



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS MATEMÁTICAS

BONDAD DE AJUSTE PARA LA
DISTRIBUCIÓN GAUSIANA INVERSA:
ESTUDIO DE POTENCIAS

TESIS

QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS

PRESENTA:

JENNER ALBERTO FRANCO TREJO

DIRECTOR DE LA TESIS: DR. FEDERICO JORGE O'REILLY TOGNO

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

MÉXICO, D.F. MARZO DE 2013



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Introducción

En bondad de ajuste se desarrollan técnicas para evaluar que tan bien una muestra de datos se ajusta a una distribución dada. En el marco formal de pruebas de hipótesis, la hipótesis nula H_0 es que una variable aleatoria dada X sigue una ley de probabilidades enunciada $F(x)$ (por ejemplo, la distribución normal o la distribución Weibull); la variable aleatoria puede venir de un proceso que está bajo investigación.

La hipótesis nula H_0 puede ser una hipótesis simple, cuando $F(x)$ está completamente especificada, por ejemplo, normal con media $\mu = 100$ y desviación estandar $\sigma = 10$; o H_0 puede dar una especificación incompleta y entonces será una hipótesis compuesta, por ejemplo, cuando enuncia sólo que $F(x)$ es normal con μ y σ no especificados.

Este trabajo está dedicado a la presentación y comparación de métodos de bondad de ajuste para la hipótesis compuesta nula Gaussian inversa (GI). Más específicamente, nuestro interés está en comparar los métodos de dos enfoques diferentes. Las técnicas de ajuste del primer enfoque comparten la característica de que están asociadas al concepto de entropía. Dentro de este enfoque encontramos la prueba de bondad de ajuste desarrollada por Vexler et al. (2011). El trabajo de estos autores fue la motivación para iniciar el desarrollo de la presente investigación, ésto debido a que en su conclusión mencionaron que la potencia de su prueba resultó mayor que las potencias de otras pruebas conocidas para un conjunto de distribuciones alternativas específicas. Es decir, en el problema de bondad de ajuste compuesto $H_0: X \sim \text{GI}(\mu, \lambda)$ vs $H_1: X$ sigue la ley de probabilidades $F(x)$ donde $F(x)$ es una función de distribución específica distinta a la GI, la prueba de Vexler et al. (2011) tuvo mayor capacidad

para rechazar la hipótesis nula cuando la alternativa era verdadera. La potencia de una prueba para una alternativa específica puede estimarse de la siguiente manera, generamos una cantidad suficientemente grande de muestras de la alternativa en cuestión, una vez obtenidas las muestras aplicamos la prueba de interés a cada una, la proporción de veces que la prueba rechaza la hipótesis nula resulta ser la potencia estimada de la prueba. Formalmente la potencia de una prueba de bondad de ajuste es la probabilidad de rechazar la hipótesis nula cuando la alternativa es verdadera.

A pesar de que la afirmación que hicieron los autores es cierta, desafortunadamente en la comparación que realizaron, no incluyeron pruebas para la GI del enfoque más utilizado en literatura para probar el ajuste de una familia paramétrica. Las pruebas de bondad de ajuste de este enfoque, se basan en la función de distribución empírica, y tienen un desarrollo teórico muy sólido que las sustenta. La función de distribución empírica (o EDF por sus siglas en inglés) es una función escalonada calculada de la muestra que siempre estima la función de distribución verdadera, cualquiera que esta sea. De la función de distribución empírica se derivan las estadísticas EDF (funcionales de la EDF) entre las que destacan la Kolmogorov, D ; la Cramér-von Mises, W^2 ; la de Kuiper, V ; la de Watson, U^2 ; y la de Anderson-Darling, A^2 . Éstas miden la discrepancia entre la función de distribución empírica y la función de distribución hipotética. Las estadísticas EDF se utilizan para probar el ajuste de la muestra a la distribución hipotética (o viceversa).

Debido a que el trabajo de Vexler et al. (2011) nos pareció incompleto, decidimos hacer una comparación de potencias más extensa. Para esto, presentamos dos métodos de bondad de ajuste para la distribución GI asociados al segundo enfoque mencionado. Uno de estos métodos presentados es el de O'Reilly y Rueda (1992), en donde la prueba se basa en la distribución asintótica de la estadística A^2 . La segunda técnica presentada es la correspondiente al trabajo de O'Reilly y Gracia-Medrano (2006). Aquí, los autores propusieron obtener el p-value condicional "exacto" de la prueba, dado el valor de la estadística suficiente minimal para los parámetros. Lo anterior es factible ya que podemos obtener una muestra de la distribución condicional de la estadística A^2 mediante un procedimiento descrito por los autores en su

trabajo. Las potencias que presentaremos para las dos técnicas de bondad de ajuste recién mencionadas se obtuvieron durante el desarrollo del presente trabajo. Éstas requieren un trabajo computacional relativamente intenso.

El capítulo uno lo dedicamos al estudio de una prueba de normalidad que fue introducida por Vasicek (1976). Ésta es una prueba basada en una caracterización de entropía. Se consideró importante presentarla ya que está fuertemente relacionada a las pruebas para la GI que se exponen en el capítulo dos. Consideramos que los lectores que no estén familiarizados con pruebas basadas en la entropía, encontrarán útil este capítulo para un entendimiento más rápido de la prueba desarrollada por Mudholkar y Tian (2002).

En el capítulo dos presentamos las pruebas para la GI de Mudholkar y Tian (2002) y Vexler et al. (2011). La prueba de Mudholkar y Tian (2002) es una prueba basada en la entropía muy similar a aquella desarrollado por Vasicek (1976). Explicamos la prueba de cociente de verosimilitudes empíricas basada en densidades de Vexler et al. (2011). En el proceso de obtención de la estadística de prueba de Vexler et al. (2011), se obtiene una estadística intermedia equivalente a la estadística de Mudholkar y Tian (2002), sin embargo, las estadísticas son obtenidas de maneras diferentes. Vexler et al. (2011) afirmaron que su estadística final de prueba es una mejora de la estadística de prueba basada en la entropía. También en este capítulo presentamos una comparación de potencias entre las pruebas asociadas a la entropía y otras pruebas conocidas que ya habían sido consideradas por Vexler et al. (2011). Finalmente, mostramos los resultados obtenidos por Vexler et al. (2011) para la bondad de ajuste de cuatro conjuntos de datos diferentes analizados por Folks y Chhikara (1978). Las pruebas de Mudholkar y Tian (2002) y Vexler et al. (2011) concuerdan en todos los casos al nivel de significancia del 5% en si se rechaza o no la supocición GI.

En el tercer capítulo ilustramos como llevar a cabo la prueba basada en la distribución asintótica de la estadística A^2 (O'Reilly y Rueda, 1992), así como la prueba basada en el p-value condicional "exacto" (O'Reilly y Gracia-Medrano, 2006). Explicamos que ambos procedimientos de prueba pueden llevarse a cabo de dos maneras,

que al menos desde un punto de vista muestral grande resultan equivalentes. Una forma es calculando la estadística A^2 usando el estimador “plug-in” para la función de distribución GI, es decir, estimamos los parámetros de la GI, por ejemplo, usando máxima verosimilitud, y los sustituimos en la función de distribución. La otra manera es obteniendo la estadística A^2 con la ayuda del estimador Rao-Blackwell para la función de distribución. Presentamos las potencias obtenidas con ambos métodos de prueba para las mismas alternativas consideradas por Vexler et al. (2011). Con base en una comparación de potencias, concluimos que las pruebas asociadas a la estadística A^2 resultan superiores a las pruebas basadas en la entropía. Al final de este capítulo evaluamos la bondad de ajuste GI usando la prueba basada en el p-value condicional “exacto” para los cuatro conjuntos de datos estudiados por Vexler et al. (2011) y hacemos una comparación de los resultados obtenidos.

Índice general

Introducción	i
1. Una Prueba de Normalidad Basada en la Entropía Muestral	1
1.1. Estimación de Vasicek de la Entropía	2
1.2. Consistencia del Estimador de Vasicek	8
1.3. Una Caracterización de la Distribución Normal	13
1.4. Prueba de Normalidad	15
2. Pruebas que Involucran la Entropía para la Gausiana Inversa	19
2.1. Prueba Basada en la Entropía Muestral	20
2.1.1. Una Caracterización de la Distribución GI	20
2.1.2. Prueba de Bondad de Ajuste para la Gausiana Inversa	21
2.2. Prueba Basada en el Cociente de Verosimilitudes Empíricas	27
2.2.1. La Técnica de la Verosimilitud Empírica Basada en la Densidad	27
2.2.2. Pruebas de Bondad de Ajuste para Hipótesis Compuestas	34
2.2.3. Prueba de Bondad de Ajuste para la Gausiana Inversa	35
3. Pruebas que Involucran la Estadística de Anderson-Darling para la Gausiana Inversa	47
3.1. Prueba Basada en la Distribución Asintótica de la Estadística A^2	48
3.1.1. Estimadores para la Función de Distribución y Bondad de Ajuste	48
3.1.2. Prueba de Bondad de Ajuste para la Gausiana Inversa	49
3.2. Prueba Basada en el p-value Condicional “Exacto”	53
3.2.1. Generación de Muestras “look-alike”	53
3.2.2. Prueba de Bondad de Ajuste para la Gausiana Inversa	57

4. Conclusiones	65
A. Algunos Resultados Elementales	67
B. Algunas Demostraciones	73
Bibliografía	87

Índice de figuras

1.1. Primera justificación del estimador de $\frac{d}{dp}F^{-1}(p)$	5
1.2. Segunda justificación del estimador de $\frac{d}{dp}F^{-1}(p)$	6

Índice de tablas

2.1.	Los valores críticos, $K_{m,n,\alpha}^*(f_Y)$, de la prueba propuesta al nivel de significancia $\alpha = 0.05$, i.e., $P_{H_0}\{K_{m,n}(f_Y) \leq K_{m,n,\alpha}^*(f_Y)\} = \alpha$	25
2.2.	Control del error de Tipo I de la prueba $K_{m,n}$ propuesta: $\alpha = 0.05$	25
2.3.	Una comparación de potencias empíricas para la prueba basada en la entropía con $\alpha = 0.05$	26
2.4.	Los valores críticos, $\log(TK_{n,\alpha}^*)$, de la prueba propuesta al nivel de significancia α , i.e., $P_{H_0}\{\log(TK_n) > \log(TK_{n,\alpha}^*)\} = \alpha$	41
2.5.	Los valores críticos, $\log(TK_{n,\alpha}^*)$, de la prueba propuesta al nivel de significancia α , i.e., $P_{H_0}\{\log(TK_n) > \log(TK_{n,\alpha}^*)\} = \alpha$	42
2.6.	Control del error de Tipo I de la prueba $\log(TK_n)$ propuesta: $\alpha = 0.05$	42
2.7.	Una comparación de potencias empíricas para la prueba de cociente de verosimilitudes empíricas basada en densidades con $\alpha = 0.05$	43
2.8.	Prueba para la GI basada en los datos introducidos por Folks y Chhikara (1978) . Nivel de significancia $\alpha = 0.05$. Comparación de la prueba de Vexler et al. (2011) contra la prueba de Mudholkar y Tian (2002)	45
3.1.	Distribución asintótica para A^2 . $\alpha =$ nivel de significancia (cola superior); $\theta = \lambda/\mu$	52
3.2.	Una comparación de potencias empíricas para la prueba basada en la distribución asintótica de la estadística A^2 , con el nivel de significancia $\alpha = 0.05$. La estadística A^2 se obtuvo usando el estimador “plug-in” para la función de distribución GI.	54

-
- 3.3. Una comparación de potencias empíricas para la prueba basada en la distribución asintótica de la estadística A^2 , con el nivel de significancia $\alpha = 0.05$. La estadística A^2 se obtuvo usando el estimador Rao-Blackwell para la función de distribución GI. 55
- 3.4. Una comparación de potencias para la prueba basada en el p-value condicional “exacto”, con el nivel de significancia $\alpha = 0.05$. La estadística A^2 se obtuvo usando el estimador “plug-in” para la función de distribución GI. 60
- 3.5. Una comparación de potencias para la prueba basada en el p-value condicional “exacto”, con el nivel de significancia $\alpha = 0.05$. La estadística A^2 se obtuvo usando el estimador Rao-Blackwell para la función de distribución GI. 61
- 3.6. Prueba para la GI basada en los datos introducidos por Folks y Chhikara (1978). Nivel de significancia $\alpha = 0.05$. Comparación de la prueba de O’Reilly y Gracia-Medrano (2006) contra las pruebas de Vexler et al. (2011) y Mudholkar y Tian (2002). La estadística A^2 se obtuvo usando el estimador “plug-in” para la función de distribución GI. 62
- 3.7. Prueba para la GI basada en los datos introducidos por Folks y Chhikara (1978). Nivel de significancia $\alpha = 0.05$. Comparación de la prueba de O’Reilly y Gracia-Medrano (2006) contra las pruebas de Vexler et al. (2011) y Mudholkar y Tian (2002). La estadística A^2 se obtuvo usando el estimador Rao-Blackwell para la función de distribución GI. 63

Capítulo 1

Una Prueba de Normalidad Basada en la Entropía Muestral

El presente capítulo está dedicado a una prueba de normalidad que fue introducida por Vasicek (1976). Aunque el objetivo de este trabajo es hacer una comparación de potencias para distintas pruebas de bondad de ajuste para la distribución GI, se consideró importante introducir la prueba de normalidad basada en la entropía (Vasicek, 1976), ya que está fuertemente relacionada a las pruebas que se exponen en el capítulo dos. En la primera sección recordamos la definición matemática de la entropía de una distribución. Una vez definida la entropía, procedemos a hacer un desarrollo detallado para obtener un estimador de ésta. El estimador de la entropía de Vasicek (1976), es de vital importancia para la prueba de normalidad, puesto que la estadística de prueba se basa fundamentalmente en éste. La segunda sección está dedicada a argumentar la consistencia del estimador obtenido para la entropía. Tal consistencia fue utilizada por Vasicek (1976) para demostrar que la potencia de su prueba converge a 1 si el tamaño de la muestra, n , tiende a infinito. En la tercera sección se demuestra que entre todas las distribuciones que poseen una densidad y tienen una varianza dada, la entropía es maximizada por la distribución normal. Tal resultado fue probado primero por Shannon (1949) y fue utilizado por Vasicek (1976) para introducir su prueba de normalidad. La cuarta y última sección explica cómo se construye la estadística de prueba de Vasicek (1976) así como el procedimiento de prueba.

1.1. Estimación de Vasicek de la Entropía

La entropía fue introducida en teoría de la información como una medida de incertidumbre (Shannon, 1948) y se define como sigue.

Definición 1.1. *La entropía de una distribución F con una función de densidad de probabilidad (fdp) f está definida como*

$$H(f) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx. \quad (1.1)$$

Claramente

$$H(f) = \int_{-\infty}^{\infty} \log\left\{\frac{1}{f(x)}\right\} f(x) dx,$$

por lo que si hacemos $x = F^{-1}(p)$ tenemos

$$F(x) = p, \quad f(x) dx = dp,$$

cuando $x \rightarrow \infty$, $F(x) = p \rightarrow 1$,

cuando $x \rightarrow -\infty$, $F(x) = p \rightarrow 0$, y así

$$\begin{aligned} H(f) &= \int_{-\infty}^{\infty} \log\left\{\frac{1}{f(x)}\right\} f(x) dx \\ &= \int_0^1 \log\left\{\frac{1}{f(F^{-1}(p))}\right\} dp \\ &= \int_0^1 \log\left\{\frac{d}{dp} F^{-1}(p)\right\} dp, \end{aligned} \quad (\text{por (A.1)})$$

es decir, la entropía (1.1) se puede expresar en la forma

$$H(f) = \int_0^1 \log\left\{\frac{d}{dp}F^{-1}(p)\right\}dp. \quad (1.2)$$

Enseguida construiremos un estimador de (1.2) siguiendo la metodología de Vasicek (1976). Sea X_1, X_2, \dots, X_n , $n \geq 3$, una muestra de la distribución F . Sabemos que la función de distribución empírica F_n , no es una función uno a uno, pero si suponemos que

$$F_n(X_{(k)}) = \frac{k}{n} \Rightarrow F_n^{-1}\left(\frac{k}{n}\right) = X_{(k)},$$

entonces tenemos que para $(i-1)/n < p \leq i/n$,

$$\begin{aligned} \frac{d}{dp}F^{-1}(p) &\approx \frac{F^{-1}\left(\frac{i+m}{n}\right) - F^{-1}\left(\frac{i-m}{n}\right)}{\frac{i+m}{n} - \frac{i-m}{n}} && \text{(por (A.2))} \\ &\approx \frac{F_n^{-1}\left(\frac{i+m}{n}\right) - F_n^{-1}\left(\frac{i-m}{n}\right)}{\frac{i+m}{n} - \frac{i-m}{n}} && \text{(porque } F_n \approx F) \\ &= \frac{X_{(i+m)} - X_{(i-m)}}{\frac{2m}{n}}, \end{aligned}$$

es decir,

$$\frac{d}{dp}F^{-1}(p) \approx \frac{n}{2m}(X_{(i+m)} - X_{(i-m)}),$$

donde $i = m+1, \dots, n-m$ y $m < n/2$.

Una justificación geométrica de la idea anterior está dada en la Figura 1.1.

Además, para $p \leq m/n$ tenemos

$$\frac{d}{dp}F^{-1}(p) \approx \frac{n}{2m}(X_{(i+m)} - X_{(1)}),$$

y para $p > (n - m)/n$

$$\frac{d}{dp}F^{-1}(p) \approx \frac{n}{2m}(X_{(n)} - X_{(i-m)}).$$

La Figura 1.2 ayuda a entender el por qué de las aproximaciones dadas.

Por las ideas recién expresadas encontramos que

$$\log\left\{\frac{d}{dp}F^{-1}(p)\right\} \approx \log\left\{\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})\right\},$$

donde $X_{(i)} = X_{(1)}$, si $i < 1$, y $X_{(i)} = X_{(n)}$, si $i > n$, y así, un estimador H_{mn} de $H(f)$ está dado por

$$H_{mn} = n^{-1} \sum_{i=1}^n \log\left\{\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})\right\}, \quad (1.3)$$

donde $X_{(i)} = X_{(1)}$, si $i < 1$, y $X_{(i)} = X_{(n)}$, si $i > n$.

A continuación investigamos el comportamiento del estimador (1.3). En virtud de (A.2) tenemos

$$F'(X_{(i)}) \approx \frac{F(X_{(i+m)}) - F(X_{(i-m)})}{X_{(i+m)} - X_{(i-m)}},$$

es decir,

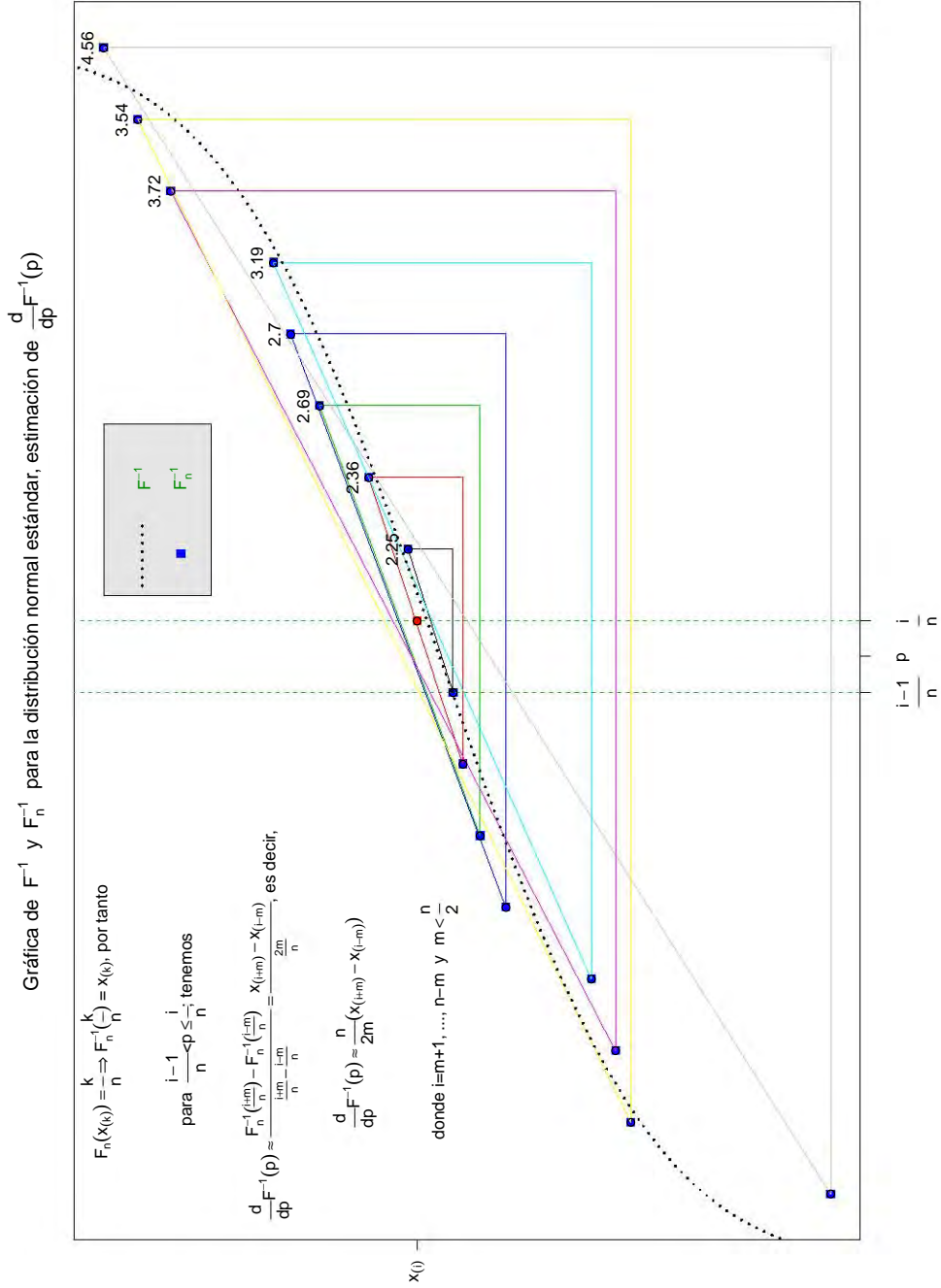


Figura 1.1 Se tomó una muestra aleatoria de tamaño $n = 17$ de la distribución normal estándar para generar los puntos $(\frac{i}{n}, F_n^{-1}(\frac{i}{n}))$. Luego se calculó $\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})$ para $i = 9$ fijo y $m = 1, \dots, 8$.

