

# Probabilidad y Estadística

George Canavos

Apuntes por FODE  
(Christian L. Paredes Aguilera.)

# Contents

<b>1</b>	<b>Introducción y estadística descriptiva</b>	<b>3</b>
1.3	Medidas numéricas descriptivas . . . . .	3
<b>2</b>	<b>Conceptos en probabilidad</b>	<b>6</b>
2.5	Desarrollo axiomático de la probabilidad . . . . .	6
2.6	Probabilidad conjunta, marginal y condicional . . . . .	9
2.7	Eventos estadísticamente independientes . . . . .	9
2.8	El teorema de Bayes . . . . .	9
2.9	Ejercicios . . . . .	10
<b>3</b>	<b>Variables aleatorias y distribución de probabilidad</b>	<b>11</b>
3.1	El concepto de variables aleatorias . . . . .	11
3.2	Distribuciones de probabilidad de variables aleatorias discretas . . . . .	12
3.3	Distribuciones de probabilidad de variables aleatorias continuas . . . . .	12
3.4	Valor esperado de una variable aleatoria . . . . .	13
3.4.1	Propiedades . . . . .	13
3.5	Momentos de una variable aleatoria . . . . .	14
3.6	Otras medidas de tendencia central y dispersión . . . . .	17
3.7	Funciones generadoras de momentos . . . . .	18
3.8	Ejercicios . . . . .	20
<b>4</b>	<b>Algunas distribuciones discretas de probabilidad</b>	<b>21</b>
4.2	La distribución binomial . . . . .	21
4.3	La distribución de Poisson . . . . .	25
4.4	La distribución hipergeométrica . . . . .	30
4.5	La distribución binomial negativa . . . . .	31
<b>5</b>	<b>Algunas distribuciones continuas de probabilidad</b>	<b>33</b>
5.1	La distribución normal . . . . .	33
5.2	La distribución uniforme . . . . .	38
5.3	La distribución beta . . . . .	40
5.4	La distribución gama . . . . .	42
5.5	La distribución de Weibull . . . . .	45
5.6	La distribución exponencial negativa . . . . .	47
5.7	La distribución de una función de variable aleatoria . . . . .	49
5.8	Conceptos básicos en la generación de números aleatorios por computadora . . . . .	50
5.8.1	Distribución uniforme sobre el intervalo (a,b) . . . . .	50
5.8.2	La distribución de Weibull . . . . .	50
5.8.3	La distribución de Erlang . . . . .	51
5.8.4	La distribución normal . . . . .	51
5.8.5	La distribución binomial . . . . .	51
5.8.6	La distribución Poisson . . . . .	52

<b>6</b>	<b>Distribuciones conjuntas de probabilidad</b>	<b>53</b>
6.1	Distribución de probabilidad bivariadas . . . . .	53
6.2	Distribuciones marginales de probabilidad . . . . .	54
6.3	Valores esperados y momentos para distribuciones bivariadas . . . . .	55
6.4	Variables aleatorias estadísticamente independientes . . . . .	57
6.5	Distribuciones de probabilidad condicional . . . . .	58
6.6	Análisis bayesiano: las distribuciones a priori y a posteriori. . . . .	60
6.7	La distribución normal bivariada . . . . .	60
<b>7</b>	<b>Muestras aleatorias y distribuciones de muestreo</b>	<b>62</b>
7.3	Distribuciones de muestreo de estadísticas . . . . .	63
7.4	La distribución de muestreo de $\bar{X}$ . . . . .	65
7.5	La distribución de muestreo de $S^2$ . . . . .	69
7.6	La distribución t de Student . . . . .	71
7.7	La distribución de la diferencia entre dos medias muestrales . . . . .	74
7.8	La distribución F . . . . .	75
<b>8</b>	<b>Estimación puntual y por intervalo</b>	<b>78</b>

# Introducción y estadística descriptiva

## 1.3 Medidas numéricas descriptivas

**Definición 1.1** La media de las observaciones  $x_1, x_2, \dots, x_n$  es el promedio aritmético de éstas y se denota por

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} \quad (1.1)$$

**Definición 1.2** La mediana de un conjunto de observaciones es el valor para el cual, todas las observaciones se ordenan de manera creciente, la mitad de éstas es menor que este valor y la otra mitad mayor.

**Definición 1.3** La moda de un conjunto de observaciones es el valor de la observación que ocurre con mayor frecuencia en el conjunto.

$$\bar{x} = \sum_{i=1}^k \frac{f_i x_i}{n} \quad (1.2)$$

$$Mediana = L + c \left( \frac{j}{f_m} \right) \quad (1.3)$$

**Definición 1.4** La varianza de las observaciones  $x_1, x_2, \dots, x_n$  es, en esencia, el promedio del cuadrado de las distancias entre cada observación y la media del conjunto de observaciones. La varianza se denota por

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} \quad (1.4)$$

**Definición 1.5** La raíz cuadrada positiva de la varianza recibe el nombre de la desviación estándar y se denota por

$$s = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}} \quad (1.5)$$

El uso de la ecuación (1.4) puede dar origen a errores grandes por redondeo. Con un poco de álgebra se obtiene, a partir de (1.4), una fórmula computacional más exacta para esas condiciones:

$$\begin{aligned} s^2 &= \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} = \frac{\sum (x_i^2 - 2\bar{x}x_i + \bar{x}^2)}{n-1} = \frac{\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2}{n-1} = \frac{\sum x_i^2 - \frac{2(\sum x_i)(\sum x_i)}{n} + \frac{n(\sum x_i)^2}{n^2}}{n-1} = \\ s^2 &= \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1} \end{aligned} \quad (1.6)$$

$$s = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1}} \quad (1.7)$$

Para datos agrupados, puede calcularse el valor aproximado de la varianza mediante el uso de la fórmula

$$s^2 = \frac{\sum_{i=1}^k f_i(x_i - \bar{x})^2}{n-1} \quad (1.8)$$

6

$$s^2 = \frac{\sum_{i=1}^k f_i x_i^2 - \frac{\left(\sum_{i=1}^k f_i x_i\right)^2}{n}}{n-1} \quad (1.9)$$

La fórmula para la desviación estándar es

$$s = \sqrt{\sum_{i=1}^k \frac{f_i(x_i - \bar{x})^2}{n-1}} \quad (1.10)$$

**Definición 1.6** La desviación media es el promedio de los valores absolutos de las diferencias entre cada observación y la media de las observaciones. La desviación media está dada por

$$D.M. = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

Para datos agrupados, el valor de la desviación media se aproxima por

$$D.M. = \frac{\sum_{i=1}^k f_i |x_i - \bar{x}|}{\sum_{i=1}^k f_i} \quad (1.11)$$

**Definición 1.7** La desviación mediana es el promedio de los valores absolutos de las diferencias entre cada observación y la mediana de éstas. Esta dada por:

$$D.M. = \frac{\sum_{i=1}^n |x_i - D.Md|}{n} \quad (1.12)$$

## Conceptos en probabilidad

**Definición 2.1** Si un experimento que está sujeto al azar resulta de  $n$  formas igualmente probables y mutuamente excluyentes y si  $n_A$  de estos resultados tienen un atributo  $A$ , la probabilidad de  $A$  es la proporción de  $n_A$  co respecto a  $n$ .

**Definición 2.2** Si un experimento se repite  $n$  veces bajo las mismas condiciones y  $n_B$  de los resultados son favorables a un atributo  $B$ , el límite de  $n_B/n$  conforme  $n$  se vuelve grande, se define como la probabilidad del atributo  $B$ .

**Definición 2.3** El conjunto de todos los posibles resultados de un experimento aleatorio recibe el nombre de espacio muestral.

### 2.5 Desarrollo axiomático de la probabilidad

**Definición 2.4** Se dice que un espacio muestral es discreto si su restado puede ponerse en una correspondencia uno a uno con el conjunto de los enteros positivos.

**Definición 2.5** Se dice que un espacio muestral es continuo si sus resultados consisten de un intervalo de números reales.

**Definición 2.6** un evento del espacio muestral es un grupo de resultados contenidos en éste, cuyos miembros tienen una característica en común.

**Definición 2.7** El evento que contiene a ningún resultado del espacio muestral recibe el nombre de evento nulo o vacío.

**Definición 2.8** El evento formado por todos los posibles resultados en  $E_1$  o  $E_2$  o en ambos, recibe el nombre de unión de  $E_1$  y  $E_2$  y se denota por  $E_1 \cup E_2$ .

**Definición 2.9** El evento formado por todos los resultados comunes tanto a  $E_1$  como a  $E_2$  recibe el nombre de intersección de  $E_1$  y  $E_2$  y se denota por  $E_1 \cap E_2$ .

**Definición 2.10** Se dice que los eventos  $E_1$  y  $E_2$  son mutuamente excluyentes o disjuntos si no tienen resultados en común; en otras palabras  $E_1 \cap E_2 = \emptyset =$  evento vacío.

**Definición 2.11** Si cualquier resultado de  $E_2$  también es un resultado de  $E_1$ , se dice que el evento  $E_2$  está contenido en  $E_1$ , y se denota por  $E_2 \subset E_1$

**Definición 2.12** El complemento de un evento  $E$  con respecto al espacio muestral  $S$ , es aquel que contiene a todos los resultados de  $S$  que no se encuentran en  $E$ , y se denota por  $\bar{E}$ .



**Definición 2.13** Sean  $S$  cualquier espacio muestral y  $E$  cualquier evento de éste. Se llamará función de probabilidad sobre el espacio muestral  $S$  a  $P(E)$  si satisface los siguientes axiomas:

1.  $P(E) \geq 0$ .
2.  $P(S) = 1$ .
3. Si, para los eventos  $E_1, E_2, E_3, \dots$   $E_i \cap E_j = \emptyset$  para todo  $i \neq j$  entonces

$$P(E_1 \cup E_2 \cup \dots) = P(E_1) + P(E_2) + \dots$$

**Teorema 2.1**  $P(\emptyset) = 0$ .

2.1

Demostración.-

$$S \cup \emptyset = S \quad y \quad S \cap \emptyset = \emptyset.$$

por el axioma 3,

$$P(S \cup \emptyset) = P(S) + P(\emptyset);$$

pero por el axioma 2,  $P(S) = 1$ , y de esta manera  $P(\emptyset) = 0$ .

■

**Teorema 2.2** Para cualquier evento  $E \subset S$ ,  $0 \leq P(E) \leq 1$ .

2.2

Demostración.- Por el axioma 1,  $P(E) \geq 0$ ; de aquí sólo es necesario probar que  $P(E) \leq 1$ .

$$E \cup \bar{E} = S \quad y \quad E \cap \bar{E} = \emptyset.$$

Por los axiomas 2 y 3,

$$P(E \cup \bar{E}) = P(E) + P(\bar{E}) = P(S) = 1;$$

dado que  $P(\bar{E}) \geq 0$ ,  $P(E) \leq 1$ .

■

**Teorema 2.3** Sea  $S$  un espacio muestral que contiene a cualesquiera dos eventos  $A$  y  $B$ ; entonces,

2.3

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

■

## 2.6 Probabilidad conjunta, marginal y condicional

**Definición 2.14** Sean  $A$  y  $B$  cualesquiera dos eventos que se encuentran en un espacio muestral  $S$  de manera tal que  $P(B) > 0$ . La probabilidad condicional de  $A$  al ocurrir el evento  $B$ , es el cociente de la probabilidad conjunto de  $A$  y  $B$  con respecto a la probabilidad marginal de  $B$ ; de esta manera se tiene

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0. \quad (2.1)$$

$$P(A \cap B) = P(B)P(A|B). \quad (2.2)$$

La definición 2.14 puede extenderse para incluir cualquier número de eventos que se encuentren en el espacio muestral. Por ejemplo, puede demostrarse que para tres eventos  $A, B$  y  $C$

$$P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)}, \quad P(B \cap C) > 0. \quad (2.3)$$

$$P(A \cap B|C) = \frac{P(A \cap B \cap C)}{P(C)}, \quad P(C) > 0. \quad (2.4)$$

## 2.7 Eventos estadísticamente independientes

**Definición 2.15** Sean  $A$  y  $B$  dos eventos cualesquiera de un espacio muestral  $S$ . Se dice que el evento  $A$  es estadísticamente independiente del evento  $B$  si  $P(A|B) = P(A)$ .

$$P(A|B) = P(A)$$

**Definición 2.16** Los eventos  $A_1, A_2, \dots, A_k$  de un espacio muestral  $S$  son estadísticamente independientes si y sólo si la probabilidad conjunta de cualquier  $2, 3, \dots, k$  de ellos es igual al producto de sus respectivas probabilidades marginales.

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

## 2.8 El teorema de Bayes

**Teorema 2.4** Si  $B_1, B_2, \dots, B_n$  son  $n$  eventos mutuamente excluyentes, de los cuales uno debe ocurrir, es decir  $\sum_{i=1}^n P(B_i) =$

1 entonces,

$$P(B_j \setminus A) = \frac{P(B_j)P(A \setminus B_j)}{\sum_{i=1}^n P(B_i)P(A \setminus B_i)} \quad (2.5)$$

■

## 2.9 Ejercicios

2.2. Demostración.- sabiendo que  $P(\bar{A} \cap B) = P(B) - P(A \cap B)$ , entonces

$$\frac{P(A \setminus B)}{P(A)} + \frac{P(\bar{A} \setminus B)}{P(B)} = \frac{P(A \cap B) + P(B) - P(A \cap B)}{P(B)} = 1$$

2.3. Demostración.- Supongamos que los eventos  $A$  y  $B$  son no vacíos, por definición de evento independientes y mutuamente excluyentes tenemos que

$$P(A \cap B) = P(A) \cdot P(B) \neq \emptyset$$

Luego se cumple que dos eventos independientes son, también mutuamente excluyentes si por lo menos uno de los eventos es vacío.

## Variables aleatorias y distribución de probabilidad

### 3.1 El concepto de variables aleatorias

**Definición 3.1** Sea  $S$  un espacio muestral sobre el que se encuentra definida una función de probabilidad. Sea  $X$  una función de valor real definida sobre  $S$ , de manera que transforme los resultados de  $S$  en puntos sobre la recta de los reales. Se dice entonces que  $X$  es un variable aleatoria.

**Definición 3.2** Se dice que una variable aleatoria  $X$  es discreta si el número de valores que puede tomar es contable (ya sea finito o infinito), y si éstos pueden arreglarse en una secuencia que corresponde con los enteros positivos.

**Definición 3.3** Se dice que una variable aleatoria  $X$  es continua si sus valores consisten en uno o más intervalos de la recta de los reales.

## 3.2 Distribuciones de probabilidad de variables aleatorias discretas

**Definición 3.4** Sea  $X$  una variable aleatoria discreta. Se llamará a  $p(x) = P(X = x)$  función de probabilidad de la variable aleatoria  $X$ , si satisface las siguientes propiedades:

1.  $p(x) \geq 0$  para todos los valores  $x$  de  $X$ ;
2.  $\sum_x p(x) = 1$ .

**Definición 3.5** La función de distribución acumulativa de la variable aleatoria  $X$  es la probabilidad de que  $X$  sea menor o igual a un valor específico de  $x$  y está dada por:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

En general, la función de distribución acumulativa  $F(x)$  de una variable aleatoria discreta es una función no decreciente de los valores de  $X$ , de tal manera que:

1.  $0 \leq F(x) \leq 1$  para cualquier  $x$ ;
2.  $F(x_i) \geq F(x_j)$  si  $x_i \geq x_j$ ;
3.  $P(X > x) = 1 - F(x)$ .
4.  $P(X = x) = F(x) - F(x - 1)$ ;
5.  $P(x_i \leq X \leq x_j) = F(x_j) - F(x_i - 1)$

## 3.3 Distribuciones de probabilidad de variables aleatorias continuas

**Definición 3.6** 1.  $f(x) \geq 0, -\infty < x < \infty$ ,

2.  $\int_{-\infty}^{\infty} f(x) dx$  y

3.  $P(a \leq X \leq b) = \int_a^b f(x) dx$

Para cualquier  $a$  y  $b$ , entonces  $f(x)$  es la función de densidad de probabilidad de la variable aleatoria continua  $X$ .

Para la función de distribución acumulativa  $F(x)$  se tiene:

$$P(X \leq x) = F(x) = \int_{-\infty}^x f(t) dt$$

Dado que para cualquier variable aleatoria continua  $X$ ,

$$P(X = x) = \inf_x^x f(t) dt = 0, \quad \implies \quad P(X \leq x) = P(X < x) = F(x)$$

La distribución acumulativa  $F(x)$  es una función lisa no decreciente de los valores de la v.a. con las siguientes propiedades:

1.  $F(-\infty) = 0$ ;
2.  $F(\infty) = 1$ ;
3.  $P(a < X < b) = F(b) - F(a)$
4.  $dF(x)/dx = f(x)$ .

### 3.4 Valor esperado de una variable aleatoria

**Definición 3.7** El valor esperado de una variable aleatoria  $X$  es el promedio o valor medio de  $X$  y está dado por:

$$E(X) = \sum_x x p(x) \quad \text{Si } x \text{ es discreta, o}$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad \text{Si } x \text{ es continua.}$$

En donde  $p(x)$  y  $f(x)$  son las funciones de probabilidad y de densidad de probabilidad, respectivamente.

En general, el valor esperado de una función  $g(x)$  de la variables aleatoria  $X$ , está dado por:

$$E(g(X)) = \sum_x g(x) p(x) \quad \text{Si } x \text{ es discreta, o}$$

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) x f(x) dx \quad \text{Si } x \text{ es continua.}$$

#### 3.4.1 Propiedades

1. El valor esperado de una constante  $c$  es el valor de la constante.

$$E(c) = \int_{-\infty}^{\infty} c f(x) dx = c \int_{-\infty}^{\infty} f(x) dx = c$$

2. El valor esperado de la cantidad  $aX + b$ , en donde  $a$  y  $b$  son constantes, es el producto de  $a$  por el valor esperado de  $x$  más  $b$ .

$$E(aX + b) = \int_{-\infty}^{\infty} (ax + b) f(x) dx = a \int_{-\infty}^{\infty} x f(x) dx + b \int_{-\infty}^{\infty} f(x) dx = aE(X) + b$$

3. El valor esperado de la suma de dos funciones  $g(X)$  y  $h(X)$  de  $X$  es la suma de los valores esperados de  $g(X)$  y  $h(X)$ .

$$E[g(X) + h(X)] = \int_{-\infty}^{\infty} [g(x) + h(x)] dx = \int_{-\infty}^{\infty} g(x)f(x) dx + \int_{-\infty}^{\infty} h(x)f(x) dx = E[g(X)] + E[h(X)]$$

### 3.5 Momentos de una variable aleatoria

Los momentos de una variable aleatoria  $X$  son los valores esperados de ciertas funciones de  $X$ .

**Definición 3.8** Sea  $X$  una variable aleatoria. El  $r$ -ésimo momento de  $X$  alrededor de cero se define por:

$$\mu'_r = E(X^r) = \sum_x x^r p(x) \quad \text{Si } x \text{ es discreta, o}$$

$$\mu'_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx \quad \text{Si } x \text{ es continua.}$$

El primer momento al rededor del cero es la media o valor esperado de la variable aleatoria. y se denota por  $\mu$ ; de ésta manera se tiene que  $\mu'_1 = \mu = E(X)$ .

**Definición 3.9** Sea  $X$  una variable aleatoria. El  $r$ -ésimo momento central de  $X$  o el  $r$ -ésimo momento alrededor de la media de  $X$  se define por:

$$\mu_r = E(X - \mu)^r = \sum_x (x - \mu)^r p(x) \quad \text{Si } x \text{ es discreta, o}$$

$$\mu_r = E(X - \mu)^r = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx \quad \text{Si } x \text{ es continua.}$$

El momento central de cero de cualquier variable aleatoria es uno, dado que:

$$\mu_0 = E(X - \mu)^0 = E(1) = 1$$

De manera similar, el primer momento central de cualquier variables aleatoria es cero, dado que:

$$\mu_1 = E(X - \mu) = E(X) - \mu = 0$$

El segundo momento central:

$$\mu_2 = E(X - \mu)^2$$

Recibe el nombre de varianza de la variable aleatoria. Puesto que:

$$\begin{aligned} \mu_2 = Var(X) &= E[(X - \mu)^2] \\ &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= \mu'_2 - \mu^2 \end{aligned}$$

La varianza de cualquier variable aleatoria es el segundo momento alrededor del origen menos el cuadrado de la media. Generalmente se denota por  $\sigma^2$

Es útil notar que la varianza de una variable aleatoria  $X$  es invariable; es decir,  $Var(X + b) = Var(X)$  para cualquier constante  $b$ . De manera más general, se demostrará que  $Var(aX + b) = a^2 Var(X)$  para cualquiera dos constantes  $a$  y  $b$ . Por definición,

$$\begin{aligned} Var(aX + b) &= E(aX + b)^2 - E^2(aX + b) \\ &= E(a^2X^2 - 2abX + b^2) - [aE(X) + b]^2 \\ &= a^2E(X)^2 + 2abE(X) + b^2 - a^2E^2(X) - 2abE(X) - b^2 \\ &= a^2E(X)^2 - a^2E^2(X) \\ &= a^2 [E(X)^2 - E^2(X)] \\ &= a^2 Var(X) \end{aligned}$$

Una medida que compara la dispersión relativa de dos distribuciones de probabilidad es el coeficiente de variación, que está definido por:

$$V = \frac{\sigma}{\mu}$$

Expresa la magnitud de la dispersión de una variable aleatoria con respecto a su valor esperado.

