

Probabilidad y Estadística

George Canavos

Apuntes por FODE
(Christian L. Paredes Aguilera.)

Contents

1	Introducción y estadística descriptiva	5
1.3	Medidas numéricas descriptivas	5
2	Conceptos en probabilidad	8
2.5	Desarrollo axiomático de la probabilidad	8
2.6	Probabilidad conjunta, marginal y condicional	11
2.7	Eventos estadísticamente independientes	11
2.8	El teorema de Bayes	11
2.9	Ejercicios	12
3	Variables aleatorias y distribución de probabilidad	13
3.1	El concepto de variables aleatorias	13
3.2	Distribuciones de probabilidad de variables aleatorias discretas	14
3.3	Distribuciones de probabilidad de variables aleatorias continuas	14
3.4	Valor esperado de una variable aleatoria	15
3.4.1	Propiedades	15
3.5	Momentos de una variable aleatoria	16
3.6	Otras medidas de tendencia central y dispersión	19
3.7	Funciones generadoras de momentos	20
3.8	Ejercicios	22
4	Algunas distribuciones discretas de probabilidad	23
4.2	La distribución binomial	23
4.3	La distribución de Poisson	27
4.4	La distribución hipergeométrica	32
4.5	La distribución binomial negativa	33
5	Algunas distribuciones continuas de probabilidad	35
5.1	La distribución normal	35
5.2	La distribución uniforme	40
5.3	La distribución beta	42
5.4	La distribución gama	44
5.5	La distribución de Weibull	47
5.6	La distribución exponencial negativa	49
5.7	La distribución de una función de variable aleatoria	51
5.8	Conceptos básicos en la generación de números aleatorios por computadora	52
5.8.1	Distribución uniforme sobre el intervalo (a,b)	52
5.8.2	La distribución de Weibull	52
5.8.3	La distribución de Erlang	53
5.8.4	La distribución normal	53
5.8.5	La distribución binomial	53
5.8.6	La distribución Poisson	54

6	Distribuciones conjuntas de probabilidad	55
6.1	Distribución de probabilidad bivariadas	55
6.2	Distribuciones marginales de probabilidad	56
6.3	Valores esperados y momentos para distribuciones bivariadas	57
6.4	Variables aleatorias estadísticamente independientes	59
6.5	Distribuciones de probabilidad condicional	60
6.6	Análisis bayesiano: las distribuciones a priori y a posteriori.	62
6.7	La distribución normal bivariada	62
7	Muestras aleatorias y distribuciones de muestreo	64
7.3	Distribuciones de muestreo de estadísticas	65
7.4	La distribución de muestreo de \bar{X}	67
7.5	La distribución de muestreo de S^2	71
7.6	La distribución t de Student	73
7.7	La distribución de la diferencia entre dos medias muestrales	76
7.8	La distribución F	77
8	Estimación puntual y por intervalo	80
8.2	Propiedades deseables de los estimadores puntuales	80
8.2.1	Estimadores insesgados	81
8.2.2	Estimadores consistentes	82
8.2.3	Estimadores insesgados de varianza mínima	83
8.2.4	Estadísticas suficientes	84
8.3	Métodos de estimación puntual	84
8.3.1	Método por máxima verosimilitud	84
8.3.2	Método de los momentos	86
8.3.3	Estimación por máxima verosimilitud para muestras censuradas	87
8.4	Estimación por intervalo	88
8.4.1	Intervalos de confianza para μ cuando se muestrea una distribución normal con varianza conocida	89
8.4.2	Intervalos de confianza para μ cuando se muestrea una distribución normal con varianza desconocida	90
8.4.3	Intervalo de confianza para la diferencia de medias cuando se muestrean dos distribuciones normales independientes	91
8.4.4	Intervalo de confianza para σ^2 cuando se muestrean dos distribución normal con media desconocida	92
8.4.5	Intervalos de confianza para el cociente de dos varianzas cuando se muestrean dos distribuciones normales independientes	92
8.4.6	Intervalos de confianza para el parámetro de proporción p cuando se muestrea una distribución binomial	93
8.5	Estimación bayesiana	94
8.5.1	Estimación puntual bayesiana	94
8.5.2	Estimación bayesiana por intervalo	94
8.6	Límites estadísticos de tolerancia	95
9	Prueba de hipótesis estadísticas	96
9.1	Introducción	96
9.2	Conceptos básicos para la prueba de hipótesis estadísticas	96
9.3	Tipos de regiones críticas y la función de potencia	98
9.4	Las mejores pruebas	98
9.5	Principios generales para probar una H_0 simple contra una H_1 uni o bilateral	99
9.5.1	Principios generales para el caso 1	99
9.5.2	Principios generales para el caso 2	99
9.5.3	Principios generales para el caso 3	100
9.6	Prueba de hipótesis con respecto a las medias cuando se muestrean distribuciones normales	100
9.6.1	Pruebas para una muestra	100

9.6.2	Pruebas para dos muestras	102
9.6.3	Reflexiones sobre las suposiciones y sensibilidad	104
9.6.4	Prueba sobre las medias cuando las observaciones están pareadas	104
9.7	Prueba de hipótesis con respecto a las varianzas cuando se muestrean distribuciones normales	104
9.7.1	Pruebas para una muestra	105
9.7.2	Pruebas para dos muestras	105
9.8	Inferencias con respecto a las proporciones de dos distribuciones binomiales independientes	106
10	Pruebas de bondad de ajuste y análisis de tablas de contingencia	107
10.1	La prueba de bondad de ajuste chi-cuadrada	107
10.2	La prueba de Kolmogorov-Smirnov	108

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.

Si el número de observaciones en el conjunto es impar, la mediana es el valor de la observación que se encuentra a la mitad del conjunto ordenando. Si el número es par se considera la mediana como el promedio aritmético de los valores de las dos observaciones que se encuentran a la mitad del conjunto ordenado. La mediana es el percentil cincuenta.

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

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

ó

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

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 μ ; 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 σ^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 μ y desviación estándar σ 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, μ'_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 μ_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 μ'_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 λ , 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 θ , 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 θ se genera mediante el empleo de los límites $(0, 2\pi)$ para θ 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 σ .

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 σ** . 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 μ y σ 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 μ y desviación estándar σ .

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 α como β 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 β .

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 α 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 α .

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 α 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 α 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 α 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 α -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 α 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 α con $v/2$ y el parámetro de escala θ 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:

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

La distribución Weibull puede aproximarse, de manera adecuada, por una distribución normal cada que el factor de forma α 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 θ 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 θ , 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 univaluadas 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 univaluadas, 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 θ 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 α es un entero positivo. Recuérdese que la variable aleatoria de Erlang es la suma de α variables aleatorias independientes distribuidas exponencialmente. Por lo tanto, un número aleatorio de Erlang es la suma de α 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 ρ 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 6.1 Sean X_1, X_2, \dots, X_n n variables aleatorias con una función de densidad conjunta de probabilidad $f(x_1, x_2, \dots, x_n)$. 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

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

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 ρ 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 μ_X y varianza σ_X^2 . De manera similar, la densidad marginal de Y es normal con media μ_Y y varianza σ_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 θ .

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 θ , 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 θ 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 θ , entonces T recibe el nombre de **estimador** de θ , y el valor específico de t como un resultado de los datos muestrales recibe el nombre de **estimación** de θ . 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 generadoras 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

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

■

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

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

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 μ y σ^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 μ y varianza σ^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 μ y varianza σ^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.- Recuérdese 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 θ/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 α con n y θ con θ/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 θ y θ^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 Γ^m , muestrear una distribución exponencial con parámetro θ , \bar{X} tiene una distribución gama con media θ , 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 μ y varianza σ^2 finita. El promedio muestral $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$ tiene una distribución con media μ y varianza σ^2/n que tiende hacia una distribución normal conforme n tiende a ∞ . 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 μ cuando se conoce el valor de la varianza población σ^2 .

7.5 La distribución de muestreo 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 μ es conocida y σ^2 no. Así, σ^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 μ y varianza σ^2 desconocida. Para determinar una distribución de muestreo que permita hacer inferencias sobre σ^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 μ y varianza σ^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 μ . De acuerdo con lo anterior, si se muestra una distribución normal con media μ y varianza σ^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

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i - \mu - \bar{X} + \mu)^2 \\ &= \sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2 \\ &= \sum_{i=1}^n \left[(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2 \right] \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu) + n(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2(\bar{X} - \mu) n(\bar{X} - \mu) + n(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2. \end{aligned}$$

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 σ^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 σ , la distribución de $Z = (\bar{X} - \mu) / (\sigma / \sqrt{n})$ es $N(0, 1)$. Desde un punto de vista práctico, la necesidad de conocer σ impide formular inferencias con respecto a μ 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 σ 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 μ y σ^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

$$f(t|x) = \frac{(x/v)^{1/2}}{\sqrt{2\pi}} e^{-xt^2/2v}, \quad -\infty < t < \infty, \quad x > 0.$$

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 μ 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 σ^2 se conoce del teorema 7.3, se sabe que la distribución de \bar{X} es normal con media μ_X y varianza $\frac{\sigma^2}{n_X}$. La distribución de \bar{Y} también es normal pero con media μ_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 σ^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 σ^2 era conocido. Sin embargo, es poco probable conocer el valor de σ^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,

$$\frac{\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma} \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}{\sqrt{\frac{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}{\sigma^2}} \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{(n_X - 1) S_X^2 + (n_Y - 1) S_Y^2}{n_X + n_Y - 2} \left(\frac{1}{n_X} + \frac{1}{n_Y} \right)}}$$

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 σ^2** .

Si las varianzas σ_X^2 y σ_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 σ^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 μ_X y varianza σ_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 μ_Y y varianza σ_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

8.2 Propiedades deseables de los estimadores puntuales

Se conoce la familia de distribuciones a partir de la cual se obtiene la muestra, pero no puede identificarse el miembro de interés de esta, ya que no se conoce el valor del parámetro. Este último tiene que estimarse con base en los datos de la muestra.

El estimador de un parámetro θ debe tener una distribución de muestro concentrada alrededor de θ y la varianza del estimador debe ser la menor posible. Para ampliar las propiedades anteriores, considérese la siguiente. Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n proveniente de una distribución con función de densidad $f(x, \theta)$, y sea $T = u(X_1, X_2, \dots, X_n)$ cualquier estadística. El problema es encontrar una función u que sea la que proporcione la mejor estimación de θ . Al buscar el mejor estimador de θ se hará uso de una cantidad muy importante que recibe el nombre de error cuadrático medio de un estimador.

Definición 8.1 Sea T cualquier estimador de un parámetro desconocido θ . Se define el error cuadrático medio de T como el valor esperado del cuadrado de la diferencia entre T y θ .

Para cualquier estadística T , se denotará el error cuadrático medio por $ECM(T)$; de esta forma

$$ECM(T) = E(T - \theta)^2.$$

Puede verse la razón del porque el error cuadrático medio es una cantidad importante para enjuiciar a los posibles estimadores de θ mediante el desarrollo de $ECM(T) = E(T - \theta)^2$; este es,

$$\begin{aligned} ECM(T) &= E(T^2 - 2\theta T + \theta^2) \\ &= E(T^2) - 2\theta E(T) + \theta^2 \\ &= \text{Var}(T) + E^2(T) - 2\theta E(T) + \theta^2 \\ &= \text{Var}(T) + [\theta - E(T)]^2. \end{aligned}$$

El error cuadrático medio de cualquier estimador es la suma de dos cantidades no negativas: una es la varianza del estimador y la otra es el cuadrado del sesgo del estimador. Estas dos cantidades se encuentran relacionadas en forma directa con las propiedades deseables de un estimador. La varianza de un estimador debe ser lo más pequeña posible mientras que la distribución de muestreo debe concentrarse del valor del parámetro. Por lo que debemos seleccionar el mejor estimador de θ ; es decir, la estadística que tenga el error cuadrático medio más pequeño posible entre todos los estimadores factibles de θ .

