Miscelánea Matemática **73** (2021) 31-38 SMM

DOI: https://doi.org/10.47234/mm.7303

El uso de distribuciones condicionales dada una estadística suficiente

José M. González-Barrios
Departamento de Probabilidad y Estadística, IIMAS
Universidad Nacional Autónoma de México
gonzaba@sigma.iimas.unam.mx
y
Raúl Rueda
Departamento de Probabilidad y Estadística, IIMAS
Universidad Nacional Autónoma de México
pinky@sigma.iimas.unam.mx

1. Introducción y Resultados Básicos

En este artículo implementamos ideas de Federico O'Reilly acerca del uso de distribuciones condicionales, o densidades condicionales dada una estadística suficiente T_n, de una muestra aleatoria $X_1, X_2, \ldots, X_n$, definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, cuya función de distribución pertenece a una familia paramétrica $\{F(x;\theta) \mid \theta \in \Theta\}$, donde $\Theta \subset \mathbb{R}^k$ es el espacio parametral. La idea de usar estadísticas suficientes, que fueron definidas primero por Ronald Fisher, es que una estadística suficiente de un parámetro θ incluye toda la información posible acerca del parámetro, el cuál es la única cantidad desconocida en una familia paramétrica. Así, el uso de una distribución condicional o una densidad condicional, puede resolver claramente varios problemas en la técnica de Bondad-de-Ajuste en Estadística.

Recordar que una familia paramétrica es un conjunto de la forma $\{F(x;\theta) \mid \theta \in \Theta\}$ donde Θ es llamado el espacio parametral, ejemplo de estas familias paramétricas son los conjuntos $\{\text{Poisson}(x;\lambda) \mid x \in \{0,1,\ldots\}, \lambda \in (0,\infty)\}$, la familia Poisson, $\{\text{Bin}(x;(k,p)) \mid x \in \{0,1,2,\ldots,k\}, k \in \mathbb{N}, p \in [0,1]\}$ la familia binomial, o $\{\text{Normal}(x;(\mu,\sigma^2)) \mid x \in \mathbb{R}, \mu \in \mathbb{R}, \text{and}\, \sigma^2 \in [0,\infty)\}$ la familia normal. Si sabemos que una variable aleatoria X tiene una distribución

Palabras clave: Distribuciones condicionales, suficiencia, bondad de ajuste.

que pertenece a una familia paramétrica $\{F(x;\theta) \mid \theta \in \Theta\}$, el valor del parámetro θ determina únicamente la distribución of X.

Recordemos que si $T_n : \mathbb{R}^n \to \mathbb{R}^k$ es un estimador suficiente de θ, entonces si definimos la distribución condicional de la muestra dada T_n, se obtiene:

$$\hat{F}_n(x_1, \ldots, x_i) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_i \leq x_i \mid T_n = t) \qquad (1)$$

donde $(x_1, \ldots, x_i) \in \mathbb{R}^i$ para toda $i \in \{1, 2, \ldots, n\}$. En particular si $i = 1$, (1) se conoce como el estimador Rao-Blackwell de la función de distribución $F(x;\theta)$.

Si estamos tratando con variables aleatorias que son discretas, con valores en un conjunto numerable $M \subset \mathbb{R}$, parece más fácil trabajar con las familias de densidades condicionales dadas T_n. Sea $\{f(x;\theta) \mid \theta \in \Theta\}$, donde suponemos que $\Theta \subset \mathbb{R}$, y si T_n es una estadística suficiente de θ, usamos la densidad condicional discreta definida como

$$\hat{f}_n(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T_n = t), \quad (2)$$

donde $(x_1, \ldots, x_n) \in M^n$. De acuerdo a la teoría de estadísticas suficientes, véase por ejemplo [1], los valores de $\hat{f}_n$ no dependen de θ, y en muchos casos como veremos abajo, podemos encontrar expresiones cerradas para estas densidades.

