UNIVERSIDAD DE SEVILLA
Facultad de Matemáticas
Dpto. de Estadística e Investigación Operativa

TESTS DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN POISSON BIVARIANTE

Tesis Doctoral

Francisco Novoa Muñoz

Directora:

María Dolores Jiménez Gamero
Dpto. de Estadística e Investigación Operativa
Facultad de Matemáticas
Universidad de Sevilla

Sevilla, Julio de 2013

UNIVERSIDAD DE SEVILLA
Facultad de Matemáticas
Dpto. de Estadística e Investigación Operativa

TESTS DE BONDAD DE AJUSTE PARA LA DISTRIBUCIÓN POISSON BIVARIANTE

V^oB^o:	DIRECTORA DEL TRABAJO	Memoria presentada por Francisco Novoa Muñoz, para optar al grado de Doctor en Matemáticas.
Fdo.	María Dolores Jiménez Gamero. Profesora Titular de Universidad.	**Fdo.** Francisco Novoa Muñoz.

Sevilla, Julio de 2013

Agradecimientos

La culminación de este proyecto es gracias a la ayuda directa o indirecta de un gran número de personas. Sin lugar a dudas, un rol fundamental lo ha desempeñado mi Directora de Tesis, la profesora Doctora María Dolores Jiménez Gamero a quien doy las gracias por su tiempo, explicaciones, sugerencias, numerosas lecturas y correcciones de esta memoria, por su apoyo constante, alto nivel de conocimiento y claridad que tiene sobre los temas de matemática y de estadística.

Quiero destacar y agradecer el apoyo y gran ayuda del profesor Don Francisco Guillén González por sus innumerables explicaciones de algunos tópicos en matemática.

Agradezco la gran disposición y voluntad de todo el plantel de profesores del Departamento de Estadística e Investigación Operativa de la Universidad de Sevilla. Gracias por la claridad con que fueron impartidas las clases del curso de doctorado, así como también la importancia y pertinencia de los temas tratados.

Doy las gracias a mis colegas, quienes debieron ocuparse de mi carga académica en mi ausencia. También agradezco a las autoridades de la Universidad del Bío-Bío por el constante apoyo que me brindaron y al programa MECESUP, del gobierno de Chile, por otorgarme la beca de doctorado para "académicos en el extranjero". Todos ellos han permitido mi perfeccionamiento académico.

Agradezco y felicito la gran labor de secretaría llevada a cabo por Francisco Méndez y su colega Francisco Valverde. Siempre están dispuestos a ayudar y dar respuesta a las consultas.

Gracias doy también a mi familia porque siempre está conmigo y a Dios, que me otorgó la salud y fortaleza para terminar esta tesis.

A Florencia y a mis hijos:
Cristian y Constanza.

Índice general

Índice de tablas

Introducción

Los datos de conteo pueden aparecer bajo diferentes circunstancias. En un marco univariante, la distribución Poisson es la distribución que con mayor frecuencia ha sido empleada para modelar tales datos (ver por ejemplo, Haight (1967, pp. 100–107) [12], Johnson y Kotz (1969, pp. 88–90) [18], Sahai y Khurshid (1993) [41]).

En la práctica, los datos de conteo bivariantes surgen en varias disciplinas diferentes y la distribución Poisson bivariante (DPB), siendo una generalización de la distribución Poisson, juega un rol importante al momento de modelarlos, siempre que dichos datos presenten una correlación no negativa.

Esta distribución ha sido usada para modelar datos que aparecen en un amplio rango de campos incluyendo medicina, para mediciones en el pretratamiento y en el postratamiento en los mismos pacientes, o el número de accidentes por trabajador en una factoría durante dos intervalos dados de tiempo (ver por ejemplo, Hamdan (1972) [13]); en marketing, para el número de compras de diferentes productos; en epidemiología, para el número de incidentes de distintos tipos de muertes en una serie de distritos; en deportes, para el número de goles marcados por cada uno de los dos equipos oponentes en un partido de balompié (ver por ejemplo, Maher (1982) [32], Karlis y Ntzoufras (2000) [21], (2003a) [22], (2003b) [23], (2005) [24], Rue y Salvesen (2000) [39]); en biología, para el número de semillas y plantas que crecen en una parcela (ver por ejemplo, Lakshminarayana, Pandit y Rao (1999) [28]); en econometría, para el número de cambios de trabajo voluntarios e involuntarios (ver por ejemplo, Jung y Winkelmann (1993) [20]); en datos de turismo (ver por ejemplo, Berkhout y Plug (2004) [4]); en seguros de coches (ver por ejemplo, Bermúdez (2009) [5]); en sanidad (ver por ejemplo, Karlis y Ntzoufras (2005) [24], Karlis y Tsiamyrtzis (2008) [25]); en la industria textil, para el número de dos tipos de defectos en muestras de fibras de textil (ver por ejemplo, Ho y Singer (2001) [15]), entre muchos otros.

En el caso multivariante, la distribución Poisson ha sido usada para modelar redes sociales multi-relacionales (ver por ejemplo, Dai, Chua y Lim (2012) [7]).

Contrastar la bondad de ajuste de las observaciones dadas con un modelo probabilístico es un aspecto crucial del análisis de datos. Para el caso univariante, se han construido muchos tests de bondad de ajuste con la finalidad de comprobar si los datos provienen de una distribución Poisson (para una revisión detallada, ver Gürtler y Henze, (2000) [11]). En comparación, la literatura sobre tests de bondad de ajuste para la DPB es más bien escasa. Hasta donde conocemos, podemos mencionar el test propuesto por Crockett (1979) [6], el test desarrollado por Loukas y Kemp (1986) [30], que se basa en una extensión del índice de dispersión univariante, y el test sugerido por Rayner y Best (1995) [37], que consiste en una modificación del test dado por Loukas y Kemp (1986) [30]. La principal desventaja de estos tests de bondad de ajuste es que no son consistentes contra cada alternativa fija.

El objetivo de esta memoria es proponer y estudiar tests de bondad de ajuste para la DPB, que sean consistentes. Dado que la función generatriz de probabilidad (fgp) caracteriza la distribución de un vector aleatorio y se puede estimar consistentemente por la función generatriz de probabilidad empírica (fgpe), los tests que proponemos son funciones de la fgpe. El primer test estadístico compara la fgpe de los datos con un estimador de la fgp de la DPB. Luego, mostramos que la fgp de la DPB es la única fgp que satisface cierto sistema de ecuaciones diferenciales parciales, lo cual nos lleva a proponer dos tests estadísticos basados en el análogo empírico de dicho sistema, uno de ellos de tipo Cramér-von Mises y el otro se basa en los coeficientes de los polinomios de la versión empírica. Los tests que proponemos pueden ser vistos como extensiones al caso bivariante de algunos tests de bondad de ajuste diseñados para el caso univariante.

Con el fin de decidir cuándo rechazar la hipótesis nula, debemos conocer la distribución nula del test estadístico o, al menos, una aproximación de la misma. Puesto que, para los tests propuestos, no es posible obtener las distribuciones nulas para tamaños de muestra finito, las aproximamos por las asintóticas, las cuales resultaron depender de cantidades desconocidas, por lo tanto no son útiles como estimaciones de la distribución nula. Así, para aproximar la distribución nula, proponemos un estimador bootstrap paramétrico.

En cuanto a la potencia, obtuvimos que los tests que proponemos son consistentes contra alternativas fijas. Además, analizamos el comportamiento asintótico de los tests estadísticos bajo alternativas contiguas y encontramos que son capaces de detectar alternativas que convergen a la nula a razón de $n^{-1/2}$.

Todas las propiedades estudiadas de los tests introducidos en esta memoria son asintóticas, es decir, describen el comportamiento de los tests para muestras de tamaño grande. Con la finalidad de evaluar el comportamiento de los tests propuestos para muestras de tamaño finito, realizamos un estudio de simulación. En todos los casos considerados, el método bootstrap proporciona una buena aproximación a la distribución nula. En cuanto a la potencia, a diferencia de los tests estadísticos diseñados por los investigadores Crockett (1979) [6], Loukas y Kemp (1986) [30], y Rayner y Best (1995) [37], los tests que proponemos fueron capaces de detectar todas las alternativas seleccionadas.

Con el propósito de mantener la notación tan simple como sea posible, desarrollamos el análisis teórico para el caso bivariante, pero los métodos se pueden extender de manera natural al caso multivariante.

La presente memoria se organiza de la siguiente manera. En el Capítulo 1 presentamos algunos resultados preliminares que nos servirán en los capítulos siguientes, también damos la definición de la DPB con algunas de sus propiedades y además, mostramos contrastes de bondad de ajuste para la distribución Poisson tanto para datos univariantes como para datos bivariantes.

El Capítulo 2 contiene los dos primeros tests estadísticos que proponemos, que son de tipo Cramér-von Mises. Aquí, también mostramos la distribución asintótica nula de los tests estadísticos y proporcionamos estimadores bootstraps consistentes. En la parte final, estudiamos la potencia de los tests propuestos frente a alternativas fijas y locales.

En el Capítulo 3 presentamos el tercer estadístico que proponemos y describimos sus características. Estudiamos su distribución asintótica nula y aproximamos su distribución nula por medio de un estimador bootstrap consistente. También, analizamos su potencia frente a alternativas fijas y contiguas.

El Capítulo 4 está dedicado a mostrar los resultados de un estudio de simulación y la aplicación de los tests propuestos a dos conjuntos de datos reales. Dicho estudio de simulación fue realizado con el objetivo de evaluar el comportamiento de los tests que proponemos y comparar la potencia tanto entre ellos como con otros tests que encontramos en la literatura estadística.

En el Capítulo 5 entregamos las expresiones matemáticas de los tests que hemos desarrollado y damos algunos detalles técnicos que son muy útiles al momento de implementar algoritmos o subrutinas en algún lenguaje de programación.

El Capítulo 6 muestra cómo los tests propuestos se pueden extender al caso multivariante general.

Capítulo 1

Resultados previos y definiciones

1.1. Notación

- Todos los vectores a utilizar en esta memoria serán vectores filas y $v^\top$ es el traspuesto del vector fila v.
- Para cualquier vector v, v_k denota su $k-$ésima coordenada.
- $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$.
- $\Theta = \{\theta = (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3 \ : \ \theta_1 > \theta_3, \ \theta_2 > \theta_3, \ \theta_3 > 0\}$.
- $\langle \cdot , \cdot \rangle_V$ denota el producto escalar en el espacio vectorial V.
- $\| \cdot \|_V$ denota la norma en el espacio vectorial V y $\| \cdot \|$ denota la norma Euclidea de $\mathbb{R}^d$, $d \in \mathbb{N}$.
- $I\{A\}$ denota la función indicadora del conjunto A.
- P_ϑ denota la ley de probabilidades de una DPB con parámetro ϑ.
- E_ϑ denota la esperanza con respecto a la función de probabilidad P_ϑ.
- $\xrightarrow{L}$ denota convergencia en ley (o en distribución).
- $\xrightarrow{P}$ denota convergencia en probabilidad.
- $\xrightarrow{c.s.}$ denota convergencia casi segura (c.s.).

- Si $\{C_n\}$ es una sucesión de variables aleatorias (v.a.) y $\epsilon \in \mathbb{R}$, entonces
  - $C_n = O_{_P}(n^{-\epsilon})$ significa que $n^{\epsilon} C_n$ está acotada en probabilidad, cuando $n \to \infty$.
  - $C_n = o_{_P}(n^{-\epsilon})$ significa que $n^{\epsilon} C_n \xrightarrow{P} 0$, cuando $n \to \infty$.
  - $C_n = o(n^{-\epsilon})$ significa que $n^{\epsilon} C_n \xrightarrow{c.s.} 0$, cuando $n \to \infty$.

1.2. Resultados preliminares

Para demostrar la Proposición 1.2.2, dada a continuación, primero presentaremos un lema, que es una extensión del Lema 8.2.6 en Athreya y Lahiri (2006, p. 242) [1]. Previamente, consideremos la siguiente notación: para un conjunto arbitrario S, ∂S e $intS$ denotan los conjuntos de puntos frontera y puntos interiores de S, respectivamente.

Lema 1.2.1 *Sean $\{f_n\}_{n\geq 1}$ y f una colección de funciones reales y no decrecientes definidas sobre $Q = [b_1, c_1] \times [b_2, c_2] \times \cdots \times [b_d, c_d] \subseteq \mathbb{R}^d$, con $-\infty < b_j \leq c_j < \infty$, $j = 1, 2, \ldots, d$. Sea $D = D_1 \cup D_2 \cup D_3$, donde D_1 es el conjunto de vértices de Q, D_2 es un conjunto denso en ∂Q y D_3 es un conjunto denso en $intQ$. Si f es continua en Q y*

$$|f_n(x) - f(x)| \to 0, \quad \forall x \in D,$$

entonces

$$\sup_{x\in Q} |f_n(x) - f(x)| \to 0.$$

Demostración Sea $\varepsilon > 0$ arbitrario pero fijo. Como f es una función continua y Q es un conjunto compacto, entonces f es uniformemente continua en Q y por tanto

$$\exists \delta = \delta(\varepsilon) > 0 \text{ tal que } \forall x, y \in Q : \|x - y\| < \delta \implies |f(x) - f(y)| < \varepsilon. \tag{1.1}$$

Sean

$$H_i = [u_{1i}, v_{1i}] \times [u_{2i}, v_{2i}] \times \cdots \times [u_{di}, v_{di}], \quad i = 1, 2, \ldots, M,$$

con $\|u_i - v_i\| < \delta$, $u_i = (u_{1i}, u_{2i}, \ldots, u_{di}) \in D$ y $v_i = (v_{1i}, v_{2i}, \ldots, v_{di}) \in D$, $1 \leq i \leq M$, de modo que

$$Q \subseteq \bigcup_{i=1}^{M} H_i.$$

Consideremos

$$\Delta_n = \max_{x \in V} |f_n(x) - f(x)|, \tag{1.2}$$

donde $V = \{u_i, v_i : i = 1, 2, \ldots, M\}$. Por la convergencia puntual de $f_n(\cdot)$ a $f(\cdot)$ en D, obtenemos

$$\Delta_n \to 0. \tag{1.3}$$

Por otro lado, para cada $x \in Q$, existe $i \in \{1, 2, \ldots, M\}$, tal que $x \in H_i$, con lo cual, $u_i \leq x \leq v_i$. La monoticidad de las funciones f_n y f, junto con (1.1) y (1.2), implican

$$f_n(x) - f(x) \leq f_n(v_i) - f(u_i) = f_n(v_i) - f(v_i) + f(v_i) - f(u_i) \leq \Delta_n + \varepsilon.$$

Análogamente

$$f_n(x) - f(x) \geq -\Delta_n - \varepsilon.$$

De (1.3) y puesto que $\varepsilon > 0$ es arbitrario se logra

$$\sup_{x \in Q} |f_n(x) - f(x)| \to 0. \ \square$$

Feuerverger (1989) [8] demostró que la función generatriz de momentos empírica convergía c.s. a la función generatriz de momentos (fgm), en conjuntos compactos, para v.a.

Aunque existe una relación uno a uno entre la fgm y la fgp, hay problemas en el origen, por esta razón, probaremos a continuación, que la fgpe converge c.s. a la fgp, para el caso multivariante en general, en conjuntos de la forma $Q = [b_1, c_1] \times [b_2, c_2] \times \cdots \times [b_d, c_d]$, donde $0 \leq b_j \leq c_j < \infty,\ 1 \leq j \leq d,\ d \in \mathbb{N}$.

También probaremos, que las derivadas de la fgpe convergen c.s. a las derivadas de la fgp, en conjuntos Q como los definidos en el párrafo anterior. Sea $d \in \mathbb{N}$. Para cualquier función $h : S \subseteq \mathbb{R}^d \to \mathbb{R}$, denotaremos

$$D^{k_1 k_2 \cdots k_d} h(u) = \frac{\partial^k}{\partial u_1^{k_1} \partial u_2^{k_2} \cdots \partial u_d^{k_d}}\ h(u),$$

$\forall k_1, k_2, \ldots, k_d \in \mathbb{N}_0$ tal que $k = k_1 + k_2 + \cdots + k_d$.

Sean $\boldsymbol{X}_1 = (X_{11}, \ldots, X_{d1}), \boldsymbol{X}_2 = (X_{12}, \ldots, X_{d2}), \ldots, \boldsymbol{X}_n = (X_{1n}, \ldots, X_{dn})$ vectores aleatorios independientes e idénticamente distribuidos (iid) definidos sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y que toman valores en $\mathbb{N}_0^d$. En lo que sigue, sea

$$g_n(u) = \frac{1}{n}\sum_{i=1}^{n} u_1^{X_{1i}} u_2^{X_{2i}} \cdots u_d^{X_{di}}, \quad u \in W,$$

la fgpe de $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$, para algún conjunto apropiado $W \subseteq \mathbb{R}^d$. De manera más formal,

$$g_n(u,\omega) = \frac{1}{n}\sum_{i=1}^{n} u_1^{X_{1i}(\omega)} u_2^{X_{2i}(\omega)} \cdots u_d^{X_{di}(\omega)}, \quad u \in W, \quad \omega \in \Omega,$$

pero la dependencia de ω es usualmente suprimida.

Proposición 1.2.2 *Sean $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid de $\boldsymbol{X} = (X_1, \ldots, X_d) \in \mathbb{N}_0^d$. Sea $g(u) = E\Big(u_1^{X_1} u_2^{X_2} \cdots u_d^{X_d}\Big)$ la fgp de $\boldsymbol{X}$, definida sobre $W \subseteq \mathbb{R}^d$. Además, sean $0 \le b_j \le c_j < \infty,\ 1 \le j \le d$, tal que $Q = [b_1, c_1] \times [b_2, c_2] \times \cdots \times [b_d, c_d] \subseteq W$, entonces*

$$\sup_{u \in Q} |g_n(u) - g(u)| \overset{c.s.}{\longrightarrow} 0. \tag{1.4}$$

Si $D^{k_1 k_2 \cdots k_d} g(u)$ existe en Q, entonces

$$\sup_{u \in Q} \left| D^{k_1 k_2 \cdots k_d} g_n(u) - D^{k_1 k_2 \cdots k_d} g(u) \right| \overset{c.s.}{\longrightarrow} 0. \tag{1.5}$$

Demostración Sea D un conjunto denso numerable en Q, de acuerdo al Lema 1.2.1. Por la ley fuerte de los grandes números, existe un conjunto $A_D \in \mathcal{A}$ tal que $P(A_D) = 1$ y para todo $\omega \in A_D$:

$$g_n(u,\omega) \overset{c.s.}{\longrightarrow} g(u), \forall u \in D.$$

Puesto que $X_i \ge 0$ y $u_i \ge 0$, $1 \le i \le d$, tenemos que $g_n(u,\omega)$ y $g(u)$ son funciones no decrecientes.

Además, g es una función continua definida sobre Q.

Ahora, (1.4) se sigue del Lema 1.2.1.

La demostración de (1.5) sigue pasos similares a los dados para demostrar (1.4). □

Para probar algunos de nuestros resultados aplicaremos los lemas que establecemos a continuación, los que presentamos aquí para facilitar la lectura de nuestras demostraciones.

Lema 1.2.3 (Método Delta) *(Lehmann y Romano(2005, p. 436) [29]) Supongamos que X y $X_1, X_2, \ldots$ son vectores aleatorios en $\mathbb{R}^k$. Supongamos que $\tau_n(X_n - \mu) \xrightarrow{L} X$ donde μ es un vector constante y τ_n es una sucesión de constantes tales que $\tau_n \to \infty$.*

(*a*) *Supongamos que g es una función desde $\mathbb{R}^k$ a $\mathbb{R}$ que es diferenciable en μ con gradiente (vector de primeras derivadas parciales) de dimensión $1 \times k$ en μ igual a $\dot{g}(\mu) \neq \mathbf{0}$. Entonces,*

$$\tau_n \left[g(X_n) - g(\mu) \right] \xrightarrow{L} \dot{g}(\mu) X^\top.$$

En particular, si X es normal multivariante en $\mathbb{R}^k$ con vector de media $\mathbf{0}$ y matriz de covarianzas Σ, entonces

$$\tau_n \left[g(X_n) - g(\mu) \right] \xrightarrow{L} N\left(\mathbf{0}, \dot{g}(\mu)\, \Sigma\ \dot{g}(\mu)^\top\right).$$

(*b*) *Más generalmente, supongamos que $g = (g_1, g_2, \ldots, g_q)^\top$ es una función desde $\mathbb{R}^k$ a $\mathbb{R}^q$, donde g_i es una función desde $\mathbb{R}^k$ a $\mathbb{R}$ que es diferenciable en μ. Sea D una matriz no nula de orden $q \times k$, donde el elemento (i,j) es $\dfrac{\partial}{\partial y_j} g_i(y_1, y_2, \ldots, y_k)$ evaluada en μ. Entonces*

$$\tau_n \left[g(X_n) - g(\mu) \right] = \tau_n \left[g_1(X_n) - g_1(\mu), \ldots, g_q(X_n) - g_q(\mu) \right]^\top \xrightarrow{L} D X^\top.$$

En particular, si X es normal multivariante en $\mathbb{R}^k$ con vector de media $\mathbf{0}$ y matriz de covarianzas Σ, entonces

$$\tau_n \left[g(X_n) - g(\mu) \right] \xrightarrow{L} N\left(\mathbf{0}, D \Sigma D^\top\right).$$

Para los lemas siguientes, con H denotaremos un espacio de Hilbert de dimensión infinita, separable y real.

Lema 1.2.4 *(Teorema central del límite en espacios de Hilbert, van der Vaart y Wellner (1996, pp. 50–51)[44]) Si $X_1, X_2, X_3, \ldots$ son elementos aleatorios iid medibles Borel en un espacio de Hilbert H con media cero (es decir, $E\left(\langle X_1, h \rangle_H\right) = 0$ para cada h), y $E\left(\|X_1\|_H^2\right) < \infty$, entonces la sucesión $\dfrac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ converge en distribución a la variable Gaussiana G. La distribución de G está determinada por la distribución de sus marginales $\langle G, h \rangle_H$, que se distribuyen según una ley $N\left(0, E\left(\langle X, h \rangle_H^2\right)\right)$ para cada $h \in H$.*

Lema 1.2.5 *(Teorema 1.1 en Kundu et al. (2000) [27]) Sea* $\{e_k : k \geq 0\}$ *una base ortonormal de* H*. Para cada* $n \geq 1$*, sea* $Y_{n1}, Y_{n2}, \ldots, Y_{nn}$ *un arreglo triangular de elementos aleatorios independientes* H*-valuados con medias cero y momentos segundos finitos, sea* $Y_n = \sum_{j=1}^{n} Y_{nj}$*. Sea* C_n *el operador de covarianza de* Y_n*. Supongamos que se cumplen las siguientes condiciones:*

(i) $\lim_{n\to\infty} \langle C_n e_k, e_l \rangle_H = a_{kl}$*, existe para todo* $k, l \geq 0$*.*

(ii) $\lim_{n\to\infty} \sum_{k=0}^{\infty} \langle C_n e_k, e_k \rangle_H = \sum_{k=0}^{\infty} a_{kk} < \infty$*.*

(iii) $\lim_{n\to\infty} L_n(\varepsilon, e_k) = 0$ *para cada* $\varepsilon > 0$ *y cada* $k \geq 0$*, donde, para* $b \in H$*,*

$$L_n(\varepsilon, b) = \sum_{j=1}^{n} E\left(\langle Y_{nj}, b \rangle_H^2 \, I\{|\langle Y_{nj}, b \rangle_H| > \varepsilon\} \right).$$

Entonces

$$Y_n \xrightarrow{L} N(0, C),$$

en H*, donde el operador de covarianza* C *es caracterizado por* $\langle Cf, e_l \rangle_H = \sum_{k=0}^{\infty} \langle f, e_k \rangle_H a_{kl}$*, para cada* $l \geq 0$*.*

1.3. Definición de la distribución Poisson bivariante

Se han dado varias definiciones para la DPB (ver por ejemplo, Kocherlakota y Kocherlakota (1992, pp. 87–90) [26], para una revisión detallada). En esta memoria, consideraremos la siguiente, que es la que ha recibido más atención en la literatura estadística (ver por ejemplo, Holgate (1964) [16]; Johnson, Kotz y Balakrishnan (1997) [19]).

Definición 1.3.1 *(Johnson, Kotz y Balakrishnan (1997, pp. 124–125) [19]) Sean*

$$X_1 = Y_1 + Y_3 \quad \text{y} \quad X_2 = Y_2 + Y_3,$$

donde Y_1*,* Y_2 *e* Y_3 *son v.a. Poisson, mutuamente independientes con medias dadas por* $\theta_1' = \theta_1 - \theta_3 > 0$*,* $\theta_2' = \theta_2 - \theta_3 > 0$ *y* $\theta_3 \geq 0$*, respectivamente.*

A la distribución conjunta del vector (X_1, X_2) *se le denomina* **distribución Poisson bivariante** *(DPB) con parámetro* $\theta = (\theta_1, \theta_2, \theta_3)$*, lo cual denotaremos mediante* $(X_1, X_2) \sim PB(\theta_1, \theta_2, \theta_3)$ *o simplemente* $(X_1, X_2) \sim PB(\theta)$*.*

La función de probabilidad conjunta de X_1 *y* X_2 *está dada por*

$$P_\theta(X_1 = x_1, X_2 = x_2) = \exp(\theta_3 - \theta_1 - \theta_2) \sum_{i=0}^{\min\{x_1,x_2\}} \frac{(\theta_1 - \theta_3)^{x_1-i}\,(\theta_2 - \theta_3)^{x_2-i}\,\theta_3^i}{(x_1-i)!\,(x_2-i)!\,i!},$$

donde $x_1, x_2 \in \mathbb{N}_0$*.*

Observación 1.3.2 *Si* $\theta_3 = 0$*, entonces* $X_1 = Y_1$ *y* $X_2 = Y_2$*, y por tanto* X_1 *y* X_2 *son independientes, pues* Y_1 *e* Y_2 *son v.a. Poisson mutuamente independientes. En los Capítulos 2 y 3 supondremos que* $\theta_3 > 0$*. El caso* $\theta_3 = 0$ *será descrito en el Capítulo 6.*

Tal como sucede en el caso univariante, una de las formas como se obtuvo la función de probabilidad conjunta de la DPB fue como el límite de la distribución binomial bivariante.

En primer lugar daremos la definición de la distribución binomial bivariante y luego presentaremos el resultado que aproxima la distribución binomial bivariante a la DPB, cuya demostración puede verse en Hamdan y Al-Bayyati (1969) [14].

Definición 1.3.3 *(Johnson, Kotz y Balakrishnan (1997, p. 125) [19]) Supongamos que cada individuo de una población es clasificado ya sea como* A *o* A^c *y simultáneamente como* B *o* B^c*, con probabilidades dadas por*

	B	B^c	
A	p_{11}	p_{10}	p_1
A^c	p_{01}	p_{00}	q_1
	p_2	q_2	1

Consideremos una muestra aleatoria de tamaño n*, seleccionada con reemplazo de la población anterior. Sean las v.a.:*

X_1 = *número de individuos que son clasificados como* A*,*

X_2 = *número de individuos que son clasificados como* B*.*

Estas v.a. tienen distribución binomial bivariante conjunta, con función de probabilidad

$$P(X_1 = x_1, X_2 = x_2) = \sum_{k=0}^{\min\{x_1,x_2\}} \frac{n!\ p_{11}^{k}\, p_{10}^{x_1-k}\, p_{01}^{x_2-k}\, p_{00}^{n+k-x_1-x_2}}{k!\ (x_1-k)!\ (x_2-k)!\ (n+k-x_1-x_2)!}\,,$$

lo que se representará por $(X_1, X_2) \sim BB(n; p_{10}, p_{01}, p_{11})$.

Teorema 1.3.4 *(Hamdan y Al-Bayyati (1969) [14])*
Sea $(X_1, X_2) \sim BB(n; p_{10}, p_{01}, p_{11})$*. Suponer que* $p_{10}, p_{01}, p_{11} \to 0$*, cuando* $n \to \infty$*, de modo que* $np_{10} = \theta_1 - \theta_3$*,* $np_{01} = \theta_2 - \theta_3$ *y* $np_{11} = \theta_3$*. Entonces*

$$\begin{aligned}\lim_{n\to\infty} & P(X_1 = x_1, X_2 = x_2) \\ &= \exp(\theta_3 - \theta_1 - \theta_2) \sum_{k=0}^{\min\{x_1,x_2\}} \frac{(\theta_1-\theta_3)^{x_1-k}(\theta_2-\theta_3)^{x_2-k}\,\theta_3^{k}}{(x_1-k)!\,(x_2-k)!\,k!},\ \ x_1, x_2 \in \mathbb{N}_0.\end{aligned}$$

1.4. Algunas características y propiedades de la DPB

A continuación presentamos ciertos resultados de la DPB que nos serán de utilidad en el desarrollo de esta memoria. Quizás el de mayor importancia es el dado a continuación, pues es la base de nuestro trabajo.

Para ello consideremos el vector aleatorio $(X_1, X_2) \sim PB(\theta)$, como el establecido en la Definición 1.3.1.

1.4.1. Función generatriz de probabilidad

Para deducir la fgp de la DPB, recordemos que la fgp de una v.a. X que se distribuye Poisson (univariante) con parámetro $\lambda > 0$, se define y calcula mediante

$$E_\lambda\big(t^X\big) = \sum_{x=0}^{\infty} t^x \frac{\lambda^x \exp(-\lambda)}{x!} = \exp(-\lambda) \sum_{x=0}^{\infty} \frac{(\lambda t)^x}{x!} = \exp\{\lambda(t-1)\},\ \forall\, t \in \mathbb{R}.$$

Ahora, como las v.a. Y_1, Y_2 e Y_3 se distribuyen según una ley de Poisson, entonces sus fgp se pueden expresar por

$$E_{\theta_1'}\big(t^{Y_1}\big) = \exp\{\theta_1'(t-1)\},\ \ E_{\theta_2'}\big(t^{Y_2}\big) = \exp\{\theta_2'(t-1)\},\ \ E_{\theta_3}\big(t^{Y_3}\big) = \exp\{\theta_3(t-1)\},\ \forall\, t \in \mathbb{R}.$$

Además, como las v.a. Y_1, Y_2 e Y_3 son mutuamente independientes, entonces, la fgp conjunta, $g(u;\theta)$, de la DPB se obtiene mediante

$$
\begin{aligned}
g(u;\theta) &= E_\theta\big(u_1^{X_1}\, u_2^{X_2}\big)\,, \\
&= E_\theta\big(u_1^{Y_1+Y_3}\, u_2^{Y_2+Y_3}\big) = E_{\theta_1'}\big(u_1^{Y_1}\big)\, E_{\theta_2'}\big(u_2^{Y_2}\big)\, E_{\theta_3}\big\{(u_1u_2)^{Y_3}\big\} \\
&= \exp\{\theta_1'(u_1-1)+\theta_2'(u_2-1)+\theta_3(u_1u_2-1)\}, \\
&= \exp\{(\theta_1-\theta_3)(u_1-1)+(\theta_2-\theta_3)(u_2-1)+\theta_3(u_1u_2-1)\}, \\
&= \exp\{\theta_1(u_1-1)+\theta_2(u_2-1)+\theta_3(u_1-1)(u_2-1)\}, \qquad (1.6)
\end{aligned}
$$

$\forall\, u=(u_1,u_2)\in\mathbb{R}^2$, $\forall\,\theta\in\Theta$.

1.4.2. Función generatriz de momentos

Recordemos que la fgm de una v.a. X que se distribuye según una ley de Poisson (univariante) con parámetro $\lambda>0$, se define y calcula mediante

$$
M_X(t) = E_\lambda\{\exp(tX)\} = \exp(-\lambda)\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} = \exp\big\{\lambda\left(e^t-1\right)\big\}\,,\ \forall\, t\in\mathbb{R}.
$$

Puesto que las v.a. Y_1, Y_2 e Y_3 se distribuyen según una ley de Poisson, entonces sus fgm se pueden escribir como

$$
M_{Y_1}(t)=\exp\big\{\theta_1'\big(e^t-1\big)\big\},\ M_{Y_2}(t)=\exp\big\{\theta_2'\big(e^t-1\big)\big\},\ M_{Y_3}(t)=\exp\big\{\theta_3\big(e^t-1\big)\big\}, \forall\, t\in\mathbb{R}.
$$

Además, como las v.a. Y_1, Y_2 e Y_3 son mutuamente independientes, entonces, la fgm conjunta, se obtiene como

$$
\begin{aligned}
M_{(X_1,X_2)}(u_1,u_2) &= E_\theta\{\exp(u_1X_1+u_2X_2)\} \\
&= E_\theta[\exp\{u_1Y_1+u_2Y_2+(u_1+u_2)Y_3\}] \\
&= E_{\theta_1'}\big(e^{u_1Y_1}\big)\, E_{\theta_2'}\big(e^{u_2Y_2}\big)\, E_{\theta_3}\big\{e^{(u_1+u_2)Y_3}\big\} \\
&= \exp\big\{\theta_1'(e^{u_1}-1)+\theta_2'(e^{u_2}-1)+\theta_3(e^{u_1+u_2}-1)\big\}\,, \\
&= \exp\big\{(\theta_1-\theta_3)(e^{u_1}-1)+(\theta_2-\theta_3)(e^{u_2}-1)+\theta_3\big(e^{u_1+u_2}-1\big)\big\}\,, \\
&= \exp\{\theta_1(e^{u_1}-1)+\theta_2(e^{u_2}-1)+\theta_3(e^{u_1}-1)(e^{u_2}-1)\}\,,
\end{aligned}
$$

$\forall\, u = (u_1, u_2) \in \mathbb{R}^2$, $\forall\, \theta \in \Theta$.

1.4.3. Momentos

Recordemos que si una v.a. X se distribuye según una ley de Poisson (univariante) con parámetro $\lambda > 0$, entonces su $k-$ésimo momento en torno al cero es

$$\mu_k^{'} = E_\lambda\big(X^k\big) = \sum_{i=0}^{k} \lambda^i\, S(k,i),$$

donde $S(a,b)$ es llamado **número Stirling de segundo tipo** y satisface las relaciones

$$S(n,0) = 0,\ S(0,0) = S(n,1) = S(n,n) = 1,\ \text{para } n \in \mathbb{N},$$

$$S(n,j) = S(n-1,j-1) + jS(n-1,j),\ \text{para } j = 1,2,\ldots,n-1.$$

En particular, para $k = 1,2,3$, obtenemos

$$E_\lambda(X) = \lambda, \quad E_\lambda\big(X^2\big) = \lambda + \lambda^2, \quad E_\lambda\big(X^3\big) = \lambda + 3\lambda^2 + \lambda^3.$$

El correspondiente momento central de X es

$$\mu_k = E_\lambda\big\{(X-\lambda)^k\big\} = \sum_{i=0}^{k} \binom{k}{i} (-\lambda)^{k-i}\, \mu_i^{'}.$$

Para el modelo Poisson bivariante, el $r = (r_1, r_2)-$ésimo momento en torno al origen es

$$\begin{aligned}
\mu_{r_1,r_2}^{'} &= \mu_r^{'}(X_1,X_2) = E_\theta(X_1^{r_1} X_2^{r_2}) = E_\theta\{(Y_1+Y_3)^{r_1}(Y_2+Y_3)^{r_2}\} \\
&= E_\theta\left\{ \sum_{i_1=0}^{r_1} \binom{r_1}{i_1} Y_1^{i_1}\ Y_3^{r_1-i_1} \sum_{i_2=0}^{r_2} \binom{r_2}{i_2} Y_2^{i_2}\ Y_3^{r_2-i_2} \right\} \\
&= \sum_{i_1=0}^{r_1} \sum_{i_2=0}^{r_2} \binom{r_1}{i_1}\binom{r_2}{i_2} E_{\theta_1'}\big(Y_1^{i_1}\big)\ E_{\theta_2'}\big(Y_2^{i_2}\big)\ E_{\theta_3}\big(Y_3^{r_1+r_2-i_1-i_2}\big) \\
&= \sum_{i_1=0}^{r_1} \sum_{i_2=0}^{r_2} \binom{r_1}{i_1}\binom{r_2}{i_2} \sum_{j_1=0}^{i_1} (\theta_1-\theta_3)^{j_1} S(i_1,j_1) \sum_{j_2=0}^{i_2} (\theta_2-\theta_3)^{j_2} S(i_2,j_2) \\
&\qquad \times \sum_{j_3=0}^{r_1+r_2-i_1-i_2} \theta_3^{j_3} S(r_1+r_2-i_1-i_2, j_3).
\end{aligned}$$

En particular,

$$
\begin{aligned}
E_\theta(X_k) &= \theta_k, \quad \text{para } k = 1, 2,\\
E_\theta(X_k^2) &= \theta_k + \theta_k^2, \quad \text{para } k = 1, 2,\\
E_\theta(X_1\, X_2) &= \theta_1\, \theta_2 + \theta_3,\\
E_\theta(X_1^2\, X_2) &= \theta_1\, \theta_2 + \theta_1^2\, \theta_2 + 2\, \theta_1\, \theta_3 + \theta_3,\\
E_\theta(X_1\, X_2^2) &= \theta_1\, \theta_2 + \theta_1\, \theta_2^2 + 2\, \theta_2\, \theta_3 + \theta_3,\\
E_\theta(X_1^2\, X_2^2) &= \theta_1\, \theta_2 + \theta_1\, \theta_2^2 + \theta_1^2\, \theta_2 + \theta_1^2\, \theta_2^2 + 4\, \theta_1\, \theta_2\, \theta_3 + 2\, \theta_1\, \theta_3 + 2\, \theta_2\, \theta_3 + \theta_3 + 2\, \theta_3^2.
\end{aligned}
$$

Observación 1.4.1 *De los resultados dados en las ecuaciones anteriores se sigue que la DPB tiene momentos de todos los órdenes.*

Escribiremos el correspondiente $r = (r_1, r_2)-$ésimo momento central para el modelo Poisson bivariante mediante

$$
\begin{aligned}
\mu_{r_1,r_2} = \mu_r(X_1, X_2) &= E_\theta[\{X_1 - E_\theta(X_1)\}^{r_1}\{X_2 - E_\theta(X_2)\}^{r_2}]\\
&= E_\theta[\{X_1 - \theta_1\}^{r_1}\{X_2 - \theta_2\}^{r_2}]\\
&= \sum_{i=0}^{r_1}\sum_{j=0}^{r_2}\binom{r_1}{i}\binom{r_2}{j}(-\theta_1)^{r_1-i}(-\theta_2)^{r_2-j}E_\theta\big(X_1^i X_2^j\big)\\
&= \sum_{i=0}^{r_1}\sum_{j=0}^{r_2}\binom{r_1}{i}\binom{r_2}{j}(-\theta_1)^{r_1-i}(-\theta_2)^{r_2-j}\mu_{i,j}^{'}.
\end{aligned}
$$

Por otra parte, como las v.a. X_1 y X_2 se distribuyen según una ley de Poisson (univariante), entonces sus varianzas son

$$\operatorname{var}(X_k) = \theta_k,\ \ k = 1, 2.$$

Además, como Y_1, Y_2 e Y_3 son variables independientes, la covarianza entre X_1 y X_2 es

$$
\begin{aligned}
\operatorname{cov}(X_1, X_2) &= \operatorname{cov}(Y_1 + Y_3, Y_2 + Y_3)\\
&= \operatorname{cov}(Y_1, Y_2) + \operatorname{cov}(Y_1, Y_3) + \operatorname{cov}(Y_3, Y_2) + \operatorname{cov}(Y_3, Y_3)\\
&= \operatorname{var}(Y_3) = \theta_3\,.
\end{aligned}
$$

Por lo tanto, el coeficiente de correlación entre X_1 y X_2 es

$$\rho = \operatorname{corr}(X_1, X_2) = \frac{\operatorname{cov}(X_1, X_2)}{\sqrt{\operatorname{var}(X_1)\ \operatorname{var}(X_2)}} = \frac{\theta_3}{\sqrt{\theta_1\,\theta_2}}\,.$$

Este coeficiente de correlación no puede exceder de $\theta_3\,(\text{mín}\{\theta_1,\theta_2\})^{-1}$, porque

$$\text{mín}\,\{\theta_1,\theta_2\} \leq \theta_k,\ \ k=1,2,\ \text{ luego } \theta_1\,\theta_2 \geq (\text{mín}\,\{\theta_1,\theta_2\})^2\,.$$

O bien, como lo señaló Holgate (1964) [16],

$$0 < \rho \leq \text{mín}\left(\sqrt{\frac{\theta_1}{\theta_2}}, \sqrt{\frac{\theta_2}{\theta_1}}\right),$$

pues, de la Definición 1.3.1

$$\sqrt{\frac{\theta_1}{\theta_2}} = \sqrt{\frac{\theta_1}{\theta_2}\frac{\theta_1}{\theta_1}} = \frac{\theta_1}{\sqrt{\theta_1\theta_2}} = \frac{\theta_1' + \theta_3}{\sqrt{\theta_1\theta_2}} \geq \rho,$$

de igual forma se consigue que $\rho \leq \sqrt{\frac{\theta_2}{\theta_1}}$.