8.2.1 Estimadores insesgados

El término $\theta - E(T)$ recibe el nombre de sesgo del estimador.

Definición 8.2 Se dice que la estadística $T = u(X_1, X_2, \dots, X_n)$ es un estimador insesgado del parámetro θ , si $E(T) = \theta$ para todos los posibles valores de θ . De esta forma, para cualquier estimador insesgado de θ , la distribución de muestreo de T se encuentra centrada alrededor de θ y $ECM(T) = Var(T)$.

En la sección 7.4, demostramos que sin importar la distribución de la población de interés, $E(\bar{X}) = \mu$. Por lo tanto, la media muestral es un estimador insesgado de la media de la población μ para todo los valores de μ . De hecho, si X_1, X_2, \dots, X_n es una muestra aleatoria de la distribución de X con media μ , entonces cualquier X_i de la muestra es un estimador insesgado de μ , dado que $E(X_i) = \mu$ para todo $i = 1, 2, \dots, n$. Además si una estadística T es cualquier combinación lineal de las variables aleatorias de la muestra de manera tal que

$$T = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$

en donde $\sum_{i=1}^n a_i = 1$, entonces T es un estimador insesgado de μ dado que

$$\begin{aligned} E(T) &= \sum_{i=1}^n a_i E(X_i) \\ &= \sum_{i=1}^n a_i \mu \\ &= \mu. \end{aligned}$$

En la sección 7.5 se demostró que si la varianza muestral S^2 está dada por $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n - 1)$. Entonces, cuando se muestrea una distribución normal, $E(S^2) = \sigma^2$. Lo que demostraremos que es insesgado sin importar cuál sea la distribución de la población de interés. Sea X_1, X_2, \dots, X_n una muestra aleatoria de alguna distribución con una función de densidad no específica. De esta manera, $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ para toda $i = 1, 2, \dots, n$. Entonces,

$$\begin{aligned} E(S^2) &= E \left[\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n - 1} \right] \\ &= (n - 1)^{-1} E \left\{ \sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2 \right\} \\ &= (n - 1)^{-1} E \left\{ \sum_{i=1}^n [(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2] \right\} \\ &= (n - 1)^{-1} \left[\sum_{i=1}^n E(X_i - \mu)^2 - 2n E(\bar{X} - \mu)^2 \right] \end{aligned}$$

pero por definición $E(X_i - \mu)^2 = Var(X_i) = \sigma^2$ y $E(\bar{X} - \mu)^2 = Var(\bar{X}) = \sigma^2 / n$. Por lo tanto,

$$\begin{aligned} E(S^2) &= (n - 1)^{-1} \left[\frac{n\sigma^2 - (n\sigma^2)}{n} \right] \\ &= \frac{\sigma^2(n - 1)}{n - 1} \\ &= \sigma^2. \end{aligned}$$

En otras palabras, S^2 es un estimador insesgado de σ^2 sólo cuando el divisor es igual a $n - 1$. Es la razón por la que se divide por $n - 1$ en lugar de n . S no hará un estimador insesgado de σ .

8.2.2 Estimadores consistentes

Es razonable esperar que un buen estimador de un parámetro θ sea cada vez menor conforme crece el tamaño de la muestra. Se tendrá un mejor estimador de θ si se basa en 30 observaciones que si lo hace con sólo cinco.

Definición 8.3 Sea T el estimador de un parámetro θ , y sea T_1, T_2, \dots, T_n una secuencia de estimadores que representan a T con base en muestras de tamaño $1, 2, \dots, n$, respectivamente. Se dice que T es un estimador consistente (sencillo) para θ si

$$\lim_{n \rightarrow \infty} P(|T_n - \theta| \leq \epsilon) = 1$$

para todos los valores de θ y $\epsilon > 0$. El requisito constituye lo que se denomina convergencia en probabilidad. Es decir, si un estimador es consistente, converge en probabilidad al valor del parámetro que está intentando estimar conforme el tamaño de la muestra crece. Esto implica que la varianza de un estimador consistente T_n disminuye conforme n crece, y la media de T_n tiende hacia donde n crece. Para demostrar que \bar{X} es un estimador consistente de μ , primero se enuncia el teorema de desigualdad de Tchebysheff.

Teorema 8.1 Sea X una variable aleatoria con una función (densidad) de probabilidad $f(x)$ de manera tal que tanto $E(X) = \mu$ como $Var(X) = \sigma^2$ tienen un valor finito. Entonces

$$P(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}$$

o

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

para cualquier constante $k \geq 1$. ■

Este teorema permite determinar los límites de las probabilidades de variables aleatorias discretas y continuas sin tener que especificar sus funciones (densidades) de probabilidad. Asegura que la probabilidad de que una variable aleatoria se aleje no más de k desviaciones estándar de la media, es mayor o igual a $1/k^2$ para algún valor de $k \geq 1$. Por ejemplo

$$P(|X - \mu| \leq 2\sigma) \geq 1 - \frac{1}{4}$$

y

$$P(|X - \mu| \leq 3\sigma) \geq 1 - \frac{1}{9}.$$

para cualquier variable aleatoria X con media μ y varianza σ^2 finitas.

Teorema 8.2 Sean X_1, X_2, \dots, X_n , n variables aleatorias IID, tales que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ tienen un valor finito para $i = 1, 2, \dots, n$. Entonces, $\bar{X}_n = \sum_{i=1}^n X_i / n$ es un estimador consistente de μ .

Demostración.- Se quiere demostrar que

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \leq \epsilon) = 1.$$

Dado que \bar{X}_n es una variable aleatoria tal que $E(\bar{X}_n) = \mu$ y $Var(\bar{X}_n) = \sigma^2/n$, se deduce del teorema de Tchebysheff que

$$P(|\bar{X}_n - \mu| > k\sigma/\sqrt{n}) \leq 1/k^2.$$

Sea k una constante positiva igual a $\epsilon\sqrt{n}/\sigma$, en donde ϵ es un número real positivo. Entonces,

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Dado que σ^2 tiene un valor finito, tomando el límite de esta expresión conforme n tiende al infinito se tiene

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \epsilon) = 0.$$

Por lo tanto, se concluye que

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \leq \epsilon) = 1.$$

y \bar{X}_n es un estimador consistente de μ . ■

El teorema 8.2 también se conoce como la ley de los grandes números.

8.2.3 Estimadores insesgados de varianza mínima

Para un parámetro que posee un error cuadrático medio mínimo es difícil determinar un estimador para todo los posibles valores del parámetro. Pero podremos analizar uno que tenga un error cuadrático medio mínimo. Si se tiene una varianza mínima para todos los valores posibles de θ . Este estimador recibe el nombre de estimador insesgado de varianza mínima uniforme (VMU)

Definición 8.4 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución cuya función (densidad) de probabilidad es $f(x; \theta)$. Sea la estadística $T = u(X_1, X_2, \dots, X_n)$ un estimador de θ tal que $E(T) = \theta$ y $Var(T)$ es menor que la varianza de cualquier otro estimador insesgado de θ para todos los posibles valores de θ . Se dice entonces que T es un estimador insesgado de varianza mínima de θ .

La varianza de un estimador insesgado es la cantidad más importante para decidir que tan bueno es el estimador para estimar un parámetro θ . Por ejemplo, sean T_1 y T_2 cualesquiera dos estimadores insesgados de θ . Se dice que T_1 es un estimador más eficiente de θ que T_2 si $Var(T_1) \leq Var(T_2)$. Es muy común utilizar el cociente $Var(T_1)/Var(T_2)$ para determinar la eficiencia relativa de T_2 respecto de T_1 . Si los estimadores son sesgados, se emplea sus errores cuadráticos medios para determinar las eficiencias relativas.

Teorema 8.3 Cota inferior de Cramér-Rao. Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución con una función (densidad) de probabilidad $f(x; \theta)$. Si T es un estimador insesgado de θ , entonces la varianza de T debe satisfacer la siguiente desigualdad

$$Var(T) \geq \frac{1}{n E \left[\left(\frac{\partial \ln f(X; \theta)}{\partial \theta} \right)^2 \right]}.$$

El teorema 8.3 establece un límite inferior para la varianza de un estimador de θ . Para un estimador insesgado cuya varianza se apeg a la cota inferior de Cramér-Rao, se tiene la siguiente definición. ■

Definición 8.5 Si T es cualquier estimador insesgado del parámetro θ tal que

$$\text{Var}(T) \geq \frac{1}{n E \left[\left(\frac{\partial \ln f(X; \theta)}{\partial \theta} \right)^2 \right]},$$

entonces se dice que T es un estimador eficiente de θ .

8.2.4 Estadísticas suficientes

Una estadística suficiente para un parámetro θ es aquella que utiliza toda la información contenida en la muestra aleatoria con respecto a θ . La utilidad de una estadística suficiente recae en el hecho de que si un estimador insesgado de un parámetro θ es una función de una estadística suficiente, entonces tendrá la varianza más pequeña de entre todos los estimadores insesgados de θ que no se encuentren basados en una estadística suficiente. De hecho, si existe el estimador eficiente de θ , se encontrará que este es una estadística suficiente. Un criterio para terminar una estadística suficiente está dado por el siguiente teorema, el cual se conoce como teorema de factorización de Neyman.

Teorema 8.4 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución con una función de densidad de probabilidad $f(x; \theta)$. Se dice que la estadística $T = u(X_1, X_2, \dots, X_n)$ es una estadística suficiente para θ si y sólo si la función de verosimilitud puede factorizarse de la siguiente manera:

$$L(x_1, x_2, \dots, x_n; \theta) = h(t; \theta) g(x_1, x_2, \dots, x_n)$$

para cualquier valor $t = u(x_1, x_2, \dots, x_n)$ de T y en donde $g(x_1, x_2, \dots, x_n)$ no contiene al parámetro θ . ■

8.3 Métodos de estimación puntual

Se estudiará cómo obtener estimadores que, de manera general, tenga buenas propiedades. Específicamente se considerarán los métodos de máxima verosimilitud y el de momentos.

8.3.1 Método por máxima verosimilitud

El método de estimación por máxima verosimilitud, selecciona como estimador a aquél valor del parámetro que tiene la propiedad de maximizar el valor de la probabilidad de la muestra aleatoria observada. En otras palabras, el método de máxima verosimilitud consiste en encontrar el valor del parámetro que maximiza la función de verosimilitud.

Definición 8.6 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución con función (densidad) de probabilidad $f(x; \theta)$, y sea $L(x_1, x_2, \dots, x_n; \theta)$ la verosimilitud de la muestra como función de θ . Si $t = u(x_1, x_2, \dots, x_n)$ es el valor de θ para el cual el valor de la función de verosimilitud es máxima, entonces $T = u(X_1, X_2, \dots, X_n)$ es el estimador de máxima verosimilitud de θ , y t es el estimador de máxima verosimilitud.