El tercer momento central

$$\mu_3 = E(X - \mu)^3$$

esta relacionado con la asimetría de la distribución de probabilidad de  $X$ .

Cualquier momento central de una variable aleatoria  $X$  puede expresarse en términos de los momentos de ésta, alrededor de cero. Por definición:

$$u_r = E(X - \mu)^r$$

pero la expansión de  $(X - \mu)^r$  puede expresarse como:

$$(X - \mu)^r = \sum_{i=0}^r (-1)^i \frac{r!}{(r-i)!i!} \mu^i x^{r-i} = \sum_{i=0}^r (-1)^i \frac{r!}{(r-i)!i!} \mu^i E(X^{r-i}) = \sum_{i=0}^r (-1)^i \frac{r!}{(r-i)!i!} \mu^i \mu'_{r-i}$$

En particular,

$$\mu_3 = \mu'_3 - 3\mu\mu'_2 + 2\mu^3$$

Para las distribuciones de probabilidad que presentan un sólo pico, si  $\mu_3 < 0$ , se dice que la distribución es asimétrica negativamente, si  $\mu_3 > 0$ , la distribución es asimétrica positivamente y si  $\mu_3 = 0$ , la distribución recibe el nombre de simétrica.

Una medida más apropiada de la asimetría, es el tercer momento estandarizado, dado por:

$$\alpha_3 = \frac{\mu_3}{(\mu_2)^{3/2}}$$

Que recibe el nombre de coeficiente de asimetría. Una distribución de probabilidad es asimétrica positiva, negativa o simétrica si  $\alpha_3 > 0$ ,  $\alpha_3 < 0$  o  $\alpha_3 = 0$  respectivamente.



El cuarto momento central,

$$\mu_4 = E(X - \mu)^4 = \mu'_4 - 4\mu\mu'_3 + 6\mu^2\mu'_2 - 3\mu^4$$

Es una medida de qué tan puntiaguda es la distribución de probabilidad y recibe el nombre de curtosis. Al igual que para el tercer momento, es preferible emplear el cuarto momento estandarizado,

$$\alpha_4 = \frac{\mu_4}{\mu_2^2}$$

Si  $\alpha_4 > 3$ , la distribución de probabilidad presenta un pico relativamente alto y recibe el nombre de letocúrtica, si  $\alpha_4 < 3$ , la distribución es relativamente plana y recibe el nombre de platicúrtica y si  $\alpha_4 = 3$ , la distribución no presenta un pico muy alto i muy bajo y recibe el nombre de mesocúrtica.

En este momento se considerará el concepto de variable aleatoria estandarizada. Sea  $X$  cualquier variable aleatoria con media  $\mu$  y desviación estándar  $\sigma$  la cantidad

$$Y = \frac{(x - \mu)}{\sigma}$$

define una variable aleatoria  $Y$  con media cero y desviación estándar uno. Esta variable recibe el nombre de **variable aleatoria estandarizada** correspondiente a  $X$

El valor esperado de  $Y$  es cero, puesto que:

$$E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}E(X - \mu) = 0$$

De hecho, puesto que  $E(Y) = 0$ , el  $r$ -ésimo momento central de  $Y$  es:

$$\begin{aligned}\mu_r(Y) &= E(Y^r) \\ &= E\left(\frac{X - \mu}{\sigma}\right)^r \\ &= \frac{1}{\sigma^r}E(X - \mu)^r \\ &= \frac{\mu_r(X)}{\sigma^r}\end{aligned}$$

De esta manera se tiene que:

$$\mu_r(Y) = \frac{\mu_r(X)}{[\mu_2(X)]^{r/2}}$$

Ya que

$$\sigma^r = \sqrt{[\mu_2(X)]^r}$$

es decir,

$$(\sigma^r)^2 = [\mu_2(X)]^r \implies (\sigma^2)^r = [\mu_2(X)]^r \implies (\sigma^2)^r = (\sigma^2)^r$$

de donde se tiene que  $Var(Y) = \mu_2(Y) = 1$ . En particular, nótese que  $\alpha_3(Y) = \alpha_3(X)$  y  $\alpha_4(Y) = \alpha_4(X)$ . La estandarización de una variable aleatoria afecta a la media y a la varianza, pero no a los factores de forma.

### 3.6 Otras medidas de tendencia central y dispersión

**Definición 3.10** Para cualquier variable aleatoria  $X$ , se define a la mediana  $x_{0.5}$  de  $X$ , para ser:

$$P(X < x_{0.5}) \leq 1/2 \quad y \quad P(X \leq x_{0.5}) \geq 1/2 \quad \text{si } X \text{ es discreta, o}$$

$$P(X \leq x_{0.5}) = 1/2 \quad \text{si } X \text{ es continua}$$

**Definición 3.11** Para cualquier variable aleatoria  $X$ , se define la moda como el valor  $x_m$  de  $X$  que maximiza la función de probabilidad si  $X$  es discreta o la función de densidad si  $X$  es continua.

**Definición 3.12** Para cualquier variable aleatoria  $X$ , el valor cuantil  $X_q$ , de orden  $q$ .  $0 < q < 1$ , es el valor de  $X$  tal que:

$$P(X < x_q) \leq q \quad y \quad P(X \leq x_q) \geq q \quad \text{si } X \text{ es discreta, o}$$

$$P(X \leq x_q) = q \quad \text{si } X \text{ es continua}$$

**Definición 3.13** La desviación media de una variable aleatoria  $X$  es el valor esperado de la diferencia absoluta entre  $X$  y su media, y está dado por:

$$E|X - \mu| = \sum |x - \mu| p(x) \quad \text{si } X \text{ es discreta, o}$$

$$E|X - \mu| = \int_{-\infty}^{\infty} |x - \mu| dx \quad \text{si } X \text{ es continua}$$

### 3.7 Funciones generadoras de momentos

**Definición 3.14** Sea  $X$  una variable aleatoria. El valor esperado de  $\exp(tX)$  recibe el nombre de función generadora de momentos, y se denota por  $m_X(t)$ , si el valor esperado existe para cualquier valor de  $t$  en algún intervalo  $-c < t < c$  en donde  $c$  es un número positivo. En otras palabras:

$$m_X(t) = E[\exp(tX)] = \sum_x e^{tx} p(x) \quad \text{si } X \text{ es discreta, o}$$

$$m_X(t) = E[\exp(tX)] = \int_{-\infty}^{\infty} e^{tx} f(x) dx \quad \text{si } X \text{ es continua}$$

Si la función generadora de momentos existe, puede demostrarse que es única y que determina por completo la distribución de probabilidad de  $X$ . En otras palabras **si dos variables aleatorias tienen la misma función generadora de momentos, entonces tienen la misma distribución de probabilidad.**

Si la función generadora de momentos existe para  $-c < t < c$  entonces existen las derivadas de ésta de todas las órdenes para  $t = 0$ . Lo anterior asegura que  $m_X(t)$  generará todos los momentos de  $X$  al rededor del origen. Para demostrar lo anterior se diferencia  $m_X(t)$  con respecto a  $t$ , y se evalúa la derivada en  $t = 0$ . Suponiendo que pueden intercambiarse los símbolos de diferenciación y esperanza, se tiene:

$$\begin{aligned} \left. \frac{d^2 m_X(t)}{dt^2} \right|_{t=0} &= \left. \frac{d}{dt} E[e^{tX}] \right|_{t=0} \\ &= E \left[ \frac{d}{dt} (e^{tX}) \right] \\ &= E (X \cdot e^{tX}) \Big|_{t=0} \\ &= E(X) \\ &= \mu \end{aligned}$$

Al tomar la segunda derivada y evaluar en  $t = 0$ .

$$\begin{aligned}
\left. \frac{d^2 m_X(t)}{dt^2} \right|_{t=0} &= \left. \frac{d^2}{dt^2} E[e^{tX}] \right|_{t=0} \\
&= E \left[ \frac{d^2}{dt^2} (e^{tX}) \right] \\
&= E \left[ \frac{d}{dt} (X \cdot e^{tX}) \right] \\
&= E[X^2 e^{tX}] \Big|_{t=0} \\
&= E(X^2) \\
&= \mu_2'
\end{aligned}$$

Al continuar este proceso de diferenciación se puede deducir que se obtiene el

$$\begin{aligned}
\frac{d^r m_X(t)}{dt^r} &= \left. \frac{d^r}{dt^r} E[e^{tX}] \right|_{t=0} \\
&= E \left[ \frac{d^r}{dt^r} (e^{tX}) \right] \\
&= E[X^r e^{tX}] \Big|_{t=0} \\
&= E(X^r) \\
&= \mu_r'
\end{aligned}$$

mismo resultado si se reemplaza la función exponencial por su expansión en serie de potencias.

$$E[e^{tX}] = E \left( 1 + tX \frac{t^2 X^2}{2!} + \dots + \frac{t^r X^r}{r!} + \dots \right)$$

**Definición 3.15** Sea  $X$  una variable aleatoria. El valor esperado de  $\exp[t(X - \mu)]$  recibe el nombre de función generadora de momentos central y denota por  $m_{X-\mu}(t)$ , si el valor esperado existe para cualquier  $t$  en algún intervalo  $-c < t < c$  en donde  $c$  es un número positivo.

$$m_{X-\mu}(t) = E \exp[t(X - \mu)] = \sum_x \exp[t(x - \mu)] p(x) \quad \text{si } X \text{ es discreta, o}$$

$$m_{X-\mu}(t) = E \exp[t(X - \mu)] = \int_{-\infty}^{\infty} \exp[t(x - \mu)] f(x) dx \quad \text{si } X \text{ es continua}$$

### 3.8 Ejercicios

Los ejercicios se encuentra en el archivo *Ej\_Cap3.Rmd*

## Algunas distribuciones discretas de probabilidad

### 4.2 La distribución binomial

Llámesese éxito a la ocurrencia del evento y fracaso a su no ocurrencia.

Las dos suposiciones claves para la distribución binomial son:

- i) La probabilidad de éxito  $p$  permanece constante para cada ensayo.
- ii) Los  $n$  ensayos son independientes entre sí.

Para obtener la función de probabilidad de la distribución binomial, primero se determina la probabilidad de tener, en  $n$  ensayos,  $x$  éxitos consecutivos seguidos de  $n-x$  fracasos consecutivos. Dado que, por hipótesis, los  $n$  ensayos son independientes de la definición 2.15, se tiene:

$$p \cdot p \cdots p \cdot (1-p) \cdot (1-p) \cdots (1-p) = p^x (1-p)^{n-x}$$

**Definición 4.1** **Distribución binomial con función de probabilidad.** Sea  $X$  una variable aleatoria que representa el número de éxitos en  $n$  ensayos y  $p$  la probabilidad de éxito con cualquiera de éstos. Se dice entonces que  $X$  tiene una distribución binomial con función de probabilidad.

$$p(x; n, p) \begin{cases} \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} & x = 0, 1, 2, \dots, n. \\ 0 \text{ para cualquier otro valor.} & 0 \leq p \leq 1. \text{ para } n \text{ entero.} \end{cases}$$

El nombre distribución binomial proviene del hecho de que los valores de  $p(x; n, p)$  para  $x = 1, 2, \dots, n$  son los términos sucesivos de la expansión binomial de  $[(1-p) + p]^n$ ; esto es,

$$\begin{aligned} [(1-p) + p]^n &= (1-p)^n + n(1-p)^{n-1}p + \frac{n(n-1)}{2!}(1-p)^{n-2}p^2 + \cdots + p^n \\ &= \sum_{x=0}^n \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n p(x; n, p) \end{aligned}$$

Pero dado que  $[(1-p) + p]^n = 1$  y  $p(x; n, p) \geq 0$  para  $x = 0, 1, 2, \dots, n$ , este hecho también verifica que  $p(x; n, p)$  es una función de probabilidad.

La probabilidad de que una variable aleatoria  $X$  sea menor o igual a un valor específico de  $x$ , se determina por la **función de distribución acumulativa**.

$$P(X \leq x) = F(x; n, p) = \sum_{i=0}^x \binom{n}{i} p^i (1-p)^{n-i}$$

Nótese que si  $n = 1$ , la función de probabilidad binomial se deduce a

$$p(x; n, p) = \begin{cases} p^x (1-p)^{1-x} & x = 0, 1 \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Que es la **función de probabilidad de la distribución puntual o Bernoulli**.

Por definición 3.8, el primer momento alrededor del cero de la variable aleatoria binomial  $X$  es el valor esperado de  $X$ .

$$\begin{aligned} E(X) &= x \sum_{x=0}^n \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \\ &= x \sum_{x=1}^n \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^n \frac{n!}{(n-x)!(x-1)!} p^x (1-p)^{n-x} \end{aligned}$$

en donde se ha escrito la suma desde uno hasta  $n$ , dado que cuando  $x = 0$  el primer término es cero y se cancela la  $x$  del numerador con la  $x$  en  $x!$ . Factorizando  $n$  y  $p$ , se tiene:

$$E(X) = np \sum_{x=1}^n \frac{(n-1)!}{(n-x)!(x-1)!} p^{x-1} (1-p)^{n-x}$$

Si  $y = x - 1$  y  $m = n - 1$ , entonces:

$$E(X) = np \sum_{y=0}^m \frac{m!}{(m-y)!y!} p^y (1-p)^{n-y}$$

De donde se sabe que  $p(y; m, p) = \frac{m!}{(m-y)!y!} p^y (1-p)^{n-y}$  es la función de probabilidad de una variable aleatoria binomial  $Y$  con parámetros  $m = n - 1$  y  $p$ ; de ésta manera  $\sum_{y=0}^m p(y; m, p) = 1$  y **la media de una variable aleatoria binomial es:**

$$E(X) = \mu = np.$$

Para obtener la varianza, se necesita el segundo momento alrededor de cero,  $\mu'_2$ , o:

$$E(X^2) = \sum_{x=0}^n x^2 p(x; n, p)$$

pero, el término  $x^2/x!$  se cancelará una sola  $x$  en el numerador, y la que resta evitará que la suma se manipule de la misma forma en que se determinó la media. La alternativa es escribir  $x^2$  como:

$$x^2 = x(x-1) + x;$$

de esta manera se tiene:

$$E(X^2) = E[X(X-1)] + E(X).$$

Dado que  $E(X)$  ya se ha determinado, puede usarse el mismo procedimiento para evaluar  $E(X(X-1))$ :

$$\begin{aligned} E[X(X-1)] &= \sum_{x=0}^n x(x-1) \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \\ &= \sum_{x=2}^n x(x-1) \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \\ &= \sum_{x=2}^n \frac{n!}{(n-x)(x-2)!} p^x (1-p)^{n-x} \\ &= n(n-1)p^2 \sum_{x=2}^n \frac{(n-2)!}{(n-x)!(x-2)!} p^{x-2} (1-p)^{n-x} \end{aligned}$$

Sea  $y = x - 2$  y  $m = n - 2$ , entonces:

$$\begin{aligned} E[X(X-1)] &= n(n-1)p^2 \sum_{y=0}^m \frac{m!}{(m-y)!y!} p^y (1-p)^{m-y} \\ &= n(n-1)p^2 \sum_{y=0}^m p(y; m, p) \\ &= n(n-1)p^2 \end{aligned}$$

Así,

$$E(X^2) = \mu'_2 = n(n-1)p^2 + np$$

De esta manera, la **varianza de una variable aleatoria binomial** es:

$$Var(X) = \mu'_2 - \mu^2 = n(n-1)p^2 + np - n^2p^2 = np[(n-1)p + 1 - np] = np(1-p).$$

Para obtener el tercer momento alrededor del cero, se determina  $E[X(X-1)(X-2)]$  dado que:

$$E[X(X-1)(X-2)] = \mu'_3 - 3\mu'_2 + 2\mu$$



$$\begin{aligned}
E[X(X-1)(X-2)] &= \sum_{x=0}^n x(x-1)(x-2) \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \\
&= \sum_{x=3}^n \frac{n!}{(n-x)!(x-3)!} p^x (1-p)^{n-x} \\
&= n(n-1)(n-2)p^3 \sum_{x=3}^n \frac{(n-3)!}{(n-x)!(x-3)!} p^{x-3} (1-p)^{n-x} \\
&= n(n-1)(n-2)p^3 \sum_{x=3}^n \frac{m!}{(m-y)!y!} p^y (1-p)^{m-y} \\
&= n(n-1)(n-2)p^3
\end{aligned}$$

Así,

$$\begin{aligned}
\mu'_3 - 3\mu'_2 + 2\mu &= n(n-1)(n-2)p^3 \\
\mu'_3 &= n(n-1)(n-2)p^3 + 3[n(n-1)p^2 + np] - 2np \\
&= n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np
\end{aligned}$$

El tercer momento central  $\mu_3$  puede ser determinado por  $\mu_3 = \mu'_3 - 3\mu\mu'_2 + 2\mu^3$ .

$$\mu_3 = n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np - 3np[n(n-1)p^2 + np] + 2n^3p^3$$

Lo que se traduce como:

$$\mu_3 = np(1-p)(1-2p)$$

Por tanto de  $\alpha_3 = \frac{\mu_3}{(\mu_2)^{3/2}}$  el tercer momento estandarizado de la distribución binomial es:

$$\begin{aligned}
\alpha_3 &= \frac{np(1-p)(1-2p)}{[np(1-p)]^{3/2}} \\
&= \frac{np(1-p)(1-2p)}{np(1-p)[np(1-p)]^{1/2}} \\
&= \frac{1-2p}{[np(1-p)]^{1/2}}
\end{aligned}$$

De manera similar, para el cuarto momento alrededor del cero se evalúa  $E[X(X-1)(X-2)(X-3)]$  dado que:

$$E[X(X-1)(X-2)(X-3)] = \mu'_4 - 6\mu'_3 + 11\mu'_2 - 6\mu.$$

es decir,

$$\begin{aligned}
E[X(X-1)(X-2)(X-3)] &= \sum_{x=4}^n x(x-1)(x-2)(x-3) \cdot \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \\
&= n(n-1)(n-2)(n-3)p^4 \sum_{x=4}^n \frac{(n-4)!}{(n-x)!(x-4)!} p^{x-4} (1-p)^{n-x} \\
&= n(n-1)(n-2)(n-3)p^4 \sum_{y=0}^m \frac{m!}{(m-y)!y!} p^y (1-p)^{m-y} \\
&= n(n-1)(n-2)(n-3)p^4
\end{aligned}$$

Sustituir en  $E[X(X-1)(X-2)(X-3)] = \mu'_4 - 6\mu'_3 + 11\mu'_2 - 6\mu$ . y para resolver  $\mu'_4$  se tiene:

$$\mu'_4 = n(n-1)(n-2)(n-3)p^4 + 6[n(n-1)(n-2)p^2 + 3n(n-1)p^2 + np] - 11[n(n-1)p^2 + np] + 6np$$

Luego de acuerdo a  $\mu_4 = \mu'_4 - 4\mu\mu'_3 + 6\mu^2\mu'_2 - 3\mu^4$  el cuarto momento central es:

$$\mu_4 = np(1-p)\{3np(1-p) + [1 - 6p(1-p)]\}.$$

Y de de acuerdo con  $\alpha_4 = \mu_4/\mu_2^2$

$$\alpha_4 = \frac{np(1-p)\{3np(1-p) + [1 - 6p(1-p)]\}}{n^2p^2(1-p)^2} = 3 + \frac{[1 - 6p(1-p)]}{np(1-p)}$$

La varianza de una variable aleatoria binomial siempre es menor que el valor de su media.

De acuerdo con la definición 3.14 la función generadora de momentos para la distribución binomial es:

$$\begin{aligned} m_X(t) = E(e^{tX}) &= \sum_{x=0}^n e^{tx} \frac{n!}{(n-x)x!} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \frac{n!}{(n-x)x!} (e^t p) (1-p)^{n-x} \\ &= (1-p)^n + n(1-p)^{n-1} (e^t p) + \frac{n(n-1)}{2!} (1-p)^{n-2} (e^t p)^2 + \dots + (e^t p)^n \\ &= [(1-p) + e^t p]^n \end{aligned}$$

Al tomar las dos primeras derivadas de  $[(1-p) + e^t p]^n$  con respecto de  $t$ , se obtiene:

$$\frac{dm_X(t)}{dt} = ne^t p [(1-p) + e^t p]^{n-1}$$

y

$$\frac{d^2 m_X(t)}{dt^2} = n(n-1)(e^t p)^2 [(1-p) + e^t p]^{n-2} + ne^t p [(1-p) + e^t p]^{n-1}$$

Si  $t = 0$ , obtiene el primero y segundo momento al rededor del cero,

$$\left. \frac{dm_X(t)}{dt} \right|_{t=0} = np[(1-p) + p]^{n-1} = np$$

y

$$\left. \frac{d^2 m_X(t)}{dt^2} \right|_{t=0} = n(n-1)p^2[(1-p) + p]^{n-2} + np[(1-p) + p]^{n-1} = n(n-1)p^2 + np.$$

### 4.3 La distribución de Poisson

La distribución es el principal modelo de probabilidad empleado para analizar problemas de líneas de espera. Además, ofrece una aproximación excelente a la función de probabilidad binomial cuando  $p$  es pequeño y  $n$  es grande.

