



ANALISIS ESTADISTICO DE VALORES EXTREMOS

Aplicaciones en hidrología

Gloria Elena Maggio

RESUMEN

El objetivo de este trabajo es ofrecer una guía práctica para el estudio estadístico de tendencias de valores extremos que incluye expresiones de cálculo, métodos de ajuste, gráficos y tablas.

Luego de una introducción referida a conceptos generales de estadística clásica, se pasa a desarrollar la teoría de valores extremos para continuar con una exposición detallada de la aplicación de la función de Gumbel a este tipo de problemas y una síntesis del empleo de las funciones aplicadas a los logaritmos de los datos, es decir Galton, Fréchet y log-Pearson. Por último, se extraen conclusiones de tipo general y se presenta un ejemplo concreto vinculado con el estudio del emplazamiento de una central nuclear.

Las técnicas aquí expuestas pueden ser aplicadas a fenómenos meteorológicos, tales como lluvias y vientos, a caudales o cotas hidrométricas de cursos fluviales y a cualquier otro problema en el que requiera conocer el comportamiento de los máximos o mínimos de una población, en lugar del típico análisis de valores medios.

TABLA DE CONTENIDO

1. CONCEPTOS BASICOS DE ESTADISTICA	1
1.1 PROBABILIDAD	1
1.2 SUCECOS SIMULTANEOS.....	2
1.2.1 Regla de multiplicación	3
1.2.2 Regla de adición	3
1.3 FUNCIONES DE DISTRIBUCION.....	4
1.4 PARAMETROS ESTADISTICOS.....	6
1.4.1 Momentos.....	6
1.4.1.1 Función generatriz	7
1.4.2 Posición	7
1.4.2.1 Valor medio	8
1.4.2.2 Modo.....	8
1.4.2.3 Mediana	8
1.4.3 Dispersión.....	8
1.4.3.1 Desviación estándar	9
1.4.3.2 Recorrido intercuartil.....	9
1.4.3.3 Coeficiente de variación.....	9
1.4.4 Asimetría.....	10
1.4.4.1 Coeficiente de asimetría relativa	10
1.4.4.2 Coeficiente de Pearson.....	10
1.4.5 Aplanamiento.....	10
1.4.6 Estimadores	11
1.5 DISTRIBUCIONES TEORICAS	11
1.5.1 Prueba de Bernoulli.....	12
1.5.2 Distribución binomial.....	12
1.5.3 Distribución de Poisson.....	13
1.5.4 Distribución normal.....	14
1.5.5 Distribución Chi Cuadrado	16
1.5.6 Distribución Gamma.....	17
1.6 PRUEBAS ESTADISTICAS	18
1.6.1 Prueba de Chi Cuadrado.....	19
1.6.2 Prueba de Kolmogorof - Smirnof.....	21
2. DISTRIBUCIONES EXTREMAS.....	23
2.1 PROBABILIDADES Y PERIODOS DE RETORNO	25
2.2 FACTORES DE FRECUENCIA.....	25
2.3 DISTRIBUCIONES TEORICAS	26
2.3.1 Función de Gumbel.....	26
2.3.2 Función de Fréchet.....	29
2.3.3 Función de Pearson tipo III.....	29
2.3.4 Función de Galton	30
2.4 POSICIONES GRAFICAS.....	31
2.5 AJUSTE DE CURVAS.....	32
2.5.1 Método de máxima verosimilitud.....	32
2.5.2 Método de los momentos.....	33
2.5.3 Regresión mínimo cuadrática	34
2.6 RIESGO	37
3. DISTRIBUCION DE GUMBEL.....	39
3.1 PARAMETROS ESTADISTICOS.....	39
3.2 DETERMINACION DE LA RECTA DE AJUSTE.....	40
3.2.1 Método de Gumbel.....	40
3.2.1.1 Ordenamiento de los datos.....	41
3.2.1.2 Asignación de probabilidades	41
3.2.1.3 Valores de la variable reducida.....	44
3.2.1.4 Graficado de los datos.....	44
3.2.1.5 Coeficientes de ajuste	46

3.2.2	<i>Método de Lieblein</i>	50
3.2.2.1	Selección de subgrupos	50
3.2.2.2	Cálculo de los estimadores	50
3.2.2.3	Cálculo de la varianza	52
3.3	INTERVALOS DE CONFIANZA	54
3.3.1	<i>Método de Gumbel</i>	55
3.3.2	<i>Método de Lieblein</i>	57
3.3.3	<i>Método de Kaczmarek</i>	58
3.3.4	<i>Método de Bernier-Veron</i>	59
3.3.5	<i>Comparación entre métodos</i>	61
3.4	PRUEBAS DE HIPOTESIS	62
4.	DISTRIBUCIONES LOGARITMICAS	66
4.1	DISTRIBUCION DE FRECHET	67
4.1.1	<i>Parámetros estadísticos</i>	68
4.1.2	<i>Recta de ajuste</i>	68
4.1.3	<i>Intervalos de confianza</i>	69
4.1.4	<i>Pruebas de hipótesis</i>	70
4.2	DISTRIBUCION DE GALTON	71
4.2.1	<i>Parámetros estadísticos</i>	71
4.2.2	<i>Recta de ajuste</i>	72
4.2.3	<i>Intervalos de confianza</i>	73
4.2.4	<i>Pruebas de hipótesis</i>	74
4.3	DISTRIBUCION LOG PEARSON	76
4.3.1	<i>Recta de ajuste</i>	76
4.3.2	<i>Intervalos de confianza</i>	78
4.3.3	<i>Pruebas de hipótesis</i>	79
5.	CONCLUSIONES SOBRE LAS FUNCIONES DE AJUSTE	81
6.	APLICACIONES	85
6.1	ANALISIS DE CRECIDAS	87
6.2	ANALISIS DE ESTIAJES	89
6.3	CONCLUSIONES ACERCA DEL ESTUDIO	92

1. CONCEPTOS BASICOS DE ESTADISTICA

En este capítulo se expondrán algunos conceptos básicos de estadística necesarios para comprender los temas más específicos desarrollados en capítulos posteriores. Se trata de un conjunto de definiciones y de una apretada síntesis de los fundamentos de la teoría de probabilidades presentados sólo como una guía general. Estos conceptos pueden ser ampliados en cualquier obra sobre estadística y probabilidad en general, por ejemplo en las referencias (1), (2) y (3).

La teoría de los valores extremos se aplica principalmente a fenómenos meteorológicos o hidrológicos tales como velocidades de vientos, caudales o cotas hidrométricas de cursos de agua y registros pluviométricos, aunque también puede utilizarse para estudiar problemas de políticas de "stock", longevidad o fatiga de materiales. (12)

Tomando como ejemplo el caudal máximo de un río, puede comprobarse fácilmente que su valor variará de un día a otro sin que el mismo pueda predecirse con exactitud, aunque sí sea posible establecer *a priori* un rango de valores, resultado del experimento "tomar la lectura del máximo caudal diario". Desde este punto de vista el análisis del fenómeno no sería distinto del correspondiente al experimento "tomar nota del resultado de arrojar un dado". Este tipo de experiencias en las cuales los resultados varían de una realización a otra se denominan *experimentos aleatorios*, mientras que las variables con ellos relacionadas son conocidas como *variables aleatorias*. Así, la velocidad del viento, el nivel de las precipitaciones pluviales o el resultado de extraer una carta de un mazo de naipes son variables aleatorias.

En lo que sigue de este trabajo se operará exclusivamente con experimentos y variables aleatorios.

1.1 PROBABILIDAD

Según la teoría clásica, para un experimento aleatorio que arroje, sobre un total de resultados c , un conjunto de valores favorables a , existirá una probabilidad de éxito igual al cociente:

$$p = \frac{a}{c}$$

Repitiendo un número elevado de veces el experimento en cuestión, todos los casos posibles tenderán a presentarse en una cantidad proporcional a su probabilidad. Aceptando esta característica, puede ejecutarse n veces el mismo experimento; si en ellas aparecen f casos favorables, la relación f/n será aproximadamente igual a la probabilidad de ocurrencia del suceso, coincidiendo con ella para n tendiendo a infinito, es decir:

$$\lim_{n \rightarrow \infty} \frac{f}{n} = \frac{a}{c} = p$$

El número de casos favorables f es la frecuencia del suceso y el cociente f/n es su razón frecuencial. La definición clásica tiene el inconveniente de tornar dificultoso el

determinar, por ejemplo, si los casos posibles son o no igualmente probables. Esto dió lugar a diversas controversias que fueron solucionándose a medida que la teoría de probabilidades fue profundizándose, en especial gracias a la obra de De Moivre y Bernoulli. Actualmente se tiende a la creación de modelos matemáticos que permitan explicar los fenómenos probabilísticos. De esta forma puede postularse la existencia de un número p que represente la idealización de la razón frecuencial f/n , así como una constante física puede ser tomada como la idealización de mediciones empíricas. En este caso dicho número será, por definición, la probabilidad matemática del suceso en estudio, siendo su frecuencia relativa o razón frecuencial, una medida experimental de la misma (3).

En forma independiente del tipo de definición a que se recurra, resulta obvio que el campo de variación de la frecuencia relativa está dado por la siguiente desigualdad:

$$0 \leq \frac{f}{n} \leq 1$$

Si f/n representa la razón frecuencial de la aparición de un suceso A en un cierto experimento aleatorio, al hacer n grande su valor se acercará asintóticamente al de la probabilidad de ocurrencia de A . Luego, esta probabilidad tomará valores en un intervalo comprendido entre cero y la unidad.

Si A es un "hecho cierto", se presentará en cada repetición del experimento aleatorio, siendo entonces $f = n$ y $p(A) = 1$. Por consiguiente la probabilidad de un suceso cierto es unitaria.

Si A , en cambio, es un "hecho imposible", no se presentará para valor alguno de n , resultando $f = p(A) = 0$. Por lo tanto, la probabilidad de un suceso imposible es nula.

$$0 \leq p(A) \leq 1$$

1.2 SUCESOS SIMULTANEOS

Si se tienen dos sucesos A y B relacionados con un cierto experimento aleatorio, puede resultar necesario conocer la probabilidad de ocurrencia de *ambos simultáneamente*, este hecho será expresado como $p(A \ B)$, leyéndose como "probabilidad de A y B" y siendo equivalente a la intersección de dos conjuntos.

Si la aparición de uno de los dos sucesos afecta o condiciona la aparición del otro, podrá hablarse de probabilidades condicionales, expresadas como $p(A / B)$ que se lee como "probabilidad de A condicionada a que se haya producido B ". Si los sucesos no se afectan entre sí se dice que son independientes, resultando lo siguiente:

$$p(A / B) = p(A) \qquad p(B / A) = p(B) \qquad (1 - 1)$$

Si los sucesos A y B son la extracción de dos naipes de un cierto palo, al tomarlos de mazos distintos serán sucesos independientes mientras que si se los extrae de un único

mazo, la probabilidad del segundo suceso se verá condicionada por el resultado del primero.

En otros casos, se desea evaluar la probabilidad de que *al menos* uno de los sucesos simultáneos se presente. Este hecho se denota como $p(A + B)$ y se lee "probabilidad de *A* o de *B*", tal como la operación unión de conjuntos. Como caso particular deben considerarse los sucesos mutuamente excluyentes donde la aparición de uno de ellos implica la imposibilidad del otro.

Si se extraen, nuevamente, dos naipes de un mazo y si *A* y *B* representan la extracción de una carta en especial (número y palo), ambos sucesos serán mutuamente excluyentes. Es decir $p(A B) = 0$.

1.2.1 Regla de multiplicación

La probabilidad de que dos sucesos *A* y *B* se produzcan simultáneamente es igual a la probabilidad del suceso *A* multiplicada por la probabilidad de ocurrencia del suceso *B* bajo el supuesto de que *A* se haya presentado previamente. O, asimismo, es igual a la probabilidad de *B* por la probabilidad de *A* condicionada a la ocurrencia de *B*.

$$p(AB) = p(A) p(B / A) = p(B) p(A / B) \quad (1 - 2)$$

Si los sucesos *A* y *B* son independientes, la expresión anterior se simplifica de la siguiente manera:

$$p(AB) = p(A) p(B) \quad (1 - 3)$$

La expresión (1-2) puede extenderse a *n* sucesos simultáneos.

$$p(A_1 A_2 \cdots A_n) = p(A_1) p(A_2 / A_1) p(A_3 / A_1 A_2) \cdots p(A_n / A_1 A_2 \cdots A_{n-1}) \quad (1 - 4)$$

Si todos ellos son independientes, resulta:

$$p(A_1 A_2 \cdots A_n) = p(A_1) p(A_2) \cdots p(A_n) \quad (1 - 5)$$

1.2.2 Regla de adición

La probabilidad de que se presente al menos uno de dos sucesos *A* o *B* está dada por la suma de las probabilidades de aparición de cada uno de ellos menos la probabilidad de presentación simultánea.

$$p(A + B) = p(A) + p(B) - p(AB) \quad (1 - 6)$$

Como caso particular se consideran los sucesos mutuamente excluyentes en los que el último término de la ecuación (1 - 6) es nulo.

$$p(A + B) = p(A) + p(B) \quad (1 - 7)$$

La extensión de (1 - 6) a *n* sucesos lleva a:

$$\begin{aligned}
p(A_1 + A_2 + \dots + A_n) &= p(A_1) + p(A_2) + \dots + p(A_n) - p(A_1A_2) - p(A_1A_3) - \dots \\
&+ p(A_{n-1}A_n) + p(A_1A_2A_3) + p(A_1A_2A_4) + \dots + p(A_{n-2}A_{n-1}A_n) \dots (-1)^{n-1} p(A_1A_2 \dots A_n)
\end{aligned}
\tag{1 - 8}$$

Para n sucesos mutuamente excluyentes la ecuación anterior se reduce a la siguiente:

$$p(A_1 + A_2 + \dots + A_n) = p(A_1) + p(A_2) + \dots + p(A_n) \tag{1 - 9}$$

1.3 FUNCIONES DE DISTRIBUCION

El establecimiento de modelos matemáticos implica el conocimiento de la distribución de probabilidades para todos los valores de la variable aleatoria x . En particular, es útil conocer la probabilidad de que esta variable tome un valor menor o igual a un cierto nivel de referencia X , situación que se expresa mediante la siguiente notación:

$$F(x) = p(x \leq X) \tag{1 - 10}$$

Siendo $F(x)$ la *función de distribución* de la variable aleatoria x , la cual, si es conocida para todo el intervalo de variación de x , describe completamente la distribución de probabilidad del fenómeno estudiado. Por otra parte, de acuerdo a lo dicho en el punto 1.1, la función de distribución cumple con las siguientes condiciones:

$$\begin{aligned}
\lim_{x \rightarrow +\infty} F(x) &= 1 & \lim_{x \rightarrow -\infty} F(x) &= 0
\end{aligned}$$

La primera representa la probabilidad de un hecho cierto y la segunda la de un hecho imposible.

Si la variable aleatoria vinculada a un determinado experimento sólo puede tomar ciertos valores, tal como números enteros, tanto la variable como la función de distribución serán del tipo discreto. La Figura 1.1(a) muestra un ejemplo que corresponde a una distribución binomial.

En cambio, si la variable aleatoria puede tomar cualquier valor dentro de su campo de variación, tanto ella como su función de distribución serán de tipo continuo. La Figura 1.1(b) representa la función de distribución para una variable que toma valores continuos entre a y b .

En el primer caso, la figura puede interpretarse como una distribución de “masa” ubicada en posiciones fijas del eje de abscisas, mientras que en el segundo caso dicha “masa” se distribuye uniformemente entre a y b .

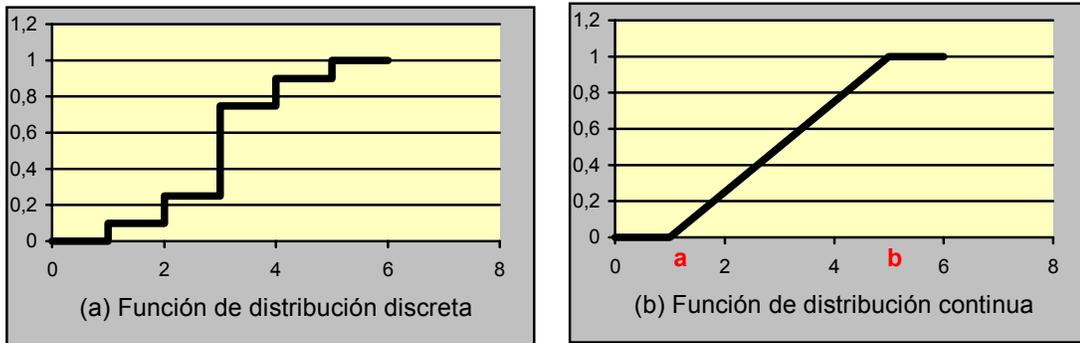


Figura 1.1: Funciones de distribución

En particular, la probabilidad de que la variable aleatoria tome valores comprendidos entre dos límites A y B , tal que definan un intervalo inferior al campo de variación de dicha variable, está dada por la diferencia entre los valores de su función de distribución en esos puntos.

$$p(A < x \leq B) = F(B) - F(A) \quad (1 - 11)$$

De esta forma la cantidad de “masa” comprendida en un intervalo infinitesimal $x, x + dx$, será también una medida de la probabilidad de que la variable aleatoria tome un valor dentro de ese intervalo. Por lo tanto, puede establecerse una función $f(x)$ que represente la densidad de distribución de la “masa” o, lo que lo mismo, la frecuencia relativa de aparición de valores de x en ese intervalo para experimentos repetidos. Así definida, esta función es la llamada densidad de probabilidad o *función de frecuencia* de la variable aleatoria y resulta ser la derivada de la función de distribución.

$$f(x) = F'(x) \quad (1 - 12)$$

Las Figuras 1.2 (a y b) ilustran las funciones de frecuencia correspondientes a las funciones de distribución representadas en la Figura 1.1. La masa completa de la distribución representa la probabilidad de que la variable aleatoria tome un valor cualquiera dentro de su campo de validez, o sea la probabilidad de un hecho cierto.

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (1 - 13)$$

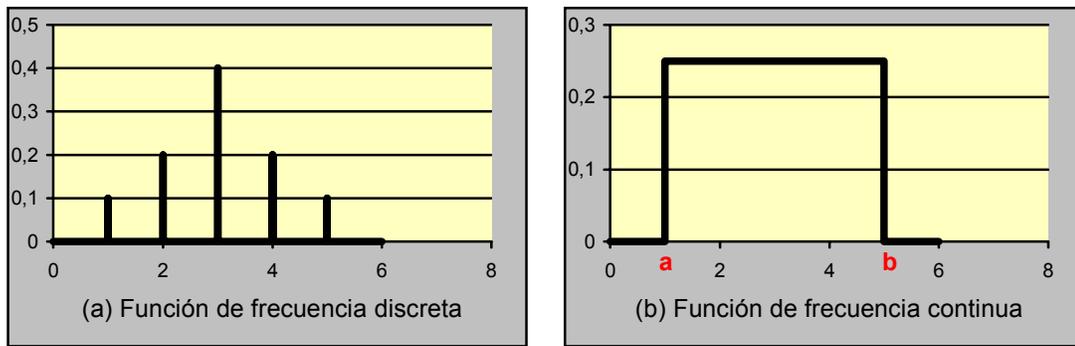


Figura 1.2: Funciones de frecuencia

1.4 PARAMETROS ESTADISTICOS

Si bien el conocimiento de las funciones de frecuencia o de distribución brinda una información completa acerca del experimento aleatorio estudiado, a veces resulta conveniente o necesario describir una distribución estadística por medio de unos pocos valores representativos. Estos reciben el nombre de *características* de la distribución o *parámetros estadísticos* y permiten mensurar ciertas particularidades.

1.4.1 Momentos

Así como se efectuó una analogía entre masa y probabilidad, también es posible aplicar el concepto físico de *momento* a una función probabilística. El momento k -ésimo de la variable aleatoria x , está definida de la siguiente forma:

$$\alpha_k = \int_{-\infty}^{\infty} x^k f(x) dx \quad (1 - 14)$$

Y en particular, para una distribución discreta:

$$\alpha_k = \sum_i x_i^k p(x_i) \quad (1 - 15)$$

Las ecuaciones (1 - 14) y (1 - 15) representan momentos con respecto al origen. Si en ellas se reemplaza la variable aleatoria x por la diferencia $x - c$, se obtendrán los momentos con respecto al punto c , siendo dichas expresiones casos particulares para $c = 0$. Resulta de importancia especial el caso en que c coincide con α_1 , es decir con el momento de primer orden dado que éste determina la abscisa del centro de gravedad de la distribución.

$$\alpha_1 = \int_{-\infty}^{\infty} x f(x) dx = m \quad (1 - 16)$$

Los momentos con respecto de m son llamados *momentos centrales*.

$$\mu_k = \int_{-\infty}^{\infty} (x - m)^k f(x) dx \quad (1 - 17)$$

O bien:

$$\mu_k = \sum_i (x - m)^k p(x_i) \quad (1 - 18)$$

1.4.1.1 Función generatriz

Para una función continua, se define como *función generatriz* de los momentos a la siguiente expresión:

$$\psi(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \quad (1 - 19)$$

Y para variable discreta:

$$\psi(t) = \sum e^{tx} p(x) \quad (1 - 20)$$

Para $t = 0$, tanto la integral como la sumatoria son siempre convergentes resultando según (1 - 13):

$$\psi(0) = 1$$

En cambio para $t \neq 0$, ni la integral ni la sumatoria serán necesariamente convergentes, pudiendo darse el caso de que lo sean para un cierto rango de esa variable.

Una interesante propiedad de la función generatriz es que, de existir el momento α_k , el mismo puede ser obtenido por derivación sucesiva de $\psi(t)$ haciendo $t = 0$ luego de derivar.

$$\alpha_k = \psi^{(k)}(0) \quad (1 - 21)$$

Además, en el caso antedicho (existencia del momento de orden k), el desarrollo de la función generatriz en serie de Mc Laurin será:

$$\psi(t) = \sum_{k=0}^{\infty} \frac{\alpha_k}{k!} t^k \quad (1 - 22)$$

1.4.2 Posición

El parámetro “posición” es un valor de abscisa que determina un punto central de la función alrededor del cual se distribuyen todos sus demás valores. Como ejemplo se definirán los parámetros de posición más frecuentemente usados: el valor medio, el modo y la mediana. En los puntos siguientes, las definiciones se referirán al caso

general de variables continuas; sin embargo, los conceptos son también aplicables a variables discretas, para cuales las integrales se convierten en sumatorias.

1.4.2.1 Valor medio

El *valor medio* está definido como la abscisa del centro de gravedad de la función de frecuencia y coincide con el momento de primer orden con respecto al origen. Representa el baricentro de la función.

$$m = \int_{-\infty}^{\infty} x f(x) dx = \alpha_1 \quad (1 - 23)$$

En el caso de observaciones indirectas en las que exista una relación lineal entre dos variables ($y = a + b x$), el valor medio también responde a una relación lineal que surge de la aplicación directa de la (1 - 23).

$$m(y) = a + b m(x) \quad (1 - 24)$$

Para observaciones indirectas en las que una variable es suma de otras dos ($z = x + y$), el valor medio también es igual a la suma de los valores medios individuales.

$$m(z) = m(x) + m(y) \quad (1 - 25)$$

1.4.2.2 Modo

El modo está definido como el valor más probable de la variable aleatoria coincidiendo, por lo tanto, con el máximo de la función de frecuencia. Para su cálculo debe resolverse la siguiente ecuación:

$$\text{Si } \frac{df(x)}{dx} = 0 \Rightarrow x = \text{MODO} \quad (1 - 26)$$

1.4.2.3 Mediana

La *mediana* es la abscisa que divide la “masa” total en dos porciones exactamente iguales. Puede determinarse fácilmente a partir de la función de distribución:

$$\text{MEDIANA} = F(0,5) \quad (1 - 27)$$

En distribuciones absolutamente simétricas, el valor medio, el modo y la mediana coinciden.

1.4.3 Dispersión

El parámetro dispersión indica la medida en que la variable aleatoria se distribuye alrededor del valor de posición. Da una idea del grado de repetibilidad de los resultados de un experimento aleatorio. Las principales medidas de dispersión son la desviación estándar y el recorrido intercuartil.

1.4.3.1 Desviación estándar

Cuando se utiliza el valor medio como medida de posición es lógico trabajar con la *desviación estándar* como parámetro de dispersión. Su valor coincide con la raíz cuadrada del momento central de segundo orden.

$$s = \sqrt{\mu_2} \quad (1-28)$$

Se recuerda la expresión que define al momento central de segundo orden:

$$\mu_2 = \int_{-\infty}^{\infty} (x - m)^2 f(x) dx \quad (1-29)$$

El cuadrado de la desviación estándar se denomina *varianza* de la distribución. En el caso de observaciones indirectas en las que exista una relación lineal entre dos variables ($y = a + b x$), la varianza puede hallarse aplicando la (1 - 29).

$$s^2(y) = b^2 s^2(x) \quad (1-30)$$

Para observaciones indirectas en las que una variable es suma de otras dos ($z = x + y$), la varianza también es igual a la suma de las varianzas individuales.

$$s^2(z) = s^2(x) + s^2(y) \quad (1-31)$$

Nótese que las propiedades no se aplican a la desviación estándar sino a la varianza.

1.4.3.2 Recorrido intercuartil

Así como la mediana representa la abscisa que divide la “masa” total en dos porciones iguales, los *cuartiles* (primero, segundo y tercero) la dividen en cuartos. El segundo cuartil coincide con la mediana.

$$\xi_1 = F(0,25) \quad ; \quad \xi_2 = F(0,5) \quad ; \quad \xi_3 = F(0,75)$$

Cuando se trabaja con la mediana como parámetro de posición, suele recurrirse como medida de dispersión al *recorrido intercuartil*. Este está definido como la diferencia entre los cuartiles tercero y primero.

$$\xi_{3-1} = F(0,75) - F(0,25) \quad (1-32)$$

1.4.3.3 Coeficiente de variación

Es también una medida de dispersión, aunque, a diferencia de las anteriores, está expresado en forma adimensional y no en unidades de la variable aleatoria.

El *coeficiente de variación* está definido como el cociente entre la desviación estándar y el valor medio y suele relacionárselo con el error cometido al realizar una serie de mediciones.

$$c_v = \frac{s}{m} \quad (1 - 33)$$

1.4.4 Asimetría

Permite evaluar la medida en que una función de frecuencia se aparta de la simetría perfecta, caso para el cual este parámetro es nulo. Un alto valor positivo de asimetría significa una larga cola para valores crecientes de la variable aleatoria, mientras que un valor negativo indica lo opuesto, es decir cola a la izquierda.

1.4.4.1 Coeficiente de asimetría relativa

Este parámetro se basa en el hecho de que todos los momentos centrales de orden impar son nulos para distribuciones simétricas. Aprovechando esta característica, toma el momento central de tercer orden dividido por el cubo de la desviación estándar para dar origen a un coeficiente adimensional.

$$\gamma_1 = \frac{\mu_3}{s^3} \quad (1 - 34)$$

1.4.4.2 Coeficiente de Pearson

En este caso se establece la diferencia entre valor medio y el modo de la distribución y se la divide por la desviación estándar para dar lugar también a un parámetro adimensional.

$$\gamma_p = \frac{m - MODO}{s} \quad (1 - 35)$$

1.4.5 Aplanamiento

Este parámetro refleja el grado en que una función de frecuencia resulta achatada en sus valores centrales. Se basa en el momento central de cuarto orden al que se le divide por la cuarta potencia de la desviación estándar a efectos de hacerlo adimensional.