Gráfica de F^{-1} y F_n^{-1} para la distribución normal estándar, estimación de $\frac{d}{dp}F^{-1}(p)$

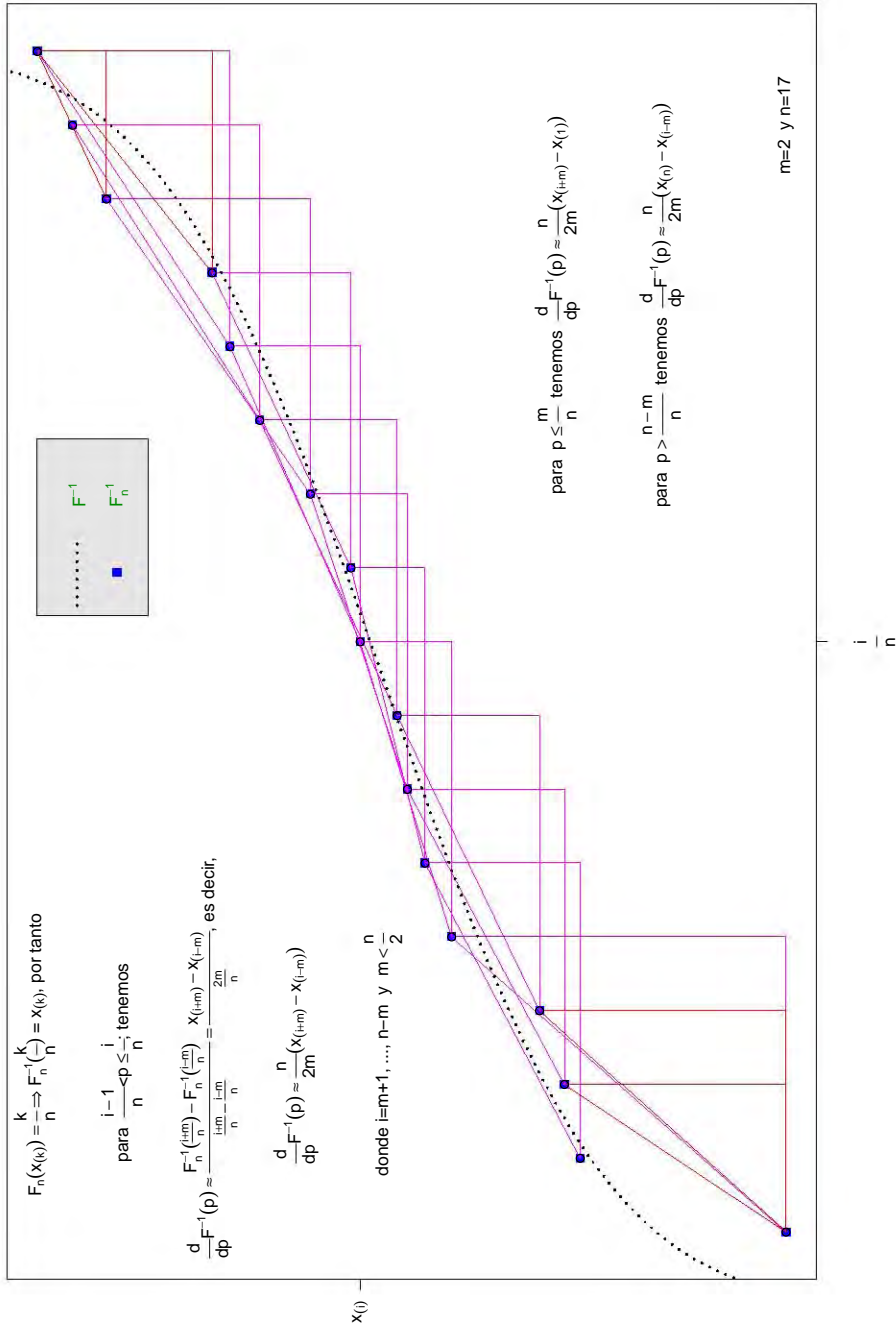


Figura 1.2 Se tomó una muestra aleatoria de tamaño $n = 17$ de la distribución normal estándar para generar los puntos $(\frac{i}{n}, F_n^{-1}(\frac{i}{n}))$. Para $m = 2$, las alturas y bases de los triángulos en color fucsia están dadas por $X_{(i+m)} - X_{(i-m)}$ y $\frac{2m}{n}$ respectivamente donde $i = 3, \dots, 15$.

$$f(X_{(i)}) \approx \frac{F(X_{(i+m)}) - F(X_{(i-m)})}{X_{(i+m)} - X_{(i-m)}},$$

o bien,

$$1 \approx \frac{F(X_{(i+m)}) - F(X_{(i-m)})}{f(X_{(i)})\{X_{(i+m)} - X_{(i-m)}\}}.$$

Así,

$$\begin{aligned} H_{mn} &= n^{-1} \sum_{i=1}^n \log\left\{\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})\right\} \\ &\approx n^{-1} \sum_{i=1}^n \log\left\{\frac{n}{2m}(X_{(i+m)} - X_{(i-m)}) \cdot \frac{F(X_{(i+m)}) - F(X_{(i-m)})}{f(X_{(i)})\{X_{(i+m)} - X_{(i-m)}\}} \cdot \frac{F(X_{(i+m)}) - F(X_{(i-m)})}{f(X_{(i)})\{X_{(i+m)} - X_{(i-m)}\}}\right\} \\ &= n^{-1} \sum_{i=1}^n \log\left\{\frac{1}{f(X_{(i)})} \cdot \frac{F(X_{(i+m)}) - F(X_{(i-m)})}{f(X_{(i)})\{X_{(i+m)} - X_{(i-m)}\}} \cdot \frac{n}{2m}\{F(X_{(i+m)}) - F(X_{(i-m)})\}\right\} \\ &= n^{-1} \sum_{i=1}^n \log\left[\frac{1}{f(X_{(i)})}\right] + n^{-1} \sum_{i=1}^n \log\left[\frac{F(X_{(i+m)}) - F(X_{(i-m)})}{f(X_{(i)})\{X_{(i+m)} - X_{(i-m)}\}}\right] + \\ &\quad n^{-1} \sum_{i=1}^n \log\left[\frac{n}{2m}\{F(X_{(i+m)}) - F(X_{(i-m)})\}\right], \end{aligned}$$

haciendo

$$\begin{aligned} V_{mn} &= n^{-1} \sum_{i=1}^n \log\left[\frac{F(X_{(i+m)}) - F(X_{(i-m)})}{f(X_{(i)})\{X_{(i+m)} - X_{(i-m)}\}}\right], \\ U_{mn} &= n^{-1} \sum_{i=1}^n \log\left[\frac{n}{2m}\{F(X_{(i+m)}) - F(X_{(i-m)})\}\right], \end{aligned}$$

tenemos

$$H_{mn} \approx -n^{-1} \sum_{i=1}^n \log f(X_i) + V_{mn} + U_{mn}. \quad (1.4)$$

El primer término en el lado derecho de (1.4) no depende de m y representa la media muestral estimada de $H(f) = E\{-\log f(x)\}$ asumiendo que el valor de f en los puntos X_1, X_2, \dots, X_n es conocido. Si la varianza de $-\log f(x)$ es finita, este es el estimador insesgado de varianza mínima de $H(f)$ dado los valores de f en los puntos muestrales. Los dos términos restantes representan dos fuentes de error de estimación adicional. El término V_{mn} es debido a estimación de f por diferencias finitas. Para n fijo, su efecto decrece con valores decrecientes de m . El término U_{mn} corresponde a el error debido a estimar incrementos de F por incrementos de F_n . Los incrementos son tomados sobre intervalos $(X_{(i-m)}, X_{(i+m)})$ cuya longitud crece con m , y por lo tanto la perturbación debida a U_{mn} es más pequeña mientras más grande es el valor de m . Cuando $n \rightarrow \infty$, reducción simultanea del efecto de estos dos términos de ruido requiere que $m \rightarrow \infty$, $m/n \rightarrow 0$. Una elección óptima de m para un n dado, sin embargo, depende de la distribución (que es desconocida) de F . En general, mientras más suave es la densidad de F , más grande es tal valor óptimo de m .

1.2. Consistencia del Estimador de Vasicek

Antes de demostrar que (1.3) es un estimador consistente de la entropía demostraremos el siguiente Lema.

Lema 1.1. *La variable aleatoria U_{mn} converge a cero en probabilidad cuando $n \rightarrow \infty$, $m \rightarrow \infty$, $m/n \rightarrow 0$.*

Demostración. Sea $\epsilon > 0$ arbitraria pero fija. Luego,

$$\begin{aligned}
0 &\leq P(|U_{mn}| \geq \epsilon) \\
&\leq \frac{E(|U_{mn}|)}{\epsilon} && \text{(por la Desigualdad de Chebychev)} \\
&= -\frac{E(U_{mn})}{\epsilon} && \text{(por la Proposición B.3(a))} \\
&\rightarrow 0 && \text{(por la Proposición B.3(c))}
\end{aligned}$$

cuando $n \rightarrow \infty$, $m \rightarrow \infty$, $m/n \rightarrow 0$. Es decir,

$$\lim_{n \rightarrow \infty, m \rightarrow \infty} P(|U_{mn} - 0| \geq \epsilon) = 0.$$

□

Ya que $F(X_{(1)}), F(X_{(2)}), \dots, F(X_{(n)})$ se distribuye como una muestra ordenada de tamaño n de la distribución uniforme sobre $(0, 1)$, la distribución de U_{mn} no depende de F , así, el sesgo debido a la presencia de U_{mn} en (1.4) puede ser eliminado al usar

$$H'_{mn} = H_{mn} - E(U_{mn}),$$

en lugar de H_{mn} como un estimador de la entropía. El siguiente teorema afirma la consistencia de H'_{mn} (y por el Lema 1.1 también la de H_{mn}).

Teorema 1.1. *Sea X_1, X_2, \dots, X_n una muestra de una distribución F con una fdp f y una varianza finita. Entonces*

$$H'_{mn} \xrightarrow{P} H(f) \quad \text{cuando } n \rightarrow \infty, m \rightarrow \infty \text{ y } m/n \rightarrow 0.$$

Demostración. Primero notemos que

$$\begin{aligned}
H_{mn} &= \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right\} \\
&= \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{X_{(i+m)} - X_{(i-m)}}{F(X_{(i+m)}) - F(X_{(i-m)})} \frac{n}{2m} \{F(X_{(i+m)}) - F(X_{(i-m)})\} \right\} \\
&= \frac{1}{n} \sum_{i=1}^n \log \left[\frac{X_{(i+m)} - X_{(i-m)}}{F(X_{(i+m)}) - F(X_{(i-m)})} \right] + \\
&\quad \frac{1}{n} \sum_{i=1}^n \log \left[\frac{n}{2m} \{F(X_{(i+m)}) - F(X_{(i-m)})\} \right] \\
&= \frac{1}{n} \sum_{i=1}^n \log \left[\frac{X_{(i+m)} - X_{(i-m)}}{F(X_{(i+m)}) - F(X_{(i-m)})} \right] + \\
&\quad U_{mn} \\
&= -(2m)^{-1} \sum_{i=1}^n \log \left[\frac{F(X_{(i+m)}) - F(X_{(i-m)})}{X_{(i+m)} - X_{(i-m)}} \right] \frac{2m}{n} + \\
&\quad U_{mn},
\end{aligned}$$

es decir,

$$\begin{aligned}
H_{mn} &= -(2m)^{-1} \sum_{i=1}^n \log \left[\frac{F(X_{(i+m)}) - F(X_{(i-m)})}{X_{(i+m)} - X_{(i-m)}} \right] \{F_n(X_{(i+m)}) - F_n(X_{(i-m)})\} + \\
&\quad U_{mn}.
\end{aligned}$$

Ahora, las $2m$ clases de residuos módulo $2m$ representadas por los residuos $j = 1, 2, \dots, 2m$ están dadas por

$$\begin{aligned}
C_1 &= \{ \dots, 1 - 3 \cdot 2m, 1 - 2 \cdot 2m, 1 - 1 \cdot 2m, 1, 1 + 1 \cdot 2m, 1 + 2 \cdot 2m, 1 + 3 \cdot 2m, \dots \} \\
C_2 &= \{ \dots, 2 - 3 \cdot 2m, 2 - 2 \cdot 2m, 2 - 1 \cdot 2m, 2, 2 + 1 \cdot 2m, 2 + 2 \cdot 2m, 2 + 3 \cdot 2m, \dots \} \\
&\quad \vdots \\
C_{2m} &= \{ \dots, 2m - 3 \cdot 2m, 2m - 2 \cdot 2m, 2m - 1 \cdot 2m, 2m, 2m + 1 \cdot 2m, 2m + 2 \cdot 2m, \dots \}
\end{aligned}$$

Ya que las $2m$ clases de residuos anteriores forman una partición de los enteros, si hacemos

$$S_j = - \sum_{i \in C_j} \log \left[\frac{F(X_{(i+m)}) - F(X_{(i-m)})}{X_{(i+m)} - X_{(i-m)}} \right] \{F_n(X_{(i+m)}) - F_n(X_{(i-m)})\}$$

tenemos que

$$\begin{aligned} H_{mn} &= -(2m)^{-1} \sum_{i=1}^n \log \left[\frac{F(X_{(i+m)}) - F(X_{(i-m)})}{X_{(i+m)} - X_{(i-m)}} \right] \{F_n(X_{(i+m)}) - F_n(X_{(i-m)})\} + \\ &\quad U_{mn} \\ &= (2m)^{-1} \sum_{j=1}^{2m} S_j + U_{mn}, \end{aligned}$$

luego,

$$H'_{mn} = (2m)^{-1} \sum_{j=1}^{2m} S_j + U_{mn} - E(U_{mn}).$$

Cuando $X_{(i-m)}, X_{(i+m)}$ pertenecen a un intervalo en el cual $f(x)$ es positiva y continua, entonces existe un valor $X'_i \in (X_{(i-m)}, X_{(i+m)})$ tal que

$$\frac{F(X_{(i+m)}) - F(X_{(i-m)})}{X_{(i+m)} - X_{(i-m)}} = f(X'_i),$$

así,

$$S_j = - \sum_{i \in C_j} \log [f(X'_i)] \{F_n(X_{(i+m)}) - F_n(X_{(i-m)})\},$$

es decir, S_j es una suma de Stieltjes de la función $-\log f(x)$ con respecto a la

medida F_n sobre la suma de intervalos de continuidad de f en los cuales $f(x) > 0$. Por el teorema de Glivenko–Cantelli $F_n(x) \rightarrow F(x)$ c.s. uniformemente sobre x , así cuando $n \rightarrow \infty$

$$\begin{aligned} S_j &= - \sum_{i \in C_j} \log[f(X'_i)] \{F_n(X_{(i+m)}) - F_n(X_{(i-m)})\} \\ &\rightarrow - \sum_{i \in C_j} \log[f(X'_i)] \{F(X_{(i+m)}) - F(X_{(i-m)})\} \quad \text{uniformemente c.s. sobre } C_j \\ &= - \sum_{i \in C_j} \log[f(X'_i)] \Delta F_{i,m}, \end{aligned}$$

donde $\Delta F_{i,m} = F(X_{(i+m)}) - F(X_{(i-m)})$. Ya que $X_{(i+m)} - X_{(i-m)} \rightarrow 0$ c.s. cuando $m/n \rightarrow 0$,

$$\begin{aligned} S_j &= - \sum_{i \in C_j} \log[f(X'_i)] \Delta F_{i,m} \\ &\rightarrow - \int_{-\infty}^{\infty} [\log f(x)] dF(x) \quad \text{uniformemente c.s. sobre } C_j \\ &= - \int_{-\infty}^{\infty} [\log f(x)] f(x) dx \\ &= - \int_{-\infty}^{\infty} f(x) \log f(x) dx \\ &= H(f), \end{aligned}$$

cuando $n, m \rightarrow \infty$ y $m/n \rightarrow 0$.

Como

$$S_j \rightarrow H(f) \quad \text{uniformemente c.s. sobre } C_j,$$

tenemos que

$$(2m)^{-1} \sum_{j=1}^{2m} S_j \rightarrow H(f) \quad \text{c.s.} \quad (1.5)$$

Es claro que

$$\begin{aligned} H_{mn} &= (2m)^{-1} \sum_{j=1}^{2m} S_j + U_{mn} \\ &\xrightarrow{P} H(f) \\ &\text{cuando } n \rightarrow \infty, m \rightarrow \infty \text{ y } m/n \rightarrow 0, \quad (\text{por (1.5) y el Lema 1.1}) \end{aligned}$$

se sigue que

$$\begin{aligned} H'_{mn} &= H_{mn} - E(U_{mn}) \\ &= [(2m)^{-1} \sum_{j=1}^{2m} S_j + U_{mn}] - E(U_{mn}) \\ &\xrightarrow{P} H(f) \quad \text{cuando } n \rightarrow \infty, m \rightarrow \infty \text{ y } m/n \rightarrow 0. \end{aligned}$$

□

1.3. Una Caracterización de la Distribución Normal

Proposición 1.1. *Si F es una distribución con fdp f y varianza σ^2 finita, entonces la entropía de F es finita o bien $-\infty$.*

Demostración. Si f^* y g son funciones de densidad de probabilidad, entonces

$$\int_{-\infty}^{\infty} f^*(x) \log \left[\frac{g(x)}{f^*(x)} \right] dx \leq 0,$$

(Lehmann y Casella, 1998, p. 47). Luego, para f tenemos

$$\int_{-\infty}^{\infty} f(x)[\log g(x) - \log f(x)]dx \leq 0,$$

es decir,

$$-\int_{-\infty}^{\infty} f(x) \log f(x)dx \leq -\int_{-\infty}^{\infty} f(x) \log g(x)dx,$$

o bien,

$$H(f) \leq -\int_{-\infty}^{\infty} f(x) \log g(x)dx.$$

Sea g la fdp normal con la misma media y varianza que f , es decir,

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\},$$

por tanto

$$\begin{aligned} H(f) &\leq -\int_{-\infty}^{\infty} f(x) \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}\right]dx \\ &= -\int_{-\infty}^{\infty} f(x)\left[-\log \sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(x - \mu)^2\right]dx \\ &= \log \sqrt{2\pi\sigma^2} + \frac{1}{2\sigma^2} \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx \\ &= \log \sqrt{2\pi\sigma^2} + \frac{1}{2} \\ &= \log \sqrt{2\pi\sigma^2} + \log \sqrt{e} \\ &= \log\{\sqrt{2\pi e}\sigma\}. \end{aligned}$$

En virtud de la desigualdad

$$H(f) \leq \log\{\sqrt{2\pi e}\sigma\} \quad (1.6)$$

concluimos que $H(f)$ es o bien finita o bien $-\infty$. □

Corolario 1.1. *Entre todas las distribuciones que poseen una función de densidad f y tienen una varianza dada σ^2 , la entropía $H(f)$ es maximizada por la distribución normal. La entropía de la distribución normal con varianza σ^2 es $\log\{\sqrt{2\pi e}\sigma\}$.*

1.4. Prueba de Normalidad

Vasicek (1976) se basó en la propiedad del Corolario 1.1 para introducir una prueba de la hipótesis compuesta de normalidad. Para tal propósito utilizó el estimador H_{mn} . Enseguida ilustraremos la metodología seguida por Vasicek (1976).