1.4.4. Fórmulas de recursión

Para calcular numéricamente los valores de $P(r,s;\theta) = P_\theta(X_1 = r, X_2 = s)$, son útiles las relaciones de recurrencia presentadas por Johnson, Kotz y Balakrishnan (1997, p. 125) [19] que se enuncian como sigue

Proposición 1.4.2 *(Johnson et al. (1997, p. 125) [19]) Si* $(X_1, X_2) \sim PB(\theta_1,\theta_2,\theta_3)$, *entonces*

$$\begin{aligned} rP(r,s;\theta) &= (\theta_1-\theta_3)P(r-1,s;\theta) + \theta_3\,P(r-1,s-1;\theta),\\ sP(r,s;\theta) &= (\theta_2-\theta_3)\,P(r,s-1;\theta) + \theta_3\,P(r-1,s-1;\theta), \qquad (1.7)\\ &\textit{Si } r<0 \quad \textit{o} \ \ s<0,\ \textit{entonces considerar } P(r,s;\theta)=0. \end{aligned}$$

Además de las relaciones dadas en la Proposición anterior, también nos serán de gran utilidad las relaciones de recurrencia dadas en Johnson, Kotz y Balakrishnan (1997, p. 127) [19], que se enuncian como sigue.

Proposición 1.4.3 *Sea* $(X_1, X_2) \sim PB(\theta_1, \theta_2, \theta_3)$, *entonces*

$$\begin{aligned}
\frac{\partial}{\partial\theta_1}P(r,s;\theta) &= P(r-1,s;\theta) - P(r,s;\theta),\\
\frac{\partial}{\partial\theta_2}P(r,s;\theta) &= P(r,s-1;\theta) - P(r,s;\theta) \qquad (1.8)\\
\frac{\partial}{\partial\theta_3}P(r,s;\theta) &= P(r-1,s-1;\theta) - P(r-1,s;\theta) - P(r,s-1;\theta) + P(r,s;\theta).
\end{aligned}$$

1.4.5. Estimación puntual

Los estimadores más usados comúnmente son los obtenidos por los siguientes métodos:

- Método de máxima verosimilitud (Kocherlakota y Kocherlakota (1992, pp. 43-45) [26]).

  Propiedades Según Kocherlakota y Kocherlakota (1992, p. 45) [26], el estimador de máxima verosimilitud, $\hat{\theta}_{MV}$, es consistente, asintóticamente normal y asintóticamente eficiente para θ.

  - El estimador máximo verosímil de θ debe satisfacer

    $$\hat{\theta}_k = \bar{X}_k = \frac{1}{n}\sum_{i=1}^{n} X_{ki}\,, \quad k = 1, 2 \quad \text{y} \quad \bar{R} = 1,$$

    donde

    $$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} \frac{P(X_{1i}-1, X_{2i}-1;\theta)}{P(X_{1i}, X_{2i};\theta)}\,.$$

    Observación 1.4.4 *El estimador máximo verosímil del parámetro* θ_3 *no tiene una forma explícita y por lo tanto debe ser calculado por métodos numéricos utilizando la ecuación* $\bar{R} = 1$. *En nuestro caso, para calcular el estimador* $\hat{\theta}_3$ *empleamos el método iterativo de Newton-Raphson estándar.*

    La matriz de varianzas y covarianzas asintótica de los estimadores máximo verosímiles, Σ_{MV}, está dada por

    $$\Sigma_{MV} = \begin{pmatrix} \theta_1 & \theta_3 & \theta_3 \\ \theta_3 & \theta_2 & \theta_3 \\ \theta_3 & \theta_3 & \dfrac{\theta_3^2\,(\theta_1+\theta_2-2\theta_3)\,(Q-1) - \theta_3^2 + (\theta_1-2\theta_3)(\theta_2-2\theta_3)}{(\theta_1\theta_2-\theta_3^2)(Q-1) - \theta_1 - \theta_2 + 2\theta_3} \end{pmatrix},$$

donde

$$Q = \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \frac{P^2(r-1, s-1; \theta)}{P(r, s; \theta)}. \tag{1.9}$$

- Por las propiedades de los estimadores máximo verosímiles, se tiene

$$\sqrt{n}\left(\hat{\theta}_{MV} - \theta\right) \xrightarrow{L} N_3(\mathbf{0}, \Sigma_{MV}).$$

- Método de los momentos (Kocherlakota y Kocherlakota (1992, pp. 34-35) [26]).

    - El estimador de θ por el método de los momentos, $\hat{\theta}_{MM}$, es consistente y está dado por

    $$\hat{\theta}_k = \bar{X}_k,\ k = 1, 2 \quad \text{y} \quad \hat{\theta}_3 = s_{11} = \frac{1}{n}\sum_{i=1}^{n} X_{1i}X_{2i} - \bar{X}_1\bar{X}_2.$$

    Por otro lado, la matriz de varianzas y covarianzas asintótica de los estimadores por el método de los momentos, Σ_{MM}, está dada por

    $$\Sigma_{MM} = \begin{pmatrix} \theta_1 & \theta_3 & \theta_3 \\ \theta_3 & \theta_2 & \theta_3 \\ \theta_3 & \theta_3 & \theta_1\theta_2 + \theta_3 + \theta_3^2 \end{pmatrix}.$$

    - Por ser $\hat{\theta}_{MM}$ función de momentos muestrales, aplicando el Método Delta (Teorema 1.2.3) se tiene que

    $$\sqrt{n}\left(\hat{\theta}_{MM} - \theta\right) \xrightarrow{L} N_3(\mathbf{0}, \Sigma_{MM}).$$

- Método del doble cero (Kocherlakota y Kocherlakota (1992, pp. 42-43) [26]).

    - El estimador de θ por el método del doble cero, $\hat{\theta}_{DC}$, es consistente y está dado por

    $$\hat{\theta}_k = \bar{X}_k,\ k = 1, 2 \quad \text{y} \quad \hat{\theta}_3 = \bar{X}_1 + \bar{X}_2 + \log\hat{\phi},$$

    donde $\phi = \exp(\theta_3 - \theta_1 - \theta_2)$ y $\hat{\phi}$ es la proporción observada de los datos $(0, 0)$ en la muestra.

    La matriz de varianzas y covarianzas asintótica de los estimadores por el método del doble cero, Σ_{DC}, está dada por

    $$\Sigma_{DC} = \begin{pmatrix} \theta_1 & \theta_3 & -\theta_1\,\phi \\ \theta_3 & \theta_2 & -\theta_2\,\phi \\ -\theta_1\,\phi & -\theta_2\,\phi & \phi\,(1-\phi) \end{pmatrix}.$$

- Por ser $\hat{\theta}_{DC}$ función de momentos muestrales, aplicando el Método Delta (Teorema 1.2.3) se tiene que

$$\sqrt{n}\left(\hat{\theta}_{DC} - \theta\right) \xrightarrow{L} N_3(\mathbf{0}, \Sigma_{DC}).$$

- Método de los puntos pares (Kocherlakota y Kocherlakota (1992, pp. 40-41) [26]).

  - El estimador de θ por el método de los puntos pares, $\hat{\theta}_{PP}$, es consistente y está dado por

$$\hat{\theta}_1 = \bar{X}_1, \quad \hat{\theta}_2 = \bar{X}_2 \quad \text{y} \quad \hat{\theta}_3 = \frac{1}{2}\left(\bar{X}_1 + \bar{X}_2\right) + \frac{1}{4}\log\left(\frac{2\hat{A}}{n} - 1\right), \text{ si } \hat{A} > \frac{n}{2},$$

donde $2A = n\left[1 + \exp\{-2(\theta_1 + \theta_2 - 2\theta_3)\}\right]$ y $\hat{A}$ es la suma de las frecuencias de ocurrencia de los pares (X_{1i}, X_{2i}), $i = 1, 2, \ldots, n$ para los cuales el valor de ambas variables tiene la misma paridad, es decir, en ambas variables el valor es par o en ambas variables el valor es impar.

La matriz de varianzas y covarianzas asintótica de los estimadores por el método de los puntos pares, Σ_{PP}, está dada por

$$\Sigma_{PP} = \begin{pmatrix} \theta_1 & \theta_3 & \theta_3 \\ \theta_3 & \theta_2 & \theta_3 \\ \theta_3 & \theta_3 & \frac{1}{4}(6\theta_3 - \theta_1 - \theta_2) + \frac{1}{16}[\exp\{4(\theta_1 + \theta_2 - 2\theta_3)\} - 1] \end{pmatrix}.$$

  - Por ser $\hat{\theta}_{PP}$ función de momentos muestrales, aplicando el Método Delta (Teorema 1.2.3) se tiene que

$$\sqrt{n}\left(\hat{\theta}_{PP} - \theta\right) \xrightarrow{L} N_3(\mathbf{0}, \Sigma_{PP}).$$

- Método de los puntos pares condicionados (Papageorgiou y Loukas (1988) [34]).

  - El estimador de θ por el método de los puntos pares condicionados, $\hat{\theta}_{PC}$, es consistente y está dado por

$$\hat{\theta}_1 = \bar{X}_1, \quad \hat{\theta}_2 = \bar{X}_2 \quad \text{y} \quad \hat{\theta}_3 = \bar{X}_2 + \frac{1}{2}\log\left(\frac{2\hat{A}}{\hat{A} + \hat{B}} - 1\right), \text{ si } \hat{A} > \hat{B},$$

donde $2A = B\left[1 + \exp\{2(\theta_3 - \theta_2)\}\right]$, A y B son las sumas de las frecuencias observadas en los puntos $(0, 2s)$ y $(0, 2s+1)$, $s \in \mathbb{N}_0$, respectivamente.

La matriz de varianzas y covarianzas asintótica de los estimadores por el método de los puntos pares condicionados, Σ_{PC}, está dada por

$$\Sigma_{PC} = \frac{1}{4}\begin{pmatrix} C & D & E \\ D & F & F-G+H \\ E & F-G+H & F+2(H-G)+J \end{pmatrix},$$

donde

$$\begin{aligned}
&\alpha = \exp(-\theta_1), && \beta = \exp\{2(\theta_3-\theta_2)\}, \\
&C = \alpha(\beta+1)\{2+\alpha(\beta+1)\}, && D = \alpha^2\left(\beta^2-1\right), \\
&E = \frac{D(\theta_1+\alpha^2\beta)+2\theta_3\alpha^2(\beta+1)-\alpha^2\beta H}{\alpha^2\beta}, && F = \alpha(1-\beta)\{2+\alpha(\beta-1)\}, \\
&G = \beta^{-1}\theta_1(\beta-1)^2, && H = (\beta+1)\{\theta_3(\beta^{-1}+1)-2\theta_2\}, \\
&J = \frac{\theta_1(\beta-1)^2+2\theta_3(\beta^2-1)+\theta_2(\beta+1)^2}{\alpha^2\beta^2}.
\end{aligned} \tag{1.10}$$

- Por ser $\hat{\theta}_{PC}$ función de momentos muestrales, aplicando el Método Delta (Teorema 1.2.3) se tiene que

$$\sqrt{n}\left(\hat{\theta}_{PC}-\theta\right) \xrightarrow{L} N_3(\mathbf{0}, \Sigma_{PC}).$$

Eficiencia relativa asintótica

La eficiencia relativa asintótica es la razón entre la varianza generalizada (determinante de la matriz de varianzas y covarianzas) de los estimadores máximo verosímiles y la del estimador bajo consideración. Las expresiones de las varianzas generalizadas de los estimadores discutidos anteriormente son:

$$|\Sigma_{MV}| = \frac{(\theta_1-\theta_3)^2(\theta_2-\theta_3)^2}{(\theta_1\theta_2-\theta_3^2)(Q-1)-\theta_1-\theta_2+2\theta_3},$$

$$|\Sigma_{MM}| = \theta_1^2\theta_2^2+\theta_1\theta_2\theta_3-(\theta_1+\theta_2)\theta_3^2+\theta_3^3-\theta_3^4,$$

$$|\Sigma_{DC}| = \left(\theta_1\theta_2-\theta_3^2\right)\{\exp(\theta_1+\theta_2-\theta_3)-1\}-\theta_1\theta_2(\theta_1+\theta_2-2\,\theta_3),$$

$$\begin{aligned}|\Sigma_{PP}| = \frac{1}{16}\Big[&\left(\theta_1\theta_2-\theta_3^2\right)[\exp\{4(\theta_1+\theta_2-2\theta_3)\}-1] \\ &+4\left\{2\theta_3(\theta_1-\theta_3)(\theta_2-\theta_3)-\theta_1(\theta_2-\theta_3)^2-\theta_2(\theta_1-\theta_3)^2\right\}\Big],\end{aligned}$$

$$|\Sigma_{PC}| = \frac{1}{64}\left[F\{CJ+E(2D-E)\}-(G-H)\{2D(E-D)+C(G-H)\}-D^2(F+J)\right],$$

donde Q está definido en (1.9), y C, D, E, F, G, H y J están definidas en (1.10).

Un análisis más detallado sobre la eficiencia relativa asintótica del estimador de θ se puede encontrar en los trabajos desarrollados por los investigadores que han tratado el tema, ver por ejemplo, Holgate (1964) [16], Loukas et al. (1986) [31], Papageorgiou y Loukas (1988) [34], Paul y Ho (1989) [35], y Kocherlakota y Kocherlakota (1992) [26]. Cabe señalar que la eficiencia relativa asintótica para el estimador de θ por el método de los puntos pares condicionados está solamente tratado en Papageorgiou y Loukas (1988) [34].

1.5. Bondad de Ajuste

Un aspecto crucial en cualquier análisis de datos es contrastar la bondad de ajuste de las observaciones con el modelo probabilístico supuesto, es decir, se contrasta si los datos provienen de la población que se supone.

Dadas las observaciones $X_1, X_2, \ldots, X_n$ iid, con distribución F, el objetivo es contrastar la hipótesis nula H_0 : "$F = F_0$". La hipótesis alternativa será H_1 : "$F \neq F_0$", F_0 puede ser una distribución totalmente especificada, o bien, especificada salvo por un número finito de parámetros.

1.5.1. Tests de bondad de ajuste en dimensión uno

En este apartado nuestro interés es contrastar

$$\begin{aligned} &H_0 : X \sim P(\vartheta), \text{ para algún } \vartheta > 0, \\ &H_1 : X \text{ no se distribuye } P(\vartheta), \ \forall\, \vartheta > 0. \end{aligned} \tag{1.11}$$

Sean $X_1, X_2, \ldots, X_n$ observaciones independientes de una v.a. X que toma valores enteros no negativos. Sean

$$F(k\,; \vartheta) = \exp(-\vartheta) \sum_{j=0}^{k} \frac{\vartheta^j}{j!}, \ k \in \mathbb{N}_0,$$

la función de distribución de la distribución Poisson $P(\vartheta)$ y F la función de distribución desconocida de X. Además, $\hat{\vartheta}_n = \hat{\vartheta}_n(X_1, X_2, \ldots, X_n) = \bar{X}_n$ denota el estimador máximo verosímil de ϑ.

Como es bien sabido (ver por ejemplo, Gürtler y Henze (2000) [11]), el estadístico χ^2, que es una herramienta clásica para contrastar la bondad de ajuste, presenta dos inconvenientes al tratar de comprobar la hipótesis que los datos se distribuyen Poisson: (a) selección de celdas y (b) estimaciones del parámetro ϑ para datos agrupados.

Gürtler y Henze (2000) [11] presentan una variedad de procedimientos para contrastar (1.11). La distribución nula de los estadísticos considerados dependen del verdadero y desconocido valor del parámetro. Ni siquiera su distribución asintótica es libre, es decir, también depende del parámetro. Por ello, estos autores emplean un bootstrap paramétrico para estimar los valores críticos o los $p-$valores.

En las siguientes subsecciones estudiaremos algunos de los contrastes presentados por Gürtler y Henze (2000) [11].

1.5.1.1. Índice de dispersión de Fisher. Estadístico $\hat{U}_{n2}^2$

Índice de dispersión de Fisher

Puesto que la media y la varianza de la distribución Poisson son iguales, entonces el cociente entre sus estimadores debería estar cercano a 1. Específicamente,

$$\frac{\hat{\sigma}^2}{\hat{\vartheta}} = \frac{1}{n}\frac{\sum_{j=1}^{n}(X_j - \bar{X}_n)^2}{\bar{X}_n} \approx 1. \tag{1.12}$$

Un procedimiento para contrastar el test estadístico dado en (1.11) es rechazar H_0 para valores pequeños o grandes del índice de dispersión de Fisher, dado por

$$D_n = \sum_{j=1}^{n}\frac{(X_j - \bar{X}_n)^2}{\bar{X}_n}.$$

Estadístico $\hat{U}_{n2}^2$

De (1.12) y del índice de dispersión de Fisher, resulta

$$D_n \approx n \quad \Longleftrightarrow \quad (D_n - n)^2 \approx 0.$$

Por lo tanto, para contrastar las hipótesis en (1.11), se rechaza H_0 para valores "grandes" del estadístico

$$\hat{U}_{n2}^2 = \left\{\frac{1}{\sqrt{2n}}(D_n - n)\right\}^2.$$

Como
$$\hat{U}_{n2}^2 = \frac{1}{2n}(D_n - n)^2 \xrightarrow[H_0]{L} \chi_1^2,$$
(veáse por ejemplo Rayner, Thas y Best (2009, p. 156) [38]).

Así, se rechaza $H_0 : X \sim P(\vartheta)$ si $P_{H_0}\Big(\hat{U}_{n2}^2 > c\Big) = \alpha$, donde $c = \chi^2_{1,1-\alpha}$. Por lo tanto, se rechaza $H_0 : X \sim P(\vartheta)$ si
$$\frac{1}{2n}\left(\sum_{j=1}^{n}\frac{(X_j - \bar{X}_n)^2}{\bar{X}_n} - n\right)^2 > \chi^2_{1,1-\alpha}\,.$$

Observación 1.5.1 *El test basado en el estadístico $\hat{U}_{n2}^2$ no es un test consistente pues está construido en base a momentos, específicamente se centra en la propiedad que la media y la varianza para la distribución Poisson son iguales.*

1.5.1.2. Contrastes basados en la función generatriz de probabilidad empírica

Estadístico $R_{n,a}$

Puesto que la distribución de X está caracterizada por su fgp, obtenemos que si X se distribuye según una ley de Poisson, entonces su fgp está dada por
$$g(t;\vartheta) = E_\vartheta\big(t^X\big) = \exp\{\vartheta(t-1)\},$$
la cual puede ser estimada por su fgpe
$$g_n(t) = \frac{1}{n}\sum_{j=1}^{n} t^{X_j} \quad \text{de} \quad X_1, X_2, \ldots, X_n.$$

Por tanto, parece natural basar un contraste de H_0 sobre
$$G_n(t) = \sqrt{n}\left(g_n(t) - g(t\,;\hat{\vartheta}_n)\right),\ 0 \le t \le 1. \tag{1.13}$$

Nótese que, bajo H_0, $G_n(t)$ es la diferencia de dos estimadores consistentes de $g(t;\vartheta)$. Por lo que un test razonable para contrastes debería rechazar H_0 para valores "grandes" de $R_{n,a}$ definido por
$$R_{n,a} = \int_0^1 G_n^2(t) t^a dt,$$

donde $a \geq 0$ es una constante. Elegir un valor grande de a significa colocar más peso cerca del punto extremo $t = 1$.

Observación 1.5.2 *Para $a = 0$, se obtiene el estadístico sugerido por Rueda et al. (1991) [40] y el caso general, cuando $a \geq 0$, fue propuesto por Baringhaus et al. (2000) [2].*

Estadístico $T_{n,a}$

Es otro estadístico para contrastar H_0, el cual está basado en el hecho que $g(t;\vartheta)$ es la única fgp que satisface la ecuación diferencial

$$\vartheta\, g(t) - g'(t) = 0,$$

donde $g(t)$ denota la fgp de la variable aleatoria X.

Si H_0 es cierta, entonces

$$\bar{X}_n\, g_n(t) - g_n'(t) \approx 0, \ \forall t.$$

Por lo que un test razonable para contrastar la hipótesis nula debería rechazar H_0 para valores "grandes" de $T_{n,a}$ definido por

$$T_{n,a} = n \int_0^1 \left\{ \bar{X}_n\, g_n(t) - g_n'(t) \right\}^2 t^a\, dt,$$

donde $a \geq 0$ es una constante. Elegir un valor grande de a significa colocar más peso cerca del punto extremo $t = 1$.

La distribución de los estadísticos $R_{n,a}$ y $T_{n,a}$ es desconocida para tamaños de muestra finito. Un modo clásico de aproximar la distribución nula es mediante la distribución asintótica nula.

La convergencia débil de los estadísticos $R_{n,a}$ y $T_{n,a}$ está dada en el resultado siguiente, cuya prueba sigue pasos similares a los dados en la demostración del Teorema 2.1 en Baringhaus y Henze (1992) [3].

Teorema 1.5.3 *(Gürtler y Henze (2000) [11]) Sea* $\{X_{n,1}, X_{n,2}, \ldots, X_{n,n}\}$, $n \geq 1$, *un arreglo triangular de v.a. iid en cada fila, tales que* $X_{n,1} \sim P(\vartheta_n)$, *donde* $0 < \vartheta = \lim_{n\to\infty} \vartheta_n$ *existe. Sean* $G_n(t)$ *definido como en (1.13) y*

$$\tilde{G}_n(t) = \sqrt{n}\left(\bar{X}_n\, g_n(t) - g_n'(t)\right),\ 0 \leq t \leq 1,$$

donde

$$g_n(t) = \frac{1}{n}\sum_{j=1}^{n} t^{X_{n,j}}, \quad g(t\,;\hat{\vartheta}_n) = \exp\Big\{\hat{\vartheta}_n(t-1)\Big\}$$

$$\hat{\vartheta}_n = \hat{\vartheta}_n(X_{n,1}, X_{n,2}, \ldots, X_{n,n}) = \bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_{n,j}.$$

Entonces se cumple lo siguiente:

(i) $R_{n,a} \xrightarrow{L} \int_0^1 Z^2(t)\, t^a dt$, $a \geq 0$, *donde* $Z(\cdot)$ *es un proceso Gaussiano centrado con núcleo de covarianza*

$$K(u,v) = \exp\{\vartheta(uv-1)\} - \{1+\vartheta(u-1)(v-1)\}\exp\{\vartheta(u+v-2)\},\ 0 \leq u, v \leq 1.$$

(ii) $T_{n,a} \xrightarrow{L} \int_0^1 \tilde{Z}^2(t)\, t^a dt$, $a \geq 0$, *donde* $\tilde{Z}(\cdot)$ *es un proceso Gaussiano centrado con núcleo de covarianza*

$$\tilde{K}(u,v) = \vartheta\,\{1+\vartheta(1-u)(1-v)\}\ \exp\{\vartheta(uv-1)\} - \vartheta\exp\{\vartheta(u+v-2)\},$$

donde $0 \leq u, v \leq 1$.

Además, Gürtler y Henze (2000) [11] demuestran que los test basados en los estadísticos $R_{n,a}$ y $T_{n,a}$ son consistentes frente a alternativas fijas.

Un modo alternativo de aproximar la distribución nula de los estadísticos antes citados es usar un estimador bootstrap paramétrico, que se describe como sigue:

Denotemos por W_n a uno cualquiera de los estadísticos anteriores ($R_{n,a}$ o $T_{n,a}$). Sea $H_{n,\vartheta}(t) = P_\vartheta(W_n \leq t)$ la función de distribución de la distribución nula de W_n cuando ϑ es el verdadero valor del parámetro. El método bootstrap estima el cuantil de $H_{n,\vartheta}(t)$ mediante el $(1-\alpha)$-cuantil de $H_{n,\hat{\vartheta}_n}$, que se aproxima mediante simulación, empleando los siguientes pasos:

- Generar B seudo-muestras aleatorias de tamaño n con distribución $P(\hat{\vartheta}_n)$, es decir, generar $X^*_{j1}, X^*_{j2}, \ldots, X^*_{jn}, j=1,2,\ldots,B$ iid de acuerdo a $P(\hat{\vartheta}_n)$.

- Calcular $W^*_{j,n} = W_n(X^*_{j1}, X^*_{j2}, \ldots, X^*_{jn})$, para $j = 1,2,\ldots,B$.

- Sea $H^*_{n,B}(t) = \dfrac{1}{B}\displaystyle\sum_{j=1}^{B} I\{W^*_{j,n} \leq t\}$ para la función de distribución empírica de $W^*_{1,n}, W^*_{2,n}, \ldots, W^*_{B,n}$ y $W^*_{1:B} \leq W^*_{2:B} \leq \ldots \leq W^*_{B:B}$ para sus estadísticos de orden.

  Finalmente, el valor crítico, $c^*_{n,B}$, está dado por

$$c^*_{n,B} = \begin{cases} W^*_{B(1-\alpha):B}\,, & \text{si } B(1-\alpha) \text{ es un entero,} \\ W^*_{[B(1-\alpha)]+1:B}\,, & \text{en otro caso,} \end{cases}$$

  donde $[y]$ es la parte entera de y.

Estadístico V_n

Motivados por el hecho que $\frac{\partial^2}{\partial t^2}\log g(t;\vartheta) \equiv 0$, Nakamura y Pérez-Abreu (1993) [33] propusieron la suma de los cuadrados de los coeficientes del polinomio $g^2_n(t)\,\frac{\partial^2}{\partial t^2}\log g_n(t)$ como un estadístico para contrastar (1.11).

Como

$$\frac{\partial^2}{\partial t^2}\log g_n(t) = \frac{g_n(t)\dfrac{\partial^2}{\partial t^2}g_n(t) - \left\{\dfrac{\partial}{\partial t}g_n(t)\right\}^2}{g^2_n(t)}, \tag{1.14}$$

llamando $N_n(t)$ al numerador y puesto que $g_n(t) = \displaystyle\sum_{i=1}^{n} t^{X_i}$, obtenemos

$$\begin{aligned} N_n(t) &= g_n(t)\frac{\partial^2}{\partial t^2}g_n(t) - \left\{\frac{\partial}{\partial t}g_n(t)\right\}^2 \\ &= \frac{1}{n^2}\sum_{i,j=1}^{n} X_i(X_i - X_j - 1)t^{X_i+X_j-2}. \end{aligned}$$

Debido a que $g^2_n(t) > 0, \forall t$, entonces de (1.14)

$$\frac{\partial^2}{\partial t^2}\log g_n(t) \equiv 0 \iff N_n(t) = 0,\ \forall t.$$

Haciendo $X_{(n)} = \text{máx}\{X_1, X_2, \ldots, X_n\}$ y notando que $N_n(t)$ es un polinomio aleatorio en t de grado $d_n = 2X_{(n)} - 2$, obtenemos

$$N_n(t) = \sum_{k=0}^{d_n} a_k t^k,$$

donde

$$a_k = \frac{1}{n^2} \sum_{i,j=1}^{n} X_i(X_i - X_j - 1) I\{X_i + X_j - 2 = k\}, \ 0 \leq k \leq d_n.$$

Por lo tanto

$$N_n(t) = 0, \ \forall t \iff a_k = 0, \ 0 \leq k \leq d_n.$$

Así, un estadístico para contrastar H_0 es considerar la suma de los cuadrados de los coeficientes polinomiales, esto es,

$$\begin{aligned} V_n &= \sum_{k=0}^{d_n} a_k^2 \\ &= \frac{1}{n^4} \sum_{i,j,k,l=1}^{n} X_i(X_i - X_j - 1) X_k(X_k - X_l - 1) I\{X_i + X_j = X_k + X_l\}, \end{aligned}$$

que expresa a V_n como un V-estadístico con un núcleo de orden 4.

La distribución nula y la distribución asintótica nula de V_n son ambas desconocidas. No obstante, Nakamura y Pérez-Abreu (1993) [33] observaron numéricamente que la distribución nula de

$$V_n^* = \frac{nV_n}{(\bar{X}_n)^{1.45}}$$

es aproximadamente independiente de ϑ.

1.5.2. Tests de bondad de ajuste en dimensión dos

El objetivo de este apartado es contrastar

$$\begin{aligned} &H_0 : (X_1, X_2) \sim PB(\theta_1, \theta_2, \theta_3), \ \text{para algún } \theta_1, \theta_2, \theta_3 > 0, \\ &H_1 : (X_1, X_2) \ \text{no se distribuye } PB(\theta_1, \theta_2, \theta_3), \ \forall\, \theta_1, \theta_2, \theta_3 > 0. \end{aligned} \tag{1.15}$$

Hasta donde conocemos, existen tres tests estadísticos para contrastar (1.15). A continuación los describiremos brevemente y expondremos lo esencial de cada uno de ellos.

1.5.2.1. Test T de Crockett

El estadístico propuesto por Crockett (1979) [6], T, está basado en una forma cuadrática en $Z_{X_1} = S^2_{X_1} - \bar{X}_1$ y $Z_{X_2} = S^2_{X_2} - \bar{X}_2$, donde $\bar{X}_1, \bar{X}_2, S^2_{X_1}$ y $S^2_{X_2}$ son las medias y varianzas muestrales, respectivamente. El objetivo es encontrar la matriz de varianzas y covarianzas de Z_{X_1} y Z_{X_2}.

Se tiene que $\mathrm{var}(Z_{X_1}) = \mathrm{var}(S^2_{X_1}) + \mathrm{var}(\bar{X}_1) - 2\ \mathrm{cov}(S^2_{X_1}, \bar{X}_1)$. De Stuart y Ord (1987, Volumen I, p. 338) [43], obtenemos que

$$\mathrm{var}(S^2_{X_1}) = \frac{\mu_4 - \mu_2^2}{n},$$

donde μ_k es el k-ésimo momento central definido en la Subsección 1.4.3.

Como $\mu_1 = 0$, de Stuart y Ord (1987, Volumen I, ejemplo 10.2, p. 323) [43]

$$\mathrm{cov}(S^2_{X_1}, \bar{X}_1) = \frac{\mu_3}{n}.$$

Así,

$$\mathrm{var}(Z_{X_1}) = \frac{\theta_1 + 3\theta_1^2 - \theta_1^2}{n} + \frac{\theta_1}{n} - 2\frac{\theta_1}{n} = 2\frac{\theta_1^2}{n}$$

y similarmente,

$$\mathrm{var}(Z_{X_2}) = 2\frac{\theta_2^2}{n}.$$

Puesto que,

$$\mathrm{cov}(Z_{X_1}, Z_{X_2}) = \mathrm{cov}(S^2_{X_1}, S^2_{X_2}) + \mathrm{cov}(\bar{X}_1, \bar{X}_2) - \mathrm{cov}(S^2_{X_1}, \bar{X}_2) - \mathrm{cov}(\bar{X}_1, S^2_{X_2}),$$

de la primera ecuación de (2.1.19) en Kocherlakota y Kocherlakota (1992, p. 40) [26]

$$\mathrm{cov}(S^2_{X_1}, S^2_{X_2}) = \mathrm{cov}(m_{2,0}, m_{0,2}) = \frac{\mu_{2,2} - \mu_{2,0}\mu_{0,2}}{n},$$

donde $\mu_{r,s}$ es el (r,s)-ésimo momento central, descrito en la Subsección 1.4.3, además

$$m_{r,s} = \frac{1}{n}\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)^r (X_{2i} - \bar{X}_2)^s.$$

De Kocherlakota y Kocherlakota (1992, p. 43) [26]

$$\mathrm{cov}(\bar{X}_1, \bar{X}_2) = \frac{1}{n}\,\mathrm{cov}(X_1, X_2) = \frac{\theta_3}{n}$$

y de (2.1.17) en Kocherlakota y Kocherlakota (1992, p. 40) [26]

$$\operatorname{cov}(S_{X_1}^2, \bar{X}_2) = \operatorname{cov}(m_{2,0}, m_{0,1}) = \frac{\mu_{2,1} - \mu_{2,0}\mu_{0,1} + 2\mu_{1,1}\mu_{1,0}\mu_{0,0} - \mu_{2,1}\mu_{0,0} - 2\mu_{1,0}\mu_{1,1}}{n}.$$

Haciendo los cálculos respectivos obtenemos

$$\operatorname{cov}(Z_{X_1}, Z_{X_2}) = 2\frac{\theta_3^2}{n}.$$

Así, si V denota la matriz de varianzas y covarianzas de $Z = (Z_{X_1}, Z_{X_2})$, entonces los resultados de esta sección muestran que

$$V = \frac{2}{n}\begin{pmatrix} \theta_1^2 & \theta_3^2 \\ \theta_3^2 & \theta_2^2 \end{pmatrix}.$$

Usando los estimadores $\bar{X}_1, \bar{X}_2$ y $m_{1,1} = S_{X_1X_2}$ (covarianza muestral) para θ_1, θ_2 y θ_3, respectivamente, Crockett (1979) [6] demuestra que, bajo H_0,

$$T = Z\hat{V}^{-1}Z^\top \xrightarrow{L} Y \sim \chi_2^2,$$

con lo cual se consigue el estadístico propuesto por dicho autor, dado por

$$T = \frac{n}{2}\frac{\bar{X}_2^2\left(S_{X_1}^2 - \bar{X}_1\right)^2 - 2S_{X_1X_2}^2\left(S_{X_1}^2 - \bar{X}_1\right)\left(S_{X_2}^2 - \bar{X}_2\right) + \bar{X}_1^2\left(S_{X_2}^2 - \bar{X}_2\right)^2}{\bar{X}_1^2\bar{X}_2^2 - S_{X_1X_2}^4}.$$

Así, se rechaza H_0 si

$$T \geq \chi_{2,\alpha}^2,$$

donde $\chi_{2,\alpha}^2$, para $0 < \alpha < 1$, denota el percentil α superior de la distribución χ^2 con 2 grados de libertad.

1.5.2.2. Test I_B de Loukas y Kemp

Loukas y Kemp (1986) [30] desarrollaron un test para la DPB basado en lo que llamaron índice de dispersión bivariante y denotaron por I_B, pues es una extensión del índice de dispersión univariante.

Con el objetivo de obtener I_B, consideremos ahora una muestra aleatoria simple (X_{11}, X_{21}), $(X_{12}, X_{22}), \ldots, (X_{1n}, X_{2n})$, tal que $(X_{1k}, X_{2k}) \sim PB(\theta_1, \theta_2, \theta_3)$, para $k = 1, 2, \ldots, n$, donde $\theta_1 > \theta_3$, $\theta_2 > \theta_3$ y $\theta_3 > 0$.

Si θ_1, θ_2 y θ_3 son conocidos, entonces el índice de dispersión bivariante toma la forma

$$I_B = \frac{1}{1-\rho^2} \sum_{i=1}^{n} \left(W_{1i}^2 - 2\rho W_{1i} W_{2i} + W_{2i}^2 \right),$$

donde $W_{ji} = \dfrac{X_{ji} - \theta_j}{\sqrt{\theta_j}}, j = 1, 2, \ i = 1, 2, \ldots, n$ y $\rho = \dfrac{\theta_3}{\sqrt{\theta_1 \theta_2}}$.

Bajo las condiciones del Teorema 1 en Loukas y Kemp (1986) [30], dichos autores muestran que I_B se distribuye aproximadamente como una variable χ^2 con $2n$ grados de libertad.

En la situación práctica usual, cuando θ_1, θ_2 y θ_3 son desconocidos, entonces Loukas y Kemp (1986) [30] definen el índice de dispersión bivariante como

$$\begin{aligned} I_B &= \frac{\displaystyle\sum_{i=1}^{n} \left\{ \frac{(X_{1i} - \bar{X}_1)^2}{\bar{X}_1} - \frac{2S_{12}(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\bar{X}_1 \bar{X}_2} + \frac{(X_{2i} - \bar{X}_2)^2}{\bar{X}_2} \right\}}{1 - \dfrac{S_{12}^2}{\bar{X}_1 \bar{X}_2}} \\ &= \frac{n(\bar{X}_2 S_1^2 - 2S_{12}^2 + \bar{X}_1 S_2^2)}{\bar{X}_1 \bar{X}_2 - S_{12}^2}, \end{aligned}$$

donde $\bar{X}_1$ y $\bar{X}_2$ son las medias muestrales, S_1^2 y S_2^2 son la varianzas muestrales y S_{12} es la covariaza muestral.

Como en este proceso se debe estimar el parámetro θ, entonces se pierden tres grados de libertad y por tanto, este nuevo I_B se distribuye aproximadamente como una χ^2 con $2n - 3$ grados de libertad, como lo mencionan Loukas y Kemp (1986) [30].

Así, se rechaza H_0 si

$$I_B \geq \chi^2_{2n-3,\alpha},$$

donde $\chi^2_{2n-3,\alpha}$, para $0 < \alpha < 1$, denota el percentil α superior de la distribución χ^2 con $2n - 3$ grados de libertad.

