# Capítulo 3

# Elementos de Probabilidad y Estadística

## 3.1. Introducción

En este capítulo se presentan conceptos básicos de Probabilidad y Estadística, ya que dentro del diseño y planeación de una obra hidráulica juegan un papel importante en el análisis hidrológico de eventos futuros. Un ejemplo sería el diseño de vertedores de una presa de almacenamiento en donde se necesita predecir lluvias o avenidas de diseño con una determinada frecuencia o periodo de retorno.

## 3.2. Conceptos básicos de probabilidad

### 3.2.1. Experimento

Un experimento es toda ***acción*** que se realiza con el fin de ***observar su resultado***. La experimentación es una etapa fundamental para llegar al conocimiento científico. La teoría de la probabilidad ha sido motivada por diversas circunstancias de la vida real en las que se realiza un experimento y el investigador observa un resultado. Un ejemplo que se puede considerar es la determinación de la carga de ruptura de una varilla de acero que se prueba en una cierta máquina y es sometida a un esfuerzo de tensión. En este experimento la ***acción*** consiste en someter a la varilla a una prueba de ruptura por tensión; después de aplicar incrementos de carga, la varilla falla por tensión, que sería el ***resultado*** de la acción; y la ***observación*** es la lectura en la máquina para determinar la carga de ruptura que ocasionó que la varilla fallara (Olivera, S.A., 1987).

Un experimento es determinista si se puede predecir con certeza su resultado antes de que éste se realice. Un experimento es aleatorio cuando no es posible asegurar el resultado que se va a presentar al realizarlo; es decir que a pesar de repetirlo en condiciones aproximadamente idénticas, sus resultados no son esencialmente los mismos.

### 3.2.2. Espacio muestral

Es el conjunto de todos los posibles resultados de un experimento aleatorio y generalmente se representa a dicho conjunto con las letras ***S***, L, **Ω**.

### 3.2.3. Evento

Un evento A es una colección de puntos muestrales contenidos en el espacio muestral ***S*** de un experimento aleatorio.

***Eventos mutuamente excluyentes***
Son eventos que no pueden ocurrir simultáneamente. Si se tienen dos eventos A y B cualesquiera, tales que la ocurrencia de uno implica que no puede ocurrir el otro, esto es, $A \cap B = \Phi$ entonces A y B son eventos mutuamente excluyentes.

***Eventos colectivamente exhaustivos***
A los eventos $A_1, A_2, \ldots, A_n$ se les llama eventos conjuntamente exhaustivos cuando su unión forma todo el espacio muestral, es decir:

$$S = \left(A_1 \cup A_2 \cup \ldots \cup A_n\right)$$

### 3.2.4. Definición axiomática de probabilidad

Sea ***S*** un espacio muestral, sea *L* la clase de todos los eventos y sea *P* una función de valores reales definida en *L*. Entonces *P* se llama una función de probabilidad, y ***P(A)*** se denomina la probabilidad del evento A, si y sólo sí se cumplen los siguientes axiomas (Borras, H., 1985).

***Axioma I***
La probabilidad de un evento es un número mayor o igual que cero y menor o igual que uno:

$$0 \leq P(A) \leq 1 \tag{3.1}$$

***Axioma II***
La probabilidad del evento ***S*** es la unidad:

$$P(S) = 1 \tag{3.2}$$

Donde ***S*** es el evento formado por todo el espacio muestral

***Axioma III***
La probabilidad de un evento que sea la unión de los eventos A y B mutuamente excluyentes, es la suma de las probabilidades de estos dos eventos:

$$P(A \cup B) = P(A) + P(B) \tag{3.3}$$

Para cualquier secuencia infinita de eventos mutuamente excluyentes $A_1$, $A_2$, $A_3$,..., se tiene

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_n) \tag{3.4}$$

### 3.2.5. Teoremas derivados de la definición axiomática

***Teorema 1.1***
Sea el evento A un evento cualquiera de S y A' el complemento de A, entonces:

$$P(A') = 1 - P(A) \tag{3.5}$$

Demostración:

Por el axioma II:

$$P(S) = 1$$

tomando en cuenta que:

$$(A \cup A') = S$$

obteniendo su probabilidad de ambos lados:

$$P(A \cup A') = P(S)$$

aplicando el axioma III

$$P(A) + P(A') = 1$$

$$\therefore \quad P(A') = 1 - P(A)$$

***Teorema 1.2***
Sea $\boldsymbol{\Phi}$ el evento imposible, entonces:

$$P(\phi) = 0 \tag{3.6}$$

Demostración:

haciendo

$$S = S \cup \phi$$

obteniendo su probabilidad de ambos lados:

$$P(S) = P(S) + P(\phi)$$

por el axioma II:

$$P(S) = 1$$

sustituyendo

$$1 = 1 + P(\phi)$$

$$\therefore P(\phi) = 0$$

***Teorema 1.3***
Si A y B son dos eventos de S y A⊂B (A es subconjunto de B)

$$P(A) \leq P(B) \tag{3.7}$$

Demostración:

El evento de B de la Fig. 3.1, se puede representar como la unión de los eventos mutuamente excluyentes A y A'∩B, esto es, B = {A U (A'∩ B)}.

del axioma III

$$P(B) = P(A) + P(A' \cap B)$$

y por el axioma I

$$P(A' \cap B) \geq 0$$

$$\therefore P(B) \geq P(A)$$

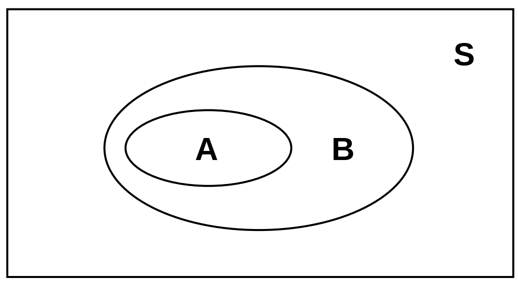


**Figura 3.1. El conjunto A es subconjunto de B.**

***Teorema 1.4***
Si $A_1$, $A_2$,..., $A_n$ es una colección de eventos mutuamente excluyentes y conjuntamente exhaustivos, como se muestra en la Fig. 3.2.

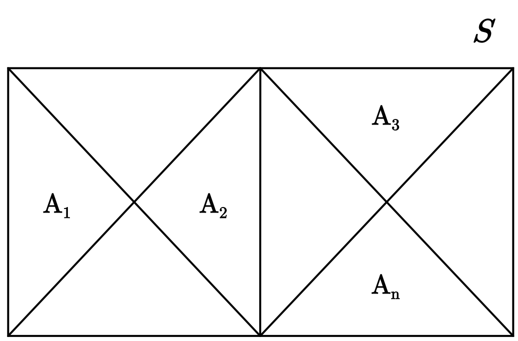


**Figura 3.2. Eventos mutuamente excluyentes y conjuntamente exhaustivos.**

Entonces:

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i)$$

$$con \quad A_i \cap A_j = \phi \quad para \quad i \neq j$$

Este teorema es solo una forma general del axioma III.

***Teorema 1.5***
Para dos eventos cualesquiera A y B:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{3.8}$$

Demostración:

analizando la Fig. 3.3 se tiene

$$P(A \cup B) = P(A - B) + P(A \cap B) + P(B - A) \tag{3.9}$$

por otra parte

$$P(A - B) = P(A) - P(A \cap B) \tag{3.10}$$

$$P(B - A) = P(B) - P(A \cap B) \tag{3.11}$$

sustituyendo las ecuaciones (3.10) y (3.11) en la ecuación (3.9):

$$P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B)$$
$$\therefore \; P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

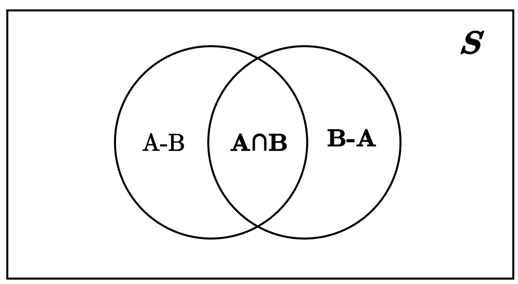


**Figura 3.3. P(AUB)**

## 3.3. Variables aleatorias

### 3.3.1. Introducción

Para dar solución a algunos problemas de ingeniería, se genera un cierto grado de incertidumbre debido a la variación de los fenómenos presentes en la naturaleza. Como resultado de las incertidumbres, el ingeniero no puede predecir con exactitud el evento que pueda ocurrir en el futuro, para solucionarlo utiliza modelos estadísticos y probabilistas en donde la variable esencial se llama ***Variable aleatoria*** (Benjamín et al., 1970).

### 3.3.2. Concepto de variable aleatoria

Dado un experimento aleatorio *E*, entonces una variable aleatoria $X$ se define como una función real de variable real cuyo dominio es el espacio muestral ***S*** de probabilidades, contenido por eventos simples $e_n$, y por otro lado el codominio de la función es un espacio que contiene al conjunto de los números reales *IR* con valores $x_n = X(e_n)$ (Montgomery et al., 2000).

