

División de Ciencias Básicas e Ingeniería

**Maestría en Ciencias Matemáticas Aplicadas e
Industriales**



**Cálculo del p-valor en pruebas de bondad de
ajuste**

Tesis que para obtener el título de:

Maestro en Ciencias

Presenta:

Jesús Iván Beltrán Beltrán

Asesores:

Dr. Federico J. O'Reilly Togno

Dra. Blanca Rosa Pérez Salvador

México, D.F. a 10 de diciembre de 2012

Lo más sabio es el tiempo, porque aclara todo.

Tales de Mileto





Índice

Índice	v
Resumen	ix
Introducción	xi
1. Pruebas de hipótesis	1
1.1. Hipótesis simple vs alternativa simple	4
1.2. Hipótesis compuestas	6
2. Pruebas de bondad de ajuste para distribuciones continuas	9
2.1. La prueba χ^2 (o X^2)	10
2.2. Pruebas basadas en la función de distribución empírica EDF	11
2.2.1. Definición de la función de distribución empírica	11
2.2.2. Propiedades de la función de distribución empírica	12
2.2.3. Estadísticas de prueba EDF	14
2.3. Proceso empírico	16
2.3.1. Proceso empírico	17
2.3.2. Proceso empírico con parámetros estimados	18
2.3.3. Procesos Gaussiano y Browniano atado	19



2.3.4. Familias de localización y/o escala	19
2.4. Simulación del p-valor (o p-value) para el caso simple	20
2.5. Simulación del p-valor para familias de localización y/o escala	21
2.6. Ejemplo	22
3. Uso del estimador Rao-Blackwell para la generación de muestras-*	25
3.1. Estimador Rao-Blackwell	26
3.2. Generación de muestras Look-Alike	27
3.2.1. Simulación del p-valor condicional	29
3.3. Distribución exponencial truncada	31
3.3.1. Ejemplo	33
3.4. Distribución Gaussiana inversa	34
3.4.1. Ejemplo	35
4. Pruebas de bondad de ajuste para distribuciones discretas	37
4.1. Proceso empírico asociado a la función generadora de probabilidades	38
4.1.1. Estadística de prueba $d_n(\theta)$	39
4.1.2. Propuesta de estadística de prueba $d_n^*(\theta)$	40
4.2. Estadísticas de prueba Cramér-von Mises	42
4.2.1. Estadísticas de prueba Cramér-von Mises caso general	43
4.3. Estadística de prueba de razón de verosimilitudes generalizadas	45
4.3.1. Divergencia Kullback-Leibler	46
4.4. Uso de la probabilidad condicional	47
4.4.1. Resultados teóricos	47
4.4.2. Descripción de la prueba de bondad de ajuste	51
4.5. Distribución binomial negativa generalizada	53

4.5.1. Distribución series de potencias modificada	54
4.5.2. Probabilidad condicional	55
4.6. Generación de muestras Look-Alike	55
4.6.1. Distribución Rao-Blackwell para algunas distribuciones	57
4.7. Aplicaciones	59
4.7.1. Nidos de gorrión (Sparrow nests)	59
4.7.2. Ciclones tropicales de categoría 4 y 5 en el Atlántico	60
4.7.3. Mosquito transmisor de malaria	61
Conclusiones y perspectivas	63
Bibliografía	65
Apéndice	69
A.1.	69





Resumen

En este trabajo se exponen algunos métodos para las pruebas de bondad de ajuste continuo y discreto donde se hace uso del llamado proceso empírico y también se utilizan pruebas recientes de la literatura. Se realizan simulaciones de muestras condicionales con el uso de la función de distribución Rao-Blackwell, como en el caso de la Gaussiana inversa en O'Reilly y Gracia-Medrano (2006) o como lo es en los casos discretos discutidos por González Barrios et al., (2006), donde se utilizan las herramientas computacionales existentes hoy en día. Se comenta la distribución Rao-Blackwell para la distribución de series de potencia, así como para sus casos particulares; los cuales son, la binomial, la binomial negativa y la Poisson. Para las pruebas de bondad de ajuste discreto se propone una extensión de estadística de prueba, que es función de la generadora de probabilidades. También se desarrolla la función de distribución Rao-Blackwell para la binomial negativa generalizada. Aunado a lo anterior se ejemplifica con una aplicación de bondad de ajuste para la distribución binomial negativa generalizada.

Palabras clave: Prueba de bondad de ajuste, Distribución binomial negativa generalizada, Proceso empírico, Distribución Rao-Blackwell, p-valor condicional.



Introducción

En inferencia estadística, con un conjunto de datos muestrales, se plantea inferir sobre la población. Esas inferencias o decisiones que se pueden tomar son referentes a estimación de parámetros o prueba de hipótesis. Por ejemplo, cuando se hace el supuesto sobre la variable de interés asociada a los elementos de una población, que sigue una distribución particular; ó en el caso de regresión lineal, cuando se tiene una variable explicativa x , y se supone un modelo que relaciona a la variable de interés Y , como $Y = f(x) + \epsilon$ considerando que $\epsilon \sim N(0, \sigma^2)$.

Estos supuestos pueden ser ciertos o no, los cuales se llaman hipótesis estadísticas. Al procedimiento mediante el cual se investiga la verdad o falsedad de una hipótesis se le llama prueba de hipótesis.

Las pruebas de bondad de ajuste son un caso particular de pruebas de hipótesis; la particularidad radica que en las pruebas de hipótesis clásicas, se supone que tanto en la hipótesis nula como en la hipótesis alternativa es conocida la forma funcional de la densidad; y en el caso de las pruebas de bondad de ajuste conocemos la forma funcional en la hipótesis nula, pero en la hipótesis alternativa se desconoce la función de densidad, y por ello, el conjunto de posibilidades puede ser muy grande.

En un fenómeno de interés, es importante verificar la suposición de que los datos experimentales se ajustan, o no, a cierto modelo probabilístico. En este caso se está ante un problema de bondad de ajuste. Las pruebas de bondad de ajuste tienen por objetivo determinar si los datos se generaron o se pueden modelar con una determinada función distribución.

Una teoría de bondad de ajuste para el caso continuo muy desarrollada es la que se



basa en la función de distribución empírica, y para el caso discreto, existe un símil.

Con las facilidades de cómputo que se tienen en la actualidad se ha propuesto, además del uso del remuestreo paramétrico (bootstrap), la simulación de valores de la estadística de prueba, pero la simulación no se hace con la función de distribución de la estadística de prueba (que depende de los parámetros) sino que se hace con la distribución condicional dada una estadística suficiente. Con este procedimiento, se evalúa el p-valor (condicional), tan exacto como el número de simulaciones de muestras condicionales. Hay trabajos relativamente recientes como el de Engen y Lillegard, (1997), Lindqvist y Taraldsen, (2005), O'Reilly y Gracia-Medrano, (2006), González-Barrios et al., (2010) y otros ahí citados, en que la propuesta (digamos el paradigma), es poder simular el p-valor condicional.

El presente trabajo se basa principalmente en el desarrollo de algunas técnicas de bondad de ajuste en el caso discreto, mencionando el caso continuo; haciendo uso de simulaciones para el cálculo de los p-valores asociados a las estadísticas de prueba. Para esto se utilizan las herramientas computacionales existentes hoy en día.

Esta tesis está estructurada en 4 capítulos, iniciamos con los preliminares donde se da una breve introducción a la teoría básica de las pruebas de hipótesis paramétricas.

En el segundo capítulo se presenta una introducción a las pruebas de bondad de ajuste continuo. Se define la función de distribución empírica, así como también las propiedades más importantes que cumple, se dan las estadísticas de prueba más usadas y el procedimiento para calcular el p-valor, por último se proporciona un ejemplo.

En el tercer capítulo se revisa la definición de la función de distribución Rao-Blackwell, por medio de la cual se proporciona la teoría para la generación de muestras condicionales, así como el procedimiento para el cálculo de los p-valores condicionales (asociados), y por último se dan algunas aplicaciones.

En el cuarto capítulo se presenta una introducción a las pruebas de bondad de ajuste discreto. Se comenta un proceso empírico, símil al del caso continuo, basado en la función de generatriz de probabilidades. Se proporcionan algunas estadísticas de prueba, para la simulación de sus p-valores condicionales. También se presenta una prueba de bondad de ajuste discreto que se basa en el uso de la probabilidad condicional; la cual se plantea



en González-Barrios et al. (2006). La aportación de este trabajo, es proponer una nueva estadística de prueba, extensión de la propuesta en O'Reilly y Rueda (1999), para el caso discreto. Además, se estudia la binomial negativa generalizada, para la cual se encuentra la probabilidad condicional. Al final se proporcionan unas aplicaciones.

Por último se dan las conclusiones y perspectivas futuras para continuar la investigación.



Capítulo 1

Pruebas de hipótesis

El material presentado en este capítulo fue consultado en el libro Mood et al. (1974).

Con base en una muestra, un ingeniero tiene que decidir, si la duración promedio de cierto tipo de neumático es cuando menos de 22,000 millas, o si basado en experimentos, un agrónomo tiene que decidir si un determinado fertilizante produce mayor rendimiento de frijol de soya que otro fertilizante; en ambos casos se están enfrentando a problemas de pruebas de hipótesis estadísticas.

En inferencia estadística, hay dos problemas centrales, la estimación de parámetros y las pruebas de hipótesis. Una hipótesis estadística es una suposición que se hace acerca de la distribución de una variable aleatoria, por ejemplo, que cierta distribución tiene media 1.5. Una prueba estadística de una hipótesis es un procedimiento en el cual se usa una muestra con el fin de determinar cuándo podemos rechazar la hipótesis, es decir, actuar como si fuera falsa o cuándo no debiéramos rechazar ésta, es decir, actuar como si fuera verdadera.

Con mucha frecuencia tenemos que tomar decisiones en las que el azar juega un papel muy importante. Si tenemos que elegir de entre dos posibilidades, la decisión entre ambas se debería basarse en el resultado de una prueba estadística.

Se necesita introducir algo de lenguaje, notación y dar algunas definiciones. Se asume que se tiene una muestra aleatoria X_1, \dots, X_n de alguna función de densidad $f(\cdot; \theta)$. Una hipótesis estadística será una hipótesis sobre la distribución de la población.



Definición 1 (Hipótesis estadística) *Una hipótesis estadística es una aseveración o conjetura acerca de una distribución de una o más variables aleatorias. Si la hipótesis estadística está totalmente especificada se tiene el llamado caso simple, en otro caso es referido como caso compuesto.*

Notación 1 *Se denota a la hipótesis estadística, por la letra \mathcal{H} , seguido de dos puntos que a su vez es seguida por la afirmación que especifica la hipótesis.*

Ejemplo. Sea

$$X_1, \dots, X_n$$

una muestra aleatoria de una distribución normal con media θ y varianza 25; La hipótesis estadística es que la media es menor o igual a 18, la cual se denota por, $\mathcal{H} : \theta \leq 18$, caso compuesto (ya que la distribución no está completamente especificada). Por otro lado, la hipótesis $\mathcal{H} : \theta = 18$ es un caso simple ya que la distribución está completamente especificada.

Definición 2 (Prueba de una hipótesis estadística) *Una prueba de una hipótesis estadística \mathcal{H} es una regla o procedimiento para decidir o no, si se rechaza \mathcal{H} .*

Notación 2 *Se denota a la prueba de la hipótesis estadística, por la letra Υ .*

Ejemplo. Sea

$$X_1, \dots, X_n$$

una muestra aleatoria de una distribución normal con media θ y varianza 25. Considere $\mathcal{H} : \theta \leq 18$. Una posible prueba Υ es la siguiente: Rechazar \mathcal{H} si y sólo si $\bar{X} > 18 + \frac{5}{\sqrt{n}}(1.96)^1$.

Una prueba puede ser aleatorizada o no ser aleatorizada. La prueba Υ dada en el ejemplo anterior es una prueba no aleatorizada. Si Υ' es la estadística de prueba discreta, por ejemplo,

$$T \sim \text{Bin}(n = 5, p) \quad y$$

¹El número 1.96 corresponde a una prueba con nivel 0.025.

$$\mathcal{H}_0 : p = \frac{1}{2} \text{ vs. } \mathcal{H}_1 : p > \frac{1}{2}.$$

Como T sólo toma valores en $\{0, 1, 2, 3, 4, 5\}$. Si usásemos rechace si y sólo si $T > 4$, la probabilidad del error tipo I es $\frac{1}{32} = 0.03125$ y si usásemos rechace si y sólo si $T > 3$, la probabilidad del error tipo I sería $\frac{6}{32} = \frac{3}{16} = 0.1875$ (ver definición 5).

Cuando por necesidad (matemática) se requiere una prueba que tenga probabilidad del error tipo I exactamente igual a 0.05, entonces se procede ALEATORIZANDO; esto es se usa el procedimiento de rechazo con $T > 4$ con probabilidad q y el otro procedimiento con $T > 3$ con probabilidad $1 - q$ y se encuentra q que satisfaga

$$q \left(\frac{1}{32} \right) + (1 - q) \left(\frac{3}{16} \right) = 0.05.$$

Esto nos da una prueba que con probabilidad 1 rechaza si $t = 5$, con probabilidad 0.05 rechaza si $t = 4$ y con probabilidad 0 rechaza si $t = 0, 1, 2, 3$. Esa es una prueba aleatorizada. Las pruebas aleatorizadas no se verán más, pero es importante tener una idea de dicha prueba, a continuación se da la definición de ambas pruebas.

Notación 3 Sea \mathcal{X} el espacio muestral de las observaciones; esto es, $\mathcal{X} = \{(x_1, \dots, x_n) : \text{es un posible valor de } (X_1, \dots, X_n)\}$.

Definición 3 (Prueba no aleatorizada y región crítica) Sea una prueba Υ de una hipótesis estadística \mathcal{H} que se define como sigue: Se rechaza \mathcal{H} si y sólo si $(x_1, \dots, x_n) \in C_\Upsilon$, donde C_Υ es un subconjunto de \mathcal{X} ; luego Υ es llamada prueba no aleatorizada, y C_Υ es llamada región crítica de la prueba Υ .

Definición 4 (Prueba aleatorizada) Sea una prueba Υ de una hipótesis estadística \mathcal{H} que es definida como prueba aleatorizada si Υ es definida por la función $\psi_\Upsilon(x_1, \dots, x_n) = P[\mathcal{H} \text{ es rechazada} | (x_1, \dots, x_n) \text{ es observada}]$. La función $\psi_\Upsilon(\cdot, \dots, \cdot)$ es llamada función crítica de la prueba Υ .

En muchas problemas de pruebas de hipótesis dos hipótesis son discutidas: La primera, la hipótesis que está siendo probada, es llamada hipótesis nula, denotada por \mathcal{H}_0 , y la segunda es llamada hipótesis alternativa, denotada por \mathcal{H}_1 .



Definición 5 (Tipos de errores y tamaño del error) Rechazar \mathcal{H}_0 cuando es verdadera es llamado el error tipo I, y aceptar \mathcal{H}_0 cuando es falsa es llamado el error tipo II. El tamaño del error tipo I es definido como la probabilidad de cometer el error tipo I, y similarmente el tamaño del error tipo II es la probabilidad de cometer el error tipo II.

Definición 6 (Función potencia) Sea Υ la prueba para la hipótesis nula \mathcal{H}_0 . La función potencia de la prueba Υ , denotada por $\pi_{\Upsilon}(\theta)$, es definida como la probabilidad de que \mathcal{H}_0 sea rechazada cuando el verdadero valor del parámetro es θ .

La función potencia jugará un papel importante en las pruebas de hipótesis similar al que el error cuadrático medio juega en la teoría de la estimación. Si Υ es una prueba no aleatorizada, se tiene que $\pi_{\Upsilon}(\theta) = P_{\theta}[(X_1, \dots, X_n) \in C_{\Upsilon}]$, donde C_{Υ} es la región crítica asociada con la prueba Υ .

Definición 7 (Tamaño de la prueba) Sea Υ la prueba para la hipótesis nula $\mathcal{H}_0 : \theta \in \Theta_0$, donde $\Theta_0 \subset \Theta$; es decir, Θ_0 es un subconjunto del espacio paramétrico Θ . El tamaño de la prueba Υ de \mathcal{H}_0 se define como $\sup_{\theta \in \Theta_0} [\pi_{\Upsilon}(\theta)]$.

Muchos escritores utilizan los términos nivel de significancia y el tamaño de la prueba de manera indistinta.

1.1. Hipótesis simple vs alternativa simple

En esta sección se considera las pruebas con hipótesis nula simple contra la hipótesis alternativa simple, este caso no es muy usado en aplicaciones estadísticas, pero sirve para una introducción a la teoría de la pruebas de hipótesis.

Se asume que se tiene una muestra que viene dada de una de dos distribuciones especificadas completamente, el objetivo es determinar cuál es. Es decir, se asume que



una muestra aleatoria X_1, X_2, \dots, X_n tiene densidad $f_0(x)$ ó $f_1(x)$, y se quiere probar $\mathcal{H}_0 : X_i \sim f_0(\cdot)$, contra $\mathcal{H}_1 : X_i \sim f_1(\cdot)$. Si sólo se tiene la observación x_1 , $f_0(\cdot)$ y $f_1(\cdot)$, se podría racionalmente pensar que la observación viene de $f_0(\cdot)$ si $f_0(x_1) > f_1(x_1)$, y viceversa, decidir que la observación viene de $f_1(\cdot)$, si $f_1(x_1) > f_0(x_1)$. Esta idea intuitiva lleva a la siguiente definición.

Definición 8 (Prueba de razón de verosimilitudes) Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de $f_0(\cdot)$ o $f_1(\cdot)$. Una prueba Υ de $\mathcal{H}_0 : X_i \sim f_0(\cdot)$, contra $\mathcal{H}_1 : X_i \sim f_1(\cdot)$ se define la prueba de razón de verosimilitudes:

Se rechaza \mathcal{H}_0 si $\lambda < k$,

Se acepta \mathcal{H}_0 si $\lambda \geq k$,

donde

$$\lambda = \lambda(x_1, \dots, x_n) = \frac{\prod_{i=1}^n f_0(x_i)}{\prod_{i=1}^n f_1(x_i)} = \frac{L_0(x_1, \dots, x_n)}{L_1(x_1, \dots, x_n)} = \frac{L_0}{L_1},$$

y k es una constante no negativa. ($L_j = L_j(x_1, \dots, x_n)$ es la función de verosimilitud para la muestra de la densidad $f_j(\cdot)$).