Sean X_1, X_2, \dots, X_n una muestra de una distribución F que tiene una fdp f y una varianza σ^2 finita, y $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ las estadísticas de orden. Sea m un entero positivo menor que $n/2$ y haga $X_{(i)} = X_{(1)}$ para $i < 1$, $X_{(i)} = X_{(n)}$ para $i > n$. Recordemos que

$$\begin{aligned} H_{mn} &= n^{-1} \sum_{i=1}^n \log\left\{\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})\right\} \\ &= \sum_{i=1}^n \log\left\{\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})\right\}^{1/n} \\ &= \log\left[\prod_{i=1}^n \left\{\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})\right\}^{1/n}\right] \\ &= \log\left[\frac{n}{2m} \left\{\prod_{i=1}^n (X_{(i+m)} - X_{(i-m)})\right\}^{1/n}\right], \end{aligned}$$

y definamos

$$\begin{aligned}
K_{mn} &= \frac{1}{s} \exp(H_{mn}) \\
&= \frac{1}{s} \exp\left(\log\left[\frac{n}{2m} \left\{ \prod_{i=1}^n (X_{(i+m)} - X_{(i-m)}) \right\}^{1/n}\right]\right) \\
&= \frac{n}{2ms} \left\{ \prod_{i=1}^n (X_{(i+m)} - X_{(i-m)}) \right\}^{1/n},
\end{aligned}$$

donde $s^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

La prueba K_m de la hipótesis compuesta de normalidad introducida por Vasicek (1976), es la prueba con región crítica

$$K_{mn} \leq K_{mn\alpha}^*, \quad (1.7)$$

donde $K_{mn\alpha}^*$ es el valor crítico de K_{mn} al nivel de significancia α . Sabemos que

$$s \xrightarrow{P} \sigma$$

cuando $n \rightarrow \infty$, y

$$\exp(H_{mn}) \xrightarrow{P} \exp(H(f)), \quad (\text{por el Teorema 1.1})$$

cuando $n \rightarrow \infty$, $m \rightarrow \infty$, $m/n \rightarrow 0$, luego,

$$\frac{1}{s} \exp(H_{mn}) \xrightarrow{P} \frac{1}{\sigma} \exp(H(f)), \quad (\text{por el Teorema A.5})$$

cuando $n \rightarrow \infty$, $m \rightarrow \infty$, $m/n \rightarrow 0$, es decir,

$$K_{mn} \xrightarrow{P} \frac{1}{\sigma} \exp(H(f)). \quad (1.8)$$

Ahora bien, bajo la hipótesis nula tenemos que

$$\begin{aligned} K_{mn} &\xrightarrow{P} \frac{1}{\sigma} \exp(H(f)) \\ &= \frac{1}{\sigma} \exp(\log\{\sqrt{2\pi e}\sigma\}) && \text{(por el Corolario 1.1)} \\ &= \sqrt{2\pi e}, \end{aligned}$$

y bajo una distribución alternativa

$$\begin{aligned} K_{mn} &\xrightarrow{P} \frac{1}{\sigma} \exp(H(f)) \\ &< \sqrt{2\pi e}. && \text{(por el Corolario 1.1)} \end{aligned}$$

Esto significa que la prueba K_m es consistente para tales alternativas. No hay necesidad sin embargo, para restringir el uso de la prueba a distribuciones con una densidad y un segundo momento finito, como se establece en el siguiente teorema cuya demostración puede encontrarse en Vasicek (1976).

Teorema 1.2. *La prueba K_m de cualquier tamaño $\alpha > 0$ es una prueba consistente, cuando $n \rightarrow \infty$, $m \rightarrow \infty$, $m/n \rightarrow 0$, para todas las alternativas sin una parte continua singular.*

Capítulo 2

Pruebas que Involucran la Entropía para la Gaussiana Inversa

El presente capítulo se divide en dos secciones, en la primera sección presentamos la prueba de bondad de ajuste basada en la entropía de [Mudholkar y Tian \(2002\)](#) para la Gaussiana inversa. Para desarrollar esta prueba, los autores obtuvieron una caracterización basada en la entropía para la distribución GI. Tal caracterización es análoga a aquella de la distribución normal enunciada en el [Corolario 1.1](#), es decir, en el caso de la Gaussiana inversa no es la distribución misma con la que se alcanza la entropía máxima, sino la distribución de una transformación de la variable aleatoria original. Así, dada una muestra de variables aleatorias no-negativas i.i.d., para probar la bondad de ajuste a la distribución GI siguiendo la metodología de [Mudholkar y Tian \(2002\)](#), primero que nada debemos aplicar la transformación a la muestra, y luego realizar un procedimiento muy similar al introducido por [Vasicek \(1976\)](#) para probar normalidad. La estadística de prueba de [Mudholkar y Tian \(2002\)](#), al igual que la de [Vasicek \(1976\)](#), depende de un parámetro entero m , cuya única especificación es que debe ser menor que $n/2$, donde n es el tamaño de la muestra. En su artículo, los autores afirmaron que la elección óptima de m para un n dado, es aún un problema abierto. Sin embargo, basados en sus estudios de potencia, recomendaron $m = 2$ para $n = 10$, $m = 3$ para $n = 20$ y $m = 4$ para $n = 50$ como los valores óptimos. [Vasicek \(1976\)](#) hizo la misma recomendación.

La segunda sección de este capítulo, está dedicada a la prueba de cociente de verosimilitudes empíricas basada en densidades. Tal prueba fue desarrollada por Vexler et al. (2011). Es una prueba “libre de distribución” (distribution-free), en la cual su estadística aproxima la estadística de cociente de verosimilitudes paramétricas. En el proceso de obtención de la estadística de prueba de Vexler et al. (2011), se obtiene una estadística que al igual que la de Mudholkar y Tian (2002), depende de un parámetro entero m , cuya única restricción es que debe ser menor que $n/2$. Esta estadística intermedia es equivalente a la estadística de Mudholkar y Tian (2002), sin embargo, las estadísticas son obtenidas de maneras diferentes. Para obtener la estadística de cociente de verosimilitudes empíricas basada en densidades, Vexler et al. (2011) lograron eliminar la dependencia del parámetro entero m , de la estadística mencionada anteriormente, por tal razón, los autores consideran su estadística final de prueba como una mejora a la estadística basada en la entropía.

2.1. Prueba Basada en la Entropía Muestral

2.1.1. Una Caracterización de la Distribución GI

Comenzamos la presente sección definiendo un parámetro para variables aleatorias no-negativas, tal parámetro nos ayudará a definir el conjunto de distribuciones sobre el cual la distribución de la transformación adecuada de una variable aleatoria GI alcanza entropía máxima.

Definición 2.1. *Para cualquier variable aleatoria no-negativa Y , defina el parámetro ξ^2 por*

$$\xi^2(Y) = E(Y^2) - 1/E(Y^{-2}), \quad (2.1)$$

es decir, ξ^2 es la diferencia entre las medias aritmética y armónica de Y^2 .

La siguiente proposición afirma la existencia de un estimador consistente para ξ^2 .

Proposición 2.1. *Sean Y_1, Y_2, \dots, Y_n variables aleatorias no-negativas i.i.d., la estadística w^2 definida por*

$$w^2 = \frac{1}{n-1} \left[\sum_{i=1}^n Y_i^2 - n^2 \left(\sum_{i=1}^n Y_i^{-2} \right)^{-1} \right], \quad (2.2)$$

es un estimador consistente de ξ^2 .

El siguiente teorema cuya demostración puede consultarse en Mudholkar y Tian (2002) establece la caracterización basada en la entropía para la distribución Gaussiana inversa.

Teorema 2.1. *La variable aleatoria X con distribución $GI(\mu, \lambda)$ está caracterizada por la propiedad de que $1/\sqrt{X}$ alcanza entropía máxima entre todas las variables aleatorias Y no-negativas, absolutamente continuas sujetas a las restricciones*

$$\begin{aligned} E[Y^{-2}] &= \mu, \\ E[Y^2] &= 1/\mu + 1/\lambda. \end{aligned}$$

Corolario 2.1. *Suponga que X es la variable aleatoria $GI(\mu, \lambda)$, y sea $Y = 1/\sqrt{X}$. Entonces la entropía de Y con fdp (B.7) está dada por*

$$H_Y(f) = \log(\xi \sqrt{2\pi e}/2),$$

donde $\xi^2 = 1/\lambda = E(Y^2) - 1/E(Y^{-2})$.

Corolario 2.2. *La variable aleatoria X con distribución Gaussiana inversa está caracterizada por la propiedad de que $1/\sqrt{X}$ alcanza la entropía máxima entre todas las variables aleatorias Y no-negativas, absolutamente continuas con un valor dado en $\xi^2(Y) = E(Y^2) - 1/E(Y^{-2})$.*

2.1.2. Prueba de Bondad de Ajuste para la Gaussiana Inversa

Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con fdp $f(x)$ cuyo soporte es $(0, \infty)$, $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ las estadísticas de orden respectivas, y $Y_{(i)} =$

$1/\sqrt{X_{(n-i+1)}}$, $i = 1, \dots, n$. Denotemos por $H_{mn}(f_Y)$ al estimador de $H(f_Y)$ para la distribución de $Y = 1/\sqrt{X}$, es decir,

$$H_{mn}(f_Y) = n^{-1} \sum_{i=1}^n \log \left\{ \frac{n}{2m} (Y_{(i+m)} - Y_{(i-m)}) \right\},$$

donde $Y_{(i)} = Y_{(1)}$, si $i < 1$, y $Y_{(i)} = Y_{(n)}$, si $i > n$.

En virtud del Corolario 2.2, e imitando la metodología seguida por Vasicek (1976) para probar normalidad, Mudholkar y Tian (2002) proponen rechazar la hipótesis compuesta nula $H_0 : X \sim GI(\mu, \lambda)$ si

$$K_{m,n}(f_Y) \leq K_{m,n,\alpha}^*(f_Y), \quad (2.3)$$

donde

$$K_{m,n}(f_Y) = \exp(H_{mn}(f_Y))/(w/2), \quad (2.4)$$

y el valor crítico $K_{m,n,\alpha}^*(f_Y)$ está determinado por el nivel de significancia.

Ahora bien, sabemos que

$$2/w \xrightarrow{P} 2/\xi, \quad (\text{por la Proposición 2.1})$$

cuando $n \rightarrow \infty$, y que

$$\exp(H_{mn}(f_Y)) \xrightarrow{P} \exp(H(f_Y)) \quad (\text{por el Teorema 1.1})$$

cuando $n \rightarrow \infty$, $m \rightarrow \infty$ y $m/n \rightarrow 0$, en consecuencia

$$\exp(H_{mn}(f_Y))/(w/2) \xrightarrow{P} \exp(H(f_Y))/(\xi/2), \quad (\text{por el Teorema A.5})$$

es decir,

$$K_{m,n}(f_Y) \xrightarrow{P} \exp(H(f_Y))/(\xi/2). \quad (2.5)$$

Bajo la hipótesis nula tenemos que

$$\begin{aligned} K_{m,n}(f_Y) &\xrightarrow{P} \exp(H(f_Y))/(\xi/2) \\ &= \exp(\log(\xi\sqrt{2\pi e}/2))/(\xi/2) \quad (\text{por el Corolario 2.1}) \\ &= \sqrt{2\pi e} \end{aligned}$$

cuando $n \rightarrow \infty$, $m \rightarrow \infty$ y $m/n \rightarrow 0$.

Por otra parte, si X_1, X_2, \dots, X_n son de cualquier distribución alternativa con ξ^2 finita, entonces

$$\begin{aligned} K_{m,n}(f_Y) &\xrightarrow{P} \exp(H(f_Y))/(\xi/2) \\ &< \sqrt{2\pi e} \quad (\text{por el Corolario 2.2}) \end{aligned}$$

cuando $n \rightarrow \infty$, $m \rightarrow \infty$ y $m/n \rightarrow 0$.

Teorema 2.2. *La prueba basada en la entropía de la hipótesis compuesta Gaussiana inversa es consistente contra cualquier distribución alternativa concentrada sobre la recta real positiva y que no tiene parte singular.*

Distribución Nula

Mudholkar y Tian (2002) mencionan que es generalmente reconocido que incluso la distribución asintótica de la estadística de prueba $K_{m,n}$ (que incluye estimadores

de un parámetro de perturbación, por ejemplo, s de σ en el caso normal, y w de ξ en el caso GI), es analíticamente difícil. Por tal razón, deciden tabular los percentiles de Monte Carlo de la distribución nula, para luego poder implementar su prueba. Para obtener los percentiles de la distribución nula de $K_{m,n}$ realizan lo siguiente:

1. generan 10,000 muestras de tamaño n de la distribución GI(1,1),
2. con las muestras generadas, obtienen 10,000 valores de la estadística $K_{m,n}$ donde $m < n/2$,
3. finalmente, usan los valores generados de $K_{m,n}$ para determinar los valores críticos $K_{m,n,0.05}^*$.

Una selección de los resultados del experimento son presentados en la Tabla 2.1.

El control del error de Tipo I usando los percentiles del 5% de la estadística $K_{m,n}$, lo evaluán al simular muestras aleatorias de poblaciones GI. Una selección de los resultados obtenidos es presentada en la Tabla 2.2.

Propiedades de Potencia

Para entender las propiedades de potencia de la prueba de bondad de ajuste basada en la entropía para la distribución GI, Mudholkar y Tian (2002) realizaron el siguiente experimento de Monte Carlo: generaron 10,000 muestras de tamaño $n = 10, 20, 30$ de cada una de poblaciones estudiadas en Edgeman et al. (1988) y Edgeman (1990), específicamente

- exponencial con media 1,
- uniforme(0,1),
- Weibull(1,2) con parámetro de escala 1 y parámetro de forma 2,
- lognormal(0.5,1) con media e y desviación estándar $e\sqrt{e-1}$,

la estadística de prueba $K_{m,n}$ fue calculada para cada una de las muestras, y se registró la proporción de rechazos para H_0 al nivel de significancia 0.05. Una selección de las potencias estimadas que resultaron en comparación con los resultados disponibles está dada en la Tabla 2.3.

Tabla 2.1 Los valores^a críticos, $K_{m,n,\alpha}^*(f_Y)$, de la prueba propuesta al nivel de significancia $\alpha = 0.05$, i.e., $P_{H_0}\{K_{m,n}(f_Y) \leq K_{m,n,\alpha}^*(f_Y)\} = \alpha$.

n	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$
3	0.8477				
4	0.9426				
5	1.0896	1.5490			
6	1.2225	1.6813			
7	1.3778	1.7848	1.8568		
8	1.4787	1.8703	1.9872		
9	1.5776	1.9765	2.0554	2.0356	
10	1.6913	2.0678	2.1516	2.1433	
12	1.8367	2.2359	2.3001	2.2874	2.2347
14	1.9484	2.3719	2.4354	2.4150	2.3734
16	2.0616	2.4699	2.5569	2.5269	2.5058
18	2.1194	2.5452	2.6542	2.6393	2.5942
20	2.1951	2.6403	2.7333	2.7209	2.6843
25	2.3231	2.7843	2.8912	2.8919	2.8820
30	2.4272	2.8790	3.0080	3.0424	3.0231
35	2.4903	2.9613	3.0968	3.1308	3.1304
40	2.5604	3.0320	3.1682	3.2082	3.2044
45	2.5913	3.0801	3.2287	3.2735	3.2791
50	2.6286	3.1182	3.2636	3.3149	3.3341

^a Estimaciones por simulación basadas en 10,000 repeticiones.

Tabla 2.2 Control^a del error de Tipo I de la prueba $K_{m,n}$ propuesta: $\alpha = 0.05$.

Tamaño muestral	GI(1,0.5)	GI(1,2)	GI(1,4)	GI(1,8)
10	0.043	0.049	0.054	0.050
20	0.051	0.050	0.055	0.055
30	0.048	0.052	0.056	0.052

^a Estimaciones por simulación basadas en 10,000 repeticiones.

Tabla 2.3 Una comparación de potencias^a empíricas para la prueba basada en la entropía con $\alpha = 0.05$.

Distribución	n	$K_{m=2,n}$ ^b	$K_{m=3,n}$ ^b	$KS1$ ^c	$KS2$ ^d
Exponencial	10	0.210	—	0.262	0.280
	20	—	0.466	0.518	0.525
	30	—	0.665	0.654	0.668
Uniforme(0,1)	10	0.484	—	0.342	0.356
	20	—	0.880	0.616	0.630
	30	—	0.982	0.776	0.782
Weibull(1,2)	10	0.119	—	0.06	0.074
	20	—	0.260	0.128	0.080
	30	—	0.395	0.168	0.111
LogNor(0.5,1)	10	0.050	—	0.048	0.055
	20	—	0.057	0.068	0.080
	30	—	0.073	0.082	0.111

^a Estimaciones por simulación basadas en 10,000 repeticiones.

^b Prueba basada en la entropía.

^c Prueba de Kolmogorov-Smirnov modificada (Edgeman et al., 1988).

^d Prueba de Kolmogorov-Smirnov usando transformación (Edgeman, 1990).

2.2. Prueba Basada en el Cociente de Verosimilitudes Empíricas

2.2.1. La Técnica de la Verosimilitud Empírica Basada en la Densidad

Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución con fdp f completamente desconocida. Defina la función de verosimilitud por

$$L_f = \prod_{i=1}^n f(X_i). \quad (2.6)$$

Vexler y Gurevich (2010) introducen un método “libre de distribución” para estimar L_f , a éste lo llaman técnica de la verosimilitud empírica basada en la densidad. Vamos a aplicar la metodología de verosimilitud empírica máxima de Vexler y Gurevich (2010) para estimar (2.6), para este fin reescribimos (2.6) como

$$\begin{aligned} L_f &= \prod_{i=1}^n f(X_i) \\ &= \prod_{i=1}^n f(X_{(i)}) \\ &= \prod_{i=1}^n f_i, \quad \text{donde } f_i = f(X_{(i)}), \text{ para } i = 1, \dots, n. \end{aligned}$$

El método consiste en derivar valores de f_i , $i = 1, \dots, n$, que maximicen L_f y satisfagan una restricción empírica correspondiente a $\int f(u)du = 1$. Para conseguir tal restricción empírica defina

$$\Gamma_m = \frac{1}{2m} \sum_{j=1}^n \int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx, \quad (2.7)$$

por la Proposición B.1 tenemos

$$\begin{aligned}\Gamma_m &= \frac{1}{2m} \sum_{j=1}^n \int_{X_{(j-m)}}^{X_{(j+m)}} f(x) dx \\ &= \int_{X_{(1)}}^{X_{(n)}} f(x) dx - \frac{1}{2m} \sum_{k=1}^{m-1} (m-k) \left[\int_{X_{(n-k)}}^{X_{(n-k+1)}} f(x) dx + \int_{X_{(k)}}^{X_{(k+1)}} f(x) dx \right],\end{aligned}$$

y claramente

$$\Gamma_m \leq \int_{X_{(1)}}^{X_{(n)}} f(x) dx \leq \int_{-\infty}^{\infty} f(x) dx = 1. \quad (2.8)$$

También

$$\begin{aligned}\Gamma_m &= \int_{X_{(1)}}^{X_{(n)}} f(x) dx - \frac{1}{2m} \sum_{k=1}^{m-1} (m-k) [F(X_{(n-k+1)}) - F(X_{(n-k)}) + \\ &\quad F(X_{(k+1)}) - F(X_{(k)})] \\ &\approx \int_{X_{(1)}}^{X_{(n)}} f(x) dx - \frac{1}{2m} \sum_{k=1}^{m-1} (m-k) [F_n(X_{(n-k+1)}) - F_n(X_{(n-k)}) + \\ &\quad F_n(X_{(k+1)}) - F_n(X_{(k)})] \\ &= \int_{X_{(1)}}^{X_{(n)}} f(x) dx - \frac{1}{2m} \sum_{k=1}^{m-1} (m-k) \left[\frac{n-k+1}{n} - \frac{n-k}{n} + \frac{k+1}{n} - \frac{k}{n} \right] \\ &= \int_{X_{(1)}}^{X_{(n)}} f(x) dx - \frac{1}{2m} \frac{2}{n} \sum_{k=1}^{m-1} (m-k) \\ &= \int_{X_{(1)}}^{X_{(n)}} f(x) dx - \frac{1}{mn} [(m-1)m - \frac{m-1}{2}m] \\ &= \int_{X_{(1)}}^{X_{(n)}} f(x) dx - \left[\frac{m-1}{n} - \frac{m-1}{2n} \right] \\ &= \int_{X_{(1)}}^{X_{(n)}} f(x) dx - \frac{m-1}{2n} \leq 1 - \frac{m-1}{2n},\end{aligned}$$

por lo que podemos estimar empíricamente Γ_m via

$$\begin{aligned}\widehat{\Gamma}_m &= 1 - \frac{1}{n} - \frac{m-1}{2n} \\ &= F_n(X_{(n)}) - F_n(X_{(1)}) - \frac{m-1}{2n} \\ &\approx F(X_{(n)}) - F(X_{(1)}) - \frac{m-1}{2n} \\ &= \int_{X_{(1)}}^{X_{(n)}} f(x)dx - \frac{m-1}{2n} \\ &\approx \Gamma_m.\end{aligned}$$

Por lo tanto podemos obtener que $\Gamma_m \approx 1$ cuando $m, n \rightarrow \infty$ y $m/n \rightarrow 0$. Por simplicidad, aplicando el teorema de integración del valor medio, obtenemos la aproximación

$$\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx \approx (X_{(j+m)} - X_{(j-m)})f_j,$$

así

$$\Gamma_m = \frac{1}{2m} \sum_{j=1}^n \int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx \approx \frac{1}{2m} \sum_{j=1}^n f_j(X_{(j+m)} - X_{(j-m)}), \quad (2.9)$$

por lo que definiendo

$$\widetilde{\Gamma}_m = \frac{1}{2m} \sum_{j=1}^n f_j(X_{(j+m)} - X_{(j-m)}), \quad (2.10)$$

tenemos

$$\widetilde{\Gamma}_m \approx \Gamma_m. \quad (2.11)$$

Ahora podemos representar la condición (2.8) en la forma empírica

$$\tilde{\Gamma}_m \leq 1. \quad (2.12)$$

Encontraremos entonces los valores de f_1, \dots, f_n que maximizan $\log(L_f)$ bajo la restricción empírica

$$\tilde{\Gamma}_m = \frac{1}{2m} \sum_{j=1}^n f_j (X_{(j+m)} - X_{(j-m)}) = 1. \quad (2.13)$$

Para lograr lo anterior, según el teorema del multiplicador de Lagrange, debemos examinar los puntos críticos de la función

$$\sum_{j=1}^n \log(f_j) - \lambda \left[\frac{1}{2m} \sum_{j=1}^n f_j (X_{(j+m)} - X_{(j-m)}) - 1 \right],$$

donde λ es un multiplicador de Lagrange. Así, debemos resolver el sistema de ecuaciones simultaneas

$$\begin{aligned} \frac{1}{f_1} - \lambda \frac{1}{2m} (X_{(j+m)} - X_{(j-m)}) &= 0 \\ \frac{1}{f_2} - \lambda \frac{1}{2m} (X_{(j+m)} - X_{(j-m)}) &= 0 \\ &\vdots \\ \frac{1}{f_n} - \lambda \frac{1}{2m} (X_{(j+m)} - X_{(j-m)}) &= 0 \\ \frac{1}{2m} \sum_{j=1}^n f_j (X_{(j+m)} - X_{(j-m)}) &= 1 \end{aligned}$$

para f_1, \dots, f_n y λ .

