JOURNAL OF BASIC SCIENCES

Ley Tracy - Widom en Pruebas de Hipótesis para la Covarianza Poblacional

Didier Cortez Elizalde, Addy Margarita Bolivar Cimé*
División Académica de Ciencias Básicas, Universidad Juárez Autónoma de Tabasco,
Cunduacán, C.P. 86690, Tabasco, México.
*addy.bolivar@ujat.mx

Recibido: 14 de junio de 2018. Aceptado: 5 de septiembre de 2018. Publicado: 1 de diciembre de 2018.

Resumen

Es bien conocido en Análisis Multivariado que la estimación de la matriz de covarianza es complicada en el caso en que la dimensión de los datos es mayor o igual al tamaño de la muestra (caso de dimensión alta). En este trabajo se muestra cómo la distribución Tracy- Widom, que aparece en la Teoría de Matrices Aleatorias, puede ser utilizada para llevar a cabo pruebas de hipótesis de esfericidad o de igualdad de la matriz de covarianza poblacional a una matriz especificada, considerando datos normales de dimensión alta. Se muestra, mediante simulaciones, que estas pruebas estadísticas tienen un buen comportamiento en términos del tamaño de la prueba, tanto en el caso de dimensión alta como en el caso en que el tamaño de la muestra es mayor a la dimensión de los datos (caso clásico). También estudiamos, mediante simulaciones, el comportamiento de la potencia de la prueba.

Palabras clave: *Datos de dimensión alta, Distribución Tracy-Widom, Prueba de esfericidad, Simulación, Teoría de Matrices Aleatorias.*

Abstract

It is well known in Multivariate Analysis that the estimation of the population covariance matrix is complicated in the case when the dimension of the data is greater or equal than the sample size (high dimensional case). In this work, we show how to use the Tracy-Widom distribution, that appears in Random Matrix Theory, to carry out a sphericity test or testing that the population covariance matrix equals a specified matrix, considering high dimensional data from the normal distribution. We show, by simulations, that these hypothesis testing have a good behavior in terms of the size of the test, in the high dimensional case and the case where the sample size is larger tan the dimension of the data (classical case). We also study, by simulations, the behavior of the power of the test.

Keywords: *High dimensional data, Tracy-Widom distribution, Sphericity test, Simulation, Random Matrix Theory.*

---

1. Introducción

La Teoría de Matrices Aleatorias (TMA) estudia matrices cuyos elementos son variables aleatorias (o equivalentemente, variables aleatorias que toman valores en un espacio de matrices). La primera distribución propuesta para matrices aleatorias fue

Journal of Basic Sciences, Vol. 4 (12), pp. 1-13, sep-dic 2018

la *distribución Wishart*, nombrada así en honor a John Wishart, quien en 1928 la introdujo como una generalización de la distribución chi-cuadrada (ver [8]). A inicios de los 50's los físicos usaron modelos de matrices aleatorias para estudiar fenómenos cuánticos, a partir de esto se desarrolló una buena parte de la Teoría de Matrices Aleatorias que se conoce hoy en día, algunos pioneros fueron Eugene Wigner, Vladimir Marchenko y Leonid Pastur (ver [6]). Posteriormente esta teoría ha sido ampliamente desarrollada también por matemáticos y estadísticos (ver [2]).

Los datos multivariados de dimensión mayor o igual al tamaño de la muestra (datos multivariados de dimensión alta) aperecen en diversos campos, algunos de ellos son genética, análisis funcional, finanzas, análisis de imágenes médicas, climatología, reconocimiento de texto, entre otros (ver [4]). Cabe mencionar que en el contexto de datos multivariados de dimensión alta la estimación de la matriz de covarianza poblacional es complicada, ya que se tienen que estimar muchos parámetros con pocos datos, por lo que la estimación de esta matriz y pruebas de hipótesis acerca de ella requieren técnicas estadísticas diferentes a las del caso clásico donde el tamaño de la muestra es mayor a la dimensión de los datos.

La distribución Wishart es la distribución de la matriz de covarianza muestral de una muestra aleatoria de la distribución normal multivariada (ver [1]), debido a esto el estudio de la distribución Wishart ha sido de gran importancia en Análisis Multivariado. Un resultado proporcionado por Johnstone en [4], muestra que la distribución asintótica del eigenvalor más grande de una matriz aleatoria con distribución Wishart es la distribución Tracy-Widom. Mediante este resultado es posible llevar a cabo pruebas de hipótesis acerca de la matriz de covarianza poblacional de un conjunto de datos normales multivariados de dimensión alta, pero también es posible aplicar estas pruebas de hipótesis en el caso clásico.

En este trabajo de divulgación se muestra la forma en que la distribución Tracy-Widom se utiliza para llevar a cabo pruebas de hipótesis de la matriz de covarianza poblacional de datos normales multivariados. Se lleva a cabo un estudio de simulación, en términos del tamaño de la prueba y su potencia, para evaluar su comportamiento, tanto en el caso de dimensión alta como en el caso clásico. También se comparan la prueba de hipótesis basadas en la distribución Tracy-Widom con la prueba de razón de verosimilitud comúnmente usada en el caso clásico.

