

Universidad de Sevilla
Facultad de Matemáticas

043
360 (70/0)

ESTIMACIÓN INSESGADA APLICADA A LA
APROXIMACIÓN Y CARACTERIZACIÓN
DE DISTRIBUCIONES

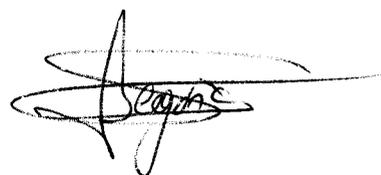
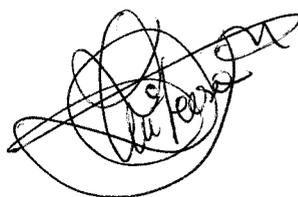
Visado en Sevilla, Diciembre de 2000.

Memoria dirigida por:

Prof. Dr. D. Fernando López Blázquez.

Prof. Dra. Dña. Teresa Gómez Gómez.

Memoria presentada para
optar al grado de Doctor
en Ciencias Matemáticas.



Fdo.: Begoña Salamanca Miño.

UNIVERSIDAD DE SEVILLA
FACULTAD DE MATEMÁTICAS
BIBLIOTECA

UNIVERSIDAD DE SEVILLA
NEGOCIADO DE TESIS

Queda registrado este Título de Doctor al
folio... 57 ... número... 174 ... del libro
correspondiente.

Sevilla, 13 DIC. 2000

El Jefe del Negociado.

Reina Raffello

UNIVERSIDAD DE SEVILLA

Depositado en el D^{to} de Estad. e Inv. Químicas
de la F. de Matemáticas

de esta Universidad desde el día 15-12-00
hasta el día 17 de Enero - 2001

Sevilla 18 de Enero del 2001

EL DIRECTOR DE

Martín

Índice

Introducción	4
1 Polinomios Ortogonales y la Familia Exponencial Natural	9
1.1 La familia NEF	9
1.2 Obtención de los polinomios ortogonales	12
1.2.1 Método 1. El método de Gram-Schmidt	12
1.2.2 Método 2. Determinantes de las matrices de momentos	14
1.2.3 Método 3. Método matricial	15
1.3 Estimación insesgada	16
1.4 Obtención de los coeficientes de Fourier	19
1.5 Propiedades límites de los polinomios ortogonales	22
1.6 Propiedades límites de los estimadores insesgados	26
1.7 Cotas inferiores para la varianza de estimadores insesgados	30
1.8 Caracterización del conjunto de las funciones estimable umvu	31
1.9 Comparación entre el emv y el umvue	36
1.10 Una extensión de la clase \mathcal{U}	37
2 Algoritmo para la obtención de estimadores insesgados en distribuciones de series de potencia	41
2.1 Introducción	41

2.2	Estimación insesgada en la familia de distribuciones en series de potencia.....	43
2.3	Polinomios ortonormales.....	49
2.4	Algoritmo.....	51
2.5	Ejemplos.....	53
2.6	Conclusiones.....	54
3	Aproximación de Poisson a la probabilidad Binomial	55
3.1	Introducción.....	55
3.2	Aproximación a la probabilidad binomial.....	56
3.3	Propiedades de las aproximaciones.....	63
4	Aproximación Binomial a la probabilidad Hipergeométrica	70
4.1	Introducción.....	70
4.2	Desarrollo de Krawtchouk para probabilidades hipergeométricas.....	71
4.3	Aproximación a la probabilidad hipergeométrica.....	75
4.4	Propiedades de la aproximación.....	79
5	Aproximación entre las probabilidades Hipergeométrica Negativa y Binomial Negativa	83
5.1	Introducción.....	83
5.2	Aproximación a la probabilidad hipergeométrica negativa.....	87
5.3	Propiedades de las aproximaciones.....	88
5.4	Probabilidades hipergeométricas negativas acumuladas.....	91
5.5	Conclusiones.....	94

6 Una propiedad martingala inversa que caracteriza a la familia exponencial nautural con función de varianza cuadrática	95
6.1 Introducción	95
6.2 Demostración	99
6.2.1 Estimación insesgada en las familias exponenciales	99
6.2.2 Demostración	103
6.3 Ejemplos	105
7 Una relación entre la esperanza condicionada de los records y la estimación insesgada en la distribución exponencial	108
7.1 Estimación insesgada en la distribución exponencial.....	108
7.2 Records.....	110
7.3 Regresión lineal de records	112
7.4 Regresión lineal de estadísticos ordenados.....	116
Bibliografía	118

Introducción

Los estimadores insesgados son utilizados en diversas áreas de la Estadística tales como en los modelos lineales, teoría de muestreo, inferencia paramétrica clásica, etc. Cabe mencionar también que algunas técnicas como el Jackknife o los modernos métodos de remuestreo como el Bootstrap tienen su origen en el intento de reducir el sesgo de los estimadores.

La presente memoria se centra en uno de los problemas más básicos en esta área, la obtención de estimadores insesgados de mínima varianza en familias paramétricas.

Trabajos pioneros son entre otros, los de Barakin (1949), Stein (1950), Washio et al. (1956), Tate (1959). La Teoría básica de la estimación insesgada puede encontrarse en libros tales como *The theory of statistical inference*, Zacks (1971, cap. 3) o *Theory of point estimation*, Lehman (1983, cap. 2)

Ésta teoría básica, se basa fundamentalmente, en los conocidos teoremas de Blackwell-Rao y Lehman-Scheffé. Sin embargo esta teoría no resuelve totalmente algunas cuestiones importantes:

- i) No nos permite resolver problemas sobre las propiedades de los estimadores, como pueden ser propiedades asintóticas o estudio de cotas para la varianza.
- ii) Resulta paradójico que para la obtención de un estimador insesgado de mínima varianza se necesite condicionar un insesgado a un estimador suficiente y completo.

Por estos motivos se hace necesario desarrollar una teoría más amplia que permita resolver estos problemas. Una posible solución es considerar los

estimadores como elementos de espacios de funciones de cuadrado integrable (\mathcal{L}^2). Estos espacios son ricos en propiedades matemáticas, por lo que se podrá obtener de una forma unificada muchas propiedades de los estimadores.

El primer problema que se puede resolver con esta teoría es la de calcular los estimadores insesgados en términos de las bases ortogonales de estos espacios. En las familias exponenciales es posible obtener bases relativamente sencillas formadas por un sistema de polinomios ortogonales. Los primeros trabajos en los que se encuentran los estimadores insesgados expresados en términos de polinomios ortogonales son los de Abbey y David (1970) y Morris (1982,1983).

En la memoria nos centraremos solo en la familia exponencial natural, aunque esta teoría se puede aplicar a otras familias de distribuciones, como por ejemplo en familias no regulares, Barranco Chamorro (1996), familias aunque regulares no exponenciales, López-Blázquez (2000) o en situaciones de censuramiento, López-Blázquez y Salamanca Miño (2000).

En el primer capítulo de la memoria, que es una introducción, presentamos los elementos básicos de esta teoría. El resultado más novedoso de este capítulo es la obtención de las posibles distribuciones límites de los estimadores insesgados en familias exponenciales. Este resultado extiende el de López-Blázquez y Castaño-Martínez (1999), que era específico para la familia exponencial con función de varianza cuadrática. Otro resultado interesante es la caracterización de las funciones paramétricas que admiten estimadores umvu.

Una de las ventajas de utilizar polinomios ortogonales es la obtención de estimadores insesgados de forma algorítmica. En el segundo capítulo se muestra como es esto posible para familias de series de potencia.

Esta familia fue introducida por Kosambi (1949) y Noak (1950). El procedimiento clásico para la obtención de estimadores insesgados fue obtenido por Roy y Mitra (1957). Si bien esta fórmula da lugar a una solución formal para la obtención de estimadores existen ciertas limitaciones, como pueden ser:

- i) La inestabilidad numérica que se produce cuando el estimador suficiente toma valores relativamente grandes.
- ii) No se obtiene ningún procedimiento para calcular la varianza de los estimadores.
- iii) El problema más importante, es que para la aplicación de éste procedimiento es necesario obtener la distribución del estadístico suficiente y completo.

Este último punto, puede ser realmente complicado sobre todo en el caso en el que se consideren familias de distribuciones truncadas. Basándose en esta fórmula, diversos autores como Joshi y Park (1974), Charalambides (1974), Jani (1977,1978), Kumar y Consul (1980) y Voinov (1986), obtuvieron expresiones para estimadores insesgados en ciertas subfamilias de interés, como la Binomial truncada, la Poisson decapitada y la binomial negativa truncada. Normalmente estas fórmulas están basadas en relaciones recurrentes que implican la utilización de números de Stirling Generalizados, que son específicos para cada una de las distribuciones.

En este capítulo se muestra un procedimiento algorítmico para la obtención de estimadores insesgados basados en los desarrollos ortogonales. Las ventajas de este algoritmo son:

- i) No es necesario obtener la distribución del estadístico suficiente y completo
- ii) No son necesarios números de Stirling Generalizados.
- iii) Es posible obtener un estimador de la varianza de los estimadores insesgados, lo cual nos posibilitará la realización de inferencias posteriores, como por ejemplo la obtención de intervalos de confianza.

La teoría de la estimación insesgada desarrollada en el primer capítulo, es muy rica en propiedades matemáticas. En los sucesivos capítulos veremos

como algunos de los resultados de esta teoría pueden utilizarse para resolver problemas que no son propiamente de inferencia estadística. Concretamente veremos unas aplicaciones a la aproximación y caracterización de funciones de distribución.

Así en el capítulo 3 estudiamos la aproximación de Poisson a la probabilidad binomial. Utilizando la propiedad de unicidad (casi seguro) de los estimadores insesgados hemos obtenido una relación exacta entre cualquier distribución binomial y cualquier distribución de Poisson.

La truncación de la serie obtenida y una elección adecuada de los parámetros permite obtener aproximaciones a la distribución binomial. Casos particulares de nuestra aproximación son la aproximación clásica de Poisson (1937), la aproximación de Kolmogorov y la de Burr (1973).

La teoría de la estimación insesgada permite también obtener diversas propiedades de estas aproximaciones. Es posible incluso mejorar algunas cotas que han sido ya propuestas en la literatura, tales como la de Simons y Johnson (1971), Vervaat (1970) y extender propiedades como las de Anderson y Samuel (1967) y Poor (1991).

Estudios similares han sido desarrollados en el capítulo 4, en el que se estudia la aproximación binomial a la probabilidad hipergeométrica y en el capítulo cinco, que se estudia la aproximación de la hipergeométrica negativa a la binomial negativa.

Destacamos nuevamente, que nuestros resultados se obtienen a partir de la unicidad de los estimadores insesgados. En el capítulo 4 hemos obtenido una mejora en las aproximaciones de Ord (1968) y Bennett (1965).

Las aproximaciones obtenidas en el capítulo 5, aún habiendo sido obtenidas bajo el mismo principio, pueden considerarse novedosas ya que no han sido encontrado resultados similares.

Los últimos dos capítulos de esta memoria ilustran como la teoría de la estimación insesgada puede utilizarse para la caracterización de distribuciones.

Los polinomios asociados a las distribuciones exponenciales satisfacen una propiedad martingala, ver Schoutens y Teugels (1998). La cuestión que nos

planteamos en el capítulo 6 es caracterizar aquellas familias de distribuciones exponenciales cuyos polinomios satisfacen una propiedad martingala inversa. El resultado fundamental obtenido es que tal propiedad solo es satisfecha por la familia exponencial natural con función de varianza cuadrática. La demostración de este resultado está basada en la caracterización de ciertas familias de funciones estimable umvu.

En el último capítulo, utilizaremos propiedades de la estimación insesgada en la distribución exponencial para la caracterización de distribuciones mediante la regresión lineal de estadísticos de orden y records. El origen de este problema se remonta a un trabajo de Ferguson (1967), en el cual se caracterizan las distribuciones con regresión lineal entre estadísticos ordenados adyacentes. El problema para el caso no adyacente permaneció abierto hasta 1997 (ver López-Blázquez y Moreno (1997)).

La caracterización dada por estos autores requiere ciertas hipótesis de suavidad sobre las funciones de distribución implicadas. El resultado principal que obtenemos en el capítulo 7 muestra como pueden ser obtenidos estos resultados sin necesidad de tales hipótesis de suavidad.

Capítulo 1

Polinomios Ortogonales y la familia Exponencial Natural

1.1 La familia NEF

Sea ν una medida σ -finita definida en la σ -álgebra de Borel de \mathbb{R} . Sea $\mathcal{L}_\nu(\theta)$ la transformada de Laplace de ν definida como

$$\mathcal{L}_\nu(\theta) = \int \exp(x\theta)\nu(dx),$$

y

$$\Theta = \{\theta \in \mathbb{R} : \mathcal{L}_\nu(\theta) < \infty\}.$$

Supondremos que Θ es un conjunto abierto no vacío. Sea

$$\psi(\theta) = \log \mathcal{L}_\nu(\theta).$$

Entonces, la familia de funciones de densidad (con respecto a ν) definida como

$$f(x; \theta) = \exp\{x\theta - \psi(\theta)\}, \quad \theta \in \Theta, \quad (1.1)$$

es la conocida Familia Exponencial Natural Regular (NEF) generada por la medida ν , (ver Brown (1986)). Si X es una variable aleatoria que sigue una

densidad en la familia (1.1), se denotará $X \sim NEF(\nu; \theta)$. En este trabajo, supondremos que todas las familias exponenciales son regulares.

Sea X_1, \dots, X_n una muestra aleatoria simple de una $NEF(\nu; \theta)$, entonces la media muestral, \bar{X}_n , es suficiente y completo con densidad

$$f_n(s, \theta) = \exp\{n(s\theta - \psi(\theta))\}$$

con respecto a la medida convolución $\nu_n = (\nu * \dots * \nu)/n$. La función generatriz de momentos (fgm) de \bar{X}_n es

$$M_n(t; \theta) = \exp\{n(\psi(\theta + t) - \psi(\theta))\}.$$

Como $\theta \in \Theta$ y Θ es un conjunto abierto, $M_n(t; \theta)$ es una función analítica en $t = 0$, por tanto todos los momentos de \bar{X}_n existen. Sea $m_{k,n}(\theta) = E_\theta[\bar{X}_n^k]$, $k \geq 0$. Se puede comprobar que los momentos satisfacen

$$m_{k,n}(\theta) = n^{-k} \frac{D_\theta^k \exp\{n\psi(\theta)\}}{\exp\{n\psi(\theta)\}}, \quad (1.2)$$

y las relaciones de recurrencia

$$\begin{aligned} m_{k+1,n}(\theta) &= m_{k,n}(\theta) D_\theta \psi(\theta) + \frac{1}{n} D_\theta m_{k,n}(\theta), \quad k \geq 0 \\ m_{0,n}(\theta) &= 1, \end{aligned} \quad (1.3)$$

donde D_θ denota el operador derivada con respecto a θ .

En particular, $\mu = \mu(\theta) = E_\theta[X] = D_\theta \psi(\theta)$ se denomina función de medias de la familia. La varianza de una variable aleatoria que sigue una $NEF(\nu; \theta)$ es $Var_\theta(X) = D_\theta^2 \psi(\theta) > 0$. Así, la función media es estrictamente creciente y el parámetro μ es frecuentemente usado para parametrizar la NEF. La varianza considerada como una función de la media es la función de varianza de la familia y se denotará por $V(\mu)$.

Los momentos centrales $\mu_{k,n}(\theta) = E_\theta[(\bar{X}_n - \mu)^k]$ también existen para todo $k \geq 0$.

Lema 1.1 *Los momentos $\{m_{k,n}(\theta)\}_{k \geq 0}$ determinan unívocamente la distribución de \bar{X}_n .*

Demostración. Como la fgm de \bar{X}_n , $M_n(t; \theta) = \sum_{k \geq 0} m_{k,n}(\theta) t^k / k!$, es analítica en cierto disco $|t| < R$, $R > 0$, tenemos

$$\sum_{k \geq 0} m_{2k,n}(\theta) |t|^{2k} / (2k)! \leq \sum_{k \geq 0} |m_{k,n}(\theta) t|^k / k! < \infty, \quad |t| < R,$$

entonces

$$\overline{\lim}_k (m_{2k,n}(\theta) / (2k)!)^{1/(2k)} < \infty. \quad (1.4)$$

Usando la fórmula de Stirling $(2k)! \sim 2^{2k+1} \sqrt{\pi k} 2k^{2k+1/2} \exp(-2k)$ y (1.4), deducimos que

$$\overline{\lim}_k m_{2k,n}^{1/(2k)}(\theta) / (2k) < \infty,$$

y esta es una condición suficiente para que el problema de los momentos tenga una única solución, ver Shirayev (1984), p.294. \square

Para cada $\theta \in \Theta$, consideramos el espacio de funciones

$$\mathcal{L}_{n,\theta}^2 = \left\{ T(s) : \int T^2(s) f_n(s, \theta) \nu_n(ds) < \infty \right\}.$$

Como es usual en la teoría de los espacios \mathcal{L}^2 consideraremos que dos funciones, $T_1, T_2 \in \mathcal{L}_{n,\theta}^2$, serán equivalentes (o iguales) si $T_1 = T_2$, ν_n -c.s. Con el producto escalar $\langle T_1, T_2 \rangle_{n,\theta} = E_\theta[T_1(\bar{X}_n)T_2(\bar{X}_n)]$, el espacio $(\mathcal{L}_{n,\theta}^2, \langle \cdot, \cdot \rangle_\mu)$ (o simplemente $\mathcal{L}_{n,\theta}^2$) es un espacio de Hilbert para cualquier $\theta \in \Theta$. Denotaremos por $\|T\|_{n,\theta}^2 = \langle T, T \rangle_{n,\theta}$. En $\mathcal{L}_{n,\theta}^2$, la sucesión de funciones $\{1, s, s^2, \dots\}$ es un conjunto completo. Entonces, esta sucesión puede ser ortogonalizada (usando el método de Gram-Schmidt por ejemplo) y obtener un sistema de polinomios mónicos ortogonales (SPMO), $\{P_{k,n}(s; \theta)\}_{k \geq 0}$. El número de polinomios ortogonales en el SPMO es d_n , la dimensión de $\mathcal{L}_{n,\theta}^2$. Además, d_n es finita si y solo si el soporte de la medida ν_n contiene un número finito de puntos y en este caso d_n es el cardinal del soporte. En otro caso, ésto es, si el soporte de ν_n contiene infinitos puntos d_n es infinito.

Algunos métodos para la obtención de polinomios ortogonales serán descritos en las siguientes secciones.

Cualquier función $T_n \in \mathcal{L}_{n,\theta}^2$ admite un único desarrollo en términos de los polinomios ortogonales de la forma

$$T_n(s) = \sum_{k=0}^{d_n-1} a_{k,n}(\theta) P_{k,n}(s; \theta), \quad (1.5)$$

donde los coeficientes de Fourier están determinados por las relaciones

$$a_{k,n}(\theta) = \langle T_n, P_{k,n} \rangle_{n,\theta} / \|P_{k,n}\|_{n,\theta}^2. \quad (1.6)$$

Dada una función T_n , ésta pertenece a $\mathcal{L}_{n,\theta}^2$ si y solo si los coeficientes de Fourier definidos en (1.6) satisfacen

$$\sum_{k=0}^{d_n-1} a_{k,n}^2(\theta) \|P_{k,n}(s; \theta)\|_{n,\theta}^2 < \infty,$$

y en este caso la serie dada en (1.5) es convergente en el sentido de $\mathcal{L}_{n,\theta}^2$.

1.2 Obtención de los polinomios ortogonales

En esta sección describiremos diferentes métodos para la obtención de polinomios ortogonales asociados a la densidad de \bar{X}_n (con respecto a la medida convolución ν_n), $f_n(s, \theta) = \exp\{n(s\theta - \psi(\theta))\}$, $\theta \in \Theta$.

1.2.1 Método 1. El método de Gram-Schmidt

La aplicación del método de Gram-Schmidt (MGS) a una sucesión completa en $\mathcal{L}_{n,\theta}^2$ produce una sucesión de funciones ortogonales. Algunas sucesiones completas de polinomios en $\mathcal{L}_{n,\theta}^2$ son:

- (i) $\{s^k, k \geq 0\}$.
- (ii) $\{(s - \mu)^k, k \geq 0\}$.
- (iii) $\{G_{k,n}(s; \theta) = n^{-k} D_\theta^k f_n(s, \theta) / f_n(s, \theta), k \geq 0\}$.
- (iv) $\{J_{k,n}(s; \theta) = (V(\mu)/n)^k D_\mu^k f_n(s, \theta) / f_n(s, \theta), k \geq 0\}$. Nótese que aquí las derivadas son con respecto a μ .

La completitud de las sucesiones de polinomios descritas en (iii) y (iv) se obtienen a partir de los resultado de Abbey and David (1970) .

La aplicación del método de Gram-Schmidt a la sucesión dada en (ii) da

$$\begin{aligned} P_{0,n}(s; \theta) &= 1; \\ P_{1,n}(s; \theta) &= s - \mu; \\ P_{2,n}(s; \theta) &= (s - \mu)^2 - D_\mu V(\mu)(s - \mu)/n - V(\mu)/n, \end{aligned} \quad (1.7)$$

con normas,

$$\begin{aligned} \|P_{0,n}\|_{n,\theta}^2 &= 1; \\ \|P_{1,n}\|_{n,\theta}^2 &= V(\mu)/n; \\ \|P_{2,n}\|_{n,\theta}^2 &= 6V^2(\mu)/n - 4V^2(\mu)/n^2 + V(\mu)D_\mu V(\mu)/n^3. \end{aligned}$$

Los polinomios dados en (iii) pueden ser obtenidos por la relación de recurrencia

$$\begin{aligned} G_{k+1,n}(s; \theta) &= (s - \mu)G_{k,n}(s; \theta) + \frac{1}{n}D_\theta G_{k,n}(s; \theta), \quad k \geq 0 \\ G_{0,n}(s; \theta) &= 1. \end{aligned}$$

Ejemplo 1.1 *Los polinomios dados en (iv) han sido estudiados por Morris (1982, 1983) en el caso particular en que la función de varianza sea cuadrática, (QVF), i.e., $V(\mu) = v_2\mu^2 + v_1\mu + v_0$. La familia NEF-QVF consiste, esencialmente, de seis subfamilias de distribuciones bien conocidas: normal (con varianza conocida), binomial, Poisson, gamma, binomial negativa, y secante hiperbólica generalizada. Para estas distribuciones los polinomios $\{J_{k,n}(s; \mu)\}_{k \geq 0}$ forman un sistema de polinomios mónicos ortogonales con respecto a la medida $f_n(s; \mu)\nu_n(ds)$ y satisfacen la fórmula de recurrencia (omitiendo argumentos por simplicidad)*

$$\begin{aligned} J_{k+1,n} &= (s - \mu - kD_\mu V)J_{k,n} - k\{1 + (k - 1)v_2/n\}VJ_{k-1,n}/n; \\ J_{0,n} &= 1; \quad J_{1,n} = s - \mu. \end{aligned} \quad (1.8)$$

Estos polinomios son los de Hermite (normal), Krawtchouk (binomial), Charlier (Poisson), Laguerre generalizado (gamma), Meixner (binomial negativa) y Pollaceck (secante hiperbólica). La relación de ortogonalidad es

$$E_{\theta} [J_{k,n}(\bar{X}_n; \mu) J_{j,n}(\bar{X}_n; \mu)] = \delta_{kj} \rho_{k,n} (V(\mu)/n)^k, \quad (1.9)$$

con

$$\rho_{k,n} = k! \prod_{j=0}^{k-1} (1 + jv_2/n), \quad j \geq 1, \quad \rho_{0,n} = 1$$

y δ_{kj} la delta de Kronecker. Para la NEF general, los polinomios $\{J_{k,n}(s; \mu)\}_{k \geq 0}$ no son ortogonales. Una condición necesaria y suficiente para la ortogonalidad es que la función de varianza sea cuadrática. Ver Pomneret (1996)

1.2.2 Método 2. Determinantes de las matrices de momentos

Consideremos la matriz de momentos $\mathcal{M}_{k,n}(\theta) = (m_{i+j,n}(\theta))_{i,j=0,\dots,k}$ con $m_{k,n}(\theta) = E_{\theta}[\bar{X}_n^k]$, $k \geq 0$. Sea $d_n = \dim(\mathcal{L}_{n,\theta}^2)$. Una condición necesaria y suficiente para la existencia de un SPMO es que $\Delta_{k,n}(\theta) = \det \mathcal{M}_{k,n}(\theta) \neq 0$, $k = 0, \dots, d_n - 1$, ver Chihara (1979). En este caso el k -ésimo polinomio ortogonal es

$$P_{k,n}(s; \theta) = \Delta_{k,n}^{-1}(\theta) \det \begin{bmatrix} 1 & m_{1,n}(\theta) & \cdots & m_{k,n}(\theta) \\ m_{1,n}(\theta) & m_{2,n}(\theta) & \cdots & m_{k+1,n}(\theta) \\ \vdots & \vdots & \ddots & \vdots \\ m_{k-1,n}(\theta) & m_{k,n}(\theta) & \cdots & m_{2k-1,n}(\theta) \\ 1 & s & \cdots & s^k \end{bmatrix}, \quad (1.10)$$

para $k = 0, \dots, d_n - 1$. Las normas de estos polinomios son dadas por

$$\|P_{k,n}\|_{n,\theta}^2 = \Delta_{k,n}(\theta) / \Delta_{k-1,n}(\theta), \quad k \geq 0$$

($\Delta_{-1,n}(\theta) = 0$) y la fórmula de recurrencia de tres términos es

$$\begin{aligned} P_{k+1,n} &= (s - c_{k+1,n})P_{k,n} - \lambda_{k+1,n}P_{k-1,n}; \\ P_{0,n} &= 1; \quad P_{1,n} = s - \mu; \end{aligned}$$

con

$$\lambda_{k+1,n} = \lambda_{k+1,n}(\theta) = \Delta_{k-2,n}(\theta)\Delta_{k,n}(\theta)/\Delta_{k-1,n}^2(\theta)$$

y

$$c_{k+1,n} = c_{k+1,n}(\theta) = \langle s, P_{k,n}^2 \rangle_{n,\theta} / \|P_{k,n}\|_{n,\theta}^2.$$

1.2.3 Método 3. Método matricial

Este método nos permite calcular simultáneamente los $m + 1$ primeros polinomios ortonormales. Supongamos que el j -ésimo polinomio ortonormal tiene una expresión de la forma

$$O_{j,n}(s; \theta) = \sum_{r=0}^j \delta_{j,r,n}(\theta)(s - \mu)^r, \quad j = 0, \dots, m; \quad \text{con } m \leq d_n - 1.$$

Usando notación matricial y omitiendo argumentos, tenemos

$$\begin{bmatrix} O_{0,n} \\ O_{1,n} \\ \vdots \\ O_{m,n} \end{bmatrix} = \begin{bmatrix} \delta_{0,0,n} & 0 & 0 & \cdots & 0 \\ \delta_{1,0,n} & \delta_{1,1,n} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_{m,0,n} & \delta_{m,1,n} & \delta_{m,2,n} & \cdots & \delta_{m,m,n} \end{bmatrix} \begin{bmatrix} 1 \\ s - \mu \\ \vdots \\ (s - \mu)^m \end{bmatrix},$$

o bien

$$\mathbf{o}_{m,n}(s; \theta) = \mathbb{D}_{m,n}(\theta) \mathbf{c}_m(s; \theta). \quad (1.11)$$

Notar que $\mathcal{R}_{m,n}(\theta)$ es una matriz triangular inferior que contiene los coeficientes de los polinomios ortonormales hasta el grado m . De (1.11), obtenemos

$$\mathbf{o}_{m,n}(\overline{X}_n; \theta) \mathbf{o}_{m,n}^t(\overline{X}_n; \theta) = \mathbb{D}_{m,n}(\theta) \mathbf{c}_m(\overline{X}_n; \theta) \mathbf{c}_m^t(\overline{X}_n; \theta) \mathbb{D}_{m,n}^t(\theta)$$

(t denota la traspuesta de una matriz). Tomando valores esperados, como los polinomios son ortonormales,

$$\mathbb{I}_{m+1} = \mathbb{D}_{m,n}(\theta) \mathbf{C}_{m,n}(\theta) \mathbb{D}_{m,n}^t(\theta)$$

o

$$\mathbb{C}_{m,n}(\theta) = (\mathbb{D}_{m,n}(\theta))^{-1} (\mathbb{D}_{m,n}^t(\theta))^{-1},$$

donde $\mathbb{C}_{m,n}(\theta) = (E_\theta(\bar{X}_n - \mu)^{i+j})_{i,j=0,\dots,m}$. Entonces, un método para obtener la matriz $\mathbb{D}_{m,n}(\theta)$ es el siguiente

1. Calcular la matriz de momentos centrales $\mathbb{C}_{m,n}(\theta)$. Los momentos centrales pueden obtenerse a partir de 1.3 y de

$$\mu_k = \sum_{j=0}^k (-1)^k \binom{k}{j} m_{k-j,n} \mu^j.$$

2. Buscar la descomposición de Cholesky de $\mathbb{C}_{m,n}(\theta)$, esto es, calcular $\mathbb{T}_{m,n} = \mathbb{T}_{m,n}(\theta)$ tal que $\mathbb{C}_{m,n} = \mathbb{T}_{m,n} \mathbb{T}_{m,n}^t$.
3. La fila k -ésima de la matriz $\mathbb{D}_{m,n}(\theta) = (\mathbb{T}_{m,n}(\theta))^{-1}$, contiene los coeficientes del k -ésimo polinomio ortonormal en potencias crecientes de $(s - \mu)$

1.3 Estimación insesgada

Sea X_1, \dots, X_n una muestra aleatoria simple de una $NEF(\nu; \theta)$. Una función del parámetro, $h(\theta)$, se dice umvu-estimable para un tamaño de muestra n si existe una función, T_n , satisfaciendo:

(i) $T_n(s) \in \bigcap_{\theta \in \Theta} \mathcal{L}_{n,\theta}^2$

(ii) $E_\theta[T_n(\bar{X}_n)] = h(\theta)$, para $\theta \in \Theta$.

Nótese que esta definición da lugar a los estimadores de mínima varianza que se obtienen en la teoría de Blackwell-Rao-Lehman-Scheffe. La condición (ii) es la condición de insesgades de una función del suficiente y completo. La condición (i) expresa el hecho de que nos limitamos a estimadores con varianza finita. Esto no supone ninguna restricción práctica. Así que $T_n(\bar{X}_n)$ es

el estimador insesgado de mínima variación (umvue) de $h(\theta)$ para un tamaño de muestra n . Denotaremos por \mathcal{U}_n al conjunto de todas las funciones estimables umvu para el tamaño de muestra n . El conjunto \mathcal{U}_n es un subconjunto del conjunto de las funciones analíticas en Θ , ver Lehmann (1983).