Encontremos primero el valor de λ . Del sistema anterior tenemos

$$\begin{aligned}
 1 - \lambda \frac{1}{2m} f_1(X_{(j+m)} - X_{(j-m)}) &= 0 && \text{se multiplicó por } f_1 \\
 1 - \lambda \frac{1}{2m} f_2(X_{(j+m)} - X_{(j-m)}) &= 0 && \text{se multiplicó por } f_2 \\
 &\vdots \\
 1 - \lambda \frac{1}{2m} f_n(X_{(j+m)} - X_{(j-m)}) &= 0 && \text{se multiplicó por } f_n \\
 \frac{1}{2m} \sum_{j=1}^n f_j(X_{(j+m)} - X_{(j-m)}) &= 1,
 \end{aligned}$$

sumando las primeras n ecuaciones tenemos

$$\begin{aligned}
 n - \lambda \frac{1}{2m} \sum_{j=1}^n f_j(X_{(j+m)} - X_{(j-m)}) &= 0 \\
 \frac{1}{2m} \sum_{j=1}^n f_j(X_{(j+m)} - X_{(j-m)}) &= 1,
 \end{aligned}$$

así,

$$\lambda = n.$$

Luego, de

$$\frac{1}{f_j} - \frac{n}{2m} (X_{(j+m)} - X_{(j-m)}) = 0,$$

obtenemos

$$f_j = \frac{2m}{n(X_{(j+m)} - X_{(j-m)})} \quad \text{para } j = 1, \dots, n.$$

Finalmente, el estimador de L_f , usando la técnica no-paramétrica de la verosimilitud empírica máxima basada en la densidad es

$$\tilde{L}_f = \prod_{j=1}^n \frac{2m}{n(X_{(j+m)} - X_{(j-m)})}. \quad (2.14)$$

Mejoramiento del Estimador \tilde{L}_f

Según Vexler et al. (2011), mejoran el estimador \tilde{L}_f al eliminar la dependencia del parámetro entero m . Su razonamiento, inspirado por Vexler y Gurevich (2010), se muestra enseguida. Sabemos que $\Gamma_m = \frac{1}{2m} \sum_{j=1}^n \int_{X_{(j-m)}}^{X_{(j+m)}} f(x) dx \leq 1$, para toda m , por lo tanto, por (2.11), podemos restringir los valores de f_j para que satisfagan $\tilde{\Gamma}_m = \frac{1}{2m} \sum_{j=1}^n f_j (X_{(j+m)} - X_{(j-m)}) \leq 1$ para toda m . Ahora bien, es claro que para cada m_0 fija se cumple

$$\{f_1, \dots, f_n : \tilde{\Gamma}_m \leq 1, \text{ para toda } m\} \subset \{f_1, \dots, f_n : \tilde{\Gamma}_{m_0} \leq 1\},$$

así

$$\max_{\{f_1, \dots, f_n : \tilde{\Gamma}_m \leq 1, \text{ para toda } m\}} \prod_{j=1}^n f_j \leq \max_{\{f_1, \dots, f_n : \tilde{\Gamma}_{m_0} \leq 1\}} \prod_{j=1}^n f_j,$$

y por lo tanto

$$\max_{\{f_1, \dots, f_n : \tilde{\Gamma}_m \leq 1, \text{ para toda } m\}} \prod_{j=1}^n f_j \leq \min_{m_0} \max_{\{f_1, \dots, f_n : \tilde{\Gamma}_{m_0} \leq 1\}} \prod_{j=1}^n f_j. \quad (2.15)$$

También, si $f_j, j = 1, \dots, n$ satisfacen $\tilde{\Gamma}_r \leq 1$, para alguna r , entonces

$$1 \geq \tilde{\Gamma}_r \approx \Gamma_r \approx \int_{Y(1)}^{Y(n)} f(y) dy \approx \Gamma_k \approx \tilde{\Gamma}_k,$$

y así, $\tilde{\Gamma}_k \lesssim 1$. Esto es, si f_1, \dots, f_n satisfacen $\tilde{\Gamma}_r \leq 1$, podemos esperar que f_1, \dots, f_n esten sujetas a $\tilde{\Gamma}_k \leq 1$, para $k \neq r$ también. Por lo tanto, para $k \neq r$ podemos esperar que

$$\{f_1, \dots, f_n : \tilde{\Gamma}_r \leq 1\} \approx \{f_1, \dots, f_n : \tilde{\Gamma}_k \leq 1\},$$

así

$$\{f_1, \dots, f_n : \tilde{\Gamma}_m \leq 1, \text{ para toda } m\} \approx \{f_1, \dots, f_n : \tilde{\Gamma}_{m_0} \leq 1, \text{ para alguna } m_0\},$$

luego

$$\max_{\{f_1, \dots, f_n : \tilde{\Gamma}_m \leq 1, \text{ para toda } m\}} \prod_{j=1}^n f_j \approx \max_{\{f_1, \dots, f_n : \tilde{\Gamma}_{m_0} \leq 1, \text{ para alguna } m_0\}} \prod_{j=1}^n f_j,$$

y ya que

$$\max_{\{f_1, \dots, f_n : \tilde{\Gamma}_{m_0} \leq 1, \text{ para alguna } m_0\}} \prod_{j=1}^n f_j \geq \min_{m_0} \max_{\{f_1, \dots, f_n : \tilde{\Gamma}_{m_0} \leq 1\}} \prod_{j=1}^n f_j,$$

tenemos

$$\max_{\{f_1, \dots, f_n : \tilde{\Gamma}_m \leq 1, \text{ para toda } m\}} \prod_{j=1}^n f_j \gtrsim \min_{m_0} \max_{\{f_1, \dots, f_n : \tilde{\Gamma}_{m_0} \leq 1\}} \prod_{j=1}^n f_j. \quad (2.16)$$

De las desigualdades (2.15) y (2.16) Vexler y Gurevich (2010) al igual que Vexler et al. (2011) concluyeron que la aproximación (2.14) a la verosimilitud L_f puede ser mejorada como

$$\min_m \max_{\{f_1, \dots, f_n: \tilde{\Gamma}_m \leq 1\}} \prod_{j=1}^n f_j,$$

por lo que proponen el estimador

$$\tilde{L}_f^{Mej} = \min_{1 \leq m < n^\delta} \prod_{j=1}^n \frac{2m}{n(X_{(j+m)} - X_{(j-m)})}, \quad \delta \in (0, 1) \quad (2.17)$$

para crear estadísticas de prueba que aproximen no paraméricamente cocientes de verosimilitudes paramétricos potentes en diferentes problemas estadísticos.

2.2.2. Pruebas de Bondad de Ajuste para Hipótesis Compuestas

Dada una muestra aleatoria X_1, X_2, \dots, X_n de una población con fdp f y una varianza finita, considere el problema de probar

$$H_0 : f = f_{H_0} \quad \text{vs} \quad H_1 : f = f_{H_1} \quad (2.18)$$

donde, bajo la hipótesis alternativa, f_{H_1} es completamente desconocida, mientras que, bajo la hipótesis nula, $f_{H_0}(x) = f_{H_0}(x; \boldsymbol{\theta})$ es conocida hasta el vector de parámetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ ($d \geq 1$). Si tanto la hipótesis nula H_0 como la alternativa H_1 estuviesen completamente especificadas y en consecuencia las verosimilitudes L_{H_0} y L_f bajo H_0 y H_1 respectivamente, fuesen totalmente conocidas, el Lema de Neyman-Pearson garantizaría que el cociente de verosimilitudes L_f/L_{H_0} es la estadística de prueba más potente. Motivados por esta idea, considerando (2.18), Vexler y Gurevich (2010) sugieren la estadística de prueba

$$G_n = \min_{1 \leq m < n^\delta} \frac{\tilde{L}_f^{Mej}}{\prod_{j=1}^n f_{H_0}(X_j, \hat{\theta})} = \min_{1 \leq m < n^\delta} \frac{\prod_{j=1}^n \frac{2m}{n(X_{(j+m)} - X_{(j-m)})}}{\prod_{j=1}^n f_{H_0}(X_j, \hat{\theta})}, \quad (2.19)$$

donde $\delta \in (0, 1)$ y $\hat{\theta}$ estima a θ (por ejemplo, $\hat{\theta}$ es el estimador de máxima verosimilitud de θ). Según Vexler y Gurevich (2010) al igual que Vexler et al. (2011), (2.19) es una estadística mejorada de la estadística

$$G_{mn} = \frac{\tilde{L}_f}{\prod_{j=1}^n f_{H_0}(X_j, \hat{\theta})} = \frac{\prod_{j=1}^n \frac{2m}{n(X_{(j+m)} - X_{(j-m)})}}{\prod_{j=1}^n f_{H_0}(X_j, \hat{\theta})}. \quad (2.20)$$

Vexler y Gurevich (2010) analizan pruebas basadas en (2.19) para normalidad y uniformidad, mientras que Vexler et al. (2011) se enfocan en pruebas para la hipótesis compuesta nula GI, este último caso es el que nos interesa en este trabajo.

2.2.3. Prueba de Bondad de Ajuste para la Gaussiana Inversa

Considere el problema de probar la hipótesis nula de que X_1, X_2, \dots, X_n vienen de una distribución GI con parámetros desconocidos μ y λ . Siguiendo la idea de Mudholkar y Tian (2002) expuesta en la sección anterior, Vexler et al. (2011) transforman las observaciones como $Y_{(i)} = 1/\sqrt{X_{(n-i+1)}}$, $i = 1, \dots, n$ (realizan la transformación para mostrar que su prueba es un mejoramiento de la prueba de Mudholkar y Tian (2002)). Por el Lema B.1, la hipótesis a probar es equivalente a la hipótesis H_0 de que Y_1, Y_2, \dots, Y_n vienen de una distribución con fdp de la forma

$$f_Y(y) = \frac{2}{\sqrt{2\pi\xi}} \exp\left(-\frac{(y - \nu/y)^2}{2\xi^2}\right), \quad y \geq 0,$$

donde, $\nu = 1/\mu$ y $\xi^2 = 1/\lambda$, es decir,

$$f_Y(y) = \left(\frac{2\lambda}{\pi}\right)^{1/2} \exp\left(\frac{-\lambda(y^{-1} - \mu y)^2}{2\mu^2}\right). \quad (2.21)$$

Es relativamente fácil demostrar que los estimadores de máxima verosimilitud $\hat{\mu}$ y $\hat{\lambda}$ de μ y λ son

$$\hat{\mu} = n^{-1} \sum_{i=1}^n Y_i^{-2} \quad (2.22)$$

y

$$\hat{\lambda} = n\hat{\mu}^2 \left[\sum_{i=1}^n (Y_i^{-1} - \hat{\mu}Y_i)^2 \right]^{-1}, \quad (2.23)$$

por lo que el estimador “plug-in” de la verosimilitud L_{H_0} está dado por

$$\begin{aligned} \hat{L}_{H_0} &= \max_{(\mu, \lambda)} \prod_{i=1}^n f_{H_0}(Y_{(i)} | (\mu, \lambda)) \\ &= \max_{(\mu, \lambda)} \prod_{i=1}^n \left(\frac{2\lambda}{\pi} \right)^{1/2} \exp\left(-\frac{\lambda(Y_{(i)}^{-1} - \mu Y_{(i)})^2}{2\mu^2} \right) \\ &= \max_{(\mu, \lambda)} \left(\frac{2\lambda}{\pi} \right)^{n/2} \exp\left(\frac{-\lambda}{2\mu^2} \sum_{i=1}^n (Y_{(i)}^{-1} - \mu Y_{(i)})^2 \right) \\ &= \left(\frac{2\hat{\lambda}}{\pi} \right)^{n/2} \exp\left(\frac{-\hat{\lambda}}{2\hat{\mu}^2} \sum_{i=1}^n (Y_{(i)}^{-1} - \hat{\mu}Y_{(i)})^2 \right) \\ &= \left(\frac{2\hat{\lambda}}{\pi} \right)^{n/2} \exp\left(\frac{-\hat{\lambda}}{2\hat{\mu}^2} \cdot \frac{n\hat{\mu}^2}{\hat{\lambda}} \right) \quad (\text{por (2.23)}) \\ &= \left(\frac{2\hat{\lambda}}{\pi} \right)^{n/2} \exp\left(-\frac{n}{2} \right) \\ &= \left(\frac{2\hat{\lambda}}{\pi e} \right)^{n/2}. \end{aligned}$$

El estimador de L_f obtenido siguiendo la metodología de verosimilitud empírica máxima basada en la densidad es

$$\tilde{L}_f = \prod_{i=1}^n \frac{2m}{n(Y_{(i+m)} - Y_{(i-m)})}, \quad (2.24)$$

cuya versión mejorada según el enfoque que estamos considerando es

$$\tilde{L}_f^{Mej} = \min_{1 \leq m < n^\delta} \prod_{j=1}^n \frac{2m}{n(Y_{(i+m)} - Y_{(i-m)})}, \quad \delta \in (0, 1). \quad (2.25)$$

En vista de lo anterior, Vexler et al. (2011) proponen la estadística de prueba

$$TK_n = \frac{\tilde{L}_f^{Mej}}{\hat{L}_{H_0}} = \frac{\min_{1 \leq m < n^\delta} \prod_{i=1}^n \frac{2m}{n(Y_{(i+m)} - Y_{(i-m)})}}{(\frac{2\hat{\lambda}}{\pi e})^{n/2}}, \quad \delta = 0.5. \quad (2.26)$$

para probar la hipótesis compuesta GI como un mejoramiento de la estadística

$$TK_{nm} = \frac{\tilde{L}_f}{\hat{L}_{H_0}} = \frac{\prod_{i=1}^n \frac{2m}{n(Y_{(i+m)} - Y_{(i-m)})}}{(\frac{2\hat{\lambda}}{\pi e})^{n/2}}. \quad (2.27)$$

La prueba de cociente de verosimilitudes empíricas basada en densidades propone rechazar la hipótesis compuesta nula GI si

$$\log(TK_n) > C, \quad (2.28)$$

donde C es un umbral de prueba.

Proposición 2.2. *Bajo H_0 , $n^{-1} \log(TK_n) \xrightarrow{P} 0$, mientras que, bajo H_1 , si $E[(\log f_{H_1}(Y_1))^2] < \infty$, entonces $n^{-1} \log(TK_n) \xrightarrow{P} E[\log(\frac{f_{H_1}(Y_1)}{f_{H_0}(Y_1; \mathbf{a})})] > 0$, cuando $n \rightarrow \infty$, donde $\mathbf{a} = (E(Y_1^{-2}), 1/[E(Y_1^2) - (E(Y_1^{-2}))^{-1}])$.*

La Proposición 2.2 demuestra que $P_{H_1}(\log(TK_n) > C_\alpha) \rightarrow 1$, donde C_α es un valor crítico que satisface el error de Tipo I α cuando $n \rightarrow \infty$. Por lo tanto, la prueba

propuesta es consistente (i.e., de potencia 1 asintóticamente).

Vexler et al. (2011) afirman que la estadística TK_{nm} es equivalente a la estadística $K_{m,n}(f_Y)$ propuesta por Mudholkar y Tian (2002) para la GI. En su artículo los autores presumen la estadística de prueba de Mudholkar y Tian (2002) como un producto intermedio de su método de cociente de verosimilitudes empíricas basada en densidades. Experimentos de Monte Carlo realizados en el presente trabajo determinaron que la prueba que rechaza la hipótesis compuesta nula si

$$\log(TK_{nm}) > C,$$

donde C es un umbral de prueba, tiene las mismas potencias empíricas que la prueba

$$K_{m,n}(f_Y) \leq K_{m,n,\alpha}^*(f_Y),$$

desarrollada por Mudholkar y Tian (2002) contra las mismas alternativas consideradas. Lo anterior nos conduce a pensar que es en este sentido que Vexler et al. (2011) afirman la equivalencia entre $K_{m,n}(f_Y)$ y TK_{nm} . Ahora bien, escriba (2.23) como

$$\hat{\lambda} = \left(\sum_{i=1}^n Y_i^2/n - n \left(\sum_{i=1}^n Y_i^{-2} \right)^{-1} \right)^{-1}, \quad (2.29)$$

luego

$$\begin{aligned}
TK_{nm} &= \frac{\prod_{i=1}^n \frac{2m}{n(Y_{(i+m)} - Y_{(i-m)})}}{\left(\frac{2\hat{\lambda}}{\pi e}\right)^{n/2}} \\
&= \frac{\prod_{i=1}^n \left(\frac{2m}{n(Y_{(i+m)} - Y_{(i-m)})}\right)^{-1/n}}{\left(\frac{2\hat{\lambda}}{\pi e}\right)^{-1/2}} \\
&= \frac{\prod_{i=1}^n \left(\frac{n(Y_{(i+m)} - Y_{(i-m)})}{2m}\right)^{1/n}}{(\hat{\lambda}-1)^{1/2} \left(\frac{2}{\pi e}\right)^{-1/2}} \\
&= \frac{\frac{n}{2m} \prod_{i=1}^n (Y_{(i+m)} - Y_{(i-m)})^{1/n}}{\left(\sum_{i=1}^n Y_i^2/n - n(\sum_{i=1}^n Y_i^{-2})^{-1}\right)^{1/2} \left(\frac{\pi e}{2}\right)^{1/2}} \quad (\text{sustituyendo (2.29)}) \\
&= \frac{\frac{n}{m} \prod_{i=1}^n (Y_{(i+m)} - Y_{(i-m)})^{1/n}}{2\left(\frac{\pi e}{2}\right)^{1/2} \left(\sum_{i=1}^n Y_i^2/n - n(\sum_{i=1}^n Y_i^{-2})^{-1}\right)^{1/2}}
\end{aligned}$$

mientras que

$$H_{mn}(f_Y) = \log\left[\frac{n}{2m} \prod_{i=1}^n (Y_{(i+m)} - Y_{(i-m)})^{1/n}\right],$$

y así

$$\begin{aligned}
K_{m,n}(f_Y) &= \exp(H_{mn}(f_Y))/(w/2) \\
&= \frac{2}{w} \exp\left(\log\left[\frac{n}{2m} \prod_{i=1}^n (Y_{(i+m)} - Y_{(i-m)})^{1/n}\right]\right) \\
&= \frac{n}{mw} \prod_{i=1}^n (Y_{(i+m)} - Y_{(i-m)})^{1/n} \\
&= \frac{\frac{n}{m} \prod_{i=1}^n (Y_{(i+m)} - Y_{(i-m)})^{1/n}}{\left(\sum_{i=1}^n Y_i^2/(n-1) - n^2(\sum_{i=1}^n Y_i^{-2})^{-1}/(n-1)\right)^{1/2}}.
\end{aligned}$$

De hecho, Vexler et al. (2011) afirman que

$$TK_{nm} = \frac{\frac{n}{2m} \prod_{i=1}^n (Y_{(i+m)} - Y_{(i-m)})^{1/n}}{(\sum_{i=1}^n Y_i^2/n - n(\sum_{i=1}^n Y_i^{-2})^{-1})^{1/2}/2},$$

(ver detalles en el Apéndice B de Vexler et al. (2011)).

Distribución Nula

Vexler et al. (2011) mencionan que es generalmente aceptado que la distribución asintótica de la estadística de prueba $\log(TK_n)$ (que incluye estimadores $\hat{\mu}$ y $\hat{\lambda}$ de los parámetros de perturbación μ y λ respectivamente, en el caso GI), es analíticamente difícil. Por lo anterior, y motivados por aplicaciones prácticas, deciden considerar valores críticos para tamaños muestrales fijos. Tabulan los percentiles de Monte Carlo de la distribución nula realizando el siguiente experimento:

1. generan 50,000 muestras de tamaño n de la distribución $GI(1,1)$,
2. con las muestras generadas, obtienen 50,000 valores de la estadística $\log(TK_n)$,
3. finalmente, usan los valores generados de la estadística de prueba $\log(TK_n)$ para determinar los valores críticos $\log(TK_{n,\alpha}^*)$, donde α representa el nivel de significancia.

Los resultados del experimento son presentados en las Tablas 2.4 y 2.5.

Para evaluar la exactitud de los valores críticos obtenidos, generan muestras aleatorias de poblaciones GI y estiman el error de Tipo I usando los percentiles del 5% de la estadística de prueba $\log(TK_n)$ ($\alpha = 0.05$). El control del error de Tipo I estimado se ilustra en la Tabla 2.6.

Propiedades de Potencia

Para investigar las propiedades de potencia de la prueba de bondad de ajuste GI propuesta, Vexler et al. (2011), realizaron el siguiente estudio de Monte Carlo. Generaron 10,000 muestras de tamaño $n = 10, 20, 30$ de las siguientes poblaciones:

Tabla 2.4 Los valores^a críticos, $\log(TK_{n,\alpha}^*)$, de la prueba propuesta al nivel de significancia α , i.e., $P_{H_0}\{\log(TK_n) > \log(TK_{n,\alpha}^*)\} = \alpha$.

