
Introducción de la ley de Hermite en la Wikipedia

TRABAJO FINAL DE GRADO

Estefany Uría Mundo

UNIVERSIDAD AUTÓNOMA DE BARCELONA
GRADO EN ESTADÍSTICA APLICADA
TUTOR: PERE PUIG
JULIO, 2013

Índice

1. Objetivo	2
2. Introducción	2
3. Historia	4
4. Definición	5
4.1. Función de Probabilidad	5
4.2. Notación	6
4.3. Gráficas	7
5. Propiedades	8
5.1. Cumulantes y momentos	8
5.2. Coeficiente de simetría	9
5.3. Curtosis	9
5.4. Función Característica	10
5.5. Función de Distribución	10
5.6. Otras Propiedades	10
6. Estimación de Parámetros	12
6.1. Método de los momentos	12
6.2. Método de Máxima Verosimilitud	12
6.3. Método de la Frecuencia de Ceros y Media Muestral	14
6.4. Ejemplo	14
7. Tests para la distribución de Poisson	16
7.1. Test de la razón de verosimilitud	16
7.2. Test del “score” o de los multiplicadores de Lagrang	16
7.3. Ejemplo	17
8. Modelos que conducen a la distribución	18
8.1. Modelo Poisson-Binomial	18
8.2. La distribución de la suma de dos variables de Poisson correlacionadas	19
8.3. Modelo Poisson-Normal	19
9. Ejemplo: Registro de defunciones diarias en Madrid (1995)	21
10. Introducción de la ley de Hermite en la Wikipedia	23
A. Código R	24
Bibliografía	26

1. Objetivo

El objetivo de este trabajo es estudiar las propiedades de la distribución discreta de Hermite para así poder subir toda la información a Internet, concretamente en la enciclopedia libre Wikipedia. Muchas distribuciones están descritas en Wikipedia pero no así ocurre con la distribución de Hermite. Pensamos que sería útil incluir esta información, dadas las muchas aplicaciones de la citada distribución.

2. Introducción

En teoría de probabilidad y estadística, la distribución de probabilidad de una variable aleatoria es una función que a cada suceso definido sobre esta variable le asigna una probabilidad de que dicho suceso ocurra. En estadística las variables aleatorias se pueden diferenciar en 2 categorías; las discretas y las continuas. Este trabajo se centrará únicamente en las variables aleatorias discretas, que son aquellas cuya función de probabilidad solo tomará valores positivos en un conjunto de valores discreto finito o infinito. Como ejemplo de las distribuciones de probabilidad discreta más comunes se pueden nombrar: La distribución binomial, la distribución binomial negativa, la distribución Poisson y la distribución uniforme discreta, entre otras. Pero una de las limitaciones de estas distribuciones es que no permiten modelar datos de recuento cuando estos datos no tienen una única moda, son multimodales. En la siguiente figura se puede ver un ejemplo de este tipo de datos.

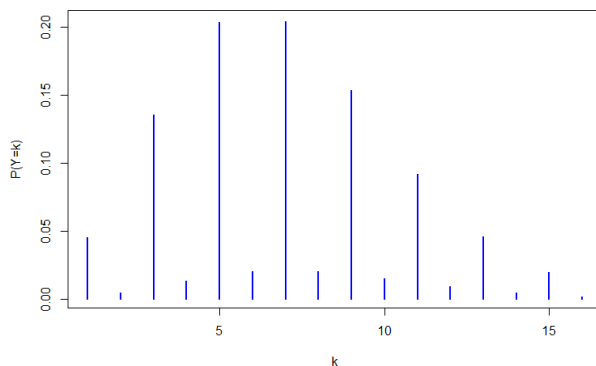


Figura 1: Distribución de probabilidad de datos multimodales

La distribución de Hermite es flexible a la hora de modelar datos de conteo (“*count data*”) que presenten una cierta multimodalidad, y también cuando presentan un exceso de ceros, a esto último se le denomina cero-inflación. En concreto, la distribución de Hermite es cero-inflada respecto a la distribución de Poisson, lo cual significa que la probabilidad del cero en la Hermite es siempre superior a la probabilidad del cero de la Poisson, cuando ambas distribuciones tienen la misma esperanza. También permite modelizar datos en los que aparece el fenómeno de sobredispersión, esto ocurre cuando la varianza es mayor que la esperanza, siendo esta sobredispersión moderada. La distribución Binomial de Poisson para

el caso en que ($n = 2$), coincide con la distribución de Hermite. Se la conoce con este nombre porque los matemáticos Kemp and Kemp [1] la llamaron así ya que tiene la propiedad de que la función de probabilidad y la función generadora de momentos se pueden expresar en términos de los polinomios de Hermite modificados.

En este trabajo se estudian las propiedades de la distribución Hermite, así como los métodos de estimación para los parámetros y algunos tests de ajuste del modelo.

3. Historia

La distribución Hermite apareció por primera vez en una publicación de A.G. McKendrick en 1926, "*Applications of Mathematics to Medical Problems*" [6]. En esta publicación el autor explica diferentes métodos matemáticos que se pueden aplicar a investigaciones médicas, siendo uno de estos métodos el uso de la distribución de Poisson Bivariada, permitiendo de esta manera considerar observaciones de Poisson correlacionadas. McKendrick demostró que la distribución de la suma de dos variables de Poisson correlacionadas siguen una distribución que más tarde sería conocida como distribución Hermite.

Como ejemplo práctico McKendrick modeló la distribución de recuentos de bacterias en leucocitos de un experimento fagocítico. Para estos datos utilizó el método de los momentos para poder estimar los parámetros y obtuvo un modelo más satisfactorio que modelando estos mismos datos con la distribución de Poisson.

Formalmente la distribución de Hermite fue introducida por C.D.Kemp and Adrienne W.Kemp, en 1965 en el trabajo "*Some Properties of 'Hermite' Distribution*". [1] Este trabajo se centra en las propiedades de la distribución de Hermite, tales como la estimación de los parámetros mediante el método de máxima verosimilitud, el cálculo de la función de probabilidad y su relación con los polinomios de Hermite.

Uno de los ejemplos que han utilizado en esta publicación es la distribución de recuentos de bacterias en leucocitos que utilizo McKendrick, pero Kemp and Kemp estiman los parámetros del modelo mediante el método de máxima verosimilitud.

Estos autores publicaron también en 1966 el trabajo "*An Alternative Derivation of the Hermite Distribution*"[8]. En este trabajo se muestra que la distribución de Hermite se puede conseguir mediante la mixtura de una distribución de Poisson y una distribución Normal, es decir, que vendría generada por una variable aleatoria de Poisson cuyo parámetro sería a su vez una variable aleatoria gaussiana.