En este trabajo se proponen pruebas de bondad de ajuste del tipo

$$H_0 : X_i \rightsquigarrow f(x;\theta) \text{ para algún } \theta \in \Theta \qquad vs. \;\; H_1 : X_i \not\rightsquigarrow f(x;\theta) \text{ para ningún } \theta \in \Theta. \quad (3)$$

También se hace una nueva propuesta para probar

$$H_0 : X_i \rightsquigarrow f(x;\theta) \text{ para algún } \theta \in \Theta \qquad vs. \;\; H_1 : X_i \rightsquigarrow g(x;\gamma) \text{ para algún } \gamma \in \Gamma, \quad (4)$$

donde Γ es otro espacio parametral de una densidad discreta $g(x;\gamma)$. El primer autor que usó la densidad condicional dada una estadística suficiente minimal fue Fisher en [2].

Los resultados obtenidos para variables discretas se encuentran en una muy amplia familia de distribuciones llamada la familia de **distribuciones de series de potencias**, que se pueden escribir como:

$$P(X = x) = \frac{a_x \theta^x}{\eta(\theta)} \quad \text{para} \quad x = 0, 1, 2, \ldots, \qquad (5)$$

donde $\theta > 0$, $a_x \geq 0$ y $\eta(\theta) = \sum_{x=0}^{\infty} a_x \theta^x$, la función $\eta(\cdot)$ se llama **función serie**. Por supuesto, a_x puede ser cero para $x \geq n$, dando lugar a variables aleatorias discretas finitas. Es conocido, véase [6], que la suma de n variables aleatorias independientes de la misma serie de potencias, tiene una distibución de la misma clase con función serie $[\eta(\theta)]^n$

Hay cuatro familias muy conocidas y usadas en la práctica que forman parte de la familia de series de potencias:
i) Si X es una variable geómetrica con parámetro $0 \leq p \leq 1$, entonces $0 \leq 1-p \leq 1$ es la probabilidad de un fracaso, y X mide el número de fracasos antes del primer éxito, en eventos independientes. y como $\eta(1-p) = \sum_{x=0}^{\infty}(1-p)^x = \frac{1}{1-(1-p)} = \frac{1}{p}$, entonces si tomamos $a_x = 1$ para cada $x \in \{0,1,2\ldots\}$, entonces

$$P(X=x) = \frac{a_x(1-p)^x}{\eta(1-p)} = p(1-p)^x \quad \text{para cada} \quad x = 0,1,2,\ldots,$$

que es una variable geómetrica con parámetro $(1-p)$ denotado por $X \rightsquigarrow G(p)$.
ii) Si X es una variable binomial con parametros $k \in \mathbb{N}$ y $0 < p < 1$ y $x \in \{0,1,2,\ldots,k\}$. Entonces

$$\sum_{x=0}^{k} \left(\begin{array}{c} k \\ x \end{array} \right) \left(\frac{p}{1-p}\right)^x (1)^{k-x} = \left(1 + \frac{p}{1-p}\right)^k = \frac{1}{(1-p)^k}.$$

Por lo tanto, si tomamos $a_x = \left(\begin{array}{c} k \\ x \end{array} \right)$ para $x = 0,1,2\ldots,k$. Por lo que

$$\begin{aligned} P(X=x) &= \frac{a_x \left(\frac{p}{1-p}\right)^x}{\sum_{x=0}^{k} \left(\begin{array}{c} k \\ x \end{array} \right) \left(\frac{p}{1-p}\right)^x (1)^{k-x}} \\ &= \frac{\left(\begin{array}{c} k \\ x \end{array} \right) \left(\frac{p}{1-p}\right)^x}{\frac{1}{(1-p)^k}} = \left(\begin{array}{c} k \\ x \end{array} \right) p^x (1-p)^{k-x}, \end{aligned}$$

que es una variable binomial con parámetros k y $0 < p < 1$.
iii) Si X es una variable aleatoria Poisson con parámetro $\lambda > 0$, si tomamos $a_x = \frac{1}{x!}$ para $x \in \{0,1,2\ldots\}$ entonces

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}.$$

Por lo que

$$P(x=x) = \frac{\frac{1}{x!}\lambda^x}{\sum_{x=0}^{\infty} \frac{\lambda^x}{x!}} = \frac{\lambda^x e^{-\lambda}}{x!},$$

que es una variable Poisson con parámetro $\lambda > 0$, denotado por $X \rightsquigarrow P(\lambda)$.