Una variable aleatoria $X$ es una función real de variable real que ***asigna*** a cada evento simple $e_i \in S$ un valor numérico real $x_n \in IR$, mejor dicho $X(e_n) = x_n$. En la Fig. 3.4 se aprecia la naturaleza de la transformación de $X : S \to IR$.

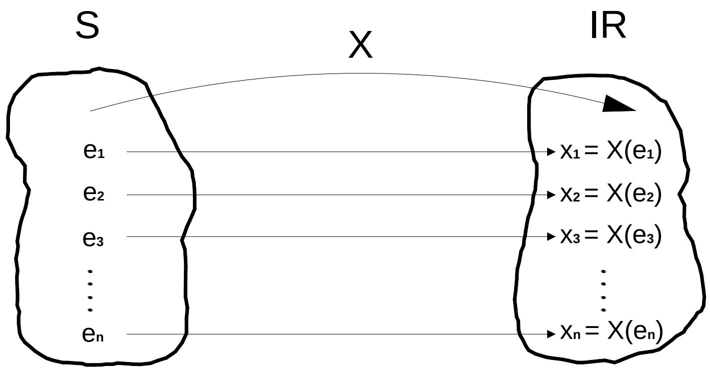


**Figura 3.4. Concepto de variable aleatoria.**

Con respecto a la notación, se consideran letras mayúsculas $X$ para denotar variables aleatorias y letras minúsculas $x$ para los valores particulares que puedan tomar. En la literatura es común encontrar denotadas con mayúsculas a las ***funciones de distribución de probabilidad acumulada F(x)*** y con minúsculas a las ***funciones densidad de probabilidad f(x)*** (Borras, H., 1985).

El comportamiento de una variable aleatoria "$X$" se describe mediante su ley de probabilidades y se caracteriza principalmente por medio de una ***distribución de probabilidad*** de la variable aleatoria. Al conjunto de parejas ordenadas $(x_i, p_i)$, se le conoce como distribución de probabilidad de la variable aleatoria $X$. Las variables aleatorias pueden ser discretas o continuas.

Como podemos observar una variable aleatoria se caracteriza en forma analítica por medio de una distribución de probabilidad y a su vez queda definida por una función de probabilidad. A continuación en la Fig. 3.5 se muestra un cuadro sinóptico para comprender la caracterización y definición de una variable aleatoria.

***Variable Aleatoria***

***Una variable aleatoria discreta se caracteriza por medio de una distribución de probabilidad y está definida por:***

- ***a) Función de probabilidad o función masa de probabilidad p(x).***
- ***b) Función de distribución de probabilidad acumulada o función de distribución para variables aleatorias discretas F(x).***

***Una variable aleatoria continua se caracteriza por medio de una distribución de probabilidad y está definida por:***

- ***c) Función densidad de probabilidad f(x)***
- ***d) Función de distribución de probabilidad acumulada o función de distribución para variables aleatorias continuas F(x).***

**Figura 3.5. Características de una variable aleatoria.**

### 3.3.3. Variables aleatorias discretas

- ***Concepto de variable aleatoria discreta***

Si el dominio de definición de una variable aleatoria es un intervalo finito, y los resultados del experimento aleatorio al que está asociada dicha variable aleatoria sólo definen algunos de los valores comprendidos dentro de ese intervalo, se dice que la variable aleatoria es discreta.

Obsérvese que la distribución de probabilidad $\{(x_i, p_i)\}$ de la variable aleatoria, define completamente el fenómeno al que está asociada, ya que dichos valores determinan los resultados del experimento y sus correspondientes probabilidades y la forma de cómo se presentan esos resultados (Benjamín et al., 1970).

- ***Función de probabilidad o función masa de probabilidad p(x)***

Una distribución de probabilidad analíticamente queda definida mediante una función de probabilidad $p(x_i)$, la cual relaciona los resultados de un experimento aleatorio con su probabilidad de ocurrencia. Cuando el resultado del experimento aleatorio es un número, generalmente se le asigna el mismo valor; o sea, la variable aleatoria $X$ puede tomar los valores de $x_i$. Lo anterior se puede representar mediante la siguiente expresión. En la Fig. 3.6 se representa gráficamente la función de probabilidad (Montgomery et al., 2000).

$$p(x_i) = P(X = x_i) \tag{3.12}$$

La condiciones que debe de cumplir cualquier función $p(x_i)$ para que sea función de probabilidad de una variable aleatoria $X$, y así poder definir una distribución de probabilidad, debe cumplir los primeros dos axiomas de la teoría de la probabilidad.

$$0 \leq P(x_i) \leq 1\,, \quad \forall i \tag{3.13}$$

$$\sum_{i=1}^{x} P(x_i) = 1 \tag{3.14}$$

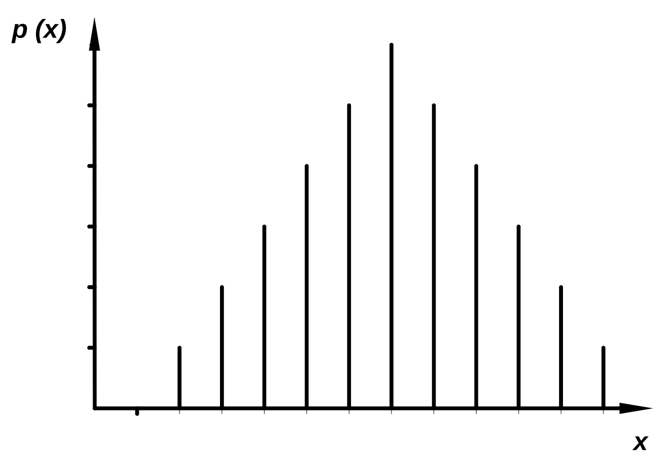


**Figura 3.6. Función de probabilidad.**

### 3.3.4. Variables aleatorias continuas

En el análisis estadístico en hidrología generalmente se trabaja con variables aleatorias continuas, por lo que se analizarán con precaución.

#### *3.3.4.1. Concepto de variable aleatoria continua*

Si el dominio de definición de una variable aleatoria es un intervalo infinito, y los resultados del experimento aleatorio al que está asociada dicha variable definen cualquier valor real comprendido dentro del intervalo, se dice que la variable aleatoria es continua.

Este tipo de variables se presentan con más frecuencia en los sistemas físicos debido a que la masa, el calor, el volumen, el tiempo y otras características, pueden tomar cualquier valor, dentro de una escala continua. Sin embargo, formalmente puede definirse la distribución de probabilidades de una variable aleatoria continua $X$ por medio de su ***función densidad de probabilidad*** (Benjamín et al., 1970).

#### *3.3.4.2. Función densidad de probabilidad f(x)*

En el caso de una variable aleatoria continua no es posible establecer explícitamente las parejas ordenadas (valor de la variable aleatoria, y su correspondiente probabilidad) que definen una distribución de probabilidad continua, por no ser numerable el conjunto de valores de la variable aleatoria continua. Una variable aleatoria continua $X$ puede tomar una infinidad de valores en un pequeño intervalo; en consecuencia, la probabilidad de que $X$ tome un valor en particular es prácticamente cero y para estudiar el comportamiento de las variables se debe de hablar de subintervalos, dentro de los cuales puede encontrarse el valor de $X$ (Montgomery et al., 2000).

Sea $X$ una variable aleatoria continua en el intervalo $-\infty \leq x \leq \infty$, o sea, que puede tomar cualquier valor real en ese intervalo, por lo que no es posible definir su distribución de probabilidad en forma tabular; no es posible expresar explícitamente el conjunto infinito de parejas ordenadas $\{(x_i, p_i)\}$ que se tiene para todo valor de x en el intervalo. En este caso deberá definirse la distribución de probabilidad x en forma gráfica o analítica.

Por la continuidad de los valores de x en el intervalo, en este caso el polígono de probabilidad se convierte en una curva continua de ecuación $y = f(x)$, siendo $f(x)$ una función real de la variable aleatoria $X$ en su intervalo de definición. A la función $f(x)$ se le llama ***función densidad de probabilidad*** de la variable aleatoria $X$.

La función densidad de probabilidad se define de tal manera que la probabilidad del evento $P(x_1 \leq X \leq x_2)$ sea igual al área bajo la curva de la función $f(x)$ entre $x = x_1$ y $x = x_2$, como se muestra en la Fig. 3.7.

entonces:

$$\boxed{P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)\, dx} \qquad (3.15)$$

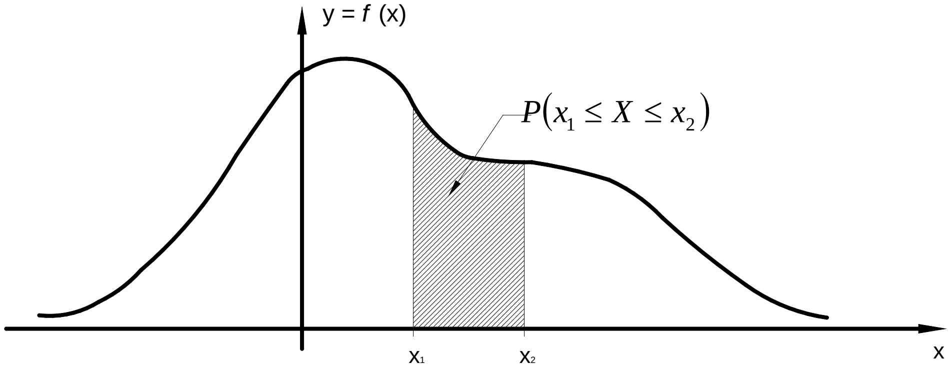


**Figura 3.7. Función densidad de probabilidad.**

Es importante darse cuenta de que $f(x)$ no representa la probabilidad de nada, solo cuando la función integra entre dos puntos se produce una probabilidad.