A continuación se describe la forma en que se divide este trabajo. En la Sección 2 se presenta la definición y algunos resultados importantes de la distribución Wishart; en la Sección 3 se presenta la distribución Tracy-Widom como la distribución asintótica del eigenvalor más grande de una matriz Wishart; en la Sección 4 se presentan las pruebas de hipótesis para la covarianza poblacional basadas en la distribución Tracy-Widom; en la Sección 5 se lleva a cabo un estudio de simulación para evaluar el comportamiento de estas pruebas; en la Sección 6 se compara la prueba basada en la ley Tracy-Widom con la prueba de razón de verosimilitud, considerando el caso clásico únicamente; finalmente en la Sección 7 se presentan las conclusiones.

2. Distribución Wishart

La distribución de Wishart, propuesta por John Wishart en [8], fue la primera distribución propuesta para matrices aleatorias. A continuación se presenta la definición

Journal of Basic Sciences, Vol. 4 (12), pp. 1-13, sep-dic 2018

de esta distribución.

Definición 1. Sea $A = Z'Z$, donde las filas de la matriz Z de $n \times p$ son independientes e idénticamente distribuidas (i.i.d.) con distribución $N_p(\mathbf{0}, \Sigma)$, entonces se dice que A tiene *distribución Wishart* con n grados de libertad y matriz de covarianza Σ. La notación utilizada es $A \sim W_p(n, \Sigma)$.

Puede verse que la matriz A de la definición anterior se puede escribir de la forma

$$A = \sum_{\alpha=1}^{n} Z_\alpha Z'_\alpha, \tag{1}$$

donde los vectores Z_α son i.i.d. $N_p(\mathbf{0}, \Sigma)$. De hecho los vectores Z_α son las filas de la matriz Z.

En el caso en que $n \geq p$ la función de densidad de la distribución $W_p(n, \Sigma)$ existe y está dada por el siguiente teorema, cuya demostración puede consultarse en [1].

Teorema 1. Si $A \sim W_p(n, \Sigma)$ con $n \geq p$, entonces la función de densidad de A es

$$f(B) = \frac{1}{2^{pn/2}\Gamma_p\left(\frac{1}{2}n\right)(\det\Sigma)^{n/2}} \operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}B\right)(\det B)^{(n-p-1)/2}$$

para $B > 0$ (B definida positiva), donde $\Gamma_p(\cdot)$ es la función gamma multivariada y $\operatorname{etr}(\cdot) = \exp(\operatorname{tr}(\cdot))$ es la composición de la exponencial con la traza.

El siguiente corolario nos proporciona la función de densidad de la matriz de covarianza muestral cuando los datos provienen de la distribución $N_p(\mu, \Sigma)$, y nos dice también que la matriz de covarianza muestral tiene distribución $W_p(n, \Sigma/n)$.

Corolario 1. Si $X_1, ..., X_N$ son vectores aleatorios independientes $N_p(\mu, \Sigma)$ y $N > p$, la función de densidad de la matriz de covarianza muestral

$$S_n = \frac{1}{n}\sum_{i=1}^{N}(X_i - \overline{X})(X_i - \overline{X})', \tag{2}$$

donde $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ y $n = N - 1$, es

$$f(S) = \frac{1}{\Gamma_p\left(\frac{1}{2}n\right)(\det\Sigma)^{n/2}}\left(\frac{1}{2}n\right)^{pn/2} \operatorname{etr}\left(-\frac{1}{2}n\Sigma^{-1}S\right)(\det S)^{(n-p-1)/2},$$

para $S > 0$ (S definida positiva).

Prueba. Notemos que podemos escribir S_n como

$$S_n = \sum_{i=1}^{n}[(1/\sqrt{n})Z_i][(1/\sqrt{n})Z_i]',$$

donde $(1/\sqrt{n})Z_1, ..., (1/\sqrt{n})Z_n$ son i.i.d. $N_p(\mathbf{0}, \Sigma/n)$. Así por la representación (1) de la distribución Wishart tenemos que $S_n \sim W_p(n, \Sigma/n)$. ■

Como caso particular del corolario anterior, para $p = 1$ la distribución Wishart es igual a una distribución ji-cuadrada con n grados de libertad (χ^2_n). De este modo la distribución Wishart puede considerarse una generalización matricial de la distribución ji-cuadrada, de hecho la distribución Wishart tiene propiedades similares a la de una ji-cuadrada (ver, por ejemplo, [1]).

3. Distribución Tracy-Widom

El siguiente teorema, debido a Johnstone [4], muestra cuál es la distribución asintótica del eigenvalor más grande de una matriz Wishart. Este resultado será utilizado en la siguiente sección para realizar pruebas de hipótesis acerca de la matriz de covarianza poblacional de datos normales multivariados.

