UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

AJUSTE A ALGUNAS DISTRIBUCIONES DE PROBABILIDAD USADAS EN HIDROLOGIA SUPERFICIAL

T E S I S

QUE PARA OBTENER EL TITULO DE

INGENIERO CIVIL

P R E S E N T A :

IGNACIO ARTURO CALDIÑO VILLAGOMEZ

MEXICO, D.F. 1982

FACULTAD DE INGENIERIA
EXAMENES PROFESIONALES
60-1-290

Al Pasante señor IGNACIO ARTURO CALDINO VILLAGOMEZ,
P r e s e n t e .

En atención a su solicitud relativa, me es grato transcribir a usted a continuación el tema que aprobado por esta Dirección propuso el Profesor M. en I. Oscar A. Fuentes Mariles, para que lo desarrolle como tesis en su Examen Profesional de Ingeniero CIVIL.

"AJUSTE A ALGUNAS DISTRIBUCIONES DE PROBABILIDAD USADAS EN HIDROLOGIA SUPERFICIAL"

1. Introducción.
2. Fundamentos de probabilidad y estadística.
3. Funciones de distribución de probabilidad.
4. Estimadores.
5. Métodos de ajuste.
6. Bondad de ajuste.
7. Aplicaciones.
8. Conclusiones y recomendaciones.

Ruego a usted se sirva tomar debida nota de que en cumplimiento de lo especificado por la Ley de Profesiones, deberá prestar Servicio Social durante un tiempo mínimo de seis meses como requisito indispensable para sustentar Examen Profesional; así como de la disposición de la Dirección General de Servicios Escolares en el sentido de que se imprima en lugar visible de los ejemplares de la tesis, el título del trabajo realizado.

A t e n t a m e n t e
"POR MI RAZA HABLARA EL ESPIRITU"
Cd. Universitaria, 4 de noviembre de 1981.
EL DIRECTOR

ING. JAVIER JIMENEZ ESPRIU

JJE/OBLH/sen

I N D I C E

1. INTRODUCCION

El diseño de muchas obras hidráulicas requiere de la determinación de la magnitud de un evento relacionado con fenómenos naturales como precipitaciones o avenidas cuyos valores son variables con respecto al tiempo y tienen carácter aleatorio.

En el diseño de determinados tipos de obras hidráulicas, el concepto del período de retorno de las avenidas o gastos máximos posee gran importancia ya que tal período está asociado a la probabilidad de ocurrencia (riesgo) de un evento de mayor magnitud que el de diseño durante la vida útil de la obra.

La magnitud del evento hidrológico de diseño debe ser tal que satisfaga las condiciones de seguridad y economía de una obra de Ingeniería Civil. Para determinar esta magnitud, el ingeniero puede recurrir a registros de algunos eventos (gastos, precipitaciones, volúmenes,...) medidos en estaciones para tal efecto.

El ingeniero, por lo tanto, se enfrenta por lo menos a tres tipos de incertidumbre*: La natural o esencial debida a la consideración de

* Referencia 2.

que los eventos hidrológicos son eventos aleatorios; la estadística, relacionada con una determinación imprecisa de los parámetros del modelo; y la de modelo, al no poder tener la seguridad de que el modelo escogido es la mejor elección.

La incertidumbre natural no puede evitarse debido a la misma naturaleza de los eventos hidrológicos al carecer de una información completa de ellos. Lo que en éste trabajo se trata, está relacionado con la incertidumbre estadística y con la incertidumbre de modelo.

El uso de modelos probabilísticos y de la e s t a d í s t i c a para inferir el comportamiento de los eventos hidrológicos data de mucho tiempo y está respaldado por una amplia teoría.

En este trabajo, se presenta la teoría en la que se basan las inferencias a partir de las cuales se obtienen eventos de diseño, así como su aplicación a los gastos máximos anuales registrados en una estación hidrométrica.

Se presentan ejemplos, algunos de ellos, con la finalidad de visualizar mejor conceptos teóricos y otros que, haciendo uso de la teoría de Probabilidad y Estadística, pretenden hacer notar la importancia que tiene el hacer ciertas consideraciones a priori, las cuales pueden conducir a errores de mayor magnitud en la obtención de un evento de diseño.

Se trata también, con la información disponible, de reducir las incertidumbres estadística y de modelo al aplicar la teoría, además de cuantificarlas.

El capítulo 2 se refiere a las distribuciones empíricas y la teoría de Probabilidad, además de conceptos utilizados en la estimación de eventos como el período de retorno y valores extremos.

En el capítulo 3, se estudian varios modelos usados en Hidrología superficial; conociendo sus características y con la experiencia del ingeniero puede reducirse la tarea de seleccionar o rechazar un modelo.

En el capítulo 4 se tratan las propiedades de los estimadores de los parámetros; en el capítulo 5, los métodos de estimación de esos parámetros así como sus propiedades al ser estimados por los distintos métodos.

El capítulo 6 se dedica a la verificación o selección del modelo propuesto para lo cual se usan pruebas de bondad de ajuste; se tratan varios criterios, que se relacionan básicamente con la incertidumbre de modelo.

En el capítulo 7 se presenta la aplicación de los conceptos de los capítulos anteriores a los gastos máximos anuales registrados en una estación hidrométrica; los resultados se analizan y se hacen observaciones.

Finalmente, en el capítulo 8, se plantean las conclusiones del desarrollo de este trabajo y se hacen algunas recomendaciones que pretenden ser de utilidad.

2. FUNDAMENTOS DE PROBABILIDAD Y ESTADISTICA

2.1. Estadística

2.1.1. *Población y muestra.*

A la totalidad de posibles observaciones o datos que se obtienen cuando se lleva a cabo un experimento en forma exhaustiva se le llama población.

Una muestra es un grupo de datos obtenidos de la pohlación de una determinada forma; el número de elementos que la componen se llama tamaño de la muestra y generalmente se denota por n.

2.1.2. *Distribuciones de frecuencia.*

Una vez que se ha extraído una muestra de n datos (observaciones), es conveniente ordenarlos de acuerdo a su valor (de mayor a menor o viceversa) cuando pueden cuantificarse. De acuerdo al número de datos y al criterio del ingeniero, es posible determinar intervalos a los que se les llama intervalos de clase, definidos por límites llamados límites de clase.

Los puntos medios de los intervalos se llaman marcas de clase y el número de datos que corresponden al intervalo, frecuencia.

Cuando los intervalos de clase y sus correspondientes frecuencias se presentan en forma ordenada en una tabla, se tiene una distribución de frecuencias. Un histograma es una representación gráfica de la distribución de frecuencias, en la que, en el eje horizontal (abscisas) se representa la característica del fenómeno estudiado por medio de intervalos de clase y en el eje vertical (ordenadas), la frecuencia con que se presenta esa característica. Esto da lugar a una serie de rectángulos cuyas bases son los intervalos de clase y sus alturas, las correspondientes frecuencias.

Si se divide la frecuencia de cada intervalo entre el número total de observaciones (n), se obtienen frecuencias relativas, a cuya distribución se le llama distribución de frecuencia relativas.

Un histograma puede ser de frecuencias o de frecuencias relativas. La suma de las áreas de los rectángulos para un histograma de frecuencias relativas es igual a la unidad.

Si en un histograma de frecuencias se unen los puntos medios de los lados superiores de sus rectángulos, a la línea que resulta de esa unión se le llama polígono de frecuencias. Similarmente, puede obtenerse un polígono de frecuencias relativas a partir del histograma de frecuencias relativas. Los puntos tienen como abscisas las marcas de clase y como ordenadas las frecuencias o las frecuencias relativas.

Si se suman parcial y sucesivamente las frecuencias, la frecuencia total de todos los valores menores que un cierto límite de un intervalo se llama frecuencia acumulada, si se presenta en una tabla, se tiene una distribución de frecuencias acumuladas y si se hace para las frecuencias relativas, la distribución es de frecuencias relativas acumuladas.

Las gráficas correspondientes a la distribución de frecuencias acumuladas y a la de frecuencias relativas acumuladas se construyen por medio de puntos que tienen como abscisas los límites reales de clase y como ordenadas las sumas parciales de las frecuencias acumuladas y las frecuencias relativas acumuladas respectivamente. Las gráficas obtenidas al unir los puntos se llaman polígono de frecuencias acumuladas y polígono de frecuencias relativas acumuladas. El punto inicial del polígono de frecuencias relativas tiene como ordenada 0 y el punto final, 1.

2.1.3. *Medidas de tendencia central, de dispersión y de asimetría.*

De una muestra, pueden calcularse valores representativos de ella llamados estadísticos muestrales o sólo estadísticos. A los correspondientes valores de la población se les llama parámetros poblacionales o sólo parámetros. A continuación se describen algunos importantes estadísticos.

Media. Es el promedio aritmético de todos los datos de la muestra. Si los valores observados se denotan por $X_1, X_2, \ldots, X_n$ y la media por $\bar{X}$, se tiene:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2.1}$$

Mediana. Es el valor para el cual, la mitad de los datos son menores que él y la mitad, mayores. Si el número de datos es par, la mediana es el promedio aritmético de los dos valores medios cuando los datos están ordenados de mayor a menor o viceversa.

Moda. Es el valor con más frecuencia de los datos de la muestra. Es posible que no exista o que no sea única.

Variancia. Es una medida de dispersión de los valores observados; indica como éstos se encuentran agrupados alrededor de la media. Se representa por S^2.

$$S_x^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \tag{2.2}$$

La división entre *n* se hace para obtener un promedio de las desviaciones y de tratar así de evitar la dependencia del tamaño de la muestra.

Desviación estándar. Se define como la raíz cuadrada positiva de la variancia.

$$Sx = \sqrt{Sx^2} \tag{2.3}$$

Coeficiente de variación. Se define como el cociente de la desviación estándar de la muestra entre su media. Resulta útil cuando se comparan las dispersiones de varias muestras. Se representa por V.

$$Vx = \frac{Sx}{\bar{X}} \tag{2.4}$$

Coeficiente de asimetría. Es precisamente una medida de asimetría de la distribución de frecuencias de la muestra. Si el histograma de frecuencias (o el de frecuencias relativas) tiene un eje vertical de simetría, este coeficiente vale cero; se define como:

$$g = \frac{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^3}{S_x^3}$$

(2.5)

La división entre el cubo de la desviación estándar es para hacer adimencional el coeficiente.

La media, la mediana y la moda son llamadas medidas de tendencia central, la variancia, la desviación estándar y el coeficiente de variación, medidas de dispersión; el coeficiente de asimetría, como se mencionó, es una medida de asimetría.

2.1.4. *Momentos de una muestra.*

Dados los valores o datos $X_1, X_2, \ldots, X_n$ con media $\bar{X}$, el momento de orden r con respecto a la media está dado por:

$$Ur = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^r$$

(2.6)

De acuerdo a esta expresión, la variancia (ec. 2.2), representa el momento de segundo orden con respecto a la media. Asímismo, el dividendo en (2.5) es el momento de tercer orden de la muestra.

2.1.5. *Medidas de correlación.*

Cuando se tienen dos muestras del mismo tamaño, es posible establecer --

una correlación entre ellas, es decir una relación uno a uno entre los datos. Por medio de la covariancia y el coeficiente de correlación es posible cuantificar la correlación.

Covariancia. Si se representan por X_i y $\bar{X}$ respectivamente los datos y la media de una muestra y por y_i y $\bar{y}$ los datos y la media de otra, la covariancia de esas dos muestras se expresa como:

$$S_{x,y} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(y_i - \bar{y}) \tag{2.7}$$

Coeficiente de correlación. Al igual que la covariancia correlaciona- - dos muestras de igual tamaño. Siendo S_x y S_y las correspondientes des-- viaciones estándar de las muestras formadas por los valores X_i y y_i, el coeficiente de correlación está dado por:

$$r_{x,y} = \frac{S_{x,y}}{S_x S_y} \tag{2.8}$$

El coeficiente de correlación toma valores en el intervalo $[-1, 1]$. Los valores extremos ± 1 indican que los puntos (X_i, y_i) se encuentran - - exactamente sobre una línea recta cuando estos se dibujan en un sistema de coordenadas X Y. El signo del coeficiente indica la pendiente aproxi mada que siguen los puntos en conjunto.

2.2. Probabilidad

2.2.1. Experimento, evento, axiomas.

Los problemas de ingeniería son en parte también problemas de incertidumbre debido al conocimiento incompleto del fenómeno estudiado y a la falta de información suficiente (datos). Esta incertidumbre es tratada por la teoría de probabilidad.

Definiendo como experimento a cualquier proceso de observación, al conjunto de todos los posibles resultados de un experimento se le llama espacio muestral, el cual consiste en una serie de puntos llamados puntos muestrales, cada uno de las cuales está asociado con un solo resultado definido.

Un evento es un conjunto de puntos muestrales; si tiene un solo elemento se llama evento simple y si tiene dos o más, evento compuesto. Cuando dos eventos no tienen puntos muestrales en común son llamados mutuamente exclusivos.

Un espacio muestral continuo es aquel en el que sus puntos pueden tomar cualquier valor real entre ciertos límites (que pueden ser $-\infty$ y $+\infty$). Un espacio muestral discontinuo o discreto es aquel en el que sus elementos toman sólo valores enteros.

Es importante definir los siguientes axiomas de probabilidad.

1. La probabilidad de un evento A es mayor o igual que 0 y menor o igual que 1.

$$0 \leq P\left[A\right] \leq 1$$

$P\left[A\right]$: significa probabilidad de A (2.9)

2. La probabilidad de un evento seguro S es 1

$$P\left[S\right] = 1 \tag{2.10}$$

3. La probabilidad de un evento, resultado de la unión de dos eventos -- A y B mutuamente exclusivos es igual a la suma de sus respectivas - probabilidades

$$P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) \tag{2.11}$$

3. Del axioma 2, si S es la unión de n eventos simples y estos son -- mutuamente exclusivos:

$$\sum_{i=1}^{n} P\left(E_i\right) = 1 \tag{2.12}$$

Donde E_i son eventos simples asociados a puntos muestrales.

2.2.2 *Probabilidad condicional e independencia.*

Probabilidad condicional. La probabilidad condicional del evento A dado que el evento B ha ocurrido, se define como:

$$P\left(A|B\right) = \frac{P\left[A \cap B\right]}{P\left(B\right)} \qquad P\left(B\right) \neq 0$$

$\left(A \cap B\right)$ es la intersección de los eventos A y B (2.13)

Independencia. Dos eventos A y B son mutuamente independientes si

$$P\left(A|B\right) = P\left(A\right) \tag{2.14}$$

Igualando (2.13) y (2.14)

$$\frac{P\left[A \cap B\right]}{P\left[B\right]} = P\left[A\right] \tag{2.15}$$

Por lo que para eventos independientes se tiene que:

$$P\left[A \cap B\right] = P\left[A\right]\ P\left[B\right] \tag{2.16}$$

y también:

$$P\left[B|A\right] = P\left[B\right] \tag{2.17}$$

En general, si los eventos A,B,...,N son mutuamente independientes, debe cumplirse que:

$$P\left[A \cap B \cap \ldots \cap N\right] = P\left[A\right]\ P\left[B\right] \ldots P\left[N\right] \tag{2.18}$$

2.2.3 *Variables aleatorias.*

En la teoría de probabilidad, resulta de gran utilidad el concepto matemático de variable y de función; en dicha teoría se manejan variables -- aleatorias y funciones de esas variables.

Una variable aleatoria es una variable numérica cuyo valor no puede predecirse con precisión antes de un experimento. Existe un valor de la -- variable aleatoria asociado a cada evento simple del espacio muestral, pero diferentes eventos simples pueden tener el mismo valor asociado de la variable aleatoria, es decir, una variable aleatoria asigna un valor numérico a cada posible resultado.

Cuando el número de valores (finito o infinito) que una variable aleatoria puede tomar es contable (0,1,2,3,....,), a la variable se le llama discreta, de otra forma, variable aleatoria continua.

2.2.4 *Funciones de probabilidad.*

La ley de probabilidad de una variable discreta es llamada función masa de probabilidad (fmp) y tiene la forma:

$$P_X(x) = P(X = x) \tag{2.19}$$

La letra mayúscula X representa una variable aleatoria, la misma letra pero mínuscula (x), el valor que puede tomar.

Para una variable aleatoria continua, su ley de probabilidad es llamada función de densidad de probabilidad (fdp).

Una variable continua puede tomar cualquier valor en los ejes de coordenadas reales por lo que la probabilidad de que la variable aleatoria tome un valor específico es cero, es decir, $P_X(X=x)=0$. Debido a lo anterior se toma un intervalo diferencial de x a $x + dx$ tal que la probabilidad de que X esté en él es $f_X(x)\,dx$, que es la función de densidad de probabilidad.

Como los eventos de diferentes intervalos son mutuamente exclusivos, la probabilidad de que una variable aleatoria tome un valor en un intervalo de longitud finita es la integral de la fdp en ese intervalo. En forma gráfica, si el intervalo está limitado por x_1 y x_2, la probabilidad de que X asuma un valor dentro del intervalo es el área bajo la curva --- fx (x) dx; es decir:

$$P\left[x_1 \leq X \leq x_2\right] = \int_{x_1}^{x_2} fx\ (x)\ dx \tag{2.20}$$

Como se mencionó, a fx (x), se le llama función de densidad de probabilidad y su valor no es propiamente una probabilidad, como su nombre lo indica es sólo una medida de densidad en un punto. En la figura 2.1. Se presenta esta función.

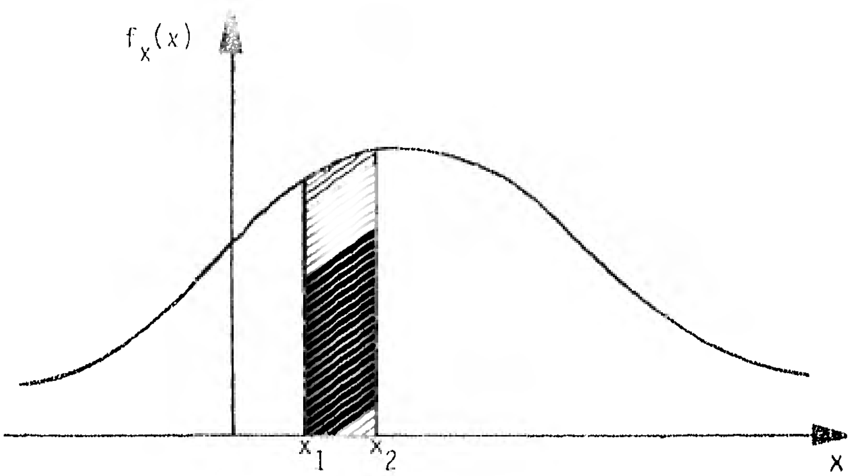


Figura 2.1. Función de densidad de probabilidad.

Una fdp debe cumplir dos condiciones:

$$f_X\ (x) \geq 0 \text{ para toda } x$$

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1 \tag{2.21}$$

Otra forma de describir la distribución de probabilidad de una variable aleatoria la representa la función de distribución acumulada (fda) dada por:

$$F_X(x) = P_X(X \leq x) \tag{2.22}$$

Para una variable discreta:

$$Fx(x) = \sum_{i=1}^{x_0} P_X(x_i) \tag{2.23}$$

y para una variable continua

$$F_X(x) = P(-\infty \leq X \leq x) = \int_{-\infty}^{x} f_X(x)\, dx \tag{2.24}$$

es decir:

$$\frac{dF_X(x)}{dx} = f_X(x)\, dx \tag{2.25}$$

En la figura 2.2 se presenta la función de distribución acumulada.

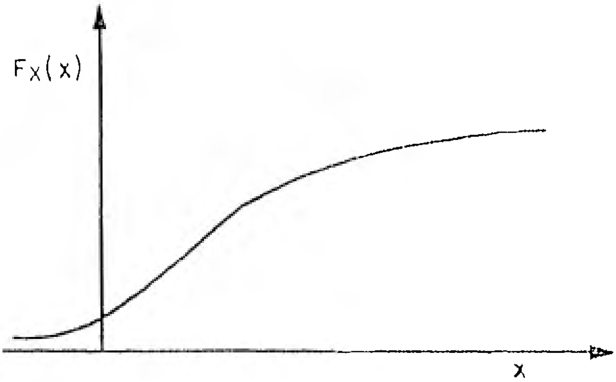


Figura 2.2. Función de distribución acumulada.

2.2.5 *Esperanzas o valores esperados.*

Para una variable aleatoria, dada su función de probabilidad, la esperanza matemática se define como la convergencia a un valor constante de la suma de los productos de la variable por su respectiva probabilidad cuando el número de productos tiende a infinito.

$$E\{X\} = \int_{-\infty}^{\infty} x f_X(x)\, dx \qquad \text{si } x \text{ es contínua} \tag{2.26}$$

y

$$E\{X\} = \lim_{n \to \infty} \sum_{i=1}^{n} x_i P_X(x_i) \quad \text{si } x \text{ es discreta} \tag{2.27}$$

X: es la variable aleatoria

E{X}: es la esperanza, esperanza matemática, valor esperado o media de X.

Para una función de X; g(X):

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x)\ f_X(x)dx \quad \text{si } X \text{ es contínua} \tag{2.28}$$

y

$$E\{g(X)\} = \lim_{n\to\infty} \sum_{i=1}^{n} g(x_i)\ P_X(x_i) \text{ si } X \text{ es discreta} \tag{2.29}$$

2.2.6 *Momentos de una variable aleatoria*

Una función de distribución de probabilidad puede ser definida en términos de sus momentos. Los momentos representan parámetros que tienen significado físico o geométrico al igual que en mecánica de los sólidos.

Simbólicamente, los momentos de una variable aleatoria con respecto al origen se definen como:

$$\mu_X^{(r)'} = \int_{-\infty}^{\infty} x^r\ f_X\ (x)\ dx \quad \text{si } X \text{ es contínua} \tag{2.30}$$

$$\mu_X^{(r)'} = \sum_{i=1}^{n} x_i^r\ \ P_X(x_i) = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ si } X \text{ es discreta} \tag{2.31}$$

Donde:

X: es la variable aleatoria

$\mu_X^{(r)'}$: es el momento de orden r de X con respecto al origen

r: es el orden del momento: 1, 2, 3,...

El momento de orden 1, de primer orden o primer momento con respecto al origen es la media o promedio aritmético de X, se denotará para m_X.

Es igual al valor esperado de X expresado por (2.26) ó (2.27) según el caso. Por lo tanto:

$$\mu_x^{(1)'} = E\{X\} = mx = \int_{-\infty}^{\infty} X \, f_X(x)\, dx \quad \text{si } x \text{ es continua}$$

(2.32)

$$\mu_x^{(1)'} = E\{X\} = mx = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{si } x \text{ es discreta}$$

(2.33)

Los momentos con respecto a la media se llaman momentos centrales y se definen como los momentos de una función de la variable aleatoria: $g(x) = X - \mu_x^{(1)'} = X - mx$:

$$\mu^{(r)} = \int_{-\infty}^{\infty} (X - mx)^r \, f_X(x)\, dx \quad \text{si } X \text{ es continua}$$

(2.34)

$$\mu_x^{(r)} = \frac{1}{n}\sum_{\substack{i=1 \\ n\to\infty}}^{n} (x_i - m_X)^r \quad \text{si } X \text{ es discreta}$$

(2.35)

$\mu_X^{(r)}$: es el momento de orden r respecto a la media ó r-ésimo mo-- mento central.

Al igual que el primer momento respecto al origen, el segundo momento central es de gran impotancia ya que representa la variancia de la población y se representa por σ^2:

Sustituyendo $(x - m_X)$ por $g(x)$ en (2.28) y (2.29), se tiene:

$$\mu_x^{(2)} = \sigma_X^2 = \text{Var}\{X\} = E\{X - m_X\}^2 = \int_{-\infty}^{\infty} (X - m_X)^2 f_X(x)\, dx \quad \text{si } X \text{ es continua}$$

(2.36)

$$\mu_x^{(2)} = \sigma_X^2 = \text{Var}\{X\} = E\{X - m_X\}^2 = \sum_{\substack{i=1 \\ n\to\infty}}^{n} (x_i - m_X)^2 \quad \text{si } X \text{ es discreta}$$

(2.37)

$E\left[X - m_X\right]$: es el valor esperado de la función $g(x) = X - m_X$

La raíz cuadrada positiva de la variancia en (2.36) y (2.37) es llamada desviación estándar.

$$\sigma_X = +\sqrt{\sigma_X^2} \qquad (2.38)$$

El coeficiente de variación de una variable aleatoria formaliza, al igual que para una muestra, la comparación entre la desviación estándar y la media. Se define por:

$$V_X = \frac{\sigma_X}{m_X} \qquad (2.39)$$

El tercer momento central es necesario para definir el coeficiente de asimetría γ_1:

$$\gamma_1 = \frac{\mu_X^{(3)}}{\sigma_X^3} \qquad (2.40)$$

Además de γ_1, en función de los momentos centrales pueden obtenerse:

$$\left.\begin{aligned} \gamma_2 &= \frac{\mu_X^{(4)}}{{\mu_X^{(2)}}^2} \\ \gamma_3 &= \frac{\mu_X^{(5)}}{{\mu_X^{(2)}}^{5/2}} \\ \gamma_4 &= \frac{\mu_X^{(6)}}{{\mu_X^{(2)}}^3} \end{aligned}\right\} \qquad (2.41)$$

(Ref. 9)

γ_2 es llamado coeficiente de curtosis y al igual que γ_3 y γ_4, llega a resultar de utilidad en aplicaciones de probabilidad y estadística.

2.2.7 *Probabilidad conjunta*

Cuando dos o más variables aleatorias son consideradas en forma simultánea, su comportamiento conjunto sigue una ley de probabilidad conjunta que al igual que para el caso de una sola variable, puede ser descrito por una función de densidad de probabilidad, una función masa de probabilidad o por funciones de distribución acumulada.

Para variables discretas:

$$Px,y\ (x,y) = P\left[(X = x) \cap (Y = y)\right] \tag{2.42}$$

$$Fx,y\ (x,y) = P\left[(X \leq x) \cap P\,(Y \leq y)\right] = \sum_{x_i \leq x}\ \sum_{y_i \leq y} P_Y\ (x_i, y_i) \tag{2.43}$$

x,y : son las variables aleatorias consideradas en forma conjunta.

En forma análoga, para variables continuas:

$$P\left[x_1 \leq X \leq x_2\right] \text{ y } P\left[y_1 \leq Y \leq y_2\right] = \int_{x_1}^{x_2}\int_{y_1}^{y_2} fx,y\ (x,y)\ dx\ dy \tag{2.44}$$

$$Fx\ y\ (x,y) = P\left[(X \leq x) \text{ y } (Y \leq y)\right] = P\left[(-\infty \leq X \leq x) \text{ y} (-\infty \leq Y \leq y)\right]$$

$$= \int_{-\infty}^{x}\int_{-\infty}^{y} fx,y\ (x,y)\ dy\ dx \tag{2.45}$$

En forma análoga a (2.25):

$$fx,y\ \ (x\ y) = \frac{\partial^2}{\partial x\ \ \partial y} Fx,y\ (x,y) \tag{2.46}$$

Si las variables aleatorias X y Y son independientes, es decir, si los eventos relacionados con X son independientes de los eventos relacionados con Y, se tiene (Ref 2):

$$fx|y\ (x,y) = f_x\ (x)$$
$$fy|x\ (y,x) = f_y\ (y)$$

$$f_{x,y}(x,y) = f_x(x)\, f_y(y)$$
$$F_{x,y}(x,y) = F_x(x)\, F_y(y)$$
$$F_{x|y}(x,y) = F_x(x)$$

(2.47)

También es posible definir los momentos de una función de probabilidad conjunta. Aquí se tratará sólo uno llamado covariancia, que es una medida de correlación entre X y Y, dada por:

$$\sigma_{x,y} = \mathrm{Cov}(X,Y) = E(X - m_x)(Y - m_y)$$
$$= \iint_{-\infty}^{\infty} (x - m_x)(y - m_y)\, f_{x,y}(x,y)\, dx\, dy$$

para X,Y continuas.

(2.48)

$$\sigma_{x,y} = \mathrm{Cov}(x,y) = \sum_{\substack{\text{todas} \\ \text{las } x}} \sum_{\substack{\text{todas} \\ \text{las } y}} (x - m_x)(y - m_y)\, P_{x,y}(x,y)$$

para x,y discretas

(2.49)

Otra medida de la correlación, la representa el coeficiente de correlación de dos variables aleatorias.

$$\rho_{x,y} = \frac{\mathrm{Cov}[X,Y]}{\sigma x \ \sigma y}$$

(2.50)

Donde:

$\rho_{x,y}$: es el coeficiente de correlación

$\sigma x, \sigma y$: son las respectivas variancias de las variables aleatorias X,Y

Al igual que el coeficiente de correlación para dos muestras, el coeficiente de correlación de dos variables aleatorias, tiene como límites los valores $\pm$ 1 y su magnitud el mismo significado.

Aunque, aparentemente, un histograma y un polígono de frecuencias son si milares a una función de densidad de probabilidad y a una función de distribución acumulada respectivamente, los primeros se refieren a datos empíricos observados y los segundos, a modelos de probabilidad teóricos.

A partir de los datos empíricos, el ingeniero puede proponer un modelo matemático como representativo del fenómeno estudiado Generalmente es difícil encontrar un modelo exacto del fenómeno por lo que pueden proponerse varias funciones de probabilidad y de acuerdo a ciertos criterios y pruebas, que se estudiarán posteriormente, elegir un modelo representativo con el cual puedan hacerse "predicciones".

En resumen, el ingeniero debe adoptar un modelo matemático y posteriormente probar su validez con las observaciones hechas o datos disponibles, para lo cual deben fijarse ciertas reglas y así aceptar o rechazar un modelo particular ó elegir uno de entre varios posibles

2.2.8 *Período de retorno, riesgo y valores extremos.*

Período de retorno. El período de retorno T de eventos hidrológicos se define como el tiempo promedio entre dos ocurrencias de un evento mayor que una determinada magnitud (excedencias) Esto no significa que una excedencia ocurra cada T años, sino que el tiempo promedio entre dos excedencias es de T años El concepto de período de retorno también es aplicable a eventos de menor magnitud (mínimos).

Si el período de retorno se refiere a eventos de mayor magnitud que un cierto valor, puede relacionarse con una probabilidad de excedencia. Si el tiempo promedio entre dos excedencias es T, la probabilidad de que el evento (excedencia) ocurra en cualquier año es 1/T Así:

$$T = \frac{1}{p} \; ; \quad p = p\left[X > x\right] \tag{2.51}$$

Siendo p la probabilidad de que el evento sea mayor que un cierto valor (excedencia).

Cuando se realizan n observaciones, la probabilidad observada de cada -- evento es:

$$p = \frac{i}{n} \tag{2.52}$$

i: es el rango o número progresivo (1,2,3,...) del evento cuando han sido ordenados de mayor a menor.

En hidrología, en vez de (2.52) se usa:

$$p = \frac{i}{n+1} \tag{2.53}$$

Esto se debe a que cundo se dibujan las observaciones (puntos) y las probabilidades acumuladas en papeles especiales, la escala correspondiente a las probabilidades no es de 0 a 1. Por otro lado $i/(n+1)$ es un estimador insesgado de la probabilidad observada para cada punto. La propiedad de insesgo será estudiada en capítulo 4.

Sustituyendo (2.53) en (2.51), se tiene:

$$T = \frac{n+1}{i} \tag{2.54}$$

La probabilidad de no excedencia está dada por (2.22)

$$F_x(x) = p\left[X \le x\right]$$

de donde:

$$P\ (X > x) = 1 - P(X \leq x) = 1 - Fx\ (x) \tag{2.55}$$

Sustiruyendo en (2.51)

$$Tx\ (x) = \frac{1}{1 - Fx\ (x)} \tag{2.56}$$

ó

$$F_x\ (x) = 1 - \frac{1}{Tx(x)} \tag{2.57}$$

Riesgo. Se define como la probabilidad de que el evento de diseño sea excedido al menos una vez durante la vida útil de diseño.

La probabilidad de no excedencia está dada por (2.22)

$$Fx\ (x) = P\left[X \leq x\right]$$

Si los eventos son independientes, la probabilidad de no excedencia en cualquiera de los n años de vida útil es:

$$P\left[X_1 \leq x\right]\ P\left[X_2 \leq x\right] \cdots P\left[X_n \leq x\right] = \left[Fx\ (x)\right]^n \tag{2.58}$$

Por lo tanto, la probabilidad de excedencia es:

$$P(\ X > x\ ) = 1 - Fx\ (x)^n \tag{2.59}$$

A esta probabilidad se le llama riesgo, se representa por R. Sustituyendo (2.57) en (2.59) se tiene finalmente:

$$R = 1 - \left[1 - \frac{1}{Tx\ (x)}\right]^n \tag{2.60}$$

Generalmente el riesgo es dado en porciento.

Valores extremos. Los valores extremos de una serie de variables aleatorias son de gran interés para llevar a cabo el diseño de obras hidráulicas. Los valores extremos son también variables aleatorias. La distribución de los valores extremos depende del tamaño de la muestra y de la distribución que sigue la población.

Si una muestra de tamaño n tiene como elementos $x_1, x_2, \ldots, x_n$, siendo Y el valor máximo, $F_y(y) = P\left[Y \leq y\right]$, $Fx_i(x) = P\left[X_i \leq x\right]$ y además $f_y(y)$ y $f_{x_i}(x)$ son las respectivas fdp de Y y de x_i:

$$Fy(y) = P\left[Y \leq y\right] = P\left[\underset{\text{todas}}{x_i \leq y}\right] \tag{2.61}$$

si las X_i son variables independientes:

$$Fy(y) = Fx_1(y)\; Fx_2(y) \ldots Fx_n(y) = \left[Fx(y)\right]^n \tag{2.62}$$

Aplicando (2.25):

$$Fy(y) = \frac{d\,Fy(y)}{dy} = n\left[Fx(y)\right]^{n-1} f_x(y) \tag{2.63}$$

De (2.63) puede observarse que $f_y(y)$ depende del tamaño de la muestra n y de la distribución de la población de donde se obtiene la muestra.

2.2.9 *Muestreo y distribuciones muestrales*

Muestreo. Una condición que debe tener una muestra para que a partir de ella puedan hacerse "predicciones" de otros valores de la población, es la de ser representativa de ésta, es decir, que sin contener a todos los elementos de la población, sea del tamaño suficiente, de tal forma que contenga todas las variedades de la característica estudiada; para satis facer ésto, todos los elementos deben tener la misma oportunidad de ser seleccionados y además ser independientes. Esto último significa que, al extraer un elemento, no debe afectarse la oportunidad de selección de los demás. Una muestra representativa también es llamada muestra -- aleatoria.

Al proceso de obtención de muestras de una población se le llama muestreo. Ya que las muestras deben ser representativas, es de especial interés el muestreo aleatorio. En términos matemáticos, el muestreo aleatorio se define como aquel en que las n observaciones $x_1, x_2, \ldots x_n$ de un experimento son independientes y siguen la misma función de probabilidad.

El muestreo puede ser con reemplazo y sin reemplazo Es con reemplazo cuando cada elemento de la población puede ser seleccionado una o más veces; si sólo puede ser seleccionado una vez, es sin reemplazo. Por lo tanto, cuando se extrae una muestra con reemplazo, sus observaciones son independientes.

El tamaño de una población puede ser finito o infinito dependiendo del experimento realizado. Cuando el muestreo que se hace de una población finita, se lleva a cabo con reemplazamiento puede considerarse esta última como de tamaño infinito.

El muestreo aleatorio sólo puede obtenerse si el muestreo es con reemplazamiento en poblaciones finitas o si se hace de poblaciones infinitas. En la práctica, el muestreo de poblaciones muy grandes puede considerarse como aleatorio

Distribuciones muestrales. Si se toman de una población, varias muestras de igual tamaño, al compararlas entre sí, se obtienen series diferentes de valores y por consiguiente, diferentes histogramas; esto se debe a la naturaleza aleatoria del fenómeno estudiado y origina por lo tanto, estadísticos aleatorios. Si a cada muestra se calcula el mismo estadístico, la distribución que éste sigue es llamada distribución muestral del estadístico y puede estudiarse como la distribución de una variable aleatoria.

3. FUNCIONES DE DISTRIBUCION DE PROBABILIDAD

Como se mencionó en el capítulo anterior, para poder llevar a cabo predicciones a partir de una muestra, debe adoptarse un modelo representativo del fenómeno bajo estudio. Los modelos matemáticos o, brevemente, modelos, para el caso de este trabajo, son leyes o funciones de probabilidad, también llamadas funciones de distribución, distribuciones de probabilidad o solamente distribuciones.

En el presente capítulo se describen varios modelos probabilísticos. En el capítulo 6 se estudiarán algunos criterios para seleccionar el "mejor" modelo.

3.1 Distribución normal

Es la distribución de probabilidad más utilizada en el campo de la Probabilidad y Estadística. La representación gráfica de su función de densidad de probabilidad (fdp) es una curva simétrica en forma de campana llamada Curva normal o Gaussiana (fig. 3.1). Con frecuencia otras distribuciones pueden aproximarse a la curva normal. Muchos métodos estadísticos parten de una consideración de normalidad.

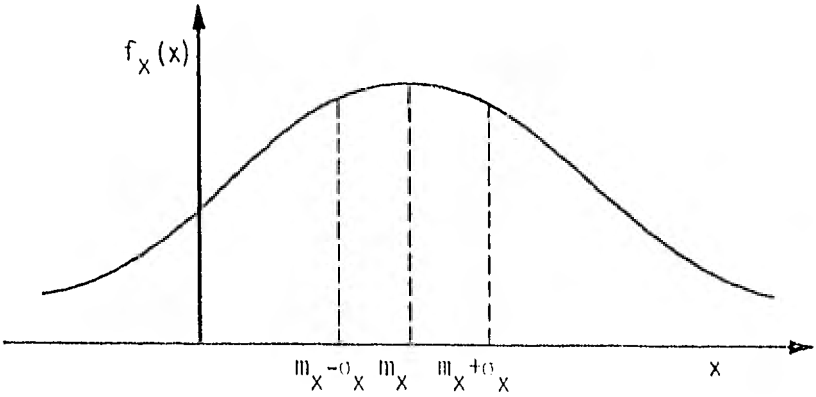


Figura 3.1. fdp Normal

3.1.1 *Función de densidad, función de distribución acumulada y parámetros.*

La fdp de la Distribución Normal está dada por la siguiente expresión:

$$f_X(x)= \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{x-m_X}{\sigma_X}\right)^2 \right\} \qquad -\infty < x < \infty \tag{3.1}$$

Donde:

X: es la variable aleatoria y x es su valor

m_X, σ_X : son respectivamente la media y la desviación estándar de la variable X

π = 3.14159...

$\exp[\ ] = e^{()}$: es la base de los logaritmos naturales.

La variable:

$$Z = \frac{X - m_X}{\sigma_X} \tag{3.2}$$

es llamada variable normal estandarizada, tiene media cero y desviación estándar igual a uno, por lo que la expresión (3.1) en función de Z, se escribe:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right) \tag{3.3}$$

Los parámetros de la Distribución Normal son de media m y la desviación estándar σ por lo que puede representarse por N (m, σ). Cuando se usa la variable normal estandarizada, se tiene N(0,1).

La probabilidad de que una variable aleatoria normal esté en un intervalo está dada por la integral de la fdp para ese intervalo. También puede obtenerse a partir de las probabilidades acumuladas. No existe una expresión sencilla de la función de distribución acumulada (fda) para la distribución normal pero su evaluación numérica se encuentra en la tabla 1 del apéndice .

Considerando (2.22)

$$F_X(x)=P(X\leq x) = P\left[Z \leq \frac{x-m_X}{\sigma_X}\right] = F_Z\left[\frac{x-m_X}{\sigma_X}\right]= F_Z(z)$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} \exp(-z^2/2)dz \qquad -\infty \leq z \leq \infty \tag{3.4}$$

Las letras mayúsculas X, Z se refieren a variables aleatorias y las minúsculas x,z a sus correspondientes valores.