Tiene la propiedad deseable de proporcionar estimadores que son funciones de estadísticas suficientes, siempre y cuando el estimador MV sea único. Además, el método MV proporciona el estimador eficiente, si es que existe. Sin embargo los estimadores MV son generalmente sesgados. El procedimiento para obtener este tipo de estimadores es relativamente directo. Debido a la naturaleza de la función

de verosimilitud se escoge, por lo común, maximizar el logaritmo natural de $L(\theta)$. Esto es, en muchas ocasiones más fácil obtener el estimador MV maximizando $\ln L(\theta)$ que $L(\theta)$.

Ejemplo 8.6 En un experimento binomial se observan $X = x$ éxitos en n ensayos. Obtener el estimador de máxima verosimilitud del parámetro binomial p .

Respuesta.- En este caso la función de verosimilitud es idéntica a la probabilidad de que $X = x$; de esta forma

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

Entonces,

$$\ln L(x; p) = \ln(n!) - \ln[(n-x)!] - \ln(x!) + (n-x) \ln(1-p).$$

Para encontrar el valor de p , para el cual $\ln L(x; p)$ tiene un valor máximo, se toma la primera derivada con respecto a p y se iguala a cero:

$$\frac{\partial [\ln L(x; p)]}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = 0.$$

Después de resolver para p , se obtiene el estimador MV de p el cual recibe el nombre de proporción muestral X/n , y el estimador MV es $p = x/n$. Para confirmar que este valor maximiza a $\ln L(x; p)$, se toma la segunda derivada con respecto a p y se evalúa en x/n :

$$\frac{\partial^2 [\ln L(x; p)]}{\partial p^2} = -\frac{np(1-p) + (x-np)(1-2p)}{[p(1-p)]^2}$$

y

$$\left. \frac{\partial^2 [\ln L(x; p)]}{\partial p^2} \right|_{x/n} = -\frac{x}{(x/n)^2 [1 - (x/n)]} \quad x/n < 1,$$

lo que confirma el resultado, dado que la segunda derivada es negativa. Para un ejemplo específico, si se observan $x = 5$ con base en 25 ensayos independientes, el estimador MC de p es 0.2. ■

Ejemplo 8.7 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución normal con una función de densidad de probabilidad

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Determinar los estimadores de μ y σ^2 .

Respuesta.- Para este problema se procederá de la misma forma que en el caso de un solo parámetro. Dado que la función de verosimilitud depende tanto de μ como de σ^2 , los estimadores MV de μ y σ^2 son los valores para los cuales la función de verosimilitud tiene un valor máximo. De acuerdo con lo anterior

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \dots e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}, \end{aligned}$$

y

$$\ln [L(x_1, x_2, \dots, x_n; \mu, \sigma^2)] = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Después de obtener las primeras derivadas parciales con respecto a μ y con respecto a σ^2 e igualándolas a cero, se tiene

$$\frac{\partial [\ln L(\mu, \sigma^2)]}{\partial \mu} = -\frac{2}{2\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0$$

y

$$\frac{\partial [\ln L(\mu, \sigma^2)]}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0.$$

Resolviendo la primera ecuación para μ , substituyendo este valor en la segunda y resolviendo para σ^2 , se tiene

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

y

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

■

Si tomamos la segunda derivada notaremos que estos valores maximizan la función de verosimilitud, ellos son los estimadores MV de μ y σ^2 .

Notemos que se ha introducido la acostumbrada notación de sombrero, $\hat{\cdot}$, para denotar un estimador MV. También notemos que el estimador MV de σ^2 es sesgado, confirmándose de esta manera un comentario anterior en el sentido en el que los estimadores MV no necesariamente son insesgados.

El método de máxima verosimilitud posee otra propiedad deseable conocida como propiedad de invarianza. Sea $\hat{\theta} = u(X_1, X_2, \dots, X_n)$ el estimador de máxima verosimilitud de θ . Si $g(\theta)$ es una función univaluada de θ . Por ejemplo, dado que, cuando se muestrea una distribución normal, el estimador MV de σ^2 es

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

por la propiedad de invarianza, el estimador MV de la desviación estándar σ es

$$\hat{\sigma} = \left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{\frac{1}{2}}.$$

Como ejemplo adicional de la propiedad de invarianza, el estimador MV de la función de confiabilidad Weibull es

$$\hat{R}(t) = e^{-\frac{t^\alpha}{\hat{\theta}}}$$

donde $\hat{\theta}$ es el estimador MV del parámetro de escalar θ .

8.3.2 Método de los momentos

Definición 8.7 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución con función (densidad) de probabilidad $f(x; \theta)$. El r -ésimo momento alrededor de cero se define como

$$M'_r = \frac{1}{n} \sum_{i=1}^n X_i^r.$$

El método de los momentos proporciona una alternativa razonable cuando no se pueden determinar los estimadores de máxima verosimilitud. Recuérdese que los parámetros son, en general, funciones de los momentos teóricos. En esencia, el método se implementa igualando tantos momentos muestrales con los correspondientes momentos teóricos tantas veces como sea necesario para determinar un estimador de momentos para un parámetro desconocido.

8.3.3 Estimación por máxima verosimilitud para muestras censuradas

En algunas situaciones de muestreo, en forma especial en las pruebas de duración, el procedimiento de prueba puede terminar antes de proporcionar una muestra aleatoria completa. Una prueba típica de duración consiste en artículos iguales (tales como componentes eléctricos o mecánicos) seleccionados en forma aleatoria de un proceso y operando en un medio cuidadosamente controlado hasta que el artículo falla. Si la prueba de duración se termina sólo cuando todas las unidades de la muestra han fallado, se dice que la muestra aleatoria de tiempos está completa. Pero por varias razones la prueba termina ya sea después de un lapso de tiempo predeterminado x_0 o después de que falla un determinado número de unidades $m \leq n$. Las dos condiciones producen muestras censuradas. Si X_0 es un lapso de fijo de tiempo, el número de unidades que fallan de las n , desde el comienzo de la prueba hasta el tiempo x_0 es una variable aleatoria; esta constituye una muestra censurada de tipo *I*. Si m es fijo y el tiempo de terminación X_m es la variable aleatoria, se dice que la muestra es de tipo *II*.

Supongamos que los tiempos de duración de las unidades son variables aleatorias X_1, X_2, \dots, X_n independientes exponencialmente distribuidas, con una función

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

El interés recae en encontrar el estimador de máxima verosimilitud del parámetro θ . La función de verosimilitud para un muestreo censurado del tipo *II* es la probabilidad conjunta de que fallen m unidades en los tiempos x_1, x_2, \dots, x_m en ese orden, y sobrevivan $n - m$ unidades con un tiempo de supervivencia igual a x_m . La parte de la función de verosimilitud que corresponde a las m unidades que han fallado en los tiempos x_1, x_2, \dots, x_m es $f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_m; \theta)$. Pero esta sólo una de las posibles formas en que pueden fallar m unidades de un total de n . El número total de formas es $n! / (n - m)!$. La probabilidad de que $n - m$ unidades sobrevivan un tiempo x_m , está dada por la función de confiabilidad a tiempo x_m ; de esta forma, para la distribución exponencial,

$$P(X > x_m) = e^{-\frac{x_m}{\theta}}.$$

Por lo tanto, la función de verosimilitud Es

$$\begin{aligned} L(x_1, x_2, \dots, x_m; \theta) &= \frac{n!}{(n - m)!} \left[\frac{1}{\theta} e^{-\frac{x_1}{\theta}} \cdot \dots \cdot \frac{1}{\theta} e^{-\frac{x_1}{\theta}} e^{-\frac{x_1}{\theta}} \cdot \dots \cdot e^{-\frac{x_1}{\theta}} \right] \\ &= \frac{n!}{(n - m)!} \left[\frac{1}{\theta^m} e^{-\frac{1}{\theta} \sum_{i=1}^m x_i} \cdot e^{-\frac{n-m}{\theta} x_m} \right] \\ &= \frac{n!}{(n - m)!} \left(\frac{1}{\theta} e^{-\frac{1}{\theta} T_m} \right). \end{aligned}$$

Donde

$$T_m = \sum_{i=1}^m x_i + (n - m)x_m.$$

Tomando el logaritmo natural de L , se tiene

$$\ln L(x_1, x_2, \dots, x_m; \theta) = \ln(n!) - \ln[(n - m)!] - m \ln \theta - \frac{1}{\theta} T_m.$$

Entonces

$$\frac{d [\ln L(x_1, x_2, \dots, x_m; \theta)]}{d\theta} = -\frac{m}{\theta} + \frac{1}{\theta^2} T_M$$

e igualando la derivada a cero, el estimador de máxima verosimilitud de θ es

$$\hat{\theta} = \frac{\sum_{i=1}^m x_i + (n-m)x_m}{m}.$$

8.4 Estimación por intervalo

Una media muestral, por ejemplo $\bar{x} = 200$ unidades, es un estimador puntual de un parámetro desconocido. Este estimador, ¿implica que la demanda media desconocida no sea mayor de 250 ni menor de 150?. En este punto es difícil saberlo, ya que no se tiene ninguna indicación del posible error en el estimador puntual. El error en el estimado puntual se mide en términos de la variación muestral del correspondiente estimador.

Por ejemplo, supóngase que la desviación estándar de la media muestral \bar{X} es 60 unidades. De acuerdo con el teorema central del límite puede argumentarse que $\bar{X} \rightarrow N(\mu, 60)$, conforme $n \rightarrow \infty$. De esta forma, la probabilidad de que \bar{X} se encuentre dentro de dos desviaciones estándar alrededor de μ , es de, aproximadamente, 0.95. En otras palabras, para n grande,

$$P(|\bar{X} - \mu| < 120) = 0.95,$$

o

$$P(-120 < \bar{X} - \mu < 120) = 0.95.$$

Restando X y multiplicando por -1 en el interior de los paréntesis, se tiene

$$P(\bar{X} - 120 < \mu < \bar{X} + 120) = 0.95.$$

Si se sustituye el estimado para $\bar{x} = 200$, se tiene

$$P(80 < \mu < 320) = 0.95,$$

lo que sugiere que es enteramente posible que la demanda sea tan grande como 250 unidades o tan pequeña como 150 unidades, siempre que d.e. (\bar{X})=60. Por otro lado, supóngase que la desviación estándar de \bar{X} es igual a 10. Entonces, la expresión correspondiente a dos desviaciones estándar, es

$$P(\bar{X} - 20 < \mu < \bar{X} + 20) = 0.95,$$

y para $\bar{X} = 200$,

$$P(180 < \mu < 220) = 0.95.$$

En este caso es poco probable que μ sea tan grande como 250 o tan pequeño como 150. En ambos casos la clave está en la desviación estándar del estimador puntual. En esencia, para la estimación del intervalo se considera, tanto el estimador puntual del parámetro θ , como su distribución de muestreo.