**Definición 4.2** Sea  $X$  una variable aleatoria que representa el número de eventos aleatorios independientes que ocurren a una rapidez constante sobre el tiempo o el espacio. Se dice entonces que la variable aleatoria  $X$  tiene una distribución de Poisson con función de probabilidad.

$$p(x; \lambda) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & x = 0, 1, 2, \dots; \quad \lambda > 0 \\ 0 & \text{para cualquier otro valor} \end{cases}$$

El parámetro de la distribución de Poisson es  $\lambda$ , el número promedio de ocurrencias del evento aleatorio por unidad de tiempo.

Verificamos que es una función de probabilidad. Puesto que  $p(x; \lambda) > 0$  para  $x = 0, 1, 2, \dots$ ,

$$\sum_{x=0}^{\infty} p(x; \lambda) = \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} \left( 1 + \lambda + \frac{\lambda^2}{2!} + \dots \right) = e^{-\lambda} e^{\lambda} = 1$$

Nótese que la variable aleatoria de Poisson tiene un valor entero, y que pueden usarse los valores de las probabilidades acumuladas para determinar las probabilidades individuales mediante el empleo de la relación:

$$p(x; \lambda) = F(x; \lambda) - F(x - 1; \lambda)$$

### Deducción de la función de probabilidad de Poisson.-

Sea  $p(x; t)$  la probabilidad de tener, de manera exacta,  $X$  ocurrencias en un intervalo  $t$ , y supóngase lo siguiente:

1. En este intervalo, los eventos ocurren de manera independiente.
2. La probabilidad de una sola ocurrencia, en un intervalo muy pequeño  $dt$  es  $vdt$ , en donde  $v$  es la frecuencia constante de ocurrencia y ( $v > 0$ ).
3. El intervalo  $dt$  es tan pequeño, que la probabilidad de tener más de una ocurrencia en  $dt$  es despreciable.

El evento que en el tiempo  $t + dt$  ha ocurrido exactamente  $x$  veces, puede llevarse a cabo de dos maneras diferentes y excluyentes:

1. Existen  $x$  ocurrencias por tiempo  $t$ , con probabilidad  $p(x; t)$  y ninguna en  $dt$ , con probabilidad  $(1 - vdt)$ . Dada la suposición de independencia, la probabilidad conjunta es  $p(x; t)(1 - vdt)$ .
2. Existen  $x - 1$  ocurrencias por tiempo  $t$ , con probabilidad  $p(x - 1; t)$  y una durante  $dt$ , con probabilidad  $vdt$ . Otra vez, dada la suposición de independencia, la probabilidad conjunta es:  $p(x - 1; t)vdt$ .

Esto es:

$$p(x; t + dt) = p(x; t)(1 - vdt) + p(x - 1; t)vdt.$$

Después de multiplicar, transportar  $p(x; t)$  al primer miembro, y dividir por  $dt$  se tiene:

$$\frac{p(x; t + dt) - p(x; t)}{dt} = v[p(x - 1; t) - p(x; t)].$$

Si se toma el límite conforme  $dt \rightarrow 0$ , por definición se tiene:

$$\frac{dp(x;t)}{dt} = v[p(x-1;t) - p(x;t)]$$

Si se toma el límite conforme  $dt \rightarrow 0$ , por definición se tiene:

$$\frac{dp(x;t)}{dt} = v[p(x-1;t) - p(x;t)]$$

que es una ecuación diferencial lineal con respecto a  $t$  y una ecuación de diferencias finitas de primer orden, con respecto a  $x$ . Si  $x = 0$ , la ecuación se convierte en

$$\frac{dp(0;t)}{dt} = v[p(-1;t) - p(0;t)] = -vp(0;t)$$

dado que  $p(-1;t)$  tiene que ser cero. La solución general de la ecuación diferencial lineal

$$\frac{dp(0;t)}{dt} = -vp(0;t)$$

se obtiene mediante separación de variables e integración en ambos miembros, lo que da como resultado:

$$\ln[p(0;t)] = \ln(c) - vt,$$

o

$$p(0;t) = ce^{-vt}$$

Dado que la probabilidad de tener cero ocurrencias en un intervalo  $t = 0$ , debe ser 1,  $c = 1$ , y

$$p(0;t) = e^{-vt}.$$

Si  $x = 1$ , la ecuación  $\frac{dp(x;t)}{dt} = v[p(x-1;t) - p(x;t)]$  se convierte en

$$\frac{dp(1;t)}{dt} = v[p(0;t) - p(1;t)]$$

o

$$\frac{dp(1;t)}{dt} + vp(1;t) = ve^{-vt}$$

Esta ecuación es una ecuación diferencial no homogénea con la condición inicial de que  $p(1;0) = 0$  dado que la probabilidad de tener exactamente una ocurrencia en  $t = 0$  debe ser cero. La solución es

$$p(1;t) = (vt)e^{-vt}$$

De manera similar, para  $x = 2$  y  $p(2;0) = 0$ ,  $\frac{dp(x;t)}{dt} = v[p(x-1;t) - p(x;t)]$  se reduce a

$$\frac{dp(2;t)}{dt} + vp(2;t) = v^2te^{-vt}$$

cuya solución es

$$p(2;t) = \frac{(vt)^2e^{-vt}}{2!}.$$

Al continuar este proceso puede deducirse que la probabilidad de tener exactamente  $x$  ocurrencias en  $t$  es

$$p(x;t) = \frac{(vt)^xe^{-vt}}{x!}, \quad x = 0, 1, 2, \dots$$

siempre que  $p(x;0) = 0$ . Si se sustituye  $\lambda = vt$  en esta última ecuación, el resultado es la función de probabilidad de Poisson.

La probabilidad de que una variable aleatoria de Poisson  $X$  sea menor o igual a un valor de  $x$  se determina por la función de distribución acumulativa.

$$P(X \leq x) = F(x; \lambda) = \sum_{i=0}^x \frac{e^{-\lambda} \lambda^i}{i!}$$

se podría usar también la relación

$$p(x; \lambda) = F(x; \lambda) - F(x-1; \lambda).$$

**Teorema 4.1** Sea  $X$  una variable aleatoria con distribución binomial y función de probabilidad:

$$p(x; n, p) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n.$$

Si para  $n = 1, 2, \dots$  la relación  $p = \lambda/n$  es cierta para alguna constante  $\lambda > 0$ , entonces

$$\lim_{n \rightarrow \infty, p \rightarrow 0} p(x; n, p) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

Demostración.- Al multiplicar numerador y denominador por  $n^x$  y sustituir  $n!/(n-x)! = n(n-1)(n-2) \dots (n-x+1)$ , la función de probabilidad binomial es:

$$\begin{aligned} p(x; n, p) &= \frac{n(n-1)(n-2) \dots (n-x+1)}{n^x x!} (np)^x (1-p)^{n-x} \\ &= \frac{n(n-1)(n-2) \dots (n-x+1)}{n^x x!} \frac{\lambda^x}{x!} (1-p)^{n-x} \\ &= 1 \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{x-1}{n}\right) \frac{\lambda^x}{x!} (1-p)^{n-x} \\ &= \frac{\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{x-1}{n}\right)}{(1-p)^x} \frac{\lambda^x}{x!} (1-p)^n \end{aligned}$$

Dado que:

$$(1-p)^n = [(1-p)^{-1/p}]^{-np} = [(1-p)^{-1/p}]^{-\lambda}$$

y por definición

$$\lim_{z \rightarrow 0} (1+z)^{1/z} = e,$$

mediante el cambio de variable  $z = -p$ , se tiene

$$\lim_{p \rightarrow 0} [(1-p)^{-1/p}]^{-\lambda} = e^{-\lambda}$$

Además,

$$\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{x-1}{n}\right) = 1$$

Al sustituir se tiene,

$$\lim_{n \rightarrow \infty, p \rightarrow 0} p(x; n, p) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

Los momentos de la variable de Poisson se determinan mediante los mismo procedimientos utilizados para obtener los momentos de la variable aleatoria binomial. **Si  $X$  es una variable aleatoria de Poisson, su valor esperado es:**

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = \lambda.$$

Para la varianza  $X$ :

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda} \lambda^x}{x!} = \lambda^2 e^{-\lambda} \sum_{y=2}^{\infty} \frac{\lambda^{y-2}}{(y-2)!} = \lambda^2 e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = \lambda^2$$

Luego por el hecho de que  $E(X^2) = E[X(X-1)] + E(X)$  tenemos,

$$E(X^2) = \mu'_2 = \lambda^2 + \lambda,$$

**Y la varianza de  $X$  de una variable aleatoria de Poisson es:**

$$Var(X) = \mu'_2 - \mu^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

De esta manera, una característica distintiva de la variable aleatoria de Poisson es que su media es igual a su varianza.

Para el tercer momento central, se tiene:

$$E[X(X-1)(X-2)] = \lambda^3.$$

De donde

$$\mu'_3 = \lambda^3 + 3\lambda^2 + \lambda,$$

**Y el tercer momento central es:**

$$\mu_3 = \lambda$$

Como resultado, **el coeficiente de asimetría** se determina por:

$$\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{1}{\sqrt{\lambda}}.$$

Par el cuarto momento central puede emplearse el mismo procedimiento para demostrar que:

$$E[X(X-1)(X-2)(X-3)] = \lambda^4,$$

De donde

$$\mu'_4 = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda.$$

Así el **cuarto momento central** es:

$$\mu_4 = 3\lambda^2 + \lambda,$$

**Y el cuarto momento estandarizado para la distribución de Poisson** lo establece:

$$\lambda_4 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{\lambda}$$

La función generadora de momentos para la distribución de Poisson se determina por:

$$m_x(t) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{[\lambda(e^t-1)]}.$$

En conclusión, la distribución de Poisson es leptocúrtica con un sesgo positivo y se emplea para modelar el número de eventos aleatorios independientes que ocurren a una rapidez constantes ya sea sobre el tiempo o el espacio.

#### 4.4 La distribución hipergeométrica

**Definición 4.3** Sea  $N$  el número total de objetos en una población finita, de manera tal que  $k$  de estos es de un tipo  $N - k$  de otros. Si se selecciona una muestra aleatoria de la población constituida por  $n$  objetos de la probabilidad de que  $x$  sea de un tipo exactamente y  $n - x$  sea de otro, está dada por la función de probabilidad hipergeométrica:

$$p(x; N, n, k) = \begin{cases} \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}, & x = 0, 1, 2, \dots, n; \quad x \leq k, \quad n - x \leq N - k; \quad N, n, k, \text{ enteros positivos,} \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

La **distribución acumulada** estará dada por

$$P(X \leq x) = F(x; N, n, k) = \sum_{i=0}^x \frac{\binom{k}{i} \binom{N-k}{n-i}}{\binom{N}{n}}.$$

Si la proporción de artículos defectuosos en el lote es  $p = k/N$ , puede escribirse la función de probabilidad hipergeométrica como:

$$\lim_{N \rightarrow \infty} p_H(x; N, n, p) = p_B(x; n, p)$$

en donde  $p_B(x; n, p)$  es la función de probabilidad binomial.

Para determinar la **media de la distribución hipergeométrica** se sigue un procedimiento análogo al empleado para la distribución binomial. Si la función de probabilidad está dada por la definición de función de probabilidad hipergeométrica entonces,

$$E(X) = \sum_{x=0}^n x \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}} = \sum_{x=1}^n \frac{k!}{(k-x)!x!} \frac{\binom{N-k}{n-x}}{\binom{N}{n}} = k \sum_{x=1}^n \frac{\binom{k-1}{x-1} \binom{N-k}{n-x}}{\binom{N}{n}}$$

Pero puede demostrarse que:

$$\binom{N}{n} = \frac{N}{n} \binom{N-1}{n-1}, \quad \text{o} \quad \frac{N!}{(N-n)!n!} = \frac{N}{n} \left[ \frac{(N-1)!}{(N-n)!(n-1)!} \right].$$

Entonces:

$$E(X) = k \sum_{x=1}^n \frac{\binom{k-1}{x-1} \binom{N-k}{n-x}}{\frac{N}{n} \binom{N-1}{n-1}} = \frac{nk}{N} \sum_{x=1}^n \frac{\binom{k-1}{x-1} \binom{N-k}{n-x}}{\binom{N-1}{n-1}}$$

Si  $M = N - 1$ ,  $r = k - 1$ ,  $s = n - 1$  y  $y = x - 1$ ,

$$E(X) = \frac{nk}{N} \sum_{y=0}^s \frac{\binom{r}{y} \binom{M-r}{s-y}}{\binom{M}{s}} = \frac{nk}{N}.$$

Con el mismo procedimiento podemos demostrar que **la varianza** de una distribución hipergeométrica es:

$$Var(X) = \frac{nk(N-k)}{N^2} \cdot \frac{N-n}{N-1}.$$

#### 4.5 La distribución binomial negativa

Sea un escenario binomial en que se observa una secuencia de ensayos independientes; la probabilidad de éxito en cada ensayo es constante e igual a  $p$ . En lugar de fijar el número de ensayos en  $n$  y observar el número de éxitos, supóngase que se continúan los ensayos hasta que han ocurrido exactamente  $k$  éxitos. Esta situación lleva a lo que se conoce como la distribución binomial negativa.

**Definición 4.4** Sea  $X + k$ , el número de ensayos independientes necesarios para alcanzar, de manera exacta,  $k$  éxitos en un experimento binomial en donde la probabilidad de éxito en cada ensayo es  $p$ . Se dice entonces que  $X$  es una variable binomial negativa con función de probabilidad

$$p(x; k, p) = \begin{cases} \binom{k+x-1}{k-1} p^k (1-p)^x & x = 0, 1, 2, \dots, \quad k = 1, 2, \dots, \quad 0 \leq p \leq 1, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Debe notarse que si  $k = 1$  en la definición 4.4 entonces se conoce con el nombre de distribución geométrica:

$$p(x; p) = p(1-p)^x, \quad x = 0, 1, 2, 3, \dots, \quad 0 \leq p \leq 1$$

La variable aleatoria geométrica representa el número de fallas que ocurren antes de que se presente el primer éxito.

Puede demostrarse que si  $X$  es una variable aleatoria binomial negativa con función de probabilidad dada en 4.4, entonces:

$$P(X \leq x) = P(Y \geq k),$$

en donde  $Y$  es una variable aleatoria binomial con parámetros  $n = k + x$  y  $p$ . Esto es:



$$F_{NB}(x; k, p) = 1 - F_S(k - 1; k + x, p)$$

en donde  $F_{NB}(x; k, p)$  es la **distribución binomial negativa acumulada** y  $F_S(k - 1; k + x, p)$  es la distribución acumulada.

Si  $X$  es una variable aleatoria binomial con función de probabilidad entonces,

$$E(X) = \frac{k(1-p)}{p}$$

$$Var(X) = \frac{k(1-p)}{p^2}$$

$$\alpha_3 = \frac{2-p}{[k(1-p)]^{1/2}}$$

$$\alpha_4 = 3 + \frac{(p^2 - 6p + 6)}{k(1-p)}.$$

La función generadora de momentos se obtiene de la siguiente manera:

$$\begin{aligned} E(e^{tX}) &= \sum_{x=0}^{\infty} e^{tx} \binom{k+x-1}{k-1} p^k (1-p)^x \\ &= \sum_{x=0}^{\infty} \frac{(k+x-1)!}{(k-1)!x!} p^k [(1-p)e^t]^x \\ &= p^k + kp^k[(1-p)e^t] + \frac{k(k+1)}{2!} p^k [(1-p)e^t]^2 + \dots, \end{aligned}$$

pero esta es la expansión binomial de  $\left[ \frac{1}{p} - \frac{(1-p)e^t}{p} \right]^{-k}$ ; por lo tanto, **función generadora de momentos** está dada por:

$$m_x(t) = \frac{p^k}{[1 - (1-p)e^t]^k}$$

## Algunas distribuciones continuas de probabilidad

### 5.1 La distribución normal

**Definición 5.1** se dice que una variable aleatoria  $X$  se encuentra normalmente distribuida si su función de densidad de probabilidad está dada por

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$$

Si se obtienen las dos primeras derivadas de  $f(x; \mu, \sigma)$  con respecto a  $x$  y se igualan a cero, se tiene que el valor máximo de  $f(x; \mu, \sigma)$  ocurre cuando  $x = \mu$ , y los valores  $x = \mu \pm \sigma$  son las abscisas de los dos puntos de inflexión de la curva.

**Demostrar que la definición 5.1 es una función de densidad de probabilidad.**

**Demostración.-** El que la función sea no negativa se satisface, ya que  $f(x; \mu, \sigma) > 0$  para  $-\infty < x < \infty$ ,  $-\infty < \mu < \infty$  y  $\sigma > 0$ . Para demostrar que:

$$\int_{-\infty}^{\infty} f(x; \mu, \sigma) dx = 1.$$

Sea

$$I = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

el valor de la integral y aplíquese la transformación lineal  $y = (x - \mu)/\sigma$  de manera tal que  $x = \sigma y + \mu$  y  $dx = \sigma dy$ . Esto da como resultado:

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy.$$

Si puede demostrarse que  $I^2 = 1$ , puede deducirse que  $I = 1$  puesto que  $f(x; \mu, \sigma)$  tiene un valor positivo. De acuerdo con lo anterior:

$$I^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(y^2+z^2)}{2}} dy dz,$$

en donde se ha escrito el producto de las dos integrales como una doble integral ya que las funciones de  $z$  son contantes con respecto a  $y$  como también de manera viceversa. Al cambiar de coordenadas rectangulares representadas por  $x$  e  $y$ , a coordenadas polares  $r$  y  $\theta$ , en donde  $y = r \cos \theta$  y  $z = r \sin \theta$ . Esto es:

$$y^2 + z^2 = r^2 \cos^2 \theta + r^2 \sin^2 \theta = r^2$$

y el elemento de área  $dydz$ , en coordenadas rectangulares se reemplaza por  $rdrd\theta$  en coordenadas polares. Dado que los límites  $(-\infty, \infty)$  tanto para  $y$  como para  $z$  generan el plano completo  $yz$ , el plano correspondiente a  $r$  y a  $\theta$  se genera mediante el empleo de los límites  $(0, 2\pi)$  para  $\theta$  y  $(0, \infty)$  para  $r$ . De esta forma se tiene:

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty e^{-r^2/2} r dr d\theta = \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_0^\infty e^{-r^2/2} r dr = \frac{\theta}{2\pi} \Big|_0^{2\pi} \cdot [-e^{-r^2/2}]_0^\infty = 1.$$

**La media de una variable aleatoria distribuida normalmente** se encuentra definida por:

$$E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Se pretende demostrar que  $E(X) = \mu$ . Supóngase que a  $E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$  se suma y se resta

$$\frac{\mu}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

La identidad se mantiene, pero después de reacomodar términos se tiene

$$\begin{aligned} E(X) &= \int_{-\infty}^\infty (x - \mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \frac{\mu}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty (x - \mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu \end{aligned}$$

dado que el valor de la segunda integral es uno. Al afectar un cambio de variable de integración de manera tal que  $y = \frac{x - \mu}{\sigma}$ ,  $x = \sigma y + \mu$  y  $dx = \sigma dy$ , se tiene:

$$E(X) = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^\infty y e^{-y^2/2} dy + \mu = -\frac{\sigma}{\sqrt{2\pi}} e^{-y^2/2} \Big|_{-\infty}^\infty + \mu = \mu.$$

El lector recordará de sus cursos de cálculo que la última integral es cero porque el integrando es una función impar y la integración se lleva a cabo sobre un intervalo simétrico alrededor de cero.

Si el valor máximo de la función de densidad de probabilidad normal ocurre cuando  $x = \mu$  este es la media, la mediana y la moda de cualquier variable aleatoria distribuida aleatoriamente.

Para encontrar los demás momentos, se determinará la función generadora de momentos. Por definición:

$$m_{X-\mu}(t) = E[e^{t(X-\mu)}] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty e^{t(x-\mu)} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty e^{-\frac{1}{2\sigma^2}[(x-\mu)^2 - 2\sigma^2 t(x-\mu)]} dx.$$

de donde se completa el cuadrado en el interior del paréntesis rectangular y se tiene:

$$(x - \mu)^2 - 2\sigma^2 t(x - \mu) = (x - \mu)^2 - 2\sigma^2 t(x - \mu) + \sigma^4 t^2 - \sigma^4 t^2 = (x - \mu - \sigma^2 t)^2 - \sigma^4 t^2.$$

Por lo que,

$$m_{X-\mu}(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{\frac{\sigma^2 t^2}{2}} e^{-\frac{x-(\mu+\sigma^2 t)^2}{2\sigma^2}} dx = e^{\frac{\sigma^2 t^2}{2}},$$

dado que el integrando junto con el factor  $\frac{1}{\sqrt{2\pi}\sigma}$  es una función de densidad de probabilidad normal con parámetros  $\mu + \sigma^2 t$  y  $\sigma$ .

Al desarrollar  $e^{\frac{\sigma^2 t^2}{2}}$  en serie de potencias se tiene:

$$m_{X-\mu}(t) = 1 + \frac{(\sigma t)^2}{2} + \frac{(\sigma t)^4}{4 \cdot 2!} + \frac{(\sigma t)^6}{8 \cdot 3!} + \frac{(\sigma t)^8}{16 \cdot 4!} + \dots$$

Cuando las potencias impares de  $t$  no se encuentran presentes, todos los momentos centrales de  $X$  de orden impar son cero, de esta forma se asegura la simetría de la curva.