Tamaño muestral	α							
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
10	7.1106	6.5853	6.2781	6.0392	5.8605	5.7016	5.5559	5.4365
15	8.4823	7.8319	7.4314	7.1291	6.9042	6.7207	6.5589	6.4104
20	9.1504	8.4316	7.9990	7.6968	7.4769	7.2840	7.1336	6.9831
25	9.6813	8.9327	8.5005	8.1984	7.9547	7.7582	7.5782	7.4365
30	10.3156	9.5426	9.1027	8.7600	8.4807	8.2622	8.0810	7.9190
35	10.7687	9.9797	9.5197	9.1735	8.9102	8.6944	8.4981	8.3395
40	11.1102	10.3427	9.8696	9.5152	9.2475	9.0397	8.8636	8.7066
45	11.4799	10.7279	10.2458	9.8953	9.6262	9.3990	9.2114	9.0504
50	11.7863	11.0235	10.5751	10.2299	9.9838	9.7577	9.5647	9.3945
55	12.2253	11.4291	10.9412	10.5814	10.3017	10.0863	9.8780	9.6964
60	12.4767	11.6485	11.1331	10.7850	10.5217	10.3099	10.1081	9.9422
65	12.7206	11.8821	11.4015	11.0314	10.7672	10.5457	10.3398	10.1695
70	13.0735	12.2269	11.7186	11.3325	11.0447	10.7979	10.5989	10.4137
75	13.3515	12.5373	12.0057	11.6373	11.3509	11.1066	10.9014	10.7156
80	13.6866	12.8380	12.3101	11.9135	11.5901	11.3451	11.1249	10.9333
85	13.8912	12.9608	12.5013	12.1090	11.8101	11.5501	11.3309	11.1470
90	14.1715	13.2555	12.7261	12.3453	12.0278	11.7836	11.5773	11.3801
95	14.3226	13.3931	12.8802	12.5106	12.2058	11.9523	11.7369	11.5396
100	14.5004	13.5990	13.0714	12.6594	12.3785	12.1138	11.8898	11.6932
120	15.3873	14.4181	13.8550	13.4787	13.1542	12.8996	12.6652	12.4551
150	16.3815	15.3858	14.8241	14.3910	14.0375	13.7557	13.5141	13.2955
200	17.6377	16.7383	16.1220	15.6491	15.3042	15.0077	14.7280	14.5083
250	19.0457	18.0246	17.3325	16.8719	16.4938	16.1581	15.8500	15.5890
300	20.0134	18.8813	18.1871	17.7225	17.3100	16.9705	16.6827	16.4113

^a Estimaciones por simulación basadas en 50,000 replicaciones de datos para cada n y α .

Tabla 2.5 Los valores^a críticos, $\log(TK_{n,\alpha}^*)$, de la prueba propuesta al nivel de significancia α , i.e., $P_{H_0}\{\log(TK_n) > \log(TK_{n,\alpha}^*)\} = \alpha$.

Tamaño muestral	α					
	0.09	0.1	0.2	0.25	0.3	0.5
10	5.3403	5.2464	4.5818	4.3537	4.1537	3.5341
15	6.2808	6.1572	5.3583	5.0968	4.8658	4.1954
20	6.8429	6.7233	5.8916	5.6168	5.3909	4.6884
25	7.3021	7.1786	6.3610	6.0822	5.8399	5.0826
30	7.7906	7.6603	6.8005	6.5054	6.2528	5.4606
35	8.2009	8.0782	7.1993	6.8813	6.6158	5.7813
40	8.5511	8.4212	7.4996	7.1925	6.9176	6.0574
45	8.9051	8.7682	7.8237	7.4973	7.2154	6.3309
50	9.2391	9.0930	8.1252	7.7896	7.5094	6.5824
55	9.5344	9.3925	8.3903	8.0455	7.7480	6.7965
60	9.7673	9.6370	8.6481	8.3002	8.0015	7.0258
65	10.0041	9.8645	8.8596	8.5034	8.1955	7.1992
70	10.2722	10.1242	9.1086	8.7367	8.4193	7.3990
75	10.5448	10.3932	9.3044	8.9402	8.6249	7.5767
80	10.7531	10.6066	9.5426	9.1530	8.8196	7.7530
85	10.9747	10.8349	9.7473	9.3604	9.0177	7.8937
90	11.2128	11.0538	9.9233	9.5282	9.1813	8.0560
95	11.3760	11.2151	10.0996	9.7067	9.3485	8.2065
100	11.5174	11.3670	10.2358	9.8300	9.4805	8.3179
120	12.2712	12.0964	10.8948	10.4702	10.0877	8.8112
150	13.1050	12.9199	11.6778	11.2222	10.8089	9.4228
200	14.3100	14.1224	12.7396	12.2268	11.7565	10.1912
250	15.3619	15.1536	13.6324	13.0803	12.5864	10.8283
300	16.1683	15.9364	14.3229	13.7106	13.1711	11.2812

^a Estimaciones por simulación basadas en 50,000 repeticiones de datos para cada n y α .

Tabla 2.6 Control^a del error de Tipo I de la prueba $\log(TK_n)$ propuesta: $\alpha = 0.05$.

Tamaño muestral	GI(1,0.5)	GI(1,2)	GI(1,4)	GI(1,8)
10	0.0445	0.0532	0.0601	0.0654
20	0.0424	0.0503	0.0541	0.0573
30	0.0486	0.0534	0.0524	0.0553
40	0.0504	0.0514	0.0529	0.0585
50	0.0446	0.0527	0.0520	0.0489

^a Estimaciones por simulación basadas en 10,000 repeticiones.

- exponencial con media 1,
- uniforme(0,1),
- Weibull(1,2) con parámetro de escala 1 y parámetro de forma 2,
- lognormal(0.5,1) con media e y desviación estándar $e\sqrt{e-1}$.

Por ejemplo, con las 10,000 muestras de tamaño 10, de la distribución $\exp(1)$, obtuvieron 10,000 replicaciones de la estadística $\log(TK_n)$, y posteriormente calcularon la proporción de replicaciones que excedieron el valor crítico 5.8605 ($\alpha = 0.05$) de la Tabla 2.4. De esta manera estimaron la potencia 0.2060 de la prueba de bondad de ajuste EL dada la alternativa ya mencionada (ver Tabla 2.7). La Tabla 2.7 muestra una comparación de estas potencias contra las de otras pruebas incluyendo la de Mudholkar y Tian (2002).

Tabla 2.7 Una comparación de potencias^a empíricas para la prueba de cociente de verosimilitudes empíricas basada en densidades con $\alpha = 0.05$.

Distribución	n	$K_{m=2,n}$	$K_{m=3,n}$	$K_{m=4,n}$	$K_{m=5,n}$	Z	KS1	KS2	TK_n
Exponencial	10	0.2075 ^b	0.1991	0.1499	0.0643	0.0206	0.262	0.280	0.2060
	20	0.4668	0.4675 ^b	0.4347	0.3781	0.0226	0.518	0.525	0.4636
	30	0.6347	0.6597 ^b	0.6563	0.6169	0.0328	0.654	0.668	0.6446
Uniforme(0,1)	10	0.4768 ^b	0.4759	0.3987	0.1857	0.1814	0.342	0.356	0.5078
	20	0.8682	0.8802 ^b	0.8645	0.8407	0.426	0.616	0.630	0.8826
	30	0.9687	0.9789 ^b	0.9833	0.9772	0.5764	0.776	0.782	0.9815
Weibull(1,2)	10	0.1251 ^b	0.1256	0.1064	0.0428	0.0721	0.06	0.074	0.1354
	20	0.2536	0.2707 ^b	0.2332	0.2105	0.1611	0.128	0.080	0.2565
	30	0.3564	0.3895 ^b	0.4028	0.3636	0.277	0.168	0.111	0.3847
LogNor(0.5,1)	10	0.0467 ^b	0.0464	0.0437	0.0318	0.0407	0.048	0.055	0.0460
	20	0.0588	0.0535 ^b	0.0441	0.0351	0.042	0.068	0.080	0.0541
	30	0.0707	0.0679 ^b	0.0591	0.0517	0.0397	0.082	0.111	0.0547

^a Estimaciones por simulación basadas en 10,000 replicaciones. $K_{m,n}$ -prueba basada en la entropía (Mudholkar y Tian, 2002)-; Z-prueba caracterizada por independencia (Mudholkar et al., 2001)-; KS1-prueba de Kolmogorov-Smirnov modificada (Edgeman et al., 1988)-; y KS2-prueba de Kolmogorov-Smirnov usando transformación (Edgeman, 1990)-.

^b Valores correspondientes a los m óptimos encontrados empíricamente por Mudholkar y Tian (2002) dadas alternativas conocidas.

Ejemplo

Como ejemplo para demostrar la aplicabilidad de la prueba de bondad de ajuste de cociente de verosimilitudes empíricas basada en densidades, Vexler et al. (2011) usan su método propuesto para evaluar la adecuación de la distribución GI a datos de cuatro diferentes estudios que fueron analizados en Folks y Chhikara (1978). Específicamente, los Conjuntos de datos llamados por Vexler et al. (2011) Conjunto de datos 1, 2, 3 y 4 fueron los siguientes.

Vida útil (en días) de un producto comestible:

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.

Resistencia a la fractura de la soldadura MIG (soldadura de metal con gas inerte):

54.4, 62.6, 63.2, 67.0, 70.2, 70.5, 70.6, 71.4, 71.8, 74.1,
74.1, 74.3, 78.8, 81.8, 83.0, 84.4, 85.3, 86.9, 87.3.

Precipitación (en pulgadas) en el Puente Jug de Maryland:

1.01, 1.11, 1.13, 1.15, 1.16, 1.17, 1.17, 1.20, 1.52, 1.54, 1.54, 1.57, 1.64,
1.73, 1.79, 2.09, 2.09, 2.57, 2.75, 2.93, 3.19, 3.54, 3.57, 5.11, 5.62.

Cantidad de escorrentía en el Puente Jug de Maryland:

0.17, 0.19, 0.23, 0.33, 0.39, 0.39, 0.40, 0.45, 0.52, 0.56, 0.59, 0.64, 0.66,
0.70, 0.76, 0.77, 0.78, 0.95, 0.97, 1.02, 1.12, 1.24, 1.59, 1.74, 2.92.

La Tabla 2.8 presenta los p-values obtenidos usando la prueba de bondad de ajuste EL y la prueba basada en la entropía de Mudholkar y Tian (2002) para los datos mencionados.

En base a los resultados de la Tabla 2.8, la prueba de Vexler et al. (2011) y la prueba de Mudholkar y Tian (2002) proporcionan las mismas conclusiones acerca de la prueba de bondad de ajuste de la distribución GI al nivel de significanci $\alpha = 0.05$. La distribución GI es rechazada para los Conjuntos de datos 1 y 3.

Tabla 2.8 Prueba para la GI basada en los datos introducidos por Folks y Chhikara (1978). Nivel de significancia $\alpha = 0.05$. Comparación de la prueba de Vexler et al. (2011) contra la prueba de Mudholkar y Tian (2002).

Conjunto de datos	Tamaño muestral	Prueba ^a	p-value
1	26	TK_{26}	0.0063
		$K_{m=2,26}$	0.0061
		$K_{m=3,26}$	0.0058
		$K_{m=4,26}$	0.0090
		$K_{m=5,26}$	0.0041
2	19	TK_{19}	0.1791
		$K_{m=2,19}$	0.1565
		$K_{m=3,19}$	0.1516
		$K_{m=4,19}$	0.2663
		$K_{m=5,19}$	0.3379
3	25	TK_{25}	0.0099
		$K_{m=2,25}$	0.0197
		$K_{m=3,25}$	0.0159
		$K_{m=4,25}$	0.0098
		$K_{m=5,25}$	0.0063
4	25	TK_{25}	0.9271
		$K_{m=2,25}$	0.9393
		$K_{m=3,25}$	0.9538
		$K_{m=4,25}$	0.8738
		$K_{m=5,25}$	0.8001

^a TK_n —prueba de cociente de verosimilitudes empíricas basada en densidades (Vexler et al., 2011)—; $K_{m,n}$ —prueba basada en la entropía (Mudholkar y Tian, 2002)—.

Capítulo 3

Pruebas que Involucran la Estadística de Anderson-Darling para la Gausiana Inversa

En este capítulo, consideramos el enfoque más utilizado en literatura para probar el ajuste de una familia paramétrica. Este enfoque, que se basa en la distribución empírica se utiliza para la Gausiana inversa; en particular para pruebas asociadas a la estadística de Anderson-Darling. Veremos que para las distribuciones alternativas que ya hemos analizado, las potencias de las pruebas obtenidas aquí, superan a las de las pruebas basadas en la entropía en la mayoría de los casos. La primera sección de este capítulo, la dedicamos al estudio de la prueba para la GI basada en la distribución asintótica de la A^2 . Esta prueba fue presentada por O'Reilly y Rueda (1992) quienes utilizaron el hecho de que cuando los parámetros de la GI son estimados por algún método eficiente asintóticamente, la distribución asintótica de la A^2 depende únicamente del cociente de los parámetros. A la luz de este resultado, los autores encontraron un conjunto de puntos porcentuales para la distribución asintótica de la A^2 para diferentes valores del cociente de parámetros (de la GI). El conjunto de valores críticos encontrados es suficiente para llevar a cabo pruebas para la GI.

En la segunda sección, introducimos una variante para realizar pruebas de ajuste a la GI que no utiliza ninguna aproximación asintótica. Esta técnica fue propuesta por

O'Reilly y Gracia-Medrano (2006) y la parte medular consiste en simular muestras denominadas “look-alike”, dada la muestra de interés. Una vez generadas las muestras “look-alike” se calcula la estadística de Anderson-Darling con cada una de ellas, así como con la muestra a la cual se le desea probar el ajuste. La proporción de A^2 's “look-alike” que exceden el valor de la estadística de Anderson-Darling obtenida con la muestra original representa el p-value condicional “exacto”. Lo anterior es cierto ya que bajo la hipótesis nula, los valores de A^2 de las muestras “look-alike”, forman una muestra de la distribución condicional de la estadística de Anderson-Darling. Esta sección incluye también una comparación final de potencias contra cada una de las pruebas para la GI estudiadas en este trabajo. Finalmente concluimos con una comparación de la prueba en cuestión contra las asociadas al concepto de entropía, tal comparación se basó en los resultados obtenidos al evaluar el ajuste de la distribución GI a los datos introducidos por Folks y Chhikara (1978).

3.1. Prueba Basada en la Distribución Asintótica de la Estadística A^2

Para probar el ajuste de la distribución Gausiana inversa con parámetros desconocidos, O'Reilly y Rueda (1992) estudian la distribución asintótica de la estadística de Anderson-Darling A^2 , cuando la muestra es de una Gausiana inversa. Los autores mencionan que de entre todas las estadísticas EDF, en su artículo sólo estudian la A^2 , debido principalmente a sus buenas propiedades de potencia reportadas en general.

3.1.1. Estimadores para la Función de Distribución y Bondad de Ajuste

Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con función de distribución $F(x, \theta)$, donde el vector de parámetros θ es desconocido. Asuma que T_n es la estadística suficiente minimal para θ .

Cuando $\hat{\theta}$ es un estimador apropiado para θ , siendo una función de T_n , $F_n(x, \hat{\theta})$ es llamado el estimador “plug-in” para la función de distribución. Mientras que

$\tilde{F}_n(x) = P(X_i \leq x | T_n)$ (donde i puede ser $1, \dots, n$ y el estimador es el mismo debido al hecho de que condicionalmente sobre T_n , las variables aleatorias X_1, X_2, \dots, X_n son intercambiables) es conocido como el estimador Rao-Blackwell para $F(x, \boldsymbol{\theta})$.

Las pruebas frecuentes de ajuste, en el caso compuesto, están basadas en el proceso empírico con parámetros estimados que es:

$$\hat{\xi}_n(x) = \sqrt{n}\{F_n(x) - F_n(x, \hat{\boldsymbol{\theta}})\},$$

donde $F_n(x)$ es la función de distribución empírica. Es decir, frecuentemente las pruebas son funcionales de $\hat{\xi}_n$. Sin embargo, es sabido que en muchos casos (incluido el caso de la GI), las pruebas de ajuste pueden ser construidas con base en el estimador Rao-Blackwell en vez del estimador “plug-in” usando el proceso

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

que a veces es denominado el proceso empírico Rao-Blackwelizado.

3.1.2. Prueba de Bondad de Ajuste para la Gaussiana Inversa

Sea X_1, X_2, \dots, X_n una muestra de una distribución cuya fdp es $h(x)$. La hipótesis a ser probada es que $h(x)$ es la densidad Gaussiana inversa $GI(x|\mu, \lambda)$, con μ y λ desconocidas.

La función de distribución Gaussiana inversa puede ser escrita como

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

donde $R = R(x; \mu, \lambda) = -\left(\frac{\lambda}{x}\right)^{\frac{1}{2}} + \left(\frac{\lambda x}{\mu^2}\right)^{\frac{1}{2}}$, $L = L(x; \mu, \lambda) = -\left(\frac{\lambda}{x}\right)^{\frac{1}{2}} - \left(\frac{\lambda x}{\mu^2}\right)^{\frac{1}{2}}$ y Φ es la función de distribución normal estándar.

El estimador Rao-Blackwell correspondiente, disponible en literatura, es

$$\tilde{F}_n(x) = \begin{cases} 0, & \text{si } x < l, \\ 1, & \text{si } x > u, \\ G_{n-2}(B) + \frac{n-2}{n} \left[1 + \frac{4(n-1)\hat{\lambda}}{n^2\hat{\mu}} \right]^{\frac{n-3}{2}} G_{n-2}(-B'), & \text{en otra parte,} \end{cases} \quad (3.1)$$

con $\hat{\mu}$ y $\hat{\lambda}$ los estimadores de máxima verosimilitud, ambos funciones de la estadística suficiente minimal

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

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

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

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

y l y u determinan el intervalo donde el argumento dentro de la raíz cuadrada que define a C es positivo.

Teóricamente, para obtener la estadística de Anderson-Darling, primero hay que mapear las observaciones en el intervalo $(0,1)$, y luego calcular A^2 con los puntos muestrales transformados. Es decir, se deben obtener

$$Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n)},$$

donde

$$Z_{(i)} = F(X_{(i)}; \mu, \lambda), \quad (3.2)$$

y calcular A^2 como

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) \{\log Z_{(i)} + \log(1 - Z_{(n-i+1)})\}. \quad (3.3)$$

Como μ y λ son desconocidos, en la práctica no es posible obtener (3.2). La solución común es estimar los parámetros y sustituir los estimadores en (3.2). Otra solución es reemplazar $F(x; \mu, \lambda)$ por el estimador Rao-Blackwell.

O'Reilly y Rueda (1992) demuestran formalmente que cuando μ y λ son estimados por máxima verosimilitud u otro método eficiente asintóticamente, la distribución asintótica de A^2 depende únicamente de $\theta = \lambda/\mu$. Los autores también demuestran que si en lugar de estimar y sustituir los parámetros, se usa el estimador Rao-Blackwell, la distribución asintótica para A^2 sigue siendo la misma. Los puntos porcentuales de la distribución asintótica para A^2 , encontrados por O'Reilly y Rueda (1992) para diferentes valores de θ , incluyendo los casos límite $\theta = 0$ y $\theta = \infty$, se muestran en la Tabla 3.1. Los puntos críticos varían suavemente con el logaritmo de θ , por lo tanto, una forma fácil y precisa de encontrar puntos críticos para valores de θ no incluidos en la Tabla 3.1 es interpolando.

La Prueba

Para probar la hipótesis compuesta nula GI, usando los valores críticos de la distribución asintótica para A^2 , dados en la Tabla 3.1, hacemos lo siguiente:

1. estimamos los parámetros μ y λ , usando máxima verosimilitud por ejemplo, y con los estimadores obtenemos $\hat{\theta} = \hat{\lambda}/\hat{\mu}$,
2. calculamos A^2 , o bien usando el estimador "plug-in", o bien usando el estimador Rao-Blackwell para la función de distribución Gaussiana inversa,

Tabla 3.1 Distribución asintótica para A^2 . $\alpha =$ nivel de significancia (cola superior); $\theta = \lambda/\mu$.

θ	Valor crítico				
	$\alpha=0.20$	0.10	0.05	0.025	0.01
0	0.90	1.19	1.50	1.83	2.27
2^{-10}	0.89	1.18	1.49	1.82	2.26
2^{-9}	0.89	1.18	1.48	1.81	2.24
2^{-8}	0.88	1.17	1.47	1.79	2.22
2^{-7}	0.86	1.15	1.45	1.76	2.19
2^{-6}	0.84	1.12	1.41	1.71	2.13
2^{-5}	0.81	1.07	1.35	1.64	2.03
2^{-4}	0.77	1.01	1.27	1.53	1.90
2^{-3}	0.72	0.93	1.16	1.40	1.73
2^{-2}	0.66	0.85	1.05	1.26	1.54
2^{-1}	0.61	0.78	0.95	1.13	1.37
2^0	0.57	0.72	0.87	1.02	1.23
2^1	0.54	0.68	0.82	0.96	1.14
2^2	0.53	0.66	0.79	0.92	1.09
2^3	0.52	0.65	0.77	0.90	1.07
2^4	0.52	0.64	0.76	0.89	1.05
2^5	0.51	0.64	0.76	0.88	1.05
∞	0.51	0.63	0.76	0.88	1.04

3. elegido un nivel de significancia α , interpolamos el valor crítico de la Tabla 3.1 linealmente en $\log \hat{\theta}$,
4. finalmente, rechazamos la hipótesis GI si $\log(A^2)$ es mayor que el valor crítico que se obtuvo por interpolación.