Para cada valor diferente de k se tienen diferentes pruebas. Para cada k fijo en la prueba se rechaza \mathcal{H}_0 si la razón (o cociente) de verosimilitudes es pequeña.

Pruebas más potentes

Sea X_1, \dots, X_n una muestra aleatoria que viene de la densidad $f_0(\cdot)$ ó de la densidad $f_1(\cdot)$. Vamos a escribir $f_0(x) = f(x; \theta_0)$ y $f_1(x) = f(x; \theta_1)$; entonces X_1, \dots, X_n es una muestra aleatoria de la familia paramétrica $\{f(x; \theta) : \theta = \theta_0 \text{ ó } \theta = \theta_1\}$. Donde $\Theta = \{\theta_0, \theta_1\}$ es la familia paramétrica, con θ_0 y θ_1 conocidos. Se quiere probar $\mathcal{H}_0 : \theta = \theta_0$ contra $\mathcal{H}_1 : \theta = \theta_1$. Asociado a cada prueba Υ de \mathcal{H}_0 contra \mathcal{H}_1 , queda definida su función potencia $\pi_\Upsilon(\theta)$. Una buena prueba es una para la cual

$$\pi_\Upsilon(\theta_0) = P[\text{rechazar } \mathcal{H}_0 | \mathcal{H}_0 \text{ es verdadera}]$$

es pequeña y $\pi_\Upsilon(\theta_1) = P[\text{rechazar } \mathcal{H}_0 | \mathcal{H}_0 \text{ es falsa}]$ fuera grande. Se podrían utilizar razonablemente los dos valores $\pi_\Upsilon(\theta_0)$ y $\pi_\Upsilon(\theta_1)$ para establecer criterios para definir la mejor

prueba. Siguiendo lo anterior, si fijamos el tamaño del error tipo I y minimizamos el tamaño del error tipo II, se sigue la siguiente definición.

Definición 9 (Pruebas más potentes) *Sea una prueba Υ^* de $\mathcal{H}_0 : \theta = \theta_0$ contra $\mathcal{H}_1 : \theta = \theta_1$, se dice ser una prueba más potente de tamaño α ($0 < \alpha < 1$) si y sólo si*

(i) $\pi_{\Upsilon^*(\theta_0)} = \alpha$.

(ii) $\pi_{\Upsilon^*(\theta_1)} \geq \pi_{\Upsilon(\theta_1)}$ para cualquier otra prueba Υ para la cual $\pi_{\Upsilon(\theta_0)} \leq \alpha$.

Una prueba Υ^* es más potente de tamaño α , si tiene tamaño α y entre todas las pruebas de tamaño menor o igual a α , es la que tiene mayor potencia.

El siguiente teorema es útil para encontrar una prueba más potente de tamaño α .

Teorema 1 (Lema de Neyman-Pearson) *Sea X_1, \dots, X_n una muestra aleatoria de $f(x; \theta)$, donde θ es uno de los valores θ_0 ó θ_1 , y sea $0 < \alpha < 1$ fijo.*

Sea k^ una constante positiva y C^* un subconjunto de \mathcal{X} que satisfice:*

(i) $P_{\theta_0}[(X_1, \dots, X_n) \in C^*] = \alpha$,

(ii) $\lambda = \frac{L(\theta_0; x_1, \dots, x_n)}{L(\theta_1; x_1, \dots, x_n)} = \frac{L_0}{L_1} \leq k^*$, si $(x_1, \dots, x_n) \in C^*$,

y $\lambda > k^$ si $(x_1, \dots, x_n) \in \overline{C^*}$. Luego, la prueba Υ^* correspondiente a la región crítica C^* es una prueba más potente de tamaño α de $\mathcal{H}_0 : \theta = \theta_0$ contra $\mathcal{H}_1 : \theta = \theta_1$. ($\overline{C^*}$ es el complemento de C^*).*

1.2. Hipótesis compuestas

En la sección anterior se consideró las pruebas de hipótesis simple contra alternativa simple. Ahora se ve el caso más general del problema de prueba de hipótesis que son las pruebas de hipótesis compuestas. Se asume que se tiene una muestra aleatoria de $f(x; \theta)$,



$\theta \in \Theta$, y se quiere probar $\mathcal{H}_0 : \theta \in \Theta_0$ contra $\mathcal{H}_1 : \theta \in \Theta_1$, donde $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$, y Θ_0, Θ_1 son disjuntos. Por lo general $\Theta_1 = \Theta - \Theta_0 (= \Theta \cap \Theta_0^c)$. A continuación se dan las definiciones para el caso de las pruebas de hipótesis compuestas.

Para una muestra aleatoria X_1, \dots, X_n de una densidad $f(x, ; \theta)$, $\theta \in \Theta$, se busca una prueba para $\mathcal{H}_0 : \theta \in \Theta_0$ contra $\mathcal{H}_1 : \theta \in \Theta_1 = \Theta - \Theta_0$, para eso se da la siguiente definición.

Definición 10 (Razón (o cociente) de verosimilitudes generalizadas)

Sea $L(\theta; x_1, \dots, x_n)$ la función de verosimilitud para la muestra X_1, \dots, X_n con densidad conjunta $f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta)$, donde $\theta \in \Theta$. La razón de verosimilitudes generalizadas, denotada por λ o λ_n , se define por

$$\lambda = \lambda_n = \lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta; x_1, \dots, x_n)}{\sup_{\theta \in \Theta} L(\theta; x_1, \dots, x_n)}.$$

Vea que λ es función de x_1, \dots, x_n , es decir, $\lambda(x_1, \dots, x_n)$. Cuando las observaciones son reemplazadas por sus correspondientes variables aleatorias X_1, \dots, X_n , se escribe Λ en lugar de λ . De hecho, Λ es una estadística ya que no depende de parámetros desconocidos. El denominador de Λ es la función de verosimilitud evaluada en los estimadores de máxima verosimilitud, y el numerador corresponde a la función de verosimilitud evaluada en los estimadores máximos verosímiles restringidos a \mathcal{H}_0 .

El valor λ de la estadística Λ es usada para resolver $\mathcal{H}_0 : \theta \in \Theta_0$ contra $\mathcal{H}_1 : \theta \in \Theta_1 = \Theta - \Theta_0$; usando que \mathcal{H}_0 es rechazada si y sólo si $\lambda \leq \lambda_0$, donde λ_0 es una constante fija que satisface $0 \leq \lambda_0 \leq 1$. La prueba de razón de verosimilitudes generalizadas tiene muy buenas propiedades.

Pruebas uniformemente más potentes

Al igual que las pruebas más potentes, se define en el caso de hipótesis compuesta las pruebas uniformemente más potentes.



Definición 11 (Pruebas uniformemente más potentes) *Sea una prueba Υ^* para $\mathcal{H}_0 : \theta \in \Theta_0$ contra $\mathcal{H}_1 : \theta \in \Theta - \Theta_0$, que se dice ser una prueba uniformemente más potente de tamaño α si y sólo si*

(i) $\sup_{\theta \in \Theta_0} \pi_{\Upsilon^*(\theta)} = \alpha.$

(ii) $\pi_{\Upsilon^*(\theta)} \geq \pi_{\Upsilon(\theta)}$ para toda $\theta \in \Theta - \Theta_0$ y para cualquier otra prueba Υ con tamaño menor o igual a α .

Una prueba Υ^* es uniformemente más potente de tamaño α si tiene tamaño α y entre todas las pruebas de tamaño menor o igual a α es la que tiene mayor potencia para todos los valores alternativos de θ .

Al igual que el Lema de Neyman-Pearson para el caso simple, se puede obtener un teorema análogo, en el caso de la familia exponencial para pruebas de hipótesis compuestas.

Teorema 2 *Sea X_1, \dots, X_n una muestra aleatoria de $f(x; \theta)$, $\theta \in \Theta$, donde Θ es algún intervalo. Se asume que $f(x; \theta) = a(\theta)b(x) \exp[c(\theta)d(x)]$, y un conjunto $t(x_1, \dots, x_n) = \sum_{i=1}^n d(x_i)$.*

(i) *Si $c(\theta)$ es monótona, creciente y si existe k^* tal que*

$$P_{\theta_0}[t(x_1, \dots, x_n) > k^*] = \alpha,$$

entonces la prueba Υ^ con región crítica $C^* = \{(x_1, \dots, x_n) : t(x_1, \dots, x_n) > k^*\}$ es una prueba uniformemente más potente de tamaño α de $\mathcal{H}_0 : \theta \leq \theta_0$ contra $\mathcal{H}_1 : \theta > \theta_0$ o de $\mathcal{H}_0 : \theta = \theta_0$ contra $\mathcal{H}_0 : \theta > \theta_0$.*

(ii) *Si $c(\theta)$ es monótona, decreciente y si existe k^* tal que*

$$P_{\theta_0}[t(x_1, \dots, x_n) < k^*] = \alpha,$$

entonces la prueba Υ^ con región crítica $C^* = \{(x_1, \dots, x_n) : t(x_1, \dots, x_n) < k^*\}$ es una prueba uniformemente más potente de tamaño α de $\mathcal{H}_0 : \theta \leq \theta_0$ contra $\mathcal{H}_1 : \theta > \theta_0$ o de $\mathcal{H}_0 : \theta = \theta_0$ contra $\mathcal{H}_0 : \theta > \theta_0$.*

Capítulo 2

Pruebas de bondad de ajuste para distribuciones continuas

Una familia paramétrica de funciones de distribución es el conjunto de funciones que depende de un parámetro θ ; el parámetro puede ser un número real o un vector y la familia se denota como:

$$\mathcal{F}_0 = \{F(\cdot; \theta) : \theta \in \Theta\}.$$

Una prueba de bondad de ajuste es un problema de prueba (o contraste) de hipótesis que se plantea como "probar" $H_0 : F \in \mathcal{F}_0$ vs. $H_1 : F \in (\mathcal{F} - \mathcal{F}_0)$, donde \mathcal{F} tiene dos posibilidades, es decir, es la familia no paramétrica de todas las distribuciones continuas (por ejemplo en \mathbb{R} o \mathbb{R}^+), o es la familia no paramétrica de todas las distribuciones discretas (por ejemplo en $\{0, 1, 2, \dots\}$). En el supuesto que la familia no paramétrica \mathcal{F} sea la de todas las distribuciones continuas, al problema se le llama prueba de bondad de ajuste continuo (o prueba de bondad de ajuste para distribuciones continuas) y para el caso que la familia no paramétrica \mathcal{F} sea la todas las distribuciones discretas, al problema se le llama prueba de bondad de ajuste discreto (o prueba de bondad de ajuste para distribuciones discretas).

Cuando en la prueba de bondad de ajuste el parámetro θ es conocido al problema se le llama prueba de bondad de ajuste simple y cuando θ es desconocido se le llama prueba de bondad de ajuste compuesto o en presencia de parámetros desconocidos.

En las siguientes secciones se van a presentar diferentes pruebas de bondad de ajuste para las distribuciones continuas que se encuentran en literatura.



2.1. La prueba χ^2 (o X^2)

Caso simple (parámetros conocidos)

Sea x_1, x_2, \dots, x_n una muestra aleatoria de $F(x)$, y sea $F_0(x)$ una distribución totalmente conocida. Se quiere probar

$$H_0 : F(x) = F_0(x) \quad \text{vs.}$$

$$H_1 : F(x) \neq F_0(x).$$

En 1900 Karl Pearson propuso un método de solución para el problema. Se considera que (a, b) es el recorrido natural de la variable aleatoria de interés; a y b no tienen restricciones; por ejemplo $a = -\infty$ y/o $b = \infty$. Se toma $k - 1$ número reales $y_1 < \dots < y_{k-1}$, que particionan al intervalo (a, b) en k subintervalos: $(y_0, y_1]$, $(y_1, y_2]$, \dots , $(y_{k-1}, y_k]$, donde $a = y_0 < y_1 < y_2 < \dots < y_{k-1} < y_k = b$.

Enseguida se cuenta cuantos elementos de la muestra pertenecen a cada subintervalo $(y_{i-1}, y_i]$, a este número se le denota n_i . Se calcula la probabilidad bajo F_0 de cada subintervalo; esta probabilidad se denota como p_i ,

$$p_i = F_0(y_i) - F_0(y_{i-1}), \quad i = 1, \dots, k.$$

Finalmente se define la estadística de prueba X^2 de Pearson como

$$X^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i},$$

donde n es el tamaño de la muestra y desde luego $n = n_1 + \dots + n_k$. Cuando la hipótesis nula es verdadera la distribución asintótica de X^2 converge a una distribución ji-cuadrada (χ^2) con $k-1$ grados de libertad¹. El procedimiento es rechazar H_0 cuando $X^2 > \chi_{k-1, 1-\alpha}^2$, donde $\chi_{k-1, 1-\alpha}^2$ es el percentil $1 - \alpha$ de la distribución ji-cuadrada con $k - 1$ grados de libertad, y α es el nivel de significancia elegido.

¹La estadística de prueba X^2 se puede utilizar para distribuciones continuas y discretas.



Caso compuesto (parámetros desconocidos)

En este caso se sustituyen los parámetros desconocidos por ciertos estimadores², al hacer esto bajo H_0 y bajo condiciones de regularidad la estadística converge a una χ^2 con $k - s - 1$ grados de libertad, donde s es el número de parámetros desconocidos. El procedimiento es rechazar H_0 cuando $X^2 > \chi_{k-s-1, 1-\alpha}^2$, con un nivel de significancia α .

El problema de determinar el número de intervalos k , y los números reales y_i es muy importante, esto es lo que le da a la prueba X^2 una gran debilidad, ya que la convergencia depende fuertemente de estos números. Cuando la muestra es demasiado pequeña no es recomendable usar esta prueba, en este caso existen otras pruebas que son mejores, las cuales revisaremos en las siguientes secciones.

2.2. Pruebas basadas en la función de distribución empírica EDF

2.2.1. Definición de la función de distribución empírica

En esta sección se da la definición de la función de distribución empírica y algunas de sus propiedades más importantes. La función de distribución empírica depende de la muestra y es:

²Son los estimadores de mínima ji-cuadrada.

Definición 12 Sea X_1, X_2, \dots, X_n una muestra aleatoria con función de distribución común F , se define la función de distribución empírica o EDF (por sus siglas en inglés empirical distribution function) por

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{[X_i \leq x]},$$

donde I_A es la función indicadora del evento A , es decir

$$I_A(x) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{en otro caso.} \end{cases}$$

La función de distribución empírica $F_n(x)$, da la proporción de valores de la muestra que son menores e iguales a x . Con un cálculo sencillo se muestra que la función de distribución empírica es igual a:

$$F_n(x) = \begin{cases} 0, & \text{si } x < x_{(1)}, \\ i/n & \text{si } x_{(i)} \leq x < x_{(i+1)}, \quad i = 1, \dots, n-1, \\ 1 & \text{si } x \geq x_{(n)}, \end{cases}$$

en donde las $x_{(i)}$'s son la muestra observada pero ordenada de menor a mayor.

Una aplicación importante de la función de distribución empírica es en las pruebas de bondad de ajuste debido a que la función de distribución empírica converge uniformemente y casi seguramente a la verdadera función de distribución F cuando $n \rightarrow \infty$. Cualquier medida de discrepancia entre $F_n(x)$ y una hipotética F_0 puede ser usada como una estadística de prueba para probar que $F(x)$ es igual a $F_0(x)$.

2.2.2. Propiedades de la función de distribución empírica

$F_n(x)$ es un estimador insesgado de $F(x)$

Como X_1, X_2, \dots, X_n es una muestra aleatoria de una distribución F , entonces $I_{[X_i \leq x]}$ son variables aleatorias independientes, donde cada $I_{[X_i \leq x]}$ tiene una distribución Bernoulli

con probabilidad de éxito,

$$P[I_{[X_i \leq x]}] = P[X_i \leq x] = F(x).$$

Por lo tanto

$$nF_n(x) \sim Bin(n, F(x)),$$

y tenemos que

$$\begin{aligned} E[F_n(x)] &= E \left[\frac{1}{n} \sum_{i=1}^n I_{[X_i \leq x]} \right] \\ &= \frac{1}{n} \sum_{i=1}^n E [I_{[X_i \leq x]}] \\ &= \frac{1}{n} \sum_{i=1}^n P[X \leq x] \\ &= \frac{nF(x)}{n} \\ &= F(x). \end{aligned}$$

Con lo que se prueba que $F_n(x)$ es un estimador insesgado para $F(x)$.

$F_n(x)$ es un estimador consistente para $F(x)$

Para determinar la consistencia de $F_n(x)$ como estimador de $F(x)$ se calcula la varianza de $F_n(x)$, para esto primero calculamos

$$\begin{aligned} Var[I_{[X_i \leq x]}] &= E [I_{[X_i \leq x]}^2] - E [I_{[X_i \leq x]}]^2 \\ &= E [I_{[X_i \leq x]}] - E [I_{[X_i \leq x]}]^2 \\ &= F(x) - F(x)^2 \\ &= F(x)(1 - F(x)). \end{aligned}$$

Por lo que

$$\begin{aligned} Var[F_n(x)] &= \frac{1}{n^2} \sum_{i=1}^n Var [I_{[X_i \leq x]}] \\ &= \frac{1}{n^2} \sum_{i=1}^n [F(x)(1 - F(x))] \\ &= \frac{F(x)(1 - F(x))}{n}. \end{aligned}$$



De aquí se sigue que $\lim_{n \rightarrow \infty} V[F_n(x)] = 0$, por lo que se puede concluir que $F_n(x)$ es un estimador consistente de $F(x)$.

$F_n(x)$ “converge” a una variable aleatoria normal con media $F(x)$ y varianza $F(x)(1 - F(x))$

Tomando en cuenta que $F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{[X_i \leq x]}$, aplicando el teorema central del límite para cada x fijo tenemos que

$$\sqrt{n}\{F_n(x) - F(x)\} \longrightarrow N(0, F(x)(1 - F(x))).$$

$F_n(x)$ converge a $F(x)$ uniformemente en x

Unos de los resultados más importantes que se cumple uniformemente en x , es el Teorema de Glivenko-Cantelli (1933) que algunos autores le llaman el teorema fundamental de la estadística.