En 1971, Y.C.Patel [5] en su tesis doctoral hizo un estudio comparativo de varios procedimientos de estimación para esta distribución. Este incluye el método de máxima verosimilitud, la estimación por el método de los momentos y la estimación por el método de la frecuencia de ceros y media muestral.

En 1974, Gupta and Jain en la publicación "*A Generalized Hermite Distribution And Its Properties*" estudiaron una generalización de la distribución Hermite, entre otras cosas mirando la relación existente con los polinomios de Hermite.

4. Definición

4.1. Función 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 variable aleatoria $Y = X_1 + 2X_2$ sigue una distribución de Hermite con parámetros a_1 y a_2 , cuya función de probabilidad es la siguiente [4]:

$$p_n = P(Y = n) = \exp[-(a_1 + a_2)] \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{a_1^{n-2j} a_2^j}{(n-2j)! j!} \quad (1)$$

Donde:

- $n = 0, 1, 2, 3, \dots$
- $a_1, a_2 \geq 0$
- $(n-2j)!$ y $j!$ son el factorial de $(n-2j)$ y j respectivamente
- $\lfloor n/2 \rfloor$ es la parte entera de $n/2$

Demostración: La función de probabilidad de la Poisson es la siguiente,

$$f(k; \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Si $Y = X_1 + 2X_2$ donde $X_1 \sim \text{Poiss}(a_1)$ y $X_2 \sim \text{Poiss}(a_2)$, entonces,

$$\begin{aligned} P(Y = n) &= \sum_{k=0}^{\lfloor n/2 \rfloor} P[X_1 + 2X_2 = n | X_2 = k] P(X_2 = k) \\ &= \sum_{k=0}^{\lfloor n/2 \rfloor} P(X_1 = n - 2k) P(X_2 = k) \\ &= \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{e^{-a_1} a_1^{n-2k}}{(n-2k)!} \frac{e^{-a_2} a_2^k}{k!} \\ &= e^{-a_1} e^{-a_2} \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{a_1^{n-2k} a_2^k}{(n-2k)! k!} \\ &= e^{-(a_1+a_2)} \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{a_1^{n-2k} a_2^k}{(n-2k)! k!} \end{aligned}$$

□

La *función generatriz de probabilidad* es útil para tratar con variables aleatorias discretas que toman valores no negativos. Esta función se puede utilizar para generar las probabilidades de la distribución así como sus momentos.

Una variable aleatoria que sigue una distribución de Hermite tiene la siguiente *función generatriz de probabilidad* [4] :

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

Demostración: La función generatriz de probabilidad para una variable discreta Y esta definida como,

$$G_Y(s) = \sum_{n=0}^{\infty} p_n s^n = E[s^Y]$$

Si X_1 y X_2 son v.a. independientes, donde $X_1 \sim Poiss(a_1)$ y $X_2 \sim Poiss(a_2)$, entonces

$$G_{X_1+2X_2}(s) = E[s^{X_1+2X_2}] = E[s^{X_1} s^{2X_2}] = E[s^{X_1}] E[s^{2X_2}] = E[s^{X_1}] E[(s^2)^{X_2}] = G_{X_1}(s) G_{X_2}(s^2)$$

La función generatriz de probabilidad de la Poisson es,

$$G_X(s) = \exp[\lambda(s-1)]$$

Entonces,

$$\begin{aligned} G_{X_1+2X_2}(s) &= \exp(a_1(s-1)) \exp(a_2(s^2-1)) \\ &= \exp[a_1(s-1) + a_2(s^2-1)] \end{aligned}$$

□

4.2. Notación

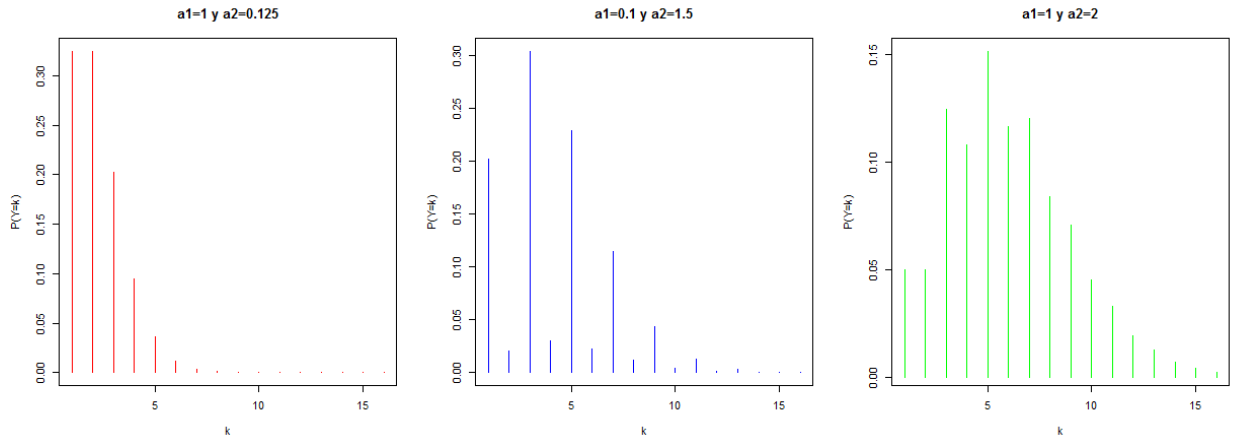
Dada una variable aleatoria $Y = X_1 + 2X_2$ que sigue una distribución Hermite, donde X_1 y X_2 son dos variables independientes Poisson con parámetros a_1 y a_2 respectivamente, entonces escribiremos