Propiedades de Potencia

En este trabajo, investigamos las propiedades de potencia de la prueba basada en la distribución asintótica de la estadística A^2 , para esto, realizamos el siguiente experimento de Monte Carlo. Generamos 10,000 muestras de tamaño $n = 10, 20, 30$ de las poblaciones consideradas por Vexler et al. (2011) y por Mudholkar y Tian (2002). Para cada población y cada tamaño muestral n , realizamos la prueba GI descrita anteriormente 10,000 veces, la proporción de rechazos obtenidos representa la potencia estimada de la prueba. Una comparación de potencias empíricas de la

prueba de O’Reilly y Rueda (1992) contra otras pruebas conocidas, incluyendo la de Vexler et al. (2011) y la de Mudholkar y Tian (2002), se muestra en las Tablas 3.2 y 3.3. Es clara la superioridad de la prueba de O’Reilly y Rueda (1992) en cuanto a potencia se refiere contra las alternativas ya mencionadas. Es importante señalar que Gracia-Medrano y O’Reilly (2004) ya habian comparado las potencias de la prueba de O’Reilly y Rueda (1992) contra las potencias de la prueba de Edgeman et al. (1988) para las alternativas consideradas y varios niveles de significancia, resultando el método de O’Reilly y Rueda (1992) con las potencias más altas.

3.2. Prueba Basada en el p-value Condicional “Exacto”

O’Reilly y Gracia-Medrano (2006) recomendaron el uso de la distribución condicional de la prueba de bondad de ajuste, dado el valor de la estadística suficiente minimal para los parámetros, en el problema de probar el ajuste de una distribución conocida sólo en su forma. En este contexto, ya que los parámetros no son de interés, y por tanto, considerados molestia, condicionar parece ser apropiado. Bajo este enfoque no hay necesidad de usar tablas, en vez de esto, se emplea un algoritmo basado en una simulación especial que produce el p-value condicional “exacto”. Por esta razón, en el caso continuo, se considera un procedimiento n -finito, de nivel α exacto. El procedimiento se puede usar también en el caso discreto, pero el nivel sería aproximado.

3.2.1. Generación de Muestras “look-alike”

Considere el problema de probar la hipótesis nula de que X_1, X_2, \dots, X_n vienen de una distribución $F(x, \theta)$ donde el vector de parámetros θ es desconocido. Sea T_n la estadística suficiente minimal para θ y $\tilde{F}_n(x)$ el estimador Rao-Blackwell para la función de distribución ($\tilde{F}_n(x) = P(X_i \leq x | T_n)$ donde i puede ser $1, \dots, n$ y el estimador es el mismo debido al hecho de que condicionalmente sobre T_n , las variables aleatorias X_1, X_2, \dots, X_n son intercambiables).

Tabla 3.2 Una comparación de potencias^a empíricas para la prueba basada en la distribución asintótica de la estadística A^2 , con el nivel de significancia $\alpha = 0.05$. La estadística A^2 se obtuvo usando el estimador “plug-in” para la función de distribución GI.

Distribución	n	$K_{m=2,n}$	$K_{m=3,n}$	$K_{m=4,n}$	$K_{m=5,n}$	Z	KS1	KS2	TK_n	A^2 O & R
Exponencial	10	0.2075 ^b	0.1991	0.1499	0.0643	0.0206	0.262	0.280	0.2060	0.3445
	20	0.4668	0.4675 ^b	0.4347	0.3781	0.0226	0.518	0.525	0.4636	0.6348
	30	0.6347	0.6597 ^b	0.6563	0.6169	0.0328	0.654	0.668	0.6446	0.7970
Uniforme(0,1)	10	0.4768 ^b	0.4759	0.3987	0.1857	0.1814	0.342	0.356	0.5078	0.5168
	20	0.8682	0.8802 ^b	0.8645	0.8407	0.426	0.616	0.630	0.8826	0.8673
	30	0.9687	0.9789 ^b	0.9833	0.9772	0.5764	0.776	0.782	0.9815	0.9711
Weibull(1,2)	10	0.1251 ^b	0.1256	0.1064	0.0428	0.0721	0.06	0.074	0.1354	0.1746
	20	0.2536	0.2707 ^b	0.2332	0.2105	0.1611	0.128	0.080	0.2565	0.3706
	30	0.3564	0.3895 ^b	0.4028	0.3636	0.277	0.168	0.111	0.3847	0.5297
LogNor(0.5,1)	10	0.0467 ^b	0.0464	0.0437	0.0318	0.0407	0.048	0.055	0.0460	0.0633
	20	0.0588	0.0535 ^b	0.0441	0.0351	0.042	0.068	0.080	0.0541	0.1099
	30	0.0707	0.0679 ^b	0.0591	0.0517	0.0397	0.082	0.111	0.0547	0.1451

^a Estimación por simulación basada en 10,000 repeticiones. A^2 O & R—prueba basada en la distribución asintótica de la estadística A^2 (O’Reilly y Rueda, 1992)-; TK_n —prueba de cociente de verosimilitudes empíricas basada en densidades (Vexler et al., 2011)-; $K_{m,n}$ —prueba basada en la entropía (Mudholkar y Tian, 2002)-; Z—prueba caracterizada por independencia (Mudholkar et al., 2001)-; $KS1$ —prueba de Kolmogorov-Smirnov modificada (Edgeman et al., 1988)-; y $KS2$ —prueba de Kolmogorov-Smirnov usando transformación (Edgeman, 1990)-.

^b Valores que corresponden a los m óptimos encontrados empíricamente por Mudholkar y Tian (2002) dadas alternativas conocidas.

Tabla 3.3 Una comparación de potencias^a empíricas para la prueba basada en la distribución asintótica de la estadística A^2 , con el nivel de significancia $\alpha = 0.05$. La estadística A^2 se obtuvo usando el estimador Rao-Blackwell para la función de distribución GI.

Distribución	n	$K_{m=2,n}$	$K_{m=3,n}$	$K_{m=4,n}$	$K_{m=5,n}$	Z	KS1	KS2	TK_n	A^2	O & R
Exponencial	10	0.2075 ^b	0.1991	0.1499	0.0643	0.0206	0.262	0.280	0.2060	0.3343	
	20	0.4668	0.4675 ^b	0.4347	0.3781	0.0226	0.518	0.525	0.4636	0.6235	
	30	0.6347	0.6597 ^b	0.6563	0.6169	0.0328	0.654	0.668	0.6446	0.8031	
Uniforme(0,1)	10	0.4768 ^b	0.4759	0.3987	0.1857	0.1814	0.342	0.356	0.5078	0.5131	
	20	0.8682	0.8802 ^b	0.8645	0.8407	0.426	0.616	0.630	0.8826	0.8587	
	30	0.9687	0.9789 ^b	0.9833	0.9772	0.5764	0.776	0.782	0.9815	0.9687	
Weibull(1,2)	10	0.1251 ^b	0.1256	0.1064	0.0428	0.0721	0.06	0.074	0.1354	0.2099	
	20	0.2536	0.2707 ^b	0.2332	0.2105	0.1611	0.128	0.080	0.2565	0.3891	
	30	0.3564	0.3895 ^b	0.4028	0.3636	0.277	0.168	0.111	0.3847	0.5384	
LogNor(0.5,1)	10	0.0467 ^b	0.0464	0.0437	0.0318	0.0407	0.048	0.055	0.0460	0.0845	
	20	0.0588	0.0535 ^b	0.0441	0.0351	0.042	0.068	0.080	0.0541	0.1199	
	30	0.0707	0.0679 ^b	0.0591	0.0517	0.0397	0.082	0.111	0.0547	0.1431	

^a Estimación por simulación basada en 10,000 repeticiones. A^2 O & R—prueba basada en la distribución asintótica de la estadística A^2 (O'Reilly y Rueda, 1992); TK_n —prueba de cociente de verosimilitudes empíricas basada en densidades (Vexler et al., 2011); $K_{m,r}$ —prueba basada en la entropía (Mudholkar y Tian, 2002); Z —prueba caracterizada por independencia (Mudholkar et al., 2001); $KS1$ —prueba de Kolmogorov-Smirnov modificada (Edgeman et al., 1988); y $KS2$ —prueba de Kolmogorov-Smirnov usando transformación (Edgeman, 1990).

^b Valores que corresponden a los m óptimos encontrados empíricamente por Mudholkar y Tian (2002) dadas alternativas conocidas.

En muchas aplicaciones la estadística T_n es lo que se denomina doblemente transitiva, que 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 en familias dentro de la Familia Exponencial Natural (FEN) donde T_n representa una suma de n variables aleatorias i.i.d. Si la estadística es doblemente transitiva y x_1, x_2, \dots, x_n una realización de X_1, X_2, \dots, X_n , entonces es posible generar una muestra $x_1^*, x_2^*, \dots, x_n^*$ que será una realización independiente condicionalmente dado $t_n = T_n(x_1, \dots, x_n)$ de la distribución condicional de X_1, X_2, \dots, X_n . La muestra generada cumplirá que $T_n(x_1^*, \dots, x_n^*) = T_n(x_1, \dots, x_n)$. O'Reilly y Gracia-Medrano (2006) llaman a $x_1^*, x_2^*, \dots, x_n^*$ muestra “look-alike” y es posible obtenerla siguiendo el procedimiento descrito en el Teorema 3.1. El procedimiento para obtener muestras “look-alike”, esencialmente sólo necesita el uso del estimador Rao-Blackwell $\tilde{F}_n(x)$, sin embargo, antes de considerar el Teorema 3.1, necesitamos hacer una observación que toma en cuenta la redundancia de la distribución condicional de la muestra completa X_1, X_2, \dots, X_n dado T_n .

En la presencia de parámetros, digamos k , usualmente la dimensión de T_n coincidirá con k , como en la FEN, así, aún si es correcto hablar de la distribución condicional de la muestra completa dado T_n , en realidad se identifica la distribución condicional de $n - k$ términos de la muestra dado T_n y se observa que los k términos restantes satisfacen un sistema de ecuaciones que los define hasta la incertidumbre de una permutación. Es decir, si se da $t_n = T_n(x_1, \dots, x_n)$ y se genera digamos, $x_1^*, x_2^*, \dots, x_{n-k}^*$, entonces para obtener los términos restantes $x_{n-k+1}^*, \dots, x_n^*$ que completan la muestra \mathbf{x}^* que produce el mismo valor $T_n(x_1^*, \dots, x_n^*) = t_n$, debemos resolver un sistema de ecuaciones.

Asuma que el entero máximo i ($\leq n$) para el cual la distribución condicional de X_1, X_2, \dots, X_n dado T_n resulta en una distribución sin redundancias, es $n - k$. Es decir, la distribución condicional de la muestra está realmente en un espacio de dimensión $n - k$ y no en uno de dimensión menor. Por facilidad de notación y sin pérdida de generalidad, los últimos $n - k$ términos de la muestra \mathbf{x}^* , aquellos con subíndices superiores, se generarán primero, y la observación hecha sobre el sistema con k valores desconocidos se usará para producir los k términos restantes.

Teorema 3.1. *Bajo doble transitividad, teniendo a $n - k$ como el número máximo de términos, para la distribución condicional de los miembros muestrales dado t_n que no tienen redundancias, el procedimiento para obtener los $n - k$ términos superiores de la muestra \mathbf{x}^* que tienen exactamente la misma distribución condicional que cualquiera de los $n - k$ términos de la muestra original \mathbf{X} es el siguiente:*

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

Paso 1 Sea u_n una realización de una v.a. $U(0,1)$, entonces defina $x_n^* = \tilde{F}_n^{-1}(u_n)$ y recalcule t_{n-1} de t_n y x_n^* , usando doble transitividad, y denote 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 notación obvia para su inversa, sea u_{n-1} otra realización de una v.a. $U(0,1)$ independientemente seleccionada de u_n , y defina el número $x_{n-1}^* = \tilde{F}_{n-1}^{-1}(u_{n-1})$ y recalcule t_{n-2}^* de t_{n-1}^* y x_{n-1}^* ; continuar de esta manera hasta,

Paso (n-k) Genere x_{k+1}^* de u_{k+1} , seleccionada de una $U(0,1)$, independientemente de las anteriores u_j 's usando \tilde{F}_{k+1}^{-1} adecuadamente.

Estos $n - k$ términos de la muestra \mathbf{x}^* es una realización de la distribución condicional para cualesquiera $n - k$ términos de la muestra \mathbf{X} dado t_n .

Finalmente, encontrar los k términos restantes x_1^*, \dots, x_k^* como cualquier permutación de la solución única (ordenada) derivada del requerimiento de que $T_n(x_1, \dots, x_n)$ debe ser igual a $T_n(x_1^*, \dots, x_n^*)$.

3.2.2. Prueba de Bondad de Ajuste para la Gaussiana Inversa

El estimador Rao-Blackwell $\tilde{F}_n(x)$ en (3.1) está bien definido para n tan pequeño como 3, así, k del Teorema 3.1 es 2. El hecho de que no hay expresión explícita para la inversa del estimador Rao-Blackwell no importa, ya que se calcula con una búsqueda iterativa simple.

Prueba Basada en el p-value Condicional “Exacto”

Para obtener el p-value condicional “exacto” de la prueba para el modelo GI dada una muestra X_1, X_2, \dots, X_n , realizamos lo siguiente:

1. calculamos la estadística de Anderson-Darling y la llamamos A_0^2 ,
2. generamos una cantidad suficientemente grande de muestras “look-alike”, y con cada muestra calculamos A^2 también,
3. la proporción de las A^2 “look-alike” que exceden A_0^2 es el p-value condicional.

Bajo la hipótesis nula, los valores de A^2 de las muestras “look-alike”, hacen una muestra de la distribución condicional de la estadística de Anderson-Darling.

Como ya hemos mencionado antes, el proceso empírico de Rao-Blackwell puede ser usado para construir pruebas EDF, como las basadas en el proceso empírico con parámetros estimados. Ambos procesos conducen a la misma distribución asintótica de las pruebas EDF, así, al menos desde un punto de vista muestral grande, no debe haber diferencias debido a usar uno en lugar de otro. Por tanto, en el procedimiento recién descrito para obtener el p-value condicional, los valores de la estadística A^2 pueden ser obtenidos o bien usando el estimador “plug-in” o bien usando el estimador Rao-Blackwell para la función de distribución.

Hemos ilustrado el uso de la distribución condicional de una estadística particular, pero el procedimiento puede ser adaptado para simular la distribución condicional exacta de cualquier estadística empleada para probar el ajuste.

Propiedades de Potencia

Hemos investigado las propiedades de potencia de la prueba basada en el p-value condicional “exacto” (para la distribución GI), para tal efecto realizamos el siguiente experimento de Monte Carlo. Generamos 1,000 muestras de tamaño $n = 10, 20, 30$ de cada una de las poblaciones consideradas por Mudholkar y Tian (2002) y por Vexler et al. (2011). Para cada población y cada tamaño muestral n , realizamos la prueba (1,000 veces) para la GI descrita anteriormente, cada prueba se basó en 1,000

muestras “look-like”. La proporción de rechazos obtenidos representa la potencia de la prueba. Una comparación de potencias de la prueba de O’Reilly y Gracia-Medrano (2006) contra otras pruebas conocidas (incluyendo las pruebas ya consideradas de O’Reilly y Rueda (1992), Vexler et al. (2011), y Mudholkar y Tian (2002)) se muestra en las Tablas 3.4 y 3.5. Es evidente la superioridad del método propuesto por O’Reilly y Gracia-Medrano (2006) en cuanto a potencia se refiere. Nótese también la similitud de potencias entre los métodos de O’Reilly y Gracia-Medrano (2006) y O’Reilly y Rueda (1992).

Ejemplo

Siguiendo a Vexler et al. (2011) decidimos usar la prueba basada en el p-value condicional “exacto” para evaluar el ajuste de la distribución GI a los datos introducidos por Folks y Chhikara (1978). Los Conjuntos de datos nombrados por Vexler et al. (2011) como Conjunto de datos 1, 2, 3 y 4 fueron los siguientes: Vida útil (en días) de un producto comestible, resistencia a la fractura de la soldadura MIG (soldadura de metal con gas inerte), precipitación (en pulgadas) en el Puente Jug de Maryland, cantidad de escorrentía en el Puente Jug de Maryland.

En las Tablas 3.6 y 3.7 se han añadido los p-values obtenidos usando el método de O’Reilly y Gracia-Medrano (2006) a la Tabla 2.8 considerada por Vexler et al. (2011). Cuando se ha usado el estimador “plug-in” para obtener los valores de la estadística de Anderson-Darling, las conclusiones acerca de rechazar o no la suposición GI para los datos coinciden para las tres pruebas consideradas al nivel de significancia $\alpha = 0.05$, sin embargo, no ocurre así si el nivel de significancia se cambia a 0.01. Cuando se ha usado el estimador Rao-Blackwell para obtener los valores de la estadística de Anderson-Darling, la prueba basada en el p-value condicional “exacto” acepta la suposición GI (al nivel de significancia $\alpha = 0.05$) para el Conjunto de datos 3, discrepando de las conclusiones de las pruebas de Mudholkar y Tian (2002) y Vexler et al. (2011). Los resultados obtenidos usando la prueba para la GI basada en la estadística A^2 y el p-value condicional “exacto” concuerdan con los resultados obtenidos por O’Reilly y Rueda (1992), quienes ya habían obtenido p-values para su prueba aplicada a los Conjuntos de datos 1, 3 y 4.

Tabla 3.4 Una comparación de potencias^a para la prueba basada en el p-value condicional “exacto”, con el nivel de significancia $\alpha = 0.05$. La estadística A^2 se obtuvo usando el estimador “plug-in” para la función de distribución GI.

Distribución	n	$K_{m=2,n}$	$K_{m=3,n}$	$K_{m=4,n}$	$K_{m=5,n}$	Z	KS1	KS2	TK_n	A^2 O & R	A^2 “look-alike”	
Exponencial	10	0.2075 ^b	0.1991	0.1499	0.0643	0.0206	0.262	0.280	0.2060	0.3445	0.3540	
	20	0.4668	0.4675 ^b	0.4347	0.3781	0.0226	0.518	0.525	0.4636	0.6348	0.6510	
	30	0.6347	0.6597 ^b	0.6563	0.6169	0.0328	0.654	0.668	0.6446	0.7970	0.7900	
	Uniforme(0,1)	10	0.4768 ^b	0.4759	0.3987	0.1857	0.1814	0.342	0.356	0.5078	0.5168	0.5530
		20	0.8682	0.8802 ^b	0.8645	0.8407	0.426	0.616	0.630	0.8826	0.8673	0.8720
		30	0.9687	0.9789 ^b	0.9833	0.9772	0.5764	0.776	0.782	0.9815	0.9711	0.9730
Weibull(1,2)		10	0.1251 ^b	0.1256	0.1064	0.0428	0.0721	0.06	0.074	0.1354	0.1746	0.2020
		20	0.2536	0.2707 ^b	0.2332	0.2105	0.1611	0.128	0.080	0.2565	0.3706	0.3690
		30	0.3564	0.3895 ^b	0.4028	0.3636	0.277	0.168	0.111	0.3847	0.5297	0.5440
	LogNor(0.5,1)	10	0.0467 ^b	0.0464	0.0437	0.0318	0.0407	0.048	0.055	0.0460	0.0633	0.0770
		20	0.0588	0.0535 ^b	0.0441	0.0351	0.042	0.068	0.080	0.0541	0.1099	0.1290
		30	0.0707	0.0679 ^b	0.0591	0.0517	0.0397	0.082	0.111	0.0547	0.1451	0.1430

^a La prueba de interés se realizó 1, 000 veces, y cada una se basó en 1, 000 muestras “look-like”, mientras que las otras potencias se estimaron en base a 10, 000 repeticiones. A^2 “look-alike”-prueba basada en el p-value condicional “exacto” (O’Reilly y Gracia-Medrano, 2006)-; A^2 O & R-prueba basada en la distribución asintótica de la estadística A^2 (O’Reilly y Rueda, 1992)-; TK_n -prueba de cociente de verosimilitudes empíricas basada en densidades (Vexler et al., 2011)-; $K_{m,n}$ -prueba basada en la entropía (Mudholkar y Tian, 2002)-; Z -prueba caracterizada por independencia (Mudholkar et al., 2001)-; $KS1$ -prueba de Kolmogorov-Smirnov modificada (Edgeman et al., 1988)-; y $KS2$ -prueba de Kolmogorov-Smirnov usando transformación (Edgeman, 1990)-.
^b Valores que corresponden a los m óptimos encontrados empíricamente por Mudholkar y Tian (2002) dadas alternativas conocidas.

Tabla 3.5 Una comparación de potencias^a para la prueba basada en el p-value condicional “exacto”, con el nivel de significancia $\alpha = 0.05$. La estadística A^2 se obtuvo usando el estimador Rao-Blackwell para la función de distribución GI.