La segunda derivada de  $m_{X-\mu}(t)$  evaluada en  $t = 0$  es **la varianza** y está dada por:

$$Var(X) = \left. \frac{d^2 m_{X-\mu}(t)}{dt^2} \right|_{t=0} = \sigma^2 + \frac{12t^2\sigma^4}{4 \cdot 2!} + \frac{30t^4\sigma^6}{8 \cdot 3!} + \dots \Big|_{t=0} = \sigma^2;$$

De esta manera **la desviación estándar es  $\sigma$** . De manera similar, la cuarta derivada de  $m_{X-\mu}(t)$  evaluada en  $t = 0$  es el cuarto momento central, el cual es:

$$\mu_4 = \left. \frac{d^4 m_{X-\mu}(t)}{dt^4} \right|_{t=0} = 3\sigma^4 + \frac{360t^2\sigma^6}{8 \cdot 3!} + \dots \Big|_{t=0} = 3\sigma^4$$

De acuerdo con lo anterior, para cualquier distribución normal el coeficiente de asimetría es  $\alpha_3(X) = 0$ , mientras que la curtosis relativa es  $\alpha_4(X) = \frac{3\sigma^4}{\sigma^4} = 3$ . Para momentos alrededor del cero, puede determinarse la función generadora de momentos centrales o viceversa. Dado que

$$m_{X-\mu}(t) = E \left[ e^{t(X-\mu)} \right] = e^{-\mu t} E \left[ e^{tX} \right] = e^{-\mu t} m_X(t),$$

para una distribución normal

$$e^{-\mu t} m_X(t) = e^{\frac{\sigma^2 t^2}{2}}$$

y

$$m_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}.$$

La probabilidad de que una variable aleatoria normalmente distribuida  $X$  sea menor o igual a un valor específico,  $x$  está dada por **función de distribución acumulativa**

$$P(X \leq x) = F(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$

Sea  $Z$  una variable aleatoria definida por la siguiente relación:

$$Z = \frac{(X - \mu)}{\sigma}$$

en donde  $\mu$  y  $\sigma$  son la media y la desviación estándar de  $X$ , respectivamente. De acuerdo con lo anterior,  $Z$  es una variable aleatoria estandarizada con media cero y desviación estándar uno. Así,

$$P(X \leq x) = P[X \leq (x - \mu)/\sigma] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{(x-\mu)/\sigma} e^{-\frac{z^2}{2}} \sigma dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-\frac{z^2}{2}} dz.$$

El integrando junto con el factor  $1/\sqrt{2\pi}$  es la **función de densidad de probabilidad de la variable aleatoria normal estandarizada Z**. De donde

$$F_X(x; \mu, \sigma) = F_Z(z; 0, 1)$$

Para cualquier valor específico de  $z$ , el correspondiente valor en la tabla es la probabilidad de que la variable aleatoria normal estándar  $Z$  sea menor o igual a  $z$ ; esto es

$$P(Z \leq z) = F_Z(z; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt.$$

La notación  $X \sim N(\mu, \sigma)$  denotará que la variable  $X$  se encuentra distribuida normalmente con media  $\mu$  y desviación estándar  $\sigma$ .

Determinaremos la probabilidad de que un valor de  $X$  se encuentre entre  $a$  y  $b$ , si  $X \sim N(\mu, \sigma)$ . Por definición:

$$P(a \leq X \leq b) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx,$$

pero, mediante el empleo de  $E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x (x - \mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu$  se tiene:

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{z^2}{2}} dz = F_Z\left(\frac{b-\mu}{\sigma}; 0, 1\right) - F_Z\left(\frac{a-\mu}{\sigma}; 0, 1\right)$$

**Teorema 5.1** Sea  $X$  una variable aleatoria binomial con media  $np$  y desviación estándar  $\sqrt{np(1-p)}$ . La distribución de la variable aleatoria tiende a la normal

$$Y = \frac{X - np}{\sqrt{np(1-p)}}$$

estándar conforme el número de ensayos independientes  $n \rightarrow \infty$ .

**Demostración.-** La demostración que aquí se presenta se basa en el hecho de que una función generadora de momentos define, de manera única, a una distribución. Se demostrará que la función generadora de momentos de  $Y$  tiende a una distribución normal conforme  $n \rightarrow \infty$ .  $X$  es una variable aleatoria binomial:

$$m_X(t) = [(1-p) + pe^t]^n$$

Entonces:

$$m_Y(t) = E(e^{tY}) = E\left[e^{\frac{t(X-np)}{\sqrt{np(1-p)}}}\right] = e^{\frac{np t}{\sqrt{np(1-p)}}} \cdot E\left[e^{\frac{tX}{\sqrt{np(1-p)}}}\right]$$

donde  $E\left[e^{\frac{tX}{\sqrt{np(1-p)}}}\right]$  es la función generadora de momentos de  $X$  con argumento  $\frac{t}{\sqrt{np(1-p)}}$ . De esta forma se tiene:

$$m_Y(t) = e^{\frac{-np t}{\sqrt{np(1-p)}}} \left[ (1-p) + pe^{\frac{t}{\sqrt{np(1-p)}}} \right]^n ;$$

pero:

$$e^{-\frac{np t}{\sqrt{np(1-p)}}} = \left( e^{-\frac{p t}{\sqrt{np(1-p)}}} \right)^n$$

y:

$$m_Y(t) = \left[ (1-p)e^{-\frac{p t}{\sqrt{np(1-p)}}} + p e^{\frac{t}{\sqrt{np(1-p)}} - \frac{p t}{\sqrt{np(1-p)}}} \right]^n = \left[ (1-p)e^{-\frac{p t}{\sqrt{np(1-p)}}} + p e^{\frac{(1-p)t}{\sqrt{np(1-p)}}} \right]^n.$$

En la última expresión, al expandir ambas funciones exponenciales en una serie de potencias, se tiene:

$$\begin{aligned} (1-p)e^{-\frac{p t}{\sqrt{np(1-p)}}} &= (1-p) - \frac{(1-p)p t}{\sqrt{np(1-p)}} + \frac{(1-p)p^2 t^2}{2np(1-p)} + \text{términos en } (-1)^k \left(\frac{1}{n}\right)^{k/2}, k = 3, 4, \dots \\ &= (1-p) - \frac{(1-p)p t}{\sqrt{np(1-p)}} + \frac{p t^2}{2n} + \text{términos en } (-1)^k \left(\frac{1}{n}\right)^{k/2}, k = 3, 4, \dots \end{aligned}$$

y

$$\begin{aligned} p e^{-\frac{(1-p)t}{\sqrt{np(1-p)}}} &= p + \frac{(1-p)p t}{\sqrt{np(1-p)}} + \frac{(1-p)p t^2}{2np(1-p)} + \text{términos en } \left(\frac{1}{n}\right)^{k/2}, k = 3, 4, \dots \\ &= p + \frac{(1-p)p t}{\sqrt{np(1-p)}} + \frac{(1-p)t^2}{2n} + \text{términos en } \left(\frac{1}{n}\right)^{k/2}, k = 3, 4, \dots \end{aligned}$$

Al sustituir los resultados anteriores en  $m_Y(t)$  y agrupar términos,

$$m_Y(t) = \left[ 1 + \frac{t^2}{2n} + \text{términos en } \left(\frac{1}{n}\right)^{k/2} \right]^n, k = 3, 4, \dots$$

Dado que todos los términos que contiene a  $(1/n)^{k/2}$ ,  $k = 3, 4, \dots$ , tienen exponentes mayores que uno, puede factorizarse el término  $1/n$ . De esta forma se tiene que:

$$m_Y(t) = \left[ 1 + \frac{1}{n} \left( \frac{t^2}{2} + \text{términos en } \left(\frac{1}{n}\right)^{(k-2)/2} \right) \right]^n, k = 3, 4, \dots$$

Por definición:

$$\lim_{n \rightarrow \infty} \left( 1 + \frac{u}{n} \right)^n = e^u$$

entonces, conforme  $n \rightarrow \infty$ , la última expresión para  $m_Y(t)$  es idéntica a esta forma, con  $u$  representando a todo lo que se encuentra entre paréntesis de esta expresión. Pero conforme  $n \rightarrow \infty$ , todos los términos de  $u$ , excepto el primero, tienen un valor de cero, dado que todos tienen potencias positivas de  $n$  en sus denominadores. De acuerdo con lo anterior.

$$\lim_{n \rightarrow \infty} m_Y(t) = e^{t^2/2},$$

que es la función generadora de momentos de la distribución normal estándar.

■

La aproximación del teorema anterior es adecuada tanto como  $np > 5$  cuando  $p \leq 1/2$ , o cuando  $n(1-p) > 5$  para  $p > 1/2$ . Esto es

$$P(a \leq X_B \leq b) = P\left(\frac{a - np}{\sqrt{np(1-p)}} \leq Z_N \leq \frac{b - np}{\sqrt{np(1-p)}}\right)$$

en donde  $Z_N$  es  $N(0, 1)$ . Como también

$$P(X_B = x) \approx P\left(\frac{x - np - 1/2}{\sqrt{np(1-p)}} \leq Z_N \leq \frac{x - np + 1/2}{\sqrt{np(1-p)}}\right)$$

por lo que se puede modificar la expresión de desigualdad de la siguiente manera:

$$P(a \leq X_B \leq b) = P\left(\frac{a - np - 0.5}{\sqrt{np(1-p)}} \leq Z_N \leq \frac{b - np + 0.5}{\sqrt{np(1-p)}}\right)$$

## 5.2 La distribución uniforme

**Definición 5.2** Se dice que una variable aleatoria  $X$  está distribuida uniformemente sobre el intervalo  $(a, b)$  si su función de densidad de probabilidad está dada por:

$$f(x; a, b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

La **función de distribución acumulativa** se determina de manera fácil y está dada por

$$P(X \leq x) = F_{x;a,b} = \frac{1}{b-a} \int_a^x dt = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & a \leq x \leq b, \\ 1 & x > b. \end{cases}$$

Se sigue entonces que para cualquier subintervalo  $(a_1, b_1)$  interior a  $(a, b)$ :

$$P(a_1 \leq X \leq b_1) = F(b_1; a, b) - F(a_1; a, b) = \frac{b_1 - a_1}{b - a}.$$

Este resultado ilustra que la probabilidad de que  $X$  tome valores del subintervalo  $(a_1, b_1)$ , es  $\frac{1}{b-a}$  por la longitud del subintervalo y de esta forma, igual a la probabilidad de que  $X$  tome un valor en cualquier otro subintervalo de la misma longitud.

El **valor esperado** de una variable aleatoria distribuida de manera uniforme es

$$E(X) = \frac{1}{b-a} \int_a^b x \, dx = \frac{a+b}{2}.$$

Para obtener los momentos superiores de  $X$ , es más fácil trabajar con la variable aleatoria  $Y = X - \frac{a+b}{2}$ , que desplaza la media a cero, dado que  $E(Y) = E(X) - \frac{a+b}{2}$ . De esta forma:

$$f(y; \theta) = \frac{1}{\theta}, \quad -\frac{\theta}{2} \leq y \leq \frac{\theta}{2},$$

en donde  $\theta = b - a$ . De acuerdo con lo anterior, el  $r$ -ésimo momento central de  $Y$  es igual al  $r$ -ésimo momento central alrededor del cero, esto es:

$$\mu_r(Y) = \mu'_r(Y) = \theta^{-1} \int_{-\theta/2}^{\theta/2} y^r dy = \left( \frac{1}{\theta} \right) \frac{y^{r+1}}{r+1} \Big|_{-\theta/2}^{\theta/2} = \begin{cases} 0 & \text{si } r \text{ es impar} \\ \theta^r / [(r+1)2^r] & \text{si } r \text{ es par} \end{cases}$$

Dado que ni la varianza ni los factores de forma se ven afectados por el cambio de localización, la varianza, el coeficiente de asimetría y la curtosis relativa de la variable aleatoria distribuida uniformemente se encuentran a partir de  $\begin{cases} 0 & \text{si } r \text{ es impar} \\ \theta^r / [(r+1)2^r] & \text{si } r \text{ es par} \end{cases}$  y están determinadas por:

$$Var(X) = \frac{(b-a)^2}{12}$$

$$\alpha_3(X) = 0$$

$$\alpha_4(X) = \frac{\frac{(b-a)^4}{80}}{\left[ \frac{(b-a)^2}{12} \right]^2} = \frac{9}{5}.$$

Puede emplearse  $f(y; \theta) = \frac{1}{\theta}$ ,  $-\frac{\theta}{2} \leq y \leq \frac{\theta}{2}$ , para determinar la **desviación media** de la siguiente manera:

$$E|Y| = \theta^{-1} \int_{-\theta/2}^{\theta/2} y dy = 2\theta^{-1} \int_0^{\theta/2} y dy = \frac{\theta}{4} = \frac{b-a}{4}.$$

No tiene moda y su **mediana** es igual a la media. Los valores cuantiles  $x_q$ , correspondientes a la proporción acumulativa  $q$ , son de manera tal que:

$$F(x_q, a, b) = q$$

Los que por  $\begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & a \leq x \leq b, \\ 1 & x > b. \end{cases}$  son:

$$x_q = a + (b-a)q.$$



### 5.3 La distribución beta

**Definición 5.3** Se dice que una variable aleatoria  $X$  posee una distribución beta si su función de densidad está dada por:

$$f(x; \alpha, \beta) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & 0 < x < 1, \quad \alpha, \beta > 0, \\ 0 & \text{para cualquier otro valor} \end{cases}$$

Recordemos que  $\Gamma(n) = \int_0^\infty u^{n-1} e^{-u} du$ ,  $n > 0$ .

Si tanto  $\alpha$  como  $\beta$  son menores que uno, la distribución beta tiene un perfil en forma de  $U$ . Si  $\alpha < 1$  y  $\beta \geq 1$ , la distribución tiene un perfil de  $J$  transpuesta, y si  $\beta < 1$  y  $\alpha \geq 1$ , el perfil es una  $J$ . Finalmente cuando  $\alpha = \beta$  la distribución es simétrica. Nótese que si en la definición se reemplaza por  $x - 1$ , se obtiene la siguiente relación de simetría:

$$f(1-x; \beta, \alpha) = f(x; \alpha, \beta)$$

El nombre de esta distribución proviene de su asociación con la función beta que se encuentra definida por

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$$

Puede demostrarse que las funciones beta y gama se encuentran relacionadas por la expresión

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

Mediante el empleo de estas últimas fórmulas, es obvio que la definición es una función de densidad de probabilidad. Esto es:

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} B(\alpha, \beta) = 1,$$

y puesto que  $f(x; \alpha, \beta)$  es no negativa, entonces la definición es una función de densidad de probabilidad.

La **función de distribución acumulativa** se encuentra definida por:

**Definición 5.4**

$$F(x; \alpha, \beta) = \begin{cases} 0 & x \leq 0, \\ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x t^{\alpha-1} (1-t)^{\beta-1} dt & 0 < x < 1, \\ 1 & x \geq 1. \end{cases}$$

La integral que aparece es la función beta incompleta:

$$B_x(\alpha, \beta) = \int_0^x t^{\alpha-1} (1-t)^{\beta-1} dt.$$

De esta forma, la función de distribución beta puede expresarse como un cociente de funciones beta incompletas,

$$F(x; \alpha, \beta) = \frac{B_x(\alpha, \beta)}{B(\alpha, \beta)} = \frac{\int_0^x t^{\alpha-1} (1-t)^{\beta-1} dt}{\int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx} = I_x(\alpha, \beta), \quad 0 < x < 1,$$

Con el fin de encontrar los valores cuantiles correspondientes a puntos de alto porcentaje, considérese lo siguiente:

$$P(X \leq x) = P(1 - X \geq 1 - x) = 1 - P(1 - X \leq 1 - x);$$

entonces, por la relación de simetría  $f(1 - x; \beta, \alpha) = f(x; \alpha, \beta)$ :

$$F(x; \alpha, \beta) = 1 - F(1 - x; \beta, \alpha)$$

o

$$I_x(\alpha, \beta) = 1 - I_{1-x}(\beta, \alpha).$$

Obtenemos los momentos de la variable aleatoria beta como sigue.

$$\begin{aligned} \mu'_r &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+r-1} (1-x)^{\beta-1} dx \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} B(\alpha + r, \beta) \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha + r)\Gamma(\beta)}{\Gamma(\alpha + \beta + r)} \\ &= \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + r)}{\Gamma(\alpha)\Gamma(\alpha + \beta + r)} \end{aligned}$$

Como resultado,

$$E(X) = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + 1)}{\Gamma(\alpha)\Gamma(\alpha + \beta + 1)} = \frac{\alpha}{\alpha + \beta}.$$

Y

$$Var(x) = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} - \frac{\alpha^2}{(\alpha + \beta)^2} = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Al seguir este procedimiento y después de efectuar el álgebra necesaria, el **coeficiente de asimetría** y la **curtosis relativa** para la distribución beta están dadas por:

$$\alpha_3(X) = \frac{2(\beta + \alpha)\sqrt{\alpha + \beta + 1}}{\sqrt{\alpha\beta}(\alpha + \beta + 2)}$$

Y

$$\alpha_4(X) = \frac{3(\alpha + \beta + 1) [2(\alpha + \beta)^2 \alpha \beta (\alpha + \beta - 6)]}{\alpha \beta (\alpha + \beta + 2)(\alpha + \beta + 3)}$$

Puede demostrarse que si la suma de los parámetros que determinan el perfil de la distribución beta es, de manera relativa, grande, la función de distribución acumulativa beta se puede aproximar de manera adecuada por la diferencia de dos funciones de distribución normal estándar. Esto es:

$$F(x; \alpha, \beta) \approx F_N(z_u; 0, 1) - F_N(z_t; 0, 1),$$

En donde:

$$z_u = \frac{[\beta] - 0.5 - (\alpha + \beta - 1)(1 - x)}{[x(\alpha + \beta - 1)(1 - x)]^{1/2}},$$

$$z_t = -\frac{(\alpha + \beta - 1)(1 - x) + 0.5}{[x(\alpha + \beta - 1)(1 - x)]^{1/2}},$$

donde  $[\beta]$  denota el entero más grande que no excede a  $\beta$ .

#### 5.4 La distribución gama

Supóngase que una pieza metálica se encuentra sometida a cierta fuerza, de manera que se romperá después de aplicar un número específico de ciclos de fuerza. Si los ciclos ocurren de manera independiente y a una frecuencia promedio, entonces el tiempo que debe transcurrir antes de que el material se rompa es una variable aleatoria que cumple con la distribución gama.

**Definición 5.5** Se dice que la variable aleatoria  $X$  tiene una distribución gama si su función de densidad de probabilidad está dada por:

$$f(x; \alpha, \theta) = \begin{cases} \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta} & x > 0, \alpha, \theta > 0 \\ 0 & \text{para cualquier otro resultado.} \end{cases}$$

en donde  $\Gamma(\alpha)$  es la función gama.

Demostraremos que la función dada es función de densidad de probabilidad. Para hacerlo, considérese un cambio de variable de integración, tal que  $u = x/\theta$ ,  $x = \theta u$  y  $dx = \theta du$ ; entonces:

$$\frac{1}{\Gamma(\alpha)\theta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x/\theta} dx = \frac{1}{\Gamma(\alpha)\theta^\alpha} \int_0^\infty (\theta u)^{\alpha-1} e^{-u} \theta du = \frac{1}{\Gamma(\alpha)} \int_0^\infty u^{\alpha-1} e^{-u} du = 1.$$

Dado que  $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$ .

Con este procedimiento similar se demuestra que el  $r$ -ésimo momento alrededor del cero es:

$$\mu'_r = \frac{1}{\Gamma(\alpha)\theta^\alpha} \int_0^\infty x^{\alpha+r-1} e^{-x/\theta} dx = \frac{\theta^{\alpha+r}}{\Gamma(\alpha)\theta^\alpha} \int_0^\infty u^{\alpha+r-1} e^{-u} du = \frac{\theta^r \Gamma(\alpha + r)}{\Gamma(\alpha)}$$

Se sigue, por tanto que:

$$E(X) = \alpha\theta.$$

$$Var(X) = \alpha\theta^2.$$

Ademas el **coeficiente de asimetría** es:

$$\alpha_3(X) = 2\sqrt{\alpha}.$$

Y la **curtosis relativa** está dada por:

$$\alpha_4(X) = 3 \left( 1 + \frac{2}{\alpha} \right).$$

Para valores grandes de  $\alpha$  la distribución gama puede aproximarse, en algún grado, por una distribución normal. Esto es, la variable aleatoria

$$Z = \frac{X - \alpha\theta}{\theta\sqrt{\alpha}}$$

es, de manera aproximada, igual a la normal estándar para valores grandes de  $\alpha$ .

