UNIVERSIDAD DEL BÍO-BÍO
FACULTAD DE EDUCACIÓN Y HUMANIDADES
DEPARTAMENTO DE CIENCIAS DE LA EDUCACIÓN
ESCUELA DE PEDAGOGÍA EN EDUCACIÓN MATEMÁTICA

ESTUDIO DE LA DISTRIBUCIÓN DE HERMITE BIVARIADA

MEMORIA PARA OPTAR AL TITULO DE PROFESOR DE EDUCACIÓN MEDIA EN EDUCACIÓN MATEMÁTICA

AUTOR: GONZÁLEZ ALBORNOZ, PABLO ÁNDRES
Profesor Guía: Dr. Novoa Muñoz, Francisco

CHILLÁN, 2016

DEDICATORIA

A mi esposa Karen. Quien me apoyo y alentó para continuar, cuando parecía que me iba a rendir.

A mi madre Gladys. Por haberme apoyado en todo momento, por sus consejos, sus valores, por la motivación constante que me ha permitido ser una persona de bien, pero más que nada, por su amor.

A mi padre Axcel. Por los ejemplos de perseverancia y constancia que lo caracterizan y que me ha infundado siempre, por el valor mostrado para salir adelante y por su amor. Para ellos es esta dedicatoria de tesis, pues es a ellos a quienes se las debo por su apoyo incondicional

AGRADECIMIENTOS

A Dios

Mi agradecimiento se dirige a quien ha forjado mi camino y me ha dirigido por el sendero correcto, a Dios, el que en todo momento está conmigo ayudándome a aprender de mis errores y a no cometerlos otra vez. Eres quien guía el destino de mi vida.

Mi esposa Karen

La ayuda que me has brindado ha sido sumamente importante, estuviste a mi lado inclusive en los momentos y situaciones más tormentosas, siempre ayudándome. No fue sencillo culminar con éxito este proyecto, sin embargo siempre fuiste muy motivadora y esperanzadora, me decías que lo lograría perfectamente. Me ayudaste hasta donde te era posible, incluso más que eso. Muchas gracias, amor.

Papá, Mamá, Axcel y Lorena

Suponen los cimientos de mi desarrollo, todos y cada uno de ustedes, han destinado tiempo para enseñarme nuevas cosas, para brindarme aportes invaluables que servirán para toda la vida.

Especialmente estuvieron presentes en la evolución y posterior desarrollo total de mi tesis, les agradezco con creces. Los amo

ÍNDICE GENERAL

Dedicatoria I

Agradecimientos II

Lista de figuras V

1. Introducción 1

1.0.1. Distribución de Hermite 2

2. Estudio Probabilistico 4

2.1. Introducción 4

2.1.1. Funciones Generadoras 5

2.1.2. Distribuciones Condicional y Marginal 7

2.2. Función generadora de Probabilidad 8

2.3. Distribución Marginal 10

2.4. Función de Probabilidad 10
2.5. Relación de Recurrencia 11
2.6. Momentos y Cumulantes 13
2.7. Independencia 13
2.8. Distribución Condicional y Regresión 14

3. Estudio inferencial **16**

3.1. Introducción 16
3.1.1. Estimación 17
3.2. Estimación Puntual 26
3.2.1. Método de los momentos 27
3.2.2. Método de Event-Point 27
3.2.3. Método de Maxima Verosimilitud 30

Conclusiones **32**

Bibliografía **33**

ÍNDICE DE FIGURAS

2.1. Población y muestra 4

3.1. Algunos valores de los parámetros para que $E > 0,7$ para el método even-point 30

CAPÍTULO 1

INTRODUCCIÓN

En la teoría de la probabilidad y estadística, la distribución de probabilidad de una variable aleatoria es una función que permite obtener probabilidad de ocurrencia [9]. Las distribuciones más conocidas son la distribución Normal, distribución T-Student, distribución Binomial, distribución de Poisson, etc. Estas distribuciones son herramientas estadísticas que permiten modelar fenómenos reales, en los cuales se desee realizar un análisis estadísticos e inferencial.

La distribución de Hermite resulta al sumar dos variables de Poisson (X_1 y X_2) dependientes, de la forma $Y = X_1 + 2X_2$ [5]. La distribución es flexible a la hora de modelar datos de conteo que presenten una multimodalidad, junto con presentar varios ceros, lo que se denomina cero-inflación. Támbien permite modelar datos en los que la sobredispersión es moderada; esto es que la varianza es mayor a la esperanza. La distribución de Hermite para el caso en que n = 2 coincide con la distribución Binomial de Poisson (n: nº de ocurrencias del evento o fenómeno). Recibe el nombre porque los matemáticos Kemp and Kemp la llamaron así , debido a que la función de probabilidad y la función generadora de momentos se pueden expresar en términos de los polinomios de Hermite modificados [1].

En esta tesis realizaremos una investigación a la distribución de hermite bivariada, obtenida por kocherlakota (1988) [11] a través de una variable de Poisson bivariada con parametros $\tau\lambda_1$ y $\tau\lambda_2$, de igual forma por Kemp y Papageorgio (1982) [10] con un reparametrización.

Fue A.G. McKendrick en 1926 , en *applications of mathematics to medical problems* [12] quien publicó por primera vez sobre la distribución de Hermite. Intentando explicar diversos métodos matemáticos que se pueden aplicar en investigaciones médicas. McKendrick demostró, lo que posteriormente seria llamado la distribución de Hermite, al sumar dos variables de Poisson dependientes. Como aplicación McKendrick modeló un experimento fagocítico (recuento de bacterias en leucocitos) a través de la distribución de Hermite, obteniendo un modelo mas satisfactorio que con la distribución de Poisson.

C.D. Kemp y A.W. Kemp en 1965, quienes en *"Some Properties of Hermite Distribution"* [1] introdujeron formalmente la distribución de Hermite. Esta investigación se concentra en el cálculo de la función de probabilidad, su relación con los polinomios de Hermite y la estimación de parámetros con el método de máxima verosimilitud.

El estudio de la generalización de la distribución de Hermite lo hicieron Gupta y Jain en .*A Generalized Hermite Distribution and Its Properties"*.(1974) [13]

A C.D Kemp y H. Papageorgiou en 1982 [10] estudiaron varias formas en la que la distrbución de Hemite bivariada puede ser expresada por procesos de reparametrización.

En 1983 H. Papageorgiou, C.D. Kemp. y S. Loukas publicaron en Biometrika: *Some Methods of Estimation For the Bivariate Hermite Distribution"* [4] donde exponen tres métodos de estimación: estimación por punto de equilibrio, por momentos y otra con frecuencia de cero.

1.0.1. Distribución de Hermite

Funcion de probabilidad

Dadas dos variables independientes de Poisson, X_1 y X_2 con parámetros a_1 y a_2 respectivamente, la distribución de la varibale aleatoria $Y = X_1 + 2X_2$ sigue una distribución de Hermite con parámetros a_1 y a_2 cuya función de probabilidad es:

$$p_n = P(Y = n) = exp[-(a_1 + a_2)] \sum_{j=0}^{[n/2]} \frac{a_1^{n-2j} a_2^j}{(n-2j)!j!} \tag{1.1}$$

Donde:

- $n = 0, 1, 2, 3, ...$
- $[n/2]$ es la parte entera de $n/2$
- $a_1, a_2 \geq 0$

Función generatriz de probabilidad

Esta función es útil para generar las probabilidades de la distribución y sus momentos.

$$G_Y(s) = \sum_{n=0}^{\infty} P_n s^n = exp[a_1(s-1) + a_2(s^2-1)] \tag{1.2}$$

$$G_Y(s) = \sum_{n=0}^{\infty} P_n s^n = exp[-(a_1+a_2) + a_1 s + a_2 s^2] \tag{1.3}$$

Observación 1.1 (Notación)**.** *De una variable $Y = X_1 + 2X_2$ que siga una distribución de Hermite escribiremos*

$$Y \sim Herm(a_1, a_2)$$

Función generadora de momentos

$$M(t) = G(e^t) = exp[-(a_1+a_2) + a_1 e^t + a_2 e^{2t}] \tag{1.4}$$

Función de distribución

$$F(x; a_1, a_2) = P(X \leq x) = exp[-(a_1+a_2)] \sum_{i=0}^{[x]} \sum_{j=0}^{[i/2]} \frac{a_1^{i-2j} a_2^j}{(i-2j)!j!} \tag{1.5}$$

Observación 1.2 (Propiedades)**.**

- *No es una distribución unimodal*
- *Distribución cerrada bajo convoluciones: Sea $X_1 \sim Herm(a_1, a_2)$, $X_2 \sim Herm(a_3, a_4)$ y $Y = X_1 + X_2$, entonces Y sigue una distribución Hermite*
- *Permite modelar datos que tienen un exceso de ceros, esto se denomina cero-inflación*
- *La distribución de Hermite tiene sobdredispersión moderada; es decir que la varianza es mayor que la esperanza.*

CAPÍTULO 2

ESTUDIO PROBABILISTICO

2.1. Introducción

En el mundo de la estadística, uno de los conceptos principales es el muestreo. En casi todos los problemas de estadística, de un cuerpo de datos llamado población, se toma un número más pequeño de mediciones , es decir, una muestra. Como se expresa en la Figura 2.1

Nuestro principal interés esta en la población, pero ésta puede ser difícil o imposible de enumerar. Imagine tratar de registrar la estatura de todas las personas sanas del mundo o de la preferencia presidencial de todo votante registrado en Chile. En cambio, tratamos de describir o inferir el comportamiento de la población en base a la información obtenida de una muestra representativa de esa población.