Teorema 2. Sea $A \sim W_p(n, I_p)$ y sea l_1 el eigenvalor más grande de A. Si $p/n \to \gamma > 0$ cuando $n \to \infty$, entonces

$$\frac{l_1 - \mu_{np}}{\sigma_{np}} \xrightarrow[dist]{} F_1,$$

donde las constantes de centralización y escala son

$$\mu_{np} = \left(\sqrt{n-1} + \sqrt{p}\right)^2,$$
$$\sigma_{np} = \left(\sqrt{n-1} + \sqrt{p}\right)\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{1/3},$$

y F_1 es la función de distribución de la *ley Tracy-Widom (T-W)* definida como

$$F_1(s) = \exp\left(-\frac{1}{2}\int_s^\infty q(x) + (x-s)q^2(x)dx\right), \quad s \in \mathbb{R},$$

donde q es solución de la ecuación diferencial de Painlevé II

$$q''(x) = xq(x) + 2q^3(x), \quad q(x) \sim \mathrm{Ai}(x), \quad \text{cuando} \quad x \to +\infty,$$

y $\mathrm{Ai}(x)$ es la función de Airy, la cual satisface la ecuación diferencial $\mathrm{Ai}''(x) - x\mathrm{Ai}(x) = 0$, $\mathrm{Ai}(x) \to 0$ cuando $x \to \infty$ y $\mathrm{Ai}(x) \sim \frac{1}{2\sqrt{\pi}x^{1/4}} e^{-\frac{2}{3}x^{3/2}}$ cuando $x \to \infty$.

La función de distribución F_1 proviene de una familia de distribuciones F_β, donde $\beta = 1, 2, 4$. La función F_β con $\beta = 1, 2$ y 4 aparece como la distribución del eigenvalor más grande de los conjuntos *Gaussian Othogonal Ensemble (GOE)*, *Gaussian Unitary Ensemble (GUE)* y *Gaussian Symplectic Ensemble (GSE)*, respectivamente, los cuales aparecen en la Teoría de Matrices Aleatorias (ver [6]).

En [5] se sugiere usar las *constantes de centralización y escala de segundo orden*

Journal of Basic Sciences, Vol. 4 (12), pp. 1-13, sep-dic 2018

para una convergencia más rápida, dadas por

$$\mu'_{np} = \left(\sqrt{n-\frac{1}{2}} + \sqrt{p-\frac{1}{2}}\right)^2, \tag{3}$$

$$\sigma'_{np} = \left(\sqrt{n-\frac{1}{2}} + \sqrt{p-\frac{1}{2}}\right)\left(\frac{1}{\sqrt{n-\frac{1}{2}}} + \frac{1}{\sqrt{p-\frac{1}{2}}}\right)^{1/3}. \tag{4}$$

El Teorema 2 es válido usando las constantes μ'_{np} y σ'_{np}, es decir, se tiene el siguiente resultado.

Teorema 3. Bajo las mismas condiciones del teorema anterior, si $p/n \to \gamma > 0$ cuando $n \to \infty$, entonces

$$\frac{l_1 - \mu'_{np}}{\sigma'_{np}} \xrightarrow[dist]{} F_1,$$

donde μ'_{np} y σ'_{np} son las constantes de centralización y escala definidas anteriormente.

4. Pruebas de hipótesis para la covarianza poblacional basada en la ley T-W

En esta sección se verá la aplicación de la distribución Tracy-Widom para llevar a cabo pruebas de hipótesis sobre la matriz de covarianza poblacional de datos normales multivariados.

4.1 Prueba de hipótesis para $H_0 : \Sigma = I_p$

Como se muestra en [4], el resultado del Teorema 2 proporciona una herramienta para llevar a cabo pruebas de hipótesis de la matriz de covarianza a partir del eigenvalor más grande de la matriz de covarianza muestral (eigenvalor muestral más grande), en particular para probar

$$H_0 : \Sigma = I_p \quad \text{vs} \quad H_1 : \Sigma \neq I_p. \tag{5}$$

Debido al Teorema 3, si tenemos las observaciones de una muestra aleatoria de tamaño n de una normal multivariada $N_p(\mathbf{0}, \Sigma)$ y l_1 es el eigenvalor muestral más grande, una prueba de hipótesis de nivel α para contrastar las hipótesis (5) es rechazar H_0 si

$$\frac{nl_1 - \mu'_{np}}{\sigma'_{np}}$$

es mayor que el punto porcentual superior α de la distribución Tracy-Widom F_1, denotado por $F_1(\alpha)$, donde μ'_{np} y σ'_{np} son las constantes de centralización y escala dadas por (3) y (4), respectivamente.

