Miscelánea Matemática **73** (2021) 5-20 SMM

DOI: https://doi.org/10.47234/mm.7301

Bondad de ajuste: una visita

Raúl Rueda
Instituto de Investigaciones en
Matemáticas Aplicadas y en Sistemas
Universidad Nacional Autónoma de México
pinky@sigma.iimas.unam.mx

Para Ana María, María, Eugenia y Federico

1. Introducción

Desde su tesis de doctorado titulada *On goodness-of-fit tests based on Rao-Blackwell distribution function estimators*, Federico O'Reilly trabajó en varios aspectos del problema de Bondad de Ajuste, sobre todo en el uso de estimadores de la función de distribución, para el caso compuesto, en lugar de usar los estimadores máximo verosímiles (O'Reilly y Quesenberry, 1972,1973). En la decada de 1980, trabajó con M. Stephens de la Universidad Simon Fraser en varios temas relacionados como caracterizaciones (O'Reilly y Stephens, 1982), el uso de espacios normalizados (Lockhart *et al.* 1986a, 1986b) y censura (O'Reilly y Stephens, 1988), entre otros.

El primer trabajo sobre el tema que se tenga reportado, de acuerdo con David y Edwards (2001) es el de Pearson (1900). Kolmogorov (1933) introdujo el uso de la distribución empírica y el uso del proceso empírico para desarrollar procedimientos de prueba.

La siguiente sección es una descripción del problema de bondad de ajuste, desde sus objetivos y diferentes soluciones propuestas, todas ellas basadas en la distribución empírica. Recordemos que cuando Federico comenzó a trabajar en estos problemas, el uso de algoritmos de simulación no estaba muy extendido, las soluciones en esos años consistían en encontrar la distribución asintótica del proceso empírico relacionada e invertir numéricamente su función característica para encontrar algunos cuantiles que permitieran llevar a cabo la prueba. Esta es la teoría que revisaremos en este trabajo.

Palabras clave: Bondad de ajuste, Proceso empírico, Anderson - Darling.

2. El problema de bondad de ajuste

Sea $\{X_1, \ldots, X_m\}$ una muestra aleatoria de una distribución $F(X;\theta)$. El problema de Bondad de Ajuste, consiste en probar la hipótesis

$$H_0 : F(X;\theta) = F_0(X;\theta), \tag{1}$$

en donde F_0 es una distribución cuya forma es conocida, pero posiblemente el parámetro θ no lo sea. Cuando θ es conocido, se tiene el caso simple; si es desconocido, se tiene el caso compuesto.

De todos los procedimientos que existen en la literatura, los que se basan en una medida de la discrepancia entre F_0 y F_m, la distribución empírica de la muestra, definida como

$$F_m(x) = \frac{1}{m}\sum_{i=1}^{m} I_{(-\infty,x]}(X_i) \quad x \in \Re,$$

donde I_A denota a la función indicadora del conjunto A, son los más estudiados en la literatura (véase D'Agostino & Stephens, 1986). Se han propuesto una diversidad de funciones de discrepancia, pero sobresalen dos entre ellas: las del tipo supremo y las cuadráticas. Asociadas a la primera clase se encuentran las estadísticas de Kolmogorov − Smirnov (Kolmogorov, 1933; Smirnov, 1939). La familia de Cramér − Von Mises (Cramér, 1928; Von Mises, 1931) se encuentra entre las del segundo tipo. Ambas medidas de discrepancia están fuertemente relacionadas con funciones de distancia o métricas definidas en espacios de funciones, la métrica del supremo y la métrica cuadrática. Estas discrepancias son

$$D_m = \sup_{x\in(-\infty,\infty)} |F_m(x) - F_0(x)|, \tag{2}$$

$$\mathcal{W}_m^2 = n\int_{-\infty}^{\infty} (F_m(x) - F_0(x))^2 \psi(x)dx, \tag{3}$$

donde ψ es una función de peso no negativa.

Existen algunos resultados sobre la distribución de estas estadísticas, basadas en el proceso empírico $\sqrt{m}(F_m(x) - F_o(x))$, para m finita. Pero los esfuerzos fueron encaminados a buscar la distribución asintótica cuando la hipótesis se supone verdadera y F_0 continua.

En el caso simple, el procedimiento de prueba se simplifica enormemente, al considerar la transformación $u_i = F_0(X_i;\theta)$ para toda $i \in J_m$, pues es bien sabido que estas nuevas variables aleatorias, $u_1, \ldots, u_m$ se distribuyen, si la hipótesis es verdadera, como uniformes en el intervalo $(0, 1)$. Así, en el caso simple, la hipótesis (1) es equivalente a probar

$$u_1, \ldots, u_m \sim U(u|0, 1)$$

con $u_i = F_0(X_i;\theta)$ para toda $i \in J_m$, independiente de quién sea F_0.