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

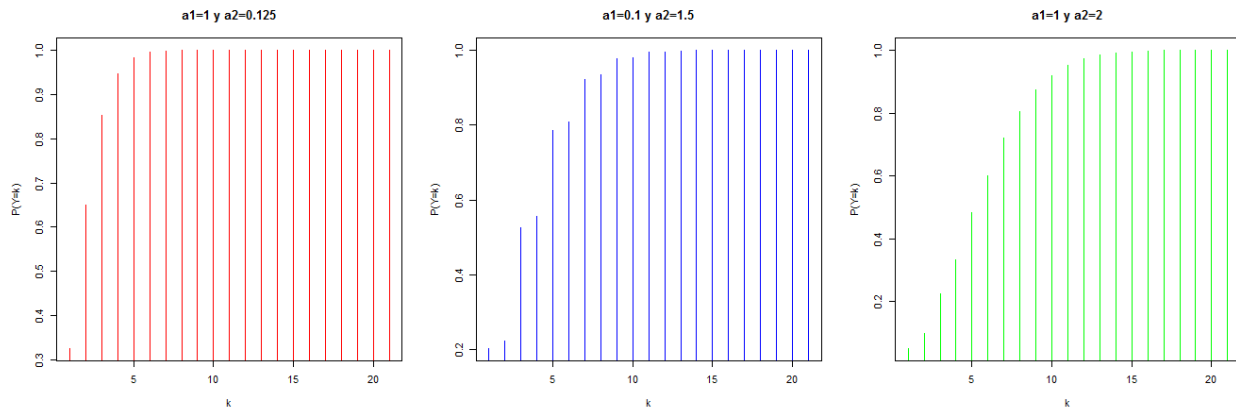
4.3. Gráficas

La figura 2(a) muestra gráficamente algunas funciones de probabilidad de la distribución de Hermite. Se ha calculado esta función con diferentes valores de los parámetros ($Y \sim Herm(1, 0.125)$, $Y \sim Herm(0.1, 1.5)$, $Y \sim Herm(1, 2)$) para observar el cambio en los perfiles. Se puede ver que la función de probabilidad en general es asimétrica, y que en algunos casos no es unimodal.

Mientras que la figura 2(b) muestra gráficamente la función de distribución de probabilidad de Hermite para los mismos valores considerados de los parámetros.



(a) Función de Probabilidad



(b) Función de Distribución

Figura 2

5. Propiedades

5.1. Cumulantes y momentos

Para poder calcular los cumulantes de la distribución primero se calcula la *función generadora de cumulantes* $K(t)$ que se define como el logaritmo de la función generadora de momentos:

$$K(t) = \log(M(t)) = \sum_{n=1}^{\infty} \frac{k_n}{n!} t^n.$$

Los cumulantes son precisamente los coeficientes k_n de su desarrollo en serie de potencias. El primer cumulante k_1 coincide siempre con el valor esperado de la distribución, k_2 es su varianza y k_3 es el momento de orden tres centrado en la esperanza.

Por lo tanto, utilizando la *función generadora de momentos* $M(t)$ de la distribución de Hermite,

$$M(t) = G(e^t) = \exp[a_1(e^t - 1) + a_2(e^{2t} - 1)] \quad (3)$$

Se obtiene que la *función generadora de cumulantes* es [7],

$$K(t) = \log(M(t)) = a_1(e^t - 1) + a_2(e^{2t} - 1) \quad (4)$$

Para obtener los cumulantes hay que ir derivando la función anterior,

$$\begin{aligned} K_1'(t) &= a_1 e^t + 2a_2 e^{2t} \\ K_2''(t) &= a_1 e^t + 4a_2 e^{2t} \\ K_3'''(t) &= a_1 e^t + 8a_2 e^{2t} \\ K_4''''(t) &= a_1 e^t + 16a_2 e^{2t} \\ &\vdots \end{aligned}$$

Se sustituye $t=0$,

$$\begin{aligned} k_1'(0) &= a_1 + 2a_2 \\ k_2''(0) &= a_1 + 4a_2 \\ k_3'''(0) &= a_1 + 8a_2 \\ k_4''''(0) &= a_1 + 16a_2 \\ &\vdots \end{aligned}$$

Por lo tanto se obtiene la siguiente expresión para el n -ésimo cumulante:

$$k_n = a_1 + 2^n a_2 \quad (5)$$

Los cumulantes están relacionados con los momentos a través de la siguiente fórmula [9],

$$k_n = \mu'_n - \sum_{m=1}^{n-1} \binom{n-1}{m-1} k_m \mu'_{n-m}$$

La media y los momentos centrados en la media de orden 2, 3 y 4 de la distribución Hermite son:

Orden	Momento	Cumulante
1	$\mu_1 = k_1 = a_1 + 2a_2$	μ
2	$\mu_2 = k_2 = a_1 + 4a_2$	σ^2
3	$\mu_3 = k_3 = a_1 + 8a_2$	k_3
4	$\mu_4 = k_4 + 3k_2^2 = a_1 + 16a_2 + 3(a_1 + 4a_2)^2$	k_4

5.2. Coeficiente de simetría

El coeficiente de simetría se define como el tercer momento centrado en la media dividido entre la potencia 3/2 de la varianza. En el caso de la distribución de Hermite es [7]:

$$\gamma_1 = \frac{\mu_3^2}{\mu_2^{3/2}} = \frac{(a_1 + 8a_2)^2}{(a_1 + 4a_2)^{3/2}} \quad (6)$$

- Siempre se cumple que $\gamma_1 > 0$, por lo tanto la distribución es asimétrica positiva o a la derecha.

5.3. Curtosis

La curtosis se define como el cuarto momento centrado en la media dividido entre la varianza al cuadrado. En el caso de la distribución de Hermite es [7]:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{a_1 + 16a_2 + (a_1 + 4a_2)^2}{(a_1 + 4a_2)^2} = \frac{a_1 + 16a_2}{(a_1 + 4a_2)^2} + 3 \quad (7)$$

- Siempre se cumple que $\beta_2 > 3$, por lo tanto la distribución es más apuntada y con mayor cola que la distribución Normal.

5.4. Función Característica

La función característica de una variable aleatoria es una función de variable real que toma valores complejos y permite la aplicación de métodos analíticos (análisis funcional) en el estudio de la probabilidad. Esta función, para una variable aleatoria X con distribución discreta a valores enteros no negativos, se define como,

$$\phi(t) = E[e^{itX}] = \sum_{j=0}^{\infty} e^{ijt} P[X = j]$$

De la definición se deduce que la función generadora de momentos está relacionada con la función característica mediante la identidad, $\phi(t) = M(it)$.

Por lo tanto para la distribución de Hermite, utilizando la expresión (3) tenemos que la *función característica* es [1],

$$\phi(t) = e^{[a_1(e^{it}-1)+a_2(e^{2it}-1)]} \quad (8)$$

5.5. Función de Distribución

La función de distribución describe la probabilidad de que una variable aleatoria X , sujeta a cierta ley de distribución de probabilidad, se sitúe en la zona de valores menores o iguales a x .

Por lo tanto, la *función de distribución de Hermite* está definida como [1],

$$\begin{aligned} F(x; a_1, a_2) &= P(X \leq x) \\ &= \exp[-(a_1 + a_2)] \sum_{i=0}^{\lfloor x \rfloor} \sum_{j=0}^{\lfloor i/2 \rfloor} \frac{a_1^{i-2j} a_2^j}{(i-2j)! j!} \end{aligned} \quad (9)$$

5.6. Otras Propiedades

- No es una distribución unimodal.

La distribución Hermite puede tener cualquier número de modas. Esta propiedad se comprueba fácilmente dibujando la función de probabilidad de Hermite con los parámetros $a_1 = 0.01$ y $a_2 = 1.22$, como hizo McKendrick en su publicación [6], y se puede observar que las primeras cinco probabilidades son 0.2923, 0.0029, 0.3566, 0.0036, 0.2175.