La **función generadora de momentos** para la variable aleatoria gama  $X$  está dada por:

$$E[e^{tX}] = \frac{1}{\Gamma(\alpha)\theta^\alpha} \int_0^\infty x^{\alpha-1} e^{-\frac{1-\theta t}{\theta}x} dx.$$

Sea  $u = \frac{(1-\theta t)x}{\theta}$ ,  $x = \frac{u\theta}{1-\theta t}$  y  $dx = \left[ \frac{\theta}{1-\theta t} \right] du$ , entonces:

$$\begin{aligned} E[e^{tX}] &= \frac{1}{\Gamma(\alpha)\theta^\alpha} \int_0^\infty \frac{u^{\alpha-1}\theta^{\alpha-1}}{(1-\theta t)^{\alpha-1}} e^{-u} \frac{\theta}{1-\theta t} du \\ &= \frac{1}{\Gamma(\alpha)(1-\theta t)^\alpha} \int_0^\infty u^{\alpha-1} e^{-u} du \\ &= (1-\theta t)^{-\alpha}, \quad 0 \leq t < \frac{1}{\theta}. \end{aligned}$$

La **función de distribución acumulativa** se determina por la siguiente expresión:

$$F(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} \int_0^x t^{\alpha-1} e^{-\frac{t}{\theta}} dt, \quad x > 0.$$

Si se efectua el cambio de variable  $u = \frac{t}{\theta}$  de manera tal que  $t = \theta u$  y  $dt = \theta du$ , entonces la función de distribución acumulativa toma la siguiente forma:

$$F(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} \int_0^{x/\theta} (\theta u)^{\alpha-1} e^{-u} du = \frac{1}{\Gamma(\alpha)} \int_0^{x/\theta} u^{\alpha-1} e^{-u} du.$$

La integral  $\int_0^{x/\theta} u^{\alpha-1} e^{-u} du$  se conoce como la función gama incompleta y se denota generalmente por  $\gamma(x/\theta; \alpha)$ . De acuerdo con lo anterior la función gama de distribución acumulativa se escribe como:

$$P(X \leq x) = F(x; \alpha, \theta) = \frac{\gamma(x/\theta; \alpha)}{\Gamma(\alpha)}$$

Cuando el parámetro de forma  $\alpha$  es igual a 1, la distribución de Erlang (gama) se reduce a lo que se conoce como la distribución exponencial negativa.

También se tiene la equivalencia de

$$I(u, p) = F(x; \alpha, \theta)$$

donde  $u = \frac{x}{\theta}$  y  $p = \alpha - 1$ . Debe notarse que si el parámetro de forma  $\alpha$  es un entero positivo,  $I(u, p) = F(x; \alpha, \theta)$  se puede expresar, en forma cerrada:

$$F(x; \alpha, \theta) = 1 - \left[ 1 + \frac{x}{\theta} + \frac{1}{2!} \left( \frac{x}{\theta} \right)^2 + \dots + \frac{1}{(\alpha - 1)!} \left( \frac{x}{\theta} \right)^{\alpha-1} \right] e^{-x/\theta}$$

Cuando  $\alpha$  es entero positivo, la distribución gama también se conoce como **distribución de Erlang**. Existe una asociación entre los modelos de probabilidad de Poisson y Erlang. Esto es, la probabilidad de que ocurra a lo más  $\alpha - 1$  eventos de Poisson en un tiempo  $x$  a una frecuencia constante  $1/\theta$  se desprende de

$P(X \leq x) = F(x; \lambda) = \sum_{i=0}^x \frac{e^{-\lambda} \lambda^i}{i!}$  y está dada por:

$$F_P(\alpha - 1; x/\theta) = \left[ 1 + \frac{x}{\theta} + \frac{1}{2!} \left( \frac{x}{\theta} \right)^2 + \dots + \frac{1}{(\alpha - 1)!} \left( \frac{x}{\theta} \right)^{\alpha-1} \right] e^{-x/\theta}.$$

Por otro lado, si se supone que el tiempo de espera sigue el modelo de Erlang, la probabilidad de que el tiempo de espera hasta que ocurra el  $\alpha$ -ésimo evento exceda un lapso  $x$  específico, está determinado por:

$$\begin{aligned} P(X > x) &= 1 - F_P(x; \alpha, \theta) \\ &= 1 - \left\{ 1 - \left[ 1 + \frac{x}{\theta} + \frac{1}{2!} \left( \frac{x}{\theta} \right)^2 + \dots + \frac{1}{(\alpha - 1)!} \left( \frac{x}{\theta} \right)^{\alpha-1} \right] e^{-x/\theta} \right\} \\ &= \left[ 1 + \frac{x}{\theta} + \frac{1}{2!} \left( \frac{x}{\theta} \right)^2 + \dots + \frac{1}{(\alpha - 1)!} \left( \frac{x}{\theta} \right)^{\alpha-1} \right] e^{-x/\theta} \\ &= F_P(\alpha - 1; x/\theta) \end{aligned}$$

Cuando el parámetro de forma  $\alpha$  es igual a uno, la distribución de Erlang (gama) se reduce a lo que se conoce como la **distribución exponencial negativa**. Esta distribución se emplea de manera extensa para representar lapsos aleatorios de tiempo. Nótese que la variable aleatoria de una distribución exponencial negativa puede pensarse como el lapso que transcurre hasta el primer evento de Poisson. De acuerdo con lo anterior, la variable de Erlang es la suma de variables aleatorias independientes distribuidas exponencialmente.

Otro caso especial del modelo de probabilidad gama es la distribución chi-cuadrado. Si se reemplaza en

$$f(x; \alpha, \theta) = \begin{cases} \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta} & x > 0, \alpha, \theta > 0 \\ 0 & \text{para cualquier otro resultado.} \end{cases}$$

el parámetro de forma  $\alpha$  con  $v/2$  y el parámetro de escala  $\theta$  con 2, el resultado es la **función de densidad de probabilidad de una variable aleatoria chi-cuadrado** y se determina por:

$$f(x; \alpha, \theta) = \begin{cases} \frac{1}{\Gamma(v/2)2^{v/2}} x^{v/2-1} e^{-x/2} & x > 0, \alpha, \theta > 0 \\ 0 & \text{para cualquier otro resultado.} \end{cases}$$

La distribución chi-cuadrado se encuentra caracterizada por un solo parámetro  $v$ , que recibe el nombre de grados de libertad. Su notación es:

$$X \sim \chi_v^2$$

La **función de distribución acumulativa** está dada por:

$$P(X \leq x) = \frac{1}{\Gamma(v/2)2^{v/2}} \int_0^x t^{v/2-1} e^{-t/2} dt, \quad x > 0.$$

Los **cuantiles**  $x_{1-\alpha, v}$  es

$$P(X \leq x_{1-\alpha, v}) = \int_0^{x_{1-\alpha, v}} f(x; v) dx = 1 - \alpha.$$

Los momentos de distribución chi-cuadrado se obtienen a partir de  $E(X) = \alpha\theta$ ,  $Var(X) = \alpha\theta^2$ ,  $\alpha_3(X) = 2/\sqrt{\alpha}$  y  $\alpha_4(X) = 3(1 + 2/\alpha)$ , dado por:

$$\begin{aligned} E(X) &= v \\ Var(X) &= 2v \\ \alpha_3(X) &= \frac{4}{\sqrt{2v}} \\ \alpha_4(X) &= 3 \left( 1 + \frac{4}{v} \right) \end{aligned}$$

Análogamente y a partir de  $E(e^{tX}) = (1 - \theta t)^{-\alpha}$  la **función generadora de momentos** para la distribución chi-cuadrado es:

$$m_X(t) = (1 - 2t)^{-v/2}, \quad 0 \leq t < \frac{1}{2}.$$

Su varianza es dos veces el valor de su media. Presenta un sesgo positivo y un pico mayor que el de una distribución normal, pero el coeficiente de asimetría tiende a cero y a una curtosis relativa igual a tres conforme a  $v$  tiende al infinito.

## 5.5 La distribución de Weibull

El esfuerzo al que se someten los materiales pueden modelarse de manera adecuada mediante el empleo de esta distribución. En los últimos 25 años esta distribución se empleó como modelo para situaciones del tipo tiempo-falla y con el objetivo de lograr una amplia variedad de componentes mecánicos y eléctricos.

**Definición 5.6** Se dice que una variable aleatoria  $X$  tiene una distribución de Weibull si su función de densidad de probabilidad está dada por:

$$f(x; \alpha, \beta) = \begin{cases} \frac{\alpha}{\theta^\alpha} x^{\alpha-1} e^{-(x/\theta)^\alpha} & x > 0, \alpha, \beta > 0 \\ 0 & \text{para cualquier otro resultado.} \end{cases}$$

Se puede introducir un parámetro adicional al reemplazar la variable aleatoria de Weibull  $X$  por  $X - a$  en donde  $a$  es un parámetro de localización que representa un valor umbral o tiempo de garantía.

La **función de distribución acumulativa de Weibull** es:

$$F(x; \alpha, \theta) = \frac{\alpha}{\theta^\alpha} \int_0^x t^{\alpha-1} e^{-(t/\theta)^\alpha} dt$$

puede obtenerse en forma cerrada mediante la evaluación directa de la integral anterior. Esto es:

$$F(x; \alpha, \theta) = \frac{\alpha}{\theta^\alpha} \left( -\frac{\theta^\alpha}{\alpha} \right) e^{-(t/\theta)^\alpha} \Big|_0^x = 1 - e^{-(x/\theta)^\alpha}, \quad x \geq 0.$$

De esta fórmula mencionada, el **valor cuantil**  $x_q$  es:

$$1 - e^{(x_q/\theta)^\alpha} = q \quad \Rightarrow \quad x_q = -\theta [\ln(1 - q)]^{1/\alpha} = \theta \left[ \ln \left( \frac{1}{1 - q} \right) \right]^{1/\alpha}$$

En particular, la **mediana de una variable aleatoria de Weibull** es:

$$x_{0.5} \theta [\ln(2)]^{1/\alpha}$$

Los momentos y los factores de una variable aleatoria de Weibull se encuentran primero al determinar el  $r$ -ésimo momento central alrededor de cero:

$$u'_r = E(X^r) = \int_0^\infty x^r f(x; \alpha, \theta) dx = \frac{\alpha}{\theta^\alpha} \int_0^\infty x^{\alpha+r-1} e^{-(x/\theta)^\alpha} dx$$

Luego sea  $u = (x/\theta)^\alpha$ ; entonces  $x = \theta u^{1/\alpha}$  y  $dx = (\theta/\alpha) u^{1/\alpha-1} du$ . El resultado es:

$$u'_r = \frac{\alpha}{\theta^\alpha} \int_0^\infty (\theta u^{1/\alpha})^{\alpha+r-1} e^{-u} \frac{\theta}{\alpha} u^{1/\alpha-1} du = \theta^r \int_0^\infty u^{r/\alpha} e^{-u} du = \theta^r \Gamma \left( 1 + \frac{r}{\alpha} \right)$$

Por lo que la **media** es:

$$E(X) = \theta \Gamma \left( 1 + \frac{1}{\alpha} \right)$$

Y la **varianza** es:

$$Var(X) = \theta^2 \left[ \left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right]$$

La distribución Weibull puede aproximarse, de manera adecuada, por una distribución normal cada que el factor de forma  $\alpha$  se encuentre en 3.6 y 5.83.

Cuando el parámetro de forma es igual a uno, la distribución de Weibull (al igual que la gama), se reduce a la distribución exponencial negativa. Cuando  $\alpha = 2$  y el parámetro de escala  $\theta$  se reemplaza por  $\sqrt{2}\sigma$ , la función de densidad de Weibull se reduce a

$$f(x; \sigma^2) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}, \quad x > 0,$$

que es la función de densidad de probabilidad de lo que se conoce como **distribución de Rayleigh**.

## 5.6 La distribución exponencial negativa

Es un caso especial de la distribución de Gama (Erland). La variable exponencial es el tiempo que transcurre hasta que se da el primer evento de Poisson. Es decir, la distribución exponencial se puede modelar el lapso entre dos eventos consecutivos de Poisson que ocurren de manera independiente y a una frecuencia constante. Se emplea al modelar problemas del tipo tiempo-falla y como modelo para el intervalo en problemas de líneas de espera. Esta distribución no tiene memoria. Es decir, la probabilidad de ocurrencia de eventos presentes o futuros no depende de lo que hayan ocurrido en el pasado. De esta forma, la probabilidad de que una unidad falle en un lapso específico depende nada más de la duración de este, no del tiempo en que la unidad ha estado en operación.

**Definición 5.7** Si una variable aleatoria  $X$  tiene una distribución exponencial, su función de densidad de probabilidad está dada por:

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & x > 0, \theta > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

La distribución exponencial se caracteriza por un parámetro  $\theta$ , que representa el lapso promedio de tiempo entre dos eventos independientes de Poisson. En contexto de fiabilidad recibe el nombre de promedio entre fallas y  $1/\theta$  es la frecuencia de falla.

La **función acumulativa** se obtiene directamente de los modelos de Weibull o de Erland,

$$P(X \leq x) = F(x; \theta) = 1 - e^{-x/\theta}$$

Las expresiones para los valores cuantiles, momentos y factores de forma, se obtienen de la distribución Weibull con  $\alpha = 1$ . Esto es:



$$E(X) = \theta$$

$$Var(X) = \theta^2$$

$$\alpha_3(X) = 2$$

$$\alpha_4(X) = 9.$$

**Definición 5.8** Sea  $T$  una variable aleatoria que representa el tiempo de vida de un sistema y sea  $f(t)$  la función de densidad de probabilidad de  $T$ . La función de confiabilidad del sistema a tiempo  $t$ ,  $R(t)$ , es la probabilidad de que el lapso de duración del sistema sea mayor que un tiempo  $t$  dado. De acuerdo con lo anterior,

$$R(t) = P(T > t) = 1 - F(t), \quad t > 0.$$

Otra cantidad muy útil para seleccionar una función de densidad de probabilidad para el lapso de vida medio de una unidad (o sistema) es la frecuencia de falla o función de riesgo, que se define de la siguiente forma:

**Definición 5.9** Sean  $f(t)$  y  $R(t)$  las funciones de densidad de probabilidad y de confiabilidad, respectivamente, de una unidad en un tiempo dado  $t$ . La frecuencia de falla  $h(t)$  se define como la proporción de unidades que fallan en el intervalo  $(t, t + dt)$  con respecto a las que siguen funcionando a tiempo  $t$ , y está determinada por:

$$h(t) = \frac{f(t)}{R(t)}.$$

Si se conoce la frecuencia de falla, es posible determinar la función de densidad de probabilidad de la variable aleatoria. Dado que  $R(t) = 1 - F(t)$ , mediante diferenciación con respecto a  $t$ , se tiene que  $R'(t) = -F'(t)$ , pero  $F'(t) = f(t)$ . Como resultado se tiene que la frecuencia de falla puede expresarse como:

$$h(t) = -\frac{R'(t)}{R(t)}.$$

Integrando ambos miembros desde 0 hasta  $t$ , se tiene:

$$\int_0^t h(x) dx = -\int_0^t [R'(x)R(x)] dx = -\ln [R(t)] + \ln [R(0)] = -\ln [R(t)],$$

donde  $x$  es una variable muda de integración. Dado que:

$$-\ln [R(t)] = \int_0^t h(x) dx,$$

se tiene

$$R(t) = e^{\left[-\int_0^t h(x) dx\right]}.$$

Mediante el empleo de  $h(t) = \frac{-R'(t)}{R(t)}$ , la función de densidad de probabilidad es:

$$f(t) = h(t)e^{\left[-\int_0^t h(x) dx\right]}.$$

Si la frecuencia de falla  $\frac{1}{\theta}$ , es constante, la función de densidad de probabilidad del tiempo de vida medio es la exponencial negativa. Esto es, si  $h(t) = \frac{1}{\theta}$ , entonces de la anterior ecuación se tiene:

$$f(t) = \frac{1}{\theta} \cdot e^{-\int_0^t \frac{1}{\theta} dx} = \frac{1}{\theta} e^{-\frac{t}{\theta}}$$

Dado que la función de confiabilidad a tiempo  $t$  para un tiempo de vida medio distribuido exponencialmente es:

$$R(t) = e^{-\frac{t}{\theta}}, \quad t > 0.$$

La frecuencia de falla está dada por:

$$h(t) = \frac{\frac{1}{\theta} e^{-\frac{t}{\theta}}}{e^{-\frac{t}{\theta}}} = \frac{1}{\theta}, \quad t > 0$$

## 5.7 La distribución de una función de variable aleatoria

Se examinará una técnica para determinar la distribución de una función de variable aleatoria, considerando el caso de una variable aleatoria continua. Sea  $X$  una variable aleatoria con función de densidad de probabilidad  $f_X(x)$ , y sea  $Y = g(X)$  una función definida de  $X$ . Supóngase que es posible resolver  $y = g(x)$  para  $x$  obteniendo de esta forma la función inversa  $x = g^{-1}(y)$ . Si  $g(x)$  y  $g^{-1}(y)$  son funciones univalueadas de  $x$  e  $y$ , respectivamente, se dice que la transformación es uno a uno. Esto es, a cada punto en el espacio muestral de  $X$  le corresponde un punto único del espacio muestral de  $Y$  y viceversa. Si se supone la existencia de una transformación uno a uno y además que  $y = g(x)$  es una creciente y diferenciable en  $x$ , se puede determinar la función de densidad de probabilidad de  $X$  en la siguiente forma:

Debido a la existencia de una transformación uno a uno:

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq y] = P[X \leq g^{-1}(y)].$$

Entonces:

$$F_Y(y) = F_X[g^{-1}(y)].$$

Al establecer la diferencia con respecto a  $y$  y mediante el empleo de la regla de la cadena se tiene:

$$f_Y(y) = \frac{dF_X[g^{-1}(y)]}{dx} \cdot \frac{dx}{dy} = f_X[g^{-1}(y)] \cdot \frac{dx}{dy}.$$

Si  $g(x)$  es una función decreciente de  $x$ , el resultado que se obtiene es el mismo con excepción de que la derivada de una función decreciente es negativa. De esta manera se puede formular la siguiente proposición:

**Teorema 5.2** Sea  $X$  una variable aleatoria continua con función de densidad de probabilidad  $f_X(x)$  y definase  $Y = g(X)$ . Si  $y = g(x)$  y  $x = g^{-1}(y)$  son funciones univalueadas, continuas y diferenciables y si  $y = g(x)$  es una función creciente o decreciente de  $x$ , la función de densidad de probabilidad de  $Y$  está determinada por:

$$f_Y(y) = f_X[g^{-1}(y)] \left| \frac{dx}{dy} \right|,$$

en donde la cantidad  $J = |dx/dy|$  recibe el nombre de Jacobiano de la transformación. ■

## 5.8 Conceptos básicos en la generación de números aleatorios por computadora

**Teorema 5.3** Para cualquier variable aleatoria continua  $X$ , la función de distribución acumulativa  $F(x; \theta)$  con parámetro  $\theta$  se puede representar por una variable aleatoria  $U$ , la cual encuentra uniformemente distribuida sobre el intervalo unitario.

Demostración.- Dado que por definición la función de distribución acumulativa de  $X$  está dada por:

$$F(x; \theta) = \int_{-\infty}^{\infty} f(t; \theta) dt,$$

a cada valor de  $x$  le corresponde un valor de  $F(x; \theta)$  que necesariamente se encuentra en el intervalo  $(0, 1)$ . Además,  $f(X; \theta)$  también es una variable aleatoria en virtud de la aleatoriedad de  $X$ . Para cada valor  $u$  de la variable aleatoria  $U$ , la función  $u = F(x; \theta)$  define una correspondencia uno a uno entre  $U$  y  $X$  siendo la relación inversa  $x = F^{-1}(u)$ . Al tener  $du = dF(x; \theta) = f(x; \theta) dx$ , el Jacobiano de la transformada es:

$$J = \left| \frac{dx}{dy} \right| = [f(x; \theta)]^{-1} = \left[ f(F^{-1}(u); \theta) \right]^{-1}.$$

La función de densidad de probabilidad de la variable aleatoria  $U$ , mediante el empleo de

$$f_Y(y) = f_X \left[ g^{-1}(y) \right] \left| \frac{dx}{dy} \right|,$$

es:

$$g(u) = f \left[ F^{-1}(u); \theta \right] \left[ F^{-1}(u); \theta \right]^{-1} = 1, \quad 0 \leq u \leq 1.$$

■

La esencia del teorema recae en el hecho de que, para muchos casos, es posible determinar de manera directa el valor de  $x$  que corresponde al valor de  $u$  de las variables aleatorias  $X$  y  $U$ , respectivamente, de manera tal que  $F(x; \theta) = u$ .

### 5.8.1 Distribución uniforme sobre el intervalo (a,b)

La función de densidad de probabilidad es:

$$f(x; a, b) = \frac{1}{b-a}, \quad a \leq x \leq b.$$

Para generar un número aleatorio  $x$ ,  $a \leq x \leq b$ , primero se genera un valor aleatorio  $u$  a partir de  $(0, 1)$ , se iguala a la función de distribución acumulativa, se integra y se resuelve para el límite superior  $x$ . De forma que:

$$(b-a)^{-1} \int_a^x dt = u \Rightarrow \frac{x-a}{b-a} = u, \quad \text{o} \quad x = u(b-a) + a, \quad a \leq x \leq b.$$

### 5.8.2 La distribución de Weibull

La función de densidad de probabilidad es:

$$f(x; \alpha, \theta) = \frac{\alpha}{\theta^\alpha} x^{\alpha-1} e^{-(x/\theta)^\alpha}, \quad x > 0.$$

Para generar números aleatorios de Weibull  $x > 0$ , se resuelve la ecuación

$$\frac{\alpha}{\theta^\alpha} \int_0^x t^{\alpha-1} e^{-(t/\theta)^\alpha} dt = u \Rightarrow \left(\frac{\alpha}{\theta^\alpha}\right) \left(-\frac{\theta^\alpha}{\alpha}\right) e^{-(x/\theta)^\alpha} = u$$

$$o \quad 1 - e^{-(x/\theta)^\alpha} = u, \quad y \quad x = \theta \left[ \ln \left( \frac{1}{1-u} \right) \right]^{1/\alpha}.$$

Dado que para  $\alpha = 1$ , la distribución de Weibull se reduce a la exponencial, donde pueden generarse números aleatorios para una distribución.

