TRATAMIENTO DE DATOS DE UNA MAGNITUD QUE PRESENTA DISTRIBUCIÓN EXPONENCIAL TOMADOS CON ERROR INSTRUMENTAL NO DESPRECIABLE. ESTIMACIÓN DE LA CONSTANTE DE DECAIMIENTO.

J.J. Blostein

Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires
Ciudad Universitaria, Pab 1, (1428) Buenos Aires, Argentina
e-mail: blostein@ciudad.com.ar

Se desarrolla un formalismo que permite estimar el parámetro de una variable aleatoria que presenta distribución exponencial para datos obtenidos con incerteza de truncamiento.

A formalism is developed so that it allows us to estimate the parameter of an aleatory variable that presents exponential distribution for data obtained with rounding uncertainty.

I. INTRODUCCIÓN

Existe una gran variedad de magnitudes físicas que presentan distribución exponencial, también conocida como distribución de decaimiento exponencial. Sin embargo, no siempre resulta inmediato determinar experimentalmente el parámetro característico de la distribución. En general aunque la magnitud en cuestión adopta valores continuos, luego de cada medición el instrumento informa sólo valores discretos. El resultado que en cada medición informa el instrumento no coincide con el verdadero valor de la magnitud en cuestión, sino que se encuentra afectado por el redondeo o truncamiento que inevitablemente realizan los instrumentos digitales. Se supone en este trabajo que el instrumento discretiza la magnitud en cuestión, y que se realiza un histograma contando durante intervalos regulares de ancho T la cantidad de eventos registrados. Una de las situaciones más delicadas se presenta cuando la incerteza debida al truncamiento del sistema de medición, T, es comparable al parámetro característico de decaimiento de la magnitud medida, τ. En estos casos las mediciones se encuentran sensiblemente afectadas por T. En el presente trabajo se encara el problema de la estimación del parámetro τ teniendo en cuenta la incerteza con que se obtiene cada medición producto de la discretitud del instrumento.

II. DESARROLLO

El desarrollo matemático parte de suponer una variable aleatoria X distribuida exponencialmente con parámetro $\lambda=1/\tau$ que se desea determinar. La función de densidad de probabilidad (f.d.p.) asociada a la variable aleatoria X es entonces

$$f_\lambda(x)=\lambda e^{-\lambda x} \quad , \quad x>0, \tag{1}$$

donde el subíndice λ indica la dependencia paramétrica de la función f.

El instrumento de medición se modela de la siguiente manera: tiene divisiones equiespaciadas a intervalos regulares de ancho T. Cuando se realiza una medición la variable aleatoria X adopta un valor x. Si x pertenece al rango del instrumento se obtiene una lectura. Esta lectura será aquella división del instrumento inmediatamente mayor a x.

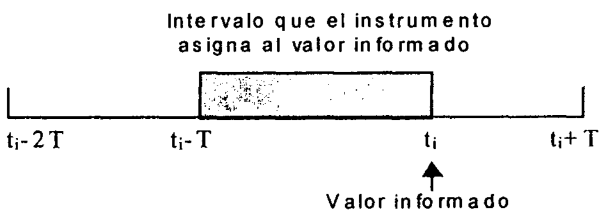

Figura 1. : Esquema del instrumento de medición.

Tomando las lecturas del instrumento como eventos de interés, la probabilidad de obtener una lectura t_i cualquiera, es la probabilidad de que X adopte cualquier valor dentro del intervalo [ti -T, ti]:

$$g_{\lambda,T}(t_i)=\int_{t_i-T}^{t_i}\lambda e^{-\lambda\xi}d\xi \ . \tag{2}$$

Al introducir la variable ω=-$\lambda\xi$, la integral (2) se reduce a

$$g_{\lambda,T}(t_i)=\int_{-\lambda t_{i-1}}^{-\lambda t_i}e^{\omega}d\omega=e^{-\lambda t_{i-1}}-e^{-\lambda t_i} \ ; \tag{3}$$