Teorema 3 (Glivenko-Cantelli) Sea X_1, X_2, \dots, X_n variables aleatorias independientes con distribución común $F(x)$, y $F_n(x)$ la distribución empírica de la muestra. Sea $D_n = \sup_x |F_n(x) - F(x)|$, entonces

$$P \left[\lim_{n \rightarrow \infty} D_n = 0 \right] = 1.$$

El resultado está relacionado con la ley fuerte de los grandes números, y trata sobre la convergencia de la distribución empírica a la verdadera función de distribución.

2.2.3. Estadísticas de prueba EDF

Se les llama estadísticas de prueba EDF, a las que resulta ser una medida de la discrepancia entre $F_n(x)$ y $F_0(x)$ considerando una métrica ó seudométrica particular $\delta_n(F_n, F_0)$. Estas métricas miden la discrepancia entre la función de distribución empírica F_n y la función de distribución hipotética F_0 . Si se logra obtener la distribución de $\delta_n(F_n, F_0)$ bajo

H_0 , puede entonces evaluarse la probabilidad de que $\delta_n(F_n, F_0)$ tome un valor grande. El cómputo para obtener las distribuciones exactas de una métrica o pseudométrica suele ser difícil, lo cual hizo imprescindible el uso de las distribuciones asintóticas.

Estadísticas Kolmogorov y Kuiper

Una de las pruebas tradicionales de bondad de ajuste que utiliza a la EDF, es la de Kolmogorov, la cual se dio a conocer por el trabajo de Kolmogorov (1933). Para la prueba de hipótesis nula $H_0 : F = F_0$ vs $H_1 : F \neq F_0$, la estadística de prueba Kolmogorov está dada por:

$$D_n = \sup_x |F_n(x) - F_0(x)|.$$

Esta métrica calcula la diferencia absoluta más grande entre F_n y F_0 . La estadística también puede ser escrita como:

$$D_n = \max(D^+, D^-),$$

donde $D^+ = \sup_x \{F_n(x) - F_0(x)\}$ y $D^- = \sup_x \{F_0(x) - F_n(x)\}$.

El cálculo del supremo de una función no diferenciable típicamente requiere la evaluación de muchos puntos. Sin embargo, como F_n es una función escalonada, y como F_0 es una función monótona creciente, denotando a $x_{(i)}$, como ya se mencionó, como la estadística de orden i de la muestra, tenemos que la estadística D^+ equivale a

$$D^+ = \max_{1 \leq i \leq n} \left(\frac{i}{n} - z_{(i)} \right), \text{ donde } z_{(i)} = F_0(x_{(i)}).$$

De igual forma tenemos que

$$D^- = \max_{1 \leq i \leq n} \left(z_{(i)} - \frac{i-1}{n} \right), \text{ donde } z_{(i)} = F_0(x_{(i)}).$$

Otra estadística de prueba es la de Kuiper que se define como:

$$V = D^+ + D^-.$$

Estadísticas de la familia Cramér-von Mises

Hay otras dos estadísticas de prueba que se derivan de la familia de Cramér-von Mises definida por:

$$E = n \int_{-\infty}^{\infty} \{F_n(x) - F_0(x)\}^2 \Psi(x) dF_0(x),$$

en donde $\Psi(x)$ es una función que le da el peso o ponderación al cuadrado de las diferencias $\{F_n(x) - F_0(x)\}^2$ (ver D'Agostino y Stephens, 1986; Villers, 2011).

Las estadísticas de prueba que se derivan de dicha familia son:

Cramér-von Mises

$$W_n^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F_0(x)\}^2 dF_0(x).$$

Anderson-Darling

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{\{F_n(x) - F_0(x)\}^2}{\{F_0(x)\}\{1 - F_0(x)\}} dF_0(x).$$

Una versión modificada de la familia Cramér-von Mises es la estadística de prueba Watson, la cual es,

$$U_n^2 = n \int_{-\infty}^{\infty} \left\{ F_n(x) - F_0(x) - \int_{-\infty}^{\infty} \{F_n(x) - F_0(x)\} dF_0(x) \right\}^2 dF_0(x).$$

Como F_n es escalonada, y F_0 es monótona creciente, y denotando a las $z_{(i)}$'s como antes, aplicando cálculo tenemos que las tres estadísticas de prueba equivalen a (ver D'Agostino y Stephens, 1986; Villers, 2011):

- $W_n^2 = \sum_{i=1}^n \left\{ z_{(i)} - \frac{2i-1}{2n} \right\}^2 + \frac{1}{12n}.$
- $A_n^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i-1) \log z_{(i)} + (2n+1-2i) \log(1-z_{(i)})].$
- $U_n^2 = W^2 - n(\bar{z} - 0.5)^2,$ donde $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i.$

2.3. Proceso empírico

En general, es muy difícil obtener las distribución exactas de las estadísticas de prueba EDF, por lo cual se han utilizado las distribuciones asintóticas (ver O'Reilly, 1990).



En esta sección se define el **proceso empírico** en general, cuando F_0 es continua y totalmente conocida; dicho proceso converge al Browniano atado (ver sección 2.3.3). Cuando F_0 es continua y tiene parámetros desconocidos, se define el **proceso empírico con parámetros estimados**. Para el caso en que la familia F_0 sea una familia con parámetros de localización y/o escala y cumpla ciertas condiciones de regularidad (ver Cramér, 1945), y el estimador $\hat{\theta}_n$ de θ cumpla ciertas propiedades de eficiencia (ver Durbin, 1973), el proceso empírico con parámetros estimados converge a un proceso Gaussiano³ con media cero y cierta función de covarianza que no depende de los parámetros.

2.3.1. Proceso empírico

Se define el **proceso empírico** en general,

$$\xi_n(x) = \sqrt{n}\{F_n(x) - F_0(x; \theta)\}_{x \in \mathbb{R}},$$

en el caso en que tenemos θ conocido y F_0 continua, la ley límite del proceso empírico corresponde a la de un Browniano atado. Las estadísticas EDF se pueden poner en función de $\xi_n(x)$ como sigue:

$$\begin{aligned} D_n &= \sup_x |\xi_n(x)|, \\ W_n^2 &= \int_{-\infty}^{\infty} \xi_n^2(x) dF_0(x), \\ A_n^2 &= \int_{-\infty}^{\infty} \frac{\xi_n^2(x)}{F_0(x)[1 - F_0(x)]} dF_0(x), \\ U_n^2 &= \int_{-\infty}^{\infty} \left\{ \xi_n(x) - \int_{-\infty}^{\infty} \xi_n(x) dF_0(x) \right\}^2 dF_0(x). \end{aligned}$$

Una vez que se tiene el valor de la estadística de prueba EDF, lo siguiente es ver el valor en una tabla de la correspondiente distribución asintótica. La velocidad de convergencia de las distribuciones exactas (n finito) a la límite es muy buena, por ejemplo, para el caso de la estadística de prueba A_n^2 , se recomienda el uso de la distribución límite si $n \geq 5$. Para una referencia de las distribuciones tabuladas ver D'Agostino y Stephens (1986). Eligiendo un nivel de significancia α , el procedimiento para probar H_0 es el siguiente (ver O'Reilly, 1990):

³En la sección 2.3.3 se define el proceso Gaussiano.

1. Transforme cada x_1, \dots, x_n con F_0 , por lo que tenemos z_1, \dots, z_n después ordenamos $z_{(1)} \leq \dots \leq z_{(n)}$.
2. Sea d_{obs} el valor de la estadística EDF que haya elegido.
3. Vea el valor de la tabla correspondiente llamado d_α , cuando $d_{obs} > d_\alpha$ se rechaza H_0 .

2.3.2. Proceso empírico con parámetros estimados

Cuando F_0 es continua y tiene parámetros desconocidos, donde $\hat{\theta}_n$ es un estimador de θ , se define el proceso empírico con parámetros estimados como:

$$\hat{\xi}_n(x) = \sqrt{n}\{F_n(x) - F_0(x; \hat{\theta}_n)\}_{x \in \mathbb{R}},$$

cuando la familia de distribución F_0 es de localización y/o escala y cumple las propiedades de regularidad (ver Cramér, 1945), donde el estimador $\hat{\theta}_n$ de θ se obtiene por un método apropiado, que depende de la familia en particular a probar (ver Durbin, 1973), el proceso empírico con parámetros estimados no depende del valor particular del parámetro θ , cuya ley límite es un proceso Gaussiano con media cero y cierta función de covarianza (no depende del valor del parámetro).

Por lo general, la mayoría de las familias con parámetros de localización y/o escala (ver sección 2.3.4) cumplen con las condiciones de regularidad. Un contraejemplo de una familia de localización y/o escala que no cumple las condiciones de regularidad es la familia uniforme en $[0, \theta]$, ya que una de las condiciones es que el recorrido de la variable aleatoria no dependa del parámetro.

El procedimiento para probar H_0 es análogo que para el caso simple, usando las tablas que corresponden a la familia particular de localización y/o escala (ver D'Agostino y Stephens, 1986).

En las siguientes secciones se dan las definiciones necesarias para ver el proceso límite de un proceso empírico.



2.3.3. Procesos Gaussiano y Browniano atado

El proceso estocástico continuo $\{W_t\}_{t \geq 0}$, se dice que es un proceso Gaussiano si para cualesquiera $\{t_1, t_2, \dots, t_n\}$ se tiene que

$$(W(t_1), W(t_2), \dots, W(t_n)) \sim \mathcal{N}(\mu, \Sigma).$$

Una vez que se dice que $\{W_t\}_{t \geq 0}$ es un proceso Gaussiano basta con decir cuáles son sus momentos para identificar específicamente el proceso Gaussiano.

Sea $\mathcal{C} = \mathcal{C}[0, 1]$ el espacio de las funciones continuas en $[0, 1]$, donde la distancia entre dos puntos x e y se define como:

$$\rho(x, y) = \sup_{t \in [0, 1]} |x(t) - y(t)|.$$

El proceso estocástico $\{B_t\}_{t \geq 0}$, se dice ser un Browniano, si es un proceso Gaussiano con trayectorias en $\mathcal{C}[0, 1]$, con $E[B_t] = 0$ & $Cov(B_s, B_t) = \min\{s, t\}$.

El proceso estocástico $\{B_t^0\}_{t \geq 0}$ se dice que es un Browniano atado, si

$$B_t^0 = B_t - tB_1, \text{ para } t \in [0, 1].$$

Obsérvese que $\{B_t^0\}_{t \geq 0}$ es un proceso gaussiano con trayectorias en $\mathcal{C}[0, 1]$, tal que:

$$E[B_t^0] = 0 \quad \forall t \quad \& \quad Cov(B_s^0, B_t^0) = s(1-t) \quad \forall s \leq t.$$

2.3.4. Familias de localización y/o escala

Sea Z una variable aleatoria con distribución conocida. A la colección de distribuciones de probabilidad de las variables aleatorias X que se pueden definir de la forma

$$X = \mu + \sigma Z \quad \mu \in R \text{ y } \sigma > 0,$$

se la denomina familia de localización y escala construida a partir de la distribución de Z (ver Casella y Berger, 1990).



En particular, si Z es variable aleatoria absolutamente continua con función de densidad $f_*(\cdot)$, entonces la familia de funciones de densidad

$$\left\{ f(x; \mu, \sigma) = \frac{1}{\sigma} f_*\left(\frac{x - \mu}{\sigma}\right) : \mu \in \mathbb{R}, \sigma > 0 \right\},$$

forman la familia de localización y escala de $f_*(\cdot)$. El parámetro de escala dilata la distribución si $\sigma > 1$ y la contrae si $\sigma < 1$. El parámetro de localización μ traslada la densidad $|\mu|$ unidades a la derecha si $\mu > 0$ o traslada la densidad $|\mu|$ unidades a la izquierda si $\mu < 0$.

Algunos ejemplos de familias de distribución de localización y/o escala, o que están relacionadas, son las funciones de distribución: normal, lognormal, exponencial, Pareto, logística, Cauchy, Weibull con parámetro de forma conocido, valor extremo máximo y valor extremo mínimo (ver Villers, 2011).

2.4. Simulación del p-valor (o p-value) para el caso simple

Se tiene que la hipótesis a probar es:

$$H_0 : F(x) = F_0(x) \quad \text{vs.}$$

$$H_1 : F(x) \neq F_0(x).$$

El p-valor es la probabilidad, calculada al asumir que H_0 es cierta, de que la estadística de prueba tome valores tan extremos o más extremos que los calculados con la muestra. El p-valor no se fija a priori, sino que se determina a partir de la muestra. Sea D la estadística de prueba EDF, y d_{obs} el valor de la estadística de prueba calculado bajo la muestra, se define el p-valor como (ver D'Agostino y Stephens, 1986):

$$p := P[D > d_{obs} : \text{Dado } H_0 \text{ es cierta}].$$

Para calcular dicha probabilidad se pregunta, ¿cuál es la distribución de D bajo H_0 ?, y como se comentó anteriormente es en general difícil saber cuál es la distribución exacta,

por lo que en el caso simple (F_0 totalmente conocida), podemos simular un gran número de muestras (bajo H_0) y con ellas calcular el valor que toma D .

El procedimiento para generar el p-valor es el siguiente: sea d_{obs} el valor de la estadística de prueba EDF calculado con la muestra; se genera un número grande de muestras de la distribución F_0 , digamos M muestras; para cada muestra se calcula la estadística de prueba EDF. Tendremos M observaciones simuladas de la estadística de prueba EDF; d_1, \dots, d_M . El p-valor (aproximado), es la proporción de valores d_i 's que exceden a d_{obs} (ver D'Agostino y Stephens, 1986), es decir

$$p \approx \frac{\#\{j | d_j > d_{obs}\}}{M},$$

entre más grande sea M , mejor será la aproximación del p-valor. Si hubiéramos elegido un nivel de significancia (nivel de rechazo) α , el criterio de rechazo sería el siguiente:

- si $p < \alpha$, se rechaza H_0 ;
- si $p \geq \alpha$, no se rechaza H_0 .

2.5. Simulación del p-valor para familias de localización y/o escala

Se tiene que el contraste a probar es:

$$H_0 : F(x) = F_0(x; \theta) \quad vs.$$

$$H_1 : F(x) \neq F_0(x; \theta).$$

Como ya se comentó anteriormente si F_0 es una familia con parámetros de localización y/o escala, donde θ se estima mediante un procedimiento adecuado (el cual depende de la familia a probar), el proceso empírico con parámetros estimados (n finito) no depende del valor del parámetro desconocido, por lo que fijando un valor para θ arbitrario, podemos encontrar un p-valor exacto por simulación (ver O'Reilly y Gracia-Medrano, 2010).

El procedimiento para calcular el p-valor es análogo al caso simple sólo que ahora se sustituye el parámetro θ por un valor arbitrario fijo (uno podría poner el valor del estimador).

En la tesis Villers (2011), se da el procedimiento adecuado para estimar θ para las siguientes familias: normal, lognormal, exponencial, Pareto, logística, Cauchy, Weibull, valor extremo máximo y valor extremo mínimo. Donde las familias anteriores son de localización y/o escala o tienen alguna relación con alguna familia de localización y/o escala. Por ejemplo, la familia Weibull se considera con parámetro de forma conocido. En la tesis mencionada, se agrega un programa en Matlab con dichas distribuciones para el cálculo de las estadísticas de prueba y la determinación de sus correspondientes p-valores (ver Villers, 2011).

2.6. Ejemplo

Una muestra aleatoria de tamaño 10 es obtenida: $x_1 = 0.621, x_2 = 0.503, x_3 = 0.203, x_4 = 0.477, x_5 = 0.710, x_6 = 0.581, x_7 = 0.329, x_8 = 0.480, x_9 = 0.554$ y $x_{10} = 0.382$ (ver Conover, 1999). Se quiere probar la siguiente hipótesis

$$H_0 : F(x) = F_0(x) \quad \text{vs.}$$

$$H_1 : F(x) \neq F_0(x),$$

donde la función de distribución de la hipótesis nula es la función de distribución uniforme en $(0, 1)$. La función de distribución de la hipótesis nula es:

$$F_0(x) = \begin{cases} 0 & \text{si } x < 0, \\ x & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

Utilizando la estadística de prueba de Kolmogorov, la cual es:

$$D_n = \max(D^+, D^-),$$

donde

$$D^+ = \max_{1 \leq i \leq n} \left(\frac{i}{n} - z_{(i)} \right),$$



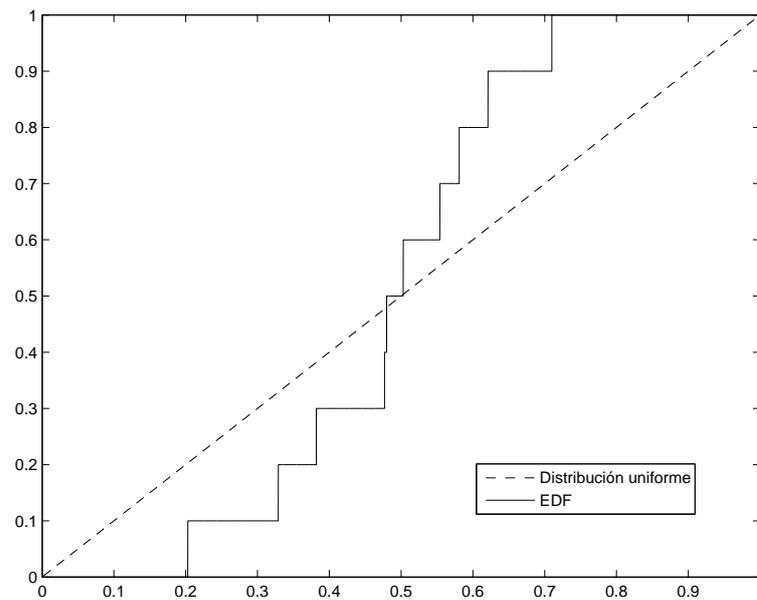


Figura 2.1: EDF y la función de distribución uniforme.

y

$$D^- = \max_{1 \leq i \leq n} \left(z_{(i)} - \frac{i-1}{n} \right).$$

La gráfica 2.1 muestra la función uniforme y la función de distribución empírica.