Como

$$G_m(u) = \frac{1}{m}\sum_{i=1}^{m} I_{[0,u]}(U_i),$$

donde $I_{[0,u]}$ denota a la función indicadora del intervalo $[0,u]$ y $\{U_1,\ldots,$ $U_m\}$ a la muestra ordenada, puede demostrarse fácilmente que

$$\mathrm{E}(G_m(u)) = u \text{ y } \mathrm{Cov}(G_m(u), G_m(u')) = \frac{u(1-u')}{m}, \quad u < u',$$

de donde puede deducirse que el proceso

$$y_m(u) = \sqrt{m}(G_m(u) - u),\ u \in [0,1], \tag{4}$$

tiene media cero y función de covarianza dada por

$$\mathrm{Cov}(y_m(u), y_m(u')) = \mathrm{mín}(u,u') - uu'. \tag{5}$$

Usando el teorema central de límite, puede demostrarse que el vector $(y_m(u_1), y_m(u_2),\ldots,\ y_m(u_k))$ se distribuye asintóticamente como una normal multivariada con media cero y matriz de varianzas calculada a partir de (3), para todo valor de k, lo que significa que las distribuciones finito dimensionales del proceso $\{y_m\}$ convergen a las distribuciones finito dimensionales de un proceso Gaussiano con media cero y función de covarianza dada por (3). La demostración de que un proceso con las características anteriores existe en $C[0,1]$ puede encontrarse en Billingsley (1968).

Denótese por $\omega(u)$ al proceso browniano, es decir, al proceso definido en el intervalo $[0,1]$ que cumple con

i. $P[\omega(0) = 0] = 1$

ii. $p(\omega(u)) = N(\omega(u)|0,u)$ para cada $u \in [0,1]$

iii. $\omega(u_1) - \omega(u_0), \omega(u_2) - \omega(u_1), \ldots, \omega(u_k) - \omega(u_{k-1})$
son independientes para cada $0 = u_0 < u_1 < \ldots < u_k \leq 1$. (6)

Considérese ahora al proceso

$$y(u) = \omega(u) - u\ \omega(1),$$

llamado *browniano atado*, para el que puede demostrarse a partir de (12), que tiene media cero y función de covarianza dada por (3).

Sin embargo la convergencia de las distribuciones finito dimensionales no es suficiente para asegurar que, por ejemplo, $\sup|y_m(u)| \to \sup|y(u)|$, para esto es necesario el concepto de convergencia débil o convergencia en distribución (*e.g.* Billingsley, 1968).

Sea Ω un espacio métrico con métrica d que contiene a las funciones ϕ definidas en el $[0,1]$, denótese con $\mathcal{A}$ a la σ-álgebra generada por los abiertos en (Ω, d) y considérese a $(\Omega, \mathcal{A})$. Sea $\{\eta_m : m \in \mathbf{N}\}$ una sucesión de funciones medibles definidas sobre Ω y η otra función medible definida también sobre Ω. Se dice que $\{\eta_m\}$ converge débilmente a $\{\eta\}$ en $(\Omega, \mathcal{A})$ si

$$\lim_{n\to\infty} \mathrm{E}_{P_m}[f(\eta_m)] = \mathrm{E}_P[f(\eta)]$$

para toda f continua y acotada en el espacio métrico (Ω, d), denotándolo por $\{\eta_m\} \Rightarrow \{\eta\}$, donde $\{P_m : m \in \mathbf{N}\}$ y P son las correspondientes distribuciones de probabilidad definidas en $(\Omega, \mathcal{A})$. Esto es equivalente a decir que $\{P_m : m \in \mathbf{N}\}$ converge débilmente a P y se denota con $P_m \Rightarrow P$.

Un resultado muy importante en convergencia débil, es el que asegura que si g es una función definida en Ω que es continua con respecto a la métrica d y $\{\eta_m\} \Rightarrow \{\eta\}$ entonces $\{g(\eta_m)\} \Rightarrow \{g(\eta)\}$.

Dado el resultado anterior, bastaría demostrar que las funciones definidas en Ω como

$$g_1(\eta) = \sup_u |\eta(u)|,$$

$$g_2(\eta) = \int_0^1 \eta^2(u)\psi(u)du, \tag{7}$$

son continuas respecto a una métrica d.

Como el proceso empírico $\{y_m\}$ no tiene trayectorias continuas en el intervalo $[0,1]$, (mientras que las del proceso browniano atado sí son continuas), el espacio natural de definición del proceso es D, el espacio de todas las funciones en el $[0,1]$ continuas por la derecha y cuyo límite izquierdo existe para todo u en $[0,1]$. Este espacio puede ser dotado de diferentes métricas, por ejemplo, la métrica uniforme definida para toda η_1, η_2 en D como

$$d_u(\eta_1, \eta_2) = \sup_{0\le u\le 1} |\eta_1(u) - \eta_2(u)|.$$

El espacio métrico (D, d_u) es completo pero no separable (*e.g.* Billingsley, 1968), lo que presenta ciertas dificultades técnicas para demostrar la convergencia débil de ciertos procesos. Sin embargo, puede demostrarse (Skorokhod, 1956) que existe una métrica d_0 definida sobre $D \times D$ tal que el espacio métrico (D, d_0) es separable y completo. La introducción de esta métrica en el estudio de la convergencia débil no es una condición necesaria, sin embargo facilita enormemente la teoría (Shorack y Wellner, 1986).

Si se considera al espacio $(D, \mathcal{D})$, donde $\mathcal{D}$ es la σ-álgebra generada por los abiertos de (D, d_u) y se extiende el dominio de definición de los procesos browniano y browniano atado a $(D, \mathcal{D})$, puede demostrarse que el proceso $\{y_m\}$ definido en (2) es un elemento de D. A partir de la convergencia de las distribuciones finito dimensionales de $\{y_m\}$ a las de $\{y\}$, Doob (1949) mediante un argumento heurístico, demostró que $\{y_m\} \Rightarrow \{y\}$. Donsker (1952) dió una demostración formal de esta convergencia en $(D, \mathcal{D}, d_u)$ (véase Pollard, 1984).