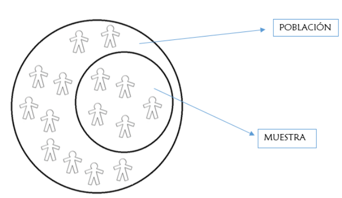

Figura 2.1: Población y muestra

La probabilidad se emplea como herramienta; permite que usted evalúe la confiabilidad de sus conclusiones acerca de la población.

2.1.1. Funciones Generadoras

En el estudio de variables aleatorias existe una variedad de funciones que se han introducido para facilitar el resumen de las propiedades en una forma funcional compacto pero manejable. Hay una serie de tales propiedades que son de interés en general. Por lo tanto de manera correspondiente, hay una serie de definiciones para generar estas funciones, cada uno dando lugar a una una propiedad específica de la variable aleatoria. En esta sección vamos a definir estas funciones útiles para el desarrollo de las relaciones entre las diversas características de las variables aleatorias.

Definición 2.1 (Función Generadora de Probabilidad (fgp))**.** *La función generadora de probabilidad (fgp) del par de variables aleatorias* (X, Y) *con función de probabilidad* $f(x, y)$ *es la* $E[t_1^x t_2^y]$.

Designaremos a la función generadora de probabilidad como $\Pi(t_1, t_2)$. *Desde el definición se aprecia fácilmente que*

$$\Pi(t_1, t_2) = \sum_{(x,y)\in T} t_1^x t_2^y f(x, y) \tag{2.1}$$

de la ecuación (2.1) es evidente que al ser una serie geométrica es absolutamente convergente en el rectángulo unido $|t_1| \leq 1, |t_2| \leq 1$. *Como tal, la función generadora de probabilidad es infinitamente diferenciable a las variables* (x, y) *en* $(0, 0)$ *. También es única en el sentido de una relación de uno a uno con la función de probabilidad. Por lo cual, de la ffunción generadora de probabilidad se puede obtener la función de probabilidad de las variables.*

La función generadora de probabilidad da lugar a la función de probabilidad de las variables aleatorias. Sin embargo, nuestro mayor interés es en los momentos conjuntos $\mu'_{r,s} = E(X^r Y^s)$ *de la variables aleatorias* X *e* Y. *Para ello tenemos la definición de la función generadora de momentos (fgm), la cual depende de la existencia de los momentos. Para asegurarse de que no existe, se utiliza la definición restringida dada por Hogg y Craig (1978, p. 77) [14].*

Definición 2.2 (Función Generadora de Momentos (fgm))**.** *La función generadora de momentos del par de variables aleatorias X e Y es el $E[exp(t_1X + t_2Y)]$ siempre que exista (t_1, t_2), situada en el rectángulo $|t_1| \leq h_1, |t_2| \leq h_2$ donde h_1 y h_2 son números reales positivos. La función generadora de momentos está representado por $M(t_1, t_2)$. A partir de la definición:*

$$M(t_1, t_2) = \sum_{x,y} exp[t_1^x + t_2^y]f(x, y) \tag{2.2}$$

Observación 2.1. *Como se ha mencionado, esta definición de la función generadora de momentos asume que existen todos los momentos. Sin embargo, si la función generadora de momentos no existe, es necesario la función caracteríística definida sobre el plano complejo.*

La ampliación de las funciones exponenciales en (2.2), es posible escribir

$$M(t_1, t_2) = \sum_{x,y} \frac{t_1^r}{r!}\frac{t_2^s}{s!}\mu'_{r,s} \tag{2.3}$$

De esta expansión y el hecho de que el orden de la diferenciación y la suma puede ser intercambiado para establecer las siguientes reglas para la determinación de los momentos, dada la función generadora de momentos

- $\mu'_{r,s}$ *es el coeficiente de* $\frac{t_1^r}{r!}\frac{t_2^s}{s!}$ *en la expansión de la función generadora de momentos en potencia de t_1 y t_2.*
- $\mu'_{r,s}$ *viene dada por la derivado parcial mixta*

$$\frac{\partial^{r+s}}{\partial t_1^r \partial t_2^s} = M(t_1, t_2)|_{t_1=0, t_2=0} \tag{2.4}$$

Al comparar las funciones (2.1) y (2.2) obtenemos la siguiente relación

$$M(t_1, t_2) = \Pi[expt_1, expt_2] \tag{2.5}$$

Definición 2.3 (Función Generadora de cumulantes)**.** *La función generadora cumulantes se define, siendo el vector aleatorio bidimensional $X = (X_1, X_2)$ con función generadora de momentos $M(t_1, t_2)$. Suponiendo que existen todos los momentos para X. Podemos considerar el logaritmo de la función generadora de momentos y desarrollarlo en serie de potencias, dando lugar a la función generatriz de cumulantes bidimensionales*

$$K(t_1, t_2) = logM[t_1, t_2] = \sum_{r}\sum_{s}{}_{(r,s)\neq(0,0)} \frac{t_1^r}{r!}\frac{t_2^s}{s!}\kappa_{r,s} \tag{2.6}$$

En la cantidad $\kappa_{r,s}$ se llama el cumulante de orden (r, s)

2.1.2. Distribuciones Condicional y Marginal

Es interesante tener en cuenta las distribuciones marginales y condicionales, La distribución marginal describe el comportamiento individual de cada una de las variables aleatorias. Por otro lado, las distribuciones condicionales lleva a cabo la interdependencia de las variables aleatorias. Por lo tanto la distribución condicional puede dar interpretaciones simples y útiles.

Es posible relacionar las funciones generadoras de probabilidad de las distribuciones condicionales con la función generadora de probabilidadde la distribución conjunta. Aunque el método habitual de la determinación de la fp condicional directamente de la articulación de las funciones de probabilidad marginales se puede utilizar, esto implicará la determinación de la función de probabilidad conjunta. Desafortunadamente, en el caso de las distribuciones discretas bivariadas esto no siempre es un problema trivial. También el función de probabilidad condicional resultante no se presta a la interpretación útil, siendo bastante intratable en la mayoría de los casos que se presentan en la práctica. Por otro lado, se verá que el uso de la función generadora de probabilidad en la determinación de la distribución condicional sin tener que encontrar la función de probabilidad bivariante es mucho más general. También es útil para dar una mayor comprensión de la naturaleza de las distribuciones condicionales. La función generadora de probabilidad resultante se ve en la mayoría de situaciones útil, pero simple.

Definición 2.4 (Distribuciones Marginales)**.** *Dada la función de probabilidad conjunta de* (X, Y), $f(x, y)$. *A continuación, las funciones de probabilidad marginales son* $g(x) = \sum_y f(x, y)$ *y* $h(y) = \sum_x f(x, y)$. *Las funciones de generadoras de las distribuciones marginales se pueden determinar de las función de probabilidad conjunta. Así, las funciones generadoras de probabilidad marginales son:*

$$\begin{aligned} \Pi_x(t) &= \sum_x g(x)t^x = \sum_x t^x \sum_y f(x, y) = \sum_x \sum_y f(x, y)t^x = \Pi(t, 1) \\ \Pi_y(t) &= \sum_y g(x)t^y = \sum_y t^y \sum_x f(x, y) = \sum_y \sum_x f(x, y)t^y = \Pi(1, t) \end{aligned} \tag{2.7}$$

Observación 2.2. *En forma análoga a partir de la definición de la función generadora de momentos es posible ver que las funciones generadoras de momentos de las distribuciones marginales son* $M_x(t) = M(t, 0)$ *y* $M_y(t) = M(0, t)$.

Definición 2.5 (Distribución Condicionada)**.** *Por las definiciones, la función de probabilidad condicional de* Y, *dado* $X = x$, *es:*

$$f(y|x) = \frac{f(x, y)}{g(x)} \tag{2.8}$$

Teorema 2.1. *Sea* $\Pi(t_1, t_2)$ *la función generadora de probabilidad conjunta de (X, Y). Entonces, la función generadora de probabilidad de la distribución condicional de* Y *dado* $X = x$ *es:*

$$\Pi_y(t|x) = \frac{\Pi^{(x,0)}(o,t)}{\Pi(x,0)(0,1)} \tag{2.9}$$

Este teorema se puede utilizar para determinar la regresión de Y *en* X.

Regresión de Y *en* X *es*

$$E[Y|X = x] = \frac{\Pi^{(x,1)}(o,1)}{\Pi(x,0)(0,1)} \tag{2.10}$$

Definición 2.6 (Independencia)**.** *Dadas dos variables aleatorias* X *e* Y*, se dirán que son independientes si, y sólo si* $f(x, y) = g(x)h(y)$ *para todo* $(x, y) \in T$

Una propiedad de la función generadora de probabilidad es :

Teorema 2.2. *Sea la propiedad de la función generadora de probabilidad conjunta de las variables aleatoria* X *e* Y *es* $\Pi(t_1, t_2)$ *con sus funciones generadoras de probabilidad marginales* $\Pi_x(t)$ *y* $\Pi_y(t)$ *respectivamente. Entonces* X *e* Y *son independiente si, y sólo si* $\Pi(t_1, t_2) = \Pi_x(t)\Pi_y(t)$

2.2. Función generadora de Probabilidad

Kocherlakota [3] presenta 3 formas que se ha demostrado que surge la distribución de Hermite bivariada:

- *Utilizando el resultado de Kocherlakota (1998) donde* $\underline{X} \underset{\tau}{\wedge} N(\mu, \sigma^2)$*, donde* $\underline{X}$ *se distribuye como una bivariables de Poisson con parametros* $\tau\lambda_1$*,* $\tau\lambda_2$ *y* $\tau\lambda_3$*, donde la función de probabilidad es*

  $$\Pi(t_1, t_2) = exp(\mu u + \frac{1}{2}\sigma^2\mu^2) \tag{2.11}$$

  con

  $$u = \lambda_1(t_1 - 1) + \lambda_2(t_2 - 1) + \lambda_3(t_1t_2 - 1)$$

- *Kemp y Papageorgiou (1982) consideran que una forma en que la distribución (2.11) se puede expresar mediante un proceso de reparametrización es:*

$$\begin{gathered}\Pi(t_1, y_2) = exp[a_1(t_1-1) + a_2(t_1^2-1) + a_3(t_2-1) + a_4(t_2^2-1)] \\ +a_5(t_1t_2-1) + a_6(t_1^2t_2-1) + a_7(t_1t_2^2-1) + a_8(t_1^2t_2^2-1)\end{gathered} \tag{2.12}$$

  y la alternativa

$$\begin{gathered}\Pi(t_1, y_2) = exp[b_1t_1^* + b_2t_1^{2*} + b_{t_1^*} + b_4t_2^{2*} + \\ b_5t_1^*t_2^* + b_6t_1^{2*}t_2^* + b_7t_1^*t_2^{2*} + b_8t_2^{2*}t_2^{2*}]\end{gathered} \tag{2.13}$$

  Cuando $t_1^* = t_1 - 1$ *y* $t_2^* = t_2 - 1$. *Para ambas ecuaciones los coeficientes son funciones de los cinco parámetros,* λ_1, λ_2, λ_3, μ *y* σ. *En (2.12) con* $\tau = \mu + \sigma^2\gamma$, *estos parámetros son*

  - $a_1 = \lambda_1\tau$
  - $a_2 = \frac{1}{2}\sigma^2\lambda_1^2$
  - $a_3 = \lambda_2\tau$
  - $a_4 = \frac{1}{2}\sigma^2\lambda_2^2$
  - $a_5 = \lambda_3\tau + \sigma^2\lambda_1\lambda_2$
  - $a_6 = \sigma^2\lambda_1\lambda_3$
  - $a_7 = \sigma^2\lambda_2\lambda_3$
  - $a_8 = \frac{1}{2}\sigma^2\lambda_3^2$

  Con $\gamma = -(\lambda_1 + \lambda_2 + \lambda_3)$

  Observación 2.3. *Si el parámetro* λ_3 *se fija igual a cero en la función generadora de probabilidad, los coeficientes* a_6, a_7 *y* a_8 *son todos iguales a cero. en ese caso la forma de la fgp* $\pi(t_1, t_2)$ *se simplifica a sólo los primeros cinco términos.*

- *Consael (1952) ha examinado el modelo de Poisson-normal bivariada como: Bivariada Poisson* $(\lambda_1, \lambda_2, \lambda_3 = 0)\underset{\lambda_1,\lambda_2}{\wedge}$ *bivariada normal*$(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

  que produce la función generadora de probabilidad con cinco parámetros

$$\Pi(t_1, t_2) = exp[\mu_1(t_1-1) + \frac{1}{2}\sigma_1^2(t_1-1)^2 + \mu_2(t_2-1) + \frac{1}{2}\sigma_2^2(t_2-1)^2 + \rho\sigma_1\sigma_2(t_1-1)(t_2-1)] \tag{2.14}$$

  Con las restricciones $\mu_i - \sigma_i^2 \geq \sigma_1\sigma_2$, *i=1,2 y* $\rho \geq 0$

2.3. Distribución Marginal

A partir de la función generadora de probabilidad conjunta (2.12) o (2.13), se puede determinar la función generadora de probabilidad marginal de X_1, estableciendo a $t_2 = 1$

$$\Pi_{x_1}(t_1) = exp[\mu(\lambda_1 + \lambda_3)(t-1) + \frac{1}{2}\sigma^2(\lambda_1 + \lambda_3)^2(t_1 - 1)]$$

o

$$\Pi_{x_1}(t_1) = exp[\alpha_1(t-1) + \alpha_2(t^2 - 1)]$$

Donde

- $\alpha_1 = (\lambda_1 + \lambda_3)[\mu - \sigma^2(\lambda_1 + \lambda_3)]$
- $\alpha_2 = \frac{1}{2}\sigma^2(\lambda_1 + \lambda_3)^2$

En forma análoga

$$\Pi_{x_1}(t_1) = exp[\beta_1(t-1) + \beta_2(t^2 - 1)]$$

Donde

- $\beta_1 = (\lambda_2 + \lambda_3)[\mu - \sigma^2(\lambda_2 + \lambda_3)]$
- $\alpha_2 = \frac{1}{2}\sigma^2(\lambda_2 + \lambda_3)^2$

2.4. Función de Probabilidad

Observamos que en el momento de generación de la $exp[\mu t + \frac{\sigma^2 t^2}{2}]$ distribución normal se puede diferenciar r veces con respecto a t produciendo:

$$M^{(r)}(t) = M(t)P_r(t)$$

Cuando $P_r(t)$ es un polinomio de grado r en t. Obtenemos la función de probabilidad

$$f(r,s) = \frac{\lambda_1^r \lambda_2^s}{r!s!} M(\gamma) \sum_{k=0}^{min(r,s)} \binom{r}{k}\binom{s}{k} k! P_{r+s-k}(\gamma)\delta^k \tag{2.15}$$

Donde $\gamma = -(\lambda_1 + \lambda_2 + \lambda_3)$ *y* $\delta = \frac{\lambda_3}{[\lambda_1 \lambda_2]}$.

Observación 2.4. *Si* $\lambda_3 = 0$ *entonces la función de probabilidad se reduce a*

$$f(r,s) = \frac{\lambda_1^r \lambda_2^s}{r!s!} M[-(\lambda_1 + \lambda_2)] P_{r+s}[-(\lambda_1 + \lambda_2)] \tag{2.16}$$

2.5. Relación de Recurrencia

En el caso de la distribución de Hermite bivariada las relaciones de recurrencia adquieren una forma especial. Tenga en cuenta que, en este caso, el coeficiente diferencial de r-ésimo de $M(t)$ satisface la relación de recurrencia [3].

$$M^{(r)}(t) = (r-1)\sigma^2 M^{(r-2)}(t) + (\mu + \sigma^2 t) M^{(r-1)}(t)$$

Al aplicar, tenemos

$$M^{(r)}(\gamma) = (r-1)\sigma^2 M^{(r-2)}(\gamma) + (\mu + \sigma^2 t) M^{(r-1)}(\gamma)$$

Por lo tanto, las relaciones de recurrencia para las probabilidades de $f(r,0)$ y $f(0,s)$ son

$$\left.\begin{array}{ll} r: & f(r,0) = (\lambda_1\sigma)^2 f(r-2,0) + (\mu+\sigma^2\gamma)\lambda_{1f(r-1,0)} \quad r \geq 1 \\ s: & f(0,s) = (\lambda_2\sigma)^2 f(0,s-2) + (\mu+\sigma^2\gamma)\lambda_{2f(0,s-1)} \quad s \geq 1 \end{array}\right\} (2.17)$$

Más general tenemos

$$\left.\begin{array}{ll} f(r,s) = \frac{(r+1)\lambda_2}{s\lambda_1} f(r+1,s-1) + \frac{(r-s+1)\lambda_3}{s\lambda_1} f(r,s-1) & r \geq s \\ f(r,s) = \frac{(s+1)\lambda_1}{r\lambda_2} f(r-1,s+1) + \frac{(s-r+1)\lambda_3}{r\lambda_2} f(r-1,s) & s \geq r \end{array}\right\} \tag{2.18}$$

Observación 2.5. *Se pueden calcular rápidamente las siguientes probabilidades con* $\tau = \mu + \sigma^2\gamma$

- $f(0,0) = m(\gamma)$
- $f(1,0) = \lambda_1 \tau M(\gamma)$
- $f(0,1) = \lambda_2 \tau M(\gamma)$

Kemp y Papageorgiou (1982) han obtenido la siguiente desigualdades para cualquier par de r y s.

$$P\{X_1 \leq r, X_2 \leq s\} \geq P\{X_1 \leq r\}P\{X_2 \leq s\}$$

$$P\{X_1 \geq r, X_2 \geq s\} \geq P\{X_1 \geq r\}P\{X_2 \geq s\}$$

2.6. Momentos y Cumulantes

La función de generatriz de cumulantes factoriales [3] es

$$G(t_1, t_2) = b_{t1} + b_2 t_1^2 + b_3 t_2 + b_4 t_2^2 + b_5 t_1 t_2 + b_6 t_1^2 t_2 + b_7 t_1 t_2^2 + b_8 t_1^2 t_2^2$$

A partir del cual, los cumulantes factoriales $\kappa_{[r,s]}$ se determinan fácilmente como los coeficientes de $\frac{t_1^2 t_2^2}{r!s!}$:

$$\left.\begin{array}{llll} \kappa_{[1,0]} = b_1 & \kappa_{[2,0]} = 2b_2 & \kappa_{[0,1]} = b_3 & \kappa_{[0,2]} = 2b_4 \\ & & & \\ \kappa_{[1,1]} = b_5 & \kappa_{[2,1]} = 2b_6 & \kappa_{[1,2]} = 2b_7 & \kappa_{[2,2]} = 4b_8 \end{array}\right\} \tag{2.19}$$

Todo orden superior de cumulante factoriales son cero.