Sean $X_1, X_2, ..., X_N$ vectores aleatorios i.i.d. con distribución $N_p(\mu, \Sigma)$. Supongamos ahora que estamos interesados en el juego de hipótesis

$$H_0 : \Sigma = \sigma^2 I_p \quad \text{vs} \quad H_1 : \Sigma \neq \sigma^2 I_p. \tag{6}$$

La hipótesis H_0 anterior es conocida como *hipótesis de esfericidad*. Sea S_n la matriz de covarianza muestral dada por (2). Si σ^2 es conocido, entonces bajo H_0 se tiene que $A = nS_n/\sigma^2 \sim W_p(n, I_p)$ (ver Sección 2). Por lo que en este caso, para contrastar las hipótesis (6) se puede aplicar la prueba de hipótesis basada en la ley T-W utilizando al eigenvalor más grande de A. Como se menciona en [9], en el caso en que σ^2 es desconocido puede ser estimado con $\widehat{\sigma^2} = trS_n/p$, y al considerar que $\widehat{A} = nS_n/\widehat{\sigma^2}$ tiene distribución aproximada $W_p(n, I_p)$, se puede aplicar la prueba hipótesis basada en la ley T-W utilizando el eigenvalor más grande de $\widehat{A}$ para contrastar el juego de hipótesis (6).

4.2 Prueba de hipótesis para $H_0 : \Sigma = \Sigma_0$.

Sean $X_1, X_2, ..., X_N$ vectores aleatorios i.i.d. con distribución $N_p(\mu, \Sigma)$. Consideremos ahora el juego de hipótesis

$$H_0 : \Sigma = \Sigma_0 \quad \text{vs} \quad H_1 : \Sigma \neq \Sigma_0, \tag{7}$$

donde Σ_0 es una matriz positiva definida especificada. Definamos $Y_i = \Sigma_0^{-1/2} X_i$, $i = 1, 2, \ldots, N$, los cuales son vectores aleatorios i.i.d. con distribución normal multivariada $N_p(\Sigma_0^{-1/2}\mu, \Sigma_0^{-1/2}\Sigma\Sigma_0^{-1/2})$. Observemos que bajo la hipótesis nula, las Y_i son i.i.d. con distribución $N_p(\Sigma_0^{-1/2}\mu, I_p)$. Debido a lo anterior, para contrastar las hipótesis (7) se puede aplicar la prueba de hipótesis basada en la ley T-W utilizando al eigenvalor más grande de la matriz de covarianza muestral de las Y_i, que como se mostrará en el siguiente lema, es igual al eigenvalor más grande de $\Sigma_0^{-1} S_n$, donde S_n es la matriz de covarianza muestral de las X_i.

Lema 1. Sean X_i y Y_i, con $i = 1, 2, \ldots, N$, como anteriormente. Denotamos por $l_1(A)$ al eigenvalor más grande de una matriz A. Sea S_n la matriz de covarianza muestral de las X_i. Entonces la matriz de covarianza muestral de las Y_i está dada por $\Sigma_0^{-1/2} S_n \Sigma_0^{-1/2}$ y

$$l_1(\Sigma_0^{-1} S_n) = l_1(\Sigma_0^{-1/2} S_n \Sigma_0^{-1/2}).$$

Prueba. Veamos que la matriz de covarianza muestral de las Y_i está dada por $\Sigma_0^{-1/2} S_n \Sigma_0^{-1/2}$. Tenemos que

$$\begin{aligned}
\Sigma_0^{-1/2} S_n \Sigma_0^{-1/2} &= \Sigma_0^{-1/2} \frac{1}{n} \sum_{i=1}^{N} (X_i - \overline{X})(X_i - \overline{X})' \Sigma_0^{-1/2} \\
&= \frac{1}{n} \sum_{i=1}^{N} \Sigma_0^{-1/2} (X_i - \overline{X})(X_i - \overline{X})' (\Sigma_0^{-1/2})' \\
&= \frac{1}{n} \sum_{i=1}^{N} (\Sigma_0^{-1/2} X_i - \Sigma_0^{-1/2} \overline{X})(\Sigma_0^{-1/2} X_i - \Sigma_0^{-1/2} \overline{X})'.
\end{aligned}$$

Journal of Basic Sciences, Vol. 4 (12), pp. 1-13, sep-dic 2018

Notar que $\Sigma_0^{-1/2}\overline{X} = \frac{1}{N}\sum_{i=1}^{N}\Sigma_0^{-1/2}X_i = \overline{Y}$, por lo que la última expresión de arriba es igual a

$$\widehat{S} = \frac{1}{n}\sum_{i=1}^{N}(Y_i - \overline{Y})(Y_i - \overline{Y})',$$

la cual es la matriz de covarianza muestral de las Y_i. Ahora mostraremos que

$$l_1(\Sigma_0^{-1}S_n) = l_1(\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}).$$

Observemos que si λ es un eigenvalor de $\Sigma_0^{-1}S_n$, entonces para algún $v \neq \mathbf{0}$ se tiene que

$$\Sigma_0^{-1}S_n v = \lambda v \iff \Sigma_0^{-1/2}S_n v = \lambda\Sigma_0^{1/2}v \iff \Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}(\Sigma_0^{1/2}v) = \lambda(\Sigma_0^{1/2}v).$$

