Asymptotic Robustness of Hotteling's Statistic

Gneri, Mario Antonio and Pimentel Barbosa, Emanuel
IMECC, UNICAMP, Campinas, SP, Brasil

Abstract

We show that under quite general distributional conditions the asymptotic distribution of Hotteling's T^2 statistic is the chi-square distribution, which is also the asymptotic distribution of T^2 under the assumption of normality.

Keywords and phrases: Hotteling's T^2, nonnormality, robustness.

1 Introduction

1.1 Hotteling's T^2 statistic under normality

Let X be a p-variate random vector with mean μ and variance-covariance matrix Σ. A sample of n independent and identically distributed observations $X_1, X_2, ..., X_n$ of vector X is taken and the sample mean vector $\bar{X}(n)$ and variance-covariance matrix $S(n)$ are calculated, where:

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

Hotteling's T^2 statistic is defined by:

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

It is well known that if the sample vectors have a common multivariate $\text{Normal}_p(\mu,\Sigma)$ distribution the statistic T^2 is distributed as $\frac{(n-1)p}{(n-p)}\, F_{p,n-p}$ and that when n tends to infinity the limit distribution of T^2 is chi-square with p degrees of freedom.

We will show that, replacing the normal distribution by any continuous distribution with finite second order moments, the asymptotic distribution of Hotteling's T^2 statistic is also the chi-square distribution with p degrees of freedom.

There exists a recent work - as in G. Willems and others (2002) - on alternative robust versions of the Hotteling's T^2 statistic in the finite sample context. Also the asymptotic robustness of T^2 is often mentioned (as in the book of Johnson and Wichern, (1998), page 187), but we believe that a rigorous proof of this fact is still lacking.

1.2 Notation, norms and convergence

- $\mathbb{N}$ and $\mathbb{R}$ denote the sets of natural and real numbers, respectively.

- Let $m \in \mathbb{N}$ and $k \in \mathbb{N}$, if $A \in \mathbb{R}^{m\times k}$, then A' denotes the transpose of A.
- Define the norm of $A = [a_{ij}] \in \mathbb{R}^{m\times k}$ by:

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

It is easy to see that if $A \in \mathbb{R}^{m\times k}$ and $B \in \mathbb{R}^{k\times t}$, then

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

for m, k and t natural numbers.

- As usually, we say that a sequence A_n in $\mathbb{R}^{m\times k}$ converges to $A \in \mathbb{R}^{m\times k}$ if and only if $\| A_n - A \|$ converges to zero. We remark that all metrics induced by norms are equivalent in $\mathbb{R}^{m\times k}$, in particular that convergence in the norm $\| \, \|$ is equivalent to coordinatewise convergence.
- If Z is a p-variate random vector, then PZ^{-1} denotes the probability distribution induced by Z in $\mathbb{R}^p$.

2 Tightness in general spaces of probability measures and convergence in probability in $\mathbb{R}^{m\times k}$

2.1 Tightness and relative-compactness of a family of distributions

Let $\mathcal{F}$ be the space of finite measures defined on the σ-algebra of the Borel sets of a complete separable metric space (Ω, d) and denote by $\mathcal{T}$ the topology of convergence in distribution (or the same, in law) in $\mathcal{F}$.

Prohorov (1957)[2] proved that exists a metric π in $\mathcal{F}$ such that it induces $\mathcal{T}$ and that $(\mathcal{F}, \pi)$ is separable and complete.

Prohorov also caracterized the relatively compact sets of $(\mathcal{F}, \mathcal{T})$. In the case of a family $\mathcal{P}$ of probability measures, Prohorov proved that $\mathcal{P}$ is relatively compact if and only if the family is tight, that is, for any $\epsilon > 0$ there exists a compact set $K = K(\epsilon, \mathcal{P})$ in (Ω, d) such that

$$P(K) > 1 - \epsilon \quad \text{for all} \quad P \quad \text{in} \quad \mathcal{P}.$$

We use this property in the proof of the Proposition of Subsection 2.1.

2.2 Convergence in probability in $\mathbb{R}^{m\times k}$

In this subsection we prove de equivalence between coordinatewise convergence in probability and convergence in probability in the sense of de norm $\| \, \|$.