El intervalo $[\bar{X} - 120, \bar{X} + 120]$ es un intervalo aleatorio, y la probabilidad de que este intervalo contenga el valor esperado de μ es de 0.95. En otras palabras, si se obtuviesen muestras del mismo tamaño en forma repetida de una población, y cada vez que estas se seleccionan, se calculan los valores específicos para el intervalo aleatorio $[\bar{X} - 120, \bar{X} + 120]$; entonces debe esperarse que un 95% de estos intervalos contengan el valor de la media desconocida μ . La probabilidad de 0.95 para el intervalo aleatorio sugiere que la confianza en que el intervalo (80, 320) contenga el valor de la media desconocida μ es alta. Así, cuando se escribe

$$P(80 < \mu < 320) = 0.95,$$

no se está formulando ninguna proposición probabilística en el sentido clásico, sino más bien se expresa un grado de confianza. El intervalo $(80, 320)$ recibe el nombre de intervalo de confianza del 95% para μ .

En términos generales, la construcción de un intervalo de confianza para un parámetro desconocido θ , consiste en encontrar una estadística suficiente T y relacionarla con otra variable aleatoria $X' = f(T; \theta)$, donde X involucra a θ pero la distribución de X no contiene a θ , así como tampoco a ningún otro parámetro desconocido. Entonces se seleccionan dos valores x_1 y x_2 tales que

$$P(x_1 < X < x_2) = 1 - \alpha,$$

donde $1 - \alpha$ recibe el nombre de **coeficiente de confianza**. Mediante una manipulación algebraica de las dos expresiones, se puede modificar el contenido entre paréntesis y expresarlo como

$$P(h_1(T) < \theta < h_2(T)) = 1 - \alpha,$$

donde $h_1(T)$ y $h_2(T)$ son funciones de la estadística T y de esta forma, variables aleatorias. Al seguirse el desarrollo podemos encontrar intervalos de confianza unilaterales, de la forma

$$P[g_1(T) < \theta] = 1 - \alpha$$

o

$$P[\theta < g_2(T)] = 1 - \alpha.$$

8.4.1 Intervalos de confianza para μ cuando se muestrea una distribución normal con varianza conocida

El interés recae en la construcción de un intervalo de confianza de un $100(1 - \alpha)\%$ sobre μ y donde α es un número pequeño, tal que $0 < \alpha < 1$. La construcción de un intervalo de confianza se hace con base en el mejor estimador de μ , explícitamente de la media muestral \bar{X} .

Sea

$$P(-120 < \bar{X} - \mu < 120) = 0.95.$$

Sumando μ , se tiene

$$P(\mu - 120 < \bar{X} < \mu + 120) = 0.95.$$

De esta forma, los límites $\mu - 120$ y $\mu + 120$ son funciones de los posibles valores de μ . Por lo tanto, y en general, se puede escribir

$$P[g_1(\mu) < \bar{X} < g_2(\mu)] = 1 - \alpha$$

de manera tal que

$$\int_{-\infty}^{g_1(\mu)} f(\bar{x}; \mu) d\bar{x} = \frac{\alpha}{2}$$

y

$$\int_{g_2(\mu)}^{\infty} f(\bar{x}; \mu) d\bar{x} = \frac{\alpha}{2},$$

donde $f(\bar{x}; \mu)$ es la función de densidad de la distribución de muestreo de \bar{X} , y $g_1(\mu)$ y $g_2(\mu)$ son funciones de μ las cuales no contiene a ningún otro parámetro desconocido.

Determinemos $g_1(\mu)$ y $g_2(\mu)$. Dado que $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$, la normal estándar $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, y

$$P[g_1(\mu) < \bar{X} < g_2(\mu)] = P\left[\frac{g_1(\mu) - \mu}{\sigma/\sqrt{n}} < Z < \frac{g_2(\mu) - \mu}{\sigma/\sqrt{n}}\right] = 1 - \alpha.$$

Pero ya que $P(z_{\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha$, en donde los valores cuantiles $z_{\alpha/2}$ y $z_{1-\alpha/2}$ son tales que $P(Z < z_{\alpha/2}) = \alpha/2$ y $P(Z < z_{1-\alpha/2}) = 1 - \alpha/2$, respectivamente. Se sigue que

$$\frac{g_1(\mu) - \mu}{\sigma/\sqrt{n}} = z_{\alpha/2}$$

y

$$\frac{g_2(\mu) - \mu}{\sigma/\sqrt{n}} = z_{1-\alpha/2}.$$

Dando solución en términos de $g_1(\mu)$ y $g_2(\mu)$, respectivamente, se obtienen

$$g_1(\mu) = \mu - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

y

$$g_2(\mu) = \mu + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

(Dado que para la normal estándar $z_{\alpha/2} = -z_{1-\alpha/2}$ puede sustituirse $-z_{1-\alpha/2}$ para $z_{\alpha/2}$ en la ecuación de arriba). Por lo tanto,

$$P\left(\mu - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Al manipular las desigualdades que se encuentran dentro de los paréntesis se tiene

$$P\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Si se reemplaza la variable aleatoria X por el estimador \bar{x} calculando a partir de los datos de una muestra de tamaño n , un intervalo de confianza del $100(1 - \alpha)\%$ para μ , es

$$\bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}},$$

en donde $\bar{x} - z_{1-\alpha/2}(\sigma/\sqrt{n})$ y $\bar{x} + z_{1-\alpha/2}(\sigma/\sqrt{n})$ reciben el nombre de límites de confianza inferiores y superiores, respectivamente, para μ .

Vemos que mientras más grande es el tamaño de la muestra, más pequeño es el ancho del intervalo; o para un coeficiente de confianza $1 - \alpha$ más grande, mayor es el ancho de intervalo. Es decir, un tamaño grande de la muestra disminuirá la varianza del estimador, y un coeficiente de confianza grande incrementa el valor cuantil dando como resultado un intervalo más amplio.

8.4.2 Intervalos de confianza para μ cuando se muestrea una distribución normal con varianza desconocida

Recordemos que cuando se muestrea una $N(\mu, \sigma)$, donde tanto μ como σ^2 son desconocidos, la variables aleatoria

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene distribución t de Student con $n - 1$ grados de libertad. Por lo tanto, es posible determinar el valor cuantil $t_{1-\alpha/2, n-1}$ de T , para el cual

$$P(-t_{1-\alpha/2, n-1} < T < t_{1-\alpha/2, n-1}) = 1 - \alpha,$$

en donde el valor cuantil es tal que $P(T < t_{1-\alpha/2, n-1}) = \alpha/2$ y $P(T < t_{1-\alpha/2, n-1}) = 1 - \alpha/2$. Al sustituir para T , se tiene

$$P\left(-t_{1-\alpha/2, n-1} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{1-\alpha/2, n-1}\right) = 1 - \alpha.$$

y

$$P\left(\bar{X} - t_{1-\alpha/2, n-1} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{1-\alpha/2, n-1} \frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Por lo tanto, el intervalo $\bar{X} \pm t_{1-\alpha/2, n-1} S/\sqrt{n}$ es un intervalo aleatorio y la probabilidad de que este contenga el valor verdadero de μ , es $1 - \alpha$. De esta forma, dados n datos de una muestra aleatoria de tamaño n a partir de las cuales se calculan los estimados \bar{x} y s^2 . Un intervalo de confianza del $100(1 - \alpha)\%$ para μ es

$$\bar{x} \pm t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}}.$$

8.4.3 Intervalo de confianza para la diferencia de medias cuando se muestrean dos distribuciones normales independientes

Sean X_1, X_2, \dots, X_{n_X} e Y_1, Y_2, \dots, Y_{n_Y} dos muestras aleatorias de dos distribuciones normales independientes, con medias μ_X y μ_Y y varianzas σ_X^2 y σ_Y^2 , respectivamente. Se desea construir un intervalo de confianza para la diferencia $\mu_X - \mu_Y$. Supóngase que se conocen los valores de las varianzas. Entonces la variable aleatoria

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$$

es $N(0, 1)$. De esta forma es posible encontrar el valor cuantil $z_{1-\alpha/2}$ tal que

$$P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha.$$

De donde se tiene

$$P\left(\bar{X} - \bar{Y} - z_{1-\alpha/2} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} = 1 - \alpha < \mu_X - \mu_Y < \bar{X} - \bar{Y} + z_{1-\alpha/2} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\right) = 1 - \alpha,$$

que es un intervalo aleatorio que no contiene parámetros desconocidos. Un intervalo de confianza del $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$ es

$$\bar{x} - \bar{y} \pm z_{1-\alpha/2} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}},$$

en donde el valor cuantil $z_{1-\alpha/2}$ es tal que $P(Z < z_{1-\alpha/2}) = 1 - \alpha/2$.

Si las varianzas σ_X^2 y σ_Y^2 se desconocen pero son iguales, entonces la variable aleatoria

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$$

tiene una distribución t de Student con $k = n_X + n_Y - 2$ grados de libertad. Al seguir el procedimiento anterior, se tiene que un intervalo de confianza del $100(1 - \alpha)\%$ para $\mu_X - \mu_Y$, es

$$\bar{x} - \bar{y} \pm t_{1-\alpha/2, n_X + n_Y - 2} s_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}},$$

en donde el estimado combinando de la varianza común es

$$s_p^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}.$$

8.4.4 Intervalo de confianza para σ^2 cuando se muestrean dos distribución normal con media desconocida

Recordemos que la distribución de muestreo de $(n-1)S^2/\sigma^2$ es chi-cuadrada con $n-1$ grados de libertad. Entonces, es posible determinar los valores cuantiles $\chi^2_{\alpha/2, n-1}$ y $\chi^2_{1-\alpha/2, n-1}$, tales que

$$P \left[\chi^2_{\alpha/2, n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\alpha/2, n-1} \right] = 1 - \alpha.$$

De donde

$$P \left[\frac{1}{\chi^2_{\alpha/2, n-1}} > \frac{\sigma^2}{(n-1)S^2} > \frac{1}{\chi^2_{1-\alpha/2, n-1}} \right] = 1 - \alpha$$

Entonces el intervalo

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} \right]$$

es un intervalo aleatorio el cual contiene a σ^2 y a parámetros conocidos con una probabilidad de $1 - \alpha$. De esta forma, con base en los datos de una muestra aleatoria de tamaño n , se calcula el estimado s^2 y un intervalo de confianza del $100(1 - \alpha)\%$ para σ^2 , es de $(n-1)s^2/\chi^2_{1-\alpha/2, n-1}$ a $(n-1)s^2/\chi^2_{\alpha/2, n-1}$. Es interesante notar que la variable aleatoria pivotal es $(n-1)S^2/\sigma^2$ ya que su función de densidad, dada por

$$f(y; n-1) = \begin{cases} \frac{1}{\Gamma[(n-1)/2] 2^{(n-1)/2}} y^{[(n-1)/2]-1} e^{-y/2} & y > 0, \\ 0 & \text{para cualquier otro valor.} \end{cases}$$

no contiene ningún parámetro desconocido.

