

Fórmulas sobre Probabilidad y Variable Aleatoria

1. Probabilidad

1.1 σ -álgebra de eventos

Dado un espacio muestral S y una familia \mathcal{A} de subconjunto de S , esta familia será llamada σ -álgebra de eventos si tiene las propiedades siguientes:

(\mathcal{A}_1) Si $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

(\mathcal{A}_2) $\emptyset \in \mathcal{A}$

(\mathcal{A}_3) Si $A_1, A_2, \dots, A_k, \dots$ es cualquier sucesión numerable de eventos en \mathcal{A} , entonces $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$

Propiedad 1 $S \in \mathcal{A}$

Propiedad 2 Si A_1, A_2, \dots, A_N es una sucesión finita de eventos en \mathcal{A} , entonces $\bigcup_{k=1}^N A_k \in \mathcal{A}$

$$\bigcup_{k=1}^N A_k \in \mathcal{A}$$

Propiedad 3 Si $A_1, A_2, \dots, A_k, \dots$ es cualquier sucesión numerable de eventos en \mathcal{A} , entonces $\bigcap_{k=1}^{\infty} A_k \in \mathcal{A}$

Propiedad 4 Si A_1, A_2, \dots, A_N es una sucesión finita de eventos en \mathcal{A} , entonces $\bigcap_{k=1}^N A_k \in \mathcal{A}$

Una familia definida según (\mathcal{A}_1) a (\mathcal{A}_3), contiene a todos los eventos que podamos construir por unión e intersección de conjuntos, o sea \mathcal{A} es "cerrada" bajo estas operaciones.

1.2 Definición Axiomática de Probabilidad (Axiomas de Kolmogorov)

Sea S un espacio muestral asociado a un experimento aleatorio \mathcal{E} , y sea \mathcal{A} una σ -álgebra de eventos en S . Una probabilidad P definida sobre los eventos de S es una función que a cada evento A de la σ -álgebra, le asigna un número real, denotado $P(A)$ y llamado Probabilidad de A , de modo que se satisfacen los axiomas:

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

(2) $P(S) = 1$

(3) Si $A_1, A_2, \dots, A_n, \dots$ es una sucesión de eventos mutuamente excluyentes, entonces

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Propiedades de la Probabilidad

Proposición 1: $P(\emptyset) = 0$

Proposición 2: Si A_1, A_2, \dots, A_N es una sucesión de N eventos mutuamente excluyentes, entonces

$$P\left(\bigcup_{i=1}^N A_i\right) = \sum_{i=1}^N P(A_i)$$

Proposición 3: $P(A) + P(A^c) = 1$

Proposición 4: Si A y B son eventos arbitrarios, entonces:

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

Proposición 5: Si A y B son eventos tales que $A \subseteq B$, entonces:

$$P(A) \leq P(B)$$

Corolario: En el contexto de la proposición anterior, se cumple:

$$0 \leq P(A) \leq P(B) \leq 1$$

1.3 Probabilidad Geométrica

Sea un experimento aleatorio \mathcal{E} consistente en tomar un punto al azar de un conjunto geométrico S que tiene una medida $m(S)$ y sea A un evento del espacio muestral resultante. Si $m(A)$ denota la medida de este evento, entonces la probabilidad de A es:

$$P(A) = \frac{m(A)}{m(S)}$$

1.4 Reglas de Conteo

• Principio de la Multiplicación

Si una 'operación' A puede realizarse u ocurrir de a maneras diferentes y otra 'operación' B puede realizarse de b maneras diferentes, entonces la operación compuesta $A \times B$ consistente en realizar A primero y luego realizar B , se puede realizar de $(a \times b)$ maneras distintas.

• Principio de la Adición

Si una 'operación' A puede realizarse u ocurrir de a maneras diferentes y otra 'operación' B puede realizarse de b maneras diferentes, siendo ambas operaciones excluyentes, entonces la operación compuesta A ó B consistente en realizar A o realizar B pero no ambas, se puede realizar de $(a+b)$ maneras distintas.

Permutaciones y Combinaciones

Factorial de un entero $N! = N \times (N-1) \times (N-2) \times \dots \times 3 \times 2 \times 1$. Adicionalmente definimos $0! = 1!$

Proposición

El número total de Permutaciones de tamaño r tomadas de un conjunto con n elementos distintos es: $P_r^n = n!/(n-r)!$

Proposición

El número total de de Combinaciones (subconjuntos) de tamaño r formadas a partir de un conjunto con n elementos es $C_r^n = n!/(n-r)!r!$

Serie geométrica: Si $-1 < r < 1 \Rightarrow \sum_{k=1}^{\infty} r^k = \frac{r}{1-r}$ y $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$

Binomio de Newton: $(a+b)^n = \sum_{k=0}^n C_k^n a^k b^{n-k}$

1.5 Probabilidad Condicional $P(B|A) = \frac{P(B \cap A)}{P(A)}$ = la probabilidad de B cuando S se reduce al evento A .

Regla del Producto $P(A \cap B) = P(B|A)P(A)$ y $P(A \cap B \cap C) = P(C|A \cap B)P(B|A)P(A)$

1.6 Independencia Dos eventos A y B se dicen Independientes si y sólo se cumple que $P(A \cap B) = P(A)P(B)$

1.7 Probabilidad Total

Sean A_1, A_2, \dots, A_N eventos mutuamente excluyentes, todos con probabilidad positiva y tales que $\bigcup_j A_j = S$. Sea B otro evento de S. Entonces se cumple:

$$P(B) = \sum_j P(B | A_j) P(A_j)$$

1.8 Teorema de Bayes

En el contexto del Teorema de Probabilidad Total, si además $P(B) > 0$, entonces se cumple:

$$P(A_k | B) = \frac{P(B | A_k) P(A_k)}{\sum_j P(B | A_j) P(A_j)} \quad \forall k$$

2. Variable Aleatoria

2.1 Definición Si S es un espacio muestral, una variable aleatoria X, definida sobre S, es una función cuyo dominio es S y cuyo rango es un conjunto de números reales que denotaremos R_X .

2.1.2 Clasificación

Según R_X de una variable aleatoria X es **Continua** si R_X es un intervalo y X es **Discreta** si R_X es un conjunto finito o numerable

2.2 Función de Probabilidad $P_X(x)$

Si X es v.a. discreta, la función de probabilidad de X, es $P_X(x) = P(X=x)$

Propiedades de $P_X(x)$.

- (a) $0 \leq P_X(x) \leq 1 \quad \forall x$
- (b) $\sum_{x \in R_X} P_X(x) = 1$ donde $x \in R_X$ indica que la suma se hace sobre todos los x que pertenecen a R_X
- (c) $P(X \in A) = \sum_{x \in A} P_X(x) \quad \forall A \subseteq R_X$

2.3 Función de Densidad

Si X es una v.a. continua, una función de densidad de X, denotada $f_X(x)$, es una función no negativa y continua, tal que para todo intervalo $[a, b] \subseteq R_X$ se cumple:

$$P(a < X \leq b) = \int_a^b f_X(x) dx$$

Propiedades.

- (a) $0 \leq f_X(x) \quad \forall x$
- (b) $\int_{-\infty}^{+\infty} f_X(x) dx = \int_{R_X} f_X(x) dx = 1$
- (c) $P(X \in A) = \int_A f_X(x) dx$

2.4 Función de Distribución Acumulativa

Si X es una v.a., se define la Función de Distribución Acumulativa de X, denotada F_X , mediante $F_X(x) = P(X \leq x) \quad \forall x$ real

Propiedades de $F_X(x)$

- (1) $0 \leq F_X(t) \leq 1$ para todo t real.
- (2) $a < b \Rightarrow F_X(a) \leq F_X(b)$
- (3) F_X es continua a la derecha ('diestro continua'), ie.

$$F_X(t) = \lim_{h \rightarrow 0^+} F_X(t+h)$$
 para todo t real
- (4) $\lim_{t \rightarrow -\infty} F_X(t) = 0$ y $\lim_{t \rightarrow +\infty} F_X(t) = 1$
- (5) $P[a < X \leq b] = F_X(b) - F_X(a)$
- (6) $P[X = b] = F_X(b) - \lim_{h \rightarrow 0^+} F_X(b+h)$

Propiedades adicionales:

- (1) Si X es discreta, entonces $F_X(b) = \sum_{x_j | x_j \leq b} P_X(x_j)$
- (2) Si X es continua, entonces $\int_{-\infty}^t f_X(x) dx = F_X(t)$ para todo t real
- Si X es discreta con $R_X = \{x_1, x_2, \dots, x_{N-1}, x_N, \dots\}$ donde $x_1 < x_2 < \dots < x_{N-1} < x_N \dots$
Entonces $F_X(x_N) = \sum_{j=1}^N P_X(x_j)$ y también $P_X(x_N) = F_X(x_N) - F_X(x_{N-1})$
- Si X es continua, entonces $f_X(x) = F'_X(x)$

2.5 Valor Esperado o Esperanza Matemática

Sea X variable aleatoria y $H(X)$ una función de X:

$$E[H(X)] = \begin{cases} \sum_{x \in R_X} H(x) P_X(x) & \text{si } X \text{ es Discreta} \\ \int_{-\infty}^{+\infty} H(x) f_X(x) dx & \text{si } X \text{ es Continua} \end{cases}$$

La Media Poblacional $\mu = \mu_X = E(X)$

La Varianza Poblacional $\sigma^2 = \sigma_X^2 = V(X) := E[(X - \mu_X)^2]$

Desigualdad de Tchebychev

X v.a. con media μ_X y desviación estándar σ_X . Sea k una constante positiva dada, entonces:

$$P[|X - \mu_X| < k\sigma_X] \geq 1 - \frac{1}{k^2}$$

Propiedad 1 $E(c) = c$.