Sea $\theta_0 \in \Theta$ arbitrario. Como $T_n(s) \in \mathcal{L}_{n,\theta_0}^2$, es el estimador insesgado de una función paramétrica $h(\theta_0) \in \mathcal{U}_n$, admite un desarrollo en términos de polinomios mónicos ortogonales de la forma

$$T_n(\bar{X}_n) = \sum_{k \geq 0} a_{k,n}(\theta_0) P_{k,n}(\bar{X}_n; \theta_0),$$

donde

$$a_{k,n}(\theta_0) = E_{\theta_0} [T_n(\bar{X}_n) P_{k,n}(\bar{X}_n; \theta_0)] / \|P_{k,n}(\cdot; \theta_0)\|_{n,\theta_0}^2, \quad k \geq 0. \quad (1.12)$$

La condición (i) y la ortogonalidad de los polinomios implica que

$$E_{\theta_0} [T_n^2(\bar{X}_n)] = \sum_{k \geq 0} a_{k,n}^2(\theta_0) \|P_{k,n}(\cdot; \theta_0)\|_{n,\theta_0}^2 < \infty, \quad \forall \theta_0 \in \Theta. \quad (1.13)$$

Como $P_{0,n}(s; \theta_0) = 1$, es inmediato que

$$a_{0,n}(\theta_0) = h(\theta_0) \quad (1.14)$$

entonces, de (1.13) y (1.14) obtenemos

$$Var_{\theta_0}(T_n(\bar{X}_n)) = \sum_{k \geq 1} a_{k,n}^2(\theta_0) \|P_{k,n}(\cdot; \theta_0)\|_{n,\theta_0}^2 < \infty. \quad (1.15)$$

Nótese que θ_0 ha sido escogido arbitrariamente en Θ . Por tanto, escribiremos a partir de ahora

$$T_n(\bar{X}_n) = \sum_{k \geq 0} a_{k,n}(\theta) P_{k,n}(\bar{X}_n; \theta), \quad \text{para todo } \theta \in \Theta. \quad (1.16)$$

Es bien conocido el resultado de que el estimador umvu basado en un estadístico suficiente y completo es ($\nu_n - c.s.$) único. La expresión (1.16) refleja el

hecho de que aunque el umvu estimador sea único, éste admite infinitos desarrollos en términos de los polinomios ortogonales, uno para cada valor de $\theta \in \Theta$. En otras palabras, el lado derecho de (1.16) no depende del valor de θ escogido.

De la relación de insegadez dada en (ii), obtenemos

$$\int T_n(s) \exp\{ns\theta\} \nu_n(ds) = \exp\{n\psi(\theta)\} h(\theta), \text{ para todo } \theta \in \Theta. \quad (1.17)$$

Derivando j veces respecto a θ en ambos lados de (1.17), usando la fórmula de Leibnitz para derivar un producto y (1.2), obtenemos

$$n^j \int T_n(s) s^j \exp\{ns\theta\} \nu_n(ds) = \exp\{n\psi(\theta)\} \sum_{r=0}^j \binom{j}{r} n^{j-r} m_{j-r,n}(\theta) D_\theta^r h(\theta),$$

para todo $\theta \in \Theta$, o equivalentemente

$$E_\theta[T_n(\bar{X}_n) \bar{X}_n^j] = \sum_{r=0}^j \binom{j}{r} m_{j-r,n}(\theta) D_\theta^r h(\theta) / n^r, \text{ para todo } \theta \in \Theta. \quad (1.18)$$

A partir de (1.18) y de la expresión explícita de los polinomios ortogonales no es difícil deducir que $E_\theta[T_n(\bar{X}_n) P_{k,n}(\bar{X}_n; \theta)]$ es una función de los coeficientes del k -ésimo polinomio ortogonal, los primeros k momentos de \bar{X}_n y las k primeras derivadas de la función $h(\theta)$. Esto implica que los coeficientes de Fourier $a_{k,n}(\theta) = E_\theta[T_n P_{k,n}] / \|P_{k,n}\|_{n,\theta}^2$ en el desarrollo ortogonal del estimador, pueden ser obtenidos sin un conocimiento explícito de $T_n(\bar{X}_n)$. Este resultado puede encontrarse en Abbey y David (1970).

Ejemplo 1.2 *Estimadores insesgados en la NEF-QVF.* Sea X_1, \dots, X_n una muestra aleatoria simple de una NEF-QVF y $h(\theta) \in \mathcal{U}_n$. Consideremos la función de medias $\mu = D_\theta \psi(\theta)$ y $g(\mu) = (h \circ (D_\theta \psi)^{-1})(\mu)$, donde $(D_\theta \psi)^{-1}$ denota la función inversa de la función media. Nótese que $h(\theta) = g(\mu(\theta))$, esto es, g es una reparametrización en términos de μ de la función $h(\theta)$. La condición de insesgadez es

$$\int T_n(s) f_n(s, \theta) \nu_n(ds) = g(\mu), \quad (1.19)$$

derivando k veces en ambos miembros de (1.19) con respecto a μ ,

$$(n/V(\mu))^k \int T_n(s) J_{k,n}(s, \theta) f_n(s, \theta) \nu_n(ds) = D_\mu^k g(\mu)$$

de (1.12) y el resultado del ejemplo 1.1, tenemos que el k -ésimo coeficiente de Fourier es

$$a_{k,n}(\theta) = D_\mu^k g(\mu) / \rho_{k,n}.$$

Observamos, pues que en el caso particular de la familia NEF-QVF los coeficientes de Fourier son particularmente simples.

1.4 Obtención de los coeficientes de Fourier

Vamos a dar ahora un método matricial para obtener los $m + 1$ primeros coeficientes de Fourier. Previamente, vamos a estudiar algunas propiedades de los polinomios $G_{k,n}(s; \theta) = n^{-k} D_\theta^k f_n(s, \theta) / f_n(s, \theta)$, $k \geq 0$. Supongamos que estos polinomios tienen la expresión explícita

$$G_{j,n}(s; \theta) = \sum_{r=0}^j \gamma_{j,r,n}(\theta) (s - \mu)^r, \quad j = 0, \dots, m; \quad (1.20)$$

Usando notación matricial, tenemos

$$\begin{bmatrix} G_{0,n} \\ G_{1,n} \\ \vdots \\ G_{m,n} \end{bmatrix} = \begin{bmatrix} \gamma_{0,0,n} & 0 & 0 & \cdots & 0 \\ \gamma_{1,0,n} & \gamma_{1,1,n} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{m,0,n} & \gamma_{m,1,n} & \gamma_{m,2,n} & \cdots & \gamma_{m,m,n} \end{bmatrix} \begin{bmatrix} 1 \\ s - \mu \\ \vdots \\ (s - \mu)^m \end{bmatrix},$$

o

$$\mathbf{g}_{m,n}(s; \theta) = \mathbb{G}_{m,n}(\theta) \mathbf{c}_m(s; \theta). \quad (1.21)$$

Nótese que $\mathbb{G}_{m,n}(\theta)$ es una matriz triangular inferior con elementos diagonales $\gamma_{j,j,n} = 1$, por tanto, tiene matriz inversa $\mathbb{G}_{m,n}^{-1}(\theta)$ la cual es también triangular inferior.

Lema 1.2 Consideremos la matriz triangular inferior $\mathbb{F}_{m,n}(\theta) = (f_{r,j,n}(\theta))_{r,j=0,\dots,m}$, con

$$f_{r,j,n}(\theta) = \begin{cases} \binom{j}{r} \mu_{j-r,n}(\theta), & 0 \leq r \leq j \leq m \\ 0 & j > m \end{cases}$$

Entonces, $\mathbb{F}_{m,n}(\theta) = \mathbb{G}_{m,n}^{-1}(\theta)$.

Demostración. Sea $T = T(s)$ una función arbitraria de $\bigcap_{\theta \in \Theta} \mathcal{L}_{n,\theta}^2$ con

$$\int T(s) f_n(s, \theta) \nu_n(ds) = h(\theta), \text{ para todo } \theta \in \Theta. \quad (1.22)$$

Derivando k veces en ambos miembros de (1.22) respecto de θ ,

$$\int T(s) G_{k,n}(s; \theta) f_n(s, \theta) \nu_n(ds) = D_\theta^k h(\theta) / n^k. \quad (1.23)$$

De (1.18), se puede probar fácilmente que

$$\int T(s) (s - \mu)^j f_n(s, \theta) \nu_n(ds) = \sum_{r=0}^j \binom{j}{r} \mu_{j-r,n}(\theta) D_\theta^r h(\theta) / n^r, \text{ para todo } \theta \in \Theta. \quad (1.24)$$

Sustituyendo (1.20) en el lado izquierdo de (1.23) y usando (1.24), obtenemos

$$\sum_{r=0}^k \left\{ \sum_{j=r}^k \binom{j}{r} \mu_{j-r,n}(\theta) \gamma_{k,j,n}(\theta) \right\} D_\theta^r h(\theta) / n^r = D_\theta^k h(\theta) / n^k, \quad (1.25)$$

como T fue escogido arbitrariamente, obtenemos de (1.25),

$$\sum_{j=r}^k \binom{j}{r} \mu_{j-r,n}(\theta) \gamma_{k,j,n}(\theta) = \delta_{r,k},$$

y esto implica $\mathbb{F}_{m,n}(\theta) = \mathbb{G}_{m,n}^{-1}(\theta)$. \square

Consideremos los $m + 1$ primeros polinomios mónicos ortogonales

$$P_{j,n}(s; \theta) = \sum_{r=0}^j \pi_{j,r,n}(\theta) (s - \mu)^r, \quad j = 0, \dots, m; \quad (1.26)$$

en notación matricial

$$\begin{bmatrix} P_{0,n} \\ P_{1,n} \\ \vdots \\ P_{m,n} \end{bmatrix} = \begin{bmatrix} \pi_{0,0,n} & 0 & 0 & \cdots & 0 \\ \pi_{1,0,n} & \pi_{1,1,n} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_{m,0,n} & \pi_{m,1,n} & \pi_{m,2,n} & \cdots & \pi_{m,m,n} \end{bmatrix} \begin{bmatrix} 1 \\ s - \mu \\ \vdots \\ (s - \mu)^m \end{bmatrix},$$

o

$$\mathbf{p}_{m,n}(s; \theta) = \mathbb{P}_{m,n}(\theta) \mathbf{c}_m(s; \theta). \quad (1.27)$$

De (1.21) y lema 1.2, tenemos

$$\mathbf{p}_{m,n}(s; \theta) = \mathbb{B}_{m,n}(\theta) \mathbf{g}_{m,n}(s; \theta), \quad \text{con } \mathbb{B}_{m,n}(\theta) = \mathbb{P}_{m,n}(\theta) \mathbb{F}_{m,n}(\theta).$$

Dada una función $h(\theta) \in \mathcal{U}_n$ con estimador insesgado $T_n(\bar{X}_n)$, consideremos el vector $\mathbf{a}_{m,n}(\theta) = (a_{0,n}(\theta), \dots, a_{m,n}(\theta))^t$, donde $a_{k,n}(\theta)$ es el k -ésimo coeficiente de Fourier de T_n . Entonces,

$$\begin{aligned} \mathbf{a}_{m,n}(\theta) &= \text{diag}(\|P_{k,n}\|_{n,\theta}^{-2}, k = 0, \dots, m) E_\theta[T_n(\bar{X}_n) \mathbf{p}_{m,n}(\bar{X}_n; \theta)] = \\ &= \text{diag}(\|P_{k,n}\|_{n,\theta}^{-2}, k = 0, \dots, m) \mathbb{B}_{m,n}(\theta) E_\theta[T_n(\bar{X}_n) \mathbf{g}_{m,n}(\bar{X}_n; \theta)] = \\ &= \mathbb{A}_{m,n}(\theta) \mathbf{d}_m(\theta), \end{aligned}$$

donde

$$\mathbb{A}_{m,n}(\theta) = \text{diag}(\|P_{k,n}\|_{n,\theta}^{-2}, k = 0, \dots, m) \mathbb{B}_{m,n}(\theta)$$

y

$$\mathbf{d}_m(\theta) = (h(\theta), D_\theta h(\theta)/n, \dots, D_\theta^m h(\theta)/n^m)^t.$$

Nótese que la matriz $\mathbb{A}_{m,n}(\theta)$ depende de cantidades relacionadas con la familia exponencial, pero no de la función a estimar $h(\theta)$. De nuevo, vemos que los coeficientes de Fourier pueden ser obtenidos sin conocimiento explícito de T_n . Tan solo se necesitan las derivadas de la función $h(\theta)$. Nótese que la primera columna de $\mathbb{A}_{m,n}(\theta)$ es el vector $(1, 0, \dots, 0)^t$ (de dimensión $m + 1$), entonces para $k \geq 1$, el k -ésimo coeficiente de Fourier, $a_{k,n}(\theta)$, es una

combinación lineal de $D_\theta h(\theta), \dots, D_\theta^k h(\theta)$, i.e., en esta combinación lineal no aparece $h(\theta)$.

En las siguientes secciones veremos que el desarrollo en términos de los polinomios ortogonales es muy útil para obtener muchas propiedades de los estimadores en la familia exponencial natural como cotas inferiores de la varianza para los estimador, propiedades límites, comparación entre estimadores, etc.

1.5 Propiedades límites de los polinomios ortogonales

Para investigar las propiedades límites, necesitamos algunos resultados previos. Sea η_n una sucesión de medidas σ -finita definidas en $\mathcal{B}(\mathbb{R})$ e $Y_n \sim NEF(\eta_n; \theta)$ con espacio paramétrico común Θ , un subconjunto abierto de \mathbb{R} . Sea Y una variable aleatoria con todos los momentos finitos, tal que la sucesión $m_k = E[Y^k]$, $k \geq 0$, caracteriza la distribución de Y . Usando la caracterización de la convergencia en distribución mediante funciones características, se obtiene de forma inmediata el siguiente resultado.

Lema 1.3 *Son equivalentes:*

- (i) $Y_n \xrightarrow{d} Y$, cuando n tiende a infinito, para cualquier $\theta \in \Theta$.
- (ii) $m_{k,n}(\theta) = E_\theta[Y_n^k] \rightarrow m_k = E[Y^k]$, cuando n tiende a infinito, para cualquier $\theta \in \Theta$ y $k \geq 0$.

Una consecuencia inmediata del lema 1.3 es:

Corolario 1.1 *Sea g una función continua definida en \mathbb{R}^r (o un subconjunto de \mathbb{R}^r). $Y_n \sim NEF(\eta_n; \theta)$ e Y en las condiciones arriba expuestas. Entonces*

$$\lim_n g(m_{1,n}(\theta), \dots, m_{r,n}(\theta)) = g(m_1, \dots, m_r), \text{ para todo } \theta \in \Theta.$$

O más generalmente, si $\{g_n\}_{n \geq 1}$ es una sucesión de funciones continuas tal que g_n convergen (puntualmente) a una función g , entonces

$$\lim_n g_n(m_{1,n}(\theta), \dots, m_{r,n}(\theta)) = g(m_1, \dots, m_r), \quad \text{para todo } \theta \in \Theta.$$

Para nuestro propósito, una aplicación importante del corolario 1.1 es la convergencia de los determinantes de las matrices de momentos de Y_n al correspondiente determinante de la matriz de momentos de Y . Consecuentemente,

Teorema 1.1 Sean $Y_n \sim NEF(\eta_n; \theta)$ e Y en las condiciones anteriormente expuestas. Sean $\mathbb{M}_{k,n}(\theta) = (m_{i+j,n}(\theta))_{i,j=0,\dots,k}$ y $\mathbb{M}_k = (m_{i+j})_{i,j=0,\dots,k}$ las matrices de momentos asociadas a Y_n e Y respectivamente. Si $\det \mathbb{M}_k \neq 0$, entonces

$$\lim_n P_{k,n}(s; \theta) = P_k(s; \theta)$$

y

$$\lim_n \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2 = \|P_k(\cdot)\|^2,$$

donde $P_{k,n}(s; \theta)$ y $P_k(s)$ denotan los polinomios ortogonales mónicos de grado k de Y_n e Y , respectivamente.

Demostración. Inmediata a partir de la expresión de los polinomios ortogonales mónicos dados en (1.10). \square

Sea $\{X_j\}_{j \geq 1}$ una sucesión de variables aleatorias independientes que siguen una $NEF(v; \theta)$. El teorema central del límite implica que ,

$$Z_n(\theta) = \sqrt{n/V(\mu)} \{\bar{X}_n - \mu\} \xrightarrow{d} Z \sim \mathcal{N}(0, 1),$$

cuando n tiende a infinito, para cualquier $\theta \in \Theta$.

Sean $\{R_{k,n}(z; \theta)\}_{k \geq 0}$ y $\{P_{k,n}(s; \theta)\}_{k \geq 0}$ los SPMO asociados a la densidad de $Z_n(\theta)$ y \bar{X}_n , respectivamente. Se comprueba fácilmente que

$$R_{k,n}(z; \theta) = (n/V(\mu))^{k/2} P_{k,n}(\mu + z\sqrt{V(\mu)/n}; \theta), \quad (1.28)$$

y

$$\|R_{k,n}(\cdot; \theta)\|_{n,\theta}^2 = (n/V(\mu))^k \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2. \quad (1.29)$$

(Por supuesto, las normas son calculadas en sus respectivos espacios \mathcal{L}^2).

Consideremos los polinomios de Hermite $\{He_k(z)\}_{k \geq 0}$ definidos recurrentemente como

$$\begin{aligned} He_{k+1}(z) &= zHe_k(z) - kHe_{k-1}(z), \quad k \geq 1; \\ He_0(z) &= 1; \quad He_1(z) = z, \end{aligned}$$

con expresión explícita

$$He_k(z) = k! \sum_{j=0}^{\lfloor k/2 \rfloor} (-1)^j \frac{z^{k-2j}}{2^{2j} j! (k-2j)!}.$$

El coeficiente de la m -ésima potencia de z en $He_k(z)$ será denotado como $coef(m, He_k)$. En particular,

$$coef(0, He_k) = \begin{cases} 0, & k = 2r - 1, \quad r \geq 1 \\ (2r - 1)!!, & k = 2r, \quad r \geq 1 \end{cases} \quad (1.30)$$

donde $(2r - 1)!! = (2r - 1)(2r - 3) \dots 1$.

Los polinomios de Hermite son mónicos y ortogonales con respecto a la densidad de la distribución $\mathcal{N}(0, 1)$. Como esta distribución está unívocamente determinada por la sucesión de momentos, del teorema 1.1, tenemos

$$\lim_n R_{k,n}(z; \theta) = He_k(z), \quad \text{para } \theta \in \Theta, \quad k \geq 0, \quad z \in \mathbb{R},$$

y

$$\lim_n \|R_{k,n}(\cdot; \theta)\|_{n,\theta}^2 = \|He_k\|^2 = k!.$$

Entonces, obtenemos el siguiente resultado:

Teorema 1.2 *Sea $\{X_j\}_{j \geq 1}$ una sucesión de variables aleatorias independientes siguiendo una NEF($v; \theta$) con función de media y varianza $\mu(\theta)$ y $V(\mu)$, respectivamente, y $\{P_{k,n}(s; \theta)\}_{k \geq 0}$ el SPMO asociado a la densidad de \bar{X}_n . Entonces:*

(i) $\lim_n (n/V(\mu))^{k/2} P_{k,n}(\mu + z\sqrt{V(\mu)/n}; \theta) = He_k(z)$, para cualquier $\theta \in \Theta$, $k \geq 0$ y $z \in \mathbb{R}$.

(ii) $\lim_n (n/V(\mu))^k \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2 = k!$, para cualquier $\theta \in \Theta$, $k \geq 0$.

Supongamos que el polinomio ortogonal de grado k con respecto a la densidad de \bar{X}_n se escribe como $P_{k,n}(s; \theta) = \sum_{j=0}^k \pi_{k,j,n}(\theta)(s - \mu)^j$. El siguiente resultado es una aplicación directa del teorema 1.2:

Corolario 1.2 *Se verifica:*

$$\lim_n (n/V(\mu))^{(k-j)/2} \pi_{k,j,n}(\theta) = \text{coef}(j, He_k), \quad \forall \theta \in \Theta \quad \text{y} \quad 0 \leq j \leq k.$$

Como una consecuencia de este corolario, dado que $\pi_{k,k,n}(\theta_0) = 1$, se tiene:

$$\begin{aligned} \lim_n P_{k,n}(s; \theta_0) &= \sum_{j=0}^k \lim_n (n/V(\mu_0))^{(k-j)/2} \pi_{k,j,n}(\theta_0) \lim_n (V(\mu_0)/n)^{(k-j)/2} (s - \mu_0)^j \\ &= (s - \mu_0)^k, \end{aligned}$$

o más generalmente, si $\{s_n\}_n$ es una sucesión de números reales tal que $\lim_n s_n = \mu$, entonces

$$\lim_n P_{k,n}(s_n; \theta_0) = (\mu - \mu_0)^k. \quad (1.31)$$

Ejemplo 1.3 *Lindsay (1989) investigó los comportamientos límites de los determinantes de las matrices de momentos. Usando su notación, sea F_n la función de distribución de $\sqrt{n}\bar{X}$, con \bar{X} la media de n observaciones i.i.d. de F , una distribución arbitraria con media = 0 y varianza finita $\sigma^2 > 0$. Sea $v_k = \det(m_{i+j})_{i,j=0,\dots,k}$, donde $m_r = E[(\sqrt{n}\bar{X})^r]$.*

Él interpretó la cantidad $\Lambda_{k,n} = v_k v_{k-2} / v_{k-1}^2$ como el crecimiento en la impredecibilidad lineal del k -ésimo momento (ver, p.713). Sea $\{Q_{k,n}(s)\}_{k \geq 0}$ un SPMO con respecto a la densidad de F_n . De acuerdo con los resultados

mostrados en la subsección 1.2.2, $\Lambda_{k,n}$ es el coeficiente de $Q_{k-1,n}$ en la fórmula de recurrencia de tres términos para los polinomios mónicos ortogonales:

$$Q_{k+1,n} = (s - c_{k+1})Q_{k,n} - \Lambda_{k,n}Q_{k-1,n}, \quad k \geq 1. \quad (1.32)$$

El TCL implica que $\sqrt{n}\bar{X}$ converge en distribución a $\mathcal{N}(0, \sigma^2)$. Los polinomios mónicos ortogonales con respecto a la distribución $\mathcal{N}(0, \sigma^2)$ son $\{P_k(s) = \sigma^k He_k(s/\sigma)\}_{k \geq 0}$ y satisfacen la relación de recurrencia

$$P_{k+1} = sP_k - k\sigma^2 P_{k-1}, \quad k \geq 1. \quad (1.33)$$

La convergencia de $Q_{k,n}(s)$ a $P_k(s)$ cuando n tiende a infinito implica que $\lim_n \Lambda_{k,n} = k\sigma^2$, basta para ello comparar (1.32) y (1.33)

1.6 Propiedades límites de los estimadores insesgados

Derivaremos algunas propiedades límites de los estimadores insesgados a partir de los desarrollos ortogonales. Primero de todo, nótese que si $h(\theta)$ es una función estimable umvu para un tamaño de muestra n_0 con estimador insesgado $T_{n_0}(\bar{X}_{n_0})$, entonces $T_n(\bar{X}_n) = E[T_{n_0}(\bar{X}_{n_0}) | \bar{X}_n]$ es insesgado de $h(\theta)$ para un tamaño de muestra $n \geq n_0$ y

$$Var_\theta(T_n(\bar{X}_n)) = Var_\theta(E[T_{n_0}(\bar{X}_{n_0}) | \bar{X}_n]) \leq Var_\theta(T_{n_0}(\bar{X}_{n_0})) < \infty.$$

Así, la sucesión $\{\mathcal{U}_n\}_{n \geq 1}$ es no decreciente en el sentido de la inclusión de conjuntos. Más aún, $\lim_n Var_\theta(T_n(\bar{X}_n)) = 0$, y esto significa que la sucesión de estimadores insesgados $\{T_n(\bar{X}_n)\}_{n \geq n_0}$ es consistente.

Ahora, estudiaremos el comportamiento límite de los coeficientes de Fourier en el desarrollo ortogonal de los estimadores insesgados.

Teorema 1.3 Sea $h(\theta) \in \mathcal{U}_{n_0}$, para cierto entero positivo n_0 , y $T_n(\bar{X}_n) = \sum_{k \geq 0} a_{k,n}(\theta) P_{k,n}(\bar{X}_n; \theta)$ el estimador insesgado para el tamaño de muestra

$n \geq n_0$. Sea $g(\mu) = h \circ (D_\theta \psi)^{-1}(\mu)$, donde $\mu = D_\theta \psi(\theta)$ es la función de medias de la NEF. Entonces

$$\lim_n a_{k,n}(\theta) = D_\mu^k g(\mu)/k!, \quad \text{para todo } k \geq 0 \text{ y } \theta \in \Theta. \quad (1.34)$$

Demostración. Sea $\{X_j\}_{j \geq 1}$ una sucesión de variables aleatorias independientes que siguen una NEF($\nu; \theta_1$). La consistencia del estimador insesgado implica que $p \lim_n T_n(\bar{X}_n) = h(\theta_1)$. La serie $\sum_{k \geq 0} a_{k,n}(\theta) P_{k,n}(\bar{X}_n; \theta)$ es absolutamente convergente en el sentido cuadrático, entonces es absolutamente convergente en probabilidad (para cualquier $\theta \in \Theta$). Este hecho justifica los cambios entre el sumatorio y el límite (en probabilidad) que se harán a lo largo de la demostración de este teorema.

Escogiendo arbitrariamente un $\theta_0 \in \Theta$, el estimador insesgado puede ser escrito como

$$T_n(\bar{X}_n) = \sum_{k \geq 0} a_{k,n}(\theta_0) P_{k,n}(\bar{X}_n; \theta_0) \quad (1.35)$$

y tomando límite en probabilidad (bajo θ_1), usando (1.31), tenemos

$$\begin{aligned} g(\mu_1) &= h(\theta_1) = p \lim_n T_n(\bar{X}_n) = \sum_{k \geq 0} \lim_n a_{k,n}(\theta_0) p \lim_n P_{k,n}(\bar{X}_n; \theta_0) = \\ &= \sum_{k \geq 0} \left\{ \lim_n a_{k,n}(\theta_0) \right\} (\mu_1 - \mu_0)^k, \end{aligned} \quad (1.36)$$

donde $\mu_1 = \mu(\theta_1)$ y $\mu_0 = \mu(\theta_0)$.

Desarrollando la función $g(\mu)$ en serie de Taylor alrededor de $\mu = \mu_0$,

$$g(\mu_1) = \sum_{k \geq 0} \frac{D_\mu^k g(\mu_0)}{k!} (\mu_1 - \mu_0)^k, \quad (1.37)$$

para μ_1 suficientemente cerca de μ_0 . Igualando coeficientes en (1.36) y (1.37) obtenemos el resultado. \square

Veamos ahora el comportamiento límite de $P_{k,n}(\bar{X}_n; \theta)$.

Teorema 1.4 Para todo $\theta \in \Theta$,

$$(i) \{n/V(\mu)\}^{k/2} P_{k,n}(\bar{X}_n; \theta) \xrightarrow{d} He_k(Z), \quad Z \sim \mathcal{N}(0, 1).$$

$$(ii) \overline{\lim}_n \{n/(2V(\mu) \log \log n)\}^{k/2} P_{k,n}(\bar{X}_n; \theta) = 1.$$

Demostración. Este resultado es directo a partir del teorema 1.2, el teorema central del límite y la ley del logaritmo iterado, respectivamente. \square

Con estos ingredientes podemos establecer la distribución límite del estimador insesgado. Supongamos que $\{X_j\}_{j \geq 1}$ es una sucesión de variables aleatorias independientes siguiendo una $NEF(\nu; \theta_0)$, donde θ_0 puede ser interpretado como el verdadero pero desconocido valor del parámetro. Sea $\mu_0 = D_\theta \psi(\theta_0)$ la verdadera media. Sea $h(\theta) \in \mathcal{U}_{n_0}$ y $g(\mu) = h \circ (D_\theta \psi)^{-1}(\mu)$ (la función $g(\mu)$ es la función $h(\theta)$ reparametrizada en términos de la media). Los operadores diferenciales D_θ y D_μ están estrechamente relacionados, pues

$$D_\theta = V(\mu) D_\mu,$$

$$D_\theta^2 = VV' D_\mu + V^2 D_\mu^2,$$

$$D_\theta^3 = (VV'^2 + V^2 V'') D_\mu + 3V^2 V' D_\mu^2 + V^3 D_\mu^3,$$

y así sucesivamente. (Aquí V' , V'' , \dots , denotan las derivadas con respecto a μ).

Teorema 1.5 Sea $h(\theta) \in \mathcal{U}_{n_0}$ para cierto entero positivo n_0 , consideremos $g(\mu) = h \circ (D_\theta \psi)^{-1}(\mu)$ y $j = j(\mu_0) = \min\{k > 0 : D_\mu^k g(\mu_0) \neq 0\}$, i.e., j es el orden de la primera derivada de g (con respecto a μ) en μ_0 no nula. Entonces

$$n^{j/2} \{T_n(\bar{X}_n) - h(\theta_0)\} \xrightarrow{d} \frac{D_\mu^j g(\mu_0)}{j!} V^{j/2}(\mu_0) He_j(Z), \quad n \rightarrow \infty, \quad (1.38)$$

con $Z \sim \mathcal{N}(0, 1)$ y

$$\overline{\lim}_n \{n/(2 \log \log n)\}^{j/2} \{T_n(\bar{X}_n) - h(\theta_0)\} = \frac{D_\mu^j g(\mu_0)}{j!} V^{j/2}(\mu_0). \quad (c.s) \quad (1.39)$$

Demostración. Consideremos el desarrollo ortogonal del estimador insesgado de $h(\theta)$,

$$T_n(\bar{X}_n) = \sum_{k \geq 0} a_{k,n}(\theta_0) P_{k,n}(\bar{X}_n; \theta_0). \quad (1.40)$$

Se vió al final de la sección 1.4 que para $k \geq 1$, el k -ésimo coeficiente de Fourier, $a_{k,n}(\theta_0)$, es una combinación lineal de $D_\theta h(\theta_0), \dots, D_\theta^k h(\theta_0)$, i.e., en esta combinación lineal no aparece $h(\theta_0)$. De la relación entre los operadores diferenciales D_θ y D_μ dados arriba es fácil comprobar que $D_\theta^k h(\theta_0) = 0$, para $1 \leq k < j$ y $D_\theta^k h(\theta_0) = V^k(\mu_0) D_\mu^k g(\mu_0)$. Entonces $a_{k,n}(\theta_0) = 0$, para $k = 1, \dots, j-1$ y $a_{j,n}(\theta_0) \neq 0$. De (5.10),

$$\begin{aligned} T_n(\bar{X}_n) - h(\theta_0) &= a_{j,n}(\theta_0) P_{j,n}(\bar{X}_n; \theta_0) + \sum_{k \geq j+1} a_{k,n}(\theta_0) P_{k,n}(\bar{X}_n; \theta_0) \\ &= W_{1,n} + W_{2,n}. \end{aligned}$$

De los Teoremas 1.3 y 1.4, tenemos

$$n^{j/2} W_{1,n} \xrightarrow{d} \frac{D_\mu^j g(\mu_0)}{j!} V^{j/2}(\mu_0) He_j(Z), \quad n \rightarrow \infty, \quad (1.41)$$

y

$$n^{j/2} W_{2,n} \xrightarrow{P_{\mu_0}} 0, \quad n \rightarrow \infty. \quad (1.42)$$

usando (1.41) y (1.42), se sigue (1.38). La demostración de (1.39) es similar. \square

En particular, si $j = 1$, del teorema 1.5, obtenemos la normalidad asintótica del umvu estimador. Si $j = 2$, notando que $He_2(z) = z^2 - 1$, tendremos

$$\frac{2n}{D_\mu^2 g(\mu_0) V(\mu_0)} \{T_n(\bar{X}_n) - h(\theta_0)\} + 1 \xrightarrow{d} \chi_1^2.$$

El estimador de máxima verosimilitud (emv) de $g(\mu)$ es $g(\bar{X}_n)$. Del teorema 3.1B de Serfling (1980), p.119, se tiene que

$$n^{j/2} \{g(\bar{X}_n) - g(\mu_0)\} \xrightarrow{d} \frac{D_\mu^j g(\mu_0)}{j!} V^{j/2}(\mu_0) Z^j, \quad n \rightarrow \infty. \quad (1.43)$$

Comparando (1.43) y el resultado del Teorema 1.5, vemos que para $j = 1$ el emv y el umvue tienen igual distribución, pero esta propiedad no se verifica para $j \geq 2$. Más comparaciones entre los estimadores umvu y umv serán estudiadas en la sección 1.9.

Las distribuciones límites obtenidas en el Teorema 1.5, habían sido obtenidas previamente por López-Blázquez y Castaño-Martínez (1999) para la familia NEF-QVF. Estas distribuciones límites aparecen también cuando se consideran estimadores insesgados basados en la media Winsorizada en la distribución geométrica, ver López-Blázquez y Salamanca-Miño (2000), e incluso pueden aparecer en familias no exponenciales, tal como la distribución χ^2 -no centrada, ver López-Blázquez (2000).

