si $n \geq n_0$

$$\begin{aligned} &Probabilidad\ \left(\{\| (Y_n)' A_n\ Y_n \| \leq \epsilon\}\right) \geq \\ &Probabilidad\ \left(\left\{\| Y_n \| \leq M\right\} \bigcap \left\{\| A_n \| \leq \frac{\epsilon}{M^2 p^2}\right\}\right) \geq 1 - \delta. \end{aligned}$$

ii) Caso general: es fácil verificar que

$$(Y_n)' A_n\ Y_n = (Y_n)'(A_n - A)\ Y_n + (Y_n)' A\ Y_n.$$

Observemos que el primer término del segundo miembro de la igualdad abajo ($(Y_n)'(A_n - A)\ Y_n$) converge en probabilidad a 0 (por i)) y que el segundo ($(Y_n)'A\ Y_n$) converge en distribución a $Y'A\ Y$ (porque $Y_n \to Y$ en distribución y la función $g(Y) = Y'A\ Y$ es continua en Y).

La prueba se completa si aplicamos a las variables aleatorias

$$(Y_n)'(A_n - A) Y_n \text{ e } (Y_n)' A\ Y_n$$

el teorema de Slutzky antes mencionado •

4.2 Distribución asintótica de T^2

Teorema 3 *Sea X un vector aleatorio p-dimensional con distribución continua, vector de medias μ, matriz de covariancia positiva definida Σ y momentos finitos de orden 4. Considere una muestra de n observaciones independientes e idénticamente distribuídas $X_1, X_2, ..., X_n$ del vector X. Sean $\bar{X}(n)$ y $S(n)$ el vector de medias y matriz de covariancia muestrales, respectivamente (definidos en 1.1). Sea T^2 la estadística de Hotelling,o sea:*

$$T^2 = n(\bar{X}(n) - \mu)'(S(n))^{-1}(\bar{X}(n) - \mu).$$

Entonces la distribución asintótica de T^2 es una chi-cuadrado con p grados de libertad.

Demostración: i) Por hipótesis, la distribución de X es continua y la matriz de covariancia Σ es definida positiva, entonces

$$\text{Probabilidad } (\{S(n) \text{ es definida positiva }\}) = 1$$

como puede ser visto en Seber (1984), p.8 y 522. Por lo tanto no nos preocuparemos con la inversibilidad de $S(n)$.

La existencia de Σ implica la finitud de $\Sigma_{ij}, 1 \leq i \leq p, 1 \leq j \leq p$ y también sabemos que $E((S(n))_{ij}) = \Sigma_{ij}$. Entonces obtenemos:

$$(S(n))_{ij} \rightarrow \Sigma_{ij} \quad \text{en probabilidad, } \ i = 1, 2, ..., p \quad \text{and} \quad j = 1, 2, ..., p.$$

por aplicación de la la ley de los grandes números versión Khinchin, que afirma que si $Y_1, Y_2, ..., Y_n$ son variables aleatorias independientes e identicamente distribuídas con media finita τ entonces la media muestral $\bar{Y}(n)$ converge en probabilidad a τ (veja Gnedenko (1978)) .

Por lo tanto, dado que la función $A \rightarrow A^{-1}$ $(\mathbb{R}^{p \times p} \rightarrow \mathbb{R}^{p \times p})$ es continua en la matriz inversible Σ, tenemos:

$$((S(n))^{-1})ij \rightarrow ((\Sigma)^{-1})ij \quad \text{en probabilidad,} \tag{4}$$

$i =$ 1, 2,..., p and $j =$ 1, 2,...,p.

Como consecuencia de (4) y del Lema de 1.4, tenemos que:

$$\| ((S(n))^{-1}) - ((\Sigma)^{-1}) \| \rightarrow 0 \quad \text{en probabilidad.} \tag{5}$$

ii) Por otra parte, el Teorema Central del Límite permite decir que:

$$n^{\frac{1}{2}}(\bar{X}(n) - \mu) \rightarrow Normal_p(0, \Sigma) \quad \text{en distribución.} \tag{6}$$

iii) Finalmente, dados (5), (6) y la Proposición 2 de 4.1, tenemos que:

$$T^2 = n^{\frac{1}{2}}(\bar{X}(n) - \mu)'(S(n))^{-1}n^{\frac{1}{2}}(\bar{X}(n) - \mu) \to W'\Sigma^{-1}W,$$

donde W tiene distribución Normal$_p$(0,Σ) y entonces T^2 tiene distribución chi-cuadrado con p grados de libertad •

References

Anderson, T. W., 1971. An Introduction to Multivariate Statistical Analysis, John Wiley and Sons.

Gnedenko, B., 1978. The Theory of Probability, Mir Publishers, Moscow.

Hotelling, H., 1931. The generalization of Student´s ratio, *Annals of Mathematical Statistics* 2, 330–378.

Johnson, R. A. and Wichern, D. W., 1998. Applied Multivariate Analysis, Prentice Hall.

Prohorov, Yu. V., 1956. Convergence of random processes and limit theorems in probability theory, *Theor. Probability Appl.* 1, 157–214.

Seber, G. A. F., 1984. Multivariate Observations, Wiley Series in probability and mathematical statistics, John Wiley and Sons.

Depto Estatística, IMECC-UNICAMP, Cidade Universitária Zeferino Vaz,
Barão Geraldo, 13083-859, Campinas, SP, Brasil