Una función $f(x)$ debe de cumplir con las condiciones siguientes para que sea función densidad de probabilidad de una variable aleatoria continua:

a) La función $f(x)$ es positiva para todo valor de x:

$$f(x) \geq 0 \tag{3.16}$$

b) El área bajo la curva de la función $f(x)$ en el intervalo $-\infty \leq x \leq \infty$ vale uno:

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \tag{3.17}$$

### *3.3.4.3. Función de distribución de probabilidad acumulada F(x)*

Otra forma de caracterizar el comportamiento de una variable aleatoria es mediante la distribución de probabilidad acumulada y puede definirse formalmente por medio de una ***función de distribución de probabilidad acumulada*** $F(x_0)$, como:

$$F(x_0) = P(x \leq x_0) \tag{3.18}$$

La función de distribución acumulada de cualquier tipo de variable aleatoria discreta o continua, tiene las siguientes propiedades y su representación gráfica se visualiza en la Fig. 3.8 (Montgomery et al., 2000).

$$\begin{aligned}
&F(x_0) \geq 0 \quad \forall\ x_0 \\
&F(-\infty) = 0 \\
&F(\infty) = 1 \\
&F(x_0 + \varepsilon) \geq F(x_0), \quad para\, toda\, cons\tan te\, \varepsilon > 0 \\
&P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)
\end{aligned}$$

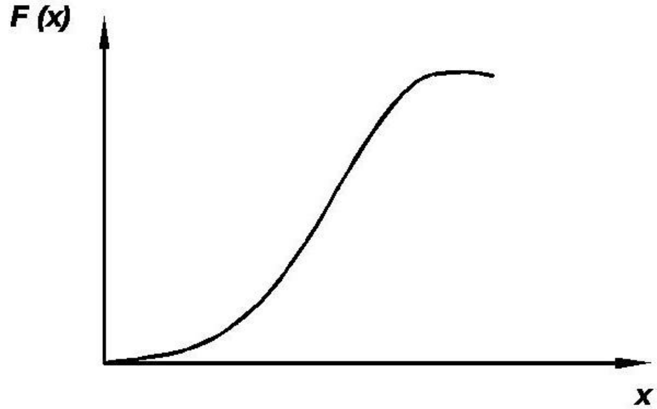


**Figura 3.8. Función de distribución de probabilidad acumulada.**

#### *3.3.4.4. Función de distribución de probabilidad acumulada para variables aleatorias continuas*

Si la variable aleatoria es continua y $f(x)$ es su función densidad de probabilidad, la función $F(x)$ es:

$$F(x_0) = P(x \leq x_0) = \int_{-\infty}^{x_0} f(x)\, dx$$

$$\boxed{F(x_0) = \int_{-\infty}^{x_0} f(x)\, dx} \tag{3.19}$$

### 3.3.5. Esperanza matemática o valor esperado de variables aleatorias continuas

En general, en la literatura se le considera a la esperanza matemática como valor esperado o valor promedio de la variable aleatoria $X$. Puede ser interpretada como un ***promedio ponderado*** de una función *g(x)* y como se verá posteriormente es de gran utilidad para definir algunos ***parámetros poblacionales*** que caracterizan el comportamiento de la variable aleatoria (Borras, H., 1985).

- ***Esperanza matemática de una función de una variable aleatoria continua***

Si la variable aleatoria es de tipo continua, y su función densidad de probabilidad es $f(x)$, la esperanza matemática de *g(x)* se define como:

$$\boxed{E\{g(x)\} = \int_{-\infty}^{\infty} g(x)\, f(x)\;\; dx} \tag{3.20}$$

### 3.3.6. Parámetros poblacionales de la distribución de variables aleatorias continuas

Como se había comentado, los parámetros más importantes a calcular son la media y la varianza. Para su estudio se presentaran dos casos particulares de la esperanza matemática que son: los momentos con respecto al origen y los momentos con respecto a la media (Olivera, S.A., 1987).

***Momentos de orden n con respecto al origen***

Para una distribución de probabilidad, se define al momento de orden $n$ con respecto al origen de la variable aleatoria $X$, a la esperanza matemática de la función *g(x)* considerando lo siguiente:

tomando en cuenta que:

$$g(x) = X^n \tag{3.21}$$

sustituyendo la ecuación (3.21) en la ecuación (3.20), resulta

$$\boxed{E\{X^n\} = \int_{-\infty}^{\infty} x^n \; f(x)\,dx} \tag{3.22}$$

donde: $n$ es un número entero positivo.

### *3.3.6.1. Momentos de orden n con respecto a la media*

Se llama momento de orden n con respecto a la media de la distribución de probabilidad de la variable aleatoria $X$, a la esperanza matemática de la función *g(x)* considerando lo siguiente:

si

$$g(x) = (X - \mu_X)^n \tag{3.23}$$

donde: $n$ es un entero positivo y $\mu_x$ la media de la distribución,

sustituyendo la ecuación (3.23) en la ecuación (3.20)

$$\boxed{E\{(X - \mu_X)^n\} = \int_{-\infty}^{\infty} (x - \mu_X)^n \; f(x)\,dx} \tag{3.24}$$

### *3.3.6.2. Media de la distribución poblacional*

Tomando en cuenta el momento de orden n respecto al origen para el caso de variables aleatorias discretas, se tiene:

$$E\{X^n\} = \sum_{\forall x} x^n p(x) \tag{3.25}$$

y obteniendo el primer momento con respecto al origen, n = 1

$$E\{X\} = \sum_{\forall x} x \;\; p(x) \tag{3.26}$$

si hacemos

$$\mu_X = E\{X\} \tag{3.27}$$

sustituyendo la ecuación (3.26) en la ecuación (3.27)

$$\mu_X = \sum_{\forall x} x \;\; p(x) \tag{3.28}$$

Si consideramos que cada $x_i$ ocurre con la misma probabilidad $p_i$, entonces cada $p_i$ tiene valor de:

$$p_i = \frac{1}{n} \tag{3.29}$$

sustituyendo la ecuación (3.29) en la ecuación (3.28)

$$\boxed{\mu_X = \frac{1}{n}\sum_{i=1}^{n} x_i} \tag{3.30}$$

A la expresión anterior se le llama ***media de la distribución***. La media es uno de los parámetros más representativos de la distribución, ya que conociendo la media o valor esperado de la distribución se tiene una idea del valor central de la distribución.

Una analogía del concepto físico del primer momento, que es la media de la distribución de probabilidad de la variable aleatoria  es semejante al concepto de centroide que se estudia en mecánica. Observando la Fig. 3.9, la media es la abscisa del centroide de la función de probabilidad (Olivera, S.A., 1987).

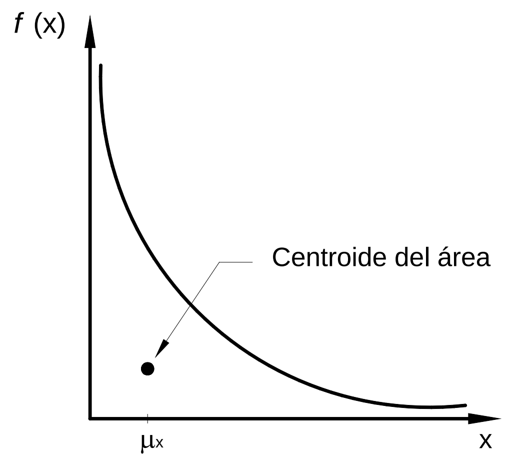


**Figura 3.9. La media es la abscisa del centroide.**

### *3.3.6.3. Varianza de la distribución poblacional*

Si n = 0, se tiene el momento de orden cero con respecto a la media, y tanto en el caso discreto como en el continuo vale uno.

Si n = 1, se tiene el momento con respecto a la media, y tanto en el caso discreto como en el continuo vale cero.

Si n = 2, se tiene el segundo momento con respecto a la media.

de la ecuación (24)

$$E\left\{(X-\mu_X)^2\right\} = \int_{-\infty}^{\infty} (x-\mu_X)^2 \, f(x)\, dx \tag{3.31}$$

haciendo

$$\sigma_X^2 = E\left\{(X-\mu_X)^2\right\}$$

Por lo tanto el momento de orden dos, n = 2, con respecto a la media, recibe el nombre de *varianza* y se expresa de la siguiente manera:

$$\boxed{\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 \; f(x)\, dx} \tag{3.32}$$

Si se considera a la esperanza matemática como un promedio y a la diferencia $(X - \mu_X)$ como la distancia entre un valor particular de $X$ y su media, ***entonces se interpreta a la varianza como el promedio de los cuadrados de las diferencias entre el valor medio ($\mu_x$) y los posibles valores de la variable aleatoria*** (Olivera, S.A., 1987).