Propiedad 2 $E[aH_1(X) + bH_2(X)] = aE[H_1(X)] + bE[H_2(X)]$.

Corolarios

- (1) $V(X) = E(X^2) - \mu_X^2$.
- (2) Si $Y = a + bX \quad \forall X$, entonces $E(Y) = a + bE(X)$ y $V(Y) = b^2V(X)$.

Función Generatriz de Momentos $M_X(t) = E(e^{tX})$

Propiedades

(1) $M_X^{(k)}(0) = E(X^k)$, si existe el valor esperado

(2) X e Y variable aleatorias, entonces $M_X(t) = M_Y(t) \Leftrightarrow F_X = F_Y$

Cálculo del valor esperado por desarrollo asintótico

Sea X v.a. con $E(X) = \mu$ y $V(X) = \sigma^2$. Sea $H(X)$ función al menos dos veces diferenciable en $X = \mu$. Entonces se cumplen:

a) $E[H(X)] \cong H(\mu) + \frac{H''(\mu)}{2} \sigma^2$

b) $V[H(X)] \cong [H'(\mu)]^2 \sigma^2$

Fórmulas sobre las Principales Distribuciones

1. Distribución Binomial $B(x; n, p)$

$$P_X(x) = P(X = x) = C_x^n p^x q^{n-x} \quad \text{donde } x = 0, 1, 2, \dots, n$$

$$E(X) = \mu_X = np; \quad V(X) = \sigma_X^2 = npq \quad \text{y} \quad M_X(t) = (pe^t + q)^n \quad -\infty < t < \infty$$

Uso

X = # de veces que ocurre un determinado evento A sobre un total fijo de n repeticiones u observaciones independientes de un experimento.

Binomio de Newton: $(a+b)^n = \sum_{k=0}^n C_k^n a^k b^{n-k}$

2. Distribución Geométrica $G(x; p)$

$$P_X(x) = P(X = x) = pq^{x-1} \quad x = 1, 2, 3, \dots$$

$$E(X) = \mu_X = 1/p; \quad V(X) = \sigma_X^2 = q/p^2 \quad \text{y} \quad M_X(t) = \frac{pe^t}{1 - qe^t} \quad t < -\ln q.$$

Uso

X = # de veces que se debe repetir un experimento hasta lograr que ocurra un determinado suceso A por primera vez.

Serie geométrica: Si $-1 < r < 1 \Rightarrow \sum_{k=1}^{\infty} r^k = \frac{r}{1-r}$ y $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$

3. Distribución de Pascal o Binomial Negativa $Bn(x; r, p)$

$$P_X(x) = P(X = x) = C_{r-1}^{x-1} p^r q^{x-r} \quad x = r, r+1, r+2, \dots$$

$$E(X) = \mu_X = r/p \quad \text{y} \quad V(X) = \sigma_X^2 = rq/p^2.$$

Uso

X = # de veces que se debe repetir un experimento hasta lograr que ocurra un determinado suceso A por r -ésima vez.

4. Distribución Hipergeométrica $H(x; N, M, n)$

$$P_X(x) = P(X = x) = \frac{C_x^M C_{n-x}^{N-M}}{C_n^N} \quad x = 0, 1, \dots, n$$

$$E(X) = \mu_X = n \frac{M}{N} \quad \text{y} \quad V(X) = \sigma_X^2 = n \left(\frac{M}{N} \right) \left(\frac{N-M}{N} \right) \left(\frac{N-n}{N-1} \right). \quad M_X(t) \text{ es poco útil.}$$

Uso

X = # de casos con una característica A de interés en una muestra de n casos tomados al azar y sin reemplazo de una población de tamaño N de los cuales M en total tienen esa característica.

5. Distribución de Poisson $P(x; \lambda)$

$$P_X(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, 3, \dots$$

$$E(X) = \mu_X = \lambda; \quad V(X) = \sigma_X^2 = \lambda \quad \text{y} \quad M_X(t) = e^{\lambda(e^t - 1)} \quad -\infty < t < \infty$$

Uso

• Si $X \sim B(x; n, p)$, $n \rightarrow \infty$ y $p \rightarrow 0 \Rightarrow np \rightarrow \lambda$, entonces $B(x; n, p) \cong P(x; \lambda)$.

• E evento que se presenta en puntos aleatorios del tiempo (o del espacio), tal que:

- (1) Para todo intervalo de longitud dt suficientemente pequeña, la probabilidad de observar una vez E es proporcional a dt , i.e.:
 $P(E \text{ ocurre una vez en } [t, t+dt]) = w dt \quad \forall t \text{ real } (w > 0).$
- (2) Para todo intervalo de longitud dt suficientemente pequeña, la probabilidad de observar más de una vez E es nula i.e.:
 $P(E \text{ ocurre más de una vez en } [t, t+dt]) = 0.$
- (3) Intervalos disjuntos son independientes en relación a la ocurrencia de E .

Si $t > 0$ es un valor dado y definimos X = # de veces que ocurre E en el intervalo $[0, t] \Rightarrow X \sim P(x; \lambda = wt)$ donde w = tasa de ocurrencias de E por unidad de tiempo.

6. Distribución Exponencial $Exp(x; \beta)$

$$f_X(x) = \beta e^{-\beta x} \quad x > 0$$

$$E(X) = \mu_X = 1/\beta; \quad V(X) = \sigma_X^2 = 1/\beta^2 \quad \text{y} \quad M_X(t) = \left(\frac{\beta}{\beta - t} \right) \quad t < \beta$$

Uso

- Como un modelo en general
- En un proceso de Poisson, si T := Tiempo que transcurre hasta que ocurre E por primera vez, entonces $T \sim Exp(t; \beta = w)$, donde w es la tasa de ocurrencias de E por unidad.

7. Distribución Gamma $\Gamma(x; \alpha, \beta)$

$$f_X(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \quad x > 0$$

$$E(X) = \mu_X = \alpha\beta \quad ; \quad V(X) = \sigma_X^2 = \alpha\beta^2 \quad \text{y} \quad M_X(t) = \frac{1}{(1-\beta t)^\alpha} \quad \text{si} \quad t < \frac{1}{\beta}$$

Uso

- Como un modelo en general

- En un proceso de Poisson, si $T :=$ Tiempo que transcurre hasta que ocurre E por k -ésima vez, entonces $T \sim \Gamma(x; \alpha = k, \beta = 1/w)$

Función matemática Gamma $\Gamma(p) = \int_0^\infty y^{p-1} e^{-y} dy \quad p > 0$

(1) $\Gamma(p) = (p-1)\Gamma(p-1) \quad p > 1$

(2) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

8. Distribución Normal $N(\mu, \sigma^2)$

$$f_X(x) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \quad -\infty < x < \infty$$

$$E(X) = \mu_X = \mu \quad , \quad V(X) = \sigma_X^2 = \sigma^2 \quad \text{y} \quad M_X(t) = e^{i\mu t + \frac{1}{2}\sigma^2 t^2} \quad -\infty < t < \infty$$

Propiedad

$$X \sim N(\mu, \sigma^2) \quad \text{e} \quad Y = a + bX \Rightarrow Y \sim N(a + b\mu, b^2\sigma^2)$$

Uso

- Como un modelo en general
- **Teorema del Límite Central:** $X_1, X_2, \dots, X_n, \dots$ variables aleatorias independientes con medias $\mu_1, \mu_2, \dots, \mu_n, \dots$ y varianzas $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2, \dots$ y $T := \sum_{j=1}^n X_j$

donde el número n de sumandos es grande ($n \geq 30$), $\Rightarrow T \sim N(\mu_T, \sigma_T^2)$, donde

$$\mu_T = \sum_{j=1}^n \mu_j \quad \text{y} \quad \sigma_T^2 = \sum_{j=1}^n \sigma_j^2$$

- $X \sim B(x; n, p)$ y n es "grande" $\Rightarrow P(X \leq k) \approx P(Z \leq \frac{k - np}{\sqrt{npq}})$

9. Distribución Lognormal $\text{LogN}(\mu, \sigma^2)$

$$f_X(x) = \frac{e^{-(\ln x - \mu)^2/2\sigma^2}}{x\sqrt{2\pi}\sigma} \quad x > 0$$

$$X \sim \text{LogN}(\mu, \sigma^2) \Leftrightarrow \ln X \sim N(\mu, \sigma^2)$$

$$\mu_X = E(X) = e^{\mu + \frac{1}{2}\sigma^2} \quad ; \quad \sigma_X^2 = V(X) = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2} \quad \text{y} \quad E(X^t) = e^{t\mu + \frac{t^2}{2}\sigma^2}$$

Uso

- Como un modelo en general
- **Teorema del Límite Central para productos:** $W_1, W_2, \dots, W_n, \dots$ variables aleatorias positivas e independientes tales que existe $E(\ln W_j)$ y $V(\ln W_j) \quad \forall j$ y

sea $W := \prod_{j=1}^n W_j$ donde el número n de factores es grande ($n \geq 30$), \Rightarrow

$$\ln W \sim N(\mu, \sigma^2), \quad \text{donde} \quad \mu = \sum_{j=1}^n E(\ln W_j) \quad \text{y} \quad \sigma^2 = \sum_{j=1}^n V(\ln W_j)$$

10. Distribución Uniforme $U(x; \alpha, \beta)$

$$f_X(x) = \frac{1}{\beta - \alpha} \quad \alpha \leq x \leq \beta$$

$$E(X) = \mu_X = \frac{\alpha + \beta}{2} \quad ; \quad V(X) = \sigma_X^2 = \frac{(\beta - \alpha)^2}{12} \quad . \quad M_X(t) \text{ es poco útil}$$

Uso

- Como un modelo en general
- Se toma un punto al azar de $[\alpha, \beta]$ y $X = \text{Valor obtenido} \Rightarrow X \sim U(x; \alpha, \beta)$
- Si $Y \sim f_Y(y)$ y $X := F_Y(X) \Rightarrow X \sim U(x; 0, 1)$

11. Distribución beta $Beta(x; \alpha, \beta)$

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 \leq x \leq 1$$

$$E(X) = \mu_X = \frac{\alpha}{\alpha + \beta} \quad ; \quad V(X) = \sigma_X^2 = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)} \quad . \quad M_X(t) \text{ es poco útil}$$

Uso

- Como un modelo para proporciones.