Observación 1.5.4 *Si θ_1, θ_2 y θ_3 son desconocidos, entonces Loukas y Kemp (1986) [30] aproximan θ_1 y θ_2 por sus estimadores máximo verosímiles, esto es, $\hat{\theta}_1 = \bar{X}_1$ y $\hat{\theta}_2 = \bar{X}_2$, y el estimador de θ_3 que, como debe calcularse por métodos numéricos, lo aproximan por la covarianza muestral, es decir, $\hat{\theta}_3 = S_{12}$.*

1.5.2.3. Test $\boldsymbol{NI_B}$ de Rayner y Best

Rayner y Best (1995) [37] expresaron el estadístico de Loukas y Kemp (1986) [30], I_B, como sigue

$$I_B = \frac{n}{1-\hat{\rho}^2}\left(\frac{S^2_{X_1}}{\bar{X}_1} - 2\,\frac{S^2_{X_1X_2}}{\bar{X}_1\bar{X}_2} + \frac{S^2_{X_2}}{\bar{X}_2}\right), \tag{1.16}$$

donde $\hat{\rho} = \dfrac{S_{X_1X_2}}{\sqrt{\bar{X}_1\bar{X}_2}}$ es un estimador de ρ, $\bar{X}_1$ y $\bar{X}_2$ son las medias muestrales, $S^2_{X_1}$ y $S^2_{X_2}$ son la varianzas muestrales, y $S_{X_1X_2}$ es la covariaza muestral.

De (1.16) se observa que

$$\text{si}\ \ \hat{\rho}^2 = \frac{S^2_{X_1X_2}}{\bar{X}_1\bar{X}_2} > \frac{1}{2}\left(\frac{S^2_{X_1}}{\bar{X}_1} + \frac{S^2_{X_2}}{\bar{X}_2}\right),\ \text{entonces}\ \ I_B < 0.$$

Por lo tanto, cuando ρ crece hay una probabilidad creciente que I_B sea negativo y su distribución no sea bien aproximada por una χ^2. Para remediar esta situación, proponen utilizar un nuevo estadístico, lo llaman NI_B y lo definen mediante la expresión

$$NI_B = \frac{n}{1-r^2}\left(\frac{S^2_{X_1}}{\bar{X}_1} - 2\,r^2\sqrt{\frac{S^2_{X_1}S^2_{X_2}}{\bar{X}_1\bar{X}_2}} + \frac{S^2_{X_2}}{\bar{X}_2}\right),$$

donde, el estimador de ρ es el coeficiente de correlación muestral, dado por

$$r = \frac{\sum_{i=1}^{n}(X_{1i}-\bar{X}_1)(X_{2i}-\bar{X}_2)}{\sqrt{\sum_{i=1}^{n}(X_{1i}-\bar{X}_1)^2\ \sum_{i=1}^{n}(X_{2i}-\bar{X}_2)^2}}.$$

Al igual que I_B, bajo H_0, $NI_B \cong \chi^2_{2n-3}$. Por lo tanto, se rechaza H_0 si

$$NI_B \geq \chi^2_{2n-3,\alpha},$$

donde $\chi^2_{2n-3,\alpha}$, para $0<\alpha<1$, denota el percentil α superior de la distribución χ^2 con $2n-3$ grados de libertad.

Observación 1.5.5 *Los tests estadísticos T, I_B y NI_B no son consistentes, pues están construidos en base a los momentos, específicamente se basan en que los dos primeros momentos poblacionales son iguales.*

Observación 1.5.6 *Para efectos de programar los estadísticos* T, I_B *y* NI_B*, Rayner y Best (1995) [37] hacen notar que Crockett (1979) [6] usa* $(n-1)$ *como divisor para calcular* $S_{X_1}^2$ *y* $S_{X_2}^2$*, y divisor* n *en* $S_{X_1X_2}$*, en cambio Loukas y Kemp (1986) [30] usan el divisor* n *para calcular* $S_{X_1}^2, S_{X_2}^2$ *y* $S_{X_1X_2}$.

Capítulo 2

Estadísticos tipo Cramér-von Mises

2.1. Tests estadísticos

Sean $\boldsymbol{X}_1 = (X_{11}, X_{21}), \boldsymbol{X}_2 = (X_{12}, X_{22}), \ldots, \boldsymbol{X}_n = (X_{1n}, X_{2n})$ vectores aleatorios iid de $\boldsymbol{X} = (X_1, X_2) \in \mathbb{N}_0^2$. Basándonos en la muestra aleatoria $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$, nuestro objetivo es contrastar la hipótesis

$$H_0 : (X_1, X_2) \sim PB(\theta_1, \theta_2, \theta_3), \text{ para algún } (\theta_1, \theta_2, \theta_3) \in \Theta,$$

contra la alternativa

$$H_1 : (X_1, X_2) \nsim PB(\theta_1, \theta_2, \theta_3), \ \forall (\theta_1, \theta_2, \theta_3) \in \Theta.$$

Con este propósito, aprovecharemos algunas de las propiedades de la fgp que nos permitirán proponer los siguientes dos tests estadísticos.

1. De acuerdo a la Proposición 1.2.2, un estimador consistente de la fgp es la fgpe. Si H_0 es verdadera y $\hat{\theta}_n$ es un estimador consistente de θ, entonces $g(u; \hat{\theta}_n)$ estima consistentemente la fgp de los datos.

   Como la distribución de $\boldsymbol{X}$ es determinada de forma única por su fgp, $g(u)$, $u \in [0, 1]^2$, un test razonable para contrastar H_0 debería rechazar la hipótesis nula para valores "grandes" de $R_{n,w}(\hat{\theta}_n)$ definido por

$$R_{n,w}(\hat{\theta}_n) = \int_0^1 \int_0^1 G_n^2(u; \hat{\theta}_n) w(u) du, \qquad (2.1)$$

donde

$$G_n(u;\hat{\theta}_n) = \sqrt{n}\left\{g_n(u) - g(u;\hat{\theta}_n)\right\},$$

$\hat{\theta}_n = \hat{\theta}_n(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n) = (\hat{\theta}_{1n}, \hat{\theta}_{2n}, \hat{\theta}_{3n})$ es un estimador consistente de θ y $w(u)$ es una función medible de peso tal que $w(u) \geq 0,\ \forall u \in [0,1]^2$, y

$$\int_0^1 \int_0^1 w(u)du < \infty. \tag{2.2}$$

Este último supuesto sobre w asegura que la integral doble en (2.1) es finita para cada n fijo.

$R_{n,w}(\hat{\theta}_n)$ es una extensión bivariante de los estadísticos propuestos por Rueda et al. (1991) [40] y Baringhaus et al. (2000) [2] para contrastar la bondad de ajuste a la distribución Poisson univariante, tal como se vió en la Subsección 1.5.1.2.

2. Puesto que la fgp de la distribución Poisson univariante con parámetro λ es la única fgp que satisface $g'(t) = \lambda g(t)$, Baringhaus y Henze (1992) [3] propusieron un test estadístico que se basa en el análogo empírico de esta ecuación, presentado en la Subsección 1.5.1.2.

   Con el fin de extender este test al caso bivariante, primero tenemos que encontrar una ecuación o un sistema de ecuaciones cuya única solución sea la fgp de la DPB. La siguiente proposición establece dicho sistema.

Proposición 2.1.1 *Sea $G_2 = \{g : [0,1]^2 \to \mathbb{R}$, tal que g es una fgp y $\frac{\partial}{\partial u_i} g(u_1,u_2)$, $i = 1,2$, existen $\forall (u_1,u_2) \in [0,1]^2\}$. Sea $g(u_1,u_2;\theta)$ como en (1.6). Entonces $g(u_1,u_2;\theta)$ es la única fgp en G_2 que satisface el siguiente sistema*

$$D_i(u;\theta) = 0, \quad i = 1,2, \quad \forall u \in [0,1]^2, \tag{2.3}$$

donde

$$\begin{aligned} D_1(u;\theta) &= \frac{\partial}{\partial u_1} g(u_1,u_2) - \{\theta_1 + \theta_3(u_2 - 1)\} g(u_1,u_2), \\ D_2(u;\theta) &= \frac{\partial}{\partial u_2} g(u_1,u_2) - \{\theta_2 + \theta_3(u_1 - 1)\} g(u_1,u_2). \end{aligned}$$

Demostración Sea (X_1, X_2) un vector aleatorio y sea $g(u_1,u_2) = E\left(u_1^{X_1} u_2^{X_2}\right)$ su fgp. Entonces, de la primera expresión en (2.3)

$$\frac{\partial}{\partial u_1} \log g(u_1,u_2) = \theta_1 + \theta_3(u_2 - 1). \tag{2.4}$$

Integrando (2.4) sobre u_1, obtenemos

$$\begin{aligned} g(u_1,u_2) &= \exp\{\phi_1(u_2)+\theta_1 u_1+\theta_3(u_2-1)u_1\}, \\ &= \exp\{\phi(u_2)+\theta_1(u_1-1)+\theta_3(u_1-1)(u_2-1)\}, \end{aligned}$$

donde $\phi(u_2)=\phi_1(u_2)+\theta_1+\theta_3(u_2-1)$.

Procediendo similarmente, de la segunda ecuación en (2.3) obtenemos

$$g(u_1,u_2)=\exp\{\varphi(u_1)+\theta_2(u_2-1)+\theta_3(u_1-1)(u_2-1)\}.$$

Así, necesariamente $\varphi(u_1)=\theta_1(u_1-1)$ y $\phi(u_2)=\theta_2(u_2-1)$, en otras palabras, la fgp de la DPB es la única solución de (2.3). □

Por la Proposición 1.2.2, $g(u)$ y sus derivadas pueden ser estimadas consistentemente por la fgpe y las derivadas de la fgpe, respectivamente. Así, si H_0 fuera cierta, entonces las funciones

$$\begin{aligned} D_{1n}(u;\hat{\theta}_n) &= \frac{\partial}{\partial u_1} g_n(u_1,u_2)-\left\{\hat{\theta}_{1n}+\hat{\theta}_{3n}(u_2-1)\right\} g_n(u_1,u_2), \\ D_{2n}(u;\hat{\theta}_n) &= \frac{\partial}{\partial u_2} g_n(u_1,u_2)-\left\{\hat{\theta}_{2n}+\hat{\theta}_{3n}(u_1-1)\right\} g_n(u_1,u_2), \end{aligned} \tag{2.5}$$

deberían estar cerca de 0, donde $\hat{\theta}_n$ es un estimador consistente de θ. Así, para contrastar H_0 consideramos el siguiente test estadístico

$$S_{n,w}(\hat{\theta}_n)=n\int_0^1\int_0^1\left\{D_{1n}^2(u;\hat{\theta}_n)+D_{2n}^2(u;\hat{\theta}_n)\right\} w(u)du, \tag{2.6}$$

donde la función de peso $w(u)$ es como la definida en el ítem anterior.

Notar que la fgp de $\boldsymbol{X}$ está en G_2 sí y sólo si $E(X_i)$ existe, $i=1,2$.

En los dos casos anteriores, un test razonable para contrastar H_0 debería rechazar la hipótesis nula para valores grandes de cada test estadístico. Ahora, para determinar qué son los valores grandes en cada caso, debemos calcular la distribución nula de cada test estadístico o al menos una aproximación de cada una de ellas.

Puesto que las distribuciones nulas son desconocidas, trataremos de aproximarlas. Un modo clásico de estimar la distribución nula es mediante la distribución asintótica nula. En la siguiente sección estudiaremos esta situación.

2.2. Aproximación de la distribución nula

2.2.1. Distribución asintótica nula

Como un primer intento de aproximar la distribución nula de $R_{n,w}(\hat{\theta}_n)$ y de $S_{n,w}(\hat{\theta}_n)$ obtendremos sus distribuciones asintóticas nulas. Con este propósito, supondremos que el estimador $\hat{\theta}_n$ satisface la siguiente condición de regularidad.

Supuesto 2.2.1 *Bajo H_0, si $\theta = (\theta_1, \theta_2, \theta_3) \in \Theta$ denota el verdadero valor del parámetro, entonces*

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\ell}\left(\boldsymbol{X}_i; \theta\right) + \boldsymbol{o}_{P}(1),$$

donde $\boldsymbol{\ell} : \mathbb{N}_0^2 \times \Theta \longrightarrow \mathbb{R}^3$ es tal que

$$E_\theta\big\{\boldsymbol{\ell}(\boldsymbol{X}_1; \theta)\big\} = \mathbf{0} \in \mathbb{R}^3$$

y

$$J(\theta) = E_\theta\Big\{\boldsymbol{\ell}(\boldsymbol{X}_1; \theta)^\top \boldsymbol{\ell}(\boldsymbol{X}_1; \theta)\Big\} < \infty.$$

Aquí, $\boldsymbol{o}_{P}(1)$ es un vector que consta de 3 elementos $o_{P}(1)$.

Observación 2.2.2 *El Supuesto 2.2.1 no es restrictivo pues lo cumplen los estimadores más usados comúnmente, como son los citados en la Subsección 1.4.5.*

Para obtener la distribución asintótica nula de $R_{n,w}(\hat{\theta}_n)$ y de $S_{n,w}(\hat{\theta}_n)$, la plataforma de trabajo que usaremos será el espacio de Hilbert separable $\mathcal{H} = L^2\left([0,1]^2, w\right)$ definido por

$$\mathcal{H} = \left\{\varphi : [0,1]^2 \to \mathbb{R},\ \text{medible, tal que } \|\varphi\|_{\mathcal{H}}^2 = \int_0^1 \int_0^1 \varphi(u)^2 w(u) du < \infty\right\},$$

con producto escalar $\langle \phi, \psi \rangle_{\mathcal{H}} = \int_0^1 \int_0^1 \phi(u)\psi(u)w(u)du$.

En dicho espacio de Hilbert $\mathcal{H}$, tenemos que

- $R_{n,w}(\hat{\theta}_n) = \|G_n(\hat{\theta}_n)\|_{\mathcal{H}}^2$, con $G_n(u; \hat{\theta}_n) = \dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=1}^{n} G(\boldsymbol{X}_i; \hat{\theta}_n; u)$, $u \in [0,1]^2$,

  donde
  $$G(\boldsymbol{X}_i; \hat{\theta}_n; u) = u_1^{X_{1i}}\, u_2^{X_{2i}} - g(u; \hat{\theta}_n),\ i = 1, 2, \ldots, n.$$

- $S_{n,w}(\hat{\theta}_n) = \|Z_{1n}\|^2_{\mathcal{H}} + \|Z_{2n}\|^2_{\mathcal{H}}$, con $Z_{kn}(u) = \dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=1}^{n} V_k(\boldsymbol{X}_i; \hat{\theta}_n; u)$, $k = 1, 2$, $u \in [0, 1]^2$, donde

$$V_1(\boldsymbol{X}_i; \hat{\theta}_n; u) = X_{1i}\, I\{X_{1i} \geq 1\}\, u_1^{X_{1i}-1}\, u_2^{X_{2i}} - \left\{\hat{\theta}_{1n} + \hat{\theta}_{3n}(u_2 - 1)\right\} u_1^{X_{1i}}\, u_2^{X_{2i}},$$

$$V_2(\boldsymbol{X}_i; \hat{\theta}_n; u) = X_{2i}\, I\{X_{2i} \geq 1\}\, u_1^{X_{1i}}\, u_2^{X_{2i}-1} - \left\{\hat{\theta}_{2n} + \hat{\theta}_{3n}(u_1 - 1)\right\} u_1^{X_{1i}}\, u_2^{X_{2i}},$$

$i = 1, 2, \ldots, n$.

El siguiente resultado proporciona la distribución asintótica nula de $R_{n,w}(\hat{\theta}_n)$ y $S_{n,w}(\hat{\theta}_n)$.

Teorema 2.2.3 *Sean $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid de $\boldsymbol{X} = (X_1, X_2) \sim PB(\theta)$. Supongamos que se cumple el Supuesto 2.2.1. Entonces*

(a) $R_{n,w}(\hat{\theta}_n) = \|R_n\|^2_{\mathcal{H}} + r_n$,

donde $P_\theta(|r_n| > \varepsilon) \to 0$, $\forall \varepsilon > 0$, $R_n(u) = \dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=1}^{n} R^0(\boldsymbol{X}_i; \theta; u)$, *con*

$$R^0(\boldsymbol{X}_i; \theta; u) = u_1^{X_{1i}} u_2^{X_{2i}} - g(u; \theta) \left\{1 + \boldsymbol{\ell}\,(\boldsymbol{X}_i; \theta)\,(u_1 - 1, u_2 - 1, (u_1 - 1)(u_2 - 1))^\top\right\},$$

(b)

$$S_{n,w}(\hat{\theta}_n) = \|S_{1n}\|^2_{\mathcal{H}} + \|S_{2n}\|^2_{\mathcal{H}} + s_n, \tag{2.7}$$

donde $P_\theta(|s_n| > \varepsilon) \to 0$, $\forall \varepsilon > 0$, $S_{kn}(u) = \dfrac{1}{\sqrt{n}} \displaystyle\sum_{i=1}^{n} S_k^0(\boldsymbol{X}_i; \theta; u)$, $k = 1, 2$, *con*

$$\begin{aligned} S_1^0(\boldsymbol{X}_i; \theta; u) = {} & X_{1i}\, I\{X_{1i} \geq 1\}\, u_1^{X_{1i}-1}\, u_2^{X_{2i}} - \{\theta_1 + \theta_3(u_2 - 1)\}\, u_1^{X_{1i}}\, u_2^{X_{2i}} \\ & - g(u; \theta)\boldsymbol{\ell}\,(\boldsymbol{X}_i; \theta)\,(1, 0, u_2 - 1)^\top, \\ S_2^0(\boldsymbol{X}_i; \theta; u) = {} & X_{2i}\, I\{X_{2i} \geq 1\}\, u_1^{X_{1i}}\, u_2^{X_{2i}-1} - \{\theta_2 + \theta_3(u_1 - 1)\}\, u_1^{X_{1i}}\, u_2^{X_{2i}} \\ & - g(u; \theta)\boldsymbol{\ell}\,(\boldsymbol{X}_i; \theta)\,(0, 1, u_1 - 1)^\top. \end{aligned}$$

Además,

$$R_{n,w}(\hat{\theta}_n) \xrightarrow{L} \sum_{j \geq 1} \lambda_j^R \chi^2_{1j}.$$

$$S_{n,w}(\hat{\theta}_n) \xrightarrow{L} \sum_{j\geq 1} \lambda_j^S \chi_{1j}^2.$$

donde $\chi_{11}^2, \chi_{12}^2, \ldots$ *son v.a. independientes* χ^2 *con 1 grado de libertad y los conjuntos* $\{\lambda_j^R\}_{j\geq 1}$ *y* $\{\lambda_j^S\}_{j\geq 1}$ *son los autovalores no nulos de los operadores* $C_R(\theta)$ *y* $C_S(\theta)$*, respectivamente, definidos sobre el espacio de funciones*

$\{\tau : \mathbb{N}_0^2 \to \mathbb{R},$ *tal que* $E_\theta\{\tau^2(\boldsymbol{X})\} < \infty, \forall\theta \in \Theta\}$*, como sigue*

$$C_R(\theta)\tau(x) = E_\theta\{h_R(x, \boldsymbol{Y}; \theta)\tau(\boldsymbol{Y})\} \quad y \quad C_S(\theta)\tau(x) = E_\theta\{h_S(x, \boldsymbol{Y}; \theta)\tau(\boldsymbol{Y})\} \tag{2.8}$$

donde

$$h_R(x, y; \theta) = \int_0^1 \int_0^1 R^0(x; \theta; u)R^0(y; \theta; u)w(u)du,$$

$$h_S(x, y; \theta) = \int_0^1 \int_0^1 \sum_{k=1}^{2} S_k^0(x; \theta; u)S_k^0(y; \theta; u)w(u)du. \tag{2.9}$$

Demostración Solamente presentaremos la demostración de la parte (b), pues la demostración de la parte (a) sigue pasos similares.

Por definición, $S_{n,w}(\hat{\theta}_n) = \|Z_{1n}\|_{\mathcal{H}}^2 + \|Z_{2n}\|_{\mathcal{H}}^2$, con

$$Z_{kn}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_k(\boldsymbol{X}_i; \hat{\theta}_n; u), \;\; k = 1, 2.$$

Por desarrollo en serie de Taylor,

$$Z_{kn}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_k(\boldsymbol{X}_i; \theta; u) + \frac{1}{n} \sum_{i=1}^{n} Q_k^{(1)}(\boldsymbol{X}_i; \theta; u) \sqrt{n}(\hat{\theta}_n - \theta)^\top + q_{kn}, \tag{2.10}$$

$k = 1, 2$, donde

$$q_{kn} = \frac{1}{2}\sqrt{n}\big(\hat{\theta}_n - \theta\big)\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n} Q_k^{(2)}\big(\boldsymbol{X}_i; \widetilde{\theta}; u\big) \sqrt{n}\big(\hat{\theta}_n - \theta\big)^\top,$$

$\widetilde{\theta} = \alpha\hat{\theta}_n + (1-\alpha)\theta$, para algún $0 < \alpha < 1$ y $Q_k^{(2)}\big(x; \vartheta; u\big)$ es la matriz de orden 3×3 que contiene las derivadas de segundo orden de $V_k\big(x; \vartheta; u\big)$ con respecto a ϑ, para $k = 1, 2$,

$$Q_k^{(1)}(x; \vartheta; u) = \left(Q_{k1}^{(1)}(x; \vartheta; u), Q_{k2}^{(1)}(x; \vartheta; u), Q_{k3}^{(1)}(x; \vartheta; u)\right),$$

donde $Q_{kj}^{(1)}(x; \vartheta; u) = \frac{\partial}{\partial\vartheta_j} V_k(x; \vartheta; u)$, para $k = 1, 2,\ j = 1, 2, 3$.

Como las primeras derivadas son

$$Q_1^{(1)}(x;\vartheta;u) = -u_1^{x_1}u_2^{x_2}(1,0,u_2-1), \quad Q_2^{(1)}(x;\vartheta;u) = -u_1^{x_1}u_2^{x_2}(0,1,u_1-1), \tag{2.11}$$

entonces, $\left|Q_{kj}^{(1)}(\boldsymbol{X}_1;\theta;u)\right| \leq 1$, para $k=1,2$, $j=1,2,3$, $\forall u \in [0,1]^2$.

Así, considerando (2.2), resulta

$$E_\theta\left\{\left\|Q_{kj}^{(1)}(\boldsymbol{X}_1;\theta;u)\right\|_{\mathcal{H}}^2\right\} < \infty, \; k=1,2, \; j=1,2,3. \tag{2.12}$$

Como, para $k=1,2$, $j=1,2,3$, tenemos que

$$E_\theta\left[\left\{\frac{1}{n}\sum_{i=1}^n Q_{kj}^{(1)}(\boldsymbol{X}_i;\theta;u)\right\}^2\right] = \frac{1}{n}E_\theta\left[\left\{Q_{kj}^{(1)}(\boldsymbol{X}_1;\theta;u)\right\}^2\right] + \frac{n-1}{n}\left[E_\theta\left\{Q_{kj}^{(1)}(\boldsymbol{X}_1;\theta;u)\right\}\right]^2,$$

entonces

$$\begin{aligned}
E_\theta&\left[\left\{\frac{1}{n}\sum_{i=1}^n Q_{kj}^{(1)}(\boldsymbol{X}_i;\theta;u) - E_\theta\left\{Q_{kj}^{(1)}(\boldsymbol{X}_1;\theta;u)\right\}\right\}^2\right] \\
&= \frac{1}{n}E_\theta\left[\left\{Q_{kj}^{(1)}(\boldsymbol{X}_1;\theta;u)\right\}^2\right] - \frac{1}{n}\left[E_\theta\left\{Q_{kj}^{(1)}(\boldsymbol{X}_1;\theta;u)\right\}\right]^2 \\
&\leq \frac{1}{n}E_\theta\left[\left\{Q_{kj}^{(1)}(\boldsymbol{X}_1;\theta;u)\right\}^2\right].
\end{aligned}$$

Usando la desigualdad de Markov y (2.12), se logra

$$\begin{aligned}
P_\theta&\left[\left\|\frac{1}{n}\sum_{i=1}^n Q_{kj}^{(1)}(\boldsymbol{X}_i;\theta;u) - E_\theta\left\{Q_{kj}^{(1)}(\boldsymbol{X}_1;\theta;u)\right\}\right\|_{\mathcal{H}} > \varepsilon\right] \\
&\leq \frac{1}{n\,\varepsilon^2}\,E_\theta\left[\left\|Q_{kj}^{(1)}(\boldsymbol{X}_1;\theta;u)\right\|_{\mathcal{H}}^2\right] \to 0,
\end{aligned}$$

para $k=1,2$, $j=1,2,3$.

Así, obtenemos que

$$\frac{1}{n}\sum_{i=1}^n Q_k^{(1)}(\boldsymbol{X}_i;\theta;u) \xrightarrow{P} E_\theta\left\{Q_k^{(1)}(\boldsymbol{X}_1;\theta;u)\right\} = \begin{cases} -g(u;\theta)(1,0,u_2-1), & \text{si } k=1, \\ -g(u;\theta)(0,1,u_1-1), & \text{si } k=2, \end{cases} \tag{2.13}$$

en $\mathcal{H}$.

De (2.11) se sigue que $Q_k^{(2)}\big(x;\vartheta;u\big) = \mathbf{0}$ (la matriz nula de orden 3×3), $k=1,2$ y así $q_{kn}=0$.

Ahora, por el Supuesto 2.2.1 y como $q_{kn}=0$, entonces, (2.10) se puede escribir como

$$Z_{kn}(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[V_k(\boldsymbol{X}_i;\theta;u) + E_\theta\Big\{Q_k^{(1)}(\boldsymbol{X}_1;\theta;u)\Big\}\boldsymbol{\ell}\,(\boldsymbol{X}_i;\theta)^\top\Big]$$

$$+ E_\theta\Big\{Q_k^{(1)}(\boldsymbol{X}_1;\theta;u)\Big\}\ \boldsymbol{o}_{P}(1)^\top.$$

Así, de esta última ecuación, y de (2.12) y (2.13)

$$Z_{kn}(u) = S_{kn}(u) + s_{kn},$$

donde $\|s_{kn}\|_{\mathcal{H}} = o_{P}(1)$, $k=1,2$.

Por otra parte, observar que

$$\|S_{1n}\|_{\mathcal{H}}^2 + \|S_{2n}\|_{\mathcal{H}}^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} h_S(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta),$$

donde, $h_S(x,y;\theta)$ es como el definido en (2.9) y satisface $h_S(x,y;\theta) = h_S(y,x;\theta)$, $E_\theta\left\{h_S^2(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)\right\} < \infty$ y $E_\theta\left\{|h_S(\boldsymbol{X}_1,\boldsymbol{X}_1;\theta)|\right\} < \infty$.

Además, por la Proposición 2.1.1 y el Supuesto 2.2.1

$$E_\theta\{h_S(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)\} = \int_0^1\int_0^1\sum_{k=1}^{2} E_\theta\big\{S_k^0(\boldsymbol{X}_1;\theta;u)\big\}\,E_\theta\big\{S_k^0(\boldsymbol{X}_2;\theta;u)\big\}\,w(u)\,du = 0.$$

Por último, como h_S es degenerado, $E_\theta\{h_S(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)/\boldsymbol{X}_1\} = 0$, entonces, por el Teorema 6.4.1.B en Serfling (1980) [42]

$$\|S_{1n}\|_{\mathcal{H}}^2 + \|S_{2n}\|_{\mathcal{H}}^2 \xrightarrow{L} \sum_{j\geq 1}\lambda_j^S\,\chi_{1j}^2\,, \tag{2.14}$$

donde $\chi_{11}^2, \chi_{12}^2, \ldots$ y el conjunto $\{\lambda_j^S\}_{j\geq 1}$ son como los definidos en el Teorema.

De (2.14) se desprende que $\|S_{1n}\|_{\mathcal{H}}^2 + \|S_{2n}\|_{\mathcal{H}}^2 = O_P(1)$.

Ahora, como

$$\|Z_{kn}\|_{\mathcal{H}}^2 = \|S_{kn}\|_{\mathcal{H}}^2 + \|s_{kn}\|_{\mathcal{H}}^2 + 2\langle S_{kn}, s_{kn}\rangle_{\mathcal{H}},\ k=1,2,$$

y puesto que $\langle S_{kn}, s_{kn}\rangle_{\mathcal{H}} \leq \|S_{kn}\|_{\mathcal{H}}\, \|s_{kn}\|_{\mathcal{H}} = o_{P}(1)$, entonces

$$S_{n,w}(\hat{\theta}_n) \xrightarrow{L} \sum_{j\geq 1} \lambda_j^S\, \chi^2_{1j}\,.$$

Lo cual concluye la demostración del resultado. $\square$

La distribución asintótica nula de $R_{n,w}(\hat{\theta}_n)$ y de $S_{n,w}(\hat{\theta}_n)$ depende del desconocido verdadero valor del parámetro θ, por tanto, en la práctica, no proporcionan una solución útil al problema de estimar la distribución nula de los respectivos tests estadísticos. Esto podría solucionarse al reemplazar θ por $\hat{\theta}_n$.

Pero una dificultad mayor es determinar los conjuntos $\{\lambda_j^R\}_{j\geq 1}$ y $\{\lambda_j^S\}_{j\geq 1}$, puesto que, en la mayoría de los casos, calcular los autovalores de un operador no es una tarea simple y en nuestro caso, debemos obtener también las expresiones de $h_R(x, y; \theta)$ y $h_S(x, y; \theta)$, que no son fáciles de encontrar, pues dependen de la función $\boldsymbol{\ell}$, que por lo general no tiene una expresión sencilla.

Así, en la siguiente subsección consideramos otra forma de aproximar la distribución nula de los tests estadísticos, el método bootstrap paramétrico.

2.2.2. Aproximación bootstrap de la distribución nula

Un modo alternativo de estimar la distribución nula es mediante el método bootstrap paramétrico.

Sean $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid que toman valores en $\mathbb{N}_0^2$. Supongamos que $\hat{\theta}_n = \hat{\theta}_n(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n) \in \Theta$.

Sean $\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \ldots, \boldsymbol{X}_n^*$ vectores aleatorios iid de una población que se distribuye según la ley $PB(\hat{\theta}_{1n}, \hat{\theta}_{2n}, \hat{\theta}_{3n})$, dado $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ y sea $R^*_{n,w}(\hat{\theta}^*_n)$ la versión bootstrap de $R_{n,w}(\hat{\theta}_n)$ obtenida al reemplazar $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ y $\hat{\theta}_n = \hat{\theta}_n(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$ por $\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \ldots, \boldsymbol{X}_n^*$ y $\hat{\theta}^*_n = \hat{\theta}_n(\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \ldots, \boldsymbol{X}_n^*)$, respectivamente, en la expresión de $R_{n,w}(\hat{\theta}_n)$.

Además, sea P_* la ley de probabilidad condicional bootstrap, dado $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$.

En forma análoga se describe $S^*_{n,w}(\hat{\theta}^*_n)$, la versión bootstrap de $S_{n,w}(\hat{\theta}_n)$.

Para probar que el método bootstrap se puede usar para aproximar consistentemente la distribución nula de $R_{n,w}(\hat{\theta}_n)$, o de $S_{n,w}(\hat{\theta}_n)$, supondremos las siguientes hipótesis que son un poco más fuertes que el Supuesto 2.2.1.

Supuesto 2.2.4 *Se cumple el Supuesto 2.2.1 y las funciones* $\boldsymbol{\ell}$ *y* J *satisfacen*

1. $\sup_{\vartheta\in\Theta_0} E_\vartheta\Big[\|\boldsymbol{\ell}(\boldsymbol{X};\vartheta)\|^2 I\{\|\boldsymbol{\ell}(\boldsymbol{X};\vartheta)\|>\gamma\}\Big]\to 0$, *cuando* $\gamma\to\infty$, *donde* $\Theta_0\subseteq\Theta$ *es una vecindad abierta de* θ.

2. $\boldsymbol{\ell}(\boldsymbol{X};\vartheta)$ *y* $J(\vartheta)$ *son continuas como funciones de* ϑ *en* $\vartheta=\theta$ *y* $J(\vartheta)$ *es finita* $\forall\vartheta\in\Theta_0$.

Tal como se estableció en la Observación 2.2.2, el Supuesto 2.2.4 no es restrictivo pues lo cumplen estimadores comúnmente usados.

Para mostrar que el método bootstrap aproxima consistentemente a la distribución nula de $R_{n,w}(\hat{\theta}_n)$, o de $S_{n,w}(\hat{\theta}_n)$, nos será útil el siguiente resultado, que lo volveremos a usar más adelante.

Lema 2.2.5 *Supongamos que se verifica el Supuesto 2.2.4 y que* $\hat{\theta}_n\xrightarrow{c.s.}\theta$, *para algún* $\theta\in\Theta$. *Entonces*

(a) $$\sum_{r,s\geq 0} r\left|\boldsymbol{\ell}_i(r,s;\hat{\theta}_n)P(r,s;\hat{\theta}_n)-\boldsymbol{\ell}_i(r,s;\theta)P(r,s;\theta)\right|=o(1),\ \ i=1,2,3,$$

(b) $$\sum_{r,s\geq 0} s\left|\boldsymbol{\ell}_i(r,s;\hat{\theta}_n)P(r,s;\hat{\theta}_n)-\boldsymbol{\ell}_i(r,s;\theta)P(r,s;\theta)\right|=o(1),\ \ i=1,2,3,$$

donde $P(r,s;\vartheta)=P_\vartheta(X_1=r,X_2=s)$, $\forall\vartheta\in\Theta_0$.

Demostración Solamente presentaremos la demostración de la parte (b), pues la demostración de la parte (a) sigue pasos similares.

Como
$$\sum_{r,s\geq 0}=\sum_{r,s\geq 0;\,\|\boldsymbol{\ell}(r,s;\vartheta)\|>\gamma}+\sum_{r,s\geq 0;\,\|\boldsymbol{\ell}(r,s;\vartheta)\|\leq\gamma},$$
consideraremos por separado cada uno de estos sumandos.

<u>Caso $\|\boldsymbol{\ell}(r,s;\vartheta)\|>\gamma$</u> Tengamos presente que $E_\theta\left(X_2^2\right)=\theta_2+\theta_2^2$. Sea $\varepsilon>0$ arbitrario pero fijo. Por el Supuesto 2.2.4 (1), $\exists\gamma=\gamma(\varepsilon)>0$ tal que

$$\sup_{\|\boldsymbol{\ell}(r,s;\vartheta)\|>\gamma}\sum_{r,s}\|\boldsymbol{\ell}(r,s;\vartheta)\|^2P(r,s;\vartheta)<\frac{\varepsilon^2}{4\left(\theta_2+\theta_2^2\right)},\quad\forall\vartheta\in\Theta_0 \tag{2.15}$$

y por tanto

$$\sup_{\|\boldsymbol{\ell}(r,s;\vartheta)\|>\gamma} \sum_{r,s} s\left|\boldsymbol{\ell}_i(r,s;\vartheta)P(r,s;\vartheta) - \boldsymbol{\ell}_i(r,s;\theta)P(r,s;\theta)\right| < \varepsilon, \tag{2.16}$$

$i = 1, 2, 3$, pues, por la desigualdad de Cauchy-Schwarz

$$\begin{aligned}
&\sum_{r,s} s\left|\boldsymbol{\ell}_i(r,s;\vartheta)P(r,s;\vartheta) - \boldsymbol{\ell}_i(r,s;\theta)P(r,s;\theta)\right| \\
&\quad\leq \sum_{r,s} s\left|\boldsymbol{\ell}_i(r,s;\vartheta)\right| P(r,s;\vartheta) + \sum_{r,s} s\left|\boldsymbol{\ell}_i(r,s;\theta)\right| P(r,s;\theta) \\
&\quad\leq \left(\theta_2 + \theta_2^2\right)^{\frac{1}{2}} \left\{\sum_{r,s} \|\boldsymbol{\ell}(r,s;\vartheta)\|^2 P(r,s;\vartheta)\right\}^{\frac{1}{2}} + \left(\theta_2 + \theta_2^2\right)^{\frac{1}{2}} \left\{\sum_{r,s} \|\boldsymbol{\ell}(r,s;\theta)\|^2 P(r,s;\theta)\right\}^{\frac{1}{2}}.
\end{aligned}$$

De lo anterior y considerando (2.15), obtenemos (2.16).

<u>Caso $\|\boldsymbol{\ell}(r,s;\vartheta)\| \leq \gamma$</u> Para esta situación, primero mostraremos la validez de (b) cuando $r, s \geq M$ y luego para $r, s < M$, donde M es una constante positiva que determinaremos más adelante.

Como $\sum_{r,s\geq 0} sP(r,s;\vartheta) = E_\vartheta(X_2) = \vartheta_2$ y las $P(r,s;\vartheta)$ son continuas como funciones de ϑ, entonces dado $\varepsilon_1 > 0$, $\exists M = M(\varepsilon_1) > 0$, de modo que

$$\sum_{r,s<M} sP(r,s;\vartheta) > \vartheta_2 - \frac{\varepsilon_1}{2\gamma},\ \forall\vartheta \in \Theta_1 \subseteq \Theta_0,$$

siempre que $\|\vartheta - \theta\| < \delta_1$, para cierto $\delta_1 > 0$. Por tanto, si $\|\boldsymbol{\ell}(r,s;\vartheta)\| \leq \gamma$, entonces

$$\sum_{r,s\geq M} s\left|\boldsymbol{\ell}_i(r,s;\vartheta)P(r,s;\vartheta) - \boldsymbol{\ell}_i(r,s;\theta)P(r,s;\theta)\right| < \varepsilon_1,\ \forall\vartheta \in \Theta_0 \cap \left\{\|\vartheta - \theta\| < \delta_1\right\}, \tag{2.17}$$

$i = 1, 2, 3$, pues

$$\vartheta_2 = \sum_{r,s\geq 0} sP(r,s;\vartheta) = \sum_{r,s\geq M} sP(r,s;\vartheta) + \sum_{r,s<M} sP(r,s;\vartheta) > \sum_{r,s\geq M} sP(r,s;\vartheta) + \vartheta_2 - \frac{\varepsilon_1}{2\gamma},$$

de donde

$$\sum_{r,s\geq M} sP(r,s;\vartheta) < \frac{\varepsilon_1}{2\gamma}.$$

Considerando esta última desigualdad y el hecho que $|\boldsymbol{\ell}_i(r,s;\vartheta)| \leq \|\boldsymbol{\ell}(r,s;\vartheta)\| \leq \gamma$, para $i = 1, 2, 3$, entonces

$$\sum_{r,s \geq M} s \left| \boldsymbol{\ell}_i(r,s;\vartheta) P(r,s;\vartheta) - \boldsymbol{\ell}_i(r,s;\theta) P(r,s;\theta) \right|$$

$$\leq \gamma \left\{ \sum_{r,s \geq M} sP(r,s;\vartheta) + \sum_{r,s \geq M} sP(r,s;\theta) \right\} < \varepsilon_1,$$

lo cual prueba (2.17).