Si el exponente *n* toma cualquier otro valor par, se tendrá también un indicador de la dispersión de los datos y éste será siempre positivo, mientras que para una potencia impar, el momento puede ser positivo, negativo o cero.

La raíz cuadrada de la varianza define el concepto de ***desviación estándar*** de la variable aleatoria $X$ y se representa por $\sigma_X$, donde:

$$\sigma_X = \sqrt{\sigma_X^2} \tag{3.33}$$

La desviación estándar explica la dispersión promedio de los valores posibles de la variable aleatoria con respecto a su media.

## 3.4. Conceptos básicos de estadística

### 3.4.1. Introducción

La estadística ha llegado a ser un instrumento de uso cotidiano para todos los profesionistas que están en contacto con fenómenos de naturaleza aleatoria, y que a partir del conocimiento de ciertos datos cuantitativos del fenómeno, deben tomar decisiones sobre su comportamiento general.

### 3.4.2. Concepto de Estadística

La Estadística se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los datos, siempre y cuando la variabilidad e incertidumbre sea una causa intrínseca de los mismos; así como de realizar inferencias a partir de ellos, con la finalidad de ayudar a la toma de decisiones y en su caso formular predicciones.

Para su estudio, la estadística se divide en: estadística descriptiva, cuando los resultados del análisis no pretenden ir más allá del conjunto de datos, e inferencia estadística cuando el objetivo del estudio es derivar las conclusiones obtenidas a un conjunto de datos más amplio (Borras, H., 1985).

### 3.4.3. Estadística descriptiva

***Definición***

Ciencia dedicada a describir las características existentes en un conjunto de datos (la muestra), utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos. Las tareas de la estadística descriptiva son: La organización de los datos numéricos de la muestra a través de las tablas y las representaciones gráficas (Borras, H., 1985).

***Población***
Es un conjunto finito o infinito de objetos, llamados elementos, que tienen en común una o varias características particulares que se desean estudiar.

***Muestra***
Es un subconjunto de la población. La selección de una muestra es una etapa muy importante dentro del estudio estadístico, debido a que la información que presenta la muestra es la base para hacer suposiciones o inferencias sobre lo que ocurre en la población.

#### 3.4.3.1. Parámetros estadísticos de una muestra

Como ya se vio anteriormente los parámetros poblacionales se representan mediante letras griegas $\mu$, $\sigma$, $\sigma^2$, $\nu$, $\gamma$. En el caso de una distribución de frecuencias también se pueden establecer medidas descriptivas, conocidos como ***parámetros estadísticos o estadísticos muestrales***, y para distinguirlas de los parámetros poblacionales se usan las letras latinas $\bar{x}$, $S$, $S^2$, $C_v$, $g$.

En el análisis hidrológico se recomienda el uso de los estadísticos no sesgados, ya que generalmente se trabaja con muestras relativamente pequeñas.

Los datos de una muestra pueden caracterizarse numéricamente mediante los siguientes grupos de parámetros estadísticos:

- ***Medidas de tendencia central***

***Media*** $(\bar{x})$

Se define como el promedio aritmético de los datos de una muestra y está dada por:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (3.34)$$

donde:

$\bar{x}$ media de la muestra
$x_i$ valores de la muestra
$n$ número total de valores

- ***Medidas de dispersión***

Como su nombre lo indica, las medidas de dispersión reflejan la separación o alejamiento de los elementos de una muestra (Borras, H., 1985).

***Varianza ($S_x^2$)***

En el tema de variables aleatorias se definió a la varianza de una variable aleatoria como el segundo momento con respecto a la media. De manera análoga, para una distribución de frecuencias se define el momento k-ésimo con respecto a la media como:

$$m_k = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^k \tag{3.35}$$

y en consecuencia, la ***varianza muestral*** (insesgada) se define como:

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \tag{3.36}$$

y la ***desviación estándar ($S_X$)*** como:

$$S_X = \sqrt{S_X^2} \tag{3.37}$$

***Coeficiente de variación (Cv)***

Para una distribución de frecuencias, se define como el cociente de la desviación estándar muestral entre la media muestral; esto es:

$$C_V = \frac{S_X}{\bar{x}} \tag{3.38}$$

- ***Medidas de asimetría***

Para medir en forma adimensional la asimetría de una distribución de frecuencias, se utiliza el ***coeficiente de asimetría*** por momentos ***g***, que se define como el cociente del tercer momento con respecto a la media entre la raíz cuadrada del segundo momento con respecto a la media elevada al cubo; esto es:

$$g = \frac{m_3}{\left(m_2^3\right)^{1/2}}$$

de donde resulta

$$g = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{\left[\frac{(n-1)}{n}S_X^2\right]^{3/2}} \tag{3.39}$$

## 3.5. Generalidades de inferencia estadística

### 3.5.1. Introducción

En estadística se le llama inferencia al proceso de inducir o bien deducir las características o parámetros poblacionales a partir de la información muestral, midiendo con probabilidades la incertidumbre (Olivera, S.A., 1987).

Para hacer una inferencia estadística, se pueden ***estimar los parámetros*** descriptivos de la población a partir de la información obtenida en una muestra, ya sea como un valor puntual, como un intervalo, o bien establecer valores hipotéticos de los parámetros y probar estadísticamente si son válidas las hipótesis.

Un estimador es un valor aproximado de un parámetro poblacional, y es determinado de los estadísticos muestrales obtenidos de la población. Existen dos formas para estimar los parámetros: ***puntuales o por intervalos de confianza***.

***Estimación puntual:*** cuando solamente se da un valor del parámetro desconocido.
***Estimación por intervalos de confianza:*** cuando se fijan dos valores entre los cuales se encuentra el parámetro desconocido.

### 3.5.2. Métodos para determinar la estimación puntual de parámetros poblacionales

A continuación se presentan algunos métodos para la estimación de parámetros más comunes en hidrología, estos son: método de momentos y método de máxima verosimilitud.

#### *3.5.2.1. Método de momentos*

El método de los momentos es un procedimiento muy sencillo para encontrar un estimador de uno o más parámetros poblacionales. Consiste básicamente en igualar los momentos poblacionales con los muestrales con lo que se genera un sistema de ecuaciones, cuyo tamaño depende del número de parámetros a estimar (Montgomery et al., 2000).

Supóngase que $X$ es una variable aleatoria ***continua*** con función densidad de probabilidad

$$f(x;\theta_1,\theta_2,\dots\theta_k) \tag{3.40}$$

Donde dicha función tiene parámetros desconocidos $\theta_1,\theta_2,\dots\theta_k$. Sea $X_1,X_2,\dots,X_n$ una muestra aleatoria de tamaño $n$ de $X$.

Por otra parte definidos los primeros ***k-ésimos momentos de una muestra*** con respecto al origen, resulta:

$$m_t = \frac{\sum_{i=1}^{n} X_i^t}{n} \qquad t = 1,2,\dots,k \tag{3.41}$$

Los primeros ***k momentos de la población*** con respecto al origen son:

$$\mu_t = E(X^t) = \int_{-\infty}^{\infty} x^t f(x;\theta_1,\theta_2,\ldots\theta_k)\,dx$$
$$t = 1,2,\ldots k \qquad X \text{ es continua} \tag{3.42}$$

Al igualar los momentos de la muestra con los momentos de la población se producirán $k$ ecuaciones con $k$ incógnitas, esto es:

$$\mu_t = m_t$$

$$\boxed{\begin{array}{c}\displaystyle\int_{-\infty}^{\infty} x^t f(x;\theta_1,\theta_2,\ldots\theta_k)\,dx = \frac{\sum_{i=1}^{n} X_i^t}{n} \\ t = 1,2,\ldots,k\end{array}} \tag{3.43}$$

Con la solución de la ecuación anterior, se encuentran los ***estimadores*** de momento $\hat{\theta}_1,\hat{\theta}_2,\ldots\hat{\theta}_k$ de los *parámetros* $\theta_1,\theta_2,\ldots\theta_k$.

### *3.5.2.2. Método de máxima verosimilitud*

Uno de los mejores métodos para obtener un estimador puntual es el de máxima verosimilitud (Montgomery et al., 2000).

Supóngase que $X$ es una variable aleatoria continua con función densidad de probabilidad $f(x;\theta)$, donde $\theta$ es un parámetro desconocido único y asociado a la distribución de probabilidad de $X$. Sean $X_1, X_2, \ldots, X_n$ los valores observados en una muestra aleatoria de tamaño *n*. Su función densidad conjunta se describe como:

$$f(x_1, x_2, \ldots, x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta) \tag{3.44}$$

A la expresión anterior es conocida como ***función de verosimilitud*** y se representa con la letra *L*.

$$L = \prod_{i=1}^{n} f(x_i;\theta) \tag{3.45}$$

Para poder encontrar el estimador de máxima verosimilitud $\hat{\theta}$ del parámetro $\theta$, $\hat{\theta}$ debe de maximizar a la función de verosimilitud

$$\hat{\theta} = \text{máx } f\left(X_1, X_2, \ldots, X_n;\tilde{\theta}\right) \tag{3.46}$$

***"El procedimiento para estimar los parámetros en donde la función de verosimilitud alcanza su máximo, implica realizar la derivada parcial de L con respecto a*** $\theta$ ***e igualando a cero"*** (Montgomery et al., 2000).