Función matemática Beta $B(\alpha, \beta) = \int_0^1 y^{\alpha-1} (1-y)^{\beta-1} dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$

Fórmulas sobre Vector Aleatorio Bidimensional

3.1 Definición y Clasificación

Definición

Un vector aleatorio (X,Y) es un vector cuyas componentes son variables aleatorias X e Y definidas conjuntamente sobre el mismo espacio muestral. El conjunto de posibles parejas (X,Y) se denota R_{XY} y se llama Rango del vector (X,Y)

Clasificación

(X,Y) se dice discreto si sus componentes son v.a. discretas

(X,Y) se dice continuo si sus componentes son v.a. continuas

(X,Y) es mixtos, si tiene una componente discreta y otra continua

3.2 Distribuciones Conjunta, Marginales y Condicionales

3.2.1 Caso discreto

Función de Probabilidad Conjunta

Si (X,Y) es vector aleatorio discreto, la función de probabilidad conjunta de (X,Y) , denotada $P_{XY}(x,y)$, se define mediante:

$$P_{XY}(x,y) = P[(X=x) \cap (Y=y)]$$

Propiedades

(a) $P_{XY}(x,y) \geq 0$

(b) $\sum_x \sum_y P_{XY}(x,y) = 1$

(c) $P[(X,Y) \in A] = \sum_{(x,y) \in A} P_{XY}(x,y)$

Función de Probabilidad Marginal

(X,Y) v.a.d, con función de probabilidad conjunta $P_{XY}(x,y)$, entonces

La Función de Probabilidad Marginal de X es $P_X(x) = \sum_y P_{XY}(x,y)$

La Función de Probabilidad Marginal de Y es $P_Y(y) = \sum_x P_{XY}(x,y)$

Función de Probabilidad Condicional

Función de Probabilidad Condicional de Y dado que $X=x$, denotada $P_{Y|X}(y|x)$ es

$$P_{Y|X}(y|x) = \frac{P_{XY}(x,y)}{P_X(x)} \quad \text{con } x \text{ valor dado tal que } P_X(x) > 0$$

Función de Probabilidad Condicional de X dado que $Y=y$, denotada $P_{X|Y}(x|y)$ es

$$P_{X|Y}(x|y) = \frac{P_{XY}(x,y)}{P_Y(y)} \quad \text{con } y \text{ valor dado tal que } P_Y(y) > 0$$

3.2.2 Caso continuo

Función de Densidad Conjunta

Si (X,Y) es vector aleatorio continuo con rango R_{XY} , la función de densidad conjunta de (X,Y) , denotada $f_{XY}(x,y)$ es una función continua tal que cumple:

$$f_{XY}(x,y) \geq 0; \quad \iint_{R^2} f_{XY}(x,y) = 1 \quad \text{y} \quad P[(X,Y) \in A] = \iint_A f_{XY}(x,y) \quad \forall A \subseteq R_{XY}$$

Nota (Integrales dobles)

1) Recordemos que si una función $f_{XY}(x,y)$ es continua sobre una región $A \subset R^2$ dada por $A = \{(x,y) | a \leq x \leq b, c \leq y \leq d\}$, entonces la integral doble $\iint_A f_{XY}(x,y)$

se calcula como una integral iterada, primero sobre y luego sobre x , o también en el orden inverso (Teorema de Fubini). Esto es:

$$\iint_A f_{XY}(x,y) = \int_a^b \left[\int_c^d f_{XY}(x,y) dy \right] dx = \int_c^d \left[\int_a^b f_{XY}(x,y) dx \right] dy.$$

2) También, si $A \subset R^2$ se puede escribir como una región de fronteras definidas en términos de funciones, como $A = \{(x,y) | a \leq x \leq b, h_1(x) \leq y \leq h_2(x)\}$, entonces:

$$\iint_A f_{XY}(x,y) = \int_a^b \left[\int_{h_1(x)}^{h_2(x)} f_{XY}(x,y) dy \right] dx.$$

3) Análogamente, si $A \subset R^2$ es $A = \{(x,y) | g_1(y) \leq x \leq g_2(y), c \leq y \leq d\}$, entonces:

$$\iint_A f_{XY}(x,y) = \int_c^d \left[\int_{g_1(y)}^{g_2(y)} f_{XY}(x,y) dx \right] dy$$

Nótese finalmente que aunque en la definición se integra sobre todo R^2 , en la práctica la integral es sólo sobre R_{XY} , pues fuera de R_{XY} , $f_{XY}(x,y) = 0$.

Función de Densidad Condicional

Función de Densidad Condicional de Y dado que $X=x$, denotada $f_{Y|X}(y|x)$ es

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \quad \text{donde } x \text{ es valor dado tal que } f_X(x) > 0$$

Función de Densidad Condicional de X dado que $Y=y$, denotada $f_{X|Y}(x|y)$ es

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad \text{donde } y \text{ es valor dado tal que } f_Y(y) > 0$$

3.2.3 Independencia

X e Y v.a. son independientes si:

$$P_{XY}(x,y) = P_X(x)P_Y(y) \quad \forall (x,y) \quad (\text{caso discreto})$$

$$f_{XY}(x,y) = f_X(x)f_Y(y) \quad \forall (x,y) \quad (\text{caso continuo})$$

3.3 Valor Esperado

3.3.1 Definición

Si (X,Y) es vector aleatorio y $H(X,Y)$ es una función de (X,Y) , se define el Valor Esperado de $H(X,Y)$, denotado $E[H(X,Y)]$, mediante

$$E[H(X,Y)] = \begin{cases} \sum_x \sum_y H(x,y) P_{XY}(x,y) & \text{si } (X,Y) \text{ es discreto} \\ \int \int H(x,y) f_{XY}(x,y) dx dy & \text{si } (X,Y) \text{ es continuo} \end{cases}$$

3.3.2 Casos Especiales

• $\mu_X := E[X]$ y $\mu_Y := E[Y]$ (medias poblacionales)

• $\sigma_X^2 := V(X) := E[(X - \mu_X)^2]$ y $\sigma_Y^2 := V(Y) := E[(Y - \mu_Y)^2]$ (varianzas poblacionales)

- $\sigma_{XY} \equiv \text{Cov}(X, Y) := E[(X - \mu_X)(Y - \mu_Y)]$ (Covarianza entre X e Y). Mide asociación lineal entre X e Y

Interpretación de σ_{XY} :

- $\sigma_{XY} > 0$ X e Y están asociadas directamente.
- $\sigma_{XY} < 0$ X e Y están asociadas inversamente.
- Si $\sigma_{XY} = 0$, no hay relación lineal entre X e Y , aunque puede haber una relación no lineal

Coefficiente de Correlación de Pearson $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$, mide lo mismo que σ_{XY} pero carece de unidades y está acotado entre -1 y 1

Propiedades de ρ_{XY}

- $0 \leq |\rho_{XY}| \leq 1$
- $\rho_{XY} > 0$ indica asociación directa o positiva entre X e Y
- $\rho_{XY} < 0$ indica asociación inversa o negativa entre X e Y
- $|\rho_{XY}| = 1 \Leftrightarrow$ Existen constantes α y β tales que $Y = \alpha + \beta X$
- $|\rho_{XY}| \approx 1$ indica que entre X e Y hay asociación (lineal) 'fuerte'
- $|\rho_{XY}| \approx 0$ indica que entre X e Y hay asociación (lineal) 'débil'

3.3.3 Propiedades del Valor Esperado

- $E[C] = C$ para toda constante C
- Si $\alpha_1, \alpha_2, \dots, \alpha_n$ son constantes (o variables no aleatorias) y $H_1(X, Y), H_2(X, Y), \dots, H_n(X, Y)$ son funciones de (X, Y) , entonces $E[\sum_{j=1}^n \alpha_j H_j(X, Y)] = \sum_{j=1}^n \alpha_j E[H_j(X, Y)]$
- Si X e Y son independientes entonces $E[H(X)G(Y)] = E[H(X)]E[G(Y)]$
- $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$
- Si X e Y son independientes, entonces $\text{Cov}(X, Y) = 0$
- Si $W = \alpha X + \beta Y$, donde α y β son constantes, entonces $\mu_W = E[W] = \alpha E[X] + \beta E[Y]$ y $\sigma_W^2 = V[W] = \alpha^2 V[X] + \beta^2 V[Y] + 2\alpha\beta \text{Cov}(X, Y)$

Propiedad:

Sean $T = (X + Y)$ y $D = (X - Y)$ entonces:

- $\mu_T = \mu_X + \mu_Y$ y $\sigma_T^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y$
- $\mu_D = \mu_X - \mu_Y$ y $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho_{XY}\sigma_X\sigma_Y$

3.3.4 Esperanza Condicional $E[Y|X]$

$$E[H(X, Y) | X = x] = \begin{cases} \sum_y H(x, y) P_{Y|X}(y | x) & \text{si } (X, Y) \text{ es discreto} \\ \int_{-\infty}^{+\infty} H(x, y) f_{Y|X}(y | x) & \text{si } (X, Y) \text{ es continuo} \end{cases}$$

Casos especiales

- $\mu_{Y|X} = E[Y | X = x]$, es el valor esperado de Y en la distribución condicional de Y dado $X = x$. También se llama Función de Regresión de Y sobre X y

suele ser una función de X que se usa para pronosticar el valor de Y cuando se conoce el valor de X .