### 5.8.3 La distribución de Erlang

La función de densidad de probabilidad es:

$$f(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, \quad x > 0,$$

en donde  $\alpha$  es un entero positivo. Recuérdese que la variable aleatoria de Erlang es la suma de  $\alpha$  variables aleatorias independientes distribuidas exponencialmente. Por lo tanto, un número aleatorio de Erlang es la suma de  $\alpha$  valores aleatorios exponenciales, en donde cada valor se genera mediante

$$x = \theta \left[ \ln \left( \frac{1}{1-u} \right) \right]^{1/\alpha}.$$

### 5.8.4 La distribución normal

La función de distribución acumulativa normal es:

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = u$$

no puede resolverse, en forma cerrada, para  $x$ . De manera alternativa, puede demostrarse que si  $U_1$  y  $U_2$  son dos variables aleatorias independientes con distribución uniforme sobre el intervalo unitario, entonces

$$Z_1 = (-2 \ln U_1)^{1/2} \sin(2\pi U_2) \quad y \quad Z_2 = (-2 \ln U_1)^{1/2} \cos(2\pi U_2)$$

son dos variables aleatorias normales estandarizadas e independientes.

### 5.8.5 La distribución binomial

Para generar números aleatorios a partir de una distribución binomial con función de probabilidad se considerará lo siguiente: la variable aleatoria binomial es vista como la suma de  $n$  resultados de un proceso de Bernoulli descrito por:

$$p(x; n, p) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n$$

$$y = \begin{cases} 1 & \text{si } 0 \leq u \leq p, \\ 0 & \text{si } p < u \leq 1. \end{cases}$$

donde  $u$  es un número aleatorio uniforme sobre el intervalo unitario. Esto es, se generan  $n$  números aleatorios a partir del intervalo unitario, se convierten a unos y ceros de acuerdo con  $y$  y la suma de los unos en esta secuencia es el número aleatorio binomial.

### 5.8.6 La distribución Poisson

Recuérdese que la probabilidad de tener  $x$  ocurrencias en un intervalo de tiempo  $t$  está definida por:

$$p(x; t) = \frac{(vt)^x e^{-vt}}{x!}, \quad x = 0, 1, 2, \dots,$$

donde  $v$  es la frecuencia constante de ocurrencia, y  $\lambda = vt$  es el número promedio de éstas. Como la ocurrencia en el tiempo de dos eventos independientes de Poisson se encuentra distribuida exponencialmente, se puede generar un número aleatorio de Poisson  $x$  mediante la generación sucesiva de números aleatorios exponenciales por  $x = \theta \left[ \ln \left( \frac{1}{1-u} \right) \right]^{1/\alpha}$  para  $\alpha = 1$ . El proceso se continúa hasta que la suma de los valores  $x + 1$  sea mayor que el intervalo de tiempo  $t$ . Por lo tanto, el número aleatorio de Poisson es  $x$ .

## Distribuciones conjuntas de probabilidad

### 6.1 Distribución de probabilidad bivariadas

**Definición 6.1** Sean  $X$  e  $Y$  dos variables aleatorias discretas. La probabilidad de que  $X = x$  e  $Y = y$  está determinada por la función de probabilidad bivariada

$$p(x, y) = P(X = x, Y = y),$$

en donde  $P(x, y) \geq 0$  para toda  $x, y$ , de  $X, Y$ , y  $\sum_x \sum_y p(x, y) = 1$ .

La **función de distribución acumulativa bivariada** es la probabilidad conjunta de que  $X \leq x$  y  $Y \leq y$ , dada por

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_i \leq y} p(x_i, y_i).$$

La **función de distribución trinomial** viene dado por:

$$p(x, y, n, p_1, p_2) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}$$

y su generalización llamada **función de distribución multinomial** viene dada por:

$$p(x_1, x_2, \dots, x_{k-1}; n, p_1, p_2, \dots, p_{k-1}) = \frac{n!}{x_1!x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}, \quad x_i = 0, 1, \dots, n \text{ para } i = 1, 2, \dots, k$$

en donde  $x_k = n - x_1 - x_2 - \dots - x_{k-1}$  y  $p_k = 1 - p_1 - p_2 - \dots - p_{k-1}$ .

**Definición 6.2** Sean  $X$  e  $Y$  dos variables aleatorias continuas. Si existe una función  $f(x, y)$  tal que la probabilidad conjunta:

$$P(a < X < b, c < Y < d) = \int_a^b \int_c^d f(x, y) \, dy \, dx$$

para cualquier valor de  $a, b, c$  y  $d$  en donde  $f(x, y) \geq 0$ ,  $-\infty < x, y < \infty$  y  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dy \, dx = 1$ , entonces  $f(x, y)$  es la función de densidad de probabilidad bivariada de  $X$  e  $Y$ .

La **función de distribución bivariada acumulativa** de  $X$  e  $Y$  es la probabilidad conjunta de que  $X \leq x$  e  $Y \leq y$ , dada por:

$$P(X \leq x, Y \leq y) = F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, dv \, du.$$

Por lo tanto, la función de densidad bivariada se encuentra diferenciando  $F(x, y)$  con respecto a  $x$  e  $y$ ; es decir,

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

## 6.2 Distribuciones marginales de probabilidad

Es posible determinar varias distribuciones marginales para cualquier distribución de probabilidad que contenga más de dos variables aleatorias.

**Definición 6.3** Sean  $X$  e  $Y$  dos variables aleatorias discretas con una función de probabilidad conjunta  $p(x, y)$ . Las funciones marginales de probabilidad de  $X$  y de  $Y$  están dadas por

$$p_X(x) = \sum_y p(x, y) \quad \text{y} \quad p_Y(y) = \sum_x p(x, y),$$

respectivamente.

**Definición 6.4** Sean  $X$  e  $Y$  dos variables aleatorias continuas con una función de densidad de probabilidad conjunta  $f(x, y)$ . Las funciones de densidad de probabilidad de  $X$  e  $Y$  están dadas por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy \quad \text{y} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, dx,$$

respectivamente.

Para variables aleatorias continuas conjuntas, si se conoce **la función de distribución acumulativa**  $F(x, y)$ , las distribuciones acumulativas marginales de  $X$  e  $Y$  se obtienen de la siguiente forma:

$$P(X \leq x) = F_X(x) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(t, y) dy dt, \quad y \quad F_X(x) = \int_{-\infty}^x f_X(t) dt = F(x, \infty)$$

De manera similar

$$P(Y \leq y) = F_Y(y) = \int_{-\infty}^y \int_{-\infty}^{\infty} f(x, t) dx dt = \int_{-\infty}^y f_Y(t) dt = F(\infty, y).$$

Así puede determinarse la distribución acumulativa marginal de  $X$  dejando que  $Y$  tome un valor igual al límite superior de la función de distribución conjunta de  $X$  e  $Y$ .

### 6.3 Valores esperados y momentos para distribuciones bivariadas

**Definición 6.5** Sean  $X$  e  $Y$  dos variables aleatorias que se distribuyen conjuntamente. El valor esperado de una función de  $X$  y de  $Y$ ,  $g(x, y)$ , se define como

$$E[g(X, Y)] = \sum_x \sum_y g(x, y) p(x, y)$$

si  $X$  e  $Y$  son discretas, o

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dy dx$$

si  $X$  e  $Y$  son continuas, en donde  $p(x, y)$  y  $f(x, y)$  son las funciones de probabilidad y de densidad de probabilidad conjuntas, respectivamente.

Como consecuencia de la definición anterior, el **r-ésimo momento de  $X$  alrededor del cero** es

$$E(X^r) = \int_{-\infty}^{\infty} x^r f(x, y) dy dx = \int_{-\infty}^{\infty} x^r f_X(x) dx.$$

De manera similar

$$E(Y^r) = \int_{-\infty}^{\infty} y^r f_Y(y) dy.$$

El **r y s-ésimo momento producto de  $X$  e  $Y$  alrededor del origen** es:

$$E(X^r Y^s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s f(x, y) dy dx.$$

y **alrededor de las medias** es



$$E[(X - \mu_X)^r (Y - \mu_Y)^s] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^r (y - \mu_Y)^s f(x, y) dy dx.$$

De particular importancia es el momento producto alrededor de las medias cuando  $r = s = 1$ . Este momento producto recibe el nombre de **covarianza de X e Y**, y se encuentra definido por

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

Al igual que la varianza, que es una medida de dispersión de una variable aleatoria, la covarianza es una medida de la variabilidad conjunta de X e Y. De esta forma, la covarianza es una medida de asociación entre los valores de X e Y y sus respectivas dispersiones.

Desarrollando el miembro derecho de  $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$  se tiene

$$E[(X - \mu_X)(Y - \mu_Y)] = E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] = E(XY) - \mu_X\mu_Y;$$

de esta forma

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Si la covarianza de X e Y se divide por el producto de las desviaciones estándar de X e Y, el resultado es una cantidad sin dimensiones que recibe el nombre de **coeficiente de correlación** y que se denota por  $\rho(X, Y)$ :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Se puede demostrar que el coeficiente de correlación se encuentra contenido en el intervalo  $-1 \leq \rho \leq 1$ . De hecho  $\rho$  es la covarianza de dos variables aleatorias estandarizadas  $X'$  e  $Y'$  en donde

$$X' = \frac{X - \mu_X}{\sigma_X} \quad \text{e} \quad Y' = \frac{Y - \mu_Y}{\sigma_Y}.$$

Esto significa que el coeficiente de correlación es sólo una medida estandarizada de la asociación lineal que existe entre las variables aleatorias X e Y en relación con sus dispersiones. El valor  $\rho = 0$  indica la ausencia de cualquier asociación lineal, mientras que los valores  $-1$  y  $1$  indican relaciones lineales perfectas negativas y positivas, respectivamente.

## 6.4 Variables aleatorias estadísticamente independientes

**Definición 6.6** Sean  $X$  e  $Y$  dos variables aleatorias con una distribución conjunta. Se dice que  $X$  e  $Y$  son estadísticas independientes si y sólo si,

$$p(x, y) = p_X(x)p_Y(y) \text{ si } X \text{ e } Y \text{ son discretas}$$

o bien

$$f(x, y) = f_X(x)f_Y(y) \text{ si } X \text{ e } Y \text{ son continuas,}$$

para todo  $x$  e  $y$ , en donde  $p(x, y)$  y  $f(x, y)$  son las funciones bivariadas de probabilidad y de densidad de probabilidad, respectivamente, y en donde  $p_X(x)$ ,  $p_Y(y)$ ,  $f_X(x)$  y  $f_Y(y)$  son las funciones de probabilidad marginal o de densidad de probabilidad marginal apropiadas.

Se desprende de esta definición que si  $X$  e  $Y$  son estadísticamente independientes, la probabilidad conjunta

$$\begin{aligned} P(a < X < b, c < Y < d) &= \int_a^b \int_c^d f(x, y) \, dy \, dx \\ &= \int_a^b \int_c^d f_X(x)f_Y(y) \, dy \, dx \\ &= \int_a^b f_X(x) \, dx \int_c^d f_Y(y) \, dy \\ &= P(a < X < b)P(c < Y < d). \end{aligned}$$

Por la misma condición,

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y) \, dy \, dx \\ &= \int_a^b \int_c^d xyf_X(x)f_Y(y) \, dy \, dx \\ &= \int_c^d xf_X(x) \, dy \int_a^b yf_Y(y) \, dx \\ &= E(X)E(Y). \end{aligned}$$

Si  $X$  e  $Y$  son estadísticamente independientes, entonces  $\text{Cov}(X, Y) = \rho(X, Y) = 0$ . Sin embargo la proposición inversa no es cierta.

Sean  $X$  e  $Y$  dos variables aleatorias continuas con una función de densidad conjunta de probabilidad  $f(x, y)$ . El **valor esperado de una función lineal** de  $X$  e  $Y$  es

$$\begin{aligned} E(aX + bY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by)f(x, y) \, dy \, dx \\ &= a \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y) \, dy \, dx + b \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x, y) \, dy \, dx \\ &= aE(X) + bE(Y). \end{aligned}$$

para cualquier valor de las constantes  $a$  y  $b$ .

La varianza de una función lineal de  $X$  e  $Y$  es

$$\begin{aligned}\text{Var}(aX + bY) &= E(aX + bY)^2 - E^2(aX + bY) \\ &= E(a^2X^2 + 2abXY + b^2Y^2) - [aE(X) + bE(Y)]^2 \\ &= a^2E(X^2) + 2abE(XY) + b^2E(Y^2) - a^2E^2(X) - 2abE(X)E(Y) - b^2E^2(Y) \\ &= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).\end{aligned}$$

Además si  $X$  e  $Y$  son estadísticamente independientes,

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y).$$

La generalización de estos resultados a  $n$  variables aleatorias se hace por inducción y se establece en el siguiente teorema:

**Teorema** Sean  $X_1, X_2, \dots, X_n$   $n$  variables aleatorias con una función de densidad conjunta de probabilidad  $f(x_1, x_2, \dots, x_n)$ .  
6.1 Entonces

$$\begin{aligned}E\left[\sum_{i=1}^n a_i X_i\right] &= \sum_{i=1}^n [a_i E(X_i)] \\ \text{Var}\left[\sum_{i=1}^n a_i X_i\right] &= \sum_{i=1}^n a_i^2 \text{Var}(X_i) + \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_i, X_j)\end{aligned}$$

para cualquier constante  $a_i$ ,  $i = 1, 2, \dots, n$ . ■

## 6.5 Distribuciones de probabilidad condicional

Considere la función

$$\frac{f(x, y)}{f_Y(y)},$$

en donde  $f_Y(y)$  es la densidad marginal de  $Y$ . Si se mantiene constante a la variable aleatoria  $Y$  en el valor observado y de manera tal que  $f_Y(y) > 0$ , entonces  $\frac{f(x, y)}{f_Y(y)}$  define una función no negativa de  $X$  cuya integral es 1, dado que por definición

$$\int_{-\infty}^{\infty} \frac{f(x, y)}{f_Y(y)} dx = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(x, y) dx = \frac{f_Y(y)}{f_Y(y)} = 1.$$

De esta forma,  $\frac{f(x, y)}{f_Y(y)}$  es una función de densidad de probabilidad y la probabilidad de que  $a < X < b$ , dado que el nivel de concentración de  $Y$  es  $y$ , está dada por:

$$P(a < X < b/y) = \int_a^b \frac{f(x, y)}{f_Y(y)} dx.$$

**Definición** Sean  $X$  e  $Y$  dos variables aleatorias con una función de densidad conjunta de probabilidad  $f(x, y)$ .  
**6.7** La función de densidad de probabilidad condicional de la variable aleatoria  $X$ , denotada por  $f(x|y)$ , para un valor fijo  $y$  de  $Y$ , está definida por

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

En donde  $f_Y(y)$  es la función de densidad de probabilidad de  $Y$  de manera tal que  $F_Y(y) > 0$ .

De manera análoga, la función de densidad de probabilidad condicional de  $Y$  para un valor fijo  $x$  de  $X$  se define como

$$f(y|x) = \frac{f(x, y)}{f_X(x)} \quad f_X(x) > 0.$$

En donde  $f_X(x)$  es la densidad marginal de  $X$ .

Nótese que si la densidad condicional  $f(x|y)$  por ejemplo, no contiene a  $y$ , entonces  $X$  es estadísticamente independientes de  $Y$ . Esto es, si  $X$  son estadísticamente independientes entonces

$$f(x, y) = f_X(x)f_Y(y)$$

y

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x).$$

De manera similar, si

$$f(x, y) = f_X(x)f_Y(y),$$

entonces

$$f(y|x) = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y)$$

Los **valores esperados condicionales** se definen de manera análoga a la señalada en la definición 6.5. Por ejemplo, los valores esperados condicionales de  $X$  puesto que  $Y = y$ , y de  $Y$ , ya que  $X = x$ , se definen como

$$E(X|y) = \int_{-\infty}^{\infty} xf(x|y) dx.$$

Y

$$E(Y|x) = \int_{-\infty}^{\infty} yf(y|x) dy.$$

De manera similar,

$$Var(X|y) = E(X^2|y) - E^2(X|y).$$

Y

$$Var(Y|x) = E(Y^2|x) - E^2(Y|x).$$

En donde

$$E(X^2|y) = \int_{-\infty}^{\infty} x^2 f(x|y) dx \quad y \quad E(Y^2|x) = \int_{-\infty}^{\infty} y^2 f(y|x) dy.$$

## 6.6 Análisis bayesiano: las distribuciones a priori y a posteriori.

**Teorema 6.2** Sea  $p_Y(y)$  o  $f_Y(y)$  la función de probabilidad o de densidad de probabilidad a priori de  $Y$ , respectivamente, y sea  $f(x|y)$  la función de verosimilitud. Entonces la probabilidad a posteriori o función de probabilidad a posteriori de  $Y$  dada la evidencia muestral  $x$ , es

$$p(y|x) = \frac{f(x|y)p_Y(y)}{\sum_Y f(x|y)p_Y(y)} \text{ si } Y \text{ es discreta ,}$$

$$p(y|x) = \frac{f(x|y)p_Y(y)}{\int_{-\infty}^{\infty} f(x|y)p_Y(y) dy} \text{ si } Y \text{ es continua .}$$

■

Es interesante notar que el denominador de es la función de densidad de probabilidad marginal o no condicional de  $X$ ; esto es,

$$f_X(x) = \sum_Y f(x|y)p_Y(y)$$

o

$$f_X(x) = \int_Y f(x|y)f_Y(y) dy.$$

Dependiendo de cuando  $Y$  es discreta o continua. Además, el numerador es el producto de la función de verosimilitud y la función de probabilidad a priori y, de ésta manera, es la probabilidad conjunta de  $X$  e  $Y$  expresado como

$$f(x, y) = f(x|y)p_Y(y) \text{ si } Y \text{ es discreta,}$$

o

$$f(x, y) = f(x|y)f_Y(y) dy \text{ si } Y \text{ es continua.}$$

Nótese que para la penúltima ecuación la función  $f(x, y)$  es una mezcla bivariada de una variable aleatoria continua y otra discreta.

## 6.7 La distribución normal bivariada

**Definición 6.8** Se dice que las variables aleatorias  $X$  e  $Y$  tienen una distribución normal bivariada si su función de densidad conjunta de probabilidad está dada por

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[ \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \left( \frac{x-\mu_X}{\sigma_X} \right) \left( \frac{y-\mu_Y}{\sigma_Y} \right) + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right]}$$

con  $-\infty < x, y < \infty$ . Donde

$$\mu_X = E(X), \quad \mu_Y = E(Y), \quad \sigma_X^2 = \text{Var}(X), \quad \sigma_Y^2 = \text{Var}(Y),$$

y  $\rho$  es el coeficiente de correlación de  $X$  e  $Y$ .

Es interesante notar que, a pesar de que  $\rho = 0$  es una condición necesaria de independencia, para la distribución normal bivariada también es una condición suficiente. Eso es, si  $\rho = 0$  entonces

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2} \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - \frac{1}{2} \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}} \\ &= f_X(x)f_Y(y). \end{aligned}$$

Donde,  $f_X(x)$  e  $f_Y(y)$  son las densidad normales univariadas de  $X$  e  $Y$ , respectivamente.

Se puede demostrar que, mediante el empleo de la definición 6.8 e integrando con respecto a  $y$ , la densidad marginal de  $X$  es normal con media  $\mu_X$  y varianza  $\sigma_X^2$ . De manera similar, la densidad marginal de  $Y$  es normal con media  $\mu_Y$  y varianza  $\sigma_Y^2$ . por la definición 6.7, la densidad de probabilidad condicional de  $X$  dado el valor  $y$  de  $Y$  es

$$f(x|y) = \frac{1}{\sqrt{2\pi\sigma_X^2(1-\rho^2)}} \times e^{\left\{ -\frac{1}{2\sigma_X^2(1-\rho^2)} \left[ x - \mu_X - \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y) \right]^2 \right\}}.$$

Esta expresión es una función de densidad de probabilidad normal con

$$E(X|y) = \mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y) \quad \text{y} \quad \text{Var}(X|y) = \sigma_X^2(1 - \rho^2).$$

Se puede obtener una expresión similar para la densidad condicional de  $Y$  dado el valor  $x$  de  $X$ .

## Muestras aleatorias y distribuciones de muestreo

**Definición 7.1** Si las variables aleatorias  $X_1, X_2, \dots, X_n$  tiene la misma función (densidad) de probabilidad que la de la distribución de la población y su función (distribución) conjunta de probabilidad es igual al producto de las marginales, entonces  $X_1, X_2, \dots, X_n$  forman un conjunto de  $n$  variables aleatorias independientes e idénticamente distribuidas (IID) que constituyen una muestra aleatoria de la población.