Por otro lado, puede demostrarse (Durbin, 1973a) que g_1 y g_2 definidas en (5) son continuas en (D, d_u), por lo que

$$\sup_{0 \le u \le 1} |y_m(u)| \Rightarrow \sup_{0 \le u \le 1} |y(u)| \quad \text{y}$$

$$\int_0^1 y_m^2(u)\psi(u)du \Rightarrow \int_0^1 y^2(u)\psi(u)du.$$

Así, bastaría determinar las distribuciones de los términos de la derecha de la expresión anterior, para tener caracterizadas a las distribuciones asintóticas de interés.

Antes de continuar, vale la pena hacer un paréntesis en el tema de convergencia débil.

Una demostración interesante de que $\{y_m\} \Rightarrow \{y\}$ en $(D, \mathcal{D}, d_u)$, es presentada por Pollard (1984): dada la convergencia de las distribuciones finito dimensionales, demuestra que las trayectorias del proceso $\{y_m\}$ pueden aproximarse por procesos *locales* que implican la convergencia débil del proceso original. Esta condición es equivalente al concepto de *tensión* que se definirá un poco más adelante. Sin embargo, Pollard (1984) menciona que este tipo de condiciones para construir aproximaciones simples al proceso no son muy utilizadas debido, posiblemente, a que debe demostrarse en forma independiente la existencia del proceso límite. El tipo de demostración más utilizado requiere desarrollar otro tipo de herramientas (*e.g.* Billingsley, 1968; Dudley, 1978).

Sea $(S, \mathcal{L})$ con S un espacio métrico y $\mathcal{L}$ la σ-álgebra generada por los abiertos de S. Sea $\mathcal{P}$ una familia de medidas de probabilidad definidas sobre $(S, \mathcal{L})$. Se dice que la familia $\mathcal{P}$ es *tensa* si para toda $\varepsilon \in \mathbf{R}^+$ existe un conjunto compacto $K \in \mathcal{L}$ tal que $P(K) > 1 - \varepsilon$ para toda $P \in \mathcal{P}$.

La importancia de este concepto queda establecida en el siguiente resultado (Billingsley, 1968): sea $\{\eta_m : m \in \mathbf{N}\}$ una sucesión contenida en $(D, \mathcal{D})$ y denótese por $\{P_m : m \in \mathbf{N}\}$ la correspondiente sucesión de distribuciones de probabilidades. Si las distribuciones finito dimensionales del proceso convergen a las del proceso $\{\eta\}$ y la familia $\{P_m : m \in \mathbf{N}\}$ es tensa, entonces $P_m \Rightarrow P$ donde P es la distribución de probabilidad asociada a $\{\eta\}$.

La condición de tensión puede relajarse, si el espacio de definición de los procesos es separable. Sea $\mathcal{P}$ una familia de medidas de probabilidad definidas en el espacio métrico $(S, \mathcal{L})$. Se dice que la familia $\mathcal{P}$ es *relativamente compacta* si para toda sucesión $\{P_m : m \in \mathbf{N}\}$ contenida en $\mathcal{P}$ existe una subsucesión $\{P_{n'}\}$ que converge débilmente a una medida P definida en $(S, \mathcal{L})$ (no necesariamente contenida en $\mathcal{P}$).

Puede mostrarse que si $\mathcal{P}$ es una familia relativamente compacta definida en $(S, \mathcal{L})$ y S es separable y completo entonces $\mathcal{P}$ es tensa (Billingsley,1968).

Este último resultado, es la razón de que se use el espacio métrico separable (D, d_0), pues en este, basta demostrar la convergencia de las distribuciones finito dimensionales y que la familia de medidas de probabilidad asociadas sea relativamente compacta, para tener la convergencia débil.

Regresando al problema simple de Bondad de Ajuste, se tiene, en resumen, que para probar la hipótesis $H_0 : F(x) = F_0(x|\theta)$ <u>con θ conocida</u>, primero hay que transformar la muestra ordenada $\{X_1, X_2, \ldots, X_m\}$, con $u_i = F_0(X_i|\theta)$, encontrar la distribución del proceso $\{y_m\}$ y finalmente la de la transformación correspondiente.

La distribución asintótica de D_m en la ecuación (2), fue encontrada por Kolmogorov (1933) y una derivación de ella puede verse en Durbin (1973a). El resultado de Kolmogorov es

$$P[\sqrt{m}D_m \geq d] \to 2\sum_{k=1}^{\infty} e^{-2k^2d^2}.$$

La obtención de la distribución asintótica de $\mathcal{W}_m^2$, ecuación (3), para cada función ψ se esboza a continuación:

Si $\xi_m(u) = \sqrt{\psi(u)y_m(u)}$, entonces

$$\mathcal{W}_m^2 = \int_0^1 \xi_m^2(u)du$$

y además,

$$\begin{aligned} \mathrm{E}(\xi_m(u)) = 0 \;\; \text{y} \\ \kappa(u, u') = \operatorname{cov}(\xi(u), \xi(u')) = \sqrt{\psi(u)\psi(u')}(\text{mín}(u, u') - uu'). \end{aligned} \tag{8}$$

La idea es representar a $\mathcal{W}_m^2$ como una suma ponderada de variables aleatorias no correlacionadas con media cero y varianza uno. El problema es análogo a cuando se tiene una suma finita de variables al cuadrado con media cero y cierta matriz de varianzas: diagonalizar la matriz usando los eigenvalores y eigenvectores correspondientes.