- $\sigma_{Y|X}^2 = V[Y | X = x] = E[Y^2 | X = x] - (E[Y | X = x])^2$, es la varianza condicional y mide la variación de Y alrededor de la función de regresión $E[Y | X] = \varphi(X)$

Propiedad

- $E(E[Y | X]) = E[\varphi(X)] = E[Y]$
- $V(Y) = E(V[Y | X]) + V(E[Y | X])$

3.3.5 Vector de Medias y Matriz de Varianza Covarianza

Vector de Medias y Matriz de Varianza-Covarianza

Sea $\begin{pmatrix} X \\ Y \end{pmatrix}$ vector aleatorio columna, el vector de medias de $\begin{pmatrix} X \\ Y \end{pmatrix}$ es $\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$ y

la matriz de varianza covarianza de $\begin{pmatrix} X \\ Y \end{pmatrix}$ es $\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}$

Nota:

Se puede definir en general $E\left[\begin{pmatrix} X \\ Y \end{pmatrix}\right] = \begin{pmatrix} E(X) \\ E(Y) \end{pmatrix}$ y si $M = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}$ es matriz

aleatoria, definimos $E(M) = \begin{pmatrix} E(X_{11}) & E(X_{12}) \\ E(X_{21}) & E(X_{22}) \end{pmatrix}$.

Propiedad

- $\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix} = E\left[\begin{pmatrix} X - E(X) \\ Y - E(Y) \end{pmatrix} \begin{pmatrix} X - E(X) & Y - E(Y) \end{pmatrix}\right]$
- $U_{2 \times 1} = \begin{pmatrix} X \\ Y \end{pmatrix}$ vector aleatorio con vector de medias μ_U y matriz de varianza

covarianza Σ_U ; sean $A_{2 \times 2} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ y $B_{2 \times 1} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ matrices dadas. Si

definimos el vector $V_{2 \times 1} = AU + B$ (o sea $V_{2 \times 1} = \begin{pmatrix} W \\ Z \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$),

entonces el vector de medias de V es $\mu_V = A\mu_U + B$ y su matriz de varianza covarianza es $\Sigma_V = A\Sigma_U A^T$ respectivamente.

3.4 Vector Aleatorio n-dimensional

Definición

Un vector cuyas componentes son n variables aleatorias definidas conjuntamente

- Un vector aleatorio n-dimensional (fila) es de la forma $\mathbf{X}_{1,m} = (X_1, X_2, \dots, X_n)$
- Un vector aleatorio n-dimensional (columna) es de la forma $\mathbf{X}_{n,1} = (X_1, X_2, \dots, X_n)^T$.

Matriz aleatoria

Una matriz aleatoria M de orden $n \times m$ es una matriz $M = (M_{ij})_{n \times m}$ de $n \times m$ variables aleatorias M_{ij} $i=1,2,\dots,n$ $j=1,2,\dots,m$ definidas conjuntamente sobre el mismo espacio muestral S .

Vector particionado

Cuando se tiene un vector aleatorio $X_{n,1} = (X_1, X_2, \dots, X_n)^T$, a veces es útil separar partes de él, por ejemplo $X_{n,1} = (X_1, \dots, X_p, X_{p+1}, \dots, X_n)^T$ donde las primeras p componentes forman un subvector $X_{p,1} = (X_1, \dots, X_p)^T$ de orden $p \times 1$ y el resto de $(n-p)$ componentes un subvector $X_{(n-p),1} = (X_{p+1}, \dots, X_n)^T$. En este contexto se escribe $X_{n,1} = (X_1 : X_2)^T$ donde $X_1 = (X_1, \dots, X_p)^T$ y $X_2 = (X_{p+1}, \dots, X_n)^T$. O más explícitamente, en formato de vector columna:

$$X_{n,1} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \\ \cdots \\ X_{p+1} \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} X_1 \\ \cdots \\ X_2 \end{pmatrix} \text{ donde } X_1 = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} \text{ y } X_2 = \begin{pmatrix} X_{p+1} \\ \vdots \\ X_n \end{pmatrix}.$$

3.4.1 Distribuciones

Las nociones de distribución conjunta $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$; distribuciones marginales $f_{X_1}(x_1), f_{X_2}(x_2), \dots, f_{X_n}(x_n)$; y condicionales $f_{X_i|X_j}(x_i|x_j)$, son extensiones directas del caso bivariado. En particular es importante:

Definición de Independencia.

X_1, X_2, \dots, X_n se dicen **independientes** si:

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n) \equiv \prod_{j=1}^n f_{X_j}(x_j)$$

Nota

En verdad, se necesita que esta regla se cumpla con todos los subconjuntos de componentes del vector, por ejemplo, debe cumplirse que

$$f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2); \quad f_{X_1, X_2, X_3}(x_1, x_2, x_3) = f_{X_1}(x_1) f_{X_2}(x_2) f_{X_3}(x_3); \text{ etc.}$$

En términos de vector particionado se escribe

- La función de distribución conjunta $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \equiv f_{X_{n,1}}(x_{n,1}) \equiv f_{X_1, X_2}(x_1, x_2)$, donde en la última expresión $X_1 \in R^p$ y $X_2 \in R^{n-p}$ son subvectores
- La distribución marginal del subvector $X_1 \in R^p$ como $f_{X_1}(x_1, \dots, x_p) \equiv f_{X_{p,1}}(x_{p,1}) \equiv f_{X_1}(x_1)$
- La distribución marginal del subvector $X_2 \in R^{n-p}$ como $f_{X_2}(x_{p+1}, \dots, x_n) \equiv f_{X_{(n-p),1}}(x_{(n-p),1}) \equiv f_{X_2}(x_2)$
- La distribución condicional del subvector $X_1 \in R^p$, dado el subvector $X_2 \in R^{n-p}$, como $f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)}$
- $X_1 \in R^p$ y $X_2 \in R^{n-p}$ son independientes si $f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \quad \forall x_1, \forall x_2$

3.4.2 Valor Esperado

- **Media Poblacional** de la componente X_j de \mathbf{X} $\mu_j \equiv \mu_{X_j} := E[X_j]$
- **Varianza Poblacional** de la componente X_j de \mathbf{X} $\sigma_j^2 \equiv \sigma_{X_j}^2 := V[X_j] = E[(X_j - \mu_j)^2]$
- **Covarianza Poblacional** entre las componentes X_i y X_j de \mathbf{X} : $\sigma_{ij} \equiv \sigma_{X_i, X_j} \equiv \text{Cov}(X_i, X_j) := E[(X_i - \mu_i)(X_j - \mu_j)]$
- **Coefficiente de Correlación Lineal de Pearson** entre X_i y X_j $\rho_{ij} \equiv \rho_{X_i, X_j} := \frac{\sigma_{ij}}{\sigma_i \sigma_j}$
- **Vector de Medias** del vector (columna) $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$:

$$\mu_{\mathbf{X}} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}_{n \times 1}$$

- **Matriz de Varianza-covarianza** del vector (columna) $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$:

$$\Sigma_{\mathbf{X}} = (\sigma_{ij})_{n \times n} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}_{n \times n}$$

- **Varianza Total** $VT := \text{tr}(\Sigma_{\mathbf{X}}) = \sum_{j=1}^n \sigma_j^2$
- **Varianza Generalizada** $VG := \text{Det}(\Sigma_{\mathbf{X}}) = |\Sigma_{\mathbf{X}}|$

- **Vector de Medias del vector particionado** $X_{n,1} = \begin{pmatrix} X_1 \\ \cdots \\ X_2 \end{pmatrix}$. Se define mediante

$$\mu_{\mathbf{X}} = E(X_{n,1}) = \begin{pmatrix} E(X_1) \\ \cdots \\ E(X_2) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \cdots \\ \mu_2 \end{pmatrix} \text{ donde } \mu_1 = E(X_1) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix} \text{ y } \mu_2 = E(X_2) = \begin{pmatrix} E(X_{p+1}) \\ \vdots \\ E(X_{n-p}) \end{pmatrix}$$

• **Matriz de Varianza-covarianza del vector particionado** $X_{n \times 1} = \begin{pmatrix} X_1 \\ \vdots \\ X_2 \end{pmatrix}$

$$\Sigma_X = (\sigma_{ij})_{n \times n} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}_{n \times n} \text{ donde } \Sigma_{11} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22}^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}, \quad \Sigma_{12} = \begin{pmatrix} \sigma_{1(p+1)} & \cdots & \sigma_{1n} \\ \sigma_{2(p+1)} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots \\ \sigma_{p(p+1)} & \cdots & \sigma_{pn} \end{pmatrix}$$

$$\Sigma_{21} = \begin{pmatrix} \sigma_{(p+1)1} & \cdots & \sigma_{(p+1)p} \\ \sigma_{(p+2)1} & \cdots & \sigma_{(p+2)p} \\ \vdots & \vdots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{np} \end{pmatrix} \text{ y } \Sigma_{22} = \begin{pmatrix} \sigma_{(p+1)(p+1)}^2 & \sigma_{(p+1)(p+2)} & \cdots & \sigma_{(p+1)n} \\ \sigma_{(p+2)(p+1)} & \sigma_{(p+2)(p+2)}^2 & \cdots & \sigma_{(p+2)n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n(p+1)} & \sigma_{n(p+2)} & \cdots & \sigma_{nn}^2 \end{pmatrix}. \text{ Nótese que } \Sigma_{11} \text{ es}$$

la matriz de varianza-covarianza de X_1 , Σ_{22} la matriz de varianza-covarianza de X_2 y $\Sigma_{12} = \Sigma_{21}^T$ es la matriz de covarianzas entre X_1 y X_2 .