8.4.5 Intervalos de confianza para el cociente de dos varianzas cuando se muestrean dos distribuciones normales independientes

Supongamos que se tienen muestras aleatorias provenientes de dos distribuciones normales con medias y varianzas desconocidas. Sean n_X y n_Y , el tamaño de las muestras y S_X^2 y S_Y^2 las varianzas muestrales. El interés se centra en construir un intervalo de confianza para el cociente σ_Y^2/σ_X^2 de las dos variables poblacionales. Recordemos que la variable aleatoria $(S_X^2/\sigma_X^2)(S_Y^2/\sigma_Y^2)$ tiene una distribución F con $n_X - 1$ y $n_Y - 1$ grados de libertad. Entonces puede escribirse

$$P \left(a < \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} < b \right) = 1 - \alpha.$$

en donde a y b son los valores cuantiles inferiores y superiores de una distribución F tal que

$$a = \frac{1}{f_{1-\alpha/2, n_Y-1, n_X-1}} \quad \text{y} \quad b = f_{1-\alpha/2, n_X-1, n_Y-1}.$$

La proposición de probabilidad se puede expresar como

$$P \left(a < \frac{S_X^2}{S_Y^2} \cdot \frac{\sigma_Y^2}{\sigma_X^2} < b \right) = 1 - \alpha$$

o

$$P \left(\frac{aS_Y^2}{S_X^2} < \frac{\sigma_Y^2}{\sigma_X^2} < \frac{bS_Y^2}{S_X^2} \right) = 1 - \alpha$$

De esta manera, un intervalo de confianza del $100(1 - \alpha)\%$ para σ_Y^2/σ_X^2 está dado por

$$\left(\frac{as_Y^2}{s_X^2}, \frac{bs_Y^2}{s_Y^2} \right).$$

8.4.6 Intervalos de confianza para el parámetro de proporción p cuando se muestrea una distribución binomial

El número de unidades defectuosas es una variable aleatoria binomial, si se supone una probabilidad constante e independiente. En una muestra aleatoria de tamaño n el parámetro p que representa la proporción de artículos defectuosos es desconocido. Se desea determinar un intervalo de confianza para p . Se demostró que el estimador de máxima verosimilitud de p , denotado por \hat{P} , es

$$\hat{P} = \frac{X}{n}.$$

en donde X es binomial con parámetros n y p . Nótese que \hat{P} es un estimador insesgado de p , ya que

$$E(\hat{P}) = \frac{1}{n}E(X) = \frac{np}{n} = p.$$

La varianza de \hat{P} se puede obtener de la siguiente forma:

$$\begin{aligned} \text{Var}(\hat{P}) &= \text{Var}\left(\frac{X}{n}\right) \\ &= \frac{1}{n^2} [np(1-p)] \\ &= \frac{p(1-p)}{n}. \end{aligned}$$

Recuérdese que para n grande, la variable aleatoria $(X - np)/\sqrt{np(1-p)}$ es aproximadamente $N(0, 1)$. Entonces puede demostrarse que la distribución De

$$\frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}$$

también tiende a $N(0, 1)$ para n grande. De esta forma, la probabilidad de intervalo aleatorio

$$\left[\hat{P} - z_{1-\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}, \hat{P} + z_{1-\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \right]$$

es en forma aproximada, $1 - \alpha$ para n grande. De acuerdo con lo anterior, un intervalo de confianza aproximado del $100(1 - \alpha)\%$ para el parámetro de proporción p , es

$$\left[\hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right],$$

en donde el estimador de máxima verosimilitud $\hat{p} = x/n$ se obtiene de la muestra aleatoria de tamaño n .

Con respecto al muestreo de una distribución binomial, un problema que surge, en forma frecuente, es el de estimar el tamaño de la muestra necesario de manera tal que con una confiabilidad de $100(1 - \alpha)\%$ aproximadamente, el estimado del parámetro de proporción se encuentre a no más de ϵ unidades de p . Dado el estimador de máxima verosimilitud X/n , puede expresarse

$$\left[\hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

como

$$P\left(\left|\frac{X}{n} - p\right| < \epsilon\right) = 1 - \alpha.$$

en donde

$$\epsilon = z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

Al resolver para n , se obtiene

$$n = \frac{z_{1-\alpha/2}^2 p(1-p)}{\epsilon^2}.$$

Utilizaremos el valor máximo de p , es decir $p = 1/2$. Esto será válido siempre que el tamaño de la muestra sea suficientemente grande.

Para ilustrar el uso de la distribución t de Student sigue siendo válido para inferencias con respecto a las medias, aún a pesar de que se muestree una distribución que no es normal.

8.5 Estimación bayesiana

8.5.1 Estimación puntual bayesiana

Definición 8.8 Sea $f_{\Theta}(\theta)$ la función de densidad a priori de un parámetro Θ , y $L(x_1, x_2, \dots, x_m | \theta)$ la función de máxima verosimilitud de una muestra aleatoria de n variables aleatorias IID condicionadas sobre la realización θ de Θ . Además, sea $f(\theta | \underline{x})$ la función de densidad a posteriori de Θ y sea $l(t, \Theta)$ la función de pérdida. El estimador Bayes de θ , $T = u(X_1, X_2, \dots, X_n)$ es aquél para el cual el valor esperado de la función de pérdida dada por

$$E[l(t, \Theta)] = \int_{\Theta} l(t, \theta) f(\theta | \underline{x}) d\theta$$

es mínimo.

8.5.2 Estimación bayesiana por intervalo

Definición 8.9 Sea $f(\theta | \underline{x})$ la función de densidad a posteriori de Θ condicionada sobre el resultado muestral $\underline{x} = \{x_1, x_2, \dots, x_n\}$ sean a y b límites tales que

$$P(a < \Theta < b | \underline{x}) = \int_a^b f(\theta | \underline{x}) d\theta = \gamma.$$

En donde a y b son funciones del resultado muestral \underline{x} . Entonces el intervalo (a, b) es un intervalo bayesiano tal que la probabilidad de que θ se encuentre contenido en (a, b) es γ .

8.6 Límites estadísticos de tolerancia

Definición 8.10 Si F es la proporción de observaciones de la variable aleatoria que se encuentra entre los límites L_1 y L_2 que son funciones univalueadas de las observaciones de manera tal que

$$D = \int_{L_1}^{L_2} f(x; \theta) dx = F_X(L_2; \theta) - F_X(L_1; \theta),$$

entonces L_1 y L_2 reciben el nombre de límites estadísticos de tolerancia.

Prueba de hipótesis estadísticas

9.1 Introducción

La prueba de una hipótesis estadística tiene una fuerte relación con el concepto de estimación.

Una hipótesis estadística es una afirmación con respecto a alguna característica desconocida de una población de interés. La esencia de probar una hipótesis estadística es el de decidir si la afirmación se encuentra apoyada por la evidencia experimental que se obtiene a través de una muestra aleatoria.

9.2 Conceptos básicos para la prueba de hipótesis estadísticas

La idea principal es determinar si el valor de un parámetro es igual a un valor específico o si es diferente de ese valor. A la afirmación igual a un valor específico se le llama *hipótesis nula* y se escribe como:

$$H_0 : \mu = 10.$$

Debemos tomar en cuenta las consecuencias que puede originarse como resultado del verdadero estado de la naturaleza al tomar alguna de las posibles decisiones. Es decir, existe dos posibles decisiones con respecto a H_0 : Rechazar o equivocarse al rechazar H_0 . Sin embargo, cada una de estas decisiones tiene las siguientes dos consecuencias con respecto al estado de la naturaleza:

$$\begin{array}{cc} \text{Rechazar } H_0 \left\{ \begin{array}{l} \text{cuando } H_0 \text{ es cierta} \\ \text{cuando } H_0 \text{ es falsa} \end{array} \right. & \text{Equivocarse al rechazar } H_0 \left\{ \begin{array}{l} \text{cuando } H_0 \text{ es cierta} \\ \text{cuando } H_0 \text{ es falsa} \end{array} \right. \end{array}$$

Cuando se toma una decisión con respecto a una hipótesis nula, dos de las posibles consecuencias relativas al verdadero estado de la naturaleza conducen a errores inferenciales. El rechazo de la hipótesis H_0 cuando en realidad es cierta, constituye lo que se denomina *error de tipo I*. Equivocarse al rechazar H_0 cuando en realidad H_0 es falsa, constituye lo que se denomina *error de tipo II*. Debemos tomar en cuenta que sólo es posible el error de tipo I cuando la decisión es el rechazar la hipótesis nula, mientras que el error de tipo II sólo es posible cuando la decisión es el no rechazar H_0 . En otras palabras, si la hipótesis nula es falsa, sólo puede cometerse un error de tipo I; si la hipótesis nula realmente es falsa, sólo puede cometerse un error de tipo II.

Es necesario tener alguna cantidad que mida la posibilidad de cometer alguno de estos errores. Esta medida es una probabilidad.

Definición 9.1 La probabilidad de rechazar H_0 , dado que H_0 es cierta, se define como la probabilidad (o tamaño) de error de tipo I y se denota por α , $0 \leq \alpha \leq 1$.

Definición 9.2 La probabilidad de no rechazar H_0 , dado que H_0 es falsa, se define como la probabilidad (o tamaño) de error de tipo II y se denota por β , $0 \leq \beta \leq 1$.

Por lo tanto, las probabilidades de error de tipo I y tipo II están dadas por las proposiciones

$$P(\text{rechazar } H_0 | H_0 \text{ es cierta}) = \alpha.$$

$$P(\text{no poder rechazar } H_0 | H_0 \text{ es falsa}) = \beta.$$

Cuando una afirmación se incorpora en la proposición de la hipótesis nula, se necesita una regla que indique que decisión tomar con respecto a H_0 una vez que se encuentra disponible la evidencia muestral. Esta regla recibe el nombre de prueba de una hipótesis estadística.

Definición 9.3 Una *prueba* de una hipótesis estadística con respecto a alguna característica desconocida de la población de interés es cualquier regla para decidir si se rechaza la hipótesis nula con base en una muestra aleatoria de la población.

Por ejemplo, se decide rechazar $H_0 : \mu = 10$. Para un tamaño n dado de la muestra, supóngase que se decide rechazar H_0 si se observa un valor de la media muestral \bar{X} que sea más grande que 12. Entonces, \bar{X} es la estadística de prueba, el valor $\bar{X} = 12$ es el *valor crítico*, y el conjunto de valores mayores que 12 constituyen la región crítica de la prueba.

El área de la región crítica es igual al tamaño del error de tipo I. Es decir,

$$P(\bar{X} > 12 | \mu = 10) = \alpha.$$

La probabilidad de α es sólo una referencia con respecto a la región $\bar{X} > 12$. Pero la decisión de rechazar H_0 se tomará con base en una sola muestra de tamaño n , a partir de la cual se calculará el estimador de \bar{x} . De esta forma, si $\bar{x} > 12$, esto no significa que la probabilidad de que H_0 sea correcta es α ; más bien, esto implica una interpretación de frecuencia para α cuando se toman muchas muestras. En otras palabras, si el valor de μ es realmente 10, y si se tomasen en forma repetida muestras de tamaño n de la población, debe esperarse que en un $100\alpha\%$ de las veces, se encuentre un valor de la estadística de prueba \bar{X} mayor que 12 y de esta forma debe rechazarse la hipótesis nula. Sólo en este sentido puede decirse que la confiabilidad al rechazar H_0 , cuando el estimador $\bar{X} > 12$ es igual al complemento del error α de tipo I, o $1 - \alpha$.

También es necesario establecer una *hipótesis alternativa* que refleje el valor posible o intervalo de valores del parámetro de interés si la hipótesis nula (H_0) es falsa. Esto es, H_1 representa alguna forma de negación de H_0 .

Cabe mencionar que el rechazo de la hipótesis nula implicaría que ha sido capaz encontrar evidencia suficiente para garantizar la hipótesis alternativa. Por otro lado, si no se presenta evidencia sustancial, se aceptará la hipótesis nula. Esta decisión no implica necesariamente que "el acusado sea inocente", más bien hace énfasis en la falta de evidencia sustancial necesaria para condenar al acusado. Por lo tanto, en cierto sentido un veredicto de culpable (rechazo de H_0) debe ser considerado como una decisión más fuerte que un veredicto de inocente (no rechazar H_0). A lo que, el error de tipo I se considera un error mucho más grave que el error de tipo II.