Tomando $\widehat{v} = \Sigma_0^{1/2}v$ en la expresión anterior, tenemos que

$$\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}\widehat{v} = \lambda\widehat{v}.$$

Ahora veamos que $\widehat{v} \neq \mathbf{0}$. Debido a que $\Sigma_0^{1/2}$ es una matriz invertible sus columnas son linealmente independientes, por lo que si $\widehat{v} = \Sigma_0^{1/2}v = \mathbf{0}$, entonces $v = \mathbf{0}$, lo cual es una contradicción ya que habíamos supuesto que $v \neq \mathbf{0}$. Así

$$\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}\widehat{v} = \lambda\widehat{v}, \quad \text{con} \quad \widehat{v} \neq \mathbf{0},$$

por lo tanto λ es un eigenvalor de $\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}$. Análogamente se puede mostrar que si λ es un eigenvalor de $\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}$ entonces también es un eigenvalor de $\Sigma_0^{-1/2}S_n$. De este modo los eigenvalores de $\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2}$ y $\Sigma_0^{-1/2}S_n$ son equivalentes y se tiene que $l_1(\Sigma_0^{-1}S_n) = l_1(\Sigma_0^{-1/2}S_n\Sigma_0^{-1/2})$. ■

Como se mencionó anteriormente, el problema de probar la hipótesis nula en (7) se traduce en probar la hipótesis nula de que la matriz de covarianza de las Y_i es la identidad. Para realizar la prueba de hipótesis con la ley T-W se debe calcular entonces el eigenvalor más grande de la matriz de covarianza muestral de las Y_i, que por el lema anterior es precisamente $l_1(\Sigma_0^{-1}S_n)$. A partir de aquí el procedimiento para aplicar la prueba de hipótesis usando la ley T-W es como se describió en la Sección 4.1.

En resumen, para contrastar las hipótesis (7) lo que se hace es lo siguiente:

- Se calcula $l_1(\Sigma_0^{-1}S_n)$, que es igual al eigenvalor más grande de la matriz de covarianza muestral de las Y_i.
- Debido a que

$$\frac{nl_1(\Sigma_0^{-1}S_n) - \mu'_{np}}{\sigma'_{np}} \underset{dist}{\longrightarrow} F_1, \tag{8}$$

  donde μ'_{np} y σ'_{np} son las constantes de centralización y escala de segundo orden del Teorema 3, entonces rechazamos H_0 con un nivel de significancia α si el lado izquierdo de (8) es mayor que el punto porcentual superior α de la distribución Tracy-Widom F_1, denotado por $F_1(\alpha)$.

Journal of Basic Sciences, Vol. 4 (12), pp. 1-13, sep-dic 2018

5. Estudio de simulación para evaluar las pruebas de hipótesis basadas en la ley T-W

Por medio de simulaciones se evaluó qué tan buena es la prueba basada en la ley Tracy-Widom para determinar si muestras aleatorias normales multivariadas tienen una cierta matriz de covarianza.

Se consideró el juego de hipótesis (5) debido a que es el más simple. Para distintos valores de n y p se simularon $M = 1000$ muestras aleatorias de tamaño $N = n+1$ con distribución $N_p(\mathbf{0}, I_p)$ y se calculó la proporción de veces que se rechazó la hipótesis nula (el tamaño de la prueba) para varios niveles de significancia, si la prueba es buena estas proporciones deben parecerse a los niveles de significancia. No se usó un valor mayor de M ya que al tomar valores grandes de p o n las simulaciones eran muy tardadas, además se vió (cosiderando algunos casos) que al tomar un valor de M más grande a 1000 los resultados no variaban de forma significativa. Se consideraron casos en que $\gamma = p/n > 1$ (caso de dimensión alta) y en que $\gamma = p/n \leq 1$ (caso clásico). Los niveles de significancia considerados fueron $\alpha = 0.1,\ 0.05,\ 0.01$. Los resultados se exponen en los Cuadros 1–4.

α	n=20, p=10	n=200, p=100	n=1000, p=500	n=2000, p=1000
0.1	0.1	0.091	0.102	0.097
0.05	0.052	0.044	0.044	0.04
0.01	0.009	0.013	0.013	0.006

Cuadro 1. Tamaño de la prueba para el caso $\gamma = 1/2$.

α	n=10, p=10	n=100, p=100	n=500, p=500	n=1000, p=1000
0.1	0.097	0.102	0.11	0.102
0.05	0.047	0.045	0.061	0.051
0.01	0.016	0.009	0.013	0.009

Cuadro 2. Tamaño de la prueba para el caso $\gamma = 1$.