1.7 Cotas inferiores para la varianza de estimadores insesgados

Sea $h \in \mathcal{U}_n$ y $V = V(X_1, \dots, X_n)$ un estimador insesgado de $h(\theta)$ (no necesariamente basado en el estimador suficiente y completo \bar{X}_n). Como $T_n(\bar{X}_n)$ es el estimador insesgado de mínima varianza, de (1.15),

$$Var_{\theta}(V) \geq Var_{\theta}(T_n(\bar{X}_n)) \geq \sum_{k=1}^m a_{k,n}^2(\theta) \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2 = B_{m,n}(h, \theta), \quad (1.44)$$

para todo $\theta \in \Theta$ y $m < d_n$. Entonces, la sucesión $\{B_{m,n}(h, \theta)\}_{m \geq 1}$ es un conjunto no decreciente de cotas inferiores para la varianza del estimador insesgado de $h(\theta)$ para un tamaño de muestra n . Además, si $g = h \circ (D_{\theta}\psi)^{-1}$, tenemos $B_{1,n}(h, \theta) = (D_{\mu}g(\mu))^2 V(\mu)/n$ que es la cota inferior de Cramer-Rao. En general, $B_{m,n}(h, \theta)$ es la m -ésima cota de Battacharyya. De (1.44), tenemos que $B_{m,n}(h, \theta)$ es alcanzable si y solo si $T_n(\bar{X}_n)$ es un polinomio de grado m en \bar{X}_n . Este resultado fue dado por Fend (1959), pero obtenido de forma distinta.

Una cuestión interesante es si la varianza del umvue es el límite de las cotas de Battacharyya. Por ejemplo, Khan (1980) investigó este pro-

blema en el caso particular de la familia NEF-QVF. Consideremos el caso $d_n = \infty$ (en el caso $d_n < \infty$ la sucesión de cotas de Bhattacharyya es finita, y una pregunta acerca del límite no tiene sentido). Bien, si $h \in \mathcal{U}_n$, la serie $\sum_{k=1}^{\infty} a_{k,n}^2(\theta) \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2$ es finita para cualquier $\theta \in \Theta$, y por tanto es obvio que

$$\lim_m B_{m,n}(h, \theta) = \sum_{k=1}^{\infty} a_{k,n}^2(\theta) \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2 = \text{Var}_{\theta}(T_n(\bar{X}_n)),$$

y como consecuencia, el límite de las cotas de Battacharyya es la varianza del estimador insesgado en cualquier NEF.

Algunas propiedades límites de las cotas de Battacharyya cuando el tamaño de muestra tiende a infinito pueden derivarse de algunos resultados de la sección 1.6, por ejemplo

$$\lim_n n^{m+1}(B_{m+1,n}(h, \theta) - B_{m,n}(h, \theta)) = \frac{(D_{\mu}^{m+1}g(\mu))^2}{(m+1)!} V^{m+1}(\mu).$$

1.8 Caracterización del conjunto de las funciones umvu estimables

Para decidir si una función analítica dada en Θ , $h(\theta)$, es una función umvu estimable para un tamaño de muestra n , lo que puede hacerse es obtener los coeficientes de Fourier $a_{k,n}(\theta)$ (los cuales pueden ser determinados por las derivadas de $h(\theta)$) y comprobar si

$$\sum_{k \geq 1} a_{k,n}^2(\theta) \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2 \tag{1.45}$$

es finito para cualquier $\theta \in \Theta$. Debemos distinguir dos casos.

Si d_n es finito, cualquier función $h(\theta)$ que pertenezca a \mathcal{U}_n tiene un estimador insesgado que puede ser escrito como

$$T_n(\bar{X}_n) = \sum_{k=0}^{d_n-1} a_{k,n}(\theta) P_{k,n}(\bar{X}_n; \theta).$$

Así $T_n(\bar{X}_n)$ es un polinomio de grado $d_n - 1$ en \bar{X}_n y el conjunto \mathcal{U}_n es el conjunto de las funciones tales que son los valores esperados de los polinomios de grado $d_n - 1$ en \bar{X}_n .

En el caso en que d_n sea infinito, (1.45) es una serie de términos positivos y una condición necesaria y suficiente para la convergencia es

$$\overline{\lim}_k (a_{k,n}^2(\theta) \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2)^{1/k} < 1, \text{ para } \theta \in \Theta. \quad (1.46)$$

Una condición suficiente para (1.46) es

$$\overline{\lim}_k \frac{a_{k+1,n}^2(\theta) \|P_{k+1,n}(\cdot; \theta)\|_{n,\theta}^2}{a_{k,n}^2(\theta) \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2} < 1, \text{ para } \theta \in \Theta.$$

Ejemplo 1.4 Sea X_1, \dots, X_n una muestra aleatoria simple procedente de una distribución gamma con parámetro $p > 0$ (conocido) y $\lambda > 0$, denotada $Ga(p, \lambda)$. La media y varianza de esta distribución son $\mu = p/\lambda$ y $V(\mu) = \mu^2/p$ respectivamente. Entonces, la $Ga(p, \lambda)$ es una NEF-QVF. De los ejemplos 1.1 y 1.2, la serie (1.45) es

$$\sum_{k=1}^{\infty} \frac{(D_{\mu}^k g(\mu))^2}{k!(np)_{[k]}} \mu^{2k}, \quad (1.47)$$

donde $(np)_{[k]} = (np)(np+1)\cdots(np+k-1)$, $k \geq 1$ y $(np)_{[0]} = 1$. Una función de la media, $g(\mu)$, es umvu estimable para un tamaño de muestra n si y solo si (1.47) es convergente para cualquier $\mu > 0$. En este caso el estimador umvu para un tamaño de muestra n viene dado por

$$T_n(\bar{X}_n) = \sum_{k=0}^{\infty} \frac{(-1)^k \mu^k D_{\mu}^k g(\mu)}{(np)_{[k]}} L_k^{(np-1)}(np\bar{X}_n/\mu), \quad (1.48)$$

donde

$$L_k^{(np-1)}(x) = \sum_{j=0}^k \binom{k+np-1}{k-j} (-x)^j / j!$$

es el k -ésimo polinomio generalizado de Laguerre, ver Chihara (1978), p.145. En particular, utilizando los criterios anteriores una función de la forma

$g(\mu) = \mu^{-\alpha}$, $\alpha > 0$, pertenece a \mathcal{U}_n si $n > 2\alpha/p$. Análogamente las funciones $g(\mu) = \exp(\beta\mu)$, son estimables mvu para cualquier real β y cualquier tamaño de muestra.

Nótese que la convergencia de la serie (1.45) es una condición necesaria y suficiente para la convergencia en sentido cuadrático de la serie

$$\sum_{k \geq 0} a_{k,n}(\theta) P_{k,n}(\bar{X}_n; \theta). \quad (1.49)$$

Pero puede ocurrir que la serie (1.45) diverja y (1.49) converja en otro sentido. Por ejemplo, para establecer una condición suficiente para la convergencia de (1.49) en el sentido \mathcal{L}^1 , notamos que

$$\int |P_{k,n}(s; \theta)| f_n(s; \theta) \nu_n(ds) \leq \|P_{k,n}(\cdot; \theta)\|_{n,\theta}, \quad \text{para todo } k \geq 0 \text{ y } \theta \in \Theta.$$

Así, (1.49) converge en el sentido \mathcal{L}^1 a cierta variable aleatoria $T_n(\bar{X}_n)$ si

$$\sum_{k \geq 0} |a_{k,n}(\theta)| \|P_{k,n}(\cdot; \theta)\|_{n,\theta} < \infty, \quad \text{para todo } \theta \in \Theta, \quad (1.50)$$

y en este caso $E_\theta[T_n(\bar{X}_n)] = h(\theta)$, para cualquier $\theta \in \Theta$, aunque no necesariamente $Var_\theta[T_n(\bar{X}_n)]$ es finita.

Ejemplo 1.5 Sea X_1, \dots, X_n una muestra aleatoria simple procedente de una NEF - QVF $(\nu; \theta)$, con $d_n = \infty$ (esto es, una NEF-QVF que no es del tipo binomial). Una condición suficiente para (1.50) es

$$\overline{\lim}_k k^{-\left(\frac{1}{2} + \frac{1}{4k}\right)} \left\{ \prod_{j=0}^{k-1} (n + v_{2j}) \right\}^{-1/(2k)} |D_\mu^k g(\mu)| < e^{1/2} V^{-1/2}(\mu), \quad \text{para todo } \mu \quad (1.51)$$

donde $g = h \circ (D_\theta \psi)^{-1}$. En particular, en la distribución gamma, ver ejemplo 1.4, las funciones $g(\mu) = \mu^{-\alpha}$, $\alpha > 0$ satisfacen (1.51) para cualquier $n > \alpha/p$. Entonces, comparando este resultado con el obtenido en el ejemplo 1.4, existen funciones paramétricas que admiten estimadores insesgados cuya varianza no es finita.

Nos hemos interesado también en la convergencia puntual de (1.49). En este sentido, tenemos el siguiente resultado para el caso $d_n = \infty$.

Teorema 1.6 *Si h es analítica en $\theta_0 \in \Theta$ y*

(i) *Existe $\lim_n a_{k,n}(\theta_0) = a_k(\theta_0)$, para todo $k \geq 0$,*

(ii) *La serie $\sum_{k \geq 0} a_k(\theta_0)(\mu - \mu_0)^k$ es absolutamente convergente para todo μ en un entorno de $\mu_0 = D_\mu \psi(\theta_0)$,*

entonces, existe una variable aleatoria

$$T_n(\bar{X}_n) = \sum_{k=0}^{\infty} a_{k,n}(\theta_0) P_{k,n}(\bar{X}_n; \theta_0), \text{ (c.s.)}$$

y tal que

$$T_n(\bar{X}_n) \xrightarrow{P_{\theta_0}} h(\theta_0), \quad (1.52)$$

cuando n tiende a infinito para todo $\theta \in \Theta$ en cierto entorno de θ_0 .

Demostración. Sea $\{X_j\}_{j \geq 0}$ una sucesión de v.a. i.i.d. de $NEF(\nu; \theta)$. Considérese

$$R_{m,n}(\bar{X}_n) = \sum_{k=m}^{\infty} a_{k,n}(\theta_0) P_{k,n}(\bar{X}_n; \theta_0),$$

y obsérvese que

$$\begin{aligned} 0 &\leq \overline{p \lim}_n |R_{m,n}(\bar{X}_n)| \leq \overline{p \lim}_n \sum_{k=m}^{\infty} |a_{k,n}(\theta_0)| |P_{k,n}(\bar{X}_n; \theta_0)| \\ &\leq \sum_{k=m}^{\infty} \lim_n |a_{k,n}(\theta_0)| \overline{p \lim}_n |P_{k,n}(\bar{X}_n; \theta_0)| = \sum_{k=m}^{\infty} a_k(\theta_0) |\mu - \mu_0|^k \rightarrow 0, \end{aligned}$$

cuando m tiende a infinito, para todo μ en cierto entorno de μ_0 . Esto implica que para un n suficientemente grande, la serie $\sum_{k=0}^{\infty} a_{k,n}(\theta_0) P_{k,n}(\bar{X}_n; \theta_0)$

converge absolutamente a cierta variable aleatoria $T_n(\bar{X}_n)$ y

$$\begin{aligned}
p \lim_n T_n(\bar{X}_n) &= p \lim_n \sum_{k=0}^{\infty} a_{k,n}(\theta_0) P_{k,n}(\bar{X}_n; \theta_0) = \\
&= \sum_{k=0}^{\infty} \lim_n a_{k,n}(\theta_0) p \lim_n P_{k,n}(\bar{X}_n; \theta_0) = \\
&= \sum_{k=0}^{\infty} a_k(\theta_0) (\mu - \mu_0)^k. \tag{1.53}
\end{aligned}$$

En particular, como $a_{0,n}(\theta_0) = h(\theta_0)$, de (1.53) escogiendo $\mu = \mu_0$ ($\theta = \theta_0$), obtenemos (1.52). \square

En la sección 1.6, se mostró que la sucesión de conjuntos $\{\mathcal{U}_n\}_{n \geq 1}$ es no decreciente en el sentido de la inclusión de conjuntos. Sea $\mathcal{U} = \bigcup_{n=1}^{\infty} \mathcal{U}_n$. Diremos que \mathcal{U} es el conjunto de las funciones estimable mvu de la $NEF(\nu; \theta)$.

Para caracterizar las funciones pertenecientes a \mathcal{U} , de nuevo, distinguiémos dos casos.

La medida ν tiene soporte finito. En este caso, $d_n = \dim(\mathcal{L}_{n,\theta}^2) < \infty$, para todo $n \geq 1$, y \mathcal{U} es el conjunto de funciones que tienen como estimadores insesgados polinomios en \bar{X}_n . Una función h pertenece a \mathcal{U} si y sólo si h es una combinación lineal finita (con coeficientes reales que no dependen de θ) de los momentos $m_{k,n}(\theta) = E_{\theta}[\bar{X}_n^k]$.

La medida ν tiene soporte infinito. En este caso, $d_n = \dim(\mathcal{L}_{n,\theta}^2) = \infty$, y el SPMO es infinito para todo $n \geq 1$. Dada una función h que pertenece a \mathcal{U}_{n_0} , consideraremos su estimador insesgado para un tamaño de muestra $n \geq n_0$, $T_n(\bar{X}_n) = \sum_{k=0}^{\infty} a_{k,n}(\theta) P_{k,n}(\bar{X}_n; \theta)$. Entonces h es analítica en Θ y la serie de términos positivos

$$\sum_{k=1}^{\infty} a_{k,n}^2(\theta) \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2$$

es finita para cualquier $\theta \in \Theta$ y $n \geq n_0$. Tenemos,

$$0 = \lim_n \text{Var}_{\theta}(T_n(\bar{X}_n)) = \lim_n \sum_{k=1}^{\infty} a_{k,n}^2(\theta) \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2$$

$$\begin{aligned}
&= \sum_{k=1}^{\infty} \lim_n a_{k,n}^2(\theta) \lim_n \|P_{k,n}(\cdot; \theta)\|_{n,\theta}^2 = \sum_{k=1}^{\infty} (D_{\mu}^k g(\mu)/k!)^2 \lim_n k! (V(\mu)/n)^k \\
&= \lim_n \sum_{k=1}^{\infty} (D_{\mu}^k g(\mu))^2 (V(\mu)/n)^k / k!,
\end{aligned}$$

entonces, la serie de términos positivos $\sum_{k=1}^{\infty} (D_{\mu}^k g(\mu))^2 (V(\mu)/n)^k / k!$ converge (absolutamente) para todo $\theta \in \Theta$ y cualquier n mayor que cierto entero positivo n_1 . El criterio de la raíz para la convergencia absoluta de una serie, da

$$\overline{\lim}_k \{(D_{\mu}^k g(\mu))^2 (V(\mu)/n)^k / k!\}^{1/k} < 1, \text{ para todo } \theta \in \Theta \text{ y } n \geq n_1, \quad (1.54)$$

Usando la aproximación de Stirling, $k! \sim \sqrt{2\pi} k^{k+1/2} e^{-k}$, obtenemos de (1.54) que existe un número real positivo M tal que

$$\overline{\lim}_k k^{-(\frac{1}{2} + \frac{1}{2k})} (D_{\mu}^k g(\mu))^{2/k} < M/V(\mu), \text{ para todo } \theta \in \Theta. \quad (1.55)$$

Nótese que la condición (1.55) es una condición necesaria para que una función analítica en Θ , h , pertenezca a \mathcal{U} . No es difícil ver que esta condición es también suficiente.

Ejemplo 1.6 *La distribución de Poisson con parámetro μ tiene función de varianza $V(\mu) = \mu$. Usando (1.55), se puede comprobar que para cualquier tamaño de muestra las funciones $g(\mu) = \mu^{-\alpha}$, $\alpha > 0$ no son umvu estimables.*

1.9 Comparación entre el emv y el umvue

Para simplificar, consideraremos funciones paramétricas de la media de una $NEF(\nu; \theta)$, $g(\mu)$, tal que $h(\theta) = g(\mu(\theta)) \in \mathcal{U}$. El estimador de máxima verosimilitud (emv) de $g(\mu)$ es $g(\bar{X}_n)$, y el umvu estimador es (para todo n mayor que cierto entero positivo n_0),

$$T_n(\bar{X}_n) = \sum_{k \geq 0} a_{k,n}(\theta) P_{k,n}(\bar{X}_n; \theta), \text{ para todo } \theta \in \Theta. \quad (1.56)$$

Notamos que la elección de θ en el lado derecho de (1.56) es arbitraria. Si escogemos $\theta = \hat{\theta}_n$, la solución de la ecuación de máxima verosimilitud

$$\sum_{i=1}^n (X_i - \mu(\theta)) = 0,$$

como $a_{0,n}(\theta) = h(\theta)$, obtenemos

$$T_n(\bar{X}_n) = g(\bar{X}_n) + \sum_{k \geq 1} a_{k,n}(\hat{\theta}_n) P_{k,n}(\bar{X}_n; \hat{\theta}_n). \quad (1.57)$$

Ejemplo 1.7 En la distribución gamma, ver ejemplo 1.4, se substituye $\mu = \bar{X}_n$ en el lado derecho de (1.48), obtenemos

$$T_n(\bar{X}_n) = \sum_{k=0}^{\infty} \frac{(-1)^k \bar{X}_n^k D_{\mu}^k g(\bar{X}_n)}{(np)_{[k]}} L_k^{(np-1)}(np), \quad (1.58)$$

Nótese que (1.58) coincide con la expresión del estimador umvu dado por Woodward y Gray (1975). Ellos derivaron esta fórmula aplicando el método del jacknife generalizado al estimador de máxima verosimilitud y posteriormente aplicando al resultado el procedimiento de reducción de varianza de Blackwell-Rao. Las condiciones bajo las que estos resultados pueden ser aplicados no aparecen de forma clara en su trabajo, básicamente es por que en sus demostraciones la función $g(\mu)$ tiene que satisfacer ciertas condiciones de regularidad que no son especificadas. En la sección 1.10, discutiremos estas condiciones de regularidad, ver también López-Blazquez (1998).

Ejemplo 1.8 Sea X_1, \dots, X_n una muestra aleatoria simple de una distribución Bernouilli con parámetro $p \in (0, 1)$. Esta distribución pertenece a la familia NEF-QVF con $\mu = p$ y $V(\mu) = -\mu^2 + \mu$, ($v_2 = -1$). El estimador insesgado de la función $g(p) \in \mathcal{U}_n$ (un polinomio de grado a lo sumo n) dado en (1.57) es

$$T_n(\hat{p}) = \sum_{k=0}^n \frac{n^k D_p^k g(\hat{p})}{k! n(n-1) \dots (n-k+1)} \hat{\pi}_{k,0,n}, \quad \hat{p} = \bar{X}_n, \quad (1.59)$$

donde los $\hat{\pi}$ satisfacen (ver (1.8)),

$$\begin{aligned}\hat{\pi}_{k+1,0,n} &= -k\left\{(1-2\hat{p})\hat{\pi}_{k,0,n} + \frac{n-k+1}{n^2}\hat{p}(1-\hat{p})\hat{\pi}_{k-1,0,n}\right\}, \\ \hat{\pi}_{0,0,n} &= 1; \quad \hat{\pi}_{1,0,n} = 0.\end{aligned}$$

A partir de 1.26, se tiene que $P_{k,n}(\bar{X}_n; \hat{\theta}) = \pi_{k,0,n}(\hat{\theta}_n)$. Mas aún, de acuerdo con el corolario 1.2

$$\lim_n (n/V(\mu))^{k/2} \pi_{k,0,n}(\theta) = \text{coef}(0, He_k).$$

Por otra parte, del teorema 1.3, obtenemos

$$p \lim_n a_{k,n}(\hat{\theta}_n) = D^k g(\mu)/k!.$$

De (1.7), tenemos que $P_{1,n}(\bar{X}_n; \hat{\theta}_n) = 0$, y de (1.57),

$$T_n(\bar{X}_n) - g(\bar{X}_n) = \pi_{2,0,n} a_{2,n}(\hat{\theta}_n) + \sum_{k \geq 3} \pi_{k,0,n}(\hat{\theta}_n) a_{k,n}(\hat{\theta}_n),$$

por lo cual no es difícil mostrar que

$$p \lim_n n \{T_n(\bar{X}_n) - g(\bar{X}_n)\} = V(\mu) D_\mu^2 g(\mu)/2 \quad (1.60)$$

Un resultado mas general es obtenido considerando el estimador

$$C_{m,n}(\bar{X}_n) = \sum_{k=0}^m a_{k,n}(\hat{\theta}_n) P_{k,n}(\bar{X}_n; \hat{\theta}_n), \quad m < d_n - 1.$$

Este estimador podría interpretarse como una versión corregida del estimador mve.

Siguiendo argumentos similares, como arriba, obtenemos los siguientes resultados.

Teorema 1.7 Sea $g(\mu)$ tal que $h(\theta) = g(\mu(\theta)) \in \mathcal{U}$, entonces,

$$\begin{aligned}p \lim_n n^{k/2} \{T_n(\bar{X}_n) - C_{m,n}(\bar{X}_n)\} &= \\ &= \begin{cases} 0, & k = 2r - 1, \quad r \geq 1 \\ (2r - 1)!! V^r(\mu) D_\mu^{2r} g(\mu) / (2r)!, & k = 2r, \quad r \geq 1 \end{cases}\end{aligned}$$

Como consecuencia, los estimadores $C_{m,n}(\bar{X}_n)$ son asintóticamente insesgados y si $D_\mu g(\mu) \neq 0$,

$$\sqrt{n} \{C_{m,n}(\bar{X}_n) - g(\mu)\} \xrightarrow{d} \mathcal{N}(0, (D_\mu g(\mu))^2 V(\mu)), \text{ cuando } n \rightarrow \infty.$$

Incluso en el caso en que la función $g(\mu(\theta))$ no pertenezca a \mathcal{U} , pero admite k derivadas continuas, nosotros podemos usar $C_{m,n}(\bar{X}_n)$ como un estimador de $g(\mu)$, tal como se pone de manifiesto en el siguiente ejemplo.

Ejemplo 1.9 *En la distribución Bernouilli con parámetro p , ver ejemplo 1.8, la función e^p no pertenece a \mathcal{U}_n para ningún n . Para estudiar el comportamiento del estimador $C_{3,n}(\bar{X}_n)$, simulamos 1000 muestras aleatorias de tamaño $n = 10$ de una distribución $Be(0.3)$. El valor exacto $e^p = 1.3498$ fue estimado usando el estimador de máxima verosimilitud y $C_{3,10}$. En esta simulación, obtuvimos*

$$\begin{aligned} \text{sesgo}(emv) &= 0.0269; \text{ sesgo}(C_{3,10}) = 0.0150; \\ \text{ecm}(emv) &= 0.0417; \text{ ecm}(C_{3,10}) = 0.0382; \end{aligned}$$

Se puede ver que el sesgo y el error cuadrático medio son menores cuando se usa $C_{3,10}$ en vez del emv.

1.10 Una extensión de la clase \mathcal{U}

La extensión que proponemos, viene motivada a que la condición de insesgades ,

$$E_\theta[T_n(\bar{X}_n)] = h(\theta), \text{ para todo } \theta \in \Theta,$$

puede ser muy restrictiva en algunos casos de interés.

Dada una $NEF(v; \theta)$, consideremos Θ_0 un subconjunto abierto no vacío del espacio paramétrico Θ , y sea X_1, \dots, X_n una muestra aleatoria simple de una $NEF(v; \theta)$. Una función del parámetro, $h(\theta)$, se dice que es estimable umvu en Θ_0 para un tamaño de muestra n si existe una función, T_n , satisfaciendo:

$$(i) T_n(s) \in \bigcap_{\theta \in \Theta_0} \mathcal{L}_{n,\theta}^2$$

$$(ii) E_\theta[T_n(\bar{X}_n)] = h(\theta), \text{ para todo } \theta \in \Theta_0.$$

En tal caso $T_n(\bar{X}_n)$ es un estimador mvu de $h(\theta)$ para un tamaño de muestra n en Θ_0 . Denotamos por $\mathcal{U}_n(\Theta_0)$ al conjunto de las funciones estimable umvu para el tamaño de muestra n en Θ_0 . Es obvio que en el caso $\Theta_0 = \Theta$, tengamos que $T_n(\bar{X}_n)$ es el estimador mvu (usual) de $h(\theta)$.

Es importante notar aquí que todos los resultados obtenidos en este trabajo se verifican si sustituimos la frase 'para todo $\theta \in \Theta$ ' por 'para todo $\theta \in \Theta_0$ '.

Algunos otros desarrollos son posibles relajando la condición de insesgadez. Supongamos que $h(\theta)$ es una función que no pertenece a \mathcal{U} , pero que existe una sucesión de funciones $\{h_n(\theta)\}_n$ tal que $h_n \in \mathcal{U}_n$, para todo $n \geq 1$ y $\lim_n h_n(\theta) = h(\theta)$, para todo $\theta \in \Theta_0$ un subconjunto no vacío de Θ . La sucesión de los correspondientes estimadores insesgados $\{T_n(\bar{X}_n)\}_n$ (donde $h_n(\theta) = E_\theta[T_n(\bar{X}_n)]$, para todo $\theta \in \Theta_0$) es (bajo condiciones que deben ser precisadas en cada caso) asintóticamente insesgado y asintóticamente normal.

Capítulo 2

Algoritmo para la Obtención de Estimadores Insesgados en Distribuciones de Series de Potencia

2.1 Introducción

Sea $g(\theta) = \sum_{x \geq 0} g_x \theta^x$, una función analítica, con radio de convergencia $R > 0$ (posiblemente infinito) y tal que $g_x \geq 0$, para todo $x \geq 0$. Una variable aleatoria discreta X cuya función de probabilidad es de la forma

$$P_\theta [X = x] = g_x \theta^x / g(\theta), \quad x = 0, 1, \dots; \quad \theta \in (0, R),$$

se dice que sigue una distribución de series de potencia (psd), ver Kosambi (1949) y Noack (1950). Se denotará $X \sim PSD(g(\theta))$. Esta familia ha sido estudiada muy extensamente, para una lista de buenas referencias ver Johnson, Kotz y Kemp (1992) y las referencias allí incluidas. Algunas extensiones importantes de esta familia son, por ejemplo, la distribución de serie de potencia modificada (mpsd), Gupta (1974), que también ha sido ampliamente investigada. Aunque solo se considerará en este capítulo la familia psd, los

resultados que se obtienen pueden ser ampliados sin mucha dificultad a la familia mpsd.

Sea X_1, \dots, X_n una muestra aleatoria simple de una $PSD(g(\theta))$. El estadístico $S_n = \sum_{i=1}^n X_i$ es suficiente, completo y sigue una $PSD(g^n(\theta))$. Siguiendo el resultado clásico de Lehman-Scheffe-Blackwell-Rao, la unicidad del estimador umvu de una función paramétrica es obtenida como una función de S_n , si éste existe. Condiciones para la existencia del estimador umvu en la familia psd han sido investigado por Patil (1963). Un método básico para la obtención del estimador umvu, $T(S_n)$, de una función del parámetro,

$$h(\theta) = \sum_{j=0}^{\infty} h_j \theta^j, \theta \in (0, R),$$

en una $PSD(g(\theta))$ fue dado por Roy y Mitra (1957), que obtuvieron la siguiente expresión

$$T(s) = \left(\sum_{j=0}^s h_j g_{s-j,n} \right) / g_{s,n}, \text{ siempre que } g_{s,n} \neq 0. \quad (2.1)$$

Aunque (2.1) es una fórmula fácil, no puede usarse fácilmente en ciertas circunstancias, por ejemplo en el caso de que el valor del estadístico suficiente $S_n = s$ sea moderadamente grande, digamos que algunos cientos (lo cual es una situación usual). Entonces, necesitamos conocer esos cientos de coeficientes del desarrollo de $h(\theta)$ y $g^n(\theta)$ los cuales no son fáciles de obtener. Más aún, cuando s crece $g_{s,n}$ usualmente tiende a cero muy rápido, entonces la evaluación numérica del cociente en el miembro derecho de (2.1) puede dar lugar a resultados numéricamente inestables.

Es bien conocido que la truncación (derecha, izquierda o ambos) de una psd es también una psd. La estimación de mínima varianza para ciertas familias psd truncadas ha sido investigada, entre otros por, Joshi y Park (1974), Charalambides (1974), Jani (1977, 1978), Kumar y Consul (1980) y Voinov (1986). Básicamente, estos autores usaron (2.1) y los estimadores insesgados obtenidos se expresan en términos de números de Stirling Gene-

realizados definidos adecuadamente para cada situación estudiada. Desafortunadamente, estos métodos solo pueden ser aplicados en una subclase limitada (pero importante) de la familia psd, entre otras, a la serie logarítmica truncada, la Poisson decapitada y la binomial negativa truncada. Es más, ni la varianza ni un estimador de la varianza pueden obtenerse mediante los métodos desarrollados por los autores anteriormente mencionados. Este inconveniente limita la posibilidad de realizar más inferencias a partir de los datos, tal como la construcción de intervalos de confianza.

En este capítulo se verá un algoritmo general para la obtención de estimadores insesgados en las familias psd. Este método se basa en los desarrollos de los estimadores insesgados en términos de polinomios ortogonales. Hace algunos años, este método se podría haber considerado impracticable, pero hoy en día éste no es el caso, gracias a los software matemáticos (como MAPLE o MATHEMATICA entre otros), capaces de manipular expresiones algebraicas simbólicas y potentes rutinas numéricas.

2.2 Estimación insesgada en la familia psd

Como se ha mencionado arriba, nuestro algoritmo está basado en los desarrollos ortogonales para los estimadores insesgados. En esta sección se estudiarán algunos hechos básicos sobre estos desarrollos.

Dado un número real, r , y un entero positivo j , se denotará

$$r^{[j]} = r(r-1)\cdots(r-j+1),$$

y $r^{[0]} = 1$.

Sea X_1, \dots, X_n una muestra aleatoria simple de una $PSD(g(\theta))$. Es bien conocido el resultado de que $S_n = \sum_{i=1}^n X_i$ es suficiente, completo y sigue una $PSD(g^n(\theta))$. Los momentos factoriales de cualquier orden de S_n existen y son obtenidos como:

$$f_{j,n}(\theta) = E_{\theta}[S_n^{[j]}] = \frac{\theta^j D^j(g^n(\theta))}{g^n(\theta)}, \quad j \geq 0, \quad (2.2)$$

donde D denota el operador derivada con respecto a θ .

Sea

$$\text{sop}(S_n) = \{s : P_\theta[S_n = s] > 0\}$$

el soporte de S_n . Para $\theta \in (0, R)$, consideremos

$$\mathcal{L}_{n,\theta}^2 = \{T : E_\theta[T^2(S_n)] < \infty\},$$

que es el conjunto de funciones reales de cuadrado integrable definidas en $\text{sop}(S_n)$. Se sunpondra que dos funciones $T_1, T_2 \in \mathcal{L}_{n,\theta}^2$ son iguales (ó equivalentes) si y solo si $v_n \{T_1 \neq T_2\} = 0$, con v_n la medida cardinal en $\text{sop}(S_n)$. Se puede comprobar fácilmente que $\mathcal{L}_{n,\theta}^2$ es un espacio lineal de funciones cuya dimensión coincide con el cardinal del soporte de S_n .

Consideremos en $\mathcal{L}_{n,\theta}^2$ el producto escalar

$$\langle T_1, T_2 \rangle_\theta = E_\theta[T_1(S_n)T_2(S_n)].$$

Entonces $(\mathcal{L}_{n,\theta}^2, \langle \cdot, \cdot \rangle_\theta)$ es un espacio de Hilbert. El conjunto de polinomios, $\{s^{[j]}\}_{j \geq 0}$ es completo en $\mathcal{L}_{n,\theta}^2$. Este conjunto puede ser ortonormalizado, utilizando, por ejemplo, el método de Gram-Schmidt, para obtener un sistema de polinomios ortogonales, $\{P_{k,n}(s, \theta)\}_{k \geq 0}$. El número de polinomios en el sistema ortogonal es igual a $d = \dim(\mathcal{L}_{n,\theta}^2) = \#\text{sop}(S_n)$. Estos polinomios pueden ser escritos como

$$P_{k,n}(s, \theta) = \sum_{j=0}^k \alpha_{k,j}(\theta) s^{[j]}, \quad k = 1, \dots, d-1, \quad (2.3)$$

con $P_{0,n}(s, \theta) = 1$.