## 3.6. Periodo de retorno ($Tr$)

El objetivo principal del análisis estadístico de datos hidrológico es la determinación del llamado periodo de retorno de un cierto evento hidrológico.

### 3.6.1. Concepto de periodo de retorno

El ***periodo de retorno*** (***Tr***) se define como el lapso de tiempo promedio en años, en que se presente la ocurrencia de un evento igual o mayor a una magnitud dada.

La probabilidad $p = P(X \geq x_{Tr})$ de ocurrencia del evento $X \geq x_{Tr}$ en cualquier observación puede relacionarse con el periodo de retorno, de tal modo que para cada observación existen dos resultados posibles: si $X \geq x_{Tr}$ (probabilidad $p$) se tiene un éxito, en otro caso $X < x_{Tr}$ (probabilidad 1 – $p$) se presenta una falla.

La probabilidad de ocurrencia de un evento en cualquier observación es el inverso de su periodo de retorno

$$P(X \geq x_{Tr}) = \frac{1}{Tr} \tag{3.47}$$

La ecuación anterior indica que si un evento hidrológico $X$ es igual o mayor que $x$, entonces ocurre dicho evento por lo menos una vez en Tr años, de donde 1/Tr es la probabilidad de excedencia.

De la ecuación (3.47) se deriva el periodo de retorno con probabilidad de no excedencia,

$$P(X \leq x_{Tr}) = 1 - \frac{1}{Tr} \tag{3.48}$$

Usualmente cuando se tienen datos de un cierto periodo, y se desea aplacar algún método estadístico para extrapolar dichos datos a periodos de retorno mayores al de las mediciones, es necesario asignar un valor $Tr$ a cada dato registrado. Para asignar periodos de retorno a una serie de datos es común el empleo de la ley empírica de Weibull:

$$Tr = \frac{n+1}{k} \tag{3.49}$$

donde:

n número de años de registro
k número de orden del dato analizado ordenado de mayor a menor

### 3.6.2. Criterios usuales para fijar un periodo de retorno

- ***Criterios económicos***

La fijación del periodo de retorno puede llevarse a cabo por medio de criterios económicos, como pueden ser la comparación de los costos anuales de las obras con los daños producidos por avenidas.

Según la Fig. 3.10, entre mayor sea el periodo de retorno Tr, los costos de una obra crecerán de manera importante y por lo tanto los costos de los daños producidos por avenidas serán relativamente pequeñas. La suma de las curvas 1 y 2 será el costo total y el costo mínimo será el punto mínimo de dicha curva de costos totales (Monsalve, S.J., 1999).

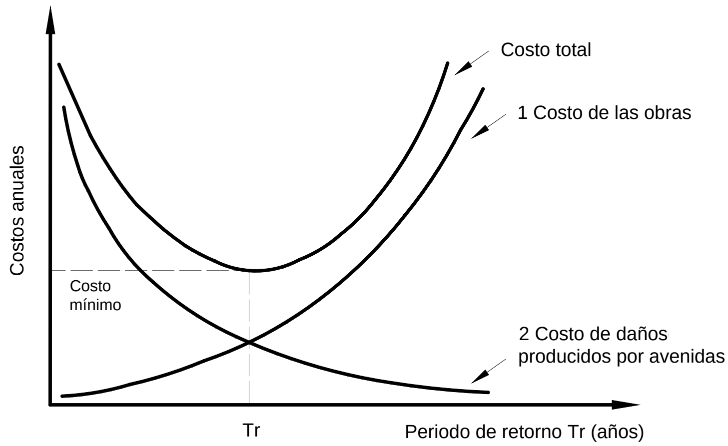


**Figura 3.10. Análisis de costos anuales de obras para la determinación de periodos de retorno.**

- ***Criterios usuales***

I. Vida útil de la obra
II. Tipo de estructura
III. Facilidad de reparación y ampliación
IV. Peligro de pérdidas y vidas humanas

## 3.7. Funciones de distribución de probabilidad de variables aleatorias continuas más usadas en hidrología

### 3.7.1. Introducción

Las funciones de densidad más comunes para el análisis hidrológico son las siguientes:

- Normal
- Lognormal
- Gamma
- Exponencial
- Pearson tipo III (Gamma de tres parámetros)
- Gumbel (Distribución general de valores extremos tipo I)
- Gumbel dos poblaciones (Gumbel mixta)

Las primeras cinco obedecen a un tipo de población, mientras que la distribución Gumbel dos poblaciones trabaja con dos tipos de población; por ejemplo los gastos producidos por lluvias ciclónicas y no ciclónicas o bien en sitios, como en el Noroeste de México, en los que se presentan volúmenes grandes debido a lluvias de invierno conocidas como "equipatas" (Domínguez et al., 2009).

Las funciones de probabilidad que generalmente se ajustan a lluvias y escurrimientos anuales son las tres primeras. Las dos últimas se desarrollaron para el análisis de eventos extremos. La función exponencial es muy poco flexible, se usa para lluvias máximas y para estudios de volúmenes almacenados.

Como ya se había comentado, los métodos estadísticos se basan en *ajustar* una muestra de gastos máximos registrados, por medio de una función de distribución de probabilidad, para posteriormente extrapolarlos a un determinado periodo de retorno.

Las distribuciones que manejaremos para ajustar las muestras con el fin de diseñar Avenidas de Diseño, son las distribuciones de valores extremos, o sea, las funciones de distribución Gumbel y Gumbel dos poblaciones. De modo que daremos una breve explicación sobre estas distribuciones en particular.

### 3.7.2. Función de distribución de valores extremos tipo I (Gumbel)

Consideremos una población en donde se tienen ***n*** muestras, cada muestra contiene ***n*** eventos. A hora seleccionemos el valor máximo ***q*** de cada muestra. Si el tamaño de las muestras es suficientemente grande, o sea, cuando ***n*** tiende a infinito la función de distribución de probabilidad acumulada o función de distribución de probabilidad de ***q*** tiende a una distribución de tipo Gumbel;

$$F(q) = e^{-e^{-\left[\frac{q-\beta}{\alpha}\right]}} \tag{3.50}$$

La función densidad de probabilidad está dada por:

$$f(q) = \frac{1}{\alpha} e^{-e^{-\left[\frac{q-\beta}{\alpha}\right]}} e^{-\left[\frac{q-\beta}{\alpha}\right]} \tag{3.51}$$

$$(-\infty < q < \infty, \alpha > 0)$$

donde:

| | |
|---|---|
| $\alpha$ | parámetro de escala |
| $\beta$ | parámetro de ubicación |
| $q$ | variable aleatoria continua (representa a los gastos máximos) |
| $F(q)$ | función de distribución de valores extremos tipo I |
| $f(q)$ | función densidad de probabilidad tipo Gumbel |

Esta función de distribución se utiliza para determinar la probabilidad de que se presenten grandes avenidas, debido a que se ha demostrado teóricamente que se ajusta a los valores máximos. La tendencia de la función de distribución Gumbel se presenta en la Fig. 3.11.

- ***Parámetros poblacionales de la distribución Gumbel***

**Media**

$$\mu = \beta + \Gamma\,\alpha \tag{3.52}$$

donde:

$\Gamma$ constante de Euler $\Gamma = 0.5772156649$

**Desviación estándar**

$$\sigma = \frac{\pi}{\sqrt{6}}\alpha \tag{3.53}$$

**Coeficiente de asimetría**

$$\gamma = 1.1396 \tag{3.54}$$

El coeficiente de asimetría es constante lo cual indica que esta distribución es sesgada a la derecha para todos los valores de $\alpha$ y $\beta$. Como la distribución es asimétrica se recomienda el método de máxima verosimilitud para estimar los parámetros $\alpha$ y $\beta$ (Berezowsky et al., 1981).