Valor esperado de una matriz aleatoria

Sea M una matriz aleatoria de orden $n \times m$, $E(M) = (E(M_{ij}))_{n \times m}$

Propiedad

Si $X_{n \times 1}$ es vector columna con vector de medias μ_X y matriz de varianza covarianza Σ_X , entonces se cumple que $\Sigma_X = E[(X - \mu_X)(X - \mu_X)^T]$.

Propiedad

Si $X_1 \in R^p$ y $X_2 \in R^{n-p}$ son vectores aleatorios columna con vectores de medias μ_1 y μ_2 respectivamente, entonces la matriz de covarianza Σ_{12} satisface $\Sigma_{12} = E[(X_1 - \mu_1)(X_2 - \mu_2)^T]$. Análogamente $\Sigma_{21} = E[(X_2 - \mu_2)(X_1 - \mu_1)^T]$

Propiedad

Σ_X es simétrica y semidefinida positiva, esto es $Y^T \Sigma_X Y \geq 0$ para todo vector no nulo $Y \in R^n$

Proposición

Sea \bar{X} vector aleatorio y sean $H_1(\bar{X}), H_2(\bar{X}), \dots, H_k(\bar{X})$ k funciones reales de \bar{X} (esto es $H_j: R^n \rightarrow R$). Sean $\alpha_1, \alpha_2, \dots, \alpha_k$ constantes (o también variables pero no aleatorias), entonces se cumple $E[\sum_{j=1}^k \alpha_j H_j(\bar{X})] = \sum_{j=1}^k \alpha_j E[H_j(\bar{X})]$

Proposición

Sea \bar{X} vector aleatorio y sean $H_1(\bar{X}), H_2(\bar{X}), \dots, H_k(\bar{X})$ k funciones vectoriales de \bar{X} (esto es $H_j: R^n \rightarrow R^m$). Sean $\alpha_1, \alpha_2, \dots, \alpha_k$ constantes (o también variables, pero no aleatorias), entonces se cumple $E[\sum_{j=1}^k \alpha_j H_j(\bar{X})] = \sum_{j=1}^k \alpha_j E[H_j(\bar{X})]$

Proposición

Sea \bar{X} vector aleatorio n -dimensional $\bar{X}_{1 \times n} = (X_1, X_2, \dots, X_n)$ de componentes independientes entonces $E[\prod_{j=1}^n H_j(X_j)] = \prod_{j=1}^n E[H_j(X_j)]$

Corolario

Sea \bar{X} vector aleatorio n -dimensional $\bar{X}_{1 \times n} = (X_1, X_2, \dots, X_n)$ tal que sus componentes son independientes, entonces se cumple $\sigma_{ij} = \text{Cov}(X_i, X_j) = 0$

Proposición

Sea \bar{X} vector aleatorio n -dimensional $\bar{X}_{1 \times n} = (X_1, X_2, \dots, X_n)$ y sean $\alpha_1, \alpha_2, \dots, \alpha_n$ constantes (o también variables no aleatorias). Entonces:

$$\bullet \quad E[\sum_{j=1}^n \alpha_j X_j] = \sum_{j=1}^n \alpha_j E[X_j] \quad \text{y} \quad V[\sum_{j=1}^n \alpha_j X_j] = \sum_{j=1}^n \alpha_j^2 V[X_j] + 2 \sum_{i < j} \alpha_i \alpha_j \text{Cov}(X_i, X_j)$$

Corolario

Si X_1, X_2, \dots, X_n son independientes, entonces $E[\sum_{j=1}^n \alpha_j X_j] = \sum_{j=1}^n \alpha_j E[X_j]$ y

$$V[\sum_{j=1}^n \alpha_j X_j] = \sum_{j=1}^n \alpha_j^2 V[X_j]. \text{ En particular } E[\sum_{j=1}^n X_j] = \sum_{j=1}^n E[X_j] \quad \text{y} \quad V[\sum_{j=1}^n X_j] = \sum_{j=1}^n V[X_j]$$

Nota:

Si X_1, X_2, \dots, X_n son independientes, entonces $E(\prod_{j=1}^n X_j) = \prod_{j=1}^n E(X_j)$ pero no ocurre lo

mismo con las varianzas, esto es: $V(\prod_{j=1}^n X_j) \neq \prod_{j=1}^n V(X_j)$

Proposición

Si $A_{n \times m}$ es matriz constante (o no aleatoria) entonces $E(A_{n \times m}) = A_{n \times m}$

Proposición

Si $A_{n \times q}$ es matriz de constante (o no aleatoria) y $M_{q \times m}$ es matriz aleatoria entonces $E(A_{n \times q} M_{q \times m}) = A_{n \times q} E(M_{q \times m})$. O en notación más simple $E(AM) = AE(M)$.

Proposición

Si $C_{p \times m}$ es matriz de constante (o no aleatoria) y $M_{n \times p}$ es matriz aleatoria entonces $E(M_{n \times p} C_{p \times m}) = E(M_{n \times p}) C_{p \times m}$. O en notación más simple $E(MC) = E(M)C$.

En general podemos escribir $E(AMC + B) = AE(M)C + B$ donde A, B y C son matrices no aleatorias y M es matriz aleatoria.

Proposición

Sea $\bar{X}_{m \times 1}$ vector columna con vector de medias $\mu_{\bar{X}}$ y matriz de varianza covarianza $\Sigma_{\bar{X}}$. Sea $\bar{Y}_{m \times 1}$ otro vector tal que $\bar{Y}_{m \times 1} = A_{m \times n} \bar{X}_{n \times 1} + B_{m \times 1}$ donde $A_{m \times n}$ y $B_{m \times 1}$ son no aleatorias. Entonces $\mu_{\bar{Y}} = A \mu_{\bar{X}} + B$ y $\Sigma_{\bar{Y}} = A \Sigma_{\bar{X}} A^T$

Propiedad

Si $\bar{X}_{m \times 1}$ tiene componentes con distribución normal con vector de medias $\mu_{\bar{X}}$ y matriz de varianza covarianza $\Sigma_{\bar{X}}$ y $\bar{Y}_{m \times 1} = A_{m \times n} \bar{X}_{n \times 1} + B_{m \times 1}$ entonces $\bar{Y}_{m \times 1}$ tiene componentes con distribución normal con $\mu_{\bar{Y}} = A \mu_{\bar{X}} + B$ y $\Sigma_{\bar{Y}} = A \Sigma_{\bar{X}} A^T$

Fórmulas sobre Muestreo y Estadísticas

4.1.1 Población

Si $X \sim f_X(x)$ variable aleatoria con rango R_X . La Población de X se define como el conjunto $\{(x, f_X(x)) | x \in R_X\}$.

4.1.2 Muestra Aleatoria

Si X es variable aleatoria, una muestra aleatoria de tamaño n (m.a.) es un vector aleatorio n -dimensional (X_1, X_2, \dots, X_n) cuyas componentes representan el proceso de repetir n veces, y de manera independiente, el experimento aleatorio que genera a X , registrando los valores obtenidos.

Se cumple

- $R_X = R_{X_j}; f_{X_j}(x_j) = f_X(x_j)$ y $F_{X_j}(x_j) = F_X(x_j) \quad \forall j$
- $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_n}(x_n) = f_X(x_1) f_X(x_2) \dots f_X(x_n) \equiv \prod_{j=1}^n f_X(x_j)$
- X_1, X_2, \dots, X_n son independiente e idénticamente distribuidas, lo que se denota i.i.d.

4.1.3 Estadística

Es una función $h(X_1, X_2, \dots, X_n)$ que sólo depende de las componentes de la muestra aleatoria (X_1, X_2, \dots, X_n) .