- Distribución cerrada bajo convoluciones.

Como sucede con la distribución de Poisson, la distribución de Hermite tiene esta propiedad. Que sea cerrada bajo convoluciones significa que, dadas 2 variables aleatorias (“count data”) $X_1(a_1, a_2)$ y $X_2(b_1, b_2)$ que siguen una distribución Hermite, entonces $X_1 + X_2$ también sigue una distribución Hermite, como explica P. Puig en su publicación [3]. A la hora de utilizar una distribución en ciertas aplicaciones prácticas, puede ser mejor utilizar modelos discretos que tengan esta propiedad.

- Esta distribución permite modelar datos cuando estos tienen exceso de ceros, a este fenómeno se le denomina *cero-inflación* [3]. La distribución de Hermite es cero-inflada respecto a la distribución de Poisson, en el sentido de que la probabilidad del cero en la Hermite es siempre superior a la probabilidad del cero de la Poisson, cuando ambas distribuciones tienen la misma esperanza. Esto es evidente ya que la probabilidad del cero para una $Herm(a_1, a_2)$ es $p_0 = \exp(-a_1 - a_2)$. En cambio la probabilidad del cero para una Poisson con la misma esperanza ($\mu = a_1 + 2a_2$) es $p_0 = \exp(-a_1 - 2a_2)$, que es inferior a la anterior.

En muchos casos cuando una muestra de datos tienen exceso de ceros ajustarlos mediante la distribución de Poisson no es lo adecuado. Para estos casos lo ideal es considerar una distribución discreta con más de un parámetro y que sea flexible ante este fenómeno.

- Sobredispersión moderada [3].

Una variable aleatoria tiene sobredispersión cuando su varianza es mayor a su esperanza. En este caso se dice que es moderada porque el índice de dispersión “ d ” se encuentra entre los valores 1 y 2, como se puede demostrar a continuación,

$$d = \frac{Var(Y)}{E(Y)} = \frac{a_1 + 4a_2}{a_1 + 2a_2} = \frac{a_1 + 2a_2 + 2a_2}{a_1 + 2a_2} = 1 + \frac{2a_2}{a_1 + 2a_2}$$

En la última ecuación se puede observar que el mínimo valor que puede tener “ d ” es 1 (cuando el parámetro a_2 vale 0) y el máximo es 2, porque la fracción siempre es ≤ 1 alcanzando el valor 1 cuando $a_1 = 0$ o cuando a_2 tiende a infinito.

6. Estimación de Parámetros

6.1. Método de los momentos

La media y la varianza de la distribución Hermite son $\mu = a_1 + 2a_2$ y $\sigma^2 = a_1 + 4a_2$.

Para obtener los estimadores de los parámetros a través del método del momento se tiene que resolver el siguiente sistema de ecuaciones:

$$\begin{cases} \bar{x} = a_1 + 2a_2 \\ s^2 = a_1 + 4a_2 \end{cases}$$

Resolviendo este sistema de ecuaciones se obtiene que los estimadores de a_1 y a_2 son,

$$\hat{a}_1 = 2\bar{x} - s^2 \quad (10)$$

$$\hat{a}_2 = \frac{s^2 - \bar{x}}{2} \quad (11)$$

Dados que los parámetros a_1 y a_2 son ambos positivos, los estimadores \hat{a}_1 y \hat{a}_2 son aceptables (es decir, sus estimadores son ≥ 0) solo si $\bar{x} < m_2 < 2\bar{x}$, siendo m_2 el momento muestral de orden 2 [5].

6.2. Método de Máxima Verosimilitud

Dada una muestra X_1, X_2, \dots, X_n con una distribución Hermite con parámetros a_1 y a_2 , se desea estimar estos parámetros basándose en los valores observados de esta muestra.

Se sabe que la media y la varianza de la distribución son $\mu = a_1 + 2a_2$ y $\sigma^2 = a_1 + 4a_2$ respectivamente. Con esta información se puede utilizar la siguiente parametrización,

$$\begin{cases} a_1 = \mu(2 - d) \\ a_2 = \frac{\mu(d - 1)}{2} \end{cases}$$

Sustituyendo estos valores de a_1 y a_2 en la formula (1), se obtiene la siguiente función de probabilidad,

$$P(X = x) = \exp\left[-\left(\mu(2 - d) + \frac{\mu(d - 1)}{2}\right)\right] \sum_{j=0}^{\lfloor x/2 \rfloor} \frac{(\mu(2 - d))^{x-2j} \left(\frac{\mu(d-1)}{2}\right)^j}{(x - 2j)! j!} \quad (12)$$

Por lo tanto el *logaritmo de la función de verosimilitud*, o *función de log-verosimilitud* de la función de distribución anterior es la siguiente [3],

$$\begin{aligned} l(x_1, \dots, x_m; \mu, d) &= \log(L(x_1, \dots, x_m; \mu, d)) \\ &= m\mu \left(-1 + \frac{d-1}{2}\right) + \log(\mu(2-d)) \sum_{i=1}^m x_i + \sum_{i=1}^m \log(q_i(\theta)) \end{aligned} \quad (13)$$

Donde:

$$\blacksquare q_i(\theta) = \sum_{j=0}^{\lfloor x_i/2 \rfloor} \frac{\theta^j}{(x_i - 2j)!j!}; \theta = \frac{d-1}{2\mu(2-d)^2}$$

A partir de la formula (13) se pueden calcular las ecuaciones de verosimilitud,

$$\frac{\partial l}{\partial \mu} = m \left(-1 + \frac{d-1}{2} \right) + \frac{1}{\mu} \sum_{i=1}^m x_i - \frac{d-1}{2\mu^2(2-d)^2} \sum_{i=1}^m \frac{q'_i(\theta)}{q_i(\theta)} \quad (14)$$

$$\frac{\partial l}{\partial d} = m \frac{\mu}{2} - \frac{\sum_{i=1}^m x_i}{2-d} - \frac{d}{2\mu(2-d)^3} \sum_{i=1}^m \frac{q'_i(\theta)}{q_i(\theta)} \quad (15)$$

■ Cálculos directos demuestran que los estimadores de máxima verosimilitud son:

- $\hat{\mu} = \bar{x}$.
- \hat{d} puede calcularse resolviendo numéricamente la ecuación,

$$\sum_{i=1}^m \frac{q'_i(\tilde{\theta})}{q_i(\tilde{\theta})} = m(\bar{x}(2-d))^2 \quad (16)$$

○ Donde, $\tilde{\theta} = \frac{d-1}{2\bar{x}(2-d)^2}$

- Se puede demostrar que la función de log-verosimilitud es estrictamente convexa en el dominio de los parámetros. Consecuentemente el estimador de máxima verosimilitud es único.