Para $r, s < M$. Como $\boldsymbol{\ell}_i(r,s;\vartheta) P(r,s;\vartheta)$ son funciones continuas en ϑ, para $i = 1, 2, 3$, se tendrá que $\forall \varepsilon_2 > 0$, $\exists \delta_2 = \delta_2(\varepsilon_2) > 0$, tal que $\|\vartheta - \theta\| < \delta_2$ implica que

$$\sup_{r,s<M} \left| \boldsymbol{\ell}_i(r,s;\vartheta) P(r,s;\vartheta) - \boldsymbol{\ell}_i(r,s;\theta) P(r,s;\theta) \right| < \frac{\varepsilon_2}{M^3}$$

y por tanto,

$$\sum_{r,s<M} s \left| \boldsymbol{\ell}_i(r,s;\vartheta) P(r,s;\vartheta) - \boldsymbol{\ell}_i(r,s;\theta) P(r,s;\theta) \right|$$

$$< M \sum_{r,s<M} \left| \boldsymbol{\ell}_i(r,s;\vartheta) P(r,s;\vartheta) - \boldsymbol{\ell}_i(r,s;\theta) P(r,s;\theta) \right| < \varepsilon_2,$$

siempre que $\vartheta \in \{\|\vartheta - \theta\| < \delta_2\} \cap \Theta_0$, para cierto $\delta_2 > 0$. Lo cual concluye la demostración, pues ε_1 y ε_2 son arbitrarios. $\square$

También nos será de utilidad el siguiente resultado.

Proposición 2.2.6 *Sean $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid de $\boldsymbol{X} = (X_1, X_2) \in \mathbb{N}_0^2$. Supongamos que se verifica el Supuesto 2.2.4 y que $\hat{\theta}_n \xrightarrow{c.s.} \theta$, para algún $\theta \in \Theta$. Entonces*

(a) $\sup_{u,v \in [0,1]^2} \left| K_n^R(u,v) - K^R(u,v) \right| \xrightarrow{c.s.} 0$, *donde*

$$K_n^R(u,v) = E_* \left\{ R^0(\boldsymbol{X}_1^*; \hat{\theta}_n; u) R^0(\boldsymbol{X}_1^*; \hat{\theta}_n; v) \right\},$$

$$K^R(u,v) = E_\theta \left\{ R^0(\boldsymbol{X}_1; \theta; u) R^0(\boldsymbol{X}_1; \theta; v) \right\},$$

con $R^0(\boldsymbol{X}_1; \theta; u)$ el definido en el Teorema 2.2.3(a).

(b) $\sup_{u,v\in[0,1]^2} \left| K_n^S(u,v) - K^S(u,v) \right| \xrightarrow{c.s.} 0$*, donde*

$$K_n^S(u,v) = \left(E_* \left\{ S_i^0(\boldsymbol{X}_1^*; \hat{\theta}_n; u) S_j^0(\boldsymbol{X}_1^*; \hat{\theta}_n; v) \right\} \right), 1 \leq i,j \leq 2.$$

$$K^S(u,v) = \left(E_\theta \left\{ S_i^0(\boldsymbol{X}_1; \theta; u) S_j^0(\boldsymbol{X}_1; \theta; v) \right\} \right), 1 \leq i,j \leq 2.$$

con $S_i^0(\boldsymbol{X}_1;\theta;u)$ *el definido en el Teorema 2.2.3(b),* $i = 1,2$.

Demostración Solamente presentaremos la demostración de la parte (b), pues la demostración de la parte (a) sigue pasos similares.

El Supuesto 2.2.4 y $\hat{\theta}_n \xrightarrow{c.s.} \theta$ implican que

$$\sup_{u,v\in[0,1]^2} \left| E_* \left\{ S_i^0(\boldsymbol{X}_1^*; \hat{\theta}_n; u) S_j^0(\boldsymbol{X}_1^*; \hat{\theta}_n; v) \right\} - E_\theta \left\{ S_i^0(\boldsymbol{X}_1; \theta; u) S_j^0(\boldsymbol{X}_1; \theta; v) \right\} \right| \xrightarrow{c.s.} 0, \tag{2.18}$$

$1 \leq i,j \leq 2$.

Para probar este resultado, primero hacemos la resta de las esperanzas en (2.18) y encontramos, por ejemplo, que

$$\begin{aligned}
&E_*\left\{ S_1^0(\boldsymbol{X}_1^*; \hat{\theta}_n; u) S_2^0(\boldsymbol{X}_1^*; \hat{\theta}_n; v) \right\} - E_\theta\left\{ S_1^0(\boldsymbol{X}_1; \theta; u) S_2^0(\boldsymbol{X}_1; \theta; v) \right\} \\
&= v_1 u_2 \left[\left\{ \hat{\theta}_{3n} + \phi_1(r;\hat{\theta}_n)\phi_2(r;\hat{\theta}_n) \right\} g(r;\hat{\theta}_n) - \left\{ \theta_3 + \phi_1(r;\theta)\phi_2(r;\theta) \right\} g(r;\theta) \right] \\
&\quad - v_1 \left\{ \phi_1(r;\hat{\theta}_n)\phi_2(v;\hat{\theta}_n) g(r;\hat{\theta}_n) - \phi_1(r;\theta)\phi_2(v;\theta) g(r;\theta) \right\} \\
&\quad - u_2 \left\{ \phi_1(u;\hat{\theta}_n)\phi_2(r;\hat{\theta}_n) g(r;\hat{\theta}_n) - \phi_1(u;\theta)\phi_2(r;\theta) g(r;\theta) \right\} \\
&\quad + \phi_1(u;\hat{\theta}_n)\phi_2(v;\hat{\theta}_n) g(r;\hat{\theta}_n) - \phi_1(u;\theta)\phi_2(v;\theta) g(r;\theta) \\
&\quad + A(u)^\top \left\{ g(u;\hat{\theta}_n) g(v;\hat{\theta}_n) J(\hat{\theta}_n) - g(u;\theta) g(v;\theta) J(\theta) \right\} B(v) \\
&\quad + \left[g(v;\hat{\theta}_n)\phi_1(u;\hat{\theta}_n) E_*\left\{ u_1^{X_1^*} u_2^{X_2^*} \boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n) \right\} - g(v;\theta)\phi_1(u;\theta) E_\theta\left\{ u_1^{X_1} u_2^{X_2} \boldsymbol{\ell}(\boldsymbol{X}_1;\theta) \right\} \right] B(v) \\
&\quad + \left[g(u;\hat{\theta}_n)\phi_2(v;\hat{\theta}_n) E_*\left\{ v_1^{X_1^*} v_2^{X_2^*} \boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n) \right\} - g(u;\theta)\phi_2(v;\theta) E_\theta\left\{ v_1^{X_1} v_2^{X_2} \boldsymbol{\ell}(\boldsymbol{X}_1;\theta) \right\} \right] A(u), \\
&\quad - \left[g(v;\hat{\theta}_n) E_*\left\{ \psi_1(u;\boldsymbol{X}_1^*;\hat{\theta}_n) \right\} - g(v;\theta) E_\theta\left\{ \psi_1(u;\boldsymbol{X}_1;\theta) \right\} \right] B(v) \\
&\quad - \left[g(u;\hat{\theta}_n) E_*\left\{ \psi_2(v;\boldsymbol{X}_1^*;\hat{\theta}_n) \right\} - g(u;\theta) E_\theta\left\{ \psi_2(v;\boldsymbol{X}_1;\theta) \right\} \right] A(u)
\end{aligned}$$

donde

$$r_1 = u_1 v_1,\ r_2 = u_2 v_2,\ r = (r_1, r_2),\quad A(u) = (1, 0, u_2 - 1)^\top,\ B(v) = (0, 1, v_1 - 1)^\top,$$

$$\phi_1(w; \vartheta) = \vartheta_1 + \vartheta_3(w_2 - 1),\quad \phi_2(w; \vartheta) = \vartheta_2 + \vartheta_3(w_1 - 1),\quad w = (w_1, w_2),$$

$$\psi_1(u; \boldsymbol{X}_1; \vartheta) = X_1\, I\{X_1 \geq 1\}\, u_1^{X_1 - 1}\, u_2^{X_2}\, \boldsymbol{\ell}(\boldsymbol{X}_1; \vartheta),$$

$$\psi_2(u; \boldsymbol{X}_1; \vartheta) = X_2\, I\{X_2 \geq 1\}\, u_1^{X_1}\, u_2^{X_2 - 1}\, \boldsymbol{\ell}(\boldsymbol{X}_1; \vartheta),$$

Las otras diferencias de (2.18) tienen expresiones análogas a la diferencia de esperanzas anterior. Las situaciones que aparecen en tales esperanzas son de alguno de los siguientes casos:

- Caso 1: sea $f(u, v; \vartheta)$ una función polinomial en las variables $u, v \in [0, 1]^2$ y $\vartheta \in \Theta_1 \subseteq \Theta_0$, donde Θ_1 es un conjunto compacto que contiene a θ. Luego, $f(u, v; \theta) g(u; \theta)$ es continua como función de u, v y θ, por lo tanto, es uniformemente continua en $[0, 1]^2 \times \Theta_1$ y como $\hat{\theta}_n \xrightarrow{c.s.} \theta$, entonces

$$\sup_{u,v \in [0,1]^2} \left| f(u, v; \hat{\theta}_n) g(u; \hat{\theta}_n) - f(u, v; \theta) g(u; \theta) \right| = o(1).$$

- Caso 2: como $g(u; \theta) g(v; \theta) J(\theta)$ es continua como función de θ, entonces es uniformemente continua en Θ_1, donde Θ_1 es un conjunto compacto que contiene a θ. Además, como $u, v \in [0, 1]^2$ y $\hat{\theta}_n \xrightarrow{c.s.} \theta$, entonces

$$\sup_{u,v \in [0,1]^2} \left| A(u)^\top \left[g(u; \hat{\theta}_n) g(v; \hat{\theta}_n) J(\hat{\theta}_n) - g(u; \theta) g(v; \theta) J(\theta) \right] B(v) \right| = o(1).$$

- Caso 3: sea $f(u, v; \vartheta)$ una función polinomial en las variables $u, v \in [0, 1]^2$ y $\vartheta \in \Theta_1 \subseteq \Theta_0$, donde Θ_1 es un conjunto compacto que contiene a θ. Por lo tanto f es continua como función de ϑ y también lo es $g(u; \vartheta)$, luego $f(u, v; \hat{\theta}_n) = f(u, v; \theta) + o(1)$ y $g(u; \hat{\theta}_n) = g(u; \theta) + o(1)$, con lo cual

$$\begin{aligned}
f(u, v; \hat{\theta}_n) g(u; \hat{\theta}_n) E_* \left\{ v_1^{X_1^*} v_2^{X_2^*} \boldsymbol{\ell}(\boldsymbol{X}_1^*; \hat{\theta}_n) \right\} - f(u, v; \theta) g(u; \theta) E_\theta \left\{ v_1^{X_1} v_2^{X_2} \boldsymbol{\ell}(\boldsymbol{X}_1; \theta) \right\} \\
= f(u, v; \theta) g(u; \theta) \left[E_* \left\{ v_1^{X_1^*} v_2^{X_2^*} \boldsymbol{\ell}(\boldsymbol{X}_1^*; \hat{\theta}_n) \right\} - E_\theta \left\{ v_1^{X_1} v_2^{X_2} \boldsymbol{\ell}(\boldsymbol{X}_1; \theta) \right\} \right] \\
+ o(1) \{ f(u, v; \theta) + g(u; \theta) + 1 \} E_* \left\{ v_1^{X_1^*} v_2^{X_2^*} \boldsymbol{\ell}(\boldsymbol{X}_1^*; \hat{\theta}_n) \right\}.
\end{aligned}$$

Usando el Lema 2.2.5 y el hecho que $\hat{\theta}_n \in \Theta_0$, resulta

$$
\begin{aligned}
E_*\Big\{v_1^{X_1^*} v_2^{X_2^*} \boldsymbol{\ell}_i(\boldsymbol{X}_1^*;\hat{\theta}_n)\Big\} &- E_\theta\{v_1^{X_1} v_2^{X_2} \boldsymbol{\ell}_i(\boldsymbol{X}_1;\theta)\} \\
&= \sum_{r,s\geq 0} v_1^r v_2^s \Big\{\boldsymbol{\ell}_i(r,s;\hat{\theta}_n)P(r,s;\hat{\theta}_n) - \boldsymbol{\ell}_i(r,s;\theta)P(r,s;\theta)\Big\} \\
&\leq \sum_{r,s\geq 0} \Big|\boldsymbol{\ell}_i(r,s;\hat{\theta}_n)P(r,s;\hat{\theta}_n) - \boldsymbol{\ell}_i(r,s;\theta)P(r,s;\theta)\Big| = o(1),
\end{aligned}
$$

para $i = 1, 2, 3$. Por lo tanto,

$$
\Big|E_*\Big\{v_1^{X_1^*} v_2^{X_2^*} \boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\Big\} - E_\theta\big\{v_1^{X_1} v_2^{X_2} \boldsymbol{\ell}(\boldsymbol{X}_1;\theta)\big\}\Big| = \mathbf{o}(1).
$$

Ahora, puesto que $J(\hat{\theta}_n) < \infty$, entonces

$$
\Big\|E_*\Big\{v_1^{X_1^*} v_2^{X_2^*} \boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\Big\}\Big\| \leq \Big\{E_*\Big(\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\|^2\Big)\Big\}^{1/2} < \infty.
$$

Además, como $|g(u;\theta)| \leq 1$ y $|f(u,v;\theta)| < \infty$, $\forall u, v \in [0,1]^2$, entonces

$$
\sup_{u,v\in[0,1]^2} \Big|\Big[\phi(u,v;\hat{\theta}_n)E_*\Big\{\varphi(v;\boldsymbol{X}_1^*;\hat{\theta}_n)\Big\} - \phi(u,v;\theta)E_\theta\{\varphi(v;\boldsymbol{X}_1;\theta)\}\Big]\,A(u)\Big| = o(1),
$$

donde $\phi(u,v;\vartheta) = f(u,v;\vartheta)g(u;\vartheta)$ y $\varphi(v;\boldsymbol{X}_1;\vartheta) = v_1^{X_1} v_2^{X_2} \boldsymbol{\ell}(\boldsymbol{X}_1;\vartheta)$.

- Caso 4: como $g(u;\hat{\theta}_n) = g(u;\theta) + o(1)$, pues es continua como función de $\vartheta \in \Theta_0$, entonces

$$
\begin{aligned}
g(u;\hat{\theta}_n)\, E_*\Big\{X_2^*\, I\{X_2^* \geq 1\}\, v_1^{X_1^*}\, v_2^{X_2^*-1} \boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\Big\}& \\
- g(u;\theta) E_\theta \left\{X_2\, I\{X_2 \geq 1\}\, v_1^{X_1}\, v_2^{X_2-1} \boldsymbol{\ell}(\boldsymbol{X}_1;\theta)\right\}& \\
= g(u;\theta)\Big[E_*\Big\{X_2^*\, I\{X_2^* \geq 1\}\, v_1^{X_1^*}\, v_2^{X_2^*-1} \boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\Big\}& \\
- E_\theta \left\{X_2\, I\{X_2 \geq 1\}\, v_1^{X_1}\, v_2^{X_2-1} \boldsymbol{\ell}(\boldsymbol{X}_1;\theta)\right\}\Big]& \\
+ o(1)\, E_*\Big\{X_2^*\, I\{X_2^* \geq 1\}\, v_1^{X_1^*}\, v_2^{X_2^*-1} \boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\Big\}&.
\end{aligned}
$$

Usando el Lema 2.2.5 y el hecho que $\hat{\theta}_n \in \Theta_0$, obtenemos

$$E_*\Big\{X_2^* I\{X_2^* \geq 1\} v_1^{X_1^*} v_2^{X_2^*-1} \boldsymbol{\ell}_i(\boldsymbol{X}_1^*; \hat{\theta}_n)\Big\} - E_\theta\big\{X_2 I\{X_2 \geq 1\} v_1^{X_1} v_2^{X_2-1} \boldsymbol{\ell}_i(\boldsymbol{X}_1; \theta)\big\}$$

$$= \sum_{r\geq 0, s\geq 1} s\, v_1^r v_2^{s-1} \Big\{\boldsymbol{\ell}_i(r,s;\hat{\theta}_n) P(r,s;\hat{\theta}_n) - \boldsymbol{\ell}_i(r,s;\theta) P(r,s;\theta)\Big\}$$

$$\leq \sum_{r,s\geq 0} s \Big|\boldsymbol{\ell}_i(r,s;\hat{\theta}_n) P(r,s;\hat{\theta}_n) - \boldsymbol{\ell}_i(r,s;\theta) P(r,s;\theta)\Big| = o(1),$$

para $i = 1, 2, 3$. Por lo tanto,

$$\Big|E_*\Big\{\psi_2(v; \boldsymbol{X}_1^*; \hat{\theta}_n)\Big\} - E_\theta\{\psi_2(v; \boldsymbol{X}_1; \theta)\}\Big| = \mathbf{o}(1),$$

donde $\psi_2(v; \boldsymbol{X}_1; \vartheta) = X_2\, I\{X_2 \geq 1\}\, v_1^{X_1}\, v_2^{X_2-1} \boldsymbol{\ell}(\boldsymbol{X}_1; \vartheta)$.

Por otra parte,

$$E_*\Big\{\psi_2(v; \boldsymbol{X}_1^*; \hat{\theta}_n)\Big\} \leq \big\{E_*\big(X_2^{*2}\big)\big\}^{1/2} \Big\{E_*\Big(\|\boldsymbol{\ell}(\boldsymbol{X}_1^*; \hat{\theta}_n)\|^2\Big)\Big\}^{1/2} < \infty,$$

con lo cual,

$$\sup_{u,v\in[0,1]^2} \Big|\Big[g(u;\hat{\theta}_n) E_*\Big\{\psi_2(v; \boldsymbol{X}_1^*; \hat{\theta}_n)\Big\} - g(u;\theta) E_\theta\{\psi_2(v; \boldsymbol{X}_1; \theta)\}\Big] A(u)\Big| = o(1).$$

La presentación de estos cuatro casos generales muestran que se verifica (2.18) y con ello se consigue el resultado. □

Teorema 2.2.7 *Sean $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid de $\boldsymbol{X} = (X_1, X_2) \in \mathbb{N}_0^2$. Supongamos que se verifica el Supuesto 2.2.4 y que $\hat{\theta}_n \xrightarrow{c.s.} \theta$, para algún $\theta \in \Theta$. Entonces*

(a) $\sup_{x\in\mathbb{R}} \Big|P_*\Big\{R_{n,w}^*(\hat{\theta}_n^*) \leq x\Big\} - P_\theta\Big\{R_{n,w}(\hat{\theta}_n) \leq x\Big\}\Big| \xrightarrow{c.s.} 0.$

(b) $\sup_{x\in\mathbb{R}} \Big|P_*\Big\{S_{n,w}^*(\hat{\theta}_n^*) \leq x\Big\} - P_\theta\Big\{S_{n,w}(\hat{\theta}_n) \leq x\Big\}\Big| \xrightarrow{c.s.} 0.$

Demostración Solamente presentaremos la demostración de la parte (b), pues la demostración de la parte (a) sigue pasos similares.

Por definición, $S_{n,w}^*(\hat{\theta}_n^*) = \|Z_{1n}^*\|_{\mathcal{H}}^2 + \|Z_{2n}^*\|_{\mathcal{H}}^2$, con

$$Z_{kn}^*(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_k(\boldsymbol{X}_i^*; \hat{\theta}_n^*; u),\ k = 1, 2.$$

Siguiendo pasos similares a los dados en la demostración del Teorema 2.2.3(b) se puede ver que

$$Z^*_{kn}(u) = S^*_{kn}(u) + s^*_{kn},\ k = 1, 2,$$

con $\|s^*_{kn}\|_{\mathcal{H}} = o_{P_*}(1)$ c.s., $k = 1, 2$, donde $S^*_{kn}(u)$ es definido como $S_{kn}(u)$ que aparece en el Teorema 2.2.3, con $\boldsymbol{X}_i$ y θ reemplazados por $\boldsymbol{X}^*_i$ y $\hat{\theta}_n$, respectivamente.

Por conveniencia analítica, ahora consideraremos el siguiente espacio de Hilbert separable:

$$\mathcal{H}_1 = \Bigg\{ \varphi : [0,1]^2 \to \mathbb{R}^2,\ \text{donde } \varphi(u) = (\varphi_1(u), \varphi_2(u)) \text{ es una función medible tal que}$$

$$\|\varphi\|^2_{\mathcal{H}_1} = \int_{[0,1]^2} \left\{ \varphi_1^2(u) + \varphi_2^2(u) \right\} w(u) du < \infty \Bigg\}.$$

con producto escalar

$$\langle (\phi_1, \phi_2), (\psi_1, \psi_2) \rangle_{\mathcal{H}_1} = \int_{[0,1]^2} \{\phi_1(u)\psi_1(u) + \phi_2(u)\psi_2(u)\} w(u) du < \infty.$$

Claramente, $\|\varphi\|^2_{\mathcal{H}_1} = \|\varphi_1\|^2_{\mathcal{H}} + \|\varphi_2\|^2_{\mathcal{H}}$.

Sea

$$Y^*_n(u) = \sum_{i=1}^n Y^*_{ni}(u)$$

donde

$$Y^*_{ni}(u) = \frac{1}{\sqrt{n}} S(\boldsymbol{X}^*_i; \hat{\theta}_n; u),\ 1 \leq i \leq n,$$

$$S(\boldsymbol{X}^*_i; \hat{\theta}_n; u) = \Big(S^0_1(\boldsymbol{X}^*_i; \hat{\theta}_n; u), S^0_2(\boldsymbol{X}^*_i; \hat{\theta}_n; u) \Big),\ 1 \leq i \leq n. \tag{2.19}$$

Observar primero que $Y^*_{ni}(u)$ tiene medias cero y momentos segundos finitos, para $1 \leq i \leq n$, siempre que $\hat{\theta}_n \in \Theta_0$, lo cual ocurre c.s. porque $\hat{\theta}_n \xrightarrow{c.s.} \theta$.

Consideremos el núcleo de covarianza $K^S_n(u, v) = E_*\big\{Y^*_n(u)^\top Y^*_n(v)\big\}$, luego

$$K^S_n(u, v) = \left(E_* \left\{ S^0_i(\boldsymbol{X}^*_1; \hat{\theta}_n; u) S^0_j(\boldsymbol{X}^*_1; \hat{\theta}_n; v) \right\} \right), 1 \leq i, j \leq 2.$$

Además, sea $K^S(u, v) = E_\theta \left\{ S(\boldsymbol{X}_1; \theta; u)^\top S(\boldsymbol{X}_1; \theta; v) \right\}$ donde S es como el definido en (2.19), es decir,

$$K^S(u, v) = \Big(E_\theta \left\{ S^0_i(\boldsymbol{X}_1; \theta; u) S^0_j(\boldsymbol{X}_1; \theta; v) \right\} \Big), 1 \leq i, j \leq 2.$$

Sea $\mathcal{Z}$ un proceso Gaussiano centrado cuyo operador de covarianza C es caracterizado por

$$\begin{aligned}\langle Cf, h\rangle_{\mathcal{H}_1} &= cov\left(\langle \mathcal{Z}, f\rangle_{\mathcal{H}_1}, \langle \mathcal{Z}, h\rangle_{\mathcal{H}_1}\right) = E_\theta\left\{\langle \mathcal{Z}, f\rangle_{\mathcal{H}_1}\langle \mathcal{Z}, h\rangle_{\mathcal{H}_1}\right\} \\ &= \int_{[0,1]^4} f(u)K^S(u,v)h(v)^\top w(u)w(v)dudv.\end{aligned} \tag{2.20}$$

Por el Lema 1.2.4, $(S_{1n}, S_{2n}) \xrightarrow{L} \mathcal{Z}$ en $\mathcal{H}_1$, cuando los datos son iid provenientes del vector aleatorio $\boldsymbol{X} \sim PB(\theta)$, donde $S_{kn}(u)$ es como el definido en el Teorema 2.2.3(b), $k = 1, 2$.

Sea $\{e_k : k \geq 0\}$ una base ortonormal de $\mathcal{H}_1$. A continuación probaremos que se cumplen las condiciones (i)-(iii) del Lema 1.2.5.

(i) Sea C_n el operador de covarianza de Y_n^*, esto es, sean $f, h \in \mathcal{H}_1$, luego

$$\begin{aligned}\langle C_n f, h\rangle_{\mathcal{H}_1} &= cov(\langle Y_n^*, f\rangle_{\mathcal{H}_1}, \langle Y_n^*, h\rangle_{\mathcal{H}_1}) = E_*\{\langle Y_n^*, f\rangle_{\mathcal{H}_1}\langle Y_n^*, h\rangle_{\mathcal{H}_1}\} \\ &= \int_{[0,1]^4} f(u)E_*\left\{Y_n^*(u)^\top Y_n^*(v)\right\}h(v)^\top w(u)w(v)dudv \\ &= \int_{[0,1]^4} f(u)K_n^S(u,v)h(v)^\top w(u)w(v)dudv.\end{aligned} \tag{2.21}$$

De (2.20), (2.21) y de la Proposición 2.2.6(b),

$$\begin{aligned}\lim_{n\to\infty}\langle C_n e_k, e_l\rangle_{\mathcal{H}_1} &= \lim_{n\to\infty}\int_{[0,1]^4} e_k(u)K_n^S(u,v)e_l(v)^\top w(u)w(v)dudv \\ &= \int_{[0,1]^4} e_k(u)K^S(u,v)e_l(v)^\top w(u)w(v)dudv = \langle Ce_k, e_l\rangle_{\mathcal{H}_1} = a_{kl}.\end{aligned}$$

(ii) De (2.20), (2.21), de la Proposición 2.2.6(b) y del ítem (i), se obtiene

$$\begin{aligned}\lim_{n\to\infty}\sum_{k=0}^{\infty}\langle C_n e_k, e_k\rangle_{\mathcal{H}_1} &= \lim_{n\to\infty}\sum_{k=0}^{\infty}\int_{[0,1]^4} e_k(u)K_n^S(u,v)e_k(v)^\top w(u)w(v)dudv \\ &= \sum_{k=0}^{\infty}\int_{[0,1]^4} e_k(u)K^S(u,v)e_k(v)^\top w(u)w(v)dudv \\ &= \sum_{k=0}^{\infty}\langle Ce_k, e_k\rangle_{\mathcal{H}_1} = \sum_{k=0}^{\infty} a_{kk} < \infty,\end{aligned}$$

pues, de la primera ecuación en (2.20) y la igualdad de Parseval

$$\sum_{k=0}^{\infty} a_{kk} = \sum_{k=0}^{\infty} \langle Ce_k, e_k \rangle_{\mathcal{H}_1} = \sum_{k=0}^{\infty} E_\theta \Big\{ \langle \mathcal{Z}, e_k \rangle^2_{\mathcal{H}_1} \Big\} = E_\theta \Big\{ \|\mathcal{Z}\|^2_{\mathcal{H}_1} \Big\} < \infty.$$

(iii) Como $|g(u;\hat{\theta}_n)| \leq 1$, entonces

$$|S_k^0(\boldsymbol{X}_i^*;\hat{\theta}_n;u)| \leq X_{ki}^* + \hat{\theta}_{kn} + \hat{\theta}_{3n} + \sqrt{2}\|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\|, \ \forall u \in [0,1]^2, \ k = 1, 2.$$

Sea $0 < M = \int_{[0,1]^2} w(u)du < \infty$, luego

$$\langle Y_{ni}^*, e_k \rangle^2_{\mathcal{H}_1} \leq \frac{4M}{n} \left(\|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| + \frac{1}{2} A_i \right)^2,$$

donde $A_i = (X_{1i}^* + \hat{\theta}_{1n} + \hat{\theta}_{3n}) + (X_{2i}^* + \hat{\theta}_{2n} + \hat{\theta}_{3n})$, siempre que $\hat{\theta}_n \in \Theta_0$, lo cual ocurre c.s. puesto que $\hat{\theta}_n \xrightarrow{c.s.} \theta$.

Ahora, como

$$\|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| + \frac{1}{2} A_i \leq \begin{cases} 2\|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\|, & \text{si } A_i \leq 2\|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| \\ A_i, & \text{si } 2\|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| < A_i \end{cases},$$

entonces

$$\varepsilon < \left| \langle Y_{ni}^*, e_k \rangle_{\mathcal{H}_1} \right| \leq \frac{2\sqrt{M}}{\sqrt{n}} \left(\|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| + \frac{1}{2} A_i \right),$$

implica que

$$\begin{aligned} I\left\{ \left| \langle Y_{ni}^*, e_k \rangle_{\mathcal{H}_1} \right| > \varepsilon \right\} &\leq I\left\{ \|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| + \frac{1}{2} A_i > \frac{\sqrt{n}\,\varepsilon}{2\sqrt{M}} \right\} \\ &= I\left\{ \|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| > \gamma, \ A_i \leq 2\|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| \right\} \\ &\quad + I\left\{ A_i > \frac{\sqrt{n}\,\varepsilon}{2\sqrt{M}}, \ A_i > 2\|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| \right\}, \end{aligned}$$

donde $\gamma = \gamma(n, \varepsilon) = \frac{\sqrt{n}\,\varepsilon}{4\sqrt{M}}$ es tal que $\gamma \to \infty$ cuando $n \to \infty$, pues $0 < M < \infty$.

Por lo tanto,

$$\begin{aligned} &\langle Y_{ni}^*, e_k \rangle^2_{\mathcal{H}_1} I\left\{ \left| \langle Y_{ni}^*, e_k \rangle_{\mathcal{H}_1} \right| > \varepsilon \right\} \\ &\qquad \leq \frac{16M}{n} \, \|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\|^2 \, I\left\{ \|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| > \gamma, \ A_i \leq 2\|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| \right\} \\ &\qquad\quad + \frac{4M}{n} A_i^2 \, I\left\{ A_i > \frac{\sqrt{n}\,\varepsilon}{2\sqrt{M}}, \ A_i > 2\|\boldsymbol{\ell}(\boldsymbol{X}_i^*;\hat{\theta}_n)\| \right\}. \end{aligned}$$

En consecuencia,

$$L_n(\varepsilon, e_k) = \sum_{i=1}^{n} E_*\Big(\langle Y_{ni}^*, e_k\rangle_{\mathcal{H}_1}^2 I\big\{\big|\langle Y_{ni}^*, e_k\rangle_{\mathcal{H}_1}\big| > \varepsilon\big\}\Big) \le L_{1n}(\varepsilon, e_k) + L_{2n}(\varepsilon, e_k),$$

donde,

$$\begin{aligned}
L_{1n}(\varepsilon, e_k) &= \frac{16M}{n}\sum_{i=1}^{n} E_*\Big[\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\|^2\, I\Big\{\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\| > \gamma,\ A_1 \le 2\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\|\Big\}\Big] \\
&= 16ME_*\Big[\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\|^2\, I\Big\{\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\| > \gamma,\ A_1 \le 2\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\|\Big\}\Big] \\
&\le 16M \sup_{\hat{\theta}_n\in\Theta_0} E_{\hat{\theta}_n}\Big[\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\|^2\, I\Big\{\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\| > \gamma\Big\}\Big]
\end{aligned}$$

y como $\gamma \to \infty$ cuando $n \to \infty$, entonces, por el Supuesto 2.2.4(1) se concluye que

$$\lim_{n\to\infty} L_{1n}(\varepsilon, e_k) = 0. \tag{2.22}$$

Considerando la desigualdad de Hölder y el hecho que la función indicadora es tal que $0 \le I\{C\} \le 1$, podemos escribir

$$\begin{aligned}
L_{2n}(\varepsilon, e_k) &= \frac{4M}{n}\sum_{i=1}^{n} E_*\left(A_1^2\, I\left\{A_1 > \frac{\sqrt{n}\,\varepsilon}{2\sqrt{M}},\ A_1 > 2\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\|\right\}\right) \\
&= 4ME_*\left(A_1^2\, I\left\{A_1 > \frac{\sqrt{n}\,\varepsilon}{2\sqrt{M}},\ A_1 > 2\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\|\right\}\right) \\
&\le 4M\left\{E_*\big(A_1^3\big)\right\}^{2/3}\left[E_*\left(I\left\{A_1 > \frac{\sqrt{n}\,\varepsilon}{2\sqrt{M}},\ A_1 > 2\|\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat{\theta}_n)\|\right\}\right)\right]^{1/3} \\
&= 4M\left\{E_*\big(A_1^3\big)\right\}^{2/3}\left\{P_*(A_1 > t)\right\}^{1/3},
\end{aligned}$$

donde $t = t(n,\varepsilon) = \frac{\sqrt{n}\,\varepsilon}{2\sqrt{M}}$ es tal que $t \to \infty$ cuando $n \to \infty$, pues $0 < M < \infty$.

Puesto que $|A_1| = A_1$ y $E_*(A_1^3) < \infty$, el Corolario 1.14 (ii) en Serfling (1980, p. 47) implica que $P_*(A_1 > t) \to 0$ cuando $t \to \infty$. Además, como $0 < M < \infty$, entonces

$$\lim_{n\to\infty} L_{2n}(\varepsilon, e_k) = 0. \tag{2.23}$$

De (2.22) y (2.23), concluimos que

$$\lim_{n\to\infty} L_n(\varepsilon, e_k) = 0,$$

de donde se cumple la condición (iii) de Kundu et al. (2000).

Así, del Lema 1.2.5, tenemos que $Y_n^* \xrightarrow{L} \mathcal{Z}$, c.s., en $\mathcal{H}_1$. Como habíamos mencionado antes, $\mathcal{Z}$ es también el límite débil de (S_{1n}, S_{2n}) cuando los datos son iid del vector aleatorio $\boldsymbol{X} \sim PB(\theta)$. Finalmente, el resultado sigue del teorema de la aplicación continua. □

Observación 2.2.8 *Es importante observar que el resultado del Teorema 2.2.7 se cumple si H_0 es verdadera o no.*

Si de hecho, H_0 es verdadera, el Teorema 2.2.7 implica que la distribución condicional de $R^*_{n,w}(\hat{\theta}^*_n)$ está cercana c.s. a la distribución nula de $R_{n,w}(\hat{\theta}_n)$.

Si H_0 es falsa, entonces el Teorema 2.2.7 nos dice que la distribución condicional de $R^*_{n,w}(\hat{\theta}^*_n)$ y la distribución de $R_{n,w}(\hat{\theta}_n)$, cuando la muestra es tomada de una población con distribución $PB(\theta)$, están c.s. cercanas, donde θ es el límite c.s. de $\hat{\theta}_n$.

Sea

$$r^*_{n,w,\alpha} = \text{ínf}\Big\{x : P_*(R^*_{n,w}(\hat{\theta}^*_n) \geq x) \leq \alpha\Big\}$$

el percentil superior α de la distribución bootstrap de $R_{n,w}(\hat{\theta}_n)$.

Del Teorema 2.2.7, la función test

$$\Psi^*_R = \begin{cases} 1, & \text{si } R_{n,w}(\hat{\theta}_n) \geq r^*_{n,w,\alpha}, \\ 0, & \text{en caso contrario,} \end{cases}$$

o equivalentemente, el test que rechaza H_0 cuando

$$p^*_R = P_*\Big(R^*_{n,w}(\hat{\theta}^*_n) \geq R_{obs}\Big) \leq \alpha,$$

es asintóticamente correcto en el sentido que, cuando H_0 es cierta

$$\lim_{n\to\infty} P_\theta(\Psi^*_R = 1) = \alpha,$$

donde R_{obs} es el valor observado del test estadístico $R_{n,w}(\hat{\theta}_n)$.

Similarmente, sea

$$s^*_{n,w,\alpha} = \text{ínf}\Big\{x : P_*(S^*_{n,w}(\hat{\theta}^*_n) \geq x) \leq \alpha\Big\}$$

el percentil superior α de la distribución bootstrap de $S_{n,w}(\hat{\theta}_n)$.

Del Teorema 2.2.7, la función test

$$\Psi_S^* = \begin{cases} 1, & \text{si } S_{n,w}(\hat{\theta}_n) \geq s^*_{n,w,\alpha}, \\ 0, & \text{en caso contrario,} \end{cases}$$

o equivalentemente, el test que rechaza H_0 cuando

$$p^*_{_S} = P_*\Big(S^*_{n,w}(\hat{\theta}^*_n) \geq S_{obs}\Big) \leq \alpha,$$

es asintóticamente correcto en el sentido que, cuando H_0 es cierta

$$\lim_{n\to\infty} P_\theta(\Psi_S^* = 1) = \alpha,$$

donde S_{obs} es el valor observado del test estadístico $S_{n,w}(\hat{\theta}_n)$.

Como es usual, en la práctica, $r^*_{n,w,\alpha}$ o $p^*_{_R}$ deben ser aproximados por simulación. Dados $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid de $\boldsymbol{X} = (X_1, X_2) \in \mathbb{N}_0^2$, siguiremos los siguientes pasos:

(1) Calcular el estimador $\hat{\theta}_n$ y calcular R_{obs}, el valor de $R_{n,w}(\hat{\theta}_n)$ para la muestra original.

(2) Generar una muestra bootstrap, digamos, $\boldsymbol{X}^*_1, \boldsymbol{X}^*_2, \ldots, \boldsymbol{X}^*_n$ iid desde la distribución $PB(\hat{\theta}_n)$.

(3) Sobre la base del bootstrap, calcular el estimador $\hat{\theta}^*_n$ y el valor del test estadístico, digamos R^*.

(4) Calcular $R_{n,w}(\hat{\theta}^*_n)$ para cada muestra bootstrap y denotar por $R^*_b,\ b = 1, 2, \ldots, B$, el respectivo valor resultante.

(5) Aproximar el $p-$valor bootstrap, $p^*_{_R}$, por medio de la expresión

$$\hat{p}_{_R} = \frac{card\{b : R^*_b \geq R_{obs}\}}{B},$$

o aproximar el punto crítico, $r^*_{n,w,\alpha}$, por $R^*_{a:B}$, donde $a = [(1-\alpha)B] + 1$, $[x]$ es la parte entera de x, y $R^*_{1:B}, R^*_{2:B}, \ldots, R^*_{B:B}$ son los valores $R^*_b,\ b = 1, 2, \ldots, B$, en orden creciente.

Observación 2.2.9 *Para aproximar $s^*_{n,w,\alpha}$ se sigue el mismo procedimiento anterior, con los cambios obvios.*

2.3. Alternativas

En esta sección estudiaremos el comportamiento de los tests propuestos Ψ_R^* y Ψ_S^* bajo alternativas fijas y locales.