Como resultado se tiene que muchas veces se selecciona con anticipación el tamaño del error de tipo I que puede tolerarse y se intenta construir un procedimiento de prueba que minimice el tamaño del error de tipo II. Es decir, no es posible fijar tanto a α como a β y diseñar alguna regla de decisión para probar H_0 contra H_1 , dada una muestra aleatoria de tamaño n . Es por esta razón que se dice "equivocación al

rechazar H_0 mas que aceptar H_0 cuando la evidencia muestral no apoya el rechazo de la hipótesis nula.

También cabe mencionar que el error de tipo II aumenta conforme el error de tipo I disminuya. Y la probabilidad α del error de tipo I también se conoce como el *nivel de significancia estadístico*. En este contexto la palabra "significancia" sólo implica que la evidencia muestral es tal que garantiza el rechazo de H_0 a un nivel dado α .

9.3 Tipos de regiones críticas y la función de potencia

Se desea estudiar los tipos de regiones críticas que pueden surgir. Considérese la hipótesis nula simple

$$H_0 : \theta = \theta_0$$

con respecto al parámetro de interés θ , cuando se muestrea una distribución cuya función de densidad de probabilidad es $f(x|\theta)$ en donde θ_0 es el valor propuesto de θ . Si la hipótesis alternativa es de la forma

$$H_1 : \theta > \theta_0 \quad \text{o} \quad H_1 : \theta < \theta_0,$$

se dice que H_1 es una hipótesis *alternativa unilateral*. La región crítica también recibe el nombre de región de *rechazo unilateral*. Vale la pena notar que la hipótesis alternativa unilateral debe formularse sólo si el valor de uno de los parámetros que se encuentre en el lado opuesto, no tiene sentido para el investigador. De otro modo, debe establecerse una *hipótesis alternativa bilateral*. Esto es, si la hipótesis alternativa no proporciona una dirección con respecto al valor propuesto de θ_0 . Es decir,

$$H_1 : \theta \neq \theta_0.$$

El tamaño del error de tipo II se obtiene como una función de los valores alternativos de θ ($\theta >$) bajo H_1 . Debe notarse que $\beta(\theta)$ se conoce como la *función característica de operación*, y cuando se grafica $\beta(\theta)$ para diversos valores de θ de H_1 , se obtiene una curva característica de operación (CO).

Dado que $\beta(\theta)$ es la probabilidad de que un valor de la estadística de prueba no se encuentre en la región crítica cuando H_0 es falsa, entonces $1 - \beta(\theta)$ representa la probabilidad de que un valor de la estadística de prueba se encuentre dentro de la región crítica cuando H_0 es falsa. Esta probabilidad se conoce como la *función potencia de la prueba*. En otras palabras, las funciones potencia y característica de operación son complementarias.

Definición 9.4 La función $P(\theta) = 1 - \beta(\theta)$ recibe el nombre de función potencia y representa la probabilidad de rechazar la hipótesis nula cuando esta es falsa; es decir, cuando el valor del parámetro de H_1 es cierto.

En esencia, la potencia de una prueba es la probabilidad de detectar que H_0 es, en forma verdadera, falsa; de aquí el uso de la palabra potencia.

9.4 Las mejores pruebas

Sea X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n de una población cuya función de densidad de probabilidad es $f(x|\theta)$, y considérese la hipótesis

$$H_0 : \theta = \theta_0$$

contra

$$H_1 : \theta = \theta_1.$$

en donde se especifican θ_0 y θ_1 . Supóngase que α es el tamaño máximo del error de tipo I que se puede tolerar. Entonces la mejor prueba para H_0 contra H_1 es aquella que tiene el tamaño más pequeño del error de tipo II (y de esta forma la mayor potencia) de entre todas las pruebas que tengan un tamaño del error de tipo I no mayor que α . Se puede determinar las regiones críticas para estas pruebas mediante el uso del siguiente teorema, el cual se conoce como el **lema de Neyman Pearson**:

Teorema 9.1 Si existe una región crítica C de tamaño α y una constante positiva k tal que

$$\frac{L_0(x_1, x_2, \dots, x_n; \theta_0)}{L_1(x_1, x_2, \dots, x_n; \theta_1)} \leq k \quad \text{interior } C,$$

$$\frac{L_0(x_1, x_2, \dots, x_n; \theta_0)}{L_1(x_1, x_2, \dots, x_n; \theta_1)} \geq k \quad \text{exterior } C,$$

entonces C es la mejor región crítica de tamaño α para probar $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$, en donde L_0 y L_1 son las funciones de verosimilitud relativa a H_0 y H_1 respectivamente. ■

9.5 Principios generales para probar una H_0 simple contra una H_1 uni o bilateral

No existen pruebas uniformes más potentes para hipótesis alterativas bilaterales a pesar de que, en forma usual, existen para hipótesis alternativas unilaterales.

9.5.1 Principios generales para el caso 1

Considérese la prueba de la hipótesis nula

$$H_0 : \theta = \theta_0$$

contra la alternativa

$$H_1 : \theta \neq \theta_0,$$

donde θ_0 es el valor propuesto de algún parámetro θ bajo H_0 . Dada una muestra aleatoria de tamaño n de la distribución de interés, el procedimiento general para probar H_0 es escoger el mejor estimador de θ , T y rechazar H_0 cuando el estimado t obtenido de la muestra, es en forma suficiente, diferente del valor propuesto de θ_0 . Esto es, si el estimado t es lo suficientemente distinto del valor propuesto θ_0 , entonces se ha observado un evento raro (y la hipótesis nula es correcta), o se ha observado un valor de la estadística que sugiere un valor θ diferente del propuesto θ_0 . Cuando el estimado t es en forma suficiente distinto de θ_0 , se asumirá la última posibilidad y se dejará el tamaño del error de tipo II igual a la probabilidad del anterior. En particular, para un tamaño preseleccionado α , del error de tipo I se obtiene una región crítica bilateral en los extremos de la distribución de muestreo de T , de esta manera tal que el área, en cualquier lado, más allá del valor crítico es igual a $\alpha/2$. Entonces se rechaza H_0 en favor de H_1 cuando el estimado t se encuentra dentro de la región crítica. Cuando el estimado t no se encuentra dentro de la región crítica, no puede rechazarse la hipótesis nula. De esta forma, cualquier diferencia con respecto al valor de θ_0 se considera causada por fluctuaciones en el muestreo del estimador T .

Este enfoque es muy similar a la construcción de un intervalo de confianza bilateral para θ . Para cualquier valor propuesto de θ_0 , que se encuentre dentro de un intervalo de confianza del $100(1 - \alpha)\%$ para θ , H_0 no será rechazada. y si se encuentra fuera se rechazara la hipótesis nula.

9.5.2 Principios generales para el caso 2

Considérese la prueba de la hipótesis nula

$$H_0 : \theta = \theta_0$$

contra la alternativa

$$H_1 : \theta > \theta_0.$$

Para un tamaño α , del error tipo I, la región crítica se encuentra localizada en el extremo superior de la distribución de muestreo T y H_0 se rechaza si el estimado t no es menor que el valor crítico.

9.5.3 Principios generales para el caso 3

Para probar la hipótesis nula

$$H_0 : \theta = \theta_0$$

contra

$$H_1 : \theta < \theta_0,$$

el procedimiento general es rechazar a H_0 cada vez que el estimado t sea, e forma suficiente menor que el valor propuesto θ_0 .

Con respecto a la prueba de hipótesis estadísticas, y debido a que se coloca un gran énfasis en el tamaño del error de tipo I generalmente se formula la hipótesis nula tal que esta se rechace si la evidencia experimental apoya esta decisión. Es decir, lo que realmente se desea es concluir que la hipótesis alternativa es la correcta.

En la práctica escogemos el tamaño del error de tipo I antes de la determinación de la muestra aleatoria. Si se obtiene como resultado que la hipótesis nula no puede rechazarse con el valor escogido de α debe evitarse aumentar el tamaño del error de tipo I con la idea de rechazar la hipótesis nula.

La discusión anterior constituye el método clásico para probar hipótesis estadísticas. El cual, se dirigieron algunas críticas directas hacia este enfoque debido a que la decisión final de rechazar o no una H_0 dada, es demasiado cortante y seca y no proporciona una medida real de que la decisión sea correcta en términos de la probabilidad. Para esto lo que se ha sugerido es el cálculo del llamado **valor p**. El valor p es la probabilidad, dado que H_0 es cierta, de que la estadística de prueba tome un valor mayor o igual que el calculado con base a la muestra aleatoria. Un valor relativamente pequeña puede sugerir que si H_0 es realmente cierta, el valor de la estadística de prueba sea poco probable. Puede entonces optarse por rechazar H_0 debido a que esta decisión tendrá una alta probabilidad de ser correcta.

Se recomienda el cálculo del valor p acoplado con el enfoque clásico de escoger un tamaño del error de tipo I antes de la determinación de la muestra aleatoria. Entonces, la decisión de rechazar o no a H_0 puede basarse en una región crítica de tamaño α , con el valor p proporcionando una medida real en términos de la probabilidad de que la decisión sea correcta. De acuerdo con lo anterior, se sugiere lo siguiente regla:

Si el valor p es menor o igual a α , se rechaza H_0 ; de otra forma no puede rechazarse la hipótesis nula.

9.6 Prueba de hipótesis con respecto a las medias cuando se muestrean distribuciones normales

9.6.1 Pruebas para una muestra

Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución normal con media μ desconocida. En este caso el interés recae en probar uno de los siguientes conjuntos de hipótesis con respecto a μ .

$$\begin{array}{lll} H_0 : \mu = \mu_0 & H_0 : \mu = \mu_0 & H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 & H_1 : \mu > \mu_0 & H_1 : \mu < \mu_0 \end{array}$$

Primero, supóngase que el valor de la varianza poblacional σ^2 es conocido. Entonces la estadística de prueba es la media muestral \bar{X} , misma que, bajo la hipótesis nula, tiene una distribución normal con media μ_0 y desviación estándar σ/\sqrt{n} . La región crítica de tamaño α para la hipótesis bilateral es de la forma

$$\text{Rechazar } H_0 \text{ si } \begin{cases} \bar{X} \geq \bar{x}_{1-\alpha/2} \\ \text{o} \\ \bar{X} \leq \bar{x}_{\alpha/2} \end{cases}$$

donde $\bar{x}_{1-\alpha/2}$ y $\bar{x}_{\alpha/2}$ son los valores cuantiles críticos de \bar{X} de manera tal que

$$P(\bar{X} \geq \bar{x}_{1-\alpha/2}) = \alpha/2 \quad \text{y} \quad P(\bar{X} \leq \bar{x}_{\alpha/2}) = \alpha/2.$$

Dado que bajo H_0 , $\bar{X} \sim N(\mu_0, \sigma/\sqrt{n})$, entonces en forma equivalente

$$P\left(Z \geq \frac{\bar{x}_{1-\alpha/2} - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha/2 \quad \text{y} \quad P\left(Z \leq \frac{\bar{x}_{\alpha/2} - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha/2.$$

o

$$z_{1-\alpha/2} = \frac{\bar{x}_{1-\alpha/2} - \mu_0}{\sigma/\sqrt{n}} \quad \text{y} \quad z_{\alpha/2} = \frac{\bar{x}_{\alpha/2} - \mu_0}{\sigma/\sqrt{n}}$$

en donde $z_{1-\alpha/2}$ y $z_{\alpha/2}$ son los correspondientes valores cuantiles de Z . Por lo tanto, se sigue que H_0 debe rechazarse cuando un valor \bar{x} de la media muestral \bar{X} es tal que

$$\bar{x} \geq \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}} + \mu_0 \quad \text{o} \quad \bar{x} \leq \frac{\sigma z_{\alpha/2}}{\sqrt{n}} + \mu_0.$$

De manera equivalente, se rechazará H_0 cuando

$$z \geq z_{1-\alpha/2} \quad \text{o} \quad z \leq z_{\alpha/2},$$

donde $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$ es el valor de la correspondiente normal estándar al valor \bar{x} de \bar{X} .