2.7. Independencia

En el caso de la familia de ocho parámetro de la distribución de Hermite, no parece que las variables aleatorias X_1 y X_2 pueden ser independientes, sin embargo, para la familia de cinco parámetros hay dos casos que podemos considerar. En el primer caso cuando la

Bivariada Poisson $(\lambda_1, \lambda_2, \lambda_3 = 0) \bigwedge_{\lambda_1, \lambda_2}$ *bivariada normal*$(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

La función generadora de probabilidad está dada por (2.14) . Aquí se ve fácilmente que las variables aleatorias son independientes si y sólo si $\rho = 0$; es decir, si y sólo si las variables aleatorias independientes de Poisson se combinan con las distribuciones normales independientes.

Por otro lado, si tenemos en cuenta la familia de cinco parámetros obtenida en (2.12) de que surge cuando la composición es

$$P(\lambda_1, \tau)P(\lambda_2\tau) \underset{\tau}{\wedge} g(\tau)$$

Siendo g $N(\mu, \sigma^2)$, *y la fgp de* (X_1, X_2) *es*

$$\Pi(t_1, t_2) = exp[\mu(\lambda_1(t_1 - 1) + \lambda_2(t_2 - 1)) + \frac{1}{2}\sigma^2\{\lambda_1(t_1 - 1) + \lambda_2(t_2 - 1)\}^2]$$

Observación 2.6. *La condición necesaria y suficiente para que* X_1 *y* X_2 *para ser independientes es que* $\sigma = 0$. *La distribución resultante es la bivariada de Poisson, que puede considerarse un caso especial de la distribución Hermite bivariada.*

2.8. Distribución Condicional y Regresión

Si la función generadora de probabilidad es de la forma en (2.12), usando el teorema (2.1) la función generadora de probabilidad de la distribución condicional de X_1 *dado* $X_2 = x_2$ *es:*

$$\Pi_{x_1}(t_1|x_2) = \frac{M^{(x_2)}[-(\lambda_2 + \lambda_3) + \lambda_1(t-1)]}{M^{(x_2)}[-(\lambda_2 + \lambda_3)]}[\frac{\lambda_2}{\lambda_2 + \lambda_3} + \frac{\lambda_3 t}{\lambda_2 + \lambda_3}]^{x_2} \tag{2.20}$$

Se puede observar que el coeficiente diferencial en el lado de la derecha se puede escribir como

$$M^{(r)}(t) = M(t)\sigma^r H_x^*[\sigma t + \frac{\mu}{\sigma}] \tag{2.21}$$

con

$$H_x^* = \sum_{r=0}^{[\frac{x}{2}]} \frac{x! t^{x-2r}}{(x-2r)! r! 2^r} \tag{2.22}$$

Los polinomios de Hermite modificados definidos por Kemp y Kemp (1965). Sustituyendo en el primer término del lado derecho de (2.20) tenemos:

$$\frac{M[\lambda_1(t-1) - \lambda_2 - \lambda_3]}{M[-\lambda_2 - \lambda_3]} \frac{H_x^*[\sigma\{\lambda_1(t-1) - \lambda_2 - \lambda_3\} + \frac{\mu}{\sigma}]}{H_x^*[\sigma(-\lambda_2 - \lambda_3) + \frac{\mu}{\sigma}} \tag{2.23}$$

El primer término de la ecuación (2.23) se puede expresar como

$$exp[\{\mu - \sigma^2(\lambda_2 + \lambda_3)\}\lambda_1(t-1) + \frac{1}{2}(\sigma\lambda_1)^2(t-1)^2]$$

Y el segundo término es:

$$\sum_{r=0}^{[\frac{x_2}{2}]} \omega_r(x_2)(Q + Pt)^{x_2-2r}$$

Donde

$$P = \frac{\sigma^2\lambda_1}{\mu - \sigma^2(\lambda_2 + \lambda_3)}; Q = 1 - P$$

y

$$\omega_r(x) = \frac{\frac{x!}{(x-2r)!r!}\alpha^{-r}}{\sum_{s=0}^{[\frac{x}{2}]} \frac{x!}{(x-2s)!s!}\alpha^{-s}}$$

Con

$$\alpha = 2(\{\frac{\mu}{\sigma} - \sigma(\lambda_2 + \lambda_3)\})^2$$

En el desarrollo anterior hemos asumido que el parámetro satisface la condición $\frac{\mu}{\sigma} > \sigma(\lambda_2 + \lambda_3)$. *Bajo este supuesto la distribución condicional de* X_1 *dado* $X_2 = x_2$ *es la convolución de*

- $Y_1 \sim B(x_2, p), p = \frac{\lambda_3}{(\lambda_2+\lambda_3)}$
- $Y_2 \sim$ *Hermite obtenida de* $P(\lambda_1\tau) \underset{\tau}{\wedge} N(\mu - \sigma^2(\lambda_2 + \lambda_3), \sigma^2)$
- Y_3 *que es una mezcla de* $B(x_2 - 2r, P), r = 0, 1, 2, ..., [\frac{x_2}{2}]$ *con el componente r-ésimo que tiene la función de peso dado anteriormente.*

Estos resultados son de Kocherlakota (1988) [11]. Kemp y Papageorgiou (1982) [10] consideran la distribución para el caso $\lambda_3 = 0$ *La regresión se puede encontrar como de costumbre de la distribución condicional como*

$$E[X_1|X_2] = x_2p + \lambda_1[\mu - \sigma^2(\lambda_2 + \lambda_3)] + \sum_{r=0}^{[\frac{x_2}{2}]} \omega_r(x_2)(x_2 - 2r)P$$

Lo cual no es lineal en el regresor x_2.

CAPÍTULO 3

ESTUDIO INFERENCIAL

3.1. Introducción

La inferencia, específicamente la toma de decisiones, desempeña un papel importante en la vida de casi todas las personas. Veamos a continuación algunas aplicaciones:

- *Un corredor financiero desea pronosticar el comportamiento del mercado de acciones.*
- *Un metalurgista desea determinar si un nuevo tipo de acero es más resistente a altas temperaturas que el actual.*

Aun cuando usted pueda pensar que su propia capacidad de tomar decisiones es muy buena, la experiencia sugiere que éste puede no ser el caso. Este capítulo trata sobre el problema general de la inferencia estadística, con especial referencia a las distribuciones discretas. Para una exposición detallada sobre problemas de estimación usaremos a Rao (1973) [15] como referencia teórica . El interés aquí es principalmente presentar los resultados sobre estimación. En este sentido, el material que se presenta aquí es una breve revisión de los métodos de estimación puntual.

3.1.1. Estimación

Bajo condiciones regulares, el método de máxima verosimilitud ha demostrado ser superior a los otros métodos. Asintóticamente produce estimadores que se distribuyen en forma normal, sin sesgo y con varianza mínima. Esta técnica, sin embargo, sufre de algunos inconvenientes, importantes, como por ejemplo las dificultad de los cálculos computacionales. Para muchas distribuciones discretas, se necesitan procedimientos iterativos para la resolución de las ecuaciones.

Kemp y Kemp (1988) [16] describen técnicas útiles para distribuciones discretas univariados, que proporciona estimaciones rápidas. A pesar de estos estimadores pueden no poseer las propiedades deseables, tales como la varianza pequeña y sin sesgo, que pueden ser útiles para un vistazo rápido a los datos o de valores de prueba iniciales para los procedimientos iterativos.

El tipo de estimación depende de la distribución particular bajo estudio y la naturaleza de sus parámetros. En algunas situaciones, varios procedimientos pueden ser utilizados con el fin de obtener una solución. Papageorgiou y Kemp (1988) [17] señalan que, para los tamaños de muestra moderados de menos de 500 observaciones, los procedimientos de estimación, a menudo pueden fallar; es decir, el valor estimado del parámetro, puede no estar dentro del espacio de parámetros. En la práctica es típico los conjuntos de datos mucho menores de 500 observaciones; por lo tanto, varios métodos pueden tener que ser utilizado para estimar los parámetros.

Las técnicas que parecen ser más útiles para la distribución de dos variables son el método de máxima verosimilitud y el método de los momentos. Vamos a discutir estos y sus propiedades en las siguientes secciones.

Método de Momentos

El método clásico de momentos consiste en igualar los momentos de la muestra a sus poblaciones equivalentes, expresada en términos de los parámetros. El número de momentos requeridos depende del número de parámetros. la elección de los momentos particulares es arbitraria, pero los momentos de orden inferior son generalmente preferidos ya que son generalmente las funciones más simple de los parámetros. Además, los momentos de orden superior tienen variaciones más grandes .

En una distribución discreta de dos variables, los datos se pueden resumir en la matriz $\{n_{ij}\}$, *la frecuencia observada correspondiente a la observación* $\{X = i, Y = j\}$ *para* $i = 1, 2, ...; j =$

$1, 2, \ldots$. *Donde el tamaño total de la muestra es*

$$\sum_i \sum_j n_{ij} = n$$

La función de probabilidad

$$f(r, s) = P\{X = r, T = s\}$$

Es una función de los parámetros $\theta_1, \theta_2, \ldots \theta_q$.

El momento de la muestra (en torno a cero) de la orden (r, s) se denota por

$$m'_{r,s} = \frac{1}{n} \sum_i \sum_j i^r j^s n_{ij} \tag{3.1}$$

En particular

$$m'_{1,0} = \bar{x} \qquad m'_{0,1} = \bar{y}$$

Los momentos centrales están representados en consecuencia por

$$m_{r,s} = \frac{1}{n} \sum_i \sum_j (i - \bar{x})^r (j - \bar{y})^s n_{ij} \tag{3.2}$$

Al tener $q = 3$ *con* c_1, c_2, c_3 *los momentos de muestra apropiados para estimar* $\theta_1, \theta_2, \ldots \theta_3$ *entonces:*

$$\left.\begin{aligned} c_1 &= u_1(\widetilde{\theta}_1, \widetilde{\theta}_2.\widetilde{\theta}_3) \\ c_2 &= u_2(\widetilde{\theta}_1, \widetilde{\theta}_2.\widetilde{\theta}_3) \\ c_3 &= u_3(\widetilde{\theta}_1, \widetilde{\theta}_2.\widetilde{\theta}_3) \end{aligned}\right\} \tag{3.3}$$

Donde $(\widetilde{\theta}_1, \widetilde{\theta}_2.\widetilde{\theta}_3)$ *son los momentos estimadores de* $(\theta_1, \theta_2, \ldots \theta_3)$.