2.3.1. Alternativas fijas

Como uno de nuestros objetivos es la consistencia de los tests de bondad de ajuste, lo próximo que haremos es estudiar este tópico para los tests que proponemos. Con este propósito, primero obtendremos el límite c.s. de $\frac{1}{n}R_{n,w}$ y $\frac{1}{n}S_{n,w}$.

Teorema 2.3.1 *Sean $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid de $\boldsymbol{X} = (X_1, X_2) \in \mathbb{N}_0^2$ con fgp $g(u)$. Si $\hat{\theta}_n \xrightarrow{c.s.} \theta$, para algún $\theta \in \mathbb{R}^3$, entonces*

(a) $\displaystyle \frac{1}{n}R_{n,w}(\hat{\theta}_n) \xrightarrow{c.s.} \int_0^1 \int_0^1 \Big\{g(u) - g(u;\theta)\Big\}^2 w(u)du = \eta(g;\theta).$

(b) Si $g(u) \in G_2$, $\displaystyle \frac{1}{n}S_{n,w}(\hat{\theta}_n) \xrightarrow{c.s.} \int_0^1 \int_0^1 \left\{D_1^2(u;\theta) + D_2^2(u;\theta)\right\} w(u)du = \xi(g;\theta).$

Demostración Solamente presentaremos la demostración de la parte (b), pues la demostración de la parte (a) sigue pasos similares.

Por definición

$$\frac{1}{n}S_{n,w}(\hat{\theta}_n) = \|D_{1n}\|_{\mathcal{H}}^2 + \|D_{2n}\|_{\mathcal{H}}^2,$$

donde $D_{kn} = D_{kn}(u;\hat{\theta}_n), k = 1,2$, están definidas en (2.5).

De las ecuaciones (2.3) y (2.5), obtenemos

$$\begin{aligned}
\Big|D_{1n}(u;\hat{\theta}_n) - D_1(u;\theta)\Big| &\leq r_1 + |\theta_1 - \hat{\theta}_{1n}| + |\theta_3 - \hat{\theta}_{3n}| + |\hat{\theta}_{1n} + \hat{\theta}_{3n}|r_0,\\
\Big|D_{2n}(u;\hat{\theta}_n) - D_2(u;\theta)\Big| &\leq r_2 + |\theta_2 - \hat{\theta}_{2n}| + |\theta_3 - \hat{\theta}_{3n}| + |\hat{\theta}_{2n} + \hat{\theta}_{3n}|r_0,
\end{aligned}$$

$\forall u \in [0,1]^2$, donde

$$r_0 = \sup_{u\in[0,1]^2} |g_n(u) - g(u)|$$

y

$$r_i = \sup_{u\in[0,1]^2} \left|\frac{\partial}{\partial u_i} g_n(u) - \frac{\partial}{\partial u_i} g(u)\right|, \; i = 1,2.$$

De la Proposición 1.2.2, $r_i = o(1)$, $i = 0, 1, 2$. El hecho que $|\theta_i - \hat{\theta}_{in}| = o(1)$, $i = 1, 2, 3$, junto con (2.2), implican que

$$\frac{1}{n} S_{n,w}(\hat{\theta}_n) = \|D_1\|^2_{\mathcal{H}} + \|D_2\|^2_{\mathcal{H}} + o(1),$$

lo que demuestra el resultado porque $\|D_1\|^2_{\mathcal{H}} + \|D_2\|^2_{\mathcal{H}} = \xi(g; \theta)$. $\square$

Como una consecuencia de los Teoremas 2.2.3, 2.2.7 y 2.3.1, el siguiente resultado da la consistencia de los tests Ψ^*_R y Ψ^*_S.

Corolario 2.3.2 *Sean $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid de $\boldsymbol{X} \in \mathbb{N}_0^2$ con fgp $g(u)$. Supongamos que se cumplen las hipótesis de los Teoremas 2.2.3 y 2.2.7.*

(a) Si $\eta(g; \theta) > 0$, entonces $P(\Psi^*_R = 1) \to 1$.

(b) Si $g(u) \in G_2$ y $\xi(g; \theta) > 0$, entonces $P(\Psi^*_S = 1) \to 1$.

Observación 2.3.3 *Notar que $\eta(g; \theta) \geq 0$ ($\xi(g; \theta) \geq 0$). Si $w > 0$ en casi todo $[0, 1]^2$, entonces $\eta(g; \theta) = 0$ ($\xi(g; \theta) = 0$) sí y sólo si H_0 es cierta.*

Por lo tanto, la consistencia de los tests propuestos está garantizada simplemente tomando una función de peso que sea positiva en casi todo $[0, 1]^2$.

2.3.2. Alternativas contiguas

El resultado de consistencia dado en el Corolario 2.3.2 no distingue entre medidas de probabilidad alternativas de P. Una mejor discriminación se obtiene considerando alternativas para las cuales la potencia tiende a valores menores o iguales que 1. Esto se logra reemplazando una alternativa fija por una sucesión de alternativas que converge a la nula a una cierta velocidad, que se suelen denominar alternativas contiguas.

Con este objetivo consideremos ahora un arreglo triangular $\boldsymbol{X}_{n,1}, \boldsymbol{X}_{n,2}, \ldots, \boldsymbol{X}_{n,n}$ de vectores aleatorios bivariantes independientes por filas que toman valores en $\mathbb{N}_0^2$ y función de probabilidad conjunta $P_n(x_1, x_2)$ dada por

$$P_n(x_1, x_2) = P_\theta(x_1, x_2) \left\{ 1 + \frac{1}{\sqrt{n}} b_n(x_1, x_2) \right\}, \tag{2.24}$$

donde $P_\theta(x_1, x_2)$ es la función de probabilidad del vector aleatorio bivariante que tiene una distribución $PB(\theta)$, para algún $\theta \in \Theta$, y $b_n(x_1, x_2)$ satisface las siguientes condiciones.

Supuesto 2.3.4 *(1)* $E_\theta\{b_n(X_1, X_2)\} = 0$, $\forall n$.

(2) $b_n(x_1, x_2) \to b(x_1, x_2)$, $\forall (x_1, x_2) \in \mathbb{N}_0^2$.

(3) $\sup_n E_\theta\big\{b_n(X_1, X_2)^4\big\} < \infty$.

El Supuesto 2.3.4(1) asegura que $\sum_{x_1, x_2 \in \mathbb{N}_0} P_n(x_1, x_2) = 1$; los Supuestos 2.3.4(2) y (3) contienen condiciones de tipo técnico que usaremos en las demostraciones.

El Teorema 2.2.3 establece que, cuando H_0 es cierta, $R_{n,w}(\hat{\theta}_n)$ converge en distribución a una combinación lineal de variables χ^2 independientes con 1 grado de libertad, donde los pesos son los autovalores del operador $C_R(\theta)$ dado en (2.8). Sea $\{\phi_j^R\}$ el conjunto de autofunciones ortonormales correspondiente a los autovalores $\{\lambda_j^R\}$ de $C_R(\theta)$.

Observar que lo expuesto anteriormente para $R_{n,w}(\hat{\theta}_n)$, también es válido para $S_{n,w}(\hat{\theta}_n)$, considerando el operador $C_S(\theta)$ con sus conjuntos $\{\phi_j^S\}$ y $\{\lambda_j^S\}$ de autofunciones ortonormales y de autovalores, respectivamente.

El siguiente Teorema da la ley límite de estos estadísticos bajo las alternativas $P_n(x_1, x_2)$ en (2.24).

Teorema 2.3.5 *Sea $\boldsymbol{X}_{n,1}, \boldsymbol{X}_{n,2}, \ldots, \boldsymbol{X}_{n,n}$ un arreglo triangular de vectores aleatorios bivariantes que son independientes por filas y que toman valores en $\mathbb{N}_0^2$, con función de probabilidad dada por $P_n(x_1, x_2)$ definida en (2.24). Supongamos que se cumplen los Supuestos 2.2.1 y 2.3.4. Entonces*

(a) $R_{n,w}(\hat{\theta}_n) \xrightarrow{L} \sum_{k=1}^{\infty} \lambda_k^R \big(Z_k + c_k^R\big)^2$, *donde* $c_k^R = \sum_{x_1, x_2} b(x_1, x_2)\, \phi_k^R(x_1, x_2)$ *y* $Z_1, Z_2, \ldots$ *son variables normales estándar independientes.*

(b) $S_{n,w}(\hat{\theta}_n) \xrightarrow{L} \sum_{k=1}^{\infty} \lambda_k^S \big(Z_k + c_k^S\big)^2$, *donde* $c_k^S = \sum_{x_1, x_2} b(x_1, x_2)\, \phi_k^S(x_1, x_2)$ *y* $Z_1, Z_2, \ldots$ *son variables normales estándar independientes.*

Demostración Sea $A_n \subseteq \mathbb{N}_0^2$ tal que $P_\theta(A_n) \to 0$, entonces

$$P_n(A_n) = P_\theta(A_n) + \frac{1}{\sqrt{n}} \sum_{(x_1,x_2)\in A_n} P_\theta(x_1,x_2) b_n(x_1,x_2).$$

De la desigualdad de Hölder y del Supuesto 2.3.4(3) se sigue que

$$\begin{aligned} &\sum_{(x_1,x_2)\in A_n} P_\theta(x_1,x_2) b_n(x_1,x_2) \\ &\qquad \le \left\{ \sum_{(x_1,x_2)\in A_n} P_\theta(x_1,x_2) \right\}^{3/4} \left\{ \sum_{(x_1,x_2)\in A_n} P_\theta(x_1,x_2) b_n(x_1,x_2)^4 \right\}^{1/4} \\ &\qquad \le \sup_n E_\theta \left\{ b_n(X_1,X_2)^4 \right\} < \infty, \end{aligned}$$

y por tanto $P_n(A_n) \to 0$, esto es, P_n es contigua a P_θ.

A continuación veremos que, bajo P_n, $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en ley a una distribución normal. Con este objetivo, sean $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid con función de probabilidad común P_θ y sea

$$l_n = \log \frac{P_n(\boldsymbol{X}_1) P_n(\boldsymbol{X}_2) \ldots P_n(\boldsymbol{X}_n)}{P_\theta(\boldsymbol{X}_1) P_\theta(\boldsymbol{X}_2) \ldots P_\theta(\boldsymbol{X}_n)}.$$

Sea $Z_{n,i} = b_n(\boldsymbol{X}_i)$, $1 \le i \le n$. Entonces, por desarrollo en serie de Taylor,

$$l_n = \sum_{i=1}^n \log \left(1 + \frac{1}{\sqrt{n}} Z_{n,i} \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{n,i} - \frac{1}{2n} \sum_{i=1}^n Z_{n,i}^2 + \varrho_n, \tag{2.25}$$

donde $\varrho_n = \varrho_n(Z_{n,1}, Z_{n,2}, \ldots, Z_{n,n})$,

$$\varrho_n = \sum_{i=1}^n \frac{2}{3!\, n\sqrt{n}\, a_{in}^3} Z_{n,i}^3,$$

con $a_{in} = a_{in}(Z_{n,i})$

$$a_{in} = 1 + \frac{\alpha_i}{\sqrt{n}} Z_{n,i},$$

para algún $0 < \alpha_i < 1$, $1 \le i \le n$. A continuación estudiamos el límite de cada uno de los sumandos en el lado derecho de (2.25).

Sea ε una constante positiva. Además, sea $M_n = \sqrt{n}/2$ y consideremos $\tilde{\varrho}_n = \varrho_n(Z_{n,1} I\{|Z_{n,1}| \le M_n\}, \ldots, Z_{n,n} I\{|Z_{n,n}| \le M_n\})$.

Notemos que $\tilde{a}_{in} = a_{in}(Z_{n,i}I\{|Z_{n,i}| \le M_n\}) \ge \frac{1}{2}, 1 \le i \le n$, y por lo tanto

$$E_\theta(|\tilde{\varrho}_n|) \le \frac{2^3}{3}\frac{1}{\sqrt{n}} \sup_m E_\theta\big\{|b_m(\boldsymbol{X}_1)|^3\big\}\,,$$

lo cual implica que

$$P_\theta(|\tilde{\varrho}_n| > \varepsilon) \to 0. \tag{2.26}$$

Como

$$\begin{aligned} P_\theta\left(|\tilde{\varrho}_n - \varrho_n| > \varepsilon\right) &\le P_\theta\left(\bigcup_{i=1}^n \{|Z_{n,i}| > M_n\}\right) \\ &\le \sum_{i=1}^n P_\theta\left(|Z_{n,i}| > M_n\right) \\ &\le \frac{n}{M_n^4} \sup_m E_\theta\big\{b_m(\boldsymbol{X}_1)^4\big\} \to 0. \end{aligned} \tag{2.27}$$

De (2.26) y (2.27), concluimos que

$$P_\theta\left(|\varrho_n| > \varepsilon\right) \to 0. \tag{2.28}$$

Por el Supuesto 2.3.4, $E_\theta(Z_{n,i}) = 0$, y

$$\sigma_n^2 = Var_\theta(Z_{n,i}) \to \sigma^2 = E_\theta\left\{b(X_1, X_2)^2\right\} < \infty.$$

Además,

$$\begin{aligned} E_\theta\left[Z_{n,i}^2\, I\{|Z_{n,i}| > na\}\right] &\le E_\theta^{1/2}\left(Z_{n,i}^4\right) P_\theta^{1/2}(|Z_{n,i}| > na) \\ &\le \frac{1}{(na)^4}\left[\sup_n E_\theta\left\{b_n(X_1, X_2)^4\right\}\right]^2 \to 0, \quad \forall a > 0. \end{aligned}$$

Por el teorema central del límite para arreglos triangulares (Teorema 1.9.3 en Serfling, 1980, pp. 31-32, [42]) se sigue que

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_{n,i} \xrightarrow{L} N(0, \sigma^2). \tag{2.29}$$

Se tiene que

$$E_\theta\left(\frac{1}{n}\sum_{i=1}^n Z_{n,i}^2\right) = \sigma_n^2 \to \sigma^2,$$

$$Var_\theta\left(\frac{1}{n}\sum_{i=1}^{n} Z_{n,i}^2\right) = \frac{Var_\theta(Z_{n,i}^2)}{n} \leq \frac{1}{n}\sup_n E_\theta\left\{b_n(X_1,X_2)^4\right\} \to 0,$$

de donde

$$\frac{1}{2n}\sum_{i=1}^{n} Z_{n,i}^2 \xrightarrow{P} \frac{1}{2}\sigma^2. \tag{2.30}$$

Ahora, del Supuesto 2.2.1, (2.25) y (2.28)–(2.30), la sucesión

$$\left(\sqrt{n}(\hat{\theta}_n - \theta), l_n\right)$$

converge en ley a una distribución normal multivariante de dimensión 4, cuando los datos provienen de P_θ. Así, por el tercer Lema de Le Cam (Corolario 12.3.2 en Lehmann y Romano (2005) [29]), concluimos que, cuando los datos provienen de P_n, $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en distribución a una ley normal, lo cual implica que es acotada en probabilidad.

Siguiendo pasos similares a los dados en la demostración del Teorema 2.2.3 (b), podemos probar que (2.7) también se cumple cuando los datos provienen de P_n, con $P_n(|s_n| > \varepsilon) \to 0$, $\forall\varepsilon > 0$.

Así, cuando los datos tienen la función de probabilidad P_n, aplicando el Teorema 2.3 en Gregory (1977) [10], obtenemos que

$$\|S_{1n}\|_{\mathcal{H}}^2 + \|S_{2n}\|_{\mathcal{H}}^2 \xrightarrow{L} \sum_{k=1}^{\infty}\lambda_k^S\left(Z_k + c_k^S\right)^2,$$

y con ello se consigue el resultado en (b). La demostración del resultado en (a) sigue los mismos pasos. □

Del Teorema 2.3.5, concluimos que el test Ψ_R^* (Ψ_S^*) es capaz de detectar alternativas como las establecidas en (2.24), que convergen a la DPB a una razón de $n^{-1/2}$.

Capítulo 3

Estadístico $W_n(\hat{\theta}_n)$

3.1. Definición del test estadístico

En este capítulo proponemos otro test para contrastar la hipótesis nula, cuyo test estadístico lo deduciremos de la Proposición 2.1.1. Para ello consideremos que se verifican las hipótesis en dicha proposición.

Sea $(X_1, X_2) \in \mathbb{N}_0^2$ un vector aleatorio y sea $g(u_1, u_2) = E\left(u_1^{X_1} u_2^{X_2}\right)$ su fgp, luego, por definición

$$g(u) = \sum_{r,s \geq 0} u_1^r u_2^s P(r, s),$$

donde $P(r, s) = P(X_1 = r, X_2 = s)$.

De la ecuación anterior y de las definiciones de $D_1(u; \theta)$ y $D_2(u; \theta)$ dadas en la Proposición 2.1.1, podemos escribir

$$D_1(u; \theta) = \sum_{r \geq 0} \sum_{s \geq 0} \{(r + 1)P(r + 1, s) - (\theta_1 - \theta_3)P(r, s) - \theta_3 P(r, s - 1)\}\, u_1^r u_2^s,$$

$$D_2(u; \theta) = \sum_{r \geq 0} \sum_{s \geq 0} \{(s + 1)P(r, s + 1) - (\theta_2 - \theta_3)P(r, s) - \theta_3 P(r - 1, s)\}\, u_1^r u_2^s.$$

Consideremos ahora las versiones empíricas de las ecuaciones anteriores, esto es, consideremos $D_{1n}(u; \hat{\theta}_n)$ y $D_{2n}(u; \hat{\theta}_n)$ definidas en (2.5). Si H_0 fuera cierta entonces $D_{1n}(u; \hat{\theta}_n)$ y $D_{2n}(u; \hat{\theta}_n)$ deberían ser próximas a 0, $\forall u \in [0, 1]^2$. Esta proximidad puede interpretarse de varias formas. En el capítulo anterior ya vimos que esto era equivalente a $\int \{D_{1n}(u; \hat{\theta}_n)^2 + D_{2n}(u; \hat{\theta}_n)^2\} w(u) du \approx 0$.

Veamos otra interpretación siguiendo un razonamiento similar al hecho en Nakamura y Pérez-Abreu (1993) [33] para el caso univariante, que presentamos al final de la Sección 1.5.1. Para ello observemos que

$$D_{1n}(u;\hat{\theta}_n) = \sum_{r\geq 0}\sum_{s\geq 0} d_1(r,s;\hat{\theta}_n)u_1^r u_2^s,$$

$$D_{2n}(u;\hat{\theta}_n) = \sum_{r\geq 0}\sum_{s\geq 0} d_2(r,s;\hat{\theta}_n)u_1^r u_2^s,$$

donde

$$d_1(r,s;\hat{\theta}_n) = (r+1)p_n(r+1,s) - (\hat{\theta}_{1n} - \hat{\theta}_{3n})p_n(r,s) - \hat{\theta}_{3n}p_n(r,s-1),$$

$$d_2(r,s;\hat{\theta}_n) = (s+1)p_n(r,s+1) - (\hat{\theta}_{2n} - \hat{\theta}_{3n})p_n(r,s) - \hat{\theta}_{3n}p_n(r-1,s),$$

y

$$p_n(r,s) = \frac{1}{n}\sum_{k=1}^{n} I\{X_{1k} = r, X_{2k} = s\}$$

es la frecuencia relativa empírica del par (r,s). Por tanto, $D_{in}(u;\hat{\theta}_n) = 0$, $\forall u \in [0,1]^2$, $i = 1,2$, sí y sólo si los coeficientes de $u_1^r u_2^s$ en las expansiones anteriores son nulos $\forall r,s \geq 0$. Esto nos lleva a considerar el siguiente estadístico para contrastar H_0:

$$W_n(\hat{\theta}_n) = \sum_{r\geq 0}\sum_{s\geq 0}\{d_1(r,s;\hat{\theta}_n)^2 + d_2(r,s;\hat{\theta}_n)^2\} = \sum_{r,s=0}^{M}\{d_1(r,s;\hat{\theta}_n)^2 + d_2(r,s;\hat{\theta}_n)^2\}, \quad (3.1)$$

donde $M = \max\{X_{1(n)}, X_{2(n)}\}$, $X_{k(n)} = \max_{1\leq i\leq n} X_{ki}$, $k = 1,2$.

Teniendo en cuenta que

$$d_k(r,s;\hat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n}\phi_{krs}(\boldsymbol{X}_i;\hat{\theta}_n), \quad k = 1,2,$$

con

$$\begin{aligned}\phi_{1rs}(x;\theta) &= (r+1)I\{x_1 = r+1, x_2 = s\} - (\theta_1 - \theta_3)I\{x_1 = r, x_2 = s\} \\ &\qquad -\theta_3 I\{x_1 = r, x_2 = s-1\}, \\ \phi_{2rs}(x;\theta) &= (s+1)I\{x_1 = r, x_2 = s+1\} - (\theta_2 - \theta_3)I\{x_1 = r, x_2 = s\} \\ &\qquad -\theta_3 I\{x_1 = r-1, x_2 = s\},\end{aligned} \quad (3.2)$$

donde $x = (x_1, x_2)$, entonces el estadístico $W_n(\hat{\theta}_n)$ puede ser expresado como sigue:

$$W_n(\hat{\theta}_n) = \frac{1}{n^2} \sum_{i,j=1}^{n} h(\boldsymbol{X}_i, \boldsymbol{X}_j; \hat{\theta}_n),$$

donde

$$\begin{aligned} h(x, y; \theta) &= h_1(x, y; \theta) + h_2(x, y; \theta), \\ h_k(x, y; \theta) &= \sum_{r \geq 0} \sum_{s \geq 0} \phi_{krs}(x; \theta) \phi_{krs}(y; \theta), \quad k = 1, 2, \end{aligned} \tag{3.3}$$

con $x = (x_1, x_2)$ e $y = (y_1, y_2)$.

Un test razonable para contrastar H_0 debería rechazar la hipótesis nula para valores grandes de $W_n(\hat{\theta}_n)$. Ahora, para determinar qué son los valores grandes, debemos calcular la distribución nula de $W_n(\hat{\theta}_n)$ o al menos una aproximación de ella.

Como la distribución nula de $W_n(\hat{\theta}_n)$ es desconocida, trataremos de estimarla empleando el modo clásico, esto es, aproximaremos la distribución nula mediante la distribución asintótica nula. En la siguiente sección estudiaremos esta situación.

3.2. Aproximación de la distribución nula de $\boldsymbol{W_n(\hat{\theta}_n)}$

3.2.1. Distribución asintótica nula

El siguiente resultado proporciona la distribución asintótica nula de $W_n(\hat{\theta}_n)$.

Teorema 3.2.1 *Sean $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ v.a. iid de $\boldsymbol{X} = (X_1, X_2) \sim PB(\theta_1, \theta_2, \theta_3)$. Supongamos que se cumple el Supuesto 2.2.1. Entonces*

$$nW_n(\hat{\theta}_n) = \frac{1}{n} \sum_{i,j=1}^{n} h_d(\boldsymbol{X}_i, \boldsymbol{X}_j; \theta) + \rho_n,$$

donde $P_\theta(|\rho_n| > \varepsilon) \to 0$, $\forall \varepsilon > 0$,

$$h_d(x, y; \theta) = h(x, y; \theta) + \boldsymbol{\ell}(x; \theta) S \boldsymbol{\ell}(y; \theta)^\top, \tag{3.4}$$

con $h(x, y; \theta)$ definido en (3.3), $x = (x_1, x_2)$, $y = (y_1, y_2)$, $S = \sum_{r,s \geq 0} A_{rs}$ y A_{rs} es la *matriz simétrica dada por*

$$A_{rs} = \begin{pmatrix} a^2 & 0 & a(b-a) \\ 0 & a^2 & a(c-a) \\ a(b-a) & a(c-a) & (b-a)^2+(c-a)^2 \end{pmatrix}, \tag{3.5}$$

con $a = P(r, s; \theta)$, $b = P(r, s-1; \theta)$, $c = P(r-1, s; \theta)$. *Además*

$$nW_n(\hat{\theta}_n) \xrightarrow{L} \sum_{j \geq 1} \lambda_j \chi^2_{1j},$$

donde $\chi^2_{11}, \chi^2_{12}, \ldots$ *son v.a. independientes* χ^2 *con 1 grado de libertad y el conjunto* $\{\lambda_j\}$ *son los autovalores no nulos del operador* $C(\theta)$ *definido sobre el espacio de funciones* $\{\tau : \mathbb{N}_0^2 \to \mathbb{R}, \textit{tal que } E_\theta\{\tau^2(\boldsymbol{X})\} < \infty, \forall \theta \in \Theta\}$, *como sigue*

$$C(\theta)\tau(x) = E_\theta\{h_d(x, \boldsymbol{Y}; \theta)\tau(\boldsymbol{Y})\}. \tag{3.6}$$

Demostración Por definición

$$W_n(\hat{\theta}_n) = \frac{1}{n^2} \sum_{i,j=1}^{n} h(\boldsymbol{X}_i, \boldsymbol{X}_j; \hat{\theta}_n).$$

Por desarrollo en serie de Taylor

$$\begin{aligned} &W_n(\hat{\theta}_n) \\ &= \frac{1}{n^2} \sum_{i,j=1}^{n} \left\{ h(\boldsymbol{X}_i, \boldsymbol{X}_j; \theta) + Q^{(1)}(\boldsymbol{X}_i, \boldsymbol{X}_j; \theta)(\hat{\theta}_n - \theta)^\top + \frac{1}{2}(\hat{\theta}_n - \theta) Q^{(2)}(\boldsymbol{X}_i, \boldsymbol{X}_j; \theta)(\hat{\theta}_n - \theta)^\top \right\} \\ &\quad + \frac{1}{3!\, n^2} \sum_{i,j=1}^{n} \sum_{k_1,k_2,k_3=1}^{3} \left. \frac{\partial^3}{\partial\vartheta_{k_1} \partial\vartheta_{k_2} \partial\vartheta_{k_3}} h(\boldsymbol{X}_i, \boldsymbol{X}_j; \vartheta) \right|_{\vartheta = \tilde{\theta}} (\hat{\theta}_{k_1 n} - \theta_{k_1})(\hat{\theta}_{k_2 n} - \theta_{k_2})(\hat{\theta}_{k_3 n} - \theta_{k_3}), \end{aligned} \tag{3.7}$$

donde $Q^{(k)}(x, y; \theta)$ representan las $k-$ésimas derivadas de $h(x, y; \vartheta)$ respecto de ϑ evaluadas en θ, $k = 1, 2$, y $\tilde{\theta} = \alpha\hat{\theta}_n + (1-\alpha)\theta$, para algún $0 < \alpha < 1$.

Las derivadas de $h(x, y; \theta)$ están dadas por

$$\frac{\partial}{\partial \vartheta_1} h(x, y; \theta) = -\sum_{r,s \geq 0} \Big[I\{x_1 = r, x_2 = s\} \phi_{1rs}(y; \theta) + I\{y_1 = r, y_2 = s\} \phi_{1rs}(x; \theta) \Big],$$

$$\frac{\partial}{\partial \vartheta_2} h(x, y; \theta) = -\sum_{r,s \geq 0} \Big[I\{x_1 = r, x_2 = s\} \phi_{2rs}(y; \theta) + I\{y_1 = r, y_2 = s\} \phi_{2rs}(x; \theta) \Big],$$

$$\begin{aligned} \frac{\partial}{\partial \vartheta_3} h(x, y; \theta) = & \sum_{r,s \geq 0} \Big(I\{x_1 = r, x_2 = s\} - I\{x_1 = r, x_2 = s-1\} \Big) \phi_{1rs}(y; \theta) \\ & + \sum_{r,s \geq 0} \Big(I\{y_1 = r, y_2 = s\} - I\{y_1 = r, y_2 = s-1\} \Big) \phi_{1rs}(x; \theta) \\ & + \sum_{r,s \geq 0} \Big(I\{x_1 = r, x_2 = s\} - I\{x_1 = r-1, x_2 = s\} \Big) \phi_{2rs}(y; \theta) \\ & + \sum_{r,s \geq 0} \Big(I\{y_1 = r, y_2 = s\} - I\{y_1 = r-1, y_2 = s\} \Big) \phi_{2rs}(x; \theta), \end{aligned}$$

$$\frac{\partial^2}{\partial \vartheta_1^2} h(x, y; \theta) = 2 \sum_{r,s \geq 0} I\{x_1 = r, x_2 = s\} I\{y_1 = r, y_2 = s\},$$

$$\frac{\partial^2}{\partial \vartheta_1 \partial \vartheta_2} h(x, y; \theta) = 0,$$

$$\begin{aligned} \frac{\partial^2}{\partial \vartheta_1 \partial \vartheta_3} h(x, y; \theta) = & -\sum_{r,s \geq 0} \Big[I\{x_1 = r, x_2 = s\} \big(I\{y_1 = r, y_2 = s\} - I\{y_1 = r, y_2 = s-1\} \big) \\ & + I\{y_1 = r, y_2 = s\} \big(I\{x_1 = r, x_2 = s\} - I\{x_1 = r, x_2 = s-1\} \big) \Big], \end{aligned}$$

$$\frac{\partial^2}{\partial \vartheta_2^2} h(x, y; \theta) = 2 \sum_{r,s \geq 0} I\{x_1 = r, x_2 = s\} I\{y_1 = r, y_2 = s\},$$

$$\begin{aligned} \frac{\partial^2}{\partial \vartheta_2 \partial \vartheta_3} h(x, y; \theta) = & -\sum_{r,s \geq 0} \Big[I\{x_1 = r, x_2 = s\} (I\{y_1 = r, y_2 = s\} - I\{y_1 = r-1, y_2 = s\}) \\ & + I\{y_1 = r, y_2 = s\} \big(I\{x_1 = r, x_2 = s\} - I\{x_1 = r-1, x_2 = s\} \big) \Big], \end{aligned}$$

$$\begin{aligned} \frac{\partial^2}{\partial \vartheta_3^2} h(x, y; \theta) = & \, 2 \sum_{r,s \geq 0} \Big[\big(I\{x_1 = r, x_2 = s\} - I\{x_1 = r, x_2 = s-1\} \big) \\ & \times \big(I\{y_1 = r, y_2 = s\} - I\{y_1 = r, y_2 = s-1\} \big) \Big] \\ & + 2 \sum_{r,s \geq 0} \Big[\big(I\{x_1 = r, x_2 = s\} - I\{x_1 = r-1, x_2 = s\} \big) \\ & \times \big(I\{y_1 = r, y_2 = s\} - I\{y_1 = r-1, y_2 = s\} \big) \Big]. \end{aligned}$$

Además,
$$\frac{\partial^3}{\partial\vartheta_i\partial\vartheta_j\partial\vartheta_k}h(x,y;\theta)=0,\ 1\le i,j,k\le 3.$$

De lo anterior se obtiene que

$$\begin{aligned}
E_\theta\left\{Q^{(1)}(\boldsymbol{X}_1,\boldsymbol{X}_1;\theta)\right\} &= 2(\theta_1-\theta_3,\theta_2-\theta_3,4\theta_3-\theta_1-\theta_2)<\infty,\\
E_\theta\left\{Q^{(1)}(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)\right\} &= \mathbf{0},\ \text{(vector nulo de } \mathbb{R}^3),\\
E_\theta\left\{Q^{(2)}(\boldsymbol{X}_1,\boldsymbol{X}_1;\theta)\right\} &= -2\begin{pmatrix} -1 & 0 & 1\\ 0 & -1 & 1\\ 1 & 1 & -4\end{pmatrix}<\infty,\\
E_\theta\left\{Q^{(2)}(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)\right\} &= 2\textstyle\sum_{r,s\ge 0}A_{rs}<\infty,
\end{aligned}$$

donde A_{rs} es la matriz definida en (3.5).

Puesto que

$$\frac{1}{n^2}\sum_{i,j=1}^{n}Q^{(k)}(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta)=\frac{1}{n^2}\sum_{i\ne j}Q^{(k)}(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta)+\frac{1}{n^2}\sum_{i=1}^{n}Q^{(k)}(\boldsymbol{X}_i,\boldsymbol{X}_i;\theta),\ k=1,2,$$

y por la ley fuerte de los grandes números para U-estadísticos (véase por ejemplo Teorema 5.4 en Serfling (1980) [42]),

$$\frac{1}{n^2}\sum_{i\ne j}Q^{(k)}(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta)\xrightarrow{c.s.}E_\theta\left\{Q^{(k)}(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)\right\}<\infty,\ k=1,2,$$

y por la ley fuerte de los grandes números,

$$\frac{1}{n}\sum_{i=1}^{n}Q^{(k)}(\boldsymbol{X}_i,\boldsymbol{X}_i;\theta)\xrightarrow{c.s.}E_\theta\left\{Q^{(k)}(\boldsymbol{X}_1,\boldsymbol{X}_1;\theta)\right\}<\infty,\ k=1,2.$$

Así, (3.7) se puede escribir como

$$W_n(\hat\theta_n)=\frac{1}{n^2}\sum_{i,j=1}^{n}h(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta)+\sqrt{n}(\hat\theta_n-\theta)\frac{1}{n}S\sqrt{n}(\hat\theta_n-\theta)^\top+o_P(1),$$

donde S es la matriz simétrica dada por $S=\frac{1}{2}E_\theta\big\{Q^{(2)}(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)\big\}$.

Considerando ahora el Supuesto 2.2.1, se obtiene

$$W_n(\hat\theta_n)=\frac{1}{n^2}\sum_{i,j=1}^{n}h_d(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta)+\varepsilon_n,$$

donde $h_d(x,y;\theta)$ es el definido en (3.4) y $\varepsilon_n = o_P(n^{-1})$.

Notar que $h_d(x,y;\theta) = h_d(y,x;\theta)$, pues $\left\{\boldsymbol{\ell}(x;\theta)S\boldsymbol{\ell}(y;\theta)^\top\right\}^\top = \boldsymbol{\ell}(y;\theta)S\boldsymbol{\ell}(x;\theta)^\top \in \mathbb{R}$, y bajo H_0, resulta

$$E_\theta\{|h_d(\boldsymbol{X}_1,\boldsymbol{X}_1;\theta)|\} < \infty \quad \text{y} \quad E_\theta\left\{h_d(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)^2\right\} < \infty.$$

Además, por el Supuesto 2.2.1 y el hecho que $E_\theta\{h(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)\} = 0$, dan como resultado $E_\theta\{h_d(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)\} = 0$.

Por último, como h_d es degenerado, $E_\theta\{h_d(\boldsymbol{X}_1,\boldsymbol{X}_2;\theta)/\boldsymbol{X}_1\} = 0$, entonces, por el Teorema 6.4.1.B en Serfling (1980) [42],

$$\frac{1}{n}\sum_{i,j=1}^{n} h_d(\boldsymbol{X}_i,\boldsymbol{X}_j;\theta) \xrightarrow{L} \sum_{j\geq 1} \lambda_j\, \chi^2_{1j},$$

con lo cual se consigue el resultado. □

Los mismos comentarios hechos al final de la Subsección 2.2.1 para $R_{n,w}(\hat{\theta}_n)$ y $S_{n,w}(\hat{\theta}_n)$ se pueden hacer para $W_n(\hat{\theta}_n)$, es decir, la distribución asintótica nula no proporciona una estimación útil a la distribución nula de $W_n(\hat{\theta}_n)$. Así, en la siguiente sección consideramos otra forma de aproximar la distribución nula del test estadístico, el método bootstrap.

3.2.2. Estimador bootstrap de la distribución nula de $\boldsymbol{W_n(\hat{\theta}_n)}$

Como mencionamos en la Sección 2.2.2, un modo alternativo de aproximar la distribución nula es mediante el método bootstrap. Para ello, sean $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n$ vectores aleatorios iid que toman valores en $\mathbb{N}_0^2$ tales que $\hat{\theta}_n = \hat{\theta}_n(\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n) \in \Theta$.

Sean $\boldsymbol{X}_1^*,\boldsymbol{X}_2^*,\ldots,\boldsymbol{X}_n^*$ v.a. iid de una población que se distribuye según la ley $PB(\hat{\theta}_{1n},\hat{\theta}_{2n},\hat{\theta}_{3n})$, dado $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n$ y sea $W_n^*(\hat{\theta}_n^*)$ la versión bootstrap de $W_n(\hat{\theta}_n)$ obtenida al reemplazar $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n$ y $\hat{\theta}_n = \hat{\theta}_n(\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n)$ por $\boldsymbol{X}_1^*,\boldsymbol{X}_2^*,\ldots,\boldsymbol{X}_n^*$ y $\hat{\theta}_n^* = \hat{\theta}_n(\boldsymbol{X}_1^*,\boldsymbol{X}_2^*,\ldots,\boldsymbol{X}_n^*)$, respectivamente, en la expresión de $W_n(\hat{\theta}_n)$.

Como en el Capítulo 2, sea P_* la ley de probabilidad condicional bootstrap, dado $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n$.

Para mostrar que el método bootstrap aproxima consistentemente a la distribución nula de $W_n(\hat{\theta}_n)$ nos serán útiles las siguientes expresiones.

Observar que del segundo sumando de (3.4) podemos escribir que

$$
\begin{aligned}
\boldsymbol{\ell}(x;\theta)A_{rs}\,\boldsymbol{\ell}(y;\theta)^\top &= \boldsymbol{\ell}(x;\theta)\begin{pmatrix} a & 0 & 0\\ 0 & a & 0\\ b-a & c-a & 0\end{pmatrix}\begin{pmatrix} a & 0 & b-a\\ 0 & a & c-a\\ 0 & 0 & 0\end{pmatrix}\boldsymbol{\ell}(y;\theta)^\top\\
&= \left(\boldsymbol{\ell}(x;\theta)(a,0,b-a)^\top, \boldsymbol{\ell}(x;\theta)(0,a,c-a)^\top\right)\begin{pmatrix}(a,0,b-a)\boldsymbol{\ell}(y;\theta)^\top\\ (0,a,c-a)\boldsymbol{\ell}(y;\theta)^\top\end{pmatrix}\\
&= \boldsymbol{\ell}(x;\theta)\,(a,0,b-a)^\top\boldsymbol{\ell}(y;\theta)\,(a,0,b-a)^\top\\
&\quad + \boldsymbol{\ell}(x;\theta)\,(0,a,c-a)^\top\boldsymbol{\ell}(y;\theta)\,(0,a,c-a)^\top.
\end{aligned}
$$

Sean

$$\phi_{3rs}(x;\theta) = \boldsymbol{\ell}(x;\theta)(a,0,b-a)^\top \quad \text{y} \quad \phi_{4rs}(x;\theta) = \boldsymbol{\ell}(x;\theta)(0,a,c-a)^\top, \tag{3.8}$$

donde $x = (x_1, x_2)$.

Con esta notación,

$$\boldsymbol{\ell}(x;\theta)A_{rs}\,\boldsymbol{\ell}(y;\theta)^\top = \phi_{3rs}(x;\theta)\phi_{3rs}(y;\theta) + \phi_{4rs}(x;\theta)\phi_{4rs}(y;\theta). \tag{3.9}$$

También nos será de utilidad el siguiente resultado.