La distancia vertical más grande de las dos gráficas en la figura 2.1 es 0.290, que ocurre cuando $x = 0.710$ ya que $F_n(0.710) = 1.000$ y $F_0(0.710) = 0.710$. En otras palabras,

$$\begin{aligned} D_n &= \sup_x |F_n(x) - F_0(x)| \\ &= |F_n(0.710) - F_0(0.710)| \\ &= 0.290. \end{aligned}$$

Utilizando la estadística de prueba Anderson-Darling

$$\begin{aligned} A_n^2 &= -n - \frac{1}{n} \sum_{i=1}^n [(2i-1) \log z_{(i)} + (2n+1-2i) \log(1-z_{(i)})] \\ &= 1.3892. \end{aligned}$$

De igual forma, calculando los otras 3 estadísticas de prueba EDF tenemos:

Anderson Darling	Cramér-vonMises	Kolmogorov	Kuiper	Watson
1.3892	0.2483	0.2900	0.5190	0.2457

Con 10000 muestras simuladas (sección 2.4), los 5 p-valores correspondientes a las 5 estadísticas de prueba son ⁴:

Anderson Darling	Cramér-vonMises	Kolmogorov	Kuiper	Watson
0.2020	0.1855	0.3033	0.0409	0.0095

De la tabla anterior podemos observar que el p-valor simulado de las estadísticas Kuiper y Watson es menor a 0.05, por lo que concluiríamos si hubiéramos elegido una de estas estadísticas, es que existe evidencia para rechazar H_0 con un nivel de significancia 0.05. Sin embargo, de haber elegido las otras 3 estadísticas de prueba concluiríamos que no podemos rechazar H_0 a menos que eligiéramos un nivel de significancia grande (mayor a 0.1855).

⁴Los cálculos de las estadísticas de prueba y sus correspondientes p-valores fueron programadas en el software MATLAB R2010a. El tiempo promedio en hacer estos cálculos es de aproximadamente 740 milisegundos.

Capítulo 3

Uso del estimador Rao-Blackwell para la generación de muestras-*

Para enfrentar el problema de bondad de ajuste de una distribución en presencia de parámetros desconocidos (caso compuesto), tradicionalmente se estudiaban las distribuciones asintóticas de las estadísticas usadas en la bondad de ajuste (funciones del proceso empírico con parámetros estimados). Eso requirió de un esfuerzo grande para identificar el comportamiento asintótico del referido proceso. En ocasiones, ese comportamiento asintótico dependía adicionalmente de los valores de los parámetros desconocidos con lo que, en el mejor de los casos, el uso de la distribución asintótica además de ser ya una aproximación (pues el tamaño n de la muestra es finito), incorporaba una aproximación adicional por usar un estimador en lugar del parámetro desconocido, para poder usar las tablas.

Para las distribuciones cuyos parámetros no se pueden considerar de localización y/o escala, existen diversos procedimientos que en general sólo son aproximaciones (ver Villers, 2011). En particular en O'Reilly y Gracia-Medrano (2006) se proporciona un método para encontrar una solución exacta cuando existe la función de distribución Rao-Blackwell y en González-Barrios et al. (2010), se analiza la exponencial truncada.

En las siguientes secciones se da la definición de la función de distribución Rao-Blackwell. Con el uso de ella, se proporciona el procedimiento para la generación de las llamadas muestras condicionales o muestras look-alike (como en O'Reilly y Gracia-Medrano, 2006), así como el cálculo de los p-valores condicionales. Por último se revisan los ejemplos

de bondad de ajuste para la distribución Gaussiana inversa y la distribución exponencial truncada.

3.1. Estimador Rao-Blackwell

Como ya se comentó anteriormente, cuando el parámetro de la distribución no es de localización y/o escala, el límite del proceso empírico con parámetros estimados, en general depende del valor de los parámetros, por lo que hay que utilizar métodos alternativos.

Una pregunta muy natural es, ¿Que se obtiene si en lugar de estimar $F_0(\cdot; \theta)$ con $F_0(\cdot; \hat{\theta}_n)$, se utiliza otro tipo de estimador?. Existe una teoría que proporciona un estimador insesgado y de varianza mínima, que se deduce a partir de la teoría de Rao-Blackwell, a dicho estimador se le conoce con el mismo nombre y se enuncia enseguida.

Sea X_1, \dots, X_n una muestra aleatoria de la función de distribución

$$F_0(\cdot; \theta); \theta \in \Theta_0 \quad \text{y sea } T_n = T_n(X_1, \dots, X_n),$$

una estadística suficiente minimal y completo, entonces el estimador Rao-Blackwell de $F_0(x; \theta)$ es (ver O'Reilly, 1990):

$$\tilde{F}_n(x) := P[X_1 \leq x \mid T_n].$$

A esta función también se le llama distribución Rao-Blackwell. En muestreo independiente T_n es simétrica, por lo que las X_i dada T_n son intercambiables,

$$\tilde{F}_n(x) = P[X_i \leq x \mid T_n],$$

con $i = 1, \dots, n$.

Bajo condiciones muy generales, el proceso $\{\tilde{F}_n(x)\}_{n \geq 1}$ resulta ser una martingala al revés (reverse martingale) respecto a la sucesión de σ álgebras (ver Berk, 1969; Bhattacharyya y Sen, 1977),

$$\{\sigma(T(X_1, \dots, X_n), X_{n+1}, X_{n+2}, \dots)\}_{n \geq 1}.$$

Por lo que el proceso

$$\sqrt{n}\{\tilde{F}_n(x) - F_0(x; \theta)\},$$

puede demostrarse que converge a un proceso Gaussiano de media cero y función de covarianza $h(x, y) = \lim_{n \rightarrow \infty} nE[\{\tilde{F}_n(x) - F_0(x; \theta)\}\{\tilde{F}_n(y) - F_0(y; \theta)\}]$ (ver Bhattacharyya y Sen, 1977; Brown, 1990).

Se define el proceso empírico tipo Rao-Blackwell como,

$$\sqrt{n}\{F_n(x) - \tilde{F}_n(x)\},$$

donde $\tilde{F}_n(x)$ es utilizado en lugar de $F_0(x; \hat{\theta}_n)$. Moore (1973), demostró que para los casos de la distribución normal y exponencial, el proceso empírico tipo Rao-Blackwell y el proceso empírico con parámetros estimados (donde $\hat{\theta}_n$ es el estimador de máxima verosimilitud de θ) convergen al mismo límite, y conjeturó que el resultado es bastante general. A eso se le ha llamado durante muchos años la conjetura de Moore¹. En Lockhart y O'Reilly (2005) se demostró, que si la familia de distribución F_0 pertenece a la familia exponencial natural, se cumple la conjetura Moore (ver Lockhart y O'Reilly, 2005).

El estimador Rao-Blackwell puede ser usado para generar muestras, las cuales tendrían la misma distribución condicional que la muestra inicial (si la hipótesis nula es verdadera), estas muestras simuladas son las "look-alike" o "muestras-*" (ver O'Reilly y Gracia-Medrano, 2006).

3.2. Generación de muestras Look-Alike

En esta sección se revisa el algoritmo para generar las llamadas muestras "look-alike" o bien "*-samples", que referiremos como "muestras-*". Dichas muestras se generan a partir de la distribución Rao-Blackwell.

Existen importantes familias que tienen parámetros que no son de localización y/o escala, por ejemplo la distribución Gamma (tiene parámetro de forma), la distribución exponencial truncada, la distribución Gaussiana inversa, etc. En estas familias la distribución límite de la estadística de prueba (que es función del proceso empírico estimado)

¹Conjetura de Moore: $n^{1-\delta} \sup_x |F_0(x; \hat{\theta}_n) - \tilde{F}_n(x)| \rightarrow 0$ en probabilidad para cualquier $\delta > 0$ fijo.

depende del valor de los parámetros, en estos casos se enfrenta el problema adicional de tener que estimar θ para poder usar tablas; y además hay que analizar el efecto de haber estimado, o sea analizar la variación en el estimador, por lo cual, hay que utilizar un método alternativo, y el uso del estimador Rao-Blackwell da un procedimiento exacto para probar bondad de ajuste.

Ahora la simulación no se hace sobre los valores de la estadística de prueba (que depende de los parámetros desconocidos) sino que se hace con la distribución condicional dada una estadística suficiente T_n . Con este procedimiento, se evalúa el p-valor (condicional), tan exacto como el número de simulaciones de muestras condicionales.

La estadística T_n cumple la propiedad de doblemente transitividad, si el par (T_n, X_n) es una correspondencia uno a uno con el par (T_{n-1}, X_n) . Lo cual significa que el conocimiento de T_n y X_n es equivalente al conocimiento de T_{n-1} y X_n . Este es el caso de las familias que pertenecen a la familia exponencial natural donde T_n es la suma de n elementos independientes e idénticamente distribuidos, es decir, $T_n = \sum_{i=1}^n X_i$. Por lo que en este caso, el conocimiento de $T_n = \sum_{i=1}^n X_i$ y X_n es equivalente al conocimiento de $T_{n-1} = \sum_{i=1}^{n-1} X_i$ y X_n . En la mayoría de las aplicaciones se tiene que la estadística T_n cumple la propiedad de doble transitividad.

Si la estadística T_n es doblemente transitiva y se quiere simular la muestra x_1^*, \dots, x_n^* de la distribución condicional de X_1, \dots, X_n dada $T_n = t_n$ (o lo que es lo mismo, dada T_n , en t_n), sólo se necesita la expresión $\tilde{F}_n(x)$.

Sea k el número de parámetros desconocidos de la familia paramétrica $F_0(\cdot; \theta)$, el siguiente resultado dice cómo simular las primeras $n - k$ componentes de la muestra-*, y para obtener los restantes k términos se resuelve un sistema de k incógnitas.

Teorema 4 (O'Reilly y Gracia-Medrano, 2006) *Bajo doble transitividad, teniendo $n - k$ como el número máximo de términos de la muestra, para la distribución condicional de esos términos dada $T_n = t_n$, el procedimiento para obtener (por ejemplo los superiores) $n - k$ términos de la x^* -muestra (look-alike) que tengan exactamente la misma distribución condicional (como cualesquiera $n - k$ términos de la muestra), es la siguiente:*

Dada $\tilde{F}_n(x)$, el estimador Rao-Blackwell basado en $t_n = T_n(x_1, \dots, x_n)$, se denota por $\tilde{F}_n^{-1}(u)$ su inversa.

Paso 1. *Sea u_n una realización de una variable aleatoria $U(0, 1)$, se define $x_n^* = \tilde{F}_n^{-1}(u_n)$, y usando la doble transitividad, se calcula t_{n-1} de t_n y x_n^* , y se denota este nuevo valor por t_{n-1}^* .*

Paso 2. *Ahora con \tilde{F}_{n-1} , el estimador Rao-Blackwell basado en t_{n-1}^* y la obvia notación para su inversa, sea u_{n-1} otra realización de una variable aleatoria $U(0, 1)$ independiente de u_n ; se define el número $x_{n-1}^* = \tilde{F}_{n-1}^{-1}(u_{n-1})$ y del conocimiento de t_{n-1}^* y x_{n-1}^* , se calcula t_{n-2}^* ; se sigue el proceso hasta que:*

Paso $(n - k)$. *Usando \tilde{F}_{k+1}^{-1} generar x_{k+1}^* con u_{k+1} , seleccionando a u_{k+1} como una realización de una variable aleatoria $U(0, 1)$, independiente de las anteriores u_j 's.*

Estos $n - k$ términos de la muestra-* son una realización de la distribución condicional para cualesquiera $n - k$ términos de la muestra dada t_n .

Finalmente, se encuentran los restantes k términos x_1^*, \dots, x_k^* que pueden ser cualquier permutación de la única solución (ordenada) generada por la condición que $T_n(x_1^*, \dots, x_n^*)$ es igual a t_n . A la muestra obtenida x_1^*, \dots, x_n^* se le llamará muestra look-alike ó muestra-*.

3.2.1. Simulación del p-valor condicional

En varios trabajos ya se ha propuesto el uso de la distribución condicional dada una estadística suficiente para probar bondad de ajuste. A continuación mencionamos algunas

de estas técnicas.

En el artículo de Lindqvist y Taraldsen, (2005) realizan Monte Carlo mediante la esperanza condicional dada una estadística suficiente. Por este método se expresa la esperanza condicional en términos de una esperanza (incondicional), la cual puede ser calculada mediante simulación Monte Carlo.

En el artículo de Lockhart et al., (2007) dan un método para probar bondad de ajuste si la muestra aleatoria admite una estadística suficiente. Proponen mediante muestreo de Gibbs generar muestras condicionales, a estas muestras le llaman muestras co-sufficient, las cuales se generan mediante la distribución condicional de la muestra dada el valor de la estadística suficiente. El método se ilustra con un ejemplo para la distribución gamma; también comparan su método contra el bootstrap paramétrico para probar bondad de ajuste para la distribución gamma, haciendo simulaciones del tamaño y potencia para varias estadísticas de prueba.

En el artículo de González-Barrios et al., (2006) proporcionan un método exacto para probar bondad de ajuste para el caso discreto, utilizando la probabilidad condicional dada una estadística suficiente.

En el trabajo de O'Reilly y Gracia-Medrano, (2006) mediante la simulación de las llamadas muestras look-alike (sección anterior), produce un método exacto para el cálculo del p-valor condicional. El procedimiento para el cálculo del p-valor condicional es análogo al cálculo del p-valor (incondicional) que se vio en las secciones 2.4 y 2.5. A continuación damos el procedimiento para el cálculo del p-valor condicional.

Sea d_{obs} el valor de la estadística de prueba calculado de la muestra; se genera un número grande de muestras "look-alike"; digamos M muestras, y para cada muestra, se calcula la estadística de prueba, por lo cual se tendrán M observaciones (simuladas) de la estadística de prueba; d_1, \dots, d_M . El p-valor condicional, denotado por p , se calcula (aproximadamente), como la proporción de valores d_i 's que exceden a d_{obs} (ver O'Reilly y Gracia-Medrano, 2006), es decir

$$p \approx \frac{\#\{j | d_j > d_{obs}\}}{M},$$

Si se quiere actuar como en la teoría de Neyman Pearson, se elige un nivel de signifi-

cancia α (probabilidad de error tipo I), y el criterio de rechazo sería el siguiente:

- cuando $p < \alpha$, se rechaza H_0 ;
- cuando $p \geq \alpha$, no se rechaza H_0 .

3.3. Distribución exponencial truncada

Sea Z una variable aleatoria con distribución exponencial con parámetro $\theta > 0$, se define la distribución exponencial truncada (en x_0), como la distribución condicional de Z dado $Z \leq x_0$, es decir, la densidad condicional de Z es (ver Deemer y Votaw, 1955),

$$f(z; \theta) = \frac{\theta e^{-\theta z}}{1 - e^{-\theta x_0}}, \quad 0 < z \leq x_0, \quad x_0 \text{ conocido}, \quad \theta > 0, \quad (3.1)$$

y cero para cualquier otro valor de z .

La función de distribución condicional de Z es,

$$F(z; \theta) = \frac{e^{-\theta z} - 1}{e^{-\theta x_0} - 1}.$$

En Lindqvist et al. (2003), se da la familia exponencial truncada en $x_0 = 1$ y mencionan que el valor de θ puede estar en todos los reales y cuando $\theta = 0$ se convierte en la distribución uniforme en $(0, 1)$.

La familia exponencial truncada en x_0 (ecuación 3.1), de forma análoga que en Lindqvist et al. (2003), el parámetro θ puede ser extendido a todos los reales y cuando $\theta = 0$ se convierte en la distribución uniforme en $(0, x_0)$.

Sea x_1, x_2, \dots, x_n una muestra aleatoria de la distribución exponencial truncada; en este caso $t_n = \sum_{i=1}^n x_i$ es suficiente minimal.

En el artículo de O'Reilly y Rueda, (2007) se menciona la distribución exponencial truncada en $x_0 = 1$ como una familia que propusieron en Lindqvist et al. (2003) como un contraejemplo a un resultado de simulación condicional de Engen y M. Lillegard, (1997).

Ahora sabemos que esa familia apareció en Deemer y Votaw desde 1955 y en González Barrios et al. (2010) se obtiene el estimador Rao-Blackwell, desconociendo que en Sathe

y Varde, (1969) ya lo habían obtenido. Se aclara que los métodos para obtenerlo fueron diferentes.

Shate y Varde, (1969) se da la distribución Rao-Blackwell basada en t_n para la exponencial truncada en x_0 , la cual es:

$$\tilde{F}_n(x|t_n) = \frac{\sum_{j=0}^{\lfloor t_n/x_0 \rfloor} (-1)^j \binom{n-1}{j} \left[1 - \frac{jx_0}{t_n}\right]^{n-1} - \sum_{j=0}^{\lfloor (t_n-x)/x_0 \rfloor} (-1)^j \binom{n-1}{j} \left[1 - \frac{jx_0+x}{t_n}\right]^{n-1}}{\sum_{j=0}^{\lfloor t_n/x_0 \rfloor} (-1)^j \binom{n}{j} \left[1 - \frac{jx_0}{t_n}\right]^{n-1}}, \quad (3.2)$$

observe que esta función está definida para $0 \leq x \leq t_n$ y $n \geq 1$.

En el artículo de Gonzalez-Barrios et al. (2010) se proporciona el estimador Rao-Blackwell basada en t_n para la exponencial truncada en $x_0 = 1$, la cual es:

$$\tilde{F}_n(x|t_n) = (n-1) \frac{\sum_{j=0}^{n-2} \binom{n-1}{j} (-1)^j \int_0^x (t-u-j)_+^{n-2} du}{\sum_{j=0}^{n-1} \binom{n}{j} (-1)^j (t-j)_+^{n-1} du}, \quad (3.3)$$

la expresión anterior en el artículo mencionado tiene un error tipográfico, ya que en la expresión en el denominador el coeficiente binomial tiene $n-1$ en lugar de n . Observé que la función 3.3 está definida para $n \geq 2$.

Con un cálculo sencillo se puede verificar que las expresiones (3.2) y (3.3) son iguales con $x_0 = 1$ y $n \geq 2$.

Una vez que se tiene la distribución Rao-Blackwell, se procede como en el teorema 4; es decir, se generan primero los $n-1$ elementos de la muestra; x_2^*, \dots, x_n^* , y por último, se obtiene $x_1^* = t_1^*$ (es decir igual a $t_2^* - x_2^*$). La muestra-* que se acaba de obtener es una muestra condicional dada t_n .