Mientras que en el caso univariado, puede no haber una ambigüedad en la elección de los momentos c_1, c_2, c_3 *, en el caso de dos variables esta elección no es única. Por lo tanto en el caso* c_1, c_2, c_3 *univariado se toman usualmente los primeros tres momentos de la muestra . En la situación de dos variables* c_1 *y* c_2 *puede ser elegido los dos primeros momentos marginales y* c_3 *se iguala al momento en mixto.*

La solución $(\widetilde{\theta}_1, \widetilde{\theta}_2.\widetilde{\theta}_3)$ *de 3.3 se puede obtener directamente o de forma iterativa. Una de estas técnicas es el método de Newton-Raphson presentado a continuación:*

El lado derecho de (3.3) se expande θ^0 *en una serie de Taylor. Donde*

$$\underline{c} = \underline{u}(\underline{\theta^0}) + U\underline{\delta}$$

. Con

$$U = (\frac{\partial u_i}{\partial \theta_1}, \frac{\partial u_i}{\partial \theta_2}, \frac{\partial u_i}{\partial \theta_3})|_{\theta=\theta^0} \quad i = 1, 2, 3.$$

y

$$\underline{\delta} = \underline{\widetilde{\theta}} - \underline{\theta^0}$$

La solución de $\underline{\delta}$*, tenemos:*

$$\underline{\delta} = U^{-1}[\underline{c} - \underline{u}(\underline{\theta^0})]. \tag{3.4}$$

Esto produce $\underline{\widetilde{\theta}} = \underline{\theta^0} + \underline{\delta}$*. el procedimiento se termina cuando* $\underline{\delta}$ *es pequeño.*

<u>*Momentos de los estimadores de momentos*</u>

Es posible determinar los momentos de los estimadores $\underline{\widetilde{\theta}}$ *mediante la expansión de (3,3) en una serie de Taylor alrededor de los valores esperados de los estimadores. Dado* $E(\underline{\widetilde{\theta}}) = \underline{\mu}$ *a continuación:*

$$\underline{c} = \underline{u}(\underline{\mu}) + V(\underline{\widetilde{\theta}} - \underline{\mu}). \tag{3.5}$$

Dónde

$$V = \left\{ \frac{\partial u_i}{\partial \theta_j} \right\} \Bigg|_{\underline{\theta}=\underline{\mu}}$$

Donde $\underline{\mu}$ *es el valor esperado de* $\underline{\widetilde{\theta}}$*, se ve fácilmente que la matriz de varianza de* $\underline{\widetilde{\theta}}$ *es:*

$$\sum_{MM} = T \sum_{\underline{c}} T' \tag{3.6}$$

Donde $T = V^{-1}$ *y* $\sum_{\underline{c}}$ *es la matriz de varianza de los momentos* $\underline{c}$*.*

Error estándar de los momentos de la muestra

El procedimiento de estimación con el método de los momentos en una distribución discreta de dos variables puede implicar los momentos factoriales de muestra $m_{[r,s]}$ *además de los momentos* $m'_{r,s}$ *y* $m_{r,s}$*. Con el fin de encontrar* $\sum_{\underline{c}}$ *en (3,6), Es necesario determinar las variaciones de los momentos de la muestra. Aquí los resultados asintóticos en los primeros dos momentos de* $m'_{r,s}$ *se derivan. A partir de estos resultados se determina la matriz de covarianza de* $m_{r,s}$*. Para los momentos factoriales sólo se examinan los momentos de orden inferior. Dado* n_{ij} *la frecuencia observada en la celda* (i, j) *correspondiente al evento* $\{X = i, \quad Y = j\}$ *con probabilidad* p_{ij} *para* $i = 0, 1, 2, ...; j = 0, 1, 2, ...$ *A continuación, los totales marginales y las probabilidades correspondientes están dadas por*

$$n_{i+} = \sum_j n_{ij} \qquad n_{+j} = \sum_i n_{ij} \qquad p_{i+} = \sum_j p_{ij} \qquad p_{+j} = \sum_j p_{ij} \tag{3.7}$$

Se conoce que los $\{n_{ij}\}$ *j tienen una distribución multimodalidad con los parámetros* n *y* $\{p_{ij}\}$*.Así :*

$$E(n_{ij}) = \sum_j n_{ij}; \qquad E(n_{ij}^2) = np_{ij}[1 + (n-1)p_{ij}] \qquad E(n_{ij}, n_{kh}) = n(n-1)p_{ij}p_{kh}$$

y

$$Var(n_{ij}) = np_{ij}(1 - p_{ij}); \qquad Cov(n_{ij}, n_{kh}) = -np_{ij}p_{kh}$$

Para $i \neq j$ *y* $k \neq h$.

Recordando la definición de $m'_{r,s}$ *tenemos:*

$$E(m'_{r,s}) = \mu' r, s \tag{3.8}$$

Dónde

$$\mu' r, s = \sum_i \sum_j i^r j^s p_{ij}$$

Además

$$E[(m'_{r,s})^2]) = E[\frac{1}{n^2}\{\sum_i \sum_j i^r j^s n_{ij}\}^2]$$

$$= E[\frac{1}{n^2}\{\sum_i \sum_j i^{2r} j^{2s} n_{ij}^2 + \sum_{i_1} \sum_{j_1} \sum_{i_2} \sum_{j_2} i_1^r i_2^r j_1^s j_2^s n_{i_1 j_1} n_{i_2 j_2}\}]$$

$$= \frac{1}{n^2}\{\sum_i \sum_j i^{2r} j^{2s}[np_{ij} + n(n-1)p_{ij}^2] + \sum_{i_1} \sum_{j_1} \sum_{i_2} \sum_{j_2} i_1^r i_2^r j_1^s j_2^s [n(n-1)p_{i_1 j_1} p_{i_2 j_2}]\}$$

Por lo tanto

$$E[(m'_{r,s})^2]) = \frac{1}{n}\mu'_{2r,2s} + \frac{n-1}{n}\mu'^2_{r,s}. \tag{3.9}$$

En forma análoga

$$E[m'_{r_1,s_1} m'_{r_2,s_2}] = E\{\frac{1}{n^2}[\sum_i \sum_j i^{r_1} j^{s_1} n_{ij}][\sum_i \sum_j i^{r_2} j^{s_2} n_{ij}]\}$$

$$= E\{\frac{1}{n^2}\{\sum_i \sum_j i^{r_1+r_2} j^{s_1+s_2} n_{ij}^2 + \frac{1}{n^2} \sum_{i_1} \sum_{j_1} \sum_{i_2} \sum_{j_2} i_1^{r_1} j_1^{s_1} i_2^{r_2} j_2^{r_2} n_{i_1 j_1} n_{i_2 j_2}\}$$

Cuyos rendimientos

$$E[m'_{r_1,s_1} m'_{r_2,s_2}] = \frac{1}{n^2}\{\sum_i \sum_j i^{r_1+r_2} j^{s_1+s_2} p_{ij} +$$

$$\frac{n-1}{n}[\sum_i \sum_j i^{r_1+r_2} j^{s_1+s_2} p_{ij}^2 + \sum_{i_1} \sum_{j_1} \sum_{i_2} \sum_{j_2} i_1^{r_1} j_1^{s_1} i_2^{r_2} j_2^{r_2} p_{i_1 j_1} p_{i_2 j_2}]$$

$$= \frac{1}{n}\mu'_{r_1+r_2,s_1+s_2} + \frac{n-1}{n}\mu'_{r_1,s_1}\mu_{r_2,s_2} \tag{3.10}$$

Combinando (3.9) y (3.10) obtenemos la matriz varianza de $(m'_{r_1,s_1}, m'_{r_2,s_2})$ *dada por*

$$\frac{1}{n}\begin{bmatrix} \mu'_{2r_1,2s_1} - \mu'^2_{r_1.s_1} & \mu'_{r_1+r_2,s_1+s_2} - \mu'_{r_1,s_1}\mu'_{r_2,s_2} \\ \\ \dots & \mu'_{2r_1,2s_1} - \mu'^2_{r_1.s_1} \end{bmatrix} \tag{3.11}$$

Para $(m'_{r_1,s_1}, m'_{r_2,s_2}, m'_{r_3,s_3})$ *la matriz de 3x3 es la extensión de (3.11)*

<u>Método de even-point</u>

Como se ha visto, en el método de los momentos de las distribuciones de dos variables, las estimaciones para dos de los parámetros se puede obtener igualando los momentos marginales con sus análogos observados. Si hay más de dos parámetros a estimar puede ser difícl escoger que momentos utilizar. Puede surgir otro tipo de problema. Al igual que en el caso univariado, los estimadores de momentos pueden no cumplir las restricciones del espacio de los parámetros. Por ejemplo, en la distribución binomial negativa univariado, si s^2 *no es mayor que la media, vamos a tener una estimación negativa para un parámetro positivo. Este tipo de problema es más común en la distribucion de dos variables, que tienen más parámetros, que representa no sólo sus medias y varianzas marginales, sino también una estructura de correlación.*