En el caso continuo, esto supone resolver la ecuación

$$\int_0^1 \kappa(u,u')\gamma(u')du' = \lambda\gamma(u), \tag{9}$$

donde λ_j y $\gamma_j(u)$, $j \in \mathbf{N}$ son los eigenvalores y las eigenfunciones del kernel $\kappa(u,u')$ que cumplen con

$$\int_0^1 \gamma_j^2(u)du = 1 \text{ y } \int_0^1 \gamma_j(u)\gamma_k(u)du = 0 \text{ para } j \neq k.$$

De esta forma, si

$$\zeta_{nj} = \lambda_j^{-1/2} \int_0^1 \gamma_j(u)\xi_m(u)du,$$

entonces

$$\mathcal{W}_m^2 = \sum_{j=1}^{\infty} \lambda_j \zeta_{nj}^2,$$

donde $\zeta_{n1}, \zeta_{n2}, \ldots$ son variables aleatorias no correlacionadas con media cero y varianza unitaria (Durbin, 1973a).

Usando este resultado y como la distribución asintótica de $\mathcal{W}_m^2$ es la misma que la de

$$\mathcal{W}^2 = \int_0^1 \psi(u)y^2(u)du,$$

donde $\{y(u) : u \in [0,1]\}$ es el proceso browniano atado, se tiene que

$$\mathcal{W}^2 = \sum_{j=1}^{\infty} \lambda_j z_j^2, \tag{10}$$

con λ_j los eigenvalores del kernel $\kappa(u,u') = \sqrt{\psi(u)\psi(u')}(\text{mín}(u,u') - uu')$ y en donde z_j son variables normales con media cero y varianza uno, que resultan de resolver (9). De manera que $\mathcal{W}^2$ es una serie ponderada de variables aleatorias independientes Ji-cuadrada con un grado de libertad, cuya función característica está dada por

$$\phi(t) = \prod_{j=1}^{\infty} (1 - 2it\lambda_j)^{-1/2}. \tag{11}$$

Así, habrá que invertir $\phi(t)$ para encontrar los cuantiles de $\mathcal{W}^2$.

En general, resolver la ecuación (9) es complicado. En estos casos lo que se hace es calcular algunos cuantiles de la distribución asintótica haciendo una aproximación a (10) e invirtiendo numéricamente la correspondiente aproximación a (11). Específicamente, supóngase que

$$\mathcal{W}^2 \approx \frac{1}{m}\sum_{j=1}^{m} \psi(u_j) y^2(u_j) \quad u_j \in [0,1],\ j \in J_m \text{ y } m \in \mathbf{N},$$

donde, en general, $J_n = \{1, 2, \ldots, n\}$.

Si V denota a la matriz cuya entrada ij-ésima está dada por

$$v_{ij} = \frac{1}{m} \ \ \kappa(u_i, u_j)\ i, j \in J_m, \tag{12}$$

con κ definido en (8), una aproximación a (10) está dada por

$$\mathcal{W}^2 \approx \sum_{j=1}^{m} \lambda_j \nu_j,$$

donde $\lambda_j,\ \ j \in J_m$ son los eigenvalores de la matriz (12) y ν_j son variables aleatorias Ji-cuadrada independientes con un grado de libertad definidas para cada j como

$$\nu_j = \lambda_j^{-1/2} \sum_{k=1}^{m} \zeta_{jk}\ y(u_j),$$

con $(\zeta_{j1}, \zeta_{j2}, \ldots, \zeta_{jm})'$ el eigenvector asociado a λ_j.

Con esta aproximación, la expresión (11) puede escribirse como

$$\phi(t) \approx \prod_{j=1}^{m} (1 - 2it\lambda_j)^{-1/2}.$$

Esta función característica puede invertirse numéricamente usando el siguiente resultado debido a Imhof (1961),

$$P[\mathcal{W}^2 > w] = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \frac{\operatorname{sen} h(v)}{v g(v)} dv \tag{13}$$

donde

$$h(v) = \frac{1}{2}\left\{ \sum_{i=1}^{m} \tan^{-1} \lambda_i v - wv \right\} \ \text{ y}$$

$$g(v) = \prod_{i=1}^{m} \left(1 + \lambda_i^2 v^2\right)^{1/4}.$$

En resumen, para encontrar algún cuantil de la distribución asintótica de $\mathcal{W}_m^2$ o equivalentemente, la distribución de $\mathcal{W}^2$, se calcula la matriz V definida en (12) para algún valor de m, se encuentran los eigenvalores y eigenvectores ortonormalizados asociados a V y se resuelve numéricamente la integral (13).

De esta manera, el problema simple de Bondad de Ajuste queda resuelto.

Sin embargo, el valor de θ es desconocido en la mayoría de las aplicaciones de Bondad de Ajuste. En este caso la transformación (2) no puede aplicarse, pues la función de distribución no está totalmente especificada. Existen varias soluciones a este problema, en el fondo todas llevan a estimar a $F_0(x|\theta)$, ya sea estimando θ y usar este estimador para transformar la muestra, o bien, estimando toda la función de distribución.