Proposición 3.2.2 *Sean $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid de $\boldsymbol{X} = (X_1, X_2) \in \mathbb{N}_0^2$. Supongamos que se verifica el Supuesto 2.2.4 y que $\hat{\theta}_n \xrightarrow{c.s.} \theta$, para algún $\theta \in \Theta$. Entonces*

$$\sup_{p,q,u,v\in\mathbb{N}_0} |K_n(p,q,u,v) - K(p,q,u,v)| \xrightarrow{c.s.} 0,$$

donde

$$K_n(p,q,u,v) = \Big(E_*\{\phi_{ipq}(\boldsymbol{X}_1^*;\hat{\theta}_n)\phi_{juv}(\boldsymbol{X}_1^*;\hat{\theta}_n)\}\Big), 1 \le i,j \le 4.$$

$$K(p,q,u,v) = \Big(E_\theta\{\phi_{ipq}(\boldsymbol{X}_1;\theta)\phi_{juv}(\boldsymbol{X}_1;\theta)\}\Big), 1 \le i,j \le 4.$$

con $\phi_{1rs}(\boldsymbol{X}_1;\vartheta)$, $\phi_{2rs}(\boldsymbol{X}_1;\vartheta)$ definidos en (3.2), y $\phi_{3rs}(\boldsymbol{X}_1;\vartheta)$ y $\phi_{4rs}(\boldsymbol{X}_1;\vartheta)$ definidos en (3.8).

Demostración El Supuesto 2.2.4 y el hecho de que $\hat{\theta}_n \xrightarrow{c.s.} \theta$ implican que

$$\sup_{p,q,u,v\in\mathbb{N}_0}\left|E_*\{\phi_{ipq}(\boldsymbol{X}_1^*;\hat{\theta}_n)\phi_{juv}(\boldsymbol{X}_1^*;\hat{\theta}_n)\} - E_\theta\{\phi_{ipq}(\boldsymbol{X}_1;\theta)\phi_{juv}(\boldsymbol{X}_1;\theta)\}\right| \xrightarrow{c.s.} 0, \tag{3.10}$$

donde $1 \leq i, j \leq 4$.

Para probar este resultado consideremos los siguientes casos:

(a) i, j tales que $1 \leq i, j \leq 2$,

(b) i, j tales que $3 \leq i, j \leq 4$ y

(c) i, j tales que $i = 1, 2$, $j = 3, 4$.

Notar primero que por el Teorema del Valor Medio

$P(r, s; \hat{\theta}_n) - P(r, s; \theta) = \nabla_\vartheta P(r, s; \tilde{\theta})(\hat{\theta}_n - \theta)^\top$, $\tilde{\theta} = \alpha\hat{\theta}_n + (1-\alpha)\theta$, para algún $0 < \alpha < 1$,

donde $\nabla_\vartheta P(r, s; \tilde{\theta})$ denota el vector gradiente de $P(r, s; \vartheta)$ evaluado en $\tilde{\theta}$. Por las relaciones de recurrencia dadas en (1.8), obtenemos que

$$\nabla_\vartheta P(r, s; \tilde{\theta}) = (a - c, b - c, d - a - b + c),$$

donde $a = P(r-1, s; \tilde{\theta}), b = P(r, s-1; \tilde{\theta}), c = P(r, s; \tilde{\theta})$ y $d = P(r-1, s-1; \tilde{\theta})$.

Como $0 \leq P(r, s; \tilde{\theta}) \leq 1$ y puesto que $\hat{\theta}_n \xrightarrow{c.s.} \theta$, resulta

$$\sup_{r,s \in \mathbb{N}_0} \left| P(r, s; \hat{\theta}_n) - P(r, s; \theta) \right| = o(1). \tag{3.11}$$

- En el caso (a), una situación es la siguiente

$$E_*\{\phi_{1pq}(\boldsymbol{X}_1^*; \hat{\theta}_n)\phi_{1uv}(\boldsymbol{X}_1^*; \hat{\theta}_n)\}$$
$$= \begin{cases} \begin{array}{r} (p+1)^2 P(p+1, q; \hat{\theta}_n) + \hat{\theta}_{3n}^2 P(p, q-1; \hat{\theta}_n) \\ +(\hat{\theta}_{1n} - \hat{\theta}_{3n})^2 P(p, q; \hat{\theta}_n) \end{array} & \text{si } u = p, v = q, \\ -(p+1)(\hat{\theta}_{1n} - \hat{\theta}_{3n}) P(p+1, q; \hat{\theta}_n), & \text{si } u = p+1, v = q, \\ -(p+1)\hat{\theta}_{3n} P(p+1, q; \hat{\theta}_n), & \text{si } u = p+1, v = q+1, \\ -p(\hat{\theta}_{1n} - \hat{\theta}_{3n}) P(p, q; \hat{\theta}_n), & \text{si } u = p-1, v = q, \\ -p\hat{\theta}_{3n} P(p, q-1; \hat{\theta}_n), & \text{si } u = p-1, v = q-1, \\ \hat{\theta}_{3n}(\hat{\theta}_{1n} - \hat{\theta}_{3n}) P(p, q; \hat{\theta}_n), & \text{si } u = p, v = q+1, \\ \hat{\theta}_{3n}(\hat{\theta}_{1n} - \hat{\theta}_{3n}) P(p, q-1; \hat{\theta}_n), & \text{si } u = p, v = q-1, \end{cases}$$

y análogo para $E_\theta\left\{\phi_{1pq}(\boldsymbol{X}_1;\theta)\phi_{1uv}(\boldsymbol{X}_1;\theta)\right\}$, con las modificaciones obvias.

Al hacer la resta de las esperanzas anteriores nos encontramos con las siguientes tres situaciones diferentes

(a.1) $t(\hat{\theta}_n)P(r,s;\hat{\theta}_n)-t(\theta)P(r,s;\theta)$,

(a.2) $r\Big\{t(\hat{\theta}_n)P(r,s;\hat{\theta}_n)-t(\theta)P(r,s;\theta)\Big\}$,

(a.3) $r^2\Big\{P(r,s;\hat{\theta}_n)-P(r,s;\theta)\Big\}$,

donde $t(\vartheta)=\vartheta_3^k$, $t(\vartheta)=(\vartheta_1-\vartheta_3)^k$, para $k=1,2$, o bien $t(\vartheta)=\vartheta_3(\vartheta_1-\vartheta_3)$.

Para la situación (a.1), puesto que $t(\vartheta)$ es continua como función de $\vartheta\in\Theta_0$ y como $\hat{\theta}_n\xrightarrow{c.s.}\theta$, entonces $t(\hat{\theta}_n)=t(\theta)+o(1)$, por lo tanto

$$t(\hat{\theta}_n)P(r,s;\hat{\theta}_n)-t(\theta)P(r,s;\theta)=t(\theta)\{P(r,s;\hat{\theta}_n)-P(r,s;\theta)\}+o(1)P(r,s;\hat{\theta}_n), \tag{3.12}$$

como $t(\theta)<\infty$, de (3.11) resulta

$$\sup_{r,s\in\mathbb{N}_0}\Big|t(\hat{\theta}_n)P(r,s;\hat{\theta}_n)-t(\theta)P(r,s;\theta)\Big|=o(1).$$

Para la situación (a.3), por desarrollo en serie de Taylor

$$r^2\Big\{P(r,s;\hat{\theta}_n)-P(r,s;\theta)\Big\}=\sum_{i=1}^{3}r^2\frac{\partial}{\partial\theta_i}P(r,s;\tilde{\theta})(\hat{\theta}_{in}-\theta_i),$$

donde $\tilde{\theta}=\alpha\hat{\theta}_n+(1-\alpha)\theta$, para algún $0<\alpha<1$.

Veamos el primer sumando y análogo el resto:

Como $\frac{\partial}{\partial\theta_1}P(r,s;\tilde{\theta})=P(r-1,s;\tilde{\theta})-P(r,s;\tilde{\theta})$ y $E_\vartheta(X_1^2)<\infty,\forall\vartheta\in\Theta_0$, entonces

$$\begin{aligned}r^2\frac{\partial}{\partial\theta_1}P(r,s;\tilde{\theta})(\hat{\theta}_{1n}-\theta_1)&=(\hat{\theta}_{1n}-\theta_1)r^2\{P(r-1,s;\tilde{\theta})-P(r,s;\tilde{\theta})\}\\&\le|\hat{\theta}_{1n}-\theta_1|\sum_{r,s\ge0}r^2\{P(r-1,s;\tilde{\theta})+P(r,s;\tilde{\theta})\}\\&\le|\hat{\theta}_{1n}-\theta_1|\left[E_{\tilde{\theta}}\{(X_1+1)^2\}+E_{\tilde{\theta}}(X_1^2)\right]=o(1).\end{aligned}$$

De donde

$$r^2\Big\{P(r,s;\hat\theta_n)-P(r,s;\theta)\Big\}$$
$$\leq\Big\{|\hat\theta_{1n}-\theta_1|+2|\hat\theta_{3n}-\theta_3|\Big\}\left[E_{\tilde\theta}\left\{(X_1+1)^2\right\}+E_{\tilde\theta}\big(X_1^2\big)\right]+2|\hat\theta_{2n}-\theta_2|E_{\tilde\theta}\big(X_1^2\big).$$

Por lo tanto

$$\sup_{r,s\in\mathbb{N}_0}\left|r^2\Big\{P(r,s;\hat\theta_n)-P(r,s;\theta)\Big\}\right|=o(1).$$

Para la situación (a.2) y considerando (3.12), obtenemos

$$r\left\{t(\hat\theta_n)P(r,s;\hat\theta_n)-t(\theta)P(r,s;\theta)\right\}$$
$$=r\,t(\theta)\left\{P(r,s;\hat\theta_n)-P(r,s;\theta)\right\}+rP(r,s;\hat\theta_n)\,o(1).$$

Puesto que $t(\theta)<\infty$, $E_\vartheta(X_1)<\infty, \forall\vartheta\in\Theta_0$ y siguiendo pasos similares a los dados en la situación (a.3) anterior se consigue que

$$\sup_{r,s\in\mathbb{N}_0}\left|r\left\{t(\hat\theta_n)P(r,s;\hat\theta_n)-t(\theta)P(r,s;\theta)\right\}\right|=o(1).$$

- Para el caso (b), tenemos, por ejemplo

$$\begin{aligned}
E_*\Big\{\phi_{3pq}(\boldsymbol{X}_1^*;\hat\theta_n)\phi_{3uv}(\boldsymbol{X}_1^*;\hat\theta_n)\Big\}&=E_*\Big\{\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat\theta_n)B(p,q;\hat\theta_n)^\top\boldsymbol{\ell}(\boldsymbol{X}_1^*;\hat\theta_n)B(u,v;\hat\theta_n)^\top\Big\}\\
&=B(p,q;\hat\theta_n)J(\hat\theta_n)B(u,v;\hat\theta_n)^\top,\\
E_\theta\{\phi_{3pq}(\boldsymbol{X}_1;\theta)\phi_{3uv}(\boldsymbol{X}_1;\theta)\}&=B(p,q;\theta)J(\theta)B(u,v;\theta)^\top,
\end{aligned}$$

donde

$$B(r,s;\vartheta)=(P(r,s;\vartheta),0,P(r,s-1;\vartheta)-P(r,s;\vartheta)). \tag{3.13}$$

Como, por el Supuesto 2.2.4, $J(\vartheta)$ es continua como función de $\vartheta\in\Theta_0$ y $\hat\theta_n\xrightarrow{c.s.}\theta$, entonces $J(\hat\theta_n)=J(\theta)+o(1)$. Además, por el Teorema del Valor Medio

$$B(r,s;\hat\theta_n)^\top-B(r,s;\theta)^\top=grad(P)(\hat\theta_n-\theta)^\top=\mathbf{o}(1), \tag{3.14}$$

pues

$$grad(P)=\begin{pmatrix}\nabla_\theta P(r,s;\tilde\theta)\\ \mathbf{0}\\ \nabla_\theta P(r,s-1;\tilde\theta)-\nabla_\theta P(r,s;\tilde\theta)\end{pmatrix}<\infty,$$

donde $\tilde{\theta} = \alpha\hat{\theta}_n + (1-\alpha)\theta$, para algún $0 < \alpha < 1$.

Por tanto

$$\begin{aligned}
&\left| E_* \Big\{ \phi_{3pq}(\boldsymbol{X}_1^*; \hat{\theta}_n) \phi_{3uv}(\boldsymbol{X}_1^*; \hat{\theta}_n) \Big\} - E_\theta \{ \phi_{3pq}(\boldsymbol{X}_1; \theta) \phi_{3uv}(\boldsymbol{X}_1; \theta) \} \right| \\
&\qquad\qquad\qquad \leq \left| \Big\{ B(p,q;\hat{\theta}_n) - B(p,q;\theta) \Big\} J(\theta) B(u,v;\hat{\theta}_n)^\top \right| \\
&\qquad\qquad\qquad\quad + \left| B(p,q;\theta) J(\theta) \Big\{ B(u,v;\hat{\theta}_n)^\top - B(u,v;\theta)^\top \Big\} \right| + o(1).
\end{aligned}$$

Puesto que $J(\vartheta) < \infty$ y $B(r,s;\vartheta) < \infty$, $\forall \vartheta \in \Theta_0, \forall r,s \in \mathbb{N}_0$, entonces se logra

$$\sup_{p,q,u,v \in \mathbb{N}_0} \left| E_* \Big\{ \phi_{3pq}(\boldsymbol{X}_1^*; \hat{\theta}_n) \phi_{3uv}(\boldsymbol{X}_1^*; \hat{\theta}_n) \Big\} - E_\theta \{ \phi_{3pq}(\boldsymbol{X}_1; \theta) \phi_{3uv}(\boldsymbol{X}_1; \theta) \} \right| \xrightarrow{c.s.} 0.$$

Aplicando el mismo procedimiento se demuestra (3.10) para $3 \leq i,j \leq 4$.

- En el caso (c), una situación es

$$E_* \Big\{ \phi_{1pq}(\boldsymbol{X}_1^*; \hat{\theta}_n) \phi_{3uv}(\boldsymbol{X}_1^*; \hat{\theta}_n) \Big\} - E_\theta \{ \phi_{1pq}(\boldsymbol{X}_1; \theta) \phi_{3uv}(\boldsymbol{X}_1; \theta) \} = A_1 + A_2 + A_3,$$

donde

$$\begin{aligned}
A_1 &= (p+1) P(p+1,q;\hat{\theta}_n) \boldsymbol{\ell}(p+1,q;\hat{\theta}_n) B(u,v;\hat{\theta}_n)^\top \\
&\qquad\qquad\qquad - (p+1) P(p+1,q;\theta) \boldsymbol{\ell}(p+1,q;\theta) B(u,v;\theta)^\top, \\
A_2 &= (\theta_1 - \theta_3) P(p,q;\theta) \boldsymbol{\ell}(p,q;\theta) B(u,v;\theta)^\top \\
&\qquad\qquad\qquad - (\hat{\theta}_{1n} - \hat{\theta}_{3n}) P(p,q;\hat{\theta}_n) \boldsymbol{\ell}(p,q;\hat{\theta}_n) B(u,v;\hat{\theta}_n)^\top, \\
A_3 &= \theta_3 P(p,q-1;\theta) \boldsymbol{\ell}(p,q-1;\theta) B(u,v;\theta)^\top \\
&\qquad\qquad\qquad - \hat{\theta}_{3n} P(p,q-1;\hat{\theta}_n) \boldsymbol{\ell}(p,q-1;\hat{\theta}_n) B(u,v;\hat{\theta}_n)^\top
\end{aligned}$$

y B es definido como en (3.13).

Veamos primero que $\sup_{r,s,u,v \in \mathbb{N}_0} |A_3| = o(1)$. De manera análoga se demuestra que $\sup_{r,s,u,v \in \mathbb{N}_0} |A_2| = o(1)$.

$$
\begin{aligned}
\hat{\theta}_{3n}P(r,s;\hat{\theta}_n)\boldsymbol{\ell}(r,s;\hat{\theta}_n)B(u,v;\hat{\theta}_n)^\top - \theta_3 P(r,s;\theta)\boldsymbol{\ell}(r,s;\theta)B(u,v;\theta)^\top \\
= (\hat{\theta}_{3n}-\theta_3)P(r,s;\hat{\theta}_n)\boldsymbol{\ell}(r,s;\hat{\theta}_n)B(u,v;\hat{\theta}_n)^\top \\
+ \theta_3 P(r,s;\theta)\boldsymbol{\ell}(r,s;\theta)\left\{B(u,v;\hat{\theta}_n)^\top - B(u,v;\theta)^\top\right\} \\
+ \theta_3\left\{P(r,s;\hat{\theta}_n)\boldsymbol{\ell}(r,s;\hat{\theta}_n) - P(r,s;\theta)\boldsymbol{\ell}(r,s;\theta)\right\}B(u,v;\hat{\theta}_n)^\top.
\end{aligned}
\tag{3.15}
$$

Como $J(\vartheta)$ es finita $\forall\vartheta \in \Theta_0$, esto implica que $|P(r,s;\vartheta)\boldsymbol{\ell}_i(r,s;\vartheta)| \leq L$, para cierta constante positiva $L > 0$, $\forall r,s \in \mathbb{N}_0$, $\forall\vartheta \in \Theta_0$, $i = 1,2,3$.

Puesto que $\hat{\theta}_n \xrightarrow{c.s.} \theta$, de (3.14) resulta

$$
\sup_{u,v\in\mathbb{N}_0}\left|B_i(u,v;\hat{\theta}_n) - B_i(u,v;\theta)\right| = o(1),\ \ i = 1,2,3. \tag{3.16}
$$

Por lo tanto

$$
\sup_{r,s,u,v\in\mathbb{N}_0}\left|(\hat{\theta}_{3n}-\theta_3)P(r,s;\hat{\theta}_n)\boldsymbol{\ell}(r,s;\hat{\theta}_n)B(u,v;\hat{\theta}_n)^\top\right| = o(1),
$$

$$
\sup_{r,s,u,v\in\mathbb{N}_0}\left|\theta_3 P(r,s;\theta)\boldsymbol{\ell}(r,s;\theta)\left\{B(u,v;\hat{\theta}_n)^\top - B(u,v;\theta)^\top\right\}\right| = o(1).
$$

Para el tercer sumando de (3.15), el Lema 2.2.5 junto con que $\hat{\theta}_n \in \Theta_0$ y el hecho que $|B_i(u,v;\hat{\theta}_n)| \leq 1$, $i = 1,2,3$, implican

$$
\sup_{r,s,u,v\in\mathbb{N}_0}\left|\theta_3\left\{P(r,s;\hat{\theta}_n)\boldsymbol{\ell}(r,s;\hat{\theta}_n) - P(r,s;\theta)\boldsymbol{\ell}(r,s;\theta)\right\}B(u,v;\hat{\theta}_n)^\top\right| = o(1).
$$

Veamos ahora que $\sup_{r,s,u,v\in\mathbb{N}_0}|A_1| = o(1)$. Para ello, notar que

$$
\begin{aligned}
rP(r,s;\hat{\theta}_n)\boldsymbol{\ell}(r,s;\hat{\theta}_n)B(u,v;\hat{\theta}_n)^\top - rP(r,s;\theta)\boldsymbol{\ell}(r,s;\theta)B(u,v;\theta)^\top \\
= r\left\{P(r,s;\hat{\theta}_n)\boldsymbol{\ell}(r,s;\hat{\theta}_n) - P(r,s;\theta)\boldsymbol{\ell}(r,s;\theta)\right\}B(u,v;\hat{\theta}_n)^\top \\
+ rP(r,s;\theta)\boldsymbol{\ell}(r,s;\theta)\left\{B(u,v;\hat{\theta}_n)^\top - B(u,v;\theta)^\top\right\}.
\end{aligned}
$$

Usando el Lema 2.2.5(a) junto con que $\hat{\theta}_n \in \Theta_0$, obtenemos

$$
\begin{aligned}
\sup_{r,s\in\mathbb{N}_0} r\left|\boldsymbol{\ell}_i(r,s;\hat{\theta}_n)P(r,s;\hat{\theta}_n) - \boldsymbol{\ell}_i(r,s;\theta)P(r,s;\theta)\right| \\
\leq \sum_{r,s\geq 0} r\left|\boldsymbol{\ell}_i(r,s;\hat{\theta}_n)P(r,s;\hat{\theta}_n) - \boldsymbol{\ell}_i(r,s;\theta)P(r,s;\theta)\right| = o(1),\ \ i = 1,2,3.
\end{aligned}
$$

De esta última relación y puesto que $|B_i(u,v;\hat\theta_n)| \le 1$, $i=1,2,3$, nos resulta

$$\sup_{r,s,u,v\in\mathbb{N}_0} \left| r\left\{ P(r,s;\hat\theta_n)\boldsymbol{\ell}(r,s;\hat\theta_n) - P(r,s;\theta)\boldsymbol{\ell}(r,s;\theta) \right\} B(u,v;\hat\theta_n)^\top \right| = o(1).$$

Por la desigualdad de Cauchy-Schwarz y puesto que $J(\theta)<\infty$, entonces

$$\begin{aligned}\sup_{r,s\in\mathbb{N}_0} r|\boldsymbol{\ell}_i(r,s;\theta)|P(r,s;\theta) &\le \sum_{r,s\ge 0} r|\boldsymbol{\ell}_i(r,s;\theta)|P(r,s;\theta)\\ &\le \left\{E_\theta\big(X_1^2\big)\right\}^{1/2}\left\{E_\theta\big(\|\boldsymbol{\ell}(\boldsymbol{X}_1;\theta)\|^2\big)\right\}^{1/2} < \infty,\ i=1,2,3.\end{aligned}$$

Esta última desigualdad junto con (3.16) implican

$$\sup_{r,s,u,v\in\mathbb{N}_0} \left| rP(r,s;\theta)\boldsymbol{\ell}(r,s;\theta)\left\{ B(u,v;\hat\theta_n)^\top - B(u,v;\theta)^\top \right\} \right| = o(1).$$

Así, concluimos que $\sup_{r,s,u,v\in\mathbb{N}_0}|A_1| = o(1)$.

Los casos (a), (b) y (c) muestran que se verifica (3.10) para $1\le i,j\le 4$, con lo cual se demuestra la proposición. □

Ahora probaremos que el método bootstrap estima consistentemente la distribución nula de $W_n(\hat\theta_n)$.

Teorema 3.2.3 *Sean $\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n$ vectores aleatorios iid de $\boldsymbol{X}=(X_1,X_2)\in\mathbb{N}_0^2$. Supongamos que se verifica el Supuesto 2.2.4 y que $\hat\theta_n \xrightarrow{c.s.} \theta$, para algún $\theta\in\Theta$. Entonces*

$$\sup_{x\in\mathbb{R}}\left| P_*\left(nW_n^*(\hat\theta_n^*)\le x\right) - P_\theta\Big(nW_n(\hat\theta_n)\le x\Big)\right| \xrightarrow{c.s.} 0.$$

Demostración Por definición

$$W_n^*(\hat\theta_n^*) = \frac{1}{n^2}\sum_{i,j=1}^n h^*(\boldsymbol{X}_i^*,\boldsymbol{X}_j^*;\hat\theta_n^*).$$

Siguiendo pasos similares a los dados en la demostración del Teorema 3.2.1 se puede ver que

$$W_n^*(\hat\theta_n^*) = W_{1n}^*(\hat\theta_n^*),$$

donde

$$W_{1n}^*(\hat\theta_n^*) = \frac{1}{n^2}\sum_{i,j=1}^n h_d^*(\boldsymbol{X}_i^*,\boldsymbol{X}_j^*;\hat\theta_n) + \xi^*,$$

con $\xi^* = o_{P_*}(n^{-1})$ c.s. y h_d^* es definido como h_d en (3.4) con $\boldsymbol{X}_i$, $\boldsymbol{X}_j$ y θ reemplazados por $\boldsymbol{X}_i^*$,$\boldsymbol{X}_j^*$ y $\hat{\theta}_n$, respectivamente.

De (3.3) y (3.9), h_d^* se puede escribir como

$$\begin{aligned} h_d^*(\boldsymbol{X}_i^*, \boldsymbol{X}_j^*; \hat{\theta}_n) &= h_1^*(\boldsymbol{X}_i^*, \boldsymbol{X}_j^*; \hat{\theta}_n) + h_2^*(\boldsymbol{X}_i^*, \boldsymbol{X}_j^*; \hat{\theta}_n) + \sum_{r,s\geq 0} \boldsymbol{\ell}(\boldsymbol{X}_i^*; \hat{\theta}_n) A_{rs}^* \boldsymbol{\ell}(\boldsymbol{X}_j^*; \hat{\theta}_n)^\top \\ &= \sum_{1\leq k\leq 4} \sum_{r,s\geq 0} \phi_{krs}(\boldsymbol{X}_i^*; \hat{\theta}_n) \phi_{krs}(\boldsymbol{X}_j^*; \hat{\theta}_n), \end{aligned}$$

con lo cual

$$W_{1n}^*(\hat{\theta}_n^*) = \sum_{i,j=1}^{n} \frac{1}{n^2} \sum_{1\leq k\leq 4} \sum_{r,s\geq 0} \phi_{krs}(\boldsymbol{X}_i^*; \hat{\theta}_n) \phi_{krs}(\boldsymbol{X}_j^*; \hat{\theta}_n).$$

La demostración se completará verificando las condiciones (i)-(iii) del Lema 1.2.5 para $W_{1n}^*(\hat{\theta}_n^*)$. Para ello, consideremos el siguiente espacio de Hilbert separable

$$\mathcal{H}_2 = \left\{ z = z(r,s) = (z_{1rs}, z_{2rs}, z_{3rs}, z_{4rs})_{r,s\geq 0} \ \text{ tal que } \ \|z\|_{\mathcal{H}_2}^2 = \sum_{1\leq k\leq 4} \sum_{r,s\geq 0} z_{krs}^2 < \infty \right\}.$$

con producto escalar $\langle x, y\rangle_{\mathcal{H}_2} = \sum_{1\leq k\leq 4} \sum_{r,s\geq 0} x_{krs} y_{krs} < \infty$.

En este espacio de Hilbert se tiene que

$$nW_{1n}^*(\hat{\theta}_n^*) = \sum_{1\leq k\leq 4} \sum_{r,s\geq 0} \left\{ \frac{1}{\sqrt{n}} \sum_{1\leq i\leq n} \phi_{krs}(\boldsymbol{X}_i^*; \hat{\theta}_n) \right\}^2 \equiv \|\Phi_n^*\|_{\mathcal{H}_2}^2,$$

donde Φ_n^* está dado por

$$\Phi_n^*(r,s) = \sum_{i=1}^{n} \Phi_{ni}^*(r,s),$$

$$\Phi_{ni}^*(r,s) = \frac{1}{\sqrt{n}} \Phi(\boldsymbol{X}_i^*; \hat{\theta}_n; r, s), 1 \leq i \leq n,$$

$$\Phi(\boldsymbol{X}_i^*; \hat{\theta}_n; r, s) = \Big(\phi_{1rs}(\boldsymbol{X}_i^*; \hat{\theta}_n), \phi_{2rs}(\boldsymbol{X}_i^*; \hat{\theta}_n), \phi_{3rs}(\boldsymbol{X}_i^*; \hat{\theta}_n), \phi_{4rs}(\boldsymbol{X}_i^*; \hat{\theta}_n)\Big), 1 \leq i \leq n, \tag{3.17}$$

con $\phi_{1rs}(x;\theta)$ y $\phi_{2rs}(x;\theta)$ definidos en (3.2), $\phi_{3rs}(x;\theta)$ y $\phi_{4rs}(x;\theta)$ definidos en (3.8), donde x y θ son reemplazados por $\boldsymbol{X}_i^*$ y $\hat{\theta}_n$, respectivamente.

Por el Supuesto 2.2.4 y por (1.7), $\Phi_{ni}^*(r,s)$ tiene medias nulas y momentos segundos finitos, $1 \leq i \leq n$, pues $\hat{\theta}_n \xrightarrow{c.s.} \theta$.

Sean $(p,q),(u,v) \in \mathbb{N}_0^2$ y consideremos el núcleo de covarianza $K_n(p,q,u,v) = E_*\left\{\Phi_n^*(p,q)^\top \Phi_n^*(u,v)\right\}$, luego

$$K_n(p,q,u,v) = \Big(E_*\{\phi_{ipq}(\boldsymbol{X}_1^*;\hat{\theta}_n)\phi_{juv}(\boldsymbol{X}_1^*;\hat{\theta}_n)\}\Big), 1 \le i,j \le 4.$$

Además, sea $K(p,q,u,v) = E_\theta\left\{\Phi(\boldsymbol{X}_1;\theta;p,q)^\top \Phi(\boldsymbol{X}_1;\theta;u,v)\right\}$ donde Φ es como el definido en (3.17), es decir,

$$K(p,q,u,v) = \Big(E_\theta\{\phi_{ipq}(\boldsymbol{X}_1;\theta)\phi_{juv}(\boldsymbol{X}_1;\theta)\}\Big), 1 \le i,j \le 4.$$

Sea $\mathcal{Z}$ la sucesión Gaussiana centrada cuyo operador de covarianza C está caracterizado por

$$\begin{aligned}\langle Cx,y\rangle_{\mathcal{H}_2} &= cov\left(\langle \mathcal{Z},x\rangle_{\mathcal{H}_2}, \langle \mathcal{Z},y\rangle_{\mathcal{H}_2}\right) = E_\theta\left\{\langle \mathcal{Z},x\rangle_{\mathcal{H}_2}\langle \mathcal{Z},y\rangle_{\mathcal{H}_2}\right\} \\ &= \sum_{p,q,u,v\ge 0} x(p,q)K(p,q,u,v)y(u,v)^\top. \end{aligned} \tag{3.18}$$

Por el Lema 1.2.4, $(\Phi_{1n},\Phi_{2n},\Phi_{3n},\Phi_{4n}) \xrightarrow{L} \mathcal{Z}$ en $\mathcal{H}_2$, cuando los datos son iid provenientes de $\boldsymbol{X} \sim PB(\theta)$, donde

$$\Phi_{kn}(r,s) = \frac{1}{\sqrt{n}}\sum_{1\le i\le n}\phi_{krs}(\boldsymbol{X}_i;\theta), 1 \le k \le 4.$$

Para nuestro caso una base ortonormal conveniente en $\mathcal{H}_2$ es la dada por el conjunto $\left\{e_{r_0s_0}^k : r_0,s_0 \ge 0, k=1,2,3,4\right\}$, donde

$$\begin{aligned} e_{r_0s_0}^1(r,s) &= \left(I\{r=r_0,s=s_0\},0,0,0\right)_{r,s\ge 0}, \\ e_{r_0s_0}^2(r,s) &= \left(0,I\{r=r_0,s=s_0\},0,0\right)_{r,s\ge 0}, \\ e_{r_0s_0}^3(r,s) &= \left(0,0,I\{r=r_0,s=s_0\},0\right)_{r,s\ge 0}, \\ e_{r_0s_0}^4(r,s) &= \left(0,0,0,I\{r=r_0,s=s_0\}\right)_{r,s\ge 0}. \end{aligned}$$

Ahora mostraremos que se verifican las condiciones (i)-(iii) del Lema 1.2.5.

(i) Sea C_n el operador de covarianza de Φ_n^*, esto es, $\forall x, y \in \mathcal{H}_2$

$$\begin{aligned}\langle C_n x, y\rangle_{\mathcal{H}_2} &= cov(\langle \Phi_n^*, x\rangle_{\mathcal{H}_2}, \langle \Phi_n^*, y\rangle_{\mathcal{H}_2}) = E_*\big\{\langle \Phi_n^*, x\rangle_{\mathcal{H}_2}\langle \Phi_n^*, y\rangle_{\mathcal{H}_2}\big\} \\ &= \sum_{p,q,u,v\geq 0} x(p,q)E_*\big\{\Phi_n^*(p,q)^\top \Phi_n^*(u,v)\big\}\, y(u,v)^\top \\ &= \sum_{p,q,u,v\geq 0} x(p,q)K_n(p,q,u,v)y(u,v)^\top. \end{aligned} \tag{3.19}$$

De (3.18), (3.19) y de la Proposición 3.2.2

$$\begin{aligned}\lim_{n\to\infty}\langle C_n e^k_{p_0q_0}, e^l_{u_0v_0}\rangle_{\mathcal{H}_2} &= \lim_{n\to\infty}\sum_{p,q,u,v\geq 0} e^k_{p_0q_0}(p,q)K_n(p,q,u,v)e^l_{u_0v_0}(u,v)^\top \\ &= \sum_{p,q,u,v\geq 0} e^k_{p_0q_0}(p,q)K(p,q,u,v)e^l_{u_0v_0}(u,v)^\top \\ &= \langle Ce^k_{p_0q_0}, e^l_{u_0v_0}\rangle_{\mathcal{H}_2} = a^{kl}_{p_0q_0u_0v_0}.\end{aligned}$$

(ii) De (3.18), (3.19), de la Proposición 3.2.2 y del ítem (i), se obtiene

$$\begin{aligned}&\lim_{n\to\infty}\sum_{1\leq k\leq 4}\sum_{p_0,q_0\geq 0}\langle C_n e^k_{p_0q_0}, e^k_{p_0q_0}\rangle_{\mathcal{H}_2} \\ &\qquad = \lim_{n\to\infty}\sum_{1\leq k\leq 4}\sum_{p_0,q_0\geq 0}\sum_{p,q,p,q\geq 0} e^k_{p_0q_0}(p,q)K_n(p,q,p,q)e^l_{p_0q_0}(p,q)^\top \\ &\qquad = \sum_{1\leq k\leq 4}\sum_{p_0,q_0\geq 0}\sum_{p,q,p,q\geq 0} e^k_{p_0q_0}(p,q)K(p,q,p,q)e^l_{p_0q_0}(p,q)^\top \\ &\qquad = \sum_{1\leq k\leq 4}\sum_{p_0,q_0\geq 0}\langle Ce^k_{p_0q_0}, e^k_{p_0q_0}\rangle_{\mathcal{H}_2} \\ &\qquad = \sum_{1\leq k\leq 4}\sum_{p_0,q_0\geq 0} a^{kk}_{p_0q_0p_0q_0} < \infty,\end{aligned}$$

pues, de la primera ecuación en (3.18) y la igualdad de Parseval

$$\begin{aligned}\sum_{1\leq k\leq 4}\sum_{p_0,q_0\geq 0} a^{kk}_{p_0q_0p_0q_0} &= \sum_{1\leq k\leq 4}\sum_{p_0,q_0\geq 0}\langle Ce^k_{p_0q_0}, e^k_{p_0q_0}\rangle_{\mathcal{H}_2} \\ &= \sum_{1\leq k\leq 4}\sum_{p_0,q_0\geq 0} E_\theta\Big\{\langle \mathcal{Z}, e^k_{p_0q_0}\rangle^2_{\mathcal{H}_2}\Big\} = E_\theta\Big\{\|\mathcal{Z}\|^2_{\mathcal{H}_2}\Big\} < \infty.\end{aligned}$$

(iii) Sean $r_0, s_0 \in \mathbb{N}_0$, fijos,

$$\langle \Phi^*_{ni}, e^k_{r_0 s_0} \rangle^2_{\mathcal{H}_2} = \frac{1}{n} \phi^2_{k r_0 s_0}(\boldsymbol{X}^*_i; \hat{\theta}_n), k = 1, 2, 3, 4. \tag{3.20}$$

Para $k = 1$, tenemos

$$|\langle \Phi^*_{ni}, e^1_{r_0 s_0} \rangle_{\mathcal{H}_2}| = \frac{1}{\sqrt{n}} |\phi_{1 r_0 s_0}(\boldsymbol{X}^*_i; \hat{\theta}_n)| \leq r_0 + 1 + |\hat{\theta}_{1n} - \hat{\theta}_{3n}| + \hat{\theta}_{3n}.$$

Por tanto, dado $\varepsilon > 0$, existirá $n_0 = n_0(r_0, \varepsilon) \in \mathbb{N}_0$ tal que

$$\frac{1}{\sqrt{n}} |\phi_{1 r_0 s_0}(\boldsymbol{X}^*_i; \hat{\theta}_n)| \leq \frac{1}{\sqrt{n}} (r_0 + 1 + |\hat{\theta}_{1n} - \hat{\theta}_{3n}| + \hat{\theta}_{3n}) \leq \varepsilon, \forall n > n_0.$$

De donde,

$$I\left\{ \frac{1}{\sqrt{n}} |\phi_{1 r_0 s_0}(\boldsymbol{X}^*_i; \hat{\theta}_n)| > \varepsilon \right\} = \left\{ \begin{array}{ll} 1, & \text{si } n \leq n_0, \\ 0, & \text{si } n > n_0. \end{array} \right. \tag{3.21}$$

Se tiene que

$$L_n(\varepsilon, e^1_{r_0 s_0}) = \frac{1}{n} \sum_{i=1}^n E_* \left[\phi^2_{1 r_0 s_0}(\boldsymbol{X}^*_i; \hat{\theta}_n) I\left\{ \frac{1}{\sqrt{n}} |\phi_{1 r_0 s_0}(\boldsymbol{X}^*_i; \hat{\theta}_n)| > \varepsilon \right\} \right].$$

Además, de (3.2)

$$\begin{aligned} \phi^2_{1 r_0 s_0}(\boldsymbol{X}^*_i; \hat{\theta}_n) &= (r_0 + 1)^2\, I\{X^*_{1i} = r_0 + 1, X^*_{2i} = s_0\} \\ &\quad + (\hat{\theta}_{1n} - \hat{\theta}_{3n})^2\, I\{X^*_{1i} = r_0, X^*_{2i} = s_0\} + \hat{\theta}^2_{3n} I\{X^*_{1i} = r_0, X^*_{2i} = s_0 - 1\}. \end{aligned}$$

Con esta última igualdad y (3.21)

$$\begin{aligned} &\sum_{i=1}^n E_* \left[\phi^2_{1 r_0 s_0}(\boldsymbol{X}^*_i; \hat{\theta}_n) I\left\{ \frac{1}{\sqrt{n}} |\phi_{1 r_0 s_0}(\boldsymbol{X}^*_i; \hat{\theta}_n)| > \varepsilon \right\} \right] \\ &= \sum_{i=1}^n \sum_{r,s \geq 0} \Big[(r_0 + 1)^2\, I\{r = r_0 + 1, s = s_0\} + (\hat{\theta}_{1n} - \hat{\theta}_{3n})^2\, I\{r = r_0, s = s_0\} \\ &\qquad + \hat{\theta}^2_{3n}\, I\{r = r_0, s = s_0 - 1\} \Big]\, I\left\{ \frac{1}{\sqrt{n}} |\phi_{1 r_0 s_0}(r, s; \hat{\theta}_n)| > \varepsilon \right\} P(r, s; \hat{\theta}_n) \\ &\leq \sum_{i=1}^{n_0} \sum_{r,s \geq 0} \Big[(r_0 + 1)^2\, I\{r = r_0 + 1, s = s_0\} + (\hat{\theta}_{1n} - \hat{\theta}_{3n})^2\, I\{r = r_0, s = s_0\} \\ &\qquad + \hat{\theta}^2_{3n}\, I\{r = r_0, s = s_0 - 1\} \Big]\; P(r, s; \hat{\theta}_n) \\ &= \sum_{i=1}^{n_0} \left\{ (r_0 + 1)^2 P(r_0 + 1, s_0; \hat{\theta}_n) + (\hat{\theta}_{1n} - \hat{\theta}_{3n})^2 P(r_0, s_0; \hat{\theta}_n) + \hat{\theta}^2_{3n} P(r_0, s_0 - 1; \hat{\theta}_n) \right\} \\ &\leq n_0 \left\{ (r_0 + 1)^2 + (\hat{\theta}_{1n} - \hat{\theta}_{3n})^2 + \hat{\theta}^2_{3n} \right\}. \end{aligned}$$

Por tanto
$$L_n(\varepsilon, e^1_{r_0s_0}) \leq \frac{n_0}{n}\Big\{(r_0+1)^2 + (\hat{\theta}_{1n} - \hat{\theta}_{3n})^2 + \hat{\theta}^2_{3n}\Big\}.$$
Puesto que r_0 y ε son fijos y $n_0 = n_0(r_0, \varepsilon)$, entonces esta última desigualdad converge a cero. Así
$$\lim_{n\to\infty} L_n(\varepsilon, e^1_{r_0s_0}) = 0.$$
Análogamente se logra que $\lim_{n\to\infty} L_n(\varepsilon, e^2_{r_0s_0}) = 0$.