Cualquier función $T(s) \in \mathcal{L}_{n,\theta}^2$ admite un único desarrollo en términos de los polinomios ortonormales de la forma

$$T(s) = \sum_{k \geq 0} c_k(\theta) P_{k,n}(s, \theta), \quad \text{con } c_k(\theta) = \langle T, P_{k,n} \rangle_\theta. \quad (2.4)$$

Usando la propiedad de ortogonalidad de los polinomios se puede comprobar que

$$E_\theta[T(S_n)] = c_0(\theta) \quad \text{y} \quad \text{Var}_\theta(T(S_n)) = \sum_{k \geq 1} c_k^2(\theta). \quad (2.5)$$

Sea $h(\theta)$ una función paramétrica que admite un estimador insesgado, $T(S_n)$, para un tamaño de muestra n con varianza finita, es decir,

$$(i) E_{\theta}[T(S_n)] = h(\theta), \text{ para todo } \theta \in (0, R).$$

$$(ii) Var_{\theta}(T(S_n)) < \infty, \text{ para todo } \theta \in (0, R).$$

En esta situación, $T(S_n)$ es el estimador umvu de $h(\theta)$. Nótese que la condición (ii) es equivalente a decir que $T \in \mathcal{L}_{n,\theta}^2$ para todo $\theta \in (0, R)$.

De (i), se tiene,

$$\sum_{s \geq 0} T(s) g_{s,n} \theta^s = g^n(\theta) h(\theta), \quad \theta \in (0, R). \quad (2.6)$$

Derivando j veces en ambos miembros de (2.6), usando la fórmula de Leibnitz para la derivada de un producto y (2.2), se obtiene

$$\sum_{s \geq 0} T(s) s^{[j]} g_{s,n} \theta^{s-j} = g^n(\theta) \sum_{m=0}^j \binom{j}{m} \theta^{m-j} f_{j-m,n}(\theta) D^m h(\theta), \quad \theta \in (0, R),$$

ó equivalentemente,

$$\langle T(s), s^{[j]} \rangle_{\theta} = \sum_{m=0}^j \binom{j}{m} \theta^m f_{j-m,n}(\theta) D^m h(\theta), \quad \theta \in (0, R). \quad (2.7)$$

De (2.4) y (2.7), obtenemos

$$\begin{aligned} c_k(\theta) &= \sum_{m=0}^k \left\{ \sum_{j=m}^k \alpha_{k,j}(\theta) \binom{j}{m} f_{j-m,n}(\theta) \right\} \theta^m D^m h(\theta), \quad k \geq 1, \quad \theta \in (0, R). \\ c_0(\theta) &= h(\theta) \end{aligned} \quad (2.8)$$

Como conclusión de (2.8), se tiene que los coeficientes en el desarrollo ortogonal del estimador umvu de $h(\theta)$ se pueden obtener sin un conocimiento explícito de T , solo se necesitan las derivadas de h . Nótese también que hay

muchas representaciones de T en series ortogonales: una para cada posible valor del parámetro θ .

El procedimiento arriba descrito para la obtención de estimadores insesgados puede parecer un poco complicado y difícil en situaciones prácticas. Sin embargo en este capítulo se mostrará lo contrario: este método puede ser fácilmente implementado en el ordenador usando un software adecuado. Aquí 'software adecuado' significa un programa con cálculo simbólico, álgebra matricial y manipulación de series de potencia.

Hay algunos casos particulares en los cuales el método descrito arriba lleva a expresiones cerradas de estimadores insesgados. Veamos algunos ejemplos de distribuciones psd en las cuales el sistema de polinomios ortogonales es bien conocido, se verá también el desarrollo ortogonal de los estimadores insesgados en estas familias.

a) *La distribución binomial.* Una distribución binomial, $Bi(N, p)$, con N un entero positivo conocido y $p \in (0, 1)$ es una $PSD((1 + \theta)^N)$ con $\theta = p/q$, $q = 1 - p$. En este caso, $d = \dim(\mathcal{L}_{n,\theta}^2) = Nn + 1$. Un sistema mónico de polinomios ortogonales en $\mathcal{L}_{n,\theta}^2$ es

$$P_{k,n}(s, \mu) = k! k_k \left(s; nN, \frac{\mu}{N} \right), k = 0, \dots, Nn, \mu = Np,$$

con $k_k(s; N, p) = \sum_{j=0}^k \binom{s}{k-j} \binom{N-s}{j} (-1)^j p^j q^{k-j}$, el k -ésimo polinomio de Krawtchouk.

El desarrollo ortogonal del estimador insesgado de una función $h(p)$ es

$$T(S_n) = \sum_{k=0}^{Nn} \frac{D^k h(p) \Gamma(Nn - k + 1) N^k}{\Gamma(Nn + 1)} k_k(s; nN, p).$$

b) *La distribución binomial negativa.* Una distribución binomial negativa, $NB(r, p)$, con $r > 0$ (conocido) y $p \in (0, 1)$ es una $PSD((1 - \theta)^{-r})$ con $\theta = 1 - p$. En este caso $d = \infty$, entonces el sistema de polinomios mónicos ortogonales en $\mathcal{L}_{n,\theta}^2$ es infinito y viene dado por

$$P_{k,n}(s, p) = (-1)^k \left(\frac{1-p}{p} \right)^k m_k(s; nr, 1-p), k \geq 0,$$

con $m_k(s; \alpha, q) = (-1)^k k! \sum_{j=0}^k \binom{s}{k-j} \binom{-s-\alpha}{k-j} q^{-j}$, el k -ésimo polinomio de Meixner de primera clase.

El desarrollo ortogonal del estimador insesgado de una función $h(p)$ es

$$T(S_n) = \sum_{k=0}^{\infty} \frac{D^k h(p) \Gamma(nr) (r(1-p))^k}{\Gamma(nr+k) p^k k!} m_k(S_n; nr, 1-p).$$

c) *La distribución de Poisson.* Una distribución de Poisson, $P(\theta)$, $\theta > 0$, es una $PSD(\exp(\theta))$. De nuevo $d = \infty$ y el sistema de polinomios mónicos ortogonales en $\mathcal{L}_{n,\theta}^2$ es infinito y viene dado por

$$P_{k,n}(s, \theta) = C_k^{(n\theta)}(s) = \sum_{j=0}^k \binom{k}{j} \binom{s}{j} j! (-n\theta)^{k-j}, k \geq 0,$$

con $C_k^{(n\theta)}$ el k -ésimo polinomio de Charlier.

El desarrollo ortogonal del estimador insesgado de una función $h(p)$ es

$$T(S_n) = \sum_{k=0}^{\infty} \frac{D^k h(\theta)}{n^k k!} C_k^{(n\theta)}(S_n).$$

Las expresiones de los estimadores insesgados dados en los ejemplos anteriores pueden ser obtenidas a partir de los resultados del capítulo 1 y de Morris (1983). Las propiedades utilizadas sobre los polinomios ortogonales pueden encontrarse en Chihara (1978).

En general, sea $h(\theta)$ una función paramétrica de una $PSD(g(\theta))$ y $T(S_n)$ un estimador insesgado para el tamaño de muestra n . Se mostró en la introducción que $T(S_n)$ admite un desarrollo ortogonal de la forma

$$T(S_n) = \sum_{k=0}^{d-1} c_k(\theta_0) P_{k,n}(S_n, \theta_0), \theta_0 \in (0, R). \quad (2.9)$$

En la práctica los siguientes problemas surgen cuando se trata de usar la expresión (2.9) para obtener un estimador insesgado de $h(\theta)$:

(i) Puede ocurrir que $d = \infty$ o d sea finito pero muy grande. Esta dificultad puede ser reducida si en vez de utilizar el estimador insesgado se usa una aproximación truncando (2.9):

$$\tilde{T}_m(S_n) = \sum_{k=0}^m c_k(\theta_0) P_{k,n}(s, \theta_0), \quad \theta_0 \in (0, R), \quad y, \quad m < d \quad (2.10)$$

con m pequeño pero lo suficientemente grande para que la diferencia entre $T(S_n)$ y $\tilde{T}_m(S_n)$ sea pequeña con probabilidad próxima a uno.

En cualquier caso, el objetivo final es implementar en el ordenador un procedimiento para calcular (2.9), necesariamente, en muchos casos, se deberá truncar el número de términos en el lado derecho de (2.9), dada la imposibilidad de trabajar con series infinitas o con un gran número de términos. Para tamaños de muestra moderado (por ejemplo n sobre 20), se observa que escogiendo $m = 4$ se obtienen aproximaciones satisfactorias.

(ii) *La obtención de polinomios ortonormales.* En el primer capítulo se dieron varios métodos para la obtención de polinomios ortonormales. Por lo tanto para la familia psd, puede usarse cualquiera de ellos, por ejemplo, el método de Gram-Schmidt. Alternativamente, en la sección 2.3 se da otro procedimiento para la obtención de polinomios ortonormales hasta un grado m .

(iii) *La obtención de los coeficientes de Fourier c_k .* Se verá en la sección 4 que estos coeficientes pueden ser obtenidos multiplicando cierta matriz \mathbb{E}_m por el vector de derivadas $\mathbb{D}_m^t = (h(\theta_0), h'(\theta_0), \dots, h^{(m)}(\theta_0))$. La matriz \mathbb{E}_m es independiente de la función $h(\theta)$, así que podrá ser reutilizada para la estimación de cualquier otra función paramétrica usando los mismos datos. Sucesivas derivadas de una función pueden ser fácilmente obtenidas por el método FFT, ver Henrici (1993), vol. 3. Afortunadamente, estas rutinas para la obtención de derivadas están implementadas en los paquetes estadísticos que se mencionaron en la introducción. Los coeficientes de Fourier son también útiles para la obtención de un estimador de la varianza del estimador insesgado (aproximado).

2.3 Polinomios ortonormales

La aproximación del estimador insesgado dado en (2.10) requiere polinomios ortonormales hasta un grado $m < d$. Un método para la obtención de $P_k = P_{k,n}(S_n, \theta_0)$, para $k = 0, \dots, m$ es el siguiente.

De (2.3), usando notación matricial y omitiendo argumentos por simplicidad, se tiene

$$\begin{bmatrix} P_0 \\ P_1 \\ P_2 \\ \vdots \\ P_m \end{bmatrix} = \begin{bmatrix} \alpha_{0,0} & 0 & 0 & \dots & 0 \\ \alpha_{1,0} & \alpha_{1,1} & 0 & \dots & 0 \\ \alpha_{2,0} & \alpha_{2,1} & \alpha_{2,2} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \alpha_{m,0} & \alpha_{m,1} & \alpha_{m,2} & \dots & \alpha_{m,m} \end{bmatrix} \begin{bmatrix} 1 \\ S_n \\ S_n^{[2]} \\ \vdots \\ S_n^{[m]} \end{bmatrix} \quad (2.11)$$

ó bien:

$$\mathbb{P}_m = \mathbb{A}_m \mathbb{S}_m. \quad (2.12)$$

Entonces, multiplicando en (2.11) por las traspuestas y tomando esperanza en ambos miembros, se obtiene

$$E_{\theta_0}[\mathbb{P}_m \mathbb{P}_m^t] = \mathbb{A}_m E_{\theta_0}[\mathbb{S}_m \mathbb{S}_m^t] \mathbb{A}_m^t. \quad (2.13)$$

Dada la ortogonalidad de los polinomios, se tiene

$$(E_{\theta_0}[\mathbb{P}_m \mathbb{P}_m^t])_{i,j} = \delta_{i,j}$$

por lo tanto, el lado izquierdo de (2.13) es la matriz identidad de dimensión $m + 1$, \mathbb{I}_{m+1} .

Denotemos por $\mathbb{G}_m = E_{\theta_0}[\mathbb{S}_m \mathbb{S}_m^t] = (g_{i,j}(\theta_0))_{i,j=0,\dots,m}$ que es una matriz simétrica y definida positiva cuyos elementos

$$g_{i,j}(\theta_0) = E_{\theta_0}[S_n^{[i]} S_n^{[j]}]$$

se pueden obtener usando la siguiente relación de recurrencia:

$$\begin{aligned}
g_{0,j}(\theta_0) &= f_{j,n}(\theta_0), & j \geq 0, \\
g_{i+1,j}(\theta_0) &= g_{i,j+1}(\theta_0) + (j-i)g_{i,j}(\theta_0), & i = 0, \dots, m; j \geq i+1 \\
g_{j,i}(\theta_0) &= g_{i,j}(\theta_0), & j < i+1.
\end{aligned} \tag{2.14}$$

Entonces, (2.13) quedaría

$$\mathbb{I}_{m+1} = \mathbb{A}_m \mathbb{G}_m \mathbb{A}_m^t.$$

de donde

$$\mathbb{G}_m = \mathbb{A}_m^{-1} (\mathbb{A}_m^{-1})^t. \tag{2.15}$$

De (2.15), se tiene que \mathbb{G}_m es una matriz positiva y como \mathbb{A}_m^{-1} es una matriz triangular inferior, la descomposición de \mathbb{G}_m como $\mathbb{A}_m^{-1} (\mathbb{A}_m^{-1})^t$ es la descomposición de Cholesky. Dada una matriz positiva \mathbb{M} con descomposición de Cholesky $\mathbb{M} = \mathbb{C}\mathbb{C}^t$, con \mathbb{C} una matriz triangular inferior, denotaremos $\mathbb{C} = \text{Cholesky}(\mathbb{M})$. Entonces, en este caso, la matriz de coeficientes de los polinomios ortonormales es

$$\mathbb{A}_m = (\text{Cholesky}(\mathbb{G}_m))^{-1}. \tag{2.16}$$

Para el cálculo de \mathbb{A}_m se tiene que invertir la matriz $\text{Cholesky}(\mathbb{G}_m)$. Afortunadamente, esta es una matriz triangular inferior y la inversión de esta clase de matrices requiere muchos menos cálculos que la inversión de una matriz completa. La descomposición de Cholesky está implementada como una subrutina en muchos paquetes matemáticos.

Un procedimiento alternativo para la obtención de polinomios ortonormales es el siguiente:

1. Se construye la $(m+1) \times (2m+2)$ matriz $\mathbb{G} = (\mathbb{G}_m | \mathbb{I}_{m+1})$, donde $\mathbb{G}_m = E_{\theta_0}[\mathbb{S}_m \mathbb{S}_m^t] = (g_{i,j}(\theta_0))_{i,j=0,\dots,m}$ e \mathbb{I}_{m+1} es la $(m+1) \times (m+1)$ matriz identidad.
2. Usando el método de eliminación por filas de Gauss se obtiene $(\mathbb{T}_{m+1} | \mathbb{B}_{m+1})$, donde \mathbb{T}_{m+1} es una matriz triangular superior de dimensión $(m+1) \times (m+1)$.

Se puede comprobar que \mathbb{B}_{m+1} es una matriz triangular inferior $(m+1) \times (m+1)$ con todos sus elementos en la diagonal iguales a 1. La fila $(k+1)$ -ésima de \mathbb{B}_{m+1} da los coeficientes en orden ascendente del polinomio ortogonal mónico de grado k , es decir, $\beta_{k,j}$ con

$$Q_k(s, \theta_0) = \sum_{j=0}^k \beta_{k,j}(\theta) s^{[j]}, \quad k = 0, \dots, m.$$

El $(k+1)$ -ésimo elemento de la diagonal de \mathbb{T}_{m+1} es $\langle Q_k, Q_k \rangle_{\theta_0}^2$. Por lo tanto, los coeficientes del polinomio ortonormal de grado k , $\beta_{k,j}(\theta)$, se pueden obtener dividiendo la fila $(k+1)$ -ésima de \mathbb{B}_{m+1} por la raíz cuadrada del k -ésimo elemento de la diagonal de la matriz \mathbb{T}_{m+1} .

2.4 Algoritmo

Los datos que se deben introducir en nuestro algoritmo son:

- a) La función analítica: $g(\theta)$;
- b) El tamaño de muestra: n ;
- c) La función a estimar: $h(\theta)$;
- d) El grado del polinomio de mayor orden que se usará en la aproximación: m ;
- e) El valor muestral de estadístico suficiente: $S_n = s$;

Los pasos a seguir en este algoritmo, serán los siguientes:

1. Escoger un valor arbitrario de θ_0 ;

Aunque el valor de θ_0 es arbitrario, se recomienda escoger $\theta_0 =$ estimador de máxima verosimilitud (emv) de θ .

2. Calcular los $(2m+1)$ primeros momentos factoriales $f_{0,n}(\theta_0), f_{1,n}(\theta_0), \dots, f_{2m,n}(\theta_0)$.

Estos momentos, pueden ser calculados desarrollando la función generatriz de probabilidad de la $PSD(g^n(\theta_0))$, (que en este caso es $\mathcal{G}_n(t; \theta_0) = (g(\theta_0 t)/g(\theta_0))^n$), en serie de potencia alrededor de $t = 1$, puesto que este desarrollo es

$$\mathcal{G}_n(t; \theta_0) = \sum_{j \geq 0} \frac{\mathcal{G}_n^{(j)}(1; \theta_0)}{j!} (t - 1)^j$$

y es bien conocido que $\mathcal{G}_n^{(j)}(1; \theta_0) = f_{j,n}(\theta_0)$. Este desarrollo se puede realizar con un software matemático adecuado, por ejemplo, en MAPLE, usando la instrucción $taylor(\mathcal{G}_n(t; \theta_0), t = 1, 2m + 1)$.

3. Construir la matriz \mathbf{G}_m . Para esto, basta usar la relación de recurrencia dada en (2.14).
4. Calcular la matriz de coeficientes del polinomio ortonormal: \mathbf{A}_m , como se vio en (2.16), y la matriz de los polinomios: \mathbf{P}_m , ver (2.12).
5. Calcular la matriz $\mathbf{E}_m = \mathbf{A}_m \mathbf{H}_m$, donde \mathbf{H}_m es la matriz triangular inferior cuyos coeficientes son

$$h_{i,j} = \begin{cases} 0, & \text{si } i < j, \\ \binom{i}{j} \theta^j f_{i-j}, & \text{si } i \geq j. \end{cases}$$

6. Calcular el vector $\mathbf{C}_m = \mathbf{E}_m \mathbf{D}_m$, donde $\mathbf{D}_m^t = (h(\theta_0), h'(\theta_0), \dots, h^{(m)}(\theta_0))^t$. El k -ésimo componente de \mathbf{C}_m es el k -ésimo coeficiente de Fourier del desarrollo ortogonal del estimador insesgado, ver (2.8).
7. Calcular la aproximación del estimador insesgado $\tilde{T}_m(S_n) = \mathbf{C}_m^t \mathbf{P}_m$, ver (2.10).

8. Si en el paso 1, θ_0 se escogió como el estimador de máxima verosimilitud de θ , entonces una aproximación del emv a la varianza del estimador insesgado es dada por $\hat{V} = \mathbf{C}_m^t \mathbf{C}_m$, véase la expresión de la varianza dada en (2.5).

2.5 Ejemplo

Este ejemplo fue dado por Newell (1965). Se obtuvieron muestras de saliva de pacientes de los cuales se sospechaba que podían padecer una enfermedad maligna. De cada muestra se obtuvieron cuatro pequeñas submuestras y se clasificaron independientemente como positiva (maligna) o negativa (no maligna). Así, cada muestra puede ser clasificada de 0 al 4 dependiendo del número de submuestras malignas. Desde el punto de vista patológico las muestras son clasificados en tres categorías:

a) muestras no malignas, las cuales dan resultado negativo en todas las submuestras.

b) muestras gravemente malignas, que darían resultado positivo en todas las submuestras.

c) muestras parcialmente malignas. Para ésta, se supone que hay una probabilidad, p , de que cualquier muestra seleccionada aleatoriamente sea clasificada como positiva, tendríamos por tanto una distribución binomial de parámetros $n = 4$ y p .

Cuando se tiene una muestra con cuatro ceros se clasifica en muestra no maligna, sin embargo no se sabe si esta podría pertenecer a la binomial ($x = 0$). Similarmente ocurriría en el caso de tener una muestra con cuatro positivos que pertenecería a las muestras gravemente malignas y no se sabría si pertenece a la binomial. Así, se tiene un modelo binomial doblemente truncado.

De 87 muestra con al menos una submuestra positiva siendo muestras no malignas, el patólogo observo:

$n_1 = 45$ muestras con exactamente 1 positiva; $n_2 = 27$ muestras con

exactamente 2 positivas; $n_3 = 15$ muestras con exactamente 3 positivas.

Con estos datos, se estima la probabilidad de cada una de las categorías de la binomial doblemente truncada, i.e.,

$$P[X = x] = \binom{4}{x} p^x q^{4-x} / (1 - p^4 - q^4), \text{ para } x = 1, 2, 3 \text{ y } q = 1 - p.$$

Se muestra en la tabla 1 la estimación para las diferentes frecuencias usando (2.10) con $m = 4, 6,$ y 10 . También se muestran los emv de la varianza de estos estimadores

	\tilde{T}_4	\hat{V}_4	\tilde{T}_6	\hat{V}_6	\tilde{T}_{10}	\hat{V}_{10}
$x = 1$	0.4795	2.294×10^{-3}	0.4795	2.294×10^{-3}	0.4796	2.294×10^{-3}
$x = 2$	0.3857	4.150×10^{-4}	0.3858	4.151×10^{-4}	0.3858	4.161×10^{-4}
$x = 3$	0.1347	7.734×10^{-4}	0.1347	7.734×10^{-4}	0.1347	7.734×10^{-4}

Se observa que la truncación en $m = 4$ da aproximadamente el mismo resultado que en el caso $m = 6$ or 10 .

2.6 Conclusiones

Se ha dado un algoritmo para la obtención de un estimador (aproximadamente) insesgado en distribuciones de series de potencia. Algunas de las ventajas del algoritmo son:

- Es válida para cualquier psd.
- No requiere el cálculo explícito de la distribución del estadístico suficiente.
- No utiliza generalizaciones de los números de Stirling.
- Es posible obtener estimaciones de la varianza de los estimadores.

El algoritmo esta basado en desarrollos ortogonales de los estimadores y pueden ser fácilmente implementados usando programas matemáticos como MATHEMATICA o MAPLE.

Capítulo 3

Aproximación de Poisson a la Probabilidad Binomial

3.1 Introducción

En 1837, el matemático francés Siméon-Denis Poisson derivó la distribución que llevaría su nombre como una aproximación a la ley binomial. Desde entonces, se han desarrollado diferentes aproximaciones y relaciones entre las distribuciones binomial y Poisson. Para una revisión sobre estas aproximaciones puede verse, por ejemplo, Johnson, Kotz y Kemp (1992) y las referencias allí incluidas.

En el resultado principal de esta sección, que se verá en el Teorema 3.1, se obtiene un desarrollo en series de polinomios de Gram-Charlier tipo B para probabilidades binomial el cual es válido para cualesquiera valores posibles de los parámetros N y p de la distribución binomial. Un hecho curioso sobre este desarrollo es que depende de un parámetro arbitrario $\lambda > 0$, así que el desarrollo obtenido refleja como una distribución binomial dada puede ser aproximada por una distribución Poisson arbitraria. Se mostrará que usando unos pocos términos de la serie y escogiendo adecuadamente el parámetro λ se obtienen algunas de las aproximaciones clásicas de la probabilidad bino-

mial a la probabilidad Poisson, así como otras posibilidades. Si se desea una precisión mayor, entonces pueden sumarse más términos para obtener una aproximación mejor. Los términos de la serie se pueden obtener fácilmente usando la fórmula de recurrencia de los polinomios de Charlier. Como una aplicación de nuestro resultado, se verá un método para aproximar probabilidades hipergeométricas a probabilidades binomiales. En la sección 3.3, se verá que cuando N tiende a infinito y $\lambda = Np > 0$ permanece fijo, la convergencia de la aproximación propuesta a la probabilidad binomial es bastante fuerte. Casos particulares de los resultados aquí obtenidos revelan algunos aspectos del comportamiento límite de la aproximación clásica de Poisson a la distribución binomial.

3.2 Aproximación de la Probabilidad Binomial

Denotaremos la función de probabilidad de las distribuciones de Poisson y binomial respectivamente por

$$p(x; \lambda) = e^{-\lambda} \lambda^x / x!, \quad x = 0, 1, \dots, \lambda > 0;$$

y

$$b(x; N, p) = \binom{N}{x} p^x (1-p)^{N-x}, \quad x = 0, 1, \dots, N,$$

con N un entero no negativo y $p \in (0, 1)$. Las probabilidades acumuladas de la Poisson y de la binomial se denotarán respectivamente por

$$P(x; \lambda) = \sum_{j=0}^x p(j; \lambda)$$

y

$$B(x; N, p) = \sum_{j=0}^x b(j; N, p).$$

Consideremos los polinomios de Charlier, ver Chiara (1978), p170,

$$C_k^{(\alpha)}(x) = \sum_{j=0}^k \binom{k}{j} \binom{x}{j} j! (-\alpha)^{k-j}, \quad k = 0, 1, \dots; \quad \alpha > 0. \quad (3.1)$$

Estos polinomios satisfacen las siguientes propiedades:

(P1) $C_k^{(\alpha)}(x) = (-\alpha)^k \nabla^k p(x; \lambda) / p(x; \lambda)$, con ∇ el operador diferencia definido como $\nabla f(x) = f(x) - f(x-1)$.

(P2) $C_{k+1}^{(\alpha)}(x) = (x - \alpha - k)C_k^{(\alpha)}(x) - k\alpha C_{k-1}^{(\alpha)}(x)$, $k \geq 1$; con

$$C_0^{(\alpha)}(x) = 1,$$

$$C_1^{(\alpha)}(x) = x - \alpha.$$

(P3) $\sum_{x=0}^{\infty} C_k^{(\alpha)}(x) C_j^{(\alpha)}(x) p(x; \lambda) = k! \alpha^k \delta_{kj}$, con δ_{kj} la delta de Kronecker.

(P4) Los ceros de $C_k^{(\alpha)}(x)$ son todos reales, simples y localizados en $(0, \infty)$.

Sea X_1, X_2, \dots, X_n una m.a.s. procedente de una Poisson de parametro λ . Consideremos $S_n = \frac{1}{n} \sum_{i=1}^n X_i$ que es un estadístico suficiente y completo, además su función de probabilidad en función de μ es

$$f(s; \mu) = e^{-n\mu} \frac{(n\mu)^{ns}}{(ns)!}, \quad s = 0, \frac{1}{n}, \frac{2}{n}, \dots \quad (3.2)$$

Considérese también la función

$$h(\mu) = p(ry; r\mu) = e^{-r\mu} \frac{(r\mu)^{ry}}{(ry)!}$$

Calcularemos el estimador umvu de $h(\mu)$, mediante la teoría de polinomios ortogonales que hemos desarrollado en el primer capítulo. Pero también podemos calcular el estimador umvu de esta función por otros métodos, por ejemplo el de Karakostas(1985). Como el estimador umvu es único (c.s.) estos dos deben coincidir. Veamos cuales serían estos estimadores por ambos métodos.

El método de Karakostas nos dice que si podemos descomponer la función de densidad de un estadístico suficiente y completo, S_n , de la forma

$$f_n(s; \mu) = \frac{1}{w(s)} f_n^*(s; \mu) h(\mu)$$

donde $f_n^*(s; \mu)$ es una función de densidad y $h(\mu)$ es la función a estimar. Entonces el estimador umvu de $h(\mu)$ es $T_n(S_n) = w(S_n)$.

En nuestro caso, podemos descomponer la función de densidad del estadístico suficiente y completo (4.1) como

$$e^{-n\mu} \frac{(n\mu)^{ns}}{(ns)!} = \frac{(ry)!(ns-ry)!n^{ns}}{(ns)!(n-r)^{ns-ry} r^{ry}} e^{-(n-r)\mu} \frac{((n-r)\mu)^{ns-ry}}{(ns-ry)!} e^{-r\mu} \frac{(r\mu)^{ry}}{(ry)!}.$$

Tomando

$$w(s) = \frac{(ns)!(n-r)^{ns-ry} r^{ry}}{(ry)!(ns-ry)!n^{ns}}$$

$$f_n^*(s; \mu) = e^{-(n-r)\mu} \frac{((n-r)\mu)^{ns-ry}}{(ns-ry)!}$$

y

$$h(\mu) = e^{-r\mu} \frac{(r\mu)^{ry}}{(ry)!},$$

se tiene

$$T_n(s) = \binom{ns}{ry} \left(\frac{r}{n}\right)^{ry} \left(1 - \frac{r}{n}\right)^{ns-ry} \quad (3.3)$$

Considerando $x = ry$, $N = ns$ y $p = r/n$. Reparametrizando tendremos

$$T_n(s) = \binom{N}{x} p^x (1-p)^{N-x} = b(x; N, p) \quad (3.4)$$

El estimador umvu de esta función, mediante los desarrollos ortogonales, según vimos en el capítulo anterior es de la forma

$$T(s) = \sum_{k=0}^{\infty} \frac{D^k h(\mu)}{k!n^k} C_k^{(n\mu)}(ns) \quad (3.5)$$

donde D^k denota la derivada k -ésima con respecto a μ .

Teniendo en cuenta la fórmula de Leibnitz para derivada de productos, obtenemos

$$D^k h(\mu) = \frac{C_k^{(r\mu)}(ry)}{\mu^k} p(ry; r\mu) \quad (3.6)$$

donde

$$p(ry; r\mu) = e^{-r\mu} \frac{(r\mu)^{ry}}{(ry)!}.$$

Substituyendo (3.6) en (3.5) tenemos

$$T(s) = p(ry; r\mu) \sum_{k=0}^{\infty} \frac{C_k^{(r\mu)}(ry)}{k! (\mu n)^k} C_k^{(n\mu)}(ns)$$

Considerando como antes $x = ry$, $N = ns$, $p = r/n$ y $\mu = \lambda$ y reparametrizando tendremos

$$T(s, \mu) = p(x; \lambda) \sum_{k=0}^{\infty} \frac{C_k^{(\lambda)}(x)}{k! (\lambda/p)^k} C_k^{(\lambda/p)}(N) \quad (3.7)$$

Igualando (3.4) y (3.7), se obtiene:

Teorema 3.1 *Para todo entero x y N , tal que $0 \leq x \leq N$, $p \in (0, 1)$ y $\lambda > 0$*

$$b(x; N, p) = p(x; \lambda) \sum_{k=0}^{\infty} \frac{C_k^{(\lambda)}(x)}{k! (\lambda/p)^k} C_k^{(\lambda/p)}(N) \quad (3.8)$$

La elección de λ en el Teorema 3.1 es irrelevante, en el sentido de que para todo $\lambda > 0$, la serie en el miembro derecho de (3.8) converge al mismo valor: $b(x; N, p)$. Nótese que el Teorema 3.1 proporciona una relación exacta entre cualquier distribución de Poisson y cualquier binomial.