En la tabla 1 del apéndice se presenta la expresión (3.4) sólo para $z \geq 0$ debido a la simetría de la fdp. Haciendo la siguiente consideración, pueden obtenerse valores de $F_Z(-z)$:

$$F_Z(-z) = 1 - F_Z(z) \tag{3.5}$$

Para encontrar la probabilidad de que una variable aleatoria normal esté en el intervalo $[X_1, X_2]$ con ayuda de la tabla 1, debe obtenerse la probabilidad acumulada hasta X_2 y la probabilidad acumulada hasta X_1, después, obtener la diferencia; siendo ésta la probabilidad en el intervalo; la expresión es:

$$P(X_1 \leq X \leq X_2) = F_X(X_2) - F_X(X_1) \tag{3.6}$$

3.1.2 *Momentos y su relación con los parámetros.*

Debido a la simetría de la Distribución Normal con respecto a su media, todos los momentos centrales de orden impar (de esta distribución) valen cero.

Los momentos de orden par deben estar relacionados sólo con la media y con la desviación estándar, ya que estos dos momentos son los parámetros de todas las Distribuciones Normales.

Los momentos centrales de orden par, están dados por (Ref. 2):

$$\mu_X^{(r)} = E\{(X-m_X)^r\} = \int_{-\infty}^{\infty} (x-m_X)^r f_X(x)\, dx$$

$$= \frac{r}{2^{r/2}\,(r/2)!}\,\sigma_X^r \quad ; \; r = 2,4,6,\ldots, \tag{3.7}$$

Donde:

$E\{\,\}$: es la esperanza o valor esperado definido por (2.28)

m_X: está dada por (2.32)

σ_X : está dada por (2.38)

El coeficiente de asimetría vale cero para la Distribución Normal.

3.2 Distribución Lognormal (dos y tres parámetros)

Esta distribución considera el caso en que la variable dependiente se expresa como el producto de dos o más variables aleatorias; si se intentara hacer un estudio por separado de estas últimas variables, resultaría más complicado y no se justificaría.

Dada:

$$y = x_0 \cdot x_1 \; \ldots \; x_n \tag{3.8}$$

Si se toman logaritmos naturales, se tiene:

$$\mathrm{Ln}\, y = \mathrm{Ln}\, x_0 + \mathrm{Ln}\, x_1 + \ldots + \mathrm{Ln}\, x_n \tag{3.9}$$

Si los logaritmos naturales de una variable aleatoria, se distribuyen normalmente, la distribución se llama lognormal.

En:

$$X = \mathrm{Ln}\, y \tag{3.10}$$

Si X se distribuye normalmente, forma una Distribución Lognormal equivalente a:

$$y = e^X \tag{3.11}$$

3.2.1 *Función de densidad, función de distribución acumulada y parámetros.*

La función de densidad de probabilidad para (3.10) está dada por:

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{x - m_X}{\sigma_X} \right)^2 \right\} \qquad -\infty \leq x \leq \infty \tag{3.12}$$

y en función de y tomando en cuenta (3.10) ó (3.11)

$$f_y(y) = \frac{1}{y\, \sigma_L \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{\mathrm{Ln}\, y - m_L}{\sigma_L} \right)^2 \right\} \qquad y \geq 0 \tag{3.13}$$

y es una variable lognormalmente distribuída ya que su logaritmo natural es normalmente distribuído.

m_L, σ_L : son respectivamente la media y la desviación estándar de los logaritmos naturales de y. El subíndice L se refiere a la Distribución Lognormal.

Las expresiones para m_L y σ_L son:

$$m_L = m_X = m_{Ln\ y} = E\,[X] = E\,[Ln\ y] \tag{3.14}$$

$$\sigma_L = \sigma_X = \sigma_{Ln\ y} = E\,(X - m_X)$$

Los parámetros de esta distribución, llamada en forma más específica *Distribución Lognormal de dos parámetros*, son m_L y σ_L. Se representa por L $(m_L\ \sigma_L)$.

Al igual que para la Distribución Normal, puede usarse la tabla de áreas bajo la curva normal estandarizada para el cálculo de probabilidades.

La variable normal estandarizada queda dada por:

$$z = \frac{Ln\ y - m_L}{\sigma_L} \tag{3.15}$$

Con los valores de z pueden obtenerse las probabilidades acumuladas y probabilidades en un intervalo en forma semejante que para la Distribución Normal.

3.2.2 Momentos y su relación con los parámetros.

Los momentos de la variable aleatoria Y están dados por:

$$\mu_y^{(r)'} = \int_0^\infty y^r \; f_y(y)\, dy \tag{3.16}$$

de cuya solución, se obtiene:

$$m_y = \exp\left(m_L + \frac{1}{2}\,\sigma_L^2\right) \tag{3.17}$$

*

$$\sigma_y^2 = m_y^2 \;\{\exp(\sigma_L^2) - 1\} \tag{3.18}$$

3.2.3 *Distribución Lognormal de tres parámetros.*

Puede introducirse un tercer parámetro en esta distribución. Esto es apropiado si existe un límite inferior (valor límite mínimo) que no es cero. Este parámetro se representa por a_L y afecta a la variable de la distribución restándose de ella: $(y-a_L)$. sustituyendo $y-a_L$ en lugar de y en (3.13):

$$f_y(y) = \frac{1}{(y-a_L)\,\sigma_L\sqrt{2\pi}}\;\exp\left\{-\frac{1}{2}\left(\frac{\operatorname{Ln}(y-a_L) - m_L}{\sigma_L}\right)^2\right\} \tag{3.19}$$

$$(y-a_L) > 0$$

Cuando se introduce a_L, la distribución se representa por L (m_L, σ_L, a_L) y es llamada Distribución Lognormal de tres parámetros; en la figura 3.2 se presenta su función de densidad de probabilidad (para el caso de dos parámetros, la curva se inicia en el origen).

Puede suceder que el límite inferior sea conocido; si es así, a_L se determina a partir de ese valor conocido; en este caso, se usan las

* Ref. 2

expresiones vistas para la distribución de dos parámetros.

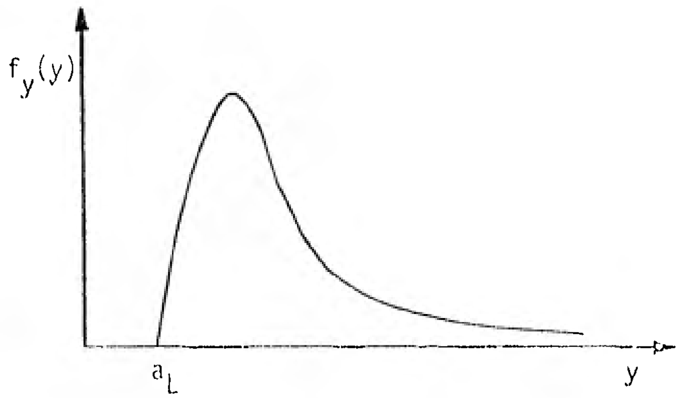


Figura 3.2. fdp Lognormal (3 parámetros)

Cuando no se conoce a priori el límite inferior, que es lo más frecuente, no pueden usarse en forma directa las expresiones correspondientes a la distribución de dos parámetros. La variancia de $(y-a_L)$ es la misma que para y pero la media cambia, lo cual implica que cambie también el coeficiente de variación: $V_y \neq V_{(y-a_L)}$.

Si a_L tiene un valor negativo, y puede tomar valores negativos siempre que se satisfaga que $(y-a_L)$ sea mayor que cero para toda y.

Se sabe de (2.39) que el coeficiente de variación está dado por:

$$V_y = \frac{\sigma_y}{m_y} \tag{3.20}$$

Para este caso, (3.20) corresponde a la Distribución Lognormal de tres parámetros.

Si a y se le resta a_L, se tiene:

$$V_{(y-a_L)} = \frac{\sigma_{(y-a_L)}}{m_{(y-a_L)}} \tag{3.21}$$

que corresponde a la distribución de dos parámetros.

Como la variancia es independiente de a_L:

$$\sigma_y = \sigma_{(y-a_L)} \tag{3.22}$$

Además:

$$m_{(y-a_L)} = m_y - a_L \tag{3.23}$$

La expresión (3.23) representa la media de una distribución de dos parámetros en función de la media de una distribución de tres parámetros. Sustituyendo $(y-a_L)$ por y en (3.17), se tiene:

$$m_{(y-a_L)} = \exp\left(m_L + \frac{1}{2}\sigma_L^2\right) \tag{3.24}$$

Sustituyendo (3.24) en (3.23) y despejando m_y:

$$m_y = \exp\left(m_L + \frac{1}{2}\sigma_L^2\right) + a_L \tag{3.25}$$

<u>Obtención del parámetro a_L</u>:

De (3.23):

$$a_L = m_y - m_{(y-a_L)} \tag{3.26}$$

De (3.20) y (3.21), considerando (3.22):

$$m_y = \frac{\sigma_y}{V_y} \tag{3.27}$$

$$m_{(y-a_L)} = \frac{\sigma_y}{V_{(y-a_L)}} \tag{3.28}$$

Sustituyendo (3.27) y (3.28) en (3.26):

$$a_L = \frac{\sigma_y}{V_y} - \frac{\sigma_y}{V_{(y-a_L)}} \tag{3.29}$$

Considerando de (3.20) que:

$$m_y = \frac{\sigma_y}{V_y} \qquad y \qquad \sigma_y = m_y V_Y$$

resulta:

$$a_L = m_y \left[1 - \frac{V_y}{V_{(y-a_L)}} \right] \tag{3.30}$$

Haciendo el cociente de (3.20) considerando (3.17) y (3.18) se llega a:

$$V_y = \frac{\sigma_y}{m_y} = \frac{m_y \{\exp(\sigma_L^2) - 1\}^{\frac{1}{2}}}{m_y} = \{\exp(\sigma_L^2) - 1\}^{\frac{1}{2}} \tag{3.31}$$

La expresión (3.31) se refiere a la Distribución Lognormal de tres parámetros. Para el caso de dos parámetros: $m_{L(y-a_L)}$ y $\sigma^2_{L(y-a_L)}$ se expresan en función de $m_{(y-a_L)}$ y $V_{(y-a_L)}$ como:

$$m_{L(y-a_L)} = \frac{1}{2} \operatorname{Ln} \left(\frac{m^2_{(y-a_L)}}{V^2_{(y-a_L)}+1} \right) = \frac{1}{2} \operatorname{Ln} \left[\frac{m^4_{(y-a_L)}}{m^2_{(y-a_L)} + \sigma^2_{(y-a_L)}} \right] * \tag{3.32}$$

---

* Referencia 3

$$\sigma^2_{L(y-a_L)} = \mathrm{Ln}\left[V^2_{(y-a_L)} + 1\right] = \mathrm{Ln}\left[\frac{\sigma^2_y + m^2_{(y-a_L)}}{m^2_{(y-a_L)}}\right] * \qquad (3.33)$$

A partir de (3.32) y (3.33) el coeficiente de asimetría γ_1, para una Distribución Lognormal de dos parámetros, puede expresarse como:

$$\gamma_1 = V^3_{(y-a_L)} + 3\, V_{(y-a_L)} \qquad (3.34)$$

Para calcular $V_{(y-a_L)}$ a partir de (3.34) debe resolverse una ecuación cúbica cuya solución es:

$$V_{(y-a_L)} = \left[\frac{\sqrt{\gamma_1^2 + 4} + \gamma_1}{2}\right]^{\frac{1}{3}} + \left[\frac{-\sqrt{\gamma_1^2 + 4} + \gamma_1}{2}\right]^{\frac{1}{3}} \qquad (3.35)$$

Sustituyendo el valor obtenido de (3.35) en (3.30) se obtiene el valor del parámetro a_L.

El 2o. y el 3er momentos dados por (3.16) son independientes de a_L.

La variable normal estandarizada z toma la forma:

$$z = \frac{\mathrm{Ln}\,(y-a_L) - m_L}{\sigma_L}$$

3.3 Distribución Gumbel

Es común en la ingeniería civil, requerir de un valor máximo o de un

*Referencia 3

valor mínimo para poder llevar a cabo el diseño de una obra, por ejemplo: conocer el valor aproximado de la avenida máxima probable de un río du-- rante la vida útil de una obra de control. La capacidad de un sistema puede depender sólo de valores extremos.

Una distribución puede restringirse al valor extremo de interés. Si se quiere conocer el límite de esa distribución para un valor máximo, disponiendo de n máximos (por ejemplo: avenidas máximas anuales), obviamente el valor extremo de interés es el máximo. Considerando que los valores máximos siguen una curva exponencial, la Distribución Gumbel, tam-- bién llamada de valores extremos tipo I, frecuentemente resulta ser una buena aproximación de la distribución de las avenidas máximas anuales de un río.

3.3.1 *Función de densidad, función de distribución acumulada y parámetros.*

La función de distribución acumulada Gumbel está dada por:

$$F_y(y) = \exp\{-\exp[-\alpha(y-\beta)]\} \quad -\infty < y < \infty \tag{3.36}$$

y la función de densidad de probabilidad (figura 3.3) por:

$$f_y(y) = \alpha \exp\{-\alpha(y-\beta) - \exp[-\alpha(y-\beta)]\} \tag{3.37}$$

Donde:

Y: es la variable aleatoria y y correspondiente valor.

α y β: son los parámetros que le dan forma a esta distribución.

α es una medida de dispersión. Se representa por G (α, β).

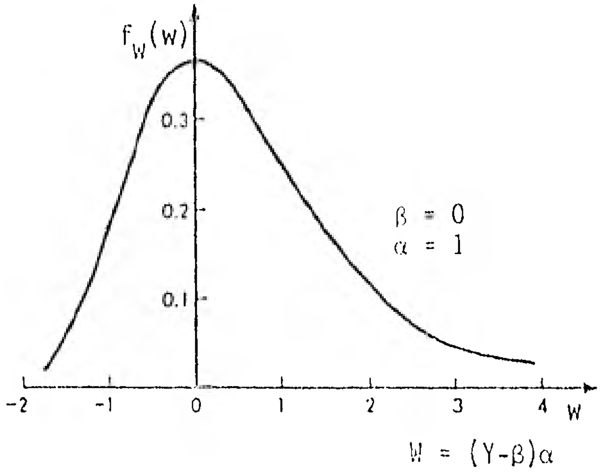


Figura 3.3. fdp Gumbel

3.2.2 *Momentos y su relación con los parámetros.*

Los momentos se obtienen a partir de la integración de (3.37) y son los siguientes:

$$m_y = \beta + \frac{\gamma}{\alpha} \qquad (3.38)$$

$$\sigma_y^2 = \frac{\pi^2}{6\alpha^2} \qquad (3.39)$$

m_y y σ_y^2 son el primer momento con respecto al origen y el segundo momento central respectivamente definidos por (2.32) y (2.36)

γ: es la constante de Euler; γ = 0.5772156...

π: 3.14159...

El coeficiente de asimetría de esta distribución es independiente de α y de β y además es positivo; está dado por:

$$\gamma_1 = \frac{\mu_X^{(3)}}{\sigma_X^3} = 1.1396 \qquad (3.40)$$

Ejemplo 3.1 Se presenta enseguida, la variación en los "parámetros" de la Distribución Gumbel al aumentar el tamaño de la "muestra"

Dados los parámetros de la Distribución Gumbel:

$\alpha = 0.004$

$\beta = 505.0000$

se calculan 100 eventos simples a partir de (3.36), para lo cual es necesario igualar esta expresión con (2.57):

$$1 - \frac{1}{T} = \exp\{ - \exp[- \alpha (y-\beta)] \} \qquad (3.a)$$

T es el periodo de retorno definido por (2.54)

$$T = \frac{n+1}{i}$$

donde n es el número de datos (en este caso 100) e i es el rango definido en 2.28.

Despejando y de (3.1) se llega a:

$$y = \beta - \frac{1}{\alpha} \text{Ln Ln } \frac{T}{T-1} \qquad (3.b)$$

Con esta expresión se calculan los 100 eventos simples y, que quedan ordenados de mayor a menor.

Una vez obtenidos los 100 eventos, se obtienen 100 números "aleatorios" entre 1 y 100 (enteros) de tal forma que ninguno se repita (sin reemplazo).

De acuerdo a esos números aleatorios, se toman los primeros veinte eventos (si por ejemplo, el primer número aleatorio es 19, se toma el evento que corresponde a ese número, cuando están ordenados de mayor a menor, lo mismo se hace para el 2o., 3o., ..., 20° número), se obtienen los parámetros α_1 y β_1, también se calcula:

$$ES = \frac{\sum_{i=1}^{n} (w_i - y_i)^2}{n_1 - 2} \qquad (3.c)$$

donde:

W_i: son los eventos estimados aplicando en forma similar (3.b)

y_i: son los eventos seleccionados aleatoriamente de los 100 valores, (ordenados de mayor a menor)

n_1: es el número de los primeros y_i (en este caso 20)

α_1 y β_1: son obtenidos de los n_1 primeros y_i

Lo mismo se hace para los primeros 21, 22,...,100 eventos; es decir, $n_1 = 20, 21, \ldots 100$.

Se hicieron cuatro ensayos. En las cuatro veces en que se "revolvieron" aleatoriamente los eventos simples, se observó lo siguiente:

Aproximadamente para $n_1 > 60$, los valores de los parámetros α_1, β_1 tienden hacia los valores dados inicialmente para calcular los 100 eventos. ($\alpha = 0.004$, $\beta = 505$)

Cuando $n_1 < 30$ (aproximadamente), los valores de α_1 y β_1 varían mucho entre sí de un ensayo a otro.

Para $n_1 = 20$

$\alpha_1 = 0.003274$ $\qquad$ $\beta_1 = 577.2930$

$\alpha_1 = 0.004235$ $\qquad$ $\beta_1 = 592.3791$

$\alpha_1 = 0.005333$ $\qquad$ $\beta_1 = 498.8725$

α_1 = 0.004118 $\qquad\qquad$ β_1 = 633.7997

Para n_1 = 60:

α_1 = 0.004090 $\qquad\qquad$ β_1 = 511.0107

α_1 = 0.004014 $\qquad\qquad$ β_1 = 532.5116

α_1 = 0.004186 $\qquad\qquad$ β_1 = 479.9319

α_1 = 0.004207 $\qquad\qquad$ β_1 = 522.6406

En cuanto a ES calculado por medio de (3.c) se observa que en los cuatro ensayos decrece cuando n_1 aumenta.

En la figura 3.4 se muestra la variación de α_1 y β_1 con respecto al tamaño n_1 y en la figura 3.5 se observa como varía ES (ec.3.c) también con respecto a n_1 para los correspondientes ensayos.

Como se mencionó, sólo se realizaron cuatro ensayos, siendo necesarios más para reafirmar las observaciones hechas.

3.3a. Distribución Doble Gumbel*

En los registros de gastos máximos anuales, en muchas ocasiones pueden observarse dos grupos de avenidas con características diferentes.

Si se considera que los gastos máximos anuales se originan por distintos procesos, es decir, si existen dos poblaciones, es difícil ajustar el registro a una sola distribución de probabilidad por ejemplo la Gumbel; sin embargo, es razonable suponer que cada grupo de avenidas por separado sigue una distribución Gumbel.

---

* Obtenida de la Referencia 12

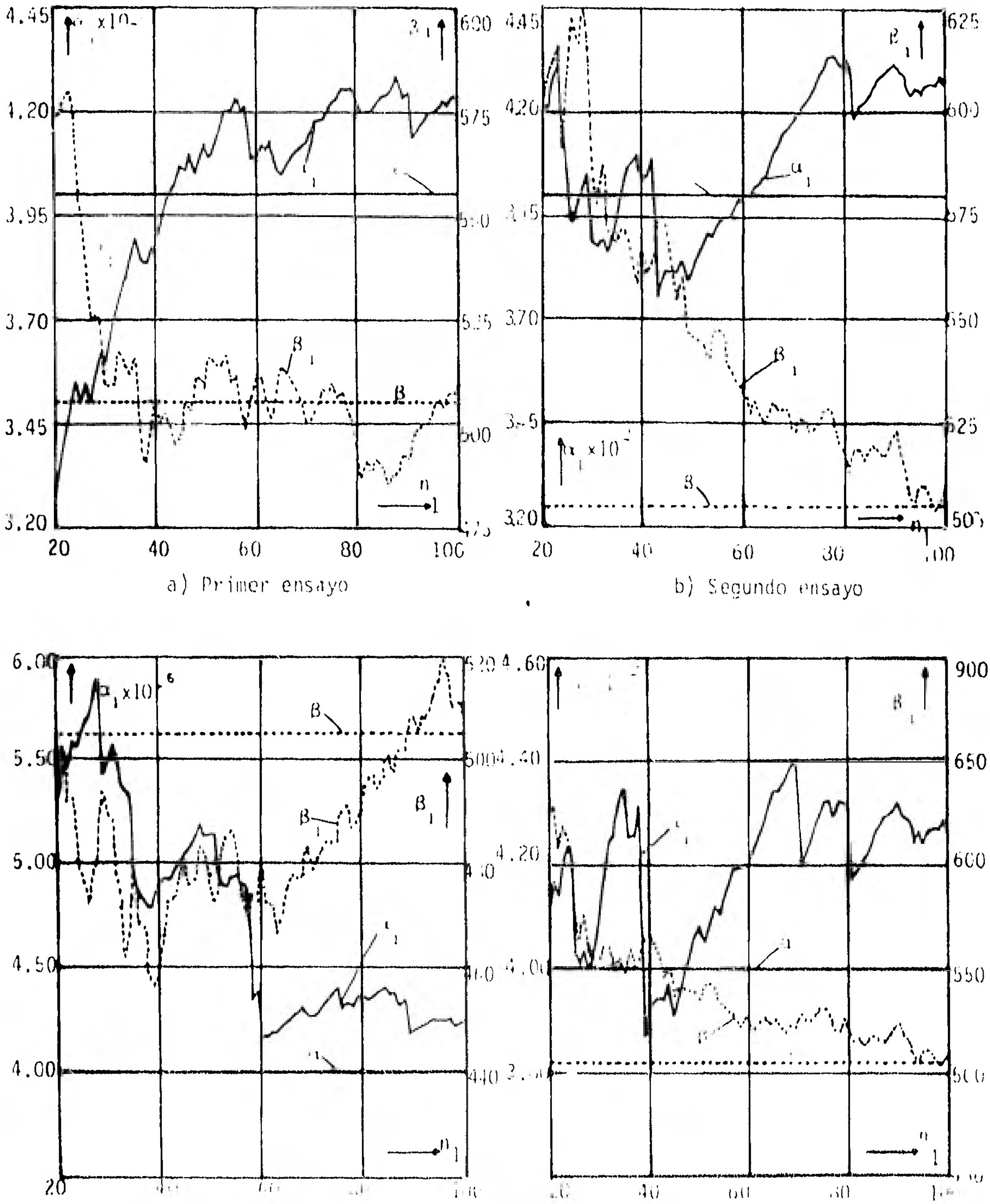
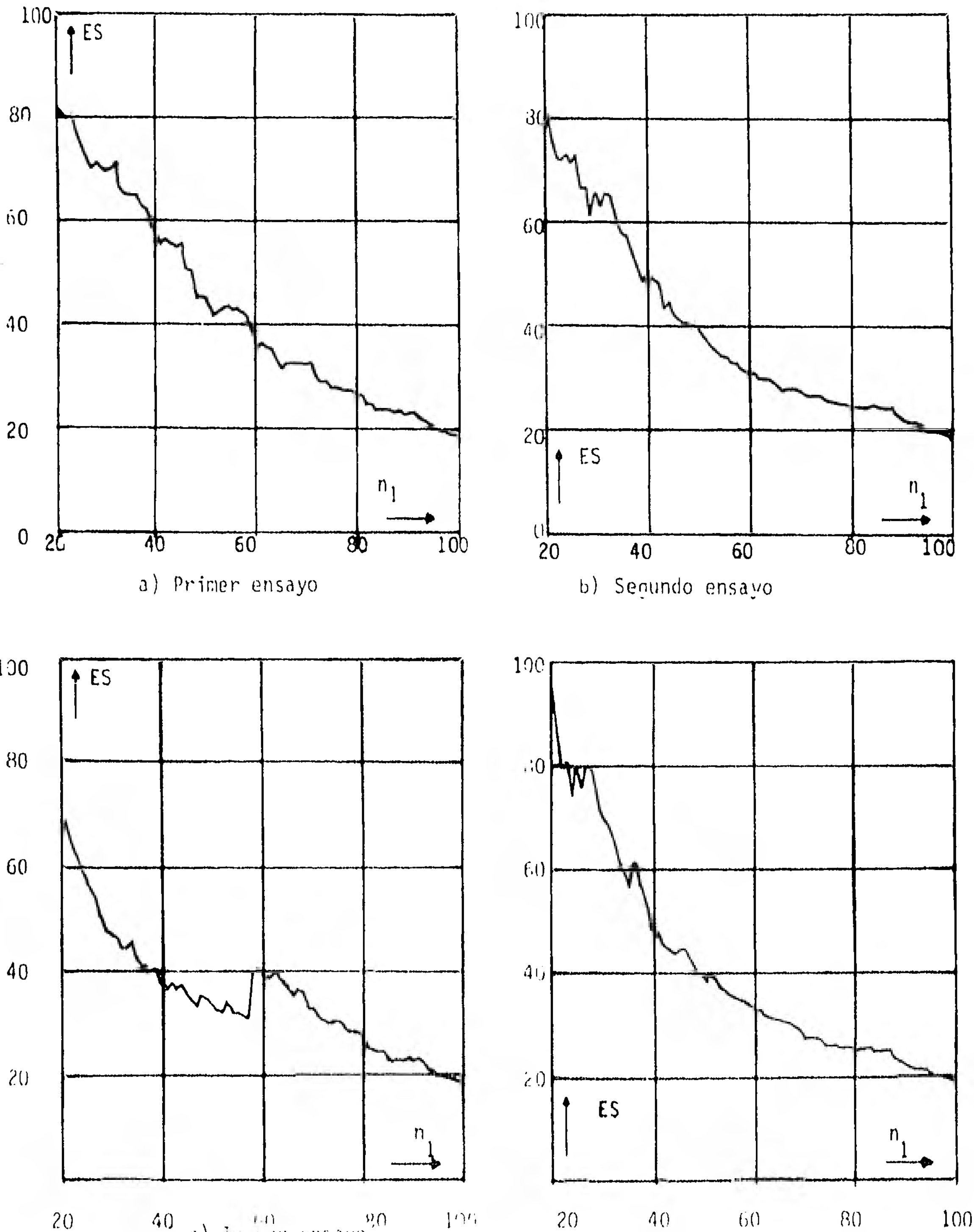


Figura 3.4. Variación de las estimas de los parámetros (α_1 y β_1) de la distribución Gumbel cuando n_1 aumenta



a) Primer ensayo

b) Segundo ensayo

c) Tercer ensayo

d) Cuarto ensayo

Figura 3.5 Variación de ES cuando n_1 aumenta

Cuando se dibujan en papel de probabilidad Gumbel, los gastos máximos de un río contra sus períodos de retorno, en muchas ocasiones pueden observarse puntos que es posible ajustar a dos segmentos rectos. El ajuste a una sola recta aunque se ha usado tradicionalmente, rara vez ocurre, esto se debe quizá a la presencia de dos poblaciones formadas por avenidas provocadas por fenómenos diferentes.

3.3a1 *Función de densidad, función de distribución acumulada y parámetros.*

Considerando que cada grupo de avenidas por separado sigue una Distribución Gumbel y que la causa de un grupo son los ciclones, los gastos máximos cuando no se presentan estos tienen parámetros α_1 y β_1. Su distribución está dada a partir de (3.36) por:

$$F_1(x) = \exp\{-\exp[-\alpha_1(x - \beta_1)]\} \qquad (3.41a)$$

Si los parámetros del otro grupo de avenidas (ciclónicas) tienen parámetros α_2 y β_2, su distribución en forma análoga a (3.41a) está dada por:

$$F_2(x) = \exp\{-\exp[-\alpha_2, (x - \beta_2)]\} \qquad (3.42a)$$

Por medio de los registros meteorológicos, es posible separar los años ciclónicos de los no ciclónicos.

La probabilidad de que en un año cualquiera se tenga un gasto máximo menor o igual que cierto valor puede escribirse como la suma de probabilidades condicionales:

$$P(X \leq x) = P(X \leq x|A)\; P(A) + P(X \leq x \mid \bar{A})\; P(\bar{A}) \qquad (3.43a)$$

donde:

X: es la variable aleatoria (gastos máximos) y x su valor.

A: representa un año no ciclónico

$\bar{A}$: representa un año ciclónico

Por definición de función de distribución acumulada (ec. 2.22), en un año no ciclónico:

$$P\ (X \leq x \mid A) = F_1\ (x)$$

Si el año es ciclónico, el gasto máximo puede presentarse debido a un ciclón o por causas "normales", aunque generalmente los ciclones ocasionan avenidas muy grandes por lo que la probabilidad de que el gasto máximo anual en un año ciclónico sea causado por un ciclón es muy cercana a uno.

$$(X \mid \bar{A}) = \text{máx}\left\{\begin{array}{l}\text{máx (gastos ciclónicos)}\\ \text{máx (gastos no ciclónicos)}\end{array}\right\}$$

La distribución de los gastos máximos X dado que el año es ciclónico está dada por:

$$P(X = x \mid \bar{A}) = P\{\text{máx (g.c.)} = x\} P\{\text{máx(g.c.)} \leq x\}$$

$$+ P\{\text{máx(g.n.c.)} = x\}\ P\{(\text{g.c.}) \leq x\} \qquad (3.44a)$$

Las funciones de densidad de probabilidad correspondientes a $F_1(x)$ y $F_2\ (x)$ son respectivamente (ec. 2.25):

$$f_1(x) = \frac{dF_1\ (x)}{dx}$$

$$f_2(x) = \frac{dF_2\ (x)}{dx}$$

Por lo tanto, considerando (3.44a)

$$f_{X}(x) = f_1(x)\,F_2(x) + f_2(x)\,F_1(x) \tag{3.45a}$$
$$(X_{máx}|\bar{A})$$

Por lo que:

$$P(X \leq x \,|\bar{A}) = \int_{-\infty}^{x} \{f_1(\xi)\,F_2(\xi) + f_2(\xi)\,F_1(\xi)\}\,d\xi \tag{3.46a}$$

Además:

$$d\{F_1(x)\,F_2(x)\} = f_1(x)\,F_2(x) + f_2(x)\,F_1(x)\,dx \tag{3.47a}$$

Por lo que (3.46a) puede calcularse como:

$$P(X \leq x\,|\bar{A}) = \{F_1(\xi)\,F_2(\xi)\}_{-\infty}^{x} \tag{3.48a}$$

Considerando que:

$$F_1(\xi) \rightarrow 0 \quad \text{y} \quad F_2(\xi) \rightarrow 0$$

$$\xi \rightarrow -\infty \qquad\qquad \xi \rightarrow -\infty$$

$$P(X \leq x\,|\bar{A}) = F_1(x)\,F_2(x) \tag{3.49a}$$

Sustituyendo (2.22) y (3.49a) en (3.43a) se tiene:

$$P(X \leq x) = F_1(x)\,P(A) + F_1(x)F_2(x)\,P(\bar{A}) \tag{3.50a}$$

Haciendo:

$p = P(A)$

(3.50a) se escribe como:

$$F(x) = F_1(x)\,p + (1-p)\,F_1(x)\,F_2(x) \tag{3.51a}$$

ó

$$F(x) = F_1(x)\,\{p + (1-p)\,F_2(x)\} \tag{3.52a}$$

Sustituyendo (3.41a) y (3.42a) en (3.52a):

$$F(x) = \exp\{-\exp\left[-\alpha_1 (x-\beta_1)\right]\}\{p + (1-p)\exp\left[-\exp(-\alpha_2(x-\beta_2))\right]\} \quad (3.53\ a)$$

que es la expresión de la función de distribución acumulada.

Esta expresión toma en cuenta distintas condiciones meteorológicas que generalmente se ignoran.

3.3a.2. *Momentos y su relación con los parámetros.*

Es válido utilizar las expresiones (2.26) ó (2.30) para la obtención de los momentos de (3.53a). Es necesario hasta el 5o. momento ya que son cinco los parámetros que se tienen ($p, \alpha_1, \alpha_2, \beta_1, \beta_2$). La estimación de parámetros será estudiada en el capítulo 5; sólo se aclara aquí en forma breve, que los métodos usados para estimar los 5 parámetros propuestos en la referencia 12 no son estudiados en este trabajo.

Inicialmente se usan las relaciones dadas por (3.38) y (3.39):

$$m_{y_1} = \beta_1 + \frac{0.5772}{\alpha_1}$$

$$\sigma^2_{y_1} = \frac{\pi^2}{6\,\alpha_1^2}$$

$$m_{y_2} = \beta_2 + \frac{0.5772}{\alpha_2}$$

$$\sigma^2_{y_2} = \frac{\pi^2}{6\,\alpha_2^2}$$

3.4 Distribución Exponencial

Es conveniente conocer como es un proceso de Poisson antes de explicar la Distribución Exponencial.

La función de probabilidad discreta para una distribución de Poisson está dada por:

$$Px\ (x) = \frac{\nu^{x}\ e^{-\nu}}{x!} \qquad x = 0,1,2,3,\ldots,\infty \tag{3.41}$$

donde:

ν = pn

p = probabilidad del suceso

n = número de pruebas o datos

e = 2.71828...

Los momentos de (3.41) están dados por:

$$E(X) = m_X = \nu$$

$$\sigma_X^{\ 2} = \nu$$

Los procesos de Poisson son funciones aleatorias de tiempo que generalmente estiman los tiempos entre llegadas de vehículos o entre eventos.

Si (3.41) se expresa como:

$$P_X(x) = \frac{(\lambda t)^{x}\ e^{-\lambda t}}{x!} \qquad x = 0,1,2,3,\ldots,\infty \tag{3.42}$$

es decir, $\nu = \lambda t$, esta forma (3.42) es fácil de asociar con un proceso de Poisson, si t representa el tiempo o es una variable similar.

Un proceso de Poisson debe cumplir las siguientes condiciones:

1. La probabilidad de un evento en un intervalo pequeño de tiempo de t a t + h es aproximadamente λh para cualquier t. ($h = \Delta t$).

2. La probabilidad de que ocurran dos o más eventos en un intervalo de tiempo corto es despreciable comparada con λh, es mucho menor que λh.

3. El número de eventos en un intervalo de tiempo es independiente del número de eventos en otro intervalo.

En una muestra aleatoria se tienen números aleatorios X: $x_1, x_2, x_3, \ldots, x_n$; en forma similar en observaciones de procesos aleatorios se tienen funciones de la muestra generalmente de tiempo;

$$X(t): x_1(t), x_2(x), \ldots, x_n(t)$$

Aplicando (3.42) para el caso en que x = 0

$$P_X(0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t} \tag{3.43}$$

También:

$$P_X(0) = 1 - F_T(t) \tag{3.44}$$

debido a que la probabilidad de que T exceda algún valor t es igual a la probabilidad de no ocurrencia de eventos en ese intervalo de longitud t.

3.4.1 *Función de densidad, función de distribución acumulada y parámetros.*

Igualando (3.43) y (3.44), se tiene:

$$1-F_T(t) = e^{-\lambda t} \qquad t \geq 0 \tag{3.45}$$

Por lo que:

$$F_T(t) = 1 - e^{-\lambda t} \qquad t \geq 0 \tag{3.46}$$

y:

$$f_T(t) = \frac{dF_T(t)}{dt} = \lambda e^{-\lambda t} \qquad t \geq 0 \tag{3.47}$$

(3.46) y (3.47) definen las funciones de distribución acumulada y la función de densidad de probabilidad respectivamente.

T es la variable aleatoria independiente y t su valor.

λ es el parámetro de esta distribución.

La Distribución Exponencial de un parámetro se representa por EX(λ) y corresponde a la primera ocurrencia de un evento de Poisson.

3.4.2. *Momentos y su relación con los parámetros.*

Los momentos de la Distribución Exponencial están dados por:

$$E\ (T) = \int_0^{\infty} t\lambda e^{-\lambda t}\, dt \quad * \tag{3.48}$$

de donde se obtiene:

$$m_T = \frac{1}{\lambda} \tag{3.49}$$

$$\sigma_T^2 = \frac{1}{\lambda^2} \tag{3.50}$$

λ representa el promedio de eventos por unidad de tiempo; $\frac{1}{\lambda}$ es el tiempo promedio entre ocurrencias de eventos.

---

* Ref. 2

3.4.3 *Distribución Exponencial de dos parámetros.*

Al igual que a la Distribución Lognormal, a esta distribución puede introducirse otro parámetro; este segundo parámetro llamado α, se resta de la variable t.

La expresión (2.46) toma la siguiente forma:

$$F_T(t) = 1-e^{-\lambda(t-\alpha)} \qquad (t-\alpha) > 0 \qquad (3.51)$$

Con este nuevo parámetro, la variancia no se modifica en su valor, pero la media sí y por lo tanto también el coeficiente de variación:

$$m_T = \frac{1}{\lambda} + \alpha \qquad (3.52)$$

$$\sigma_T^2 = \frac{1}{\lambda^2} \qquad (3.53)$$

Esta distribución de dos parámetros se representa por EX(λ,α). Cuando α vale cero, se tiene la distribución de un parámetro. En la figura 3.6 se presenta la función de densidad de probabilidad (de dos parámetros).

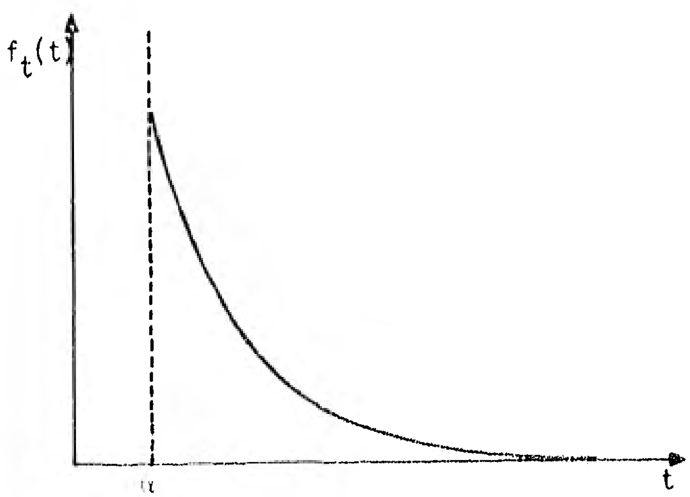


Figura 3.6. Fdp Exponencial(2 parámetros)

La estimación del tiempo entre ocurrencia de eventos es la principal aplicación de la Distribución Exponencial y es en este tipo de problemas donde cabe esperar una mejor aproximación.

En la práctica es bastante usada en problemas no relacionados con el tiempo entre ocurrencia de eventos, por ejem: en lluvias y avenidas máximas; cuando se tienen datos observados, puede llevarse a cabo el ajuste a esta distribución aunque no parezca adecuado por la naturaleza del fenómeno que se estudia ya que la distribución exponencial es más apropiada para procesos de Poisson. Muchas veces es empleada sólo como una representación conveniente de un fenómeno cuando los datos observados parecen indicarlo al ingeniero debido a experiencias similares.

3.5 Distribución Gamma

Al igual que la Distribución Exponencial, la Distribución Gamma está relacionada con procesos de Poisson.

3.5.1 Función de densidad, función de distribución acumulada y parámetros.

Considerando que los tiempos T_i, $i = 1,2,3,\ldots, k$ entre la ocurrencia de dos eventos siguen una Distribución Exponencial y son independientes, la suma de k intervalos de tiempos $X_k = T_1 + T_2 + \ldots + T_k$ tiene una distribución de la forma:

$$f_{X_k}(x) = \frac{\lambda\{\lambda(x-\delta)\}^{k-1} \exp\{-\lambda(x-\delta)\}}{(k-1)!} \quad ; \; k=1,2,\ldots; x \geq \delta; \; \lambda > 0 \qquad (3.54)$$

que es la función de densidad de probabilidad (fdp) de la Distribución Gamma

K, λ y δ son los parámetros de esta distribución por lo que es llamada Distribución Gamma de tres parámetros y se representa por $G(K,\lambda,\delta)$.