Para la hipótesis alternativa unilateral $H_1 : \mu > \mu_0$ la región crítica de tamaño α es el extremo derecho de la distribución de muestreo \bar{X} ; esta es de la forma

$$\text{Rechazar } H_0 \text{ si } \bar{X} \geq \bar{x}_{1-\alpha}.$$

en donde $\bar{x}_{1-\alpha}$ es el valor cuantil de \bar{X} , tal que $P(\bar{X} \geq \bar{x}_{1-\alpha}) = \alpha$. En forma similar, para la hipótesis alternativa $H_1 : \mu < \mu_0$, la región crítica de la forma

$$\text{Rechazar } H_0 \text{ si } \bar{X} \leq \bar{x}_{\alpha}.$$

en donde el valor \bar{x} es tal que $P(\bar{X} \leq \bar{x}_{\alpha}) = \alpha$.

Antes de resolver un ejemplo, se desarrollará una expresión general del error de tipo II para uno de los casos. Considérese la hipótesis nula $H_0 : \mu = \mu_0$ contra la alternativa $H_1 : \mu > \mu_0$. Supóngase que en realidad $\mu = \mu_1 > \mu_0$. De acuerdo a

$$\text{Rechazar } H_0 \text{ si } \bar{X} \geq \bar{x}_{1-\alpha},$$

no puede rechazarse H_0 si un valor de \bar{X} es menor que $(\sigma z_{1-\alpha}/\sqrt{n}) + \mu_0$. Dado que la probabilidad de error de tipo II es igual a la probabilidad de no rechazar un H_0 falsa, es necesario determinar

$$\beta = P\left(\bar{X} < \frac{\sigma z_{1-\alpha}}{\sqrt{n}} + \mu_0 \mid \mu = \mu_1 > \mu_0\right),$$

la que es términos de la normal estándar es

$$\beta = P \left(Z < \frac{\frac{\sigma z_{1-\alpha}}{\sqrt{n}} + \mu_0 - \mu_1}{\sqrt{n}} + \mu_0 \middle| \mu = \mu_1 \right).$$

Al sustituir cualquier valor μ_1 de la hipótesis alternativa, se puede calcular el correspondiente valor de la probabilidad de error de tipo II y de esta forma, la potencia.

Hipótesis nula	Valor de la estadística de prueba bajo H_0
$H_0 : \mu = \mu_0$	$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$
Hipótesis alternativa	Criterios de rechazo
$H_1 : \mu \neq \mu_0$	Rechazar H_0 cuando $z \leq z_{\alpha/2}$ o cuando $z \geq z_{1-\alpha/2}$
$H_1 : \mu > \mu_0$	Rechazar H_0 cuando $z \geq z_{1-\alpha}$
$H_1 : \mu < \mu_0$	Rechazar H_0 cuando $z \leq z_{\alpha}$

9.6.2 Pruebas para dos muestras

Sean X_1, X_2, \dots, X_{n_X} y Y_1, Y_2, \dots, Y_{n_Y} muestras aleatorias provenientes de dos distribuciones normales independientes con medias μ_X y μ_Y y varianzas σ_X^2 y σ_Y^2 , respectivamente. Supóngase que se desea probar la hipótesis nula

$$H_0 = \mu_X - \mu_Y = \delta_0$$

contra una de las siguientes alternativas:

$$H_1 : \mu_X - \mu_Y > \delta_0, \quad H_1 : \mu_X - \mu_Y < \delta_0, \quad H_1 : \mu_X - \mu_Y \neq \delta_0,$$

donde δ_0 es una cantidad que toma valores positivos o cero y la cual representa la diferencia propuesta entre los valores desconocidos de las medias. Supongamos que las varianzas de la población se conocen. De las secciones 7.7, 8.4, 8.4.3, es razonable concluir que la estadística de prueba apropiada es la diferencia muestral media $\bar{X} - \bar{Y}$. En particular, si un valor de $\bar{X} - \bar{Y}$ con base en la muestra aleatoria es lo suficientemente diferente, mayor o menor que δ_0 , se rechazará la hipótesis nula dependiendo de la hipótesis alternativa en cuestión.

Criterios de rechazo para la prueba de hipótesis con respecto a las medias de dos distribuciones normales e independientes con varianza conocidas

Hipótesis nula	Valor de la estadística de prueba bajo H_0
$H_0 : \mu_X - \mu_Y = \delta_0$	$z = \frac{\bar{x} - \bar{y} - \delta_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$
Hipótesis alternativa	Criterios de rechazo
$H_1 : \mu_X - \mu_Y \neq \delta_0$	Rechazar H_0 cuando $z \leq z_{\alpha/2}$ o cuando $z \geq z_{1-\alpha/2}$
$H_1 : \mu_X - \mu_Y > \delta_0$	Rechazar H_0 cuando $z \geq z_{1-\alpha}$
$H_1 : \mu_X - \mu_Y < \delta_0$	Rechazar H_0 cuando $z \leq z_{\alpha}$

Ahora, si se especifican los tamaños particulares α y β de los errores de tipo I y de tipo II, respectivamente, obtendremos una expresión para n .

Si H_0 es realmente cierta, la probabilidad de rechazarla es α ; y si H_0 es falsa en este caso supongamos ($\mu_X - \mu_Y = \sigma_1 > \sigma_0$), la probabilidad de no rechazar H_0 es β . Sea c_0 el valor crítico con respecto a la distribución de muestreo $\bar{X} - \bar{Y}$. Entonces, H_0 será rechazada cuando $\bar{x} - \bar{y} \geq c_0$, tal que

$$P(\bar{X} - \bar{Y} \geq c_0 | \mu_X - \mu_Y = \delta_0) = \alpha,$$

En términos de la normal estándar, esto es

$$P\left(Z \geq \frac{c_0 - \delta_0}{\sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}}} \middle| \mu_X - \mu_Y = \sigma_0\right) = \alpha.$$

Dado que pueden determinarse valores cuantiles $z_{1-\alpha}$ de la normal estándar tales que

$$P(Z \geq z_{1-\alpha}) = \alpha,$$

se tiene

$$\frac{c_0 - \delta_0}{\sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}}} = z_{1-\alpha}.$$

Si $\mu_X - \mu_Y > \delta_0$, entonces la probabilidad de no rechazar a H_0 es β . Por lo tanto,

$$P(\bar{X} - \bar{Y} < c_0 | \mu_X - \mu_Y = \delta_1) = \beta,$$

que en términos de la normal estándar es

$$P\left(Z < \frac{c_0 - \delta_1}{\sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}}} \middle| \mu_X - \mu_Y = \delta_1\right) = \beta.$$

Luego,

$$c_0 = z_{1-\alpha} \sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}} + \delta_0,$$

y

$$c_0 = z_\beta \sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}} + \delta_1.$$

Igualando,

$$z_{1-\alpha} \sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}} + \delta_0 = z_\beta \sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}} + \delta_1,$$

de donde

$$\sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}} (z_{1-\alpha} - z_\beta) = \delta_1 - \delta_0$$

Dado que para la normal estándar $-z_\beta = z_{1-\beta}$, se tiene

$$\sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{n}} (z_{1-\alpha} - z_{1-\beta}) = \delta_1 - \delta_0,$$

y por lo tanto

$$n = \frac{(z_{1-\alpha} - z_{1-\beta})^2 (\sigma_X^2 + \sigma_Y^2)}{(\delta_1 - \delta_0)^2}.$$

Esta expresión determina el tamaño de cada una de las dos muestras aleatorias en las dos distribuciones normales independientes, asegurando probabilidades α y β para los errores de tipo I y II, respectivamente, cuando se prueba

$$H_0 : \mu_X - \mu_Y = \delta_0 \text{ contra } H_1 : \mu_X - \mu_Y = \delta_1 > \delta_0.$$

Ahora, analicemos el caso en el que el valor de la varianza no se conoce; si las varianzas σ_X^2 y σ_Y^2 no se conocen pero se supone que son iguales, entonces para la hipótesis nula

$$H_0 : \mu_X - \mu_Y = \delta_0$$

la estadística de prueba es

$$T = \frac{\bar{X} - \bar{Y} - \delta_0}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}},$$

la cual tiene una distribución t de Student con

$$n_X + n_Y - 2$$

grados de libertad.

Criterios de rechazo para la prueba de hipótesis con respecto a las medias de dos distribuciones normales e independientes con varianza iguales pero desconocidas

Hipótesis nula	Valor de la estadística de prueba bajo H_0
$H_0 : \mu_X - \mu_Y = \delta_0$	$t = \frac{\bar{x} - \bar{y} - \delta_0}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$
Hipótesis alternativa	Criterios de rechazo
$H_1 : \mu_X - \mu_Y \neq \mu_0$	Rechazar H_0 cuando $t \leq t_{\alpha/2,m}$ o cuando $z \geq z_{1-\alpha/2,m}$, en donde $m = n_X + n_Y - 2$
$H_1 : \mu_X - \mu_Y > \mu_0$	Rechazar H_0 cuando $t \geq t_{1-\alpha,m}$
$H_1 : \mu_X - \mu_Y < \mu_0$	Rechazar H_0 cuando $t \leq t_{\alpha,m}$

9.6.3 Reflexiones sobre las suposiciones y sensibilidad

Debemos observar que la distribución t es muy robusta, o insensible a las suposiciones de normalidad, y en forma especial cuando el tamaño de la muestra es mayor o igual a 15.

9.6.4 Prueba sobre las medias cuando las observaciones están pareadas

Es deseable tener a las personas u objetos que producirán las observaciones dentro de cada nivel, tan homogéneas como sea posible. Si existe un efecto debido a factores externos, éstos pueden neutralizarse mediante la aplicación del principio de aleatorización. También es posible controlar las variaciones no deseadas controlando los factores extraños. Esto se logra tomando las observaciones en pares, donde se supone que las condiciones externas son las mismas para cada par pero pueden variar de par en par.

9.7 Prueba de hipótesis con respecto a las varianzas cuando se muestrean distribuciones normales

La inferencia con respecto a una varianza es tan importante como una con respecto a la media.

9.7.1 Pruebas para una muestra

Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución normal con media μ desconocida y varianza σ^2 desconocida. Considérese la prueba de la siguiente hipótesis

$$H_0 : \sigma^2 = \sigma_0^2$$

contra una de las siguientes alternativas

$$H_1 : \sigma^2 \neq \sigma_0^2, \quad H_1 : \sigma^2 > \sigma_0^2, \quad H_1 : \sigma^2 < \sigma_0^2,$$

donde σ_0^2 es el valor propuesto para σ^2 . La estadística de interés es la varianza muestral S^2 .