entonces la función g resulta

$$g_{\lambda,T}(t)=e^{-\lambda t}(e^{\lambda T}-1) \ . \tag{4}$$

Los subíndices indican que la función g depende paramétricamente de λ y T. Si se tiene entonces un instrumento cuya resolución es T y se admite que $i.T$ son los posibles valores de lectura ($i \in \mathrm{N}\geq 1$, donde i es el indicador del intervalo), entonces la probabilidad de obtener la lectura iT en una medición será $g_{\lambda,T}(iT)$. Esto significa que la función g al ser evaluada en los valores correspondientes a las lecturas del instrumento representa un histograma (teórico, de variable discreta) que indica la frecuencia relativa con la que se debería obtenerse cada lectura. Esto es análogo a lo que ocurre con la función g definida en la referencia1. Puede verificarse que $g_{\lambda,T}(iT)$ es mayor o igual que cero para todo i, y que

$$\sum_{i=1}^{+\infty} g_{\lambda,T}(iT) = 1 \ , \forall T \in R{>}0 \qquad (5)$$

con lo cual queda garantizado que $g_{\lambda,T}(iT)$ es una función de distribución de probabilidad (f.d.p) discreta[2].

Para encontrar la función $g_{\lambda,T}$ que mejor se ajusta al histograma experimental se aplicará el criterio de máxima verosimilitud[3, 4 y 5] (ML), y se obtendrá una expresión analítica que permite estimar λ. La probabilidad de que una lectura cualquiera, por ejemplo la primera, sea igual a T es $g_{\lambda,T}(T)$. Puesto que cada lectura se supone independiente de las demás, la probabilidad de que la segunda o cualquier otra lectura sea igual a T es también $g_{\lambda,T}(T)$; por lo tanto, la probabilidad de obtener dos lecturas T es $g_{\lambda,T}(T)^2$. Continuando con este razonamiento, la probabilidad de obtener n_1 lecturas iguales a T es $g_{\lambda,T}(T)^{n1}$. Análogamente, la probabilidad de que n_2 lecturas sean $2T$ es $g_{\lambda,T}(2T)^{n2}$, la probabilidad de que n_3 lecturas sean $3T$ es $g_{\lambda,T}(3T)^{n3}$, y así siguiendo. Antes de realizar un experimento, la probabilidad L de obtener un histograma experimental $\{n_i\}$ $i \geq 1$, es entonces

$$L = \prod_{i=1}^{+\infty} g_{\lambda,T}(iT)^{n_i} \, . \qquad (6)$$

L se conoce con el nombre de función de verosimilitud. Una vez realizado el experimento, y sabiendo T a partir de las características del instrumento, mediante la maximización de L es posible estimar el parámetro λ de la función g más verosímil.

Dado que el logaritmo es una función monótonamente creciente, maximizar L es equivalente a maximizar su logaritmo. Al aplicar logaritmos a ambos miembros de (6) se tiene

$$\ln L = \sum_{i=1}^{+\infty} n_i \ln g_{\lambda,T}(iT) \qquad (7)$$

Se define la función objetivo

$$S \equiv -\ln L = -\sum_{i=1}^{+\infty} n_i \ln g_{\lambda,T}(iT) \qquad (8)$$

que depende paramétricamente de λ y de T. Finalmente, al reemplazar en (8) la expresión de $g_{\lambda,T}(t)$ de la ecuación (4) resulta

$$S = -\sum_{i=1}^{+\infty} n_i \left[-\lambda iT + \ln(e^{\lambda T} - 1) \right] \qquad (9)$$

El objetivo es encontrar el valor de λ que minimiza S, que se tomará como estimador de λ y se notará con el símbolo Λ. Al derivar S con respecto a λ, evaluar en $\lambda = \Lambda$ e igualar a cero se obtiene

$$\left.\frac{\partial S}{\partial \lambda}\right|_{\lambda=\Lambda} = -\sum_{i=1}^{+\infty} n_i \left[-iT + \frac{Te^{\Lambda T}}{e^{\Lambda T} - 1} \right] = 0, \qquad (10)$$

o sea

$$\sum_{i=1}^{+\infty} n_i \left[-i(e^{\Lambda T} - 1) + e^{\Lambda T} \right] = 0 \, . \qquad (11)$$

Entonces

$$\sum_{i=1}^{+\infty} n_i . i . (e^{\Lambda T} - 1) = N e^{\Lambda T} \, , \qquad (12)$$

donde

$$N = \sum_{i=1}^{+\infty} n_i \, , \qquad (13)$$

es el número total de mediciones. Entonces

$$\sum_{i=1}^{+\infty} n_i . i = N \frac{e^{\Lambda T}}{(e^{\Lambda T} - 1)} \, , \qquad (14)$$

que tomando logaritmos se transforma en

$$\ln\left(\sum_{i=1}^{+\infty} i . \frac{n_i}{N}\right) = \Lambda T - \ln(e^{\Lambda T} - 1) \, . \qquad (15)$$