Como se verá más adelante la función de frecuencia normal o de Gauss normalizada tiene valor medio, mediana, modo y coeficiente de asimetría nulos y desviación estándar unitaria. Si se aplica la definición anterior a esta función resulta un coeficiente de aplanamiento o *kurtosis* igual a tres. Con el objeto de llevarlo a cero, se le restan tres unidades al mencionado valor para así obtener el coeficiente de exceso.

$$\gamma_2 = \frac{\mu_4}{s^2} - 3 \quad (1 - 36)$$

Una función con $\gamma_2 < 0$ es llamada *platocúrtica*, mientras que una $\gamma_2 > 0$ con es denominada *leptocúrtica*.

1.4.6 Estimadores

Los parámetros ya estudiados que caracterizan una cierta población son fijos, es decir que **no son** variables aleatorias, pero son desconocidos. Pueden ser evaluados a partir de una muestra tomada de la población a través del cálculo de ciertos valores característicos los cuales brindarán sólo una estimación de los verdaderos parámetros. De esta forma puede afirmarse que el parámetro α^* , calculado a partir de los resultados de un experimento aleatorio, constituye una estimación del parámetro α de la población, fijo y desconocido.

Los *estimadores* son variables aleatorias que toman valores distintos para diferentes muestras. Sin embargo, si son elegidos en forma apropiada, representan adecuadamente a los verdaderos parámetros.

Se dice que un estimador es *insesgado* cuando cumple con la condición siguiente:

$$m(\alpha^*) = \alpha \quad (1 - 37)$$

Es decir, que el valor medio de un gran número de estimaciones del mismo parámetro, calculadas a partir de distintas muestras, se aproximará asintóticamente a ese parámetro cuando el número de estimaciones tienda a infinito. Resulta, en consecuencia, muy ventajoso trabajar con estimadores insesgados al evaluar las características de una distribución probabilística.

Los estimadores insesgados para el valor medio y la varianza son, respectivamente:

$$\bar{x} = \frac{1}{n} \sum_i x_i \quad \Rightarrow \quad m \quad (1 - 38)$$

$$\sigma^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 \quad \Rightarrow \quad s^2 \quad (1 - 39)$$

Estas ecuaciones se obtuvieron aplicando la propiedad (1 -24). En ellas puede comprobarse que la expresión correspondiente al estimador insesgado del valor medio es igual al promedio aritmético de las observaciones, mientras que en el caso de la varianza en el denominador aparece $n - 1$ en lugar de n . De haber empleado este último valor, la aplicación de la (1 - 37) no habría conducido a s^2 , sino al siguiente valor:

$$\frac{n-1}{n} s^2 \neq s^2$$

Es decir que se habría producido un sesgo en la estimación de la varianza.

1.5 DISTRIBUCIONES TEORICAS

Existe un gran número de distribuciones probabilísticas representativas de experimentos aleatorios que pueden ser descritas razonablemente con ayuda de expresiones matemáticas correspondientes a distribuciones teóricas. Esto tiene la gran ventaja de permitir la aplicación de ciertas propiedades de estas distribuciones a

los resultados experimentales. A continuación se describirán las más importantes funciones teóricas.

1.5.1 Prueba de Bernoulli

Las más importantes funciones teóricas se originan a partir de un sencillo experimento conocido como *prueba de Bernoulli*. Una serie de experiencias reciben esta denominación si, en cada ensayo, es posible obtener sólo dos resultados (favorable - desfavorable) y si las probabilidades respectivas se mantienen constantes a lo largo de toda la serie. Es común llamar p a la probabilidad de *éxito* y q a la de *fracaso*, resultando:

$$p + q = 1 \quad (1 - 40)$$

1.5.2 Distribución binomial

Frecuentemente es necesario conocer la probabilidad de obtener x éxitos en n pruebas de Bernoulli, sin importar el orden de los mismos. Puede tomarse como base el caso particular en el cual los x éxitos se obtienen en las primeras x experiencias y los $n - x$ fracasos aparecen posteriormente, en forma también consecutiva. Dado que las pruebas son independientes, es factible aplicar la regla de multiplicación simplificada (1 - 5) para el cálculo de la probabilidad P de esta secuencia.

$$P = p^x q^{n-x}$$

Dado que hay otras posibles series de resultados, deben considerarse todas las posibilidades para el cálculo correcto de la probabilidad buscada. Esta resultará igual a la suma de todos los casos individuales debido a que las secuencias son mutuamente excluyentes. En consecuencia, la función de frecuencia de la *distribución binomial* estará dada por la expresión anterior multiplicada por las combinaciones de n elementos tomados de x .

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad (1 - 41)$$

La función de distribución es igual a:

$$F(x) = \sum_{i=1}^x \binom{n}{i} p^i q^{n-i} \quad (1 - 42)$$

Se recuerda la expresión que permite calcular un número combinatorio:

$$\binom{n}{x} = \frac{n!}{x! (n-x)!}$$

Los principales parámetros estadísticos de la distribución binomial son:

$$m = n p \quad (1 - 43)$$

$$s = \sqrt{n p q} \quad (1 - 44)$$

$$\gamma_1 = \frac{q - p}{\sqrt{n p q}} \quad (1 - 45)$$

$$\gamma_2 = \frac{1 - 6 p q}{n p q} \quad (1 - 46)$$

1.5.3 Distribución de Poisson

Existen numerosas aplicaciones prácticas de la distribución binomial en las cuales la probabilidad del hecho favorable es muy baja, mientras que el número de repeticiones del experimento es muy elevado, siendo el producto de ambas magnitudes constante; es decir:

$$\begin{cases} p \rightarrow 0 \\ n \rightarrow \infty \\ n p \rightarrow \lambda \end{cases} \quad (1 - 47)$$

En situaciones como la descrita, la distribución binomial puede aproximarse por otra función, también de tipo discreto, llamada *distribución de Poisson*, cuya función de frecuencia se caracteriza por la expresión matemática siguiente:

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (1 - 48)$$

Siendo su función de distribución:

$$F(x) = \sum_{i=1}^x \frac{\lambda^i}{x!} e^{-\lambda} \quad (1 - 49)$$

Las condiciones (1 - 47) se cumplen principalmente en ciertos procesos distribuidos en el tiempo tales como llamadas telefónicas, desintegración radiactiva, arribo de vehículos a un puente y accidentes.

Los parámetros principales de la distribución de Poisson son:

$$m = \lambda \quad (1 - 50)$$

$$s^2 = \lambda \quad (1 - 51)$$

$$\gamma_1 = \frac{1}{\sqrt{\lambda}} \quad (1 - 52)$$

1.5.4 Distribución normal

Otra aproximación de la distribución binomial en la que se considera una serie grande de repeticiones del experimento aleatorio sin imponer restricciones especiales respecto del valor de probabilidad es la *función normal o de Gauss*, la que, a diferencia de las anteriores, es del tipo continuo.

Antes de definir la función de frecuencia normal es conveniente introducir el concepto de *variable normalizada*. Esta normalización consiste en aplicar a la variable aleatoria original la siguiente transformación lineal:

$$z = \frac{x - m}{s} \quad (1 - 53)$$

Las definiciones de valor medio y desviación estándar aplicadas a la (1 - 53) dan por resultado los valores siguientes:

$$m(z) = 0 \quad s(z) = 1 \quad (1 - 54)$$

Es decir que una variable normalizada tiene valor medio nulo y desviación estándar unitaria, en tanto que esos mismos parámetros para la variable original son m y s respectivamente.

La función de frecuencia normal o de Gauss está caracterizada por la siguiente expresión:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (1 - 55)$$

Su representación gráfica es una campana simétrica con respecto al origen y puntos de inflexión en ± 1 . El valor medio, el modo y la mediana coinciden, siendo todos ellos nulos. La función de distribución es la siguiente:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{u^2}{2}} du \quad (1 - 56)$$

Sus parámetros estadísticos, referidos a la variable normalizada z , son:

$$m = 0 \quad (1 - 57)$$

$$s = 1 \quad (1 - 58)$$

$$\gamma_1 = 0 \quad (1 - 59)$$

$$\gamma_2 = 0$$

(1 - 60)

El área encerrada por la curva de frecuencia normal en un intervalo simétrico con respecto al origen crece rápidamente hacia la unidad al incrementarse dicho intervalo. La Tabla 1.1 presenta algunos valores para los casos más significativos desde el punto de vista de ajuste de datos. La Tabla A1 del apéndice presenta una información completa respecto del área bajo la curva normal.

TABLA 1.1: AREA DELIMITADA POR LA FUNCION NORMAL	
Intervalo	Area
$\pm 1,0$	0,6826
$\pm 1,5$	0,8664
$\pm 2,0$	0,9546
$\pm 2,5$	0,9876
$\pm 3,0$	0,9974

La Figura 1.3 muestra las aproximaciones dadas por las funciones de Poisson y Gauss para un caso particular de la distribución binomial ($p = 0,2$; $q = 0,8$; $n = 20$).

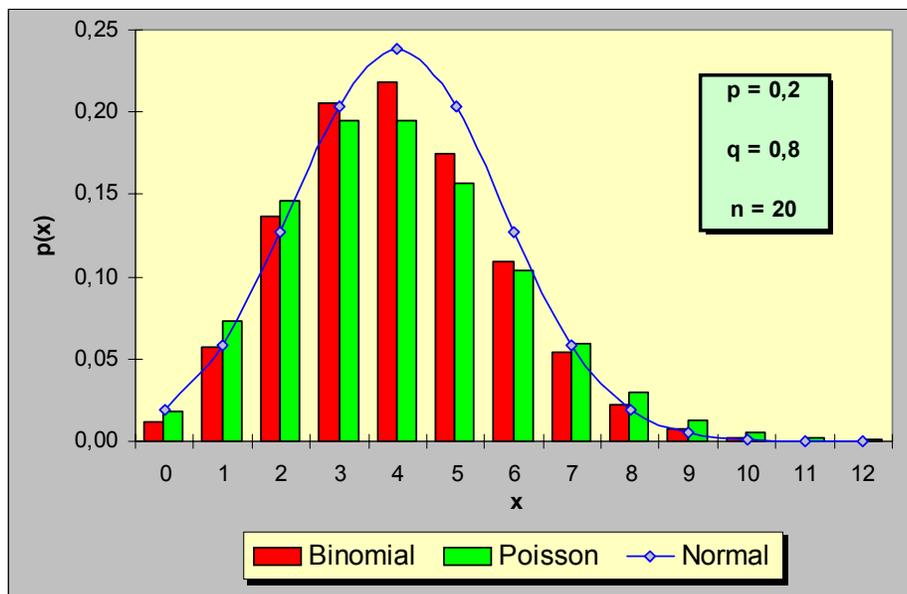


Figura 1.3: Distribución binomial y aproximaciones

1.5.5 Distribución Chi Cuadrado

La *distribución Chi Cuadrado* (χ^2) surge de considerar la suma de los cuadrados de n variables aleatorias independientes X_i , todas ellas normales con valor medio nulo y desviación estándar unitaria.

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2 \quad (1 - 61)$$

Haciendo $\chi^2 = x$, la función de frecuencia resulta ser la siguiente:

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad (1 - 62)$$

La función de frecuencia para la distribución χ^2 está definida sólo para valores positivos de x , mientras que para $x \leq 0$ es nula. El parámetro n representa el número de grados de libertad de la distribución. Para $n = 1$ y $n = 2$ la función decrece monótonamente con x , en tanto que para $n > 2$ es nula en el origen, alcanza un máximo para $x = n - 2$ y luego se acerca asintóticamente al eje de abscisas para x tendiendo a infinito. La función $\Gamma(p)$ es conocida como *función factorial* y está definida, para $p > 0$, por la integral siguiente:

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx \quad (1 - 63)$$

La función factorial tiende a infinito para p tendiendo a cero o a infinito y toma valores positivos para los restantes casos, con un mínimo para $p = 1,4616$ para el que la función vale $0,8856$. Para valores enteros de la variable independiente es válida la siguiente relación que justifica el nombre de la función:

$$\Gamma(p + 1) = p! \quad (1 - 64)$$

También cumple la siguiente propiedad:

$$\Gamma(0,5) = \sqrt{\pi}$$

La función Chi Cuadrado se caracteriza por los parámetros que se detallan a continuación:

$$m = n \quad (1 - 65)$$

$$s = \sqrt{2n} \quad (1 - 66)$$

La Figura 1.4 muestra la función chi cuadrado para varios grados de libertad.

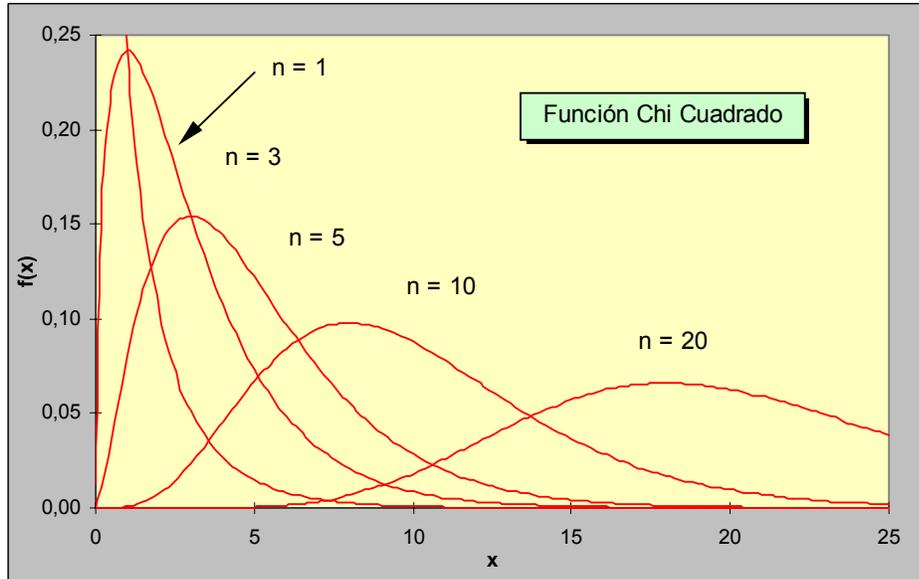


Figura 1.4: Distribución Chi Cuadrado para varios grados de libertad

1.5.6 Distribución Gamma

Está definida por la siguiente función densidad de probabilidad:

$$f(x) = \frac{e^{-x} x^{p-1}}{\Gamma(p)} \quad (1 - 67)$$

El denominador es la función factorial definida por la ecuación (1 - 63), siendo, en este caso, p el llamado factor de forma de la distribución. Este factor determina tres casos distintos, representados en la Figura 1.5, según sea $0 < p \leq 1$, $1 < p \leq 2$ o $p > 2$. Para grandes valores de p , la distribución gamma se aproxima a la normal. Sus parámetros son:

$$m = p \quad (1 - 68)$$

$$s = \sqrt{p} \quad (1 - 69)$$

Cuando se trabaja en hidrología, por lo general no se utiliza la expresión (1 - 67), que define la distribución gamma con un solo parámetro, sino que se prefiere la distribución con dos parámetros dada por la siguiente función de frecuencia:

$$f(x) = \frac{k^p}{\Gamma(p)} x^{p-1} e^{-kx} \quad (1 - 70)$$

Aquí p es un parámetro de forma y k de escala. El valor medio y la desviación estándar toman ahora los siguientes valores:

$$m = \frac{p}{k} \quad (1 - 71)$$

$$s = \frac{\sqrt{p}}{k} \quad (1 - 72)$$

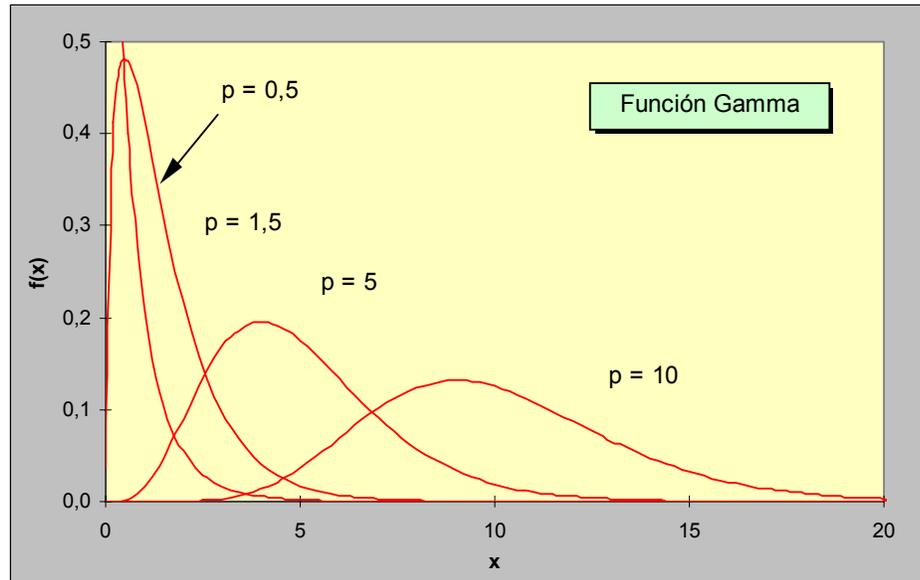


Figura 1.5: Distribución Gamma

1.6 PRUEBAS ESTADÍSTICAS

Una prueba estadística, o *dócima*, es un procedimiento cuyo objetivo es evaluar la bondad del ajuste resultante de aplicar una función teórica sobre un conjunto de datos experimentales. Estas pruebas, entre las que pueden citarse la de Pearson (Chi Cuadrado) y la de Kolmogorov-Smirnof, permiten disponer de elementos de juicio para aceptar o rechazar la función previamente seleccionada para aproximar los datos obtenidos prácticamente. Para ello se fija, un tanto arbitrariamente, *un nivel de significación* para la prueba, esto es el riesgo que se corre de rechazar la hipótesis propuesta cuando en realidad era verdadera. De aquí puede concluirse que la prueba estadística no es infalible sino que está afectada de una probabilidad de fallo. Su correcta aplicación debería disminuir el riesgo de fracaso a un mínimo.

Si se dispone de un conjunto de datos experimentales entre los que predominen, por azar, algunos pertenecientes a la cola de la distribución teórica con la que se intenta el ajuste, el valor arrojado por la prueba caerá en la zona de rechazo de la hipótesis, aunque esta sea verdadera; esto implica incurrir en un *error tipo I*, el cual está directamente relacionado con el nivel de significación de la prueba. Parecería lógico, entonces, llevar el nivel de significación a valores tan bajos como sea posible para disminuir al mínimo el riesgo de cometer un error de esta naturaleza. Sin embargo, esta estrategia conduce a incrementar la probabilidad de caer en un *error tipo II*, o sea aceptar la hipótesis propuesta para el ajuste cuando en realidad ésta es falsa. Como

solución de compromiso, se considera razonable trabajar con niveles de significación del 5% o 10%.

1.6.1 Prueba de Chi Cuadrado

Para aplicar la *prueba de Chi Cuadrado* sobre un conjunto de n datos experimentales, éstos deben agruparse en r intervalos mutuamente excluyentes con frecuencia individual f_i . Se supone que esta distribución será ajustada por una función teórica que otorga una probabilidad p_i al i -ésimo intervalo. Obviamente, se cumplirán las siguientes igualdades:

$$p_1 + p_2 + \dots + p_r = 1 \quad (1 - 73)$$

$$f_1 + f_2 + \dots + f_r = n$$

Si la hipótesis a verificar es cierta, f_i representa la frecuencia de un suceso con probabilidad p_i en una serie de n observaciones y, en consecuencia, tendrá una distribución binomial con media $n p_i$ siendo, además, asintóticamente normal. En este supuesto, las dos series de números (f_i y $n p_i$) coincidirán en la medida en que n sea lo suficientemente grande. De aquí que resulte conveniente emplear el siguiente parámetro como medida del apartamiento entre las frecuencias esperadas teóricamente y las observadas experimentalmente:

$$\chi^2 = \sum_r \frac{(f_i - n p_i)^2}{n p_i} = \sum_r z_i^2 \quad (1 - 74)$$

Siendo:

$$z_i^2 = \frac{f_i - n p_i}{\sqrt{n p_i}} \quad (1 - 75)$$

La variable z_i es asintóticamente normal con media nula, mientras que la sumatoria de sus cuadrados conduce, según lo visto en 1.5.5, a una distribución χ^2 con sólo $r - 1$ grados de libertad en virtud de la restricción lineal que liga las variables. Puede demostrarse que, si la hipótesis a verificar es cierta, la variable χ^2 definida por la (1 - 74) se aproxima, cuando n tiende a infinito, a una distribución Chi Cuadrado con $r - 1$ grados de libertad.

El procedimiento práctico para aplicar la dócima se inicia agrupando los datos en intervalos de clase tratando de que cada uno de ellos contenga al menos cinco valores. Luego se determinan la cantidad de datos que deberían haber pertenecido a cada intervalo, de ser válida la hipótesis propuesta. Finalmente se calcula el valor de variable χ^2 aplicando la (1 - 74) tal como se muestra en la Tabla 1.2.

TABLA 1.2: PRUEBA DE CHI CUADRADO			
Intervalo	Frecuencia experimental	Frecuencia teórica	Indicador de apartamiento
1	f_1	$n p_1$	z_1^2
2	f_2	$n p_2$	z_2^2
r	f_r	$n p_r$	z_r^2
Sumatorias	n	n	χ^2

Por otro lado, debe emplearse una tabla o un gráfico de la función chi cuadrado para hallar el valor de abscisa χ_p^2 que delimita a su derecha un área igual al nivel de significación asignado a la prueba, tal como $p = 5\%$. La curva a emplear debe ser la correspondiente a $r - k - 1$ grados de libertad, siendo k el número de parámetros de la función teórica calculados a partir de los datos experimentales. Esto implica que, si la función de ajuste en una distribución normal caracterizada por dos parámetros (valor medio y desviación estándar), ambos tendrán que ser calculados partiendo de los datos disponibles y estas operaciones quitarán dos grados de libertad adicionales a la distribución ($k = 2$). Si, en cambio, se ensaya un ajuste por Poisson, sólo habrá que determinar el parámetro λ resultando una reducción de un único grado de libertad ($k = 1$). La Tabla A2 del apéndice presenta los valores críticos de Chi cuadrado (χ_p^2) para diferentes valores de probabilidad y para diversos grados de libertad.

Se considerará que la hipótesis sometida a prueba es aceptable, al nivel de significación estipulado, siempre que se cumpla la siguiente desigualdad:

$$\chi^2 \leq \chi_p^2 \quad (1 - 76)$$

Un procedimiento alternativo que conduce a idéntico resultado es buscar en tablas la probabilidad $p(\chi^2)$ correspondiente al valor χ^2 calculado para $r - k - 1$ grados de libertad y compararlo con el nivel de significación de la prueba. La hipótesis puede aceptarse en el siguiente caso:

$$p(\chi^2) \geq p(\chi_p^2) = \frac{p}{100} \quad (1 - 77)$$

1.6.2 Prueba de Kolmogorof - Smirnof

En forma similar a lo requerido por la aplicación de la prueba del Chi Cuadrado sobre un conjunto de n datos experimentales, la prueba de Kolmogorov-Smirnof exige su agrupación en r intervalos mutuamente excluyentes con frecuencia individual f_i pero, a diferencia de aquella, no compara las frecuencias absolutas sino las relativas acumuladas, para lo cual toma como referencia los valores de la función de distribución teórica $F(x)$ propuesta para el ajuste. En este caso se cumplen las siguientes desigualdades:

$$F_1(x) \leq F_2(x) \leq \dots \leq F_r(x) < 1 \quad (1 - 78)$$

$$\frac{f_1}{n} \leq \frac{f_1 + f_2}{n} \leq \dots \leq \sum_{i=1}^{r-1} \frac{f_i}{n} < 1$$

El paso siguiente consiste en efectuar, para cada grupo, las diferencias, en valor absoluto, entre los valores esperados y los experimentales.

$$\Delta_k = \left| F_k(x) - \frac{1}{n} \sum_{i=1}^k f_i \right|$$

Por último, a partir de la máxima diferencia (Δ_{max}), se define la siguiente función.

$$\varphi_n(z) = P\left(\sqrt{n} \Delta_{max} \leq z\right) \quad (1 - 79)$$

Kolmogorof demostró que para n tendiendo a infinito, $\varphi_n(z)$ tiende a la función $k(z)$ que lleva su nombre.

$$\lim_{n \rightarrow \infty} \varphi_n(z) = k(z) = \sum_{-\infty}^{\infty} (-1)^k e^{-2k^2 z^2} \quad (1 - 80)$$

La prueba de Kolmogorof-Smirnof determina la aceptación de la función de ajuste propuesta toda vez que se cumpla la siguiente desigualdad:

$$\sqrt{n} \Delta_{max} \leq z \quad (1 - 81)$$

La Tabla 1.3 muestra la forma en que deben presentarse los datos para la aplicación de esta prueba estadística.

TABLA 1.3: PRUEBA DE KOLMOGOROF - SMIRNOF			
Intervalo	Frecuencia experimental	Frecuencia teórica	Indicador de apartamiento
1	f_1 / n	$F_1(x)$	Δ_1
2	$(f_1 + f_2) / n$	$F_2(x)$	Δ_2
r	$(f_1 + f_2 + \dots f_r) / n$	$F_r(x)$	Δ_r
	1	1	Δ_{max}

Antes de aplicar la (1 - 81) sobre el valor Δ_{max} calculado, se seleccionan, partir de tablas, los valores de la variable z de acuerdo con el nivel de significación asignado a la prueba. He aquí algunos ejemplos:

TABLA 1.4: FUNCION DE KOLMOGOROF	
Nivel (%)	z
10	1,22
5	1,36
1	1,63
0,5	1,95

2. DISTRIBUCIONES EXTREMAS

El enunciado del teorema central del límite dice que la función de distribución conjunta de n variables aleatorias responde a una función normal cuando n tiende a infinito. Esta propiedad es independiente de la función de distribución original de dichas variables.

Si, a partir de una población cualquiera, se toman n muestras de m elementos cada una, los valores medios individuales formarán parte, a su vez, de una nueva muestra aleatoria distribuida alrededor de la gran media, tal que:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Es decir que la gran media es una nueva variable aleatoria generada a partir de la suma de otras n variables aleatorias, siendo, por lo tanto, aplicable el teorema central del límite a esa distribución conjunta. O sea que los valores medios de n muestras de m elementos cada una se distribuyen normalmente alrededor de la gran media.

Sin embargo, si en lugar de considerar los valores medios se tienen en cuenta los máximos o mínimos de cada muestra, la distribución conjunta no será normal sino que responderá a otra función distinta. La teoría de las distribuciones extremas se ocupa de este tipo de problemas.