La truncación de la serie (3.8) y una elección adecuada de λ nos dará aproximaciones a la distribución binomial. Denotemos por

$$A_m(x; N, p, \lambda) = p(x; \lambda) \sum_{k=0}^m \left(\frac{p}{\lambda}\right)^k \frac{C_k^{(\lambda)}(x)}{k!} C_k^{(\lambda/p)}(N). \quad (3.9)$$

Los términos de la aproximación dados en (3.9) pueden ser fácilmente obtenidos usando la relación de recurrencia de los polinomios de Charlier, dada en (P2)

Vamos a ver algunos ejemplos numéricos. Por ejemplo, para la aproximación de $b(2; 10, 0.5)$ podemos escoger $\lambda = 6$ y $m = 0$. En este caso, la diferencia absoluta entre $A_0(2; 10, 0.5, 6) = 0.04462$ y $b(2; 10, 0.5) = 0.04394$ es $< 10^{-3}$, pero escogiendo $\lambda = 5$, necesitamos $m = 11$ para obtener la misma precisión. La aproximación usual a la Poisson (que no es recomendable en esta situación) da $p(2; 5) = A_0(2; 10, 0.5, 5) = 0.08422$. En una situación general, dados los valores x , N y p , sería ideal si pudiéramos determinar una elección óptima para λ , entendiendo por óptimo que el número de términos en (3.9) nos de una cierta precisión prefijada con un mínimo número de términos. Este problema no es fácil de resolver. Esta afirmación esta basada en el comportamiento aparentemente errático de $A_m(x; N, p, \lambda)$. Para ilustrar este hecho, supongamos que intentamos aproximar $b(x; N, p)$ con una cierta precisión $\varepsilon > 0$. Para un $\lambda > 0$ dado, definimos $m(\lambda, \varepsilon)$ como el menor entero tal que $|b(x, N, p) - A_m(x; N, p, \lambda)| < \varepsilon$. En el caso particular, $x = 20$, $N = 100$, $p = 0.4$ y $\varepsilon = 10^{-6}$, obtenemos los siguientes resultados para valores diferentes de λ :

λ	40.0	48.3	48.5	48.7	48.9	49.1	51
$m(\lambda, \varepsilon)$	13	8	1	1	5	10	2

Como se puede ver, pequeñas variaciones de λ producen grandes saltos en $m(\lambda, \varepsilon)$.

Entre las infinitas posibilidades para λ , podemos escoger $\lambda = \hat{\lambda} = Np$. Esta elección es particularmente interesante pues de esta forma se igualan los momentos de primer orden de la distribución binomial y de la distribución de Poisson que la aproxima. En este caso, (3.8) adopta la forma

$$b(x; N, p) = p(x; \hat{\lambda}) \sum_{k=0}^{\infty} C_k^{(Np)}(x) \gamma_k(N), \quad (3.10)$$

con

$$\gamma_k(N) = (k! N^k)^{-1} C_k^{(N)}(N).$$

De la fórmula de recurrencia dada en (P2), obtenemos

$$\begin{aligned} (k+1)N\gamma_{k+1}(N) &= -k\gamma_k(N) - \gamma_{k-1}(N), \quad k \geq 1 \\ \gamma_0(N) &= 1; \quad \gamma_1(N) = 0. \end{aligned} \quad (3.11)$$

En particular,

$$\begin{aligned} \gamma_2(N) &= -(2N)^{-1}; \\ \gamma_3(N) &= (3N^2)^{-1}; \\ \gamma_4(N) &= (8N^2)^{-1} - (4N^3)^{-1}; \\ \gamma_5(N) &= -(6N^3)^{-1} + (5N^4)^{-1}; \\ \gamma_6(N) &= -(48N^3)^{-1} + 13(72N^4)^{-1} - (6N^5)^{-1}; \end{aligned}$$

y en general, por inducción, puede probarse que

$$\gamma_k(N) = g_k N^{-\lfloor (k+1)/2 \rfloor} + o(N^{-\lfloor (k+1)/2 \rfloor}) \quad (3.12)$$

con

$$g_k = \begin{cases} \frac{(-1)^j}{2^j j!}, & k = 2j, \quad j = 1, 2, \dots \\ \frac{(-1)^{j-1}}{3(j-1)! 2^{j-1}}, & k = 2j+1, \quad j = 1, 2, \dots \end{cases}$$

La truncación de la serie en el lado derecho de (3.10) daría lugar a

$$A_m(x; N, \hat{\lambda}) = p(x; \hat{\lambda}) \sum_{k=0}^m C_k^{(\hat{\lambda})}(x) \gamma_k(N). \quad (3.13)$$

En particular, si $m = 0$ ó 1 , obtenemos la aproximación clásica de la binomial a la Poisson

$$A_1(x; N, \hat{\lambda}) = p(x; \hat{\lambda}),$$

y si $m = 2$, teniendo en cuenta que $C_2^{(\alpha)}(x) = (x - \alpha)^2 - x$, obtenemos

$$A_2(x; N, \hat{\lambda}) = p(x; \hat{\lambda}) \left\{ 1 + \frac{1}{2N} \left\{ x - (x - \hat{\lambda})^2 \right\} \right\}, \quad (3.14)$$

que es la aproximación dada por Burr (1973).

Desarrollos similares pueden obtenerse para la función de probabilidad acumulada de la Binomial. Es más, para aquellos enteros positivos x y N , tales que $0 \leq x \leq N$, $p \in (0, 1)$ y $\lambda > 0$, siguiendo un argumento similar al dado en el teorema 3.1, obtenemos

$$B(x; N, p) = P(x; \lambda) - p \left(\sum_{k=1}^{\infty} \left(\frac{p}{\lambda} \right)^{k-1} \frac{C_{k-1}^{(\lambda)}(x)}{k!} C_k^{(\lambda/p)}(N) \right) p(x; \lambda). \quad (3.15)$$

En particular, escogiendo $\hat{\lambda} = Np$ en (3.15), obtenemos el desarrollo asintótico

$$B(x; N, p) = P(x; \hat{\lambda}) + \sum_{k=2}^{\infty} (-1)^k \hat{\lambda}^k \nabla^{k-1} p(x; \hat{\lambda}) \gamma_k(N). \quad (3.16)$$

De nuevo, la truncación de la serie (3.16) da aproximaciones a la función de probabilidad acumulada de la binomial. Por ejemplo, usando dos términos de la serie (3.16) obtenemos la aproximación de Kolmogorov, ver Johnson, Kotz y Kemp (1992), p.119,

$$B(x; N, p) \doteq P(x; \hat{\lambda}) - Np^2 \nabla^2 P(x; \hat{\lambda})/2.$$

Los resultados presentados en esta sección pueden ser usados para obtener aproximaciones a la probabilidad hipergeométrica

$$h(x; n, N_1, N) = \frac{\binom{N_1}{x} \binom{N_2}{n-x}}{\binom{N}{n}} = \frac{b(x; N_1, n/N) b(n-x; N_2, n/N)}{b(n; N, n/N)}, \quad (3.17)$$

para $\max(0, n - N_2) \leq x \leq \min(n, N_1)$ y $N_2 = N - N_1$.

Sea $p = N_1/N$ y $q = 1 - p$. Tenemos la siguiente aproximación

$$b(x; N_1, n/N) \doteq p(x; np) \left(1 - \frac{C_2^{(np)}(x)}{2Np} + \frac{C_3^{(np)}(x)}{3N^2p^2} \right), \quad (3.18)$$

$$b(n-x; N_2, n/N) \doteq p(n-x; nq) \left(1 - \frac{C_2^{(nq)}(n-x)}{2Nq} + \frac{C_3^{(nq)}(n-x)}{3N^2q^2} \right), \quad (3.19)$$

$$b(n; N, n/N) \doteq p(n; n) \left(1 - \frac{C_2^{(n)}(n)}{2N} + \frac{C_3^{(n)}(n)}{3N^2} \right). \quad (3.20)$$

Multiplicando (3.18) por (3.19), dividiendo (3.20) y despreciando términos de orden $o(N^{-2})$, obtenemos,

$$\begin{aligned} h(x; n, N_1, N) &\doteq b(x; n, p) \left\{ 1 + \frac{1}{2Npq} \{ u^2 + u(q-p) + np(q-p) - np^2 \} + \right. \\ &+ \frac{1}{(Npq)^2} \{ pqu^4/4 + (q-p)u^3/3 + [(7/2-n)pq/2-1]u^2 + 2(q-p)(1/3-npq)u + \\ &\left. 2np(q-p) + (npq)^2 \right\}, \end{aligned} \quad (3.21)$$

con $u = x - np$.

Nótese que la aproximación dada por Ord (1968) esta formada por los primeros dos términos de (3.21). Mejores aproximaciones pueden obtenerse incluyendo mas términos en (3.18), (3.19) y (3.20).

3.3 Propiedades de las aproximaciones

En este apartado, investigaremos el comportamiento límite de las aproximaciones $A_m(x; N, \lambda)$, cuando N tiende a infinito con $\lambda = Np$ permaneciendo constante. Supondremos que $b(x, N, p) = 0$ si x no es un entero entre 0 y N . La llave para este estudio nos la proporciona el siguiente teorema

Teorema 3.2

$$\lim_{N \rightarrow \infty} N^{2[(m+1)/2]} \sum_{x=0}^{\infty} |b(x, N, p) - A_m(x; N, \lambda)|^2 / p(x, \lambda) = S_{m+1}(\lambda), \quad (3.22)$$

donde,

$$S_{m+1}(\lambda) = \begin{cases} (m+1)!g_{m+1}^2\lambda^{m+1} + (m+2)!g_{m+2}^2\lambda^{m+2}, & \text{si } m \text{ es par} \\ (m+1)!g_{m+1}^2\lambda^{m+1}, & \text{si } m \text{ es impar.} \end{cases}$$

Demostración. De (3.9), para enteros $x \geq 0$, $N > 0$,

$$\{b(x, N, p) - A_m(x; N, \lambda)\}^2 / p^2(x, \lambda) = \left\{ \sum_{k=m+1}^{\infty} C_k^{(\lambda)}(x) \gamma_k(N) \right\}^2, \quad (3.23)$$

entonces, usando la propiedad de ortogonalidad de los polinomios de Charlier, ver (P3), y (3.12), tenemos,

$$\begin{aligned} & \sum_{x=0}^{\infty} \{b(x, N, p) - A_m(x; N, \lambda)\}^2 / p(x, \lambda) = \\ & = \sum_{x=0}^{\infty} \left\{ \sum_{k=m+1}^{\infty} C_k^{(\lambda)}(x) \gamma_k(N) \right\}^2 p(x, \lambda) = \sum_{k=m+1}^{\infty} k! \gamma_k^2(N) \lambda^k = \\ & = N^{-2[(m+1)/2]} (m+1)! g_{m+1}^2 \lambda^{m+1} + N^{-2[(m+2)/2]} (m+2)! g_{m+2}^2 \lambda^{m+2} + \\ & + o(N^{-2[(m+2)/2]}), \end{aligned}$$

de donde (3.22) se obtendría. \square

Como muestra el teorema 3.2, la convergencia de la aproximación es bastante fuerte.

Corolario 3.1 *Sea $h(x) \geq 0$ una función tal que*

$$H(\lambda) = \sum_{x \geq 0} h^2(x) p(x, \lambda) < \infty, \quad (3.24)$$

para $\lambda > 0$. Entonces,

$$\overline{\lim}_{N \rightarrow \infty} N^{[(m+1)/2]} \sum_{x=0}^{\infty} |b(x, N, p) - A_m(x; N, \lambda)| h(x) \leq S_{m+1}^{1/2}(\lambda) H^{1/2}(\lambda). \quad (3.25)$$

Demostración. Usando la desigualdad de Cauchy-Schwarz, obtenemos

$$\sum_{x=0}^{\infty} |b(x, N, p) - A_m(x; N, \lambda)| h(x) \leq$$

$$\leq \left(\sum_{x=0}^{\infty} (b(x, N, p) - A_m(x; N, \lambda))^2 / p(x, \lambda) \right)^{1/2} \left(\sum_{x=0}^{\infty} h^2(x) p(x, \lambda) \right)^{1/2}, \quad (3.26)$$

y el resultado se sigue inmediatamente del teorema 3.2. \square

Como consecuencia, para alguna función $h(x)$ en las condiciones expuestas en el corolario 3.1 y $r < [(m + 1)/2]$,

$$\lim_{N \rightarrow \infty} N^r \sum_{x=0}^{\infty} |b(x, N, p) - A_m(x; N, \lambda)| h(x) = 0. \quad (3.27)$$

Para ilustrar la fuerte convergencia de la aproximación, podemos escoger $h(x) = \exp(\alpha x)$, $\alpha > 0$, en (3.27) o cualquier otra función que crezca rápidamente verificando las condiciones del corolario 3.1.

En el caso particular de $m = 1$, i.e., la aproximación clásica de Poisson, (3.27) nos dará

$$\lim_{N \rightarrow \infty} N^r \sum_{x=0}^{\infty} |b(x, N, p) - p(x; \lambda)| h(x) = 0, \text{ para algún } r < 1. \quad (3.28)$$

Un resultado relacionado es el de Simons y Johnson (1971) que prueban que

$$\lim_{N \rightarrow \infty} \sum_{x=0}^{\infty} |b(x, N, p) - p(x; \lambda)| h(x) = 0, \quad (3.29)$$

si y solo si $\sum_{x \geq 0} h(x) p(x, \lambda) < \infty$, para $\lambda > 0$.

También en el caso $m = 1$, de (3.25), y N suficientemente grande,

$$\sum_{x=0}^{\infty} |b(x, N, p) - p(x; \lambda)| \leq \lambda / (\sqrt{2}N). \quad (3.30)$$

Nótese que la cota en el miembro derecho de (3.30) es más fina que la cota $\{2/(1 - \lambda/N)\}^{1/2} \lambda/N$ dada por Vervaat (1970).

Un resultado más débil sobre el comportamiento límite de las aproximaciones $A_m(x; N, \lambda)$ pero no menos interesante es el siguiente:

Teorema 3.3

$$\lim_{N \rightarrow \infty} N^{[(m+1)/2]} \{b(x, N, p) - A_m(x; N, \lambda)\} / p(x; \lambda) = B_{m+1}(x, \lambda), \quad (3.31)$$

con

$$B_{m+1}(x, \lambda) = \begin{cases} g_{m+1}C_{m+1}^{(\lambda)}(x) + g_{m+2}C_{m+2}^{(\lambda)}(x), & m \text{ par} \\ g_{m+1}C_{m+1}^{(\lambda)}(x), & m \text{ impar.} \end{cases} \quad (3.32)$$

Demostración. Inmediata a partir de (3.10) y (3.12). \square

En particular, la aplicación del Teorema anterior al caso $m = 1$, resulta

$$\lim_{N \rightarrow \infty} N \left\{ \frac{b(x, N, p)}{p(x; \lambda)} - 1 \right\} = -C_2^{(\lambda)}(x)/2, \quad (3.33)$$

y como $C_2^{(\lambda)}(x) = (x - \lambda)^2 - x$, es negativo para $x \in (\lambda + 1/2 - (\lambda + 1/4)^{1/2}, \lambda + 1/2 + (\lambda + 1/4)^{1/2})$, se deduce de (3.33), que la aproximación de Poisson, para N suficientemente grande, tiende a infravalorar las probabilidades binomial en el intervalo y a sobrevalorar las probabilidades de las colas, ver Anderson y Samuels (1967). En el caso general $m > 1$, este resultado puede ser fácilmente extendido mediante el estudio del signo del polinomio en el miembro derecho de (3.32). Por ejemplo, si m es impar, el polinomio $B_{m+1}(x; \lambda)$ es un múltiplo del polinomio de Charlier $C_{m+1}^{(\lambda)}(x)$, y por (P4), éste tiene $m + 1$ raíces reales, $r_1 < \dots < r_{m+1}$. Sea $r_0 = 0$ y $r_{m+2} = \infty$. Entonces, para un N grande, la aproximación $A_m(x; N, p)$ tiende a infravalorar la probabilidad binomial para $x \in (r_j, r_{j+1})$, con $B_{m+1}(x; \lambda) > 0$. Mas aún, la magnitud de la diferencia entre la probabilidad binomial y la aproximación $A_m(x; N, p)$ es aproximadamente $B_{m+1}(x, \lambda)p(x; \lambda)/N^{[(m+1)/2]}$ para N grande.

Es también interesante estudiar el comportamiento límite de la máxima diferencia entre la probabilidad binomial y la aproximación A_m . De (3.31), se puede obtener,

$$\lim_{N \rightarrow \infty} \sup_x N^{[(m+1)/2]} |b(x, N, p) - A_m(x, \lambda)| = \sup_x p(x, \lambda) |B_{m+1}(x, \lambda)|. \quad (3.34)$$

Estudiaremos el lado derecho de (3.34) en el caso $m = 1$. Argumentos similares pueden usarse en el caso general $m > 1$.

Sea

$$w(x; \lambda) = p(x, \lambda)C_2^{(\lambda)}(x)/2, \quad x = 0, 1, \dots$$

De (3.34),

$$\begin{aligned} \lim_{N \rightarrow \infty} \sup_x N |b(x, N, p) - p(x, \lambda)| &= \\ = \sup_x |w(x; \lambda)| &= \max\{\sup_x w(x; \lambda), -\inf_x w(x; \lambda)\}, \end{aligned} \quad (3.35)$$

así, tenemos que determinar los valores extremos de $w(x; \lambda)$. Nótese que $w(x; \lambda)$ tiene un máximo (mínimo) en un entero positivo x_0 si

$$\nabla w(x_0; \lambda) \geq 0 (\leq 0) \quad \text{y} \quad \nabla w(x_0 + 1; \lambda) \leq 0 (\geq 0).$$

A partir de (P1),

$$\nabla w(x_0; \lambda) = -p(x_0, \lambda)C_3^{(\lambda)}(x_0)/(2\lambda)$$

y de (P5), la ecuación

$$C_3^{(\lambda)}(r) = 0$$

tiene tres raíces positivas, $r_{1,3}(\lambda) < r_{2,3}(\lambda) < r_{3,3}(\lambda)$. Entonces, la función $\nabla w(x; \lambda)$, (con x un entero positivo) tiene a lo sumo tres cambios de signo. Sea A el conjunto de los enteros positivos de la lista

$$\{0, [r_{i,3}(\lambda)] - 1, [r_{i,3}(\lambda)], [r_{i,3}(\lambda)] + 1, i = 1, 2, 3\}.$$

Los posibles enteros no negativos en los cuales $\nabla w(x; \lambda)$ cambia de signo, pertenecen a A . Así, de (3.35)

$$\lim_{N \rightarrow \infty} \sup_x N |b(x, N, p) - p(x, \lambda)| = \max\{|w(x; \lambda)|, x \in A\}. \quad (3.36)$$

y por lo tanto la evaluación del límite superior del miembro izquierdo de (3.36) se reduce a la evaluación de un conjunto finito.

Ejemplo 3.1 *El estudio de las raices de $C_3^{(\lambda)}(r) = 0$, para $\lambda \in (0, 3 - \sqrt{3})$ muestra que*

$$[r_{1,3}(\lambda)] = 0; \quad (3.37)$$

$$[r_{2,3}(\lambda)] = 1; \quad (3.38)$$

y

$$[r_{3,3}(\lambda)] = \begin{cases} 3, & \text{para } \lambda \in (0, 0.935822) \\ 4, & \text{para } \lambda \in (0.935822, 3 - \sqrt{3}); \end{cases}$$

donde $[\cdot]$ denota la función parte entera. Una evaluación directa del lado derecho de (3.36) da

$$\lim_{N \rightarrow \infty} \sup_x N |b(x, N, p) - p(x, \lambda)| = \begin{cases} \exp(-\lambda)\lambda^2(2 - \lambda)/2, & \text{para } \lambda < 1 \\ \exp(-\lambda)\lambda^2/2, & \text{para } \lambda \in (1, 3 - \sqrt{3}) \end{cases} \quad (3.39)$$

El resultado anterior extiende el de Poor (1991) quién obtuvo (3.39) tan solo para $\lambda \leq 2 - \sqrt{2}$.

Resultados análogos pueden ser obtenidos para probabilidades binomiales acumuladas.

A partir de (3.15) con $\lambda = Np$,

$$\lim_{N \rightarrow \infty} \sup_x N |B(x, N, p) - P(x, \lambda)| = \sup_x \lambda |C_1^{(\lambda)}(x)| p(x; \lambda) / 2. \quad (3.40)$$

De (3.40), siguiendo un argumento similar al del caso no acumulado, se obtiene

$$\lim_{N \rightarrow \infty} \sup_x N |B(x, N, p) - P(x, \lambda)| = \max\{\lambda|x - \lambda|p(x; \lambda)/2, x \in A\}, \quad (3.41)$$

donde A es el conjunto de los enteros no negativos de

$$\{0, [r_{i,2}(\lambda)] - 1, [r_{i,2}(\lambda)], [r_{i,2}(\lambda)] + 1, i = 1, 2\},$$

donde

$$r_{1,2}(\lambda) = \lambda + 0.5 - (\lambda + 0.25)^{1/2}$$

y

$$r_{2,2}(\lambda) = \lambda + 0.5 + (\lambda + 0.25)^{1/2}$$

son las raíces de la ecuación polinomial

$$C_2^{(\lambda)}(r) = (r - \lambda)^2 - r = 0.$$

Ejemplo 3.2 *Un estudio detallado de $r_{1,2}(\lambda)$ y $r_{2,2}(\lambda)$ muestra que*

$$[r_{1,2}(\lambda)] = j, \text{ para } \lambda \in (j + j^{1/2}, (j + 1) + (j + 1)^{1/2}), \text{ para } j = 0, 1, \dots$$

$$[r_{2,2}(\lambda)] = k, \text{ para } \lambda \in (k - k^{1/2}, (k + 1) - (k + 1)^{1/2}), \text{ para } k = 1, 2, \dots$$

Una evaluación directa del lado derecho de (3.41) da

$$\lim_{N \rightarrow \infty} \sup_x N |B(x, N, p) - P(x, \lambda)| = \begin{cases} \lambda^2 \exp(-\lambda)/2, & \text{para } \lambda \in (0, 2) \\ \lambda^2(\lambda - 1) \exp(-\lambda)/2, & \text{para } \lambda \in (2, 2 + 2^{1/2}) \end{cases},$$

y así sucesivamente

Capítulo 4

Aproximación Binomial a la Probabilidad Hipergeométrica

4.1 Introducción

La necesidad de una aproximación para las probabilidades hipergeométricas aparece frecuentemente en el muestreo de poblaciones finitas. Por ejemplo, cuando el muestreo es realizado sin reemplazamiento, la varianza de una proporción puede ser calculada de forma exacta utilizando la distribución hipergeométrica. Otra situación familiar en la que aparece la distribución hipergeométrica es en el análisis de tablas de contingencia 2×2 cuando se utiliza el test exacto de Fisher. Cuando los tamaños muestrales son grandes, las probabilidades hipergeométricas son numéricamente intratables debido a la aparición de factoriales de números grandes. Este hecho justifica las distintas aproximaciones que han sido desarrolladas para las probabilidades hipergeométricas. Algunas de estas aproximaciones y una buena lista de referencias pueden encontrarse en el libro de Johnson, Kotz y Kemp (1992), pp. 256-261. Muchas de estas aproximaciones están basadas en la distribución binomial.

En este capítulo obtendremos una relación exacta entre las probabilidades

hipergeométricas y las binomiales. Esta relación es dada en el Teorema 4.1 y está basada en los polinomios de Krawtchouk.

4.2 Desarrollo de Krawtchouk para probabilidades hipergeométricas

Sean N_1, N_2 , y n números enteros positivos, $N = N_1 + N_2$, y $n \leq N$. Para los enteros x satisfaciendo que

$$\max(0, n - N_2) \leq x \leq \min(N_1, n),$$

consideramos la probabilidad hipergeométrica,

$$h(x; N, N_1, n) = \frac{\binom{N_1}{x} \binom{N_2}{n-x}}{\binom{N}{n}},$$

y la probabilidad binomial

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \text{ para } x \text{ un entero tal que } 0 \leq x \leq n,$$

con $p \in (0, 1)$, $q = 1 - p$. Supondremos que $\binom{a}{b} = 0$, si $b < a$. Las probabilidades acumuladas hipergeométrica y binomial se denotarán respectivamente por

$$H(x; N, N_1, n) = \sum_{y \leq x} h(y; N, N_1, n)$$

y

$$B(x; n, p) = \sum_{y \leq x} b(y; n, p).$$

Consideremos los polinomios de Krawtchouk, ver Chihara (1978), p.161, con expresión explícita

$$k_m(s; N, p) = \sum_{j=0}^m \binom{s}{m-j} \binom{N-s}{j} (-1)^j p^j q^{m-j}, \quad m = 0, \dots, N.$$

Estos polinomios satisfacen las siguientes propiedades:

(P1) Los polinomios de Krawtchouk son ortogonales con respecto a $\psi_{N,p}$, la función de distribución escalonada, con saltos

$$d\psi_{N,p}(s) = b(s; N, p) \text{ en } s = 0, \dots, N.$$

La relación de ortogonalidad es

$$\sum_{s=0}^N k_m(s; N, p) k_r(s; N, p) b(s; N, p) = \delta_{mr} \binom{N}{m} (pq)^m,$$

donde δ_{mr} es la delta de Kronecker.

(P2) $\frac{d^m}{dp^m} \{b(s; N, p)\} = m!(pq)^{-m} k_m(s; N, p) b(s; N, p)$, $m \geq 0$.

(P3) Los polinomios de Krawtchouk satisfacen la relación de recurrencia (omitiendo argumentos):

$$(m+1)k_{m+1} = \{k_1 - m(q-p)\}k_m - Npq \{1 - (m-1)/N\} k_{m-1}, \quad m \geq 1$$

con $k_0 = 1$ y $k_1 = s - Np$.

(P4) $\nabla_s^m b(s; N, p) = (-1)^m (pq)^{-m} \binom{N+m}{N}^{-1} k_m(s; N+m, p) b(s; N+m, p)$, $m \geq 0$, donde ∇_s denota el operador diferencia,

$$\nabla_s f(s) = f(s) - f(s-1).$$

Sea X_1, \dots, X_n una muestra aleatoria simple procedente de una binomial $b(N, p)$. Consideremos $S_n = \frac{1}{n} \sum_{i=1}^n X_i$ que es un estadístico suficiente y completo, además su función de probabilidad es

$$f(s; p) = b(ns; nN, p) = \binom{nN}{ns} p^{ns} q^{n(N-s)}, \quad s = 0, \frac{1}{n}, \frac{2}{n}, \dots \quad (4.1)$$

Consideremos la función

$$h(p) = b(ry; rN, p) = \binom{rN}{ry} p^{ry} q^{r(N-y)},$$

Calculando mediante el método de Karakostas el estimador umvu de la función $h(p)$, tenemos la siguiente descomposición de $f(s; p)$

$$\begin{aligned} & \binom{nN}{ns} p^{ns} q^{n(N-s)} = \\ & = \frac{\binom{nN}{ns}}{\binom{rN}{ry} \binom{nN-rN}{ns-ry}} \binom{nN-rN}{ns-ry} p^{ns-ry} q^{(n-r)N-(ns-ry)} \binom{rN}{ry} p^{ry} q^{r(N-y)} \end{aligned}$$

y por lo tanto el estimador umvu de la función $h(p)$ será

$$T_n(s) = \frac{\binom{rN}{ry} \binom{nN-rN}{ns-ry}}{\binom{nN}{ns}} = h(ry; nN, rN, ns) \quad (4.2)$$

Por la teoría de los desarrollos ortogonales que se vió en el primer capítulo, se tiene que en este caso el estimador umvu de $h(p)$

$$T(s) = \sum_{m=0}^{nN} \frac{D^m h(p) \Gamma(nN - k + 1)}{\Gamma(nN + 1)} k_m(ns; nN, p), \quad (4.3)$$

donde $D^m h(p)$ denota la derivada m -ésima de $h(p)$ respecto a p .

Teniendo en cuenta la fórmula de Leibnitz para la derivada del producto y la propiedad (P2), se tiene

$$D^m h(p) = m! (pq)^{-m} k_m(ry; rN, p) b(ry; rN, p). \quad (4.4)$$

Sustituyendo (4.4) en (4.3) se obtiene

$$T(s) = b(ry; rN, p) \sum_{m=0}^{nN} \binom{nN}{m}^{-1} (pq)^{-m} k_m(ry; rN, p) k_m(ns; nN, p). \quad (4.5)$$

Teorema 4.1 Para enteros x, N, N_1, n , satisfaciendo que $1 \leq n, N_1 \leq N$, $\max(0, n - N + N_1) \leq x \leq \min(N_1, n)$, $y p \in (0, 1)$, $q = 1 - p$, se verifica:

$$h(x; N, N_1, n) = b(x; N_1, p) \sum_{m=0}^{N_1} \binom{N}{m}^{-1} (pq)^{-m} k_m(x; N_1, p) k_m(n; N, p). \quad (4.6)$$

Demostración. Teniendo en cuenta que el estimador umvu de cualquier función es único, igualando (4.2) y (4.5), se obtiene

$$\begin{aligned} h(ry; nN, rN, ns) &= \\ &= b(ry; rN, p) \sum_{m=0}^{nN} \binom{nN}{m}^{-1} (pq)^{-m} k_m(ry; rN, p) k_m(ns; nN, p) \end{aligned} \quad (4.7)$$

Nótese que $h(x; N, s, n)$ es (formalmente) un polinomio de grado n en la variable s , además sin más que reordenar se tiene que

$$h(x; N, s, n) = h(x; N, n, s),$$

por lo tanto $h(ry; nN, rN, ns)$ es un polinomio de grado rN . Nótese por tanto que el límite superior de la suma en (4.7) es rN y no nN , ya que $k_m(ns; rN, p)$ con $m = rN + 1, \dots, nN$ es ortogonal a cualquier polinomio de grado $\leq rN$.

Teniendo esto en cuenta, considerando $x = ry$, $N = nN$, $n = ns$ y $N_1 = rN$ y reparametrizando (4.7), obtenemos

$$h(x; N, N_1, n) = b(x; N_1, p) \sum_{m=0}^{N_1} \binom{N}{m}^{-1} (pq)^{-m} k_m(x; N_1, p) k_m(n; N, p).$$

□

Nótese que (4.6) da una relación exacta entre las probabilidades hipergeométrica y binomial. Es curioso que en este desarrollo dados x, n, N_1 , y N , la elección de $p \in (0, 1)$ en el miembro derecho de (4.6) es arbitraria, por lo tanto esta suma es siempre la misma para todo p .

4.3 Aproximación de la probabilidad hipergeométrica

Nuestro propósito es obtener aproximaciones a $h(x; N, s, n)$, para x entero tal que $\max(0, n + N_1 - N) \leq x \leq \min(N_1, n)$, N_1 , N y n enteros positivos que satisfacen $1 \leq n$, $N_1 < N$ y $n < N$. Esta aproximación será obtenida usando el desarrollo de Krawtchouk para la distribución hipergeométrica dado en (4.6). Como se comentó al final de la sección 4.2, la elección de $p \in (0, 1)$ es arbitraria. De entre las infinitas posibilidades, tomaremos $p = \hat{p} = n/N$, ($\hat{q} = 1 - \hat{p}$). Nótese que al escoger este p estamos igualando los momentos de orden uno de las distribuciones hipergeométrica y binomial consideradas.

Se define:

$$\gamma_m = \binom{N}{m}^{-1} (\hat{p}\hat{q})^{-m} k_m(n; N, \hat{p}), \quad m = 0, \dots, N_1$$

y

$$a_r = a_r(x; N_1, \hat{p}) = b(x; N_1, \hat{p}) \sum_{m=0}^r \gamma_m k_m(x; N_1, \hat{p}), \quad r = 0, \dots, N_1.$$

Proponemos la siguiente aproximación

$$h(x; N, N_1, n) \doteq a_r. \quad (4.8)$$

Los coeficientes γ_m , $m = 0, \dots, n$ pueden ser obtenidos por la fórmula de recurrencia de los polinomios de Krawtchouk dada en (P3). Es más, tenemos

$$\gamma_{m+1} = -\frac{m}{(N-m)\hat{p}\hat{q}} \{(\hat{q} - \hat{p})\gamma_m + \gamma_{m-1}\}, \quad m = 1, \dots, N_1 - 1, \quad (4.9)$$

con $\gamma_0 = 1$; $\gamma_1 = 0$.