Distribución	n	$K_{m=2,n}$	$K_{m=3,n}$	$K_{m=4,n}$	$K_{m=5,n}$	Z	KS1	KS2	TK_n	A^2 O & R	A^2 “look-alike”
Exponencial	10	0.2075 ^b	0.1991	0.1499	0.0643	0.0206	0.262	0.280	0.2060	0.3343	0.3600
	20	0.4668	0.4675 ^b	0.4347	0.3781	0.0226	0.518	0.525	0.4636	0.6235	0.6300
	30	0.6347	0.6597 ^b	0.6563	0.6169	0.0328	0.654	0.668	0.6446	0.8031	0.7980
Uniforme(0,1)	10	0.4768 ^b	0.4759	0.3987	0.1857	0.1814	0.342	0.356	0.5078	0.5131	0.5100
	20	0.8682	0.8802 ^b	0.8645	0.8407	0.426	0.616	0.630	0.8826	0.8587	0.8570
	30	0.9687	0.9789 ^b	0.9833	0.9772	0.5764	0.776	0.782	0.9815	0.9687	0.9770
Weibull(1,2)	10	0.1251 ^b	0.1256	0.1064	0.0428	0.0721	0.06	0.074	0.1354	0.2099	0.2000
	20	0.2536	0.2707 ^b	0.2332	0.2105	0.1611	0.128	0.080	0.2565	0.3891	0.3980
	30	0.3564	0.3895 ^b	0.4028	0.3636	0.277	0.168	0.111	0.3847	0.5384	0.5540
LogNor(0.5,1)	10	0.0467 ^b	0.0464	0.0437	0.0318	0.0407	0.048	0.055	0.0460	0.0845	0.0840
	20	0.0588	0.0535 ^b	0.0441	0.0351	0.042	0.068	0.080	0.0541	0.1199	0.1210
	30	0.0707	0.0679 ^b	0.0591	0.0517	0.0397	0.082	0.111	0.0547	0.1431	0.1620

^a La prueba de interés se realizó 1,000 veces, y cada una se basó en 1,000 muestras “look-like”, mientras que las otras potencias se estimaron en base a 10,000 repeticiones. A^2 “look-alike”-prueba basada en el p-value condicional “exacto” (O’Reilly y Gracia-Medrano, 2006)-; A^2 O & R-prueba basada en la distribución asintótica de la estadística A^2 (O’Reilly y Rueda, 1992)-; TK_n -prueba de cociente de verosimilitudes empíricas basada en densidades (Vexler et al., 2011)-; $K_{m,n}$ -prueba basada en la entropía (Mudholkar y Tian, 2002)-; Z -prueba caracterizada por independencia (Mudholkar et al., 2001)-; $KS1$ -prueba de Kolmogorov-Smirnov modificada (Edgeman et al., 1988)-; y $KS2$ -prueba de Kolmogorov-Smirnov usando transformación (Edgeman, 1990)-.

^b Valores que corresponden a los m óptimos encontrados empíricamente por Mudholkar y Tian (2002) dadas alternativas conocidas.

Tabla 3.6 Prueba para la GI basada en los datos introducidos por Folks y Chhikara (1978). Nivel de significancia $\alpha = 0.05$. Comparación de la prueba^a de O’Reilly y Gracia-Medrano (2006) contra las pruebas de Vexler et al. (2011) y Mudholkar y Tian (2002). La estadística A^2 se obtuvo usando el estimador “plug-in” para la función de distribución GI.

Conjunto de datos	Tamaño muestral	Prueba ^b	p-value
1	26	A^2 “look-alike”	0.0215
		TK_{26}	0.0063
		$K_{m=2,26}$	0.0061
		$K_{m=3,26}$	0.0058
		$K_{m=4,26}$	0.0090
		$K_{m=5,26}$	0.0041
2	19	A^2 “look-alike”	0.4087
		TK_{19}	0.1791
		$K_{m=2,19}$	0.1565
		$K_{m=3,19}$	0.1516
		$K_{m=4,19}$	0.2663
		$K_{m=5,19}$	0.3379
3	25	A^2 “look-alike”	0.0446
		TK_{25}	0.0099
		$K_{m=2,25}$	0.0197
		$K_{m=3,25}$	0.0159
		$K_{m=4,25}$	0.0098
		$K_{m=5,25}$	0.0063
4	25	A^2 “look-alike”	0.9275
		TK_{25}	0.9271
		$K_{m=2,25}$	0.9393
		$K_{m=3,25}$	0.9538
		$K_{m=4,25}$	0.8738
		$K_{m=5,25}$	0.8001

^a Basada en 10,000 muestras “look-alike”.

^b A^2 “look-alike”-prueba basada en el p-value condicional “exacto” (O’Reilly y Gracia-Medrano, 2006)-; TK_n -prueba de cociente de verosimilitudes empíricas basada en densidades (Vexler et al., 2011)-; $K_{m,n}$ -prueba basada en la entropía (Mudholkar y Tian, 2002)-.

Tabla 3.7 Prueba para la GI basada en los datos introducidos por Folks y Chhikara (1978). Nivel de significancia $\alpha = 0.05$. Comparación de la prueba^a de O’Reilly y Gracia-Medrano (2006) contra las pruebas de Vexler et al. (2011) y Mudholkar y Tian (2002). La estadística A^2 se obtuvo usando el estimador Rao-Blackwell para la función de distribución GI.

Conjunto de datos	Tamaño muestral	Prueba ^b	p-value
1	26	A^2 “look-alike”	0.0292
		TK_{26}	0.0063
		$K_{m=2,26}$	0.0061
		$K_{m=3,26}$	0.0058
		$K_{m=4,26}$	0.0090
		$K_{m=5,26}$	0.0041
2	19	A^2 “look-alike”	0.3867
		TK_{19}	0.1791
		$K_{m=2,19}$	0.1565
		$K_{m=3,19}$	0.1516
		$K_{m=4,19}$	0.2663
		$K_{m=5,19}$	0.3379
3	25	A^2 “look-alike”	0.0552
		TK_{25}	0.0099
		$K_{m=2,25}$	0.0197
		$K_{m=3,25}$	0.0159
		$K_{m=4,25}$	0.0098
		$K_{m=5,25}$	0.0063
4	25	A^2 “look-alike”	0.8598
		TK_{25}	0.9271
		$K_{m=2,25}$	0.9393
		$K_{m=3,25}$	0.9538
		$K_{m=4,25}$	0.8738
		$K_{m=5,25}$	0.8001

^a Basada en 10,000 muestras “look-alike”.

^b A^2 “look-alike”-prueba basada en el p-value condicional “exacto” (O’Reilly y Gracia-Medrano, 2006)-; TK_n -prueba de cociente de verosimilitudes empíricas basada en densidades (Vexler et al., 2011)-; $K_{m,n}$ -prueba basada en la entropía (Mudholkar y Tian, 2002)-.

Capítulo 4

Conclusiones

En este trabajo hemos realizado una comparación de potencias de pruebas de bondad de ajuste para la distribución Gaussiana inversa, dos de estas pruebas están asociadas al concepto de entropía y otras dos pruebas involucran a la estadística EDF de Anderson-Darling. Antes de comenzar a investigar las propiedades de potencia de las pruebas, pensamos que podíamos esperar buenos resultados para las pruebas de la hipótesis GI tanto basada en la distribución asintótica de la A^2 como en la que se basa en el p-value condicional “exacto”, puesto que una de las pruebas (Edgeman et al., 1988) contra las que comparan Mudholkar y Tian (2002) y que después retoman Vexler et al. (2011) ya había sido analizada en el trabajo de Gracia-Medrano y O’Reilly (2004), obteniéndose buenos resultados para la A^2 . Al final las suposiciones planteadas al inicio del presente trabajo fueron confirmadas, es decir, la prueba basada en la entropía, así como la prueba de cociente de verosimilitudes empíricas basada en densidades resultaron con potencias más bajas que aquellas obtenidas por las pruebas que involucran a la A^2 para alternativas conocidas, a saber: exponencial con media 1, uniforme(0,1), Weibull(1,2) con parámetro de escala 1 y parámetro de forma 2, y lognormal(0.5,1) con media e y desviación estándar $e\sqrt{e-1}$. Verificamos igual que las potencias obtenidas usando teoría asintótica para la A^2 resultan muy parecidas a las potencias “exactas”, tanto cuando se usa el estimador “plug-in” como cuando se usa el estimador Rao-Blackwell para obtener los valores de la estadística de Anderson-Darling.

Por los resultados obtenidos, recomendamos el uso de la prueba basada en el p-value condicional “exacto” sobre las pruebas que involucran a la entropía de una distribución, puesto que además de la superioridad demostrada en lo que a potencia se refiere para las alternativas ya mencionadas, este procedimiento no requiere el uso de tablas, y en vez de esto, básicamente sólo emplea el estimador Rao-Blackwell de la función de distribución para generar las así llamadas muestras “look-alike” que sirven para generar una muestra de la distribución condicional de la A^2 y con la cual se obtiene el p-value “exacto”.

Es de reconocerse que la prueba de bondad de ajuste basada en la entropía para el caso normal (con parámetros desconocidos) que introdujo Vasicek (1976), resultó ser comparable con las mejores pruebas para normalidad (incluidas las EDF), pero en el caso estudiado en este trabajo, la adaptación hecha de la prueba basada en la noción de entropía y aquella reportada por Vexler et al. (2011) no resultaron comparables con las mejores vistas hasta este momento para la Gaussiana inversa.

Apéndice A

Algunos Resultados Elementales

Teorema A.1. *Sea F una función uno a uno continua definida sobre un intervalo, y suponga que F es diferenciable en $F^{-1}(p)$, con derivada $F'(F^{-1}(p)) \neq 0$. Entonces F^{-1} es diferenciable en p , y*

$$\frac{d}{dp}F^{-1}(p) = \frac{1}{F'(F^{-1}(p))}. \quad (\text{A.1})$$

Teorema A.2. (Teorema del valor medio) *Si F es continua sobre $[a, b]$ y diferenciable sobre (a, b) , entonces hay un número x en (a, b) tal que*

$$F'(x) = \frac{F(b) - F(a)}{b - a}. \quad (\text{A.2})$$

Definición A.1. (Función digamma) *La función digamma está definida como la derivada del logaritmo de la función gamma*

$$\begin{aligned} \psi(x) &= \frac{d}{dx} \log[\Gamma(x)] \\ &= \frac{\Gamma'(x)}{\Gamma(x)}. \end{aligned}$$

Definición A.2. *Sea $[a, b]$ un intervalo dado. Por partición P de $[a, b]$ entendemos*

un conjunto finito de puntos x_0, x_1, \dots, x_n donde

$$a = x_0 \leq x_1 \leq \dots \leq x_{n-1} \leq x_n = b.$$

Definición A.3. Sea α una función monótona creciente en $[a, b]$ (como $\alpha(a)$ y $\alpha(b)$ son finitas, se deduce que α es acotada en $[a, b]$). Correspondiendo a cada partición P de $[a, b]$, escribimos

$$\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1}).$$

Es claro que $\Delta\alpha_i \geq 0$. Para toda función real f acotada en $[a, b]$ y para cada partición P de $[a, b]$ hacemos

$$\begin{aligned} M_i &= \sup f(x) && (x_{i-1} \leq x \leq x_i), \\ m_i &= \inf f(x) && (x_{i-1} \leq x \leq x_i), \\ U(P, f, \alpha) &= \sum_{i=1}^n M_i \Delta\alpha_i, \\ L(P, f, \alpha) &= \sum_{i=1}^n m_i \Delta\alpha_i. \end{aligned}$$

Definimos

$$\int_a^b f d\alpha = \inf U(P, f, \alpha), \quad (\text{A.3})$$

$$\int_a^b f d\alpha = \sup L(P, f, \alpha), \quad (\text{A.4})$$

tomando también los \inf y los \sup sobre todas las particiones.

Si los primeros miembros de (A.3) y (A.4) son iguales, representamos su valor

común por

$$\int_a^b f d\alpha \quad (\text{A.5})$$

o algunas veces por

$$\int_a^b f(x) d\alpha(x). \quad (\text{A.6})$$

Esta es la *integral de Riemann-Stieltjes* (o simplemente la *integral de Stieltjes*) de f respecto a α , sobre $[a, b]$.

Si existe (A.5), esto es si (A.3) y (A.4) son iguales, decimos que f es integrable con respecto a α , en el sentido Riemann y escribimos $f \in \mathcal{R}(\alpha)$

Teorema A.3. (*Desigualdad de Chebychev*) Sea X una variable aleatoria y sea $g(x)$ una función no-negativa. Entonces, para cada $r > 0$,

$$P[g(X) \geq r] \leq \frac{E[g(X)]}{r}.$$

Teorema A.4. La sucesión de variables aleatorias, X_1, X_2, \dots , converge en probabilidad a una constante μ si y sólo si la sucesión también converge en distribución a μ . Esto es, el enunciado

$$P(|X_n - \mu| > \varepsilon) \rightarrow 0 \quad \text{para cada } \varepsilon > 0$$

es equivalente a

$$P(X_n \leq x) \rightarrow \begin{cases} 0, & \text{si } x < \mu, \\ 1, & \text{si } x > \mu. \end{cases} \quad (\text{A.7})$$

Teorema A.5. (*Teorema de Slutsky*) Si $X_n \rightarrow X$ en distribución y $Y_n \rightarrow a$, a

constante, en probabilidad, entonces

- a. $Y_n X_n \rightarrow aX$ en distribución.
- b. $X_n + Y_n \rightarrow X + a$ en distribución.

Teorema A.6. (Teorema del Multiplicador de Lagrange) Sean $f : U \subset \mathbf{R}^n \rightarrow \mathbf{R}$ y $g : U \subset \mathbf{R}^n \rightarrow \mathbf{R}$ funciones suaves dadas. Sean $\mathbf{x}_0 \in U$ y $g(\mathbf{x}_0) = c$, y sea S el conjunto de nivel para g con valor c (recordar que éste es el conjunto de puntos $\mathbf{x} \in \mathbf{R}^n$ con $g(\mathbf{x}) = c$). Suponer que $\nabla g(\mathbf{x}_0) \neq 0$. Si $f|_S$, que denota a “ f restringida a S ”, tiene un máximo o un mínimo en S , en \mathbf{x}_0 , entonces existe un número real λ tal que

$$\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0). \quad (\text{A.8})$$

Para hallar los extremos con restricciones de f usando el teorema del multiplicador de Lagrange, debemos buscar un punto \mathbf{x}_0 y una constante λ , llamada multiplicador de Lagrange, tal que $\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0)$.

En la ecuación (A.8) se dice que las derivadas parciales de f son proporcionales a las de g . Hallar los puntos \mathbf{x}_0 en los que ocurre esto, significa resolver las ecuaciones simultáneas

$$\begin{aligned} \frac{\partial f}{\partial x_1}(x_1, \dots, x_n) &= \lambda \frac{\partial g}{\partial x_1}(x_1, \dots, x_n) \\ \frac{\partial f}{\partial x_2}(x_1, \dots, x_n) &= \lambda \frac{\partial g}{\partial x_2}(x_1, \dots, x_n) \\ &\vdots \\ \frac{\partial f}{\partial x_n}(x_1, \dots, x_n) &= \lambda \frac{\partial g}{\partial x_n}(x_1, \dots, x_n) \\ g(x_1, \dots, x_n) &= c \end{aligned}$$

para x_1, \dots, x_n y λ .

Otra manera de considerar estas ecuaciones es así: pensar en λ con una variable adicional y formar la función auxiliar $h(x_1, \dots, x_n, \lambda) = f(x_1, \dots, x_n) - \lambda[g(x_1, \dots, x_n) - c]$. En el teorema del multiplicador de Lagrange se dice que para hallar los puntos extremos de $f|S$ debemos examinar los puntos críticos de h . Éstos se encuentran resolviendo las ecuaciones

$$\begin{aligned}0 &= \frac{\partial h}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda \frac{\partial g}{\partial x_1} \\ &\vdots \\0 &= \frac{\partial h}{\partial x_n} = \frac{\partial f}{\partial x_n} - \lambda \frac{\partial g}{\partial x_n} \\0 &= \frac{\partial h}{\partial \lambda} = g(x_1, \dots, x_n) - c\end{aligned}$$

que son las mismas que las ecuaciones anteriores.

Apéndice B

Algunas Demostraciones

Proposición B.1. *Sea $f(x)$ una función de densidad de probabilidad. Entonces*

$$\sum_{j=1}^n \int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx = 2m \int_{X_{(1)}}^{X_{(n)}} f(x)dx - \sum_{k=1}^{m-1} (m-k) \int_{X_{(n-k)}}^{X_{(n-k+1)}} f(x)dx - \sum_{k=1}^{m-1} (m-k) \int_{X_{(k)}}^{X_{(k+1)}} f(x)dx$$

donde $X_{(j)} = X_{(1)}$, si $j \leq 1$, y $X_{(j)} = X_{(n)}$, si $j \geq n$.

Demostración.

$$\begin{aligned} \sum_{j=1}^n \int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx &= \sum_{j=1}^n \{F(X_{(j+m)}) - F(X_{(j-m)})\} \\ &= \sum_{j=1}^n \{[F(X_{(j+m)}) - F(X_{(j-m)})] + [F(X_{(j-m+1)}) - F(X_{(j-m+1)})] + \\ &\quad [F(X_{(j+m-1)}) - F(X_{(j+m-1)})]\} \\ &= \sum_{j=1}^n \{[F(X_{(j-m+1)}) - F(X_{(j-m)})] + [F(X_{(j+m-1)}) - F(X_{(j-m+1)})] + \\ &\quad [F(X_{(j+m)}) - F(X_{(j+m-1)})]\} \\ &= \sum_{j=1}^n \left\{ \int_{X_{(j-m)}}^{X_{(j-m+1)}} f(x)dx + \int_{X_{(j-m+1)}}^{X_{(j+m-1)}} f(x)dx + \int_{X_{(j+m-1)}}^{X_{(j+m)}} f(x)dx \right\} \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^n \int_{X_{(j-m)}}^{X_{(j-m+1)}} f(x)dx + \sum_{j=1}^n \int_{X_{(j-m+1)}}^{X_{(j+m-1)}} f(x)dx + \sum_{j=1}^n \int_{X_{(j+m-1)}}^{X_{(j+m)}} f(x)dx \\
&= \int_{X_{(1)}}^{X_{(n-m+1)}} f(x)dx + \sum_{j=1}^n \int_{X_{(j-m+1)}}^{X_{(j+m-1)}} f(x)dx + \int_{X_{(m)}}^{X_{(n)}} f(x)dx \\
&= \left\{ \int_{X_{(1)}}^{X_{(n)}} f(x)dx - \int_{X_{(n-m+1)}}^{X_{(n)}} f(x)dx \right\} + \sum_{j=1}^n \int_{X_{(j-(m-1))}}^{X_{(j+(m-1))}} f(x)dx + \\
&\quad \left\{ \int_{X_{(1)}}^{X_{(n)}} f(x)dx - \int_{X_{(1)}}^{X_{(m)}} f(x)dx \right\} \\
&= \sum_{j=1}^n \int_{X_{(j-(m-1))}}^{X_{(j+(m-1))}} f(x)dx + 2 \int_{X_{(1)}}^{X_{(n)}} f(x)dx - \int_{X_{(n-(m-1))}}^{X_{(n)}} f(x)dx - \\
&\quad \int_{X_{(1)}}^{X_{(m)}} f(x)dx.
\end{aligned}$$

Realizando los mismos pasos anteriores llegamos a que

$$\begin{aligned}
\sum_{j=1}^n \int_{X_{(j-(m-1))}}^{X_{(j+(m-1))}} f(x)dx &= \sum_{j=1}^n \int_{X_{(j-(m-2))}}^{X_{(j+(m-2))}} f(x)dx + 2 \int_{X_{(1)}}^{X_{(n)}} f(x)dx - \\
&\quad \int_{X_{(n-(m-2))}}^{X_{(n)}} f(x)dx - \int_{X_{(1)}}^{X_{(m-1)}} f(x)dx
\end{aligned}$$

Así

$$\begin{aligned}
\sum_{j=1}^n \int_{X_{(j-m)}}^{X_{(j+m)}} f(x) dx &= \sum_{j=1}^n \int_{X_{(j-(m-2))}}^{X_{(j+(m-2))}} f(x) dx + 2 \cdot 2 \int_{X_{(1)}}^{X_{(n)}} f(x) dx - \\
&\quad \left\{ \int_{X_{(n-(m-2))}}^{X_{(n)}} f(x) dx + \int_{X_{(n-(m-1))}}^{X_{(n)}} f(x) dx \right\} - \\
&\quad \left\{ \int_{X_{(1)}}^{X_{(m-1)}} f(x) dx + \int_{X_{(1)}}^{X_{(m)}} f(x) dx \right\} \\
&= \\
&\quad \vdots \\
&= 2m \int_{X_{(1)}}^{X_{(n)}} f(x) dx - \sum_{k=1}^m \int_{X_{(n-(k-1))}}^{X_{(n)}} f(x) dx - \\
&\quad \sum_{k=1}^m \int_{X_{(1)}}^{X_{(k)}} f(x) dx
\end{aligned}$$

como

$$\begin{aligned}
\sum_{k=1}^m \int_{X_{(n-(k-1))}}^{X_{(n)}} f(x) dx &= \int_{X_{(n-1)}}^{X_{(n)}} f(x) dx + \int_{X_{(n-2)}}^{X_{(n)}} f(x) dx + \\
&\quad \cdots + \int_{X_{(n-(m-1))}}^{X_{(n)}} f(x) dx \\
&= (m-1) \int_{X_{(n-1)}}^{X_{(n)}} f(x) dx + (m-2) \int_{X_{(n-2)}}^{X_{(n-1)}} f(x) dx + \\
&\quad \cdots + \int_{X_{(n-(m-1))}}^{X_{(n-(m-2))}} f(x) dx \\
&= \sum_{k=1}^{m-1} (m-k) \int_{X_{(n-k)}}^{X_{(n-k+1)}} f(x) dx
\end{aligned}$$

y

$$\begin{aligned}
\sum_{k=1}^m \int_{X(1)}^{X(k)} f(x)dx &= \int_{X(1)}^{X(2)} f(x)dx + \int_{X(1)}^{X(3)} f(x)dx + \\
&\quad \cdots + \int_{X(1)}^{X(m)} f(x)dx \\
&= (m-1) \int_{X(1)}^{X(2)} f(x)dx + (m-2) \int_{X(2)}^{X(3)} f(x)dx + \\
&\quad \cdots + \int_{X(m-1)}^{X(m)} f(x)dx \\
&= \sum_{k=1}^{m-1} (m-k) \int_{X(k)}^{X(k+1)} f(x)dx
\end{aligned}$$

tenemos que

$$\begin{aligned}
\sum_{j=1}^n \int_{X(j-m)}^{X(j+m)} f(x)dx &= 2m \int_{X(1)}^{X(n)} f(x)dx - \sum_{k=1}^{m-1} (m-k) \int_{X(n-k)}^{X(n-k+1)} f(x)dx - \\
&\quad \sum_{k=1}^{m-1} (m-k) \int_{X(k)}^{X(k+1)} f(x)dx
\end{aligned}$$

□

Teorema B.1. Denotemos por $Y_{(1)}, \dots, Y_{(n)}$ las estadísticas de orden de una muestra aleatoria, Y_1, \dots, Y_n , de una población continua con función de distribución $F(y)$ y función de densidad de probabilidad $f(y)$. Entonces la función de densidad de probabilidad conjunta de $Y_{(i)}$ y $Y_{(j)}$, $1 \leq i < j \leq n$, es

$$\begin{aligned}
f_{Y_{(i)}, Y_{(j)}}(u, v) &= \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f(u)f(v)[F(u)]^{i-1} \\
&\quad \times [F(v) - F(u)]^{j-1-i} [1 - F(v)]^{n-j}
\end{aligned}$$

para $-\infty < u < v < \infty$.