No siempre se pueden calcular las ecuaciones de verosimilitud, como muestra la siguiente proposición.

Proposición 1. *Sea x_1, x_2, \dots, x_n una muestra proveniente de una distribución de Hermite. Entonces los estimadores de máxima verosimilitud de los parámetros son $\hat{\mu} = \bar{x}$ y \hat{d} la solución de la ecuación (16), si y sólo si, $m^{(2)}/\bar{x}^2 > 1$, donde $m^{(2)} = \sum_{i=1}^n x_i(x_i - 1)/n$ es el momento factorial empírico de orden 2. [3]*

Observación 1 La condición $m^{(2)}/\bar{x}^2 > 1$ es equivalente a $\tilde{d} > 1$ donde $\tilde{d} = s^2/\bar{x}$ es el índice de dispersión empírico.

Observación 2 En la mayoría de las situaciones cuando la condición de la proposición no se satisface, el MLE es $\hat{\mu} = \bar{x}$ y $\hat{d} = 1$, y por tanto la muestra es modelizada utilizando la distribución de Poisson.

6.3. Método de la Frecuencia de Ceros y Media Muestral

La frecuencia relativa de ceros de una muestra de datos es un estimador natural de la probabilidad de ceros bajo la distribución asumida.

Para la distribución de Hermite se sabe que $f_0 = \exp[-(a_1 + a_2)]$ y $\mu = a_1 + 2a_2$. Por lo tanto, como método de estimación, es natural resolver el siguiente sistema de ecuaciones, de la misma manera que lo hace Y.C.Patel en su trabajo [5],

$$\begin{cases} \bar{x} = a_1 + 2a_2 \\ f_0 = \exp[-(a_1 + a_2)] \end{cases}$$

Donde se obtienen los estimadores para los parámetros a_1 y a_2 ,

$$\hat{a}_1 = -(\bar{x} + 2 \log(f_0)) \quad (17)$$

$$\hat{a}_2 = \bar{x} + \log(f_0) \quad (18)$$

Donde:

- $f_0 = \frac{n_0}{n}$, es la frecuencia empírica de ceros, $n_0 > 0$.

Para una distribución con una alta probabilidad de ceros, se espera que este método funcione satisfactoriamente.

Para valores aceptables de \hat{a}_1 y \hat{a}_2 (es decir, sus estimaciones son ≥ 0), se han de cumplir las siguientes desigualdades,

$$-\log\left(\frac{n_0}{n}\right) < \bar{x} < -2 \log\left(\frac{n_0}{n}\right)$$

6.4. Ejemplo

Para demostrar las aplicaciones de los 3 métodos de estimación anteriormente explicados, se realiza una simulación de la variable $Y = X_1 + 2X_2$, donde $X_1 \sim \text{Poiss}(1.2)$ y $X_2 \sim \text{Poiss}(0.25)$, y por lo tanto la variable Y seguirá una distribución de Hermite, $Y \sim \text{Herm}(1.2, 0.25)$. La tabla 1 muestra la distribución de frecuencias de la variable simulada.

Con la información disponible que se muestra a continuación, se calculan los estimadores para los 2 parámetros a través de los diferentes métodos de estimación, donde los resultados se muestran en la tabla 2,

$$n = 1000; \bar{x} = 1.723; s^2 = 2.271; \hat{d} = 1.2679; f_0 = 0.22$$

- Método de los Momentos

$$\hat{a}_1 = 2\bar{x} - s^2 = 2(1.723) - 2.271$$

$$\hat{a}_2 = \frac{s^2 - \bar{x}}{2} = \frac{2.271 - 1.723}{2}$$

Número	0	1	2	3	4	5	6	7	8	9
Frecuencia	220	303	219	141	61	31	18	5	1	1

Cuadro 1: Frecuencias observadas de la distribución simulada

■ Método de Máxima Verosimilitud

Primero se comprueba si se cumple la condición $\tilde{d} > 1$. Como $\tilde{d} = s^2/\bar{x}$, $\tilde{d} = 2.271/1.723 = 1.3180$ que es mayor a 1, y por lo tanto se pueden calcular los estimadores de máxima verosimilitud a partir de las ecuaciones de verosimilitud.

Para calcular estos parámetros a través de este método, se ha programado la función del logaritmo de la función de verosimilitud en el software estadístico R, y luego se ha utilizado el procedimiento *optimize* para maximizarla y poder obtener así el estimador de d . Este programa se puede ver en el anexo.

$$\hat{a}_1 = \mu(2 - d) = 1.723(2 - 1.2679)$$

$$\hat{a}_2 = \frac{\mu(d - 1)}{2} = \frac{1.723(1.2679 - 1)}{2}$$

■ Método de Cero Frecuencia y Media

$$\hat{a}_1 = -(\bar{x} + 2 \log(f_0)) = -(1.723 + 2 * \log(0.22))$$

$$\hat{a}_2 = \bar{x} + \log(f_0) = 1.723 + \log(0.22)$$

Método	\hat{a}_1	\hat{a}_2
Momentos	1.1777	0.2726
Máxima Verosimilitud	1.2612	0.2309
Cero Frecuencia	1.3053	0.2089

Cuadro 2: Estimadores de los parámetros

7. Tests para la distribución de Poisson

Cuando se utiliza la distribución de Hermite para modelar una muestra de datos es importante comprobar si la distribución de Poisson sería suficiente para ajustar estos datos. Siguiendo la parametrización que se ha utilizado para calcular el estimador de máxima verosimilitud, se debe de contrastar la siguiente hipótesis [3]:

$$\begin{cases} H_0 : d = 1 \\ H_1 : d > 1 \end{cases}$$

Existen diversos métodos para poder realizar este contraste, de los cuales hemos escogido los siguientes:

- Test de la razón de verosimilitud
- Test del “score” o de los multiplicadores de Lagrange

7.1. Test de la razón de verosimilitud

El estadístico ha utilizar será en este caso,

$$W = 2(l(x; \hat{x}, \hat{d}) - l(x; \hat{x}, 1))$$

donde $l()$ es la función de log-verosimilitud (13)

Como $d = 1$ pertenece a la frontera del dominio de los parámetros, bajo la hipótesis nula, el estadístico W no se distribuye como una χ_1^2 . En esta situación se puede establecer que la distribución asintótica de W es una mixtura 50:50 entre una constante igual a cero y una distribución χ_1^2 . El percentil α de la cola superior para esta mixtura es el mismo que el percentil 2α para la cola superior de una χ_1^2 . Por ejemplo, para $\alpha=0.01, 0.05$ y 0.10 serían respectivamente 5.41189, 2.70554 y 1.64237.