En particular,

$$\gamma_2 = -\frac{1}{(N-1)\hat{p}\hat{q}}$$

y

$$\gamma_3 = \frac{2(\hat{q} - \hat{p})}{(N-1)(N-2)(\hat{p}\hat{q})^2}.$$

A partir de (4.9), puede mostrarse por inducción que

$$N^{[(m+1)/2]} \gamma_m = \begin{cases} (-1)^s (\widehat{pq})^{-s} g_{2s} + o_N(1), & m = 2s, s = 1, 2, \dots \\ (-1)^{s+1} (\widehat{pq})^{-(s+1)} g_{2s+1} + o_N(1) & m = 2s + 1, s = 1, 2, \dots \end{cases} \quad (4.10)$$

donde

$$g_{2s} = (2s)! / (2^s s!), \quad s = 1, 2, \dots$$

y

$$g_{2s+1} = \sum_{j=1}^s b_j,$$

con

$$b_s = (2s)! / (2^{s-1} (s-1)!), \quad s = 1, 2, \dots$$

y

$$b_j / b_{j-1} = (2j-1) / (2(j-1)), \quad j = s, s-1, \dots, 2.$$

Para $r = 0$ (ó $r = 1$) en (4.8), obtenemos la clásica aproximación binomial

$$h(x; N, N_1, n) \doteq b(x; n, \widehat{p}),$$

y para $r = 2$,

$$h(x; N, N_1, n) \doteq b(x; N_1, \widehat{p}) \left\{ 1 - \frac{1}{2(N-1)\widehat{p}\widehat{q}} \left\{ (x - N_1\widehat{p})^2 - (\widehat{q} - \widehat{p})(x - N_1\widehat{p}) - N_1\widehat{p}\widehat{q} \right\} \right\} \quad (4.11)$$

Nótese que (4.11) es similar a la aproximación dada por Ord (1968). Dado que r (en a_r) debe ser menor o igual a N_1 , para $N_1 = 2$ la 'aproximación' dada en (4.11) es exacta. El método utilizado por Ord para obtener su resultado es diferente al nuestro. El obtuvo sus resultados extendiendo los trabajos de Hotelling y Frankel (1938) y Wasow (1956). Estudiando el cociente $h(x; N, N_1, n) / b(x; n, \widehat{p})$, Burr (1973) obtuvo una aproximación que se asemeja a (4.11).

Ejemplo 4.1 Damos algunas comparaciones entre las probabilidades hipergeométricas exacta y los valores dados por varias aproximaciones. Por ejemplo, en el caso $N = 100$, $N_1 = 30$, $p = n/N = 0.2$, obtenemos

Frecuencia	Exacta	Binomial	Ord	a_2	a_4	a_6
0	0.0003	0.0012	0.0001	0.0001	0.0003	0.0003
1	0.0035	0.0093	0.0025	0.0025	0.0033	0.0035
2	0.0189	0.0337	0.0193	0.0192	0.0185	0.0186
3	0.0598	0.0785	0.0638	0.0636	0.0603	0.0596
4	0.1268	0.1325	0.1309	0.1308	0.1281	0.1272
5	0.1918	0.1723	0.1895	0.1897	0.1917	0.192

Nótese que no hay mucha diferencia entre las aproximaciones dadas por Ord y a_2 . Obsérvese también que la aproximación de Ord y a_2 pueden dar valores negativos. Esta problema se solventará incluyendo más términos en la aproximación, es decir aumentando el valor de r

Ejemplo 4.2 En este segundo ejemplo, consideramos $N = 40$, $N_1 = 20$ y $p = n/N = 0.1$.

Frecuencia	Exacta	Binomial	Ord	a_2	a_4	a_6
0	0.0530	0.1216	0.0574	0.0558	0.0467	0.0536
1	0.2495	0.2702	0.2702	0.2702	0.2697	0.2461
2	0.3950	0.2852	0.3565	0.3583	0.38	0.4031
3	0.2495	0.1901	0.2324	0.2334	0.2373	0.2401
4	0.0530	0.0898	0.0823	0.0821	0.0688	0.0574

Aproximaciones similares pueden ser obtenidas para la función de distribución de la hipergeométrica.

Teorema 4.2 Para enteros $x, N, N_1, n \geq 2$, satisfaciendo que $1 \leq n, N_1 \leq N$, $\max(0, n - N + N_1) \leq x \leq \min(N_1, n)$, $p \in (0, 1)$, y $q = 1 - p$,

$$\begin{aligned}
H(x; N, N_1, n) &= B(x; N_1, p) - \\
-N_1 p q b(x; N_1 - 1, p) &\sum_{m=1}^n \frac{1}{m} \binom{N}{m}^{-1} (pq)^{-m} k_{m-1}(x; N_1 - 1, p) k_m(n; N, p).
\end{aligned} \tag{4.12}$$

Demostración. Es fácil probar que el estimador insesgado de la función de distribución de X , $F(x, \mu)$, es la función de distribución del estimador umvu de su función de probabilidad.

Teniendo esto en cuenta y (4.2), tenemos que el estimador umvu de la función $F(p) = B(ry; rN, p)$ es

$$T_n(s) = H(ry; nN, rN, ns) \tag{4.13}$$

Como vimos anteriormente, el estimador umvu de una función $F(p)$ mediante los desarrollos ortogonales es

$$T(s, p) = F(p) + \sum_{m=1}^{rN} \frac{D^m F(p) \Gamma(nN - k + 1)}{\Gamma(nN + 1)} k_m(ns; nN, p) \tag{4.14}$$

Nótese que

$$\frac{d}{dp} B(ry; rN, p) = -rN b(ry; rN - 1, p),$$

por lo tanto

$$D^m F(p) = -rN(m-1)! (pq)^{-(m-1)} k_{m-1}(ry; rN - 1, p) b(ry; rN - 1, p) \tag{4.15}$$

Sustituyendo (4.15) en (4.14) e igualando este último con (4.13). tenemos

$$\begin{aligned}
H(ry; nN, rN, ns) &= B(ry; rN, p) - \\
-rN p q b(ry; rN - 1, p) &\sum_{m=1}^{rN} \frac{1}{m} \binom{rN}{m}^{-1} (pq)^{-m} k_{m-1}(ry; rN - 1, p) k_m(ns; nN, p)
\end{aligned}$$

Considerando $x = ry$, $N = nN$, $n = ns$ y $N_1 = rN$, obtenemos

$$\begin{aligned}
 H(x; N, N_1, n) &= B(x; N_1, p) - \\
 -N_1 p q b(x; N_1 - 1, p) &\sum_{m=1}^{N_1} \frac{1}{m} \binom{N}{m}^{-1} (pq)^{-m} k_{m-1}(x; N_1 - 1, p) k_m(n; N, p).
 \end{aligned}
 \tag{4.16}$$

□

Si $p = \hat{p} = n/N$ en (4.12), tenemos la siguiente aproximación a la probabilidad acumulada de la hipergeométrica

$$\begin{aligned}
 A_r &= A_r(x; N_1, \hat{p}) = B(x; N_1, \hat{p}) - \\
 -N_1 \hat{p} \hat{q} b(x; N_1 - 1, \hat{p}) &\sum_{m=1}^r \gamma_m k_{m-1}(x; N_1 - 1, \hat{p}) / m, \quad r \geq 1.
 \end{aligned}$$

En particular, para $r = 1$, tenemos

$$H(x; N, N_1, n) \doteq B(x; N_1, \hat{p})$$

y para $r = 2$,

$$\begin{aligned}
 H(x; N, N_1, n) &\doteq \\
 &\doteq B(x; N_1, \hat{p}) - \frac{N_1}{2(N-1)} b(x; N_1 - 1, \hat{p}) (x - (N_1 - 1)\hat{p}).
 \end{aligned}
 \tag{4.17}$$

La aproximación dada en (4.17) es similar a la dada por Bennett (1965) la cual es citada en Johnson, Kotz y Kemp (1992), p.258.

4.4 Propiedades de la Aproximación

Vamos a estudiar el comportamiento límite de las aproximaciones dadas en las secciones anteriores cuando N tiende a infinito y la razón n/N permanece fija e igual a $p \in (0, 1)$.

Lema 4.1

$$\lim_{N \rightarrow \infty} N^{2[r/2+1]} \sum_{x=0}^{N_1} \frac{(h(x; N, N_1, n) - a_r(x, N_1, p))^2}{b(x, N_1, p)} =$$

$$= \begin{cases} \frac{(q-p)^2}{pq} \binom{N_1}{2s+1} g_{2s+1}^2 + \binom{N_1}{2s+2} g_{2s+2}^2, & r = 2s \\ \binom{N_1}{2s} g_{2s}^2, & r = 2s - 1. \end{cases}$$

Demostración. De (4.6), con $p = n/N$, tenemos

$$h(x; N, N_1, n) - a_r(x, N_1, p) = b(x, N_1, p) \sum_{m=r+1}^N \gamma_m k_m(x; N_1, p)$$

y de la ortogonalidad de los polinomios de Krawtchouk, ver (P1), obtenemos

$$\sum_{x=0}^{N_1} \frac{(h(x; N, N_1, n) - a_r(x, N_1, p))^2}{b(x, N_1, p)} = \sum_{m=r+1}^N \binom{N_1}{m} (pq)^m \gamma_m^2$$

y el resultado deseado se obtiene de (4.10). \square

Teorema 4.3 Para $\alpha < [r/2 + 1]$,

$$\lim_{N \rightarrow \infty} N^\alpha \sum_{x=0}^{N_1} |h(x; N, N_1, n) - a_r(x, N_1, p)| = 0. \quad (4.18)$$

Demostración. Usando la desigualdad de Cauchy-Schwarz, tenemos

$$\left(N^\alpha \sum_{x=0}^n |h(x; N, N_1, n) - a_r(x, N_1, p)| \right)^2 =$$

$$= N^{2\alpha} \left(\sum_{x=0}^n \frac{|h(x; N, N_1, n) - a_r(x, N_1, p)|}{b^{1/2}(x, N_1, p)} b^{1/2}(x, N_1, p) \right)^2 \leq$$

$$\leq N^{2\alpha} \sum_{x=0}^n \frac{|h(x; N, N_1, n) - a_r(x, N_1, p)|^2}{b(x, N_1, p)}, \quad (4.19)$$

por tanto (4.18) se sigue tomando límite en ambos miembros de (4.19) y del lema 4.1. \square

En particular para la aproximación 'clásica' a la hipergeométrica, se tiene que para $\alpha < 1$,

$$\lim_{N \rightarrow \infty} N^\alpha \sum_{x=0}^{N_1} |h(x; N, N_1, n) - b(x, N_1, p)| = 0. \quad (4.20)$$

Además, se verifica el siguiente resultado

$$\begin{aligned} & \lim_{N \rightarrow \infty} N^{[(r+1)/2]} \{h(x; N, N_1, n) - a_r(x, N_1, p)\} = \\ & = \begin{cases} \frac{(-1)^{s+1}}{(pq)^{s+1}} b(x; N_1, p) \{(q-p)g_{r+1}k_{r+1}(x, N_1, p) + \\ + g_{r+2}k_{r+2}(x, N_1, p)\} & r = 2s \\ \frac{(-1)^s}{(pq)^s} b(x; N_1, p) g_{r+1}k_{r+1}(x, N_1, p) & r = 2s - 1 \end{cases} \quad (4.21) \end{aligned}$$

para $s = 1, 2, \dots$

Un resultado similar se puede obtener para la probabilidades acumuladas

$$\begin{aligned} & \lim_{N \rightarrow \infty} N^{[(r+1)/2]} \{H(x; N, N_1, n) - A_r(x, N_1, p)\} = \\ & = \begin{cases} \frac{(-1)^s}{(pq)^s} N_1 b(x; N_1 - 1, p) \{(q-p)g_{r+1}k_{r+1}(x, N_1, p)/(r+1) + \\ + g_{r+2}k_{r+2}(x, N_1, p)/(r+2)\}, & r = 2s \\ \frac{(-1)^{s-1}}{(pq)^{s-1}} N_1 b(x; N_1 - 1, p) g_{r+1}k_{r+1}(x, N_1, p)/(r+1) & r = 2s - 1 \end{cases} \quad (4.22) \end{aligned}$$

En el caso particular de la aproximación binomial, i.e. $r = 1$, de (4.21) y (4.22), obtenemos

$$\lim_{N \rightarrow \infty} N \{h(x; N, N_1, n) - b(x; N_1, p)\} = -b(x; N_1, p)k_2(x; N_1, p)/(pq) \quad (4.23)$$

y

$$\lim_{N \rightarrow \infty} N \{H(x; N, N_1, n) - B(x; N_1, p)\} = -N_1 b(x; N_1 - 1, p) \{x - (N_1 - 1)p\} \quad (4.24)$$

Nótese que las raíces de

$$k_2(x; n, p) = \{(x - np)^2 - (q - p)(x - np) - npq\} / 2$$

son

$$r_1, r_2 = np + (q - p)/2 \pm \{npq + (q - p)^2/4\}^{1/2},$$

entonces estudiando el signo del miembro derecho de (4.23), se concluye que la aproximación binomial tiende a sobreestimar las colas de la probabilidad y a subestimar la probabilidad central. En el caso de la probabilidad acumulada, la aproximación binomial tiende a sobreestimar las probabilidades hipergeométricas acumuladas para valores de x mayores de $(n - 1)p$ y a subestimar en caso contrario.

Capítulo 5

Aproximación entre las Probabilidades Hipergeométrica Negativa y Binomial Negativa

5.1 Introducción

Consideremos un conjunto que contiene N elementos, los cuales están divididos en dos categorías. Sea k el número de elementos con cierta característica, así que $N - k$ elementos no tienen esta característica. Los elementos son escogidos uno a uno de tal forma que en cada extracción cada elemento tiene igual probabilidad de selección, terminando el muestreo cuando se haya obtenido el c -ésimo elemento que tenga la característica. Cuando en una extracción saquemos un elemento con la característica deseada, diremos que la extracción es un éxito, en caso contrario tendremos un fracaso. Ahora, sea

U = el número de fallos obtenidos antes de que ocurra el c -ésimo éxito.

La función de probabilidad de U es

$$f(u; N, k, c) = \begin{cases} \frac{\binom{u+c-1}{c-1} \binom{N-u-c}{k-c}}{\binom{N}{k}}, & u = 0, 1, \dots, N-k, \\ 0, & \text{otro caso.} \end{cases} \quad (5.1)$$

Una variable aleatoria con función de probabilidad como (5.1) se dice que sigue una distribución hipergeométrica negativa. Esta distribución también es llamada '*beta-binomial*', '*Markov-Pólya*' y '*Pólya-Eggenberger*'. Patil y Joshi (1968) llamaron a esta distribución '*hipergeométrica inversa*'. La hipergeométrica negativa pertenece a la familia de distribuciones hipergeométrica generalizada. Esta clase fue estudiada, entre otros, por Davies (1933, 1934), Kemp y Kemp (1956), Sibuya y Shimizu (1981). Más referencias y propiedades de esta clase y de la distribución hipergeométrica negativa pueden ser encontradas en Johnson, Kotz y Kemp (1992).

Si los elementos son seleccionados con reemplazamiento, la variable U sigue una distribución binomial negativa con parametros c y k/N . En general, si r es un entero positivo, $p \in (0, 1)$ y $q = 1-p$, la probabilidad de la binomial negativa se denotara por

$$nb(x; r, p) = \begin{cases} \binom{r+x-1}{r-1} q^x p^r, & x = 0, 1, \dots \\ 0, & \text{caso contrario.} \end{cases} \quad (5.2)$$

Cuando sea conveniente, reparametrizaremos (5.2) en términos de $\lambda = q/p$. Derivando j veces en ambos miembros de (5.2) con respecto a λ , obtenemos

$$D^j nb(x; r, p) = (-1)^j p^j m_j(x; r, q) nb(x; r, p), \quad (5.3)$$

donde $m_j(x, r, q)$ es el j -ésimo polinomio de Meixner de primera clase, con expresión explícita

$$m_j(x; r, q) = (-1)^j j! \sum_{m=0}^j \binom{x}{m} \binom{-x-r}{j-m} q^{-m}, \quad j \geq 0.$$

Estos polinomios satisfacen la relación de ortogonalidad

$$\sum_{x=0}^{\infty} m_j(x; r, q) m_i(x; r, q) n_b(x; r, p) = \delta_{ij} q^{-j} j! (r)_j, \quad i, j \geq 0, \quad (5.4)$$

donde δ_{ij} es la delta de Kronecker y

$$(r)_j = r(r+1) \cdots (r+j-1), \quad \text{para } j \geq 1,$$

y $(r)_0 = 1$. Los polinomios de Meixner de primera clase pueden ser obtenidos por la fórmula de recurrencia

$$qm_{j+1}(x; r, q) = \{-px + (1+q)j + rq\}m_j(x; r, q) - j(j+r-1)m_{j-1}(x; r, q), \quad j \geq 1, \quad (5.5)$$

con $m_{-1}(x; r, q) = 0$, y $m_0(x; r, q) = 1$.

Las propiedades de los polinomios de Meixner citadas, pueden verse en Chihara (1978), Ch. VI.

Sea X una variable aleatoria con función de probabilidad

$$g(x) = nb(x; b, p) = \binom{b+x-1}{b-1} q^x p^b \quad x = 0, 1, \dots \quad (5.6)$$

Consideremos la función

$$h(y) = nb(y; a, p) = \binom{a+y-1}{a-1} q^y p^a \quad y = 0, 1, \dots \quad (5.7)$$

Descomponiendo (5.6) de la siguiente forma

$$\binom{b+x-1}{b-1} q^x p^b = \frac{\binom{b+x-1}{b-1}}{\binom{a+y-1}{a-1} \binom{(b-a)+(x-y)-1}{(b-a)-1}} \cdot \binom{a+y-1}{a-1} q^y p^a \binom{(b-a)+(x-y)-1}{(b-a)-1} q^{x-y} p^{b-a}$$

tenemos que el estimador umvu de la función $h(p)$, según el método de Karakostas, es

$$T(x) = \frac{\binom{a+y-1}{a-1} \binom{(b-a)+(x-y)-1}{(b-a)-1}}{\binom{b+x-1}{b-1}}.$$

Tomando $k = b - 1$, $c = a$, $N = b + x - 1$ y $u = y$, tenemos

$$T(x) = \frac{\binom{u+c-1}{c-1} \binom{N-u-c}{k-c}}{\binom{N}{k}} = f(u; N, c, k) \quad (5.8)$$

Por otra parte, el estimador umvu de una función $h(p)$ mediante desarrollos ortogonales en estas condiciones, viene dado por

$$T(x) = \sum_{j=0}^{\infty} \frac{D^j h(p) q^j}{p^j (b)_j j!} m_j(x; b, q) \quad (5.9)$$

donde D^j denota la derivada j -ésima con respecto a p .

Considerando (5.3), y sustituyendo en (5.9), se obtiene

$$T(x) = nb(y; a, p) \sum_{j=0}^{\infty} \frac{q^j m_j(y, a, q)}{j! (b)_j} m_j(x, b, q), \quad (5.10)$$

ahora, considerando los cambios que realizamos anteriormente, y teniendo en cuenta que el umvu estimador de una función es único, tendremos al igualar (5.8) con (5.10)

$$f(u; N, k, c) = nb(u; c, p) \sum_{j=0}^{\infty} \frac{q^j m_j(u; c, q)}{j! (k+1)_j} m_j(N-k; k+1, q). \quad (5.11)$$

Nótese que (5.11) da una relación exacta entre las probabilidades hipergeométrica negativa y la binomial negativa para valores arbitrarios de u , N , k , c y p , donde u , N , k , c son enteros satisfaciendo $1 \leq k < N$, $c \geq 1$, $u \in \{0, \dots, N-k\}$ y $p \in (0, 1)$.

5.2 Aproximaciones a la probabilidad hipergeométrica negativa

Dado u, N, k, c con las condiciones de arriba, la elección de $p \in (0, 1)$ en el miembro derecho de (5.11) es irrelevante en el sentido de que para cualquier $p \in (0, 1)$ la serie dada en (5.11) converge al mismo valor límite: $f(u; N, k, c)$. Aproximaciones a la probabilidad hipergeométrica negativa pueden ser obtenidas escogiendo un valor particular para p y truncando la serie. Por ejemplo, tomando $p = p_0 = (k + 1)/(N + 1)$, $q_0 = 1 - p_0$ y

$$\gamma_j = \gamma_j(N, k, p_0) = q_0^j m_j(N - k; k + 1, q_0) / (j!(k + 1)_j),$$

proponemos las siguientes aproximaciones a la probabilidad hipergeométrica negativa:

$$a_m(u; N, k, c) = nb(u; c, p_0) \sum_{j=0}^m \gamma_j m_j(u; c, q_0). \quad (5.12)$$

De (5.5), es fácil comprobar que los γ_j satisfacen la fórmula de recurrencia

$$\gamma_{j+1} = \frac{j(2N - k + 1)\gamma_j - (N - k)\gamma_{j-1}}{(j + 1)(j + k + 1)(N + 1)}, \quad j \geq 1 \quad (5.13)$$

con $\gamma_0 = 1$ y $\gamma_1 = 0$.

En particular,

$$\gamma_2 = -\frac{(N - k)}{2k(N + 1)}$$

y

$$\gamma_3 = -\frac{(N - k)(2N - k + 1)}{3k(k + 1)(N + 1)^2}.$$

Para $m = 0$ (ó $m = 1$), (5.12) da la aproximación clásica de la binomial negativa a la distribución hipergeométrica negativa. Para $m = 2$, obtenemos

$$a_2(u; N, k, c) = nb(u; c, p_0) \left\{ 1 - \frac{(N - k)}{2k(N + 1)} m_2(u; c, q_0) \right\},$$

con

$$m_2(u; c, q_0) = \frac{p_0^2}{q_0^2} w^2 - \frac{p_0(1 + q_0)}{q_0^2} w - \frac{c}{q_0}, \quad w = u - \frac{cq_0}{p_0}.$$

5.3 Propiedades de las aproximaciones

En esta sección estudiaremos el comportamiento límite de las aproximaciones propuestas en (5.12). Suponemos a partir de ahora que si N tiende a infinito, el cociente $(k + 1)/(N + 1)$ permanece fijo e igual a $p_0 \in (0, 1)$. Bajo estas condiciones la probabilidad hipergeométrica negativa $f(u; N, k, c)$ converge a la respectiva binomial negativa $nb(u; c, p_0)$, ver Johnson, Kotz y Kemp (1992), p. 255.

Necesitamos algunos lemas para establecer las propiedades límites de nuestras aproximaciones. El primero concierne al comportamiento límite de los coeficientes γ_j .

Definimos

$$g_j = g_j(p_0) = \lim_N N^{[(j+1)/2]} \gamma_j, \quad j \geq 1$$

donde $[\cdot]$ denota la parte entera.

Lema 5.1 *Los g_j satisfacen la siguiente relación de recurrencia*

$$g_{2s+1} = \frac{2s(2 - p_0)}{(2s + 1)p_0} g_{2s} - \frac{1 - p_0}{(2s + 1)p_0} g_{2s-1}, \quad s \geq 1,$$

$$g_{2(s+1)} = -\frac{1 - p_0}{2(s + 1)p_0} g_{2s}, \quad s \geq 1,$$

con $g_1 = 0$ y $g_2 = -(1 - p_0)/(2p_0)$.

Demostración. La demostración es inmediata de la relación de recurrencia dada en (5.13). \square

Lema 5.2

$$\begin{aligned} & \lim_N N^{2[(m+2)/2]} \sum_{u=0}^{\infty} \frac{(f(u; N, k, c) - a_m(u; N, k, c))^2}{nb(u; c, p_0)} \\ &= \begin{cases} g_{2s+1}^2 q_0^{-(2s+1)} (2s + 1)! (c)_{2s+1} + g_{2s+2}^2 q_0^{-(2s+2)} (2s + 2)! (c)_{2s+2}, & m = 2s \\ g_{2s+2}^2 q_0^{-(2s+2)} (2s + 2)! (c)_{2s+2}, & m = 2s + 1 \end{cases} \end{aligned}$$

Demostración. Consideraremos el caso $m = 2s + 1$, $s \geq 0$. El caso $m = 2s$, $s \geq 1$ puede ser obtenido de forma similar. De (5.11) con $p_0 = (k+1)/(N+1)$, tenemos

$$\left(\frac{f(u; N, k, c) - a_m(u; N, k, c)}{nb(u; c, p_0)} \right)^2 = \left(\sum_{j=m+1}^{\infty} \gamma_j m_j(u; c, q_0) \right)^2. \quad (5.14)$$

De (5.14) y usando la propiedad de ortogonalidad de los polinomios de Meixner dada en (5.4),

$$\sum_{u=0}^{\infty} \frac{(f(u; N, k, c) - a_m(u; N, k, c))^2}{nb(u; c, p_0)} = \sum_{j=m+1}^{\infty} \gamma_j^2 q_0^{-j} j!(c)_j. \quad (5.15)$$

Para $m = 2s + 1$, de (5.15) y lema 5.1, tenemos

$$\begin{aligned} & \lim_N N^{2(s+1)} \sum_{u=0}^{\infty} \frac{(f(u; N, k, c) - a_m(u; N, k, c))^2}{nb(u; c, p_0)} \\ &= \lim_N N^{2(s+1)} \gamma_{2s+2}^2 q_0^{-(2s+2)} (2s+2)!(c)_{(2s+2)} + \sum_{j=2s+3}^{\infty} \lim_N N^{2(s+1)} \gamma_j^2 q_0^{-j} j!(c)_j \\ &= g_{2s+2}^2 q_0^{-(2s+2)} (2s+2)!(c)_{2s+2}. \end{aligned}$$

□

Ahora estamos en condiciones para establecer algunas propiedades límites de las aproximaciones propuestas.

Teorema 5.1 *Cuando N tiende a infinito con $p_0 = (k+1)/(N+1)$ fijo y para todo $\alpha < [(m+2)/2]$*

$$\lim_N N^\alpha \sum_{u=0}^{\infty} |f(u; N, k, c) - a_m(u; N, k, c)| = 0. \quad (5.16)$$

Demostración. Usando la desigualdad de Cauchy-Schwarz, y teniendo en cuenta que

$$\sum_{u=0}^{\infty} nb(u; c, p_0) = 1,$$

tenemos

$$\begin{aligned}
& N^{2\alpha} \left(\sum_{u=0}^{\infty} |f(u; N, k, c) - a_m(u; N, k, c)| \right)^2 \\
&= N^{2\alpha} \left(\sum_{u=0}^{\infty} \frac{|f(u; N, k, c) - a_m(u; N, k, c)|}{nb^{1/2}(u; c, p_0)} nb^{1/2}(u; c, p_0) \right)^2 \\
&\leq N^{2(\alpha - [(m+2)/2])} N^{2[(m+2)/2]} \sum_{u=0}^{\infty} \frac{|f(u; N, k, c) - a_m(u; N, k, c)|^2}{nb(u; c, p_0)},
\end{aligned}$$

por tanto (5.16) se tiene fácilmente. \square

En particular, para $m = 1$, obtenemos la siguiente propiedad de la aproximación binomial negativa a la probabilidad hipergeométrica negativa

$$\lim_N N^\alpha \sum_{u=0}^{\infty} |f(u; N, k, c) - nb(u; c, p_0)| = 0, \quad \text{para todo } \alpha < 1. \quad (5.17)$$

Nótese que (5.17) implica,

$$\lim_N \sup_u N^\alpha |f(u; N, k, c) - nb(u; c, p_0)| = 0, \quad \text{para todo } \alpha < 1.$$

Algunas otras propiedades límites pueden obtenerse, por ejemplo,

$$\begin{aligned}
& \lim_N N^{[(m+1)/2]} \frac{f(u; N, k, c) - a_m(u; N, k, c)}{nb(u; c, p_0)} \\
&= \begin{cases} g_{2s+1}m_{2s+1}(u; c, q_0) + g_{2s+2}m_{2s+2}(u; c, q_0), & m = 2s, s \geq 1 \\ g_{2s+2}m_{2s+2}(u; c, q_0) & m = 2s + 1, s \geq 0 \end{cases} \quad (5.18)
\end{aligned}$$

De (5.18) con $m = 1$, obtenemos

$$\lim_N N \left(\frac{f(u; N, k, c)}{nb(u; c, p_0)} - 1 \right) = -\frac{q_0}{2p_0} m_2(u; c, q_0). \quad (5.19)$$

El estudio del signo en el lado derecho de (5.19) muestra que cuando N tiende a infinito con $p_0 = (k + 1)/(N + 1)$ fijo, la aproximación binomial

negativa tiende a sobreestimar la probabilidad hipergeométrica negativa para los valores de u tales que $m_2(u; c, q_0) > 0$, o sea, la aproximación binomial negativa sobrestima la cola de las probabilidades hipergeométrica negativa y subestima las probabilidades centrales.

Otro aspecto interesante es el estudio del máximo de la diferencia absoluta entre la probabilidad hipergeométrica negativa y la aproximación binomial negativa. De los resultados obtenidos, se puede probar que

$$\lim_N \sup_u N |f(u; N, k, c) - nb(u; c, p_0)| = \frac{q_0}{2p_0} \sup_u |m_2(u; c, q_0)| nb(u; c, p_0).$$

5.4 Probabilidades hipergeométricas negativas acumuladas

Denotemos por

$$F(u; N, k, c) = \sum_{v=0}^u f(v; N, k, c)$$

y

$$NB(u; c, p) = \sum_{v=0}^u nb(v; c, p)$$

a las probabilidades acumuladas de la hipergeométrica negativa y la binomial negativa respectivamente.

Dado que el estimador umvu de la función $nb(y; a, p)$ es

$$f(y; b + x - 1, b - 1, p),$$

tenemos

$$\sum_{x=0}^{\infty} f(y; b + x - 1, b - 1, a) nb(x; b, p) = nb(y; a, p), \quad (5.20)$$

De (5.20), obtenemos

$$\sum_{x=0}^{\infty} F(y; b + x - 1, b - 1, a) nb(x; b, p) = NB(y; a, p), \quad (5.21)$$

para a, b e y enteros no negativos tales que $b > a$ y $p \in (0, 1)$. Después de algunas operaciones elementales, puede verse que

$$\frac{d}{d\lambda} NB(y; a, p) = -anb(y; a + 1, p). \quad (5.22)$$

Derivando j veces respecto a λ en el lado izquierdo de (5.21), y usando (5.3), tenemos

$$(-1)^j p^j \sum_{x=0}^{\infty} F(y; b + x - 1, b - 1, a) m_j(x; b, q) nb(x; b, p), \quad (5.23)$$

y derivando $j (\geq 1)$ veces respecto a λ en el lado derecho de (5.21), usando (5.3) y (5.22),

$$\begin{aligned} D^j NB(y; a, p) &= -aD^{j-1}nb(y; a + 1, p) = \\ &a(-1)^{j-1} p^{j-1} m_{j-1}(y; a + 1, q) nb(y; a + 1, p). \end{aligned} \quad (5.24)$$

Igualando (5.23) y (5.24), se obtiene

$$\begin{aligned} &\sum_{x=0}^{\infty} F(y; b + x - 1, b - 1, a) m_j(x; b, q) nb(x; b, p) \\ &= (a + y) nb(y; a, p) m_{j-1}(y; a + 1, q). \end{aligned} \quad (5.25)$$

Por otra parte, sabemos que cualquier función h definida en los enteros no negativos y satisfaciendo la condición $\sum_{x=0}^{\infty} h^2(x) nb(x; b, p) < \infty$, admite un desarrollo ortogonal en términos de los polinomios de Meixner de la forma

$$h(x) = \sum_{j=0}^{\infty} a_j(p) m_j(x; b, q) \quad (5.26)$$

donde

$$a_j(p) = q^j \left(j! (r)_j \right)^{-1} \sum_{x=0}^{\infty} h(x) m_j(x; b, q) nb(x; b, p), \quad j \geq 0 \quad (5.27)$$

De (5.25), (6.5) y (6.9),

$$F(y; b+x-1, b-1, a) = NB(y; a, p) + (a+y)nb(y; a, p) \sum_{j=1}^{\infty} \frac{q^j m_{j-1}(y; a+1, q)}{j!(b)_j} m_j(x; b, q), \quad (5.28)$$

ahora, si consideramos $y = u$, $N = b+x-1$, $k = b-1$ y $a = c$ en (5.28), obtenemos

$$F(u; N, k, c) = NB(u; c, p) + (c+u)nb(u; c, p) \sum_{j=1}^{\infty} \frac{q^j m_{j-1}(u; c+1, q)}{j!(k+1)_j} m_j(N-k; k+1, q), \quad (5.29)$$

lo cual nos da una relación exacta entre la probabilidad hipergeométrica negativa acumulada y la probabilidad binomial negativa acumulada. Además, podemos obtener aproximaciones a $F(u; N, k, c)$ escogiendo un valor arbitrario de $p \in (0, 1)$ y truncando la serie en el lado derecho de (5.29). Por ejemplo, escogiendo $p = p_0 = (k+1)/(N+1)$, nos da la siguiente aproximación

$$A_0(u; N, k, c) = NB(u; c, p_0),$$

$$A_m(u; N, k, c) = NB(u; c, p_0) + (c+u)nb(u; c, p_0) \sum_{j=1}^m \gamma_j m_{j-1}(u; c+1, q_0), \quad m \geq 1,$$

con γ_j definida en la sección (5.2).