3.3.1. Ejemplo

Consideremos el ejemplo de González et al. (2010). Se tiene la distribución exponencial truncada en 1, es decir

$$F_0(x) = \frac{e^{-\theta x} - 1}{e^{-\theta} - 1}, \quad 0 \leq x \leq 1.$$

En González-Barrios et al., (2010) utilizan los siguientes datos

$$0.1600, 0.1720, 0.2845, 0.9630, 0.8744$$

para probar bondad de ajuste para una distribución exponencial truncada en $x_0 = 1$. Si se quiere ver si los datos anteriores se ajustan o no a la distribución exponencial truncada en $x_0 = 1$, se tiene que el contraste a probar es:

$$H_0 : F(x) = F_0(x; \theta) \quad vs.$$

$$H_1 : F(x) \neq F_0(x; \theta).$$

Con estos datos se obtiene la estimación $\hat{\theta} = -0.11066$ (ver González-Barrios et al., 2010). Calculando las 5 estadísticas de prueba EDF (usuales) tenemos:

Anderson Darling	Cramér-vonMises	Kolmogorov	Kuiper	Watson
0.7496	0.1141	0.3041	0.5845	0.1140

Con 10000 muestras-* como se propone en O'Reilly y Gracia-Medrano (2006), los 5 p-valores correspondientes a las 5 estadísticas de prueba son ²:

Anderson Darling	Cramér-vonMises	Kolmogorov	Kuiper	Watson
0.1373	0.1208	0.1905	0.1395	0.1208

²Los cálculos de las estadísticas de prueba y sus correspondientes p-valores fueron programadas en el software MATLAB R2010a. El tiempo promedio en hacer estos cálculos es de aproximadamente 40 segundos.

El p-valor más pequeño de estas posibles estadísticas es de 0.1208, por lo que no se podría rechazar H_0 usando esas estadísticas a menos que eligiéramos un valor de significancia grande (mayor a 0.1208). Se concluye que los datos se pueden modelar adecuadamente por medio de la función de distribución exponencial truncada en 1.

3.4. Distribución Gaussiana inversa

Se dice que una variable aleatoria X tiene una distribución Gaussiana inversa con parámetros $\mu > 0$ y $\lambda > 0$, denotado por $X \sim I(\mu, \lambda)$ (ver Chhikara y Leroy, 1974), si la función de densidad es

$$f(x; \mu, \lambda) = \begin{cases} [\lambda/2\pi x^3]^{\frac{1}{2}} \exp[-\lambda(x - \mu)^2/2\mu^2 x], & \text{si } x > 0, \\ 0, & \text{si } x \leq 0. \end{cases}$$

La función de distribución de la Gaussiana inversa puede ser escrita como:

$$F(x; \mu, \lambda) = \Phi(R) + \Phi(L) \exp\left\{\frac{2\lambda}{\mu}\right\}, \quad \text{con } x > 0,$$

donde $R = R(x; \mu, \lambda) = -(\frac{\lambda}{x})^{\frac{1}{2}} + (\frac{\lambda x}{\mu^2})^{\frac{1}{2}}$, $L = L(x; \mu, \lambda) = -(\frac{\lambda}{x})^{\frac{1}{2}} - (\frac{\lambda x}{\mu^2})^{\frac{1}{2}}$ y Φ es la función de distribución normal estándar. El correspondiente estimador Rao-Blackwell, utilizado en la literatura es (ver Chhikara y Leroy, 1974):

$$\tilde{F}_n(x) = \begin{cases} 0 & \text{si } x \leq l, \\ 1 & \text{si } x \geq u, \\ G_{n-2}(W) + \frac{n-2}{n} \left[1 + \frac{4(n-1)\hat{\lambda}}{n^2\hat{\mu}}\right]^{\frac{(n-3)}{2}} G_{n-2}(-W') & \text{en otro caso,} \end{cases}$$

con $\hat{\mu}$ y $\hat{\lambda}$ los estimadores de máxima verosimilitud, es decir,

$$\hat{\mu} = \bar{X} \quad \text{y} \quad \hat{\lambda} = \frac{n}{\sum_{i=1}^n \left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right)},$$

donde G_{n-2} es la función de distribución t Student con $n - 2$ grados de libertad, y

$$W = \frac{1}{C} \sqrt{n(n-2)} \left(\frac{x}{\hat{\mu}} - 1\right), \quad W' = \frac{1}{C} \sqrt{n(n-2)} \left(1 + \frac{n-2}{n} \frac{x}{\hat{\mu}}\right),$$

$$y \quad C = \sqrt{\frac{n}{\hat{\lambda}} \left(n - \frac{x}{\hat{\mu}} \right) x - n \left(1 - \frac{x}{\hat{\mu}} \right)^2}.$$

Y por último,

$$l = \frac{\bar{X}}{2(n + V\bar{X})} \left[n(2 + V\bar{X}) - \sqrt{4n(n-1)V\bar{X} + n^2V^2\bar{X}^2} \right],$$

$$u = \frac{\bar{X}}{2(n + V\bar{X})} \left[n(2 + V\bar{X}) + \sqrt{4n(n-1)V\bar{X} + n^2V^2\bar{X}^2} \right],$$

$$y \quad V = \sum_{i=1}^n \left(\frac{1}{X_i} - \frac{1}{\bar{X}} \right).$$

Se tiene que $\hat{\mu}$ y $\hat{\lambda}$ son funciones de la estadística suficiente minimal

$$T_n = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n \frac{1}{x_i} \right).$$

3.4.1. Ejemplo

En O'Reilly y Rueda (1992) utilizando tablas asintóticas para las estadísticas EDF, se analiza el siguiente conjunto de datos de la vida útil (días) de cierto producto alimenticio:

24, 24, 26, 26, 32, 32, 33, 33, 33, 35, 41, 42, 43, 47, 48, 48, 48, 50, 52, 54, 55, 57, 57, 57, 57, 61

Si se quiere probar si los datos anteriores se ajustan o no a la distribución Gaussiana inversa, tenemos que el contraste a probar es el siguiente:

$$H_0 : X \sim I(\mu, \lambda) \text{ vs.}$$

$$H_1 : X \not\sim I(\mu, \lambda). \quad (3.4)$$

Los estimadores de máxima verosimilitud son $\hat{\mu} = 42.8846$ y $\hat{\lambda} = 484.2519$, calculando las 5 estadísticas de prueba EDF resultan en:

Anderson Darling	Cramér-vonMises	Kolmogorov	Kuiper	Watson
0.9110	0.1485	0.1763	0.2948	0.1388

Con 10000 muestras-*, los 5 p-valores correspondientes a las 5 estadísticas de prueba son³:

Anderson Darling	Cramér-vonMises	Kolmogorov	Kuiper	Watson
0.0229	0.0289	0.0426	0.0307	0.0255

El p-valor más grande es 0.0426, por lo que se puede rechazar H_0 con nivel de significancia 0.05 si usamos cualquiera de las 5 estadísticas de prueba anteriores. Por lo que concluimos que los datos no se pueden modelar adecuadamente mediante la distribución Gaussiana inversa.

³Los cálculos de las estadísticas de prueba y sus correspondientes p-valores fueron programadas en el software MATLAB R2010a. El tiempo promedio en hacer estos cálculos es de aproximadamente 434 segundos.

Capítulo 4

Pruebas de bondad de ajuste para distribuciones discretas

Sea X_1, \dots, X_n una muestra aleatoria de una distribución discreta F , supongamos sin pérdida de generalidad que los valores que toma la variable aleatoria son $0, 1, 2, \dots$; y se quiere probar $H_0 : F \in \mathcal{F}_0$ vs. $H_1 : F \notin \mathcal{F}_0$, con \mathcal{F}_0 alguna familia de distribuciones discretas con parámetro θ (conocido o desconocido) en $0, 1, 2, \dots$; como se mencionó anteriormente, cuando θ es conocido se dirá caso simple y cuando θ es desconocido se dirá compuesto o con parámetros desconocidos.

En las siguientes secciones se dan algunas estadísticas de prueba que se encuentran en la literatura. En la sección 4.1 se da la definición del proceso empírico asociado a la función generadora de probabilidades, el cual es un símil al proceso empírico del caso continuo. Posteriormente se da la definición de la estadística de prueba d_n que es función del proceso empírico. También, se da una propuesta de estadística de prueba, que es una extensión de la estadística d_n . En la sección 4.2 se dan las estadísticas de prueba Cramér-von Mises para probar bondad de ajuste en el caso simple, y se definen las estadísticas de prueba Cramér-von Mises más generales que se utilizan tanto para probar bondad de ajuste en el caso simple como para el caso compuesto. En la sección 4.3 se menciona la estadística de prueba de razón de verosimilitudes generalizadas y se da la estadística de prueba Kullback-Leibler, que es equivalente a la estadística de prueba de razón de verosimilitudes generalizadas. En la sección 4.4 se menciona el método de bondad de ajuste de González-Barrios et al., (2006), en el cual, se utiliza la probabilidad condicional. En la sección 4.5 se desarrolla la

probabilidad condicional de la distribución binomial negativa generalizada. En la sección 4.6 se menciona el procedimiento la generación de las muestras look-alike y se revisa la distribución Rao-Blackwell para las distribuciones de series de potencia, así como para sus casos particulares; los cuales son, la binomial, la binomial negativa y la Poisson. Se obtiene la distribución Rao-Blackwell para la distribución binomial negativa generalizada, la cual, no pertenece a la familia de series de potencia. Por último en la sección 4.7 se dan algunas aplicaciones.

4.1. Proceso empírico asociado a la función generadora de probabilidades

Para las pruebas de bondad de ajuste discreto, en Rueda et al., (1991) se ha utilizado la función generatriz de probabilidades y su contraparte empírica, contrastándolas con una medida de discrepancia entre la evidencia muestral y la hipótesis nula.

Se define la función generadora de probabilidades de la variable aleatoria $X \sim F(\cdot; \theta)$,

$$\phi_{\theta}(s) = E(s^X; \theta), \text{ para } |s| \leq 1, \text{ con } s \in \mathbb{R}.$$

Como X toma valores en $0, 1, 2, \dots$; se tiene que

$$\phi_{\theta}(s) = \frac{1}{n} \sum_{j=0}^{\infty} s^j E_j(\theta),$$

donde $E_j(\theta)$ es el número esperado de la observación sea igual a j , es decir, $E_j(\theta) = nP(X = j; \theta)$.

Se define la función generadora empírica,

$$\phi_n(s) = \frac{1}{n} \sum_{j=1}^n s^{x_j}, \text{ para } |s| \leq 1, \text{ con } s \in \mathbb{R},$$

donde x_1, \dots, x_n es la muestra observada. Cuando θ es conocido (caso simple) se define el **proceso empírico** asociado a la función generadora de probabilidades:

$$\eta_{n,\theta}(s) = \sqrt{n}\{\phi_n(s) - \phi_{\theta}(s)\}.$$

Marquez y Pérez-Abreu (1988) encontraron que bajo ciertas condiciones, la ley límite de dicho proceso es un Gaussiano de media cero y función de covarianza $S_\theta(s, t) = \phi_\theta(st) - \phi_\theta(s)\phi_\theta(t)$.

En el caso de parámetros desconocidos (caso compuesto), se sustituye el parámetro θ por su estimador de máxima verosimilitud $\hat{\theta}$. Se define el **proceso empírico estimado** asociado a la función generadora de probabilidad como:

$$\eta_{n,\hat{\theta}}(s) = \sqrt{n}\{\phi_n(s) - \phi_{\hat{\theta}}(s)\},$$

que bajo ciertas condiciones de regularidad converge a un Gaussiano con media cero y cierta función de covarianza (ver Rueda et al. 1991¹).

4.1.1. Estadística de prueba $d_n(\theta)$

Para el caso simple se propone la estadística de prueba (ver Rueda et al. 1991),

$$d_n(\theta) = \int_0^1 \eta_{n,\theta}(s)^2 ds, \tag{4.1}$$

y observe que $d_n(\theta)$ es similar a la estadística de prueba W_n (Crámer-von Mises) que se define en el caso continuo.

Se puede ver que la estadística de prueba $d_n(\theta)$ (ecuación 4.1) no se integra de -1 a 1. Los autores comentan que no se extiende el rango de integración, por qué en el caso de la distribución Poisson los cálculos computacionales para obtener la estadística de prueba son muy complejos.

Para el caso compuesto se propone la estadística de prueba análogo al caso simple, sólo que θ se sustituye por el estimador de máxima verosimilitud $\hat{\theta}$, es decir,

$$d_n(\hat{\theta}) = \int_0^1 \eta_{n,\hat{\theta}}(s)^2 ds.$$

En O'Reilly y Rueda, (1999) dan otra fórmula para calcular $d_n(\theta)$, la cual es más elegante y más sencilla de usar:

¹En el artículo Rueda et al. (1991) hay un error para la expresión de la función de covarianza, en O'Reilly y Rueda, (1999) se da la expresión correcta para la función de covarianza.



Se denota por $\underline{o}^T = (o_0, o_1, \dots, o_j, \dots)$ el vector de dimensión infinita, tal que,

$$o_j = \{ \text{número de } X'_i s = j \}, \quad j = 0, 1, \dots; \quad (4.2)$$

y sea $\underline{e}^T(\theta) = (e_0(\theta), e_1(\theta), \dots, e_j(\theta), \dots)$ el correspondiente vector del número esperado de \underline{o}^T , es decir,

$$e_j(\theta) = \begin{cases} nF_0(0; \theta) & \text{si } j = 0, \\ n[F_0(j; \theta) - F_0(j-1; \theta)] & \text{si } j > 0, \end{cases} \quad (4.3)$$

luego, $\phi_n(s)$ y $\phi_\theta(s)$ pueden ser expresadas

$$\phi_n(s) = \frac{1}{n} \sum_{j=0}^{\infty} s^j o_j, \quad \phi_\theta(s) = \frac{1}{n} \sum_{j=0}^{\infty} s^j e_j(\theta);$$

sustituyendo estos valores en la estadística de prueba (4.1), se tiene que,

$$d_n(\theta) = \frac{1}{n} \int_0^1 \left(\sum_{j=0}^{\infty} s^j (o_j - e_j(\theta)) \right)^2 ds,$$

y con un cálculo se muestra que

$$d_n(\theta) = \frac{1}{n} (\underline{o} - \underline{e}(\theta))^T \mathbb{M} (\underline{o} - \underline{e}(\theta)), \quad (4.4)$$

donde \mathbb{M} es la famosa matriz de Hilbert, es decir,

$$\mathbb{M} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \dots \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \dots \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

En aplicaciones la matriz de Hilbert debe ser truncada, a una matriz cuadrada de tamaño K (con K más grande, por ejemplo que un múltiplo del valor observado más grande de la muestra). Para elegir un tamaño adecuado de la matriz truncada, sólo hay que verificar que $F_0(K; \theta)$ sea muy cercano uno.

4.1.2. Propuesta de estadística de prueba $d_n^*(\theta)$

Como se comentó en la sección anterior la estadística de prueba $d_n(\theta)$ (ecuación 4.1) no se integra de -1 a 1, ya que los autores comentan que no se extiende el rango de

integración, por qué en el caso de la distribución Poisson los cálculos computacionales para obtener la estadística de prueba son muy complejos (ver Rueda et al., 1991). Pero en la actualidad no hay problema en realizar estos cálculos computaciones. Por lo que proponemos la estadística de prueba,

$$d_n^*(\theta) = \int_{-1}^1 \eta_{n,\theta}(s)^2 ds. \tag{4.5}$$

Para el caso compuesto se propone la estadística de prueba análogo al caso simple, sólo que θ se sustituye por el estimador $\hat{\theta}$.

Si queremos poner d_n^* en función de O_i y e_i como en O'Reilly y Rueda, (1999) se tiene que

$$\begin{aligned} d_n^*(\theta) &= \frac{1}{n} \int_{-1}^1 \left(\sum_{j=0}^{\infty} s^j (o_j - e_j(\theta)) \right)^2 ds \\ &= \frac{1}{n} \int_{-1}^1 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} s^{i+j} (o_i - e_i(\theta))(o_j - e_j(\theta)) ds \\ &= \frac{1}{n} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (o_i - e_i(\theta))(o_j - e_j(\theta)) \int_{-1}^1 s^{i+j} ds \\ &= \frac{1}{n} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (o_i - e_i(\theta))(o_j - e_j(\theta)) \left[\frac{1}{i+j+1} - \frac{(-1)^{i+j+1}}{i+j+1} \right]. \end{aligned}$$

Por lo que se tiene que

$$d_n^*(\theta) = \frac{2}{n} (\underline{o} - \underline{e}(\theta))^T \mathbb{N} (\underline{o} - \underline{e}(\theta)), \tag{4.6}$$

en donde \mathbb{N} es igual a,

$$\mathbb{N} = \begin{pmatrix} 1 & 0 & \frac{1}{3} & 0 & \frac{1}{5} & \dots \\ 0 & \frac{1}{3} & 0 & \frac{1}{5} & 0 & \dots \\ \frac{1}{3} & 0 & \frac{1}{5} & 0 & \frac{1}{7} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Vemos en la matriz \mathbb{N} que sus elementos, o bien coinciden con la matriz de Hilbert, ó son iguales a cero, es decir

$$[\mathbb{N}]_{ij} = \begin{cases} [\mathbb{M}]_{ij} & \text{si } i + j \text{ es par} \\ 0 & \text{en otro caso,} \end{cases} \tag{4.7}$$



para $i = 1, 2, \dots; j = 1, 2, \dots$; donde \mathbb{M} es la matriz de Hilbert.

Para aplicaciones, la matriz se trunca igual que el caso $d_n(\theta)$. Quedó pendiente en este trabajo hacer simulaciones de potencia para conocer las ventajas y desventajas relativas de la estadística de prueba $d_n^*(\theta)$.

4.2. Estadísticas de prueba Cramér-von Mises

Estadística de prueba X^2

Antes de dar las estadísticas de prueba Cramér-von Mises, recordemos que en el capítulo 2, se dio la estadística de prueba X^2 de Pearson para el caso continuo, dicha estadística de prueba también se puede utilizar para el caso de bondad de ajuste discreto, la estadística de prueba es,

$$X^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i},$$

donde k es el número de celdas y,

$$p_i = F_0(y_i) - F_0(y_{i-1}), \quad i = 1, \dots, k;$$

donde n es el tamaño de la muestra y desde luego $n = n_1 + \dots + n_k$.