iv) Si X es una variable binomial negativa con parámetros $r \in \mathbb{N}$ y $0 < p < 1$. Primero observamos que

$$\binom{r+x-1}{x} = \binom{r+x-1}{r-1},$$

y

$$\begin{aligned}\binom{r+x-1}{r-1} &= \frac{(r+x-1)!}{(r-1)!x!} = \frac{(r+x-1)(r+x-2)\cdots r}{x!} \\ &= \frac{(-1)^x(-r)(-r-1)(-r-2)\cdots(-r-x+1)}{x!} \\ &= (-1)^x \binom{-r}{x} \qquad (6)\end{aligned}$$

Así, por (6), si $q = 1-p$ entonces

$$p^{-r} = (1-q)^{-r} = \sum_{x=0}^{\infty} \binom{-r}{x} (-q)^x = \sum_{x=0}^{\infty} \binom{r+x-1}{x} (1-p)^x.$$

Por lo que, si tomamos $a_x = \binom{r+x-1}{x}$, para $x \in \{0, 1, \ldots\}$, obtenemos que

$$P(X = x) = \frac{a_x(1-p)^x}{\sum_{x=0}^{\infty} a_x q^x} = \binom{r+x-1}{x} (1-p)^x p^r,$$

que es una variable binomial negativa con parámetros $r \in \mathbb{N}$ y $0 < p < 1$.

Tenemos un importante Teorema que relaciona dsitribuciones de series de potencia con funciones de distribución condicionales dada T_n una estadística suficiente.

Teorema 1.1 *Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución de series de potencia como en la ecuación (5), donde $\theta > 0$ y $\eta(\theta) = \sum_{x=0}^{\infty} a_x \theta^x$. Sea $T_n = \sum_{i=1}^{n} X_i$. Entonces*

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T_n = t) = \frac{a_{x_1} a_{x_2} \cdots a_{x_n}}{\hat{a}_t}, \qquad (7)$$

donde $\hat{a}_t$ corresponde a la densidad de T_n, para cada $x_i \geq 0$ donde $i = 1, 2, \ldots, n$ tales que $\sum_{i=1}^{n} x_i = t$. Además, si $X_1, X_2, \ldots X_n$ tienen la distribución condicional dada $T_n = \sum_{i=1}^{n} X_i = t$, para n y t enteros no negativos de la ecuación (7). Entonces las X_i tienen la densidad dada en la ecuación (5).

Una prueba del Teorema 1.1 puede encontrarse en [3]. Para encontrar los valores de $\hat{a_x}$ se usa lo que mencionamos después de la definicion de distribucion de series de potencia en (5). Ahora damos las densidades condicionales dada $T_n = \sum_{i=1}^{n} x_i = t$ an los cuatro casos de arriba.

i) En el caso de una variable geométrica con parámetro p, i.e., $X \rightsquigarrow G(p)$,

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T_n = t) = \frac{1}{\binom{t+n-1}{n-1}}, \tag{8}$$

es decir, obtenemos una densidad *uniforme* sobre el número de posibles formas de sumar n enteros no negativos que suman t. De hecho el denominador en (8) da una fórmula para ese número de formas.
ii) En el caso de una variable binomial con parámetros k y p, i.e., $X \rightsquigarrow \mathrm{Bin}(k,p)$, tenemos

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T_n = t) = \frac{\binom{k}{x_1}\binom{k}{x_2}\cdots\binom{k}{x_n}}{\binom{nk}{t}} \tag{9}$$

es decir, la densidad conjunta es una **hipergeométrica** «multivariada», una versión de n clases de tamaño k cuya suma es t.
iii) En el caso de una variable con densidad Poisson con parámetro λ, es decir, $X \rightsquigarrow P(\lambda)$, se tiene que $a_{x_i} = 1/x_i!$ para cada $i \in \{1, 2, \ldots, n\}$ y $\hat{a}_t = 1/t!$, entonces

$$\begin{aligned} &P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T_n = t) \\ &= \binom{t}{x_n}\binom{t-x_n}{x_{n-1}}\binom{t-x_n-x_{n-1}}{x_{n-2}}\cdots\binom{t-x_n-\cdots-x_2}{x_1}\left(\frac{1}{n}\right)^t \\ &= \frac{t!}{x_1!\cdots x_n!}\left(\frac{1}{n}\right)^t, \end{aligned} \tag{10}$$