El método even-point de estimación proporciona una modificación para el método de momentos. En el caso univariado esta técnica se introdujo por primera vez por Patel (1976) [18] para la estimación de los parámetros de la distribución de Hermite univariado. Como en el método de los momentos, la primera ecuación de estimación se obtiene igualando $\bar{x}$ *a* μ *, que es una función de los parámetros desconocidos. La segunda ecuación se obtiene igualando la suma de las frecuencias relativas de los valores pares de las variables aleatorias* X *en la muestra a la suma de las probabilidades correspondientes*

$$\frac{1}{n}\sum_x n_{2x} = \sum_x f(2x;\underline{\theta})$$

Donde $f(x;\underline{\theta})$ *es la función de probabilidad de la variable aleatoria* X*. Si la fgp de* X *se denota por* $\Pi_x(t;\underline{\theta})$*. Donde Kemp y Kemp (1988) muestran que*

$$\Pi_x(0;\underline{\theta}) + \Pi_x(-1;\underline{\theta}) = \sum_x f(x;\underline{\theta})[1 + (-1)^x]$$

$$= 2\sum_x f(2x;\underline{\theta})$$

es decir, el método event-point se puede considerar como un caso especial de estimación con el fgp empírica

En el caso de dos variables $\bar{x} = \mu_x$ *y* $\bar{y} = \mu_y$ *de las dos primeras ecuaciones igualando la estimación. Para las ecuaciones de la tercera la función generadora de probabilidad conjunta* $\Pi_{X,Y}(t_1, t_2; \underline{\theta})$ *evaluada en* $(t_1, t_2) = (1,1)$ *y* $(-1,-1)$ *tenemos:*

$$\Pi_{X,Y}(1,1;\underline{\theta}) + \Pi_{X,Y}(-1,-1);\underline{\theta}) = 2\{\sum_x \sum_y f(2x, 2y; \underline{\theta}) + \sum_x \sum_y f(2x+1, 2y+1; \underline{\theta})\}.$$

Las probabilidades en el lado derecho son entonces reemplazados por las frecuencias relativas correspondientes. Si otros parámetros han de calcularse $\Pi_{X,Y}(1,1;\underline{\theta})$ *y* $\Pi_{X,Y}(-1,-1;\underline{\theta})$ *también serán necesarios.*

Papageorgiou y loukas (1988) [19] han usado una variación de esta técnica en la que incluso los estimadores puntuales se basan en la distribución condicional de X *dado* $Y = 0$

<u>Método de Maxima Versosimilitud</u>

La función de verosimilitud se denota por $L(\underline{\theta}; \underline{x}$ *con* $\underline{X}$ *que representan el vector de parámetros desconocidos y* $\underline{x}$ *de las observaciones. El principio de máxima verosimilitud [15] consiste en seleccionar para* $\underline{\theta}$ *las estimaciones obtenidas mediante la maximizacián de L con respecto a estos parámetros. Así*

$$\frac{\partial logL}{\partial \theta_i} = 0 \qquad i = 1, 2, 3, ..., q \tag{3.12}$$

puede resolverse para $\theta_1, \theta_2, ..., \theta_q$ *en términos de las observaciones* $\underline{x}$*. Las soluciones se indican mediante* $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_q$*Desafortunadamente, no siempre es posible resolver (3.12) para* $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_q$ *directamente. Hay dos métodos iterativos posibles para resolver estas ecuaciones:*

- *Método de Newton-Raphson: Las ecuacion (3.12) se pueden expandir en una serie Taylor en torno a una solución como* $\underline{\theta}^0$

$$\frac{\partial logL}{\partial \theta_i} = \left.\frac{\partial logL}{\partial \theta_i}\right|_{\underline{\theta}=\underline{\theta}^0}$$

$$+\{(\theta_1 - \theta_1^0)\frac{\partial}{\partial \theta_1} + \theta_2 - \theta_2^0)\frac{\partial}{\partial \theta_2} + ... + \theta_q - \theta_q^0)\frac{\partial}{\partial \theta_q}\}\left.\frac{\partial L}{\partial \theta_i}\right|_{\underline{\theta}=\underline{\theta}^0} \tag{3.13}$$

Si $\underline{\theta}$ son las soluciones de (3.13), entonces el lado izquierdo de (3.12) será cero; Por lo tanto, en cada una de las iteraciones de la nueva $\underline{\theta}_{j+1}$ el valor se puede encontrar a partir de $\underline{\theta}_j$ mediante la adicián de $\underline{\delta}$ a ella. El vector $\underline{\delta}$ está dado por:

$$\underline{\delta} = -A^{-1}\underline{D} \tag{3.14}$$

con A^{-1}, Y $\underline{D}$ evaluado en $\underline{\theta} = \underline{\theta}_j$ y donde :

$$A = \begin{bmatrix} \frac{\partial^2 logL}{\partial^2 \theta_1} & \frac{\partial^2 logL}{\partial \theta_1 \theta_2} & ... & \frac{\partial^2 logL}{\partial^2 \theta_1 \partial^2 \theta_q} \\ ... & \frac{\partial^2 logL}{\partial^2 \theta_2} & ... & \frac{\partial^2 logL}{\partial \theta_2 \theta_q} \\ ... & ... & ... & \frac{\partial^2 logL}{\partial^2 q} \end{bmatrix}, \qquad \underline{D} = \begin{bmatrix} \frac{\partial^2 logL}{\partial \theta_1} \\ \frac{\partial^2 logL}{\partial \theta_2} \\ ... \\ \frac{\partial^2 logL}{\partial \theta_q} \end{bmatrix} \tag{3.15}$$

Este proceso continua hasta que en $\underline{\delta}$ se hace pequeño.

- *Modo de marcar: Rao (1973) [15] define la cantidad $\frac{\partial logL}{\partial \theta_i}$ como la puntuación eficiente para $\underline{\theta}_i$. Así, el estimador de máxima verosimilitud es el valor de $\underline{\theta}_i$ para los cuales la puntuación eficiente se convierte en cero. Como lo anterior, lo podemos ampliar en torno a una solución de prueba $\underline{\theta}^0$ y predispuestas para el valor de $\underline{\delta}$ dada en (3.14) . Rao sugiere, sin embargo . que con el fin de estabilizar el valor $\underline{\delta}$, que es mejor reemplazar la matriz A por su valor esperado evaluado en $\underline{\theta}^0$. Es evidente que los valores esperados de $-A$ es la matriz de información en $\underline{\theta} = \underline{\theta}^0$.*

$$\Gamma(\underline{\theta}^0) = \left.\left\{E\left[-\frac{\partial^2 logL}{\partial_i \partial_j}\right]\right\}\right|_{\underline{\theta}=\underline{\theta}^0} \tag{3.16}$$

Donde

$$\underline{\delta} = \Gamma^{-1}(\underline{\theta}^0)\underline{D}$$

En esta técnica, el nuevo valor de $\underline{\theta}_{j+1}$ se encuentra de nuevo mediante la adición de $\underline{\delta}$ con el valor $\underline{\theta}_j$ obtenido en la iteración j-ésima.

Para una muestra grande los valores de $\underline{\theta}$ obtenido por las dos técnicas estarán muy cerca. el método de puntuación, sin embargo, tiende a converger más rápidamente.

Propiedades de los estimadores de máxima verosimilitud

La distribución exacta de los estimadores de máxima verosimilitud no está disponible en general. Sin embargo, Si el tamaño de la muestra n *es grande, se puede demostrar que el estimador* $\hat{\underline{\theta}}$ *es consistente, asintóticamente normal y eficiente para* $\underline{\theta}$*. La matriz de varianza asintótica de* $\hat{\underline{\theta}}$ *está dada por* $\Gamma^{-1}(\underline{\theta})$*.*

Los datos agrupados

En el caso especial en que los datos se clasifican en una tabla de frecuencias en ambos sentidos, como suele ser la situación de una distribucion discreta de dos variables. Sea $\cdot n_{rs}$ *la frecuencia observada en* (s, r) *de la celda para* $r = 0, 1, 2, ..., k_1$*;* $s = 0, 1, 2, ..., k_2$ *con:*

$$\sum_{r}^{k_1}\sum_{s}^{k_2} n_{r,s} = n$$

A continuación, la probabilidad se puede escribir como

$$L \propto \Pi_{r,s} f(r, s)^{n_{rs}}$$

o

$$logL = C + \sum_{r}^{k_1}\sum_{s}^{k_2} n_{r,s} log f(r, s)$$

Predispuestas para los estimadores de máxima verosimilitud, necesitamos las dos primeras derivadas del logaritmo de la probabilidad con respecto a $\underline{\theta}_i$*:*

$$\frac{\partial^2 logL}{\partial\theta_i} = \sum_{r}^{k_1}\sum_{s}^{k_2} n_{r,s}\frac{1}{f(r,s)}\frac{\partial f(r,s)}{\partial\theta_i}$$

$$\frac{\partial^2 logL}{\partial\theta_i\partial\theta_j} = \sum_{r}^{k_1}\sum_{s}^{k_2} n_{r,s}\left[\frac{1}{f(r,s)}\frac{\partial^2 f(r,s)}{\partial\theta_i\partial\theta_j} - \frac{1}{f^2(r,s)}\frac{\partial f(r,s)}{\partial_i}\frac{\partial f(r,s)}{\partial\theta_j}\right]$$