7.2. Test del “score” o de los multiplicadores de Lagrang

En este caso el estadístico ha utilizar será,

$$S_2 = 2m \left[\frac{m^{(2)} - \bar{x}^2}{2\bar{x}} \right]^2 = \frac{m(\tilde{d} - 1)^2}{2}$$

donde m = número de observaciones.

Bajo la hipótesis nula, la distribución asintótica es una distribución χ_1^2 . Es conveniente utilizar la versión con signo del test del “score”, es decir, $sign(m^{(2)} - \bar{x}^2)\sqrt{S}$ cuya distribución asintótica sería una normal estándar.

7.3. Ejemplo

Continuando con la misma simulación de datos del Ejemplo (5.4) se contrasta la siguiente hipótesis:

$$\begin{cases} H_0 : d = 1 \\ H_1 : d > 1 \end{cases}$$

Para este contraste se utiliza un nivel de significación $\alpha = 0.05$ para ambos test.

- Test de la razón de verosimilitud

A través del software estadístico R, se programa una función que calcula la función del logaritmo de verosimilitud. Se sustituyen los parámetros anteriormente calculados y se obtiene,

$$l(x; \hat{x}; \hat{d}) = l(Y; 1.723; 1.2679) = -1698.275$$

$$l(x; \hat{x}; 1) = l(Y; 1.723; 1) = -1716.742$$

Por lo tanto el estadístico es,

$$W = 2(-1698.275 + 1716.742) = 36.935$$

Como se ha explicado en el apartado (6.1) para este caso $\chi_1^2 = 2.70554$.

Entonces como $W > \chi_1^2$, rechazaríamos la H_0 .

- Test del “score”

Se calcula el estadístico de score,

$$S_2 = \frac{m(\tilde{d} - 1)^2}{2} = \frac{1000(1.3180 - 1)^2}{2} = 50.56$$

En este caso $\chi_1^2 = 3.84$. Como $S_2 > \chi_1^2$, no se puede aceptar la H_0 .

Para los dos test realizados no se puede aceptar la hipótesis nula y por lo tanto en este caso la distribución Poisson no es suficiente para modelar estos datos, decantándonos en favor de la distribución de Hermite.

8. Modelos que conducen a la distribución

Se pueden considerar 3 modelos diferentes que conducen a la distribución Hermite [1]:

- Modelo Poisson-Binomial.
- La distribución de la suma de dos variables Poisson correlacionadas.
- Modelo Poisson-Normal.

8.1. Modelo Poisson-Binomial

La distribución Poisson-Binomial [1] es un caso particular de la amplia familia de distribuciones denominadas “*Compound Poisson*”. Estas distribuciones se caracterizan por poderse expresar como la suma de un número de variables aleatorias independientes idénticamente distribuidas, donde el número de términos que se suman es una variable aleatoria que sigue una distribución de Poisson:

$$Y = \sum_{i=1}^N x_i$$

- donde X_i i.i.d y $N \sim Poisson(\lambda)$

En el caso en que las X_i siguen una distribución Binomial de parámetros n y p , la distribución de Y se denomina Poisson-Binomial. Inicialmente fue estudiada aplicándola a diversos datos ecológicos

La distribución Poisson-Binomial tiene la siguiente función generatriz de probabilidad,

$$G(s) = \exp[\lambda(q + ps)^n - \lambda]$$

- donde, $0 \leq p \leq 1$; $q = 1 - p$.

Cuando $n = 2$,

$$\begin{aligned} G(s) &= \exp[\lambda(q + ps)^2 - \lambda] \\ &= \exp[\lambda(q^2 + 2ps + p^2s^2) - \lambda] \\ &= \exp[\lambda q^2 + 2ps\lambda + p^2s^2\lambda - \lambda] \end{aligned}$$

Si se sustituye, $a_1 = 2\lambda pq$ y $a_2 = \lambda p^2$,

$$G(s) = \exp[a_1s + a_2s^2 - a_1 - a_2]$$

que es la expresión (2) correspondiente a la distribución Hermite.

8.2. La distribución de la suma de dos variables de Poisson correlacionadas

McKendrick [6] (1926) consideró la distribución bivariada de Poisson y demostró que la suma de 2 variables Poisson correlacionadas (la suma de sus dos coordenadas) siguen una distribución de Hermite.

Esto se puede ver de la siguiente manera:

Diremos que (X, Y) sigue una Poisson Bivariada $(\lambda_1, \lambda_2, \lambda_3)$ si,

$$\begin{cases} X = U + V \\ Y = Z + V \end{cases}$$

donde,

$$\begin{cases} U \sim Poisson(\lambda_1) \\ V \sim Poisson(\lambda_2) \\ Z \sim Poisson(\lambda_3) \end{cases}$$

siendo U, V, Z independientes.

Observemos que en esta distribución bivariada X e Y están correlacionadas porque la variable V aparece tanto en X como en Y .

Entonces si se suman estas 2 variables,

$$\begin{aligned} X + Y &= U + V + Z + V \\ &= U + Z + 2V \end{aligned}$$

Donde,

$$\begin{cases} U + Z \sim Poisson(\lambda_1 + \lambda_3) \\ V \sim Poisson(\lambda_2) \end{cases}$$

Por lo tanto la suma de estas 2 variables correlacionadas siguen una distribución de Hermite $(\lambda_1 + \lambda_3, \lambda_2)$.

8.3. Modelo Poisson-Normal

El modelo Poisson-Normal es un caso particular de la familia de distribuciones “*Mixed Poisson*”, donde en este caso el parámetro λ se considera una variable aleatoria que se distribuye normalmente [8].

En este caso, podemos pensar que las observaciones siguen una $Poiss(\lambda)$, donde a su vez $\lambda \sim Normal(\mu, \sigma^2)$, y entonces podemos demostrar que la distribución de la mixtura seguirá una

$$Hermite((\mu - \sigma^2), (1/2\sigma^2)).$$

La demostración parte del hecho de que la función de generatriz de probabilidad de una distribución de recuento “*Mixed Poisson*” es,

$$G(s) = \int_{-\infty}^{\infty} \exp(\lambda(s-1)) dF(\lambda).$$

Si λ es una normal con media μ y varianza σ^2 , la función de generatriz de probabilidad de esta distribución “*Mixed Poisson*” es,

$$\begin{aligned} G(s) &= (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} \exp(\lambda(s-1) - 1/2(\lambda - \mu)^2/\sigma^2) d\lambda \\ &= \exp((\mu - \sigma^2)(s-1) + 1/2\sigma(s^2 - 1)) \end{aligned}$$

Si se sustituye $a_1 = (\mu - \sigma^2)$ y $a_2 = 1/2\sigma^2$ el resultado se corresponde con la expresión (2).