que corresponde a una densidad *multinomial* donde $p_1 = p_2 = \cdots = p_n = \frac{1}{n}$, véase [5].
iv) En el caso de una variable con densidad binomial negativa con parámetros $r \in \mathbb{N}$, $0 < p < 1$, es decir, $X \rightsquigarrow \mathrm{BN}(r,p)$, aquí se tiene que $a_{x_i} = \binom{r+x_i-1}{r-1}$ para $i \in \{1, 2, \ldots, n\}$ y $\hat{a}_t = \binom{nr+t-1}{nr-1}$. Entonces

$$\begin{aligned} P\ (\ &X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T_n = t) \\ = \ &\frac{\binom{r+x_n-1}{r-1}\binom{r+x_{n-1}-1}{r-1}\cdots\binom{r+x_1-1}{r-1}}{\binom{nr+t-1}{nr-1}}, \end{aligned} \tag{11}$$

esta densidad conjunta no es muy conocida pero es similar a una hipergeométrica multivariada con n clases, pero en este caso el tamaño de cada clase varía de acuerdo a los valores de las $x_i's$, y el número de observaciones en cada clase, $r-1$ está fijo.

2. Pruebas de bondad de ajuste

Vimos que el número de formas en las que sumando n enteros no negativos nos de como resultado t, está dado por

$$\binom{t+n-1}{n-1} \tag{12}$$

Definición 2.1 *Llamaremos un* **arreglo** *de* (t,n)*, con* $n \in \mathbb{N}$ *y* t *un entero no negativo, a un vector de números enteros no negativos* $(x_n, x_{n-1}, \ldots, x_2, x_1)$ *si* $x_n \geq x_{n-1} \geq \cdots \geq x_2 \geq x_1 \geq 0$*, y además* $\sum_{i=1}^{n} x_i = t$*. En teoría combinatoria se le llama* **partición entera** *cuando* $n = t$*.*

El número de arreglos de (t,n) es considerablemente más pequeño que el número de formas de sumandos en (12). No existe una fórmula cerrada para calcular el número de arreglos de (t,n). La fórmula recursiva de Euler da la solución, solo en el caso de que $t = n$, y se denota por p_n, donde

$$p_n = \sum_{m=1}^{\infty} (-1)^{m+1} (p_{n-m(3m-1)/2} + p_{n-m(3m+1)/2}), \tag{13}$$

donde $p_0 = 1$ y $p_k = 0$ if $k < 0$.

Usando un computadora es muy sencillo generar todos los arreglos para las parejas (t,n) con n y t no muy grandes, y usando las fórmulas (8), (9), (10) y (11), es muy fácil evaluar las densidades condicionales dado que $T_n = t$ para todos los arreglos, el paso siguiente es simplemente ordenarlos de mayor a menor valor de la densidad, y utilizar estos valores, y regiones de mayor densidad, utilizando ideas del tipo de Estadística Bayesiana, para hacer pruebas del tipo

$$\begin{aligned} &H_0 : X \rightsquigarrow f(x;\theta_0) \text{ para alguna } \theta_0 \in \Theta \\ &\qquad\qquad \textit{vs.}\ H_1 : X \not\rightsquigarrow f(x;\theta) \text{ para ninguna } \theta \in \Theta, \end{aligned} \tag{14}$$

para cualquier f del tipo analizado arriba. O bien del pruebas del tipo

$$\begin{aligned} &H_0 : X \rightsquigarrow f(x;\theta_0) \text{ para alguna } \theta_0 \in \Theta \\ &\qquad\qquad \textit{vs.}\ H_1 : X \not\rightsquigarrow g(x;\gamma_1) \text{ para alguna } \gamma_1 \in \Gamma, \end{aligned} \tag{15}$$

donde g es una densidad de serie de potencias como arriba. Además, en el segundo caso (15) se puede evaluar β la probabilidad del error tipo II, de manera directa.

3. Algunas aportaciones de Federico O'Reilly Togno a la estadística

El trabajo de Federico O'Reilly, se centró en el área de bondad de ajuste, y nosotros creemos que una de sus principales aportaciones teóricas y prácticas, fue la idea de utilizar las distribuciones o densidades condicionales dada una estadística suficiente T_n, que por definición resulta ser una función de una muestra aleatoria que contiene toda la información posible acerca de un parámetro de interés.