Criterios de rechazo para la prueba de hipótesis con respecto a la varianza de una distribución normal con media desconocida

Hipótesis nula	Valor de la estadística de prueba bajo H_0
$H_0 : \sigma^2 = \sigma_0^2$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$
Hipótesis alternativa	Criterios de rechazo
$H_1 : \sigma^2 \neq \sigma_0^2$	Rechazar H_0 cuando $\chi^2 \geq \chi_{1-\alpha/2, n-1}^2$ o cuando $\chi^2 \leq \chi_{\alpha/2, n-1}^2$
$H_1 : \sigma^2 > \sigma_0^2$	Rechazar H_0 cuando $\chi^2 \geq \chi_{1-\alpha, n-1}^2$
$H_1 : \sigma^2 < \sigma_0^2$	Rechazar H_0 cuando $\chi^2 \leq \chi_{1-\alpha, n-1}^2$

9.7.2 Pruebas para dos muestras

Sean $X_1, X_2, \dots, X_{n_X}, Y_1, Y_2, \dots, Y_{n_Y}$ dos muestras aleatorias de dos distribuciones normales independientes con medias desconocidas μ_X y μ_Y y varianzas desconocidas σ_X^2 y σ_Y^2 . Considérese la prueba de la siguiente hipótesis nula

$$H_0 : \sigma_X^2 = \sigma_Y^2$$

contra una de las siguientes alternativas

$$H_1 : \sigma_X^2 \neq \sigma_Y^2, \quad H_1 : \sigma_X^2 > \sigma_Y^2, \quad H_1 : \sigma_X^2 < \sigma_Y^2.$$

Las estadísticas de interés son las varianzas muestrales S_X^2 y S_Y^2 .

Recuerde que por virtud de la independencia, las cantidades $(n_X - 1)S_X^2/\sigma_X^2$ y $(n_Y - 1)S_Y^2/\sigma_Y^2$ son dos variables aleatorias independientes chi-cuadrada con $n_X - 1$ y $n_Y - 1$ grados de libertad, respectivamente. Entonces se sigue la estadística

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$$

tiene una distribución F con $n_X - 1$ y $n_Y - 1$ grados de libertad. Pero bajo la hipótesis nula, $\sigma_X^2 = \sigma_Y^2$; de esta forma la estadística se reduce a

$$F = S_X^2/S_Y^2.$$

Criterios de rechazo para la prueba de hipótesis con respecto a la varianza de una distribución normal con media desconocida

Hipótesis nula	Valor de la estadística de prueba bajo H_0
$H_0 : \sigma_X^2 = \sigma_Y^2$	$f = \frac{S_X^2}{S_Y^2}$
Hipótesis alternativa	Criterios de rechazo
$H_1 : \sigma_X^2 \neq \sigma_Y^2$	Rechazar H_0 cuando $f \geq f_{1-\alpha/2, n_X-1, n_Y-1}$ o cuando $f \leq f_{\alpha/2, n_X-1, n_Y-1}$
$H_1 : \sigma_X^2 > \sigma_Y^2$	Rechazar H_0 cuando $f \geq f_{1-\alpha, n_X-1, n_Y-1}$
$H_1 : \sigma_X^2 < \sigma_Y^2$	Rechazar H_0 cuando $f \leq 1/f_{1-\alpha, n-1}$

9.8 Inferencias con respecto a las proporciones de dos distribuciones binomiales independientes

En valores grandes de n_1 y n_2 , la distribución de la estadística $\hat{p}_1 - \hat{p}_2$ es, en forma aproximada, normal con media y varianza,

$$E(\hat{p}_1 - \hat{p}_2) = P_1 - P_2, \quad y \quad V(\hat{p}_1 - \hat{p}_2) = \frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2},$$

respectivamente. En otras palabras, la distribución de

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}}$$

es aproximadamente $N(0, 1)$. Se sigue que para n_1 y n_2 grande, la probabilidad del intervalo aleatorio

$$[(\hat{p}_1 - \hat{p}_2) - z_{1-\alpha/2} d.e. (\hat{p}_1 - \hat{p}_2), \quad (\hat{p}_1 - \hat{p}_2) + z_{1-\alpha/2} s.d. (\hat{p}_1 - \hat{p}_2)]$$

es aproximadamente $1 - \alpha$, y el intervalo de confianza aproximado del $100(1 - \alpha)\%$ para $P_1 - P_2$ es

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}.$$

en donde $\hat{p}_1 = x/n_1$ y y/n_2 son los estimados de máxima verosimilitud de p_1 y p_2 respectivamente.

Pruebas de bondad de ajuste y análisis de tablas de contingencia

Recordemos que una hipótesis estadística es una afirmación con respecto a una característica que se desconoce de una población de interés. Se examinará en este capítulo las pruebas de hipótesis estadísticas en las que la característica que se desconoce es alguna propiedad de la forma función de la distribución que se muestrea. También se discutirá las pruebas de independencia entre dos variables aleatorias.

10.1 La prueba de bondad de ajuste chi-cuadrada

Una prueba de bondad de ajuste se emplea para decidir cuándo un conjunto de datos se apega a una distribución de probabilidad dada. Sea una muestra aleatoria de tamaño n de la distribución de una variable aleatoria X dividida en k clases exhaustivas y mutuamente excluyentes, y sea $N_i, i = 1, 2, \dots, k$, el número de observaciones en la i -ésima clase. Considérese la verificación de la hipótesis nula

$$H_0 : F(x) = F_0(x),$$

en donde el modelo de probabilidad propuesto $F_0(x)$ se encuentra especificado, de manera completa, con respecto a todos los parámetros. De esta manera se puede obtener la probabilidad p_i de obtener una observación en la i -ésima clase bajo H_0 , en donde necesariamente $\sum_{i=1}^k p_i = 1$.

Sea n_i la realización de N_i para $i = 1, 2, \dots, k$ de manera tal que $\sum_{i=1}^k n_i = n$. La probabilidad de tener, de manera exacta, n_i observaciones en la i -ésima clase es $p_i^{n_i}$ para $i = 1, 2, \dots, k$. Dado que existen k categorías mutuamente excluyentes con probabilidad p_1, p_2, \dots, p_k , entonces bajo la hipótesis nula la probabilidad de la muestra agrupada es igual a la función de probabilidad de una distribución multinomial determinada.

Para deducir una prueba estadística adecuada para H_0 , Considérese el caso en el que $k = 2$. Este es la distribución binomial con una función de probabilidad dada y en la que $x = n_1$, $p = p_1$, $n - x = n_2$ y $1 - p = p_2$. Considérese la variable aleatoria estandarizada

$$Y = \frac{N_1 - np_1}{\sqrt{np_1(1 - p_1)}}.$$

Recuérdese que para un valor de n suficientemente grande, la distribución de Y es aproximadamente igual a la normal estándar. Además, se sabe que el cuadrado de una variable aleatoria normal estándar

tiene una distribución chi-cuadrada con un grado de libertad. Entonces, la estadística

$$\begin{aligned}
 \frac{(N_i - np_1)^2}{np_1(1 - p_1)} &= \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_1 - np_1)^2}{np_2} \\
 &= \frac{(N_1 - np_1)^2}{np_1} + \frac{[n - N_2 - n(1 - p_2)]^2}{np_2} \\
 &= \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_2 - np_2)^2}{np_2} \\
 &= \sum_{i=1}^2 \frac{(N_i - np_i)^2}{np_i}.
 \end{aligned}$$

tiene aproximadamente una distribución chi-cuadrada con un grado de libertad conforme n va tomando valores cada vez más grandes.

Si se sigue este tipo de razonamiento, puede demostrarse que para $k \geq 2$ categorías distintas, la estadística

$$\sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}$$

tiene una distribución, en forma aproximada, chi-cuadrada con $k - 1$ grados de libertad, si n tiene un valor suficientemente grande. Nótese que N_i es la frecuencia observada en la i -ésima clase, y np_i es la frecuencia correspondiente que se esperaba bajo la hipótesis nula. De acuerdo con lo anterior, la estadística es la suma sobre todas las k clases de los cocientes de los cuadrados de las diferencias entre las frecuencias observadas y esperadas, y la frecuencia esperada. Esta estadística recibe el nombre de *prueba de bondad de ajuste chi-cuadrada de Pearson*. Si existe una concordancia perfecta entre las frecuencias que se observan y las que se esperan, la estadística tendrá un valor igual a cero; por otro lado, si existe gran discrepancia entre estas frecuencias, la estadística tomará un valor muy grande. Por ello se desprende que para un tamaño dado del error de tipo I, la región crítica es el extremo superior de una distribución chi-cuadrada con $k - 1$ grados de libertad. Esta prueba, es un procedimiento razonablemente adecuado para probar suposiciones de normalidad siempre y cuando el tamaño de la muestra sea grande. Podría ser n cinco veces el número de clases. Por lo que se implementa la regla de seleccionar una muestra de manera tal que toda frecuencia esperada no sea menor que cinco. La aplicabilidad de la prueba de ajustes chi-cuadrada es cuestionable cuando se tiene muestras de tamaño muy grande.

10.2 La prueba de Kolmogorov-Smirnov

Una prueba de bondad de ajuste más adecuada que la chi-cuadrada cuando $F_0(x)$ es continua, es la basada en la estadística de Kolmogorov-Smirnov. Esta no necesita que los datos estén agrupados y es aplicable a muestras de tamaño pequeño. Esta se basa en una comparación entre las funciones de distribución acumulada que se observan en la muestra ordenada y la distribución propuesta bajo la hipótesis nula. Si esta comparación revela una diferencia suficientemente grande entre las funciones de distribución muestral y propuesta, entonces la hipótesis nula de que la distribución es $F_0(x)$, se rechaza.

Considérese la hipótesis nula de que la distribución $F_0(x)$ se especifica en forma completa. Denótese por $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ a las observaciones ordenadas de una muestra aleatoria de tamaño n y definase la función de distribución acunulativa muestral como

$$S_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ k/n & x_{(k)} \leq x < x_{(k+1)}, \\ 1 & x \geq x_{(n)}. \end{cases}$$

En otras palabras, para cualquier valor ordenado x de la muestra aleatoria, $S_n(x)$ es la proporción del número de valores en la muestra que son iguales o menores a x . Ya que $F_0(x)$ se encuentra completamente especificada, es posible evaluar a $F_0(x)$ para algún valor deseado de x , y entonces comparar este último con el valor correspondiente de $S_n(x)$. Si la hipótesis nula es verdadera, entonces es lógico esperar que la diferencia sea relativamente pequeña. La estadística de Kolmogorov-Smirnov se define como

$$D_n = \max_x |S_n(x) - F_0(x)|.$$

La estadística D_n tiene una distribución que es independiente del modelo propuesto bajo la hipótesis nula.

Para un tamaño α del error de tipo I, la región crítica es de la forma

$$P\left(D_n \geq \frac{c}{\sqrt{n}}\right) = \alpha.$$

De acuerdo con lo anterior, la hipótesis H_0 se rechaza si para algún valor x observado, el valor de D_n se encuentra dentro de la región crítica de tamaño α .