En el contexto de la definición 7.1, la función (densidad) conjunta de probabilidad de  $X_1, x_2, \dots, X_n$  es la función de verosimilitud de la muestra dada por

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

en donde  $x = \{x_1, x_2, \dots, x_n\}$  denota los datos muestreados. Cuando las realizaciones  $x$  se conocen,  $L(x; \theta)$  es una función del parámetro desconocido  $\theta$ .

**Ejemplo 7.1** Se ilustrará el concepto de muestra aleatoria dada en la definición 7.1 mediante lo siguiente: Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de  $n$  variables aleatorias IID de una población cuya distribución de probabilidad es exponencial con densidad

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}, \quad 0 < x < \infty.$$

Cuando se observa  $X_1$  y se registra su realización  $x_1$ ,

$$f(x_1; \theta) = \frac{1}{\theta} e^{-x_1/\theta}, \quad 0 < x_1 < \infty.$$

Ahora se observa  $X_2$  y se registra su realización  $x_2$ . Dado que  $X_1$  y  $X_2$  son estadísticamente independientes y tienen las mismas densidades marginales,

$$f(x_2|x_1) = f(x_2; \theta) = \frac{1}{\theta} e^{-x_2/\theta}, \quad 0 < x_2 < \infty.$$

La función de densidad conjunta de  $X_1$  y  $X_2$  es

$$f(x_1, x_2; \theta) = f(x_1; \theta) f(x_2; \theta) = \frac{1}{\theta^2} e^{-(x_1+x_2)/\theta}, \quad 0 < x_i < \infty, i = 1, 2.$$

Por lo tanto, se desprende que para una muestra aleatoria de tamaño  $n$

$$L(x_1, x_2, \dots, x_n; \theta) = \frac{1}{\theta} e^{-(x_1 + x_2 + \dots + x_n)/\theta}, \quad 0 < x_i < \infty, i = 1, 2, \dots, n.$$

■

### 7.3 Distribuciones de muestreo de estadísticas

**Definición 7.2** Un parámetro es una caracterización numérica de las distribuciones de la población de manera que describe, parcial o completamente, la función de densidad de probabilidad de la característica de interés. Por ejemplo, cuando se especifica el valor del parámetro de escala exponencial  $\theta$ , se describe de manera completa la función de probabilidad

$$f(x; \theta) = \frac{1}{\theta} e^{(-x/\theta)}.$$

La oración describe de manera completa, sugiere que una vez que se conoce el valor de  $\theta$  entonces, puede formularse cualquier proposición probabilística de interés.

Dado que los parámetros son prácticamente inherentes a todos los modelos de probabilidad, es imposible calcular las probabilidades deseadas sin un conocimiento del valor de estos. Es por esta razón que la noción de una estadística y su distribución de muestreo es muy importante en inferencia estadística. Antes de dar la definición de una estadística, debe notarse que desde un punto de vista clásico (no bayesiano), un parámetro se considera como una constante fija cuyo valor se desconoce. Desde una perspectiva bayesiana un parámetro siempre es una variable aleatoria con algún tipo de distribución de probabilidad.

**Definición 7.3** Una estadística es cualquier función de las variables aleatorias que se observaron en la muestra de manera que esta función no contiene cantidades desconocidas.

De manera general, denótese una estadística por  $T = u(X)$ . Dado que  $T$  es una función de variables aleatorias, es en sí misma una variable aleatoria, y su valor específico  $t = u(x)$  puede determinarse cuando se conozcan las realizaciones  $x$  de  $X$ . Si se emplea una estadística  $T$  para estimar un parámetro desconocido  $\theta$ , entonces  $T$  recibe el nombre de **estimador** de  $\theta$ , y el valor específico de  $t$  como un resultado de los datos muestrales recibe el nombre de **estimación** de  $\theta$ . Esto es, un estimador es una estadística que identifica al mecanismo funcional por medio del cual, una vez que las observaciones en la muestra se realizan, se obtiene una estimación.

Un parámetro es una constante pero una estadística es una variable aleatoria. Además, un valor del parámetro descrito describe de manera completa un modelo de probabilidad; ningún valor de la estadística puede desempeñar tal papel si cada uno de estos depende del valor de las observaciones de las muestras.

**Definición 7.4** La distribución de muestreo de una estadística  $T$  es la distribución de probabilidad de  $T$  que puede obtenerse como resultado de un número infinito de muestras aleatorias independientes, cada una de tamaño  $n$ , provenientes de la población de interés.



La distribución de muestreo de una estadística hace posible este tipo de análisis de probabilidad, esencial para valorar el riesgo inherente cuando se formulan ingerencias.

**Teorema 7.1** Sea  $X_1, X_2, \dots, X_n$  un conjunto de  $n$  variables aleatorias independientes cada una con función generadora de momentos  $m_{X_1}(t), m_{X_2}(t), \dots, m_{X_n}(t)$ . Si

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n,$$

en donde  $a_1, a_2, \dots, a_n$  son constantes, entonces:

$$m_Y(t) = m_{X_1}(a_1 t) m_{X_2}(a_2 t) \cdots m_{X_n}(a_n t).$$

Demostración.- Mediante el empleo de la definición y la hipótesis de independencia, se tiene

$$\begin{aligned} m_Y(t) &= E \left\{ e^{[t(a_1 X_1 + a_2 X_2 + \dots + a_n X_n)]} \right\} \\ &= E \left\{ e^{[t(a_1 X_1)]} e^{[t(a_2 X_2)]} \dots e^{[t(a_n X_n)]} \right\} \\ &= E \left\{ e^{[t(a_1 X_1)]} \right\} E \left\{ e^{[t(a_2 X_2)]} \right\} \dots E \left\{ e^{[t(a_n X_n)]} \right\} \\ &= m_{X_1}(a_1 t) m_{X_2}(a_2 t) \cdots m_{X_n}(a_n t). \end{aligned}$$

■

De esta forma, la función generadora de momentos de una combinación lineal de  $n$  variables aleatorias independientes es el producto de las correspondientes funciones generadoras de momentos con argumentos iguales a las constantes de tiempo  $t$ .

**Teorema 7.2** Sea  $X_1, X_2, \dots, X_n$  un conjunto de variables aleatorias normalmente distribuidas con medias  $E(X_i) = \mu_i$  y varianzas  $Var(X_i) = \sigma_i^2$  para  $i = 1, 2, \dots, n$ . Si

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n,$$

en donde  $a_1, a_2, \dots, a_n$  son constantes, entonces  $Y$  es una variable aleatoria con distribución normal y media

$$E(Y) = a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n$$

y con varianza

$$Var(Y) = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2.$$

Demostración.- Dado que  $X_i$  se encuentra normalmente distribuida, su función generadora de momentos es

$$m_{X_i}(t) = e^{\left( \mu_i t + \frac{\sigma_i^2 t^2}{2} \right)}.$$

De acuerdo con el teorema 7.1, la función generadora de momentos de  $Y$  es

$$\begin{aligned} m_Y(t) &= m_{X_1}(a_1 t) m_{X_2}(a_2 t) \cdots m_{X_n}(a_n t) \\ &= e^{\left( \mu_1 a_1 t + \frac{a_1^2 \sigma_1^2 t^2}{2} \right)} e^{\left( \mu_2 a_2 t + \frac{a_2^2 \sigma_2^2 t^2}{2} \right)} \dots e^{\left( \mu_n a_n t + \frac{a_n^2 \sigma_n^2 t^2}{2} \right)} \\ &= e^{\left[ t \sum_{i=1}^n a_i \mu_i + \frac{t^2 \sum_{i=1}^n a_i^2 \sigma_i^2}{2} \right]}. \end{aligned}$$

Por lo tanto,  $Y$  se encuentra normalmente distribuida con media  $\sum_{i=1}^n a_i \mu_i$  y varianza  $\sum_{i=1}^n a_i^2 \sigma_i^2$ .

■

Del teorema 7.2 se desprende que si  $a_i = 1$  para  $i = 1, 2, \dots, n$ , entonces la suma de variables aleatorias independientes normalmente distribuidas también posee una distribución normal con media y varianza igual a la suma de las medias y las varianzas de cada una de las variables aleatorias. El resultado anterior se conoce como la propiedad aditiva de la distribución normal. Debe notarse que la hipótesis de normalidad no es necesaria para obtener las fórmulas de la media y la varianza de  $Y$  en el teorema 7.2. De hecho, con base en el teorema 6.1, si  $X_1, X_2, \dots, X_n$  es un conjunto de  $n$  variables aleatorias IID con medias  $E(X_i) = \mu_i$  y varianzas  $Var(X_i) = \sigma_i^2$ ,  $i = 1, 2, \dots, n$  entonces para  $Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$ ,

$$E(Y) = \sum_{i=1}^n a_i \mu_i$$

y

$$Var(Y) = \sum_{i=1}^n a_i^2 \sigma_i^2.$$

en donde  $a_1, a_2, \dots, a_n$  son constantes.

#### 7.4 La distribución de muestreo de $\bar{X}$

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria que consiste en  $n$  variables aleatorias IID tales que  $E(X_i) = \mu$  y  $Var(X_i) = \sigma^2$  para toda  $i = 1, 2, \dots, n$ . Entonces la estadística

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

se define como la media de las  $n$  variables aleatorias IID o, sencillamente, media muestral. Nótese que una vez que se conocen las realizaciones  $x_1, x_2, \dots, x_n$  de  $X_1, X_2, \dots, X_n$ , respectivamente, realización  $\bar{x}$  de  $\bar{X}$  se obtiene promediando los datos muestrales.

Si en,  $E(Y) = \sum_{i=1}^n a_i \mu_i$  y  $Var(Y) = \sum_{i=1}^n a_i^2 \sigma_i^2$ ,  $a_i = 1/n$ ,  $i = 1, 2, \dots, n$  entonces el valor esperado y la varianza de  $\bar{X}$  son

$$E(\bar{X}) = \sum_{i=1}^n \frac{1}{n} \mu = n \left( \frac{\mu}{n} \right) = \mu.$$

y

$$Var(\bar{X}) = \sum_{i=1}^n \frac{1}{n^2} \sigma^2 = n \left( \frac{\sigma^2}{n^2} \right) = \frac{\sigma^2}{n},$$

respectivamente, en donde  $\mu$  y  $\sigma^2$  son la media y la varianza de la distribución de la población a partir de la cual se obtuvo la muestra. Luego de esta última ecuación de  $Var(\bar{X})$  la desviación estándar de  $\bar{X}$  es

$$d.e.(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

la cual, en algunas ocasiones recibe el nombre de **error estándar de la media**. Si el tamaño de la muestra crece, la precisión de la media muestral para estimar la media poblacional aumenta.

**Teorema 7.3** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria que consiste de  $n$  variables aleatorias independientes y normalmente distribuidas con media  $E(X_i) = \mu$  y varianzas  $Var(X_i) = \sigma^2$ ,  $i = 1, 2, \dots, n$ . Entonces la

distribución de la media muestral  $\bar{X}$  es normal con media  $\mu$  y varianza  $\sigma^2/n$ .

Demostración.- Este teorema es un corolario del teorema 7.2. Esto es, sea  $a_i = 1/n$ ; dado que las medias y las varianzas son iguales, respectivamente, la función generadora de momentos de  $\bar{X}$  es

$$\begin{aligned} m_{\bar{X}}(t) &= e^{\left( t \sum_{i=1}^n \frac{1}{n} \mu + \frac{t^2 \sum_{i=1}^n \frac{1}{n^2} \sigma^2}{2} \right)} \\ &= e^{\left( t\mu + \frac{t^2 \sigma^2}{2n} \right)}, \end{aligned}$$

que es la función generadora de momentos de una variable aleatoria normalmente distribuida con media  $\mu$  y varianza  $\sigma^2/n$ . De esta forma, la **función de densidad de probabilidad de  $\bar{X}$  cuando se muestrea una población cuya distribución es normal**, está dada por

$$f(\bar{x}, \mu, \sigma/\sqrt{n}) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{\left[ -\frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right]}, \quad -\sigma < \bar{x} < \sigma.$$

■

**Ejemplo 7.2** Demostrar que si  $X_1, X_2, \dots, X_n$  son  $n$  variables aleatorias independientes exponencialmente distribuidas con función de densidad de probabilidad

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta} \quad x > 0,$$

entre  $\bar{X}$  tiene una distribución gama.

Demostración.- Recuerdese que la función generadora de momentos de una variable aleatoria exponencialmente distribuida es  $(1 - \theta t)^{-1}$ . De esta manera para cada  $X_i$  de la muestra,

$$m_{X_i}(t) = (1 - \theta t)^{-1}.$$

Del teorema 7.1 con  $a_i = \frac{1}{n}$ ,  $i = 1, 2, \dots, n$  se desprende que la función generadora de momentos de la media muestral  $\bar{X}$  es

$$\begin{aligned} m_{\bar{X}}(t) &= m_{X_1}\left(\frac{t}{n}\right) m_{X_2}\left(\frac{t}{n}\right) \dots m_{X_n}\left(\frac{t}{n}\right) \\ &= \left[1 - \left(\frac{\theta t}{n}\right)\right]^{-1} \left[1 - \left(\frac{\theta t}{n}\right)\right]^{-1} \dots \left[1 - \left(\frac{\theta t}{n}\right)\right]^{-1} \\ &= \left[1 - \left(\frac{\theta t}{n}\right)\right]^{-n}. \end{aligned}$$

Pero la expresión anterior es la función generadora de momentos de una distribución gama con parámetro de forma  $n$  y parámetro de escala  $\theta/n$ . De acuerdo con lo anterior, cuando se muestrea una población cuya distribución de probabilidad es exponencial, la densidad de probabilidad de  $\bar{X}$  está dada por

$$f(\bar{x}; n, \theta/n) = \frac{n^n}{\Gamma(n)\theta^n} \bar{x}^{n-1} e^{-n\bar{x}/\theta}, \quad \bar{x} > 0.$$

■

Nótese que si en las expresiones  $E(X) = \alpha\theta$  y  $\text{Var}(X) = \alpha\theta^2$  se reemplaza  $\alpha$  con  $n$  y  $\theta$  con  $\theta/n$  se obtiene

$$E(\bar{X}) = n \frac{\theta}{n} = \theta$$

y

$$\text{Var}(\bar{X}) = n \frac{\theta^2}{n^2} = \frac{\theta^2}{n},$$

como era de esperarse ya que  $\theta$  y  $\theta^2$  son la media y la varianza de una variable aleatoria con distribución exponencial.

De la sección 5.5, recuérdese que si el parámetro de forma de una distribución gama tiene un valor grande, entonces los valores de probabilidad para una variable aleatoria gama pueden aproximarse, en forma adecuada, por una distribución normal. Dado que  $\Gamma^m$ , muestrear una distribución exponencial con parámetro  $\theta$ ,  $\bar{X}$  tiene una distribución gama con media  $\theta$ , y desviación estándar  $\frac{\theta}{\sqrt{n}}$ . Entonces, para  $n$  grande

$$Z = \frac{\bar{X} - \theta}{\frac{\theta}{\sqrt{n}}}.$$

es, en forma aproximada,  $N(0, 1)$ .

Para ser que para un valor grande  $n$ , la distribución de  $\bar{X}$  es aproximadamente normal. De hecho, no importa el tipo de modelo de probabilidad a partir del cual se obtenga la muestra; muestras la media y la varianza existan, la distribución de muestreo de  $\bar{X}$  se encontrará aproximada por  $N(\mu, \sigma/\sqrt{n})$  para valores grandes de  $n$ . Lo anterior constituye uno de los más importantes teoremas en inferencia estadística, y se conoce como **teorema central del límite**.

**Teorema 7.4** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias IID con una distribución de probabilidad no especificada y que tiene una media  $\mu$  y varianza  $\sigma^2$  finita. El promedio muestral  $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$  tiene una distribución con media  $\mu$  y varianza  $\sigma^2/n$  que tiende hacia una distribución normal conforme  $n$  tiende a  $\infty$ . En otras palabras, la variable aleatoria  $(\bar{X} - \mu)/(\sigma/\sqrt{n})$  tiene como límite una distribución normal estándar.

**Demostración.-** Se quiere demostrar que la función generadora de momentos de  $(\bar{X} - \mu)/(\sigma/\sqrt{n})$  tiende a la de una distribución normal estándar conforme  $n$  tiende al infinito. Sean

$$Z_i = \frac{X_i - \mu}{\sigma}, \quad i = 1, 2, \dots, n$$

y

$$Y = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Dado que

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \right) = \frac{1}{n} \frac{1}{\frac{\sigma}{\sqrt{n}}} \sum_{i=1}^n (X_i - \mu) = \frac{1}{n} \frac{1}{\frac{\sigma}{\sqrt{n}}} (n\bar{X} - n\mu) = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Entonces,

$$Y = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i.$$

Como resultado se tiene que la función generadora de momentos de  $Y$  es igual a la función generadora de momentos de  $(1/\sqrt{n}) \sum_{i=1}^n Z_i$ . Del teorema 7.1,

$$\begin{aligned} m_Y(t) &= [m_{Z_i}(t/\sqrt{n})]^n \\ &= \left\{ E \left[ e^{\frac{tZ_i}{\sqrt{n}}} \right] \right\}^n, \end{aligned}$$

dado que las  $Z_i$  son variables aleatorias independientes.

Al expandir  $(tZ_i/\sqrt{n})$  en una serie de Taylor:

$$e^{\frac{tZ_i}{\sqrt{n}}} = 1 + \frac{t}{\sqrt{n}}Z_i + \frac{t^2}{2n}Z_i^2 + \frac{t^3}{3!n^{\frac{3}{2}}}Z_i^3 + \dots$$

Si se toma los valores esperados y se recuerda que  $E(Z_i) = 0$  y  $Var(Z_i) = 1, i = 1, 2, \dots, n$  se tiene

$$E \left[ e^{\frac{tZ_i}{\sqrt{n}}} \right] = 1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{\frac{3}{2}}}E(Z_i^3) + \dots$$

De acuerdo con lo anterior

$$\begin{aligned} m_Y(t) &= \left[ 1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{\frac{3}{2}}}E(Z_i^3) + \dots \right]^n \\ &= \left\{ 1 + \frac{1}{n} \left[ \frac{t^2}{2} + \frac{t^3}{3!\sqrt{n}}E(Z_i^3) + \dots \right] \right\}^n \\ &= \left( 1 + \frac{u}{n} \right)^n \end{aligned}$$

donde

$$u = \frac{t^2}{2} + \frac{t^3}{3!\sqrt{n}}E(Z_i^3) + \dots$$

Ahora

$$\lim_{n \rightarrow \infty} m_Y(t) = \lim_{n \rightarrow \infty} \left( 1 + \frac{u}{n} \right)^n.$$

Pero por definición

$$\lim_{n \rightarrow \infty} \left( 1 + \frac{u}{n} \right) = e^u.$$

Lo anterior da como resultado una situación idéntica a la que se tiene en la demostración del teorema 5.1. Esto es, conforme  $n \rightarrow \infty$ , todos los términos en  $u$ , excepto el primero, tienden hacia cero debido a que todos tienen potencias positivas de  $n$  en sus denominadores. Por lo tanto, se puede deducir que

$$\lim_{x \rightarrow \infty} m_Y(t) = e^{\frac{t^2}{2}},$$

o la distribución límite de  $Y = (\bar{X} - \mu)/(\sigma/\sqrt{n})$  es la normal estándar para valores grandes de  $n$ . ■

La esencia del teorema central del límite recae en el hecho de que para  $n$  grande, la distribución  $(\bar{X} - \mu)/(\sigma/\sqrt{n})$  es, en forma aproximada, normal con media cero y desviación estándar uno, sin importar cuál sea el modelo de probabilidad a partir del que se obtuvo la muestra. Debe notarse que si el modelo de probabilidad de la población es semejante a una distribución normal (esto es, si es simétrico y existe una concentración relativamente alta alrededor del punto de simetría), la aproximación normal será buena aun para pequeñas muestras. Por otro lado, si el modelo de la población tiene muy poco parecido a una distribución normal (por ejemplo existe una alta asimetría), la aproximación normal sólo será adecuada para valores relativamente grandes de  $n$ . En muchos casos, puede concluirse de forma segura, que la aproximación será buena mientras  $n > 30$ . Por lo tanto, la variable aleatoria

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

se emplea para formular inferencias acerca de  $\mu$  cuando se conoce el valor de la varianza población  $\sigma^2$ .

## 7.5 La distribución de muestro de $S^2$

Otra estadística importante empleada para formular inferencias con respecto a las varianzas de la población es la varianza denotada por  $S^2$ . Para comenzar, es necesario suponer que  $\mu$  es conocida y  $\sigma^2$  no. Así,  $\sigma^2$  se encuentra definida por

$$S^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{n},$$

en donde  $X_1, X_2, \dots, X_n$  constituye una muestra aleatoria de una distribución normal con media  $\mu$  y varianza  $\sigma^2$  desconocida. Para determinar una distribución de muestreo que permita hacer inferencias sobre  $\sigma^2$  con base en  $S^2$  definida por (7.13), se enuncia y demuestra el siguiente teorema.

**Teorema 7.5** Sean  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución normal con media  $\mu$  y varianza  $\sigma^2$ . La distribución de la variable aleatoria

$$Y = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2}$$

es del tipo chi-cuadrada con  $n$  grados de libertad.