Supóngase entonces que se tiene una muestra aleatoria $\{X_1, X_2, \ldots, X_m\}$ y se quiere probar la hipótesis

$$H_0 : F(x) = F_0(x|\theta) \text{ con } \theta \text{ desconocido.}$$

Sea $\hat{\theta}_m = \hat{\theta}(X_1, X_2, \ldots, X_m)$ un estimador de θ. Siguiendo la lógica del caso θ conocido, considérese la transformación

$$\hat{U}_i = \hat{U}(X_i) = F_0(X_i|\hat{\theta}) = \hat{F}_m(X_i) \quad \text{para toda } i \in J_m.$$

Si G_m denota a la distribución empírica de la muestra $\{\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_m\}$, el proceso empírico estimado asociado es

$$\hat{y}_m(\hat{u}) = \sqrt{n}(G_m(\hat{u}) - \hat{u}), \tag{14}$$

para el que habrá que encontrar su distribución asintótica.

Bajo ciertas condiciones, similares a las pedidas por Cramér (1946) en el desarrollo del método de máxima verosimilitud y puesto que $\hat{u} \to u$ con $u = F_0(x|\theta)$, Durbin(1973b) demostró que $\{\hat{y}_m(\hat{u})\} \Rightarrow \{\hat{y}(u)\}$ donde $\{\hat{y}(u)\}$ es un proceso Gaussiano con media cero y función de covarianza dada por

$$\mathrm{E}(\hat{y}(u), \hat{y}(u')) = \text{mín}(u, u') - uu' - h(u)'\mathcal{I}^{-1}h(u') \text{ con } u, u' \in [0, 1] \tag{15}$$

y en donde $h(u)$ es el vector cuya j-ésima entrada es $\frac{\partial}{\partial \theta_j} F(x|\theta)$ evaluado en $u = F(x|\theta)$ e $\mathcal{I}$ denota a la matriz de Fisher por unidad muestral.

En general, el estimador $\hat{\theta}_m$ que se utiliza para generar (14) es el de máxima verosimilitud, pero puede ser cualquiera que cumpla con

$$\sqrt{n}(\hat{\theta}_m - \theta) = \frac{1}{\sqrt{n}} \, \mathcal{I}^{-1} \sum_{j=1}^{n} \frac{\partial \log p(x_j|\theta)}{\partial \theta} + \varepsilon_m,$$

donde $p(x|\theta)$ denota a la función de densidad y $\varepsilon_m \to 0$ en probabilidad (Serfling, 1980).

Considerando las transformaciones g_1 y g_2 definidas en (5) se tiene que

$$\hat{\mathcal{W}}_m^2 = \int_0^1 \hat{y}_m^2(\hat{u})\psi(\hat{u})d\hat{u} \Rightarrow \int_0^1 \hat{y}^2(u)\psi(u)du$$
$$\hat{D}_m = \sup_{0\leq\hat{u}\leq 1} \sqrt{n}|\hat{y}_m(\hat{u})| \Rightarrow \sup_{0\leq u\leq 1} |\hat{y}(u)|.$$

Lilliefors (1967, 1969) encontró mediante simulación, la distribución asintótica de $\hat{D}_m$ para el caso Normal con ambos parámetros desconocidos y para el caso Exponencial; Durbin (1975) da la distribución asintótica para el caso Exponencial. Stephens (1986, y referencias incluidas), da los cuantiles asintóticos para algunas distribuciones, tanto de $\hat{D}_m$ como de $\hat{\mathcal{W}}_m^2$, para algunas formas específicas de la función ψ. Sin embargo una teoría asintótica general para $\hat{D}_m$ no está dada (D'Agostino y Stephens, 1986).

Dos miembros de la familia Cramér-Von-Mises son de especial interés: la estadística Von-Mises, que se obtiene haciendo $\psi(u) = 1,\ u \in [0,1]$ y la A^2 de Anderson-Darling (Anderson y Darling, 1952, 1954), que se obtiene haciendo $\psi(u) = [u(1-u)]^{-1},\ u \in [0,1]$. Ambas han sido ampliamente estudiadas y existe una teoría asintótica tanto para el caso de parámetros conocidos como cuando los parámetros se desconocen.

La prueba basada en la estadística A^2 ha demostrado buena potencia, cuando los parámetros se estiman por máxima verosimilitud (Stephens, 1974).

Obsérvese que la función $\psi(u) = [u(1-u)]^{-1}$ no es continua en $\{0,1\}$, por lo que el método de prueba sobre la convergencia débil esbozado anteriormente, no puede ser aplicado. Este es un caso en el que el concepto de *tensión* es útil para encontrar la distribución asintótica del proceso empírico (véase Durbin, 1973b).

Nótese que la función de covarianza (15) es la del caso simple menos una función que depende de la forma de la función de distribución y, posiblemente, del parámetro. David y Johnson (1948) demostraron que la distribución de $\hat{F}_m(x)$ no depende del parámetro si θ es un parámetro de escala y/o localización, por lo tanto la distribución finita del proceso no depende de θ, así que la distribución asintótica tampoco dependerá de θ en estos casos. Por lo tanto, la generacióm de tablas con los cuantiles de la distribución asintótica resulta más sencilla, pues solo dependerá de la forma de la función (véase Stephens, 1986). Esto es válido, tanto para la distribución asintótica de $\hat{D}_m$ como para la de $\hat{\mathcal{W}}_m^2$, incluyendo a la estadística $\hat{A}_m^2$.