De (3.20), para $k = 3$ y de (3.8)
$$\langle \Phi^*_{ni}, e^3_{r_0s_0}\rangle_{\mathcal{H}_2} = \frac{1}{\sqrt{n}}\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n) d(r_0, s_0; \hat{\theta}_n)^\top,$$
donde $d(r, s; \hat{\theta}_n) = (P(r, s; \hat{\theta}_n), 0, P(r, s-1; \hat{\theta}_n) - P(r, s; \hat{\theta}_n))$ y por tanto
$$L_n(\varepsilon, e^3_{r_0s_0})$$
$$= \frac{1}{n}\sum_{i=1}^n E_*\left[\Big\{\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n) d(r_0, s_0; \hat{\theta}_n)^\top\Big\}^2 I\Big\{\frac{1}{\sqrt{n}}\Big|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n) d(r_0, s_0; \hat{\theta}_n)^\top\Big| > \varepsilon\Big\}\right].$$
Ahora, como
$$\Big|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n) d(r, s; \hat{\theta}_n)^\top\Big| \leq \|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n)\| \|d(r, s; \hat{\theta}_n)\| \leq \sqrt{2}\|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n)\|, \tag{3.22}$$
entonces
$$\varepsilon < \frac{1}{\sqrt{n}}|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n) d(r_0, s_0; \hat{\theta}_n)^\top| \leq \frac{\sqrt{2}}{\sqrt{n}}\|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n)\|,$$
por tanto
$$\begin{aligned} I\Big\{\frac{1}{\sqrt{n}}\Big|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n) d(r_0, s_0; \hat{\theta}_n)^\top\Big| > \varepsilon\Big\} &\leq I\left\{\frac{\sqrt{2}}{\sqrt{n}}\|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n)\| > \varepsilon\right\} \\ &= I\Big\{\|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n)\| > \gamma\Big\}, \end{aligned}$$
donde $\gamma = \gamma(n) = \varepsilon\frac{\sqrt{n}}{\sqrt{2}}$ es tal que $\gamma \to \infty$ cuando $n \to \infty$.

Además, de (3.22), se tiene que $\Big(\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n) d(r_0, s_0; \hat{\theta}_n)^\top\Big)^2 \leq 2\|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n)\|^2$ y como $\boldsymbol{X}^*_1, \boldsymbol{X}^*_2, \ldots, \boldsymbol{X}^*_n$ son vectores aleatorios iid de $\boldsymbol{X}^* \sim PB(\hat{\theta}_n)$, entonces
$$\begin{aligned} L_n(\varepsilon, e^3_{r_0s_0}) &\leq \frac{2}{n}\sum_{i=1}^n E_*\Big(\|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n)\|^2 I\Big\{\|\boldsymbol{\ell}(\boldsymbol{X}^*_i; \hat{\theta}_n)\| > \gamma\Big\}\Big) \\ &\leq 2\sup_{\hat{\theta}_n\in\Theta_0} E_{\hat{\theta}_n}\Big(\|\boldsymbol{\ell}(\boldsymbol{X}^*; \hat{\theta}_n)\|^2 I\Big\{\|\boldsymbol{\ell}(\boldsymbol{X}^*; \hat{\theta}_n)\| > \gamma\Big\}\Big), \end{aligned}$$

y como $\gamma \to \infty$ cuando $n \to \infty$, entonces, por el Supuesto 2.2.4 (1) se concluye que

$$\lim_{n\to\infty} L_n(\varepsilon, e^3_{r_0 s_0}) = 0.$$

De la misma forma se prueba que $\lim_{n\to\infty} L_n(\varepsilon, e^4_{r_0 s_0}) = 0$.

Así, por el Lema 1.2.5, tenemos que $\Phi_n^* \xrightarrow{L} \mathcal{Z}$, c.s., en $\mathcal{H}_2$. Como ya habíamos mencionado, $\mathcal{Z}$ es también el límite débil de $(\Phi_{1n}, \Phi_{2n}, \Phi_{3n}, \Phi_{4n})$ cuando los datos son iid provenientes de $\boldsymbol{X} \sim PB(\theta)$. Finalmente, el resultado se obtiene por el Teorema de la aplicación continua. $\square$

Similarmente a lo expuesto en la parte final de la Sección 2.2, sea

$$w^*_{n,\alpha} = \inf\Big\{x : P_*(W_n^*(\hat{\theta}_n^*) \geq x) \leq \alpha\Big\}$$

el percentil superior α de la distribución bootstrap de $W_n(\hat{\theta}_n)$.

Del Teorema 3.2.3, la función test

$$\Psi_W^* = \begin{cases} 1, & \text{si } W_n(\hat{\theta}_n) \geq w^*_{n,\alpha}, \\ 0, & \text{en caso contrario,} \end{cases}$$

o equivalentemente, el test que rechaza H_0 cuando

$$p_W^* = P_*\Big(W_n^*(\hat{\theta}_n^*) \geq W_{obs}\Big) \leq \alpha,$$

es asintóticamente correcto en el sentido que, cuando H_0 es cierta

$$\lim_{n\to\infty} P_\theta(\Psi_W^* = 1) = \alpha,$$

donde W_{obs} es el valor observado del test estadístico $W_n(\hat{\theta}_n)$.

Observación 3.2.4 *Para aproximar $w^*_{n,\alpha}$ se sigue el mismo procedimiento presentado al final de la Sección 2.2, con los cambios obvios.*

3.3. Alternativas

En esta sección estudiaremos el comportamiento del test Ψ_W^* bajo alternativas fijas y locales.

3.3.1. Alternativas fijas

Como una de nuestras metas es obtener tests de bondad de ajuste que sean consistentes, lo próximo que haremos es estudiar este tópico para el nuevo test que proponemos. Con este fin, primero obtendremos el límite c.s. de $W_n(\hat{\theta}_n)$.

Teorema 3.3.1 *Sean $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid de $\boldsymbol{X} = (X_1, X_2) \in \mathbb{N}_0^2$, con $E(X_k^2) < \infty$, $k = 1, 2$. Sea $p(r,s) = P(X_1 = r, X_2 = s)$. Si $\hat{\theta}_n \xrightarrow{c.s.} \theta$, para algún $\theta \in \mathbb{R}^3$, entonces*

$$W_n(\hat{\theta}_n) \xrightarrow{c.s.} \sum_{r,s\geq 0} \left\{a_1(r,s;\theta)^2 + a_2(r,s;\theta)^2\right\} = \nu(P;\theta),$$

donde

$$a_1(r,s;\theta) = (r+1)p(r+1,s) - (\theta_1 - \theta_3)p(r,s) - \theta_3 p(r,s-1),$$

$$a_2(r,s;\theta) = (s+1)p(r,s+1) - (\theta_2 - \theta_3)p(r,s) - \theta_3 p(r-1,s).$$

Observación 3.3.2 *Notar que $\nu(P;\theta) \geq 0$. Además, de la Proposición 2.1.1 obtenemos que, $\nu(P;\theta) = 0$ sí y sólo si H_0 es cierta.*

Demostración del Teorema 3.3.1 Se tiene que

$$d_1(r,s;\hat{\theta}_n) = d_1(r,s;\theta) - (\hat{\theta}_{1n} - \theta_1)p_n(r,s) + (\hat{\theta}_{3n} - \theta_3)\{p_n(r,s) - p_n(r,s-1)\}. \tag{3.23}$$

También se tiene que

$$\sum_{r,s\geq 0} d_1(r,s;\theta)^2 = \frac{1}{n^2}\sum_{i\neq j} h_1(\boldsymbol{X}_i, \boldsymbol{X}_j;\theta) + \frac{1}{n^2}\sum_{i=1}^{n} h_1(\boldsymbol{X}_i, \boldsymbol{X}_i;\theta).$$

Por la ley fuerte de los grandes números para U-estadísticos (véase por ejemplo Teorema 5.4 en Serfling (1980) [42]),

$$\frac{1}{n^2}\sum_{i\neq j} h_1(\boldsymbol{X}_i, \boldsymbol{X}_j;\theta) \xrightarrow{c.s.} E\{h_1(\boldsymbol{X}_1, \boldsymbol{X}_2;\theta)\} = \sum_{r,s\geq 0} a_1(r,s;\theta)^2.$$

Como $E(X_1^2) < \infty$, por la ley fuerte de los grandes números,

$$\frac{1}{n}\sum_{i=1}^{n} h_1(\boldsymbol{X}_i, \boldsymbol{X}_i; \theta) \xrightarrow{c.s.} E\{h_1(\boldsymbol{X}_1, \boldsymbol{X}_1; \theta)\} = E\left\{\sum_{r,s\geq 0} \phi_{1rs}(\boldsymbol{X}_1; \theta)^2\right\} < \infty.$$

Por tanto,

$$\sum_{r,s\geq 0} d_1(r,s;\theta)^2 \xrightarrow{c.s.} \sum_{r,s\geq 0} a_1(r,s;\theta)^2. \tag{3.24}$$

Como $p_n(r,s)^2 \leq p_n(r,s)$, $\forall r,s \geq 0$, y $\sum_{r,s\geq 0} p_n(r,s) = 1$, se tiene que

$$(\hat{\theta}_{1n} - \theta_1)^2 \sum_{r,s\geq 0} p_n(r,s)^2 \leq (\hat{\theta}_{1n} - \theta_1)^2 = o(1), \tag{3.25}$$

y análogamente,

$$(\hat{\theta}_{3n} - \theta_3)^2 \sum_{r,s\geq 0} \{p_n(r,s) - p_n(r,s-1)\}^2 = o(1). \tag{3.26}$$

De (3.23)–(3.26),

$$\sum_{r,s\geq 0} d_1(r,s;\hat{\theta}_n)^2 \xrightarrow{c.s.} \sum_{r,s\geq 0} a_1(r,s;\theta)^2. \tag{3.27}$$

Procediendo similarmente

$$\sum_{r,s\geq 0} d_2(r,s;\hat{\theta}_n)^2 \xrightarrow{c.s.} \sum_{r,s\geq 0} a_2(r,s;\theta)^2. \tag{3.28}$$

Finalmente, el resultado se tiene de (3.27) y (3.28). □

Como una consecuencia de los Teoremas 3.2.1, 3.2.3 y 3.3.1, y la Observación 3.3.2, el siguiente resultado da la consistencia del test Ψ_W^*.

Corolario 3.3.3 *Sean $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ vectores aleatorios iid de $\boldsymbol{X} \in \mathbb{N}_0^2$. Supongamos que se cumplen las hipótesis de los Teoremas 3.2.1 y 3.2.3.*

Si H_0 no es cierta, entonces $P(\Psi_W^* = 1) \to 1$.

3.3.2. Alternativas contiguas

Como se estableció en el Teorema 3.2.1, cuando H_0 es cierta, $nW_n(\hat{\theta}_n)$ converge débilmente a una combinación lineal de v.a. χ^2 independientes con 1 grado de libertad, donde los pesos son los autovalores del operator $C(\theta)$ dado en (3.6). Sea $\{\phi_j\}$ el conjunto ortonormal de autofunciones correspondientes a los autovalores $\{\lambda_j\}$ de $C(\theta)$.

El siguiente Teorema da la ley límite del test estadístico bajo las alternativas $P_n(x_1, x_2)$ en (2.24).

Teorema 3.3.4 *Sea $\boldsymbol{X}_{n,1}, \boldsymbol{X}_{n,2}, \ldots, \boldsymbol{X}_{n,n}$ un arreglo triangular de vectores aleatorios bivariantes que son independientes por filas y que toman valores en $\mathbb{N}_0^2$, con función de probabilidad dada por $P_n(x_1, x_2)$ definida en (2.24). Supongamos que se cumplen los Supuestos 2.2.1 y 2.3.4. Entonces*

$$nW_n(\hat{\theta}_n) \xrightarrow{L} \sum_{j=1}^{\infty} \lambda_j \left(Z_j + c_j\right)^2,$$

donde $c_j = \sum_{x_1,x_2} b(x_1, x_2)\phi_j(x_1, x_2)$ y $Z_1, Z_2, \ldots$ son variables normales estándar independientes.

Demostración Sigue los mismos pasos que la demostración del Teorema 2.3.5, por lo tanto la omitiremos. □

Del Teorema 3.3.4, concluimos que el test Ψ_W^* es capaz de detectar alternativas como las establecidas en (2.24), que convergen a la DPB a una razón de $n^{-1/2}$.

Capítulo 4

Resultados numéricos

Las propiedades estudiadas en las Secciones 2.2, 2.3, 3.2 y 3.3 son asintóticas, esto es, tales propiedades describen el comportamiento de los tests propuestos para muestras de tamaño grande. Para estudiar la bondad de la aproximación bootstrap para muestras de tamaño finito, así como también para comparar la potencia de los tests propuestos tanto entre ellos como con otros tests, hemos llevado a cabo un experimento de simulación. En esta sección describimos dicho experimento y damos un resumen de los resultados que hemos obtenido.

Todos los cálculos computacionales realizados en esta memoria se llevaron a cabo mediante el uso de programas escritos en el lenguaje R [36].

Como competidores de los tests estadísticos propuestos, hemos considerado los tests dados por Crockett (1979) [6] (denotado por T), Loukas y Kemp (1986) [30] (denotado por I_B), y Rayner y Best (1995) [37] (denotado por NI_B). La expresión de cada test estadístico, así como la regiones críticas correspondientes ya las presentamos en el Capítulo 1.

Para calcular $R_{n,w}(\hat{\theta}_n)$ y $S_{n,w}(\hat{\theta}_n)$ es necesario dar una forma explícita a la función de peso w. En el caso univariante, Baringhaus et al. (2000) [2] consideraron la función de peso $w(u) = u^a$, con $a \geq 0$. Varias extensiones son posibles. Aquí hemos tenido en cuenta la siguiente

$$w(u; a_1, a_2) = u_1^{a_1} u_2^{a_2}. \tag{4.1}$$

Observar que las únicas restricciones que hemos impuesto a la función de peso son que w sea positiva casi en todas partes en $[0, 1]^2$ y la establecida en (2.2). La función $w(u; a_1, a_2)$ dada en (4.1) cumple dichas condiciones siempre que $a_i > -1$, $i = 1, 2$.

En el Capítulo 5 damos las expresiones matemáticas de los estadísticos $R_{n,w}(\hat{\theta}_n)$ y $S_{n,w}(\hat{\theta}_n)$ para la función de peso dada en (4.1). También damos la expresión para el estadístico $W_n(\hat{\theta}_n)$.

Además, mostramos los resultados que obtuvimos al aplicar los tests propuestos a dos conjuntos de datos reales.

4.1. Datos simulados

Para estudiar la bondad de la aproximación bootstrap en muestras de tamaño finito tanto para los tests propuestos, como para las aproximaciones a la distribución nula de los tests estadísticos T, I_B y NI_B, que están basados en sus distribuciones asintóticas nulas, generamos muestras de tamaño $n = 30(20)70$ de la distribución $PB(\theta_1, \theta_2, \theta_3)$, con $\theta_1 = \theta_2 =$1.0 y θ_3 tal que el coeficiente de correlación, $\rho = \theta_3/\sqrt{\theta_1\,\theta_2}$ sea igual a 0.25, 0.50 y 0.75, con el fin de examinar la bondad de las aproximaciones para datos con una correlación baja, media y alta, respectivamente.

Para estimar el parámetro θ empleamos el método de máxima verosimilitud como se describe en Kocherlakota y Kocherlakota (1992, pp. 103–105) [26]. Luego, aproximamos los $p-$valores bootstrap, p^*, de los tests propuestos, para ello, en el caso de $R_{n,w}(\hat{\theta}_n)$ y $S_{n,w}(\hat{\theta}_n)$ usamos la función de peso dada en (4.1) para $(a_1, a_2) \in \{(0, 0), (1, 0)\}$ y generamos $B = 500$ muestras bootstrap. También, calculamos los $p-$valores (asintóticos) asociados a los tests estadísticos T, I_B y NI_B.

El procedimiento anterior lo repetimos 1000 veces y calculamos la fracción de los $p-$valores estimados que resultaron ser menores o iguales que 0.05 y 0.10, que son las estimaciones de las probabilidades del error tipo I para $\alpha = 0.05$ y 0.10 (en las tablas, denotamos esto como f05 y f10), respectivamente.

Si las aproximaciones consideradas fuesen exactas, entonces los $p-$valores calculados deberían ser una muestra aleatoria de una distribución uniforme en el intervalo $(0, 1)$. Por lo tanto, para medir la bondad de las aproximaciones consideradas, calculamos el $p-$valor del test estadístico de uniformidad de Kolmogorov-Smirnov (KS) para cada conjunto de 1000 $p-$valores obtenido para cada uno de los tests estadísticos.

Repetimos el experimento anterior para $\theta_1 = 1.5$, $\theta_2 = 1.0$ y θ_3 de manera que el coeficiente de correlación, $\rho = \theta_3/\sqrt{\theta_1\,\theta_2}$, fuera aproximadamente igual a 0.25, 0.50 y 0.75. En este caso, ya que $\theta_1 \neq \theta_2$, además de $(a_1, a_2) \in \{(0, 0), (1, 0)\}$ consideramos

$(a_1, a_2) = (0, 1)$ para $R_{n,a}$ y $S_{n,a}$, con el fin de examinar el efecto de dar un peso diferente a cada uno de los componentes cuando estos tienen distintas esperanzas.

Las Tablas 4.1– 4.6 resumen los resultados obtenidos. En dichas tablas, denotamos por $R_{n,a}$ y $S_{n,a}$ a los tests estadísticos $R_{n,w}(\hat{\theta}_n)$ y $S_{n,w}(\hat{\theta}_n)$, respectivamente, cuando la función de peso w toma la forma dada por (4.1), para algún $a = (a_1, a_2)$.

Observando los valores dados en las tablas, concluimos que la aproximación asintótica de los $p-$valores de los tests estadísticos T, I_B y NI_B no da resultados satisfactorios para los casos tratados. Por el contrario, el bootstrap proporciona una aproximación precisa a la distribución nula de $R_{n,w}(\hat{\theta}_n)$ y $S_{n,w}(\hat{\theta}_n)$ en todos los casos tratados. La aproximación bootstrap a la distribución nula de $W_n(\hat{\theta}_n)$ es adecuada para $n \geq 50$.

Tabla 4.1: Resultados de simulación para la probabilidad del error tipo I. Caso $E(X_1) = E(X_2)$, $\rho = 0.25$

	$n = 30$			$n = 50$			$n = 70$		
$\theta = (1.0,1.0,0.25)$ $\rho = 0.25$	f05	f10	KS	f05	f10	KS	f05	f10	KS
$R_{n,(0,0)}$	0.035	0.091	0.818621	0.047	0.100	0.459543	0.041	0.083	0.329116
$S_{n,(0,0)}$	0.046	0.090	0.902243	0.049	0.097	0.818621	0.041	0.089	0.769894
$R_{n,(1,0)}$	0.035	0.093	0.257432	0.050	0.100	0.769894	0.042	0.086	0.413150
$S_{n,(1,0)}$	0.039	0.094	0.665399	0.048	0.098	0.995881	0.042	0.093	0.559560
W_n	0.022	0.056	1.00e-05	0.033	0.078	0.111356	0.038	0.090	0.612128
T	0.011	0.031	$<$ 2.2e-16	0.046	0.092	0.060937	0.013	0.038	$<$ 2.2e-16
I_B	0.027	0.061	$<$ 2.2e-16	0.098	0.144	0.001642	0.022	0.054	$<$ 2.2e-16
NI_B	0.010	0.034	$<$ 2.2e-16	0.068	0.111	0.003452	0.013	0.033	$<$ 2.2e-16

Tabla 4.2: Resultados de simulación para la probabilidad del error tipo I. Caso $E(X_1) = E(X_2)$, $\rho = 0.50$

	$n = 30$			$n = 50$			$n = 70$		
$\theta = (1.0,1.0,0.50)$ $\rho = 0.50$	f05	f10	KS	f05	f10	KS	f05	f10	KS
$R_{n,(0,0)}$	0.047	0.091	0.508494	0.044	0.101	0.129364	0.048	0.100	0.413150
$S_{n,(0,0)}$	0.049	0.094	0.863178	0.045	0.098	0.329116	0.046	0.100	0.459543
$R_{n,(1,0)}$	0.049	0.097	0.665399	0.039	0.096	0.863178	0.054	0.097	0.129364
$S_{n,(1,0)}$	0.045	0.092	0.459543	0.043	0.093	0.291736	0.053	0.096	0.459543
W_n	0.022	0.061	0.013476	0.032	0.077	0.111356	0.037	0.081	0.111356
T	0.064	0.095	0.018402	0.024	0.039	$<$ 2.2e-16	0.021	0.053	0.000179
I_B	0.152	0.186	$<$ 2.2e-16	0.073	0.119	$<$ 2.2e-16	0.051	0.081	$<$ 2.2e-16
NI_B	0.080	0.132	0.054004	0.018	0.049	$<$ 2.2e-16	0.007	0.035	$<$ 2.2e-16

Tabla 4.3: Resultados de simulación para la probabilidad del error tipo I. Caso $E(X_1) = E(X_2)$, $\rho = 0.75$

$\theta = (1.0,1.0,0.75)$ $\rho = 0.75$	$n = 30$			$n = 50$			$n = 70$		
	f05	f10	KS	f05	f10	KS	f05	f10	KS
$R_{n,(0,0)}$	0.045	0.097	0.769894	0.060	0.107	0.995881	0.056	0.107	0.508494
$S_{n,(0,0)}$	0.053	0.092	0.369615	0.061	0.106	0.769894	0.050	0.099	0.902243
$R_{n,(1,0)}$	0.052	0.096	0.995881	0.059	0.106	0.960002	0.050	0.109	0.769894
$S_{n,(1,0)}$	0.050	0.096	0.818621	0.059	0.104	0.934732	0.048	0.106	0.329116
W_n	0.029	0.076	0.024117	0.036	0.085	0.111356	0.038	0.088	0.129364
T	0.025	0.049	$<$ 2.2e-16	0.034	0.065	1.00e-07	0.024	0.058	5.30e-06
I_B	0.116	0.140	$<$ 2.2e-16	0.141	0.162	$<$ 2.2e-16	0.129	0.153	$<$ 2.2e-16
NI_B	0.045	0.074	6.10e-06	0.033	0.081	$<$ 2.2e-16	0.029	0.063	$<$ 2.2e-16

Tabla 4.4: Resultados de simulación para la probabilidad del error tipo I. Caso $E(X_1) \neq E(X_2)$, $\rho \approx 0.25$

	$n = 30$			$n = 50$			$n = 70$		
$\theta = (1.5,1.0,0.31)$ $\rho = 0.25311$	f05	f10	KS	f05	f10	KS	f05	f10	KS
$R_{n,(0,0)}$	0.048	0.101	0.257432	0.053	0.110	0.508494	0.047	0.100	0.989545
$S_{n,(0,0)}$	0.049	0.104	0.769894	0.052	0.104	0.172476	0.045	0.092	0.508494
$R_{n,(1,0)}$	0.046	0.097	0.413150	0.053	0.104	0.863178	0.046	0.104	0.559560
$S_{n,(1,0)}$	0.042	0.098	0.291736	0.049	0.104	0.459543	0.051	0.104	0.508494
$R_{n,(0,1)}$	0.055	0.106	0.718379	0.061	0.112	0.902243	0.048	0.090	0.863178
$S_{n,(0,1)}$	0.049	0.103	0.459543	0.051	0.102	0.329116	0.048	0.089	0.612128
W_n	0.022	0.066	0.041633	0.036	0.076	0.111356	0.037	0.082	0.111356
T	0.018	0.046	1.00e-07	0.021	0.060	0.000318	0.053	0.093	0.099411
I_B	0.031	0.060	$<$ 2.2e-16	0.013	0.028	$<$ 2.2e-16	0.108	0.161	0.000589
NI_B	0.016	0.041	$<$ 2.2e-16	0.010	0.018	$<$ 2.2e-16	0.076	0.136	0.048751

Tabla 4.5: Resultados de simulación para la probabilidad del error tipo I. Caso $E(X_1) \neq E(X_2)$, $\rho \approx 0.50$

	$n = 30$			$n = 50$			$n = 70$		
$\theta = (1.5,1.0,0.62)$ $\rho = 0.50623$	f05	f10	KS	f05	f10	KS	f05	f10	KS
$R_{n,(0,0)}$	0.050	0.094	0.329116	0.048	0.093	0.902243	0.054	0.110	0.665399
$S_{n,(0,0)}$	0.046	0.100	0.769894	0.048	0.091	0.769894	0.045	0.094	0.226206
$R_{n,(1,0)}$	0.049	0.093	0.718379	0.046	0.092	0.769894	0.050	0.104	0.718379
$S_{n,(1,0)}$	0.043	0.095	0.612128	0.048	0.087	0.818621	0.050	0.097	0.508494
$R_{n,(0,1)}$	0.050	0.089	0.612128	0.043	0.100	0.718379	0.054	0.100	0.665399
$S_{n,(0,1)}$	0.052	0.091	0.459543	0.043	0.099	0.459543	0.052	0.089	0.508494
W_n	0.026	0.055	0.003013	0.037	0.071	0.111356	0.039	0.079	0.111356
T	0.056	0.088	0.000526	0.050	0.104	0.011917	0.049	0.096	0.001109
I_B	0.147	0.201	< 2.2e-16	0.169	0.223	< 2.2e-16	0.147	0.196	< 2.2e-16
NI_B	0.094	0.152	0.000622	0.082	0.145	0.006666	0.076	0.120	0.078967

Tabla 4.6: Resultados de simulación para la probabilidad del error tipo I. Caso $E(X_1) \neq E(X_2)$, $\rho \approx 0.75$

	$n = 30$			$n = 50$			$n = 70$		
$\theta = (1.5,1.0,0.92)$ $\rho = 0.75118$	f05	f10	KS	f05	f10	KS	f05	f10	KS
$R_{n,(0,0)}$	0.059	0.102	0.369615	0.053	0.094	0.508494	0.046	0.091	0.960002
$S_{n,(0,0)}$	0.055	0.105	0.612128	0.049	0.092	0.769894	0.048	0.088	0.508494
$R_{n,(1,0)}$	0.051	0.096	0.413150	0.051	0.099	0.459543	0.050	0.091	0.863178
$S_{n,(1,0)}$	0.053	0.101	0.459543	0.051	0.092	0.769894	0.045	0.086	0.665399
$R_{n,(0,1)}$	0.058	0.104	0.718379	0.049	0.098	0.459543	0.046	0.089	0.863178
$S_{n,(0,1)}$	0.058	0.108	0.863178	0.050	0.091	0.559560	0.046	0.094	0.329116
W_n	0.037	0.081	0.000714	0.042	0.079	0.111356	0.037	0.083	0.149677
T	0.029	0.059	1.70e-06	0.057	0.094	0.008821	0.078	0.109	0.065401
I_B	0.091	0.116	$<$ 2.2e-16	0.209	0.239	$<$ 2.2e-16	0.196	0.220	$<$ 2.2e-16
NI_B	0.021	0.051	$<$ 2.2e-16	0.089	0.152	0.001554	0.094	0.149	0.003483

Para estudiar la potencia repetimos el experimento anterior para muestras de tamaño $n = 50$ para varias alternativas: la distribución binomial bivariante $BB(m; p_1, p_2, p_3)$, donde $p_1 + p_2 - p_3 \leq 1$, $p_1, p_2 \geq p_3 > 0$; la distribución binomial negativa bivariante $BNB(\nu; \gamma_0, \gamma_1, \gamma_2)$, donde $\nu \in \mathbb{N}, \gamma_0, \gamma_1 > \gamma_2 > 0$; mixturas de la DPB de la forma $pPB(\theta) + (1-p)PB(\lambda), 0 < p < 1$, denotada por $PPB(p; \theta, \lambda)$; la distribución Neyman tipo A bivariante $NTAB(\lambda; \lambda_1, \lambda_2, \lambda_3)$, donde $0 < \lambda_1 + \lambda_2 + \lambda_3 \leq 1$; y la distribución serie logarítmica bivariante $SLB(\lambda_1, \lambda_2, \lambda_3)$, donde $0 < \lambda_1 + \lambda_2 + \lambda_3 < 1$ (ver, por ejemplo, Kocherlakota y Kocherlakota [26] para una descripción de estas distribuciones).

Para generar las muestras aleatorias de las distribuciones bivariantes que usamos como alternativas, implementamos algoritmos computacionales siguiendo los procedimientos de simulación dados en Kocherlakota y Kocherlakota [26].

La distribución binomial bivariante y la distribución Neyman tipo A bivariante han sido utilizadas como alternativas en otros artículos relacionados (véase, por ejemplo, Loukas y Kemp (1986) [30], Rayner y Best (1995) [37]). Las otras distribuciones alternativas empleadas son distribuciones análogas a las utilizadas por Gürtler y Henze (2000) [11] en el caso univariante.

Los parámetros para estas alternativas fueron elegidos de manera que el índice de dispersión de cada componente del vector aleatorio estuvieran cercanos de 1, concretamente, de tal manera que $|var(X_i)/E(X_i) - 1| < 1$, $i = 1, 2$. En este sentido, las alternativas consideradas están "cerca" de la DPB.

Por otra parte, consideramos $a_1 = a_2 = 0$, porque para esta elección de a_1 y a_2 los test estadísticos $R_{n,w}(\hat{\theta}_n)$ y $S_{n,w}(\hat{\theta}_n)$ que resultan ocupan menos tiempo computacional que para otras combinaciones de a.

Las Tablas 4.7 y 4.8 muestran las alternativas consideradas y la potencia estimada para el nivel de significación nominal $\alpha = 0.05$. Estas tablas también muestran el índice de dispersión de cada componente del vector aleatorio, así como el coeficiente de correlación. Al examinar los resultados en estas tablas llegamos a la conclusión de que los tests que hemos propuesto son capaces de detectar todas las alternativas tratadas, mientras que, en general, los otros tests no pueden detectar la mayoría de las alternativas, especialmente los tests basados en I_B y NI_B.

De acuerdo a los resultados de las Tablas 4.7 y 4.8 podemos decir que las potencias de $R_{n,w}(\hat{\theta}_n)$ y $S_{n,w}(\hat{\theta}_n)$ están muy cerca, siendo $R_{n,w}(\hat{\theta}_n)$ un poco más potente que $S_{n,w}(\hat{\theta}_n)$ en la mayoría de los casos tratados. El test basado en $W_n(\hat{\theta}_n)$ es algo menos potente que los otros dos tests propuestos, en las alternativas ensayadas.

Tabla 4.7: Resultados de simulación para la potencia. Alternativas: $BB(m; p_1, p_2, p_3)$, $BNB(\nu; \gamma_0, \gamma_1, \gamma_2)$ y $PPB(p; \theta, \lambda)$

Alternativa	$\frac{var(X_1)}{E(X_1)}$	$\frac{var(X_2)}{E(X_2)}$	ρ	$R_{n,(0,0)}$	$S_{n,(0,0)}$	W_n	T	I_B	NI_B
$BB(1; 0.41, 0.02, 0.01)$	0.590	0.980	0.026	0.863	0.881	0.829	0.111	0.000	0.000
$BB(1; 0.41, 0.03, 0.02)$	0.590	0.970	0.092	0.845	0.866	0.779	0.115	0.000	0.000
$BB(2; 0.61, 0.01, 0.01)$	0.390	0.990	0.080	0.988	0.953	0.948	0.931	0.004	0.004
$BB(1; 0.61, 0.03, 0.02)$	0.390	0.970	0.020	1.000	1.000	0.999	0.945	0.000	0.000
$BB(2; 0.71, 0.01, 0.01)$	0.290	0.990	0.064	1.000	0.996	1.000	1.000	0.000	0.000
$BNB(1; 0.92, 0.97, 0.01)$	1.920	1.970	0.486	0.928	0.890	0.524	0.843	0.622	0.974
$BNB(1; 0.97, 0.97, 0.01)$	1.970	1.970	0.493	0.933	0.884	0.526	0.860	0.633	0.975
$BNB(1; 0.97, 0.97, 0.02)$	1.970	1.970	0.493	0.936	0.901	0.518	0.846	0.616	0.975
$BNB(1; 0.98, 0.98, 0.01)$	1.980	1.980	0.495	0.944	0.906	0.530	0.855	0.607	0.980
$BNB(1; 0.99, 0.99, 0.01)$	1.990	1.990	0.498	0.932	0.893	0.510	0.850	0.585	0.973
$PPB(0.35; (0.2, 0.2, 0.1); (0.9, 1.0, 0.6))$	1.170	1.202	0.762	0.850	0.845	0.622	0.528	0.000	0.000
$PPB(0.37; (0.2, 0.2, 0.1); (0.9, 0.9, 0.8))$	1.178	1.178	0.956	0.748	0.754	0.642	0.458	0.001	0.001
$PPB(0.37; (0.2, 0.2, 0.1); (0.9, 1.0, 0.2))$	1.178	1.212	0.461	0.920	0.901	0.582	0.653	0.000	0.000
$PPB(0.40; (0.2, 0.2, 0.1); (0.9, 1.0, 0.3))$	1.190	1.226	0.566	0.964	0.941	0.556	0.759	0.000	0.000
$PPB(0.40; (0.2, 0.3, 0.1); (1.0, 0.9, 0.1))$	1.226	1.131	0.370	0.957	0.918	0.516	0.765	0.000	0.000

Tabla 4.8: Resultados de simulación para la potencia. Alternativas: $NTAB(\lambda;\lambda_1,\lambda_2,\lambda_3)$ y $SLB(\lambda_1,\lambda_2,\lambda_3)$

Alternativa	$\frac{var(X_1)}{E(X_1)}$	$\frac{var(X_2)}{E(X_2)}$	ρ	$R_{n,(0,0)}$	$S_{n,(0,0)}$	W_n	T	I_B	NI_B
$NTAB(0.41;0.01,0.01,0.98)$	1.990	1.990	0.995	0.934	0.922	0.853	0.681	0.001	0.880
$NTAB(0.50;0.01,0.01,0.98)$	1.990	1.990	0.995	0.931	0.916	0.831	0.685	0.001	0.875
$NTAB(0.70;0.01,0.01,0.98)$	1.990	1.990	0.995	0.937	0.928	0.777	0.727	0.001	0.893
$NTAB(0.71;0.01,0.01,0.98)$	1.990	1.990	0.995	0.945	0.943	0.796	0.736	0.000	0.903
$NTAB(0.75;0.01,0.01,0.98)$	1.990	1.990	0.995	0.954	0.937	0.776	0.732	0.000	0.911
$SLB(0.25,0.15,0.10)$	0.690	0.779	0.104	1.000	1.000	0.995	0.354	0.034	0.031
$SLB(5d/7,d/7,d/7)^*$	1.000	1.000	0.289	0.945	0.999	0.973	0.258	0.177	0.171
$SLB(3d/4,d/8,d/8)^*$	1.000	1.000	0.267	0.950	1.000	0.981	0.274	0.183	0.164
$SLB(7d/9,d/9,d/9)^*$	1.000	1.000	0.250	0.948	0.999	0.979	0.270	0.169	0.166
$SLB(0.51,0.01,0.02)$	0.668	0.981	0.098	1.000	1.000	0.999	0.429	0.046	0.041

$^*\ d = 1-\exp(-1)\approx 0.63212.$

Tabla 4.9: Tiempo de CPU promedio (en segundos).

	$n = 30$	$n = 50$	$n = 70$
$R_{n,(0,0)}$	39678.35	42275.29	44567.67
$S_{n,(0,0)}$	3019.81	7296.59	14202.57
W_n	1452.07	1807.28	2142.86

También comparamos los tests estadísticos $R_{n,w}(\hat{\theta}_n)$, $S_{n,w}(\hat{\theta}_n)$ y $W_n(\hat{\theta}_n)$ desde un punto de vista computacional. La Tabla 4.9 muestra el tiempo de CPU promedio consumido por cada uno de los tests estadísticos. En este aspecto, $W_n(\hat{\theta}_n)$ es mucho más eficiente que $R_{n,w}(\hat{\theta}_n)$ y $S_{n,w}(\hat{\theta}_n)$. Los cálculos fueron realizados en un servidor que tiene las siguientes características técnicas: Intel(R) Xeon(R) CPU X3430 @ 2.40GHz.

4.2. Conjuntos de datos reales

Ahora, aplicaremos los tests propuestos a dos conjuntos de datos reales. El primer conjunto de datos comprende el número de plantas de las especies *Lacistema aggregatum* y *Protium guianense* en cada uno de 100 cuadrantes contiguos. El primer autor que analizó estos datos y los presentó en detalle, fue Holgate (1966) [17], después fueron analizados por Gillings (1974) [9], Crockett (1979) [6], Loukas y Kemp (1986) [30], Kocherlakota y Kocherlakota (1992) [26] y también por Rayner y Best (1995) [37]. Crockett (1979) [6], Loukas y Kemp (1986) [30] y Rayner y Best (1995) [37] examinaron dichos datos para averiguar si correspondían a un modelo Poisson bivariante, concluyeron que la DPB no modelaba correctamente los datos mencionados.

El segundo conjunto de datos se refiere a la demanda de atención en sanidad en Australia, que fueron analizados por Karlis y Tsiamyrtzis (2008) [25]. Los datos se refieren a la encuesta de Salud en Australia en el periodo 1977–1978. El tamaño de la muestra es bastante grande ($n = 5190$). Estos autores utilizaron dos variables: el número de consultas con un médico o un especialista (X_1) y el número total de medicamentos prescritos y no prescritos utilizados en los últimos dos días previos a la encuesta (X_2). Los datos se analizaron bajo el supuesto de que (X_1, X_2) tenía una DPB.