La probabilidad de que la variable aleatoria tome un valor inferior o igual a X es, por definición (1 - 10) la función de distribución:

$$P(x \leq X) = F(x)$$

Si se consideran n observaciones independientes, la probabilidad $\Phi(x)$ de que el máximo valor obtenido o, lo que es igual, que todos los n valores resulten inferiores o iguales a X puede hallarse recurriendo a la regla de multiplicación para sucesos independientes (1 - 5):

$$\Phi(x) = P(x_1 \leq X) P(x_2 \leq X) \cdots P(x_n \leq X)$$

O sea:

$$\Phi(x) = F^n(x) \quad (2 - 1)$$

La función de frecuencia de los máximos de n observaciones independientes está definida por la derivada de esta expresión.

$$\varphi(x) = n F^{n-1}(x) f(x)$$

Un razonamiento análogo permite determinar la probabilidad de que el menor valor de x sea menor que X , cuando se hacen n observaciones.

$$\Phi_1(x) = 1 - [1 - F(x)]^n$$

Su función de frecuencia es:

$$\varphi_1(x) = n [1 - F(x)]^{n-1} f(x)$$

Si la función de frecuencia inicial es simétrica, se tendrá la siguiente relación entre las funciones de frecuencia extremas para máximos y mínimos:

$$\varphi_1(-x) = \varphi(x)$$

De las expresiones planteadas anteriormente, surgen dos conclusiones inmediatas:

- Conociendo la distribución inicial de $F(x)$ es posible obtener la distribución correspondiente a valores extremos, $\Phi(x)$.
- La función $\Phi(x)$ depende del número de observaciones, n .

Ahora bien, dado que, por lo general, la expresión matemática de $F(x)$ no se conoce con exactitud, Fréchet en 1927 y Fisher y Tippett en 1928 estudiaron un conjunto de funciones asintóticas conocidas como *distribuciones de valores extremos*, las que, partiendo de ciertas funciones iniciales, brindan un ajuste adecuado para grandes valores de n , como se verá más adelante.

Si se extraen de una población, n muestras de m elementos cada una y luego se toman los n "más grandes valores", el mayor de todos éstos será también el más grande del conjunto de nm elementos considerados. Entonces, de acuerdo con Fisher y Tippett, la distribución estadística de los valores extremos en una muestra de tamaño nm debería de ser la misma que la correspondiente al mayor valor en una muestra de tamaño n excepto una transformación lineal en la variable aleatoria x (6) (8) (12).

$$F^m(x) = F(a_m x + b_m)$$

Para esta ecuación se han propuesto tres soluciones distintas que contemplan otras tantas familias de distribuciones.

Las funciones de distribución que convergen hacia la unidad al menos tan rápidamente como una exponencial pertenecen al *tipo I o exponencial*, del cual esta función es su prototipo. Esta familia de distribuciones iniciales está integrada por un conjunto de importantes funciones tales como la normal (Gauss), la log-normal (Galton) y la chi cuadrado (Pearson). Se caracterizan por tener definidos todos sus momentos aunque esto no significa que cualquier función que cumpla con esta condición pertenezca al tipo I.

Las distribuciones iniciales *tipo II o de Cauchy*, cuyo prototipo es esta misma función, no poseen momentos superiores a los de un cierto orden. Tal como en el caso anterior no todas las funciones que carezcan de momentos superiores están incluidas en el tipo II.

Al *tipo III o truncadas* pertenecen aquellas distribuciones iniciales con campo de variación de la variable limitado a un cierto valor.

Cada uno de los tres tipos de distribuciones iniciales conduce a diferentes distribuciones asintóticas, tal como se verá posteriormente.

2.1 PROBABILIDADES Y PERIODOS DE RETORNO

Cuando la variable aleatoria considerada es una magnitud relacionada con algún fenómeno natural (caudales, velocidades de viento), es conveniente referirse a *periodos de retorno* en lugar de a probabilidades de ocurrencia. Si p es la probabilidad de que una variable x supere un dado valor X en un cierto lapso (por lo general un año), el período de retorno T representará el número de unidades de tiempo que transcurrirán en promedio entre dos oportunidades en que la variable supere dicho valor, es decir:

$$p = P(x \geq X) = \frac{1}{T} \quad (2 - 2)$$

Por lo tanto, es equivalente especificar un período de retorno o recurrencia de 100 años o una probabilidad anual de 0,01.

El análisis estadístico consiste en hallar la función que mejor represente el comportamiento de la variable aleatoria x , para luego asignar a cada valor X una probabilidad o un período de recurrencia. Si $\Phi(x)$ es la función de distribución, resulta que, a partir de (1 - 10) y (2 - 2):

$$p = 1 - \Phi(x) \quad (2 - 3)$$

2.2 FACTORES DE FRECUENCIA

Para el caso especial de los fenómenos hidrológicos que responden a una distribución teórica de valores extremos (crecidas y estiajes) no existe una función que se adapte a todos los casos sino que cada uno debe ser analizado individualmente para aplicar luego la ley que mejor lo represente.

No obstante esto, Ven Te Chow (4) demostró que una variable aleatoria hidrológica x , puede ser representada por una combinación lineal de su valor medio y su desviación estándar de la siguiente manera:

$$x = \bar{x} + k \sigma_x \quad (2 - 4)$$

Para ello se basó en que cada valor de x puede expresarse como la media aritmética más un desplazamiento Δx proporcional a la desviación estándar.

$$x = \bar{x} + \Delta x$$

La expresión (2 - 4) es conocida como “expresión general para el análisis hidrológico de frecuencias”. Resulta evidente que ahora el problema consiste en determinar la función que mejor represente al factor de frecuencia k para cada caso. En general éste depende del período de retorno T , existiendo tablas y gráficos que dan la relación entre ambos para las distribuciones de uso más extendido. A este efecto pueden consultarse las referencias (5), (6) y (10).

2.3 DISTRIBUCIONES TEORICAS

A continuación se describirán las funciones de distribución más empleadas en hidrología detallando sus características principales y sus parámetros estadísticos. La Figura 2.1 ilustra algunas de ellas comparándolas con la función de frecuencia normal.

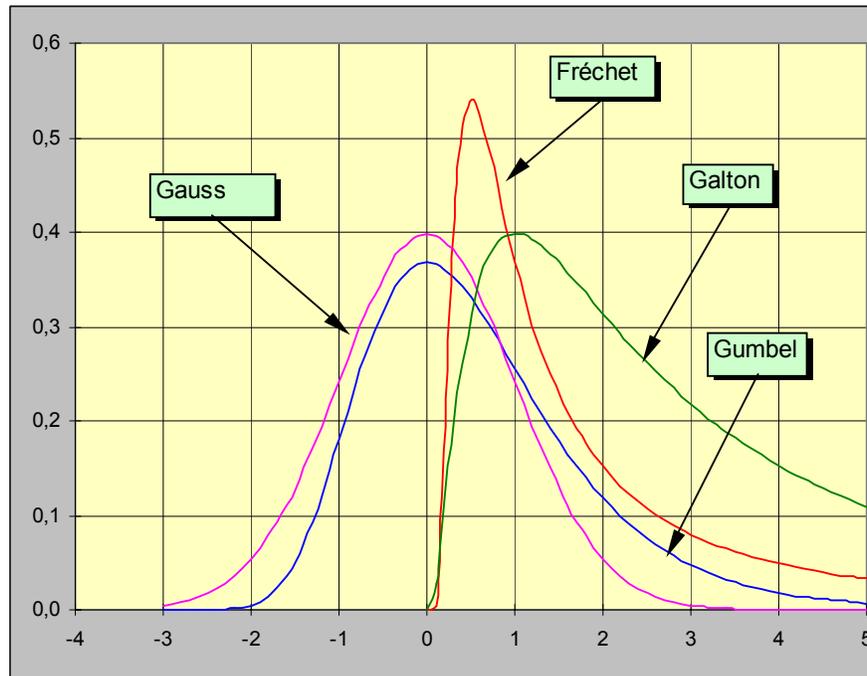


Figura 2.1: Funciones de frecuencia para variables extremas

2.3.1 Función de Gumbel

Si la función de distribución inicial converge hacia una exponencial para x tendiendo a infinito, es aplicable la ley de valores extremos tipo I (Gumbel) cuya expresión matemática es la siguiente:

$$\Phi(y) = e^{-e^{-y}} \quad (2 - 5)$$

Siendo y la variable reducida de Gumbel que es, a su vez, función lineal de la variable aleatoria original x .

$$y = \alpha_0 (x - u_0) \quad (2 - 6)$$

El campo de variación de x se extiende entre $-\infty$ y $+\infty$. Las constantes α_0 y u_0 se determinan a partir de los datos para lograr su óptimo ajuste. El valor medio y la desviación estándar de la variable reducida son fijos e independientes de la muestra.

$$\bar{y} = \gamma \quad (2 - 7)$$

$$\sigma_y = \frac{\pi}{\sqrt{6}} \quad (2 - 8)$$

Siendo γ la constante de Euler, definida por la expresión siguiente:

$$\gamma = \lim_{n \rightarrow \infty} \left(\sum_{i=1}^n \frac{1}{i} - \ln n \right) = 0,577\dots$$

Teniendo en cuenta la relación lineal que existe entre las variables x e y pueden calcularse fácilmente el valor medio y la desviación estándar para la variable aleatoria original. También es sencillo comprobar la validez de la siguiente igualdad:

$$k = \frac{x - \bar{x}}{\sigma_x} = \frac{y - \bar{y}}{\sigma_y}$$

Esto implica que, despejando y de la (2 - 5), puede hallarse la relación $k-T$ para una distribución de Gumbel. Si se tiene en cuenta, además, la vinculación existente entre la función de distribución y el período de retorno dada por las expresiones (2 - 2) y (2 - 3) se llega a la siguiente conclusión:

$$k = -\frac{\sqrt{6}}{\pi} \left[\gamma + \ln \left(\ln \frac{T}{T-1} \right) \right] \quad (2 - 9)$$

Otro aspecto interesante a considerar es la tendencia asintótica de la función de Gumbel cuando el período de retorno tiende a infinito. Este punto reviste particular importancia debido a que el objetivo principal del análisis estadístico es precisamente predecir el comportamiento de la variable bajo estudio (caudal, velocidad del viento, nivel de precipitaciones u otras) para grandes períodos de retorno. A partir de las expresiones (2 - 2), (2 - 3) y (2 - 5) se llega fácilmente a la siguiente igualdad:

$$y = \ln \frac{1}{\ln \frac{T}{T-1}} \quad (2 - 10)$$

Por otra parte, desarrollando en serie la función $e^{-\frac{1}{T}}$ resulta que, para T tendiendo a infinito, se llega a la siguiente aproximación:

$$e^{-\frac{1}{T}} \approx 1 - \frac{1}{T}$$

O, lo que es igual:

$$\frac{1}{T} \approx \ln \frac{T}{T-1} \quad (2 - 11)$$

Entonces, reemplazando (2 - 11) en (2 - 10), se obtiene la siguiente expresión válida para grandes períodos de recurrencia:

$$y \approx \ln T$$

Finalmente, la expresión completa toma la forma siguiente:

$$u = u_0 + \frac{1}{\alpha_0} \ln T \quad (2 - 12)$$

Es decir que el valor predicho por Gumbel para la variable de interés crece, aproximadamente, con el logaritmo del período de retorno. Para $T = 10$ el error cometido es del orden del 2%, en tanto que para $T = 100$ alcanza apenas el 0,1%.

Por último, las ecuaciones (2 - 2) y (2 - 10) permiten completar la Tabla 2.1 que relaciona probabilidades, períodos de retorno y valores de la variable reducida.

TABLA 2.1: FUNCION DE GUMBEL VARIABLE REDUCIDA Y PERIODO DE RETORNO		
Probabilidad (p)	Período de retorno (T)	Variable reducida (y)
0,500	2	0,367
0,200	5	1,500
0,100	10	2,250
0,050	20	2,970
0,020	50	3,902
0,010	100	4,600
0,005	200	5,296
0,002	500	6,214
0,001	1000	6,907

2.3.2 Función de Fréchet

Cuando la distribución inicial responde a una función de Cauchy, los valores extremos se ajustan por una función tipo II, de Fréchet. Su expresión matemática es similar a la de Gumbel, pero la variable reducida está vinculada con la variable aleatoria original en forma logarítmica.

$$\Phi(y) = e^{-e^{-y}} \quad (2 - 13)$$

$$y = \alpha_0 (\ln x - u_0) \quad (2 - 14)$$

El campo de variación se extiende entre 0 y $+\infty$. Presentada de esta forma, la ley de Fréchet resulta de fácil aplicación ya que pueden emplearse los métodos desarrollados para Gumbel utilizando para los cálculos el logaritmo de los datos (8) (9).

Reemplazando (2 - 14) en (2 -13) y agrupando constantes, se obtiene una expresión alternativa para la función tipo II.

$$\Phi(x) = e^{-(\Theta x)^{-\alpha_0}} \quad (2 - 15)$$

Partiendo de esta expresión, puede estudiarse la tendencia para grandes valores de período de retorno.

$$(\Theta x)^{-\alpha_0} = \ln \frac{T}{T-1}$$

Para T tendiendo a infinito resulta, si se tiene en cuenta la (2 - 11), la siguiente aproximación:

$$x \approx \frac{1}{\Theta} T^{1/\alpha_0} \quad (2 - 16)$$

Como puede comprobarse fácilmente, el empleo de la función de Fréchet produce un ajuste con un crecimiento mucho más rápido de la variable aleatoria en función del período de retorno que el correspondiente a un ajuste por Gumbel. Sin embargo, puede demostrarse que cuando α_0 tiende a infinito, la función de Fréchet converge hacia la de Gumbel (8).

2.3.3 Función de Pearson tipo III

Esta distribución, que también se aplica al logaritmo de los datos, tiene la siguiente expresión:

$$p(x) = p_0 \left(1 - \frac{x}{a}\right)^c e^{-cx/2} dx \quad (2 - 17)$$

El campo de variación está comprendido entre $-\infty$ y ε con $\varepsilon < \infty$. Los coeficiente p_0 , a y c son muy complicados de determinar dado que dependen de los momentos de

segundo y tercer orden de la distribución inicial y de la función Gamma del parámetro c , consecuentemente resulta más práctico el empleo de la fórmula de Chow (2 - 4) transformada logarítmicamente (5).

$$\log x = \bar{x}_{\log} + k \sigma_{\log}$$

Los coeficientes representan el valor medio y la desviación estándar del logaritmo de los datos. El factor de frecuencia k es función del período de retorno y del coeficiente de asimetría de los datos y puede obtenerse de tablas (5). Para los casos en que el coeficiente de asimetría es nulo, el ajuste por log-Pearson coincide con el dado por log-normal, razón por la cual esta distribución goza de cierta popularidad a pesar de sustentarse en escasas bases teóricas (5).

Para grandes valores de T , la variable aleatoria crece con el logaritmo del período de retorno, tal como ocurre con la distribución de Gumbel (8).

2.3.4 Función de Galton

Es una modificación de la función de normal o de Gauss en la cual la relación entre la variable reducida y la variable aleatoria original es logarítmica y, por ello es llamada log-normal. Fue estudiada originalmente por Galton en 1875 y, si bien no fue concebida como una ley para valores extremos, su marcada asimetría positiva permite lograr, en ciertos casos, un excelente ajuste de máximos o mínimos (6) (10). Su función de frecuencia es la siguiente:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{1}{2}\left(\frac{z-\bar{z}}{\sigma_z}\right)^2} \quad (2 - 18)$$

Donde tanto la variable reducida z como su valor medio y su desviación estándar deben evaluarse a partir de los logaritmos de la variable original.

$$z = \ln x$$

$$\bar{z} = \frac{1}{n} \sum_i \ln x_i \quad (2 - 19)$$

$$\sigma_z^2 = \frac{1}{n-1} \sum_i (\ln x_i - \bar{z})^2$$

Para grandes períodos de retorno, puede demostrarse que el comportamiento asintótico de la función de Galton es el siguiente:

$$x \approx C e^{\sqrt{\ln T}} \quad (2 - 20)$$

Siendo C , una constante arbitraria.

2.4 POSICIONES GRAFICAS

El análisis estadístico en hidrología tiene por objeto asignar a cada cota hidrométrica o caudal una probabilidad de ser alcanzado o superado o, lo que es lo mismo, atribuirle un período de retorno. Dichas probabilidades surgen del ajuste de los registros históricos, base de todo análisis, por medio de una función de distribución conocida. Por ello resulta necesario ubicar cada dato histórico en un gráfico probabilidad-caudal antes de iniciar el ajuste correspondiente.

Fueron propuestos varios métodos que otorgan un valor de “*probabilidad empírica*” a cada uno de los datos. En todos los casos la muestra (conjunto de datos históricos) debe ordenarse en forma decreciente sin importar la fecha de ocurrencia de cada suceso, asignando a cada uno de los n elementos un número de orden i , tal que $1 \leq i \leq n$.

TABLA 2.2: PROBABILIDADES EMPIRICAS		
Nombre	Fecha	Expresión
California	1923	$\frac{i}{n}$
Hazen	1930	$\frac{2i - 1}{2n}$
Weibull	1939	$\frac{i}{n + 1}$
Beard	1943	$1 - \left(\frac{1}{2}\right)^{\frac{i}{n}}$
Gringorten	1963	$\frac{1 - 0,44}{n + 0,12}$

En el Tabla 2.2, tomada de la referencia 4, se resumen las expresiones propuestas para otorgar una probabilidad a cada dato. Las dos primeras fueron usadas en los albores del análisis estadístico pero posteriormente se extendió el empleo de la fórmula de Weibull que asigna al mayor de los n datos históricos un período de retorno de $(n + 1)$ años y al más pequeño un período de $(n + 1) / n$ años, es decir aproximadamente unitario si el registro es lo suficientemente extenso.

Un estudio comparativo llevado a cabo por Benson entre la ecuaciones de Hazen, Beard y Weibull ha demostrado que esta última es la que brinda resultados más acordes con la experiencia práctica. Sin embargo, Gringorten afirma que su expresión es la que permite un mejor examen visual de los valores extremos cuando éstos son graficados en papel probabilístico.

De todas maneras, la totalidad los métodos propuestos para determinar las posiciones gráficas dan valores similares en la zona media de la distribución notándose, en cambio, las principales discrepancias en las “colas”.

Finalmente, es conveniente aclarar que, dado que las probabilidades así calculadas sólo dependen del número de orden en una tabla, a datos iguales corresponden posiciones gráficas ligeramente distintas.

2.5 AJUSTE DE CURVAS

Una vez en posesión del conjunto de datos históricos y habiendo asignado a cada uno de ellos una posición gráfica, debe seleccionarse la función de distribución más apropiada y luego determinar los valores de los coeficientes de ajuste de la misma. Obviamente, cuanto mayor sea la coincidencia entre la función teórica y los datos experimentales será factible efectuar extrapolaciones más confiables.

Si no se tiene una idea “a priori” de la función a emplear, lo mejor puede ser graficar los datos en papeles probabilísticos que respondan a distintas funciones y trazar una línea de ajuste “a ojo” seleccionando aquella que parezca brindar los mejores resultados. Esta tarea resulta sencilla dado que es posible obtener papeles con escalas de abscisas y ordenadas adaptados a distintas funciones de distribución en los cuales ésta aparezca como una recta.

De todas maneras, aunque se conozca de antemano la función teórica que va a ser utilizada, es una buena práctica graficar los datos y trazar en forma manual la recta de ajuste antes de iniciar cálculos numéricos. Si bien ésta adolecerá de los lógicos errores de este método, permitirá tener una idea bastante aproximada del resultado final.

Para hallar los coeficientes de ajuste de la función de distribución elegida, puede emplearse cualquiera de los métodos provistos por la estadística clásica: máxima verosimilitud, momentos o regresión mínimo cuadrática.

2.5.1 Método de máxima verosimilitud

Es el método más importante desde el punto de vista teórico; su desarrollo se debe a R.A. Fisher en el año 1912. Se basa en los conceptos siguientes.

Si se extrae una muestra de tamaño n de una población caracterizada por una función de frecuencia $f(x)$ con un único parámetro desconocido, α , la probabilidad de obtener un conjunto de valores x_1, x_2, \dots, x_n puede expresarse en función de α por medio de la siguiente expresión en la que se supuso que cada valor x_i se obtuvo en forma independiente de los demás:

$$L(x_1, x_2, \dots, x_n, \alpha) = f(x_1, \alpha) f(x_2, \alpha) \dots f(x_n, \alpha) \quad (2 - 21)$$

Esta expresión es conocida como función de máxima verosimilitud, consistiendo el método del mismo nombre en hallar el valor $\hat{\alpha}$ del parámetro desconocido α que maximice dicha función, es decir que sea raíz de la siguiente ecuación:

$$\frac{\partial L(x_1, x_2, \dots, x_n, \alpha)}{\partial \alpha} = 0 \quad (2 - 22)$$

Teniendo en cuenta la existencia de numerosas funciones de frecuencia exponenciales y, dado que el logaritmo de una función se hace máximo cuando dicha función pasa por su máximo, suele usarse la siguiente expresión alternativa:

$$\frac{\partial \ln [L(x_1, x_2, \dots, x_n, \alpha)]}{\partial \alpha} = 0 \quad (2 - 23)$$

Si la función de frecuencia tiene un segundo parámetro característico, β , la función de verosimilitud dependerá también del mismo, siendo entonces necesario recurrir a un sistema de ecuaciones para hallar su máximo absoluto.

$$\begin{cases} \frac{\partial \ln [L(x_1, x_2, \dots, x_n, \alpha, \beta)]}{\partial \alpha} = 0 \\ \frac{\partial \ln [L(x_1, x_2, \dots, x_n, \alpha, \beta)]}{\partial \beta} = 0 \end{cases} \quad (2 - 24)$$

En la práctica el método de máxima verosimilitud consiste en determinar los estimadores que hagan máxima la probabilidad de extraer la muestra considerada. Puede demostrarse (1) (3) que estos estimadores son los más eficientes (es decir que su varianza es mínima) y consistentes (la varianza tiende a cero cuando el tamaño de la muestra tiende a infinito).

Sin embargo a pesar de sus ventajas teóricas este método no es aplicado normalmente en la práctica debido a su complejidad.

2.5.2 Método de los momentos

Fue desarrollado por K. Pearson y basa en igualar los momentos muestrales con los correspondientes a la distribución. Se toman los momentos con respecto al origen y en número igual a la cantidad de parámetros a determinar originándose así un sistema de ecuaciones cuya resolución permite obtener los estimadores buscados. Por ejemplo, para determinar dos parámetros de ajuste, se tomarán los momentos de primer y segundo orden y se los igualará a los valores experimentales:

$$\begin{cases} \alpha_1(\alpha, \beta) = \frac{1}{n} \sum_{i=1}^n x_i \\ \alpha_2(\alpha, \beta) = \frac{1}{n} \sum_{i=1}^n x_i^2 \end{cases} \quad (2 - 25)$$

Este método conduce, por lo general a cálculos más sencillos que el método anterior pero los estimadores obtenidos son menos eficientes (3) en especial teniendo en cuenta que los momentos de orden superior tienden a magnificar los errores en los datos ubicados en las colas de las funciones de distribución .

El método de Lieblen que se expondrá en el próximo capítulo es una variante del método de los momentos.

2.5.3 Regresión mínimo cuadrática

El objetivo de este método es determinar la recta que mejor ajusta una nube de puntos (x_i, y_i) . Para ello considera que cada valor de ordenada puede representarse como la suma del valor esperado (dado por la recta de regresión) más un término de error que mide la diferencia entre la ordenada experimental y la teórica.

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (2 - 26)$$

Esta premisa implica que todo el error recae en la variable dependiente y considera nulo el error en abscisas.

La regresión mínimo cuadrática se basa en determinar los estimadores que minimicen la sumatoria de los cuadrados de los términos de error.

$$\sum \varepsilon_i^2 = \sum (y_i - \alpha - \beta x_i)^2 \Rightarrow MINIMA$$

La solución consiste en igualar a cero las derivadas parciales con respecto de α y de β y hallar las raíces. Se llega a las soluciones siguientes:

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \quad (2 - 27)$$

$$\hat{\beta} = \frac{\sigma_{xy}}{\sigma_x^2} \quad (2 - 28)$$

Siendo:

$$\bar{y} = \frac{1}{n} \sum y_i \quad (\text{valor medio de } y)$$

$$\bar{x} = \frac{1}{n} \sum x_i \quad (\text{valor medio de } x)$$

$$\sigma_x^2 = \frac{1}{n} \sum x_i^2 - \bar{x}^2 \quad (\text{varianza de } x)$$

$$\sigma_{xy} = \frac{1}{n} \sum x_i y_i - \bar{x} \bar{y} \quad (\text{covarianza } x - y)$$

Estos resultados conducen a la obtención de la recta de regresión de y sobre x .

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

También podría haberse partido de un razonamiento inverso, considerando que la totalidad del error está localizado en la variable independiente. Esto conduce a la obtención de la recta de regresión de x sobre y .

$$y = \hat{\alpha}^* + \beta^* x$$

Los coeficientes de regresión son los siguientes:

$$\hat{\alpha}^* = \bar{x} - \beta^* \bar{y} \quad (2 - 29)$$

$$\beta^* = \frac{\sigma_{xy}}{\sigma_y^2} \quad (2 - 30)$$

Siendo:

$$\sigma_y^2 = \frac{1}{n} \sum y_i^2 - \bar{y}^2 \quad (\text{varianza de } y)$$

La recta obtenida por aplicación de este método brinda, por lo general, un mejor ajuste que el conseguido mediante el método de los momentos (4).

En numerosos casos prácticos, no puede asegurarse que el error en una de las variables sea nulo o despreciable frente al correspondiente a la otra sino que ambos son del mismo orden. Para esta situación existe una alternativa consistente en determinar la recta de regresión ortogonal que minimiza tanto las desviaciones horizontales como las verticales.

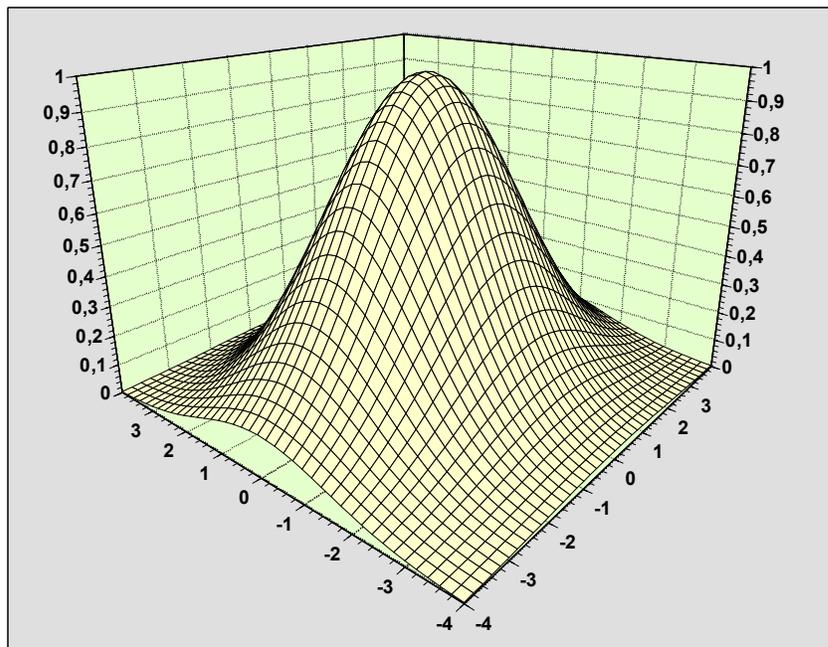


Figura 2.2: Distribución conjunta de errores

Para determinar la pendiente y la ordenada al origen de esta nueva recta, debe partirse de la base de que los errores en ambas variables siguen una distribución

normal. La distribución conjunta puede representarse, entonces, en un tercer eje dando lugar a una figura tridimensional con un máximo en el punto (\bar{x}, \bar{y}) como puede observarse en la Figura 2.2.