En forma general, K puede referirse no sólo a valores enteros, (3.54) toma entonces la forma:

$$f_X(x) = \frac{\lambda\{\lambda(x-\delta)\}^{k-1} \exp\{-\lambda(x-\delta)\}}{\Gamma(k)} \quad ; \; k>0; \; x\geq\delta; \; \lambda>0 \qquad (3.55)$$

donde:

$$\Gamma(k) = (k+1)! \qquad \text{para } k = 1,2,3,\ldots \qquad (3.56)$$

$$\Gamma(k+1) = k\Gamma(k) \qquad \text{para } k>0 \qquad (3.57)$$

$$\Gamma(k) = \int_0^{\infty} x^{k-1} e^{-x}\, dx \text{ para } k > 0 \qquad (3.58)$$

La función de distribución acumulada Gamma está dada por:

$$F_X(x) = \int_0^{x} \{\lambda^k x^{k-1} e^{-\lambda x}/\Gamma(k)\}\, dx = \frac{\Gamma\{k,\lambda(x-\delta)\}}{\Gamma(k)} \qquad (3.59)$$

El numerador en la integral es llamado función Gamma incompleta.

En la figura 3.7 se presentan varias curvas de la función de densidad gamma para 3 valores de K y $\delta = 0$.

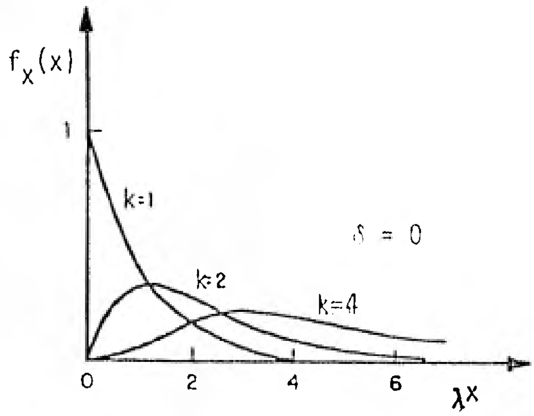


Figura 3.7. Curvas de la fdp Gamma para tres valores de k.

3.5.2. *Momentos y su relación con los parámetros.*

De la integración de (3.55) el aplicar (2.30) a (2.32) se obtienen los momentos. Estos son (Ref. 2):

$$m_X = \frac{k}{\lambda} + \delta \tag{3.60}$$

$$\sigma_X^2 = \frac{k}{\lambda^2} \tag{3.61}$$

$$\gamma_1 = \frac{E\{(x-m_X)^3\}}{\sigma_X^3} = \frac{2}{\sqrt{k}} \tag{3.62}$$

El parámetro δ representa un cambio del origen y afecta a la variable de esta distribución restándose de ella $(x-\delta)$. Si el valor de δ es cero, se tiene la Distribución Gamma de dos parámetros representada por $G(K,\lambda)$.

El cálculo de la probabilidad acumulada puede hacerse en forma alternativa a la aplicación de (3.59), de la siguiente forma:

Como:

$$y = (x-\delta)\lambda \tag{3.63}$$

sigue una distribución Xi cuadrado (X^2) (Ref.2) con $\nu = 2k$ grados de libertad y:

$$X^2 = 2y \qquad \text{(Ref.4)} \qquad (3.64)$$

se tiene que:

$$X^2 = 2\lambda\,(x-\delta) \quad * \qquad (3.65)$$

Además:

$$z = \left[\sqrt[3]{\frac{X^2}{\nu}} + \frac{2}{9\nu} - 1\right]\sqrt{\frac{9\nu}{2}} \qquad (3.66)$$

sigue una distribución aproximadamente normal con media 0, y variancia 1 para $\nu = 2k > 30$ (Ref.4). Se ha observado también, bastante aproximación para $\nu \leq 30$.

Con los valores obtenidos al aplicar (3.66), se busca en la tabla 1 del apéndice la probabilidad acumulada correspondiente a cada z.

La estimación de eventos simples x utilizando esta forma alternativa, se hace siguiendo un procedimiento inverso:

Despejando X^2 de (3.66), se tiene:

$$X^2 = \nu\left\{ z\,\frac{2}{9\nu} - \frac{2}{9\nu} + 1 \right\}^3 \qquad (3.67)$$

El valor de z se obtiene de la tabla de áreas bajo la curva normal estandarizada (tabla 1 del apéndice); se busca a partir de la probabilidad

* No debe confundirse la notación X^2 con x; X^2 se refiere a la distribución Xi cuadrado y x es la magnitud del evento asociado a la variable aleatoria de la distribución gamma.

observada que está dada por (2.57):

$$Po = 1 - \frac{1}{T} = 1 - \frac{i}{n+1}$$

donde T es el período de retorno, i el rango y n el número de datos; fueron definidos en el capítulo 2.

Una vez calculado el valor de X^2, debe calcularse y, de (3.64):

$$y = \frac{X^2}{2} \qquad (3.68)$$

También, de (3.63) puede despejarse el valor de x:

$$x = \frac{y}{\lambda} + \delta \qquad (3.69)$$

que es el valor correspondiente al evento simple estimado.

En la tabla 3.1 se presenta un resumen de las funciones de densidad de probabilidad vistas en este capítulo y sus parámetros. Para las Distribuciones Lognormal y Gamma de 3 parámetros y Exponencial de 2 parámetros, basta con hacer cero a_L, δ y α para tener las Distribuciones Lognormal y Gamma de dos parámetros y Exponencial de uno.

TABLA 3.1

DISTRIBUCION	FUNCION DE DENSIDAD DE PROBABILIDAD	
NORMAL	$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{x-m_X}{\sigma_X} \right)^2 \right]$ $-\infty < x < \infty$	m_X
LOGNORMAL (3 *parámetros*)	$f_Y(y) = \frac{1}{(y-a_L)\, \sigma_L \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left[\frac{\mathrm{Ln}(y-a_L) - m_L}{\sigma_L} \right]^2 \right\}$ $y \geq a_L$	m_L
GUMBEL	$f_Y(y) = \alpha \exp\{ -\alpha (y-\beta) - \exp[-\alpha(y-\beta)] \}$ $-\infty < y < \infty$	α
EXPONENCIAL (2 *parámetros*)	$f_T(t) = \lambda \exp\{ -\lambda (t-\alpha) \}$ $t > \alpha$	λ
GAMMA (3 *parámetros*)	$f_X(x) = \frac{\lambda \{ \lambda(x-\delta) \}^{k-1} \exp\{ -\lambda(x-\delta) \}}{\Gamma(k)}$ $x > \delta,\ k > 0$	k

4. ESTIMADORES

4.1 Introducción

Gran parte de la ingeniería está basada en el uso de funciones mediante las cuales es posible conocer el valor de una variable (dependiente) dado el valor de otra (independiente).

Cuando en una expresión $y = f(x)$, la variable independiente x se considera aleatoria, la aleatoreidad afecta a las variables que dependen funcionalmente de ella, en este caso y. Las variables hidrológicas son variables aleatorias y se tratan como tales.

Debido a su misma naturaleza, no es posible predecir el valor exacto que tomará una variable aleatoria al realizar un experimento, pero puede tenerse una aproximación a él por medio de un modelo probabilístico.

El comportamiento exacto de las variables, se conocería solamente si se pudiera establecer su función completa, es decir, conocer la población. En la práctica, ésto no es posible y puede no ser necesario.

Un modelo representativo resume solamente las características dominantes del comportamiento de la variable aleatoria; ésto es suficiente, la mayoría de las veces, para los propósitos de la ingeniería.

Como se mencionó en el capítulo 2, las cantidades que resumen y describen características de la población son llamadas parámetros y las cantidades que describen una muestra, estadísticos.

La media mx y la variancia σx^2 son ejemplos de parámetros; la media de la muestra (de esa población) $\bar{x}$ y su variancia Sx^2 lo son de estadísticos.

Un estimador es una función de la variable aleatoria en estudio aplicada en una muestra, se determina a partir de los estadísticos muestrales y se usa para inferir o "predecir" el valor de un parámetro. Al valor correspondiente del estimador se le llama estima del parámetro o brevemente, estima.

No debe esperarse que de una secuencia de observaciones finita, resulte el valor exacto del parámetro de un modelo probabilístico ya que los datos mismos son aleatorios. Es necesario, al aceptar valores de parámetros derivados de una muestra, considerar que sólo son estimaciones de sus verdaderos valores sujetos a errores. Los parámetros son valores constantes y los estadísticos, variables.

Si la estimación de un parámetro poblacional se hace por medio de un solo valor de un estadístico, es decir, por medio de una sola estima, el estimador se llama de punto ó puntual; por otra parte, si se considera que el valor del parámetro se encuentra con cierta probabilidad entre dos valores, se tiene una estimación del parámetro por medio de un intervalo de confianza.

4.2 Estimadores de punto o puntuales.

La elección del estimador a usar depende de las propiedades que posea el mismo Si de una población se extraen muestras de un mismo tamaño y se calcula el mismo estadístico a cada muestra, la distribución que sigue el estadístico se llama distribución muestral del estadístico, por ejemplo, si el estadístico calculado es la media, se tendrá una distribución muestral de medias (secc. 2.2.9).

Las características o propiedades deseables en un estimador puntual son:

4.2.1 *Insesgo.*

Un estimador es insesgado si la media de su distribución muestral es igual al parámetro poblacional Es decir, un estimador es insesgado si:

$$\underset{n\to\infty}{m_{\hat{\theta}}} = E\left[\hat{\theta}\right] = \theta \tag{4.1}$$

n: es el número de muestras de las que se obtiene $\hat{\theta}$

θ: es un parámetro poblacional

$\hat{\theta}$: es el estimador de θ

$E\left[\hat{\theta}\right]$: está definida por (2.26) ó (2.27)

4.2.2 *Eficiencia.*

Se llama estimador eficiente de un parámetro al de menor variancia de entre los estimadores insesgados de ese parámetro. La eficiencia relativa de dos estimadores insesgados está dada por:

$$\text{eficiencia relativa} = \frac{\operatorname{Var} \hat{\theta}_1}{\operatorname{Var} \hat{\theta}_2} \tag{4.2}$$

$\hat{\theta}_1$ es el estimador de menor variancia.

La eficiencia "absoluta" (Ref. 6) se obtiene de la siguiente forma:

Sea la función:

$$f\,(x_1, x_2, \ldots, x_n; \theta) = f\,(X; \theta) \tag{4.3}$$

Haciendo:

$$V = \frac{\partial}{\partial\theta} \operatorname{Ln} f(X;\theta)$$

y

$$T(\theta) = \operatorname{Var} V = E\left\{\frac{\partial}{\partial\theta} \operatorname{Ln} f\,(X;\theta)\right\}^2 = nE\left\{\frac{\partial}{\partial\theta} \operatorname{Ln} f\,(X;\theta)\right\}^2 \tag{4.4}$$

Donde:

θ es un parámetro poblacional asociado a la función $f\,(x)$

Si $\hat{\theta}$ es un estimador insesgado del parámetro θ, la eficiencia de $T\,(\theta)$ está dada por:

$$e(T) = \frac{1/T(\theta)}{\operatorname{Var} \hat{\theta}} \tag{4.5}$$

Donde $T = T\,(\theta)$ está dada por (4.4)

El rango de $e(T)$ es$(0,1)$

Un estimador de eficiencia 1 es llamado eficiente.

Eficiencia asintótica es el valor límite de la eficiencia cuando el tamaño de la muestra tiende a infinito. Para que este límite sea un número finito positivo la variancia del estimador debe ser asintótica en forma proporcional $1/n$.

Cuando la eficiencia asintótica del estimador es 1, éste es llamado asin tóticamente eficiente

4.2.3 *Consistencia*

Se llama a $\hat{\theta}$ un estimador consistente de θ si es insesgado o su sesgo-- se aproxima a cero al igual que su variancia cuando el tamaño de la muestra tiende a infinito. Dicho de otra forma y usando símbolos: $\hat{\theta}$ es-- un estimador consistente de θ sí y sólo si para un valor positivo arbitrario δ próximo a cero se cumple que:

$$P\left(|\hat{\theta}-\theta| < \delta\right) \rightarrow 1 \quad \text{cuando } n \rightarrow \infty$$

P: significa probabilidad

4.2.4 *Invariancia.*

Utilizando la misma expresión que en (4.3) para explicar el concepto de invariancia:

$$f(x_1, x_2, \ldots x_n; \theta) = f(x; \theta)$$

Si la media de la distribución es función de θ, es decir $m_x = g(\theta)$, por ejemplo $m_x = \frac{1}{\theta}$ y el estimador de la media poblacional es la media de la muestra $\bar{x}$, se tiene que:

$$g(\hat{\theta}) = \frac{1}{\hat{\theta}} = \frac{1}{1/\bar{x}} = \bar{x}$$

Si al sustituir en la función de un parámetro $g(\theta)$ el estimador ($\hat{\theta}$ por θ), el estimador no cambia, se dice que es invariante, es decir:

$$\widehat{g(\theta)} = g(\hat{\theta}) \qquad (4.6)$$

4.2.5 *Mínimo error medio al cuadrado.*

Esta medida compara no sólo estimadores insesgados sino también estimadores sesgados(entre sí y con insesgados).

Representa una medida de como se encuentra esparcido el estimador antes y después de su verdadero valor.

$$\text{Error medio al cuadrado} = E\left[\hat{\theta} - \theta\right]^2 \tag{4.7}$$

4.2.6 *Suficiencia.*

Un estimador $\hat{\theta}$ es estimador suficiente de θ si hace uso de toda la información concerniente a θ contenida en la muestra.

En la práctica, generalmente las propiedades más importantes requeridas para la estimación en problemas de Hidrología, son el insesgo y la eficiencia cuando es necesario obtener la máxima información de los datos disponibles. La eficiencia generalmente implica consistencia y suficiencia.

En el capítulo 5 se estudiarán los métodos de estimación de parámetros, cada método proporciona estimadores con algunas de las propiedades aquí vistas; estas propiedades como se verá, varían de un método a otro.

A continuación se presenta un ejemplo que explica algunas de las propiedades mencionadas.

Ejemplo 4.1 Dada una población formada por los siguientes números: 2, 3, 9, 6, 5, 7; tomando todas las posibles muestras de tamaño n=3, comprobar si la media y la mediana son estimadores insesgados de la media y cual - de los dos es un estimador eficiente, además, comprobar si la variancia de la muestra es un estimador insesgado de la variancia poblacional.

La media poblacional es :

$$mx = \frac{2 + 3 + 9 + 6 + 5 + 7}{6} = 5.3333$$

La mediana de:

2, 3, 5, 6, 7, 9

es:

$$Med_x = \frac{5 + 6}{2} = 5.50$$

La variancia:

$$\sigma_x^2 = \frac{\Sigma (xi - mx)^2}{n} = 5.5556$$

Todas las posibles muestras de tamaño n=3 que pueden obtenerse son: 6^3= 216 muestras. Obteniendo en forma exhaustiva la media, la mediana y la variancia de cada muestra, se forman distribuciones muestrales (capítulo 2) de medias, de medianas y de variancias. Calculando la media muestral de medias $m_{\bar{x}}$, la variancia muestral de medias $\sigma^2_{\bar{x}}$, la media muestral de medianas m_{med}, la variancia muestral de medianas σ^2_{med}, la media muestral de variancias m_{s^2} y la variancia muestral de variancias $\sigma^2_{Sx^2}$ se llega a-- los siguientes resultados.

$$m_{\bar{x}} = 5.3333 \qquad \sigma^2_{\bar{x}} = 1.8519$$

$$m_{med} = 5.3241 \qquad \sigma^2_{med} = 4.5153$$

$$m_{Sx^2} = 3.7037 \qquad \sigma^2_{Sx} = 8.4390$$

Puede observarse que la media de la muestra y la mediana de la muestra - son estimadores insesgados de la media poblacional ya que cumplen con -- (4.1).

Como la variancia de la distribución muestral de medias es menor que la de la distribución muestral de medianas; la media de la muestra es un estimador eficiente de la media poblacional y la mediana de la muestra es un estimador no eficiente del mismo parámetro (secc. 4.2.2)

La variancia de la muestra S_x^2 no es un estimador insesgado de la variancia poblacional ya que no cumple con (4.1), es decir:

$$E\left[S_x^2\right] = m_{S_x^2} \neq \sigma_x^2$$

Puede comprobarse también que:

$$E\left[S_x^2\right] = m_{S_x^2} = \frac{n-1}{n}\sigma_x^2 \tag{4.8}$$

ya que:

$$\frac{n-1}{n}\sigma_x^2 = \frac{2}{3}(5.5556) = 3.7037 = m_{S_x^2}$$

Por lo tanto, un estimador insesgado de la variancia poblacional es:

$$S^{2*} = \frac{n}{n-1}S^2 \tag{4.9}$$

Donde:

S^2 : es la variancia de la muestra dada por (2.2)

n : es el número de datos de la muestra.

También puede comprobarse que:

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{n}}\sigma_x \tag{4.10}$$

ya que:

$$\sigma_{\bar{x}} = \sqrt{\sigma^2_{\bar{x}}} = 1.3608$$

y

$$\frac{1}{\sqrt{n}} \sigma_x = \frac{1}{\sqrt{3}} \sqrt{5.5556} = 1.3608$$

Esta característica será presentada y utilizada más adelante en el presente capítulo.

4.3 Estimadores por intervalos de confianza.

Cuando se lleva a cabo la estimación de un parámetro, no debe esperarse que la estima sea exactamente igual al valor de éste, en cambio, es razonable esperar que su valor (del parámetro) se encuentre entre dos valores límite.

El intervalo de confianza se define como un rango del estadístico a partir del cual se obtiene el estimador. Si se pretende que el valor del parámetro se encuentre en ese intervalo, a este último debe asociársele una probabilidad $1 - \alpha, 0 \leq \alpha \leq 1$. La probabilidad $1 - \alpha$ asociada al intervalo significa que el valor del parámetro Θ se encontrará el $(1 - \alpha)$ % de las veces en el intervalo y fuera del él el α % de las veces. A $1 - \alpha$ se le llama nivel de confianza y se acostumbra darlo en porciento . A α se le llama nivel de significancia.

El intervalo de confianza varía de una muestra a otra aunque sean del mismo tamaño ya que cada muestra proporciona una diferente estima.

A los números que limitan el intervalo se les llama límites de confianza de la estima.

En símbolos, si $\hat{\theta}$ es un estimador del parámetro θ:

$$\theta = \hat{\theta} \pm \text{un error} \tag{4.11}$$

representa los límites de confianza.

$\left[\hat{\theta} - \text{error}, \hat{\theta} + \text{error}\right]$ representa el intervalo del parámetro.

A la expresión:

$$P\left[\hat{\theta} - \text{error} \leq \theta \leq \hat{\theta} + \text{error}\right] = 1-\alpha \tag{4.12}$$

se le llama intervalo de confianza. (P significa probabilidad).

La magnitud del error depende de la distribución muestral del estadístico considerado como estimador y en especial de su desviación estándar así como del nivel de confianza que se considere.

4.3.1 *Error estándar*

A la desviación estándar de un estimador cualquiera se le llama también-- error estándar.

Por lo tanto, el error que aparece en (4.11) tiene la forma:

$$\text{error} = \pm K \sigma_{\hat{\theta}} \tag{4.13}$$

donde:

$\sigma_{\hat{\theta}}$: es el error estándar del estimador $\hat{\theta}$

K : es una constante que depende del nivel de confianza y de la-- distribución que sigue $\hat{\theta}$.

Sustituyendo (4.13) en (4.11)

$$\Theta = \hat{\Theta} \pm K \hat{\Theta} \qquad (4.14)$$

Análogamente, (4.12) puede escribirse de la siguiente forma:

$$P\left[\hat{\Theta} - K\sigma_{\hat{\Theta}} \leq \Theta \leq \hat{\Theta} + K\sigma_{\hat{\Theta}}\right] = 1 - \alpha \qquad (4.15)$$

expresando así el intervalo de confianza.

Si se considera que $\hat{\Theta}$ sigue una distribución normal, el valor de K se obtiene de la tabla 1 del apéndice de areas bajo la curva normal estandarizada y es el correspondiente a z para un nivel de confianza determinado: $k = z$.

Generalmente, los niveles de confianza que se asignan son: 0.9, 0.95 y 0.99; después de esta asignación, al aplicar (4.15), como la curva normal es simétrica, se busca en la tabla el valor de z correspondiente a $(1-\alpha)/2$ ya que las áreas antes de $-z$ y después de z son $\alpha/2$ (figura 4.1).

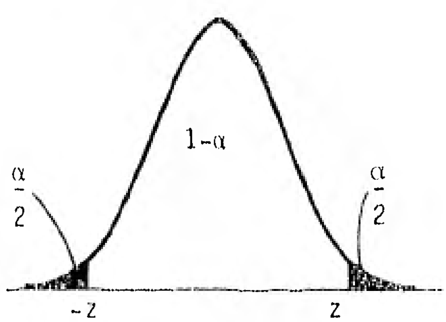


Figura 4.1. Distribución muestral de medias (Prueba de dos colas)

Es posible proceder en forma inversa, es decir dado un valor de z, obtener el nivel de confianza $1-\alpha$. Se busca en la tabla 1 el área correspondiente a z que será igual a $(1-\alpha)/2$ con lo cual se obtiene el nivel de confianza.

Sustituyendo z por k en (4.15)

$$P(\hat{\Theta} - z\,\sigma_{\hat{\Theta}} \leq \Theta \leq \hat{\Theta} + z\,\sigma_{\hat{\Theta}}) = 1 - \alpha \tag{4.16}$$

Similarmente al uso de z puede usarse "t" (Ref. 5), considerando también que $\hat{\Theta}$ se distribuye normalmente. El uso de t se prefiere el de z cuando la muestra es pequeña; esto se establece a partir de la precisión requerida. t puede obtenerse de la tabla 2 del apéndice con el nivel de confianza $1-\alpha$ y n-1 grados de libertad.

El número de grados de libertad se define como el número de observaciones independientes de n menos el número de parámetros m que se estiman a partir de la muestra.

Se denota por ν :

$$\nu = n - m \tag{4.17}$$

Sustituyendo t por k en (4.15):

$$P(\hat{\Theta} - t\,\sigma_{\hat{\Theta}} \leq \Theta \leq \hat{\Theta} + t\,\sigma_{\hat{\Theta}}) = 1-\alpha \tag{4.18}$$

4.3.2. *Distribución muestral de medias*

Para el caso de las medias muestrales, su error estándar esta dado por (4.10)

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{n}}\,\sigma_x$$

Por lo que usando (3.15) y (3.16) el intervalo y los límites de confianza de $\bar{x}$ como estimador de m_x son:

$$P(\bar{x} - t\,\sigma_{\bar{x}} \leq m_x \leq \bar{x} + t\,\sigma_{\bar{x}}) = P(\bar{x} - t\,\frac{\sigma_x}{\sqrt{n}} \leq m_x \leq \bar{x} + t\,\frac{\sigma_x}{\sqrt{n}}) = 1-\alpha \qquad (4.19)$$

Sustituyendo σ por S^*; desarrollando (4.19) y haciendo $t = t_{\nu,\alpha/2}$ se tiene:

$$P(-t_{n-m,\,\alpha/2} \leq \frac{\bar{x} - m_x}{S^*/\sqrt{n}} \leq t_{n-m,\,\alpha/2}) = 1-\alpha \qquad (4.20)$$

$S^* = \sqrt{S^{2*}}$: S^{2*} es un estimador insesgado de σ^2 dado por (4,9)

En este caso el número de grados de libertad es n-1, es decir, m=1 ya que debe estimarse un solo parámetro que es m_x aunque se usen $\bar{x}$ y S^*.

Cuando el intervalo de confianza tiene como límites, valores extremos del estimador a ambos lados del desconocido valor del parámetro como en (4.15), se tiene una prueba de dos colas; si sólo se tiene un límite ya sea superior o inferior, la prueba es de una cola. "Dos colas" significa que a cada extremo de la distribución existe un área igual a $\alpha/2$ cuyo significado fue explicado al principio de este capítulo. Si la prueba es de "una sola cola", sólo se tiene un límite (superior o inferior) y en ese extremo se tiene un área igual a α. En la figura 4.1 se presenta la distribución de un estadístico (normal) en la que se aplica una prueba de dos colas.

4.3.3. *Distribución muestral de varianzas*

El estadístico $\frac{nS^2}{\sigma^2}$ sigue una distribución χ^2 con n-1 grados de libertad

(Ref. 2). Utilizando esta característica pueden establecerse límites de confianza para la variancia σ^2 siendo el estimador S^2:

Si se realiza una prueba de una cola, se tiene:

$$P\left[\frac{nS^2}{\sigma^2} \leq X^2_{1-\alpha,\ n-1}\right] = 1 - \alpha \tag{4.21}$$

ó

$$P\left[\sigma^2 \geq \frac{n\ S^2}{X_{1-\alpha,\ n-1}}\right] = 1 - \alpha \tag{4.22}$$

que proporcionan el límite inferior para σ^2

$X^2_{1-\alpha,\ n-1}$ se obtiene, en forma similar a t, de la tabla 3 del apéndice con el nivel de confianza $1-\alpha$ y $n-1$ grados de libertad. La prueba también es válida si se sustituye en (4.22) S^2* por S^2.

Si el límite que se desea obtener es el superior, la expresión (4.22) toma la forma:

$$P\left[\sigma^2 \leq \frac{nS^2}{X^2_{\alpha,\ n-1}}\right] = 1 - \alpha \tag{4.23}$$

El valor de X^2 se obtiene de la misma tabla 3 ahora a partir de α y n-1. Cuando se realiza una prueba de dos colas, se tiene:

$$P\left[\frac{nS^2}{X^2_{1-\alpha/2,\ n-1}} \leq \sigma^2 \leq \frac{nS^2}{X^2_{\alpha/2,\ n-1}}\right] = 1 - \alpha \tag{4.24}$$

5. METODOS DE AJUSTE

5.1 Introducción

Al contar con una serie de observaciones (muestra), para poder hacer inferencias con base en ella, es necesario conocer el modelo o función de distribución de probabilidad que mejor la represente.

Debe proponerse el modelo y en seguida deben estimarse sus parámetros; también pueden proponerse y, después de estimar sus parámetros, seleccionar de acuerdo a ciertos criterios el que mejor se ajuste a las observaciones; esto último será estudiado en el capítulo 6.

En este capítulo se trata lo referente a la estimación de parámetros por tres diferentes métodos: Ajuste por Momentos, por Mínimos Cuadrados y por Máxima Verosimilitud.

También, respecto a estimaciones particulares y_T, se trata su *error estándar*; cada método de estimación suministra una forma de obtenerlo.

El *error estándar* representa una medida de la variación en la magnitud del evento estimado debido a que los parámetros estimados no son exac--

tamente los de la población por el tamaño insuficiente de la muestra. Tiene un significado similar al error estándar de un estimador (capítulo 4), y es utilizado también para obtener intervalos de confianza (de estimaciones particulares).

5.2 Momentos

5.2.1 *Descripción del método.*

La estimación de los parámetros de una función de distribución de probabilidad por el método de Ajuste por Momentos es bastante usado por su relativa sencillez en el cálculo de los estimadores y en el estudio de las distribuciones que siguen éstos.

El procedimiento consiste en igualar los momentos de los datos observados (muestra) a los correspondientes momentos de la población definida por la función de distribución de probabilidad.

El número de momentos de la muestra a usar depende del número de parámetros del modelo probabilístico; por ejemplo: si el modelo es de dos parámetros, el primero y el segundo momentos de la muestra $\bar{X}$ y S_X^2 respectivamente son necesarios y suficientes para la estimación de los dos parámetros.

Varios estimadores ajustados por el método de momentos son asintóticamente normales, es decir, su distribución muestral tiende a ser normal cuando el número de muestras crece; su media difiere de los valores de los parámetros aproximadamente en una cantidad $1/n$ siendo n el número de datos de la muestra; el sesgo $1/n$ se corrige usando estimadores insesgados.

La media y la variancia de las medias definidas por (2.1) están dadas por:

$$E\{\bar{X}\} = m_X \tag{5.1}$$

$$\mathrm{Var}\{\bar{x}\} = \sigma^2_{\bar{x}} = \frac{1}{n}\sigma^2_X \tag{5.2}$$

donde:

x : representa la variable aleatoria de la función de distribución.

m_X y σ^2_X: son la media y la variancia de la población, respectivamente.

Por lo que se refiere a la variancia muestral o segundo momento muestral, de (2.2) se tiene:

$$S^2_X = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

y de (4.8):

$$E\{S^2_X\} = m_{S^2_X} = \frac{n-1}{n}\sigma^2_X$$

De (4.9), se tiene también que un estimador insesgado de σ^2_X es:

$$S^{2*}_X = \frac{n}{n-1}S^2_X$$

ya que:

$$E\{S^{2*}_X\} = \sigma^2_X \tag{5.3}$$

Por lo tanto:

$$S_x^{2*} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \{\Sigma\, x_i^2 - n\bar{x}^2\} \qquad (5.4)$$

La variancia de S_x^{2*} está dada por (Ref. 2):

$$\mathrm{Var}\,\{S_x^{2*}\} = \frac{n^2}{(n-1)^2}\ \mathrm{Var}\,[S_x^2] \qquad (5.5)$$

$$\mathrm{Var}\,\{S_x^2\} = \frac{2(n-1)}{n^2}\ \sigma_x^4 \qquad (5.6)$$

Las raíces cuadradas de las expresiones (5.2) y (5.6) proporcionan el error estándar de los estimadores $\bar{x}$ y S_x^2 respectivamente.

5.2.2 *Error estándar de las estimaciones*

El error estándar de una estimación particular para el método de ajuste por momentos, está dado por:

$$S_T = \delta \sqrt{\frac{\sigma^2}{n}} \qquad (5.7)$$

donde:

$$\delta = \left[1 + K_0\gamma_1 + \frac{K_0^2}{4}(\gamma_2 - 1)\right]^{\frac{1}{2}} \quad \text{(Ref. 4)}. \qquad (5.8)$$

K_0; es llamado factor de frecuencia; depende del período de retorno y de los parámetros de la distribución de probabilidad empleada ya que cual--quier estimación puede expresarse como (Ref.4):

$$y_T = m_y + K_0\sigma_y \qquad (5.9)$$

y_T: es la magnitud del evento estimado para un período de retorno T.

m_y, σ_y: son la media y la desviación estándar de la población estimadas a partir de una muestra.

γ_1 en (5.8): es el coeficiente de asimetría definido por (2.40)

$$\gamma_2 = \frac{\mu^{(4)}}{\mu^{(2)^2}} \qquad \text{(capítulo 2)}$$

$\mu^{(2)}$ y $\mu^{(4)}$ son momentos centrales definidos por (2.34) ó (2.35)

γ_1 y γ_2 se estiman a partir de la muestra.

En la referencia 4 existen tablas que suministran valores del factor de frecuencia K_0 para diferentes períodos de retorno y distintas distribuciones.

Los límites de confianza para una estimación particular y_0 cuando los parámetros son estimados por el método de momentos, son:

$$\text{Límites} = \bar{y} + k_0\, S_y \pm k\, S_T \qquad (5.10)$$

donde:

$\bar{y}$ y S_y: son la media y la variancia de la muestra, respectivamente.

S_T: está dado por (5.7)

K : es una constante

El valor de K depende de la distribución que siguen las diferencias de los eventos estimados y los eventos observados (errores) y del nivel de confianza $1 - \alpha$. Si se considera que siguen una distribución normal, $K = z_{\alpha/2}$; si se considera que siguen una distribución t con $\nu = n-1-m$ grados de libertad (Ref. 1), $K = t_{\alpha/2,\nu}$; m es el número de parámetros estimados a partir de la muestra. Para el primer caso, se utiliza la tabla 1 y para el segundo, la tabla 2 (ambas del apéndice).

$0 \leq \alpha \leq 1$

5.2.3 *Propiedades de los estimadores*

Los estimadores de parámetros ajustados por momentos son asintóticamente eficientes (secc. 4.2.2) generalmente su eficiencia es menor que la unidad, para distribuciones asimétricas puede ser mucho menor que uno. Debido a que los momentos de la muestra son estimas consistentes de los momentos de la población, las estimas de los parámetros son generalmente consistentes.

A pesar de la eficiencia asintótica de los estimadores, como muchas veces es menor que uno, frecuentemente este método se usa sólo como una primera aproximación de los valores de los estimadores (estimas)

Si la distribución de frecuencias es simétrica y particularmente normal, la eficiencia puede ser igual o acercarse mucho a la unidad; en este caso, pueden descartarse otros métodos de estimación y aplicar el de momentos. Desafortunadamente, en Hidrología, la mayoría de las variables que se manejan son en mayor o menor grado sesgadas por lo que el uso de este método representa una pérdida en la estimación.

5.2.4 *Estimación de los parámetros y obtención del error estándar para*

algunas distribuciones de probabilidad.

Como fue visto en el capítulo 3, los momentos de una variable aleatoria están relacionados con los parámetros de la población (función de probabilidad que la describe). Enseguida se presenta una tabla (tabla 5.1) donde se relacionan los parámetros de las distribuciones estudiadas en el capítulo 3 con sus momentos respectivos. Los parámetros pueden estimarse al sustituir el momento correspondiente de la población por el de la muestra.

La obtención del error estándar requiere de aplicar (5.7) para lo cual se necesita el valor de δ, que puede obtenerse de tablas que existen en la referencia 4 para varias distribuciones, varios períodos de retorno y también varios tamaños de la muestra.

5.3 Mínimos cuadrados.

5.3.1. Descripción del método.

El método de estimación por mínimos cuadrados considera el ajuste de los valores observados (gastos máximos anuales para el caso de este trabajo) a un modelo lineal.

Un modelo lineal es aquel en que una variable (dependiente) y es afectada por otra u otras variables (independientes) $x_1, x_2, \ldots, x_n$ de la siguiente manera:

$$E\{y/x_1, x_2, \ldots, x_n\} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \qquad (5.11)$$

TABLA 5.1

DISTRIBUCION	RELACION DE LOS PARAMETROS CON LOS MOMENTOS
NORMAL	$\mu_y^{(1)'} = m_y$ $\mu_y^{(2)} = \sigma_y^2$
LOGNORMAL (3 *parámetros*)	$\mu_y^{(1)'} = \exp\{m_L + \frac{1}{2}\mu_L^{(2)}\} + a_L$ $\mu_y^{(2)} = \{\mu_y^{(1)'}\}^2 \{\exp(\sigma_L^2) - 1\}$ $\frac{\mu_y^{(3)}}{\{\mu_y^{(2)}\}^{3/2}} = V_{(y-a_L)} + 3V_{(y-a_L)}$ *
GUMBEL	$\mu_y^{(1)'} = \beta + \frac{\gamma}{\alpha}$ $\mu_y^{(2)} = \frac{\pi^2}{6\alpha^2}$ $\gamma = 0.5772156...$
EXPONENCIAL (2 *parámetros*)	$\mu_y^{(1)'} = \frac{1}{\lambda} + \alpha$ $\mu_y^{(2)} = \frac{1}{\lambda^2}$
GAMMA (3 *parámetros*)	$\mu_y^{(1)'} = \frac{k}{\lambda} + \delta$ $\mu_y^{(2)} = \frac{k}{\lambda^2}$ $\frac{\mu_y^{(3)}}{\{\mu_y^{(2)}\}^{3/2}} = \frac{2}{\sqrt{k}}$

* $V_{(y-a_L)}$ está dado por (3.35)

En la expresión (5.11):

$E\{\ \}$: es el valor esperado

y: es la variable dependiente

$x_1, x_2, \ldots, x_n$: son variables independientes

$\alpha, \beta_1, \beta_2, \ldots \beta_n$: son los parámetros del modelo

El modelo:

$$E\{Y\}_x = \alpha + \beta x \qquad (5.12)$$

es el caso más simple de modelos lineales, sólo correlaciona dos variables: X e Y. El segundo miembro representa una línea recta con pendiente β y ordenada al origen α. Este es el caso que se presenta en adelante.

Para llevar a cabo el ajuste de una serie de avenidas máximas anuales observadas a una función de distribución de probabilidad, resulta de utilidad correlacionar funciones del período de retorno $x_i = F(T_i)$ con las avenidas observadas $y_i = Q_{\text{máx anual } i}$, es decir, asociar a cada gasto un determinado período de retorno.

El ajuste por mínimos cuadrados obtiene estimas de α y de β, que minimizan la suma de las diferencias elevadas al cuadrado de los valores estimados y sus correspondientes valores observados.

Los estimadores de α y de β se representan por $\hat{\alpha}$ y $\hat{\beta}$; en forma semejante, a la estimación de y_i se le representa por $\hat{y}_i$; por lo tanto:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}\, x_i \qquad (5.13)$$

α y β son los verdaderos, pero desconocidos, parámetros poblacionales.

Deben encontrarse los valores de $\hat{\alpha}$ y de $\hat{\beta}$ que hagan mínima la suma:

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \{y_i - (\hat{\alpha} + \hat{\beta} x_i)\}^2 \qquad (5.14)$$

Las diferencias $y_i - \hat{y}_i$ son llamadas errores o residuos.

n:es el número de datos de la muestra.

Derivando parcialmente esta expresión con respecto a $\hat{\alpha}$ y con respecto a $\hat{\beta}$ e igualando ambas derivadas a cero; se tiene:

$$\frac{\partial}{\partial \hat{\alpha}}\left[\sum_{i=1}^{n} \{y_i - (\hat{\alpha} + \hat{\beta} x_i)\}^2\right] = \sum_{i=1}^{n} \frac{\partial}{\partial \hat{\alpha}}\{y_i - (\hat{\alpha} + \hat{\beta} x_i)\}^2$$

$$= \sum_{i=1}^{n} 2 \{y_i - (\hat{\alpha} + \hat{\beta} x_i)\} (-1) = 0 \qquad (5.15)$$

$$\frac{\partial}{\partial \hat{\beta}}\left[\sum_{i=1}^{n} \{y_i - (\hat{\alpha} + \hat{\beta} x_i)\}^2\right] = \sum_{i=1}^{n} \frac{\partial}{\partial \hat{\beta}} \{y_i - (\hat{\alpha} + \hat{\beta} x_i)\}^2$$

$$= \sum_{i=1}^{n} 2 \{y_i - (\hat{\alpha} + \hat{\beta} x_i)\} (-x_i) = 0 \qquad (5.16)$$

De (5.15):

$$- 2 \{ \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\alpha} - \sum_{i=1}^{n} \hat{\beta} x_i\} = 0 \qquad (5.17)$$

De (5.16):

$$-2 \{ \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \hat{\alpha} x_i - \sum_{i=1}^{n} \hat{\beta} x_i^2\} = 0 \qquad (5.18)$$

Ya que $\hat{\alpha}$ y $\hat{\beta}$ son constantes y como $\sum_{i=1}^{n} y_i = n \bar{y}$ y $\sum_{i=1}^{n} x_i = n\bar{x}$, se tiene:

De (5.17):

$$n\bar{y} - n\hat{\alpha} - n\hat{\beta}\bar{x} = 0 \tag{5.19}$$

De (5.18):

$$\sum_{i=1}^{n} x_i y_i - n\hat{\alpha}\bar{x} - \hat{\beta}\sum_{i=1}^{n} x_i^2 = 0 \tag{5.20}$$

Resolviendo simultáneamente (5.19) y (5.20) para $\hat{\alpha}$ y $\hat{\beta}$ se llega a:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} \tag{5.21}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{5.22}$$

Es posible expresar $\hat{\alpha}$ y $\hat{\beta}$ como funciones de la covariancia o del coefi--ciente de correlación. (Capítulo 2).

La correlación se calcula a partir de las avenidas máximas anuales obser-vadas y_i y de las funciones del período de retorno $x_i = F(T_i)$.

Como se vió en el capítulo 2, las expresiones de la variancia y de la covariancia están dadas respectivamente por:

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2\}$$

$$S_{x,y} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\}$$

por lo que (5.21) puede expresarse como:

$$\hat{\beta} = \frac{S_{x,y}}{S_x^2} \tag{5,23}$$

También se sabe de (2,8) que:

$$r_{x,y} = \frac{S_{x,y}}{S_x S_y}$$

Sustituyendo (2.8) en (5.23), $\hat{\beta}$ puede expresarse como:

$$\hat{\beta} = r_{x,y} \frac{S_y}{S_x} \tag{5.24}$$

S_x^2 y S_y^2 representan las correspondientes variancias de x y de y.