Otro procedimiento de estimación es el que utiliza al estimador de Rao-Blackwell de la función de distribución, y fue introducido por Srinivasan (1970). Recuérdese que el estimador de Rao-Blackwell de la función de distribución es

$$\tilde{F}_m(x) = \ \mathrm{E}\left[I_{[X_i \le x]} | T_m\right],$$

donde T_m es una estadística suficiente minimal y X_i es cualquier elemento de la muestra. Srinivasan (1970) considera al proceso

$$\sqrt{n}\{F_m(x) - \tilde{F}_m(x)\}, \ x \in \mathbf{R} \tag{16}$$

y encuentra, usando métodos de Monte Carlo, la distribución asintótica para el caso Normal y Exponencial, demostrando que estas distribuciones no dependen de θ.

El proceso (16) puede escribirse como

$$\tilde{y}_m(\tilde{u}) = \sqrt{n}(G_m(\tilde{u}) - \tilde{u}), \tag{17}$$

donde $\tilde{u} = \tilde{u}(x) = \tilde{F}_m(x)$ y G_m denota ahora a la función de distribución empírica de la muestra transformada $\{\tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_m\}$.

El resultado que relaciona a las distribuciones asintóticas de los procesos empíricos estimados $\{\hat{y}_m(\hat{u})\}$ y $\{\tilde{y}_m(\tilde{u})\}$, definidos en (14) y (17) respectivamente, fue establecido por Moore (1971) y es el siguiente:

Como

$$\sqrt{n}\{F_m(x) - \hat{F}_m(x)\} = \sqrt{n}\{F_m(x) - \tilde{F}_m(x)\} + \sqrt{n}\{\tilde{F}_m(x) - \hat{F}_m(x)\}$$

si

$$\sup |\tilde{F}_m(x) - \hat{F}_m(x)| \xrightarrow{P} 0 \tag{18}$$

entonces las distribuciones asintóticas de los procesos estimados $\{\hat{y}_m(\hat{u})\}$ y $\{\tilde{y}_m(\tilde{u})\}$ serán iguales.

Moore (1971) demostró que esto ocurría en los casos Normal, Exponencial y Uniforme con parámetro θ. Este último caso es un ejemplo de que la propiedad (18) no es característica de la familia exponencial. Incluso Moore (1971) conjetura que (18) debe ser cierto en general, pero no exhibe una demostración. Lockhart y O'Reilly (2005) demuestran esta conjetura para un subconjunto de la familia exponencial.

Otro procedimiento de estimación de la función de distribución ha sido propuesto por O'Reilly y Villegas (1987). Estos autores proponen el uso de la distribución predictiva, que aparece en el contexto bayesiano, como un estimador de la función de distribución y demuestran que bajo ciertas condiciones, se cumple que

$$\tilde{F}(X_m|T_m) = F^{\pi}(X_m|T_{m-1}),$$

donde π es la medida de Haar invariante por la derecha y F^{π} la distribución predictiva obtenida a partir de π.

Así, si se considera al proceso empírico estimado

$$y_m^\pi(u) = \sqrt{n}\{G_m(u) - u\},$$

con G_m la función de distribución empírica de la muestra $\{U_1^\pi, \ldots, U_m^\pi\}$, donde

$$U_i^\pi = U^\pi(X_i) = F^\pi(X_i|T_{(m-1)}) \tag{19}$$

y $T_{(m-1)}$ la estadística suficiente minimal construida con las $m-1$ observaciones restantes, se tendrá que la distribución asintótica es la misma que la del proceso $\{\tilde{y}_m\}$.

En el caso de que θ sea un parámetro de localización y/o escala, la relación (19) se cumple y entonces las distribuciones asintóticas de los tres procesos empíricos estimados coinciden.

Es importante mencionar que la transformación $\hat{U}$ no genera variables aleatorias Uniformes ni independientes (David y Johnson, 1948) y que la transformación $\tilde{U}$ sí genera variables uniformes, aunque no independientes, (O'Reilly y Quesenberry, 1973, presentan un procedimiento basado en el estimador de Rao-Blackwell que transforma variables aleatorias a subconjuntos de variables aleatorias uniformes e independientes), por lo que la transformación U^π también genera variables aleatorias uniformes no independientes, si se cumple (19).

Hoy en día, la distribución de la estadística de prueba puede simularse para cualquier valor de m, evitando el uso de tablas.

3. Mi trabajo con Federico

Conocí a Federico cuando me dio dos cursos en la maestría en 1976 y lo fui tratando cuando me lo encontraba de vez en vez, rumbo a la bilblioteca del IIMAS. En 1986, trabajando en la UAM-Iztapalapa, lo invité a que nos diera una plática en el seminario del Departamento de Matemáticas en el mes de junio y accedió. Ese fue el inicio de nuestra amistad.