Proposición B.2. Sea Y_1, Y_2, \dots, Y_n una muestra aleatoria de tamaño n de la distribución uniforme $(0,1)$. Entonces,

- (a) $Y_{(i+j)} - Y_{(i)} \sim \text{Beta}(j, n - j + 1)$,
- (b) $E(\log\{Y_{(i+j)} - Y_{(i)}\}) = \psi(j) - \psi(n + 1)$, donde ψ es la función digamma definida en el Apéndice A.

Demostración. Sea $T = Y_{(i+j)} - Y_{(i)}$.

- (a) Definamos $S = Y_{(i)}$. Derivaremos la fdp conjunta de S y T de la fdp conjunta de $Y_{(i)}$ y $Y_{(i+j)}$.

$$\begin{aligned} f_{Y_{(i)}, Y_{(i+j)}}(u, v) &= \frac{n!}{(i-1)!(j-1)!(n-j-i)!} f(u) f(v) [F(u)]^{i-1} \\ &\quad \times [F(v) - F(u)]^{j-1} [1 - F(v)]^{n-j-i} \\ &= \frac{n!}{(i-1)!(j-1)!(n-j-i)!} u^{i-1} [v - u]^{j-1} [1 - v]^{n-j-i}, \end{aligned}$$

siempre que $u < v$ y $u, v \in (0, 1)$. Del sistema de ecuaciones

$$T = Y_{(i+j)} - Y_{(i)},$$

$$S = Y_{(i)},$$

resolviendo para $Y_{(i)}$ y $Y_{(i+j)}$ obtenemos $Y_{(i)} = S$ y $Y_{(i+j)} = S + T$. El Jacobiano para esta transformación es 1. La transformación de $(Y_{(i)}, Y_{(i+j)})$ a (S, T) mapea $\{(u, v) : 0 < u < v < 1\}$ sobre el conjunto $\{(s, t) : 0 < s < 1 - t < 1\}$. Por lo tanto la fdp conjunta de (S, T) es

$$f_{S,T}(s, t) = \frac{n!}{(i-1)!(j-1)!(n-j-i)!} s^{i-1} t^{j-1} [1 - s - t]^{n-j-i}$$

Para $t \in (0, 1)$, la fdp marginal de T está dada por

$$\begin{aligned}
f_T(t) &= \int_0^{1-t} \frac{n!}{(i-1)!(j-1)!(n-j-i)!} s^{i-1} t^{j-1} [1-s-t]^{n-j-i} ds \\
&= \frac{n!}{(i-1)!(j-1)!(n-j-i)!} t^{j-1} \int_0^{1-t} s^{i-1} [1-s-t]^{n-j-i} ds \\
&= \frac{n!}{(i-1)!(j-1)!(n-j-i)!} t^{j-1} \int_0^{1-t} s^{i-1} [(1-t)-s]^{n-j-i} ds \\
&= \frac{n!}{(i-1)!(j-1)!(n-j-i)!} t^{j-1} (1-t)^{n-j-1} \int_0^{1-t} \left(\frac{s}{1-t}\right)^{i-1} \left[1 - \frac{s}{1-t}\right]^{n-j-i} ds \\
&= \frac{n!}{(i-1)!(j-1)!(n-j-i)!} t^{j-1} (1-t)^{n-j} \int_0^1 w^{i-1} [1-w]^{n-j-i} dw \\
&= \frac{n!}{(i-1)!(j-1)!(n-j-i)!} t^{j-1} (1-t)^{n-j} \frac{\Gamma(i)\Gamma(n-j-i+1)}{\Gamma(n-j+1)} \\
&= \frac{n!}{(j-1)!\Gamma(n-j+1)} t^{j-1} (1-t)^{n-j} \\
&= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} t^{j-1} (1-t)^{(n-j+1)-1},
\end{aligned}$$

es decir, $T \sim \text{Beta}(j, n-j+1)$.

(b)

$$\begin{aligned}
m_{\log\{T\}}(b) &= E(\exp[b \cdot \log\{T\}]) \\
&= E(\exp[\log\{T^b\}]) \\
&= E(T^b) \\
&= \int_0^1 t^b \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} t^{j-1} (1-t)^{(n-j+1)-1} dt \\
&= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} \int_0^1 t^{b+j-1} (1-t)^{(n-j+1)-1} dt \\
&= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} \cdot \frac{\Gamma(b+j)\Gamma(n-j+1)}{\Gamma(n+b+1)} \\
&= \frac{\Gamma(n+1)}{\Gamma(j)} \cdot \frac{\Gamma(b+j)}{\Gamma(n+b+1)},
\end{aligned}$$

luego,

$$m'_{\log\{T\}}(b) = \frac{\Gamma(n+1)}{\Gamma(j)} \cdot \frac{\Gamma(n+b+1) \cdot \Gamma'(b+j) - \Gamma(b+j) \cdot \Gamma'(n+b+1)}{\Gamma^2(n+b+1)},$$

y así

$$\begin{aligned} E(\log\{T\}) &= m'_{\log\{T\}}(0) \\ &= \frac{\Gamma(n+1)}{\Gamma(j)} \cdot \frac{\Gamma(n+1) \cdot \Gamma'(j) - \Gamma(j) \cdot \Gamma'(n+1)}{\Gamma^2(n+1)} \\ &= \frac{\Gamma(n+1) \cdot \Gamma'(j) - \Gamma(j) \cdot \Gamma'(n+1)}{\Gamma(j)\Gamma(n+1)} \\ &= \frac{\Gamma'(j)}{\Gamma(j)} - \frac{\Gamma'(n+1)}{\Gamma(n+1)}, \\ &= \psi(j) - \psi(n+1). \end{aligned}$$

□

Proposición B.3. Para la variable aleatoria U_{mn} tenemos los siguientes resultados

- (a) U_{mn} es una variable aleatoria no-positiva,
- (b) $E(U_{mn}) = \log(n) - \log(2m) + (1 - \frac{2m}{n})\psi(2m) - \psi(n+1) + \frac{2}{n} \sum_{j=1}^m \psi(j+m-1)$,
- (c) $E(U_{mn}) \rightarrow 0$, cuando $n \rightarrow \infty$, $m \rightarrow \infty$, $m/n \rightarrow 0$.

Demostración. (a) Recordemos que si a_1, \dots, a_n son números positivos, entonces

$$[a_1 a_2 \cdots a_n]^{1/n} \leq \frac{1}{n} (a_1 + a_2 + \cdots + a_n). \quad (\text{B.1})$$

También,

$$\sum_{j=1}^n \int_{X_{(j-m)}}^{X_{(j+m)}} f(x) dx \leq 2m \int_{X_{(1)}}^{X_{(n)}} f(x) dx, \quad (\text{por la Proposición B.1})$$

$$\sum_{j=1}^n \{F(X_{(j+m)}) - F(X_{(j-m)})\} \leq 2m \{F(X_{(n)}) - F(X_{(1)})\},$$

es decir,

$$\frac{1}{2m} \sum_{j=1}^n \{F(X_{(j+m)}) - F(X_{(j-m)})\} \leq F(X_{(n)}) - F(X_{(1)}). \quad (\text{B.2})$$

Luego,

$$\begin{aligned} \exp(U_{mn}) &= \exp\left(n^{-1} \sum_{j=1}^n \log\left[\frac{n}{2m} \{F(X_{(j+m)}) - F(X_{(j-m)})\}\right]\right) \\ &= \exp\left(\sum_{j=1}^n \log\left[\frac{n}{2m} \{F(X_{(j+m)}) - F(X_{(j-m)})\}\right]^{1/n}\right) \\ &= \prod_{j=1}^n \exp\left(\log\left[\frac{n}{2m} \{F(X_{(j+m)}) - F(X_{(j-m)})\}\right]^{1/n}\right) \\ &= \prod_{j=1}^n \left[\frac{n}{2m} \{F(X_{(j+m)}) - F(X_{(j-m)})\}\right]^{1/n} \\ &\leq \frac{1}{n} \sum_{j=1}^n \left[\frac{n}{2m} \{F(X_{(j+m)}) - F(X_{(j-m)})\}\right] \quad (\text{por (B.1)}) \\ &= \frac{1}{2m} \sum_{j=1}^n \{F(X_{(j+m)}) - F(X_{(j-m)})\} \\ &\leq F(X_{(n)}) - F(X_{(1)}) \quad (\text{por (B.2)}) \\ &\leq 1, \end{aligned}$$

es decir, U_{mn} es una variable aleatoria no-positiva.

(b)

$$\begin{aligned}
E(U_{mn}) &= E\left(n^{-1} \sum_{j=1}^n \log\left[\frac{n}{2m} \{F(X_{(j+m)}) - F(X_{(j-m)})\}\right]\right) \\
&= \frac{1}{n} E\left(\sum_{j=1}^n \log\left[\frac{n}{2m} \{F(X_{(j+m)}) - F(X_{(j-m)})\}\right]\right) \\
&= \frac{1}{n} \sum_{j=1}^n E\left(\log\left[\frac{n}{2m} \{F(X_{(j+m)}) - F(X_{(j-m)})\}\right]\right) \\
&= \frac{1}{n} \sum_{j=1}^n E\left(\log\left[\frac{n}{2m}\right] + \log\{F(X_{(j+m)}) - F(X_{(j-m)})\}\right) \\
&= \frac{1}{n} n E\left(\log\left[\frac{n}{2m}\right]\right) + \frac{1}{n} \sum_{j=1}^n E\left(\log\{F(X_{(j+m)}) - F(X_{(j-m)})\}\right) \\
&= \log(n) - \log(2m) + \frac{1}{n} \sum_{j=1}^n E\left(\log\{F(X_{(j+m)}) - F(X_{(j-m)})\}\right).
\end{aligned}$$

Haciendo $Y_{(i)} = F(X_{(i)})$ para $1 - m \leq i \leq n + m$, tenemos que

$$E(U_{mn}) = \log(n) - \log(2m) + \frac{1}{n} \sum_{j=1}^n E(\log\{Y_{(j+m)} - Y_{(j-m)}\}). \quad (\text{B.3})$$

Analícemos

$$\sum_{j=1}^n E(\log\{Y_{(j+m)} - Y_{(j-m)}\}).$$

Claramente

$$\begin{aligned} \sum_{j=1}^n E(\log\{Y_{(j+m)} - Y_{(j-m)}\}) &= \sum_{j=1}^m E(\log\{Y_{(j+m)} - Y_{(1)}\}) + \\ &\quad \sum_{j=m+1}^{n-m} E(\log\{Y_{(j+m)} - Y_{(j-m)}\}) + \\ &\quad \sum_{j=n-m+1}^n E(\log\{Y_{(n)} - Y_{(j-m)}\}), \end{aligned}$$

luego, por la Proposición B.2 (b) tenemos

$$\begin{aligned} \sum_{j=1}^m E(\log\{Y_{(j+m)} - Y_{(1)}\}) &= \sum_{j=1}^m E(\log\{Y_{(1+(j+m-1))} - Y_{(1)}\}) \\ &= \sum_{j=1}^m \{\psi(j+m-1) - \psi(n+1)\}, \end{aligned}$$

$$\begin{aligned} \sum_{j=m+1}^{n-m} E(\log\{Y_{(j+m)} - Y_{(j-m)}\}) &= \sum_{j=m+1}^{n-m} E(\log\{Y_{((j-m)+2m)} - Y_{(j-m)}\}) \\ &= \sum_{j=m+1}^{n-m} \{\psi(2m) - \psi(n+1)\} \end{aligned}$$

y

$$\begin{aligned} \sum_{j=n-m+1}^n E(\log\{Y_{(n)} - Y_{(j-m)}\}) &= \sum_{j=n-m+1}^n E(\log\{Y_{((j-m)+(n+m-j))} - Y_{(j-m)}\}) \\ &= \sum_{j=n-m+1}^n \{\psi(n+m-j) - \psi(n+1)\} \\ &= \sum_{j=1}^m \psi(j+m-1) - \sum_{j=n-m+1}^n \psi(n+1). \end{aligned}$$

Así

$$\begin{aligned}
\sum_{j=1}^n E(\log\{Y_{(j+m)} - Y_{(j-m)}\}) &= \sum_{j=1}^m \{\psi(j+m-1) - \psi(n+1)\} + \\
&\quad \sum_{j=m+1}^{n-m} \{\psi(2m) - \psi(n+1)\} + \\
&\quad \sum_{j=1}^m \psi(j+m-1) - \sum_{j=n-m+1}^n \psi(n+1) \\
&= 2 \sum_{j=1}^m \psi(j+m-1) - n\psi(n+1) + \\
&\quad \sum_{j=m+1}^{n-m} \psi(2m) \\
&= 2 \sum_{j=1}^m \psi(j+m-1) - n\psi(n+1) + \\
&\quad (n-2m)\psi(2m),
\end{aligned}$$

por lo tanto, de (B.3) concluimos que

$$E(U_{mn}) = \log(n) - \log(2m) + \left(1 - \frac{2m}{n}\right)\psi(2m) - \psi(n+1) + \frac{2}{n} \sum_{j=1}^m \psi(j+m-1).$$

(c) Para la función digamma, es bien sabido que

$$\psi(x+1) = \frac{1}{x} + \psi(x), \tag{B.4}$$

lo que implica que la función ψ es estrictamente creciente sobre los enteros positivos. Anderson et al. (1995) probaron que la función $f(x) = x[\log(x) - \psi(x)]$ es estrictamente decreciente y estrictamente convexa sobre $(0, \infty)$. Además,

los autores presentaron demostraciones (complicadas) para $\lim_{x \rightarrow 0} f(x) = 1$ y $\lim_{x \rightarrow \infty} f(x) = 1/2$. En virtud de estos hechos tenemos que

$$\frac{1}{2x} < \log(x) - \psi(x) < \frac{1}{x}, \quad \text{para } x > 0. \quad (\text{B.5})$$

Ahora,

$$\begin{aligned} E(U_{mn}) &= \log(n) - \log(2m) + \left(1 - \frac{2m}{n}\right)\psi(2m) - \\ &\quad \psi(n+1) + \frac{2}{n} \sum_{j=1}^m \psi(j+m-1) \\ &= \log(n) - \log(2m) + \psi(2m) - \frac{2m}{n}\psi(2m) - \\ &\quad \frac{1}{n} - \psi(n) + \frac{2}{n} \sum_{j=1}^m \psi(j+m-1) \quad (\text{usando (B.4)}) \\ &= [\log(n) - \psi(n)] - [\log(2m) - \psi(2m)] - \\ &\quad \frac{1}{n} + \left[\frac{2}{n} \sum_{j=1}^m \psi(j+m-1) - \frac{2m}{n}\psi(2m)\right]. \end{aligned}$$

Cuando n y m tienden a infinito, por (B.5), es evidente que $[\log(n) - \psi(n)]$ y $[\log(2m) - \psi(2m)]$ tienden a cero respectivamente. Como obviamente $1/n$ converge a cero cuando n tiende a infinito, sólo tenemos que demostrar que

$$\left[\frac{2}{n} \sum_{j=1}^m \psi(j+m-1) - \frac{2m}{n}\psi(2m)\right] \quad (\text{B.6})$$

converge a cero, cuando $n \rightarrow \infty$, $m \rightarrow \infty$, $m/n \rightarrow 0$.

Debido a que

$$\begin{aligned}
0 &= \left[\frac{2m}{n} \psi(2m) - \frac{2m}{n} \psi(2m) \right] \\
&\geq \left[\frac{2}{n} \sum_{j=1}^m \psi(j+m-1) - \frac{2m}{n} \psi(2m) \right] \\
&\geq \left[\frac{2m}{n} \psi(m) - \frac{2m}{n} \psi(2m) \right] \\
&= -\frac{2m}{n} [\psi(2m) - \psi(m)],
\end{aligned}$$

demostrar que (B.6) converge a cero bajo las suposiciones ya mencionadas, se reduce a demostrar que

$$-\frac{2m}{n} [\psi(2m) - \psi(m)] \rightarrow 0$$

cuando $n \rightarrow \infty$, $m \rightarrow \infty$, $m/n \rightarrow 0$.

Ahora bien, por (B.5)

$$\frac{1}{2m} - \log(m) < -\psi(m) < \frac{1}{m} - \log(m)$$

y

$$\log(2m) - \frac{1}{2m} < \psi(2m) < \log(2m) - \frac{1}{4m},$$

lo que implica que

$$\log(2) < \psi(2m) - \psi(m) < \log(2) + \frac{3}{4m},$$

y así,

$$-\frac{2m}{n}[\psi(2m) - \psi(m)] \rightarrow 0$$

cuando $n, m \rightarrow \infty$ y bajo el supuesto de que $m/n \rightarrow 0$.

Todo lo anterior demuestra que $E(U_{mn}) \rightarrow 0$, cuando $n \rightarrow \infty$, $m \rightarrow \infty$, $m/n \rightarrow 0$.

□

Lema B.1. Si Y es una variable aleatoria no-negativa con fdp

$$f_Y(y) = \frac{2}{\sqrt{2\pi\xi}} \exp\left(-\frac{(y - \nu/y)^2}{2\xi^2}\right), \quad y \geq 0, \quad (\text{B.7})$$

entonces $X = 1/Y^2$ tiene la distribución $GI(\mu, \lambda)$, donde, $\mu = 1/\nu$ y $\lambda = 1/\xi^2$.

Bibliografía

- Anderson, G. D., Barnard, R. W., Richards, K. C., Vamanamurthy, M. K., and Vuorinen, M. (1995). Inequalities for Zero-Balanced Hypergeometric Functions. *Transactions of the American Mathematical Society*, 347:1713–1723.
- Edgeman, R. L. (1990). Assessing the Inverse Gaussian Distribution Assumption. *IEEE Transactions on Reliability*, 39(33):352–355.
- Edgeman, R. L., Scott, R. C., and Pavur, R. J. (1988). A Modified-Kolmogorov Smirnov Test for the Inverse Gaussian Density with Unknown Parameters. *Commun. Statist. B*, 17:1203–1212.
- Folks, J. L. and Chhikara, R. S. (1978). The Inverse Gaussian Distribution and Its Statistical Application—A Review. *Journal of the Royal Statistical Society of Great Britain*, 40:263–289.
- Gracia-Medrano, L. and O’Reilly, F. (2004). Transformations for Testing the Fit of the Inverse-Gaussian Distribution. *Commun. Statist. Theor. Meth.*, 33(4):919–924.
- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer.
- Mudholkar, G. S., Natarajan, R., and Chaubey, Y. P. (2001). A Goodness-of-Fit Test for the Inverse Gaussian Distribution Using Its Independence Characterization. *The Indian Journal of Statistics, Series B (1960-2002)*, 63(3):362–374.
- Mudholkar, G. S. and Tian, L. (2002). An entropy characterization of the inverse Gaussian distribution and related goodness-of-fit test. *Journal of Statistical Planning and Inference*, 102:211–221.

-
- O'Reilly, F. and Gracia-Medrano, L. (2006). On the conditional distribution of goodness-of-fit tests. *Commun. Statist. A*, 35:541–549.
- O'Reilly, F. and Rueda, R. (1992). Goodness of fit for the inverse Gaussian distribution. *Can. J. Statist.*, 20(4):387–397.
- Shannon, C. E. (1948). *A Mathematical Theory of Communication*. Bell Syst. Tech. J.
- Shannon, C. E. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
- Vasicek, O. (1976). A Test for Normality Based on Sample Entropy. *Journal of the Royal Statistical Society B*, 38:54–59.
- Vexler, A. and Gurevich, G. (2010). Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy. *Computational Statistics and Data Analysis*, 50:531–545.
- Vexler, A., Shan, G., Kim, S., Tsai, W.-M., Tian, L., and Hutson, A. D. (2011). An empirical likelihood ratio based goodness-of-fit test for Inverse Gaussian distributions. *Journal of Statistical Planning and Inference*, In Press. doi:10.1016/j.jspi.2010.12.024.