- ***Variable reducida de la distribución Gumbel***

La variable reducida ***z***, de la distribución Gumbel se deduce de la ecuación (3.50) como:

$$\left(\frac{q-\beta}{\alpha}\right) = -Ln\left[Ln\left(\frac{1}{F(q)}\right)\right] \tag{3.55}$$

sustituyendo la ecuación (3.48) en la ecuación (3.55), la variable reducida ***z*** queda en términos del periodo de retorno como:

$$z = -Ln\left[Ln\left(\frac{Tr}{Tr-1}\right)\right] \tag{3.56}$$

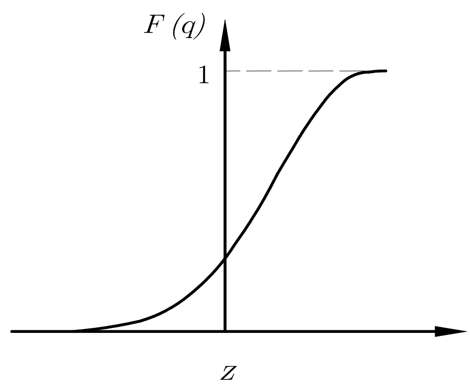


**Figura 3.11. Función distribución de Gumbel.**

- ***Estimación de parámetros estadísticos por el método de momentos***

El valor de los parámetros $\alpha$ y $\beta$ se estiman con las ecuaciones siguientes

$$\alpha = \frac{\sqrt{6}}{\pi}S \tag{3.57}$$

$$\beta = \overline{x} - \left(\frac{\sqrt{6}}{\pi}\Gamma\right)S \tag{3.58}$$

donde:

$\overline{x}$ media de muestra, calculada con la ecuación (3.34)
$S$ desviación estándar de la muestra, calculada con las ecuaciones (3.36) y (3.37)
$\Gamma$ constante de Euler $\Gamma = 0.5772156649$

- ***Estimación de parámetros estadísticos por el método de máxima verosimilitud***

El logaritmo natural de la función de verosimilitud para la distribución Gumbel es (Berezowsky et al., 1981).

$$Ln\left(L\right) = n\ Ln\left(\alpha\right) - \left[\alpha \sum_{i=1}^{n}\left(q_i - \beta\right)\right] - \left[\sum_{i=1}^{n} e^{\ -\alpha\ (q_i - \beta)}\right] \tag{3.59}$$

derivando la ecuación (3.59) parcialmente con respecto a los parámetros $\alpha$ y $\beta$ e igualando acero se tiene:

$$\frac{\partial}{\partial \beta}\left(Ln(L)\right) = n\ \alpha - \alpha \sum_{i=1}^{n} e^{\ -\alpha\ (q_i - \beta)} = 0 \tag{3.60}$$

$$\frac{\partial}{\partial \alpha}\left(Ln(L)\right) = \frac{n}{\alpha} - \sum_{i=1}^{n}\left(q_i - \beta\right) + \sum_{i=1}^{n}\left(q_i - \beta\right) e^{\ -\alpha\ (q_i - \beta)} = 0 \tag{3.61}$$

Se observa que las ecuaciones (3.60) y (3.61) forman un sistema de ecuaciones, la ecuación (3.60) puede ser escrita como:

$$e^{\alpha\ \beta} = \frac{n}{\sum_{i=1}^{n} e^{-(\alpha\ qi)}} \tag{3.62}$$

sustituyendo la ecuación (3.62) en la ecuación (3.61) resulta:

$$\frac{n}{\alpha} - \sum_{i=1}^{n} q_i + n \left[\frac{\sum_{i=1}^{n} q_i\ e^{\ \alpha\ q_i}}{\sum_{i=1}^{n} e^{\ \alpha\ q_i}}\right] = 0 \tag{3.63}$$

La ecuación (3.63) es no lineal, de modo que para darle solución se puede resolver numéricamente por medio del método de Newton-Raphson, la fórmula de recurrencia queda como:

$$\alpha_{i+1} = \frac{1}{\alpha_i} - \frac{F\left(\alpha_i\right)}{dF\left(\alpha_i\right)} \tag{3.64}$$

Donde la función y su correspondiente derivada se muestran a continuación:

$$F\left(\alpha_i\right) = n\ \alpha_i - \sum_{i=1}^{n} q_i + n \left[\frac{\sum_{i=1}^{n} q_i \cdot e^{-\left(\frac{q_i}{\alpha}\right)}}{\sum_{i=1}^{n} e^{-\left(\frac{q_i}{\alpha}\right)}}\right] \tag{3.65}$$

$$\frac{dF(\alpha_i)}{\alpha_i} = -n\ \alpha_i^2 - n\left[\frac{\left(\sum_{i=1}^{n} e^{-\left(\frac{q_i}{\alpha}\right)}\right)\left(\sum_{i=1}^{n} q_i^2 \cdot e^{-\left(\frac{q_i}{\alpha}\right)}\right) - \left(\sum_{i=1}^{n} q_i \cdot e^{-\left(\frac{q_i}{\alpha}\right)}\right)^2}{\left(\sum_{i=1}^{n} e^{-\left(\frac{q_i}{\alpha}\right)}\right)^2}\right] \tag{3.66}$$

Una vez encontrado el valor óptimo de $\alpha$, se sustituye en la ecuación (3.62) para obtener $\beta$ como:

$$\beta = \alpha\left[Ln(n) - Ln\left(\sum_{i=1}^{n} e^{-\left(\frac{q_i}{\alpha}\right)}\right)\right] \tag{3.67}$$

Los parámetros estadísticos iniciales $\alpha$ y $\beta$ que requieren en el método de Newton-Raphson, se calculan por medio de la técnica de momentos (Ecns. 3.57 y 3.58).

- ***Estimación de eventos de diseño ajustados con la función Gumbel***

Una vez obtenidos los parámetros óptimos $\alpha$ y $\beta$, y realizando el cambio de variable $q_i = Q_{calculado\ i}$ de la ecuación (3.50), se despeja $Q_{calculado\ i}$ de la siguiente manera:

Si se aplican logaritmos naturales dos veces en ambos miembros de la ecuación (3.50), desarrollando y aplicando propiedades de los logaritmos se tiene:

$$\left(\frac{Q_{calculado\ i} - \beta}{\alpha}\right) = -Ln\left[Ln\left(\frac{1}{F(Q_{calculado\ i})}\right)\right] \tag{3.68}$$

de donde los eventos de diseño se calculan como:

$$Q_{calculado\ i} = \beta - \alpha\left[Ln\left(Ln\left(\frac{1}{F(Q_{calculado\ i})}\right)\right)\right] \tag{3.69}$$

Sustituyendo la ecuación (3.48) (probabilidad de no excedencia) en la ecuación (3.69):

$$Q_{calculado\ i} = \beta - \alpha\left[Ln\left(Ln\left(\frac{T_r}{T_r - 1}\right)\right)\right] \tag{3.70}$$

### 3.7.3. Función de distribución Gumbel dos poblaciones (Gumbel mixta)

- ***Introducción***

Cuando la República Mexicana se ve afectada por ciclones, y conjuntamente por eventos hidrometeorológicos, podemos observar mediante la distribución de Gumbel que se presentan dos poblaciones totalmente diferentes; por un lado precipitaciones ocasionadas por fenómenos meteorológicos, y en segunda instancia por avenidas extremas ocasionadas por precipitaciones ciclónicas.

Para el análisis de esta distribución se supone que los fenómenos se originan por procesos diferentes, dando como resultado una distribución con dos poblaciones. La primera con una población no ciclónica y la otra corresponde a una ciclónica (Fig. 3.12).

- ***Deducción de la función de distribución de probabilidad Gumbel dos poblaciones***

Según ***González (1970)*** debido a la presencia de dos poblaciones, la función Gumbel parece ser inadecuada para determinar su frecuencia, sin embargo se puede suponer que para cada población se puede estudiar por separado analizando la distribución Gumbel.

De tal manera que para los años donde ***no se presentan ciclones***, los gastos máximos siguen una distribución de tipo Gumbel con parámetros $\alpha_1$ y $\beta_1$ de la siguiente manera:

de la ecuación (3.50) se sustituyen los parámetros $\alpha_1$ y $\beta_1$

$$F_1(q) = e^{-e^{-\left[\frac{q-\beta_1}{\alpha_1}\right]}} \tag{3.71}$$

Cuando los gastos máximos son ***provocados por ciclones***, entonces se sustituyen los parámetros $\alpha_2$ y $\beta_2$ en la ecuación (3.50) como:

$$F_2(q) = e^{-e^{-\left[\frac{q-\beta_2}{\alpha_2}\right]}} \tag{3.72}$$

La pregunta que se plantea ahora es la siguiente:

¿Cómo se puede saber o predecir la presencia de que años son ciclónicos y cuáles no lo son?

Lo anterior se logra a partir de los registros meteorológicos disponibles con que se cuenta o en un caso extremo preguntando con gente nativa del lugar.

Otra forma de saberlo es a partir de la siguiente pregunta:

¿Cuál es la probabilidad de que en cualquier año se presente un gasto máximo $Q$ que sea menor o igual que otro gasto $q$?

Para darle solución a esta pregunta se utiliza la teoría de la probabilidad elemental; al considerarse que:

Las variables que se manejan en hidrología son de tipo continuo, y se definen por medio de una función de distribución de probabilidad acumulada o función de distribución.