Donde n_{rs} *son variables aleatorias que tienen una distribución multinomial con probabilidades prescritas por la funcion de probabilidad* $f(r, s)$*,* $E(n_{rs}) = nf(r, s)$ *; por lo tanto*

$$E\{\frac{\partial^2 logL}{\partial\theta_i\partial\theta_j}\} = n\sum_r^{k_1}\sum_s^{k_2}\left[\frac{\partial^2 f(r,s)}{\partial\theta_i\partial\theta_j} - \frac{1}{f(r,s)}\frac{\partial f(r,s)}{\partial_i}\frac{\partial f(r,s)}{\partial\theta_j}\right]$$

$$= n\sum_r^{k_1}\sum_s^{k_2} f(r,s)\frac{\partial logf(r,s)}{\partial_i}\frac{\partial logf(r,s)}{\partial\theta_j}$$

o

$$\Gamma(\underline{\theta}^0) = \left\{ n\sum_r^{k_1}\sum_s^{k_2} f(r,s)\frac{\partial f(r,s)}{\partial_i}\frac{\partial f(r,s)}{\partial\theta_j} \right\}\Bigg|_{\underline{\theta}=\underline{\theta}^0} \tag{3.17}$$

Por lo tanto, utilizando el método de puntuación

$$\underline{\delta} = \Gamma^{-1}(\underline{\theta}^0)\underline{D}^*$$

donde

$$\underline{D}^* \begin{bmatrix} \sum_r^{k_1}\sum_s^{k_2}\frac{1}{f(r,s)}\frac{\partial f(r,s)}{\partial\theta_1} \\ \sum_r^{k_1}\sum_s^{k_2}\frac{1}{f(r,s)}\frac{\partial f(r,s)}{\partial\theta_2} \\ \cdots \\ \sum_r^{k_1}\sum_s^{k_2}\frac{1}{f(r,s)}\frac{\partial f(r,s)}{\partial\theta_q} \end{bmatrix}$$

Con $\underline{\theta} = \underline{\theta}^0$ *. La matriz de varianza asintótica del estimador de máxima verosimilitud* $\hat{\underline{\theta}}$ *es* $\Gamma^{-1}(\underline{\theta})$ *y se estima mediante la evaluación de* Γ^{-1} *a* $\hat{\underline{\theta}}$*.*

3.2. Estimación Puntual

Una estimación puntual del valor de un parámetro poblacional desconocido , es un número que se utiliza para aproximar el verdadero valor de dicho parámetro poblacional. A fin de realizar tal estimación, se toma una muestra de la población y calcularemos el parómetro muestral asociado. El valor de este parámetro muestral será la estimación puntual del parámetro poblacional.

3.2.1. Método de los momentos

Considere la familia de cinco parámetro de la distribución dada por las función generadora de probabilidad en las ecuaciones (2.12) y (2.13) con $a_i = b_i = 0$ *para* $i = 6, 7, 8$ *éstos producen los momentos en los términos de las* b_i *como se da (2.19). Resolviendo los parámetros en términos de los momentos de la muestra que tenemos:*

$$\left.\begin{array}{l} \tilde{b}_1 = \bar{x}_1 \\ \tilde{b}_2 = \frac{[m_{2,0} - \bar{x}_1]}{2} \\ \tilde{b}_3 = \bar{x}_2 \\ \tilde{b}_4 = \frac{[m_{0,2} - \bar{x}_2]}{2} \\ \tilde{b}_5 = m_{1,1} \end{array}\right\} \tag{3.18}$$

Reparametrización de las a_i*, tenemos las ecuaciones*

$$\begin{array}{l} \tilde{a}_1 + 2_2 +_5 = \bar{x}_1 \\ \tilde{a}_1 + 4_2 +_5 = m_{2,0} \\ \tilde{a}_3 + 2_4 +_5 = \bar{x}_2 \\ \tilde{a}_3 + 4_4 +_5 = m_{0,2} \\ \tilde{a}_5 = m_{1,1} \end{array}$$

Despejando obtenemos

$$\left.\begin{array}{l} \tilde{a}_1 = 2\bar{x}_1 - m_{2,0} - m_{1,1} \\ \tilde{a}_2 = \frac{[m_{2,0} - \bar{x}_1]}{2} \\ \tilde{a}_3 = 2\bar{x}_2 - m_{0,2} - m_{1,1} \\ \tilde{a}_4 = \frac{[m_{0,2} - \bar{x}_2]}{2} \\ \tilde{a}_5 = m_{1,1} \end{array}\right\} \tag{3.19}$$

3.2.2. Método de Event-Point

H. Papageorgiou, en su Ph.D. tesis 1977 Bradford, útilizo el método de estimación event-point para una versión de dos variables e ilustrado para una variedad de distribuciones bivariadas-Poisson generalizados. H. Papageorgiou, C.D. Kemp. y S. Loukas [4] presentan el caso particular de la distribución de Hermite bivariada, la evaluación de (2.12) a $(1, 1)$, $(-1, 1)$, $(1, -1)$ *y* $(-1, -1)$ *da las relaciones*

$$\Pi(1,1)+\Pi(-1,1)=1+exp\{-2(a_1+a_5)\}=2(\sum p_{2x,2y}+\sum p_{2x,2y+1})$$

$$\Pi(1,1)+\Pi(1,-1)=1+exp\{-2(a_3+a_5)\}=2(\sum p_{2x,2y}+\sum p_{2x+1,2y})$$

$$\Pi(1,1)+\Pi(-1,-1)=1+exp\{-2(a_1+a_3)\}=2(\sum p_{2x,2y}+\sum p_{2x+1,2y+1})$$

Para una muestra de N observaciones , denotamos por A, B, C y D las sumas de las frecuencias observadas en los puntos $(2x,2y)$, $(2x,2y+l)$, $(2x+1,2y)$ y $(2x+1,2y+1)$, respectivamente, y a $\bar{x}$ y $\bar{y}$ los medios marginales. Entonces, lo que equivale a las sumas teóricas de frecuencias observadas, y utilizando las ecuaciones de momentos $\bar{x}=\hat{a}_1+2\hat{a}_2+\hat{a}_5$ y $\bar{y}=\hat{a}_3+2\hat{a}_4+\hat{a}_5$, obtenemos los estimadores de punto de equilibro.

$$\hat{a}_1=\frac{1}{4}[log\{2N^{-1}(A+C)-1\}-log\{2N^{-1}(A+D)-1\}-log\{2N^{-1}(A+B)-1\}]$$

$$\hat{a}_2=\frac{1}{4}[log\{2N^{-1}(A+B)-1\}-2\bar{x}]$$

$$\hat{a}_3=\frac{1}{4}[log\{2N^{-1}(A+B)-1\}-log\{2N^{-1}(A+D)-1\}-log\{2N^{-1}(A+C)-1\}]$$

$$\hat{a}_4=\frac{1}{4}[log\{2N^{-1}(A+C)-1\}-2\bar{y}]$$

$$\hat{a}_5=\frac{1}{4}[log\{2N^{-1}(A+D)-1\}-log\{2N^{-1}(A+B)-1\}-log\{2N^{-1}(A+C)-1\}]$$

Para obtener soluciones finitos y satisfacer el requisito de que $\hat{a}_i>0$, deben aplicar diversas restricciones a A, B, C y D, por ejemplo $A+B$, $A+C$ y $A+D$ todo debe ser superior a $\frac{N}{2}$. Considere una sola observación aleatoria (x,y) de la distribución. Entonces;

$$
\begin{array}{lll}
A = 1 & con\ probabilidad & P_1 = \sum p_{2x,2y} \\
B = 1 & con\ probabilidad & P_1 =_{2x,2y+1} \\
C = 1 & con\ probabilidad & P_1 =_{2x+1,2y} \\
D = 1 & con\ probabilidad & P_1 =_{2x+1,2y+1}
\end{array}
$$

La función generadora de momentos conjunta de A, B, C, D, x, y para la muestra de uno es

$$
\begin{aligned}
M(t_1, ..., t_6) &= E\{exp(At_1 + Bt_2 + Ct_3 + Dt_4 + xt_5 + yt_6)\} \\
&= E\{exp(t_1 + xt_5 + yt_6)|A = 1\}P_1 + E\{exp(t_2 + xt_5 + yt_6)|B = 1\}P_2 \\
&+E\{exp(t_3 + xt_5 + yt_6)|C = 1\}P_3 + E\{exp(t_4 + xt_5 + yt_6)|D = 1\}P_4 \\
&= e^{t_1} \sum p_{2x,2y} exp(2xt_5 + 2yt_6) + e^{t_2} \sum p_{2x,2y+1} exp[2xt_5 + (2y + 1)t_6] \\
&+e^{t_3} \sum p_{2x+1,2y} exp[(2x + 1)t_5 + 2yt_6] \\
&+e^{t_4} \sum p_{2x+1,2y+1} exp[(2x + 1)t_5 + (2y + 1)t_6] \\
&= \tfrac{1}{4}e^{t_1}(\Pi_1 + \Pi_2 + \Pi_3 + \Pi_4) + \tfrac{1}{4}e^{t_2}(\Pi_1 + \Pi_2 - \Pi_3 - \Pi_4) \\
&+\tfrac{1}{4}e^{t_3}(\Pi_1 - \Pi_2 + \Pi_3 - \Pi_4) + \tfrac{1}{4}e^{t_4}(\Pi_1 - \Pi_2 - \Pi_3 - \Pi_4)
\end{aligned}
$$

Donde

$$
\begin{array}{ll}
\Pi_1 = \Pi(e^{t_5}, e^{t_6}) & \Pi_2 = \Pi(-e^{t_5}, e^{t_6}) \\
\Pi_3 = \Pi(e^{t_5}, -e^{t_6}) & \Pi_4 = \Pi(-e^{t_5}, -e^{t_6})
\end{array}
$$