$r_{x,y}$ es el coeficiente de correlación entre las funciones del período de retorno x y los gastos observados y definido en 2.1.5

Al cuadrado del coeficiente de correlación ($r_{x,y}{}^2$) se le llama coefi-- ciente de determinación; cuando se usa el método de estimación por mí-- nimos cuadrados, puede expresarse como:

$$r_{x,y}{}^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{5.25}$$

ya que corresponde a:

$$r_{x,y}{}^2 = \frac{n\, S_{\hat{y}}^2}{n\, S_y^2} = \frac{S^2_{(\hat{\alpha} + \hat{\beta} x)}}{S_y^2} \tag{5.26}$$

siendo $\hat{y} = \hat{\alpha} + \hat{\beta} x$

Una de las propiedades de las variancias es (Ref. 2):

$$\text{Var}\,\{\hat{\alpha} + \hat{\beta}\, x\} = \hat{\beta}^2\ \text{Var}\,\{x\} \tag{5.27}$$

que aplicada a una muestra corresponde a:

$$S^2_{(\hat{\alpha} + \hat{\beta}\ x)} = \hat{\beta}^2\ S_x^2 \qquad (5,28)$$

Sustituyendo (5.28) en (5.26), se tiene:

$$r_{x,y}^2 = \frac{\hat{\beta}^2\ S_x^2}{S_y^2} \qquad (5.29)$$

Sustituyendo (5.23) en (5.29):

$$r_{x,y}^2 = \left\{ \frac{S_{x,y}}{S_x^2} \right\}^2 \frac{S_x^2}{S_y^2} = \frac{S_{x,y}}{S_x^2\ S_y^2} \qquad (5.30)$$

que es el cuadrado del coeficiente de correlación definido por (2.8)

Los momentos de los estimadores obtenidos por el método de mínimos cua-- drados ($\hat{\alpha}$ y $\hat{\beta}$) tienen los primeros y segundos momentos respectivamente (Ref. 1):

$$E\ \{ \hat{\alpha} \} = \alpha \qquad (5.31)$$

$$\text{Var}\ \{ \hat{\alpha} \} = \frac{\sigma^2}{n} \left\{ 1 + \frac{\bar{x}}{S_x^2} \right\} \qquad (5.32)$$

$$E\ \{ \hat{\beta} \} = \beta \qquad (5.33)$$

$$\text{Var}\ \{ \hat{\beta} \} = \frac{\sigma^2}{n\ S_x^2} \qquad (5.34)$$

Donde:

$E\ \{\ \}$: es el valor esperado definido por (2.27) y/o (2.32).

$\text{Var}\ \{\ \}$: es la variancia definida por (2.36)

σ^2 : representa la variancia de los errores $(y_i - \hat{y}_i)$;

es llamada variancia residual,

La variancia residual σ^2 se estima por medio de S^2 de la siguiente forma:

$$S^2 = \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{n}{n-2}(1 - r^2_{x,y})\, S^2_y \qquad (5.35)$$

Al sustituir S^2 por σ^2 en (5.32) y (5.34) y luego extraer raíz cuadrada a cada expresión, se tiene:

$$S_{\hat{\alpha}} = \sqrt{\frac{S^2}{n} \{1 + \frac{\bar{x}}{S^2_x}\}} \qquad (5.36)$$

$$S_{\hat{\beta}} = \sqrt{\frac{S^2}{nS^2_x}} \qquad (5.37)$$

que son respectivamente, los errores estándar de $\hat{\alpha}$ y de $\hat{\beta}$.

5.3.2 *Error estándar de las estimaciones.*

El error estándar $\sigma_{\bar{y}_{x_0}}$ para una estimación particular $\bar{y}_0$ correspondiente a un valor x_0 está dado por:

$$\sigma_{\bar{y}_{x_0}} = \sqrt{\frac{\sigma^2}{n} \{1 + \frac{(x_0 - \bar{x})^2}{S^2_x}\}} \qquad \text{(Ref.2)} \qquad (5.38)$$

$\bar{y}_{x_0}$ es el estimador de $\alpha + \beta x_0$

Si se toman límites de confianza, éstos están dados por:

$$\text{Límites} = \hat{\alpha} + \hat{\beta} x_0 \pm k\, S_{\bar{y}_{x_0}} \qquad (5.39)$$

k es una constante que depende del nivel de confianza y de la distribución que siguen los errores.

$\sigma_{\bar{y}_{x_0}}$ es estimado por $S_{\bar{y}_{x_0}}$ al sustituir en (5.38) S^2 por σ^2.

Si se considera que los errores siguen una distribución t (Ref. 2); $k = t_{\alpha/2,\, n-2}$ se obtiene de la tabla 2 del apéndice. Si se considera una distribución normal, $k=z_{\alpha/2}$ se obtiene de la tabla 1 del apéndice.

α es el nivel de significancia y se considera de la misma forma que en el método de ajuste por momentos.

Sustituyendo la estimación de $\sigma_{\bar{y}_{x_0}}$ en (5.35), los límites de confianza para una estimación particular $\hat{y}_{x_0}$ están dados por:

$$\text{Límites} = \hat{\alpha} + \hat{\beta}\, x_0 \pm k\sqrt{\frac{S^2}{n}\left\{1 + \frac{(x_0 - \bar{x})^2}{S_x^2}\right\}} \tag{5.40}$$

que obviamente proporcionan el intervalo de confianza.

5.3.3 *Propiedades de los estimadores.*

De (5.31) y (5.33), se observa (según cap. 4), que $\hat{\alpha}$ y $\hat{\beta}$ son estimadores insesgados de α y de β.

De los estimadores lineales insesgados de β, el estimador $\hat{\beta}$ obtenido por mínimos cuadrados tiene la mínima variancia, es decir, es eficiente.*

En forma similar, $\hat{\alpha}$ es un estimador eficiente de α.

---

* Teorema de Gauss-Markov (Ref.1)

Para que las estimaciones hechas a partir del método de mínimos cuadrados se consideren eficientes, deben satisfacer las siguientes condiciones:

1a. Que los residuos $y_i - \hat{y}_i$ sean normalmente distribuídos ó cuando menos, distribuídos simétricamente.

2a. Que la variancia de la población formada por los residuos sea independiente de la posición de un grupo dado de residuos ó que las desviaciones sean independientes.

3a. Que la variancia de la población formada por los residuos a lo largo de la línea de ajuste sea constante.

Es difícil, en Hidrología, satisfacer estas condiciones, sobre todo la 2a. y la 3a. ya que, generalmente, los residuos se incrementan cuando se incrementa $\hat{y}_i$.

5.3.4. Estimación de los parámetros y obtención del error estándar para algunas distribuciones de probabilidad.

En la expresión (5.13), varía la x_i, dependiendo de la distribución a la cual se ajusta. Se sabe también, que x_i = F (T) por lo que (5.13) puede también escribirse como:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}\ F(T)$$

La función del período de retorno F(T) depende de la distribución a la cual se ajusta. Igualando (2.57) con las respectivas funciones de distribución acumulada (cap. 3), los eventos y_i, las x_i así como $\hat{\alpha}$ y $\hat{\beta}$ en función de los parámetros de cada distribución se expresan como se presenta en la tabla 5.2

TABLA 5.2

DISTRIBUCION	x_i	$\hat{\alpha}$	$\hat{\beta}$	$\hat{y}$
NORMAL	z_i	m_y	σ_y	$\hat{\alpha} + \hat{\beta}\ x_i$
LOGNORMAL (2 parámetros)	z_i	m_L	σ_L	$exp\ \{\hat{\alpha} + \hat{\beta}\ x_i\}$
GUMBEL	$-Ln Ln(\frac{Ti}{Ti-1})$	β	$\frac{1}{\alpha}$	$\hat{\alpha} + \hat{\beta}\ x_i$
EXPONENCIAL	$Ln\ Ti$	α	$\frac{1}{\lambda}$	$\hat{\alpha} + \hat{\beta}\ x_i$

Estas mismas consideraciones se utilizan para obtener el error estándar de una estimación particular, es decir, x_i se sustituye por x_o y $\bar{x}$ es la media de las x_i en (5.38)

5.4 Máxima Verosimilitud

5.4.1 Descripción del método.

Es también, un método de estimación de parámetros ampliamente usado. Además, proporciona una aproximación de la distribución del estadístico usado como estimador.

Si se tienen n observaciones aleatorias, su distribución de probabilidad conjunta está dada por:

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n | \Theta_1, \Theta_2, \ldots, \Theta_m)$$

$$= f_{X_1}(x_1)\, f_{X_2}(x_2) \cdot \cdot \cdot f_{X_n}(x_n) = \prod_{i=1}^{n} f_X(x_i | \Theta_j) \qquad (5.41)$$

donde:

$x_1, x_2, \ldots, x_n$ son las n observaciones aleatorias.

$\Theta_j = \Theta_1, \Theta_2, \ldots, \Theta_m$ son los parámetros de la distribución.

m : es el número de parámetros.

En (5.41) se consideran conocidos $\Theta_1, \Theta_2, \ldots, \Theta_m$

Si se toma la posición de que sólo se tienen los valores observados de la muestra $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$ y no se conocen los parámetros $\Theta_1, \Theta_2, \ldots, \Theta_m$ como ocurre realmente, $\prod f_X(x_i | \Theta_j)$, es vista sólo como una función de Θ_j lo cual da lugar a la relativa verosimilitud de las observaciones de la muestra como una función de Θ. En símbolos:

$$L(\Theta_j | x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i | \Theta_j) \qquad (5.42)$$

A esta expresión se le llama *función de verosimilitud* de la muestra.

A partir de la función de verosimilitud se determinan los estimadores de los parámetros $\hat{\Theta}_j$:

Los valores de los m parámetros que maximizan la función de verosimilitud $L(\hat{\Theta}_j)$ son llamados estimadores de máxima verosimilitud.

Una técnica bastante empleada para la maximización de $L(\hat{\Theta}_j)$, es la de igualar a cero la primera derivada del logaritmo natural de $L(\hat{\Theta}_j)$. La función de verosimilitud y su logaritmo natural tienen un máximo para el mismo valor de $\hat{\Theta}$. Además, debido a que muchas distribuciones de probabilidad involucran la función exponencial, la maximización de su logaritmo natural se facilita.

Si m es el número de parámetros de la distribución de probabilidad, la maximización requiere resolver una serie de m ecuaciones:

$$\frac{\partial \, LnL}{\partial \hat{\Theta}_1} = 0$$

$$\frac{\partial \, LnL}{\partial \hat{\Theta}_2} = 0 \qquad (5.43)$$

$$\vdots$$

$$\frac{\partial \, LnL}{\partial \hat{\Theta}_m} = 0$$

donde L está dada por (5.42)

5.4.2 *Error estándar de las estimaciones.*

Se tratará el caso de una distribución de tres parámetros $\hat{\Theta}_1$, $\hat{\Theta}_2$, $\hat{\Theta}_3$, estimados por máxima verosimilitud:

Para el período de retorno T, se tiene un evento simple X_T de forma que:

$$X_T = f\,(\hat{\Theta}_1, \hat{\Theta}_2, \hat{\Theta}_3) \qquad (5.44)$$

Como T no varía, se tiene que (Ref. 4):

$$S_T^2 = \left[\frac{\partial X}{\partial \hat{\Theta}_1}\right]^2 \operatorname{Var} \hat{\Theta}_1 + \left[\frac{\partial X}{\partial \hat{\Theta}_2}\right]^2 \operatorname{Var} \hat{\Theta}_2 + \left[\frac{\partial X}{\partial \hat{\Theta}_3}\right]^2 \operatorname{Var} \hat{\Theta}_3 + 2\,\frac{\partial X}{\partial \hat{\Theta}_1}\,\frac{\partial X}{\partial \hat{\Theta}_2} \operatorname{cov}(\hat{\Theta}_1, \hat{\Theta}_2)$$

$$+ 2\,\frac{\partial X}{\partial \hat{\Theta}_1}\,\frac{\partial X}{\partial \hat{\Theta}_3} \operatorname{cov}(\hat{\Theta}_1, \hat{\Theta}_3) + 2\,\frac{\partial X}{\partial \hat{\Theta}_2}\,\frac{\partial X}{\partial \hat{\Theta}_3} \operatorname{cov}(\hat{\Theta}_2, \hat{\Theta}_3). \qquad (5.45)$$

donde:

S_T^2 es el cuadrado del error estándar para una estimación correspondiente a un período de retorno T.

X: es la variable aleatoria

$\frac{\partial X}{\partial \hat{\Theta}_i}$: representan las derivadas parciales de la variable aleatoria X con respecto a Θ_i

Las derivadas parciales pueden obtenerse de la función de verosimilitud; y las variancias están dadas por (Ref. 4):

$$\operatorname{Var} \hat{\Theta}_1 = \left\{ \frac{\partial^2 L}{\partial \hat{\Theta}_2^2}\,\frac{\partial^2 L}{\partial \hat{\Theta}_3^2} - \left[\frac{\partial^2 L}{\partial \hat{\Theta}_2\,\partial \hat{\Theta}_3}\right]^2 \right\} \Big/ D \qquad (5.46)$$

donde D está dada por:

$$D = \frac{\partial^2 L}{\partial \hat{\Theta}_1^2} \left\{ \frac{\partial^2 L}{\partial \hat{\Theta}_2^2}\,\frac{\partial^2 L}{\partial \hat{\Theta}_3^2} - \left[\frac{\partial^2 L}{\partial \hat{\Theta}_2\,\partial \hat{\Theta}_3}\right]^2 \right\}$$

$$+\frac{\partial^2 L}{\partial\hat{\Theta}_1\,\partial\hat{\Theta}_2}\left[\frac{\partial^2 L}{\partial\hat{\Theta}_1\,\partial\hat{\Theta}_2}\;\frac{\partial^2 L}{\partial\hat{\Theta}_3^{\,2}}-\frac{\partial^2 L}{\partial\hat{\Theta}_2\,\partial\hat{\Theta}_3}\;\frac{\partial^2 L}{\partial\hat{\Theta}_1\,\partial\hat{\Theta}_3}\right]$$

$$-\frac{\partial^2 L}{\partial\hat{\Theta}_1\,\partial\hat{\Theta}_3}\left[\frac{\partial^2 L}{\partial\hat{\Theta}_1\,\partial\hat{\Theta}_2}\;\frac{\partial^2 L}{\partial\hat{\Theta}_2\,\partial\hat{\Theta}_3}-\frac{\partial^2 L}{\partial\hat{\Theta}_2^2}\;\frac{\partial^2 L}{\partial\hat{\Theta}_1\,\partial\hat{\Theta}_3}\right] \tag{5.47}$$

Var $\hat{\Theta}_2$ y Var $\hat{\Theta}_3$ se obtienen en forma similar a (5.46)

De esta forma es posible evaluar $S_T^{\,2}$ y por consiguiente los límites e intervalos de confianza de las estimaciones para un determinado período de retorno. Los límites son:

$$\text{Límites} = X_T \pm k\, S_T \tag{5.48}$$

donde k depende del nivel de confianza empleado y de la distribución que siguen los errores al igual que para los otros dos métodos de estimación vistos.

5.4.3 *Propiedades de los estimadores.*

Los estimadores obtenidos por el método de máxima verosimilitud no son siempre insesgados pero sí asintóticamente insesgados. Son suficientes y consistentes. Si existen estimadores eficientes este método los encuentra. Además tienen la propiedad de ser invariantes (secc.4.2.4)

Las propiedades de estos estimadores son asintóticas, es decir, sus estimas se aproximan más al valor de los parámetros cuando el número de datos de la muestra crece.

Una gran ventaja de los estimadores obtenidos por máxima verosimilitud es que sus propiedades han sido ampliamente estudiadas y en forma independiente de las funciones de distribución de probabilidad en consideración, por lo que al aplicarse a cualquier distribución, se tiene en buen grado de confiabilidad.

5.4.4 *Estimación de los parámetros y obtención del error estándar para algunas distribuciones de probabilidad.*

El procedimiento para la estimación de parámetros y el de la obtención del error estándar utilizando el Método de Máxima Verosimilitud varía de una distribución a otra:

5.4.4.1 *Distribución Lognormal (dos y tres parámetros):*

La función de densidad de probabilidad está dada por (3.19); al aplicarle (5.42) a su logaritmo natural, resulta:

$$\mathrm{Ln}L = -\sum_{i=1}^{n} \mathrm{Ln}(y_i - \hat{a}_L) - \frac{n}{2}\mathrm{Ln}(2\Pi) - n\,\mathrm{Ln}\,\hat{\sigma}_L - \frac{\sum_{i=1}^{n}\{\mathrm{Ln}(y_i - \hat{a}_L) - \hat{m}_L\}^2}{2\hat{\sigma}^2_L} \qquad (5.49)$$

y al aplicar (5.43), se llega a:

$$\hat{m}_L = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Ln}(y_i - \hat{a}_L) \qquad (5.50)$$

$$\hat{\sigma}^2_L = \frac{1}{n}\sum_{i=1}^{n}\{\mathrm{Ln}(y_i - \hat{a}_L) - \hat{m}_L\}^2 \qquad (5.51)$$

y:

$$\sum_{i=1}^{n} (y_i - \hat{a}_L)^{-1} (\hat{m}_L - \hat{\sigma}^2_L) = \sum_{i=1}^{n} (y_i - \hat{a}_L)^{-1} \operatorname{Ln} (y_i - \hat{a}_L) \qquad (5.52)$$

Al sustituir $\hat{m}_L$ y $\hat{\sigma}^2_L$ en (5.52) puede resolverse el sistema y encontrar los valores de $\hat{a}_L$, $\hat{m}_L$ y $\hat{\sigma}_L$ que son los parámetros de ésta distribución.

Para la Distribución Lognormal de dos parámetros, el cálculo se simplifica a hacer $\hat{a}_L = 0$ y aplicar (5.50) y (5.51)

El error estándar resulta de aplicar (5.45). y_T está dada por:

$$y_T = \exp (\hat{m}_L + z\hat{\sigma}_L) + \hat{a}_L \qquad (5.53)$$

por lo que (de la ec. 5.45):

$$S_T{}^2 = \operatorname{Var} \hat{a}_L + \frac{z^2 \omega^2}{4 \hat{\sigma}_L{}^2} \operatorname{Var} \hat{\sigma}^2_L + \omega^2 \operatorname{Var} \hat{m}_L + \frac{z \omega}{\sigma_L} \operatorname{cov} (\hat{a}_L, \hat{\sigma}^2_L)$$

$$+ 2\omega \operatorname{cov} (\hat{a}_L, \hat{m}_L) + \frac{z\omega^2}{\hat{\sigma}_L} \operatorname{cov} (\hat{\sigma}^2_L, \hat{m}_L) \quad * \qquad (5.54)$$

donde:

z: es la variable normal estandarizada

$$\omega = \exp (\hat{m}_L + z \hat{\sigma}_L)$$

$$\operatorname{cov}(\hat{a}_L, \hat{m}_L) = - \frac{1}{2n_D} \exp (\hat{\sigma}^2_L/2 - \hat{m}_L)$$

$$\operatorname{cov}(\hat{a}_L, \hat{\sigma}^2_L) = \frac{1}{n_D} \hat{\sigma}^2_L \exp (\hat{\sigma}^2_L/2 - \hat{m}_L)$$

$$\operatorname{cov} (\hat{m}_L, \hat{\sigma}^2_L) = - \frac{1}{n_D} \hat{\sigma}^2_L \exp (\hat{\sigma}^2_L - 2\hat{m}_L)$$

---

* Referencia 4

$$\mathrm{Var}\{\hat{a}_L\} = \frac{1}{2nD}$$

$$\mathrm{Var}'\{\hat{m}_L\} = \frac{\hat{\sigma}_L^2}{nD}\left[\frac{(\hat{\sigma}_L^2 + 1)}{2\hat{\sigma}_L^2}\exp\{2(\hat{\sigma}_L^2 - \hat{m}_L)\} - \exp(\hat{\sigma}_L^2 - 2\hat{m}_L)\right]$$

$$\mathrm{Var}'\{\hat{\sigma}_L^2\} = \frac{\hat{\sigma}_L^2}{nD}\left[(\hat{\sigma}_L^2 + 1)\exp'\{2(\hat{\sigma}_L^2 - \hat{m}_L)\} - \exp(\hat{\sigma}_L^2 - 2\hat{m}_L)\right]$$

$$D = \frac{(\hat{\sigma}_L^2 + 1)}{2\hat{\sigma}_L^2}\exp\{2(\hat{\sigma}_L^2 - \hat{m}_L)\} - \frac{(2\hat{\sigma}_L^2 + 1)}{2\hat{\sigma}_L^2}\exp(\hat{\sigma}_L^2 - 2\hat{m}_L)$$

Para la Distribución Lognormal de dos parámetros, el cálculo del error estándar se reduce a aplicar: *

$$S_T^2 = \frac{\hat{\sigma}_L^2}{n}\exp'\{2(\hat{m}_L + z\,\hat{\sigma}_L)\}(1 + z^2/2) \qquad (5.55)$$

5.4.4.2 *Distribución Gumbel:*

Aplicando (5.42) al logaritmo natural de (3.37), se obtiene:

$$\mathrm{Ln}\,L = n\,\mathrm{Ln}\,\hat{\alpha} - \hat{\alpha}\sum_{i=1}^{n}(y_i - \hat{\beta}) - \sum_{i=1}^{n}\exp\{-\hat{\alpha}(y_i - \hat{\beta})\} \qquad (5.56)$$

y aplicando (5.43)

$$\hat{\beta} = \frac{1}{\hat{\alpha}}\,\mathrm{Ln}\left[\frac{n}{\sum_{i=1}^{n}\exp(-\hat{\alpha}\,y_i)}\right] \qquad (5.57)$$

---

* Referencia 4

$$\frac{\partial \text{Ln } L}{\partial \hat{\alpha}} = F(\hat{\alpha}) = \sum_{i=1}^{n} y_i \exp(-\hat{\alpha}\, y_i) - (m_y - \frac{1}{\hat{\alpha}}) \sum_{i=1}^{n} \exp(-\hat{\alpha}\, y_i) = 0 \qquad (5.58)$$

La expresión (5.58) no puede resolverse analíticamente para $\hat{\alpha}$, sin embargo, por medio del método de Newton Raphson se puede llegar a obtener un valor bastante aproximado (de $\hat{\alpha}$).

Haciendo:

$$h = \frac{-F(\hat{\alpha}_i)}{F'(\hat{\alpha}_i)} \qquad (5.59)$$

$$\hat{\alpha}_{i+1} = \hat{\alpha}_i + h \qquad ; i = 1, 2, \ldots$$

donde:

$F'(\hat{\alpha}_1)$ es la primer derivada de $F(\hat{\alpha}_1)$:

$$F'(\hat{\alpha}_1) = -\sum_{i=1}^{n} y_i^2 \exp(-\hat{\alpha}_1 y_i) + (m_y - \frac{1}{\hat{\alpha}_1}) \cdot \sum_{i=1}^{n} y_i \exp(-\hat{\alpha}_1 y_i)$$

$$- \frac{1}{\hat{\alpha}_1^2} \sum_{i=1}^{n} \exp(-\hat{\alpha}_1\, y_i) \qquad (5.60)$$

y sustituyendo el nuevo valor α_{i+1} en (5.59) hasta que h sea menor que un valor pequeño fijado de antemano (por ejemplo 0.0001 ó 0.000001) se obtiene un valor de $\hat{\alpha}$ por este método. Sustituyendo este valor de $\hat{\alpha}$ en (5.57), se obtiene el correspondiente valor de $\hat{\beta}$.

Aplicando (5.45), el error estándar para una estimación particular está dado por:

$$S_T^2 = \frac{1}{n\,\hat{\alpha}^2} \left[1.1086 + 0.5140\, y_G + 0.6079\, y_G^2 \right] \qquad (5.61)$$

donde:

$$y_G = - \text{Ln} \left(\text{Ln} \frac{T}{T-1} \right)$$

5.4.4.3 *Distribución Gamma (dos y tres parámetros)*

Al aplicar (5.42) al logaritmo natural de la función de densidad dada por (3.55) resulta: *

$$\text{Ln } L = n\,\Gamma\,(\hat{k}) - \hat{\lambda} \sum_{i=1}^{n} (x_i - \hat{\delta}) + (\hat{k}-1) \sum_{i=1}^{n} \text{Ln}\,(x_i - \hat{\delta}) - n\hat{k}\,\text{Ln}\left(\frac{1}{\hat{\lambda}}\right) \qquad (5.62)$$

y al aplicar (5.43):

$$\frac{\partial \text{LnL}}{\partial \hat{\lambda}} = \hat{\lambda}^2 \sum_{i=1}^{n} (x_i - \hat{\delta}) - n\hat{\lambda}\,\hat{k} = 0 \qquad (5.63)$$

$$\frac{\partial \text{LnL}}{\partial \hat{k}} = n \frac{\Gamma'(\hat{k})}{\Gamma(\hat{k})} + \sum_{i=1}^{n} \text{Ln}\,(x_i - \hat{\delta}) - n\text{Ln}\,\frac{1}{\hat{\lambda}} = 0 \qquad (5.64)$$

$$\frac{\partial \text{LnL}}{\partial \hat{\delta}} = n\hat{\lambda} - (\hat{k}-1) \sum_{i=1}^{n} \frac{1}{x_i - \hat{\delta}} = 0 \qquad (5.65)$$

donde:

$$\frac{\Gamma'(\hat{k})}{\Gamma(\hat{k})} = \psi\,(\hat{k}) = \text{Ln}(\hat{k}-2) - \frac{1}{2(\hat{k}+2)} - \frac{1}{12(\hat{k}+2)^2} + \frac{1}{120(\hat{k}+2)^4}$$

$$- \frac{1}{252(\hat{k}+2)^6} - \frac{1}{\hat{k}+1} - \frac{1}{\hat{k}} \qquad (5.66)$$

---
* Referencia 4

A Ψ se le llama función digamma.

Despejando $\hat{\lambda}$ tanto de (5.63) como de (5.65); igualando y despejando después $\hat{k}$, se llega a:

$$\hat{k} = 1/\left[1 - \frac{n^2}{\sum_{i=1}^{n}(x_i-\hat{\delta})\sum_{i=1}^{n}\frac{1}{x_i-\hat{\delta}}}\right] \tag{5.67}$$

Sustituyendo (5.67) en (5.63):

$$\hat{\lambda} = \left[\frac{\sum_{i=1}^{n}(x_i-\hat{\delta})}{n} - \frac{n}{\sum_{i=1}^{n}\frac{1}{(x_i-\hat{\delta})}}\right]^{-1} \tag{5.68}$$

Al sustituir (5.67) y (5.68) en (5.64) se obtiene una expresión que sólo puede resolverse numéricamente en forma iterativa.

Haciendo: *

$$F(\hat{\delta}) = \mathrm{Ln}\,k_1 - \Psi(k_1) + \frac{1}{n}\sum_{i=1}^{n}\mathrm{Ln}\,(x_i-\delta_1) - \mathrm{Ln}\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\delta_1)\right\} \tag{5.69}$$

y

$$F'(\hat{\delta}) = \left[\frac{1}{k_1} - \Psi'(k_1)\right]\left[\left(\sum_{i=1}^{n}\frac{1}{(x_i-\delta_1)} - \frac{n^2}{\sum_{i=1}^{n}(x_i-\delta_1)}\right)\sum_{i=1}^{n}\frac{1}{(x_i-\delta_1)^2} - \sum_{i=1}^{n}\frac{1}{(x_i-\delta_1)}\left(\sum_{i=1}^{n}\frac{1}{(x_i-\delta_1)^2} - \frac{n^3}{\{\sum_{i=1}^{n}(x_i-\delta_1)\}^3}\right)\right] - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{(x_i-\delta_1)}$$

$$+ \frac{n}{\sum_{i=1}^{n}(x_i-\delta_1)} \tag{5.70}$$

---

* Referencia 4

donde:

$$\psi'(k_1) = \frac{1}{k_1+2} + \frac{1}{2(k_1+2)^2} + \frac{1}{6(k_1+2)^3} + \frac{1}{30(k_1+2)^5} + \frac{1}{42(k_1+2)^7}$$

$$- \frac{1}{30(k_1+2)^9} + \frac{1}{(k_1+1)^2} + \frac{1}{k_1^2} \qquad (5.71)$$

El valor de k_1 se obtiene aplicando (5.66), sustituyendo δ_1 por $0.98x_m$, siendo x_m el menor de todos los gastos máximos anuales (primer iteración).

Cuando $h = F(\hat{\delta})/F'(\hat{\delta})$ es menor que un determinado valor pequeño fijado de antemano (por ejem. 0.0001), se considera que la aproximación a los valores estimados de los parámetros es suficiente (Método de Newton-Raphson).

$F'(\hat{\delta})$: es la primer derivada de $F(\hat{\delta})$

$\psi'(\hat{k})$: es llamada función trigamma.

El error estándar para esta distribución (ec. 5.45), está dado por:

$$S_T^2 = \{\frac{\partial x}{\partial \lambda}\}^2 \text{Var } \lambda + \{\frac{\partial x}{\partial k}\}^2 \text{Var } k + \{\frac{\partial x}{\partial \delta}\}^2 \text{Var } \delta + 2\frac{\partial x}{\partial \lambda}\frac{\partial x}{\partial k} \text{cov } (\lambda,k)$$

$$+ 2\frac{\partial x}{\partial \lambda}\frac{\partial x}{\partial \delta} \text{cov } (\lambda,\delta) + 2\frac{\partial x}{\partial k}\frac{\partial x}{\partial \delta} \text{cov } (k,\delta) \qquad (5.72)$$

donde: *

$$\frac{\partial x}{\partial \alpha} = \left[k^{1/3} - \frac{1}{9k^{2/3}} + \frac{z}{3k^{1/6}}\right]^3 \qquad (5.73)$$

$$\frac{\partial x}{\partial k} = 3\alpha \left[k^{1/3} - \frac{1}{9k^{2/3}} + \frac{z}{3k^{1/6}}\right]^2 \left[\frac{1}{9k^{2/3}} + \frac{2}{27k^{5/3}} - \frac{z}{18k^{7/6}}\right]$$

(5.74)

---

* Referencia 4

$$\frac{\partial x}{\partial \delta} = 1 \tag{5.75}$$

$$\text{Var } \alpha = \frac{1}{n\,\alpha^2 D}\left[\frac{\Psi(k)}{k-2} - \frac{1}{(k-1)^2}\right] \tag{5.76}$$

$$\text{Var } k = \frac{2}{nD\alpha^4\,(k-2)} \tag{5.77}$$

$$\text{Var } \delta = \frac{k\Psi'(k)-1}{n\,\alpha^2\,D} \tag{5.78}$$

$$\text{cov }(\alpha,k) = \frac{-1}{n\,\alpha^3 D}\left[\frac{1}{k-2} - \frac{1}{k-1}\right] \tag{5.79}$$

$$\text{cov}(\alpha,\delta) = \frac{1}{n\alpha^2 D}\left[\frac{1}{k-1} - \Psi'(k)\right] \tag{5.80}$$

$$\text{cov}(k,\delta) = -\frac{1}{n\,\alpha^3 D}\left[\frac{k}{k-1} - 1\right] \tag{5.81}$$

$$D = \frac{1}{(k-2)\alpha^4}\left[2\Psi'(k) - \frac{2k-3}{(k-1)^2}\right] \tag{5.82}$$

$$\alpha = \frac{1}{\lambda}$$

$\Psi(k)$: está dado por (5.66)

$\Psi'(k)$ está dado por (5.71)

z está dada por (3.66)

Ejemplo 5.1.

Ajustar los valores máximos de 30,60 y 100 conjuntos de 365 números que siguen una Distribución Exponencial (máximos exponenciales) a las Distribuciones Normal, Lognormal (de dos parámetros), Gumbel y Exponencial. Calcular la Desviación Estándar de los Errores.

Los conjuntos de números con Distribución Exponencial para una media m y una desviación estandar σ de interés, se obtienen de la siguiente forma (Ref. 13):

a) Dados 365 números aleatorios con distribución uniforme entre cero y uno, estos representan la probabilidad acumulada dada por $F_y\ (y)$.

De (3.51) se tiene que:

$$F_y\ (y) = 1 - \exp\{ -\lambda\ (y-\alpha\ )\}$$

de donde puede despejarse

$$y = \alpha - \frac{1}{\lambda}\ \{ 1 - F_y(y)\ \} \qquad (5.83)$$

b) La media y la desviación estándar de la distribución se asignan como:

$$m_y = 30$$

$$\sigma_y = 10$$

c) Los parámetros λ y α se obtienen aplicando (3.53) y (3.52) respectivamente:

$$\lambda = \frac{1}{\sigma_y}$$

$$\alpha = m_y - \frac{1}{\lambda}$$

d) Para $m_y = 30$ y $\sigma_y = 10$ se obtiene que $\lambda = 0.1$ y $\alpha = 20$.

e) Sustituyendo estos valores de λ y de α en (5.83), se obtienen los 365 números y con Distribución Exponencial:

$$y = 20 - 10 \text{ Ln} \{ 1 - F_y (y)\} \qquad 0 \leq F_y (y) \leq 1$$

De 30 , 60 y 100 conjuntos de 365 números así obtenidos se escogen los máximos y se ajustan a las cuatro distribuciones.

El ajuste se efectúa utilizando el método de mínimos cuadrados (secc.5.3) para lo cual es necesario correlacionar dos variables (X e Y). Represen tando por y a los eventos exponenciales, las x son funciones del período de retorno que dependen de la distribución a la cual se ajusta:

Para la Normal: $x = z$. Se obtiene a partir de 1 - 1/T de la tabla 1 del apéndice de áreas bajo la curva normal estandarizada.

Para la Lognormal: $x = z$

Para la Gumbel: $x = -\text{Ln Ln } T/(T-1)$

Para la Exponencial: $x = \text{Ln } T$

Los estimadores por mínimos cuadrados $\hat{\alpha}$ y $\hat{\beta}$ se obtienen aplicando (5.21) y (5.22). Las estimaciones $\hat{y}i$ resultan de aplicar (5.13):

$$\hat{y}i = \hat{\alpha} + \hat{\beta} x i$$

La Desviación Estándar de los Errores S se calcula a partir de (5.35)

$$S = S^2 = \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (\hat{y}i - yi)^2$$

En la tabla 5.3 se presentan los resultados del cálculo de la Desviación Estándar de los Errores para los ajustes respectivos de 30, 60 y 100 máximos exponenciales y en las figuras 5.1, 5.2 y 5.3 las curvas de ajuste junto con los máximos exponenciales.

Se observa para este ejemplo, que cuando el número de máximos cambia (30, 60, 100),los valores de la Desviación Estándar de los Errores (D.E.) no son correspondientes en cuanto a su magnitud; es decir, si una distribución tiene la máxima D.E. para 30 máximos, no lo tiene para 60 o para 100.

TABLA 5.3

DISTRIBUCION	MAXIMOS	D.E. = S
Normal	30	1.712
Lognormal	30	1.871
Gumbel	30	2.661
Exponencial	30	4.485
Normal	60	4.204
Lognormal	60	3.285
Gumbel	60	1.831
Exponencial	60	1.385
Normal	100	3.756
Lognormal	100	2.761
Gumbel	100	1.686
Exponencial	100	2.232

La Distribución Normal presenta el menor valor de la D.E. cuando $n = 30$, en cambio, presenta el mayor cuando $n = 60$ y cuando $n = 100$

La Distribución Exponencial tiene el mayor valor de la D.E. para $n = 30$, el más bajo para n = 60 y el segundo más bajo para n = 100.

La Distribución Gumbel en cambio, presenta para n = 30 el segundo valor (más alto), para n = 60, el tercero y para n = 100 el más bajo.

Estas observaciones de los resultados del cálculo de D.E. coinciden con las de las figuras (figuras 5.1, 5.2 y 5.3):

La Distribución Normal, es la que más se asemeja a los puntos máximos para n = 30.

Entre la Distribución Gumbel y la Lognormal es difícil una elección de la que pueda representar las observaciones (puntos).

Para n = 60 la curva que siguen aproximadamente los puntos es la exponencial.

Para n = 100 la curva de ajuste Gumbel es a la que más se aproximan los puntos en los extremos, para los valores intermedios, la curva exponencial coincide con muchos.

Resulta interesante observar de este ejemplo, como la curva de la Distribución Gumbel se aproxima más a los valores máximos exponenciales cuando el número de máximos crece; en cambio, ocurre lo contrario con la Curva Normal. La Distribución Lognormal conserva aproximadamente su ajuste relativo.

Debe admitirse que este ensayo (para 30, 60 y 100 máximos) es sólo ilustrativo y que es necesaria la realización de otros más para confirmar o rechazar como válidas las consideraciones que de aquí puedan hacerse.

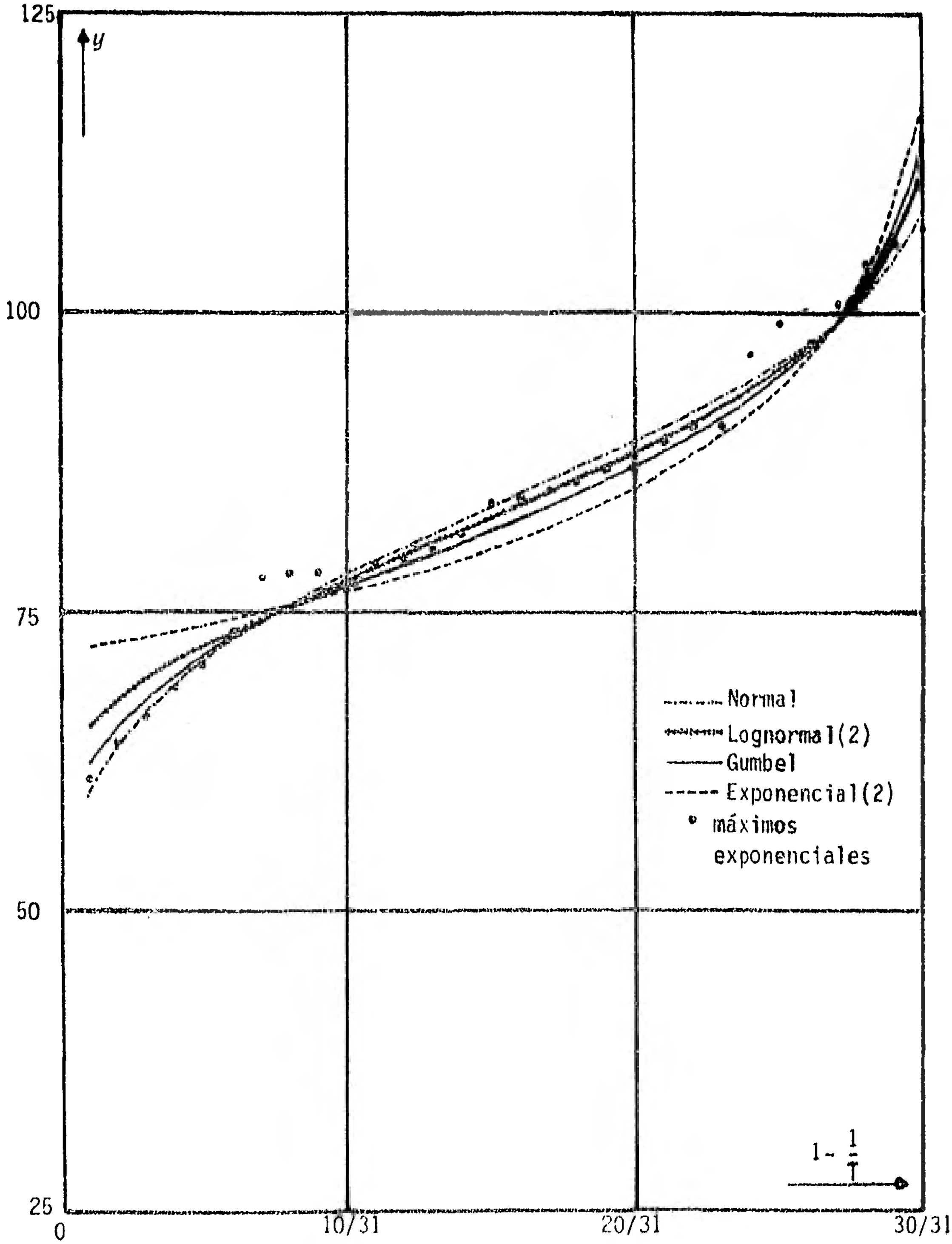


Figura 5.1. 30 eventos máximos exponenciales ajustados a 4 distribuciones.