α	n=10, p= 20	n=100, p=200	n=500, p=1000	n=1000, p=2000
0.1	0.1	0.103	0.102	0.103
0.05	0.049	0.054	0.052	0.054
0.01	0.012	0.01	0.009	0.012

Cuadro 3. Tamaño de la prueba para el caso $\gamma = 2$.

α	n=10, p=40	n=50, p=200	n=100, p=400	n=500, p=2000
0.1	0.093	0.086	0.108	0.102
0.05	0.044	0.043	0.048	0.039
0.01	0.012	0.01	0.014	0.007

Cuadro 4. Tamaño de la prueba para el caso $\gamma = 4$.

En general se observa que las proporciones en que se rechaza la hipótesis nula es cercana al nivel de significancia y no se observa ningún patrón al aumentar los valores de n y p, manteniendo fija la razón $\gamma = p/n$. Se concluye que la prueba de hipótesis basada en la ley T-W es una prueba aceptable en términos del tamaño, tanto en el caso de dimensión alta, que es el de mayor interés en este trabajo, como en el caso clásico.

Debido a que también resulta interesante estudiar la potencia de la prueba, se llevaron a cabo simulaciones para estimarla en dos casos, el primero cuando Σ es la matriz diagonal $\Sigma_1 = \text{diag}(1, 0.05, \ldots, 0.05)$; en el segundo caso se tomo a Σ como una matriz *Toeplitz* (ver [3]) de la forma $\Sigma_2 = Toeplitz(1, c, c^2, \ldots, c^{p-1})$ con $c = 0.5$, la cual es definida positiva. Se consideraron los mismos valores de M, n, p y α que para las simulaciones del tamaño de la prueba. En el primer caso, todas las proporciones de rechazo de la hipótesis nula (potencia de la prueba) fueron cero, lo cual indica que para ese caso la probabilidad del error de tipo II es prácticamente uno. Sin embargo, para el segundo caso se obtuvieron resultados muy diferentes. En los Cuadros 5–8 se presenta la potencia de la prueba para el segundo caso.

α	n=20, p=10	n=200, p=100	n=1000, p=500	n=2000, p=1000
0.1	0.735	1	1	1
0.05	0.646	1	1	1
0.01	0.463	1	1	1

Cuadro 5. Potencia con $\Sigma = Toeplitz(1, 0.5, 0.5^2, \ldots, 0.5^{p-1})$ para el caso $\gamma = 1/2$.

α	n=10, p=10	n=100, p=100	n=500, p=500	n=1000, p=1000
0.1	0.466	1	1	1
0.05	0.371	1	1	1
0.01	0.222	1	1	1

Cuadro 6. Potencia con $\Sigma = Toeplitz(1, 0.5, 0.5^2, \ldots, 0.5^{p-1})$ para el caso $\gamma = 1$.

α	n=10, p= 20	n=100, p=200	n=500, p=1000	n=1000, p=2000
0.1	0.526	1	1	1
0.05	0.427	1	1	1
0.01	0.261	1	1	1

Cuadro 7. Potencia con $\Sigma = Toeplitz(1, 0.5, 0.5^2, \ldots, 0.5^{p-1})$ para el caso $\gamma = 2$.

α	n=10, p=40	n=50, p=200	n=100, p=400	n=500, p=2000
0.1	0.573	1	1	1
0.05	0.465	1	1	1
0.01	0.278	0.998	1	1

Cuadro 8. Potencia con $\Sigma = Toeplitz(1, 0.5, 0.5^2, \ldots, 0.5^{p-1})$ para el caso $\gamma = 4$.

Se observa que, para todos los valores de γ considerados, cuando n y p crecen manteniendo fija la razón $\gamma = p/n$, la potencia de la prueba converge a uno, esto significa que la probabilidad del error de tipo II en este caso converge a cero. En conclusión, la potencia de la prueba basada en la ley T-W, dependiendo de la matriz Σ que se tome en el espacio parametral de la hipótesis alternativa, puede ser algunas veces muy mala y otras veces muy buena al hacer tender n y p a infinito manteniendo fija la razón $\gamma = p/n$.

6. Comparación con la prueba de razón de verosimilitud

6.1 Prueba de razón de verosimilitud

Una prueba muy utilizada en Análisis Multivariado clásico para contrastar las hipótesis (5) es la prueba de razón de verosimilitud. El siguiente resultado nos da la prueba de razón de verosimilitud para $H_0 : \Sigma = I_p$ (ver [7]).