4.2 Propiedades de estadísticas importantes

- $\bar{X} := \frac{\sum_{j=1}^n X_j}{n}$ es la media muestral
- $S^2 := \frac{\sum_{j=1}^n (X_j - \bar{X})^2}{n-1} = \frac{\sum_{j=1}^n X_j^2 - n\bar{X}^2}{n-1}$ es la varianza muestral

Propiedad 1

Si (X_1, X_2, \dots, X_n) es m.a. tomada de la población de una v.a. X con media μ y varianza σ^2 , entonces $E[\bar{X}] = \mu$ y $V[\bar{X}] = \frac{\sigma^2}{n}$

Observación:

$$E(M_k) = m_k \text{ y } V(M_k) = \frac{m_{2k} - (m_k)^2}{n}$$

Propiedad 2

$$E(S^2) = \sigma^2$$

Propiedad 3 (Ley de los Grandes Números para \bar{X})

Si (X_1, X_2, \dots, X_n) es m.a. tomada de la población de una v.a. X con media μ y varianza σ^2 , entonces $\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| \leq \varepsilon) = 1$

Propiedad 4 (Teorema Central del Límite para \bar{X})

Sea (X_1, X_2, \dots, X_n) m.a. tomada de la población de una v.a. X con media μ y varianza σ^2 finitas, entonces si n es suficientemente grande ($n \geq 30$), se cumple que $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma} \right) \sim N(0,1)$

4.3 Distribuciones Asociadas al Muestreo de una Distribución Normal

4.3.1 Distribución de \bar{X}

Sea (X_1, X_2, \dots, X_n) m.a. tomada de la población de una v.a. $X \sim N(\mu, \sigma^2)$. En este contexto se cumple que $\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \forall n$

4.3.2 Distribución asociada a S^2 .Distribución Ji-Cuadrado $\chi^2(k)$ y propiedades

Sea W v.a. con distribución Gamma $\Gamma(\alpha = \frac{k}{2}, \beta = 2)$ donde k es entero positivo, diremos entonces que W tiene distribución Ji-Cuadrado de parámetro k . Este caso particular de la Gamma, se denota $\chi^2(k)$, i.e. $\chi^2(k) \equiv \Gamma(\alpha = \frac{k}{2}, \beta = 2)$. El entero k es el único parámetro y se llama "Grados de libertad" de la distribución. Como se trata de un caso particular de la Gamma, tiene todas sus propiedades:

- $f_W(w) = \Gamma(\alpha = \frac{k}{2}, \beta = 2) = \chi^2(k) = \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} e^{-w/2} w^{\frac{k}{2}-1} \quad w > 0$
- $E[W] = \alpha\beta = \frac{k}{2} \cdot 2 = k, \quad V[W] = \alpha\beta^2 = \frac{k}{2} \cdot 2^2 = 2k \text{ y } M_W(t) = \left(\frac{1}{1-2t} \right)^{\frac{k}{2}} \text{ si } t < \frac{1}{2}$

Propiedad 1 (Relación con la distribución normal estándar)

Si $Z \sim N(0,1)$ y definimos $W = Z^2$, entonces $W \sim \chi^2(k=1)$

Propiedad 2 (Propiedad Reproductiva)

Si $W_1 \sim \chi^2(k_1)$ y $W_2 \sim \chi^2(k_2)$ son independientes entonces $(W_1 + W_2) \sim \chi^2(k = k_1 + k_2)$

Distribución asociada a S^2

Si de una distribución normal $N(\mu, \sigma^2)$ se toma una m.a. (X_1, X_2, \dots, X_n) de tamaño n y $W = (n-1)S^2 / \sigma^2$, entonces $W = (n-1)S^2 / \sigma^2 \sim \chi^2(k = n-1)$

Distribución de S^2

Si de una distribución normal $N(\mu, \sigma^2)$ se toma una m.a. (X_1, X_2, \dots, X_n) de tamaño n , entonces $S^2 \sim \Gamma(\alpha = \frac{n-1}{2}, \beta = \frac{2\sigma^2}{n-1})$

Corolario:

$$E[S^2] = \alpha\beta = \left(\frac{n-1}{2} \right) \left(\frac{2\sigma^2}{n-1} \right) = \sigma^2 \text{ y } V[S^2] = \alpha\beta^2 = \left(\frac{n-1}{2} \right) \left(\frac{2\sigma^2}{n-1} \right)^2 = \frac{2\sigma^4}{n-1}$$

4.4 Otras distribuciones importantes.

4.4.1 Distribución t de Student

Definición (Variable T de Student)

Sean $Z \sim N(0,1)$ y $W_1 \sim \chi^2(k)$ independientes. Se define la variable T , llamada

variable t de Student, mediante: $T = \frac{Z}{\sqrt{\frac{W}{k}}}$

Proposición (Distribución t de Student)

La función de densidad de T es $f_T(t) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})} (1 + \frac{t^2}{k})^{-\frac{k+1}{2}} \quad -\infty < t < \infty$

siendo k parámetro de la distribución.

Propiedad

Si de una distribución normal $N(\mu, \sigma^2)$ se toma una m.a. (X_1, X_2, \dots, X_n) de tamaño

n y se define la variable t mediante $t = \frac{(\bar{X} - \mu) / \frac{S}{\sqrt{n}}}{\sqrt{\frac{S^2}{n}}}$ entonces $t \sim t(k=n-1)$

4.4.2 Distribución F de Fisher

Definición (Variable F)

Sean V con distribución $\chi^2(k_1)$ y W con distribución $\chi^2(k_2)$, variables independientes. Se define la variable F mediante $F = \frac{V/k_1}{W/k_2}$.

Proposición (Distribución F de Fisher)

La función de densidad de F es $g_F(f) = \frac{\Gamma(\frac{k_1+k_2}{2})}{\Gamma(\frac{k_1}{2})\Gamma(\frac{k_2}{2})} k_1^{\frac{k_1}{2}} k_2^{\frac{k_2}{2}} \frac{f^{\frac{k_1}{2}-1}}{(k_1+k_2f)^{\frac{k_1+k_2}{2}}}, f > 0$

k_1 es llamado "grados de libertad del numerador" y k_2 es llamado "grados de libertad del denominador".

Notación: $F \sim F(k_1, k_2)$

Propiedad

F tiene distribución $F(k_1, k_2)$ si y sólo si $1/F$ tiene distribución $F(k_2, k_1)$.

Propiedad

Sean X_1 con distribución $N(\mu_1, \sigma_1^2)$ e Y_1 con distribución $N(\mu_2, \sigma_2^2)$ y sean S_1^2 y S_2^2 varianzas de respectivas muestras independientes de tamaños n y m .

Entonces $F := \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1)$.

Propiedad

Si $T \sim t(k)$, entonces $T^2 \sim F(1, k)$

Fórmulas sobre Estimación Puntual

5.1 Definiciones básicas

Sea $X \sim f_X(x; \theta)$ donde θ es parámetro o vector de parámetros de valor desconocido.

Espacio Paramétrico Θ es el conjunto de valores posibles para θ

Espacio de Información X es el conjunto de todas las muestras posibles de tamaño n (X_1, X_2, \dots, X_n)

5.2 Estimador y Valor Estimado

Estimador

Un estimador de un parámetro θ es una estadística $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ cuyo valor una vez tomada la m.a. se usa como aproximación de θ

Valor Estimado o Estimación

Es un valor particular del estimador $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$

5.3 Propiedades de un Buen Estimador

5.3.1 Insesgamiento

$\hat{\theta}$ se dice Insesgado si $E(\hat{\theta}) = \theta \quad \forall \theta \in \Theta$

5.3.2 Eficiencia

Sean $\hat{\theta}$ y $\tilde{\theta}$ dos estimadores insesgados del mismo parámetro θ . Diremos que $\hat{\theta}$ es más eficiente que $\tilde{\theta}$ si $V(\hat{\theta}) < V(\tilde{\theta})$

Mejor Estimador Lineal Insesgado (MELI)

$\hat{\theta}$ estimador de θ se dice Mejor Estimador Lineal Insesgado (MELI) de θ si:

- (1) $\hat{\theta} = \sum_{j=1}^n \alpha_j X_j$
- (2) $E[\hat{\theta}] = \theta \quad \forall \theta \in \Theta$
- (3) $\hat{\theta}$ es de **varianza mínima** en relación a cualquier otro estimador lineal e insesgado de θ

Construir el MELI equivale a resolver el Problema de optimización:

$\text{Mín} V(\sum_{j=1}^n \alpha_j X_j) \quad \text{s.a.} \quad E(\sum_{j=1}^n \alpha_j X_j) = \theta \quad \forall \theta \in \Theta$ donde las incógnitas son las constantes $\alpha_1, \alpha_2, \dots, \alpha_n$.

5.3.3 Consistencia

Sea (X_1, X_2, \dots, X_n) m.a. de tamaño n y sea $\hat{\theta}$ un estimador de θ . Diremos que $\hat{\theta}$

es Consistente si $\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| \leq \varepsilon) = 1 \quad \forall \varepsilon > 0$

En sentido estricto, como para cada n hay un estimador $\hat{\theta}_n$, debiera escribirse $\hat{\theta}_n$ y decirse que " $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \dots, X_n)$ es estimador consistente de θ si

$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| \leq \varepsilon) = 1 \quad \forall \varepsilon > 0$ "

Proposición

Si $\hat{\theta}$ es un estimador de θ tal que

- (1) $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$ ($\hat{\theta}$ es "asintóticamente insesgado") y
- (2) $\lim_{n \rightarrow \infty} V(\hat{\theta}) = 0$ ($\hat{\theta}$ es "asintóticamente eficiente")

Entonces $\hat{\theta}$ es un estimador consistente de θ .

Definición (Límite en Probabilidad)

Sea $\{W_n\}$ una sucesión de variables aleatorias y sea C una constante. Se dice que C es el límite en probabilidad de $\{W_n\}$, lo que se denota escribiendo

$$Plim W_n = C, \text{ si } \lim_{n \rightarrow \infty} P(|W_n - C| \leq \varepsilon) = 1 \quad \forall \varepsilon > 0$$

Propiedades:

Sean $\{W_n\}$ y $\{V_n\}$ sucesiones tales que existen $Plim W_n$ y $Plim V_n$ entonces

1. $Plim(W_n \pm V_n) = Plim W_n \pm Plim V_n$
2. $Plim(W_n \times V_n) = Plim W_n \times Plim V_n$
3. $Plim(W_n / V_n) = Plim W_n / Plim V_n$ si $Plim V_n \neq 0$

Proposición (Teorema de Slutsky)

Sea $\{W_n\}$ sucesión tal que existe $Plim W_n$ y sea $g(t)$ una función continua de t . Entonces $Plim g(W_n) = g(Plim W_n)$

Nota: $\hat{\theta}$ es estimador consistente de θ si y sólo si $Plim \hat{\theta} = \theta$.