Kemp y Kemp [8] comentan que, aunque no tiene un sentido estricto el considerar una *mixed Poisson*, utilizando una distribución de mixtura que pueda tomar valores negativos como ocurre con la normal, el proceso tiene sentido desde un punto de vista formal, conduciendo a una función generatriz de probabilidades propia: la de la distribución de Hermite.

9. Ejemplo: Registro de defunciones diarias en Madrid (1995)

En el artículo publicado por Pere Puig en el año 2003 [3] aparece un ejemplo práctico donde se puede aplicar la distribución de Hermite. Se trata de mirar qué tipo de distribución sigue la frecuencia del número de defunciones diarias registradas en la Comunidad Autónoma de Madrid durante el año 1995 para mujeres y hombres de 95 años o mayores. En la tabla 3 aparecen las distribuciones de frecuencias del número de defunciones para mujeres, hombres y para Mujeres y Hombres, que en este último caso son las defunciones totales sin distinguir entre hombres o mujeres. Se ha calculado el coeficiente de correlación de Pearson entre las 365 parejas de conteos de defunciones entre hombres y mujeres, y se ha obtenido un valor no significativo cercano a cero ($r=0.068$).

Es razonable utilizar una distribución invariante bajo adición para ajustar estos datos. Por lo tanto los datos han sido ajustados utilizando la distribución de Hermite. En la siguiente figura se puede ver la gráfica función de probabilidad observada para cada uno de los grupos.

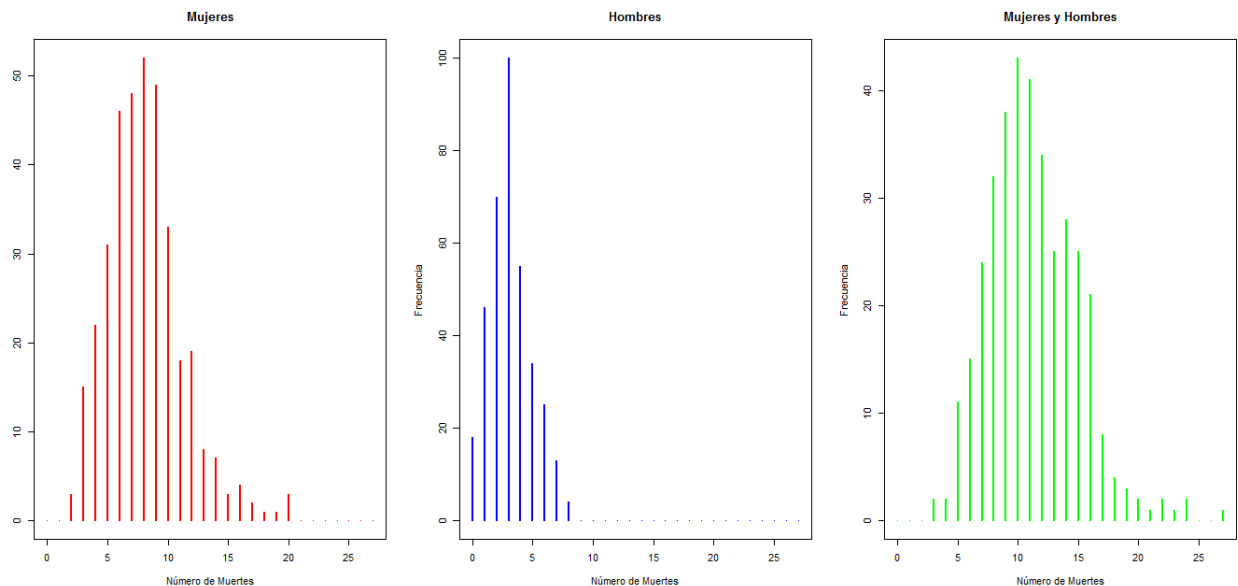


Figura 3: Frecuencias Observadas

Los estimadores para los parámetros se han calculado a través del método de Máxima Verosimilitud, sabiendo que el estimador para μ es la media poblacional y \hat{d} se calcula utilizando la función *optimize* del software R como se ha utilizado en el ejemplo del apartado 6.4 y donde el algoritmo se puede ver en el anexo.

Las correspondientes medias muestrales, índices de dispersión empíricos \tilde{d} y los estimadores de máxima verosimilitud \hat{d} para cada grupo aparecen al final de la tabla 3. Donde se puede observar que para el grupo de los hombres $\tilde{d} < 1$ y como consecuencia de la Proposición 1 el estimador de máxima verosimilitud es $\hat{d} = 1$, y por tanto los datos se ajustan utilizando la distribución de Poisson.

Para comprobar los ajustes respecto a la distribución de Poisson se realiza el test de “score” para los 3 grupos, sabiendo que $\chi_1^2 = 3.84$:

- Mujeres:

$$S_2(Mujeres) = \frac{m(\tilde{d} - 1)^2}{2} = \frac{365(1.2355 - 1)^2}{2} = 10.12$$

. Como $S_2(Mujeres) > \chi_1^2$, no se puede aceptar que los datos siguen una distribución de Poisson.

- Hombres:

$$S_2(Hombres) = \frac{m(\tilde{d} - 1)^2}{2} = \frac{365(0.9852 - 1)^2}{2} = 0.0399$$

. Como $S_2(Hombres) < \chi_1^2$, no se rechaza que los datos siguen una distribución Poisson.

- Mujeres y Hombres:

$$S_2(MyH) = \frac{m(\tilde{d} - 1)^2}{2} = \frac{365(1.2332 - 1)^2}{2} = 9.92$$

. Como $S_2(MyH) > \chi_1^2$, no se puede aceptar que los datos siguen una distribución de Poisson.

Por lo tanto, se puede ver que la distribución Poisson solo se ajusta razonablemente bien para el caso del grupo de los hombres. Para los grupos de las Mujeres y Mujeres y Hombres, la distribución Poisson será rechazada en favor de la distribución de Hermite.

Por último se realizan tests de bondad de ajuste chi-cuadrado para comprobar que las distribuciones teóricas seleccionadas funcionan bien. Los correspondientes grados de libertad y p-valores para cada grupo son: Mujeres, $\chi_{11}^2=11.186$ ($p=0.428$); Hombres, $\chi_7^2=9.810$ ($p=0.200$); y Mujeres y Hombres, $\chi_{14}^2=18.809$ ($p=0.701$). Para los grados de libertad los datos fueron agrupados considerando los grupos ≤ 2 y ≥ 15 para mujeres y ≤ 4 y ≥ 20 para mujeres y hombres.