En el trabajo que realizamos con Federico, y que se expone brevemente arriba, fue muy interesante para los tres coautores, y Federico estaba muy entusiasmado con los resultados que pudimos obtener para las técnicas de bondad-de-ajuste en el caso de variables discretas, ya que en otras ocasiones ya había trabajado en este tema, incluyendo a Raúl Rueda, como en el caso de la variable Poisson, véase [9]. En particular, las densidades que se obtuvieron en los cuatro casos que se analizaron arriba, resultaron ser muy intersantes, y sorprendentemente muy conocidas. Por otra parte los adelantos en los programas usados, y las mejoras tecnológicas en las computadoras, hicieron posible utilizar un gran número de simulaciones, para poder obtener los p-valores de una prueba, sin tener que contar necesariamente con la distribución asintótica.

El primer trabajo en el que Federico utilizó unas ideas iniciales de estas técnicas de distribuciones condicionales fue en un trabajo conjunto con Leticia Gracia-Medrano que se publicó en 2006, que se dedicó a utilizar distribuciones condicionales para resolver problemas de bondad-de-ajuste, véase [8].

Posteriormente, Federico trabajó conjuntamente con Richard Lockhart y Michael Stephens, en un trabajo teórico que publicaron conjuntamente en (2007) en la revista Biometrika, véase [7]. En este trabajo se introdujo la idea de considerar muestras, llamadas «look-alike samples» de las distribuciones condicionales dada T_n una estadística suficiente, y tuvo varias aplicaciones interesantes.

Los autores de este artículo trabajamos con Federico O'Reilly en otro trabajo que relaciona el método de Durbin llamado substitución aleatoria, en el cuál se transforma un problema de bondad-de-ajuste compuesto y se reduce a un problema simple, y usando un gran número de «look-alike samples» se puede utilizar una propuesta inversa de la transformación de Durbin para encontrar el p-valor de cualquier estadítica de prueba en un problema de bondad-de-ajuste compuesto. En este trabajo se encuentran algunas distribuciones asintóticas y algunos ejemplos de interés en la literatura, véase [4].

Este trabajo es un homenaje muy merecido para recordar la gran labor de nuestro colega y amigo Federico O'Reilly Togno, quien es ampliamente reconocido como un promotor y formador de varias generaciones de estadísticos en México, así como uno de los iniciadores de los programas de posgrado de estadística. Siempre recordaremos su gran sentido de humor y su interés en el desarrollo formal de nuestras áreas de trabajo.

Bibliografía

[1] R. Bahadur, «Sufficiency and statistical decision functions», *Ann Math Statist*, núm. 25, 1954, 423–462.

[2] R. Fisher, «The significance of deviations from expectations in a Poisson series», *Biometrics*, núm. 6, 1950, 17–24.

[3] J. González-Barrios, F. O'Reilly y R. Rueda, «Goodness of fit for discrete random variables using the conditional density», *Metrika*, núm. 64, 2006, 77–94.

[4] ———, «Durbin's random substitution and conditional Monte Carlo», *Metrika*, núm. 72, 2010, 369–383.

[5] N. Johnson, S. Kotz y N. Balakrishnan, *Discrete multivariate distributions*, Ed. Wiley, New York, 1997.

[6] N. Johnson, S. Kotz y A. Kemp, *Univariate discrete distributions*, Ed. Wiley, New York, 1992.

[7] R. Lockhart, F. O'Reilly y M. Stephens, «Use of the Gibbs sampler to obtain conditional tests, with applications», *Biometrika*, núm. 94, 2007, 992–998.

[8] F. O'Reilly y L. Gracia-Medrano, «Transformations for testing the fit of the inverse-Gaussian distribution», *Commun. Statist. Theory Methods*, vol. 33, 2006, 919–924.

[9] R. Rueda, V. Pérez-Abreu y F. O'Reilly, «Goodness-of-fit for the Poisson distribution based on the probability generating function», *Commun. Statist. Theory Methods*, vol. 20, núm. 10, 1991, 3093–3110.