Si se efectúan cortes paralelos al plano x - y y se proyectan las figuras resultantes se generarán una serie de elipses concéntricas de isoprobabilidad cuyo eje de simetría será precisamente la recta de regresión ortogonal dada por la siguiente expresión:

$$y = \hat{\alpha}^{**} + \hat{\beta}^{**} x$$

Para hallar las expresiones de los estimadores debe seguirse una secuencia de cálculo algo compleja, arribándose a las siguientes expresiones:

$$\hat{\alpha}^{**} = \bar{y} - \hat{\beta}^{**} \bar{x} \quad (2 - 31)$$

$$\hat{\beta}^{**} = \frac{\sigma_y}{\sigma_x} \quad (2 - 32)$$

Manteniendo el supuesto de que los errores en los valores de las variables siguen una distribución normal, puede demostrarse (1) que los estimadores producidos por el método de regresión mínimo cuadrática son máximo verosímiles, es decir que tienen las propiedades de eficiencia y consistencia aseguradas.

La Figura 2.3 presenta como ejemplo un conjunto de puntos arbitrario ajustados por cada una de las tres rectas mínimo cuadráticas.

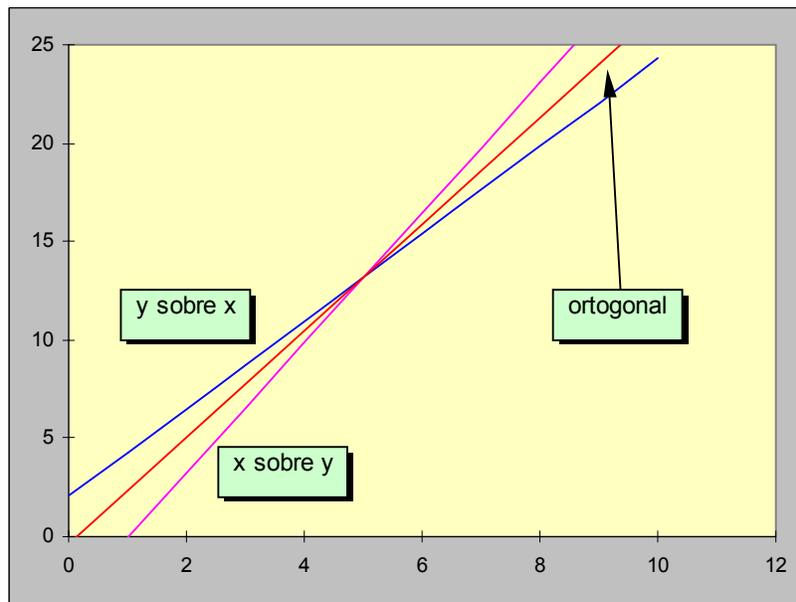


Figura 2.3: Regresiones mínimo cuadráticas

2.6 RIESGO

No debe olvidarse que, cualquiera sea la función o procedimiento de ajuste utilizado, el resultado obtenido será una relación entre la variable aleatoria y el período de retorno y que esa variable aleatoria ya sea un caudal, una cota hidrométrica, un nivel de precipitación o una velocidad de viento será empleada para un cálculo de ingeniería.

Dicho cálculo se efectuará tomando como base un cierto período de recurrencia que asegure que la probabilidad de que la variable aleatoria supere un valor de referencia sea muy baja, tal como 1% anual (período milenario). Sin embargo, aunque se trabaje con elevados períodos, la probabilidad de que la variable aleatoria supere el valor de diseño no será nula, aunque será pequeña. Esa probabilidad es denominada riesgo y su valor debe ser calculado antes de proseguir con los restantes cálculos ingenieriles.

De acuerdo con las expresiones (2 - 3) y (2 - 4), la probabilidad de que en un año genérico cualquiera **no se supere** la cota X de proyecto relativa a un período de retorno T es:

$$p(x \leq X) = 1 - \frac{1}{T} \quad (2 - 31)$$

Si N es la vida útil del emprendimiento u obra para la cual se ha efectuado el cálculo probabilístico, cada año de su vida puede ser considerado como un suceso independiente. Luego, puede aplicarse la regla de la multiplicación para determinar la probabilidad de que **en ninguno** de esos N años se supere la cota de diseño.

$$p_{TOT}(x \leq X) = \left(1 - \frac{1}{T}\right)^N \quad (2 - 32)$$

Finalmente, la probabilidad de **al menos una vez** sea superado el valor de diseño a lo largo de toda la vida útil del proyecto será una medida del *riesgo* que implica trabajar con el período de retorno utilizado para los cálculos.

$$Riesgo = 1 - \left(1 - \frac{1}{T}\right)^N \quad (2 - 33)$$

La Tabla 2.3 muestra los valores de riesgo para varias combinaciones de período de retorno de diseño y vida útil del proyecto.

TABLA 2.3: RIESGO (%)					
T \ N	10	20	30	40	50
2	99,9	100,0	100,0	100,0	100,0
5	89,3	98,8	99,9	100,0	100,0
10	65,1	87,8	95,8	98,5	99,5
20	40,1	64,2	78,5	87,1	92,3
50	18,3	33,2	45,5	55,4	63,6
100	9,6	18,2	26,0	33,1	39,5
200	4,9	9,5	14,0	18,2	22,2
500	2,0	3,9	5,8	7,7	9,5
1000	1,0	2,0	3,0	3,9	4,9

3. DISTRIBUCION DE GUMBEL

En este capítulo se expondrán algunas características particulares de la función de distribución tipo I o de Gumbel dado su importancia en el campo de la hidrología estadística. Según lo visto anteriormente, esta ley es la más apta para ajustar los valores extremos de aquellas distribuciones iniciales caracterizadas por colas factibles de ser aproximadas por exponenciales. Ejemplos de ellas son las funciones normal, log-normal y chi cuadrado.

Esto significa que si los datos provenientes, por ejemplo, de una estación de aforo tienen una distribución anual tal que responden a una función de frecuencia normal los n valores extremos (mínimos o máximos) correspondientes a n años de registro hidrométrico, se ajustarán adecuadamente por una función Gumbel. Su expresión matemática es la siguiente:

$$\Phi(y) = e^{-e^{-y}} \quad (3 - 1)$$

$$y = \alpha_0 (x - u_0)$$

Siendo $\Phi(x)$ la función de distribución de la variable reducida y , la cual está relacionada linealmente con la variable aleatoria original x por medio de los parámetros de ajuste α_0 y u_0 .

Derivando la (3-1) puede hallarse la función de frecuencia de Gumbel cuya representación gráfica se presenta en la figura 2-1.

$$\varphi(y) = \Phi'(y) = e^{-(y+e^y)} \quad (3 - 2)$$

3.1 PARAMETROS ESTADISTICOS

Derivando la expresión (3-2) e igualando a cero, puede hallarse el máximo de la función de frecuencia que no es otra cosa que el modo de la distribución de Gumbel.

$$y(max) = 0 \quad x(max) = u_0 \quad (3 - 3)$$

La mediana puede obtenerse igualando a la expresión (3-1) a 0,5 y despejando la variable reducida.

$$y(0,5) = -\ln(\ln 2) \quad x(0,5) = u_0 + \frac{\ln(\ln 2)}{\alpha_0} \quad (3 - 4)$$

La función generatriz de los momentos es:

$$\Psi(t) = \Gamma(1 - t)$$

El valor medio de las variables reducida y original es, según se mencionó anteriormente:

$$\bar{y} = \gamma \qquad \bar{x} = \alpha_0(u_0 + \gamma) \qquad (3 - 5)$$

Se recuerda que $\gamma = 0,577215\dots$ es la constante de Euler, definida en el capítulo anterior. Por otra parte, para la desviación estándar se tiene:

$$\sigma_y = \frac{\pi}{\sqrt{6}} \qquad \sigma_x = \frac{1}{\alpha_0} \frac{\pi}{\sqrt{6}} \qquad (3 - 6)$$

A manera de resumen, se comparan en la Tabla 3.1 los valores numéricos de los principales parámetros estadísticos para los casos de Gauss y Gumbel.

TABLA 3.1: VALORES NUMERICOS DE ALGUNOS PARAMETROS ESTADISTICOS PARA LAS DISTRIBUCIONES DE GAUSS Y GUMBEL		
Parámetro	Gauss	Gumbel
Variable reducida	$z = \frac{x - \bar{x}}{\sigma}$	$y = \alpha_0 (x - u_0)$
Valor medio	0	0,57721
Modo	0	0
Mediana	0	0,36651
Desviación estándar	1	1,28255
Asimetría	0	1,29858
Kurtosis	3	5,40000

Finalmente, resulta de utilidad conocer el valor del período de retorno para el valor medio de la variable reducida. El mismo puede calcularse a partir de las expresiones (2-2), (2-3) y (3-5), arribándose al valor siguiente bajo el supuesto de que se está trabajando con series anuales:

$$\bar{T} = 2,33 \text{ años}$$

3.2 DETERMINACION DE LA RECTA DE AJUSTE

Una vez en posesión del registro hidrométrico de máximos y/o mínimos anuales, debe procederse a la determinación de los parámetros de la recta de ajuste que permitan obtener la mejor aproximación de los datos originales. Para ello puede seguirse alguna de las secuencias de cálculo que se expondrán a continuación.

3.2.1 Método de Gumbel

Gumbel en sus trabajos originales (7), (12) propone desarrollar la secuencia siguiente:

- 1 - Ordenar los datos.
- 2 - Asignar probabilidades.
- 3 - Determinar los valores de la variable reducida.
- 4 - Graficar los datos (optativo).
- 5 - Hallar los coeficientes de la recta de ajuste mínimo cuadrática ortogonal.

3.2.1.1 Ordenamiento de los datos

El primer paso consiste en ordenar los datos, es decir los valores de la variable aleatoria x , en orden creciente o decreciente, según se trate de estiajes o avenidas, respectivamente. Cumplimentar este requerimiento implica, por lo general, alterar el orden cronológico del registro. Inmediatamente debe asignarse a cada dato un número de orden i , teniendo en cuenta que el número I corresponde al suceso menos probable (creciente o bajante extrema).

Tomando el caso de avenidas, la secuencia resulta ser la siguiente:

$$x_1 > x_2 > \dots > x_i > \dots > x_n$$

Como ejemplo puede observarse la Tabla 3.2 en la que se encuentran ordenados los datos de la estación hidrométrica Zárate, provincia de Buenos Aires, República Argentina, correspondientes a crecientes. La columna 2 reúne los datos ordenados, mientras que la 1 indica los números de orden y la 5 el año de ocurrencia.

3.2.1.2 Asignación de probabilidades

El paso siguiente consiste en asignar a cada dato un valor de probabilidad por medio de alguna de las expresiones presentadas en el Tabla 2.2. En este caso se recurrió a la fórmula de Weibull que se reproduce a continuación:

$$p_i = \frac{i}{n + 1} = \frac{1}{T_i} \quad (3 - 7)$$

TABLA 3.2: ESTACION HIDROMETRICA ZARATE (MAXIMOS)					
1	2	3	4	5	6
i	x_j (metros)	i / (n + 1)	y_i	año	ln x_i
1	3,45	0,020	3,922	1983	1,238
2	3,18	0,039	3,219	1959	1,157
3	3,02	0,059	2,803	1940	1,105
4	2,70	0,078	2,505	1966	0,993
5	2,63	0,098	2,271	1963	0,967
6	2,57	0,118	2,078	1958	0,944
7	2,48	0,137	1,913	1973	0,908
8	2,41	0,157	1,768	1972	0,880
9	2,22	0,176	1,639	1975	0,798
10	2,20	0,196	1,522	1962	0,788
11	2,18	0,216	1,415	1982	0,779
12	2,17	0,235	1,316	1951	0,775
13	2,14	0,255	1,223	1960	0,761
14	2,14	0,275	1,137	1977	0,761
15	2,14	0,294	1,055	1981	0,761
16	2,11	0,314	0,977	1974	0,747
17	2,11	0,333	0,903	1978	0,747
18	2,11	0,353	0,832	1979	0,747
19	2,11	0,373	0,763	1980	0,747
20	2,07	0,392	0,697	1938	0,728
21	2,07	0,412	0,634	1939	0,728
22	2,07	0,431	0,572	1957	0,728
23	2,07	0,451	0,511	1971	0,728
24	2,05	0,471	0,453	1946	0,718
25	2,05	0,490	0,395	1961	0,718
26	2,05	0,510	0,338	1965	0,718
27	2,02	0,529	0,283	1950	0,703
28	2,02	0,549	0,228	1954	0,703
29	2,00	0,569	0,173	1941	0,693
30	2,00	0,588	0,120	1942	0,693

TABLA 3.2: ESTACION HIDROMETRICA ZARATE (continuación)					
1	2	3	4	5	6
i	x _i (metros)	i / (n + 1)	y _i	año	ln x _i
31	2,00	0,608	0,066	1947	0,693
32	2,00	0,627	0,013	1948	0,693
33	2,00	0,647	-0,041	1967	0,693
34	2,00	0,667	-0,094	1976	0,693
35	1,99	0,686	-0,148	1944	0,688
36	1,98	0,706	-0,202	1968	0,683
37	1,95	0,725	-0,257	1935	0,668
38	1,95	0,745	-0,313	1936	0,668
39	1,95	0,765	-0,369	1970	0,668
40	1,93	0,784	-0,428	1952	0,658
41	1,90	0,804	-0,488	1969	0,642
42	1,85	0,824	-0,551	1955	0,615
43	1,85	0,843	-0,616	1956	0,615
44	1,84	0,863	-0,686	1945	0,610
45	1,82	0,882	-0,761	1953	0,599
46	1,78	0,902	-0,843	1943	0,577
47	1,71	0,922	-0,934	1949	0,536
48	1,58	0,941	-1,041	1937	0,457
49	1,55	0,961	-1,175	1934	0,438
50	1,54	0,980	-1,369	1964	0,432

La aplicación de la expresión (3-7) implica la asignación de un período de retorno de $n + 1$ años a la mayor creciente y aproximadamente de 1 año a la menor. En la columna 3 del Tabla 3.2 se encuentran tabulados los valores de y_i correspondientes a las cotas hidrométricas listadas en la columna 2.

Es de destacar, dado el criterio seguido para la asignación de probabilidades, la importancia de contar con un registro completo, es decir que incluya los valores extremos para todos los años. En los casos en los que no se disponga de algún valor para la estación bajo análisis, puede tratar de establecerse alguna correlación con datos de estaciones vecinas, o bien, como última posibilidad, utilizar la media aritmética de los valores disponible para reemplazar a los faltantes.

3.2.1.3 Valores de la variable reducida

Para calcular los valores de la variable reducida a partir de las probabilidades asignadas, debe partirse de las expresiones (2-3) y (2-5) mostradas a continuación:

$$p_i = 1 - \Phi(x_i) = 1 - e^{-e^{-y_i}}$$

Despejando se obtiene la expresión siguiente, a partir de la cual se completó la columna 4 de la Tabla 3.2:

$$y_i = -\ln \left(\ln \frac{1}{1 - p_i} \right) \quad (3-8)$$

Siguiendo el procedimiento explicado, se han generado n pares de valores x_i, y_i relacionados linealmente, con lo cual el problema original se ha reducido a la determinación de los coeficiente que mejor ajustan la siguiente expresión:

$$x_i = u_0 + \frac{1}{\alpha_0} y_i \quad (3-9)$$

Como alternativa, puede ser conveniente el graficado de los puntos para luego ensayar una recta de ajuste "a ojo" que dará una primera idea acerca de la bondad de la función elegida. Si la mayoría de los puntos se apartan de esta recta, resultará evidente que deberá recurrirse a otra función distinta de la ensayada, en este caso Gumbel. En particular, si para valores elevados de la variable reducida se observa una tendencia de apariencia exponencial, podría ser adecuado el empleo de una función de Fréchet la que aparecerá como una recta cuando se grafiquen los puntos en papel semilogarítmico.

3.2.1.4 Graficado de los datos

Si el conjunto de pares de valores responde a una función de Gumbel, se agruparán sobre una recta cuando sean graficados en papel lineal. Sin embargo, debe tenerse en cuenta que un gráfico trazado de esta manera sólo expondrá los valores de la variable reducida pero no las probabilidades o los períodos de retorno que son los verdaderamente importantes desde el punto de vista de ingeniería. Para ello, y con el propósito de efectuar extrapolaciones, suele trabajarse con un papel en el cual se indican dos escalas de abscisas auxiliares siguiendo esas dos variables.

En la Figura 3.1 se ilustra un diagrama para graficado de funciones de Gumbel con escala de ordenadas lineales y escala de abscisas en concordancia con la función de distribución de Gumbel. En la parte superior, otras dos escalas presentan la correspondencia con los períodos de retorno y las probabilidades de ocurrencia. Las relaciones entre las escalas están dadas por las siguientes expresiones:

$$\Phi(y) = e^{-e^{-y}} \quad T = \frac{1}{1 - \Phi(y)}$$

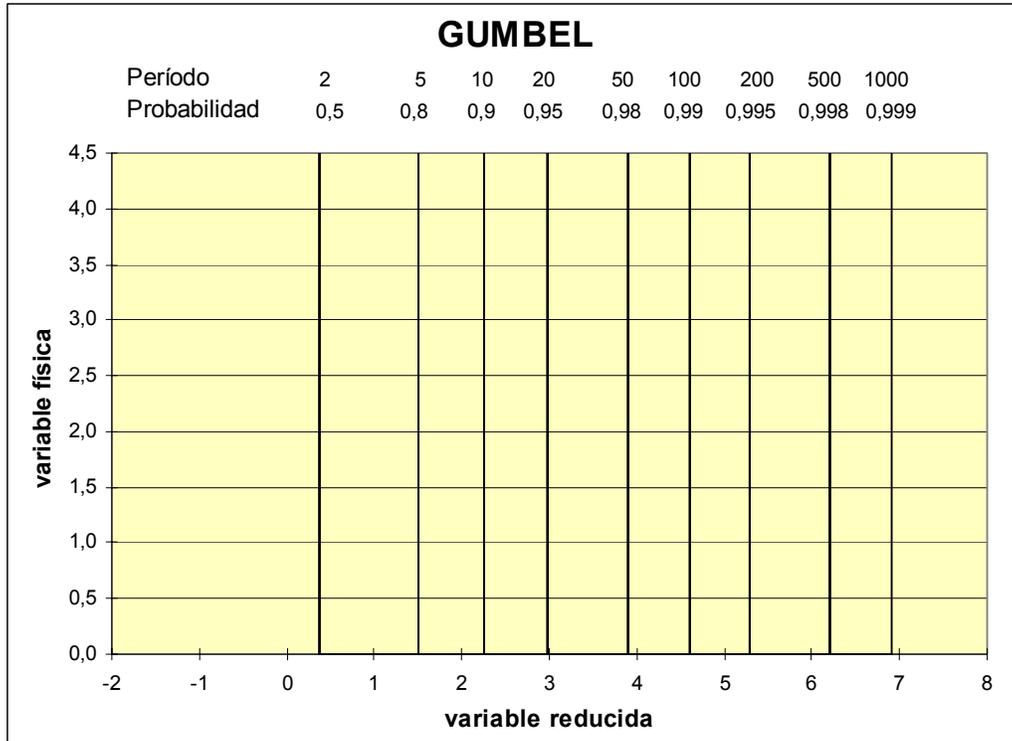


Figura 3.1: Diagrama para graficar funciones de Gumbel

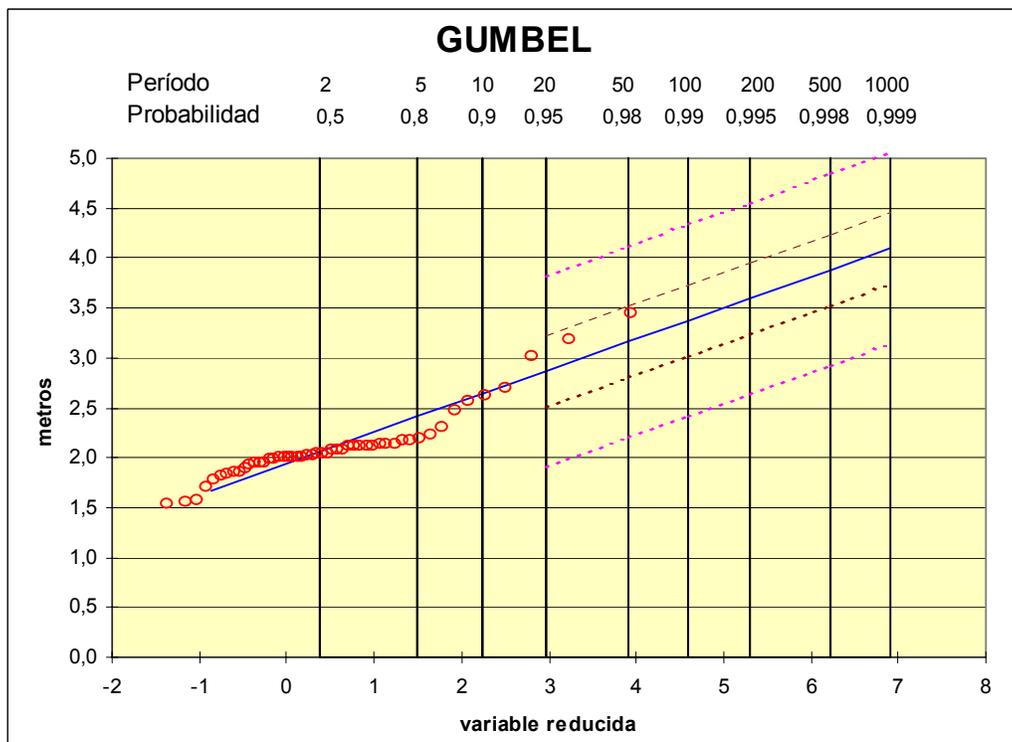


Figura 3.2: Representación de Gumbel para los datos de la Tabla 3.2

En la Figura 3.2 aparecen graficados los puntos de la Tabla 3.2, pudiendo observarse la validez de la aproximación lineal en la primera parte del mismo. También se muestran los intervalos de confianza para 68% y 95% cuyo propósito y cálculo se describirán en el punto 3.3.

3.2.1.5 Coeficientes de ajuste

El paso final consiste en la elaboración de los datos de manera de hallar los valores de los coeficientes α_0 y u_0 que mejor ajusten la recta. El método propuesto por Gumbel (7) se basa en la obtención de la recta de regresión ortogonal mediante un ajuste por cuadrados mínimos. El cálculo de la pendiente de la recta (3-9) se efectúa por medio de la expresión siguiente:

$$\frac{1}{\alpha_0^*} = \frac{\sigma_x}{\sigma_n(y)} \quad (3-10)$$

Por su parte, la ordenada al origen está dada por:

$$u_0^* = \bar{x} - \frac{\bar{y}_n}{\alpha_0^*} \quad (3-11)$$

Siendo:

α_0^* : estimador de α_0

u_0^* : estimador de u_0

\bar{x} : media aritmética de los datos observados

σ_x : desviación estándar de los datos observados

\bar{y}_n : media aritmética de los valores de la variable reducida

$\sigma_n(y)$: desviación estándar de los valores de la variable reducida

La media aritmética y la desviación estándar de los valores de la variable reducida son sólo función del número de datos (n) debido a los valores de probabilidad se obtuvieron con la expresión (3-7). En efecto:

$$\bar{y}_n = \frac{1}{n} \sum_{i=1}^n -\ln \left(\ln \frac{n+1}{n+1-i} \right) \quad (3-12)$$

$$\sigma_n^2(y) = \frac{1}{n} \sum_{i=1}^n \left[-\ln \left(\ln \frac{n+1}{n+1-i} \right) - \bar{y}_n \right]^2 \quad (3-13)$$

La Tabla 3.3 muestra los resultados de aplicar estas dos expresiones para diferentes valores de n .

TABLA 3.3: VALOR MEDIO Y DESVIACION ESTANDAR DE LA VARIABLE REDUCIDA					
n	\bar{y}_n	$\sigma_n(y)$	n	y_n	$\sigma_n(y)$
10	0,495	0,950	110	0,561	1,211
20	0,524	1,063	120	0,562	1,226
30	0,536	1,112	130	0,563	1,219
40	0,544	1,141	140	0,564	1,222
50	0,549	1,161	150	0,565	1,225
60	0,552	1,175	160	0,565	1,228
70	0,555	1,185	170	0,566	1,230
80	0,557	1,194	180	0,566	1,232
90	0,559	1,201	190	0,567	1,234
100	0,560	1,206	200	0,567	1,236
<p>Para $n \rightarrow \infty$ $\bar{y}_n \rightarrow \gamma = 0,577\dots$ $\sigma_n(y) \rightarrow \frac{\pi}{\sqrt{6}} = 1,282\dots$</p> <p>$\gamma = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{i} - \ln n = 0,5772\dots$ (<i>constante de Euler</i>)</p>					

En la Tabla 3.4 se resumen los resultados a los que se arriba a partir de los datos de la estación hidrométrica Zárate (Tabla 3.2). Allí figuran, además del valor medio y la desviación estándar de los datos, los coeficientes de asimetría y kurtosis. Como una medida de la calidad del ajuste también se calculó el coeficiente de correlación. También se exponen los intervalos de confianza cuyo cálculo se explicará más adelante en este mismo capítulo.

Es de destacar que en la Tabla 3.1 faltan datos correspondientes a ciertos años. Estos vacíos se completaron utilizando el promedio de los 45 datos disponibles de manera de alcanzar un total de 50 valores.

La mencionada Tabla 3.4 y la Figura 3.2 presentan los resultados de aplicar la recta de regresión ortogonal, de acuerdo con lo aconsejado por Gumbel. A manera de comparación la Tabla 3.5 muestra los resultados de aplicar las tres posibilidades de regresión.

TABLA 3.4: AJUSTE DE LOS DATOS DE LA TABLA 3.2 FUNCION DE GUMBEL (METODO DE GUMBEL)	
<u>Parámetros de los datos (n = 50)</u>	
Valor medio	2,11 m
Desviación estándar	0,37 m
Asimetría	1,77
Kurtosis	4,21
<u>Parámetros de la variable reducida (n = 50)</u>	
Valor medio	0,549
Desviación estándar	1,172
<u>Coefficientes de ajuste</u>	
Pendiente de la recta de regresión	0,31 m
Ordenada al origen	1,94 m
Coefficiente de correlación	0,956
<u>Extrapolaciones</u>	
Crecida centenaria (T = 100 años)	3,38 m
Intervalo de confianza del 95%	2,42 m - 4,34
Crecida milenaria (T = 1.000 años)	4,10 m
Intervalo de confianza del 95%	3,14 m - 5,06 m

TABLA 3.5: RESULTADOS DE LOS DISTINTOS TIPOS DE REGRESION APLICADOS AL AJUSTE DE GUMBEL		
Regresión	u_0	$1 / \alpha_0$
x sobre y	1,951	0,296
ortogonal	1,940	0,315
y sobre x	1,918	0,332

La expresión de la recta de ajuste se obtiene reemplazando los coeficientes regresión dados por (3-10) y (3-11) en (3-9). Ordenando términos se llega a la siguiente expresión final:

$$x = \bar{x} + \frac{y - \bar{y}_n}{\sigma_n(y)} \sigma_x$$

La cual, de acuerdo con lo expuesto en el punto 2.2, no es otra cosa que la expresión de Chow que se reproduce a continuación.

$$x = \bar{x} + k \sigma_x$$

Sin embargo, cabe aclarar que los valores de k calculados a partir de la expresión (2-9) son válidos sólo para un registro de longitud infinita. En caso contrario se hace necesario el empleo de la ecuación siguiente, en la cual el valor medio y la desviación estándar de la variable reducida deben tomarse de la Tabla 3.3.

$$k = \frac{-\ln\left(\ln\frac{T}{T-1}\right) - \bar{y}_n}{\sigma_n(y)} \quad (3-14)$$

En la Tabla 3.6 se encuentran tabulados los valores de k para varios períodos de retorno y longitudes de registro.