III. RESULTADOS

El estimador de máxima verosimilitud del parámetro λ se obtiene despejando Λ de la ecuación (15). Para ello, si se define

$$A \equiv \ln\left(\sum_{i=1}^{+\infty} i . \frac{n_i}{N}\right) \, , \qquad (16)$$

y se introduce la variable

$$\alpha = \Lambda T \, , \qquad (17)$$

la ecuación (15) puede reescribirse de la siguiente forma:

$$\alpha - \ln(e^{\alpha} - 1) - A = 0 \, . \qquad (18)$$

Si ahora introducimos la variable

$$y = e^{\alpha} - 1 \, , \qquad (19)$$

la ecuación (18) puede ponerse como

$$\ln\left(\frac{1+y}{y}\right) - A = 0 \, , \qquad (20)$$

de donde de despejar y e igualar a su definición resulta

$$y = \frac{1}{e^A - 1} = e^{\alpha} - 1. \quad (21)$$

Al despejar α de esta última ecuación tenemos

$$\alpha = \ln(1 + \frac{1}{e^A - 1}), \quad (22)$$

que reemplazando las definiciones de α y A es equivalente a

$$\Lambda = \frac{1}{T}\ln(1 + \frac{1}{\sum_{i=1}^{+\infty} i.\frac{n_i}{N} - 1}). \quad (23)$$

Puede verificarse por sustitución directa de la ecuación (23) en la ecuación (15) que el valor de Λ presentado en (23) es correcto.

Una vez realizadas las N mediciones, y en consecuencia conocidos los números naturales n_i que conforman el histograma experimental, se toma como estimador de λ aquel que surge de la ecuación (23), o sea aquel que maximiza L, asumiendo T conocido a partir de las características del instrumento. Si bien este estimador puede resultar sesgado, en el límite $N\to\infty$ tiende al parámetro λ de la distribución original, puesto que es un estimador de ML[3].

Existen condiciones que restringen la obtención de un estimador finito. Si se analiza el argumento del logaritmo de la ecuación (23), puede verse que el mismo se anula sólo cuando $\sum_{i=1}^{+\infty} i.\frac{n_i}{N} = 0$, lo cual ocurriría sólo si todos los términos de la sumatoria fuesen nulos, esto es si no se registrase lectura alguna. Asimismo, si las N lecturas informadas por el instrumento fuesen todas iguales a T, esto es si n_1 fuese igual a N y las restantes n_i (para i >1) fuesen nulas, dicho argumento sería infinito y por lo tanto la estimación de λ no sería finita.

Por su parte, resulta interesante analizar si existen condiciones bajo que se obtendría un estimador Λ nulo. El argumento del logaritmo de la ecuación (23) es igual a la unidad sólo si la serie $\sum_{i=1}^{+\infty} i.\frac{n_i}{N}$ divege; lo cual no ocurre para N finito puesto que a partir de un cierto valor de i todos los n_i son nulos y la serie converge.

Por la ley de los grandes números demostrada en la referencia[3], en el caso límite de N tendiendo a infinito, la frecuencia relativa n_i/N con que se obtiene cada valor de lectura iT tenderá a la probabilidad $g_{\lambda,T}(iT)$, donde la función $g_{\lambda,T}(t)$ está definida en (4). Si reemplazamos entonces en la ecuación (23) el valor límite

$$\frac{n_i}{N} \xrightarrow[N\to\infty]{} e^{-\lambda iT}(e^{\lambda T} - 1), \quad (24)$$

obtenemos

$$\Lambda \xrightarrow[N\to\infty]{} \frac{1}{T}\ln(1 - \frac{1}{1 - (e^{\lambda T} - 1)\sum_{i=1}^{+\infty} i(e^{-\lambda T})^i}) \quad (25)$$

Derivando la serie geométrica dentro de su rango de convergencia, se demuestra que[5] la siguiente serie converge al valor

$$\sum_{i=1}^{+\infty} i.p^i = \frac{p}{(1-p)^2}, \text{ si } |p| < 1. \quad (26)$$

Por lo tanto, como $e^{-\lambda T} < 1$ para todo λ>0, la expresión (25) para el estimador Λ en este caso límite se convierte en

$$\Lambda \xrightarrow[N\to\infty]{} \frac{1}{T}\ln(1 - \frac{1}{1 - (e^{\lambda T} - 1)\frac{e^{-\lambda T}}{(1 - e^{-\lambda T})^2}}) = \lambda \quad (27)$$

Este resultado era de esperar ya que, como se dijo anteriormente, el estimador Λ es de máxima verosimilitud y por lo tanto, como se demuestra para el caso general en la referencia[5], en el límite de infinitas mediciones debe tender al parámetro λ de la distribución original.