Figura 5.2. 60 eventos máximos exponenciales ajustados a 4 distribuciones.

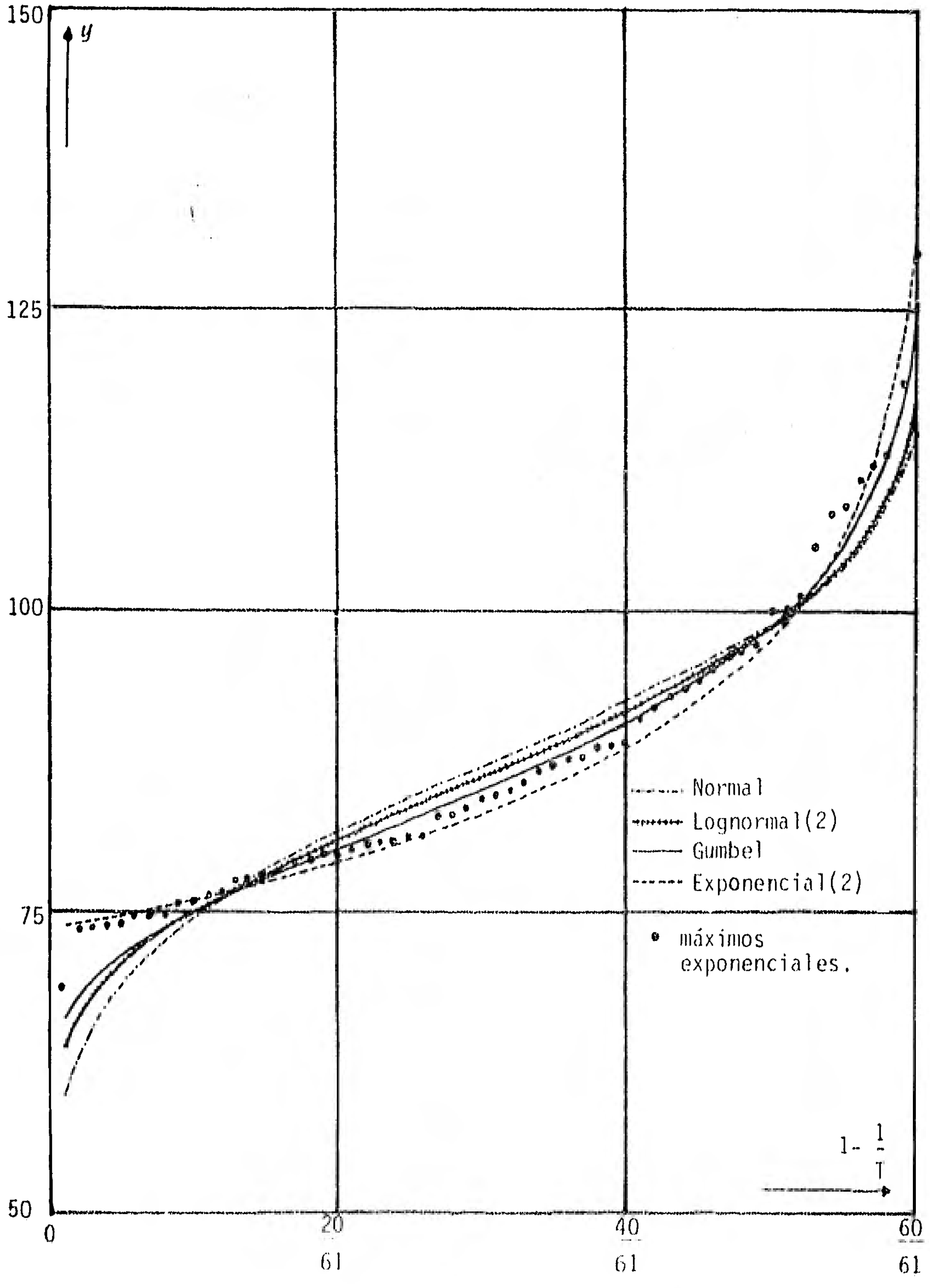
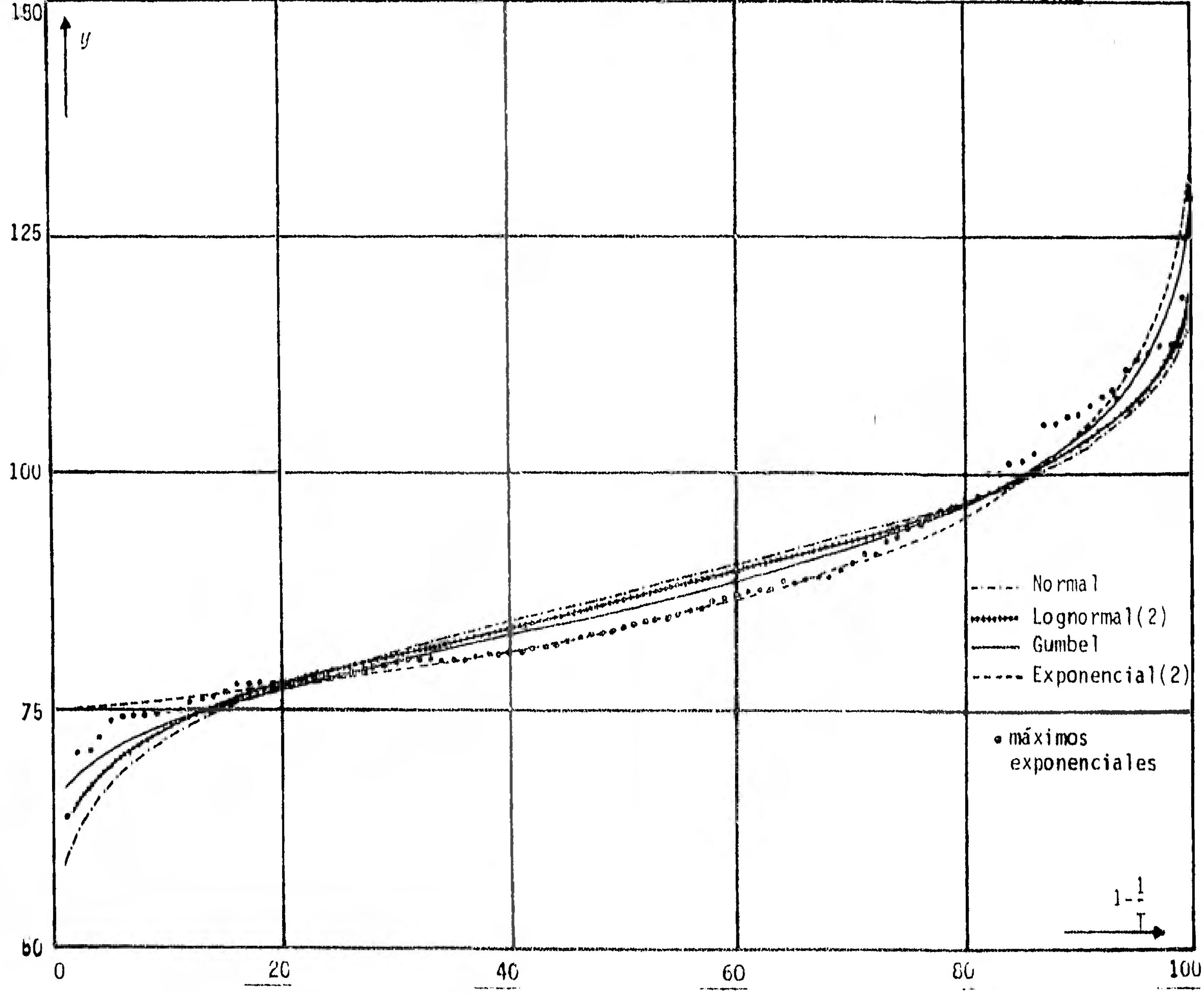


Figura 5.3 100 eventos máximos exponenciales ajustados a 4 distribuciones



6. BONDAD DE AJUSTE

6.1 Introducción

Al proponer una distribución de probabilidad como un modelo que siguen determinados eventos (gastos máximos anuales), después de estimar sus parámetros, se debe verificar su validez al compararla con la distribución que siguen las observaciones.

Resulta útil hacer primero una comparación gráfica del histograma de frecuencias relativas o su polígono con la función de densidad de probabilidad. También pueden compararse el polígono de frecuencias relativas acumuladas y la función de distribución acumulada. Cuando se usa un histograma, resulta conveniente que la función de densidad de probabilidad se presente en forma de barras que coincidan para que la comparación sea más sencilla y de mayor utilidad.

Existen papeles especiales, llamados papeles de probabilidad, en los cuales las escalas (horizontal o vertical) cambian según la distribución empleada que simplifican la comparación del polígono de frecuencias relativas acumuladas con la función de distribución acumulada. La comparación puede hacerse también, representando sólo con puntos las observaciones. El papel contiene escalas de tal forma que la función de distribución acumulada aparece como una línea recta en él. Así, la comparación se hace entre los puntos y una línea recta. Un ajuste ideal sería aquel en el que todos los puntos estuvieran sobre una línea recta.

Cuando se tiene una total incertidumbre acerca de cual modelo particular puede representar al conjunto de gastos máximos anuales, la búsqueda se inicia suponiendo que el "verdadero" modelo es una distribución de probabilidad conocida por ejemplo: Normal, Gumbel, Exponencial, etc. Una vez decididos los modelos a utilizar, después de calcular sus parámetros por los métodos de máxima verosimilitud, momentos y/o mínimos cuadrados, la selección del mejor modelo se hará de acuerdo al ajuste que presenten con las observaciones.

"Debido a la complejidad de los eventos hidrológicos, no hay un modelo probabilístico conocido al cual puedan ajustarse en forma ideal y por lo tanto tomar a éste como "verdadero", es decir, no hay eventos hidrológicos que sigan un modelo particular". Ref. (8)

Al comparar gráficamente los modelos propuestos con los datos, algunos pueden descartarse de inmediato, (es recomendable también una comparación en escalas aritméticas iguales); pero otros pueden ajustarse en forma gráfica de una manera igualmente aceptable, por lo que es necesario aplicar otros criterios de selección, es decir, pruebas de Bondad de Ajuste.

La Bondad de Ajuste usada en la aplicación de este tipo de pruebas es la forma por la cual se determina si una serie de datos puede ser expresada por los valores que toma una variable aleatoria habiendo propuesto su distribución.

Es importante al aplicar pruebas de Bondad de Ajuste hacer la siguiente consideración:

"El ajuste a distribuciones de probabilidad es considerado por conveniencia, ya que éstas están respaldadas por una firme teoría y además extraen la máxima información de una muestra por medio de técnicas propias. Si se usaran polinomios de orden alto, frecuentemente darían un ajuste mucho mejor que cualquiera de las distribuciones; no se usan porque no hay una justificación hidrológica" (Ref. 9).

El uso de Pruebas de Hipótesis* es común cuando se trata de decidir si una distribución teórica se ajusta a una serie de datos (una muestra). En términos generales, para el caso de la Bondad de Ajuste, consiste en formular una hipótesis, llamada *hipoteis nula*, acerca de que distribución siguen los datos; a esa hipótesis se le asocia una probabilidad llamada nivel de significancia (generalmente es de 0.1,0.05 ó 0.01) que es la máxima probabilidad con que se acepta tener el error de rechazar una hipótesis que debiera de aceptarse (error del tipo I). Al aceptar una hipótesis que debiera rechazarse se comete un error del tipo II.

Con el nivel de significancia (α) y con el número de datos (n) se obtiene un número llamado valor crítico que representa un límite que debe

---

* En varias referencias se trata en forma amplia este tema, para el desarrollo de este trabajo no se considera necesaria una descripción mayor.

compararse con el estadístico calculado en la prueba de Bondad de Ajuste correspondiente; en este trabajo, para las pruebas usadas, el valor del estadístico debe ser menor que el valor crítico ya que las hipótesis nulas son, en general, de la forma: "Los datos se ajustan al modelo probabilístico Θ"; cumpliendo esta condición, no se rechaza la hipótesis nula.

6.2. Prueba Xi cuadrado (X^2)

El uso de esta prueba de Bondad de Ajuste es muy generalizado para evaluar el ajuste de una serie de datos a una función de distribución.

Debe dividirse la muestra en k intervalos o categorías de datos que por facilidad es recomendable que sean los mismos del histograma.

La prueba X^2 compara el número de eventos simples observados con el número de eventos simples esperados en los intervalos correspondientes (intervalos de clase) según una determinada función de distribución de probabilidad.

El número de eventos esperados en un intervalo se obtiene multiplicando la probabilidad en ese intervalo por el número total de observaciones n y la probabilidad en un intervalo, restando la probabilidad acumulada del límite inferior de la probabilidad acumulada del límite superior, es decir, aplicando (2,20).

Debe calcularse:

$$D_1 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad (6.1)$$

donde:

D_1: es un estadístico que sigue una Distribución X^2 con n-1 grados de libertad (Ref. 2)

i: es el número de intervalo, $i = 1,2,\ldots,k$

Oi: es el número de observaciones en el intervalo i (frecuencia observada)

Ei: es el número de eventos simples esperados en el intervalo i (frecuencia esperada); depende de la distribución usada.

Si los parámetros se estiman a partir de la muestra, debe restarse un grado de libertad por cada uno.

Cuando se propone un modelo como representativo de la población a la que pertenecen los datos, se busca en la tabla 3 el valor crítico de X^2 para k-1-m grados de libertad y un nivel de significancia α (Prueba de hipótesis).

Debe cumplirse que:

$$D_1 \leq X^2_{1-\alpha,\ k-1-m} \tag{6.2}$$

m: es el número de parámetros estimados a partir de la muestra.

También es posible aplicar la prueba X^2 procediendo en forma inversa a la explicada, es decir, fijando primero un número constante de eventos simples esperados en cada intervalo y posteriormente, calcular los límites de los intervalos según la distribución empleada. En este caso, generalmente, los intervalos de clase tendrán diferentes tamaños y estarán en función de la distribución empleada.

Cuando se trata de seleccionar un modelo de varios propuestos siguiendo este criterio, se elige aquel que tiene el menor valor de D_1 al aplicar la prueba de la misma forma para todas las distribuciones. Es importante, cumplir con la condición de tener el mismo número de grados de libertad para todos los modelos al aplicar esta prueba (Ref. 8).

Ejemplo 6.1 Dados los siguientes gastos máximos anuales correspondientes a la estación Bolaños, Río Bolaños de la cuenca del Río Santiago, ajustarlos a las distribuciones normal y gumbel y aplicar la prueba X^2

AÑO	*GASTO MAXIMO* (m^3/seg)
1947	*428*
1948	*1 067*
1949	*268*
1950	*210*
1951	*363*
1952	*278*
1953	*467*
1954	*565*
1955	*777*
1956	*449*
1957	*109*
1958	*601*
1959	*447*
1960	*323*
1961	*385*
1962	*628*
1963	*812*
1964	*422*
1965	*377*
1966	*490*
1967	*1 302*
1968	*422*
1969	*188*
1970	*1 017*
1971	*804*

Llamando x_i a los gastos máximos anuales:

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{13\ 199}{25} = 527.\text{[illegible]}$$

$$S_X^2 = \frac{\Sigma(x_i-\bar{x})^2}{n} = 81\ 982.92$$

Estimando los parámetros por el método de ajuste por momentos (capítulo 5), se tiene para la Distribución Normal (secc. 3.1.1):

$$\hat{m}_X = \bar{x} = 527.96$$

$$\hat{\sigma}_X = S_X = 286.33$$

En la tabla 6.1 se presentan los intervalos de clase con su respectiva frecuencia así como los resultados y los cálculos necesarios para la aplicación de la prueba a la Distribución Normal.

TABLA 6.1

Intervalo	*frecuencia observada Oi*	*límite real de clase x*	*z*	*Probabilidad acumulada*	*Probabilidad en el intervalo*	*frecuencia esperada Ei*
		108.5	-1.465	0.0500		
109-307	5				0.1706	4.2650
		307.5	-0.770	0.2206		
308-506	11				0.2495	6.2375
		506.5	-0.075	0.4701		
507-705	3				0.2623	6.5575
		705.5	-0.620	0.7324		
706-904	3				0.1733	4.3325
		904.5	1.315	0.9057		
905-1103	2				0.0721	1.8025
		1 103.5	2.010	0.9778		
1 104-1302	1				0.0188	0.4700
		1 302.5	2.705	0.9966		
Σ	25				0.9466	

Los límites reales de clase se fijan para identificar en forma clara a que intervalo de clase pertenece un dato.

z: se obtiene al aplicar(3.2).

La probabilidad acumulada se obtiene de la tabla 1 del apéndice a partir del valor de z (secc. 3.1.1)

La probabilidad en el intervalo es la diferencia de probabilidades acumuladas de dos límites reales consecutivos y la frecuencia esperada, es el producto de esta probabilidad por el número de datos n.

De (6.1):

$$D_1 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

$$D_1 = \frac{(5.4.2650)^2}{4.2650} + \frac{(11-6.2375)^2}{6.2375} + \frac{(3-6.5575)^2}{6.5575} + \frac{(3-4.3325)^2}{4.3325}$$

$$+ \frac{(2-1.8025)^2}{1.8025} + \frac{(1-0.4700)^2}{0.4700}$$

$$D_1 = 6.7221$$

Para un nivel de significancia de 0.05 y 3 grados de libertad (ec. 6.2), se tiene de la tabla 3 del apéndice que:

$$X^2_{0.95,3} = 7.81$$

Como D_1 es menor que $X^2_{0.95,3}$, se cumple con (6.2) por lo que para este nivel, no se rechaza que los gastos máximos se ajustan a una distribución normal.

Procediendo ahora en forma inversa, es decir, fijando primero la frecuencia esperada para calcular después la frecuencia observada (tabla 6.2)

TABLA 6.2

Intervalo	frecuencia observada O_i	límite real de clase x	z	Probabilidad acumulada	Probabilidad en el intervalo	frecuencia esperada E_i
		$-\infty$	$-\infty$	0		
$-\infty$-251.94	3				0.1667	4.1667
		251.94	-0.964	0.1667		
251.95-403.69	6				0.1667	4.1667
		403.69	-0.434	0.3333		
403.70-527.96	6				0.1667	4.1667
		527.96	0	0.5000		
527.97-652.23	3				0.1667	4.1667
		652.23	0.434	0.6667		
652.24-803.98	2				0.1667	4.1667
		803.98	0.964	0.8334		
803.99-∞	5				0.1667	4.1667
		∞	∞	1.0000		
Σ	25				1.0000	25.0000

$$D_1' = \frac{1}{4.1667} \{ (3-4.1667)^2 + (6-4.1667)^2 + (6-4.1667)^2 + (3-4.1667)^2$$

$$+ (2-4.1667)^2 + (5-4.1667)^2 \}$$

$$D_1' = 3.5600$$

En este caso D_1' es menor que D_1 y también cumple con (6.2).

Se observa la gran diferencia entre los valores de D_1 y D_1' al aplicar la prueba X^2 en las dos diferentes formas a los mismos datos.

Ahora, se aplica la prueba X^2 a la Distribución Gumbel.

Ajustando por momentos los parámetros, se tiene de (3.38) y (3.39), respectivamente:

$$\beta = \bar{x} - \frac{0.577}{\alpha}$$

$$\alpha = \frac{\pi}{\sqrt{6}\, S_x}$$

Haciendo operaciones:

$$\alpha = 0.00448$$

$$\beta = 399.144$$

y calculando la probabilidad acumulada por medio de la ec. (3.36):

$$F_X(x) = \exp\{-\exp\left[-\alpha(x-\beta)\right]\}$$

En la tabla 6.3 se presenta el desarrollo de los cálculos en la obtención de la frecuencia esperada, que se realiza en forma similar a la explicada para la Distribución Normal.

TABLA 6.3

Intervalo	frecuencia observada O_i	Límite real de clase x	Probabilidad acumulada	Probabilidad en el intervalo	frecuencia esperada E_i
		108.5	0.0253		
109-307	5			0.1961	4.9033
		307.5	0.2214		
308-506	11			0.3175	7.9373
		506.5	0.5389		
507-705	3			0.2372	5.9295
		705.5	0.7761		
706-904	3			0.1252	3.1295
		904.5	0.9013		
905-1103	2			0.0570	1.4250
		1 103.5	0.9583		
1 104-1302	1			0.0244	0.6100
		1 302.5	0.98[illegible]7		

$$D_1 = \frac{(5-4.9033)^2}{4.9033} + \frac{(11-7.9373)^2}{7.9373} + \frac{(3-5.9295)^2}{5.9295} + \frac{(3-3.1295)^2}{3.1295}$$

$$+ \frac{(2-1.4250)^2}{1.4250} + \frac{(1-0.6100)^2}{0.6100}$$

$$D_1 = 3.1177$$

Como puede observarse, D_1 para la Distribución Gumbel es menor que el valor D_1 para la Distribución Normal. Para un nivel de significancia de .05 ambas distribuciones pueden considerarse como representativas de los datos siguiendo este criterio. Si se trata de elegir entre estas dos distribuciones, la que tiene el menor valor de D_1 es la que mejor se ajusta, en este caso, la Gumbel.

6.3 Prueba Kolmogorov-Smirnov.

También cuantifica la bondad de ajuste, es usada en forma alternativa con la prueba X^2. Consiste en calcular las diferencias en valor absoluto:

$$|P_x(x) - S_n(x)| \qquad (6.3)$$

y seleccionar la máxima:

$$D_2 = \max_{i=1}^{n} |P_x(x) - S_n(x)| \qquad (6.4)$$

donde:

$P_x(x)$: es la probabilidad acumulada que se obtiene a partir de la función de distribución acumulada empleada $F_x(x)$.

$S_n(x) = \frac{i}{n+1}$; i es el número de observación cuando están ordenadas de mayor a menor, n es el número total de observaciones, por lo que

(6.4) también puede expresarse como:

$$D_2 = \max_{i=1}^{n} \left| P_X(x) - \frac{i}{n+1} \right| \qquad (6.5)$$

Esta prueba también puede llevarse a cabo en forma gráfica. Si se usa papel de probabilidad, debe considerarse cuidadosamente la escala. Debe seleccionarse la desviación máxima en la escala de probabilidad de un punto a la línea teórica del modelo.

El uso de $i/(n+1)$ en lugar de i/n se debe en parte a que la escala del papel de probabilidad no está limitada de 0 a 1. En ocasiones, es conveniente dibujar $1-F_X(x)$ en lugar de $F_X(x)$. Cuando los datos se clasifican de mayor a menor su probabilidad acumulada está dada por (capítulo 2):

$$F_X(x) = 1 - \frac{1}{T} = 1 - \frac{1}{\left(\frac{n+1}{i}\right)} = 1 - \frac{i}{n+1} \qquad (6.6)$$

por lo cual también conviene el uso de $i/(n+1)$

Una vez calculado D_2, cuando se ha propuesto un modelo, debe compararse con un valor crítico d_2 que se obtiene de la tabla 4 del apéndice. Debe cumpirse que $D_2 > d_2$ a un nivel de significancia α para no rechazar que los datos se ajustan al modelo (Prueba de Hipótesis). En la tabla 4 se busca d_2 con n y α.

La prueba Kolmogorov-Smirnov tiene la ventaja sobre la X^2 de que compara los datos con el modelo en la forma en que se presentan, sin necesidad de agruparlos.

6.4 Suma de las diferencias al cuadrado.

Otra forma de cuantificar el ajuste es por medio de la suma de los cua-- drados de las diferencias de probabilidades observadas y probabilidades estimadas.

$$D_3 = \sum_{i=1}^{n} \{P_X(x) - S_n(x)\}^2 \qquad (6.7)$$

$P_X(x)$ y $S_n(x)$ son los mismos que para la prueba Kolmogorov-Smirnov.

Análogamente, puede calcularse la suma de las diferencias el cuadrado de los eventos máximos estimados y los correspondientes gastos observados:

$$\sum_{i=1}^{n} (\hat{x}_i - x_i)^2 \qquad (6.8)$$

donde:

n: es el número de datos de la muestra.

$\hat{x}_i$: es el evento estimado i obtenido a partir de la función de densidad de probabilidad o función de distribución acumulada.

x_i: es el evento observado i.

La variancia residual (ec. 3.35) está en función de (6.8) por lo que también proporciona una idea acerca de la bondad de ajuste cuando se comparan entre sí varios modelos (que modelo se ajusta mejor a los datos).

Ejemplo 6.2 Para los mismos datos del ejemplo 6.1, aplicar la prueba Kolmogorov-Smirnov y la suma de las Diferencias al Cuadrado de las probabilidades observadas y las probabilidades estimadas (acumuladas).

En la tabla 6.4 se muestra el desarrollo de las pruebas, esta tabla corresponde a la Distribución Normal; para la Distribución Gumbel, el desarrollo se efectúa en forma similar.

La primera columna (i) representa el rango del dato, es decir, su orden progresivo cuando todos los datos han sido ordenados de mayor a menor.

La segunda columna (y_i) son los gastos máximos anuales ordenados de mayor a menor.

La tercera columna contiene las probabilidades observadas calculadas a partir de (2.57).

$$P_X(x) = 1 - \frac{1}{T_{X(x)}}$$

siendo:

$$T_X(x) = T_i = \frac{n+1}{i}$$

y n, el número de datos de la muestra.

En la cuarta columna $S_n(x)$ se presentan las probabilidades estimadas, calculadas a partir de la respectiva función de distribución acumulada.

Las columnas quinta y sexta presenta respectivamente las diferencias en valor absoluto de las probabilidades observada y estimada, y las diferencias al cuadrado de las mismas probabilidades.

TABLA 6.4

i	y_i	$P_x(x)$	$S_n(x)$	$\lvert P_x(x)-S_n(x)\rvert$	$\{P_x(x)-S_n(x)\}^2$
1	1 302	0.9615	0.9966	0.0350	0.0012
2	1 067	0.9231	0.9701	0.0470	0.0022
3	1 017	0.8846	0.9562	0.0716	0.0051
4	812	0.8462	0.8394	0.0067	0.0001
5	804	0.8077	0.8325	0.0248	0.0006
6	777	0.7692	0.8078	0.0376	0.0015
7	628	0.7308	0.6366	0.0942	0.0089
8	601	0.6923	0.6007	0.0916	0.0084
9	565	0.6538	0.5515	0.1024	0.0105
10	490	0.6154	0.4473	> 0.1681 <	0.0283
11	467	0.5769	0.4157	0.1612	0.0260
12	449	0.5385	0.3914	0.1471	0.0216
13	447	0.5000	0.3887	0.1113	0.0124
14	428	0.4615	0.3635	0.0980	0.0096
15	422	0.4231	0.3557	0.0674	0.0045
16	422	0.3846	0.3557	0.0289	0.0008
17	385	0.3462	0.3088	0.0374	0.0014
18	377	0.3077	0.2990	0.0087	0.0001
19	363	0.2692	0.2827	0.0130	0.0002
20	323	0.2308	0.2371	0.0063	0.0000
21	278	0.1923	0.1913	0.0010	0.0000
22	268	0.1538	0.1820	0.0281	0.0008
23	210	0.1154	0.1334	0.0180	0.0003
24	188	0.0769	0.1176	0.0406	0.0017
25	109	0.0385	0.0717	0.0332	0.0011

Suma de las diferencias al cuadrado = 0.1473

Prueba Kolmogorov-Smirnov = 0.1681

Los resultados son los siguientes:

Para la Distribución Normal:

$$D_2 = \max_{i=1}^{n} \left| P_x(x) - \frac{i}{n+1} \right| = 0.1681$$

$$D_3 = \sum_{i=1}^{n} \left\{ P_x(x) - \frac{i}{n+1} \right\}^2 = 0.1473$$

Para la Distribución Gumbel

$$D_2 = \max_{i=1}^{n} \left| P_X (x) - \frac{i}{n+1} \right| = 0.1014$$

$$D_3 = \sum_{i=1}^{n} \{ P_X(x) - \frac{i}{n+1} \}^2 = 0.0496$$

De estos resultados se observa que tanto para la prueba Kolmogorov-Smirnov como para la Suma de las Diferencias al Cuadrado, los estadísticos correspondientes de la Distribución Gumbel son menores que los de la Normal.

Además, de la tabla 4 del apéndice se tiene que para $n = 25$ y un nivel de significancia $\alpha = 0.10$:

$$d_2 = 0.24$$

d_2 es mayor que D_2 para ambas distribuciones por lo que de acuerdo a este criterio, para un nivel de significancia del 10% tanto la Distribución Normal como la Distribución Gumbel pueden elegirse como representativas de las observaciones. Si se trata de seleccionar uno de los dos modelos, el mejor, de acuerdo a los resultados de estas pruebas es la Distribución Gumbel.

6.5 Muestras de verosimilitud.

Otro criterio de selección es el de clasificar los modelos por la función de verosimilitud de la muestra observada bajo cada modelo (Ref. 8), también llamada muestra de verosimilitud.

De la expresión (5.41), la muestra de verosimilitud dada $f_o(x)$ es:

$$f_{X_1}(x_1)\ f_{X_2}(x_2)\dots\ f_{X_n}(x_n) = \prod_{i=1}^{n} f_X(x_i|\Theta_j)$$

La muestra de verosimilitud (Ref. 2) representa la probabilidad de obtener la muestra observada como una función del verdadero estado de la naturaleza. También puede interpretarse como la relativa verosimilitud de los diversos estados de la naturaleza dadas las observaciones.

Es necesario definir probabilidad anterior, probabilidad posterior y probabilidad condicional para explicar el procedimiento para aplicar este criterio.

Las probabilidades anteriores o asignadas representan la evaluación de la información disponible antes de las observaciones que considera el ingeniero. Esta evaluación puede basarse en observaciones pasadas, en la semejanza con otros casos estudiados y también medidas subjetivas con base en la experiencia del posible modelo representativo. Cuando se proponen varios modelos, la suma de probabilidades asignadas debe ser 1; si se tiene una ignorancia total acerca del modelo que sigue el fenómeno bajo estudio, se asignan probabilidades iguales a cada uno de los propuestos. Las probabilidades anteriores se representan por $P'(\Theta)$, siendo Θ_i el modelo i; $i = 1,2,3,\dots$

Las probabilidades posteriores $P''(\Theta_i)$ se obtienen a partir de las probabilidades anteriores y de las observaciones hechas Z_k (datos).

El teorema de Bayes es utilizado en la obtención de la probabilidad posterior de un modelo Θ_i. A continuación se explica su obtención y enseguida su aplicación para obtener P" (Θ_i).

A partir de (2.13) se obtiene:

$$P(A \cap B) = P(A|B)\ P(B) \qquad P(B) \neq 0 \tag{2.13a}$$

$$P(A \cap B) = P(B|A)\ P(A) \qquad P(A) \neq 0 \tag{2.13b}$$

Si el evento B se descompone en K eventos simples, es decir:

$$B = B_1 + B_2 + \ldots + B_k = \sum_{i=1}^{k} B_i \tag{6.9}$$

De la figura 6.1 puede verse que:

$$P(A) = P\ (A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_k)$$

$$= \sum_{i=1}^{k} P(A \cap B_i) \tag{6.10}$$

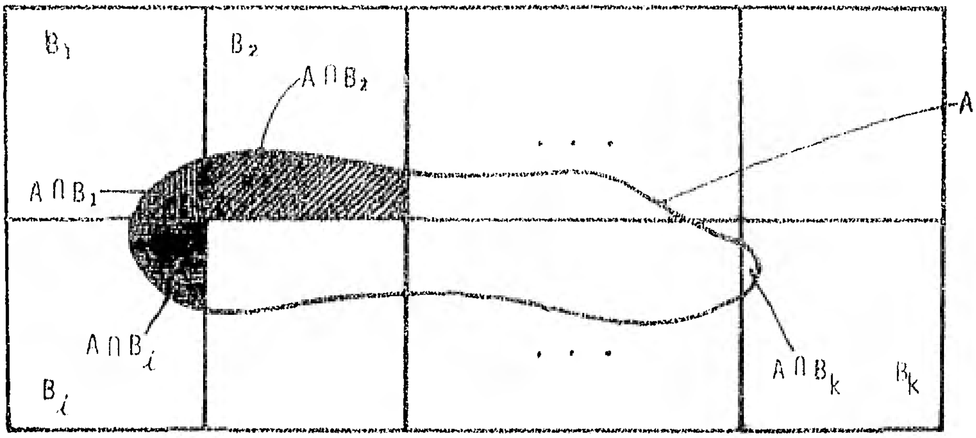


Figura 6.1 Ilustración del teorema de probabilidad total.

Igualando (2.13b) y (6.10):

$$P(A) = \sum_{i=1}^{k} P(A|B_i)\ P\ (B_i) \tag{6.11}$$

que es el teorema de probabilidad total.

Es claro que:

$$P(B_j \cap A) = P(A \cap B_j) \tag{6.12}$$

De (2.13)

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A \cap B_j)}{P(A)} \tag{6.13}$$

Sustituyendo (2.13) y (6.11) en (6.13) se tiene finalmente:

$$P(B_j|A) = \frac{P(A|B_j)\ P\ (B_j)}{\sum_{i=1}^{k} P(A|B_i)\ P(B_i)} \qquad j=1,2,\ldots\ \text{ó}\ k \tag{6.14}$$

A (6.14) se le conoce como el Teorema de Bayes.

Igualando (2.13a) y (2.13b):

$$P(A|B)\ P(B) = P(B|A)\ P(A) \qquad P(A) \neq 0,\ P(B) \neq 0 \tag{6.15}$$

Por lo que también puede considerarse:

$$P\ (O_i|\ z_k)\ P\ (z_k) = P\ (z_k|O_i)\ P'(O_i) \tag{6.16}$$

$P(z_k|O_i)$ es la probabilidad de que el resultado del experimento sea z_k cuando se considera O_i como el verdadero modelo.

$$P(\Theta_i|z_k) = \frac{P(z_k|\Theta_i)\ P'(\Theta_i)}{P(z_k)} = p''\ (\Theta_i) \tag{6.17}$$

Por el teorema de probabilidad total (Ec.6.11):

$$P(z_k) = \sum_{j=1}^{n} P(z_k|\Theta_j)\ P'(\Theta_j) \tag{6.18}$$

Por lo tanto:

$$P''(\Theta_i) = \frac{P(z_k|\Theta_i)\ P'(\Theta_i)}{\sum_{j=1}^{n} P(z_k|\Theta_j)\ P'(\Theta_j)} \qquad i = 1,2,\ldots\ \text{ó}\ n \tag{6.19}$$

Haciendo:

$$N = \frac{1}{\sum_{j=1}^{n} P(z_k|\Theta_j)\ P'(\Theta_j)} \tag{6.20}$$

se tiene:

$$P''(\Theta_i) = N\ \ P(z_k|\Theta_i)\ P'(\Theta_i) \tag{6.21}$$

A N se le llama factor de normalización y únicamente asegura que P" (Θ) forma una serie propia de probabilidades que suman 1.

La unión de la información nueva (observaciones) con la anterior (asignada) se hace por medio del producto $P(z_k|\Theta_i)\ P'(\Theta_i)$ que es llamado muestra de verosimilitud o probabilidad condicional del modelo Θ_i dadas las observaciones z_k. Para obtener $P\ (z_k|\Theta_i)$ debe obtenerse la probabilidad en el intervalo de clase (Ec. 2.20).

Los intervalos pueden ser los mismos que para la prueba X^2.

Si Θ_i es un modelo relativamente improbable para ser asociado con las observaciones z_k, esto se reflejará en $P''(\Theta_i)$ que será la más pequeña que $P'(\Theta_i)$.

De los modelos propuestos, el que se selecciona usando este criterio es aquel que presenta la máxima probabilidad posterior. La probabilidad anterior que se asigna es la misma para cada modelo.

El uso de Muestras de Verosimilitud para la selección de un modelo requiere tomar en cuenta que las probabilidades posteriores son muy sensibles a errores numéricos del factor de normalización N por lo que es necesario tener especial cuidado al realizar su cálculo; un error puede incrementar demasiado la probabilidad posterior de un modelo.

Ejemplo 6.3. Para los mismos datos del ejemplo 6.1 aplicar el criterio de muestras de verosimilitud para seleccionar una de las dos distribuciones (Normal o Gumbel).

Se utilizan por facilidad, los mismos intervalos de clase que en la prueba X^2; en la tabla 6.5 se presentan las probabilidades en los intervalos obtenidos en el ejemplo 6.1

Los modelos Θ_i en este caso son: Θ_1 = Normal y Θ_2 = Gumbel. Como se tiene una ignorancia total del modelo que siguen los gastos máximos anuales de este ejemplo, a cada distribución se le asigna la misma probabilidad anterior: $P'(\Theta_1) = P'(\Theta_2) = 0.5$

TABLA 6.5

	modelos Θ_i / resultados zk	Θ_1 = normal	Θ_2 = gumbel
z_1	< 109	0.0500	0.0253
z_2	109 - 307	0.1706	0.1961
z_3	308 - 506	0.2495	0.3175
z_4	507 - 705	0.2623	0.2372
z_5	706 - 904	0.1733	0.1252
z_6	905 -1 103	0.0721	0.0570
z_7	1 104 -1 302	0.0188	0.0244
z_8	> 1 302	0.0034	0.0173
	Σ	1.0000	1.0000

De (6.15) y (6.17):

$$P''(\Theta_i) = P(\Theta_i|z_k) = \frac{P(z_k|\Theta_i)\ P'(\Theta_i)}{\sum_{j=1}^{n} P(z_k|\Theta_i) P'(\Theta_j)}$$

Cálculo de $P(z_k|\Theta_i)$

$$P\{(z_2, z_3,\ldots,z_7) \mid \Theta_1\} = P(z_2 \mid \Theta_1)\ P(z_3 \mid \Theta_1)\ldots P(z_7 \mid \Theta_1)$$

$$= (0.1706)\ (0.2495)\ldots(0.0188) = 2.6226 \times 10^{-6}$$

$$P\{(z_2, z_3,\ldots,z_7) \mid \Theta_2\} = P(z_2 \mid \Theta_2)\ P(z_3 \mid \Theta_2)\ \ldots\ P(z_7|\Theta_2)$$

$= 2.5716 \times 10^{-6}$

ya que se trata de observaciones independientes.

Cálculo de $\sum_{j=1}^{n} P(z_k|\Theta_i)\, P'(\Theta_j)$:

$$\sum_{j=1}^{n} P\,(z_k|\Theta_i)\, P'(\Theta_j) = (2.6226 + 2.5716) \times 10^{-6} \times 0.5 = 2.5971 \times 10^{-6}$$

Aplicando (6.17)

$$P''(\Theta_1) = \frac{2.6226 \times 10^{-6}\ (0.5)}{2.5971 \times 10^{-6}} = 0.5049$$

$$P''(\Theta_2) = \frac{2.5716 \times 10^{-6}\ (0.5)}{2.5971 \times 10^{-6}} = 0.4951$$

Puede observarse que empleando este criterio, es mayor la probabilidad posterior de Θ_1 = normal en una cantidad muy pequeña por lo que la decisión de cuál distribución representa los datos resulta difícil utilizando esta prueba.

Ejemplo 6.4. Sabiendo que 30 eventos siguen una distribución de probabilidad, ajustar a la Distribución Normal, a la Lognormal de dos parámetros, a la Gumbel y a la Exponencial (de dos parámetros). Obtener la Desviación Estándar de los Errores (para cada ajuste), aplicar las pruebas Kolmogorov-Smirnov, Suma de las Diferencias al Cuadrado de las probabilidades "observada" y "estimada" (acumuladas), Xi Cuadrado y Muestras de Verosimilitud. Considérese que los 30 eventos tienen distribución a) Normal. b) Lognormal de dos parámetros. c)Gumbel. d) Exponencial.