Para conocer la probabilidad del evento ($Q \leq q$), se aplica la probabilidad total

Sea ($Q \leq q$), un evento en un espacio muestral $S$ y sean $A$ y $\overline{A}$, eventos colectivamente excluyentes cuya unión es el espacio $S$. Entonces:

$$P(Q \leq q) = P(A) \cdot P(Q \leq q \mid A) + P(\overline{A}) \cdot P(Q \leq q \mid \overline{A}) \tag{3.73}$$

donde:

$A$ año no ciclónico
$\overline{A}$ año ciclónico

Al tomar en cuenta la probabilidad para años no ciclónicos

$$P(Q \leq q \mid A) = F_1(q) \tag{3.74}$$

y la probabilidad de años ciclónicos (gastos máximo) es

$$P(Q_{máx} \leq q \mid \overline{A}) = F_1(q)\, F_2(q) \tag{3.75}$$

sustituyendo las ecuaciones (3.74) y (3.75) en la ecuación (3.73)

$$P(Q \leq q) = F_1(q)\, P(A) + F_1(q)\, F_2(q)\, P(\overline{A}) \tag{3.76}$$

si por otro lado se hace

$$p = P(A) \tag{3.77}$$

$$F(q) = P(Q \leq q) \tag{3.78}$$

sustituyendo la ecuaciones (3.77) y (3.78) en la ecuación (3.76), además de aplicar el Teorema 1.1 de probabilidad (Ecn. 3.5) resulta:

$$F(q) = F_1(q)\left[p + (1 - p)\, F_2(q)\right] \tag{3.79}$$

finalmente, la función de distribución de probabilidad Gumbel dos poblaciones para gastos máximos anuales está dada por:

$$F(q) = e^{-e^{-\left(\frac{q-\beta_1}{\alpha_1}\right)}}\left[p + \left((1-p)\, e^{-e^{-\left(\frac{q-\beta_2}{\alpha_2}\right)}}\right)\right] \tag{3.80}$$

y la función densidad de probabilidad es:

$$f(q) = e^{-e^{-\left(\frac{q-\beta_1}{\alpha_1}\right)}}\left\{\frac{p}{\alpha_1}\, e^{-\left(\frac{q-\beta_1}{\alpha_1}\right)} + \frac{(1-p)}{\alpha_1\alpha_2}\, e^{-e^{-\left(\frac{q-\beta_2}{\alpha_2}\right)}}\left[\alpha_2\, e^{-\left(\frac{q-\beta_1}{\alpha_1}\right)} + \alpha_1\, e^{-\left(\frac{q-\beta_2}{\alpha_2}\right)}\right]\right\} \tag{3.81}$$

para:
$$F(q)\begin{cases} q > 0 \\ \alpha_i > 0 \\ 0 < p < 1 \end{cases}$$

donde:

$q$ gasto máximo anual, en ($m^3/s$)
$\alpha_1$ parámetro de escala de la población no ciclónica, en ($m^3/s$)
$\beta_1$ parámetro de ubicación de la población no ciclónica, en ($m^3/s$)
$\alpha_2$ parámetro de escala de la población ciclónica, en ($m^3/s$)
$\beta_2$ parámetro de ubicación de la población ciclónica, en ($m^3/s$)
$p$ probabilidad de que se presenten eventos no ciclónicos

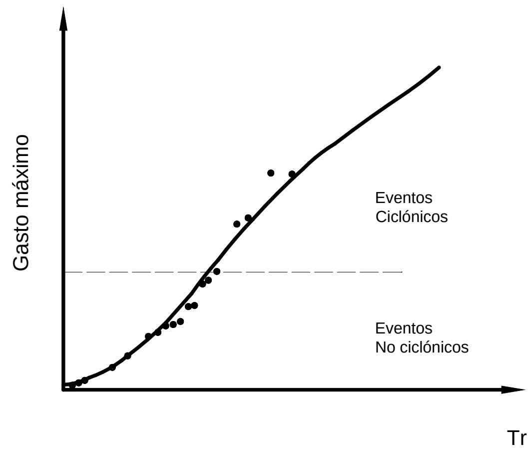


**Figura 3.12. Distribución de probabilidad Gumbel dos poblaciones.**

- ***Estimación de parámetros estadísticos por medio del algoritmo de optimización no lineal de Rosenbrock***

Una primera aproximación de los cinco parámetros estadísticos $\alpha_1, \beta_1, \alpha_2, \beta_2$ y $P$, se obtiene aplicando la técnica de momentos (Ecns. 3.57 y 3.58) para cada población por separado, es decir:

$$\alpha_1 = \frac{\sqrt{6}}{\pi} S_1 \qquad \beta_1 = \overline{x}_1 - \left( \frac{\sqrt{6}}{\pi} \Gamma \right) S_1$$
$$\alpha_2 = \frac{\sqrt{6}}{\pi} S_2 \qquad \beta_2 = \overline{x}_2 - \left( \frac{\sqrt{6}}{\pi} \Gamma \right) S_2 \tag{3.82}$$

Para estimar los estadísticos muestrales $\overline{x}_i$ y $S_i$ de la función G2P, primero se debe definir el **número de Gastos Máximos Ciclónicos *nqc***, conocido este valor, se establece el rango de valores que tendrá la población ciclónica (2) y no ciclónica (1).

Según ***Campos*** (2006) una ecuación para estimar la probabilidad $p$ en que se presentan eventos no ciclónicos, es la siguiente:

$$p = \frac{n - nqc}{n} \tag{3.83}$$

donde:

$n$ tamaño de la muestra de gastos máximos anuales
$nqc$ número de gastos máximos anuales producidos por ciclones

González (1970), propuso para optimizar los cinco parámetros estadísticos, minimizar la función de errores cuadráticos pesados $E\left(\alpha_1,\beta_1,\alpha_2,\beta_2,P\right)$, de la siguiente manera:

La distribución de probabilidad empírica de gastos máximos anuales se estima mediante la ley de Weibull (para muestras ordenadas de menor a mayor)

$$F\left(q_{empírico}\right)=\frac{k}{n+1} \tag{3.84}$$

donde:

$F\left(q_{emp}\right)$ distribución de probabilidad empírica de Weibull para eventos máximos
$k$ número de orden del gasto máximo anual
$n$ tamaño de la muestra de gastos máximos anuales

De tal forma que la función de errores cuadráticos pesados se compone por la suma que resulta de la diferencia de los eventos calculados con la distribución G2P (Ecn. 3.80) y los eventos empíricos calculados con la ley de Weibull (Ecn. 3.84)

$$E\left(\alpha_1,\beta_1,\alpha_2,\beta_2,P\right)=\sum_{i=1}^{n}\left[F(q)-F\left(q_{empírico}\right)\right]^2 w_i \tag{3.85}$$

donde:

$E\left(\alpha_1,\beta_1,\alpha_2,\beta_2,P\right)$ función de errores cuadráticos pesados
$w_i$ peso asignado al error cometido en el evento i

La función de errores cuadráticos pesados es una función no lineal, unimodal y dependiente de múltiples variables, por lo tanto para minimizar la función se tendrá que aplicar un método de optimización no lineal. El algoritmo que se aplicará para minimizar la función es el de ***Rosenbrock para variables no restringidas***.

En el ***Apéndice A*** se muestra el código fuente programado en lenguaje FORTRAN 2003, incluidas sus actualizaciones que ha sufrido el presente algoritmo. En ese mismo apartado se especifican diferentes bibliografías sobre las modificaciones de tal método.

El algoritmo de Rosenbrock requiere condiciones iniciales, algunas de ellas son por ejemplo los parámetros estadísticos, determinados mediante la técnica de momentos como se muestra en las ecuaciones (3.82) y (3.83).

- ***Estimación de eventos de diseño ajustados con la Función Gumbel dos poblaciones (G2P)***

Si ya se dispone de los parámetros estadísticos óptimos de la función, el siguiente paso será estimar los gastos máximos de diseño para diferentes periodos de retorno, el procedimiento es el siguiente:

A la función de distribución de probabilidad G2P (Ecn. 3.80) se le asigna una probabilidad de no excedencia para cada evento ***i***, mejor dicho, igualamos las ecuaciones (3.48) con (3.80)

$$P(Q \leq q) = 1 - \frac{1}{Tr} = e^{-e^{-\left(\frac{q-\beta_1}{\alpha_1}\right)}} \left[ p + \left( (1-p)\, e^{-e^{-\left(\frac{q-\beta_2}{\alpha_2}\right)}} \right) \right] \tag{3.86}$$

realizando el cambio de variable $q_i = Q_{calculado\ i}$ resulta:

$$F(Q_{calculado\ i}) = \left\{ e^{-e^{-\left(\frac{Q_{calculado\ i}-\beta_1}{\alpha_1}\right)}} \left[ p + \left( (1-p)\, e^{-e^{-\left(\frac{Q_{calculado\ i}-\beta_2}{\alpha_2}\right)}} \right) \right] \right\} - \left[ 1 - \left( \frac{1}{T_r} \right) \right] \tag{3.87}$$

de donde:

$$\left\{ e^{-e^{-\left(\frac{Q_{calculado\ i}-\beta_1}{\alpha_1}\right)}} \left[ p + \left( (1-p)\, e^{-e^{-\left(\frac{Q_{calculado\ i}-\beta_2}{\alpha_2}\right)}} \right) \right] \right\} - \left[ 1 - \left( \frac{1}{T_r} \right) \right] = 0 \tag{3.88}$$

La expresión anterior es una ecuación trascendente y la variable $Q_{calculado\ i}$ no se puede obtener de forma explícita, de modo que se resuelve numéricamente. Un método numérico para determinar cada evento $Q_{calculado\ i}$ es el de Newton-Raphson.

la derivada de la función es:

$$\begin{aligned} \frac{dF(Q_{calculado\ i})}{dQ_{calculado\ i}} = e^{-e^{-\left(\frac{Q_{calculado\ i}-\beta_1}{\alpha_1}\right)}} & \left\{ \frac{p}{\alpha_1}\, e^{-\left(\frac{Q_{calculado\ i}-\beta_1}{\alpha_1}\right)} + \frac{(1-p)}{\alpha_1\alpha_2}\, e^{-e^{-\left(\frac{Q_{calculado\ i}-\beta_2}{\alpha_2}\right)}} * \right. \\ & \left. * \left[ \alpha_2\, e^{-\left(\frac{Q_{calculado\ i}-\beta_1}{\alpha_1}\right)} + \alpha_1\, e^{-\left(\frac{Q_{calculado\ i}-\beta_2}{\alpha_2}\right)} \right] \right\} \end{aligned} \tag{3.89}$$

Finalmente para conocer diferentes eventos de diseño $Q_{calculado\ i}$ por medio de la función G2P, se resuelve la ecuación (3.88) para cada evento ***i***, para un determinado periodo de retorno Tr. El código fuente para estimar los $Q_{calculado\ i}$ se muestra en el ***Apéndice B*** y fue programado en lenguaje FORTRAN 2003.