En septiembre de 1986 recibí una llamada en casa, era Federico. Lo habían invitado a dar una plática en un *seminario bayesiano* y me propuso que trabajáramos en encontrar intervalos, yo de probabilidad y él de confianza, para el cociente de parámetros en la distribución Gaussiana inversa. Esta distribución era muy interesante, pues sus parámetros no son ni de localización ni de escala. Federico había desarrollado una prueba de bondad de ajuste compuesta y la distribución asintótica dependía de los parámetros a través de este cociente. Como en aquella época se generaban tablas, él había construido una para diferentes valores del cociente. La manera de utilizarlas, era calcular los estimadores máximo verosímiles y entrar al renglón correspondiente. El objetivo de los intervalos era para tener una idea de la variabilidad al entrar a la tabla. Como el seminario trataba de temas bayesianos, se le ocurrió que yo obtuviera intervalos de probabilidad. A partir de esa fecha, acudí al IIMAS dos tardes a la semana a trabajar con él.

La distribución Gaussiana inversa es parte de la familia exponencial, pero no es regular. Los intervalos frecuentistas se volvieron fiduciales pues presentaban una masa positiva en el cero. Contestar por qué ocurría esto, nos llevó a tratar de entender más a la estadística fiducial.

En septiembre de 1987 realicé una estancia sabática en el IIMAS trabajando con Federico. Trabajamos en varias líneas que a él le interesaban: estadística fiducial, bondad de ajuste para distribuciones discretas y el problema de la Gaussiana inversa. En este último, el reto era probar que la conjetura de Moore (1973) se cumplía para la Gaussiana inversa, esto es, que el proceso empírico que resultaba de usar los estimadores máximo verosímiles y el que resultaba de usar el estimador de Rao - Blackwell tenían el mismo límite, cosa que pudimos demostrar (O'Reilly y Rueda, 1992).

En el caso de las distribuciones discretas, en lugar de usar la distribución empírica, usamos la función empírica generadora de probabilidades, esta era una idea que Federico había platicado con Víctor Pérez Abreu del CIMAT: Federico decía que el proceso límite debería tener una función de covarianza muy similar al caso continuo, pero con la matriz de derivadas de la función generadora. Pero el proceso límite resultó muy diferente (Rueda *et al.*, 1991). Después descubrimos que habíamos incurrido en un pequeño error, que al corregir nos daba ¡lo que Federico intuía! (Rueda y O'Reilly, 1999).

Esta siempre fue una característica de Federico. Su intuición, su experiencia y todo lo que sabía sobre un problema en particular, le servía para intuir lo que se debería obtener. Pocas veces esto no ocurrió.

A principios de este milenio, una colega del Instituto llegó con Federico con la siguiente duda: En una imagen digitalizada que contenía decenas de miles de pixels en diferentes tonos de grises, quería probar si a estos pudiera ajustárseles una distribución normal. Al hacer un histograma, parecía obvio que sí. Sin embargo al usar la prueba de Anderson-Darling, ¡rechazamos la hipótesis! El problema es que las mediciones estaban hechas con una resolución tal que agrupaba los datos en diferentes clases.

El efecto del redondeo alteraba el proceso empírico bajo estudio. Este fue un trabajo que a Federico le gustó mucho, le dedicó mucho tiempo y tuvimos que adaptar muchas cosas. Desafortunadamente, no tuvo el impacto que hubiéramos querido. (O'Reilly *et al.*, 2003).

En otro trabajo sobre la distribución Levy (O'Reilly y Rueda, 1998), surgió la idea de simular de una distribución condicional a las estadísticas suficientes, que después Federico mejoró en lo que le llamó *look alike samples*. Federico redescubrió algunos algoritmos de simulación y los utilizó de manera intensiva a partir de 2004.

Trabajar con Federico siempre fue aprendizaje y diversión. Tenía un excelente sentido del humor y siempre tenía a la mano un chiste para cualquier situación. Tenía mucha paciencia para que sus alumnos entendiéramos lo que él hacía y era excelente explicando todo.

Bibliografía

[1] T. Anderson y D. Darling, «Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes», *Ann. Math. Statist.*, núm. 23, 1952, 193–212.

[2] ———, «A test of goodness of fit», *J. Amer. Statist. Assoc.*, 1954, 765–769.

[3] P. Billingsley, *Convergence of Probability Measures*, New York: Wiley, 1968.

[4] H. Cramér, *Mathematical methods in statistics*, Princeton: University Press, 1946.

[5] H. Cramér, «On the composition of elementary errors», *Scandinavian Actuarial Journal*, vol. 1928, núm. 1, 1928, 141–180.

[6] R. D'Agostino y M. Stephens, *Goodnes-of-fit techniques*, New York: Marcel Dekker, Inc., 1986.

[7] F. N. David y N. L. Johnson, «The Probability Integral Transformation When Parameters are Estimated from the Sample», *Biometrika*, vol. 35, núm. 1/2, 1948, 182–190.

[8] H. David y A. Edwards, *Annotated readings in the history of statistics*, New York: Springer, 2001.

[9] M. D. Donsker, «Justification and Extension of Doob's Heuristic Approach to the Kolmogorov- Smirnov Theorems», *The Annals of Mathematical Statistics*, vol. 23, núm. 2, 1952, 277–281.

[10] J. L. Doob, «Heuristic Approach to the Kolmogorov-Smirnov Theorems», *The Annals of Mathematical Statistics*, vol. 20, núm. 3, 1949, 393–403.

[11] R. M. Dudley, «Central Limit Theorems for Empirical Measures», *The Annals of Probability*, vol. 6, núm. 6, 1978, 899–929.

[12] J. Durbin, *Distribution theory for tests based on the sample distribution function*, Philadelphia: SIAM, 1973, Reg. Conf. Ser. Appl. Math., 9.