Teorema 4. Sean $X_1, ..., X_N$ vectores aleatorios independiente con distribución $N_p(\mu, \Sigma)$ y sea

$$A = nS_n = \sum_{i=1}^{N} (X_i - \overline{X})(X_i - \overline{X})', \quad (n = N - 1).$$

La prueba de razón de verosimilitud de tamaño α para $H_0 : \Sigma = I_p$, rechaza H_0 si $\Lambda \leqslant c_\alpha$, donde

$$\Lambda = \left(\frac{\mathrm{e}}{N}\right)^{pN/2} \mathrm{etr}(-A/2)(\det A)^{N/2}.$$

En [1, 7] puede verse que la prueba de razón de verosimilitud es sesgada, sin embargo haciendo una ligera modificación al estadístico de razón de verosimilitud, la prueba es insesgada. Tal modificación del estadístico es

$$\Lambda^* = \left(\frac{\mathrm{e}}{n}\right)^{pn/2} \mathrm{etr}(-A/2)(\det A)^{n/2},$$

y es obtenido de Λ reemplazando el tamaño de muestra N por los grados de libertad n. Notar que

$$-2\log\Lambda^* = n[\mathrm{tr}(S_n) - \log(\det(S_n)) - p].$$

El siguiente resultado, que puede ser consultado en [7, pag. 359], proporciona una aproximación a la distribución de $-2\log\Lambda^*$ a través de la distribución ji-cuadrada.

Teorema 5. Cuando la hipótesis $H_0 : \Sigma = I_p$ es verdadera, la distribución de $-2\log\Lambda^*$ sigue aproximadamente una distribución ji-cuadrada con $f = p(p+1)/2$ grados de libertad, cuando n es grande, es decir,

$$\mathbb{P}(-2\log\Lambda^* \leqslant x) \approx \mathbb{P}(\chi_f^2 \leqslant x), \quad \forall x \in \mathbb{R}.$$

Utilizando esta aproximación, una prueba de hipótesis de nivel α para contrastar (5) es rechazar H_0 si $-2\log\Lambda^* > \chi_f^2(\alpha)$, donde $\chi_f^2(\alpha)$ es el punto porcentual superior α de la distribución ji-cuadrada con f grados libertad. Esta prueba se utiliza únicamente cuando $n \geq p$. Nótese que en el caso en que $n < p$ el determinante de S_n es cero y por lo tanto $\Lambda^* = 0$.

Journal of Basic Sciences, Vol. 4 (12), pp. 1-13, sep-dic 2018

Si nos interesa probar $H_0 : \Sigma = \Sigma_0$, donde Σ_0 es una matriz especificada, como se vio anteriormente esto es equivalente a mostrar que la matriz de covarianza de los datos transformados $Y_i = \Sigma_0^{-1/2} X_i$, $i = 1, 2, \ldots, N$, es la identidad.

6.2 Estudio de simulación para la comparación de las pruebas de hipótesis

A continuación se presenta un estudio de simulación para comparar el desempeño de la prueba de razón de verosimilitud con la prueba de hipótesis basada en la ley T-W. Recordemos que esta última prueba puede ser empleada cuando $n < p$ y cuando $n \geq p$.

Para contrastar el juego de hipótesis (5) se consideraron varios valores de p y n, y se simularon $M = 10000$ muestras aleatorias de tamaño $N = n + 1$ con distribución $N_p(\mathbf{0}, I_p)$. Los escenarios en las simulaciones contemplan $p = 10$ (un valor pequeño) y $p = 100$ (un valor grande), y para cada uno de estos valores se consideraron cinco valores de n tales que las mismas razones p/n son tomadas en cuenta con los dos valores de p, con el fin de ver el comportamiento de las pruebas hipótesis al hacer crecer tanto p como n, pero manteniendo la misma razón $\gamma = p/n$. En estas simulaciones se tomó ese valor de M ya que los valores considerados de n y p no eran muy grandes. Para cada prueba se calculó el tamaño y la potencia de la prueba, esta última considerando los mismos dos tipos de matriz de covarianza que en la Sección 5, $\Sigma_1 = \text{diag}(1, 0.05, \ldots, 0.05)$ y $\Sigma_2 = toeplitz(1, 0.5, 0.5^2, \ldots, 0.5^{p-1})$. Se tomó el nivel de significancia $\alpha = 0.05$.

Los resultados de las simulaciones se exponen en los Cuadros 9 y 10. En estos cuadros la prueba χ_f^2 se refiere a la prueba de razón de verosimilitud (basada en $-2 \log \Lambda^*$) y la prueba T-W se refiere a la prueba de hipótesis basada en la ley T-W.

	Tamaño		Potencia con Σ_1		Potencia con Σ_2	
$p = 10$	χ_f^2	T-W	χ_f^2	T-W	χ_f^2	T-W
$n = 11$	0.8656	0.0471	1	0	0.9942	0.3970
$n = 15$	0.5399	0.0456	1	0	0.9921	0.5132
$n = 20$	0.3293	0.0498	1	0	0.9965	0.6394
$n = 50$	0.1136	0.0521	1	0	1	0.9727
$n = 100$	0.0689	0.0497	1	0	1	0.9999

Cuadro 9. Comparación de las pruebas considerando p pequeña

	Tamaño		Potencia con Σ_1		Potencia con Σ_2	
$p = 100$	χ_f^2	T-W	χ_f^2	T-W	χ_f^2	T-W
$n = 110$	1	0.0495	1	0	1	1
$n = 150$	1	0.0462	1	0	1	1
$n = 200$	1	0.0494	1	0	1	1
$n = 500$	0.9753	0.0520	1	0	1	1
$n = 1000$	0.5408	0.0496	1	0	1	1