Lemma

Let $\{A_n, n \in \mathbb{N}\}$ a sequence in $\mathbb{R}^{m\times k}$, $A \in \mathbb{R}^{m\times k}$. Then: $A_n \rightarrow A$ coordinatewise in probability if and only if $\| A_n - A \| \rightarrow 0$ in probability.

Proof

i) First, because of the invariance under translations of the distances induced by norms, it suffices to prove that:
$A_n \to 0$ coordinatewise in prob. if and only if $\| A_n \| \to 0$ in prob. ;

ii) Now, suppose that $A_n \to 0$ coordinatewise in probability, that is that $| (A_n)_{ij} | \to 0$ in prob., for $i = 1, 2, ..., n$ and $j = 1, 2, .., n$

let $\epsilon > 0$ and $\delta > 0$ be given, we must prove that exists $n_0 \in \mathbb{N}$ such that if $n \geq n_0$ then

$$Probability(maximum\{| (A_n)_{ij} |,\ i = 1, ..., m, j = 1, ..., k\} \geq \epsilon) \leq \delta \ ;$$

by the hipothesis of coordinatewise convergence, given $\epsilon > 0$ and $\delta > 0$ there exist natural numbers $n_{ij},\ i = 1, ..., m, j = 1, ..., k$ such that if n $\geq n_{ij}$ then

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

now call $n_0 = maximum\{n_{ij},\ i = 1, ..., m, j = 1, ..., k\}$,
then if $n \geq n_0$

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

hence

$$Probability(maximum\{| (A_n)_{ij} |,\ i = 1, ..., m, j = 1, ..., k\} < \epsilon) =$$

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

iii) Finally, suppose that $\| A_n \| \to 0$, that is,
$maximum\{ \ | (A_n)_{ij} |, i = 1, ..., m, j = 1, ..., k\} \to 0$ in probability ,
now we must prove that $| (A_n)_{uv} | \to 0$ for $u = 1, ..., m, v = 1, ..., k$;
let $\epsilon > 0$ and $\delta > 0$ be given, then by hypothesis there exists a natural number n_0 such that if $n \geq n_0$ then

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

and this implies that for $n \geq n_0$

$Probability(\{| (A_n)_{uv} |\} \geq \epsilon\}) \leq \delta$ for any pair (u,v) with $u = 1, ..., m, v = 1, ..., k$ •

3 Asymptotic normality of T^2

3.1 A consequence of Slustzky's Theorem

The following proposition is a consequence of the well known Slutzky's Theorem about convergence of random variables.

Proposition

Let $\{Y_n, n \in \mathbb{N}\}$ and Y be p-dimensional random vectors, $\{A_n, n \in \mathbb{N}\}$ a random subset of $\mathbb{R}^{p\times p}$ and A a fixed element of $\mathbb{R}^{p\times p}$.

If $Y_n \to Y$ in law and $A_n \to A$ in probability, then

$$(Y_n)' A_n \; Y_n \to Y' A \; Y \quad \text{in law} .$$

Proof

i) Case $A = 0$:

We must prove here that $(Y_n)' A_n Y_n \to 0$ in law, which is the same that $(Y_n)' A_n \; Y_n \to 0$ in probability (because 0 is a constant);

equivalently, we must prove that given $\epsilon > 0$ and $\delta > 0$, there exists a natural number n_0 such that

$$Probability(\{\| \; (Y_n)' A_n \; Y_n \; \| \leq \epsilon\}) = Probability(\{| \; (Y_n)' A_n \; Y_n \; | \leq \epsilon\}) \geq 1 - \delta \quad \text{for all} \quad n \geq n_0;$$

by hypothesis, the sequence $(Y_n, n \in \mathbb{N})$ is convergent in law, hence the family $\{P(Y_n)^{-1}, n \in N\}$ is a relatively compact set in the space of probability measures defined on de Borel sets of $\mathbb{R}^p$ and then tight; this implies that, given $\delta > 0$, there exists a compact set K in $\mathbb{R}^p$ such that

$$(P(Y_n)^{-1})(K) \geq 1 - \frac{\delta}{2} \quad \text{for all} \quad n \in \mathbb{N};$$

since compact sets are bounded in metric spaces, there exists a real number $M > 0$ such that the set K is contained in the set $[-M, M]^p$ and then