[13] J. Durbin, «Weak Convergence of the Sample Distribution Function when Parameters are Estimated», *The Annals of Statistics*, vol. 1, núm. 2, 1973, 279–290.

[14] ———, «Kolmogorov-Smirnov Tests when Parameters are Estimated with Applications to Tests of Exponentiality and Tests on Spacings», *Biometrika*, vol. 62, núm. 1, 1975, 5–22.

[15] J. P. Imhof, «Computing the Distribution of Quadratic Forms in Normal Variables», *Biometrika*, vol. 48, núm. 3/4, 1961, 419–426.

[16] A. Kolmogorov, «Sulla determiniazone empirica di una legge distribuzione», *Giorn. Ist. Ital. Attuari.*, núm. 4, 1933, 83 – 91.

[17] H. W. Lilliefors, «On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown», *Journal of the American Statistical Association*, vol. 62, núm. 318, 1967, 399–402.

[18] ———, «On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown», *Journal of the American Statistical Association*, vol. 64, núm. 325, 1969, 387–389.

[19] R. Lockhart y F. O'Reilly, «A note on Moore's conjecture», *Statistics & Probability Letters*, vol. 74, núm. 2, 2005, 212–220.

[20] R. Lockhart, F. O'Reilly y M. Stephens, «Tests for the extreme-value and weibull distributions based on normalized spacings», *Naval Research Logistics Quarterly*, vol. 33, 1986, 413 – 421.
[21] R. Lockhart, F. O'Reilly y M. Stephens, «Tests of Fit Based on Normalized Spacings», *Journal of the Royal Statistical Society. Series B*, vol. 48, 1986, 344 – 352.
[22] D. S. Moore, «A note on Srinivasan's goodness-of-fit test», *Biometrika*, vol. 60, núm. 1, 1973, 209–211.
[23] F. J. O'Reilly, R. Rueda y M. Garza-Jinich, «How Important Is the Effect of Rounding in Goodness-of-Fit», *Communications in Statistics - Simulation and Computation*, vol. 32, núm. 3, 2003, 953–976.
[24] F. O'Reilly y L. Gracia-Medrano, «On the Conditional Distribution of Goodness-of-Fit Tests», *Communications in Statistics-Theory and Methods*, vol. 35, 2006, 541–549.
[25] F. O'Reilly y C. Quesenberry, «Uniform Strong Consistency of Rao-Blackwell Distribution Function Estimators», *The Annals of Mathematical Statistics*, vol. 43, 1972, 1678 – 1679.
[26] ———, «The Conditional Probability Integral Transformation and Applications to Obtain Composite Chi-Square Goodness-of-Fit Tests», *The Annals of Statistics*, vol. 1, 1973, 74 – 83.
[27] F. O'Reilly y R. Rueda, «A note on the fit for the Levy distribution», *Communications in Statistics - Theory and Methods*, vol. 27, 1998, 1811–1821.
[28] F. O'Reilly y M. Stephens, «Transforming Censored Samples for Testing Fit», *Technometrics*, vol. 30, 1988, 79–86.
[29] F. O'Reilly y C. Villegas, «Conditional and predictive probability integral transforms», *Canadian Journal of Statistics*, vol. 15, 2008, 151 – 158.
[30] F. J. O'Reilly y R. Rueda, «Goodness of Fit for the Inverse Gaussian Distribution», *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, vol. 20, núm. 4, 1992, 387–397.
[31] F. J. O'Reilly y M. A. Stephens, «Characterizations and Goodness of Fit Tests», *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 44, núm. 3, 1982, 353–360.
[32] K. Pearson, « On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supossed to have arisen from random sampling», *Phil. Magazine*, 1900, 157 – 175.
[33] D. Pollard, *Convergence of Stochastic Processes*, New York: Springer-Verlag, 1984.
[34] R. Rueda y F. O'Reilly, «Tests of fit for discrete distributions based on the probability generating function», *Communication in Statistics- Simulation and Computation*, vol. 32, 1999, 953 – 976.
[35] R. Rueda, V. Abreu y F. O'Reilly, «Goodness of fit for the Poisson distribution based on the probability generating function», *Communications in Statistics - Theory and Methods*, vol. 20, 1991, 3093–3110.
[36] R. Serfling, *Approximation theorems of mathematical statistics*, New York: Wiley, 1980.
[37] G. Shorack y J. Wellner, *Empirical processes with applications to statistics*, New York: Wiley, 1986.
[38] A. Skorokhod, « Limit theorems for stochastic processes», *Theor. Prob. Appl.*, núm. 1, 1956, 261–290.
[39] N. Smirnov, «Sur les écarts de la courbe de distribution empirique», *Mat. Sbornik*, núm. 48, 1939, 3 – 26.
[40] R. Srinivasan, «An approach to testing the goodness of fit of incompletely specified distributions», *Biometrika*, núm. 57, 1970, 605 – 611.
[41] M. Stephens, «Tests based on EDF statistics», en *Goodnes-of-fit techniques*, New York: Marcel Dekker, Inc., 1986, 97–193.
[42] M. Stephens, «EDF statistics for goodness-of-fit and some comparisons», *J. Amer. Statist. Assoc.*, núm. 69, 1974, 730–737.

[43] R. Von MIses, *Wahrscheinlichkeitsrechnung*, Leipzig: Wein, 1931.