En el capítulo 2, se dieron las estadísticas de prueba Cramér-von Mises para probar bondad de ajuste en el caso continuo. En Choulakian et al. (1994) se dan las estadísticas de prueba Cramér-von Mises para cuando la **distribución es discreta y totalmente conocida**.

Se procede a revisar dichas estadísticas de prueba Cramér-von Mises. Sea p_i definido como arriba y k es el número de celdas elegidas. Se define $e_i = np_i$, $S_j = \sum_{i=1}^j n_i$, $T_j = \sum_{i=1}^j e_i$ y $H_j = T_j/n$. Supongamos que $Z_j = S_j - T_j$, para $j = 1, \dots, k$. Se definen las estadísticas de prueba Cramér-von Mises W^2 , U^2 y A^2 para las distribuciones discretas como:

$$W^2 = n^{-1} \sum_{j=1}^k Z_j^2 p_j,$$

$$U^2 = n^{-1} \sum_{j=1}^k (Z_j - \bar{Z})^2 p_j,$$

$$A^2 = n^{-1} \sum_{j=1}^k \frac{Z_j^2 p_j}{H_j(1 - H_j)},$$

donde $\bar{Z} = \sum_{j=1}^k Z_j p_j$. Como se puede ver fácilmente $Z_k = 0$, por lo que el último término de W^2 es cero. Y el último término de A^2 es de la forma $0/0$, por lo cual se fija igual a cero.

Distribución uniforme discreta

En el caso donde la distribución a probar es la distribución uniforme (totalmente conocida), en el artículo de Choulakian et al., (1994) encontraron la distribución asintótica de las tres estadísticas de prueba Cramér-von Mises (W^2 , U^2 y A^2), donde proporcionan tablas para los percentiles de dichas estadísticas de prueba.

4.2.1. Estadísticas de prueba Cramér-von Mises caso general

En Spinelli y Stephens, (1997) se dio una definición más general de las estadísticas de prueba Cramér-von Mises para probar bondad de ajuste discreto (ya sea el caso simple o el caso compuesto). La cual se da a continuación.

Caso simple

Sea x_1, x_2, \dots, x_n una muestra aleatoria de una variable aleatoria X , y sea o_j y e_j definidas como (4.2) y (4.3), respectivamente. Es decir o_j el número de valores observados de la muestra iguales a j y $e_j = nP(X = j; \theta) = np_j$, el número esperado de valores iguales a j ; se define $Z_j = \sum_{i=0}^j (o_i - e_i)$ y $H_j = \sum_{i=0}^j p_i$ donde $i, j = 0, 1, \dots$; las estadísticas Cramér-von Mises W^2 y A^2 se definen por

$$W^2 = n^{-1} \sum_{j=0}^{\infty} Z_j^2 p_j, \quad (4.8)$$

$$A^2 = n^{-1} \sum_{j=0}^{\infty} \frac{Z_j^2 p_j}{H_j(1 - H_j)}. \quad (4.9)$$

La omisión de p_j en (4.8) puede dar un mayor peso a las desviaciones de las colas, y se define la estadística

$$W_m^2 = n^{-1} \sum_{j=0}^{\infty} Z_j^2, \quad (4.10)$$

la cual puede dar un mejor potencia en contra de otras distribuciones de cola más pesadas.

Caso compuesto

En el caso compuesto, se puede sustituir θ por su estimador de máxima verosimilitud $\hat{\theta}$. Sea $\hat{p}_j = P(X = j; \hat{\theta})$, luego se calculan \hat{e}_j , \hat{Z}_j y \hat{H}_j reemplazando p_j con \hat{p}_j . Las estadísticas de prueba Cramér-von Mises pueden ser calculadas como en (4.8) y (4.9) usando \hat{e}_j , \hat{Z}_j y \hat{H}_j .

En aplicaciones, se necesita que la suma en las estadísticas de prueba sea finita, por lo que se propone que la suma sea truncada por abajo y por arriba; es decir, para truncar por abajo, se elige un número M_l , tal que $o_j = 0$ y $\hat{p}_j < \frac{10^{-3}}{n}$ para $j < M_l$. De igual forma se trunca por arriba, se elige un número M_u , tal que $o_j = 0$ y $\hat{p}_j < \frac{10^{-3}}{n}$ para $j > M_l$.

Cuando se esté en el caso compuesto, las estadísticas de prueba Cramér-von Mises son:

$$W^2 = n^{-1} \sum_{j=M_l}^{M_u} \hat{Z}_j^2 \hat{p}_j, \quad (4.11)$$

$$A^2 = n^{-1} \sum_{j=M_l}^{M_u} \frac{\hat{Z}_j^2 \hat{p}_j}{\hat{H}_j(1 - \hat{H}_j)}, \quad (4.12)$$

y la estadística de prueba (4.10),

$$W_m^2 = n^{-1} \sum_{j=M_l}^{M_u} \hat{Z}_j^2.$$

El truncamiento para las estadísticas de prueba en el caso simple es análogo al caso compuesto. En la siguiente sección se da la estadística de prueba de razón de verosimilitudes generalizadas y la estadística de prueba Kullback-Leibler, la cual es equivalente a la prueba de razón de verosimilitudes generalizadas.

4.3. Estadística de prueba de razón de verosimilitudes generalizadas

Uno se pregunta si los métodos clásicos desarrollados para las pruebas de hipótesis (paramétricas), sirven para probar la bondad de ajuste. Por lo anterior recordemos la definición de cociente (o razón) de verosimilitudes generalizadas para las pruebas de hipótesis paramétricas.

Sea X_1, \dots, X_n una muestra de $F(\cdot; \theta)$; $\theta \in \Theta$ y sea $f(x; \theta)$ su densidad. Bajo una muestra observada $\mathbf{x} = (x_1, \dots, x_n)$, se define la verosimilitud de θ por

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta).$$

Si se quiere probar lo siguiente,

$$H_0 : \theta \in \Theta_0, \quad (\Theta_0 \subset \Theta) \text{ vs.}$$

$$H_1 : \theta \in \Theta_1 \quad (\Theta_1 = \Theta - \Theta_0),$$

el criterio de decisión es rechazar H_0 , si el cociente,

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta_1} L(\theta; \mathbf{x})},$$

es pequeño. El criterio de rechazo es cuando la verosimilitud de H_0 es pequeña comparada con la de H_1 .

Si se quiere extender el concepto de razón de verosimilitudes generalizadas en el caso de pruebas de bondad de ajuste. Para probar $H_0 : F \in \mathcal{F}_0$ vs. $H_1 : F \notin \mathcal{F}_0$. Cuando la familia paramétrica \mathcal{F}_0 es continua, la opción $H_1 : F \notin \mathcal{F}_0$ sigue siendo no paramétrica y muy grande como familia de distribuciones; por lo que al maximizar bajo H_1 , el supremo es ∞ . Por lo que la razón de verosimilitudes generalizadas para el caso continuo no está definido.

Pero para el caso de las familias discretas, al maximizar bajo H_1 , el máximo está acotado, por lo que el uso de la razón de verosimilitudes generalizadas está bien definido. A continuación se da la estadística de razón de verosimilitudes generalizadas para el caso simple y caso compuesto en el caso discreto.



Caso simple

En el denominador de Λ para cada j observado alcanza su máximo en $\frac{O_j}{n}$ por lo que el denominador de Λ es,

$$\prod_{j \in J} \left(\frac{O_j}{n} \right)^{O_j},$$

donde J son los valores que toma la muestra y O_j es el número de valores igual a j . El numerador de Λ , es,

$$\prod_{j \in J} (p(j; \theta))^{O_j}, \text{ donde } p(j; \theta) := P(X = j; \theta).$$

Por lo que la estadística de prueba de razón de verosimilitudes generalizadas bajo la muestra \mathbf{x} es,

$$\Lambda(\mathbf{x}) = \prod_{j \in J} \left(\frac{np(j; \theta)}{O_j} \right)^{O_j}.$$

Caso compuesto

En este caso el numerador de Λ , es la máxima verosimilitud bajo H_0 , es decir,

$$\prod_{j \in J} (p(j; \hat{\theta}))^{O_j}.$$

Por lo que la estadística de prueba de razón de verosimilitudes generalizadas bajo la muestra \mathbf{x} es,

$$\Lambda(\mathbf{x}) = \prod_{j \in J} \left(\frac{np(j; \hat{\theta})}{O_j} \right)^{O_j}.$$

4.3.1. Divergencia Kullback-Leibler

En teoría de la probabilidad, la divergencia de Kullback-Leibler es un indicador de la similitud entre dos distribuciones. Dentro de la teoría de la información también se le conoce como divergencia de la información, ganancia de la información o entropía relativa. La divergencia de Kullback-Leibler entre dos funciones de probabilidades discretas P y Q , se define como:

$$D_{KL}(P||Q) = \sum_i P(i) \log \left(\frac{P(i)}{Q(i)} \right),$$

esta expresión está bien definida si $Q(i) > 0$ para cualquier i , tal que $P(i) > 0$. Si la cantidad $0 \log(0)$ aparece en la expresión, se interpreta como cero.

Se puede ver que $D_{KL}(P||Q) \geq 0$ y es nula si y sólo si $P = Q$, pero no es una métrica por no ser simétrica, es decir, no siempre se cumple que

$$D_{KL}(P||Q) = D_{KL}(Q||P).$$

En el caso de bondad de ajuste discreto con parámetros desconocidos, se define la estadística de prueba Kullback-Leibler, tomando la distribución P que representa a la empírica y a la distribución Q la hipotética, es decir,

$$D_{KL} = \sum_{j \in J} \left(\frac{O_j}{n} \right) \log \left[\frac{\frac{O_j}{n}}{p(j; \hat{\theta})} \right],$$

donde $\hat{\theta}$ es el estimador de máxima verosimilitud. Observe que la estadística de prueba está bien definida, ya que un valor observado j (esto es $O_j > 0$) con probabilidad hipotética $p(j; \hat{\theta}) = 0$ no es razonable, pues esto es como para rechazar automáticamente esta hipótesis.

Con un cálculo sencillo se puede ver que

$$D_{KL} = -\frac{1}{n} \log(\Lambda).$$

Por lo que es equivalente usar la estadística de prueba Kullback-Leibler o la estadística de prueba de razón de verosimilitudes generalizadas.

4.4. Uso de la probabilidad condicional

González-Barrios et al. (2006) haciendo uso de la probabilidad condicional encontraron un nuevo procedimiento para probar bondad de ajuste en el caso discreto. En las siguientes secciones describimos el método.

4.4.1. Resultados teóricos

Sea X_1, \dots, X_n una muestra aleatoria cuya distribución pertenece a la familia paramétrica $\{F(x; \theta) | \theta \in \Theta\}$ donde $\Theta \subset \mathbb{R}^k$. Sea $T_n : \mathbb{R}^n \rightarrow \mathbb{R}^k$ una estadística suficiente para θ ,



luego T_n da toda la información relevante sobre θ . Dada la estadística suficiente T_n podemos definir la distribución condicional de los primeros i términos de la muestra por:

$$\tilde{F}_n(x_1, \dots, x_i) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_i \leq x_i | T_n = t),$$

donde $(x_1, x_2, \dots, x_i) \in \mathbb{R}^i$ y $i = 1, \dots, n$. En un caso particular, $i = 1$, la distribución condicional es la función de distribución Rao-Blackwell de $F(x; \theta)$ que se definió en el capítulo 3.

Si X_1, \dots, X_n son variables aleatorias discretas que toman valores en $M \subset \mathbb{R}$, cuya función de probabilidad pertenece a la familia $\{f(x; \theta) | \theta \in \Theta\}$, donde $\Theta \subset \mathbb{R}$ y T_n es una estadística suficiente para θ , se puede usar la **probabilidad condicional discreta**, definida como:

$$\tilde{f}_n(x_1, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | T_n = t),$$

donde $(x_1, \dots, x_n) \in M^n$, y $T_n(x_1, \dots, x_n) = t$; ya que si $T_n(x_1, \dots, x_n) \neq t$ se tiene que $\tilde{f}_n(x_1, \dots, x_n) = 0$. Como T_n es una estadística suficiente, la probabilidad condicional discreta no depende de θ .

Una propiedad importante es que dentro de la condicional las variables aleatorias son intercambiables, es decir, para cualquier permutación σ de un conjunto $\{1, 2, \dots, n\}$, se tiene que

$$P(X_1 = x_1, \dots, X_n = x_n | T_n = t) = P(X_1 = x_{\sigma(1)}, \dots, X_n = x_{\sigma(n)} | T_n = t). \quad (4.13)$$

El primer autor, del que se tiene conocimiento, que usó la probabilidad condicional dada una estadística suficiente para bondad de ajuste fue Fisher (1950). Dicho autor utilizó la probabilidad condicional para probar bondad de ajuste para el caso de la distribución Poisson.

Primero se da la notación para las distribuciones Poisson, binomial y binomial negativa, para luego dar sus correspondientes probabilidades condicionales.

Se dice que X tiene una distribución binomial con parámetros $k \geq 1$ y $0 < p < 1$, con k conocido, denotado por $X \sim Bin(k, p)$ si y sólo si

$$P(X = x; p) = \binom{k}{x} p^x (1-p)^{k-x} \quad \text{si } x = 0, 1, \dots, k.$$

Se dice que X tiene una distribución Poisson con parámetro $\lambda > 0$, denotado por $X \sim \text{Poisson}(\lambda)$ si y sólo si

$$P(X = x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{si } x = 0, 1, \dots$$

Se dice que X tiene una distribución binomial negativa con parámetros $r \geq 1$ y $0 < p < 1$, con r conocido, denotado por $X \sim \text{NB}(r, p)$ si y sólo si

$$P(X = x; p) = \binom{r+x-1}{x} p^r (1-p)^x \quad \text{si } x = 0, 1, \dots$$

Resulta que las tres distribuciones definidas anteriormente son casos particulares de la distribución de series de potencia, definida por:

$$P(X = x; \theta) = \frac{a_x \theta^x}{b(\theta)} \quad \text{para } x = 0, 1, \dots \quad (4.14)$$

en donde $\theta > 0$, $a_x \geq 0$ y $b(\theta) = \sum_{x=0}^{\infty} a_x \theta^x$. Nótese que para una muestra de esta distribución, la estadística suficiente minimal es $\sum_{i=1}^n X_i$.

Recordando la definición de la función generadora de probabilidad,

$$\phi_{\theta}(s) = E(s^X; \theta), \quad \text{para } |s| \leq 1, \quad \text{donde } s \in \mathbb{R}$$

La función generadora de la probabilidad para distribución (4.14) es,

$$\begin{aligned} \phi_{\theta}(s) &= \sum_{j=0}^{\infty} s^j \frac{a_j \theta^j}{b(\theta)} \\ &= \frac{1}{b(\theta)} \sum_{j=0}^{\infty} a_j (s\theta)^j \\ &= \frac{b(s\theta)}{b(\theta)}. \end{aligned}$$

En González-Barrios et al. (2006) se encuentra la probabilidad condicional discreta de (4.14), a continuación damos dicho resultado.



Teorema 5 (González-Barrios et al., 2006) Sea X_1, X_2, \dots, X_n una muestra aleatoria de la distribución de series de potencia

$$P(X = x; \theta) = \frac{a_x \theta^x}{b(\theta)} \quad \text{para } x = 0, 1, \dots \quad (4.15)$$

en donde $\theta > 0$, $a_x \geq 0$ y $b(\theta) = \sum_{x=0}^{\infty} a_x \theta^x$. Sea $T_n = \sum_{i=1}^n X_i$. Entonces,

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | T_n = t) = \frac{a_{x_1} a_{x_2} \cdots a_{x_n}}{\hat{a}_t}, \quad (4.16)$$

donde \hat{a}_t corresponde a la probabilidad de T_n (que también pertenece a una distribución de series de potencia), y para los $x_i \geq 0$, $i = 1, 2, \dots, n$ tales que $\sum_{i=1}^n x_i = t$. El recíproco también es válido, es decir, si X_1, \dots, X_n tiene una distribución condicional dada $T_n = \sum_{i=1}^n X_i = t$ para cualquier n y t enteros no negativos como la ecuación (4.16), entonces X_i tiene una distribución como la dada en (4.15).

A continuación se ejemplifica la probabilidad condicional para la distribución Poisson, binomial y binomial negativa.

En el caso de la distribución Poisson con parámetro λ , $a_{x_i} = 1/x_i!$, para cada $i = 1, 2, \dots, n$, y $\hat{a}_t = 1/t!$. Luego la probabilidad condicional de (X_1, \dots, X_n) dada $T_n = t$ es

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

la cual corresponde a una distribución multinomial donde $p_1 = p_2 = \cdots = p_n = \frac{1}{n}$.

En el caso de la distribución binomial con parámetro $k \geq 1$, el cual asumimos que es conocido, y $0 < p < 1$, tenemos que $a_{x_i} = \binom{k}{x_i}$, para cada $i = 1, 2, \dots, n$, y $\hat{a}_t = \binom{nk}{t}$. La probabilidad condicional de (X_1, \dots, X_n) dada $T_n = t$ es

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | T_n = t) = \frac{\binom{k}{x_n} \binom{k}{x_{n-1}} \cdots \binom{k}{x_1}}{\binom{nk}{t}},$$

la cual corresponde a la versión multivariada de la distribución hipergeométrica, con n clases de tamaño k cada una.

En el caso de la distribución binomial negativa con parámetro $r \geq 1$, el cual asumimos que es conocido, y $0 < p < 1$, tenemos que $a_{x_i} = \binom{r + x_i - 1}{r - 1}$, para cada $i = 1, 2, \dots, n$, y $\hat{a}_t = \binom{nr + t - 1}{nr - 1}$. La probabilidad condicional de (X_1, \dots, X_n) dada $T_n = t$ es

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | T_n = t) \\ = \frac{\binom{r + x_n - 1}{r - 1} \binom{r + x_{n-1} - 1}{r - 1} \cdots \binom{r + x_1 - 1}{r - 1}}{\binom{nr + t - 1}{nr - 1}}. \end{aligned}$$

4.4.2. Descripción de la prueba de bondad de ajuste

En González-Barrios et al. (2006) se describe un método de prueba de bondad de ajuste para el caso Poisson, y comentan que se puede utilizar para cualquier distribución que cumpla el Teorema 5; a continuación se describe dicho método.