5.4 Métodos de Estimación

5.4.1 Método de Momentos

Definiciones Básicas

Sean $X \sim f_X(x; \theta)$ una v.a. y $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ vector de parámetros por estimar. La forma de $f_X(x; \theta)$ no necesariamente es conocida.

Momentos Poblacionales

El k-ésimo momento poblacional m_k se define mediante $m_k = E(X^k)$, si existe el correspondiente valor esperado

Momentos Muestrales

Si (X_1, X_2, \dots, X_n) es m.a. tomada de la población de X , el k-ésimo momento

$$\text{muestral, } M_k \text{ se define mediante } M_k = \frac{\sum_{j=1}^n X_j^k}{n}$$

Propiedad

M_k es estimador insesgado de m_k y si existe m_{2k} , además M_k es estimador Consistente de m_k

Definición

Si m_k es función de $\theta_1, \theta_2, \dots, \theta_p$, o sea $m_k = h_k(\theta_1, \theta_2, \dots, \theta_p)$ $k=1, 2, \dots, p$, de modo que se cumple:

$$\begin{aligned} m_1 &= h_1(\theta_1, \theta_2, \dots, \theta_p) \\ m_2 &= h_2(\theta_1, \theta_2, \dots, \theta_p) \\ &\vdots \\ m_p &= h_p(\theta_1, \theta_2, \dots, \theta_p) \end{aligned}$$

Diremos que $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p$ estimadores de $\theta_1, \theta_2, \dots, \theta_p$ respectivamente, son **estimadores obtenidos mediante el Método de Momentos**, si son solución al sistema de ecuaciones:

$$\begin{aligned} M_1 &= h_1(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p) \\ M_2 &= h_2(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p) \\ &\vdots \\ M_p &= h_p(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p) \end{aligned}$$

5.4.2 Método de Máxima Verosimilitud

Función de Verosimilitud

Sea $X \sim f_X(x; \theta)$ donde θ es parámetro o vector de parámetros de valor desconocido, definimos la función de verosimilitud de θ , $L(\theta)$ como la distribución conjunta de la muestra vista como función del parámetro θ . Es

$$\text{decir } L(\theta) = \prod_{j=1}^n f_{X_j}(x_j; \theta) \quad \theta \in \Theta$$

Estimador de Máxima Verosimilitud

El estimador de θ obtenido por el Método de Máxima Verosimilitud, es el estimador $\hat{\theta}$ que **maximiza** $L(\theta)$ o sea $\hat{\theta}$ es solución al problema:

$$\underset{\theta}{\text{Máx}} L(\theta) \quad \text{s.a } \theta \in \Theta$$

Nota:

$L(\theta)$ y $\ln(L(\theta))$ tienen los mismos puntos críticos, pero $\ln(L(\theta))$ suele tener una estructura más simple, por eso **es común obtener el estimador MV de θ maximizando $\ln(L(\theta))$ en lugar de $L(\theta)$** . La función $l(\theta) = \ln(L(\theta))$ se llama "función logverosimilitud"

5.4.3 Método de Mínimos Cuadrados

Supongamos a Y , X y ε tales que X es variable observable no aleatoria ("variable matemática"), Y es variable aleatoria observable y ε es variable aleatoria no observable. Supongamos que estas variables están relacionadas mediante la ecuación $Y = \varphi(X; \theta) + \varepsilon$, donde $\varphi(X; \theta)$ es función bien especificada (de forma conocida) y θ es un parámetro o vector de parámetros por estimar. **La función $\varphi(X; \theta)$ puede considerarse una función de "enlace" entre la componente aleatoria Y del modelo y el residuo no sistemático y aleatorio ε** . También se dice que $\varphi(X; \theta)$ es la "componente sistemática" del modelo y ε la "componente aleatoria" del mismo.

Supuestos Clásicos

Dado el modelo $Y = \varphi(X; \theta) + \varepsilon$ y una m.a. de n parejas de observaciones $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ que satisfacen la correspondiente relación $Y_j = \varphi(X_j; \theta) + \varepsilon_j$, asumiremos que:

- (1) $E(\varepsilon_j) = 0 \quad \forall j$
- (2) $V(\varepsilon_j) = \sigma^2 \quad \forall j$ (Homogeneidad de varianzas u Homocedasticidad)
- (3) $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$ (No autocorrelación)
- (4) $\varphi(X; \theta)$ es una función sin error de especificación (de forma conocida y correcta)
- (5) X_j es variable no aleatoria o siéndolo tiene sus valores ya dados.

Definición (Estimador de Mínimos Cuadrados Ordinarios MCO o LS)

Bajo los supuestos (1) a (5), definimos el estimador $\hat{\theta}$ obtenido mediante el Método de Mínimos Cuadrados Ordinarios (MCO) como el valor $\hat{\theta}$ que es solución al problema:

$$\underset{\theta}{\text{Min}} Q(\theta) = \sum_{j=1}^n [Y_j - \varphi(X_j; \theta)]^2 \quad \text{s.a. } \theta \in \Theta$$

Modelo de Regresión Lineal sin intercepto $Y_j = \theta X_j + \varepsilon_j$

$$\hat{\theta} = \frac{\sum_{j=1}^n X_j Y_j}{\sum_{j=1}^n X_j^2} \quad \text{es el estimador MCO de } \theta$$

Modelo de Regresión Lineal con intercepto $Y_j = \alpha + \beta X_j + \varepsilon_j$

Los estimadores MCO de β y α son respectivamente

$$\hat{\beta} = \frac{\sum_{j=1}^n X_j Y_j - n \bar{X} \bar{Y}}{\sum_{j=1}^n X_j^2 - n \bar{X}^2} \quad \text{y}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$$

Fórmulas alternativas para $\hat{\beta}$ y $\hat{\alpha}$

$$\hat{\beta} = \frac{\sum_{j=1}^n X_j Y_j - n \bar{X} \bar{Y}}{\sum_{j=1}^n X_j^2 - n \bar{X}^2} = \frac{\sum_{j=1}^n (X_j - \bar{X}) Y_j}{\sum_{j=1}^n (X_j - \bar{X})^2} = \frac{\sum_{j=1}^n (X_j - \bar{X}) (Y_j - \bar{Y})}{\sum_{j=1}^n (X_j - \bar{X})^2} = \frac{\sum_{j=1}^n \left(\frac{(X_j - \bar{X})}{\sum_{j=1}^n (X_j - \bar{X})^2} \right) Y_j}{1}$$

$$\hat{\alpha} = \frac{\bar{Y} \sum_{j=1}^n X_j^2 - \bar{X} \sum_{j=1}^n X_j Y_j}{\sum_{j=1}^n X_j^2 - n \bar{X}^2}$$

Proposición (Teorema de Gauss-Markov)

En el modelo $Y_j = \alpha + \beta X_j + \varepsilon_j$ y bajo los supuestos clásicos, los estimadores MCO $\hat{\alpha}$ y $\hat{\beta}$ son MELI de α y β respectivamente.

Observación: Mejor Estimador Lineal Afín

Un estimador de forma $c_0 + \sum_{j=1}^n c_j Y_j$ ("función lineal afín") donde c_0 se convierte en una incógnita más del problema. Por lo general en el MELI con una función lineal afín, resulta que $c_0 = 0$ y por tanto es irrelevante la distinción entre función lineal $\sum_{j=1}^n c_j Y_j$ y función lineal afín $c_0 + \sum_{j=1}^n c_j Y_j$, pero para ciertos modelos pueden presentarse diferencias y estimadores distintos.

5.4.4 Propiedades adicionales de estimadores

Invarianza

Sean $\theta \in R^p$ y $\gamma \in R^q$ ($q \leq p$) parámetros tales que $\gamma = h(\theta)$. Sean $\hat{\theta}$ y $\hat{\gamma}$ estimadores de θ y γ respectivamente, obtenidos mediante un método M . Diremos que el método M tiene la propiedad de invarianza si se cumple $\hat{\gamma} = h(\hat{\theta})$

Propiedad 1

En el contexto de la definición anterior, si la función $\gamma = h(\theta)$ además tiene inversa $\theta = h^{-1}(\gamma)$, entonces los métodos de Momentos, Máxima Verosimilitud y Mínimos Cuadrados tienen la propiedad de invarianza.