Los 30 eventos (ajustados) están en función de la distribución de probabilidad que siguen y del período de retorno:

$$F_X(x) = 1 - \frac{1}{T_X(x)}$$

donde:

$T_X(x)$: es el período de retorno definido por (2.54):

$T_X(x) = T_i = \frac{n+1}{i}$; i es el rango definido en 2.2.8

$F_X(x)$ corresponde en este caso a la función de distribución acumulada del modelo.

Para la estimación de los parámetros al ajustar los 30 eventos a las cuatro distribuciones mencionadas se usa el método de mínimos cuadrados (secc.5.3), para lo cual deben correlacionarse dos variables (X e Y). Representando por y a los 30 eventos, los valores de x dependen de la distribución a la cual se ajusta (Tabla 5.2):

Para la Normal: $x = z$

Para la Lognormal: $x = z$ (se correlaciona con $\ln y$)

Para la Gumbel: $x = -\mathrm{Ln}\,\mathrm{Ln}\,\frac{T}{T-1}$

Para la Exponencial: $x = \mathrm{Ln}\,T$

Los estimadores por mínimos cuadrados $\hat{\alpha}$ y $\hat{\beta}$ se obtienen aplicando (5.21) y (5.22). Los valores estimados $\hat{y}_i$ resultan de aplicar (5.13):

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$$

La desviación estándar de los errores S se calcula aplicando (5.35):

$$S^2 = \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)$$

La prueba Kolmogorov-Smirnov consiste en aplicar (6.4)

$$D_2 = \max_{i=1}^{n} | P_x(x) - S_n(x) |$$

donde:

$P_x(x)$ y $S_n(x)$ son las probabilidades acumuladas "observada" y "estimada" respectivamente definidas en la sección 6.3.

La suma de las diferencias al cuadrado de las probabilidades "observada" y "estimada" acumuladas se obtiene al aplicar (6.7)

$$D_3 = \sum_{i=1}^{n} [P_x(x) - S_n(x)]^2$$

$P_x(x)$ y Sn (x) son los mismos que para la prueba Xolmogorov-Smirnov.

Para aplicar la prueba Xi cuadrado (x^2), debe calcularse (6.1)

$$D_1 = \sum_{i}^{k} \frac{(O_i - E_i)^2}{E_i}$$

El procedimiento se explicó en la sección 6.2, donde también se definieron i, k, Oi y Ei.

La Prueba Muestras de Verosimilitud se aplicó en la misma forma que en el ejemplo 6.3

Los cálculos de este ejemplo, se realizaron utilizando programas de computadora; los resultados se muestran en la tabla 6.6

Se observa de los resultados, que cuando los eventos se ajustan a la misma distribución con que fueron generados, en todos los casos, el ajuste es perfecto según los criterios de la Desviación Estándar de los Errores, Kolmogorov-Smirnov y Suma de las Diferencias al Cuadrado de las probabilidades "observadas" y probabilidades "estimadas"; por lo que respecta a la prueba X_i cuadrado y Muestras de Verosimilitud (M.V.), sólo los eventos que siguen una Distribución Gumbel se ajustan mejor a una distribución distinta (la Lognormal) a aquella con que fueron generados, debe notarse que las diferencias, correspondientes a estas pruebas, entre las distribuciones de ajuste es pequeña:

Lognormal	$D_1 = 1.0198$	M.V. = 0.2702
Gumbel	$D_1 = 1.0806$	M.V. = 0.2669
Normal		M.V. = 0.2641

Cuando los eventos ajustados siguen una Distribución Normal, según los resultados de las pruebas, el ajuste a una Distribución Exponencial es muy pobre ya que los valores D_1, D_2 y D_3 son muy altos así como el de la desviación estándar de los Errores. Para el caso de la Prueba X_i cuadrado, un intervalo resultó con probabilidad cero (Exponencial) y por consiguiente, frecuencia esperada cero, por lo que se calculó un sumando de (6.1) como 0_i^2; al aplicar el criterio de Máxima Verosimilitud, esta distribución (exponencial), tuvo por lo tanto, una probabilidad posterior igual a cero.

Sucede algo similar cuando los eventos ajustados siguen una Distribución Normal y se ajustan a una Exponencial, aunque los valores de los estadísticos de las pruebas de Bondad de Ajuste indican un mejor ajuste que en el caso de eventos normales ajustados a una Distribución Exponen cial.

En general, se observa para este ejemplo que la Distribución Gumbel es la que presenta mejor ajuste a los 30 eventos después de la distribu-- ción que se sabe siguen éstos.

De la tabla 4 del apéndice, para $n = 30$ y un nivel de significancia $\alpha = 0.10$, se obtiene un valor critico $d_2 = 0.22$ que al compararlo con los valores D_2 que resultan de la prueba Kolmogorov-Smirnov se observa que es mayor en todos los casos, lo cual indica según este critero y para este nivel de significancia que los 30 eventos se ajustan aceptable-- mente a las cuatro distribuciones. Sucede algo similar al comparar el valor crítico $X^2_{0.90,3} = 6.25$ con los estadísticos D_1 (prueba X_i cuadrado) ya que sólo se rechazaría, para este ejemplo, que los eventos normales se ajustan a una Distribución Exponencial y que los eventos exponenciales se ajustan a una Distribución Normal, para los demás casos $D_1 < X^2_{0.90,3}$

Lo anterior sugiere el uso de un critero de selección cuando se tienen dos o más modelos que son aceptados como representativos de una muestra. Para el caso de la prueba X_i Cuadrado puede seleccionarse el modelo con menor valor del estadístico D_1 de todos los que no son rechazados (Ref. 8) y lo mismo puede hacerse en el caso de las otras pruebas ya que en general, todas coinciden en elegir el mismo modelo (valores correspon-- dientes) como el de mejor ajuste, o el que mejor se ajusta después de la distribución que se sabe siguen los 30 eventos así como en indicar cuál es el que peor se ajusta.

En el caso del criterio "Muestras de Verosimilitud", es propiamente un criterio de selección por lo que el que tiene mayor probabilidad posterior es el que resulta escogido de entre varios modelos propuestos.

TABLA 6.6

EVENTOS	DISTRIBUCION DE AJUSTE	PARAMETROS ESTIMADOS	VARIANCIA RESIDUAL	PRUEBA K-5	SUMA DE LAS DIFERENCIAS AL CUADRADO	PRUEBA X^2	MUESTRAS DE VEROSIMILITUD
NORMALES	NORMAL	m_y = 800 σ_y = 333.333	0	0	0	0.7345	0.4073
	LOGNORMAL	m_L = 6.5961 σ_L = 0.4841	102.2878	0.0811	0.0793	3.8199	0.3106
	GUMBEL	α = 0.00376 β = 657.5482	57.2423	0.0573	0.0418	3.1090	0.2821
	EXPONENCIAL	α = 481.9070 λ = 0.002972	109.7133	0.1613	0.1964	15.1037	0
LOGNORMALES	NORMAL	m_y = 787.8063 σ_y = 311.0782	61.5586	0.0646	0.0507	4.6644	0.2515
	LOGNORMAL	m_L = 6.6046 σ_L = 0.4001	0	0	0	1.1883	0.2662
	GUMBEL	α = 0.00388 β = 649.5926	7.1355	0.0098	0.0009	1.3740	0.2640
	EXPONENCIAL	α = 467.2561 λ = 0.002949	44.8504	0.1219	0.0626	1.4711	0.2185
GUMBEL	NORMAL	m_y = 789.3455 σ_y = 315.0722	55.0457	0,0572	0.0395	3.0965	0.2641
	LOGNORMAL	m_L = 6.6041 σ_L = 0,4101	10.6495	0.01045	0,0013	1.0198	0,2702
	GUMBEL	α = 0,0038476 β = 649.9823	0	0	0	1.0806	0.2669
	EXPONENCIAL	α = 467.4970 λ = 0.002937	52.2756	0.1290	0.0744	2.5910	0.1988
EXPONENCIALES	NORMAL	m_y = 781.7931 σ_y = 289.1791	101.7112	0.1172	0.1721	11.6601	0.1843
	LOGNORMAL	m_L = 6.6066 σ_L = 0.3487	58.9451	0.0723	0.0583	4.0546	0.2355
	GUMBEL	α = 0.0040534 β = 649.5046	50.3968	0.1021	0.0611	4.3998	0.2906
	EXPONENCIAL	α = 0.003 λ = 466.6667	0	0	0	0,6235	0.2896

7. APLICACIONES

En los capítulos anteriores se han tratado las bases teóricas sobre los que se apoyan las inferencias que se realizan a partir de una muestra (observaciones); en algunos casos se han presentado ejemplos numéricos que pretenden explicar lo tratado en los capítulos correspondientes. A excepción de los ejemplos 6.1, 6.2 y 6.3, que se refieren a gastos máximos anuales registrados por una estación hidrométrica, los demás tratan con números supuestos (ejemplo 3.1) y con números aleatorios. Del desarrollo y de los resultados, se hacen comentarios que resultan de utilidad para llegar a conclusiones que se mencionarán en el siguiente capítulo.

En el presente capitulo, se aplica la teoría descrita a los gastos máximos anuales registrados en la Estación Hidrométrica El Puente del Río Omitlán en el Estado de Guerrero.*

---

* Boletín Hidrológico y Climatológico No. 11 C.F.E. (Cuencas Pacífico Sur.

Los gastos máximos anuales (eventos simples máximos) son:

AÑO	GASTO (m^3/seg).	AÑO	GASTO (m^3/seg)
1953	285	1966	531
1954	816	1967	1 325
1955	639	1968	477
1956	594	1969	734
1957	224	1970	720
1958	319	1971	339
1959	246	1972	501
1960	637	1973	1 151
1961	1 052	1974	1 591
1962	428	1975	639
1963	1 086	1976	1 231
1964	521	1977	379
1965	430	1978	475

a) Clasificación de datos y asignación del período de retorno.

Se procede a ordenar los datos de mayor a menor asignando a cada uno su rango, período de retorno y consiguiente probabilidad observada (capítulo 2); esto se presenta en la tabla 7.1.

El rango se representa por i, los gastos máximos anuales por y_i, el período de retorno correspondiente, por T_i y la probabilidad observada acumulada por $P(y_i)$

TABLA 7.1

i	y_i	T_i	$P(y_i)$
1	1 591	27	0.9630
2	1 325	13.5	0.9259
3	1 231	9	0.8889
4	1 151	6.75	0.8519
5	1 086	5.40	0.8148
6	1 052	4.50	0.7778
7	816	3.86	0.7407
8	734	3.38	0.7037
9	720	3.00	0.6667
10	639	2.70	0.6296
11	639	2.45	0.5926
12	637	2.25	0.5556
13	594	2.08	0.5185
14	531	1.93	0.4815
15	521	1.80	0.4444
16	501	1.69	0.4074
17	477	1.59	0.3704
18	475	1.50	0.3333
19	430	1.42	0.2963
20	428	1.35	0.2593
21	379	1.29	0.2222
22	339	1.23	0.1852
23	319	1.17	0.1481
24	285	1.13	0.1111
25	246	1.08	0.0741
26	224	1.04	0.0370

b) *Cálculo de estadísticos*

La media, la variancia y el coeficiente de asimetría de estos datos se obtienen respectivamente a partir de las ecs. (2.1), (2.2) y (2.5). Se llega a los siguientes resultados:

$\bar{y} = 668.0769$

$S_y^2 = 126\ 338.9443;$ $S_y = 355.441862$

$g = 0.9594$

c) *Distribuciones de probabilidad empleadas y su ajuste.*

Las distribuciones a las cuales se ajustan los datos son estudiados en el capítulo 3:

Normal

Lognormal de dos parámetros

Lognormal de tres parámetros

Gumbel

Exponencial de dos parámetros

Gamma de dos parámetros

Gamma de tres parámetros

El ajuste se lleva a cabo por el método de Máxima Verosimilitud para las distribuciones Lognormal (de 2 y de 3 parámetros), Gumbel y Gamma (de 2 y de 3 parámetros); y por el método de Mínimos Cuadrados para las distribuciones Normal y Exponencial (capítulo 5).

d) *Estimación de los parámetros y cálculo de la desviación estándar de los errores.*

Los cálculos para la estimación de los parámetros y para la obtención de la desviación estándar de los errores $S = \{1/(n-2).(\hat{y}_i - y)^2\}^{1/2}$ (capítulo 5), se hicieron utilizando programas de computadora que se presentan en el apéndice. Los eventos observados una vez que han sido ordenados de mayor a menor y los períodos de retorno asignados fueron presentados en la tabla 7.1.

Los resultados obtenidos son:

Para la Distribución Normal:

$m_y = 668.0769$ $S = 114.2130$

$\sigma_y = 379.3653$

Para la Distribución Lognormal de dos parámetros:

$m_L = 6.3699$ $S = 73.1142$

$\sigma_L = 0.5190$

Para la Distribución Lognormal de tres parámetros:

$m_L = 6.1430$ $S = 64.1255$

$\sigma_L = 0.6467$

$a_L = 100.5804$

Para la Distribución Gumbel:

$\alpha = 0.0039$ $S = 98.1536$

$\beta = 509.3597$

Para la Distribución Exponencial de dos parámetros:

$\alpha = 265.5298$ $S = 51.5022$

$\lambda = 0.0023342$

Para la Distribución Gamma de dos parámetros:

$\lambda = 0.006433$ $S = 96.8558$

$K = 4.2976$

Para la Distribución Gamma de tres parámetros:

$\lambda = 0.002788$ $S = 56.2222$

$K = 1.2556$

$\delta = 217.7643$

e) *Error estándar e intervalos de confianza.*

Para los eventos estimados por Mínimos Cuadrados, el error estándar se calcula a partir de (5.38), donde x_0 es la x particular obtenida a partir del período de retorno asignado (De la tabla 5.2 $x = z$ para la Distribución Normal y x = Ln T para la Distribución Exponencial).

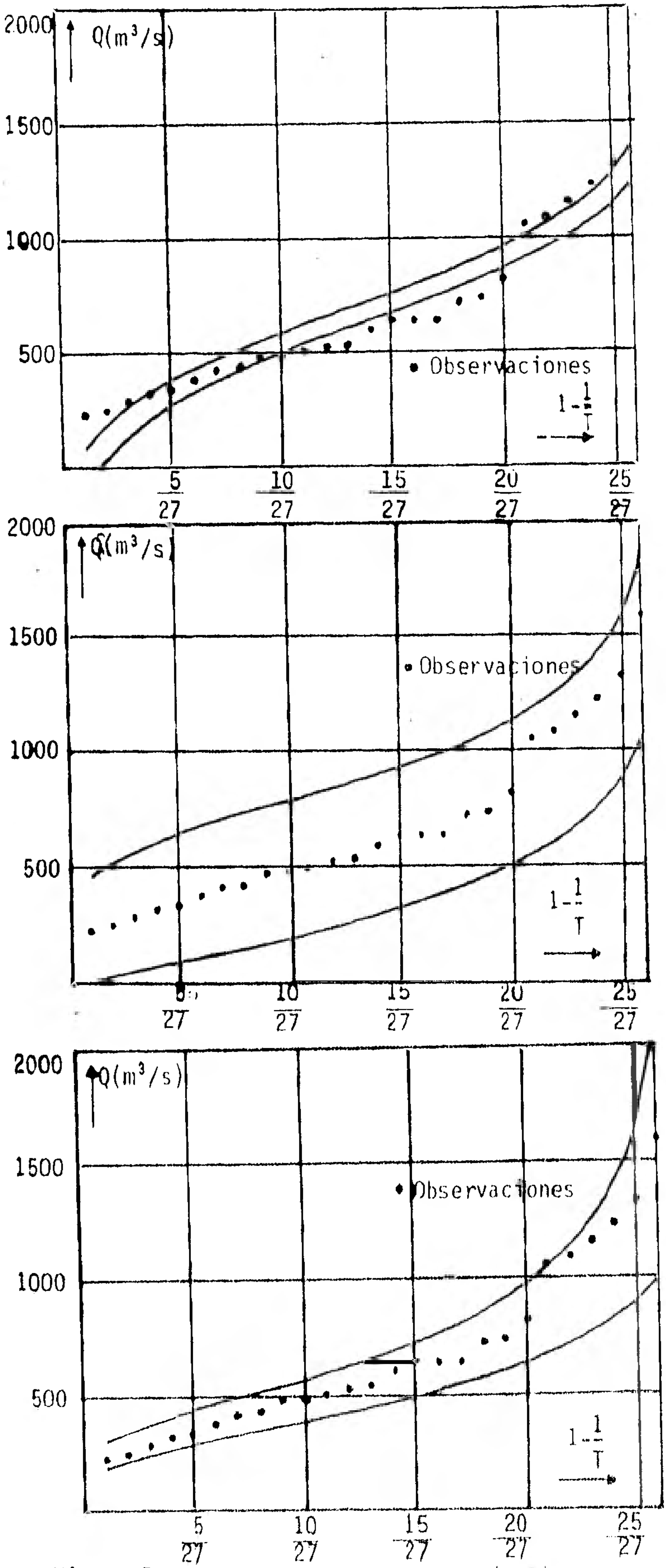
Para los eventos estimados por Máxima Verosimilitud, la obtención del error estándar requiere de aplicar (5.45); el desarrollo de esta expresión para las distribuciones utilizadas se presentó en la sección 5.4.4.

Los límites e intervalos de confianza están dados por (5.40) ó (5.48).

Los cálculos se realizaron también utilizando un programa de computadora.

En las tablas 7.2 a 7.8 se presentan los períodos de retorno asignados con los correspondientes eventos observados, eventos estimados y el error estándar. Se presentan también los eventos estimados para períodos de retorno de 10, 20, 50, 100, 1000 y 10 000 años con su respectivo error estándar e intervalo de confianza (90%).

En las figuras 7.1 y 7.2 se presentan los intervalos de confianza para un nivel del 90% (sección 4.2) para el cual t = 1.71 (de la tabla 2 del apéndice); estos intervalos son para seis distribuciones. Los intervalos para la Distribución Gamma de dos parámetros son muy grandes y no se dibujan.

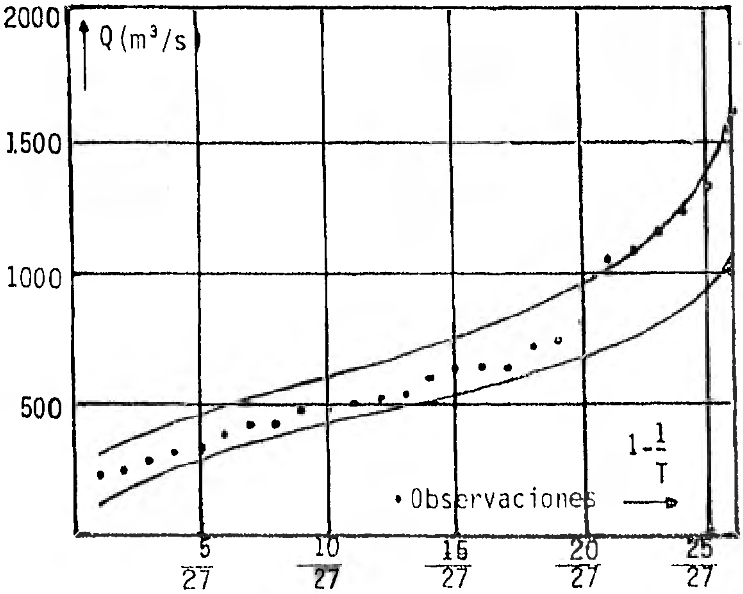


a) Distribución Normal

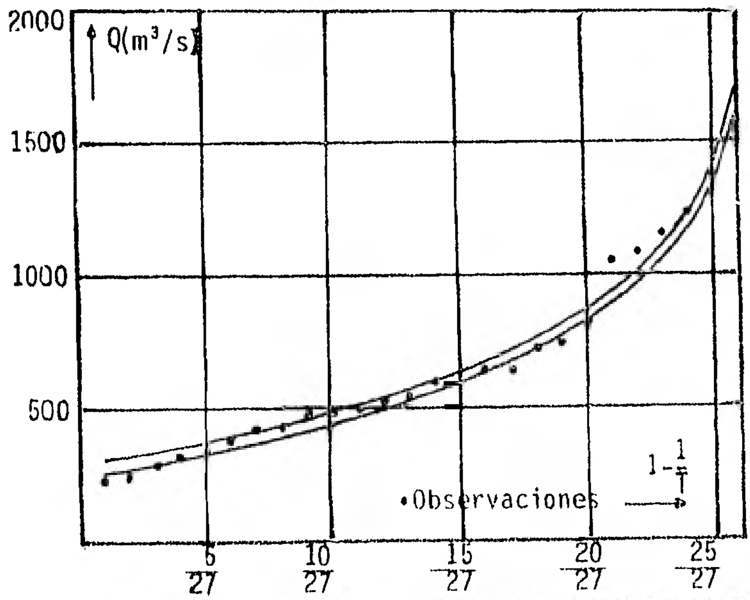
b) Distribución Lognormal de dos parámetros

c) Distribución Lognormal de tres parámetros

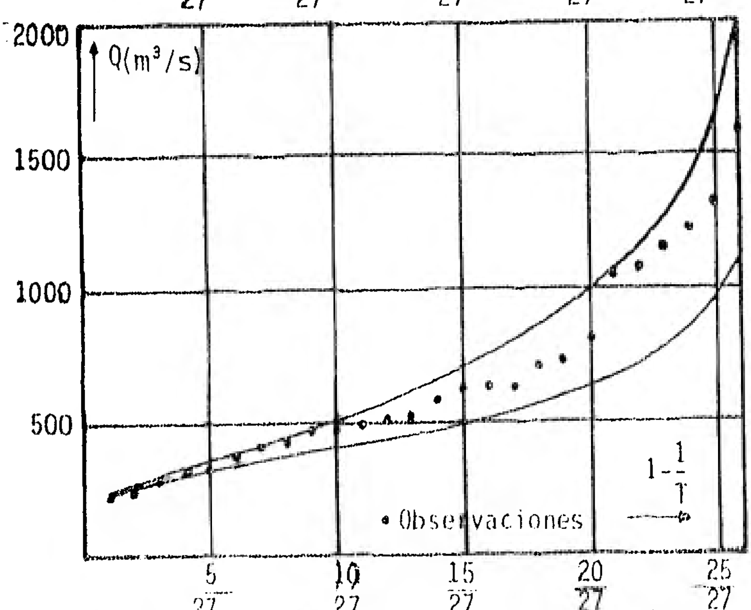
Figura 7.1 Intervalos de confianza (90%)



d) Distribución Gumbel



e) Distribución Exponencial de dos parámetros



f) Distribución Gamma de 3 parámetros

Intervalos de confianza (90%)

TABLA 7.2

GASTOS [illegible] ESTACION [illegible] RIO OMITLAN

DISTRIBUCION NORMAL

PERIODO DE RETORNO	EVENTO OBSERVADO	EVENTO ESTIMADO	ERROR ESTANDAR
[illegible].0000	1591.00	[illegible]	50.18
[illegible].5000	[illegible]5.00	[illegible]	42.70
[illegible]	1,[illegible]1.00	1131.20	[illegible]
[illegible]	[illegible]1.00	[illegible]	34.51
5.[illegible]	[illegible]	[illegible]7.86	31.74
4.[illegible]	[illegible]	[illegible].09	29.51
[illegible]	[illegible].00	912.08	27.65
[illegible]750	[illegible]	[illegible]	26.12
[illegible]000	[illegible]	[illegible]	24.87
[illegible].7000	[illegible]	[illegible]	23.89
[illegible].4545	[illegible]	[illegible]	23.16
[illegible]500	[illegible]	[illegible]	[illegible]
[illegible].0769	594.00	[illegible]	[illegible]
1.[illegible]26	531.00	[illegible]	[illegible]
1.[illegible]00	521.00	[illegible].51	[illegible]
1.[illegible]75	501.[illegible]	[illegible]	[illegible]
1.[illegible]2	477.[illegible]	[illegible]	23.16
[illegible]	475.00	541.7[illegible]	23.89
[illegible]	430.00	[illegible]04.[illegible]	24.[illegible]
1.3500	42[illegible]	[illegible]	[illegible]
1.2857	[illegible].00	[illegible]	27.65
1.2273	[illegible]	[illegible]	29.51
1.1739	317.00	[illegible]	31.76
1.[illegible]250	[illegible]	271.[illegible]	34.51
1.[illegible]	[illegible]	[illegible]04.[illegible]	[illegible].99
1.[illegible]385	[illegible]1.00	117.37	42.7[illegible]

PERIODO DE RETORNO :	10	[illegible]	[illegible]	100	[illegible]00	[illegible]000
EVENTO ESTIMADO :	[illegible]	1,[illegible].21	[illegible]	[illegible].78	[illegible].51	[illegible]
ERROR ESTANDAR :	[illegible]	[illegible]3	[illegible].25	[illegible]	[illegible].84	[illegible]
INTERVALO DE CONFIANZA (90 PORCIENTO)	1[illegible]	[illegible]	[illegible]	[illegible]	[illegible].83	[illegible]43.4[illegible]
	10[illegible]	[illegible].75	[illegible]	[illegible]	170[illegible].1[illegible]	1[illegible]14.[illegible]2

TABLA 7.3

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION LOGNORMAL DE 2 PARAMETROS

PERIODO DE RETORNO	EVENTO OBSERVADO	EVENTO ESTIMADO	ERROR ESTANDAR
27.0000	1591.00	1477.14	244.03
13.5000	1325.00	1232.?3	201.15
9.0000	1231.00	1100.5?	188.65
6.7500	1151.00	1004.26	184.27
5.4000	100?.00	929.65	182.7?
4.5000	105?.00	868.46	182.23
3.8571	?31.00	816.3?	18?.11
3.3750	[illegible]	770.??	18?.07
3.0000	7??.00	730.17	181.??
2.7000	6??.00	693.27	181.??
2.4545	6??.00	652.38	181.36
2.2500	???.00	6?7.85	180.7?
2.0769	[illegible]	59?.?4	180.08
1.9286	[illegible]	[illegible]	179.10
1.8000	[illegible]	54?.??	177.95
1.6875	5??.00	51?.??	176.58
1.5882	4??.00	??1.??	174.97
1.5000	4??.00	467.14	173.11
1.4211	4??.00	442.50	170.96
1.3500	42?.00	417.81	168.46
1.2857	[illegible]	?22.75	165.5?
1.2273	[illegible]	366.90	16?.1?
1.1739	[illegible]	339.64	15?.?4
1.1250	2??.00	309.93	152.71
1.0800	246.00	27?.69	145.69
1.0385	[illegible]	2??.07	134.66

PERIODO DE RETORNO =	10	20	50	100	1000	10000
EVENTO ESTIMADO =	113?.?6	1?71.73	16??.11	195?.8?	2904.83	4024.71
ERROR ESTANDAR	1?1.3?	?21.2?	?06.18	40?.5?	873.99	1583.21
INTERVALO DE CONFIANZA = (90 PORCIENTO)	14??.88	1751.47	2219.?7	[illegible]	4???.70	6?34.73
	80?.?1	9??.99	1172.?4	[illegible]	1408.95	[illegible].68

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION LOGNORMAL DE 3 PARAMETROS

PERIODO DE RETORNO	EVENTO OBSERVADO	EVENTO ESTIMADO	ERROR ESTANDAR
27.0000	1591.00	1578.29	331.36
13.5000	1325.0	1286.49	225.51
9.0000	1231.00	1125.51	175.52
6.7500	[illegible]	1015.04	145.31
5.4000	[illegible]	931.18	124.84
4.5000	1052.00	863.62	110.02
3.8571	816.00	807.03	98.78
3.3750	734.00	758.26	89.95
3.0000	720.00	715.33	82.83
2.7000	639.00	676.88	76.95
2.4545	639.00	641.98	71.99
2.2500	607.00	[illegible]	67.71
2.0769	594.00	[illegible]	63.96
1.9286	[illegible]	552.28	60.61
1.8000	[illegible]	[illegible]	57.55
1.6875	501.00	[illegible]	54.71
1.5882	477.00	[illegible]	52.05
1.5000	475.00	[illegible]	49.50
1.4211	430.00	[illegible]	47.03
1.3500	420.00	[illegible]	44.62
1.2857	379.00	[illegible]	42.25
1.2273	339.00	361.38	39.90
1.1739	312.00	337.47	37.60
1.1250	285.00	311.93	35.41
1.0800	246.00	283.71	33.58
1.0385	[illegible]	247.17	33.01

PERIODO DE RETORNO :	10	20	50	100	1000	10000
EVENTO ESTIMADO :	[illegible]	[illegible]	[illegible]	2196.17	3534.53	5256.78
ERROR ESTANDAR :	[illegible]	[illegible]	[illegible]	604.94	[illegible]	2514.37
INTERVALO DE CONFIANZA (95 POR CIENTO)	[illegible]	1932.29	[illegible]	3231.20	[illegible]	[illegible]
	[illegible]	766.13	1091.46	1161.13	[illegible]	954.74

TABLA 7.5

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION GUMBEL

PERIODO DE RETORNO	EVENTO OBSERVADO	EVENTO ESTIMADO	ERROR ESTANDAR
27.0000	1591.00	1347.86	153.21
13.5000	1325.00	1165.55	127.19
9.0000	1231.00	105[illegible]	112.09
6.7500	1151.00	977.73	10[illegible].44
5.4000	1086.00	915.12	[illegible]
4.5000	1052.00	862.74	[illegible]
3.8571	816.00	817.54	81.[illegible]
3.3750	734.00	774.77	76.[illegible]
3.0000	720.00	740.[illegible]	72.16
2.7000	639.00	705.[illegible]	[illegible]
2.4545	639.00	673.09	65.36
2.2500	637.00	645.33	62.54
2.0769	594.00	616.93	60.05
1.9286	571.00	589.58	57.86
1.8000	521.00	562.99	55.95
1.6875	501.00	536.91	54.00
1.5882	477.00	513.09	52.[illegible]
1.5000	475.00	485.[illegible]	51.[illegible]
1.4211	430.00	459.24	50.96
1.3500	426.00	432.58	50.41
1.2857	379.00	404.92	50.18
1.2273	339.00	375.64	50.33
1.1739	319.00	343.84	50.94
1.1250	285.00	307.93	52.17
1.0800	246.00	264.60	54.35
1.0385	224.00	204.18	58.53

PERIODO DE RETORNO :	10	20	50	100	1000	10000
EVENTO ESTIMADO :	108[illegible]	1262.37	1507.78	1686.44	2276.78	2866.08
ERROR ESTANDAR :	116.00	141.93	176.49	202.[illegible]	291.15	380.23
INTERVALO DE CONFIANZA : (90 PORCIENTO)	[illegible]	1512.19	1809.76	2033.40	2774.93	3516.64
	882.30	1026.55	1205.81	1339.40	1778.63	2215.52

TABLA 7.6

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION EXPONENCIAL

PERIODO DE RETORNO	EVENTO OBSERVADO	EVENTO ESTIMADO	ERROR ESTANDAR
27.0000	1591.00	1677.52	30.68
13.5000	1325.00	1380.56	22.80
9.0000	1231.00	1206.86	18.47
6.7500	1151.00	1083.61	15.63
5.4000	1086.00	988.01	13.65
4.5000	1052.00	909.90	12.25
3.8571	816.00	843.86	11.29
3.3750	734.00	786.65	10.66
3.0000	720.00	736.19	10.29
2.7000	639.00	691.05	10.12
2.4545	639.00	650.22	10.11
2.2500	637.00	612.95	10.22
2.0769	594.00	578.65	10.42
1.9286	531.00	546.90	10.68
1.8000	521.00	517.[illegible]	10.99
1.6875	501.00	489.70	11.32
1.5882	477.00	463.73	11.68
1.5000	475.00	439.24	12.05
1.4211	430.00	416.07	12.42
1.3500	428.00	394.10	12.80
1.2857	372.00	373.[illegible]	13.18
1.2273	339.00	353.[illegible]	13.55
1.1739	319.00	334.[illegible]	13.92
1.1250	285.00	[illegible]	14.[illegible]
1.0800	246.00	298.50	14.65
1.0385	224.00	281.[illegible]	15.00

PERIODO DE RETORNO :	10	25	50	100	1000	10000
EVENTO ESTIMADO :	1251.[illegible]	1588.95	[illegible]	[illegible]	[illegible]	[illegible]
ERROR ESTANDAR :	[illegible]	[illegible].22	[illegible].91	46.18	[illegible]	102.18
INTERVALO DE CONFIANZA :	1285.47	[illegible]	2064.37	2317.[illegible]	[illegible]	[illegible].21
(90 PORCIENTO)	121[illegible]	150[illegible].38	1876.64	[illegible]	309[illegible].20	40[illegible].56

TABLA 7.7

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO ONITLAN

DISTRIBUCION GAMMA DE [illegible] PARAMETROS

PERIODO DE RETORNO	EVENTO OBSERVADO	EVENTO ESTIMADO	ERROR ESTANDAR
27.0000	1591.00	1340.90	2389.54
13.5000	1325.00	1173.92	[illegible]
9.0000	1231.00	1071.20	2302.08
6.7500	1151.00	995.21	2130.15
5.4000	1088.00	933.98	2005.38
4.5000	1052.00	882.12	1892.99
3.8571	916.00	836.20	1794.65
3.3750	731.00	795.96	1706.49
3.0000	720.00	758.75	1626.02
2.7000	639.00	724.[illegible]	1551.47
2.4545	639.00	691.87	1481.58
2.2500	637.00	661.[illegible]	1415.36
2.0769	598.00	631.86	1352.05
1.9286	581.00	[illegible]	1290.98
1.8000	571.00	575.93	1231.51
1.6875	501.00	548.[illegible]	1172.12
1.5882	477.00	521.95	1115.34
1.5000	475.00	495.12	1057.67
1.4211	430.00	468.08	999.57
1.3500	428.00	440.52	940.41
1.2857	379.00	412.06	879.38
1.2273	339.00	382.18	815.37
1.1739	319.00	350.09	746.71
1.1250	285.00	314.10	670.52
1.0800	246.00	272.46	581.01
1.0385	224.00	216.51	462.00

PERIODO DE RETORNO :	10	[illegible]	50	100	1000	10000
EVENTO ESTIMADO :	1098.34	1362.72	1482.87	1637.43	2126.83	2596.03
ERROR ESTANDAR :	2262.01	2734.61	3198.85	3535.98	4605.72	5633.81
INTERVALO DE CONFIANZA : (90 PORCIENTO)	5139.66	5948.55	6956.00	7687.38	10007.08	12235.31
	-2942.99	-3409.11	-3990.27	-4412.53	-5753.41	-7043.24

TABLA 7.8

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION GAMMA DE 3 PARAMETROS

PERIODO DE RETORNO	EVENTO OBSERVADO	EVENTO ESTIMADO	ERROR ESTANDAR
27.0000	1591.00	1570.7[illegible]	257.14
13.5000	1325.00	1305.57	201.80
9.0000	1231.00	1150.37	169.78
6.7500	1151.00	1039.89	147.20
5.4000	1056.00	951.75	129.77
4.5000	1052.00	883.03	115.56
3.8571	816.00	822.[illegible]	10[illegible].57
3.3750	734.00	770.[illegible]	93.21
3.0000	720.00	[illegible]1.59	84.09
2.7000	[illegible]	681.45	75.93
2.4545	[illegible]	642.97	68.57
2.2500	637.00	607.49	61.86
2.0769	574.00	574.49	55.69
1.9286	531.00	541.56	50.00
1.8000	521.00	[illegible]	44.71
1.6875	501.00	486.5[illegible]	39.77
1.5882	477.00	45[illegible].[illegible]3	35.16
1.5000	475.00	429.35	30.84
1.4211	430.00	407.60	26.80
1.3500	428.00	385.52	23.03
1.2857	379.00	361.93	19.54
1.2273	337.00	338.64	16.34
1.1739	317.00	315.43	13.44
1.1250	295.00	292.01	10.82
1.0800	246.00	267.95	[illegible]
1.0385	224.00	[illegible]12.48	3.4[illegible]

PERIODO DE RETORNO :	10	20	50	100	1000	10000
EVENTO ESTIMADO :	1150.74	1455.[illegible]	1867.2[illegible]	2074.99	2981.45	3917.70
ERROR ESTANDAR :	[illegible]	[illegible]	307.01	363.90	566.91	762.88
INTERVALO DE CONFIANZA : (95 PORCIENTO)	1455.[illegible]	1854.[illegible]	2332.55	2687.62	3933.74	5227.[illegible]5
	885.[illegible]	1057.06	1281.97	1459.[illegible]	2025.17	2612.44

f) Bondad de ajuste

El cálculo de las probabilidades estimadas acumuladas se hizo aplicando las correspondientes funciones de distribución acumulada (capítulo 3); la probabilidad observada acumulada está dada por (2.57).

A partir de las probabilidades estimadas y observadas (acumuladas) se aplican las pruebas Kolmogorov-Smirnov (ec.6.5) y suma de las Diferencias al Cuadrado (ec.6.7), la primera de Bondad de Ajuste y la segunda, relacionada con ésta; en las tablas 7.10 a 7.16 se presentan las probabilidades acumuladas (observadas y estimadas) y los correspondientes eventos observados, también se presentan los resultados de las pruebas y la suma de las diferencias al cuadrado de eventos estimados y eventos observados (ec.6.8).

Los cálculos se hacen usando un programa de computadora que se presenta en el apéndice.

Los resultados obtenidos se presentan en la siguiente tabla (tabla 7.9):

TABLA 7.9

DISTRIBUCION	PRUEBA KOLMOGOROV-SMIRNOV (D_2)	SUMA DE LAS DIFERENCIAS AL CUADRADO (D_3)	$\Sigma(\hat{y}_i - y_i)^2$
NORMAL	0,1602	0,1587	313 070.71
LOGNORMAL de 2 parámetros	0,0938	0,0333	128 296.38
LOGNORMAL de 3 parámetros	0,0878	0,0259	98 690.05
GUMBEL	0,1092	0,0625	231 219.15
EXPONENCIAL de 2 parámetros	0,0741	0,0320	63 659.41

TABLA 7.9 (CONTINUACION)

DISTRIBUCION	PRUEBA KOLMOGOROV-SMIRNOV (D_2)	SUMA DE LAS DIFERENCIAS AL CUADRADO (D_3)	$\Sigma(\hat{y}_i - y_i)^2$
GAMMA de 2 parámetros	0.1026	0.0835	225 144.95
GAMMA de 3 parámetros	0.0787	0.0273	75 862.53

La aplicación de la prueba x^2 (secc.6.2) requiere formar intervalos de clase e identificar los límites reales de clase.

La secuencia en el cálculo fue ilustrada en el ejemplo 6.1

Los intervalos de clase no son los mismos para las distribuciones de tres parámetros ya que debe satisfacerse la condición (para comparar resultados) de tener el mismo número de grados de libertad para todos los modelos.