TABLA 3.6: VALORES DE "k" (ECUACION DE CHOW) PARA LA FUNCION DE AJUSTE DE GUMBEL											
T \ n	20	30	40	50	60	70	80	90	100	200	∞
2	-0,148	-0,153	-0,155	-0,157	-0,158	-0,159	-0,159	-0,160	-0,160	-0,162	-0,164
5	0,916	0,866	0,838	0,820	0,807	0,797	0,790	0,784	0,779	0,755	0,719
10	1,625	1,541	1,495	1,466	1,446	1,430	1,419	1,409	1,401	1,362	1,305
20	2,302	2,188	2,126	2,086	2,059	2,038	2,021	2,008	1,998	1,944	1,866
50	3,179	3,026	2,943	2,889	2,852	2,824	2,802	2,784	2,770	2,698	2,592
100	3,836	3,653	3,554	3,491	3,446	3,413	3,387	3,366	3,349	3,263	3,137
200	4,490	4,279	4,164	4,090	4,038	4,000	3,970	3,945	3,925	3,826	3,679
500	5,354	5,104	4,968	4,881	4,820	4,774	4,738	4,710	4,686	4,569	4,395
1.000	6,007	5,728	5,576	5,479	5,411	5,360	5,320	5,288	5,261	5,130	4,936

3.2.2 Método de Lieblein

En 1954 J. Lieblein propuso lo que él denominó "un nuevo método de análisis para valores extremos" el cual se expondrá a continuación de acuerdo con descripción dada en la referencia (9).

3.2.2.1 Selección de subgrupos

Los n datos originales, ordenados en forma cronológica, son divididos en k subgrupos de m elementos cada uno, pudiendo restar un remanente de m' datos. Ahora bien, esta partición no debe ser arbitraria sino que debe seleccionarse de manera de alcanzar un resultado óptimo. La Tabla 3.7 indica la mejor distribución para diversos tamaños de muestra.

TABLA 3.7: SUBGRUPOS PARA MAXIMA EFICIENCIA (METODO DE LIEBLEIN)							
n	(k x m) + m'	n	(k x m) + m'	n	(k x m) + m'	n	(k x m) + m'
10	(2x5)+0	20	(4x5)+0	30	(5X6)+0	40	(6X6)+4
11	(1x6)+5	21	(3x6)+3	31	(5X5)+6	41	(6X6)+5
12	(2x6)+0	22	(3X6)+4	32	(5X6)+2	42	(7X6)+0
13	(2x5)+3	23	(3X6)+5	33	(5X6)+3	43	(8X5)+3
14	(2x5)+4	24	(4X6)+0	34	(5X6)+4	44	(7X6)+2
15	(3x5)+0	25	(5X5)+0	35	(5X6)+5	45	(7X6)+3
16	(2x6)+4	26	(4X6)+2	36	(6X6)+0	46	(7X6)+4
17	(2x6)+5	27	(4X6)+3	37	(7X5)+2	47	(7X6)+5
18	(3x6)+0	28	(4X6)+4	38	(6X6)+2	48	(8X6)+0
19	(3x5)+4	29	(4X6)+5	39	(6X6)+3	49	(9X5)+4
						50	(8X6)+2

3.2.2.2 Cálculo de los estimadores

A partir de los valores de k , m y m' pueden obtenerse los siguientes factores de proporcionalidad para el cálculo del valor medio y de la varianza.

$$\bar{t} = \frac{k m}{n} \quad (3-15)$$

$$t' = \frac{m'}{n} \quad (3-16)$$

$$\bar{q} = \frac{(\bar{t})^2}{k} \quad (3-17)$$

$$q' = (t')^2 \quad (3-18)$$

En este momento debe confeccionarse una matriz de $k \times m$ completando cada fila con un subgrupo. En cada fila los elementos deben estar ordenados en forma creciente (cuando se trabaja con máximos) con lo cual el orden cronológico original resultará alterado.

$$x_{i,j} \leq x_{i,j+1}$$

El siguiente paso consiste en sumar los elementos de cada columna.

$$S_i = \sum_{j=1}^k x_{i,j} \quad (3-19)$$

Finalmente, se calculan los estimadores de los parámetros de ajuste.

$$\bar{u}_0 = \frac{1}{k} \sum_{i=1}^m a_{m_i} S_i = u_0^* \quad (3-20)$$

$$\frac{1}{\alpha_0} = \frac{1}{k} \sum_{i=1}^m b_{m_i} S_i = \frac{1}{\alpha_0^*} \quad (3-21)$$

Las expresiones (3-20) y (3-21) coincidirán con los estimadores sólo si $m' = 0$. En caso contrario, es decir si existe un grupo remanente, deben efectuarse algunas operaciones adicionales comenzando por resolver las sumatorias siguientes:

$$u'_0 = \sum_{i=1}^{m'} a_{m_i} x_i \quad (3-22)$$

$$\frac{1}{\alpha'_0} = \sum_{i=1}^{m'} b_{m_i} x_i \quad (3-23)$$

Por último, los estimadores en caso de existir un grupo remanente, se determinan de la siguiente manera:

$$u_0^* = \bar{t} \bar{u}_0 + t' u'_0 \quad (3-24)$$

$$\frac{1}{\alpha_0^*} = \bar{t} \frac{1}{\alpha_0} + t' \frac{1}{\alpha'_0} \quad (3-25)$$

Los valores de a_{mi} y b_{mi} son función de m en las ecuaciones (3-20) y (3-21) y de m' en las (3-22) y (3-23). Pueden ser obtenidos en la Tabla 3.8.

TABLA 3.8: FACTORES DE PESO (METODO DE LIEBLEIN)						
a _{mi}						
i m o m'	1	2	3	4	5	6
2	0,916	0,084	-	-	-	-
3	0,656	0,256	0,088	-	-	-
4	0,511	0,264	0,154	0,071	-	-
5	0,419	0,246	0,168	0,109	0,058	-
6	0,355	0,225	0,166	0,121	0,083	0,049
b _{mi}						
i m o m'	1	2	3	4	5	6
2	-0,721	0,721	-	-	-	-
3	-0,630	0,256	0,375	-	-	-
4	-0,559	0,086	0,224	0,249	-	-
5	-0,503	0,006	0,130	0,182	0,185	-
6	-0,459	-0,036	0,073	0,127	0,149	0,146

3.2.2.3 Cálculo de la varianza

El método de Lieblein permite calcular las varianzas para cada valor de la variable mediante la siguiente expresión:

$$\sigma^2(x) = \bar{q} Q_m + q' Q_{m'} \quad (3-26)$$

Estando Q_m y $Q_{m'}$ (genéricamente Q_n) definidos por medio de:

$$Q_n = \left(A_n y^2 + B_n y + C_n \right) \left(\frac{1}{\alpha_0^*} \right)^2 \quad (3-27)$$

La Tabla 3.9 presenta los valores de A_n , B_n y C_n en función de m (para Q_m) o de m' (para $Q_{m'}$).

TABLA 3.9: COEFICIENTES PARA EL CALCULO DE VARIANZAS (METODO DE LIEBLEIN)			
m o m'	A_n	B_n	C_n
2	0,71186	-0,12864	0,65955
3	0,34472	0,04954	0,40286
4	0,22528	0,06938	0,29346
5	0,16665	0,06798	0,23140
6	0,13196	0,06275	0,19117

Por último, en la Tabla 3.10 se exponen los resultados de aplicar este método a los datos de la Tabla 3.2 (estación Zárate). Esta información puede compararse con los resultados obtenidos aplicando el método de Gumbel presentados en la Tabla 3.4.

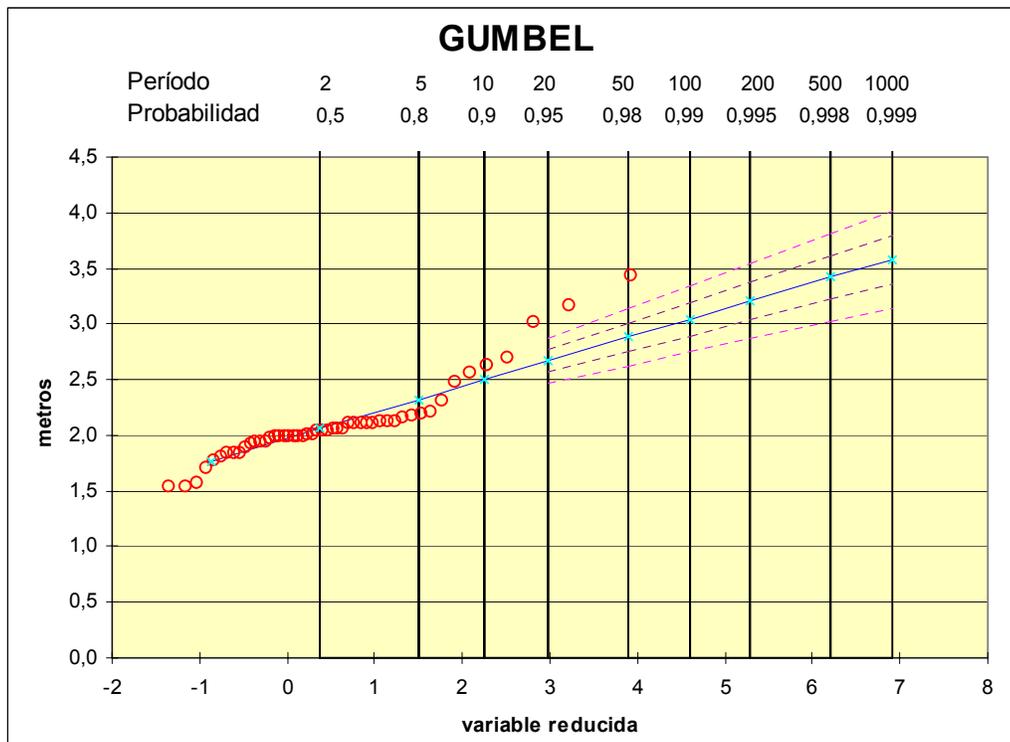


Figura 3.3: Ajuste de Lieblein (función de Gumbel) para los datos de la Tabla 3.2

La Figura 3.3 ilustra la recta de ajuste de los datos de la estación Zárate obtenida por medio del Lieblein.

TABLA 3.10: AJUSTE DE LOS DATOS DE LA TABLA 3.2 FUNCION DE GUMBEL (METODO DE LIEBLEIN)	
<u>Parámetros de los datos (n = 50)</u>	
Valor medio	2,11 m
Desviación estándar	0,37 m
Asimetría	1,75
Kurtosis	4,08
<u>Partición</u>	
n = 50 k = 8 m = 6 m' = 2	
<u>Coefficientes de ajuste</u>	
Pendiente de la recta de regresión	0,23 m
Ordenada al origen	1,97 m
Coefficiente de correlación	0,956
<u>Extrapolaciones</u>	
Crecida centenaria (T = 100 años)	3,05 m
Intervalo de confianza del 95%	2,75 m - 3,35
Crecida milenaria (T = 1.000 años)	5,58 m
Intervalo de confianza del 95%	3,15 m - 4,01 m

3.3 INTERVALOS DE CONFIANZA

Antes de entrar de lleno en el cálculo de los intervalos de confianza para un ajuste que responda a la distribución de Gumbel, es conveniente repasar algunos conceptos básicos acerca de este tema.

Ya se mencionó en el punto 1.4.6 que los parámetros de una población estadística son valores fijos, aunque desconocidos, que pueden ser evaluados mediante *estimadores* o *estadísticos* de la muestra. Estos últimos son realmente variables aleatorias ya que su valor depende de la muestra tomada.

De la misma manera que es necesario evaluar los errores cometidos al realizar cualquier tipo de medición, también resulta de suma importancia cuantificar el nivel de confianza que merecen los estimadores.

Si α^* es el estimador de un parámetro α de la población, pueden encontrarse dos números δ y ϵ , tal que la probabilidad de que el parámetro verdadero α quede incluido entre los límites $\alpha^* \pm \delta$ sea igual a $1 - \epsilon$ (3), siendo este el nivel de confianza de la estimación.

$$P(\alpha^* - \delta < \alpha < \alpha^* + \delta) = 1 - \epsilon \quad (3-28)$$

Esto quiere decir que la probabilidad de que el intervalo $\alpha^* \pm \delta$ contenga el valor buscado es igual a $1 - \epsilon$. O sea que, por ejemplo, para $\epsilon = 0,05$, el intervalo tendrá una confianza el 95%, lo que implica que de cada 20 muestras tomadas el intervalo estimado resultará correcto en 19 oportunidades, aunque por supuesto, no puede saberse cual es el incorrecto.

Por otra parte, el valor de ϵ es una medida del riesgo que se corre de cometer un error cuando se acepta el intervalo elegido. En los casos en que este riesgo es expresado en forma porcentual se lo denomina nivel de significación de la estimación. En el ejemplo anterior resultaría ser del 5%.

Se podría pensar entonces que, reduciendo ϵ infinitamente, el riesgo de cometer un error tenderá a cero mejorando la estimación. Esto es cierto, pero una disminución en ϵ implica un incremento en δ con lo cual se aumenta la confianza pero se disminuye la exactitud, dado que el intervalo resulta más amplio. En el límite para un nivel de significación tendiendo a cero, la amplitud del intervalo de confianza tenderá a infinito.

En términos prácticos, el problema consiste en fijar un nivel de significación y luego determinar los valores de δ para dicho nivel. En el caso de la distribución de Gumbel se han propuesto diversos métodos para lograr este fin, algunos de ellos se describirán a continuación.

3.3.1 Método de Gumbel

Gumbel propone un mecanismo muy simple para la determinación de intervalos de confianza, consistente en el trazado a ambos lados de la recta de ajuste de lo que él denomina "*líneas de control*", válidas para grandes períodos de retorno.

El área encerrada por la función de frecuencia normal entre los puntos $\bar{x} \pm \sigma$ es, según se expuso anteriormente, aproximadamente 0,68. El intervalo simétrico que encierra un área similar para una función de Gumbel está limitada por los puntos $y = -1,14$ e $y = 1,14$ como se ilustra en la Figura 3.4, es decir:

$$\Phi(1,14) - \Phi(-1,14) = 0,68$$

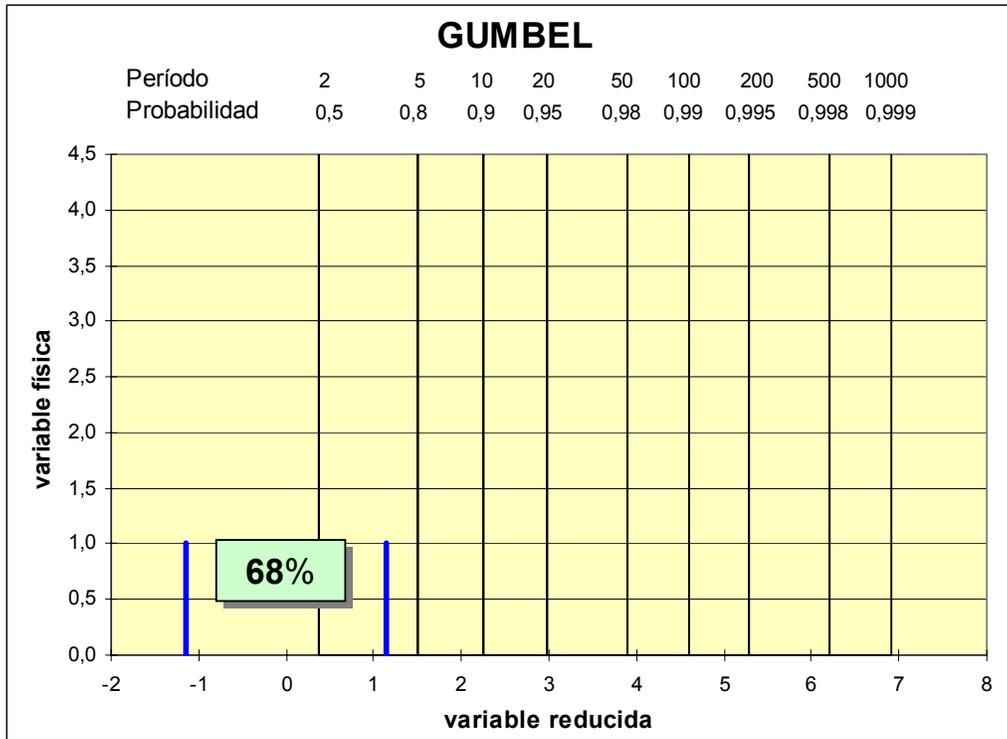


Figura 3.4: Intervalo simétrico del 68% alrededor del origen

Como ya se dijo, el método sugerido por Gumbel consiste en trazar dos líneas paralelas a la recta de ajuste, tal que para un valor dado x_0 de la variable aleatoria, dichas rectas delimiten un intervalo de la variable reducida dado por $y_1 = y_0 - 1,14$ e $y_2 = y_0 + 1,14$. La idea es que, para un período de retorno especificado, puede calcularse la variable reducida y luego sumarse y restarse 1,14 a la misma. Las dos rectas de pendiente igual a la de ajuste que pasen por esos puntos serán las líneas de control de Gumbel para un nivel de confianza del 68%.

Según Gumbel, este criterio es válido para grandes valores del período de retorno donde se cumple la igualdad siguiente (ver punto 2.3.1):

$$y = \ln T$$

Entonces resulta, para el caso de y_1 :

$$y_0 - y_1 = \ln T_0 - \ln T_1 = \ln \frac{T_0}{T_1} = 1,14$$

Con lo cual:

$$T_1 = 0,32 T_0 \quad (3-29)$$

Un razonamiento similar aplicado a y_2 permite arribar a la siguiente igualdad.

$$T_2 = 3,13 T_0 \quad (3-30)$$

Igualmente puede obtenerse el intervalo de confianza del 95% trazando rectas paralelas a la recta de ajuste que pasen por puntos $y_1 = y_0 - 3,07$ e $y_2 = y_0 + 3,07$. Los períodos correspondientes son:

$$T_1 = 0,05 T_0 \quad (3-31)$$

Un razonamiento similar aplicado a y_2 permite arribar a la siguiente igualdad.

$$T_2 = 21,5 T_0 \quad (3-32)$$

Ambos intervalos de confianza está representados en la Figura 3.2.

De acuerdo con Gumbel, si todos los puntos graficados según el método expuesto en 2.4 caen dentro de la región delimitada por las rectas de control del 68% de confianza, el ajuste puede considerarse correcto.

Para un dado valor de período de retorno, para el cual deba calcularse el intervalo de confianza, el método más simple consiste en sumar y restar a la variable reducida el desplazamiento correspondiente al intervalo buscado y calcular el valor de la variable aleatoria por medio de la ecuación de ajuste. Por ejemplo, para una crecida centenaria ($T = 100$ años) la variable aleatoria, de acuerdo con la Tabla 2.1, es igual a 4,6 en tanto que los límites de confianza del 68% se obtienen sumando y restando 1,14 a ese valor.

La Tabla 3.4, que presenta los resultados del ajuste en la estación Zárate, indica que la pendiente de la recta de ajuste es de 0,31 m, siendo la ordenada al origen 1,94 m. Para dicho caso, los intervalos se calculan como sigue.

$$x(T = 100) = u_0 + \frac{1}{\alpha_0} y = 1,94 \text{ m} + 0,31 (4,6) \text{ m} = 3,38 \text{ m}$$

$$x_1(T = 100) = u_0 + \frac{1}{\alpha_0} (y_1 - 1,14) = 1,94 \text{ m} + 0,31 (4,6 - 1,14) \text{ m} = 3,02 \text{ m}$$

$$x_2(T = 100) = u_0 + \frac{1}{\alpha_0} (y_1 + 1,14) = 1,94 \text{ m} + 0,31 (4,6 + 1,14) \text{ m} = 3,73 \text{ m}$$

En la Tabla 3.4 se trabajó con un desplazamiento de 3,07, en lugar de 1,14, equivalente a un intervalo de confianza del 95%. Se recuerda que, para una crecida milenaria ($T = 1.000$ años), la variable reducida toma un valor de 6,907.

3.3.2 Método de Lieblein

Cuando el ajuste de los datos experimentales se realiza siguiendo los pasos desarrollados en el punto 3.2.2, los intervalos de confianza se determinan en base a la varianza de cada punto calculada por medio de la expresión (3-26). Sumando y restando a los valores extrapolados una desviación estándar se obtiene el intervalo del

68% de confianza, en tanto que si toman dos desviaciones queda determinado el intervalo del 95%.

$$x_0 \pm \sigma_x \quad (68\%) \quad (3-33)$$

$$x_0 \pm 2 \sigma_x \quad (95\%) \quad (3-34)$$

Es de destacar que, a diferencia del método de Gumbel donde se trabaja sobre la variable reducida, en este caso se opera directamente sobre los resultados (cota hidrométrica en el ejemplo).

De esta forma se calcularon los resultados que aparecen en la Tabla 3.10 y en la Figura 3.3.

3.3.3 Método de Kaczmarek

Este método, propuesto en 1947, se basa en que la distribución de los momentos muestrales es asintóticamente normal, aproximándose a una función de Gauss cuando el tamaño de la muestra tiende a infinito.

Como regla general en estadística, el intervalo de confianza del 100 - p% para el valor medio está dado por:

$$\bar{x} \pm t_p \frac{\sigma_x}{\sqrt{n-1}} \quad (3-35)$$

Donde t_p es la abscisa de la función t de Student que deja en cada cola de la distribución un área de 0,5 p% y cuyo valor depende del tamaño de la muestra (n). Para un nivel de confianza del 95%, t_p tiende a 1,96 cuando n tiende a infinito, coincidiendo con el valor correspondiente a la abscisa de la función de frecuencia normal que encierra la misma área en ambas colas. La Tabla A3 del apéndice presenta los valores de t_p para distintos casos.

Kaczmarek emplea esta ecuación, ligeramente modificada, para el cálculo de la desviación estándar teniendo en cuenta que aquí se pretenden realizar extrapolaciones a partir de un ajuste inicial. Aplicando este concepto, el intervalo para el valor estimado genérico x_T es el siguiente:

$$x_T \pm t_p \xi(n, T) \frac{\sigma_x}{\sqrt{n}} \quad (3-36)$$

La Tabla 3.11 muestra los valores de $\xi(n, T)$ y de t_p para distintos tamaños de muestra, períodos de retorno y niveles de confianza. (6)

TABLA 3.11 COEFICIENTES DE KACZMAREK (ξ)							
$n \backslash T$	10	20	25	30	50	75	100
15	2,476	3,233	3,409	3,604	4,113	4,525	4,818
29	2,400	3,075	3,292	3,468	3,968	4,362	4,643
25	2,350	3,007	3,218	3,391	3,874	4,259	4,533
30	2,317	2,460	3,166	3,336	3,811	4,187	4,455
40	2,272	2,898	3,099	3,264	3,725	4,093	4,353
50	2,244	2,857	3,056	3,217	3,671	4,031	4,288
60	2,224	2,830	3,025	3,185	3,633	3,989	4,242
75	2,201	2,800	2,976	3,150	3,592	3,943	4,194
100	2,181	2,769	2,959	3,114	3,549	3,896	4,142
%	50	68	80	90	95	99	
t_p	0,674	1,000	1,282	1,645	1,960	2,576	

Para períodos superiores a los 100 años puede emplearse la siguiente expresión aproximada:

$$\xi(n, T) \approx \sqrt{1 + 1,16 k(n, T) + 1,1 k^2(n, T)} \quad (3-37)$$

Siendo $k(n, T)$ el coeficiente de la ecuación de Chow (2-4) definido por la expresión (3-14) y tabulado en la Tabla 3.6.

3.3.4 Método de Bernier-Veron

Bernier y Veron desarrollaron en 1964 un método que conduce a intervalos de confianza asimétricos aplicables a las distribuciones de Gumbel y Fréchet. La expresión a aplicar es la siguiente:

$$x_T - P_2 \sigma_x \leq x_T \leq x_T + P_1 \sigma_x \quad (3-38)$$

El cálculo de los coeficientes P_1 y P_2 se efectúa, también en este caso, a partir de los valores del tamaño muestral y período de retorno considerado con la ayuda de dos gráficos elaborados por los mencionados autores y que se reproducen en la Figura 3.5 (nivel de confianza del 70%) y en la Figura 3.6 (nivel de confianza del 95%). Nótese que en abscisas se representa la raíz de cuadrada del número de datos. Según la referencia 6, los intervalos del 95% frecuentemente resultan exagerados en tanto que los intervalos del 70% suelen ser adecuados para la generalidad de los casos.