Cuadro 10. Comparación de las pruebas considerando p grande

Notemos que en las simulaciones solo se consideró el caso en que $\gamma = p/n < 1$ y se omitió el caso en que $\gamma > 1$ (caso de dimensión alta), debido a que cuando $p > n$

Journal of Basic Sciences, Vol. 4 (12), pp. 1-13, sep-dic 2018

algunos de los eigenvalores de la matriz S_n son cero y el estadístico Λ^* cero, por tanto $-2\log\Lambda^*$ no existe.

El tamaño de la prueba de razón de verosimilitud tiende a ser bueno cuando n es lo suficientemente grande con respecto a p, pero es muy malo en caso contrario, ya que no se aproxima al nivel de significancia considerado ($\alpha = 0.05$). Los peores resultados de esta prueba se observan cuando $p = 100$ y $n = 110,\ 150,\ 200$ (p grande y n no tan alejada de p), en estos casos el tamaño de la prueba es uno, mucho más grande que el nivel de significancia considerado. En el caso en que $p = 10$ y $n = 11,\ 15,\ 20$ (p pequeña y n no tan alejada de p), el tamaño es también más grande que el nivel de significancia considerado pero es un poco más pequeño que en los casos mencionados anteriormente.

Con respecto a la prueba basada en la ley T-W, el tamaño de la prueba se aproxima al nivel de significancia considerado aún cuando n no es tan grande con respecto a p. Al hacer crecer p y n manteniendo la misma razón $\gamma = p/n$, los resultados no varían mucho en mayoría de los casos. Vemos que la prueba de razón de verosimilitud tiene mayor tamaño a la basada en la ley T-W.

Al comparar la potencia de las pruebas, vemos que la potencia de la prueba de razón de verosimilitud es mayor o igual a la potencia de la prueba basada en la ley T-W, para los dos casos considerados de Σ. De hecho la potencia de la prueba de razón de verosimilitud es uno o cercana a uno en la mayoría de los casos, mientras que la potencia de la prueba basada en la ley T-W es cero para Σ_1 y todos los valores de n y p considerados, y mejora para Σ_2 principalmente cuando n y p son grandes, que es cuando vale uno.

7. Conclusiones

Las simulaciones llevadas a cabo en este trabajo para verificar el comportamiento de las pruebas de hipótesis basadas en la ley T-W indican que esta prueba es aceptable en términos del tamaño de la prueba, tanto en el caso de dimensión alta como en el caso clásico, ya que los tamaños de las pruebas fueron cercanos a los niveles de significancia considerados. Sin embargo, la potencia de la prueba puede ser muy buena o muy mala, dependiendo de la matriz de covarianza que se tome en el espacio parametral de la hipótesis alternativa. También se realizaron simulaciones, considerando únicamente el caso clásico, para comparar la prueba de hipótesis basada en la ley T-W con la prueba de razón de verosimilitud. En estas simulaciones se observó que la prueba basada en la ley T-W resultó ser mejor, en términos del tamaño, que la prueba de razón de verosimilitud en los casos considerados, en especial cuando la dimensión es cercana al tamaño de la muestra. Sin embargo, la potencia de la prueba de razón de verosimilitud es mejor que la potencia de la prueba basada en la ley T-W. Se observa que la aproximación a la ley T-W dada por el Teorema 3 es mucho mejor que la aproximación a la distribución ji-cuadrada dada por el Teorema 5, al considerar n y p grandes con $p < n$.

Journal of Basic Sciences, Vol. 4 (12), pp. 1-13, sep-dic 2018

Referencias

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (John Wiley & Sons, Inc., New York, 1984).

[2] Z. D. Bai, J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices* (Springer, New York, 2010).

[3] A. Böttcher, S. M. Grudsky, *Toeplitz Matrices, Asymptotic Linear Algebra, and Functional Analysis* (Birkhäuser, 2012).

[4] I. M. Johnstone, On the Distribution of the Largest Eigenvalue in Principal Components Analysis *The Annals of Statistics*, (2) **29** (2001) 295–327.

[5] I. M. Johnstone, High Dimensional Statistical Inference and Random Matrices, *Proceedings of the International Congress of Mathematicians*, **1** (2007) 307–333.

[6] M. L. Mehta, *Random Matrices* (Academic press, 2004).

[7] R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (John Wiley & Sons, 2005).

[8] J. Wishart, The generalized product moment distribution in samples from a normal multivariate population, *Biometrika*, **20A** (1928) 32–43.

[9] Z. Ma, Accuracy of the Tracy-Widom Limits for the Extreme Eigenvalues in White Wishart Matrices, *Bernoulli*, (1) **18** (2012) 322–359.