En las tablas 7.17 a 7.23 se presenta el desarrollo seguido en el cálculo del estadístico D_1 dado por (6.1)

Los resultados obtenidos son para las distribuciones utilizadas se presentan a continuación:

DISTRIBUCION	D_1
Normal	8.0006
Lognormal de 2 parámetros	3.6215
Lognormal de 3 parámetros	4.2566
Gumbel	4.8330
Exponencial	2.9380
Gamma de 2 parámetros	4.7184
Gamma de 3 parámetros	4.1385

T ABLA 7.10

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION NORMAL

EVENTOS OBSERVADOS	PROB. OBSERVADA	PROB. ESTIMADA
1591.00	0.9630	[illegible]
[illegible]	0.925[illegible]	[illegible]
11[illegible]0.00	0.888[illegible]	[illegible]
[illegible]151.[illegible]0	0.851[illegible]	[illegible]
108[illegible].00	0.8148	[illegible]
1055.00	[illegible]	[illegible]
[illegible]	0.[illegible]	0.6842
[illegible]	0.7[illegible]	0.6317
[illegible]	0.[illegible]	0.[illegible]
[illegible]	0.6[illegible]	0.[illegible]
[illegible]	0.5[illegible]	[illegible]
[illegible]	0.5[illegible]	0.[illegible]
[illegible]	0.51[illegible]	0.4[illegible]7[illegible]
[illegible].00	0.[illegible]	0.4[illegible]
[illegible].00	0.4[illegible]	0.[illegible]
[illegible]	0.4[illegible]	0.[illegible]421
477.00	0.[illegible]	0.[illegible]3
4[illegible].00	0.33[illegible]	0.[illegible]
[illegible]	0.29[illegible]	0.30[illegible]
[illegible]	0.[illegible]	0.[illegible]
[illegible]	0.[illegible]	0.[illegible]
[illegible]	0.1852	0.1[illegible]
[illegible]	0.1481	0.1[illegible]
28[illegible]	0.111[illegible]	[illegible]
[illegible]	0.0741	0.1[illegible]
[illegible]	0.0[illegible]	0.1[illegible]

NUMERO DE DATOS = 26

PRUEBA KOLMOGOROV-SMIRNOV, DIFERENCIA MAYOR = 0.1[illegible]

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE EVENTOS ESTIMADOS Y EVENTOS OBSERVADOS = .313070706E+06

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE PROBABILIDADES ESTIMADAS Y OBSERVADAS = 0.1587

TABLA 7.11

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION LOGNORMAL DE 2 PARAMETROS

EVENTOS OBSERVADOS	PROB. OBSERVADA	PROB. ESTIMADA
1591.00	0.9630	0.9732
1375.00	0.9259	0.9428
1241.00	0.8889	0.9246
1151.00	0.8519	0.9044
1096.00	0.8148	0.8840
1052.00	0.7778	0.8716
[illegible]	0.7407	0.7404
[illegible]	0.7037	0.6702
720.00	0.6667	0.6566
639.00	0.6296	0.5688
639.00	0.5926	0.5688
637.00	0.5556	0.5664
594.00	0.5185	0.5130
531.00	0.4815	0.4272
[illegible]	0.4444	0.4127
501.00	0.4074	0.3838
477.00	0.3704	0.3483
475.00	0.3333	0.3453
430.00	0.2963	0.2776
428.00	0.2593	0.2746
379.00	0.2222	0.2024
339.00	0.1852	0.1473
319.00	0.1481	0.1220
285.00	0.1111	[illegible]
246.00	0.0741	0.0473
[illegible]	0.0370	0.0324

NUMERO DE DATOS = 26

PRUEBA KOLMOGOROV-SMIRNOV, DIFERENCIA MAYOR = 0.09[illegible]

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE EVENTOS ESTIMADOS Y EVENTOS OBSERVADOS = [illegible]E+06

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE PROBABILIDADES ESTIMADAS Y OBSERVADAS = 0.03[illegible]

<u>TABLA 7.12</u>

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION LOGNORMAL DE 3 PARAMETROS

EVENTOS OBSERVADOS	PROB. OBSERVADA	PROB. ESTIMADA
1591.00	0.9630	0.9641
1425.00	0.9259	0.9326
1231.00	0.8889	0.9150
1151.00	0.8519	0.8959
1086.00	0.8148	0.8770
1052.00	0.7778	0.8656
816.00	0.7407	0.7469
734.00	0.7037	0.6832
720.00	0.6667	0.6708
639.00	0.6296	0.5891
639.00	0.5926	0.5891
637.00	0.5556	0.5869
574.00	0.5185	0.5360
531.00	0.4815	0.4519
521.00	0.4444	0.4375
501.00	0.4074	0.4080
477.00	0.3704	0.3714
475.00	0.3333	0.3683
430.00	0.2963	0.2965
428.00	0.2593	0.2933
379.00	0.2222	0.2134
339.00	0.1852	0.1505
319.00	0.1481	0.1210
285.00	0.1111	0.0761
246.00	0.0741	0.0360
224.00	0.0370	0.0201

NUMERO DE DATOS = 26

PRUEBA KOLMOGOROV-SMIRNOV, DIFERENCIA MAYOR = 0.0870

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE EVENTOS ESTIMADOS Y EVENTOS OBSERVADOS = .986900544E+05

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE PROBABILIDADES ESTIMADAS Y OBSERVADAS = 0.0259

TABLA 7.13

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION GUMBEL

EVENTOS OBSERVADOS	PROB. OBSERVADA	PROB. ESTIMADA
1591.00	0.9630	0.9855
1325.00	0.9259	0.9596
1231.00	0.8889	0.9421
1151.00	0.8519	0.9218
1066.00	0.8148	0.9003
1052.00	0.7778	0.8870
816.00	0.7407	0.7396
734.00	0.7037	0.6599
720.00	0.6667	0.6447
639.00	0.6296	0.5474
639.00	0.5926	0.5474
637.00	0.5556	0.5449
594.00	0.5185	0.4876
531.00	0.4815	0.3990
521.00	0.4444	0.3846
501.00	0.4074	0.3559
477.00	0.3704	0.3215
475.00	0.3333	0.3186
430.00	0.2963	0.2557
428.00	0.2593	0.2530
379.00	0.2222	0.1893
339.00	0.1852	0.1428
319.00	0.1481	0.1219
285.00	0.1111	0.0904
246.00	0.0741	0.0609
224.00	0.0370	0.0473

NUMERO DE DATOS = 26

PRUEBA KOLMOGOROV-SMIRNOV, DIFERENCIA MAYOR = 0.1092

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE EVENTOS ESTIMADOS Y EVENTOS OBSERVADOS = .23121915E+06

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE PROBABILIDADES ESTIMADAS Y OBSERVADAS = 0.0625

TABLA 7.14

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION EXPONENCIAL

EVENTOS OBSERVADOS	PROB. OBSERVADA	PROB. ESTIMADA
1591.00	[illegible]	0.9547
1325.00	[illegible]	0.9157
1231.00	0.8889	0.8950
1151.00	0.8519	0.8734
1084.00	0.8148	0.8527
[illegible]	0.7778	[illegible]
[illegible]	0.7407	0.7233
734.00	0.7037	0.6650
720.00	0.6667	[illegible]
639.00	0.6296	[illegible]
639.00	0.5926	[illegible]
637.00	0.5556	[illegible]
594.00	0.5185	0.5355
531.00	0.4815	0.4619
521.00	0.4444	0.4492
501.00	0.4074	0.4228
477.00	0.3704	0.3896
475.00	[illegible]	0.3867
430.00	[illegible]	0.3186
428.00	[illegible]	0.3156
379.00	[illegible]	[illegible]
337.00	[illegible]	[illegible]
319.00	0.1481	0.1172
[illegible]	0.1111	0.0484
[illegible]	0.0741	0.0000
[illegible]	[illegible]	0.0000

NUMERO DE DATOS = 26

PRUEBA KOLMOGOROV-SMIRNOV: DIFERENCIA MAYOR = 0.0741

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE EVENTOS ESTIMADOS Y EVENTOS OBSERVADOS = .6365941 06E+05

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE PROBABILIDADES ESTIMADAS Y OBSERVADAS = 0.0320

TABLA 7.15

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION GAMMA DE 2 PARAMETROS

EVENTOS OBSERVADOS	PROB. OBSERVADA	PROB. ESTIMADA
1591.00	0.9630	0.9877
1325.00	0.9259	0.9604
1231.00	0.8889	0.9413
1151.00	0.8519	0.9188
1086.00	0.8148	0.8951
1052.00	0.7778	0.8804
816.00	0.7407	0.7223
724.00	0.7037	0.6402
720.00	0.6667	0.6247
639.00	0.6296	0.5276
639.00	0.5926	0.5276
637.00	0.5556	0.5250
594.00	0.5185	0.4689
531.00	0.4815	0.3830
521.00	0.4444	0.3692
501.00	0.4074	0.3416
477.00	0.3704	0.3086
475.00	0.3333	0.3059
430.00	0.2963	0.2455
420.00	0.2593	0.2429
379.00	0.2222	0.1814
339.00	0.1852	0.1361
319.00	0.1481	0.1156
285.00	0.1111	0.0843
246.00	0.0741	0.0547
224.00	0.0370	0.0411

NUMERO DE DATOS = 26

PRUEBA KOLMOGOROV-SMIRNOV, DIFERENCIA MAYOR = 0.1026

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE EVENTOS ESTIMADOS Y EVENTOS OBSERVADOS = .225144949E+06

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE PROBABILIDADES ESTIMADAS Y OBSERVADAS = 0.0835

TABLA 7.16

GASTOS MAXIMOS ANUALES ESTACION EL PUENTE RIO OMITLAN

DISTRIBUCION GAMMA DE 3 PARAMETROS

EVENTOS OBSERVADOS	PROB. OBSERVADA	PROB. ESTIMADA
[illegible]	0.9630	0.9649
1325.00	0.9259	0.9276
1231.00	0.8889	0.9100
1151.00	0.8519	0.8891
1086.00	0.8148	0.8686
1052.00	0.7778	0.8565
816.00	0.7407	0.7361
734.00	0.7037	0.6751
720.00	0.6667	0.6635
639.00	0.6296	0.5884
639.00	0.5926	0.5884
637.00	0.5556	0.5864
594.00	0.5185	0.5406
531.00	0.4815	0.4659
521.00	0.4444	0.4532
501.00	0.4074	0.4270
477.00	0.3704	0.3945
475.00	[illegible]	0.3917
430.00	0.2963	0.3271
[illegible]	0.2593	0.3241
[illegible]	0.2222	0.2492
[illegible]	0.1852	0.1858
[illegible]	0.1481	0.1538
[illegible]	0.1111	0.1001
[illegible]	0.0741	0.0419
[illegible]	0.0370	0.0170

NUMERO DE DATOS = 26

PRUEBA KOLMOGOROV-SMIRNOV, DIFERENCIA MAXIMA = [illegible]

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE EVENTOS ESTIMADOS Y EVENTOS OBSERVADOS = .75842528?E+05

SUMA DE LOS CUADRADOS DE LAS DIFERENCIAS DE PROBABILIDADES ESTIMADAS Y OBSERVADAS = [illegible]

TABLA 7.17 Aplicación de la Prueba x^2 para la Distribución Normal

Intervalo	frecuencia observada O_i	límite real de clase	z	probabilidad acumulada	probabilidad en el intervalo	frecuencia esperada E_i
		223.5	1.172	0.1206		
224 - 451	8				0.1634	4.2487
		451.5	0.571	0.2840		
452 - 679	9				0.2280	5.9273
		679.5	0.030	0.5120		
680 - 907	3				0.2240	5.8241
		907.5	0.631	0.7360		
908 - 1135	2				0.1550	4.0308
		1135.5	1.232	0.8910		
1 136 - 1363	3				0.0756	1.9645
		1363.5	1.833	0.9666		
1 364 - 1591	1				0.0259	0.6743
		1591.5	2.434	0.9925		
Σ	26				0.8719	

Los límites reales de clase (secc. 2.1.2) permiten identificar si un dato pertenece a uno u otro intervalo.

El cálculo de z se hace a partir de (3.2)

Los parámetros estimados por el método de mínimos cuadrados son:

$$\hat{m}_X = 668.0769$$

$$\hat{\sigma}_X = 379.3653$$

El cálculo de la probabilidad acumulada se hace usando la tabla 1 del apéndice como se explicó en la sección 3.1.1.

De (6.1)

$$D_1 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

$$D_1 = \frac{(8-4.2487)^2}{4.2487} + \frac{(9-5.9273)^2}{5.9273} + \frac{(3-5.8241)^2}{5.8241} + \frac{(2-4.0308)^2}{4.0308}$$

$$+ \frac{(3-1.9645)^2}{1.9645} + \frac{(1-0.6743)^2}{0.6743} .$$

$$D_1 = 8.0006$$

TABLA 7.18 Aplicación de la Prueba x^2 para la Distribución Lognormal de dos parámetros

Intervalo	frecuencia observada O_i	límite real de clase	z	probabilidad acumulada	probabilidad en el intervalo	frecuencia esperada E_i
		223.5	-1.851	0.0321		
224 - 451	8				0.2779	7.2255
		451.5	0.496	0.3100		
452 - 679	9				0.3048	7.9242
		679.5	0.292	0.6148		
680 - 907	3				0.1874	4.8713
		907.5	0.849	0.8020		
908 -1 135	2				0.0978	2.5425
		1135.5	1.281	0.8999		
1 136 -1 363	3				0.0489	1.2716
		1363.5	1.654	0.9485		
1 964 -1 591	1				0.0245	0.6357
		1591.5	1.932	0.9735		

El cálculo de z se hace por medio de (3.15):

Los parámetros estimados por el método de máxima verosimilitud son:

$$\hat{m}_L = 6.3699$$

$$\hat{\sigma}_L = 0.5190$$

El cálculo de la probabilidad acumulada se hace con la ayuda de la tabla 1 como se explicó en la sección 3.2.1.

Aplicando (6.1):

$$D_1 = \frac{(8-7.2255)^2}{7.2255} + \frac{(9-7.9242)^2}{7.9242} + \frac{(3-4.8713)^2}{4.8713} + \frac{(2-2.5425)^2}{2.5425}$$

$$+ \frac{(3-1.2716)^2}{1.2716} + \frac{(1-0.6357)^2}{0.6357}$$

$$D_1 = 3.6215$$

TABLA 7.19 Aplicación de la Prueba x^2 para la Distribución Lognormal de tres parámetros.

Intervalo	frecuencia observada O_i	límite real de clase	z	probabilidad acumulada	probabilidad en el intervalo	frecuencia esperada E_i
		223.5	2.059	0.0198		
224.0 - 418.4	6				0.2587	6.7256
		418.9	0.588	0.2784		
419.4 - 613.9	8				0.2823	7.3405
		614.4	0.153	0.5608		
614.9 - 809.3	5				0.1818	4.7268
		809.8	0.651	0.7426		
810.3 -1004.7	1				0.1054	2.7396
		1005.2	1.028	0.8479		
1005.7-1200.1	3				0.0603	1.5680
		1200.6	1.330	0.9082		
1201.1-1395.6	2				0.0350	0.9110
		1396.1	1.583	0.9433		
1396.6-1591.0	1				0.0208	0.5412
		1591.5	1.800	0.9641		

El cálculo de z se hace por medio de (3.15) al sustituir y por $y-a_L$

$$z = \frac{\mathrm{Ln}(y-a_L) - m_L}{\sigma_L}$$

Los parámetros estimados por el método de máxima verosimilitud son:

$$\hat{m}_L = 6.1430$$

$$\hat{\sigma}_L = 0.6467$$

$$\hat{a}_L = 100.5804$$

La probabilidad acumulada se obtiene siguiendo el mismo procedimiento que para la Distribución Lognormal de dos parámetros.

Aplicando (6.1)

$$D_1 = \frac{(6-6.7256)^2}{6.7256} + \frac{(8-7.3405)^2}{7.3405} + \frac{(5-4.7268)^2}{4.7268} + \frac{(1-2.7396)^2}{2.7396}$$

$$+ \frac{(3-1.5680)^2}{1.5680} + \frac{(2-0.9110)^2}{0.9110} + \frac{(1-0.5412)^2}{0.5412}$$

$$D_1 = 4.2566$$

TABLA 7.20. Aplicación de la Prueba χ^2 para la Distribución Gumbel.

Intervalo	frecuencia observada Oi	límite real de clase	probabilidad acumulada	probabilidad en el intervalo	frecuencia esperada Ei
		223.5	0.0474		
224 - 451	8			0.2382	6.1933
		451.5	0.2856		
452 - 679	9			0.3119	8.1089
		679.5	0.5975		
680 - 907	3			0.2117	5.5054
		907.5	0.8092		
908- 1135	2			0.1074	2.7937
		1135.5	0.9167		
1136-1363	3			0.0482	1.2530
		1363.5	0.9649		
1364-1591	1			0.0205	0.5339
		1591.5	0.9854		

La probabilidad acumulada se calcula a partir de la ec. (3.36):

Los parámetros estimados por el método de máxima verosimilitud son:

$$\hat{\alpha} = 0.0039$$

$$\hat{\beta} = 509.3597$$

Aplicando (6.1):

$$D_1 = \frac{(8-6.1933)^2}{6.1933} + \frac{(9-8.1089)^2}{8.1089} + \frac{(3-5.5054)^2}{5.5054} + \frac{(2-2.7937)^2}{2.7937}$$

$$+ \frac{(3-1.2530)^2}{1.2530} + \frac{(1-0.5339)^2}{0.5339}$$

$$D_1 = 4.8330$$

TABLA 7.21. Aplicación de la Prueba x^2 para la Distribución Exponencial.

ntervalo	frecuencia observada Oi	límite real de clase	probabilidad acumulada	probabilidad en el intervalo	frecuencia esperada Ei
		223.5	0		
224-451	8			0.3521	9.1558
		451.5	0.3521		
452-679	9			0.2674	6.9514
		679.5	0.6195		
680-907	3			0.1570	4.0826
		907.5	0.7765		
908-1135	2			0.0922	2.3978
		1135.5	0.8688		
1136-1363	3			0.0542	1.4083
		1363.5	0.9229		
1364-1591	1			0.0318	0.8271
		1591.5	0.9547		

La probabilidad acumulada se calcula a partir de la expresión (3.51):

Los parámetros estimados por el método de mínimos cuadrados son los siguientes:

$$\alpha = 265.5298$$

$$\lambda = 0.00233418$$

Aplicando (6.1)

$$D_1 = \frac{(8-9.1558)^2}{9.1558} + \frac{(9-6.9514)^2}{6.9514} + \frac{(3-4.0826)^2}{4.0826} + \frac{(2-2.3978)^2}{2.3978}$$

$$+ \frac{(3-1.4083)^2}{1.4083} + \frac{(1-0.8271)^2}{0.8271}$$

$$D_1 = 2.9380$$

TABLA 7.22. Aplicación de la Prueba x^2 para la Distribución Gamma de dos parámetros.

Intervalo	frecuencia observada Oi	límite real de clase	y	z	probabilidad acumulada	probabilidad en el intervalo	frecuencia esperada Ei
		223.5	2.8755	1.741	0.0385		
224-451	8					0.2298	5.9757
		451.5	5.8088	0.601	0.2683		
452-679	9					0.3043	7.9110
		679.5	8,7422	0.196	0.5726		
680-907	3					0.2209	5.7434
		907.5	11.6755	0.829	0.7935		
908-1135	2					0.1187	3.0868
		1135.5	14.6089	1.364	0.9122		
1136-1363	3					0.0536	1.3936
		1363.5	17.5423	1.830	0.9658		
1364-1591	1					0.0217	0.5635
		1591.5	20.4756	2.248	0.9875		

Las y, se obtienen a partir de la expresión (3.63)

$$y = \{ (x - \delta) - k \} \lambda$$

Los parámetros, estimados por el método de máxima verosimilitud son:

$k = 4.2976$

$\lambda = 0.006433$

$\delta = 0.0000$ (ya que la distribución es de dos parámetros)

z se obtiene al aplicar (3.66):

$$z \simeq \left\{ \sqrt[3]{\frac{X^2}{\nu}} + \frac{2}{9\nu} - 1 \right\} \sqrt{\frac{9\nu}{2}} \; ; \; x^2 = 2y \quad \nu = 2\,k.$$

Con los valores de z así obtenidos se busca en la tabla 1 del apéndice la probabilidad acumulada correspondiente.

Aplicando (6.1):

$$D_1 = \frac{(8-5.9757)^2}{5.9757} + \frac{(9-7.9110)^2}{7.9110} + \frac{(3-5.7434)^2}{5.7434} + \frac{(2-3.0868)^2}{3.0868}$$

$$+ \frac{(3-1.3936)^2}{1.3936} + \frac{(1-0.5635)^2}{0.5635}$$

$D_1 = 4.7184$

TABLA 7.23 Aplicación de la Prueba x^2 para la Distribución Gamma de tres parámetros.

Intervalo	frecuencia observada O_i	límite real de clase x	2y	z	probabilidad acumulada	probabilidad en el intervalo	frecuencia esperada Ei
		223.5	0.0320	2.279	0.0100		
224.0-418.4	6					0.2989	7.727
		418.9	1.1216	0.424	0.3086		
419.4-613.9	8					0.2529	6.5758
		614.4	2.2118	0.158	0.5615		
614.9-809.3	5					0.1695	4.4071
		809.8	3.3015	0.618	0.7310		
810.3-1004.7	1					0.1064	2.7651
		1005.2	4.3911	0.986	0.8373		
1005.7-1200.1	3					0.0649	1.6875
		1200.6	5.4807	1.296	0.9022		
1201.1-1395.6	2					0.0391	1.0175
		1396.1	6.5709	1.568	0.9414		
1396.6-1591.0	1					0.0235	0.6098
		1591.5	7.6606	1.811	0.9648		

y se calcula a partir de (3.63):

$y = 2\lambda\ (x - \delta)$

Los parámetros estimados por el método de máxima verosimilitud son:

$k = 1.2556$

$\lambda = 0.002788$

$\delta = 217.7643$

z se obtiene al aplicar (3.66) en la misma forma que para la distribución de dos parámetros y la probabilidad acumulada también en forma similar.

Aplicando (6.1):

$$D_1 = \frac{(6-7.7727)^2}{7.7727} + \frac{(8-6.5758)^2}{6.5758} + \frac{(5-4.4071)^2}{4.4071} + \frac{(1-2.7651)^2}{2.7651}$$

$$+ \frac{(3-1.6875)^2}{1.6875} + \frac{(2-1.0175)^2}{1.0175} + \frac{(1-0.6098)^2}{0.6098}$$

$$D_1 = 4.1385$$

Muestras de verosimilitud

Los modelos θ_i a los que se aplica esta prueba son las distribuciones:

θ_1 = normal

θ_2 = lognormal de dos parámetros

θ_3 = gumbel

θ_4 = exponencial

θ_5 = gamma de dos parámetros

La probabilidad anterior que se asigna a cada modelo es la misma, considerando que se tiene una ignorancia total del comportamiento del fenómeno (gasto máximos anuales):

$$P'(\theta_1) = P'(\theta_2) = \ldots P'(\theta_5) = 1/5$$

Los posibles resultados z_k son representados por los intervalos de clase. (Los mismos que para la Prueba χ^2).

De (6.17) y (6.19):

$$P''(\theta_i) = P(\theta_i|z_k) = \frac{P(z_k|\theta_i)\ P'(\theta_i)}{\sum_{i=1}^{n} P(z_k|\theta_i)\ P'(\theta_j)}$$

Para este caso $z_k = z_1, z_2, \ldots, z_6$

Cálculo de $P(z_k|\theta_i)$:

$P(z_k|\theta_i)$ representa el producto de las probabilidades de los intervalos de interés ya que se trata de eventos independientes, por lo tanto:

$$P\{(z_1, z_2, \ldots, z_6)|\theta_1\} = P(z_1|\theta_1)\ P(z_2|\theta_1) \ldots P(z_6|\theta_1)$$

$$=(0.1634)\ (0.2280)\ (0.2240)\ (0.1550)\ (0.0756)\ (0.0259) = 2.53535353 \times 10^{-6}$$

Similarmente para los demás modelos:

$$P\{(z_1(z_2 \ldots, z_6)|\theta_2\} = 1.855698 \times 10^{-6}$$

$$P\{z_1, z_2, \ldots, z_6)|\theta_3\} = 1.672801 \times 10^{-6}$$

$$P\{z_1, z_2, \ldots, z_6)|\theta_4\} = 2.349182 \times 10^{-6}$$

$$P\{z_1, z_2, \ldots, z_6)|\theta_5\} = 2.130704 \times 10^{-6}$$

Por lo tanto:

$$\sum_{i=1}^{n} P(z_k|\theta_i) = 10.54373924 \times 10^{-6}$$

Como la probabilidad anterior $P'(\theta_i)$ es la misma para cada modelo, es suficiente, para la obtención de la probabilidad posterior (ec. 6.21) sólo multiplicar $P(z_k|\theta_i)$ de cada modelo por el factor de normaliza--

ción N dado por (6.20):

$$N = \frac{1}{10.54373924 \times 10^{-6}}$$

Aplicando (6.21) se llega a los siguientes resultados:

$P''(\theta_1) = 0.2404$

$P''(\theta_2) = 0.1760$

$P''(\theta_3) = 0.1587$

$P''(\theta_4) = 0.2228$

$P''(\theta_5) = 0.2021$

g) *Análisis de resultados*

Se observa, de las figuras 7.1 a 7.7 una gran diferencia en los inter-valos. Las distribuciones cuyos parámetros se estiman por el método de Máxima Verosimilitud (M.V.) presentan, para estos datos, intervalos más grandes que aquellas en que sus parámetros se estimaron por Mínimos Cuadrados.

Para las distribuciones ajustadas por M.V. se observa también que los intervalos se hacen mayores para los valores más grandes (mayores gastos máximos observados).

La Distribución Gamma de dos parámetros presenta intervalos de confianza (para un nivel del 90%) mucho más grandes que las otras distribuciones para estos datos (no se dibujaron).

En cuanto a la desviación estándar de los errores (y consiguiente suma de las diferencias al cuadrado de eventos estimados y eventos observados correspondientes), el menor valor, comparativamente, corresponde a la Distribución Exponencial y el mayor a la Distribución Normal; la Distribución Gamma de dos parámetros tiene el tercer valor más grande por lo que se observa que este valor y los intervalos de confianza no coinciden en cuanto al orden de "clasificación de los modelos" para estos datos (según la desviación estándar de los errores y el tamaño de los intervalos de confianza).

Respecto a la Prueba Kolmogorov-Smirnov, se tiene que para un nivel de significancia del 10% (secc. 6.1) y $n = 26$, se obtiene de la tabla 4 del apéndice un valor (secc. 6.3):

$$d_2 = 0.24$$

Por este nivel de significancia, se cumple la condición de que D_2 (calculado por la ec. 6.5) sea menor que d_2 para todas las distribuciones propuestas como modelo que siguen los datos lo cual quiere decir que, según este criterio, cualquiera de estas distribuciones puede adoptarse como la de la población a la cual pertenecen los gastos máximos anuales observados. Sin embargo, si se trata de seleccionar una sola distribución, resulta razonable escoger aquella que presenta el menor valor de D_2 de todas las propuestas, en este caso la Distribución Exponencial.

De los resultados del cálculo de D_3 (Suma de las diferencias al cuadrado de las probabilidades observada y estimada acumuladas), se observa al compararlos entre sí (modelos), que en general concuerdan con los de la prueba Kolmogorov-Smirnov, por ejemplo, la Distribución Normal tiene el máximo valor tanto de D_2 como de D_3. También debe observarse que, para estos datos, las distribuciones de tres parámetros (Lognormal y Gamma, en ese orden) tuvieron los valores menores correspondientes de D_3.

En forma semejante, para la Prueba χ^2, para un nivel de significancia igual (10%) y para n - k = 3 grados de libertad (sección 4.3.1), de la tabla 3 del apéndice se obtiene el valor crítico:

$$\chi^2_{0.90,3} = 6.25$$

Para este nivel (10%), sólo la Distribución Normal es rechazada y de las restantes distribuciones, la Distribución Exponencial es la que presenta el menor valor de su correspondiente D_1; las otras distribuciones presentan valores del mismo orden de magnitud de D_1.

La prueba χ^2 es también un criterio para seleccionar un modelo de varios propuestos, por lo que en este caso se escogería la Distribución Exponencial como la distribución a la que pertenecen los gastos máximos anuales, sin embargo, debe tomarse en cuenta, que los resultados (valores de D_i) varían si se escogen otros intervalos de clase.

Por otro lado, en forma general, los resultados de esta prueba coinciden (comparativamente) con los resultados de la Prueba Kolmogorov-Smirnov; por ejemplo, la distribución con mayor valor de D_i es la que presenta también el mayor valor de D_i (Distribucion Normal)

De la aplicación del criterio Muestral de Verosimilitud resulta que el mejor modelo para los gastos máximos anuales observados, es θ_i normal, el cual presenta la mayor probabilidad posterior $P''(\theta)$. Después de θ_i,

θ_4 = exponencial es la distribución con mayor probabilidad posterior.

El hecho de que la Distribución Normal sea la mejor elección siguiendo este criterio está en desacuerdo con los resultados de las otras pruebas aplicadas, en las cuales es el "peor" modelo (de los propuestos) para estos datos. Sin embargo, el segundo mejor modelo es la Distribución Exponencial, que las otras pruebas sugieren como representativo de los datos.

Debe señalarse también, que la diferencia en la probabilidad posterior entre los modelos θ_1 y θ_4 es pequeña (menos de 0.02)

En la tabla 7.24 se presenta un resumen de las pruebas aplicadas a los gastos máximos anuales registrados.

De acuerdo a los resultados y a las observaciones de ellos hechas, el modelo seleccionado como aquel que siguen los datos y a partir del cual pueden hacerse inferencias es la Distribución Exponencial.

Enseguida se presentan las expresiones utilizadas, los parámetros estimados y los eventos estimados correspondientes a diferentes períodos de retorno con sus correspondientes intervalos de confianza para la distribución escogida:

Modelo seleccionado: Distribución Exponencial (de dos parámetros).

Función de distribución acumulada (ec. 3.5i):

$$F_Y(y) = 1 - \exp\{-[illegible](y - [illegible])\}$$

Parámetros estimados (por mínimos cuadrados):

[illegible] = 265.5298

$\lambda = 0.0023342$

Expresión de los eventos estimados y (a partir de 3.51 y 2.57)

$$y = \alpha + \frac{1}{\lambda} \text{ Ln } T$$

Eventos estimados para diferentes periodos de retorno e intervalos de confianza.

T(años)	y(m^3/seg)	$y \pm t\, S_T$	
50	1 941.50,	1 876.67,	2 006.33
100	2 238.46,	2 159.49,	2 317.43
1 000	3 224.92,	3 098.28,	3 351.56
10 000	4 211.39,	4 036.66,	4 386.12

TABLA 7.24

DISTRIBUCION	METODO DE AJUSTE	PARAMETROS ESTIMADOS	DESVIACION ESTANDAR DE LOS ERRORES	PRUEBA K - S	SUMA DE LAS DIFERENCIAS AL CUADRADO	PRUEBA X^2	MUESTRAS DE VEROSI-MILITUD.
NORMAL	mínimos cuadrados	m_y = 668.0769 σ_y = 379.3653	114.2130	0.1602	0.1587	8.0006	0.2404
LOGNORMAL de 2 parámetros	máxima verosimilitud	m_L = 6.3699 σ_L = 0.5190	73.1142	0.0938	0.0333	3.6215	0.1760
LOGNORMAL de 3 parámetros	máxima verosimilitud	m_L = 6.1430 σ_L = 0.6467 a_L = 100.5804	64.1255	0.0878	0.0259	4.2566	- - -
GUMBEL	máxima verosimilitud	α = 0.0039 β = 509.3597	98.1536	0.1092	0.0625	4.8330	0.1587
EXPONENCIAL de 2 parámetros	mínimos cuadrados	α = 265.5298 λ = 0.0023342	51.5022	0.0741	0.0320	2.9380	0.2228
GAMMA de 2 parámetros	máxima verosimilitud	λ = 0.006433 k = 4.2976	96.8558	0.1026	0.0835	4.7184	0.2021
GAMMA de 3 parámetros	máxima verosimilitud	λ = 0.002788 k = 1.2556 δ = 217.7643	56.2222	0.0787	0.0273	4.1385	----

8. CONCLUSIONES Y RECOMENDACIONES

En el desarrollo de este trabajo, tanto en la parte teórica como en las aplicaciones, se destaca la necesidad de contar con registros históricos largos, lo que conduce a reducir las incertidumbres estadística y de modelo, ésto, la mayoría de las veces no es posible para eventos hidrológicos, por lo que se debe tratar de aprovechar al máximo la información disponible.

El uso de la Probabilidad y Estadística proporciona los medios de extraer de la información las características que permiten hacer inferencias, así como formas de evaluación de resultados.

No existe un modelo probabilístico al cual una muestra se ajuste en forma perfecta, pero muchos se aproximan bastante a ella. Las distribuciones de probabilidad estudiadas, son modelos que pueden ser usados para obtener el evento de diseño (de diversas obras hidráulicas), cuando éste se obtiene utilizando registros de eventos máximos (gastos máximos anuales); habrá un modelo que presente un mejor ajuste que los otros y habrá otros que definitivamente no sean válidos para esos datos y deban descartarse.

El ejemplo 3.1 muestra la importancia que posee el tamaño de la muestra en la reducción de la incertidumbre estadística ya que aunque los valores de los estimadores variaron arriba y abajo de los valores verdaderos en una forma no consistente en los ensayos hechos, la desviación estándar de los errores siempre tuvo la tendencia a disminuir al aumentar el tamaño de la "muestra"; cabe ahora señalar que los parámetros se ajustaron por momentos. Además se observó que los valores de los estimadores de los parámetros (de la Distribución Gumbel), variaron menos y tendieron hacia los verdaderos valores (conocidos en este ejemplo), para un tamaño de la muestra del orden de 60 en el caso estudiado.

Relacionando el ejemplo 3.1 con el 5.1, en donde la estimación de parámetros se hace por dos métodos diferentes, resulta interesante la siguiente observación:

Al ajustar una serie de eventos que se sabe siguen una determinada distribución, a esa misma distribución, es de esperarse que los parámetros estimados tengan valores muy aproximados a los de los verdaderos. Para el caso de ajuste por momentos, la diferencia fue considerable en tanto que para el caso de ajuste por mínimos cuadrados, fue nula. Esto conduce a considerar una posible ventaja de un método sobre el otro.

Por lo que respecta a las propiedades de los estimadores, éstas son importantes y son usadas implícitamente al aplicar algún método de estimación; en el ejemplo 4.1, además de comprobar algunas propiedades, se subraya la necesidad de contar con estimadores que las posean. En la práctica, las propiedades insesgo y eficiencia son las primeramente requeri--

das ya que generalmente implican otras.

Los métodos de ajuste están en relación con la incertidumbre estadística. Debe destacarse la importancia de aprovechar la experiencia del ingeniero al ajustar una muestra ya que esta puede facilitar la elección del método a usar. También debe señalarse que tanto en los ejemplos como en las aplicaciones de este trabajo, se consideró siempre el caso de una población, pero no debe eliminarse la posible existencia de dos poblaciones (secc.3.3a), en cuyo caso debe recurrirse a métodos de ajuste diferentes a los estudiados aquí.

El método de ajuste por momentos tiene la venteja, sobre los otros dos aquí estudiados, de su sencillez al aplicarlo; los estimadores obtenidos por momentos son funciones de la muestra que pueden usarse como primera aproximación en el método de máxima verosimilitud que requiere en muchos casos resolver una ecuación no lineal para obtener las estimas de los parámetros.

Los métodos de mínimos cuadrados y máxima verosimilitud necesitan de cálculos para los cuales en muchas ocasiones debe recurrirse al uso de programas de computadora aunque las propiedades que poseen los estimadores obtenidos por estos métodos lo justifican ya que en muchos casos superan a las de los estimadores obtenidos por momentos.

En el ejemplo 5.1, aunque se aplica el método de estimación por mínimos cuadrados, su principal finalidad es la de observar el comportamiento de valores máximos y justificar de alguna forma el uso generalizado de algunas distribuciones como la Gumbel en la obtención de un gasto máximo de diseño.

Los resultados también confirman la importancia que tiene el tamaño de la muestra, ya que a manera que este fue mayor en los ensayos realizados. La Distribución Gumbel fue "mejorando" con respecto a las otras distribuciones empleadas de acuerdo a la desviación estándar de los errores calculada.

De la observación de las gráficas correspondientes a este ejemplo (5.1), se deduce, aunque el número de ensayos fue corto, que no obstante que algunas curvas se ajustan en forma bastante aceptable (en forma gráfica) para valores intermedios y menores, no siempre su ajuste es el mejor para los valores mayores, lo que representa una circunstancia desfavorable que también va siendo eliminada cuando el tamaño de la muestra crece.

También, de acuerdo a los resultados observados en las gráficas, puede confirmarse que es acertado el uso de estas distribuciones de probabilidad en relación con funciones del período de retorno como modelos propuestos de una serie de observaciones de valores máximos ya que ninguna curva difiere en forma exagerada de las observaciones.

De las observaciones hechas a los resultados del ejemplo 6.1, se reafirma la importancia que tienen las pruebas de bondad de ajuste y la necesidad de contar con criterios de selección. Se propone la selección del modelo que presente el menor valor de los estadísticos correspondientes a las pruebas (D_2, D_3), así como el de menor valor de S (ec. 5.35)

Las pruebas de bondad de ajuste están asociadas a la incertidumbre de modelo. Tanto en los ejemplos como en las aplicaciones, se observa que los resultados para elegir un determinado modelo coinciden en forma general, en elegir el mismo al aplicar distintas pruebas.

Algunos criterios (X^2 y Kolmogorov-Smirnov), pueden relacionarse con pruebas de Hipótesis, lo que reafirma su conveniencia en la selección de una distribución de probabilidad. Los otros criterios no están relacionados con pruebas de hipótesis, es decir, el estadístico obtenido no sigue una distribución teórica conocida o no ha sido suficientemente estudiado, aunque sus resultados, al aplicar las pruebas en este trabajo, coinciden en forma general con los de las otras (X^2 y Kolmogorov-Smirnov)

Debe considerarse la posibilidad de utilizar un modelo compuesto (Ref.8) para obtener el evento de diseño; tal modelo consiste en una combinación lineal de los modelos propuestos, siendo la probabilidad posterior (Criterio Muestras de Verosimilitud), la proporción en que interviene cada uno. El evento de diseño estaría dado por lo tanto, por la suma de porcentajes de los eventos de diseño obtenidos según cada modelo propuesto; los porcentajes representan la probabilidad posterior de cada uno (su cálculo fue ilustrado en el ejemplo 6.3).

El criterio de Muestras de Verosimilitud es muy sensible a errores numéricos por lo que debe tenerse especial cuidado en el manejo de las cifras decimales para no incrementar o restar importancia a un modelo.

Al aplicar la prueba X^2, las muestras representadas por registros de gastos máximos anuales generalmente no satisfacen algunas de las condiciones requeridas; por ejemplo: "no menos del 20% de los intervalos tengan menos de cinco eventos simples"*, lo cual conduce a una menor con--

* Referencia 8

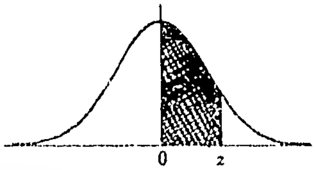
fiabilidad en la prueba.

La aplicación de las pruebas de Bondad de Ajuste a los gastos medidos en una estación hidrométrica conduce a resultados que en general coinciden en seleccionar la misma distribución como el modelo al que se ajustan mejor los datos, a partir de esa distribución puede obtenerse el evento de diseño; esta concordancia trae consigo una mayor confiabilidad en la aplicación de estas pruebas.

A P E N D I C E *

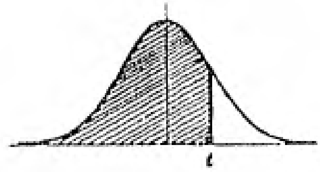
* Las tablas 1, 2 y 3 se obtuvieron de la referencia 14 y la tabla 4 de la referencia 2.

TABLA 1. Areas bajo la curva normal estandarizada de 0 a z



z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

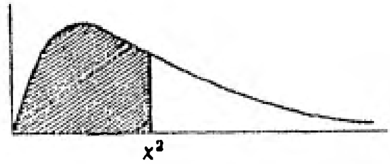
TABLA 2. Percentiles* t de la Distribución t de student con ν grados de libertad. (Area sombreada).