La Tabla 4.10 muestra los $p-$valores para contrastar la bondad de ajuste a una

DPB para los dos conjuntos de datos reales empleando los tres tests estadísticos que hemos propuesto. Para ello, en el caso de $R_{n,w}(\hat{\theta}_n)$ y $S_{n,w}(\hat{\theta}_n)$, usamos $(a_1, a_2) \in \{(0,0),(1,0),(0,1)\}$. Para los datos de las plantas, todos los tests rechazaron la hipótesis nula, esto es, que los datos no son bien modelados por una DPB. Este resultado coincide con los obtenidos por los investigadores que analizaron previamente este conjunto de datos. A la misma conclusión se llega al emplear los datos de salud.

Tabla 4.10: Resultados para los conjuntos de datos reales.

	Plantas	Salud
$R_{n,(0,0)}$	0.002	0.000
$R_{n,(1,0)}$	0.004	0.000
$R_{n,(0,1)}$	0.004	0.000
$S_{n,(0,0)}$	0.002	0.002
$S_{n,(1,0)}$	0.006	0.000
$S_{n,(0,1)}$	0.008	0.000
W_n	0.042	0.000
$\hat{\theta}_n$	$(0.64000, 0.94000, 0.19852)$	$(0.30173, 1.21830, 0.12518)$

Capítulo 5

Expresiones matemáticas de los tests y algunos aspectos computacionales

Las expresiones matemáticas en las siguientes secciones son necesarias para implementar computacionalmente los tests estadísticos que hemos propuesto.

5.1. Cálculo del test estadístico $R_{n,w}(\hat{\theta}_n)$

Para w definido como en (4.1) y para $a_1, a_2 \in \mathbb{N}_0$, de (2.1), obtenemos

$$\begin{aligned} R_{n,w}(\hat{\theta}_n) &= \frac{1}{n} \int_0^1 \int_0^1 \left[\sum_{i=1}^n \left[u_1^{X_{1i}} u_2^{X_{2i}} - e^{\hat{\theta}_{1n}(u_1-1)+\hat{\theta}_{2n}(u_2-1)+\hat{\theta}_{3n}(u_1-1)(u_2-1)} \right] \right]^2 u_1^{a_1}\, u_2^{a_2}\, du \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \int_0^1 \int_0^1 u_1^{X_{1i}+X_{1j}+a_1} u_2^{X_{2i}+X_{2j}+a_2}\, du \\ &\quad - 2\, e^{\hat{\theta}_{3n}-\hat{\theta}_{1n}-\hat{\theta}_{2n}} \sum_{i=1}^n \int_0^1 u_1^{X_{1i}+a_1}\, e^{\left\{ \left(\hat{\theta}_{1n}-\hat{\theta}_{3n}\right) u_1 \right\}} \int_0^1 u_2^{X_{2i}+a_2}\, e^{\left\{ \left(\hat{\theta}_{2n}-\hat{\theta}_{3n}+\hat{\theta}_{3n} u_1\right) u_2 \right\}}\, du \\ &\quad + n\, e^{\left\{ 2\left(\hat{\theta}_{3n}-\hat{\theta}_{1n}-\hat{\theta}_{2n}\right) \right\}} \int_0^1 u_1^{a_1}\, e^{\left\{ 2\left(\hat{\theta}_{1n}-\hat{\theta}_{3n}\right) u_1 \right\}} \int_0^1 u_2^{a_2}\, e^{\left\{ 2\left(\hat{\theta}_{2n}-\hat{\theta}_{3n}+\hat{\theta}_{3n} u_1\right) u_2 \right\}}\, du. \end{aligned} \tag{5.1}$$

Para las integrales de (5.1) que son de la forma $\displaystyle\int_0^1 t^k \exp(rt)\, dt,\ \text{con}\ k \in \mathbb{N}_0,\ r > 0,$ integrando por partes, resulta

$$\int_0^1 t^k\, e^{rt}\, dt = \sum_{i=0}^{k} \frac{(-1)^i\, k!\exp(r)}{(k-i)!\, r^{i+1}} + \frac{(-1)^{k+1}\, k!}{r^{k+1}}\,,\ \forall\, k \in \mathbb{N}_0,\ r > 0. \tag{5.2}$$

Desarrollando (5.1) y empleando (5.2) con $t = u_2$, y para k y r apropiados, obtenemos

$$\begin{aligned}
R_{n,w}(\hat{\theta}_n) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{(X_{1i}+X_{1j}+a_1+1)(X_{2i}+X_{2j}+a_2+1)}\\
&- 2\exp(\hat{\theta}_{3n}-\hat{\theta}_{1n}-\hat{\theta}_{2n})\sum_{i=1}^{n}(X_{2i}+a_2)!\left\{\sum_{k=0}^{X_{2i}+a_2}\frac{(-1)^k\ \exp(\hat{\theta}_{2n}-\hat{\theta}_{3n})}{(X_{2i}+a_2-k)!}\right.\\
&\times\left.\int_0^1\frac{u_1^{X_{1i}+a_1}\exp\{\hat{\theta}_{1n}u_1\}\ du_1}{(\hat{\theta}_{2n}-\hat{\theta}_{3n}+\hat{\theta}_{3n}u_1)^{k+1}} + (-1)^{X_{2i}+a_2+1}\int_0^1\frac{u_1^{X_{1i}+a_1}\exp\{(\hat{\theta}_{1n}-\hat{\theta}_{3n})u_1\}}{(\hat{\theta}_{2n}-\hat{\theta}_{3n}+\hat{\theta}_{3n}u_1)^{X_{2i}+a_2+1}}du_1\right\}\\
&+ n\, a_2!\exp\{2(\hat{\theta}_{3n}-\hat{\theta}_{1n}-\hat{\theta}_{2n})\}\left\{\sum_{k=0}^{a_2}\frac{(-1)^k\exp\{2(\hat{\theta}_{2n}-\hat{\theta}_{3n})\}}{(a_2-k)!\ 2^{k+1}}\right.\\
&\times\left.\int_0^1\frac{u_1^{a_1}\exp(2\hat{\theta}_{1n}u_1)}{(\hat{\theta}_{2n}-\hat{\theta}_{3n}+\hat{\theta}_{3n}u_1)^{k+1}}du_1 + \frac{(-1)^{a_2+1}}{2^{a_2+1}}\int_0^1\frac{u_1^{a_1}\exp\{2(\hat{\theta}_{1n}-\hat{\theta}_{3n})u_1\}}{(\hat{\theta}_{2n}-\hat{\theta}_{3n}+\hat{\theta}_{3n}u_1)^{a_2+1}}du_1\right\}.
\end{aligned} \tag{5.3}$$

Para las integrales de (5.3) que son de la forma $\displaystyle\int_0^1 \frac{t^k\exp(rt)}{(b+ct)^m}\, dt$, con $r, b > 0$, $c \geq 0,\ k \in \mathbb{N}_0,\ m \in \mathbb{N}$, integrando por partes, se logra:

- Para $c = 0$, salvo constantes, se obtiene una integral como la dada por (5.2)

- Para $c>0$, haciendo el cambio de variables $y=b+ct$, se obtiene

$$
\begin{aligned}
\int_0^1 \frac{t^k \exp(rt)}{(b+ct)^m}\,dt &= \frac{\exp\left(-\frac{rb}{c}\right)}{c^{k+1}} \int_b^{b+c} \frac{(y-b)^k \exp\left(\frac{r}{c}y\right)}{y^m}\,dy \\
&= \frac{\exp\left(-\frac{rb}{c}\right)}{c^{k+1}} \sum_{i=0}^{k} \binom{k}{i} (-b)^{k-i} \int_b^{b+c} \frac{y^i \exp\left(\frac{r}{c}y\right)}{y^m}\,dy \\
&= \frac{\exp\left(-\frac{rb}{c}\right)}{c^{k+1}} \sum_{i=0}^{k} \binom{k}{i} (-b)^{k-i} \begin{cases} \displaystyle\int_b^{b+c} \frac{\exp\left(\frac{r}{c}y\right)}{y^{m-i}}\,dy, & \text{si } i<m, \\ \displaystyle\int_b^{b+c} y^{i-m} \exp\left(\frac{r}{c}y\right)\,dy, & \text{si } i\geq m. \end{cases}
\end{aligned}
\tag{5.4}
$$

Para las integrales de (5.4), integrando nuevamente por partes, resulta

$$
\begin{aligned}
&\int_b^{b+c} y^{i-m} \exp\left(\frac{r}{c}y\right)\,dy \\
&\quad = (i-m)! \exp\left(\frac{rb}{c}\right) \sum_{j=0}^{i-m} \frac{(-1)^j}{(i-m-j)!} \left(\frac{c}{r}\right)^{j+1} \left\{\exp(r)(b+c)^{i-m-j} - b^{i-m-j}\right\}.
\end{aligned}
\tag{5.5}
$$

$$
\begin{aligned}
\int_b^{b+c} \frac{\exp(\frac{r}{c}y)}{y^{m-i}}\,dy &= \frac{1}{(m-i-1)!} \left[\left(\frac{r}{c}\right)^{m-i-1} \int_{\frac{rb}{c}}^{\frac{r(b+c)}{c}} \frac{\exp(t)}{t}dt + \exp\left(\frac{rb}{c}\right) \right. \\
&\left. \times \begin{cases} 0, & \text{si } m-i=1, \\ \displaystyle\sum_{j=1}^{m-i-1} \frac{(m-i-j-1)!\; r^{j-1}}{c^{j-1}\{b(b+c)\}^{m-i-j}} \left\{(b+c)^{m-i-j} - b^{m-i-j}\exp(r)\right\}, & \text{si } m-i\geq 2. \end{cases} \right],
\end{aligned}
\tag{5.6}
$$

Así, $R_{n,w}(\hat{\theta}_n)$ se puede obtener de las ecuaciones (5.2), (5.4), (5.5) y (5.6).

Claramente, (5.6) no tiene una forma cerrada, pues contiene a la integral $\int \frac{\exp(t)}{t}dt$, la cual es muy sensible para ciertas combinaciones de los valores r, b y c, de modo que el cálculo computacional de dicha integral sufre problemas de redondeo o diverge y por lo tanto no podemos obtener su valor para ciertas combinaciones de r, b y c, lo cual nos lleva a quedar sin poder calcular $R_{n,w}(\hat{\theta}_n)$.

Por razones computacionales, la siguiente representación

$$R_{n,w}(\hat{\theta}_n) = n \sum_{i,j,k,l=0}^{\infty} \frac{\Big(p_n(i,j) - p(i,j;\hat{\theta}_n)\Big)\Big(p_n(k,l) - p(k,l;\hat{\theta}_n)\Big)}{(i+k+a_1+1)(j+l+a_2+1)} \tag{5.7}$$

resultó ser más apropiada, donde $p(i,j;\theta) = P_\theta(X_1 = i, X_2 = j)$ y $p_n(i,j)$ es la frecuencia relativa del par (i,j), dada por

$$p_n(i,j) = \frac{1}{n}\sum_{k=1}^{n} I\{X_{1k} = i, X_{2k} = j\}.$$

Además, (5.7) es menos restrictiva, pues permite que $a_1 > -1$ y $a_2 > -1$.

Un truncamiento de las cuatro series infinitas en $M + 15$ arrojó valores suficientemente precisos del test estadístico y un buen rendimiento de la subrutina correspondiente, donde $M = \max\{X_{1(n)}, X_{2(n)}\}$, $X_{k(n)} = \max_{1\le i\le n} X_{ki}$, $k = 1, 2$.

5.2. Cálculo del test estadístico $S_{n,w}(\hat{\theta}_n)$

Para w definido como en (4.1), y para $a_1 > -1$ y $a_2 > -1$, de (2.6), obtenemos

$$\begin{aligned}
&S_{n,w}(\hat{\theta}_n)\\
&= \frac{1}{n}\int_0^1\int_0^1 \left[\sum_{i=1}^{n}\Big\{X_{1i}I\{X_{1i}\ge 1\}u_1^{X_{1i}-1}u_2^{X_{2i}} - \{\hat{\theta}_{1n} + \hat{\theta}_{3n}(u_2-1)\}u_1^{X_{1i}}u_2^{X_{2i}}\Big\}\right]^2 u_1^{a_1}u_2^{a_2}du\\
&\quad + \frac{1}{n}\int_0^1\int_0^1 \left[\sum_{i=1}^{n}\Big\{X_{2i}I\{X_{2i}\ge 1\}u_1^{X_{1i}}u_2^{X_{2i}-1} - \{\hat{\theta}_{2n} + \hat{\theta}_{3n}(u_1-1)\}u_1^{X_{1i}}u_2^{X_{2i}}\Big\}\right]^2 u_1^{a_1}u_2^{a_2}du\\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(S_{1ij} + S_{2ij}\right),
\end{aligned}$$

donde

$$\begin{aligned}
S_{1ij} = {}& \frac{X_{1i}\,I\{X_{1i}\ge 1\}\,X_{1j}\,I\{X_{1j}\ge 1\}}{(X_{1ij}-1)(X_{2ij}+1)} + \frac{(\hat{\theta}_{1n}-\hat{\theta}_{3n})^2}{(X_{1ij}+1)(X_{2ij}+1)}\\
&- \frac{\hat{\theta}_{3n}\Big(X_{1i}\,I\{X_{1i}\ge 1\} + X_{1j}\,I\{X_{1j}\ge 1\}\Big)}{X_{1ij}\,(X_{2ij}+2)} + \frac{2\,\hat{\theta}_{3n}(\hat{\theta}_{1n}-\hat{\theta}_{3n})}{(X_{1ij}+1)(X_{2ij}+2)}\\
&- \frac{(\hat{\theta}_{1n}-\hat{\theta}_{3n})\Big(X_{1i}\,I\{X_{1i}\ge 1\} + X_{1j}\,I\{X_{1j}\ge 1\}\Big)}{X_{1ij}\,(X_{2ij}+1)} + \frac{\hat{\theta}_{3n}^{\,2}}{(X_{1ij}+1)(X_{2ij}+3)},
\end{aligned}$$

$$
\begin{aligned}
S_{2ij} = {} & \frac{X_{2i}\, I\{X_{2i} \geq 1\}\, X_{2j}\, I\{X_{2j} \geq 1\}}{(X_{1ij}+1)(X_{2ij}-1)} + \frac{(\hat{\theta}_{2n} - \hat{\theta}_{3n})^2}{(X_{1ij}+1)(X_{2ij}+1)} \\
& - \frac{\hat{\theta}_{3n}\Big(X_{2i}\, I\{X_{2i} \geq 1\} + X_{2j}\, I\{X_{2j} \geq 1\}\Big)}{(X_{1ij}+2)\, X_{2ij}} + \frac{2\,\hat{\theta}_{3n}(\hat{\theta}_{2n} - \hat{\theta}_{3n})}{(X_{1ij}+2)(X_{2ij}+1)} \\
& - \frac{(\hat{\theta}_{2n} - \hat{\theta}_{3n})\Big(X_{2i}\, I\{X_{2i} \geq 1\} + X_{2j}\, I\{X_{2j} \geq 1\}\Big)}{(X_{1ij}+1)\, X_{2ij}} + \frac{\hat{\theta}_{3n}^{\,2}}{(X_{1ij}+3)(X_{2ij}+1)}\,,
\end{aligned}
$$

con $X_{1ij} = X_{1i} + X_{1j} + a_1$ y $X_{2ij} = X_{2i} + X_{2j} + a_2$, $1 \leq i, j \leq n$.

5.3. Cálculo del test estadístico $W_n(\hat{\theta}_n)$

De (3.1) y (3.2), obtenemos

$$
\begin{aligned}
W_n(\hat{\theta}_n) = {} & \sum_{r,s=0}^{M} \left[\sum_{i=1}^{n} \Big\{ (r+1)p_n(r+1,s) - (\hat{\theta}_{1n} - \hat{\theta}_{3n})p_n(r,s) - \hat{\theta}_{3n} p_n(r,s-1) \Big\} \right]^2 \\
& + \sum_{r,s=0}^{M} \left[\sum_{i=1}^{n} \Big\{ (s+1)p_n(r,s+1) - (\hat{\theta}_{2n} - \hat{\theta}_{3n})p_n(r,s) - \hat{\theta}_{3n} p_n(r-1,s) \Big\} \right]^2,
\end{aligned}
$$

donde $M = \text{máx}\{X_{1(n)}, X_{2(n)}\}$, $X_{k(n)} = \text{máx}_{1 \leq i \leq n} X_{ki}$, $k = 1, 2$.

Capítulo 6

Extensiones

6.1. El caso de dos variables Poisson independientes

El caso de dos variables Poisson independientes ocurre cuando $\theta_3 = 0$, que fue excluido puesto que, según la definición dada en la Sección 1.1, $\theta_3 > 0$. La razón para no considerar este importante caso es que para que la aproximación bootstrap funcione, necesitamos que θ sea un punto interior del espacio paramétrico Θ.

Si estamos interesados en un test de bondad de ajuste para un modelo Poisson independiente, entonces podemos, o bien, considerar el test estadístico $R_{n,w}(\hat{\theta}_{0n})$, $S_{n,w}(\hat{\theta}_{0n})$ o $W_n(\hat{\theta}_{0n})$, con $\hat{\theta}_{0n} = (\hat{\theta}_{1n}, \hat{\theta}_{2n}, 0)$, es decir, el mismo test estadístico como antes con $\hat{\theta}_{3n}$ fijo e igual a su valor poblacional bajo la hipótesis nula, $\theta_3 = 0$, o bien, podemos usar el Teorema de Raikov y aplicar un test de bondad de ajuste para la distribución Poisson univariante a la suma de los componentes.

Los resultados de los Capítulos 2 y 3 seguirán siendo válidos para estos tests estadísticos.

6.2. El caso general d-variante

Hasta ahora hemos asumido que los datos son bivariantes. En esta sección mostramos que los tests que hemos propuesto en esta memoria se pueden extender de manera natural al caso d-variante general, $d \in \mathbb{N}$. Sea

$$X_1 = Y_1 + Y_{d+1},\ X_2 = Y_2 + Y_{d+1},\ \ldots,\ X_d = Y_d + Y_{d+1},$$

donde $Y_1, Y_2, \ldots, Y_d, Y_{d+1}$ son v.a. Poisson (univariantes) mutuamente independientes con medias $\theta_1' = \theta_1 - \theta_{d+1} > 0, \ldots, \theta_d' = \theta_d - \theta_{d+1} > 0$ y $\theta_{d+1} \geq 0$, respectivamente.

La distribución conjunta del vector $(X_1, X_2, \ldots, X_d)$ es llamada distribución Poisson d-variante con parámetro $\theta = (\theta_1, \theta_2, \ldots, \theta_d, \theta_{d+1})$ (ver, por ejemplo, Johnson, Kotz y Balakrishnan (1997, p. 139) [19]).

La fgp conjunta de $(X_1, X_2, \ldots, X_d)$ está dada por

$$g(u;\theta) = \exp\left\{\sum_{i=1}^{d} \theta_i (u_i - 1) + \theta_{d+1}\left(\prod_{i=1}^{d} u_i - \sum_{i=1}^{d} u_i + d - 1\right)\right\}, \tag{6.1}$$

$\forall u \in \mathbb{R}^d$.

Claramente, el test basado en el estadístico $R_{n,w}(\hat{\theta}_n)$ se puede aplicar para contrastar

$$H_{0d} : (X_1, X_2, \ldots, X_d) \text{ tiene una distribución Poisson } d\text{-variante},$$

para cualquier d fijo.

El siguiente resultado muestra la extensión d-variante de la Proposición 2.1.1, lo cual nos permitirá proponer una extensión de los tests basados en $S_{n,w}(\hat{\theta}_n)$ y en $W_n(\hat{\theta}_n)$.

Proposición 6.2.1 *Sea $G_d = \{g : [0,1]^d \to \mathbb{R}$, tal que g es una fgp y $\frac{\partial}{\partial u_i} g(u_1, \ldots, u_d)$, existen $\forall u \in [0,1]^d$, $1 \leq i \leq d\}$. Sea $g(u;\theta)$ como definida en (6.1). Entonces $g(u;\theta)$ es la única fgp en G_d que satisface el siguiente sistema*

$$D_{id}(u;\theta) = 0, \quad 1 \leq i \leq d, \quad \forall u \in [0,1]^d, \tag{6.2}$$

donde

$$D_{id}(u;\theta) = \frac{\partial}{\partial u_i} g(u) - \left\{\theta_i + \theta_{d+1}\left(\prod_{j \neq i} u_j - 1\right)\right\} g(u), \quad 1 \leq i \leq d.$$

Demostración Consideremos el vector aleatorio $(X_1, X_2, \ldots, X_d) \in \mathbb{N}_0^d$, cuya fgp está dada por $g(u_1, u_2, \ldots, u_d) = E\left(u_1^{X_1} u_2^{X_2} \cdots u_d^{X_d}\right)$. Entonces, para $i = 1$ en (6.2)

$$\frac{\partial}{\partial u_1} \log g(u_1, u_2, \ldots, u_d) = \theta_1 + \theta_{d+1}\left(\prod_{j \neq 1} u_j - 1\right). \tag{6.3}$$

Integrando (6.3) sobre u_1, obtenemos

$$
\begin{aligned}
g(u_1, u_2, \ldots, u_d) &= \exp\left\{\phi_1(u_2, u_3, \ldots, u_d) + \theta_1 u_1 + \theta_{d+1}\left(\prod_{j=1}^{d} u_j - u_1\right)\right\},\\
&= \exp\left\{\varphi_1(u_2, u_3, \ldots, u_d) + \theta_1(u_1 - 1) + \theta_{d+1}\left(\prod_{j=1}^{d} u_j - \sum_{j=1}^{d} u_j + d - 1\right)\right\},
\end{aligned}
$$

donde $\varphi_1(u_2, u_3, \ldots, u_d) = \phi_1(u_2, u_3, \ldots, u_d) + \theta_1 + \theta_{d+1}\left(\sum_{j=2}^{d} u_j - d + 1\right)$.

Procediendo similarmente, de (6.2), para $i = 2, 3, \ldots, d$, obtenemos

$$
\begin{aligned}
g(u_1, u_2, \ldots, u_d) &= \exp\left\{\phi_2(u_1, u_3, \ldots, u_d) + \theta_2 u_2 + \theta_{d+1}\left(\prod_{j=1}^{d} u_j - u_2\right)\right\},\\
&= \exp\left\{\varphi_2(u_1, u_3, \ldots, u_d) + \theta_2(u_2 - 1) + \theta_{d+1}\left(\prod_{j=1}^{d} u_j - \sum_{j=1}^{d} u_j + d - 1\right)\right\},
\end{aligned}
$$

donde $\varphi_2(u_1, u_3, \ldots, u_d) = \phi_2(u_1, u_3, \ldots, u_d) + \theta_2 + \theta_{d+1}\left(\sum_{j=1, j\neq 2}^{d} u_j - d + 1\right)$.

Así, sucesivamente hasta que

$$
\begin{aligned}
g(u_1, u_2, \ldots, u_d) &= \exp\left\{\phi_d(u_1, u_2, \ldots, u_{d-1}) + \theta_d u_d + \theta_{d+1}\left(\prod_{j=1}^{d} u_j - u_d\right)\right\},\\
&= \exp\left\{\varphi_d(u_1, u_2, \ldots, u_{d-1}) + \theta_d(u_d - 1) + \theta_{d+1}\left(\prod_{j=1}^{d} u_j - \sum_{j=1}^{d} u_j + d - 1\right)\right\},
\end{aligned}
$$

donde $\varphi_d(u_1, u_2, \ldots, u_{d-1}) = \phi_d(u_1, u_2, \ldots, u_{d-1}) + \theta_d + \theta_{d+1}\left(\sum_{j=1}^{d-1} u_j - d + 1\right)$.

Por lo tanto, necesariamente debe ocurrir que

$$\varphi_1(u_2, u_3, \ldots, u_d) = \sum_{j=2}^{d} \theta_j(u_j - 1),$$

$$\varphi_2(u_1, u_3, \ldots, u_d) = \sum_{\substack{j=1 \\ j\neq 2}}^{d} \theta_j(u_j - 1),$$

$$\vdots$$

$$\varphi_d(u_1, u_2, \ldots, u_{d-1}) = \sum_{j=1}^{d-1} \theta_j(u_j - 1),$$

en otras palabras, la fgp de la DP d-variante es la única solución de (6.2). $\square$

De la Proposición 1.2.2, $g(u)$ y sus derivadas pueden ser estimadas consistentemente por medio de la fgpe y las derivadas de la fgpe, respectivamente.

Así, si H_{0d} fuera cierta, entonces las funciones

$$D_{in}(u;\hat{\theta}_n) = \frac{\partial}{\partial u_i} g_n(u_1, u_2, \ldots, u_d) - \left\{ \hat{\theta}_{i,n} + \hat{\theta}_{d+1,n} \left(\prod_{j\neq i} u_j - 1 \right) \right\} g_n(u_1, u_2, \ldots, u_d),$$

$1 \leq i \leq d$, deberían estar próximas a 0, donde $\hat{\theta}_n$ es un estimador consistente de θ.

Por lo tanto, para probar H_{0d} podríamos considerar el siguiente test estadístico

$$S_{d,n,w}(\hat{\theta}_n) = n \int_{[0,1]^d} \left\{ D_{1n}^2(u;\hat{\theta}_n) + \cdots + D_{dn}^2(u;\hat{\theta}_n) \right\} w(u)\, du,$$

donde $w(u)$ es una función de peso medible, no negativa, con integral finita en $[0,1]^d$.

Haciendo las modificaciones correspondientes, se pueden conseguir resultados similares a los establecidos en el Capítulo 2.

Para el test estadístico basado en $W_n(\hat{\theta}_n)$, consideremos que se verifican las hipótesis de la Proposición 6.2.1.

Sea $(X_1, \ldots, X_d) \in \mathbb{N}_0^d$ un vector aleatorio y sea $g(u_1, \ldots, u_d) = E\left(u_1^{X_1} u_2^{X_2} \cdots u_d^{X_d}\right)$ su fgp, luego, por definición

$$g(u) = \sum_{r_1, r_2, \ldots, r_d \geq 0} u_1^{r_1} u_2^{r_2} \cdots u_d^{r_d} P(r_1, r_2, \ldots, r_d),$$

donde $P(r_1, r_2, \ldots, r_d) = P(X_1 = r_1, X_2 = r_2, \ldots, X_d = r_d)$.

De lo anterior y de (6.2), podemos escribir

$$\begin{aligned} D_{1d}(u;\theta) = \sum_{r_1,r_2,\ldots,r_d\geq 0} \Big\{ &(r_1+1)P(r_1+1, r_2, r_3, \ldots, r_d) - (\theta_1 - \theta_{d+1})P(r_1, r_2, \ldots, r_d) \\ &-\theta_{d+1}P(r_1, r_2-1, r_3-1, \ldots, r_d-1)\Big\} u_1^{r_1} u_2^{r_2} \cdots u_d^{r_d}, \\ D_{2d}(u;\theta) = \sum_{r_1,r_2,\ldots,r_d\geq 0} \Big\{ &(r_2+1)P(r_1, r_2+1, r_3, \ldots, r_d) - (\theta_2 - \theta_{d+1})P(r_1, r_2, \ldots, r_d) \\ &-\theta_{d+1}P(r_1-1, r_2, r_3-1, \ldots, r_d-1)\Big\} u_1^{r_1} u_2^{r_2} \cdots u_d^{r_d}, \\ &\vdots \\ D_{dd}(u;\theta) = \sum_{r_1,r_2,\ldots,r_d\geq 0} \Big\{ &(r_d+1)P(r_1, r_2, \ldots, r_{d-1}, r_d+1) - (\theta_d - \theta_{d+1})P(r_1, r_2, \ldots, r_d) \\ &-\theta_{d+1}P(r_1-1, r_2-1, \ldots, r_{d-1}-1, r_d)\Big\} u_1^{r_1} u_2^{r_2} \cdots u_d^{r_d}. \end{aligned}$$

Consideremos ahora las versiones empíricas $D_{1n}(u;\hat{\theta}_n), D_{2n}(u;\hat{\theta}_n), \ldots, D_{dn}(u;\hat{\theta}_n)$, de las ecuaciones anteriores.

Si H_{0d} fuera cierta entonces $D_{1n}(u;\hat{\theta}_n), D_{2n}(u;\hat{\theta}_n), \ldots, D_{dn}(u;\hat{\theta}_n)$ deberían ser próximas a 0, $\forall u \in [0,1]^d$. Esta proximidad a cero la podemos interpretar como ya lo hicimos al inicio del Capítulo 3, para el caso bivariante. Para ello observemos que

$$D_{in}(u;\hat{\theta}_n) = \sum_{r_1,r_2,\ldots,r_d\geq 0} b_i(r_1, r_2, \ldots, r_d; \hat{\theta}_n) u_1^{r_1} u_2^{r_2} \cdots u_d^{r_d}, \quad 1 \leq i \leq d$$

donde

$$\begin{aligned} b_1(r_1, r_2, \ldots, r_d; \hat{\theta}_n) &= (r_1+1)p_n(r_1+1, r_2, r_3, \ldots, r_d) - (\hat{\theta}_{1n} - \hat{\theta}_{d+1,n})p_n(r_1, r_2, \ldots, r_d) \\ &\quad - \hat{\theta}_{d+1,n}p_n(r_1, r_2-1, r_3-1, \ldots, r_d-1), \\ b_2(r_1, r_2, \ldots, r_d; \hat{\theta}_n) &= (r_2+1)p_n(r_1, r_2+1, r_3, \ldots, r_d) - (\hat{\theta}_{2n} - \hat{\theta}_{d+1,n})p_n(r_1, r_2, \ldots, r_d) \\ &\quad - \hat{\theta}_{d+1,n}p_n(r_1-1, r_2, r_3-1, \ldots, r_d-1), \\ &\vdots \\ b_d(r_1, r_2, \ldots, r_d; \hat{\theta}_n) &= (r_d+1)p_n(r_1, r_2, \ldots, r_{d-1}, r_d+1) - (\hat{\theta}_{dn} - \hat{\theta}_{d+1,n})p_n(r_1, r_2, \ldots, r_d) \\ &\quad - \hat{\theta}_{d+1,n}p_n(r_1-1, r_2-1, \ldots, r_{d-1}-1, r_d), \end{aligned}$$

y

$$p_n(r_1, r_2, \ldots, r_d) = \frac{1}{n} \sum_{k=1}^{n} I\{X_{1k} = r_1, X_{2k} = r_2, \ldots, X_{dk} = r_d\},$$

es la frecuencia relativa empírica de la d-tupla $(r_1, r_2, \ldots, r_d)$.

Por lo tanto, $D_{in}(u; \hat{\theta}_n) = 0$, $\forall u \in [0,1]^d$, $1 \leq i \leq d$, sí y sólo si los coeficientes de $u_1^{r_1} u_2^{r_2} \cdots u_d^{r_d}$ en las expansiones anteriores son nulos $\forall r_1, r_2, \ldots, r_d \geq 0$. Esto nos lleva a considerar el siguiente estadístico para contrastar H_{0d}:

$$\begin{aligned} W_{d,n}(\hat{\theta}_n) &= \sum_{r_1, r_2, \ldots, r_d \geq 0} \left\{ \sum_{i=1}^{d} b_i(r_1, r_2, \ldots, r_d; \hat{\theta}_n)^2 \right\} \\ &= \sum_{r_1, r_2, \ldots, r_d = 0}^{M} \left\{ \sum_{i=1}^{d} b_i(r_1, r_2, \ldots, r_d; \hat{\theta}_n)^2 \right\}, \end{aligned}$$

donde $M = \max\{X_{1(n)}, X_{2(n)}, \ldots, X_{d(n)}\}$, $X_{k(n)} = \max_{1 \leq j \leq n} X_{kj}$, $1 \leq k \leq d$.

Haciendo los cambios respectivos, se pueden conseguir resultados similares a los establecidos en el Capítulo 3.

Bibliografía

[1] Athreya, K.B. y Lahiri, S.N. (2006). *Measure Theory and Probability Theory.* Springer

[2] Baringhaus, L., Gürtler, N. y Henze, N. (2000). Weighted integral test statistics and components of smooth tests of fit, *Australian & New Zealand Journal of Statistics*, **42**, 179–192.

[3] Baringhaus, L. y Henze, N. (1992). A goodness of fit test for the Poisson distribution based on the empirical generating function, *Statistics & Probability Letters*, **13**, 269–274.

[4] Berkhout, P. y Plug, E. (2004). A bivariate Poisson count data model using conditional probabilities, *Statistica Neerlandica*, **58**, 349–364.

[5] Bermúdez, L. (2009). A priori ratemaking using bivariate Poisson regression models, *Insurance: Mathematics and Economics*, **44**, 135–141.

[6] Crockett, N. G., (1979). A quick test of fit of a bivariate distribution. In D. McNeil (ed.), *Interactive Statistics*, 185–191. Amsterdam: North-Holland.

[7] Dai, B. T., Chua, F. y Lim, E. P. (2012). Structural Analysis in Multi-Relational Social Networks, *Research Collection School of Information Systems (Open Access).*

[8] Feuerverger, A. (1989). On the empirical saddlepoint approximation, *Biometrika*, **76**, 457–464.

[9] Gillings, D. B. (1974). Some further results for bivariate generalizations of the Neyman type A distribution, *Biometrics*, **30**, 619–628.

[10] Gregory, G. (1977). Large sample theory for U-statistics and tests of fit, *The Annals of Statistics*, **5**, 110–123.

[11] Gürtler, N. y Henze, N. (2000). Recent and classical goodness-of-fit tests for the Poisson distribution, *Journal of Statistical Planning and Inference*, **90**, 207–225.

[12] Haight, F. A. (1967). *Handbook of the Poisson distribution*, New York: John Wiley & Sons.

[13] Hamdan, M. A. (1972). Estimation in the Truncated Bivariate Poisson Distribution, *Technometrics*, **14**, 37–45.

[14] Hamdan, M. A. y Al-Bayyati, H. A. (1969). A note on the bivariate Poisson distribution, *The American Statistician*, **23**, No. 4, 32–33.

[15] Ho, L. L. y Singer, J. M. (2001). Generalizad Least Squares Methods for Bivariate Poisson Regression, *Communications in Statistics - Theory and Methods*, **30**(2), 263–277.

[16] Holgate, P. (1964). Estimation for the bivariate Poisson distribution, *Biometrika*, **51**, 241–245.

[17] Holgate, P. (1966). Bivariate generalizations of Neyman's Type A distribution, *Biometrika*, **53**, 241–245.

[18] Johnson, N. L. y Kotz, S. (1969). *Distributions in Statistics: Discrete Distributions*, New York: John Wiley & Sons.

[19] Johnson, N. L., Kotz, S. y Balakrishnan, N. (1997). *Discrete Multivariate Distributions*, New York: John Wiley & Sons.

[20] Jung, R. C. y Winkelmann, R. (1993). Two Aspects of Labor Mobility: A Bivariate Poisson Regression Approach, *Empirical Economics*, **18**, 543–556.

[21] Karlis, D. y Ntzoufras, I. (2000). On modelling soccer data, *Student*, **3**, 229–244.

[22] Karlis, D. y Ntzoufras, I. (2003a). Analysis of sports data by using bivariate Poisson models, *The Statistician*, **52**, Part 3, 381-393.

[23] Karlis, D. y Ntzoufras, I. (2003b). Bayesian and Non-Bayesian Analysis of Soccer Data using Bivariate Poisson Regression Models,

http://www.docstoc.com/docs/39143794/Bayesianand-Non-Ba-yesian-Analysis-of-Soccer-Data-using-Bivariate.

[24] Karlis, D. y Ntzoufras, I. (2005). Bivariate Poisson and Diagonal Inflated Bivariate Poisson Regression Models in R, *Journal of Statistical Software*, **14** (10), 1–36.

[25] Karlis, D. y Tsiamyrtzis, P. (2008). Exact Bayesian modeling for bivariate Poisson data and extensions, *Statistics and Computing*, **18**, 27–40.

[26] Kocherlakota, S. y Kocherlakota, K. (1992). *Bivariate Discrete Distributions*, New York: Marcel Dekker.

[27] Kundu, S., Majumdar, S. y Mukherjee, K. (2000). Central limits theorems revisited, *Statistics & Probability Letters*, **47**, 265–275.

[28] Lakshminarayana, J., Pandit, S. N. N. y Rao, S. (1999). On a bivariate Poisson distribution, *Communications in Statistics - Theory and Methods*, **28**(2), 267–276.

[29] Lehmann, E.L. y Romano, J.P. (2005). *Testing Statistical Hypotheses*, Springer.

[30] Loukas, S. y Kemp, C. D. (1986). The Index of Dispersion Test for the Bivariate Poisson Distribution, *Biometrics*, **42**, 941–948.

[31] Loukas, S., Kemp, C. D. y Papageorgiou, H. (1986). Even point estimation for the bivariate Poisson distribution, *Biometrika*, **73**, 222-223.

[32] Maher, M. J. (1982). Modelling association football scores, *Statistica Neerlandica*, **36**, 109–118.

[33] Nakamura, M. y Pérez-Abreu, V. (1993). Use of an Empirical Probability Generating Function for Testing a Poisson Model, *Canadian Journal of Statistics*, **21**, 149–156.

[34] Papageorgiou, H. y Loukas, S. (1988). Conditional even point estimation for bivariate discrete distributions, *Communications in Statistics–Theory and Methods*, **17**, 3403–3412.

[35] Paul, S. R. y Ho, N. I. (1989). Estimation in the bivariate Poisson distribution and hypothesis testing concerning independence, *Communications in Statistics–Theory and Methods*, **18**, 1123–1133.

[36] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, 2009, URL http://www.R-project.org.

[37] Rayner, J. C. W y Best, D. J. (1995). Smooth Tests for the Bivariate Poisson Distribution, *Australian & New Zealand Journal of Statistics*, **37**, 233–245.

[38] Rayner, J. C. W, Thas, O. y Best, D. J. (2009). *Smooth Tests of Goodness of Fit*, John Wiley & Sons.

[39] Rue, H. y Salvesen, ∅. (2000). Prediction and retrospective analysis of soccer matches in a league, *Statistician*, **49**, 399–418.

[40] Rueda, R., Pérez-Abreu, V., O'Reilly, F. (1991). Goodness of fit for the Poisson distribution based on the probability generating function, *Communications in Statistics–Theory and Methods*, **A20**, 3093–3110.

[41] Sahai, H. y Khurshid, A. (1993). Confidence Intervals for the Mean of a Poisson Distribution: A Review, *Biometrical Journal*, **35**, 857–867.

[42] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

[43] Stuart, A. y Ord, J. K. (1987). *Kendall's advanced theory of statistics*, Londres: Charles Griffin & Cía. Ltda.

[44] van der Vaart, A.W. y Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*, New York: Springer-Verlag.