En la bibliografía[3] se demuestra que si las mediciones no se encontrasen afectadas por el error de apreciación instrumental, el estimador de máxima verosimilitud del parámetro λ sería

$$\Lambda_{Trad} = \frac{1}{\bar{x}}, \quad (28)$$

donde el subíndice *Trad* hace referencia que es la forma tradicional de estimar λ, y $\bar{x}$ indica el promedio de los valores medidos de la variable aleatoria X

$$\bar{x} = \sum_{j=1}^{N} x_j. \quad (29)$$

Si se insiste en aplicar la ecuación (28) para la estimación de λ el resultado será

$$\Lambda_{Trad} = \frac{1}{\sum_{i=1}^{\infty} iT.\frac{n_i}{N}} \quad (30)$$

donde cada valor de lectura iT se ha registrado n_i veces. En el limite de infinitas mediciones ($N\to\infty$), el resultado de la estimación dada en (30) tenderá a

$$\Lambda_{Trad} \xrightarrow[N\to\infty]{} \frac{1}{\sum_{i=1}^{\infty} iT.e^{-\lambda iT}(e^{\lambda T} - 1)}, \quad (31)$$

donde simplemente se ha utilizado la expresión (24) para reemplazar las frecuencias relativas observadas. Esta última expresión puede reescribirse como

$$\frac{1}{T(e^{\lambda T} - 1)\sum_{i=1}^{\infty} i.(e^{-\lambda T})^i}, \quad (32)$$

la cual, luego de utilizar el resultado de la serie (26), se transforma en

$$\frac{1}{T(e^{\lambda T}-1)}\,\frac{(1-e^{-\lambda T})^{2}}{e^{-\lambda T}}=\frac{1-e^{-\lambda T}}{T}\neq\lambda \quad \forall\lambda>0 \tag{33}$$

Esto significa que en el caso límite de infinitas mediciones la estimación de λ por medio de la ecuación (30) es sesgada. Este último resultado se debe a que en realidad todas la mediciones se encuentran afectadas por el instrumento que se utilice. Podemos calcular explícitamente este sesgo si definimos la función $\xi(\lambda,T)$ como la diferencia entre el verdadero valor de λ y la expresión (33)

$$\xi(\lambda,T)\equiv\lambda-\frac{1-e^{-\lambda T}}{T}. \tag{34}$$

Si dividimos (34) por λ a ambos miembros tenemos

$$\frac{\xi(\lambda,T)}{\lambda}=1-\frac{1-e^{-\lambda T}}{\lambda T}\equiv\varepsilon(\lambda T). \tag{35}$$

$\varepsilon(\lambda T)$ es el error relativo que se cometería en el caso límite de infinitas mediciones al estimar λ por medio de la expresión (30). Esta función depende sólo del producto λT, y en la Figura 2 puede verse graficada para un amplio rango de valores de su argumento, mostrándose que tiende a uno cuando su argumento tiende a infinito.

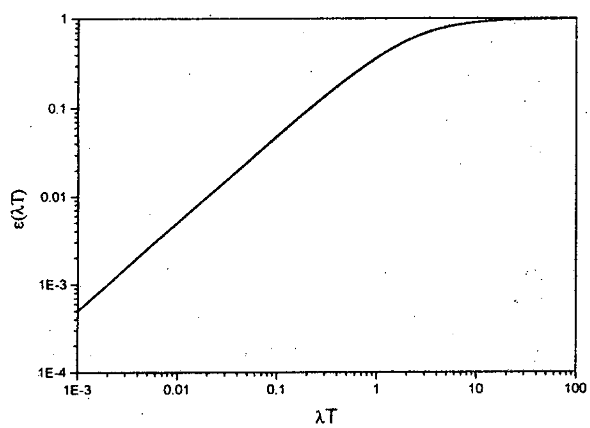

Figura 2: Gráfico de la función ε.