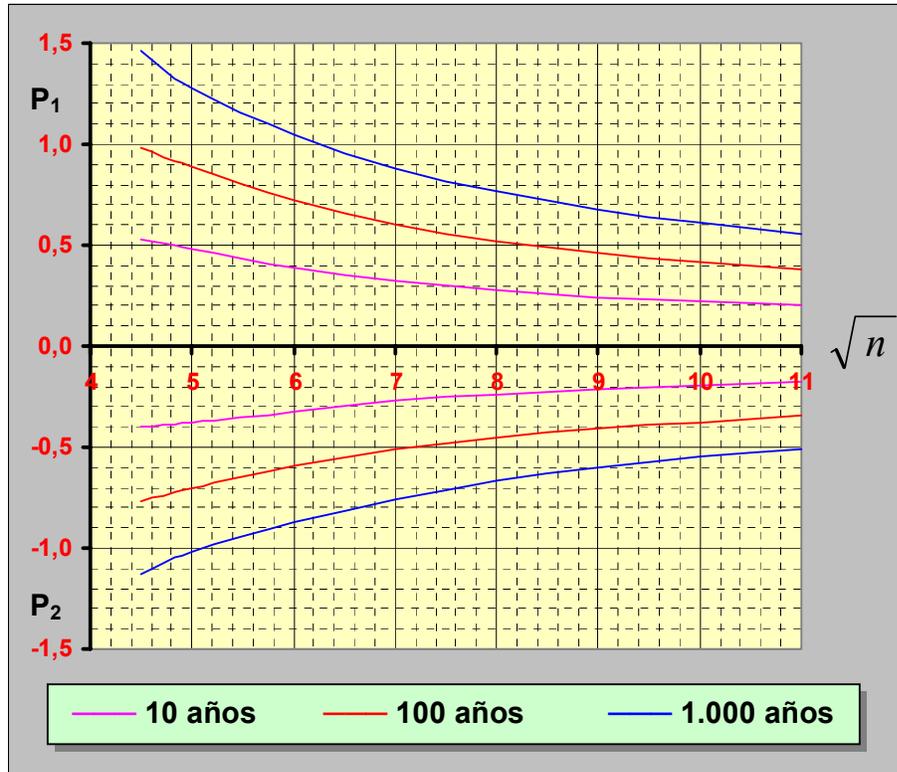


Figura 3.5: Intervalos de confianza (Bernier - Veron) para 70%

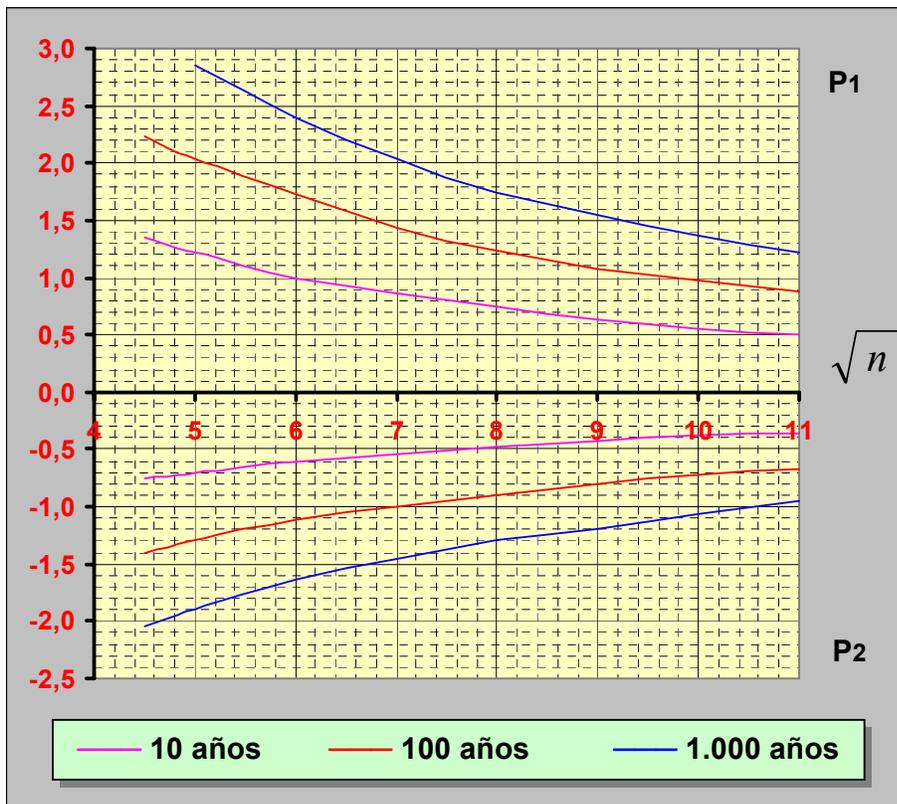


Figura 3.6: Intervalos de confianza (Bernier - Veron) para 95%

3.3.5 Comparación entre métodos

En la Tabla 3.12 se presentan las predicciones para la variable aleatoria para períodos de retorno comprendidos entre 10 años y 1.000 años de acuerdo con la recta de regresión ortogonal correspondiente al ajuste de los datos de la Tabla 3.2. En ésta, se completaron los datos faltantes con el promedio de los 45 valores originales dando lugar a un conjunto de 50 cotas hidrométricas. En dicha tabla se exponen también los intervalos de confianza del 95% calculados por tres métodos diferentes.

TABLA 3.12: PREDICCIONES E INTERVALOS DE CONFIANZA DEL 95% PARA LOS DATOS DE TABLA 3.2 (FUNCION DE GUMBEL)							
T (años)	x_T (m)	x_{MIN} (m)			x_{MAX} (m)		
		Gumbel	Kaczmarek	Bernier Veron	Gumbel	Kaczmarek	Bernier Veron
10	2,64	1,69	2,41	2,45	3,60	2,89	2,99
20	2,87	1,91	2,57	2,62	3,83	3,19	3,28
50	3,16	2,20	2,77	2,84	4,12	3,57	3,67
100	3,38	2,42	2,93	3,00	4,34	3,85	3,95
200	3,60	2,64	3,08	3,17	4,56	4,14	4,24
500	3,88	2,93	3,28	3,39	4,84	4,52	4,63
1.000	4,10	3,14	3,43	3,55	5,06	4,81	4,92

Para el cálculo de los intervalos de confianza según el método de Gumbel se aplicaron los conceptos desarrollados en el punto 3.3.1 tomando desplazamientos de $\pm 3,07$ para la variable reducida y reemplazándolos en la expresión de la recta de regresión.

Los intervalos así obtenidos son, por regla general, excesivos y de ancho constante para todos los períodos de retorno (en este caso 1,92 m), consecuencia de haber trazado rectas de control paralelas a la recta de regresión. Debido a que su fundamento teórico se basa en la aproximación logarítmica de la función de Gumbel, sólo válida para períodos de retorno elevados, su campo de aplicación estaría reducida a esa región.

Si se define el error de predicción, superior e inferior, como el cociente entre cada uno de los extremos del intervalo de confianza, respectivamente, y el valor extrapolado de la variable aleatoria, se tendrá un parámetro de evaluación relativo a la confiabilidad de la predicción para un cierto nivel de confianza. Resulta evidente que, en los casos en que los intervalos sean simétricos, ambos errores (superior e inferior) coincidirán.

Para el método de cálculo de intervalos de Gumbel se obtuvieron los siguientes errores de predicción para el ejemplo considerado:

$$T = 100 \text{ años} \quad \varepsilon = \pm 29\%$$

$$T = 1.000 \text{ años} \quad \varepsilon = \pm 24\%$$

Para el método de Kaczmarek se tomaron los valores de la Tabla 3.11 y se aplicó la expresión (3-36). Para períodos superiores a los 100 años se utilizó la fórmula aproximada (3-37). Los intervalos obtenidos son, en todos los casos, más estrechos que los dados por las líneas de control de Gumbel y van ensanchándose a medida que los períodos aumentan. El ancho relativo también se incrementa. Los errores de predicción son:

$$T = 100 \text{ años} \quad \varepsilon = \pm 14\%$$

$$T = 1.000 \text{ años} \quad \varepsilon = \pm 17\%$$

Para la aplicación del método de Bernier-Veron se trabajó con el gráfico de la Figura 3.6, interpolándose en forma logarítmica para períodos de retorno intermedios. Los intervalos de confianza hallados de esta forma siguen un comportamiento parecido a los determinados por el método de Kaczmarek pero son asimétricos con mayor incertidumbre a la derecha. Los errores son:

$$T = 100 \text{ años} \quad \varepsilon = +17\%; -12\%$$

$$T = 1.000 \text{ años} \quad \varepsilon = +19\%; -14\%$$

Por último, de acuerdo con los resultados presentados para el método de ajuste de Lieblein, los errores serían:

$$T = 100 \text{ años} \quad \varepsilon = \pm 10\%$$

$$T = 1.000 \text{ años} \quad \varepsilon = \pm 12\%$$

3.4 PRUEBAS DE HIPOTESIS

La prueba del chi cuadrado (ya presentada en el Capítulo 1) constituye una herramienta estadística que permite comprobar la veracidad de una hipótesis relativa al correcto ajuste de datos experimentales por medio de una distribución teórica. Aunque esta prueba suele dar resultados aceptables en numerosas situaciones, en el caso particular de la distribución de Gumbel no logra comportarse de manera totalmente satisfactoria.

En efecto, Gumbel, en ocasión de una conferencia dictada en París en 1956 en la que expuso su teoría de los valores extremos, decía: *"Es común entre los estadísticos utilizar el criterio conocido como chi cuadrado (χ^2). De esta manera se obtiene la probabilidad $P(\chi^2)$ de que las desviaciones entre la teoría y las observaciones sean debidas al azar. Pero este criterio depende de dos decisiones arbitrarias: la dimensión de los intervalos de clase, cuya influencia es conocida, y el comienzo del primer intervalo, cuya influencia es desconocida. Yo he demostrado por un ejemplo numérico válido para los mismos parámetros y para las mismas dimensiones de los intervalos que pequeñas variaciones en el comienzo producen modificaciones de $P(\chi^2)$ desde*

0,023 hasta 0,705. Es por esas razones que nos oponemos al empleo de este método para variables continuas" (7). A continuación comenzaba a discutir su propuesta de trazado de líneas de control descriptas en el punto 3.3.1. El ejemplo mencionado por Gumbel está expuesto en su obra "On the reliability of the classical chi square test" (Annual of Mathematics and Statistics, vol. 14, 143, pag. 253).

De acuerdo con Gumbel, la mejor manera de corroborar la calidad del ajuste es que todos los puntos graficados queden comprendidos entre ambas líneas de control.

Otros autores (6) aceptan el criterio de chi cuadrado con algunas limitaciones, consecuencia del generalmente reducido tamaño de la muestra que obliga a tomar pocos intervalos de clase lo que contribuye a reducir la efectividad de la prueba.

En principio, deberían aceptarse todas las hipótesis que arrojen valores de chi cuadrado con probabilidades determinadas por el área en la cola de la distribución superiores al 5%, mientras que deberían rechazarse aquellas que impliquen probabilidades inferiores al 1%. Casos intermedios son dudosos, por lo que convendría aguardar hasta contar con una muestra más amplia antes de tomar decisión alguna.

Es común, al realizar un análisis hidrológico de variables extremas, que se presenten casos en los cuales los valores de chi cuadrado son muy pequeños. Ellos deben ser tratados con mucha precaución dado que también pueden deberse a no disponer de una muestra lo suficientemente grande.

Para los datos de la Estación Hidrométrica Zárate (Tabla 3.2) ajustados de acuerdo a lo expuesto en la Tabla 3.4 y en la Tabla 3.12 se aplicó la dócima chi cuadrado dividiendo los datos experimentales en seis intervalos de clase. Dado que, a partir de ellos, se calcularon dos parámetros requeridos para la determinación de las frecuencias teóricas, el número final de grados de libertad será de tres. El resumen del cálculo se muestra en la Tabla 3.13.

TABLA 3.13 PRUEBA DE CHI CUADRADO APLICADA A LOS DATOS DE LA TABLA 3.2 AJUSTADOS POR MEDIO DE UNA FUNCION DE GUMBEL			
x (m)	f	n p_i	(f - n p_i)² / n p_i
< 1,5	0	0,89	0,89
1,51 - 1,80	5	9,63	2,23
1,81 - 2,00	17	11,35	2,81
2,01 - 2,20	18	10,38	5,59
2,21 - 2,40	3	7,37	2,59
> 2,40	7	10,38	1,10
	50	50	15,21
para p = 0,005 → $\chi^2_p = 7,815$ r = 6 - 1 - 2 = 3 grados de libertad			

La segunda columna contiene la frecuencia experimental, es decir la cantidad de datos de la muestra que cayeron en cada intervalo de clase indicado en la primera columna.

Para el cálculo de las frecuencias teóricas que aparecen en la tercera columna se determinaron los valores de la variable reducida que corresponden a los límites del intervalo ($x_1 - x_2$) mediante la ecuación de la recta de ajuste. Luego se cuantificó la función de distribución de Gumbel por medio de la expresión (2-5) para cada uno de los dos valores de la variable reducida (y_1 e y_2). Finalmente, las frecuencias teóricas se hallan de la siguiente manera:

$$n p_i = n [\Phi(y_2) - \Phi(y_1)]$$

En la cuarta columna aparecen los valores que, sumados, determinan χ^2 , en este caso el resultado es 15,21. Si se desea un nivel de significación del 5% para la prueba, debe buscarse en tablas (tal como la Tabla A2 del apéndice) el valor de chi cuadrado cuya probabilidad de ser excedido sea 0,05 para 3 grados de libertad, resultando:

$$p = 0,005 \rightarrow \chi^2_p = 7,815$$

El valor determinado por la prueba es superior al límite y corresponde a una probabilidad de 0,16%, sensiblemente inferior a la buscada, lo que conduciría al rechazo de la hipótesis. Sin embargo, la Figura 3.2 muestra un ajuste razonable de los puntos por medio de la recta de regresión.

Al respecto, cabe recordar lo aconsejado por el United States Geological Survey acerca de la importancia de verificar la concordancia de las hipótesis "a ojo" (6). El fundamento de este concepto aplicado a datos hidrogeológicos es que su naturaleza imperfecta, en calidad y cantidad, limita la eficiencia de cualquier modelo matemático aunque éste sea inobjetable desde el punto de vista estrictamente teórico.

Como otro elemento adicional de controversia en lo referente a las pruebas de hipótesis, se ha procedido a aplicar la prueba de Kolmogorof-Smirnof (punto 1.6.2) al mismo juego de datos y función de ajuste. Su resumen se presenta en la Tabla 3.14.

En este caso se han utilizado sólo cinco intervalos de clase, presentados en la primera columna, de manera de que ninguno quedara vacío. En la segunda columna se muestra la frecuencia experimental pero, en este caso, relativa en tanto que en la tercera columna se disponen las frecuencias relativas acumuladas cuyo valor se compara con la función de distribución de Gumbel que aparece en la cuarta columna. Esta última se determina, nuevamente, mediante la expresión (2.5) para valores de la variable reducida calculados a partir de los extremos de los intervalos utilizando la expresión de la recta de regresión.

Por último, la última columna presenta la diferencia, en valor absoluto, entre las funciones de distribución teórica y experimental y destaca su valor máximo.

TABLA 3.14 PRUEBA DE KOLMOGOROF-SMIRNOF APLICADA A LOS DATOS DE LA TABLA 3.2 AJUSTADOS POR MEDIO DE UNA FUNCION DE GUMBEL				
x (m)	f _i / n	Σ (f _i / n)	Φ _i (x)	Δ _i
< 1,80	0,10	0,10	0,21	0,11
1,81 - 2,00	0,34	0,44	0,44	0,00
2,01 - 2,20	0,36	0,80	0,64	0,16
2,21 - 2,40	0,06	0,86	0,80	0,06
> 2,40	0,14	1,00	1,00	0,00
para p = 0,05 → z _p = 1,36				

La diferencia máxima es de 0,16, en tanto que el parámetro correspondiente a un nivel de significación del 0,05 tomado de la Tabla 1.4 es 1,36. La aplicación de la prueba conduce al siguiente resultado:

$$\sqrt{n} \Delta_{\max} = 0,16 \sqrt{50} = 1,13 < z_p = 1,36$$

Lo anterior indica que, al nivel de significación del 5%, la hipótesis debería aceptarse; exactamente lo contrario de lo concluido con la aplicación de la prueba del chi cuadrado viniendo a corroborar lo relativo de las pruebas de hipótesis cuando son utilizadas para docimar ajustes de distribuciones extremas de datos hidrológicos.

4. DISTRIBUCIONES LOGARITMICAS

A diferencia de la distribución de Gumbel cuya expresión se aplica directamente a los datos, aquí se denominan distribuciones logarítmicas a las que operan sobre el logaritmo natural de los datos. Sin embargo, debe recordarse que éstos son variables físicas (cotas hidrométricas, caudales, velocidades de vientos, niveles de nieve y otras) y que, por lo tanto, deberían normalizarse de manera de transformarlos en magnitudes adimensionales sobre las que sea posible aplicar logaritmos. También debería trasladarse el "cero" de referencia, cuando ello sea necesario, para evitar números negativos.

A pesar de lo dicho suele operarse en forma directa sobre los datos, dado que finalmente, cuando se obtenga la función de ajuste deseada se aplicarán antilogaritmos para retornar a la magnitud original.

En aquellos casos en los que la solución al problema sea la obtención de una recta de regresión, el resultado adoptará una forma del tipo siguiente:

$$y = a + b \ln x$$

Cuando sea necesario utilizar parámetros estadísticos, éstos se aplicarán sobre los logaritmos de los datos.

Valor medio
$$\overline{x}_{\log} = \frac{1}{n} \sum_{i=1}^n \ln x_i \quad (4-1)$$

Desviación estándar
$$\sigma_{\log} = \sqrt{\frac{1}{n-1} \sum (\ln x - \overline{x}_{\log})^2} \quad (4-2)$$

Tomando, una vez más, la información de la Tabla 3.2, en la Tabla 4.1 se comparan los parámetros estadísticos calculados a partir de los datos originales y de los logaritmos de los datos.

TABLA 4.1: PARAMETROS ESTADISTICOS LINEALES Y LOGARITMICOS (TABLA 3.2)		
Parámetro	Lineal (x)	Logarítmico (ln x)
Valor medio	2,113	0,735
Desviación estándar	0,366	0,158
Asimetría	1,173	1,087
Kurtosis	4,209	2,426

Un rápido análisis permite concluir que la distribución logarítmica está muy desplazada hacia la izquierda, algo más dispersa (en forma relativa), es ligeramente menos asimétrica y menos plana.

Entre las distribuciones aplicadas a logaritmos más utilizadas para el ajuste de valores extremos se cuentan Fréchet, Galton (es decir, log normal) y log Pearson.

4.1 DISTRIBUCION DE FRECHET

La función de distribución de Fréchet está definida por la siguiente expresión:

$$\Phi(x) = e^{-(\Theta x)^{-\alpha_0}} \quad (4-3)$$

El campo de variación de la variable aleatoria se extiende entre 0 y $+\infty$. La expresión anterior también puede escribirse de la forma siguiente:

$$\Phi(x) = e^{-e^{-\alpha_0 \ln \Theta x}}$$

Haciendo $-\ln \Theta = u_0$ y reemplazando resulta:

$$\Phi(x) = e^{-e^{-\alpha_0 (\ln x - u_0)}} \quad (4-4)$$

Es decir:

$$\Phi(x) = e^{-e^{-y}} \quad (4-5)$$

$$y = \alpha_0 (\ln x - u_0) \quad (4-6)$$

La expresión (4-5) es la función de Gumbel aplicada a una transformación logarítmica de los datos (4-6). Se comprueba entonces que, para efectuar un ajuste por medio de una función de Fréchet, sólo es necesario seguir el procedimiento visto en el Capítulo 3 aplicado al **logaritmo natural de los datos originales**.

Para grandes valores de α_0 , del orden de 50, la función de Fréchet se aproxima notablemente a la de Gumbel (8). Además, de acuerdo a lo visto en el punto 2.3.2, para grandes períodos de retorno, la variable aleatoria resulta ser aproximadamente proporcional a una función potencial de dicho período.

$$x \approx \frac{1}{\Theta} T^{1/\alpha_0} \quad (4-7)$$

Esto significa que las probabilidades para valores elevados de la variable aleatoria decrecen más lentamente que en el caso de cualquier función exponencial. Este comportamiento puede constituir una ventaja dado que las predicciones resultantes de aplicar la distribución de Gumbel adolecen a veces de errores por defecto.

4.1.1 Parámetros estadísticos

Adoptando la (4-5) como expresión válida para la función de distribución de Fréchet, todos los parámetros estadísticos determinados en el punto 3.1 para la variable reducida serán válidos, en tanto que no lo serán aquellos valores encontrados para la variable aleatoria debido a la distinta relación matemática que las vincula.

El valor medio y la desviación estándar pueden calcularse aplicando las propiedades (1-24) y (1-30) a los parámetros de la variable reducida. Se recuerda que:

$$\bar{y} = \gamma \quad \sigma_y = \frac{\pi}{\sqrt{6}}$$

De donde surgen:

$$\overline{x_{\log}} = \frac{\gamma}{\alpha_0} - \ln \Theta \quad (4-8)$$

$$\sigma_{\log} = \frac{\pi}{\sqrt{6} \alpha_0} \quad (4-9)$$

Como se aclaró anteriormente, ambos parámetros se refieren al logaritmo natural de los datos, de acuerdo con las expresiones (4-1) y (4-2).

La abscisa de la variable aleatoria para el modo de la distribución es:

$$x_{(\max)} = \frac{1}{\Theta} \quad (4-10)$$

Mientras que la correspondiente a la mediana es:

$$\ln x_{(1/2)} = -\frac{1}{\alpha_0} \ln (\ln 2) - \ln \Theta \quad (4-11)$$

4.1.2 Recta de ajuste

Todos los pasos a seguir descritos para la distribución de Gumbel referentes a ordenamiento de datos, asignación de probabilidades, valores de la variable reducida, graficado y cálculo de los coeficientes de regresión son válidos para la distribución de Fréchet aunque aplicados a los **logaritmos naturales de los datos**. Al momento de efectuar predicciones, debe utilizarse la expresión (4-6).

Los puntos pueden ser representados sobre papel de Gumbel en el que se haya reemplazado la escala de ordenadas (variable aleatoria) por una escala logarítmica. Procediendo de esta forma la función de ajuste de Fréchet resultará una recta. A los fines de comparar ambas distribuciones puede emplearse directamente el papel de Gumbel en el cual la función de Fréchet se apartará de la recta a medida que los períodos de retorno se incrementan, prediciendo valores mayores de la variable aleatoria.

El método de Lieblein, tal como fue descrito en el capítulo anterior, es también aplicable en forma directa al logaritmo de los datos (9).

En la columna (6) de la Tabla 3-2 se aparecen los logaritmos de los datos en tanto que en la Tabla 4-1 se presentan los parámetros estadísticos de los logaritmos de los datos.

Para una correcta aplicación de la distribución de Fréchet es aconsejable efectuar una regresión directa de los datos originales mediante una función exponencial. Aplicando ese criterio, se realizó una regresión ortogonal arribándose a los siguientes resultados:

$$\frac{1}{\alpha_0} = 0,134 \quad u_0 = 0,661$$

4.1.3 Intervalos de confianza

Para el cálculo de los intervalos de confianza deben adaptarse los métodos propuestos en el capítulo anterior a una escala logarítmica. Esto generará intervalos asimétricos alrededor de los valores de predicción (6) (9).

En la Tabla 4-2 resume los resultados de aplicar el método de Venier-Bernon a los datos logarítmicos para luego transformar nuevamente los resultados. Los intervalos son, como se ve, más amplios que los mostrados en la Tabla 3-12.

TABLA 4.2: PREDICCIONES E INTERVALOS DE CONFIANZA DEL 95% PARA LOS DATOS DE TABLA 3.2 (FUNCION DE FRECHET)			
T (años)	x_{MIN} (m)	x_T (m)	x_{MAX} (m)
10	2,41	2,62	3,00
20	2,60	2,89	3,40
50	2,85	3,27	4,01
100	3,07	3,60	4,51
200	3,30	3,95	5,10
500	3,63	4,47	5,98
1.000	3,90	4,90	6,76

Los valores de incertidumbre, tal como fueran definidos anteriormente son:

$$T = 100 \text{ años} \quad \varepsilon = +25\%; -15\%$$

$$T = 1.000 \text{ años} \quad \varepsilon = +38\%; -20\%$$

4.1.4 Pruebas de hipótesis

La prueba del chi cuadrado puede llevarse a cabo en forma similar al caso de Gumbel teniendo en cuenta la necesidad de aplicarla a los logaritmos de los datos. En la Tabla 4.3 se resumen los resultados de los cálculos para intervalos idénticos a los tomados en la Tabla 3.13. Dado que no se modificaron ni los grados de libertad ni el nivel de significación de la décima, el valor de chi cuadrado crítico tampoco varió.

TABLA 4.3 PRUEBA DE CHI CUADRADO APLICADA A LOS DATOS DE LA TABLA 3.2 AJUSTADOS POR MEDIO DE UNA FUNCION DE FRECHET			
x (m)	f	n p_i	(f - n p_i)² / n p_i
< 1,5	0	0,03	0,03
1,51 - 1,80	5	8,26	1,29
1,81 - 2,00	17	14,23	0,54
2,01 - 2,20	18	11,57	3,57
2,21 - 2,40	3	7,02	2,30
> 2,40	7	8,92	0,41
	50	50	8,14
para p = 0,005 → $\chi^2_p = 7,815$ r = 6 - 1 - 2 = 3 grados de libertad			

Puesto que el valor de χ^2 obtenido es igual a 8,14 correspondiente a una probabilidad de 4,32%, es decir algo inferior al 5% buscado, la prueba caería en una zona de incertidumbre en la que se hace difícil tomar una decisión en cuanto a su validez. De todas formas es conveniente recordar lo expresado en el punto 3.4 en lo referente a la confiabilidad de estos ensayos.

TABLA 4.4 PRUEBA DE KOLMOGOROF-SMIRNOF APLICADA A LOS DATOS DE LA TABLA 3.2 AJUSTADOS POR MEDIO DE UNA FUNCION DE FRECHET				
x (m)	f_i / n	Σ (f_i / n)	Φ_i(x)	Δ_i
< 1,80	0,10	0,10	0,18	0,08
1,81 - 2,00	0,34	0,44	0,45	0,01
2,01 - 2,20	0,36	0,80	0,68	0,12
2,21 - 2,40	0,06	0,86	0,82	0,04
> 2,40	0,14	1,00	1,00	0,00
para p = 0,05 → z _p = 1,36				

Para el caso de la prueba de Kolmogorof-Smirnof (Tabla 4.4), la diferencia máxima es de 0,12, en tanto que el parámetro correspondiente a un nivel de significación del 0,05

tomado de la Tabla 1.4 es 1,36. La aplicación de la prueba conduce al siguiente resultado:

$$\sqrt{n} \Delta_{\max} = 0,12 \sqrt{50} = 0,85 < z_p = 1,36$$

Lo que implica que la hipótesis del ajuste por Fréchet debería aceptarse.

4.2 DISTRIBUCION DE GALTON

Tal como se expresó en 2.3.4, cuando el logaritmo de la variable aleatoria responde a una distribución normal, la variable original sigue una log-normal o distribución de Galton, cuya función de frecuencia es:

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma_z} e^{-\frac{1}{2} \left(\frac{z - \bar{z}}{\sigma_z} \right)^2} \quad (4-12)$$

La relación entre la variable auxiliar z y la variable original x es logarítmica.

$$z = \ln x \quad \bar{z} = \overline{x_{\log}} \quad \sigma_z = \sigma_{\log} \quad (4-13)$$

La marcada asimetría positiva de esta función la hace especialmente apta para el ajuste de valores extremos (Figura 2.1). Su campo de variación se extiende entre 0 y $+\infty$ y, para grandes períodos de retorno, la variable aleatoria puede aproximarse por la siguiente expresión:

$$x = C e^{\sqrt{\ln T}}$$

La función de Galton converge lentamente hacia la función de Gumbel.

4.2.1 Parámetros estadísticos

A partir de las expresiones (4-12) y (4-13) se obtienen las siguientes ecuaciones para el valor medio y la desviación estándar respectivamente.

$$\bar{x} = e^{\bar{z} + \sigma_z^2 / 2} \quad (4-14)$$

$$\sigma = \bar{x} \sqrt{e^{\sigma_z^2} - 1} \quad (4-15)$$

Dividiendo miembro a miembro puede calcularse el coeficiente de variación.

$$C_v = \frac{\sigma}{\bar{x}} = \sqrt{e^{\sigma_z^2} - 1} \quad (4-16)$$

El coeficiente de asimetría es función del coeficiente de variación.

$$\gamma_1 = 3 C_v + C_v^3 \quad (4-17)$$

Por último, la mediana es:

$$x_{(1/2)} = e^{\bar{x}} \quad (4-18)$$

4.2.2 Recta de ajuste

Algunos autores aconsejan utilizar las directamente expresiones (4-14) a (4-18) para obtener la función de ajuste de Galton a partir de los datos experimentales. Sin embargo, es preferible modificar ligeramente su presentación matemática a fin de disponer de un procedimiento de trabajo similar al aplicado para las otras funciones.