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

and this is equivalent to say

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

since $A_n \to 0$ in probability, given $\epsilon > 0$, $M > 0$ and p, there exists a natural number n_0 such that

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

Finally, in view of the equations (1), (2) and (3) we can say that

$$Probability\,(\{\| \; (Y_n)' A_n \; Y_n \; \| \leq \epsilon\}) \geq Probability\left(\left\{\| \; Y_n \; \| \leq M\} \bigcap \{\| \; A_n \; \| \leq \frac{\epsilon}{M^2 p^2}\right\}\right) \geq 1 - \delta.$$

ii) General case:

It is easy to see that $(Y_n)' A_n \; Y_n = (Y_n)'(A_n - A) \; Y_n + (Y_n)' A \; Y_n$;

observe now that the first term of the second member of the equality above ($(Y_n)'(A_n - A) \; Y_n$) converges in probability to 0 (by i)) and that the second one ($(Y_n)' A \; Y_n$) converges in law to $Y' A \; Y$ (because $Y_n \to Y$ in law and the function $g(Y) = Y' A \; Y$ is continuous in Y);

now apply to the random variables $(Y_n)'(A_n - A) Y_n$ and $(Y_n)' A \; Y_n$ the Slutzky's Theorem, say: if U_n, U, V are random variables, c is a constant, $U_n \to U$ in law and $V_n \to c$ in probability, then $U_n + V_n \to U + c$ in law •

3.2 Limit distribution of T^2

THEOREM

Let X be a p-variate random vector with continuous distribution, mean μ and positive definite variance-covariance matrix Σ. A sample of n independent and identically distributed observations $X_1, X_2, ..., X_n$ of vector X is taken. Let $\bar{X}(n)$ and $S(n)$ be the sample mean vector and sample variance-covariance matrix $S(n)$, respectively (defined in Subsection 1.1). Let T^2 be the Hotteling's statistic,that is:

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

Then the limit distribution of T^2 is chi-square with p degrees of freedom.

PROOF

i) By hypotheses, X has a continuous distribution and the variance-covariance matrix Σ is positive definite, then

$$\text{Probability } (\{S(n) \text{ is positive definite }\}) = 1$$

as may be seen in Seber(1984) ([3], pages 8 and 522), therefore we don't worry about the invertibility of $S(n)$;

The existence of Σ implies the finiteness of $\Sigma_{ij}, 1 \leq i \leq p, 1 \leq j \leq p$. We also have that $E((S(n))_{ij}) = \Sigma_{ij}$ and then the law of large numbers ensures that :

$$(S(n))_{ij} \rightarrow \Sigma_{ij} \text{ in probability , } i = 1, 2, ..., p \text{ and } j = 1, 2, ..., p \text{ ;}$$

hence, because the function $A \rightarrow A^{-1}$ ($\mathbb{R}^{p\times p} \rightarrow \mathbb{R}^{p\times p}$) is continuous in the invertible matrix Σ, we have:

$$((S(n))^{-1})ij \rightarrow ((\Sigma)^{-1})ij \text{ in probability } i = 1, 2, ..., p \text{ and } j = 1, 2, ..., p \text{ ,} \tag{4}$$

now, as a consequence of (4) and the Lemma of Subsection 1.4, we have:

$$\| ((S(n))^{-1}) - ((\Sigma)^{-1}) \| \rightarrow 0 \text{ in probability ;} \tag{5}$$

ii) Otherwise, by the Central Limit Theorem, we know that:

$$n^{\frac{1}{2}}(\bar{X}(n) - \mu) \rightarrow (\mu, \Sigma) \text{ in distribution ,} \tag{6}$$

iii) Finally, given (5), (6) and the Proposition of Subsection 2.1 we have:

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

where W is distributed as Normal$_p$(μ,Σ), and then T^2 is distributed as a chi-square with p degrees of freedom •

References

[1] Johnson, R. A. and Wichern, D. W. (1998). Applied Multivariate Analysis, Prentice Hall.

[2] Prohorov, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory, *Theor. Probability Appl.* 1, 157-214.

[3] Seber, G. A. F. (1984). Multivariate Observations, Wiley Series in probability and mathematical statistics, John Wiley and sons.

[4] Willems, G., Pison, G., Rousseeuw, P. J. and Van Aelst, S. (2002). A Robust Hotteling Test, *Metrika*, 125-138.