Nótese que, como $\gamma_1 = 0$, $A_0(u; N, k, c) = A_1(u; N, k, c)$ y ésta es una aproximación conocida de la binomial negativa a la probabilidad hipergeométrica negativa acumulada. Para $m = 2$, obtenemos

$$A_2(u; N, k, c) = NB(u; c, p_0) + \frac{c}{2k} nb(u; c, p_0) (u+c) \left(u - \frac{(c+1)q_0}{p_0} \right).$$

Propiedades asintóticas para estas aproximaciones se pueden obtener siguiendo argumentos similares a los usados en la sección (5.3). Por ejemplo,

tenemos

$$\lim_N N (F(u; N, k, c) - NB(u; c, p_0)) = \frac{c}{2p_0} nb(u; c, p_0)(u + c)\left(u - \frac{(c + 1)q_0}{p_0}\right),$$

de lo cual se deduce que cuando N tiende a infinito y el cociente $(k+1)/(N+1)$ permanece fijo e igual a $p_0 \in (0, 1)$, la aproximación binomial negativa tiende a sobreestimar (subestimar) a la probabilidad hipergeométrica negativa acumulada para los valores $u > (c + 1)q_0/p_0$ ($u < (c + 1)q_0/p_0$).

5.5 Conclusiones

En (5.11) damos un desarrollo ortogonal en término de los polinomios de Meixner de primera clase relacionando la probabilidad hipergeométrica negativa y probabilidad binomial negativa. Para un conjunto de valores factibles de u , N , k , c tenemos un desarrollo para la probabilidad $f(u; N, k, c)$. Un hecho curioso de este desarrollo es que depende de un parámetro arbitrario $p \in (0, 1)$. En otras palabras, se pueden obtener un número infinito de desarrollos para $f(u; N, k, c)$: uno para cada elección particular de $p \in (0, 1)$. Elecciones particulares de p y la truncación del lado derecho de la serie (5.11) nos dan aproximaciones a la probabilidad hipergeométrica negativa en términos de las probabilidades binomiales negativas. Se han estudiado algunas propiedades límites de estas aproximaciones cuando N tiende a infinito si el cociente $(k + 1)/(N + 1)$ permanece fijo e igual a cierto valor $p_0 \in (0, 1)$. En la sección 5.4 se ha extendido el estudio a las probabilidades acumuladas.

Capítulo 6

Una propiedad martingala inversa que caracteriza a la familia exponencial natural con función de varianza cuadrática

6.1 Introducción

Considérese la familia exponencial natural (NEF), con densidad

$$f(x; \theta) = \exp\{x\theta - \psi(\theta)\}$$

con respecto a cierta medida σ -finita ν . Asumiremos que el espacio paramétrico Θ es un conjunto abierto no vacío, para el cual $\int \exp(x\theta) d\nu(x)$ es finito.

Sea X una variable aleatoria cuya distribución pertenece a una NEF. La media y la varianza de X son (como se vió en el capítulo 1):

$$\mu = E_{\theta}[X] = \psi'(\theta)$$

y

$$Var_{\theta}(X) = \psi''(\theta) > 0.$$

La función de medias $\mu = \psi'(\theta)$ es 1-1 y para este estudio sería conveniente reparametrizar la familia en términos del parámetro μ . El dominio de μ se llamará espacio de medias y se notará por $\Omega = \psi'(\Theta)$. Así, denotaremos a la función de densidad de la NEF por $f(x; \mu)$ y a la función de varianza por $V(\mu) = \psi''(\theta(\mu))$.

Sea X_1, \dots, X_n una m.a.s. de una densidad $f(x; \mu)$. Se sabe que el estadístico $S_n = \sum_{i=1}^n X_i$ es suficiente y completo y que su densidad con respecto a la medida convolución $\nu_n = \nu * \dots * \nu$ (n veces) es

$$f_n(s; \mu) = \exp \{s\theta(\mu) - n\psi(\theta(\mu))\}, \mu \in \Omega.$$

Para cada $\mu \in \Omega$, consideremos el espacio de funciones

$$\mathcal{L}_{n,\mu}^2 = \left\{ T : \int T^2(s) f_n(s; \mu) \nu_n(ds) < \infty \right\}.$$

Como se vió, $\mathcal{L}_{n,\mu}^2$ es un espacio de Hilbert para cualquier $\mu \in \Omega$, considerando el producto escalar

$$\langle T_1, T_2 \rangle_{n,\mu} = E_\mu [T_1(S_n) T_2(S_n)],$$

además dos funciones T_1 y T_2 son equivalentes si $\nu_n \{T_1 \neq T_2\} = 0$.

Un sistema completo en $\mathcal{L}_{n,\mu}^2$ es

$$\begin{aligned} P_{k,n}(s; \mu) &= V^k(\mu) D_\mu^j f_n(s; \mu) / f_n(s; \mu), \quad k \geq 1, \\ P_{0,n}(s; \mu) &= 1, \end{aligned} \tag{6.1}$$

ver Abbey y David (1970).

Algunas propiedades de estas funciones son:

(P1) $P_{k,n}$ es un polinomio de grado k en s .

(P2) Los polinomios $\{P_{k,n}(s; \mu)\}_{k \geq 0}$ satisfacen la relación de recurrencia (omitiendo argumentos):

$$\begin{aligned} P_{k+1,n} &= (s - n\mu - kV')P_{k,n} + VP'_{k,n}, \quad k \geq 0 \\ P_{0,n} &= 1 \end{aligned}$$

(La prima (') denota derivar con respecto a μ). En particular, tenemos

$$P_{1,n}(s; \mu) = s - n\mu$$

y

$$P_{2,n} = (s - n\mu)^2 - V'(\mu)(s - n\mu) - nV(\mu)$$

(P3) Los polinomios $P_{0,n}$, $P_{1,n}$ y $P_{2,n}$ son mutuamente ortogonales. Es más, no es difícil probar que

(i) $E_\mu [P_{k,n}(S_n; \mu)] = \delta_{k0}$, para todo $k \geq 0$, donde δ_{k0} es la delta de Kronecker.

(ii) $E_\mu [P_{1,n}(S_n; \mu)P_{2,n}(S_n; \mu)] = 0$.

(iii) $E_\mu [P_{1,n}^2(S_n; \mu)] = nV(\mu)$.

(iv) $E_\mu [P_{2,n}^2(S_n; \mu)] = nV^2(\mu) \{V''(\mu) + 2n\}$.

(P4) En general $\{P_{k,n}(s; \mu)\}_{k \geq 0}$ no es un sistema de polinomios ortogonales. La condición necesaria y suficiente para que este sistema sea ortogonal, es que la función de varianza de la NEF sea cuadrática, es decir $V(\mu) = v_0 + v_1\mu + v_2\mu^2$, ver Morris (1982, 1983).

(P5) La siguiente *propiedad martingala en tiempo discreto* se satisface por los polinomios definidos en (6.1).

Teorema 6.1 Para cualquier entero $m \geq n \geq 1$,

$$E [P_{k,m}(S_m; \mu) | S_n] = P_{k,n}(S_n; \mu), \quad k \geq 0. \quad (6.2)$$

Demostración. Nótese que para $m > n$,

$$f_m(s; \mu) = f_n(t; \mu) f_{m-n}(s-t; \mu).$$

Usando la fórmula de Leibnitz para la derivada del producto, se obtiene

$$D_\mu^k f_m(s; \mu) = \sum_{j=0}^k \binom{k}{j} D_\mu^j f_n(t; \mu) D_\mu^{k-j} f_{m-n}(s-t; \mu),$$

y de (6.1), se tiene

$$P_{k,m}(s; \mu) = \sum_{j=0}^k \binom{k}{j} P_{j,n}(t; \mu) P_{k-j,m-n}(s-t; \mu). \quad (6.3)$$

Considérese

$$S_m = S_n + T_{m-n},$$

con $T_{m-n} = \sum_{i=n+1}^m X_i$. Nótese que S_n y T_{m-n} son independientes. Considerando en (6.3) $s = S_m$ y $t = T_{m-n}$, tomando esperanza condicionada con respecto a S_n . Usando propiedades estandar de la esperanza condicionada y la propiedad P3(i), se obtiene (6.2). \square

La propiedad martingala dada en (6.2) es satisfecha en general por los polinomios de Sheffer asociados a ciertas distribuciones, ver Schoutens y Teugels (1998).

En este capítulo se caracterizarán las funciones exponenciales para las cuales se satisface la siguiente propiedad

$$E [P_{k,n}(S_n; \mu) | S_m] = d_{k,n,m} P_{k,m}(S_m; \mu), \quad k \geq 0, \quad (6.4)$$

para cualquier $m \geq n \geq 1$, y $d_{k,n,m}$ constantes reales.

El resultado principal se da en el siguiente Teorema.

Teorema 6.2 Dada una familia exponencial natural, los polinomios definidos en (6.1) satisfacen la propiedad (6.4) si y solo si la familia tiene función de varianza cuadrática.

6.2 Demostración

La demostración del Teorema 6.2 se basa en la propiedades de los estimadores insesgados de mínima varianza (UMVUE) en las familias exponenciales. En la siguiente subsección recordaremos algunas de estas propiedades.

6.2.1 Estimadores insesgados en las familias exponenciales

Sea X_1, \dots, X_n una muestra i.i.d. procedente de una NEF con densidad $f(x; \mu)$. Siguiendo la notación del primer capítulo, se denota por \mathcal{U}_n al conjunto de funciones paramétricas, $h(\mu)$, umvu estimable con función de varianza finita.

Si $h \in \mathcal{U}_n$ para cierto n , entonces $h \in \mathcal{U}_m$ para todo $m \geq n$. Ésto es consecuencia del siguiente hecho, si $T_n(S_n)$ es el umvu estimador de $h(\mu)$ para un tamaño de muestra n , como S_m es suficiente y completo para el tamaño de muestra $m \geq n$, por el teorema de Rao-Blackwell, se tiene que $T_m(S_m) = E[T_n(S_n) | S_m]$ es el estimador umvu de $h(\mu)$ para el tamaño de muestra m . En otras palabras, se muestra que la sucesión de conjuntos $\{\mathcal{U}_n\}_n$ es creciente en el sentido de la inclusión.

Como los polinomios dados en (6.1) forman un sistema completo en $\mathcal{L}_{n,\mu}^2$ para todo $\mu \in \Omega$, cualquier función $T_n \in \bigcap_{\mu \in \Omega} \mathcal{L}_{n,\mu}^2$ admite un desarrollo de la forma

$$T_n(s) = \sum_{j \geq 0} a_{j,n}(\mu) P_{j,n}(s; \mu), \quad \mu \in \Omega. \quad (6.5)$$

En general, el sumatorio de (6.5) contiene infinitos términos. Se define la subclase $\mathcal{U}_{n,k}$ que contiene las funciones, $h(\mu)$, pertenecientes a \mathcal{U}_n , tales que su correspondiente estimador umvu, T_n , admite un desarrollo finito de la forma

$$T_n(s) = \sum_{j=0}^k a_{j,n}(\mu) P_{j,n}(s; \mu), \quad \mu \in \Omega.$$

El siguiente lema caracteriza la subclase $\mathcal{U}_{n,2}$.

Lema 6.1 Una función $h(\mu)$ pertenece a $\mathcal{U}_{n,2}$ si y solo si existen unas constantes reales c_0 , c_1 y c_2 tal que

$$h(\mu) = nc_2\{V(\mu) + n\mu^2\} + c_1\mu + c_0. \quad (6.6)$$

El estimador insesgado de (6.6) para el tamaño de muestra n es

$$\begin{aligned} T_n(S_n) &= c_2S_n^2 + c_1S_n/n + c_0 = \\ &= a_{0,n}(\mu) + a_{1,n}(\mu)P_{1,n}(S_n; \mu) + a_{2,n}(\mu)P_{2,n}(S_n; \mu), \end{aligned}$$

con

$$\begin{aligned} a_{0,n}(\mu) &= h(\mu), \\ a_{1,n}(\mu) &= h'(\mu)/n, \\ a_{2,n}(\mu) &= h''(\mu) / \{nV''(\mu) + 2n^2\}. \end{aligned} \quad (6.7)$$

Demostración. De acuerdo a la definición, una función $h(\mu)$ pertenece a $\mathcal{U}_{n,2}$ si el estimador insesgado es de la forma

$$T_n(S_n) = a_{0,n}(\mu) + a_{1,n}(\mu)P_{1,n}(S_n; \mu) + a_{2,n}(\mu)P_{2,n}(S_n; \mu), \quad \text{para todo } \mu \in \Omega. \quad (6.8)$$

La propiedad (P3) de la sección 6.1 establece que los polinomios $P_{0,n}$, $P_{1,n}$ y $P_{2,n}$ son mutuamente ortogonales, entonces los coeficientes en (6.8) pueden ser obtenidos como

$$a_{j,n}(\mu) = \langle T_n, P_{j,n} \rangle_{n,\mu} / \|P_{j,n}\|_{n,\mu}^2, \quad j = 0, 1, 2. \quad (6.9)$$

Se tiene,

$$\begin{aligned} \langle T_n, P_{j,n} \rangle_{n,\mu} &= V^j(\mu) \int T_n(s) D_\mu^j f_n(s; \mu) \nu_n(ds) \\ &= V^j(\mu) D_\mu^j \int T_n(s) f_n(s; \mu) \nu_n(ds) = V^j(\mu) D_\mu^j h(\mu). \end{aligned} \quad (6.10)$$

Usando (6.9), (6.10) y los resultados dados en P3(iii), (iv) de la sección 6.1, se obtiene

$$a_{0,n}(\mu) = h(\mu), \quad a_{1,n}(\mu) = h'(\mu)/n, \quad \text{y} \quad a_{2,n}(\mu) = h''(\mu) / \{nV''(\mu) + 2n^2\}. \quad (6.11)$$

Nótese que (6.8) es un polinomio de grado dos en S_n con coeficiente líder $a_{2,n}(\mu)$. Como $T_n(S_n)$ es un estimador, éste no debe depender de μ . Consecuentemente, $a_{2,n}(\mu)$ debe ser una constante, que se notara por c_2 . Entonces $h''(\mu) = c_2 n \{V''(\mu) + 2n\}$, por lo cual

$$h(\mu) = nc_2 \{V(\mu) + n\mu^2\} + c_1\mu + c_0, \quad (6.12)$$

para ciertas constantes c_0 y c_1 . Substituyendo (6.11) en (6.8), teniendo en cuenta (6.12) y la expresión explícita para $P_{1,n}$ y $P_{2,n}$ dado en (P2), se obtiene

$$T_n(S_n) = c_2 S_n^2 + c_1 S_n/n + c_0.$$

□

Supongamos que la densidad subyacente $f(x; \mu)$ pertenece a la NEF-QVF, con función de varianza

$$V(\mu) = v_0 + v_1\mu + v_2\mu^2.$$

Muchas de las propiedades citadas a continuación pueden encontrarse u obtenerse fácilmente de los resultados dados en Morris (1982, 1983). La familia NEF-QVF esta formada, esencialmente, por seis subfamilias: normal (con varianza conocida), binomial, Poisson, gamma, binomial negativa, y secante hiperbólica generalizada. Para la NEF-QVF, los polinomios $\{P_{k,n}(s; \mu)\}_{k \geq 0}$ definidos en (6.1) forman un sistema de polinomios mónicos ortogonales con respecto a la medida $f_n(s; \mu)\nu_n(ds)$. Estos polinomios son Hermite (normal), Krawtchouk (binomial), Charlier (Poisson), Laguerre generalizado (gamma), Meixner (binomial negativa) y Pollaceck (secante hiperbólica). La relación de ortogonalidad es

$$E_\mu [P_{k,n}(S_n; \mu)P_{j,n}(S_n; \mu)] = \delta_{kj}\alpha_{j,n}V^j(\mu), \quad (6.13)$$

con

$$\alpha_{j,n} = j! \prod_{k=0}^{j-1} (n + kv_2), \quad j \geq 1 \text{ y } \alpha_{0,n} = 1. \quad (6.14)$$

Si $h \in \mathcal{U}_n$ y $T_n(S_n)$ es el estimador insesgado para el tamaño de muestra n , el desarrollo ortogonal de T_n es

$$T_n(s) = \sum_{j \geq 0} \frac{D_\mu^j h(\mu)}{\alpha_{j,n}} P_{j,n}(s; \mu), \quad \mu \in \Omega.$$

Para un valor conocido $\mu_0 \in \Omega$, y para cualesquiera enteros $n \geq 1$ y $j \geq 0$,

$$E_\mu[P_{j,n}(S_n; \mu_0)] = \alpha_{j,n}(\mu - \mu_0)^j / j!. \quad (6.15)$$

Se puede comprobar que en el caso NEF-QVF, para todo $n \geq 1$, $\mathcal{U}_{n,k}$ es la clase de todos los polinomios en μ de grado al menos k .

Se necesitaran algunos resultados previos.

Lema 6.2 *Las siguientes afirmaciones son equivalentes:*

- (a) $\mathcal{U}_{n,2} \subseteq \mathcal{U}_{m,2}$ para ciertos enteros n, m con $1 \leq n < m$.
- (b) La familia exponencial tiene la propiedad QVF.
- (c) $\mathcal{U}_{n,k} \subseteq \mathcal{U}_{m,k}$, para todo $k \geq 0$ y para todo $n \leq m$.

Demostración.

(a) \Rightarrow (b). Considérese la función

$$h(\mu) = nV(\mu) + n^2\mu^2. \quad (6.16)$$

Claramente el estimador umvu de $h(\mu)$ para el tamaño de muestra n es $T_n(S_n) = S_n^2$. Este estimador puede ser reescrito como, (ver lema 6.1),

$$T_n(S_n) = h(\mu) + \frac{h'(\mu)}{n} P_{1,n}(S_n, \mu) + P_{2,n}(S_n, \mu), \quad \mu \in \Omega,$$

Entonces, h pertenece a $\mathcal{U}_{n,2}$ y, por hipótesis, también a $\mathcal{U}_{m,2}$. De acuerdo con el lema 6.1, deben existir constantes reales $c_{0,m}$, $c_{1,m}$ y $c_{2,m}$ tales que

$$h(\mu) = mc_{2,m} \{V(\mu) + m\mu^2\} + c_{1,m}\mu + c_{0,m}. \quad (6.17)$$

Igualando (6.16) y (6.17),

$$(n - mc_{2,m})V(\mu) = (m^2c_{2,m} - n^2)\mu^2 + c_{1,m}\mu + c_{0,m}, \text{ para todo } \mu. \quad (6.18)$$

Claramente, de (6.18), si $n - mc_{2,m} \neq 0$, $V(\mu)$ es un polinomio de grado como mucho dos en μ , es decir, la familia exponencial tiene la propiedad QVF y se obtiene el resultado.

El caso $n - mc_{2,m} = 0$ es imposible, por que si $n = mc_{2,m}$, de (6.18), se debe tener necesariamente que $m^2c_{2,m} = n^2$ y $c_{2,m} \neq 0$; combinando estas expresiones

$$n^2 = m^2c_{2,m} = m^2c_{2,m}^2, \text{ con } c_{2,m} \neq 0,$$

por lo cual $c_{2,m}$ debe ser igual a uno. Esto implicaría que $m = n$, lo cual es absurdo, pues por hipótesis $m > n$.

(b) \Rightarrow (c). Este resultado es inmediato, por que si la familia exponencial tiene la propiedad QVF, entonces la clase $\mathcal{U}_{n,k}$, para cualquier $n \geq 1$, es el conjunto de todos los polinomios en μ con grado $\leq k$.

(c) \Rightarrow (a). Trivial. □

Por último, veamos la demostración del resultado principal.

6.2.2 Demostración

Supongamos que la propiedad martingala inversa se verifica. Para cierto $n \geq 1$ y $k > 1$, consideremos una función $h \in \mathcal{U}_{n,k}$. Sea $T_n(S_n)$ su estimador umvu para el tamaño de muestra n . Entonces,

$$T_n(S_n) = \sum_{j=0}^k a_j(\mu) P_{j,n}(S_n; \mu), \text{ para todo } \mu \in \Omega.$$

Para un tamaño de muestra $m \geq n$, y usando la propiedad martingala inversa, el estimador insesgado de $h(\mu)$ es

$$T_m(S_m) = E[T_n(S_n) | S_m] = \sum_{j=0}^k a_j(\mu) d_{j,n,m} P_{j,m}(S_m; \mu), \text{ para todo } \mu \in \Omega.$$

Entonces $h \in \mathcal{U}_{m,k}$. Por tanto, se tiene que $\mathcal{U}_{n,k} \subseteq \mathcal{U}_{m,k}$, para $k > 1$, (en particular para $k = 2$), y del lema 6.2, la familia tiene la propiedad QVF.

Recíprocamente, supongamos ahora que la familia es una NEF-QVF. Sean $n < m$ dos enteros positivos. Se define

$$R_k(S_m, \mu) = E [P_{k,n}(S_n; \mu) \mid S_m], \text{ para } k \geq 0 \text{ y } \mu \in \Omega.$$

Escogiendo un $\mu_0 \in \Omega$ arbitrario. Entonces, de (6.15),

$$E_\mu [R_k(S_m, \mu_0)] = E_\mu [P_{k,n}(S_n; \mu_0)] = \alpha_{k,n}(\mu - \mu_0)^k / k!, \text{ para todo } \mu \in \Omega. \quad (6.19)$$

La expresión (6.19) se puede interpretar diciendo que

$$k!R_k(S_m, \mu_0) / \alpha_{k,n}$$

es el estimador insesgado de $(\mu - \mu_0)^k$ para el tamaño de muestra m . De nuevo, de (6.15), se tiene

$$E_\mu [k!P_{k,m}(S_m; \mu_0) / \alpha_{k,m}] = (\mu - \mu_0)^k, \text{ para todo } \mu \in \Omega, \quad (6.20)$$

es decir que $k!P_{k,m}(S_m; \mu_0) / \alpha_{k,m}$ es también el estimador umvu de $(\mu - \mu_0)^k$ para el tamaño de muestra m . Como el estimador umvu es único (c.s.), se debe tener que

$$R_k(S_m, \mu_0) = \frac{\alpha_{k,n}}{\alpha_{k,m}} P_{k,m}(S_m; \mu_0),$$

ó equivalentemente, que la propiedad (6.4) se verifica con

$$d_{j,n,m} = \prod_{k=0}^{j-1} \frac{n + kv_2}{m + kv_2}, \text{ para } j \geq 1, \quad (6.21)$$

y $d_{0,n,m} = 1$. □

Una consecuencia de (6.21) y (6.13) es que en una familia exponencial natural con función de varianza cuadrática las variables aleatorias $\{Q_{k,n}(S_n; \mu)\}_{k \geq 0}$, con

$$Q_{k,n}(s; \mu) = \frac{P_{k,n}(s; \mu)}{\|P_{k,n}(s; \mu)\|_{n,\mu}^2},$$

donde $P_{k,n}$ son los polinomios definidos en (6.1), satisfacen la *propiedad martingala inversa en tiempo discreto* $E [Q_{k,n}(S_n; \mu) \mid S_m] = Q_{k,m}(S_m; \mu)$, $k \geq 0$.

6.3 Ejemplos

Para las seis subfamilias caracterizadas por Morris (1982), se tiene:

1.- *La distribución Normal.*

Esta distribución esta asociada a la NEF-QVF pues su función de varianza es $V(\mu) = \sigma^2$, $\mu \in \mathbb{R}$. El sistema de polinomios ortogonales es

$$P_{k,m}(s, \mu) = \sigma^k m^{k/2} He_k \left((s - m\mu) / (\sigma\sqrt{m}) \right),$$

donde $He_k(y)$ denota el k -ésimo polinomio de Hermite con expresión explícita

$$He_k(y) = k! \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{(-1)^j y^{k-2j}}{j! 2^j (k-2j)!}.$$

La propiedad martingala invertida es

$$E \left[n^{-k/2} He_k \left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \right) \middle| S_m \right] = m^{-k/2} He_k \left(\frac{S_m - m\mu}{\sigma\sqrt{m}} \right), \quad (\text{c.s.}), \quad (6.22)$$

para $0 < n \leq m$, $k \geq 0$ y $\mu \in \mathbb{R}$. La propiedad (6.22) era conocida para al menos los casos $k = 1, 2$ y ha sido utilizada para caracterizar los procesos de Wiener, ver Wesolowski (1990) y Wise (1992).

2.- *La distribución Poisson.*

Esta distribución esta asociada a la NEF-QVF pues su función de varianza es $V(\mu) = \mu$, $\mu > 0$. El sistema de polinomios ortogonales es $P_{k,m}(s, \mu) = C_k^{(m\mu)}(s)$, donde

$$C_k^{(\alpha)}(y) = \sum_{j=0}^k \binom{k}{j} \binom{x}{j} j! (-\alpha)^{k-j}, \quad \alpha > 0,$$

es el k -ésimo polinomio de Charlier. La correspondiente propiedad martingala invertida es

$$E \left[n^{-k} C_k^{(n\mu)}(S_n) \middle| S_m \right] = m^{-k} C_k^{(m\mu)}(S_m), \quad (\text{c.s.}), \quad (6.23)$$

para $0 < n \leq m$, $k \geq 0$ y $\mu > 0$. La propiedad (6.23) para $k = 1, 2$ fue usada por Wesolowski(1999) para caracterizar los procesos de Poisson.

3.- *La distribución gamma.*

La distribución $Ga(p, a)$ con función de densidad $f(x; p, a) = \frac{a^p}{\Gamma(p)} \exp(-ax) x^{p-1}$, $x > 0$, $p > 0$ y $a = p/\mu > 0$, tiene media μ y varianza $V(\mu) = \mu^2/p$, $m, p > 0$, por lo tanto es NEF-QVF. El sistema de polinomios ortogonales es

$$P_{k,m}(s) = k!(-1)^k L_k^{(p-1)}(as)/a^k,$$

donde

$$L_k^{(\alpha)}(y) = \sum_{j=0}^k (-1)^j \binom{k+\alpha}{k-j} \frac{y^j}{j!}, \alpha > -1,$$

es el k -ésimo polinomio de Laguerre generalizado. La propiedad martingala invertida es

$$E \left[\frac{\Gamma(pn)}{\Gamma(pn+k)} L_k^{(p-1)}(aX_n) | X_m \right] = \frac{\Gamma(pm)}{\Gamma(pm+k)} L_k^{(p-1)}(aX_m), \text{ (c.s.)},$$

para $0 < n \leq m$, $k \geq 0$ y $p, a > 0$.

4.- *La distribución binomial.*

La distribución binomial de parámetros $p \in (0, 1)$ y $N \geq 1$, pertenece también a la familia NEF-QVF, dado que su función de varianza es $V(\mu) = \mu + \mu^2/N$. El sistema de polinomios ortogonales es

$$P_{k,m}(s) = k! k_k(s; mN, \mu/N)$$

donde $k_k(x; N, p)$ denota al k -ésimo polinomio de Krawtchouk.

$$k_k(x; N, p) = \sum_{j=0}^k (-1)^j \binom{x}{k-j} \binom{N-x}{j} p^j (1-p)^{k-j}$$

En este caso, la propiedad martingala invertida es

$$E \left[\frac{\Gamma(nN - k + 1)}{\Gamma(nN + 1)} k_k(s; nN, \mu/N) | S_m \right] = \frac{\Gamma(mN - k + 1)}{\Gamma(mN + 1)} k_k(s; mN, \mu/N), \text{ (c.s.)},$$

para $0 < n \leq m$, $k \geq 0$ y $\mu > 0$.

5.- *La distribución binomial negativa.*

La función de probabilidad de esta distribución es $P[S_m = k] = \binom{k+r-1}{k} q_m^k p_m^r$, $k = 0, 1, \dots$; con $p_m = r/(m\mu + r)$, $q_m = 1 - p_m$ y $r > 0$. La función de varianza es $V(\mu) = \mu(\mu + r)/r$, $\mu > 0$. El sistema de polinomios ortogonales asociados es

$$P_{k,m}(s) = (-1)^k (m\mu/r)^k m_k(s; r, m\mu/(m\mu + r)),$$

donde $m_k(y; r, q)$ es el k -ésimo polinomio de Meixner de primera clase con expresión explícita

$$m_k(y; r, q) = (-1)^k k! \sum_{j=0}^k \binom{y}{j} \binom{-y-r}{k-j} q^{-j}, \quad r > 0, q \in (0, 1).$$

La propiedad martingala invertida es

$$E \left[\frac{\Gamma(np)}{\Gamma(np+k)} m_k \left(S_n; r, \frac{n\mu}{n\mu+r} \right) | S_m \right] = \frac{\Gamma(mp)}{\Gamma(mp+k)} m_k \left(S_m; r, \frac{m\mu}{m\mu+r} \right), \quad (\text{c.s.}),$$

para $0 < n \leq m$, $k \geq 0$ y $\mu, r > 0$.

6.- *La distribución Secante Hiperbólica Generalizada.*

Consideremos la NEF con densidad (respecto a la medida de Lebesgue) $f(x; \theta) = \exp\{x\theta + \log \cos(\theta)\} / (2 \cosh(\pi x/2))$, $x \in \mathbb{R}$, $|\theta| < \pi/2$. La media y la varianza son respectivamente $\mu = tg(\theta)$ y $V(\mu) = 1 + \mu^2$, $\mu \in \mathbb{R}$, por tanto la densidad pertenece a la NEF-QVF y es conocida como la familia secante hiperbólica generalizada, ver Morris (1982). La propiedad martingala invertida es

$$E \left[\frac{\Gamma(n)}{\Gamma(n+k)} P_{k,n}(S_n; \mu) | S_m \right] = \frac{\Gamma(m)}{\Gamma(m+k)} P_{k,m}(S_m; \mu), \quad (\text{c.s.})$$

para $0 < n \leq m$, $k \geq 0$ y $\mu \in \mathbb{R}$, donde el sistema de polinomios ortogonales asociados en este caso, $\{P_{k,n}(y; \mu)\}_{k \geq 0}$, es el sistema de polinomios de Pollaczek definidos en términos de la fórmula de recurrencia de tres términos (omitiendo argumentos por simplicidad)

$$\begin{aligned} P_{k+1,m} &= (P_1 - 2k\mu)P_k - km\{1 + (k-1)/m\}(1 + \mu^2)P_{k-1}, \quad k \geq 1 \\ P_{0,m} &= 1; \quad P_{1,m} = x - \mu. \end{aligned}$$

Capítulo 7

Una Relación entre la Esperanza Condicionada de los Records y la Estimación Insesgada en la Distribución Exponencial

7.1 Estimación insesgada en la distribución exponencial

Sea W_1, \dots, W_n una m.a.s. procedente de una distribución exponencial con función de densidad (respecto a la medida de Lebesgue)

$$f(x; \lambda) = \lambda \exp(-\lambda x), x > 0, \lambda > 0.$$

El estadístico $S_n = \sum_{i=1}^n W_i$ es suficiente, completo y sigue una distribución gamma con parámetros n y λ . Sea \mathcal{A} un subconjunto no vacío de $(0, \infty)$.