En muchas oportunidades se acostumbra a trabajar con el parámetro $\tau=1/\lambda$ en vez de con el parámetro λ. Λ es el estimador de ML de λ, y por lo tanto tiene la propiedad invarianza[3]; el estimador de ML de τ, $\hat{\tau}$, es entonces la inversa de Λ, o sea en vista de (23)

$$\frac{1}{\Lambda}=\hat{\tau}=\frac{T}{\ln\left(1+\dfrac{1}{\sum_{i=1}^{+\infty} i.\dfrac{n_i}{N}-1}\right)}. \tag{36}$$

El estimador $\hat{\tau}$, al igual que Λ, tiene en cuenta el error producto de la discretitud de los valores de lectura del instrumento, y por las razones expuestas en el párrafo precedente a la ecuación (24) para N finito produce siempre una estimación finita. El estimador $\hat{\tau}$ es nulo cuando diverge el denominador de (36), y esto ocurre cuando las N lecturas son iguales a T.

De ignorarse en el tratamiento de los datos medidos el error producto de la discretitud en las lecturas del instrumento, el parámetro τ se estimaría en la forma tradicional dada en la bibliografía[3]

$$\hat{\tau}_{Trad}=\sum_{i=1}^{\infty} iT.\frac{n_i}{N}. \tag{37}$$

Al reemplazar en la ecuación (37) el cociente n_i/N por el valor dado en (24), al cual debe tender cuando el número total de mediciones tiende a infinito, y luego de utilizar (26) para calcular la serie involucrada resulta:

$$\hat{\tau}_{Trad}\xrightarrow[N\to\infty]{}\frac{T}{1-e^{-\frac{T}{\tau}}}. \tag{38}$$

Definamos el sesgo $\delta(\tau,T)$ como la diferencia entre el estimador $\hat{\tau}_{Trad}$ y el verdadero valor de τ

$$\delta(\tau,T)\equiv\hat{\tau}_{Trad}-\tau. \tag{39}$$

En virtud de la expresión (38) tenemos

$$\delta(\tau,T)\xrightarrow[N\to\infty]{}\frac{T}{1-e^{-\frac{T}{\tau}}}-\tau. \tag{40}$$

El límite (40) nos permite definir la función $\delta_{N\to\infty}(\tau,T)$

$$\delta_{N\to\infty}(\tau,T)\equiv\frac{T}{1-e^{-\frac{T}{\tau}}}-\tau\ , \tag{41}$$

expresión a la cual tiende el sesgo $\delta(\tau,T)$ cuando el número total de mediciones N tiende a infinito.
Resulta conveniente definir al función $\Delta(\tau/T)$

$$\Delta(\tau/T)\equiv\frac{\delta_{N\to\infty}}{T}=\frac{1}{1-e^{-\frac{T}{\tau}}}-\frac{\tau}{T} \tag{42}$$

Un gráfico de Δ en función de τ/T es mostrado en la Figura 3.

CEILAP
CITEFA - CONICET
ZUFRIATEGUI Y VARELA
1603 VILLA MARTELLI
REPUBLICA ARGENTINA

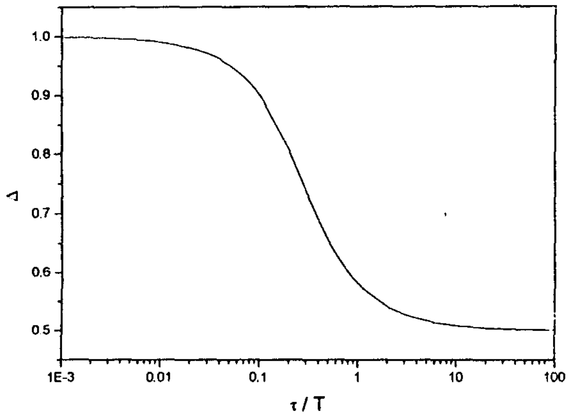

Figura 3: Gráfico de Δ en función de τ / T.