Para N observaciones independientes (x, y), *la función generadora de momentos conjunta A, B, C, D,* $\sum x$, $\sum y$ *es* $[M(t_1, ..., t_6)]^N$. *Para la matriz de covarianza de A, B, C, D,* $\sum x$, $\sum y$ *se obtiene facilmente la diferenciación, y por lo tanto, se obtienen* $\hat{p}_1$, $\hat{p}_2$, $\hat{p}_3$, $\hat{p}_4$, $\bar{x}$, $\bar{y}$ *donde* $N\hat{p}_1 = A$, *etc. Por ejemplo*

$$
cov(\hat{p}_1, \bar{x}) = -(2N)^{-1}[a_1 exp\{-2)a_1 + a_3)\} + a_5 exp\{-2(a_3 + a_5)\}
$$

$$
(a_1 + a_5) exp\{-2)a_1 + a_5)\}]
$$

La matriz de covarianza asintótica V de los estimadores event-point viene dada por $V = FQF'$ *(Kemp, 1967), donde Q es la matriz de covarianza de* $\hat{p}_1$, $\hat{p}_2$, $\hat{p}_3$, $\hat{p}_4$, $\bar{x}$, $\bar{y}$ *y F es la matriz jacobiana de la transformación en* $\hat{a}_i = f_i(\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4, \bar{x}, \bar{y})$. *Eficiencia asintótica se define como* $E = |I^{-1}|/|V|$, *donde I es la matriz de información, fórmulas de cuyos elementos están dados por Kemp y Papageorgiou (1982).*

Los cálculos de eficiencia asintótica se han llevado a cabo para una amplia selección de valores de los parámetros y resumir estos resultados mediante la búsqueda de regiones eficientes del

Figura 3.1: Algunos valores de los parámetros para que $E > 0,7$ para el método even-point

a_1	a_3	a_5	a_2	a_4	a_1	a_3	a_5	a_2	a_4
<0·2	<0·2	0·1	0·2	0·2	<0·6	<0·6	0·1	2	2
<0·2	<0·2	0·4	2	2	<0·6	<0·6	0·3	4	4

espacio de parámetros, donde $E > 0{,}7$ *Algunos ejemplos se dan en la Tabla. Parece que el método es eficiente sólo para valores pequeños de* a_1*,* a_3 *y* a_5 *. De hecho, para valores muy pequeños de* a_1*,* a_3*,* a_5 *y relativamente grandes valores de* a_2 *y* a_4 *estaba cerca de la unidad. conclusiones análogas fueron dibujadas por Patel (1976) para la distribución de Hermite univariado*

3.2.3. Método de Maxima Verosimilitud

Los estimadores de máxima verosimilitud se pueden desarrollar para las diferentes parametrizaciones. Donde n_{rs} *representan la frecuencia observada en la celda* (r, s) *y* $f(r, s)$ *la función de probabilidad correspondiente. Utilizando las definiciones de la introducción, podemos encontrar*

$$\frac{\partial logL}{\partial \theta_i} = \sum_{r,s} \frac{n_{rs}}{f(r,s)} \frac{\partial logf(r,s)}{\partial \theta_i}$$

Donde θ_i *representa los parámetros* a_i *para* $i = 1, 2, 3, 4, 5$.

Kemp y Papageorgiou (1982) [10] dan relaciones de recurrencia entre las derivadas parciales de $f(r, s)$ *con respecto a los distintos parámetros. A partir de la función generadora de probabilidad (2.12) con los cinco parámetros, tenemos.*

$$\frac{\partial logf(r,s)}{\partial a_1} = f(r-1, s) - f(r, s)$$

$$\frac{\partial logf(r,s)}{\partial a_2} = f(r-2, s) - f(r, s)$$

$$\frac{\partial logf(r,s)}{\partial a_3} = f(r, s-1) - f(r, s)$$

$$\frac{\partial logf(r,s)}{\partial a_4} = f(r, s-2) - f(r, s)$$

$$\frac{\partial logf(r,s)}{\partial a_5} = f(r-1, s-1) - f(r, s)$$

Por lo tanto

$$\left.\begin{array}{l}\frac{\partial logL}{\partial a_1} = \sum_{r,s}[\frac{f(r-1,s)}{f(r,s)} - 1] \\ \frac{\partial logL}{\partial a_2} = \sum_{r,s}[\frac{f(r-2,s)}{f(r,s)} - 1] \\ \frac{\partial logL}{\partial a_3} = \sum_{r,s}[\frac{f(r,s-1)}{f(r,s)} - 1] \\ \frac{\partial logL}{\partial a_4} = \sum_{r,s}[\frac{f(r,s-2)}{f(r,s)} - 1] \\ \frac{\partial logL}{\partial a_5} = \sum_{r,s}[\frac{f(r-1,s-1)}{f(r,s)} - 1]\end{array}\right\} \quad (3.20)$$

El elemento típico de la matriz de información está dada por

$$E[-\frac{\partial^2 logL}{\partial a_1^2}] = n\sum_{r,s} f(r,s)[\frac{f(r-1,s)}{f(r,s)} - 1]^2$$

estas ecuaciones son útiles para una solución iterativa de las ecuaciones de máxima probabilidad. las ecuaciones para el conjunto de parámetros en b_i *(o en* $\mu, \sigma^2, \lambda_1, \lambda_2 y \lambda_3$*) pueden ser escritos a lo largo de las mismas líneas*

CONCLUSIONES

El objetivo fundamental de esta tesis era la recopilación de los avances teóricos la distribución de Hermite Bivariada. De los cuales sus mayores investigadores son Kemp,C.D., Kemp,A.W., H. Papageorgiou, y S. Loukas quienes han trabajado en el area probabilistica:

- *Función Generadora de Probabilidad*
- *Función Generadora de Momentos*
- *Función Generadora de cumulantes*
- *Distribución Marginal*
- *Función de probabilidad*
- *Relación de Recurrencia*
- *Distribución Condicional*

Mientras que en el estudio inferencial se ha desarrollado su investigación hasta los métodos de la estimación puntual en los cuales se desarrollaron:

- *Método de los momentos*
- *Método Even-Point*
- *Método de Máxima Versosimilitud*

Támbien se han desarrollado modelos computacional para generar observaciones que modelen la distribución.

El ultimo avanza de la distribución de Hermite fue hecho por David Moriña, Manuel Higueras, Pedro Puig y María Oliveira quienes desarrollaron el R Package hermite, herramienta computacional que permite calcular probabilidades, generar aleatoriamente una muestra de datos , y realizar estimaciones, a través del método de máxima verosimilitud.

Este estudio nos revela las áreas de investigación que aun no se desarrollan, y quedan como propuesta de investigacińn

- *Estimación por intervalos*
- *Pruebas de Hipotesis*
- *Test de bondad de ajuste*

BIBLIOGRAFÍA

[1] *Kemp,C.D., and Kemp,A.W. (1965), Some Properties of the Hermite Distribution, Ed Biometrika.*

[2] *Johnson, N. L., Kotz, S. y Balakrishnan (1997) Discrete Multivariate Distributions. Ed. John Wiley Sons, I New York.*

[3] *Kocherlakota, S. y Kocherlakota, K. (1992)Bivariate Discrete Distributions.Ed. John Wiley Sons, I New York. 9*

[4] *H. Papageorgiou, C.D. Kemp. y S. Loukas. (1983)Some methods of estimation for the bivariate Hermite distribution. Ed. Biometrika.*

[5] *Uría Mundo, E., Puig i Casado, P. (2013), Ïntroducción de la ley de Hermite en la Wikipedia".*

[6] *D. Moriña, M. Higueras, P. Puig y M. Oliveira (2015), Generalizerd Hermite Distribution Modelling with the R Package hermite. The R Journal Vol. 7/2. pag. 263-274.*

[7] *Cresswell, W.L. Froggatt, P. (1963). "The Causation of Bus Driver Accidents". Ed. Oxford University Press.*

[8] *Kemp, A.W. Kemp, C.I. (1966). .ᴬn alternative derivation of the Hermite distribution. Ed. Biometrika 53, 627-8.*

[9] *D. A. Lind, R.D. Mason, W. G. Marchal (2001): ?Estadística para Administración y Economía?. Ed. Irwin McGraw-Hill.*

[10] *Kemp, C. D. y Papageorgiou, H. (1982). "bivariate Hermite distributions". Ed. Sankhya, A, 44, 269-280.*

[11] *Kocherlakota, S. (1988). On the compounded bivariate Poisson distribution: a unified approach".Ed. Annals of the institute of Statistical Mathematics, 40, 61-76.*

[12] A.G.McKendrick (1926),?Applications of Mathematics to Medical Problems?, Ed. Proceedings of the Edinburgh Mathematical Society, 44, 98-130.

[13] Gupta, R.P., and Jain, G.C. (1974), ?A Generalized Hermite Distribution and Its Properties?,Ed. SIAM Journal on Applied Mathematics, 27, 359-363.

[14] Hogg, R. V. y Craig, A.T. (1978).Ïntroduction to Mathematical Statistics (Fourth edition)" Ed. Macmillan Publishing Co., Inc, New York.

[15] Rao, C. R. (1973). "Linear Statistical Inference and its Applications (second edition)" Ed. John Wiley and Sons, New York.

[16] Kemp, C. D. y Kemp A. W. (1988). Rapid estimation for discrete distributions". Ed. The Statistician, 37, 243-255.

[17] Papageorgiou, H. y Kemp, C. D. (1988) .A method of estimation for some bivariate discrete distributions" Ed. Biometrical Journal, 30,993-1001.

*[18] Y.C.Patel (1976), [**?**], Ed. Biometrics, 32, 865-873.*

[19] Papageorgiou, H y Loukas, S. (1988) Çonditional even point estimation for bivariate discrete distributions" Ed. Communications in Statistics: theory and Methods, 17, 3403-3412.