ν	$t_{0.995}$	$t_{0.99}$	$t_{0.975}$	$t_{0.95}$	$t_{0.90}$	$t_{0.80}$	$t_{0.75}$	$t_{0.70}$	$t_{0.60}$	$t_{0.55}$
1	63,66	31,82	12,71	6,31	3,08	1,376	1,000	0,727	0,325	0,158
2	9,92	6,96	4,30	2,92	1,89	1,061	0,816	0,617	0,289	0,142
3	5,84	4,54	3,18	2,35	1,64	0,978	0,765	0,584	0,277	0,137
4	4,60	3,75	2,78	2,13	1,53	0,941	0,741	0,569	0,271	0,134
5	4,03	3,36	2,57	2,02	1,48	0,920	0,727	0,559	0,267	0,132
6	3,71	3,14	2,45	1,94	1,44	0,906	0,718	0,553	0,265	0,131
7	3,50	3,00	2,36	1,90	1,42	0,896	0,711	0,549	0,263	0,130
8	3,36	2,90	2,31	1,86	1,40	0,889	0,706	0,546	0,262	0,130
9	3,25	2,82	2,26	1,83	1,38	0,883	0,703	0,543	0,261	0,129
10	3,17	2,76	2,23	1,81	1,37	0,879	0,700	0,542	0,260	0,129
11	3,11	2,72	2,20	1,80	1,36	0,876	0,697	0,540	0,260	0,129
12	3,06	2,68	2,18	1,78	1,36	0,873	0,695	0,539	0,259	0,128
13	3,01	2,65	2,16	1,77	1,35	0,870	0,694	0,538	0,259	0,128
14	2,98	2,62	2,14	1,76	1,34	0,868	0,692	0,537	0,258	0,128
15	2,95	2,60	2,13	1,75	1,34	0,866	0,691	0,536	0,258	0,128
16	2,92	2,58	2,12	1,75	1,34	0,865	0,690	0,535	0,258	0,128
17	2,90	2,57	2,11	1,74	1,33	0,863	0,689	0,534	0,257	0,128
18	2,88	2,55	2,10	1,73	1,33	0,862	0,688	0,534	0,257	0,127
19	2,86	2,54	2,09	1,73	1,33	0,861	0,688	0,533	0,257	0,127
20	2,84	2,53	2,09	1,72	1,32	0,860	0,687	0,533	0,257	0,127
21	2,83	2,52	2,08	1,72	1,32	0,859	0,686	0,532	0,257	0,127
22	2,82	2,51	2,07	1,72	1,32	0,858	0,686	0,532	0,256	0,127
23	2,81	2,50	2,07	1,71	1,32	0,858	0,685	0,532	0,256	0,127
24	2,80	2,49	2,06	1,71	1,32	0,857	0,685	0,531	0,256	0,127
25	2,79	2,48	2,06	1,71	1,32	0,856	0,684	0,531	0,256	0,127
26	2,78	2,48	2,06	1,71	1,32	0,856	0,684	0,531	0,256	0,127
27	2,77	2,47	2,05	1,70	1,31	0,855	0,684	0,531	0,256	0,127
28	2,76	2,47	2,05	1,70	1,31	0,855	0,683	0,530	0,256	0,127
29	2,76	2,46	2,04	1,70	1,31	0,854	0,683	0,530	0,256	0,127
30	2,75	2,46	2,04	1,70	1,31	0,854	0,683	0,530	0,256	0,127
40	2,70	2,42	2,02	1,68	1,30	0,851	0,681	0,529	0,255	0,126
60	2,66	2,39	2,00	1,67	1,30	0,848	0,679	0,527	0,254	0,126
120	2,62	2,36	1,98	1,66	1,29	0,845	0,677	0,526	0,254	0,126
∞	2,58	2,33	1,96	1,645	1,28	0,842	0,674	0,524	0,253	0,126

* Percentiles son los valores que dividen el área en 100 partes iguales.

TABLA 3. *Percentiles de la Distribución Xi cuadrado con ν grados de libertad.*



ν	$\chi^2_{0{,}995}$	$\chi^2_{0{,}99}$	$\chi^2_{0{,}975}$	$\chi^2_{0{,}95}$	$\chi^2_{0{,}90}$	$\chi^2_{0{,}75}$	$\chi^2_{0{,}50}$	$\chi^2_{0{,}25}$	$\chi^2_{0{,}10}$	$\chi^2_{0{,}05}$	$\chi^2_{0{,}025}$	$\chi^2_{0{,}01}$	$\chi^2_{0{,}005}$
1	7,88	6,63	5,02	3,84	2,71	1,32	0,455	0,102	0,0158	0,0039	0,0010	0,0002	0,0000
2	10,6	9,21	7,38	5,99	4,61	2,77	1,39	0,575	0,211	0,103	0,0506	0,0201	0,0100
3	12,8	11,3	9,35	7,81	6,25	4,11	2,37	1,21	0,584	0,352	0,216	0,115	0,072
4	14,9	13,3	11,1	9,49	7,78	5,39	3,36	1,92	1,06	0,711	0,484	0,297	0,207
5	16,7	15,1	12,8	11,1	9,24	6,63	4,35	2,67	1,61	1,15	0,831	0,554	0,412
6	18,5	16,8	14,4	12,6	10,6	7,84	5,35	3,45	2,20	1,64	1,24	0,872	0,676
7	20,3	18,5	16,0	14,1	12,0	9,04	6,35	4,25	2,83	2,17	1,69	1,24	0,989
8	22,0	20,1	17,5	15,5	13,4	10,2	7,34	5,07	3,49	2,73	2,18	1,65	1,34
9	23,6	21,7	19,0	16,9	14,7	11,4	8,34	5,90	4,17	3,33	2,70	2,09	1,73
10	25,2	23,2	20,5	18,3	16,0	12,5	9,34	6,74	4,87	3,94	3,25	2,56	2,16
11	26,8	24,7	21,9	19,7	17,3	13,7	10,3	7,58	5,58	4,57	3,82	3,05	2,60
12	28,3	26,2	23,3	21,0	18,5	14,8	11,3	8,44	6,30	5,23	4,40	3,57	3,07
13	29,8	27,7	24,7	22,4	19,8	16,0	12,3	9,30	7,04	5,89	5,01	4,11	3,57
14	31,3	29,1	26,1	23,7	21,1	17,1	13,3	10,2	7,79	6,57	5,63	4,66	4,07
15	32,8	30,6	27,5	25,0	22,3	18,2	14,3	11,0	8,55	7,26	6,26	5,23	4,60
16	34,3	32,0	28,8	26,3	23,5	19,4	15,3	11,9	9,31	7,96	6,91	5,81	5,14
17	35,7	33,4	30,2	27,6	24,8	20,5	16,3	12,8	10,1	8,67	7,56	6,41	5,70
18	37,2	34,8	31,5	28,9	26,0	21,6	17,3	13,7	10,9	9,39	8,23	7,01	6,26
19	38,6	36,2	32,9	30,1	27,2	22,7	18,3	14,6	11,7	10,1	8,91	7,63	6,84
20	40,0	37,6	34,2	31,4	28,4	23,8	19,3	15,5	12,4	10,9	9,59	8,26	7,43
21	41,4	38,9	35,5	32,7	29,6	24,9	20,3	16,3	13,2	11,6	10,3	8,90	8,03
22	42,8	40,3	36,8	33,9	30,8	26,0	21,3	17,2	14,0	12,3	11,0	9,54	8,64
23	44,2	41,6	38,1	35,2	32,0	27,1	22,3	18,1	14,8	13,1	11,7	10,2	9,26
24	45,6	43,0	39,4	36,4	33,2	28,2	23,3	19,0	15,7	13,8	12,4	10,9	9,89
25	46,9	44,3	40,6	37,7	34,4	29,3	24,3	19,9	16,5	14,6	13,1	11,5	10,5
26	48,3	45,6	41,9	38,9	35,6	30,4	25,3	20,8	17,3	15,4	13,8	12,2	11,2
27	49,6	47,0	43,2	40,1	36,7	31,5	26,3	21,7	18,1	16,2	14,6	12,9	11,8
28	51,0	48,3	44,5	41,3	37,9	32,6	27,3	22,7	18,9	16,9	15,3	13,6	12,5
29	52,3	49,6	45,7	42,6	39,1	33,7	28,3	33,6	19,8	17,7	16,0	14,3	13,1
30	53,7	50,9	47,0	43,8	40,3	34,8	29,3	24,5	20,6	18,5	16,8	15,0	13,8
40	66,8	63,7	59,3	55,8	51,8	45,6	39,3	33,7	29,1	26,5	24,4	22,2	20,7
50	79,5	76,2	71,4	67,5	63,2	56,3	49,3	42,9	37,7	34,8	32,4	29,7	28,0
60	92,0	88,4	83,3	79,1	74,4	67,0	59,3	52,3	46,5	43,2	40,5	37,5	35,5
70	104,2	100,4	95,0	90,5	85,5	77,6	69,3	61,7	55,3	51,7	48,8	45,4	43,3
80	166,3	112,3	106,6	101,9	96,6	88,1	79,3	71,1	64,3	60,4	57,2	53,5	51,2
90	128,3	124,1	118,1	113,1	107,6	98,6	89,3	80,6	73,3	69,1	65,6	61,8	59,2
100	140,2	135,8	129,6	124,3	118,5	109,1	99,3	90,1	82,4	77,9	74,2	70,1	67,3

TABLA 4. *Valores críticos d_2 de la prueba Kolmogorov-Smirnov.*

Tamaño de la muestra.	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
5	0.51	0.56	0.67
10	0.37	0.41	0.49
15	0.30	0.34	0.40
20	0.26	0.29	0.35
25	0.24	0.26	0.32
30	0.22	0.24	0.29
40	0.19	0.21	0.25
$n > 40$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.63/\sqrt{n}$

```
LIST

10  PRINT "";
15  REM  PROGRAMA PARA ESTIMAR POR MINIMOS CUADRADOS LOS PARAMETROS DE LAS
       DISTRIBUCIONES NORMAL Y EXPONENCIAL
18  REM  ESTIMACION DE EVENTOS Y [illegible] ERROR ESTANDAR.
20  READ A$
25  DATA "NORMAL"
30  DIM T(50),X(50),Y(50),[illegible](50),PR(50),YE(50),RA(50),YS(50),YI(50),ER(50)
35  DIM [illegible](50),X1(50),PR(50),R(50),XP(50),E(50)
38  DIM X2(50)
40  PRINT "                                              DISTRIBUCION   A$
50  READ N
60  DATA  26
82  IF A$ = "NORMAL" THEN  GOSUB 530
83 S1 = 0
84 S2 = 0
86 S3 = 0
87 S4 = 0
88 S5 = 0
90  FOR I = 1 TO N
95  READ Y(I)
100 T(I) = (N + 1) / [illegible]
110  IF A$ = "NORMAL" THEN X(I) = [illegible] X1(I)
115  IF A$ = "EXPONENCIAL" THEN Y(I) = [illegible] LOG [illegible](I))
135 YC(I) = Y(I) ^ 2
160 S2 = S2 + Y(I)
175 S4 = S4 + YC(I)
190  NEXT I
192  DATA  1591,1325,1231,1151,1086,1052,816,734,720,639,639,637,599,531,5
      21,501,477,475,430,428,379,339,319,285,246,224
195  GOSUB 468
200 MX = S1 / N
210 MY = S2 / N
250 VX = 1 / N * (S3 - N * MX ^ 2)
260 VY = 1 / N * (S4 - N * MY ^ 2)
270 SXY = 1 / N * (S5 - N * MX * MY)
280 RXY2 = SXY ^ 2 / (VX * VY)
290 SC = N / (N - 2) * (1 - RXY2) * VY
295  IF SC < 0 THEN SC = - SC
300 S = SQR (SC)
310 BETA = SXY / VX
320 ALFA = MY - BETA * MX
330 A = - 1 / BETA
340 B = ALFA
350  PRINT "ALFA= "A,"BETA= "B,"MX= "MX,"MY= "MY
355  PRINT "VX= "VX,"VY= "VY,"S= "S
367 S7 = 0
370  FOR I = 1 TO N
375 E(I) = B + [illegible] X2(I)
380 YE(I) = E(I)
390 RAIZ(I) = SQR ([illegible] (1 + [illegible] MX) ^ [illegible] VX))
400 YS(I) = E(I) + 1.7 * RAIZ(I)
410 YI(I) = E(I) - 1.7 * RAIZ(I)
415 ERR(I) = (Y(I) [illegible]) ^ 2
418 S7 = S7 + ERR(I)
420  PRINT T(I),X2(I),Y(I),YE(I),RAIZ(I)," ("YI(I)":" "YS(I)" )"
430  NEXT I
440 ERS = SQR (S7 / (N - 2))
460  PRINT "DESVIACION ESTANDAR DE LOS ERRORES= "ERS,"MY= "MY
465  END
468  FOR I = 1 TO N
470 X2(I) = X(I)
475 XC(I) = X2(I) ^ 2
480 PR1(I) = X2(I) * Y(I)
490 S1 = S1 + [illegible]
495 S3 = S3 + XC(I)
500 S5 = S5 + PR1(I)
505  NEXT I
510  RETURN
530  FOR I = 1 TO N
535 T(I) = (N + 1) / I
540 PB(I) = 1 / T(I)
550 SIG = 1
560  IF PB(I) [illegible] 0.5 THEN 590
570 PB(I) = 1 - PB(I)
580 SIG = - 1
590 R(I) = SQR ([illegible] LOG (1 / [illegible]
600 XP(I) = R(I) - (2.515517 + 0.[illegible] + 0.010[illegible] * R(I) [illegible]
      + 1.432788 * R(I) + 0.189269 [illegible] .001308 * R(I) [illegible]
610 X1(I) = SIG * XP(I)
630  NEXT I
640  RETURN
```

```
PRINT "";
REM  PROGRAMA PARA ESTIMAR LOS PARAMETROS DE LA DISTRIBUCION LOGNORMAL
   POR EL METODO DE MAXIMA VEROSIMILITUD. OBTENCION DEL ERROR ESTANDAR
READ N,NP
0  DATA  26,3
  DIM X(30),EE(30),XE(30),T(30)
6  DIM XL(30),ERR(30)
7 TL = 0.000001
8 SS = 0
0  FOR I = 1 TO N
0  READ X(I)
5 SS = SS + X(I)
7 SX = SX + X(I) ^ 2
0  NEXT I
0  DATA  1591,1325,1231,1151,1086,1052,816,734,720,639,639,637,594,531,52
   1,501,477,475,430,428,379,339,319,285,246,224
5  PRINT "                                         DISTRIBUCION LOGNORMAL"
6  PRINT " "
0 M = SS / N
5 CF = 0
0 VR = 0
0  FOR I = 1 TO N
0 VR = (X(I) - M) ^ 2 + VR
5 CF = CF + (X(I) - M) ^ 3
00  NEXT I
05 R2 = SX / N - M ^ 2
10 DST =  SQR (VR / N)
20 CA = (CF / N) / DST ^ 3
30  PRINT "        MEDIA= "M,"  DESVIACION ESTANDAR= "DST
40  PRINT "  COEFICIENTE DE ,ASIMETRIA= "CA
45  IF NP = 2 THEN  GOTO 220
50 G =  SQR (4 + CA ^ 2)
60  IF CA < 0 THEN  GOTO 190
70 AL = ((CA + G) * 0.5 ^ (1 / 3) + ((CA - G) * 0.5) ^ (1 / 3)
80  GOTO 200
90 AL = ((CA + G) * 0.5) ^ (1 / 3) - ((G - CA) * 0.5) ^ (1 / 3)
200 AL = M - DST / AL
210  GOTO 230
220 AL = 0
230 G2 = (M - AL) ^ 2
240 G4 = G2 ^ 2
250 G = G2 + R2
260 ML = 0.5 *  LOG (G4 / G)
270 SL =  SQR ( LOG (G / G2))
280  PRINT "                         PARAMETROS ESTIMADOS POR MOMENTOS: "
290  PRINT "  M= "ML,"SL= "SL,"AL= "AL
300  PRINT " "
310  REM  OBTENCION DE LOS PARAMETROS POR MAXIMA VEROSIMILITUD : "
315  IF NP = 2 THEN 590
320 ITER = 1
330 VM = 1000000
335  PRINT "ITERACION","AL","F(AL)"
340  FOR I = 1 TO N
350  IF X(I) < = VM THEN VM = X(I)
360  NEXT I
370 AL = 0.98 * VM
380 A1 = AL
390 A2 = AL * 0.95
400 AM = - ( ABS (AL)) * 0.5
410 H = A2 - A1
414 A = A1
420  GOSUB 720
430 A = A2
435 F1 = F
440  GOSUB 720
445 F2 = F
450  PRINT ITER,A2,F2
460  IF  ABS (F2) < = TL THEN  GOTO 610
470 H = H * F2 / (F1 - F2)
480 A2 = A2 + H
490  IF A2 < = AM THEN  GOTO 560
  F1 = F2
  ITER = ITER + 1
  IF ITER < = 50 THEN  GOTO 430
530  PRINT "EL PROCEDIMIENTO REQUIERE MAS ITERACIONES"
540 ID = 
550  GOTO 1700
560  PRINT "LA DISTRIBUCION LOGNORMAL NO DA RESULTADOS PARA ESTOS DATOS"
570 ID = 
580  GOTO 1700
590 A2 = 0
```

```
595 A = A2
600  GOSUB 720
610 ML = S1
620 SL = SQR (S4)
630 AL = A2
640  PRINT " "
650  PRINT " PARAMETROS ESTIMADOS POR [illegible] APROXIMA[illegible]"
660  PRINT " "
700  PRINT [illegible] ML,"SL= ";SL,[illegible] AL
710  GOTO [illegible]
720  REM  CALCULO DE [illegible] NA [illegible]ON NO[illegible]
730 S1 = 0
740 S2 = 0
750 S3 = 0
760 S4 = 0
770  FOR I = 1 TO N
780 XA = X(I) - A
790 XLA(I) =  LOG (XA)
800 XA = 1 / XA
810 S1 = S1 + XLA(I)
820 S2 = S2 + XLA(I) * XA
830 S3 = S3 + XA
840  NEXT I
850 S1 = S1 / N
860  FOR I = 1 TO N
870 S4 = S4 + (XLA(I) - S1) ^ 2
880  NEXT I
890 S4 = S4 / N
900 F = S2 - (S4 - S1) * S3
920  RETURN
1000  REM  ESTIMACION DE GASTOS Y CALCULO DE [illegible] ERRORES ESTANDAR CORRESPO
      NDIENTES
1010 SQ = [illegible]
1020 VA = 0
1030 CS = 0
1040 CM = 0
1050 TF = 1.645 + 1.5241[illegible] / (N - 2) + 1.420681[illegible]
1060 SGO = SL ^ 2
1070 E1 = EXP (2 * [illegible])
1080 E2 = EXP (SGO [illegible] ML)
1090 E3 = EXP (SGO * [illegible])
1100 F = 0.5 * (1 + 1 [illegible] SGO [illegible] E1
1110 D = F [illegible] + 0.5 [illegible] SGO [illegible] E2
1120 DN = D * N
1140 FC = SGO / DN
1150  IF NP = 2 THEN  GOTO 11[illegible]
1160 VA = 1 / (2 * DN)
1170 CS = FC * E3
1180 CM = [illegible]
1190 VO = FC [illegible] (F [illegible]
1200 VR = FC [illegible]
1210 CO = [illegible] FC [illegible]
1450 SUM = 0
1500  FOR I = 1 TO N
1510 T(I) = (N [illegible]
1520 P = 1 / T(I)
1530 SIG = 1
1540  IF P [illegible] = 0.5 THEN [illegible] 1570
1550 SIG = - 1
1560 [illegible] 1 [illegible]
1570 R = SQR ( LOG (1 / [illegible] 2))
1580 [illegible] = R - (2.515517 [illegible] .802853 [illegible] .010328 [illegible] 1.432788
      * R + 0.189269 * [illegible] 2 + 0.001308 [illegible]
1590 FUN = SIG * XP
1595 XP = FUN
1600 W = EXP (ML + XP * SL)
1610 C = XP * W / SL
1620 XE(I) = W + AL
1625 [illegible] = VA + [illegible] VR [illegible] + 2 * W * CM [illegible]
      W * CO
1630 ERR(I) = [illegible]
1640 SUM = SUM + [illegible] (I [illegible]
1650 LS = XE(I) [illegible] ERR(I)
1660 LI = XE(I) [illegible] ERR(I)
1670  PRINT I,T(I),X(I),XE(I),ERR(I),[illegible]
1675  NEXT I
1680 [illegible] SQR (SUM [illegible])
1685  PRINT "SUMA DE LAS DIFE[illegible]IAS DE [illegible] [illegible]ENIDOS [illegible]I
      MADOS= ";SUM
1690  PRINT "DESVIACION [illegible]DAR DE LOS ERRORES= [illegible]
1700  END
```

```
20   PRINT "";
30   REM  PROGRAMA PARA ESTIMAR LOS PARAMETROS DE LA DISTRIBUCION GUMBEL
       POR EL METODO DE MAXIMA VEROSIMILITUD.
32   REM  OBTENCION DEL ERROR ESTANDAR.
40   PRINT "  "
60   DIM Q(30),X(30)
65   DIM T(30)
70   READ EP,N,SN
80   DATA  0.000001,26,1.0961
90   FOR I = 1 TO N
100   READ Q(I)
110   NEXT I
120   DATA  1591,1325,1231,1151,1086,1052,816,734,720,639,639,637,594,531
       ,521,501,477,475,430,428,379,339,319,285,246,224
140  SS = 0
150  S1 = 0
160   FOR I = 1 TO N
170  S1 = S1 + Q(I)
180   NEXT I
190  ME = S1 /'N
200   FOR I = 1 TO N
210  SS = SS + (Q(I) - ME) ^ 2
220   NEXT I
230  ST =  SQR'(SS / N)
240  AN = SN / ST
250  S2 = 0
260  S3 = 0
270  S4 = 0
280   FOR I = 1 TO N
290  EL =  EXP ( - AN * Q(I))
300  S2 = S2 + EL
310  S3 = S3 + Q(I) * EL
320  S4 = S4 + Q(I) * Q(I) * EL
330   NEXT I
340  FX = N / AN - S1 + N * S3 / S2
350  DX =  - N / (AN * AN) + N * ((S3 * S3 - S2 * S4) / (S2 * S2))
360  CO =  - FX / DX
370   IF  ABS (CO) <  = EP THEN 400
380  AN = AN + CO
390   GOTO 250
400  U = ( LOG (N) -  LOG (S2)) / AN
410   PRINT "  LOS PARAMETROS DE LA DISTRIBUCION GUMBEL SON  :"
420   PRINT "  "
430   PRINT "ALFA = "AN
440   PRINT "BETA= "U
450  N2 = N + 1
460   FOR I = 1 TO N
470  T = N2 / I
480  X(I) = U -  LOG ( LOG (T / (T - 1))) / AN
490   NEXT I
500   PRINT "  "
510   PRINT  SPC( 1)"I" SPC( 14)"GASTO" SPC( 14)"GASTO"
515   PRINT  SPC( 14)"REGISTRADO" SPC( 9)"CALCULADO"
520   PRINT "  "
525  TP = 1.645 + 1.5241028 / (N - 2) + 1.4206847 / (N - 2) ^ 2
530   FOR I = 1 TO N
531  T(I) = (N + 1) / I
532  Y =  -  LOG ( LOG (T(I) / (T(I) - 1)))
534  ERS = (1.1086 + 0.514 * Y + 0.6079 * Y ^ 2) / (N * AN ^ 2)
536  ST =  SQR (ERS)
537  LS = X(I) + ST * TP
538  LI = X(I) - ST * TP
540  S7 = S7 + (Q(I) - X(I)) ^ 2
550   PRINT I,T(I),Q(I),X(I),ST,"( "LS;" - "LI;" )"
560   NEXT I
561   PRINT "  "
562  V =  SQR (S7 / (N - 2))
564   PRINT "SUMA DE LAS DIFERENCIAS AL CUADRADO DE EVENTOS OBSERVADOS Y
       EVENTOS ESTIMADOS = "S7
566   PRINT "  DESVIACION ESTANDAR DE LOS ERRORES = "V
570   END
```

```
5  PRH 3
   PRINT "";
7  REM  PROGRAMA PARA ESTIMAR LOS PARAMETROS DE LA DISTRIBUCION GAMMA POR
     EL METODO DE MAXIMA VEROSIMILITUD Y OBTENER EL ERROR ESTANDAR.
8  READ N,NP
10  DATA  26,3
15  DIM X(30),EF(30),YE(30),T(30)
17 TL = 0.000001
18 SS = 0
20  FOR I = 1 TO N
30  READ X(I)
35 SS = SS + X(I)
40  NEXT I
50  DATA  1591,1325,1231,1151,1086,1052,816,734,720,[illegible],[illegible]37,637,594,531,52
     1,501,477,475,430,4[illegible],379,339,319,285,246,224
55  PRINT "                                                    DISTRIBUCION  GAMMA
     "
56  PRINT " "
60 M = SS / N
65 CF = 0
70 VR = 0
80  FOR I = 1 TO N
90 VR = (X(I) - M) ^ 2 + VR
95 CF = CF + (X(I) - M) ^ 3
100  NEXT I
110 DST =  SQR (VR / N)
120 CA = (CF / N) / DST ^ 3
130  PRINT "      MEDIA= "M,"  DESVIACION ESTANDAR= "DST
140  PRINT "  COEFICIENTE DE ASIMETRIA= "CA
150 K = (2 / CA) ^ 2
160 L =  SQR (K) / DST
170 D = M - K / L
180  PRINT "   AJUSTE POR MOMENTOS : "
190  PRINT "K= "K,"LAMBDA= "L,"DELTA= "D
200 GP =  - ( ABS (D)) * 0.5
250  REM  OBTENCION DE LOS PARAMETROS POR MAXIMA VEROSIMILITUD
280 N2 = N * N
290  IF NP = 2 THEN  GOTO [illegible]30
300 N3 = N2 * N
310  PRINT "ITERACION","F","DELTA"
315  PRINT " "
320 VM = 10000000000
330  FOR I = 1 TO N
340  IF X(I) < = VM THEN VM = X(I)
350  NEXT I
360 G1 = 0.98 * VM
370 G2 = G1
380 KO = 0
390 S1 = 0
400 S2 = 0
410 S3 = 0
420 S4 = 0
430  FOR I = 1 TO N
440 AX = X(I) - G2
450 S1 = S1 + AX
460 S2 = S2 + 1 / AX
470 S3 = S3 + 1 / AX ^ 2
480 S4 = S4 +  LOG (AX)
490  NEXT I
500 B1 = S2 / (S2 - N2 / S1)
510 A1 = S1 / N
520 W = S4 / N
530 V = [illegible] N2 / S1
540 DV = [illegible] N3 / S1 ^ 2
550 U = [illegible]
560 DU = [illegible]
570 DRB = (V * DU - [illegible] DV) [illegible]
580 DA = - S2 / N
590 C1 = B1 + [illegible]
600 TG = 1 / C1 + 0.5 / [illegible] 0.1666666[illegible] / C1 ^ 3 + 0.0333333 [illegible]
     5 + 0.02380[illegible]
     1 / B1 ^ 2
```

```
610 DG =  LOG (C1) - 0.5 / C1 - 0.083333333 / C1 ^ 2 + 0.0083333333 /
    / C1 ^ 4 - 0.003968254 / C1 ^ 6 - 1 / (B1 + 1) - 1 / B1
620 F =  LOG (B1) - DG + W -  LOG (A1)
630 DI = (1 / B1 - TG) * DRB - S2 / N + 1 / A1
640 H = F / DI
650 G2 = G1 - H
660  IF G2 < = GP THEN  PRINT "NO TERMINE AND GOTO 1800
670 KO = KO + 1
680  PRINT "    "KO,F,G2
690  IF  ABS (H) < = TL THEN  GOTO 800
700  IF KO = 20 THEN  PRINT "NO TERMINE" AND  GOTO 1800
710 G1 = G2
720  GOTO 390
730 S1 = 0
740 S2 = 0
750  FOR I = 1 TO N
760 S1 = S1 + X(I)
770 S2 = S2 + 1 / X(I)
780  NEXT I
790 G2 = 0
800 K = 1 / (1 - N2 / (S1 * S2))
810 L = (N * K) / S1
820 D = G2
830 PM = 1 / L + D
840 PD =  SQR (1 / L ^ 2 * K)
850 PS =  SQR (4 / K)
855  PRINT " "
860  PRINT "    PARAMETROS ESTIMADOS POR MAXIMA VEROSIMILITUD :"
870  PRINT " "
880  PRINT "K= "K,"L = "L,"D = "D
900  REM
1100  REM  OBTENCION DE LOS EVENTOS ESTIMADOS Y ERRORES ESTANDAR CO
     RRESPONDIENTES.
1110  PRINT "   "
1120 SD = 0
1130 VD = 0
1140 CD = 0
1150 CB = 0
1160 TP = 1.645 + 1.5241028 / (N - 2) + 1.4206847 / (N - 2) ^ 2
1170 NU = 2 * K
1180 CTE = 2 / (9 * NU)
1190 C1 = K + 2
1200 C2 = K - 2
1210 C3 = K - 1
1220 A2 = (1 / L) ^ 2
1230 A3 = A2 * 1 / L
1240 A4 = A2 ^ 2
1250 TG = 1 / C1 + 0.5 / C1 ^ 2 + 0.166666667 / C1 ^ 3 - 0.033333333 / C1 ^
     5 + 0.0238095238 / C1 ^ 7 - 0.033333333 / C1 ^ 9 + 1 / (K + 1) ^ 2 +
     1 / K ^ 2
1260 DG =  LOG (C1) - 0.5 / C1 - 0.083333333 / C1 ^ 2 + 0.0083333333 / C1 ^
     4 - 0.003968254 / C1 ^ 6 - 1 / (K + 1) - 1 / K
1270 D1 = (2 * TG - (2 * K - 3) / (C3 ^ 2)) / (C2 * A4)
1280 DN = D1 * N
1290  IF NP = 2 THEN  GOTO 1330
1300 VD = (K * TG - 1) / (DN * A2)
1310 CD = (1 / C3 - TG) / (DN * A2)
1320 CB = (1 - K / C3) / (DN * A3)
1330 VA = (DG / C2 - 1 / C3 ^ 2) / (DN * A2)
1340 VB = 2 / (DN * A4 * C2)
1350 C = (1 / C3 - 1 / C2) / (DN * A3)
1360 B1 = K ^ 0.3333333333
1370 B2 = 9 * K ^ (2 / 3)
1380 B3 = 3 * K ^ (1 / 6)
1390 B4 = 27 * K ^ 1.6666666667
1400 B5 = 18 * K ^ 1.1666666667
1450 SUM = 0
1500  FOR I = 1 TO N
1510 T(I) = (N + 1) / I
1520 P = 1 / T(I)
1530 SIG = 1
1540  IF P < = 0.5 THEN  GOTO 1570
1550 SIG = - 1
1560 P = 1 - P
1570 R =  SQR ( LOG (1 / P ^ 2))
```

```
1580 XP = R - (2.5[illegible]17 + 0.[illegible] * R + 0.010[illegible]28 * R [illegible] / (1 + 1.432788
     * R + 0.18[illegible] * R ^ 2 + 0.001308 * R [illegible])
1590 FUN = [illegible]IG [illegible]
1595 XP = FUN
1600 SHI [illegible] NU [illegible] SQR [illegible]
1610 Y = 0.5 * [illegible]
1620 YA = (B1 [illegible] / [illegible] + (P / B3) [illegible] 3
1630 XB [illegible] B2 + [illegible] / B4 [illegible] (P / B5 [illegible] YP [illegible] 3
     ) ^ 2
1640 ST = XA [illegible] VA [illegible] XB [illegible] XA * XB * C [illegible] A [illegible] CD [illegible]
     [illegible] * XB * CB
1650 XE(I) = Y / L [illegible]
1655 SUM = SUM + (XE(I) [illegible] ^ 2
1660 EE(I) = SQR [illegible]
1670 LS = XE(I) + EE(I) [illegible]
1680 LI = XE(I) - EE(I) [illegible]
1690  PRINT I,X(I),XE(I),EE(I),"[illegible]LI;"[illegible]
1700  NEXT I
1710 DI = SQR (SUM [illegible])
1720  PRINT " "
1730  PRINT " [illegible]UMA DE LAS DIFERENCIAS AL CUADRADO DE EVENTOS OBSERVADOS Y
     EVENTOS ESTIMADOS= "SUM
1740  PRINT "  DESVIACION ESTANDAR DE [illegible] ERRORES= "DI
1800  END
```

```
5  PR# 3
10  PRINT "";
15  REM  PROGRAMA PARA OBTENER PROBABILIDADES ESTIMADAS DADOS LOS VA
     LORES DE LOS PARAMETROS.
17  REM  PRUEBAS KOLMOGOROV-SMIRNOV Y SUMA DE LAS DIFERENCIAS AL CUA
     DRADO DE PROBABILIDADES OBSERVADAS Y ESTIMADAS
20  READ N
30  DATA  26
35 MN = 781.7931
40 SN = 289.1791
45 ML = 6.1430
50 SL = 0.4457
52 AL = 100.5804
55 AG = 0.004053
60 BG = 249.5045
65 AE = 466.6667
70 LE = 0.003
75 K = 1.2556
80 L = 0.002788
85 U = 317.7643
87  DIM Y(30),E1(30),E2(30),KS(30),SER(30),X1(30),XC(30)
88  DIM PB(30),F1(30),TRM(30),FAC(30),IXC(30),FC(30)
89  DIM YL(30)
90  READ H$
95  DATA "EXPONENCIAL"
96  PRINT "                                   DISTRIBUCION  "H$
97  PRINT " "
100  FOR I = 1 TO N
110  READ Y(I)
120  NEXT I
130  DATA  1591,1325,1231,1151,1086,1052,816,734,720,639,639,637,594
     ,531,521,501,477,475,430,428,379,339,319,285,246,224
140  PRINT " "
150  IF H$ = "NORMAL" THEN  GOSUB 1040
152  IF H$ = "LOGNORMAL" THEN  GOSUB 1040
153  IF H$ = "GAMMA" THEN  GOSUB 1600
154  IF H$ = "EXPONENCIAL" THEN  GOSUB 1500
155  IF H$ = "GUMBEL" THEN  GOSUB 1400
156  IF H$ = "GAMMA" THEN  GOSUB 1040
158  PRINT " "
300  PRINT " "
310  SUM = 0
320  FOR I = 1 TO N
330  IF E1(I) < 0 THEN E1(I) = 0
340 E2(I) = 1 - I / (N + 1)
350  KS(I) = ABS (E1(I) - E2(I))
360  SER(I) = (E1(I) - E2(I)) ^ 2
370  SUM = SUM + SER(I)
375  PRINT Y(I),E1(I),E2(I),KS(I),SER(I)
380  NEXT I
385  PRINT " "
390  PRINT "SUMA DE LAS DIFERENCIAS AL CUADRADO DE PROBABILIDADES OB
     SERVADAS Y ESTIMADAS = "SUM
395 MAX = KS(1)
400  FOR B = 2 TO N
405  IF KS(B) > MAX THEN MAX = KS(B)
410  NEXT B
420  PRINT " "
425  PRINT "PRUEBA KOLMOGOROV-SMIRNOV. DIFERENCIA MAYOR= "MAX
500  END
1040  REM  OBTENCION DE LA PROBABILIDAD NORMAL
1050  FOR I = 1 TO N
1055  X1(I) = (Y(I) - MN) / SN
1056  IF H$ = "LOGNORMAL" THEN X1(I) = ( LOG (Y(I) - AL) - ML) / SL
1060  IF H$ = "GAMMA" THEN Y[illegible] = ((YC(I) / NU) ^ (1 / 3) + 2 / [illegible]
     .U) - 1) * SQR ((9 * NU) / 2)
1065  F = 1
```

```
1090 IF [illegible] THEN [illegible]
1100 X1(I) = [illegible]
1110 [illegible]
1130 IF [illegible](I) [illegible] THEN [illegible]
1140 XC(I) = [illegible]
1150 PB(I) = [illegible] [illegible] XC(I) [illegible] 437
     XC(I) [illegible]
1160 GOTO 1340
1170 IF X1(I) [illegible] THEN [illegible]
1180 XC(I) [illegible]
1190 PB(I) [illegible]
1200 PT(I) [illegible]
1210 FAC(I) [illegible]
1220 DUD = [illegible]
1230 PT(I) = [illegible] 2 * FAC [illegible]
1240 TRM(I) = PT(I) [illegible]
1250 PB(I) = PB(I) + TRM(I)
1260 IF [illegible] .00007) [illegible] THEN GOTO 1300
1270 FAC(I) = FAC(I) + 1
1280 DUD = DUD + 2
1290 GOTO 1230
1300 PB(I) = 0.79788455 * X1(I) * PB(I)
1310 GOTO 1340
1320 IXC(I) = 1 / [illegible](I)
1330 PB(I) = 1 - 0.79788453 * EXP ( [illegible] X1(I) [illegible] 1 - IXC
     (I) * (1 - IXC(I) [illegible] 3 - IXC(I) * (15 [illegible] * 105)))
1340 FC(I) = 0.5 * PB(I)
1350 E1(I) = 0.5 [illegible] * FC(I)
1360 NEXT I
1364 IF H$ = "NORMAL" THEN PRINT "MN= "MN,"SN= [illegible]
1366 IF H$ = "LOGNORMAL" THEN PRINT "ML= "ML,"SL= "SL,"AL= "AL
1368 IF H$ = "GAMMA" THEN PRINT [illegible] = "K,"L= "L,"D= "D
1370 RETURN
1400 FOR I = 1 TO N
1410 E1(I) = [illegible] AG [illegible]
1420 NEXT I
1425 PRINT "AG= "AG,"BG= "BG
1430 RETURN
1500 FOR I = 1 TO N
1510 E1(I) [illegible] YF [illegible] Y [illegible] AE [illegible])
1520 NEXT I
1525 PRINT "AE= "AE, [illegible]
1530 RETURN
1600 FOR I = 1 TO N
1610 Y1(I) [illegible]
1620 XC(I) [illegible] Y1(I))
1630 NEXT I
1640 NU = [illegible]
1650 RETURN
```

BIBLIOGRAFIA Y REFERENCIAS

1. Wonnacott, T. H. and Wonnacott, R. J., 1977, *Introductory Statistics*, John Wiley & Sons. U.S.A.

2. Benjamín, J. R. and Cornell, C.A., 1970, *Probability, Statistics and Decisión for Civil Engineers*, Mc. Graw Hill, Nueva York.

3. Yevjevich, V., 1972, *Probability and Statistics in Hidrology*, Water Resources Publications, Fort Collins, Colorado.

4. Kite, G. W., 1977, *Frecuency and Risk. Analyses in Hidrology*, Water Resources Publications, Fort Collins, Colorado.

5. Haan, Charles T., 1977, *Statistical Methods in Hidrology*, Towa State University Press.

6. Lindgren, B.W., 1968, *Statistical Theory*, Macmillan Company Canada, Printed in U.S.A.

7. Olivera Salazar A. y Zúñiga Barrera S., 1977, *Serie de Probabilidad y Estadística*. IMPOS editores, S. A., México.

8. Castano, E. and Duckstein, L., 1978 (agosto), *Choice of Distribution Functions for Hidrologyc Design*, Water Resources Research.

9. Kite, G. W., 1975 (febrero), *Confidence Limits for Design Events*, Water Resources Research.

10. Freund, J. E., 1971, *Statistical Mathematical*, Prentice-Hall Inc., N. J., U.S.A.

11. Wallis, J. R., Matalas, N. C. and Slack, J. R., 1976 (junio), *Effect of Sequence Lenght n on the choice of Assumed Distribution of Floods*, Water Resources Research.

12. González Villarreal, F.J., 1970, *Contribución al Análisis de Frecuencias de Valores Extremos de los Gastos Máximos en un Río*, Instituto de Ingeniería, U.N.A.M.

13. Ramírez Valle, A.G., 1977, *Generación de Muestras Hidrológicas Sintéticas Mensuales y su Aplicación a los Datos de la Estación Hidrológica San Fernando, Tamps.*, Tesis Profesional, U.N.A.M.

14. Spiegel, M. R., 1975, *Estadística*, Serie Schaum, McGraw-Hill, México.

15. Viessman, W. Jr., Knapp, J. W., Lewis G. L. and Harbaugh, T. E., 1977, *Introduction to Hidrology*, Harper & Row Publishers, Nueva York.

16. Tung, Y. K. and Mays, L. W., 1980 (mayo), *Risk Analysis for Hydraulic Design*, Journal of the Hydraulics Division, ASCE.