**Demostración.-** Dado que  $X_i \sim N(\mu, \sigma)$ ,  $i = 1, 2, \dots, n$ ,  $Z_i = \frac{X_i - \mu}{\sigma}$  define  $n$  variables aleatorias normales estándar independientes, se tiene:

$$Y = \sum_{i=1}^n Z_i^2.$$

Del teorema 7.1,

$$\begin{aligned} m_Y(t) &= m_{Z_1^2}(t) m_{Z_2^2}(t) \dots m_{Z_n^2}(t) \\ &= (1 - 2t)^{-\frac{1}{2}} (1 - 2t)^{-\frac{1}{2}} \dots (1 - 2t)^{-\frac{1}{2}} \end{aligned}$$

dado que el cuadrado de una variable aleatoria normal estándar tiene una distribución chi-cuadrada con un grado de libertad (véase el ejemplo 5.14). De esta forma se tiene

$$m_Y(t) = (1 - 2t)^{-\frac{n}{2}}$$

que es la función generadora de momentos de una distribución chi-cuadrada con  $n$  grados de libertad. De acuerdo con lo anterior,  $Y \sim X_n^2$ . ■

Desde un punto de vista práctica, la varianza muestra tal como se encuentra definida por

$$S^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{n}$$

tiene poco uso, ya que es muy raro que se conozca el valor de la media poblacional  $\mu$ . De acuerdo con lo anterior, si se muestra una distribución normal con media  $\mu$  y varianza  $\sigma^2$ , la varianza muestral se define por

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

En el capítulo 8 se verá por qué se emplea el divisor  $(n - 1)$ . Para determinar la distribución de muestreo de  $S^2$ , y con base en una muestra aleatoria proveniente de una distribución normal, debe tomarse en cuenta el promedio de la muestra  $\bar{X}$ . Como resultado se tiene que la distribución de muestreo de  $\frac{(n - 1)S^2}{\sigma^2}$  es también una distribución chi-cuadrada con  $n - 1$  grados de libertad. Para ello probaremos

primero un teorema útil que involucra suma de dos variables aleatorias independientes chi-cuadrada y entonces se escribirá la anterior expresión en una forma equivalente, con objeto de aprovechar este teorema.

**Teorema 7.6** Si  $X_1$  y  $X_2$  son dos variables aleatorias independientes y cada una tiene una distribución chi-cuadrada con  $v_1$  y  $v_2$  grados de libertad respectivamente, entonces:

$$Y = X_1 + X_2$$

también tiene una distribución chi-cuadrada con  $v_1 + v_2$  grados de libertad.

Demostración.- Del teorema 7.1, la función generadora de momentos de  $Y$  es

$$\begin{aligned} m_Y(t) &= m_{X_1}(t)m_{X_2}(t) \\ &= (1 - 2t)^{-v_1/2}(1 - 2t)^{-v_2/2} \\ &= (1 - 2t)^{-(v_1+v_2)/2} \end{aligned}$$

que es la función generadora de momentos de una variable aleatoria chi-cuadrada con  $v_1 + v_2$  grados de libertad. ■

Ahora se deducirá la distribución de muestreo de  $\frac{(n-1)S^2}{\sigma^2}$ . De

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

se tiene que

$$(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$$

pero

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i - \mu - \bar{X} + \mu)^2 \\ &= \sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2 \\ &= \sum_{i=1}^n [(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2] \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu) + n(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2(\bar{X} - \mu)n(\bar{X} - \mu) + n(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2. \end{aligned}$$

De esta forma

$$(n-1)S^2 + n(\bar{X} - \mu)^2 = \sum_{i=1}^n (X_i - \mu)^2.$$

Al dividir ambos miembros de la expresión anterior por la varianza poblacional  $\sigma^2$  se tiene

$$\frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \left( \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$$

Del teorema 7.5 se desprende que  $\sum_{i=1}^n (X_i - \mu)^2 / \sigma^2$  tiene una distribución chi-cuadrada con  $n$  grados de libertad. De manera similar,  $[(\bar{X} - \mu) / (\sigma / \sqrt{n})]^2$  también posee una distribución chi-cuadrada con un grado de libertad, dado que  $(\bar{X} - \mu) / (\sigma / \sqrt{n})$  es  $N(0, 1)$ . Por lo tanto, si se supone que  $(n-1)S^2 / \sigma^2$  y  $[(\bar{X} - \mu) / (\sigma / \sqrt{n})]^2$  son variables aleatorias independientes, entonces por el teorema 7.6, cuando se muestrea una población cuya distribución es normal con media y varianza desconocida, la distribución de  $(n-1)S^2 / \sigma^2$ , es chi-cuadrada con  $n$ -grados de libertad. La función de densidad de probabilidad de  $Y = (n-1)S^2 / \sigma^2$  se desprende de (5.58) y está dada por:

$$f(y; n-1) = \begin{cases} \frac{1}{\Gamma[(n-1)/2] 2^{(n-1)/2}} y^{[(n-1)/2]-1} e^{-y/2} & y > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

Nótese, que dado  $Y \sim X_{n-1}^2$ ,  $E(Y) = n-1$  y  $Var(Y) = 2(n-1)$ . Además, ya que  $Y = (n-1)S^2 / \sigma^2$ ,  $S^2 = \sigma^2 Y / (n-1)$ . Por lo tanto

$$E(S^2) = E\left(\frac{\sigma^2 Y}{n-1}\right) = \frac{\sigma^2}{n-1} E(Y) = \sigma^2,$$

y

$$Var(S^2) = Var\left(\frac{\sigma^2 Y}{n-1}\right) = \frac{\sigma^4}{(n-1)^2} Var(Y) = \frac{2\sigma^4}{n-1}.$$

## 7.6 La distribución t de Student

Se recordará que cuando se muestrea una distribución normal con desviación estándar conocida  $\sigma$ , la distribución de  $Z = (\bar{X} - \mu) / (\sigma / \sqrt{n})$  es  $N(0, 1)$ . Desde un punto de vista práctico, la necesidad de conocer  $\sigma$  impide formular inferencias con respecto a  $\mu$  debido a que generalmente no se conoce el valor de la desviación estándar de la población. Dada la disponibilidad de una muestra aleatoria, el camino lógico que se sigue en este caso es reemplazar  $\sigma$  con una estimación  $s$ , que es el valor de la desviación estándar muestral  $S$ . Desafortunadamente, cuando lo anterior se lleva a cabo, la distribución  $(\bar{X} - \mu) / (S / \sqrt{n})$  no es  $N(0, 1)$ . Sin embargo es posible determinar la distribución de muestreo exacta cuando se muestra  $N(\mu, \sigma)$ , con  $\mu$  y  $\sigma^2$  desconocidos. Acá se examinará los aspectos teóricos de lo que se conoce como la distribución t de Student.

Supóngase que se realiza un experimento en que se observan dos variables aleatorias  $X$  y  $Z$ ;  $X$  tiene una distribución chi-cuadrada con  $v$  grados de libertad y  $Z$  una distribución normal con media cero y desviación estándar uno. Sea  $T$  otra variable aleatoria que es función de  $X$  y  $Z$ , de manera tal que

$$T = \frac{Z}{\sqrt{\frac{X}{v}}}.$$



Puesto que los valores de  $Z$  se encuentran entre  $(-\infty, \infty)$  y los valores de  $X$  son positivos. El valor

$$t = \frac{z}{\sqrt{\frac{x}{v}}}$$

recibe el nombre de valor de la variable aleatoria de  $t$  de Student. Lo anterior lleva al siguiente teorema.

**Teorema 7.7** Sea  $Z$  una variable aleatoria normal estándar y  $X$  una variable aleatoria chi-cuadrada con  $v$  grados de libertad. Si  $Z$  y  $X$  son independientes, entonces la variable aleatoria

$$t = \frac{z}{\sqrt{\frac{x}{v}}}$$

tiene una distribución  $t$  de Student con  $v$  grados de libertad y una función de densidad de probabilidad dada por

$$f(t, v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v} \Gamma\left(\frac{v}{2}\right)} \left[1 + \left(\frac{t^2}{v}\right)\right]^{-\frac{(v+1)}{2}}, \quad -\infty < t < \infty, v > 0.$$

Deducción.- Sea  $T$  una variable aleatoria definida por  $T = \frac{Z}{\sqrt{X/v}}$ . Considere la densidad de probabilidad de  $T$  cuando  $X$  se mantiene fija en un valor  $x$ . Dado que

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}},$$

la densidad de probabilidad condicional de

$$T = \frac{Z}{(x/v)^{1/2}}$$

se obtiene al considerar la relación inversa

$$Z = (x/v)^{1/2} T$$

y al sustituir en  $f_Z(z)$ , en donde el jacobiano de la transformación es

$$\frac{dz}{dt} = (x/v)^{1/2}.$$

De esta forma

$$f(t|x) = \frac{(x/v)^{1/2}}{\sqrt{2\pi}} e^{-xt^2/2v}, \quad -\infty < t < \infty, x > 0.$$

Se sabe que la densidad conjunta de  $T$  y  $X$  es

$$f(t, x) = f(t|x)f_X(x).$$

Dado que  $X \sim \chi_v^2$ ,

$$f_X(x) = \frac{1}{2^{v/2} \Gamma(v/2)} x^{(v-1)/2} e^{-x/2}, \quad x > 0.$$

De esta forma

$$\begin{aligned} f(t, x) &= \frac{1}{\sqrt{2\pi} 2^{v/2} \Gamma(v/2)} x^{(v-1)/2} e^{-\frac{x}{2} - \frac{xt^2}{2v}} \\ &= c_1 x^{(v-1)/2} e^{-c_2 x/2}, \end{aligned}$$

en donde  $c_1 = \frac{1}{\sqrt{2\pi v} 2^{v/2} \Gamma(v/2)}$  y  $c_2 = [1 + (t^2/v)]$ . Integrando  $f(t, x)$  con respecto a  $x$ , se obtiene la función de densidad de probabilidad de la distribución t de Student. De acuerdo con lo anterior

$$\begin{aligned}
 f_T(t) &= c_1 \int_0^\infty x^{(v-1)/2} e^{-c_2/2} dx \\
 &= c_1 \int_0^\infty (2y/c_2)^{(v-1)/2} e^{-y} (2/c_2) dy, \text{ donde } y = c_2 x/2 \text{ y } dx = (2/c_2) dy \\
 &= c_1 (2/c_2)^{(v+1)/2} \int_0^\infty y^{(v-1)/2} e^{-y} dy \\
 &= c_1 (2/c_2)^{(v+1)/2} \Gamma[(v+1)/2] \\
 &= \frac{1}{\sqrt{2\pi v} 2^{v/2} \Gamma(v/2)} \cdot \frac{2^{(v+1)/2}}{[1 + (t^2/v)]^{(v+1)/2}} \Gamma[(v+1)/2] \\
 &= \frac{\Gamma[(v+1)/2]}{\sqrt{\pi v} \Gamma[v/2]} \left[ 1 + \left( \frac{t^2}{v} \right) \right]^{-\frac{(v+1)}{2}} \quad -\infty < t < \infty.
 \end{aligned}$$

■

Conforme se tiene un número mayor de grados de libertad, la distribución t de Student tiende hacia la normal estándar.

Puede demostrarse que el valor esperado de  $T$  es

$$E(t) = 0, \quad v > 1$$

y la varianza está dada por

$$Var(T) = \frac{v}{v-2}, \quad v > 2.$$

En la tabla  $F$  se encuentra los valores cuantiles  $t_{1-\alpha, 2}$  tales que:

$$P(T \leq t_{1-\alpha, v}) = \int_{-\infty}^{t_{1-\alpha, v}} f(t; v) dt = 1 - \alpha, \quad 0 \leq \alpha \leq 1.$$

para los distintos valores de  $v$  y de las proporciones acumulativas seleccionadas  $1 - \alpha$ . Por ejemplo, si  $v = 15$ ,

$$\begin{aligned}
 P(T \leq t_{0.90, 15}) &= P(T \leq 1.341) = 0.90, \\
 P(T \leq t_{0.95, 15}) &= P(T \leq 1.753) = 0.95, \\
 P(T \leq t_{0.99, 15}) &= P(T \leq 2.602) = 0.99.
 \end{aligned}$$

Dado que la distribución t es simétrica con respecto al cero, para  $\alpha > 0.5$  los valores cuantiles  $t_{1-\alpha, v}$  serán negativos pero sus magnitudes serán las mismas que las de los correspondientes valores que se encuentran en el lado derecho. De esta forma, para  $v = 15$ ,

$$\begin{aligned}
 P(T \geq t_{0.10, 15}) &= P(T \leq -1.341) = 0.10, \\
 P(T \geq t_{0.05, 15}) &= P(T \leq -1.753) = 0.05, \\
 P(T \geq t_{0.01, 15}) &= P(T \leq -2.602) = 0.01.
 \end{aligned}$$

Desde un punto de vista práctico, es muy poca la ganancia que se tiene al emplear la distribución t de Student en lugar de la normal estándar cuando  $v \geq 30$ .

Recuérdese que para formular inferencias con respecto a  $\mu$  cuando el muestreo se lleva a cabo sobre una distribución normal con media y varianza desconocida, se necesita determinar la distribución  $(\bar{X} -$

$\mu)(S/\sqrt{n})$ . Cuando se muestrea una distribución  $N(\mu, \sigma)$  se sabe, del teorema 7.3, que la distribución de  $(\bar{X} - \mu)/(\sigma/\sqrt{n})$  es  $N(0, 1)$ . Para la misma condición, se sabe que, de

$$\frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \left( \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$$

y del teorema 7.6, la distribución de  $(n-1)S^2/\sigma^2$  es chi-cuadrada con  $n-1$  grados de libertad. Dado que puede demostrarse que  $\bar{X}$  y  $S^2$  son independientes, del teorema 7.7 se desprende que la distribución de

$$\frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \frac{1}{n-1}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \cdot \frac{\sigma}{\sqrt{S^2}}, \quad \text{o} \quad T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}},$$

es la  $t$  de Student con  $n-1$  grados de libertad.

## 7.7 La distribución de la diferencia entre dos medias muestrales

Se necesita obtener la distribución  $\bar{X}$  e  $\bar{Y}$  cuando el muestreo se lleva a cabo sobre dos poblaciones normales independientes con varianzas iguales. Si se supone que el valor de la varianza  $\sigma^2$  se conoce del teorema 7.3, se sabe que la distribución de  $\bar{X}$  es normal con media  $\mu_X$  y varianza  $\frac{\sigma^2}{n_X}$ . La distribución de  $\bar{Y}$  también es normal pero con media  $\mu_Y$  y varianza  $\frac{\sigma^2}{n_Y}$ . Dado que  $\bar{X}$  y  $\bar{Y}$  son variables aleatorias independientes normalmente distribuidas, si  $a_1 = 1$  y  $a_2 = -1$  en el teorema 7.2, la distribución de  $\bar{X} - \bar{Y}$  también es normal con media  $\mu_X - \mu_Y$  y varianza  $\left(\frac{\sigma^2}{n}\right) + \left(\frac{\sigma^2}{n_Y}\right) = \sigma^2 \left(\frac{1}{n_X} + \frac{1}{n_Y}\right)$ . Por tanto, si se conoce el valor de  $\sigma^2$ , la distribución de

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$$

es  $N(0, 1)$  con varianzas iguales.

En este desarrollo se supone que el valor de  $\sigma^2$  era conocido. Sin embargo, es poco probable conocer el valor de  $\sigma^2$  para una situación real. Así pues, debe obtenerse la distribución de  $\bar{X} - \bar{Y}$  cuando el muestreo se lleve a cabo sobre dos poblaciones normales independientes con varianzas iguales pero desconocidas. Para cada una de las dos muestras aleatorias, pueden definirse las varianzas muestrales  $S_X^2$  y  $S_Y^2$  dadas por  $S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$ . Dado que  $\frac{(n_X-1)S_X^2}{\sigma^2}$  y  $\frac{(n_Y-1)S_Y^2}{\sigma^2}$  son dos variables independientes chi-cuadrada, con  $n_X-1$  y  $n_Y-1$  grados de libertad respectivamente, por el teorema 7.6, la distribución de

$$W = \frac{(n_X-1)S_X^2}{\sigma^2} + \frac{(n_Y-1)S_Y^2}{\sigma^2}$$

también es chi-cuadrada con  $n_X + n_Y - 2$  grados de libertad. De la expresión  $T = \frac{Z}{\sqrt{X/v}}$  se desprende el hecho de que el cociente de Z en

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$$

y la raíz cuadrada de W dividida entre sus grados de libertad tiene una distribución t de Student con  $n_X + n_Y - 2$  grados de libertad. Esto es,

$$\frac{\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma} \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}{\sqrt{\frac{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}{\sigma^2} \frac{1}{n_X + n_Y - 2}}} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}{n_X + n_Y - 2} \left( \frac{1}{n_X} + \frac{1}{n_Y} \right)}}$$

o

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}},$$

en donde

$$S_p^2 = \frac{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}{n_X + n_Y - 2}.$$

Que en general recibe el nombre de **estimador combinado de la varianza común  $\sigma^2$** .

Si las varianzas  $\sigma_X^2$  y  $\sigma_Y^2$  no son iguales, pero conocen sus valores, el problema es sencillo. La distribución de

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$$

aún es  $N(0, 1)$ .

## 7.8 La distribución F

Recuérdese que las inferencias con respecto a  $\sigma^2$  cuando se muestrea una distribución normal, se formulan con base en la estadística  $\frac{(n-1)S^2}{\sigma^2}$ , la que tiene una distribución chi-cuadrada con  $n - 1$  grados de libertad. En este apartado se formulará inferencias con respecto a las varianzas de dos distribuciones normales independientes con base en las muestras aleatorias de cada uno.

Supóngase un experimento en que se observan dos variables aleatorias independientes X e Y, cada una con una distribución chi-cuadrada con  $v_1$  y  $v_2$  grados de libertad respectivamente. Sea F una variable aleatoria que es función de X e Y, de manera tal que

$$F = \frac{\frac{X}{v_1}}{\frac{Y}{v_2}}.$$

Esto lleva al siguiente teorema

**Teorema 7.8** Sean  $X$  e  $Y$  dos variables aleatorias independientes chi-cuadrada con  $v_1$  y  $v_2$  grados de libertad, respectivamente. La variable aleatoria

$$F = \frac{\frac{X}{v_1}}{\frac{Y}{v_2}}.$$

tiene una distribución  $F$  con una función de densidad de probabilidad dada por

$$g(f; v_1, v_2) = \begin{cases} \frac{\Gamma\left[\frac{v_1 + v_2}{2}\right] v_1^{\frac{v_1}{2}} v_2^{\frac{v_2}{2}}}{\Gamma\left(\frac{v_1}{2}\right) \Gamma\left(\frac{v_2}{2}\right)} f^{\frac{v_1 + v_2}{2}} & f > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

■

El **valor esperado** es

$$E(F) = \frac{v_2}{v_2 - 2}, \quad v_2 > 2.$$

y la **varianza** está dada por

$$\text{Var}(F) = \frac{v_2^2 (2v_2 + 2v_1 - 4)}{v_1 (v_2 - 2)^2 (v_2 - 4)} \quad v_2 > 4.$$

La distribución  $F$  tiene asimetría postiva para cualesquiera valores de  $v_1$  y  $v_2$ , pero esta va disminuyendo conforme  $v_1$  y  $v_2$  toman valores cada vez más grandes.

Los valores **cuantiles**  $f_{1-\alpha, v_1, v_2}$  están dados dados por

$$P(F \leq f_{1-\alpha, v_1, v_2}) = \int_0^{f_{1-\alpha, v_1, v_2}} g(f; v_1, v_2) df = 1 - \alpha, \quad 0 \leq \alpha \leq 1.$$

Regresando anuestro objetivo. Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de variables independientes y normalmente distribuidas cada una con media  $\mu_X$  y varianza  $\sigma_X^2$ . También sea  $Y_1, Y_2, \dots, Y_n$  un conjunto de  $n_Y$  variables aleatorias independientes normalmente distribuidas, cada una con media  $\mu_Y$  y varianza  $\sigma_Y^2$ . Si se supone que las  $X$  y las  $Y$  son independientes, las estadísticas

$$\frac{(n_X - 1)S_X^2}{\sigma_X^2}$$

y

$$\frac{(n_Y - 1)S_Y^2}{\sigma_Y^2}$$

son dos variables aleatorias chi-cuadrada independientes con  $n_X - 1$  y  $n_Y - 1$  grados de libertad respectivamente. Entonces, por el teorema 7.9, se desprende que la variable aleatoria

$$\frac{\frac{(n_X - 1)S_X^2}{\sigma_X^2}}{\frac{(n_Y - 1)S_Y^2}{\sigma_Y^2}} = \frac{\frac{S_X^2}{\sigma_X^2}}{\frac{S_Y^2}{\sigma_Y^2}}.$$

Si la suposición de que  $\sigma_X^2 = \sigma_Y^2$  es correcta, la estadística  $F$ , se deduce a

$$F = \frac{S_X^2}{S_Y^2}.$$

Si las dos varianzas son iguales, la probabilidad de observar un valor de  $F$  distinto, de manera suficientes, es pequeña.

tiene una distribución  $F$  con  $n_X - 1$  y  $n_Y - 1$  grados de libertad.

Para finalizar debe notarse que en esta sección, así como en las secciones 7.5 y 7.7 se desarrolló el material que se presentó bajo la hipótesis de realizar un muestreo aleatorio sobre poblaciones que tienen una distribución normal.

## Estimación puntual y por intervalo