Δ es el sesgo, medido en unidades de T, que se cometería al utilizar la expresión (37) para estimar τ cuando el número de mediciones tiende a infinito. Puede mostrarse matemáticamente los siguientes límites de Δ:

$$\Delta(\tau / T) \xrightarrow[\tau / T \to 0]{} 1, \qquad (43)$$

$$\Delta(\tau / T) \xrightarrow[\tau / T \to \infty]{} 1/2. \qquad (44)$$

Que la función ε tienda a uno cuando su argumento λT tiende a infinito es equivalente al límite de la expresión (43), y significa que en ese caso prácticamente toda la estimación se debe a la contribución del sesgo. Esto puede entenderse pensando que en casos cercanos a ese límite, el decaimiento se produce tan rápidamente que casi todas lecturas son iguales al mínimo valor T, cometiéndose en cada medición una sobreestimación de la variable aleatoria X casi igual al valor de lectura promediado en (37) para la estimación de τ. Por su parte en el otro caso extremo, cuando $\tau >> T$, la constante de decaimiento es tan grande que efecto de la discretitud del instrumento se hace despreciable, y por ello el error relativo ε que se comete en la estimación de λ tiende a anularse, tal como puede verse en la figura 2. En este límite, y cuando el número de mediciones tiende a infinito, el error absoluto cometido en la estimación de τ por medio de (36) tiende a $T/2$ lo que hace que Δ tienda a 1/2 y el error relativo en la estimación de τ tienda a anularse. En esta situación, que τ no tienda a anularse se entiende debido a que el valor de lectura del instrumento no se ha supuesto en el centro de cada intervalo, sino en su valor extremo.

IV. CONCLUSIONES

Es posible estimar el parámetro de una magnitud que presenta distribución exponencial contemplando la incerteza debida a la discretitud en los valores de lectura informados por el instrumento.

Si se utiliza el parámetro λ para describir un decaimiento exponencial, el mismo puede ser estimado mediante el estimador Λ dado en la ecuación (23). Se mostró en este trabajo que de registrarse al menos una lectura en el instrumento, y de no ser todas las lecturas iguales al mínimo valor, la estimación de λ por medio de Λ resulta finita. En todos los casos reales el número total de mediciones N es finito y por lo tanto la estimación Λ arroja siempre un resultado no nulo.

Por su parte, si se utiliza el parámetro τ (inverso del parámetro λ) para describir el decaimiento exponencial, el mismo puede ser estimado mediante el estimador $\hat{\tau}$ dado en la expresión (36) (inverso del estimador Λ). Debido a que la expresión para Λ resulta siempre distinta de cero, $\hat{\tau}$ es siempre un estimador finito. La estimación de τ por medio de la expresión (36) resulta nula de no registrarse ninguna lectura o de ser todas las lecturas iguales al mínimo valor.

Cuando el número de mediciones tiende a infinito, la convergencia de los estimadores Λ y $\hat{\tau}$ a los parámetros originales λ y τ respectivamente está garantizada por tratarse de estimadores de máxima verosimilitud, y se ha verificado este comportamiento en el resultado al que se arriba en la expresión (27). En principio es posible aproximarse tanto como se quiera al parámetro que caracteriza una distribución exponencial aumentando el número de mediciones. Ello se debe a que además de las frecuencias relativas de las lecturas, se cuenta con el conocimiento de la f.d.p. que las describe incluyendo en truncamiento. Podemos liberarnos de la limitación experimenta T a costa de interpretar cuidadosamente la información adquirida.

Finalmente se muestra lo inconveniente de estimar los parámetros λ y τ por medio de las tradicionales expresiones (30) y (36). Al tratar las lecturas como si no existiese error instrumental los estimadores obtenidos poseen sesgos que persisten aún en el caso límite de infinitas mediciones; dichos sesgos conducen a la definición de las funciones ε y Δ, que se encuentran graficadas en las Figuras 2 y 3.

Referencias

1 – Blostein J.J., Do Campo P., Moreno C., *Tratamiento de datos de una magnitud que presenta distribución normal tomados con error instrumental no despreciable*, Anales A.F.A. 1995, San Carlos de Bariloche.

2 –Feller W., *An introduction to probability theory and its aplications*, Vol 1, Cap 1, 3° Ed., Wiley, New York (1968)

3 – Meyer P.L., *Probabilidades y aplicaciones estadísticas*, Fondo Educativo Interamericano, México (1973).

4 – Freeman H., *Introducción a la inferencia estadística*, Trillas, México (1970).

5 – Cramer H., *Métodos matemáticos de estadística*, Instituto Nacional de Estadística, Madrid (1949).