## 3.8. Estimación del Error Estándar de Ajuste EEA

Con el propósito de comparar la eficiencia del ajuste realizado a la muestra con otros modelos matemáticos, se calcula el error estándar de ajuste como: (Kite, G., 1988)

$$EEA = \left[ \frac{\sum_{i=1}^{n} \left( Q_{medidos\ i} - Q_{calculados\ i} \right)^2}{n - np} \right]^{1/2} \tag{3.90}$$

donde:

*np* número de parámetros de la distribución ajustada

## 3.9. Análisis gráfico de las distribuciones de probabilidad Gumbel y Gumbel dos poblaciones.

Para dibujar las curvas de las funciones Gumbel y G2P de manera adecuada es necesario contar con papel de tipo Gumbel. El objetivo de realizar las curvas es verificar de manera visual el ajuste de valores medidos $Q_{medido\ i}$ con los calculados $Q_{calculado\ i}$. La gráfica se dibuja desarrollando el siguiente procedimiento:

En el eje de las abscisas los valores se calculan mediante la expresión de la variable reducida ***z*** (Ecn. 3.56), en el eje de las ordenadas se sitúan los valores de $Q_{medido\ i}$ y $Q_{calculado\ i}$ (Fig. 3.13).

Con la variable reducida en el eje de las abscisas, la curva tiende a generar dos rectas, una para cada población, debido a que se está utilizando un papel diseñado para generar líneas rectas en las distribuciones. Esto se hace con el fin de hacer extrapolaciones lineales y así evitar extrapolaciones no lineales (Linsley et al., 1975).

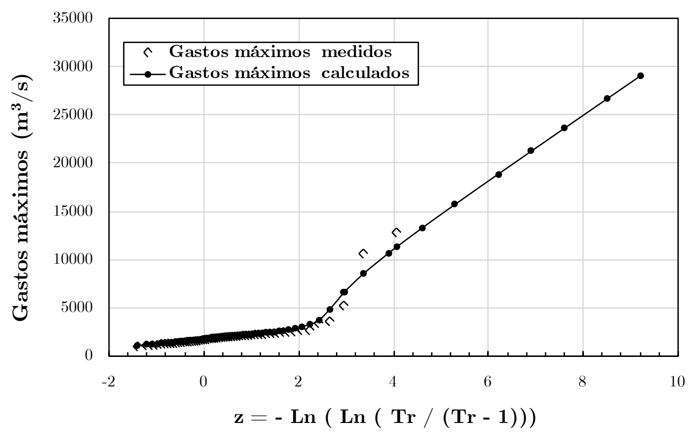


**Figura 3.13. Ajuste con la distribución de probabilidad G2P.**

Otra forma de dibujar la gráfica de la distribución G2P es en la abscisa $x_1$ la probabilidad de no excedencia, en la abscisa $x_2$ el periodo de retorno y en las ordenadas los gastos máximos observados y los calculados (Fig. 3.14).

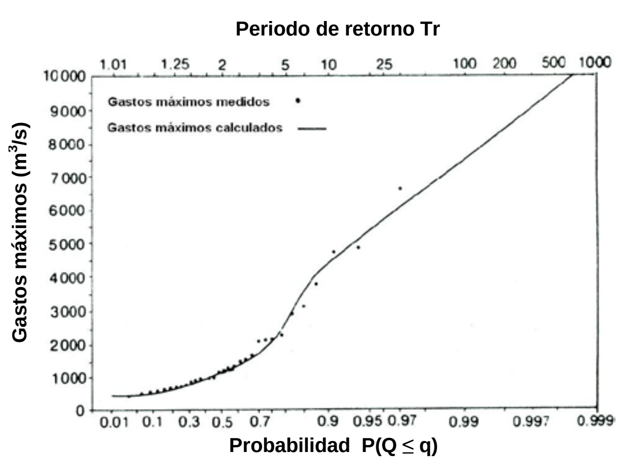


**Figura 3.14. Ajuste con la distribución de probabilidad G2P.**

# Referencias

3.1 Alvarado, C.A.J. ***"Cálculo de Avenidas de Diseño para vertedores de presas de almacenamiento"***. Tesis de Maestría, DEPFI.UNAM.1993.

3.2 Aparicio, M.F.J. ***"Fundamentos de Hidrología de superficie"***. Limusa, México, 2005.

3.3 Benjamín, J.R., Coronell, C.A. ***"Probability, statistics and decision for civil engineers"***. McGraw-Hill. New York, 1970.

3.4 Berezowsky V.M,. Fuentes M.O.***"Métodos Numéricos, Capítulo A.2.16.1 Ajuste de los parámetros de las funciones de distribución de probabilidad. Manual de Diseño de Obras Civiles"***. CFE. México, 1981.

3.5 Borras, H. ***"Apuntes de Probabilidad y Estadística"***. Facultad de Ingeniería, UNAM. 1985.

3.6 Campos, A.D.F. ***"Análisis probabilístico univariado de datos hidrológicos"***. Asociación mexicana de hidráulica: Instituto mexicano de tecnología de agua. México,2006.

3.7 Campos, A.D.F. ***"Estimación de los parámetros óptimos de la distribución Gumbel mixta por medio del algoritmo de Rosenbrock "***. Revista Ingeniería hidráulica en México, Vol. IV No 1, Enero-abril, 1989. pp 9-18.

3.8 Chow, V.T. ***"Applied Hydrology"***. McGraw-Hill, USA, 1988.

3.9 Domínguez M. R., Arganis, J. M. L., ***"Cálculo de registros sintéticos de ingresos por cuenca propia de un sistema de presas de la región noroeste de México caracterizada por eventos invernales"***. Para la Revista Ingeniería, Investigación y Tecnología. Vol. X No. 4, Octubre-diciembre, 2009. pp 353-361.

3.10 Domínguez M.R,. Fuentes M.O., Franco, V. ***"Avenida de diseño, Capítulo A.1.10 del Manual de Diseño de Obras Civiles"***. CFE. México, 1981.

3.11 Domínguez, M.R,. Carlóz, G.T. ***"Análisis Estadístico, Capítulo A.1.6 del Manual de Diseño de Obras Civiles"***. CFE. México, 1981.

3.12 Escalante, S.C., Reyes, C.L ***"Técnicas estadísticas en Hidrología"***. Facultad de Ingeniería, UNAM. 2005.

3.13 González, V.F. ***"Contribución al análisis de frecuencias de valores extremos de los gastos máximos en un río"***. Pub 277. Instituto de Ingeniería, UNAM, México, 1970.

3.14 Gumbel, E.J. ***"Statistics of extremes"***. Columbia Univ. Press, 1958.

3.15 Hines, W.W., Montgomery, C.D. ***"Probabilidad y estadística para ingeniería"***. CECSA, 2000.

3.16 Kite, G. ***"Frequency and risk analyses in Hydrology"***. Water Resources Publications, Fort Collins, Colorado, USA, 1988.

3.17 Luthe, R., Olivera, A., Schutz, F. ***"Métodos Numéricos"***. Limusa. México, 1985.

3.18 Lynsley, R.K., Kohler, M.A., Paulhus, J.L.H. ***"Hydrology for engineers"***. McGraw-Hill, 1975.

3.19 Monsalve, S.G. ***"Hidrología en la Ingeniería"***. Alfaomega, México, 1999.

3.20 Ocegueda, H.V.M. ***"Avenidas de Diseño"***. Tesis de Licenciatura Facultad de Ingeniería, UNAM. 1987.

3.21 Olivera, S.A. ***"Serie de Probabilidad y Estadística"***. 7 volúmenes, Limusa, México, 1987.

3.22 Raynal, V., Guevara J. ***"Maximum likelihood estimators for the two populations Gumbel distributions"***. Hydrological Science and Technology Journal, 13(1-4): 47-56. 1997.

3.23 Ruíz, U.M.R. "***Programa de automatización de los métodos estadísticos en hidrología"***. Facultad de Ingeniería, UNAM. 2002.

3.24 V. Balderrama. ***"Métodos Numéricos"***. Trillas, México, 1990.

3.25 Yevjevich, V. ***"Probability and Statistics in Hydrology"***. Water Resources Publications, Fort Collins, Colorado, USA, 1972.