Para ello se empleará una función de distribución normal y una variable reducida que sea una transformación lineal del logaritmo de los datos.

$$\Phi(x) = \frac{1}{2\pi} \int_{-\infty}^z e^{-u^2/2} du \quad (4-19)$$

$$u = \frac{\ln x - a}{b} \quad (4-20)$$

De existir un ajuste perfecto, las constantes a y b deberían coincidir con el valor medio y la desviación estándar logarítmicos respectivamente.

$$a = \overline{x_{\log}} \quad b = \sigma_{\log} \quad (4-21)$$

El procedimiento propuesto para el ajuste, tomando como referencia, una vez más, la Tabla 3.2, consiste en determinar, a partir de las probabilidades asignadas por la fórmula de Weibull (columna 4) los valores correspondientes de la variable reducida de la función normal (4-19). Esta operación se realiza mediante el uso de tablas para función normal buscando la abscisa que encierre un área igual a la probabilidad de Weibull. Para ello puede emplearse la Tabla A1 del apéndice. A continuación se efectúa una regresión entre estos valores y los logaritmos de la variable aleatoria.

En la Tabla 4.5 se presentan los valores de la variable reducida para diversos períodos de retorno para una función normal. La tabla fue confeccionada mediante las expresiones (2-2), (2-3) y (4-19) y resulta práctica a la hora de extrapolar.

TABLA 4.5: FUNCION DE GAUSS		
VARIABLE REDUCIDA Y PERIODO DE RETORNO		
Probabilidad (p)	Período (T)	Variable reducida (u)
0,500	2	0,000
0,200	5	0,842
0,100	10	1,282
0,050	20	1,645
0,020	50	2,054
0,010	100	2,327
0,005	200	2,576
0,002	500	2,879
0,001	1000	3,091

En la tabla 4.6 del próximo punto se resumen los resultados de las predicciones efectuadas mediante la función de Galton. Allí puede comprobarse que la misma tiende a subestimar los valores de la variable aleatoria para grandes períodos de retorno.

De manera similar a lo sucedido con la función de Fréchet, los coeficientes de la recta se obtuvieron por medio de una regresión exponencial directa lo que tiende a disminuir errores al evitar tomar logaritmos previamente. También se adoptó una regresión ortogonal cuyo coeficiente de correlación fue de 0,934, resultado del buen ajuste para bajos períodos de recurrencia.

Puede verificarse la validez de las igualdades (4-21) teniendo en cuenta que la pendiente de la recta de regresión es 0,167 en tanto que la desviación estándar logarítmica es igual a 0,158. Por su parte la ordenada al origen es 0,735 idéntica al valor medio logarítmico. Estas concordancias hacen posible el empleo de la siguiente expresión:

$$\ln x = \overline{x_{\log}} + \sigma_{\log} u \quad (4-22)$$

La ligera discrepancia entre la desviación estándar logarítmica y la pendiente se debe al uso de una regresión ortogonal.

4.2.3 Intervalos de confianza

El trazado de los intervalos de confianza podría efectuarse en forma análoga al criterio propuesto por Gumbel acerca del trazado de líneas de control paralelas a la recta de regresión. En este caso corresponderían a $\pm\sigma$ o bien a $\pm 2\sigma$ respecto del ajuste. Para un 95% de confianza se tendría:

$$\ln x = a + b u \pm 2 \sigma_{\log} \quad (4-23)$$

La aplicación de ésta da origen a los resultados de la Tabla 4.6

TABLA 4.6: PREDICCIONES E INTERVALOS DE CONFIANZA DEL 95% PARA LOS DATOS DE TABLA 3.2 (FUNCION DE GALTON)			
T (años)	x_{MIN} (m)	x_T (m)	x_{MAX} (m)
10	1,89	2,58	3,54
20	2,00	2,75	3,76
50	2,15	2,94	4,03
100	2,25	3,08	4,22
200	2,34	3,21	4,40
500	2,46	3,38	4,63
1.000	2,55	3,50	4,79

Los valores de incertidumbre, tal como fueran definidos anteriormente son:

$$T = 100 \text{ años} \quad \varepsilon = +37\%; -27\%$$

$$T = 1.000 \text{ años} \quad \varepsilon = +37\%; -27\%$$

Los intervalos son asimétricos y conservan sus valores relativos para todo el campo de existencia de la variable.

4.2.4 Pruebas de hipótesis

La Tabla 4.7 presenta los resultados de haber aplicado la prueba de chi cuadrado a la distribución de Galton.

El resultado obtenido ($\chi^2 = 12,41$) es superior al valor crítico para 5% de nivel de significación y corresponde a una probabilidad de excedencia de 0,61% que implicaría un rechazo de la hipótesis de correcto ajuste de los datos experimentales por medio de la función log-normal.

TABLA 4.7 PRUEBA DE CHI CUADRADO APLICADA A LOS DATOS DE LA TABLA 3.2 AJUSTADOS POR MEDIO DE UNA FUNCION DE GALTON			
x (m)	f	n p _i	(f - n p _i) ² / n p _i
< 1,5	0	0,90	0,90
1,51 - 1,80	5	7,78	0,99
1,81 - 2,00	17	11,00	3,27
2,01 - 2,20	18	12,28	2,66
2,21 - 2,40	3	9,01	4,01
> 2,40	7	9,33	0,58
	50	50	12,41
para p = 0,005 → $\chi^2_p = 7,815$ r = 6 - 1 - 2 = 3 grados de libertad			

La Tabla 4.8 muestra los resultados de aplicar la prueba de hipótesis de Kolmogorof-Smirnof.

TABLA 4.8 PRUEBA DE KOLMOGOROF-SMIRNOF APLICADA A LOS DATOS DE LA TABLA 3.2 AJUSTADOS POR MEDIO DE UNA FUNCION DE GALTON				
x (m)	f _i / n	Σ (f _i / n)	Φ _i (x)	Δ _i
< 1,80	0,10	0,10	0,19	0,09
1,81 - 2,00	0,34	0,44	0,40	0,04
2,01 - 2,20	0,36	0,80	0,63	0,17
2,21 - 2,40	0,06	0,86	0,80	0,06
> 2,40	0,14	1,00	1,00	0,00
para p = 0,05 → z _p = 1,36				

Lo que conduce al siguiente resultado de la prueba.

$$\sqrt{n} \Delta_{\max} = 0,17 \sqrt{50} = 1,23 < z_p = 1,36$$

Es decir que la hipótesis del ajuste de los datos por medio de la función log-normal debería de ser aceptada.

4.3 DISTRIBUCION LOG PEARSON

Esta distribución, que también se aplica al logaritmo de los datos, es muy utilizada porque permite obtener resultados de aceptable confiabilidad cuando se efectúan predicciones con grandes períodos de retorno.

La expresión de su función de frecuencia es la siguiente:

$$f(x) = \frac{\lambda^\beta (y - \varepsilon)^{\beta-1} e^{-(y-\varepsilon)}}{x \Gamma(\beta)} \quad (4 - 24)$$

Para: $y = \ln x \geq \varepsilon$

Los coeficientes están definidos por las siguientes igualdades:

$$\beta = \left(\frac{2}{C_{\log}} \right)^2 \quad (4-25)$$

$$\lambda = \frac{\sigma_{\log}}{\sqrt{\beta}} \quad (4-26)$$

$$\varepsilon = \overline{x_{\log}} - \sigma_{\log} \sqrt{\beta} \quad (4-27)$$

Donde C_{\log} es el coeficiente de asimetría del logaritmo de los datos originales y $\Gamma()$ es la función factorial. Se recuerda que:

$$\Gamma(n + 1) = n \Gamma(n)$$

El campo de variación de la función log-Pearson está comprendido entre $-\infty$ y ε con $\varepsilon < \infty$. Para grandes valores de T , la variable aleatoria crece con el logaritmo del período de retorno, tal como ocurre con la distribución de Gumbel (8).

4.3.1 Recta de ajuste

Dado que la función log-Person es analíticamente muy complicada, resulta más práctico el empleo de la fórmula de Chow (2 - 4) transformada logarítmicamente (5) o sea la expresión (4-22) que se reproduce a continuación.

$$\ln x = \overline{x_{\log}} + k \sigma_{\log} \quad (4-28)$$

Los coeficientes, como ya se vio, representan el valor medio y la desviación estándar del logaritmo de los datos. El factor de frecuencia k es función del período de retorno y del coeficiente de asimetría de los datos y puede obtenerse en la Tabla 4.9, tomada de la bibliografía (5).

TABLA 4.9: CONSTANTE k PARA LA DISTRIBUCION LOG-PEARSON								
Coefficiente de asimetría	PERIODO DE RETORNO EN AÑOS							
	1,0101	1,25	2	5	10	25	50	100
	PROBABILIDAD EN PORCENTAJE							
	99	80	50	20	10	4	2	1
3,0	-0,667	-0,636	-0,396	0,420	1,180	2,278	3,152	4,051
2,8	-0,714	-0,666	-0,384	0,460	1,210	2,275	3,114	3,973
2,6	-0,769	-0,696	-0,368	0,499	1,238	2,267	3,071	3,889
2,4	-0,832	-0,725	-0,351	0,537	1,262	2,256	3,023	3,800
2,2	-0,905	-0,752	-0,330	0,574	1,284	2,240	2,970	3,705
2,0	-0,990	-0,777	-0,307	0,609	1,302	2,219	2,912	3,605
1,8	-1,087	-0,799	-0,282	0,643	1,318	2,193	2,848	3,499
1,6	-1,197	-0,817	-0,254	0,675	1,329	2,163	2,780	3,388
1,4	-1,318	-0,832	-0,225	0,705	1,337	2,128	2,706	3,271
1,2	-1,449	-0,844	-0,195	0,732	1,340	2,087	2,626	3,149
1,0	-1,588	-0,852	-0,164	0,58	1,340	2,043	2,542	3,022
0,8	-1,733	-0,856	-0,132	0,780	1,336	1,993	2,453	2,891
0,6	-1,880	-0,857	-0,099	0,800	1,328	1,939	2,359	2,755
0,4	-2,029	-0,855	-0,066	0,816	1,317	1,880	2,261	2,625
0,2	-2,178	-0,850	-0,033	0,830	1,301	1,818	2,159	2,472
0	-2,326	-0,842	0,000	0,842	1,281	1,751	2,054	2,326
-0,2	-2,472	-0,830	0,033	0,850	1,258	1,680	1,945	2,178
-0,4	-2,615	-0,816	0,066	0,855	1,231	1,606	1,834	2,029
-0,6	-2,755	-0,800	0,099	0,857	1,200	1,528	1,720	1,880
-0,8	-2,891	-0,780	0,132	0,856	1,166	1,448	1,606	1,733
-1,0	-3,022	-0,758	0,164	0,852	1,128	1,366	1,492	1,588
-1,2	-3,149	-0,732	0,195	0,844	1,086	1,282	1,379	1,449
-1,4	-3,271	-0,705	0,225	0,832	1,041	1,198	1,270	1,318
-1,6	-3,388	-0,675	0,254	0,817	0,994	1,116	1,166	1,197
-1,8	-3,499	-0,643	0,282	0,799	0,945	1,035	1,069	1,087
-2,0	-3,605	-0,609	0,307	0,777	0,895	0,959	0,980	0,990
-2,2	-3,705	-0,574	0,330	0,752	0,844	0,888	0,900	0,905
-2,4	-3,800	-0,537	0,351	0,725	0,795	0,823	0,830	0,832
-2,6	-3,889	-0,499	0,368	0,696	0,747	0,764	0,768	0,769
-2,8	-3,973	-0,460	0,384	0,666	0,702	0,712	0,714	0,714
-3,0	-4,051	-0,420	0,396	0,636	0,660	0,666	0,666	0,667

Si se necesitan valores que no figuran en la tabla puede recurrirse a la expresión matemática (4-29) que brinda una excelente aproximación.

$$k \approx z + (z^2 - 1) \frac{C_{\log}}{6} + \frac{1}{3} (z^3 - 6z) \left(\frac{C_{\log}}{6} \right)^2 - (z^2 - 1) \left(\frac{C_{\log}}{6} \right)^3 + z \left(\frac{C_{\log}}{6} \right)^4 + \frac{1}{3} \left(\frac{C_{\log}}{6} \right)^5 \quad (4-29)$$

Si bien esta expresión puede parecer compleja es fácilmente manejable por medio de una planilla de cálculo. La variable normalizada de Gauss está representada por z en la (4-29).

El método práctico implica el cálculo de las probabilidades de que la variable aleatoria sea igual o inferior a cada período de retorno considerado.

$$P[x \leq x(T)] = 1 - \frac{1}{T} \quad (4-30)$$

A continuación se busca en tablas de función de frecuencia normal el valor de abscisa (z) que deje a su izquierda un área igual a la probabilidad (4-30). Con esa información y con el coeficiente de asimetría del logaritmo de los datos se utiliza la Tabla 4.9 o la expresión (4-29) para determinar k para cada período de retorno. Por último la (4-28) se emplea para calcular el valor del logaritmo de la variable aleatoria.

4.3.2 Intervalos de confianza

El trazado de los intervalos de confianza podría efectuarse aplicando algunos de los criterios propuestos anteriormente. En la Tabla 4.10 se exponen los resultados obtenidos por medio del método de Bernier-Veron.

TABLA 4.10: PREDICCIONES E INTERVALOS DE CONFIANZA DEL 95% PARA LOS DATOS DE TABLA 3.2 (FUNCION LOG-PEARSON)			
T (años)	x_{MIN} (m)	x_T (m)	x_{MAX} (m)
10	2,36	2,57	2,95
20	2,53	2,81	3,30
50	2,73	3,13	3,83
100	2,90	3,39	4,26
200	3,07	3,67	4,74
500	3,30	4,06	5,44
1.000	3,40	4,38	6,05

Los valores de incertidumbre, tal como fueran definidos anteriormente son:

T = 100 años $\varepsilon = +26\%; -14\%$

T = 1.000 años $\varepsilon = +38\%; -22\%$

4.3.3 Pruebas de hipótesis

La Tabla 4.11 presenta los resultados de haber aplicado la prueba de chi cuadrado a la distribución log-Pearson. El resultado obtenido ($\chi^2 = 7,66$) es inferior al valor crítico para 5% de nivel de significación por lo que la hipótesis de ajuste de los datos experimentales por medio de la función log-Pearson debería aceptarse como válida.

TABLA 4.11 PRUEBA DE CHI CUADRADO APLICADA A LOS DATOS DE LA TABLA 3.2 AJUSTADOS POR MEDIO DE UNA FUNCION LOG-PEARSON			
x (m)	f	n p_i	(f - n p_i)² / n p_i
< 1,5	0	< 0,000001	< 0,000001
1,51 - 1,80	5	8,18	1,23
1,81 - 2,00	17	14,83	0,32
2,01 - 2,20	18	11,56	3,58
2,21 - 2,40	3	7,00	2,29
> 2,40	7	8,43	0,24
	50	50	7,66
para p = 0,05 → $\chi^2_p = 7,815$ r = 6 - 1 - 2 = 3 grados de libertad			

La Tabla 4.12 muestra los resultados de aplicar la prueba de hipótesis de Kolmogorof-Smirnof.

TABLA 4.12 PRUEBA DE KOLMOGOROF-SMIRNOF APLICADA A LOS DATOS DE LA TABLA 3.2 AJUSTADOS POR MEDIO DE UNA FUNCION LOG-PEARSON				
x (m)	f_i / n	Σ (f_i / n)	Φ_i(x)	Δ_i
< 1,80	0,10	0,10	0,16	0,06
1,81 - 2,00	0,34	0,44	0,46	0,02
2,01 - 2,20	0,36	0,80	0,69	0,11
2,21 - 2,40	0,06	0,86	0,83	0,03
> 2,40	0,14	1,00	1,00	0,00
para p = 0,05 → $z_p = 1,36$				

Lo que conduce al siguiente resultado de la prueba.

$$\sqrt{n} \Delta_{\max} = 0,11 \sqrt{50} = 0,78 < z_p = 1,36$$

Es decir que este criterio también conduce a la aceptación de la hipótesis del ajuste de los datos por medio de la función log-Pearson.

5. CONCLUSIONES SOBRE LAS FUNCIONES DE AJUSTE

La Figura 5.1 muestra en forma comparativa las representaciones gráficas de las cuatro distribuciones de valores extremos presentadas en este trabajo: Gumbel, aplicada directamente sobre los datos experimentales y Fréchet, Galton y log-Pearson aplicadas sobre sus logaritmos.

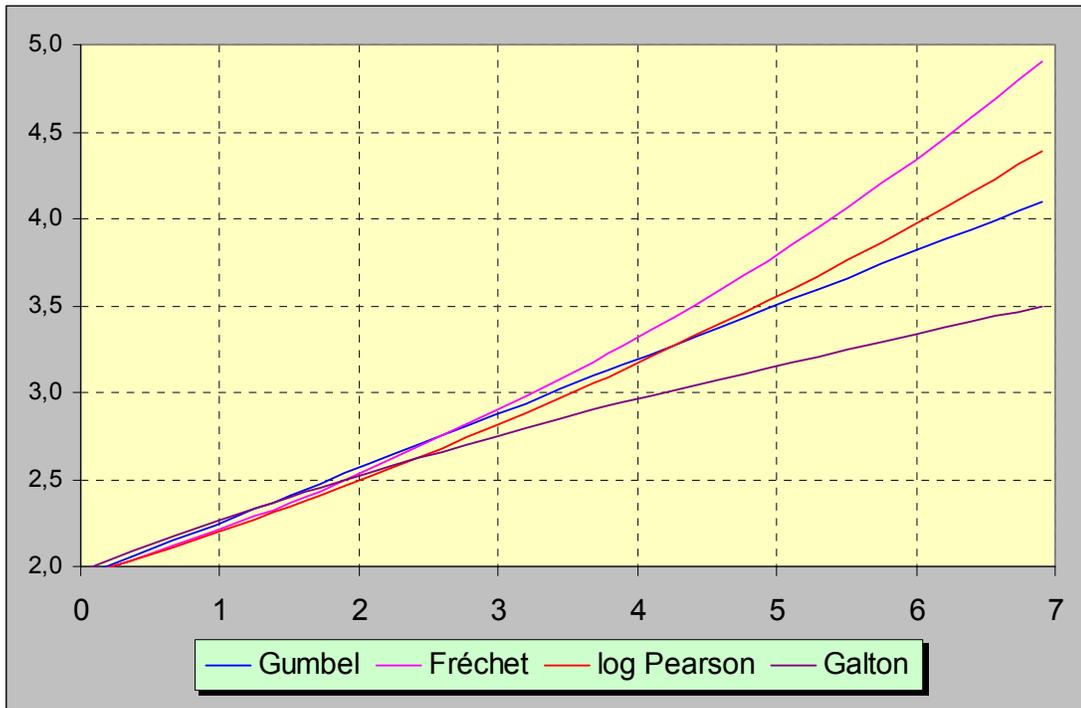


Figura 5.1: Comparación entre distribuciones de valores extremos

Los resultados expuestos corresponden a la aplicación de las cuatro funciones al ajuste de los valores presentados en la Tabla 3.2, tal como se fue desarrollando a lo largo de este trabajo. En ordenadas se representan cotas hidrométricas (en metros) y en abscisas, la variable reducida de Gumbel. Por lo tanto esta última función aparecerá como una recta y puede ser tomada como referencia.

A efectos de tener una idea más clara del significado de los valores de abscisas, puede recurrirse a la Tabla 2.1 en la que puede comprobarse que el máximo representado corresponde a un período de retorno de 1.000 años ($y = 6,9$).

La Tabla 5.1 resume los valores de predicción para las cuatro funciones.

TABLA 5.1: PREDICCIONES PARA LOS DATOS DE TABLA 3.2					
T (años)	y	Gumbel	Fréchet	Galton	log-Pearson
10	2,250	2,64	2,62	2,58	2,57
20	2,970	2,87	2,89	2,75	2,81
50	3,902	3,16	3,27	2,94	3,13
100	4,600	3,38	3,60	3,08	3,39
200	5,296	3,60	3,95	3,21	3,67
500	6,214	3,88	4,47	3,38	4,06
1.000	6,907	4,10	4,90	3,50	4,38

En la Figura 5.1 no se graficaron los datos experimentales dado que todas las funciones, dentro de sus intervalos de confianza, logran ajustarlos aceptablemente y con un buen coeficiente de correlación. También puede verse que hasta un período de 10 años las distribuciones difieren entre ellas sólo unos pocos centímetros para comenzar a separarse a partir de este punto a medida que los períodos se incrementan.

La función de Fréchet se aparta mucho de la función de Gumbel lo que, según se dijo anteriormente, podría corregir la tendencia de la última a cometer errores por defecto para elevados períodos de retorno. Sin embargo, daría la impresión de que la distribución de Fréchet predice valores excesivamente altos (un 20% por arriba de Gumbel para 1.000 años). En este contexto, log-Pearson podría constituir un compromiso razonable. Por su parte Galton suele predecir valores demasiado pequeños.

Como regla general, el desarrollo de programas de computación específicos o el empleo eficiente de planillas de cálculo para el almacenamiento y tratamiento de datos, el cálculo de parámetros estadísticos y el ajuste mediante diversas funciones es de inestimable ayuda para el tratamiento de valores extremos.

El error propio de los datos experimentales, la escasa longitud de los registros (debe tenerse en cuenta que con 50 datos o, excepcionalmente algo más 100, pretenden efectuarse pronósticos milenarios) y la falta de algunos valores intermedios hace que el empleo de técnicas de ajuste muy refinadas sea un lujo innecesario aunque teóricamente puedan ser excelentes herramientas.

La obtención de la recta de regresión por cuadrados mínimos, ya sea aplicada a los datos históricos o a sus logaritmos, aparece como el método más rápido y conveniente.

En lo referente a la calidad del ajuste, parece no haber un criterio totalmente adecuado como ocurre en el análisis estadístico clásico de valores medios. En efecto, si se comparan las cuatro funciones aplicadas a un mismo conjunto de datos, es

posible que todas estén caracterizadas por elevados coeficientes de correlación cuyos valores están determinados, principalmente, por los puntos ubicados en la zona izquierda del ajuste. Sin embargo, es seguro que todas darán pronósticos distintos para altos períodos de recurrencia, tal como sucede con el caso expuesto en la Figura 5.1 y en la Tabla 5.1.

La prueba del chi cuadrado no parece ser un indicador totalmente adecuado para la aceptación o el rechazo de hipótesis. Como se vio más arriba sólo resultó positiva en el caso de la función log-Pearson y por un margen muy estrecho. Escasez de datos, límites arbitrarios de intervalo, pocos intervalos de clase y, por lo tanto de grados de libertad, son factores que contribuyen a la ineficiencia de esta dócima para el caso de valores extremos. Su utilización cuidadosa puede servir de ayuda para comparar dos hipótesis distintas aplicadas al mismo conjunto de datos pero, aún así, resulta dependiente de factores subjetivos y arbitrarios.

En lo concerniente a la prueba de Kolmogorof-Smirnof, se trata de una prueba no paramétrica que no es la más apta para este tipo de problemas y está considerada, en líneas generales, como menos eficiente que el chi cuadrado.

El trazado de los intervalos de confianza permite contar con un método adicional de evaluación, en cuanto al hecho que los puntos experimentales caigan o no dentro de la región de confianza. Por lo general, los intervalos del 95% aparecen como excesivamente amplios, en especial en el caso de las líneas de control de Gumbel, cuyas bases teóricas parecerían ser no totalmente sólidas. Los mejores resultados se consiguen con la delimitación dada por los intervalos del 68% por los métodos de Kaczmarek y Bernier-Veron.

Debe también insistirse en el riesgo que implica el trabajar con un cierto período de retorno dado por la Tabla 2.3 al que debe sumarse la probabilidad de que la hipótesis no sea la correcta. Esta última se incrementa al tomar períodos más elevados lo que, por su parte, aumenta el riesgo.

Si se pretende efectuar un pronóstico de máxima aplicado a factores naturales (vientos, nieve, crecidas) para el diseño de un emprendimiento y para ello se elige un período de retorno elevado a fin de asegurarse de que no sea superado durante la vida útil de la obra, no sólo se incrementarán los costos de construcción sino que, además, se aumentará la probabilidad de cometer un error en la extrapolación.

Por lo antedicho, si bien los métodos probabilísticos constituyen una excelente herramienta para evaluar valores extremos, resulta conveniente comparar sus resultados con los datos provenientes de otras técnicas. En el caso particular de un estudio hidrológico, como el que se expondrá en el próximo capítulo, dos alternativas posibles son los análisis determinístico y estocástico.

El análisis determinístico, se basa en el planteo de un conjunto de ecuaciones matemáticas en las que intervienen todos los factores que puedan influir en el evento bajo análisis (por ejemplo una crecida extraordinaria) y que conducen al desarrollo de un modelo matemático de predicción. Maximizando las variables hasta alcanzar valores aceptablemente altos puede obtenerse una cota máxima probable y razonable del evento en cuestión la que podrá ser comparada con la surgida del análisis

estadístico de valores extremos. Por su parte el análisis estocástico recurre a las series temporales de datos para la elaboración de un modelo probabilístico de simulación utilizando, por ejemplo, cadenas de Markov para generar eventos naturales de manera sintética. Haciendo correr el modelo numerosas veces en computadora pueden generarse miles de años de datos artificiales que pueden manipularse luego por medio de herramientas estadísticas.

Como conclusión final puede decirse que la probabilidad de cometer errores de predicción disminuirá en la medida en que se disponga de resultados surgidos de diferentes mecanismos de análisis que puedan compararse y compatibilizarse.

6. APLICACIONES

Las aplicaciones de funciones de distribución de valores extremos están orientadas principalmente a predecir fenómenos meteorológicos tales como velocidades de vientos, nivel de precipitaciones y caudales de ríos aunque, como ya se mencionara anteriormente, existen campos diversos en los que el empleo de estas técnicas puede constituir una gran ayuda. En este capítulo se describirá el caso concreto de la evaluación de cotas hidrométricas extremas, máximas y mínimas, a lo largo del río Paraná, entre Corrientes y Buenos Aires, República Argentina, relacionado con el emplazamiento de la Central Nuclear en Atucha I (11). La Figura 6.1 muestra la zona abarcada por el estudio.



Figura 6.1: Zona comprendida en el estudio

Con el objeto de predecir avenidas y estiajes se tomaron como variables las cotas hidrométricas en varias estaciones de aforo en lugar de trabajar con caudales. Esto se debió a que el Paraná de las Palmas, o sea la parte sur del Paraná, está sometida a

un régimen mixto de mareas y vientos para el cual no es posible establecer un relación biunívoca entre cotas y caudales. A medida que el punto de medición se traslada aguas abajo, el régimen de mareas prevalece hasta dominar completamente en el Río de la Plata. Al norte de Rosario, en cambio, hubiera sido posible trabajar con caudales pero se prefirió continuar operando con alturas para mantener la coherencia del estudio.

Las estaciones de aforo existentes fueron clasificadas en base a un análisis de consistencia comparativo. Se seleccionaron, en definitiva, aquellas que contaban con registros más extensos y las que presentaban menos vacíos en su historial y mayor confiabilidad en la información reunida. En algunos casos los registros llegaban a casi 100 años ininterrumpidos de mediciones y, en todos ellos, eran superiores a los 45 años.