Diremos que una función $h(\lambda)$ es estimable en \mathcal{A} para el tamaño de muestra n , si existe una función del estadístico suficiente y completo $S_n, T_n(S_n)$, verificando

- (i) $E_\lambda[|T_n(S_n)|] < \infty$, para todo $\lambda \in \mathcal{A}$,
- (ii) $E_\lambda[T_n(S_n)] = h(\lambda)$, para todo $\lambda \in \mathcal{A}$.

Obsérvese que a diferencia de los capítulos anteriores, en este caso no se exige que la varianza del estimador sea finita. Sin embargo, si $\mathcal{A} = (0, \infty)$, y $T_n(S_n)$ tiene varianza finita podremos decir que la correspondiente función estimable es unvu-estimable en el sentido del capítulo 1. Si la función estimable admite un estimador que verifica (i) y (ii), pero con varianza infinita se puede concluir que cualquier estimador insesgado de dicha función tiene necesariamente varianza infinita.

Es posible encontrar una expresión cerrada en términos de la transformada inversa de Laplace para el estimador insesgado T_n de una función estimable h . De forma explícita, la condición de insesgaredad dada en (ii) es:

$$\frac{\lambda^n}{\Gamma(n)} \int_0^\infty T_n(s) s^{n-1} \exp(-\lambda s) ds = h(\lambda), \text{ para todo } \lambda \in \mathcal{A}. \quad (7.1)$$

La condición (i), $E_\lambda[|T_n(S_n)|] < \infty$, para todo $\lambda \in \mathcal{A}$, implica la existencia de la transformada de Laplace de $T_n(s)s^{n-1}$ para aquellos valores de λ que pertenecen a \mathcal{A} . Sea $\lambda_0 = \inf \mathcal{A}$. Entonces la transformada de Laplace de $T_n(s)s^{n-1}$ existe para cualquier $\lambda > \lambda_0$. Consecuentemente, se tiene la siguiente relación:

$$\frac{\lambda^n}{\Gamma(n)} \mathcal{L}(T_n(s)s^{n-1})(\lambda) = h(\lambda), \text{ para } \lambda > \lambda_0, \quad (7.2)$$

donde \mathcal{L} denota el operador transformada de Laplace. Por lo tanto el estimador insesgado de $h(\lambda)$ en \mathcal{A} para el tamaño de muestra n es $T_n(S_n)$, donde T_n tiene la expresión:

$$T_n(s) = \Gamma(n) \mathcal{L}^{-1}(h(\lambda)/\lambda^n)(s)/s^{n-1},$$

ver también López-Blázquez et al. (1997).

Si h es estimable y $T_n(S_n)$ es su correspondiente estimador insesgado en \mathcal{A} , puede comprobarse fácilmente que para $m > n$,

$$T_m(S_m) = E[T_n(S_n) | S_m] \quad (7.3)$$

es el estimador insesgado de h en \mathcal{A} para el tamaño de muestra m .

Mediante un cálculo elemental puede comprobarse que la densidad de S_n condicionada a $S_m = t$ es

$$f(s | t) = \frac{1}{\beta(n, m - n)} \frac{s^{n-1}(t - s)^{m-n-1}}{t^{m-1}} ds, \quad 0 < s < t, \quad (7.4)$$

y por lo tanto T_m puede obtenerse a partir de T_n de la relación (7.3):

$$T_m(t) = \frac{1}{\beta(n, m - n)} \int_0^t T_n(s) \frac{s^{n-1}(t - s)^{m-n-1}}{t^{m-1}} ds, \quad t > 0. \quad (7.5)$$

A partir de (7.2) puede obtenerse otra expresión que relaciona T_n y T_m ($n < m$):

$$\mathcal{L}(T_n(s)s^{n-1})(\lambda) = \frac{\Gamma(n)\lambda^{m-n}}{\Gamma(m)} \mathcal{L}(T_m(s)s^{m-1})(\lambda), \quad \lambda > \lambda_0. \quad (7.6)$$

7.2 Records

Sea $\{X_n\}_{n \geq 1}$ una sucesión de variables i.i.d. con función de distribución común absolutamente continua y estrictamente creciente F . De forma recurrente se define la sucesión de tiempos records, $\{L(n)\}_{n \geq 1}$, mediante

$$L(1) = 1$$

$$L(n) = \min \{j : j > L(n - 1) \text{ y } X_j > X_{L(n-1)}\}, \quad n \geq 2.$$

La variable aleatoria $X_{L(n)}$ se denomina el n -ésimo record de la sucesión original (o de la distribución F).

Nótese una ligera diferencia entre nuestra notación y la de algunos autores en la definición de los tiempos records. Mientras que nosotros comenzamos

el conteo de los records en 1, otros autores definen el récord 0-ésimo como el valor de la observación X_1 . La razón para este cambio es tener una cierta simetría en la notación: con nuestra notación el n -ésimo record de una distribución exponencial estandar (i.e. $\lambda = 1$) sigue una distribución gamma con parámetros n y 1.

Si Q es la función cuantil de F (que en este caso coincide con la función inversa) y Z es una variable aleatoria con distribución exponencial estandar, entonces

$$(X_{L(1)}, \dots, X_{L(n)}) \stackrel{d}{=} (R(Z_{L(1)}), \dots, R(Z_{L(n)})),$$

donde $R(s) = Q(1 - \exp(-s))$.

Tal como se mencionó anteriormente, el n -ésimo record de una distribución exponencial estandar sigue una distribución gamma de parámetros n y 1, de aquí que si F es una distribución para la que

$$E [|X_{L(n)}|] < \infty,$$

entonces

$$\int_0^{\infty} |R(s)| s^{n-1} \exp(-s) ds < \infty. \quad (7.7)$$

(Condiciones para la existencia de los momentos de los records pueden encontrarse en Nagaraja (1978))

Nótese que (7.7) implica la existencia de $\mathcal{L}(R(s)s^{n-1})(\lambda)$, para todo $\lambda \geq 1$.

La densidad conjunta de $Z_{L(n)}$ y $Z_{L(m)}$, $1 \leq n < m$, es

$$f_{n,m}(s, t) = \frac{1}{(n-1)!(m-n-1)!} s^{n-1} (t-s)^{m-n-1} \exp(-t), \quad 0 < s < t < \infty. \quad (7.8)$$

A partir de (7.8) puede obtenerse la densidad condicionada de $Z_{L(n)}$ a $Z_{L(m)} = t > 0$:

$$f_{n|m}(s | t) = \frac{1}{\beta(n, m-n)} \frac{s^{n-1} (t-s)^{m-n-1}}{t^{m-1}}, \quad 0 < s < t < \infty.$$

la cual coincide con (7.4).

Sea $\text{sop}(F) = \{x : 0 < F(x) < 1\}$ e $y \in \text{sop}(F)$. Supongamos que $E [|X_{L(n)}|] < \infty$ y $E [|X_{L(m)}|] < \infty$. Entonces, la esperanza condicionada de $X_{L(n)}$ a $X_{L(m)} = y$ puede calcularse como:

$$E [X_{L(n)} | X_{L(m)} = y] = \frac{1}{\beta(n, m-n)} \int_0^t R(s) \frac{s^{n-1} (t-s)^{m-n-1}}{t^{m-1}} ds, \quad (7.9)$$

donde t es tal que $Q(1 - \exp(-t)) = y$.

Sea

$$V(t) = E [X_{L(n)} | X_{L(m)} = Q(1 - \exp(-t))], \quad t > 0.$$

Entonces, con la notación de la primera sección la ecuación (7.8) tiene la siguiente interpretación: $R(S_n)$ y $V(S_m)$ son estimadores insesgados de la misma función paramétrica para los tamaños muestrales n y m respectivamente en el conjunto $\lambda > 1$.

7.3 Regresión lineal de records

La relación existente entre la esperanza condicionada de records y la estimación insesgada en la familia exponencial obtenida al final de la sección anterior será utilizada para caracterizar las distribuciones continuas para las que:

$$E [X_{L(n)} | X_{L(m)}] = bX_{L(m)} + a, \quad (7.10)$$

con $1 \leq n < m$.

Nagaraja (1988), caracterizó las distribuciones continuas que satisfacen la relación (7.10) en el caso particular de records adyacentes, es decir $m = n + 1$. El justificó el interés de este problema por sus aplicaciones en la predicción asintótica de estadísticos ordenados extremos. En el caso general en que los records considerados no sean adyacentes, y además la función de distribución que genera los records sea $m - n$ veces derivable López-Blázquez y Moreno-Rebollo (1997) probaron que las distribuciones caracterizadas por

la relación (7.10) eran precisamente las mismas tres familias de distribuciones caracterizadas por Nagaraja.

En esta sección nuestro objetivo será caracterizar las distribuciones que satisfacen (7.10) pero sin la hipótesis adicional de derivabilidad impuesta a la función de distribución F .

Nótese que si F es una distribución para la cual (7.10) se verifica necesariamente la pendiente b es positiva. Además, si se considera el cambio $Y = cX + d$, con $c > 0$, entonces

$$E [Y_{L(n)} | Y_{L(m)}] = bY_{L(m)} + a', \quad (7.11)$$

con $a' = ca + d(1 - b)$.

Por lo tanto, la clase de distribuciones para las cuales la regresión de $X_{L(n)}$ sobre $X_{L(m)}$ es lineal con pendiente b puede ser determinada salvo transformaciones de localización y escala.

El resultado principal de esta sección es el siguiente:

Teorema 7.1 Sean $1 \leq n < m$ dos enteros y F una función de distribución absolutamente continua y estrictamente creciente para la que $E [|X_{L(n)}|]$ y $E [|X_{L(m)}|]$ sean ambas finitas y $E [X_{L(n)} | X_{L(m)}]$ es lineal con pendiente $b > 0$. Entonces,

(i) Si $b \neq 1$, la función cuantil de F es (salvo transformaciones de localización y escala)

$$Q(u) = A \{\log(1/(1 - u))\}^\alpha, u \in (0, 1),$$

con $\alpha = r - n$, donde r es la única raíz positiva de la ecuación polinómica:

$$z(z + 1) \cdots (z + m - n - 1) = \frac{\Gamma(m)}{b\Gamma(n)}. \quad (7.12)$$

(ii) Si $b = 1$, la función cuantil de F es (salvo transformaciones de localización y escala)

$$Q(u) = \log \{\log(1/(1 - u))\}$$

Demostración.

(i) Caso $b \neq 1$. En este caso supondremos sin pérdida de generalidad que la ordenada en el origen es $a = 0$. Esto es así, por que si F es una distribución para la cual (7.10) se verifica, se puede considerar la variable aleatoria transformada $Y = X - (1 - b)^{-1}$, la cual satisface (7.11) con $a' = 0$.

De (7.9) y las consideraciones anteriores, la ecuación (7.10) puede ser escrita explícitamente como

$$\frac{1}{\beta(n, m - n)} \int_0^t R(s) \frac{s^{n-1}(t-s)^{m-n-1}}{t^{m-1}} ds = bR(t), \quad t > 0, \quad (7.13)$$

donde

$$R(s) = Q(1 - \exp(-s)).$$

Con la notación de la sección 1, $R(S_n)$ y $bR(S_m)$ son estimadores insesgados de la misma función paramétrica en $\lambda > 1$ para los tamaños de muestra n y m respectivamente.

Sea

$$G(\lambda) = \mathcal{L}(R(s)s^{n-1})(\lambda), \quad \lambda > 1.$$

Entonces, de (7.6) y las propiedades de la transformada de Laplace, se obtiene la siguiente ecuación diferencial

$$G(\lambda) = \frac{b\Gamma(n)}{\Gamma(m)} (-\lambda)^{m-n} G^{(m-n)}(\lambda), \quad \lambda > 1. \quad (7.14)$$

Como $G(\lambda)$ es la transformada de Laplace de $R(s)s^{n-1}$, entonces

$$G(+\infty) = \lim_{\lambda \rightarrow \infty} G(\lambda) = 0. \quad (7.15)$$

Las soluciones para $\lambda = +\infty$ de la ecuación diferencial (7.14) son de la forma

$$G(\lambda) = \lambda^{-r} \sum_{j=0}^{\infty} a_j \lambda^{-j}, \quad \text{con } r > 0 \text{ y } a_0 \neq 0. \quad (7.16)$$

Sustituyendo (7.16) en (7.14) e igualando los coeficientes del mismo grado, se obtiene

$$a_j \left(\frac{b\Gamma(n)}{\Gamma(m)} (j+r) \cdots (j+r+m-n-1) - 1 \right) = 0, \quad \text{para todo } j \geq 0. \quad (7.17)$$

En particular si $j = 0$ en (7.17), como $a_0 \neq 0$, se concluye que r debe ser una solución no negativa de la ecuación polinómica (7.12). Se puede comprobar fácilmente que (7.12) tiene una única raíz real no negativa.

En el caso $j > 0$, como r es la única raíz no negativa de (7.12), se tiene que

$$b\Gamma(n)(j+r)\cdots(j+r+m-n-1)/\Gamma(m)-1 \neq 0,$$

y necesariamente $a_j = 0$, para $j > 0$.

Entonces,

$$G(\lambda) = a_0\lambda^{-r} = \mathcal{L}(R(s)s^{n-1})(\lambda),$$

con $a_0 \neq 0$. De lo cual se obtiene

$$Q(u) = A \{\log(1/(1-u))\}^\alpha, u \in (0, 1),$$

y $\alpha = r - n$.

(ii) Caso $b = 1$. En este caso, no se puede suponer que la ordenada en el origen sea nula. En el caso, $a < 0$. La condición de regresión lineal es

$$\frac{1}{\beta(n, m-n)} \int_0^t (R(t) - R(s)) \frac{s^{n-1}(t-s)^{m-n-1}}{t^{m-1}} ds = -a, t > 0. \quad (7.18)$$

El cambio de variable $s = tw$, transforma (7.18) en

$$\frac{1}{\beta(n, m-n)} \int_0^1 (R(t) - R(tw)) w^{n-1} (1-w)^{m-n-1} dw = -a, t > 0, \quad (7.19)$$

y derivando con respecto a t en ambos miembros de (7.19), obtenemos

$$\int_0^1 (R'(t) - wR'(tw)) w^{n-1} (1-w)^{m-n-1} dw = 0, \text{ para todo } t > 0 \quad (7.20)$$

por lo cual obtenemos

$$R'(t) - wR'(tw) = 0, \text{ para todo } t > 0 \text{ y } w \in (0, 1).$$

Dado que, $R'(t) = R'(1)/t$, para todo $t > 0$, o equivalentemente, $R(t) = A \log(t) + B$, para ciertas constantes reales A y B , se tiene

$$Q(u) = A \log(\log(1/(1-u))) + B, u \in (0, 1).$$

□

Las funciones de distribución que se han caracterizado en el Teorema 7.1 son (salvo transformaciones de localización y escala):

1. Si $0 < b < 1$, puede comprobarse que r , es la única raíz no negativa de la ecuación polinómica (7.12), satisfaciendo $r > n$. Entonces, si $\theta = 1/\alpha > 0$, escogiendo $A = 1$, obtenemos

$$F(x) = 1 - \exp(-x^\theta), x > 0. \quad (7.21)$$

2. Si $b = 1$,

$$F(x) = 1 - \exp(-\exp(x)), x \in (-\infty, +\infty). \quad (7.22)$$

3. Si $b > 1$, tomando $r < n$. Entonces, si $\theta = -1/\alpha > 0$, escogiendo $A = -1$, se tiene

$$F(x) = 1 - \exp(-|x|^{-\theta}), x < 0. \quad (7.23)$$

Nótese que las distribuciones caracterizadas por regresión lineal en el caso de records no adyacentes son las mismas que caracterizó Nagaraja (1988) en el caso de records consecutivos. Estas distribuciones son las distribuciones de valor extremo, ver Gnedenko (1943).

7.4 Regresión lineal de estadísticos ordenados

Es bien conocido que existen muchas similitudes entre los estadísticos de orden y los records. Estas similitudes han sido investigadas por varios autores, entre otros, Deheuvels (1984) que afirma que, al menos en el caso continuo, no

se necesita desarrollar distintas teorías para la caracterización en términos de estadísticos de orden y en términos de records. Particularmente interesante en este sentido es también el trabajo de Gupta (1984).

Dado Y_1, \dots, Y_s una muestra i.i.d. de tamaño s , denotaremos por $Y_{(n:s)}$ al n -ésimo estadístico de orden.

Sea X una variable aleatoria con función de distribución continua y estrictamente creciente $F(x)$, y sea Y una variable aleatoria cuya función de distribución es proporcional a $-\log(1 - F(x))$.

Se puede comprobar fácilmente que la densidad condicionada de $X_{L(n)}$ dado $X_{L(m)}$, ($1 \leq n < m$) coincide con la densidad condicionada de $Y_{(n:s)}$ dado $Y_{(m:s)}$, ($1 \leq n < m \leq s$). De esta observación, tenemos que si la distribución de $X_{L(n)}$ dado $X_{L(m)}$, ($1 \leq n < m$) satisface cierta propiedad \mathcal{P} , entonces la distribución de $Y_{(n:s)}$ dado $Y_{(m:s)}$ también satisface \mathcal{P} . Como consecuencia, las distribuciones continuas para las cuales $E[Y_{(n:s)} | Y_{(m:s)}]$ son lineales con pendiente b son (salvo transformaciones de localización y escala):

1. Si $0 < b < 1$,

$$G(y) = y^\theta, \quad y \in (0, 1), \quad (7.24)$$

donde $\theta = 1/(r - n)$ y r es la única raíz positiva de la ecuación (7.12)

2. Si $b = 1$,

$$G(y) = \exp(y), \quad y \in (-\infty, 0). \quad (7.25)$$

3. Si $b > 1$,

$$G(y) = |y|^{-\theta}, \quad y \in (-\infty, -1). \quad (7.26)$$

donde $\theta = 1/(n - r)$ y r es la única raíz positiva de la ecuación (7.12).

Nótese que estas distribuciones coinciden con las caracterizadas por Ferguson(1967) en el caso de estadísticos ordenados adyacentes.

Bibliografía

- [1] ABBEY, J.L., DAVID, H. T. (1970). The construction of uniformly variance unbiased estimators for exponential distributions. *The Annals of Mathematical Statistics*, **41**(4), 1217-1222.
- [2] ABDULGHANI, A.H., ALKARBI, D.N., SHANBHAG, THABANE, L. (1997). Some structural properties of the Bhattacharyya matrices. Natural exponential families with quadratic variance functions. *Sankhyā A*, **59**(2), 232-241.
- [3] ABRAMOWITZ, M., STEGUN, I.A. (1965). *Handbook of Mathematical Functions*. Dover, New York.
- [4] ANDERSON, T.W., SAMUEL, S. M.(1967). Some inequalities among binomial and Poisson probabilities, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistical and Probability*, **1**, 1-12. Berkeley: University of California Press.
- [5] BARRANCO CHAMORRO, I. (1996). *Estimación paramétrica en familia no regulares*. Tesis doctoral.
- [6] BARANKIN, E. W. (1949), Locally best unbiased estimated. *Ann. Math. Statist.*, **20**, 477-501
- [7] BATTACHARYYA, A. (1946). On same analogues of the amount of information and their use in statistical estimation. *Sankhiā*, **8**, 1-14.

- [8] BENNET, W.S. (1965). *A New Binomial Approximation for Cumulative Hypergeometric Probabilities*. Ph. D. Thesis, American University, Washington, D.C.
- [9] BLIGHT, B.J.N., RAO, P.V. (1974). The convergence of Bhattacharyya bounds. *Biometrika*, **61**, 137-142.
- [10] BROWN, L.D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. Lecture Notes-Monograph Series, **9**, Institute of Mathematical Statistics
- [11] BURR, I.W. (1973). Some approximate relations between terms of the hypergeometric, binomial and Poisson distributions. *Commun. Stat.*, **1**(4), 297-301.
- [12] CACOULLOS, T., CHARALAMBIDES, CH. (1975). On minimum variance unbiased estimation for truncated binomial and negative binomial distributions. *Annals of the Institute of Statistical Mathematics*, **27**(2), 235-244.
- [13] CHARALAMBIDES, CH. (1974). The generalized Stirling and C-numbers, *Sankhyā, Series A*, **36**, 419-436.
- [14] CHIHARA, T. (1978). *An introduction to orthogonal polynomials*. Gordon & Breach, New York.
- [15] DAVIES, O.L. (1933). On asymptotic formulae for the hypergeometric series I, *Biometrika*, **25**, 295-322.
- [16] DAVIES, O.L. (1934). On asymptotic formulae for the hypergeometric series II, *Biometrika*, **26**, 59-107.
- [17] DEHEUVELS, P. (1984). The characterization of distributions by order statistics and record values- A unified approach. *J. Appl. Prob.*, **21**, 326-334.

- [18] FEND, A.V. (1959). On attainment of Cramer-Rao and Bhattacharya bounds for the variance of an estimate. *Ann. Math. Statist.* **30**, 381-388.
- [19] FELLER, W.(1968). *An Introduction to Probability Theory and its Application*, 3rd ed. Wiley, New York.
- [20] FERGUSON, T.S. (1967). On characterizing distributions by properties of order statistics, *Sankhyā A*, **29**, 265-278.
- [21] GHOSH, J.K., SATHE, Y.S. (1987). Convergence of Bhattacharyya bounds-revisited. *Sankhyā, A* **49**(1), 37-42.
- [22] GLASSER, G.J. (1962) Minimum variance unbiased estimators for Poisson probabilities. *Technometrics*, **4**(3), 409-418.
- [23] GNEDENKO, B.V. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire, *Ann. of Math.*, **44**, 423-453.
- [24] GRAY, H.L.; WATKINS, T.A.; SCHUCANY, W.R. (1973), On the jack-knife statistics and its relation to umvu estimators in the normal case. *Comm. Statist.*, **2**(4), 285-320.
- [25] GUPTA, R.C. (1974). Modified power series distributions and some of its applications, *Sankhyā, Series B*, **35**, 288-298.
- [26] GUPTA, R. C. (1984). Relationships between order statistics and record values and some characterization results, *J. Appl. Prob.*, **21**, 425-430.
- [27] HENRICI, P.(1993). *Applied and Computational Complex Analysis*, Vol. 3, Wiley, New York.
- [28] HODGES, J.L., LEHMANN, E.L. (1970). Deficiency. *The Annals of Mathematical Statistics*, **41**(3), 783-801.
- [29] HOTELLING, H, FRANKEL, L.R. (1938). The transformation of statistics to simplify their distributions. *Ann. Math. Statist.*, **9**, 87-96.

- [30] HWANG, T.Y; HU, C.Y. (1990), More comparisons of mle with umvue for exponential families. *Ann. Inst. Statist. Math.*, **42**(1), 65-75.
- [31] JANI, P.N. (1977). Minimum variance unbiased estimation for some left truncated modified power series distributions, *Sankhyā, Series B*, **39**(3), 258-278.
- [32] JANI, P.N. (1978). New numbers appearing in minimum variance unbiased estimation for decapited generalized, negative binomial and Poisson distribution, *Journal of the Indian Statistical Association* , **16**, 41-48.
- [33] JOHNSON, N., KOTZ, KEMP, A. (1992). Univariate discrete distributions. John Wiley and Sons, 2nd. Ed.
- [34] JOSHI, S.W., PARK, C.J. (1974). Minimum variance unbiased estimation for truncated power series distributions, *Sankhyā, Series A*, **36**, 305-314.
- [35] KARLIN, S.; STUDDEN, W.J. (1966), *Tchebycheff Systems with Applications in Analysis and Statistics*, Wiley Interscience, New York.
- [36] KHAN R.A., (1984), On umvu estimators and Bhattacharya bounds in exponential distributions, *J. Stat. Plan. Infer.*, **9** 199-206.
- [37] KARAKOSTAS, K.X. (1985). On minimum variance unbiased estimators, *The American Statistician*, **4**(1), 303-305.
- [38] KEMP, C. D., KEMP, A.W. (1956). Generalized hypergeometric distributions, *Journal of the Royal Statistical Society, Series B*, **18**, 202-211.
- [39] KOSAMBI, D.D. (1949). Characteristic properties of series distributions, *Proceedings of the National Institute for Science, India*, **15**, 109-113.
- [40] KUMAR, A., CONSUL, P.C. (1980). Minimum variance unbiased estimation for modified power series distributions, *Communications in Statistics: Theory and Methods*, **A9**(12), 1261-1275.

- [41] LEHMANN, E.L. (1959). *Testing Statistical Hypotheses*, Wiley, New York.
- [42] LEHMANN, E.L. (1983). *Theory of Point Estimation*, Wiley, New York.
- [43] LINDSAY, B.G. (1989), On the determinants of moment matrices. *Ann. Statist.*, **17**(2), 711-721.
- [44] LOPEZ-BLAZQUEZ, F., (1998), A note on umvu estimation in the transformed chi-square distribution. *Comm. Statist. Theory and Methods*, **27**(8), 2065-2073.
- [45] LOPEZ-BLAZQUEZ, F. (2000). Unbiased estimation in the Non-central Chi-Square distribution. *Journal of Multivariate Analysis*, **75**, 1-12.
- [46] LOPEZ-BLAZQUEZ, F., BARRANCO CHAMORRO, I., MORENO-REBOLLO, J.L. (1997). Umvu estimation for certain exponential distributions, *Communications in Statistics: Theory and Methods*, **26**(2), 469-482.
- [47] LOPEZ-BLAZQUEZ, F., MORENO-REBOLLO, J.L. (1997). A characterization of distribution based on linear regression of order statistics and records values. *Sankhyā A*, **59**(3), 311-323
- [48] LOPEZ-BLAZQUEZ, F., CASTANO-MARTINEZ, A. (1999). Asymptotic properties of unbiased estimators in the natural exponential family with quadratic variance function. *Sankhyā A*, **61**(2), 292-297.
- [49] LOPEZ-BLAZQUEZ, F., SALAMANCA MIÑO, B. (2000). Estimation based on the Winsorized mean in the geometric distribution. *Statistic*, **35**, 81-95.
- [50] MORRIS, C.N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.*, **10**, 65-80.

- [51] MORRIS, C.N. (1983). Natural exponential families with quadratic variance functions: Statistical theory. *Ann. Statist.*, **11**, 515-529.
- [52] NAGARAJA, H.N. (1978). On the expected value of record values, *Austral. J. Statist.*, **20**(2), 176-182.
- [53] NAGARAJA, H.N. (1984). Asymptotic linear prediction of extreme order statistics. *Ann. Inst. Statistit. Math.*, **36**(A), 289-299.
- [54] NAGARAJA, H.N. (1988). Some characterizations of continuous distributions based on regressions of adjacent order statistics and record values. *Sankhyā A*, **50**(1), 265-278.
- [55] NEWELL, D.J. (1965). Unusual frequency distributions, *Biometrics*, **21**, 159-168.
- [56] NOACK, A. (1950). A class of random variables with discrete distributions, *Annals of Mathematical Statistics*, **27**, 471-483.
- [57] ORD, J.K. (1955). Approximations to distribution functions which are hypergeometric series. *Biometrika*, **55**(1), 243-248.
- [58] PATIL, G.P. (1963). Minimum variance unbiased estimation and certain problems of additive number theory, *Annals of Mathematical Statistics*, **34**, 1050-1056.
- [59] PATIL, G.P., JOSHI, S.W. (1968). *A Dictionary and Bibliography of Discrete Distributions*. Oliver & Boyd, Edinburgh.
- [60] POISSON, S.D. (1837). *Recherches sur la Probabilité des Jugements en Criminelle et en Matière Civile, Précédées des Règles Générales du Calcul des Probabilités*. Bachelier, Imprimeur-Libraire pour les Mathématiques, la Physique, etc. Quai des Augustins, No. 55.
- [61] POMMERET, D. (1996). Orthogonal polynomials and natural exponential families. *Test*, **5**(1), 77-111.

- [62] POOR, H.V. (1991). The maximum difference between the binomial and Poisson distributions. *Stat. Probabil. Lett.*, **11**, 103-106.
- [63] PORTNOY, S, (1977), Asymptotic efficiency of minimum variance unbiased estimators. *Ann. Statist.*, **5**(3), 522-529.
- [64] RAO, C.R. (1961). Asimptotic efficiency and limiting information. *Proc. Fourth Berkeley Symp. on Math. Statist. Probab.*, **1**, 531-546. Univ. of California Press, Berkeley.
- [65] RAO, C.R. (1962). Efficiency estimates and optimum inference procedure in large samples, with discusion. *J. Roy. Statistic Soc.*, Ser. B, **24**, 46-63.
- [66] RAO, C.R. (1963). Criteria of estimation in large samples. *Sankhyā A*, **25**, 189-206.
- [67] ROY, J., MITRA, S.K. (1957). Unbiased minimum variance unbiased estimation in a class of discrete distributions, *Sankhyā*, **18**, 371-378.
- [68] SCHOUTENS, W., TEUGELS, J.L. (1998). Lévy processes, polynomials and martingales. *Stochastic Models*, **14**, (1,2), Special Issue in Honor of Marcel F. Neuts, p.335-349.
- [69] SERFLING, R.J. (1980). *Approximations Theorems of Mathematical Statistics*. John Wiley & Sons, New York.
- [70] SETH, G.R. (1949). On the variance of estimates. *Ann. Math. Statist.*, **20**, 1-27.
- [71] SHANBAG, D.N. (1972). Some characterizations based on the Bhattacharyya matrix. *J. Appl. Probab.*, **9**, 580-587.
- [72] SHIRYAYEV, A.N. (1984), *Probability*. Springer Verlag.
- [73] SIBUYA, M., SHIMIZU, R. (1981). The generalized hypergeometric family of distributions. *Ann. Inst. Statist. Math.*, **33**, 177-190.

- [74] SIMONS, G., JOHNSON, N. L. (1971). On the convergence of binomial to Poisson distributions. *Ann. Math. Statist.*, **42**(5), 1735-1736.
- [75] VERVAAT, W. (1970). Upper bounds for the distance in total variation between the binomial or the negative binomial and the Poisson distribution. *Statistica Neerlandica*, **23**, 79-86.
- [76] VOINOV, V.G. (1986). On Jani's paper on the minimum variance unbiased estimation for a left-truncated power series distribution, *Sankhyā, Series B*, **48**(1), 144-150.
- [77] WASOW, W. (1956). On the asymptotic transformation of certain distributions into the normal distribution. *Proc. 6th. Symp. Appl. Math. Soc. (Numerical Analysis)*, vol VI, 251-259, New York, McGraw Hill.
- [78] WESOŁOWSKI, J. (1990). A martingale characterization of the Wiener process, *Statistics and Probability Letters*, **10**, 213-215.
- [79] WESOŁOWSKI, J. (1999). Poisson process via martingale and related characterizations, *Journal of Applied Probability*, **36**, 919-926.
- [80] WISE, G.L. (1992). A counterexample to a martingale characterization of a Wiener process, *Statistics and Probability Letters*, **15**, 337-338.
- [81] WOODWARD, W.A.; GRAY, H.L. (1975), Minimum variance unbiased estimation in the gamma distribution, *Comm. Statist.*, **4**(10), 907-922.
- [82] ZACKS, S. (1971), *The Theory of Statistical Inference*, Wiley, New York.

UNIVERSIDAD DE SEVILLA

Reunido el Tribunal Integrado por los señores firmantes
en el día de la fecha, para juzgar la Tesis Doctoral de
D.^a Begonia Salamanca Abián
de la eliminación incespada aplicada a la aproximación
y caracterización de las distorsiones

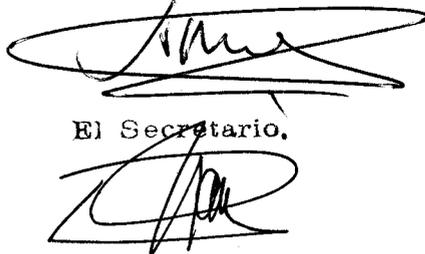
acordó otorgarle la calificación de Sobresaliente Cum Laude
por unanimidad

Sevilla, 9 de Febrero 2001

El Vocal,


El Presidente

El Vocal,


El Secretario.

El Vocal,


El Doctorado,



* 5 0 1 3 5 4 1 8 1 *

FMA C 043/360