10. Introducción de la ley de Hermite en la Wikipedia

Por último se ha introducido toda la información en la Wikipedia en inglés, como era el objetivo inicial del trabajo.

La información de momento se encuentra en el siguiente enlace, http://en.wikipedia.org/wiki/Wikipedia_talk:Articles_for_creation/Hermite_distribution, aún pendiente de revisión por parte de los programadores de Wikipedia.

En la figura 4 se puede el cuadro resumen de las propiedades de la distribución Hermite, incluyendo las gráficas de la función de probabilidad y de la función de distribución de probabilidad.

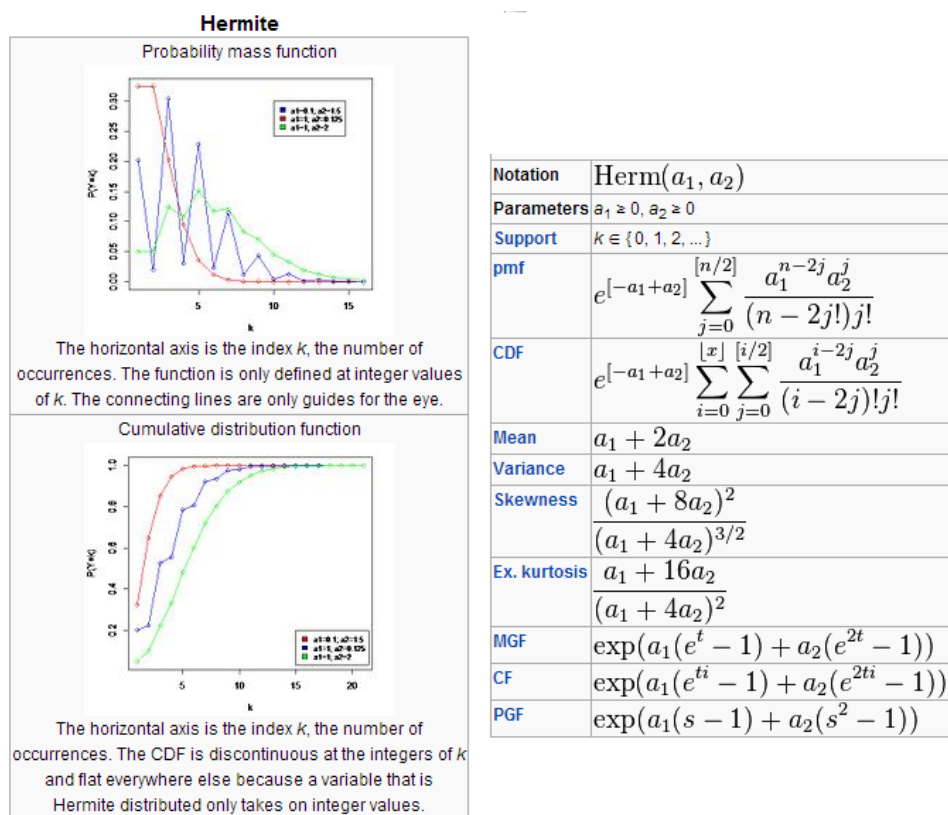


Figura 4: Cuadro resumen de Wikipedia

A. Código R

```
#####  
#### ESTIMACION MAXIMA VEROSIMILITUD ####  
#####  
  
### COMPROVAR  $d > 1$  ###  
  
d <- var(y1)/mean(y1)  
  
### FUNCION PARA ESTIMAR d ###  
  
ver3 <-function(d,mostra)  
{  
  m<-length(mostra)  
  mu=mean(mostra)  
  q=numeric(m)  
  for(i in 1:length(mostra))  
  {  
    qaux<-0  
    for(j in 0:(mostra[i]%/%2))  
    {  
      qaux<-qaux+((d-1)/(2*mu*(2-d)^2))^j/(factorial(mostra[i]-2*j)*factorial(j))  
    }  
    q[i]<-qaux  
  }  
  res<- -(m*mu*(d-3)/2) - log(mu*(2-d))*sum(mostra) - sum(log(q))  
  res  
}  
k<-sort(y1)  
d_est<-optimize(ver3,mostra=k,c(1,(3/2)))  
dest<-d_est$minimum  
  
### ESTIMACIÓN ###  
  
a11 <- mean(y1)*(2-dest)  
a22<- (mean(y1)*(dest-1))/2
```

No. de Muertes	Frec. Mujeres	Frec. Hombres	Frec. Mujeres y Hombres
0	0	18	0
1	0	46	0
2	3	70	0
3	15	100	2
4	22	55	2
5	31	34	11
6	46	25	15
7	48	13	24
8	52	4	32
9	49	0	38
10	33	0	43
11	18	0	41
12	19	0	34
13	8	0	25
14	7	0	28
15	3	0	25
16	4	0	21
17	2	0	8
18	1	0	4
19	1	0	3
20	3	0	2
21	0	0	1
22	0	0	2
23	0	0	1
24	0	0	2
25	0	0	0
26	0	0	0
27	0	0	1
$\hat{\mu}$	8.1123	3.1479	11.2603
\tilde{d}	1.2355	0.9852	1.2332
\hat{d}	1.1851	1	1.2032

Cuadro 3: Registros de defunciones en la Comunidad Autónoma de Madrid durante 1995

Referencias

- [1] Kemp,C.D., and Kemp,A.W. (1965), “Some Properties of the 'Hermite' Distribution”, *Biometrika*, 52, 381-394
- [2] Gupta, R.P., and Jain, G.C. (1974), “A Generalized Hermite Distribution and Its Properties”, *SIAM Journal on Applied Mathematics*, 27, 359-363
- [3] Puig, P. (2003), “Characterizing Additively Closed Discrete Models by a Property of Their Maximum Likelihood Estimators, With an Application to Generalized Hermite Distributions”, *Journal of the American Statistical Association*, 463, 687-692
- [4] Kotz, Samuel (1982-1989). “Encyclopedia of statistical sciences”. *John Wiley*
- [5] Y.C.Patel (1976), “Even Point Estimation and Moment Estimation in Hermite Distribution”, *Biometrics*, 32, 865-873
- [6] A.G.McKendrick (1926),“Applications of Mathematics to Medical Problems”, *Proceedings of the Edinburgh Mathematical Society*, 44, 98-130.
- [7] Kotz, Samuel (2005). “Univariate discrete distributions”. *Norman L. Johnson*
- [8] Kemp,C.D., and Kemp,A.W. (1966), “An Alternative Derivation of the Hermite Distribution”, *Biometrika*, 53,627-628.
- [9] Wikipedia, <http://en.wikipedia.org/wiki/Cumulant>