De las series de crecidas y bajantes diarias se tomaron los valores extremos anuales constituyendo, de esta forma, seis conjuntos de datos o espacios muestrales correspondientes al mismo número de estaciones seleccionadas en base a las cuales se haría el ajuste. Durante la primera parte de este estudio se trabajó con las siguientes estaciones, de norte a sur:

- Rosario
- San Pedro
- Las Palmas-Guazú
- Zárate
- Las Palmas- Las Rosas
- Riachuelo

Posteriormente se incorporaron las siguientes:

- Corrientes
- Paraná

Otras estaciones evaluadas previamente fueron Ibicuy, Carabelitas, Baradero, Canal del Este, Guazú-Brazo Largo y Las Palmas Capitán.

Las cotas hidrométricas estaban dadas, en cada caso, en forma relativa a un nivel de referencia o "cero" local por lo que, antes de iniciar el análisis estadístico, fue necesario llevar todos los datos a una referencia única seleccionándose para ello el "cero" del Riachuelo.

En la Figura 6.2 se presenta el área donde se encuentran las estaciones de aforo empleadas durante la primera parte del estudio.

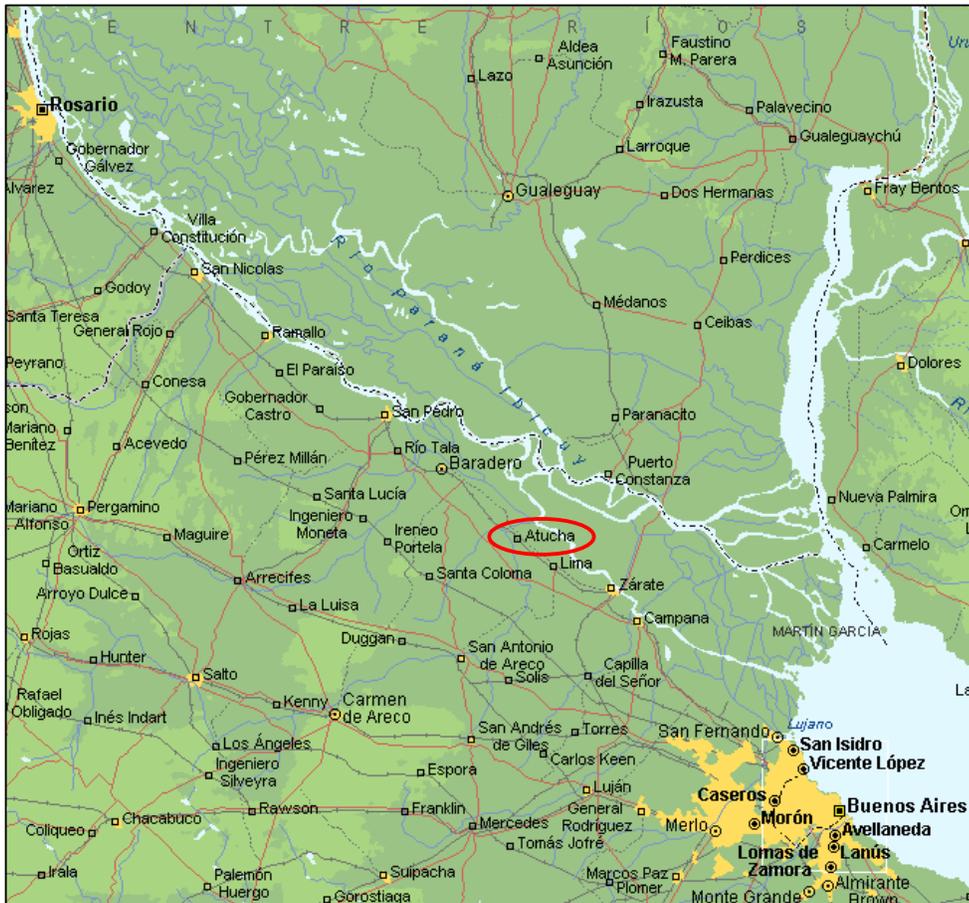


Figura 6.2: Zona de estudio durante la primera parte del trabajo

El objetivo final del trabajo expuesto era el trazado de un gráfico que representara las crecientes y bajantes con igual período de retorno en función de la ubicación de las distintas estaciones de aforo. A partir del mismo podrían hallarse por interpolación los valores correspondientes a Atucha donde se encuentran las centrales nucleares, emplazadas sobre una escarpada barranca, borde de la llanura pampeana, a 134 km al norte de la ciudad de Buenos Aires.

El conocimiento de la crecida extrema permite determinar el punto óptimo de instalación de la sala de bombas de agua para refrigeración (tomada del Paraná de las Palmas) para evitar su salida de operación ante una inundación excepcional. El máximo estiaje previsto conduce a la selección de la ubicación de las bocas de toma de agua de dichas bombas. Evidentemente, cuando menor sea la distancia entre ambos puntos más pequeña será la columna de agua y menores serán los costos.

6.1 ANALISIS DE CRECIDAS

A partir del espacio muestral correspondiente a cada estación, corregido a efectos de referirlo al mismo nivel de "cero", se ordenaron los datos en forma decreciente, se determinaron las probabilidades a través de la fórmula de Weibull y se graficaron en papel de Gumbel. En todos los casos los puntos quedaron alineados, aproximada-

mente, sobre una recta por lo que se decidió efectuar el ajuste por medio de una función de Gumbel.

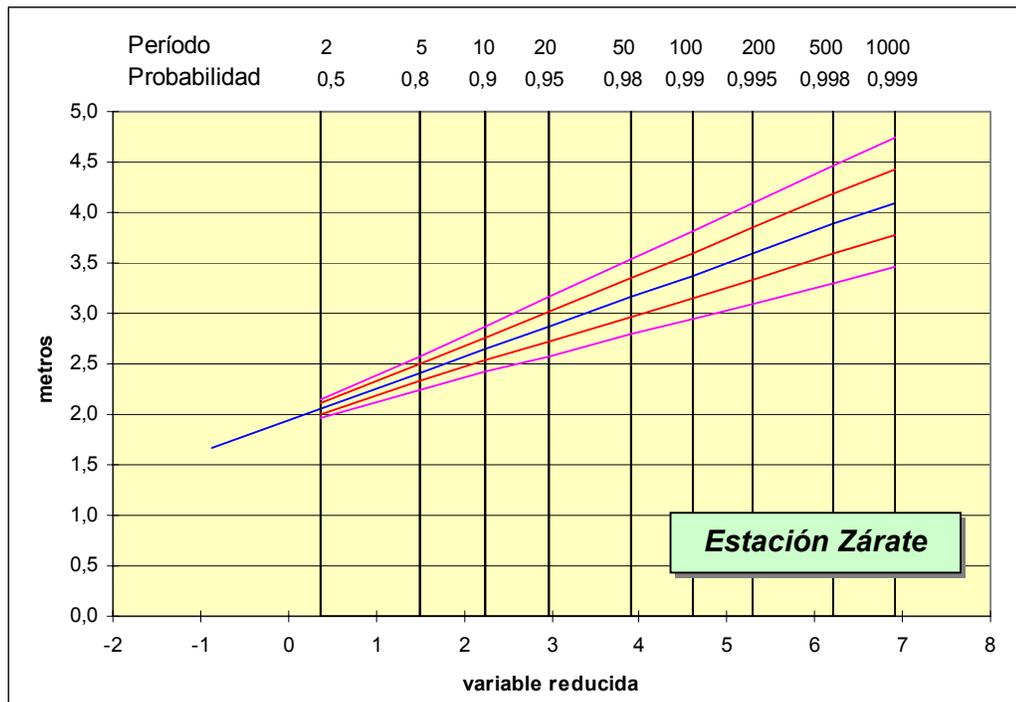


Figura 6.3: Ajuste por función de Gumbel (—) e intervalos de confianza del 68% (—) y del 95% (—)

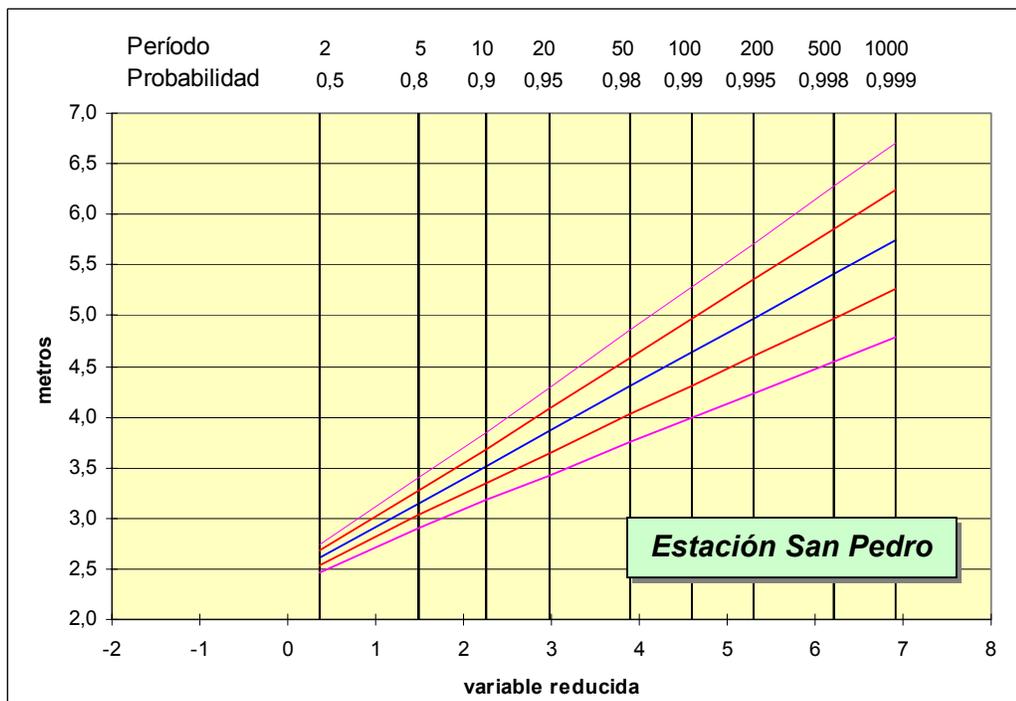


Figura 6.4: Ajuste por función de Gumbel (—) e intervalos de confianza del 68% (—) y del 95% (—)

Todos los cálculos necesarios para los ajustes se realizaron mediante un programa de computación, desarrollado por el autor, que también determina todos los parámetros estadísticos de los datos y de sus logaritmos. El mismo trabajo puede llevarse a cabo utilizando una planilla de cálculo.

En la Figura 6.3 se representa la función de ajuste de Gumbel y los intervalos de confianza de Kaczmarek para el 68% y el 95% para la estación Zárate. La Figura 6.4 muestra la misma información para San Pedro.

6.2 ANALISIS DE ESTIAJES

Para las bajantes extremas se siguió un procedimiento similar al empleado para crecidas obteniéndose una buena alineación de los puntos a lo largo de una recta para todas las estaciones ubicadas al sur de Rosario, en tanto que, para ésta y para los puntos aguas arriba se presentaron significativas diferencias. Por estas razones se decidió recurrir a la distribución de Gumbel para las primeras y buscar otra alternativa para las restantes.

Luego se efectuaron varias pruebas se comprobó que al graficar los datos provenientes de las estaciones Rosario, Paraná, Corrientes y Posadas en papel log-normal se obtenía un buen alineamiento. Dado que el coeficiente de asimetría de todos los conjuntos muestrales era próximo a cero se optó por emplear la distribución de Galton por sobre la log-Pearson que hubiera resultado mejor en caso de mayores asimetrías.

También en esta oportunidad se trabajó con un programa de computación, diseñado por el autor, para el procesamiento de la información y la obtención de los parámetros de ajuste aunque el uso de una planilla de cálculo es perfectamente válido.

Dado que hasta este momento no se ha presentado ningún ejemplo de ajuste de valores extremos mínimos, se presentará la información completa correspondiente al estudio de bajantes en la estación de aforo San Pedro.

En la Tabla 6.1 se encuentran enumeradas las mínimas anuales para la Estación Hidrométrica San Pedro, su número de orden, su probabilidad y el valor de la variable reducida correspondiente a dicha probabilidad. Nótese que la expresión de Weibull ha sido modificada teniendo en cuenta que se trata de valores extremos mínimos, estiajes en este caso.

$$p = 1 - \frac{i}{n+1} \quad (6-1)$$

El resumen de resultados se presenta en la Tabla 6-2, mientras que en la Figura 6.5 se observan los datos originales, la recta de ajuste de Gumbel y los intervalos de confianza según Kaczmarek para el 68% y para el 95%.

TABLA 6.1: ESTACION HIDROMETRICA SAN PEDRO (MINIMAS)

i	x (m)	1 - i / (n+1)	y	i	x (m)	1 - i / (n+1)	y
1	1,74	0,988	-1,478	41	0,00	0,488	0,403
2	1,24	0,975	-1,305	42	-0,03	0,475	0,440
3	1,06	0,963	-1,189	43	-0,10	0,463	0,477
4	1,05	0,950	-1,097	44	-0,14	0,450	0,514
5	0,98	0,938	-1,020	45	-0,14	0,438	0,553
6	0,90	0,925	-0,952	46	-0,16	0,425	0,592
7	0,82	0,913	-0,890	47	-0,18	0,413	0,631
8	0,68	0,900	-0,834	48	-0,20	0,400	0,672
9	0,68	0,888	-0,782	49	-0,20	0,388	0,713
10	0,62	0,875	-0,732	50	-0,20	0,375	0,755
11	0,62	0,863	-0,685	51	-0,22	0,363	0,798
12	0,60	0,850	-0,640	52	-0,23	0,350	0,842
13	0,58	0,838	-0,597	53	-0,25	0,338	0,887
14	0,58	0,825	-0,556	54	-0,34	0,325	0,934
15	0,56	0,813	-0,515	55	-0,34	0,313	0,982
16	0,50	0,800	-0,476	56	-0,34	0,300	1,031
17	0,46	0,788	-0,437	57	-0,35	0,288	1,082
18	0,44	0,775	-0,400	58	-0,40	0,275	1,134
19	0,44	0,763	-0,363	59	-0,40	0,263	1,189
20	0,40	0,750	-0,327	60	-0,40	0,250	1,246
21	0,36	0,738	-0,291	61	-0,41	0,238	1,305
22	0,32	0,725	-0,255	62	-0,42	0,225	1,367
23	0,30	0,713	-0,220	63	-0,46	0,213	1,432
24	0,30	0,700	-0,186	64	-0,50	0,200	1,500
25	0,28	0,688	-0,151	65	-0,52	0,188	1,572
26	0,28	0,675	-0,117	66	-0,57	0,175	1,648
27	0,26	0,663	-0,083	67	-0,57	0,163	1,730
28	0,20	0,650	-0,049	68	-0,60	0,150	1,817
29	0,20	0,638	-0,015	69	-0,62	0,138	1,911
30	0,18	0,625	0,019	70	-0,62	0,125	2,013
31	0,16	0,613	0,053	71	-0,66	0,113	2,126
32	0,16	0,600	0,087	72	-0,70	0,100	2,250
33	0,15	0,588	0,122	73	-0,78	0,088	2,391
34	0,10	0,575	0,156	74	-0,79	0,075	2,552
35	0,10	0,563	0,190	75	-0,82	0,063	2,740
36	0,06	0,550	0,225	76	-0,96	0,050	2,970
37	0,04	0,538	0,260	77	-0,98	0,038	3,264
38	0,00	0,525	0,295	78	-1,10	0,025	3,676
39	0,00	0,513	0,331	79	-1,30	0,013	4,376
40	0,00	0,500	0,367				

TABLA 6.2: AJUSTE DE LOS DATOS DE LA TABLA 6.1 FUNCION DE GUMBEL (METODO DE GUMBEL)	
<u>Parámetros de los datos (n = 79)</u>	
Valor medio	0,005 m
Desviación estándar	0,577 m
Asimetría	0,303
Kurtosis	0,156
<u>Parámetros de la variable reducida (n = 79)</u>	
Valor medio	0,557
Desviación estándar	1,201
<u>Coefficientes de ajuste</u>	
Pendiente de la recta de regresión ortogonal	-0,480 m
Ordenada al origen	0,272 m
<u>Extrapolaciones</u>	
Crecida centenaria (T = 100 años)	-1,94 m
Intervalo de confianza del 95% (Kaczmarek)	-2,47 m / -1,40 m
Crecida milenaria (T = 1.000 años)	-3,04 m
Intervalo de confianza del 95% (Kaczmarek)	-3,83 m / -2,26 m

Como es lógico la recta tiene pendiente negativa. Puede apreciarse en la Figura 6.5 que todos los datos recaen entre las líneas de intervalos de confianza del 95%. La incertidumbre, de acuerdo a lo definido está dada por:

$$T = 100 \text{ años} \quad \varepsilon = \pm 26\%$$

$$T = 1.000 \text{ años} \quad \varepsilon = \pm 27\%$$

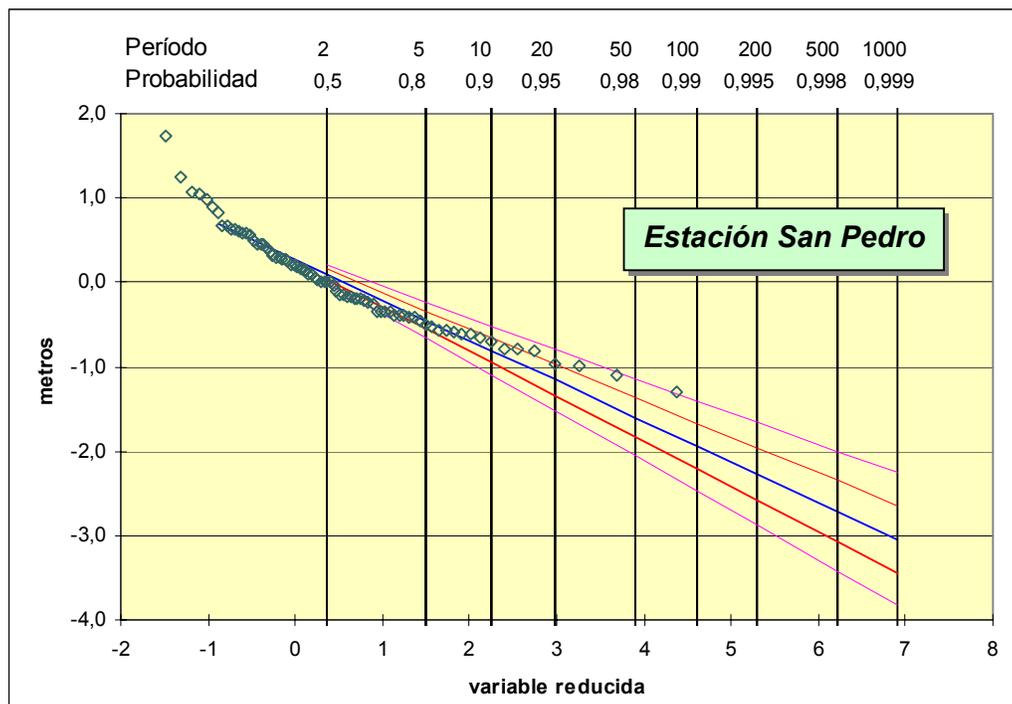


Figura 6.5: Ajuste por función de Gumbel (—) e intervalos de confianza del 68% (—) y del 95% (—)

6.3 CONCLUSIONES ACERCA DEL ESTUDIO

Para hallar los valores extremos para la localidad de Atucha (donde no se poseen datos suficientes como para encarar un análisis estadístico válido) se trazó un gráfico representando las crecientes y bajantes con igual período de retorno para distintas estaciones hidrométricas en función de la distancia. Se tomaron los resultados del análisis de valores extremos para las estaciones ubicadas al sur de Rosario y luego se interpoló para la posición de Atucha.

Los valores obtenidos para crecidas tuvieron una excelente concordancia, en Zárate y Atucha, con los calculados en ocasión de la construcción del complejo ferroviario Zárate-Brazo Largo como así también un muy buen ajuste con la cota de recurrencia milenaria tomada como base de diseño para la central Atucha I (11). En cambio para estiajes no fue posible hacer comparaciones por falta de información adecuada a la fecha de realización del estudio.

En la Figura 6.6 se graficó la diferencia entre crecientes y bajantes con igual período de retorno en función de la distancia desde el Riachuelo para el tramo Corrientes-Buenos Aires. La existencia de un mínimo en la zona de emplazamiento de las centrales nucleares permite concluir que las mismas se encuentran ubicadas en un lugar óptimo desde el punto de vista hidrométrico, puesto que el salto de agua entre

los puntos apropiados para la instalación de la sala de bombas y sus bocas de toma es mínimo.

Por último cabe aclarar que estas conclusiones estuvieron enmarcadas en un amplio estudio hidrológico llevado a cabo en la zona de Atucha. Aquí sólo se pretendió ilustrar una aplicación original con un interesante resultado a fin de presentar un ejemplo práctico relativo a la aplicación de las técnicas de análisis estadísticos de valores extremos, motivo principal del presente trabajo.

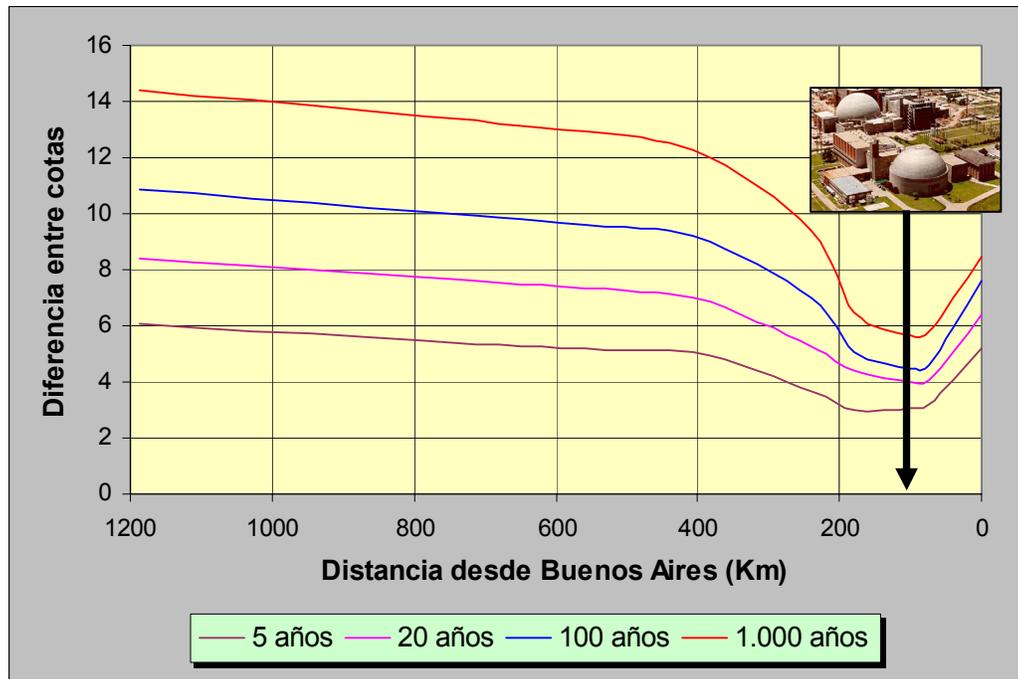


Figura 6.6: Diferencias entre cotas (metros) con igual período de retorno en función de la distancia desde Buenos Aires

APENDICES

TABLA A1: FUNCION DE FRECUENCIA NORMAL

La siguiente tabla de doble entrada representa el área bajo la curva de frecuencia normal entre 0 y x.

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
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	0,0753
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	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	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	0,2995	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 A2: FUNCION CHI CUADRADO

La probabilidad de que χ^2 tome un valor mayor que el valor crítico χ^2_p es igual a p%.

n = grados de libertad.

n	χ^2_p en función de n y p (%)													
	99	98	95	90	80	70	50	30	20	10	5	2	1	0,1
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
4	0,297	0,429	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,466
5	0,554	0,752	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086	20,515
6	0,872	1,134	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	1,239	1,564	2,167	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,321
8	1,647	2,032	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090	26,124
9	2,088	2,532	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	2,558	3,059	3,940	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	3,053	3,609	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	3,571	4,178	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	4,107	4,765	5,892	7,041	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,471	27,688	34,527
14	4,660	5,368	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,124
15	5,229	5,985	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,698
16	5,812	6,614	7,962	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000	39,252
17	6,408	7,255	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,791
18	7,015	7,906	9,390	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805	42,312
19	7,633	8,567	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191	43,819
20	8,260	9,237	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566	45,314
21	8,897	9,915	11,591	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932	46,796
22	9,542	10,600	12,338	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289	48,268
23	10,196	11,293	13,091	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,968	41,638	49,728
24	10,856	11,992	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	40,270	42,980	51,179
25	11,524	12,697	14,611	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314	52,619
26	12,198	13,409	15,379	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642	54,051
27	12,878	14,125	16,151	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	44,140	46,963	55,475
28	13,565	14,847	16,928	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	45,419	48,278	56,892
29	14,256	15,574	17,708	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	46,693	49,588	58,301
30	14,953	16,306	18,493	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	47,962	50,892	59,702

TABLA A3: FUNCION t DE STUDENT

La probabilidad de que t tome un valor mayor que el valor crítico t_p es igual a p%.

n = grados de libertad.

n	t _p en función de n y p (%)												
	90	80	70	60	50	40	30	20	10	5	2	1	0,1
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,82	63,66	636,6
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,60
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,92
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,768
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,689
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,660
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,290

BIBLIOGRAFIA

- 1) T.H. Wonnacott - R.J. Wonnacott . "Introducción a la estadística". Editorial Limusa - México (1979).
- 2) J. Feller. "An introduction to probability theory and its applications" John Willey & Sons Inc.
- 3) H. Crámer. "Teoría de probabilidades y sus aplicaciones". Aguilar - Madrid (1968).
- 4) Ven Te Chow. "Handbook of applied hydrology". Mc Graw Hill - NewYork (1965).
- 5) R.K. Linsley - M.A Kohler - J.L.H. Paulus. "Hidrología para ingenieros". Mc Graw Hill (1977).
- 6) United Nations. "Assessment of the magnitude and frequency of floods flows. Transactions of an inter-regional seminar". Water Resources Research Series N° 30 - New York (1967)
- 7) E.J. Gumbel. "Méthodes graphiques pour l'analyse des débits de crue". La Houille Blanche N°5 (1956).
- 8) J. Bernier. "Sur l'application des diverses lois limites des valeurs extrêmes au problème des débits de crue". La Houille Blanche N°5 (1956).
- 9) International Atomic Energy Agency. "Extreme meteorological events in nuclear power plants siting, excluding tropical cyclones". Safety Series N°50-SG-S11A - Viena (1981).
- 10)G. Réménieras. L'hydrologie de l'ingénieur". Eyrolles - París (1965).
- 11)H.R. Gómez - G.E. Maggio - C.R. Trípoli. "Análisis de cotas hidrométricas extremas en el emplazamiento de centrales nucleares". CNEA - NT 10/83 - Buenos Aires (1983).
- 12)E.J. Gumbel. "Statistical theory of extreme values and some practical applications" National Bureau of Standards. Applied Mathematical Series - 33 (1954).