Supongamos que tenemos una muestra aleatoria (X_1, \dots, X_n) y se quiere probar:

$$\begin{aligned} H_0 : X \sim \text{Poisson}(\lambda) \text{ para algún } \lambda > 0 \text{ vs.} \\ H_1 : X \not\sim \text{Poisson}(\lambda) \text{ para ningún } \lambda > 0, \end{aligned} \tag{4.17}$$

con un nivel α . Como se vio en la sección anterior, bajo H_0 la distribución condicional de (X_1, \dots, X_n) dada $T_n = t$ es una distribución multinomial con parámetro t y $p_1 = \cdots = p_n = \frac{1}{n}$.

Tomando los elementos del citado artículo de González-Barrios et al., (2006) se reproducen a continuación las ideas y ejemplos básicos.



Se dice que un vector \mathbf{Y} es un arreglo de (t, n) , si \mathbf{Y} es un vector de tamaño n , $\mathbf{Y} = (y_n, y_{n-1}, \dots, y_2, y_1)$ y cumple que $y_n \geq y_{n-1} \geq \dots \geq y_2 \geq y_1$ con $\sum_{i=1}^n y_i = t$. La probabilidad condicional de este arreglo \mathbf{Y} de (t, n) bajo H_0 es,

$$\begin{aligned} P(X_n = y_n, X_{n-1} = y_{n-1}, \dots, X_2 = y_2, X_1 = y_1 | T_n = t) \\ = \binom{t}{y_n} \binom{t-y_n}{y_{n-1}} \binom{t-y_n-y_{n-1}}{y_{n-2}} \dots \binom{t-y_n-\dots-y_2}{y_1} \left(\frac{1}{n}\right)^t. \end{aligned}$$

La prueba consiste en encontrar todos los arreglos de (t, n) y calcular su respectiva probabilidad condicional, luego la probabilidad condicional de cualquier permutación de un arreglo será la misma. Por ejemplo suponga que $n = 10$ y $t = 5$, luego la probabilidad condicional del arreglo $(2, 2, 1, 0, 0, 0, 0, 0, 0, 0)$ esta dado por

$$\begin{aligned} P((2, 2, 1, 0, 0, 0, 0, 0, 0, 0) | T_{10} = 5) \\ = \binom{10}{2} \binom{8}{1} \binom{7}{7} \binom{5}{2} \binom{3}{2} \binom{1}{1} \left(\frac{1}{10}\right)^5 \\ = .108, \end{aligned}$$

en donde los primeros tres coeficientes binomiales corresponden a permutaciones del arreglo.

Una vez que se encuentra todos los arreglos de (t, n) , se procede a ordenarlos en forma decreciente, según su probabilidad condicional; de la más alta a la más baja. Al evaluar la probabilidad acumulada de los arreglos, se encuentra aquel arreglo con probabilidad acumulada más cercana a $1-\alpha$. La región crítica (región de rechazo), la forman los arreglos subsecuentes cuya probabilidad es aproximadamente α .

Para ilustrar lo anterior, suponga H_0 cierta, es decir, $X \sim Poisson(\lambda)$ para algún λ y suponga $n = 10$ y con $t = 8$. En la tabla 4.1 se muestran todos los arreglos de (t, n) en orden decreciente, según sus respectivas probabilidades condicionales, que ahí aparecen, al igual que sus probabilidades acumuladas. Si se quisiera probar (4.17), con un $\alpha \approx 0.1$. De la tabla (4.1) se ve que la región crítica es el complemento del conjunto de arreglos

$$\begin{aligned} R = \{ & (2, 2, 1, 1, 1, 1, 0, 0, 0, 0), (3, 2, 1, 1, 1, 0, 0, 0, 0, 0) \\ & (2, 1, 1, 1, 1, 1, 1, 0, 0, 0), (2, 2, 2, 1, 1, 0, 0, 0, 0, 0), \\ & (3, 1, 1, 1, 1, 1, 0, 0, 0, 0), (3, 2, 2, 1, 0, 0, 0, 0, 0, 0)\}. \end{aligned}$$

Tabla 4.1: Arreglos y sus probabilidades para $T_{10} = 8$ en el caso Poisson.

Arreglo	Probabilidad bajo H_0	Probabilidad acumulada
(2,2,1,1,1,1,0,0,0,0)	0.31752	0.31752
(3,2,1,1,1,0,0,0,0,0)	0.169344	0.486864
(2,1,1,1,1,1,1,0,0,0)	0.169344	0.656208
(2,2,2,1,1,0,0,0,0,0)	0.127008	0.783216
(3,1,1,1,1,1,0,0,0,0)	0.084672	0.867888
(3,2,2,1,0,0,0,0,0,0)	0.042336	0.910224
(4,1,1,1,1,0,0,0,0,0)	0.021168	0.931392
(4,2,1,1,0,0,0,0,0,0)	0.021168	0.952560
(1,1,1,1,1,1,1,1,0,0)	0.018144	0.970704
(3,3,1,1,0,0,0,0,0,0)	0.014112	0.984816
(2,2,2,2,0,0,0,0,0,0)	0.005292	0.990108
(5,1,1,1,0,0,0,0,0,0)	0.0028224	0.9929304
(3,3,2,0,0,0,0,0,0,0)	0.002016	0.9949464
(4,3,1,0,0,0,0,0,0,0)	0.002016	0.9969624
(4,2,2,0,0,0,0,0,0,0)	0.001512	0.9984744
(5,2,1,0,0,0,0,0,0,0)	0.0012096	0.9996840
(6,1,1,0,0,0,0,0,0,0)	0.0002016	0.9998856
(5,3,0,0,0,0,0,0,0,0)	0.0000504	0.999936
(4,4,0,0,0,0,0,0,0,0)	0.0000315	0.9999675
(6,2,0,0,0,0,0,0,0,0)	0.0000252	0.9999927
(7,1,0,0,0,0,0,0,0,0)	0.0000072	0.9999999
(8,0,0,0,0,0,0,0,0,0)	0.0000001	1.0000000

El nivel condicional actual de la prueba es $\alpha_{T_{10}=8} = 0.089776$. Si por ejemplo, se observa el arreglo (3,3,1,1,0,0,0,0,0,0), se rechaza H_0 con el referido $\alpha_{T_{10}=8} = 0.089776$.

4.5. Distribución binomial negativa generalizada

Se dice que X tiene una distribución binomial negativa generalizada con parámetros (m, β, α) , tales que $m > 0$, $0 < \alpha < 1$ y $|\alpha\beta| < 1$, con m y β conocidos, denotado por $X \sim GNB(m, \beta, \alpha)$ si y sólo si (ver Jain y Consult, 1971)

$$P(X = x; \alpha) = \frac{m}{m + \beta x} \binom{m + \beta x}{x} \alpha^x (1 - \alpha)^{m + \beta x - x} \quad \text{si } x = 0, 1, \dots$$

En Consult y Gupta (1980) mostraron que la binomial negativa generalizada es en verdad una función de distribución de probabilidad, si $\beta = 0$ o bien si α está en $0 \leq \alpha \leq \frac{1}{\beta}$

($\beta \geq 1$), y no es una función de distribución para cualesquiera otros valores, además no existen los momentos de esta distribución cuando $\alpha = \frac{1}{\beta}$.

Cuando $\beta = 0$ y $\beta = 1$ la distribución se convierte en la binomial y binomial negativa, respectivamente.

En Jain y Consult, (1971) se prueba que si $X \sim GNB(m_1, \beta, \alpha)$ y $Y \sim GNB(m_2, \beta, \alpha)$ son independientes, entonces $X + Y \sim GNB(m_1 + m_2, \beta, \alpha)$.

El estimador de máxima verosimilitud de α es (ver Famoye, 1997),

$$\hat{\alpha} = \frac{\bar{x}}{m + \beta \bar{x}},$$

donde \bar{x} es la media muestral.

4.5.1. Distribución series de potencias modificada

Se define la distribución de series de potencias modificada como (ver Gupta, 1974),

$$P(X = x; \theta) = \frac{a_x [g(\theta)]^x}{b(\theta)}, \quad \text{para } x \in T \text{ con } T \subseteq \mathbb{N} \cup \{0\}, \quad (4.18)$$

donde $\theta > 0$, $a_x \geq 0$ y $b(\theta) = \sum_{x \in T} a_x [g(\theta)]^x$.

Un caso particular de la distribución de series de potencia modificada, es cuando $g(\theta)$ es invertible y $T = 0, 1, \dots$; en dicho caso se convierte en la distribución series de potencia vista en la sección 4.4.1 (ecuación 4.14).

En el caso de la binomial negativa generalizada, tomando $a_x = \frac{m}{m+\beta x} \binom{m+\beta x}{x} \geq 0$, $g(\alpha) = \alpha(1-\alpha)^{\beta-1} > 0$ y $b(\alpha) = (1-\alpha)^{-m}$. Se tiene que

$$P(X = x; \alpha) = \frac{a_x [g(\alpha)]^x}{b(\alpha)}, \quad \text{para } x = 0, 1, \dots \quad (4.19)$$

donde $b(\alpha) = \sum_{x=0}^{\infty} a_x [g(\alpha)]^x$.

Por lo que la distribución binomial negativa generalizada pertenece a la distribución de series de potencia modificada. En el caso que $\beta > 1$, la distribución binomial negativa generalizada no pertenece a la distribución de series de potencia, ya que $g(\alpha)$ no es invertible.



4.5.2. Probabilidad condicional

La probabilidad condicional de (X_1, \dots, X_n) dada $T_n = \sum_{i=1}^n x_i = t$ es

$$\begin{aligned} & P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | T_n = t) \\ &= \frac{P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, T_n = t; \alpha)}{P(T_n = t; \alpha)} \\ &= \frac{P(X_1 = x_1, X_2 = x_2, \dots, X_n = t - \sum_{i=1}^{n-1} x_i; \alpha)}{P(T_n = t; \alpha)}. \end{aligned}$$

Se puede ver que en la última expresión el numerador es diferente de cero sólo si x_1, \dots, x_{n-1} son enteros no negativos cuya suma es menor o igual a t , y el numerador es la probabilidad de eventos independientes por lo que se puede calcular fácilmente; en el denominador la probabilidad es

$$\frac{nm}{nm + \beta t} \binom{nm + \beta t}{t} \alpha^t (1 - \alpha)^{nm + \beta t - t},$$

por lo que la probabilidad condicional nos queda:

$$\begin{aligned} & P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | T_n = t) \\ &= \frac{\frac{m}{m + \beta x_n} \binom{m + \beta x_n}{x_n} \frac{m}{m + \beta x_{n-1}} \binom{m + \beta x_{n-1}}{x_{n-1}} \dots \frac{m}{m + \beta x_1} \binom{m + \beta x_1}{x_1}}{\frac{nm}{nm + \beta t} \binom{nm + \beta t}{t}}. \end{aligned}$$

Por lo que si se quiere probar que una muestra aleatoria X_1, \dots, X_n sigue una distribución binomial negativa generalizada con parámetros m y β conocidos ($\beta > 1$), se procede como en González-Barrios et al. (2006). Este resultado, en opinión del autor es un resultado nuevo.

4.6. Generación de muestras Look-Alike

Se pueden generar muestras look-alike como se vio en el capítulo 3 en el teorema de O'Reilly y Gracia-Medrano (2006). Si tenemos una muestra aleatoria X_1, \dots, X_n y

$T_n = \sum_{i=1}^n X_i$ es una estadística suficiente doblemente transitiva, es decir, el conocimiento de X_n y T_n es equivalente al conocimiento de X_n y T_{n-1} , se tiene que

$$\begin{aligned} P(X_n = x_n | T_n = t_n) &= \frac{P(X_n = x_n, T_n = t_n; \theta)}{P(T_n = t_n; \theta)} \\ &= \frac{P(X_n = x_n, T_{n-1} = t_n - x_n; \theta)}{P(T_n = t_n; \theta)} \\ &= \frac{P(X_n = x_n; \theta)P(T_{n-1} = t_n - x_n; \theta)}{P(T_n = t_n; \theta)}. \end{aligned} \quad (4.20)$$

De igual forma, aplicando doble transitividad e independencia, se tiene que la probabilidad condicional,

$$\begin{aligned} P(X_{n-1} = x_{n-1} | T_n = t_n, X_n = x_n) &= \frac{P(X_{n-1} = x_{n-1}, T_n = t_n, X_n = x_n; \theta)}{P(T_n = t_n, X_n = x_n; \theta)} \\ &= \frac{P(X_{n-1} = x_{n-1}, T_{n-1} = t_n - x_n, X_n = x_n; \theta)}{P(T_{n-1} = t_n - x_n, X_n = x_n; \theta)} \\ &= \frac{P(X_{n-1} = x_{n-1}, T_{n-1} = t_n - x_n; \theta)P(X_n = x_n; \theta)}{P(T_{n-1} = t_n - x_n; \theta)P(X_n = x_n; \theta)} \\ &= \frac{P(X_{n-1} = x_{n-1}, T_{n-1} = t_n - x_n; \theta)}{P(T_{n-1} = t_n - x_n; \theta)} \\ &= P(X_{n-1} = x_{n-1} | T_{n-1} = t_n - x_n). \end{aligned}$$

Es decir,

$$P(X_{n-1} = x_{n-1} | T_n = t_n, X_n = x_n) = P(X_{n-1} = x_{n-1} | T_{n-1} = t_n - x_n).$$

De forma análoga, se tiene que

$$\begin{aligned} P(X_{n-2} = x_{n-2} | T_n = t_n, X_n = x_n, X_{n-1} = x_{n-1}) &= P(X_{n-2} = x_{n-2} | T_{n-2} = t_n - x_n - x_{n-1}) \\ &\vdots \\ P(X_1 = x_1 | T_n = t_n, X_n = x_n, \dots, X_2 = x_2) &= P\left(X_1 = x_1 | T_1 = t_n - \sum_{i=2}^n x_i\right). \end{aligned}$$

O sea se identificaron

$$\begin{aligned} P(X_n = x_n | T_n = t_n) &\longrightarrow \tilde{f}_n(x_n | t_n), \\ P(X_{n-1} = x_{n-1} | T_n = t_n, X_n = x_n) &\longrightarrow \tilde{f}_{n-1}(x_{n-1} | t_{n-1}), \\ P(X_{n-2} = x_{n-2} | T_n = t_n, X_n = x_n, X_{n-1} = x_{n-1}) &\longrightarrow \tilde{f}_{n-2}(x_{n-2} | t_{n-2}), \\ &\vdots \\ P(X_1 = x_1 | T_n = t_n, X_n = x_n, \dots, X_2 = x_2) &\longrightarrow \tilde{f}_1(x_1 | t_1). \end{aligned}$$

Las correspondientes $\tilde{F}_1, \dots, \tilde{F}_n$, se calculan a partir de las $\tilde{f}_1, \dots, \tilde{f}_n$. Si se quiere generar una muestra condicional procedemos como en el teorema de O'Reilly y Gracia-Medrano (2006) que se vio en el capítulo 3.

4.6.1. Distribución Rao-Blackwell para algunas distribuciones

Se tiene la distribución de series de potencias

$$P(X = x; \theta) = \frac{a_x \theta^x}{b(\theta)} \quad \text{para } x = 0, 1, \dots \quad (4.21)$$

donde $\theta > 0$, $a_x \geq 0$ y $b(\theta) = \sum_{x=0}^{\infty} a_x \theta^x$.

Si $T_n = \sum_{i=1}^n X_i = t_n$, donde las X_i 's son distribuciones de series de potencia, tenemos que (ver González-Barrios et al., 2006),

$$P(T_n = t_n; \theta) = \frac{\hat{a}_{(t_n, n)} \theta^{t_n}}{b(\theta)^n},$$

donde la función $\hat{a}_{(t_n, n)}$ sólo depende de t_n y n y no de θ .

Aplicando la igualdad de la ecuación (4.20) en la distribución (4.21), tenemos que

$$\begin{aligned} P(X_n = x_n | T_n = t_n) &= \frac{\frac{a_{x_n} \theta^{x_n}}{b(\theta)} \frac{\hat{a}_{(t_n-1, n-1)} \theta^{t_n-x_n}}{b(\theta)^{n-1}}}{\frac{\hat{a}_{(t_n, n)} \theta^{t_n}}{b(\theta)^n}} \\ &= \frac{a_{x_n} \hat{a}_{(t_n-1, n-1)}}{\hat{a}_{(t_n, n)}}, \quad \text{donde } t_{n-1} = t_n - x_n. \end{aligned}$$

Por lo que la distribución Rao-Blackwell para la series de potencia es

$$\tilde{F}_n(x'_n | t_n) = \sum_{j=0}^{x'_n} \frac{a_j \hat{a}_{(t_*, n-1)}}{\hat{a}_{(t_n, n)}}, \quad \text{donde } t_* = t_n - j.$$

De igual forma aplicando la igualdad de la ecuación (4.20) para la distribución Poisson tenemos que

$$P(X_n = x_n | T_n = t_n) = \binom{t_n}{x_n} \left(\frac{1}{n}\right)^{x_n} \left(1 - \frac{1}{n}\right)^{t_n-x_n},$$

la cual es la probabilidad de una distribución binomial con parámetros t_n y $\frac{1}{n}$.

Por lo que la distribución Rao-Blackwell para la distribución Poisson es

$$\tilde{F}_n(x'_n | t_n) = \sum_{j=0}^{x'_n} \text{pbin}(j; t_n, \frac{1}{n}),$$

donde $pbm(j; t_n, \frac{1}{n})$ es la probabilidad de que una variable aleatoria binomial con parámetros t_n y $\frac{1}{n}$ tome el valor j .

Para la distribución binomial tenemos que

$$P(X_n = x_n | T_n = t_n) = \frac{\binom{k}{x_n} \binom{nk - k}{t_n - x_n}}{\binom{nk}{t_n}},$$

la cual es la probabilidad de una distribución hipergeométrica con parámetros t_n , k y nk .

Por lo que la distribución Rao-Blackwell para la distribución binomial es

$$\tilde{F}_n(x'_n | t_n) = \sum_{j=0}^{x'_n} phiper(j; t_n, k, nk),$$

donde $phiper(j; t_n, k, nk)$ es la probabilidad de que una variable aleatoria hipergeométrica con parámetros t_n , k y nk tome el valor j .

Para la distribución binomial negativa tenemos que

$$P(X_n = x_n | T_n = t_n) = \frac{\binom{r + x_n - 1}{x_n} \binom{(n-1)r + t_n - x_n - 1}{t_n - x_n}}{\binom{nr + t_n - 1}{t_n}}.$$

Por lo que su distribución Rao-Blackwell es

$$\tilde{F}_n(x'_n | t_n) = \sum_{j=0}^{x'_n} \frac{\binom{r + j - 1}{j} \binom{(n-1)r + t_n - j - 1}{t_n - j}}{\binom{nr + t_n - 1}{t_n}}.$$

Para la distribución binomial negativa generalizada tenemos que

$$\begin{aligned} & P(X_n = x_n | T_n = t_n) \\ &= \frac{\frac{m}{m + \beta x_n} \binom{m + \beta x_n}{x_n} \frac{nm - m}{nm - m + \beta(t_n - x_n)} \binom{nm - m + \beta(t_n - x_n)}{t_n - x_n}}{\frac{nm}{nm + \beta t_n} \binom{nm + \beta t_n}{t_n}}. \end{aligned}$$

La distribución Rao-Blackwell para la binomial negativa generalizada es

$$\begin{aligned} & \tilde{F}_n(x'_n | t_n) \\ &= \sum_{j=0}^{x'_n} \frac{\frac{m}{m + \beta j} \binom{m + \beta j}{j} \frac{nm - m}{nm - m + \beta(t_n - j)} \binom{nm - m + \beta(t_n - j)}{t_n - j}}{\frac{nm}{nm + \beta t_n} \binom{nm + \beta t_n}{t_n}}. \end{aligned}$$

Teniendo la función de distribución Rao-Blackwell para estas 4 distribuciones (binomial, binomial negativa, Poisson y binomial negativa generalizada), podemos generar las muestras look-alike.

4.7. Aplicaciones

4.7.1. Nidos de gorrión (Sparrow nests)

En Zar (1974) presentan el número de nido gorrión encontrados en parcelas de una hectárea. En nueve áreas no encontraron nidos, en 22 áreas encontraron un nido, en 6 áreas encontraron 2 nidos, en 2 áreas encontraron 3 nidos, y 4 nidos se observaron en una sola área.

Se quiere probar si los datos siguen una distribución Poisson con parámetro λ desconocido, es decir queremos probar

$$\begin{aligned} H_0 : X &\sim \text{Poisson}(\lambda) \text{ para algún } \lambda > 0 \text{ vs.} \\ H_1 : X &\not\sim \text{Poisson}(\lambda) \text{ para ningún } \lambda > 0. \end{aligned} \quad (4.22)$$

Utilizando las estadísticas de prueba d_{KL} (sección 4.3.1), d_n (sección 4.1.1), d_n^* (sección 4.1.2), W^2 , A^2 y W_m^2 (sección 4.2.1). En la tabla 4.2 se muestran los valores de las 6 estadísticas de prueba, así como sus correspondientes p-valores condicionales para 10000 muestras look-alike.

De la tabla 4.2 podemos observar que para 5 estadísticas de prueba podemos rechazar H_0 con un nivel de significancia $\alpha = 0.06$.

Por otro lado aplicando en método visto en González-Barríos et al. (2006) para probar (4.22). Para los datos anteriores tenemos que el tamaño de muestra igual a 40 y la suma de los datos es igual a 44, por lo que $T_{40} = 44$.

El total de arreglos de $T_{40} = 44$ es de 75168 arreglos, luego ordenando los arreglos de mayor probabilidad condicional a menor probabilidad condicional y se calcula su suma acumulada. La probabilidad acumulada correspondiente a la muestra que se observó es aproximadamente 0.806. Eso indica un p-valor condicional para este procedimiento apro-

Tabla 4.2: Nidos de gorrión. p-valor condicional con 10000 muestras look-alike. Los cálculos de las estadísticas de prueba y sus correspondientes p-valores fueron programadas en el software libre R version 2.14.0. El tiempo promedio en hacer estos cálculos es de aproximadamente 8 segundos.

Estadística de Prueba	Valor	p-valor condicional
D_{KL}	0.0771609	0.1867
d_n	0.1035778	0.0469
d_n^*	2.0004270	0.0181
W^2	0.2443600	0.0277
A^2	1.1584550	0.0520
W_m^2	0.7217253	0.0339

ximadamente de 0.2 que es parecida al p-valor condicional usando el criterio de Kullback-Leibler, que sugiere no rechazar H_0 , a un nivel de significancia menor a 0.2.

Sin embargo, con las otras 5 estadísticas utilizadas se sugiere un rechazo si el nivel es inferior a 0.06. De acuerdo a estas últimas estadísticas se concluye que se puede rechazar H_0 .

4.7.2. Ciclones tropicales de categoría 4 y 5 en el Atlántico

En el apéndice A.1, se muestran los datos de ciclones tropicales de categoría 4 y 5 ocurridos en el Atlántico Norte publicados en Carrillo-Negrete, (2010) y posteriormente usados en la tesis de López-Martínez, (2010).

Si se quiere probar si los datos siguen una distribución Poisson con parámetro λ desconocido.

En la tabla 4.3 se muestran los valores de las 6 estadísticas de prueba, así como sus correspondientes p-valores condicionales para 10000 muestras look-alike. De dicha tabla, podemos observar que para las 6 estadísticas de prueba sus p-valores condicionales son mayores a 0.84.

Por otra lado aplicando el método de González-Barrios et al. (2006). El total de arreglos con $T_{30} = 39$ es de 31118 arreglos. La probabilidad acumulada de la muestra observada correspondiente es aproximadamente 0.222. En este caso el p-valor condicional coincide

Tabla 4.3: Ciclones Tropicales. p-valor condicional con 10000 muestras look-alike. Los cálculos de las estadísticas de prueba y sus correspondientes p-valores fueron programadas en el software libre R version 2.14.0. El tiempo promedio en hacer estos cálculos es de aproximadamente 12 segundos.

Estadística de Prueba	Valor	p-valor condicional
D_{KL}	0.0309956400	0.8445
d_n	0.0001843693	0.9492
d_n^*	0.0107185897	0.9835
W^2	0.0031376021	0.9924
A^2	0.0599036220	0.9478
W_m^2	0.0252752959	0.9924

con las otras estadísticas de prueba, y no se recomienda rechazar H_0 . Por lo que concluimos que los datos se pueden modelar adecuadamente por la distribución Poisson.

4.7.3. Mosquito transmisor de malaria

En el apéndice A.1, se encuentran los datos utilizados en la tesis de López-Martínez, (2010); en este estudio se presenta el número de mosquitos transmisores de malaria en etapa esporozoíto positiva capturados con trampas de luz en una región de Sudáfrica.

Se quiere probar si los datos siguen una distribución binomial negativa generalizada con m y β conocidos, y para algún α desconocido.

Utilizando la distribución Rao-Blackwell para la binomial negativa generalizada que desarrollamos en el capítulo 4, podemos proceder como en O'Reilly y Gracia-Medrano, (2006) para generar muestras look-alike.

En la tabla 4.4, se muestran para diferentes valores de m y β , los valores de las 6 estadísticas de prueba, así como sus correspondientes p-valores condicionales para 1000 muestras look-alike. Vemos que para diferentes valores de m y β sus p-valores condicionales son relativamente grandes. Por lo que concluimos que los datos se pueden modelar adecuadamente por la distribución binomial negativa generalizada (por ejemplo con parámetros $m = 0.1$, $\beta = 1.4$ y para algún α desconocido).

Tabla 4.4: Mosquito transmisor de Malaria. Valor de la estadística de prueba y su correspondiente p-valor condicional con 1000 muestras look-alike para la distribución GNB. Los estimadores de máxima verosimilitud son $\hat{m} = 0.1237229$, $\hat{\beta} = 1.4459204$ y $\hat{\alpha} = 0.2532310$.

Valor del parámetro	D_{KL}	d_n	d_n^*	W^2	A^2	W_m^2
$m = 0.1$ y $\beta = 1.4$	0.00395 0.298	0.00048 0.697	0.00272 0.821	0.00179 0.712	0.05776 0.693	0.00407 0.812
$m = 0.1$ y $\beta = 1.6$	0.00449 0.322	0.00296 0.343	0.02219 0.420	0.01052 0.366	0.34628 0.366	0.01448 0.471
$m = 0.2$ y $\beta = 1.6$	0.00428 0.132	0.00155 0.344	0.03016 0.337	0.00576 0.397	0.16474 0.388	0.00948 0.453

Conclusiones y perspectivas

En este trabajo se abordó el problema de bondad de ajuste continuo y discreto. Teniendo una muestra aleatoria X_1, \dots, X_n y una hipótesis nula, y suponiendo que conocemos la distribución Rao-Blackwell, con las facilidades de cómputo que se tienen en la actualidad, podemos generar muestras condicionales (O'Reilly y Gracia-Medrano, 2006). Lo anterior implica que para cualquier estadística de prueba, se puede calcular su p-valor condicional sin la necesidad de tablas.

Se revisaron las pruebas de bondad de ajuste en el caso continuo, basadas en la distribución empírica (EDF). Las estadísticas de prueba EDF que se dieron fueron: Kolmogorov, Kuiper y las tres estadísticas de prueba basadas en la familia Cramér-von Mises. Se reprogramaron por separado los cálculos computacionales para bondad de ajuste de la distribución Gaussiana inversa y la distribución exponencial truncada, usando las 5 estadísticas de prueba EDF.

Se revisó la función de distribución Rao-Blackwell para la familia de distribuciones de series de potencia, así como sus casos particulares (binomial, binomial negativa y Poisson). Se extendió el resultado de la distribución Rao-Blackwell para la binomial negativa generalizada. Así que para estas distribuciones se pueden generar muestras condicionales (look-alike) como se hace en O'Reilly y Gracia-Medrano, (2006). Se desarrolló la probabilidad condicional conjunta para el caso de la binomial negativa generalizada, por lo anterior se derivó una prueba para dicha distribución, como se hizo en González-Barrios et al. (2006). Se instrumentó un programa computacional para el problema de bondad de ajuste en los casos Poisson y binomial negativa generalizada.

Por otro lado, para el caso general de bondad de ajuste discreto, la estadística de prueba d_n , que está en función del proceso empírico (un símil al caso continuo); se propuso



una nueva estadística de prueba d_n^* , que es una extensión de la estadística de prueba d_n (O'Reilly y Rueda, 1999). Se revisaron las tres estadísticas de prueba que se derivan de la familia Cramér-von Mises (Spinelli y Stephens, 1997) y la estadística de prueba Kullback-Leibler. Se desarrolló un programa computacional para el problema de bondad de ajuste en los casos de la Poisson y binomial negativa generalizada, en el cual se utilizan las 6 estadísticas de prueba.

Quedó pendiente en relación a la estadística propuesta d_n^* y también para el caso de la binomial negativa generalizada hacer simulaciones de potencia, para conocer las ventajas relativas de las distintas pruebas.

Como un último comentario, el procedimiento (equivalente a una enumeración) de González-Barrios et al. (2006), tiene el inconveniente de que si el número de arreglos (asociado a (t_n, n)) es muy grande, entonces el procedimiento no es aplicable. Queda sin embargo la posibilidad de obtener el p-valor condicional generando muestras look-alike.

Bibliografía

- [1] R.H. Berk. Strong consistency of certain sequential estimators. *Ann. Math. Statist.*, 40:1492–1495, 1969.
- [2] B.B. Bhattacharyya and P.K. Sen. Weak convergence of the Rao-Blackwell estimator of a distribution function. *The Annals of Probability*, 5(3):500–510, 1977.
- [3] B.M. Brown. Martingale central limit theorem. *Ann. Math. Statist.*, 42:59–66, 1971.
- [4] O. Carrillo-Negrete. Modelación de la frecuencia e intensidad de ciclones tropicales usando el proceso Poisson no homogéneo. *Tesis de maestría sin editar. Colegio de Postgraduados. Campus Montecillo*, 2010.
- [5] G. Casella and R.L. Berger. *Statistical inference*. Duxbury Press, 1990.
- [6] S. Chhikara and J. Leroy. Estimation of the inverse gaussian distribution function. *American Statistical Association*, 69(345):250–254, 1974.
- [7] V. Choulakian, R.A. Lockhart, and M.A. Stephens. Cramér-von Mises statistics for discrete distributions. *Can. Jour. Statist.*, 22:125–137, 1994.
- [8] W.J. Conover. *Practical Nonparametric Statistics, 3rd Edition*. Wiley Series in Probability and Statistics, 1999.
- [9] P.C. Consul and H.C. Gupta. The generalized negative binomial distribution and its characterization by zero regression. *Siam J. Appl. Math.*, 39(2), 1980.
- [10] H. Cramér. *Mathematical methods of statistics*. Princeton University Press, Princeton, 1945.

- [11] R.B. D'Agostino and M. Stephens. *Goodness-of-fit techniques*, volume 68. Statistics: a Series of Textbooks and Monographs, 1986.
- [12] W.L. Deemer and D. F. Votaw. Estimation of parameters of truncated or censored exponential distributions. *Ann. Math. Statist.*, 26:498–504, 1955.
- [13] J. Durbin. Distribution theory for tests based on the sample distribution function. *Siam J. Appl. Math.*, 1973.
- [14] S. Engen and M. Lillegard. Stochastic simulations conditioned on sufficient statistics. *Biometrika*, 84:235–240, 1997.
- [15] F. Famoye. Parameter estimation for generalized negative binomial distribution. *Comm. Statist. Sim. Comp.*, 26:269–279, 1997.
- [16] J.M. González-Barrios, F. O'Reilly, and R. Rueda. Goodness of fit for discrete random variables using the conditional density. *Metrika*, 64:77–94, 2006.
- [17] J.M. González-Barrios, F. O'Reilly, and R. Rueda. Durbin's random substitution and conditional monte carlo. *Metrika*, 72(3):369–383, 2010.
- [18] R.C. Gupta. Modified power series distribution and some of its applications. *The Indian Journal of Statistics.*, 36(3):288–298, 1974.
- [19] G.C. Jain and P.C. Consult. A generalization of negative binomial distribution. *Siam J. Appl. Math.*, 24:501–513, 1971.
- [20] B.H. Lindqvist, G. Taraldsen, M. Lillegard, and S. Engen. A counterexample to a claim about stochastic simulations. *Biometrika*, 90:489–490, 2003.
- [21] B.H. Lindqvist and G. Tarslidsen. Monte carlo conditioning on a sufficient statistic. *Biometrika*, 92(2):451–464, 2005.
- [22] R. Lockhart and F. O'Reilly. A note on Moore's conjecture. *Statistics and Probability Letters*, 74:212–220, 2005.

- [23] R.A. Lockhart, J.J. Spinelli, and M.A. Stephens. Cramér-von Mises statistics for discrete distributions with unknown parameters. *Can. Jour. Statisti.*, 35(1):125–133, 2007.
- [24] L. López-Martínez. Una prueba de razón de verosimilitudes para discriminar entre la distribución Poisson, binomial y binomial negativa. *Tesis de maestría sin editar. Colegio de Postgraduados. Campus Montecillo*, 2010.
- [25] M. Marques and V. Pérez-Abreu. Law of large numbers and central limit theorem of the empirical probability generating function of stationary random sequences and processes. *Sociedad Matemática Mexicana, Aportaciones Matemáticas*, 4:98–107, 1988.
- [26] A.M. Mood, F.A. Graybill, and D.C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, 1974.
- [27] F. O’Reilly. Algunas consideraciones sobre la inferencia estadística. *Ciencia*, 41(1):29–43, 1990.
- [28] F. O’Reilly and L. Gracia Medrano. On the conditional distribution of goodness of fit tests. *Comm. Statist. Theo. Meth.*, 35(3):541–549, 2006.
- [29] F. O’Reilly and L. Gracia Medrano. Exact goodness-of-fit tests based on sufficiency. *Springer. International Encyclopedia of Statistical Science. M. Lovric*, 2010.
- [30] F. O’Reilly and R. Rueda. Goodness of fit for the inverse gaussian distribution. *Can. Jour. Statisti.*, 20(4):387–397, 1992.
- [31] F. O’Reilly and R. Rueda. Tests of fit for discrete distributions based on the probability generating function. *Comm. Statist. Sim. Comp.*, 28(1):259–274, 1999.
- [32] F. O’Reilly and R. Rueda. Fiducial inferences for the truncated exponential distribution. *Comm. Statist. Theo. Meth.*, 36:2207–2212, 2007.
- [33] R. Rueda, V. Pérez-Abreu, and F. O’Reilly. Goodness of fit for the Poisson distribution based on the probability generating function. *Comm. Statist. Theo. Meth.*, 20:3093–3110, 1991.

- [34] Y. Sathe and S. Varde. Minimum variance unbiased estimation of reliability for the truncated exponential distribution. *Technometrics*, 11(3), 1969.
- [35] J. Spinelli and M.A Stephens. Cramér-von mises tests of fit for the Poisson distribution. *Can. Jour. Statisti.*, 25(2):257–268, 1997.
- [36] S. Villers. Programa para cálculo de p-values de estadísticas EDF para pruebas de bondad de ajuste. *Tesis de maestría sin editar. Universidad Autónoma de México.*, 2011.

Apéndice

A.1.

Tabla 4.5: Número de mosquitos de malaria en etapa esporozoíto positiva capturados con trampas de luz.

No. of sporozoite-positive mosquitoies in LTC	No. of times observed
0	761
1	21
2	5
3	0
4	2
5	0

Tabla 4.6: Número de ciclones tropicales de categoría 4 y 5 ocurridos en el Atlántico Norte.

Año	Número de Ciclones
1980	1
1981	1
1982	1
1983	0
1984	1
1985	1
1986	0
1987	0
1988	2
1989	2
1990	0
1991	1
1992	1
1993	0
1994	0
1995	3
1996	2
1997	0
1998	2
1999	4
2000	1
2001	1
2002	1
2003	2
2004	3
2005	4
2006	0
2007	2
2008	2
2009	1