Propiedad 2

En general, el método de Máxima Verosimilitud tiene la propiedad de invarianza, esto es, si $\theta \in R^p$ y $\gamma \in R^q$ ($q \leq p$) son parámetros tales que $\gamma = h(\theta)$ y $\hat{\theta}_{MV}$ y $\hat{\gamma}_{MV}$ son los estimadores máximo verosímiles de θ y γ respectivamente, entonces se cumple que $\hat{\gamma}_{MV} = h(\hat{\theta}_{MV})$

Distribución Asintótica del Estimador Máximo Verosímil

Sea X v.a. con distribución $f_X(x; \theta)$ con $\theta \in \Theta$ y sea $\hat{\theta}$ el estimador Máximo verosímil de θ . Entonces, bajo las "condiciones de regularidad" R1 a R6,

- R1: Si $\theta \neq \theta' \Rightarrow f_X(x; \theta) \neq f_X(x; \theta')$.
- R2: R_X no depende de θ .
- R3: El verdadero valor de θ es "punto interior" del espacio paramétrico Θ .
- R4: $f_X(x; \theta)$ es función tres veces diferenciable de θ ($\equiv \frac{d^3}{d\theta^3} f_X(x; \theta)$)
- R5: $\frac{d}{d\theta} \int f_X(x; \theta) dx = \int \frac{d}{d\theta} f_X(x; \theta) dx$
- R6: $\frac{d^2}{d\theta^2} f_X(x; \theta) \leq M(x)$ donde $M(x)$ es una función definida en un entorno del verdadero valor de θ :

Si el tamaño de muestra n es grande, $\hat{\theta}$ tiene distribución normal: $\hat{\theta} \sim N(\theta, \sigma_{\hat{\theta}}^2)$ donde

$$\sigma_{\hat{\theta}}^2 = \left\{ n E \left[\left(\frac{\partial}{\partial \theta} \ln f_X(x; \theta) \right)^2 \right] \right\}^{-1}$$

Corolario

Bajo las condiciones de regularidad, el estimador MV $\hat{\theta}$ de θ es consistente y asintóticamente insesgado.

Fórmulas sobre Intervalos de Confianza

6.1 Definición de Intervalos de Confianza

Sea θ un parámetro, sean L_1 y L_2 dos estadísticos y sea $(1-\alpha)$ una probabilidad conocida. Diremos que $[L_1, L_2]$ es un Intervalo de Confianza (I.C.) para θ al nivel de $100(1-\alpha)\%$ de Confianza, si se cumple $P(L_1 \leq \theta \leq L_2) = 1-\alpha \quad \forall \theta \in \Theta$

Observación:

También se suele presentar el I.C. escribiendo lo siguiente:

$$L_1 \leq \theta \leq L_2 \quad 100(1-\alpha)\% \text{ de Confianza}$$

La construcción de un I.C. busca encontrar un intervalo que tenga alta probabilidad $1-\alpha$ y que a la vez sea de longitud mínima.

6.2 Metodología (Método de la Variable Base)

Sea $X \sim f_X(x; \theta)$ y sea (X_1, X_2, \dots, X_n) m.a. tomada de la población de X . Dado un nivel de confianza $(1-\alpha)$, para construir un Intervalo de Confianza para θ :

- Determinar una variable base $W = W(X_1, X_2, \dots, X_n; \theta)$ cuya estructura contenga al parámetro θ pero cuya distribución $g_W(w)$ no dependa de θ
- Buscar en la distribución de W dos valores a y b tales que:

$$P(W(X_1, X_2, \dots, X_n; \theta) \leq a) = \alpha/2 = P(W(X_1, X_2, \dots, X_n; \theta) > b) \text{ y por tanto se cumpla}$$

$$P(a \leq W(X_1, X_2, \dots, X_n; \theta) \leq b) = 1-\alpha$$

- Dentro de la probabilidad anterior, despejar θ del interior del intervalo, de modo que se cumpla:

$$P(a \leq W(X_1, X_2, \dots, X_n; \theta) \leq b) = P(L_1(X_1, X_2, \dots, X_n; a, b) \leq \theta \leq L_2(X_1, X_2, \dots, X_n; a, b)) = 1-\alpha$$

Entonces el Intervalo de Confianza para θ al nivel $100(1-\alpha)\%$ es

$$[L_1(X_1, X_2, \dots, X_n; a, b), L_2(X_1, X_2, \dots, X_n; a, b)]$$

y se escribe

$$L_1(X_1, X_2, \dots, X_n; a, b) \leq \theta \leq L_2(X_1, X_2, \dots, X_n; a, b) \text{ al } 100(1-\alpha)\% \text{ de Confianza}$$

Fórmulas sobre Prueba o Contraste de Hipótesis

7.1 Elementos

Hipótesis Estadística

Es una afirmación acerca de la distribución poblacional asociada a una variable X . La denotamos mediante H .

Hipótesis Simple: Si H especifica totalmente la distribución de X

Hipótesis Compuesta: Si H no especifica totalmente la distribución de X

Contraste Estadístico o Prueba de Hipótesis

Un sistema de análisis a partir del cual se toma la decisión de aceptar o rechazar una hipótesis estadística H_0 , mediante el análisis de una(s) muestra(s) aleatoria(s).

Hipótesis Planteada o Nula

Es aquella que se somete a prueba. La denotamos H_0 ,

Hipótesis Alternativa

Es aquella que nos queda si se rechaza H_0 . Esta hipótesis se denotará H_1 .

Normalmente H_1 es el complemento de H_0 , dentro del contexto de investigación.

Región Crítica

Denotada C , es un evento cuya ocurrencia conduce al rechazo de H_0 .

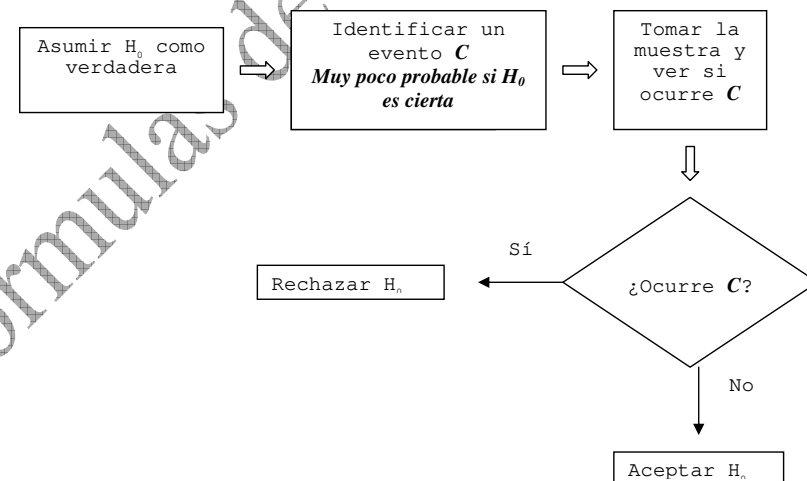
Error Tipo I

Es el error que se produce si se rechaza H_0 injustamente siendo verdadera)

Error Tipo II

Es el error que se produce si se acepta H_0 siendo falsa

Metodología general



Dada una región crítica C , es posible cometer ó el Error I ó el Error II. La probabilidad del Error I se denota α , y la probabilidad del Error II se denota β , esto es:

$\alpha = P(\text{Error I}) = P(C | H_0 \text{ es verdadera})$. Esta probabilidad α se llama también "Nivel de Significación del contraste" y mide el tamaño probabilístico de la región crítica

$$\beta = P(\text{Error II}) = P(C^c | H_0 \text{ es falsa})$$

$1 - \beta = 1 - P(\text{Error II}) = 1 - P(C^c | H_0 \text{ es falsa})$ se llama "Potencia o poder del contraste".

Nota:

- (1) α es apropiado si es menor o igual que 0.05 (o no mayor que 5%)
- (2) La región C se construye en términos de algún estadístico, llamado "estadístico de contraste".

Región Crítica Óptima

Una región crítica C se dice "óptima" si para un nivel α dado, minimiza la probabilidad β . Las regiones críticas que figuran en los textos y/o paquetes estadísticos son óptimas.

Aunque α y β no son complementarios, sí se puede demostrar que dado un tamaño de muestra n , α y β van en sentidos contrarios: O sea, si reducimos α entonces β "crece". Por eso, si deseamos que ambas probabilidades sean pequeñas, debemos hacer crecer el tamaño de muestra.

Se recomienda escribir la hipótesis de investigación como H_1 y su negación como H_0 .

7.2 Hipótesis Simples vs Hipótesis Simples: Teorema de Neyman-Pearson

Sea $X \sim f_X(x; \theta)$ y supongamos que sólo hay dos valores posibles para θ : $\theta = \theta_0$ o $\theta = \theta_1$ y tenemos que decidir entre una de las posibilidades. Si escribimos $H_0: \theta = \theta_0$ y $H_1: \theta = \theta_1$, estamos ante el caso de H_0 simple vs H_1 simple.

Supongamos ahora que el tamaño de muestra n es dado y que hemos fijado una probabilidad α conocida para el Error I.

Teorema de Neyman-Pearson

En el contexto anterior, la Región Crítica Óptima C , o sea aquella que fijada la probabilidad α de Error I, minimiza la probabilidad β de Error II, es de la forma:

$$C = \left\{ (X_1, X_2, \dots, X_n) \mid \frac{L(\theta = \theta_1)}{L(\theta = \theta_0)} \geq k \right\}$$

donde k satisface la ecuación

$$P\left(\frac{L(\theta = \theta_1)}{L(\theta = \theta_0)} \geq k \mid H_0 \text{ es verdadera}\right) = \alpha$$

7.3 Hipótesis Simples vs Hipótesis Compuestas: Pruebas Uniformemente Más Poderosas UMP

Sea $X \sim f_X(x; \theta)$ y supongamos que deseamos contrastar $H_0: \theta = \theta_0$ vs $H_1: \theta \in \omega$

donde ω es un conjunto conocido de valores alternativos para θ . Estamos ante el caso de H_0 simple vs H_1 compuesta.

Supongamos ahora que el tamaño de muestra n es dado y que hemos fijado una probabilidad α conocida para el Error I.

Para esta situación no existe un teorema como el de Neyman-Pearson que nos proporcione una región crítica C óptima que minimice la probabilidad β de Error II, pues esto depende de cómo es ω

Hipótesis Pseudosimple: Se define mediante $H_1: \theta = \theta_1$ ($\theta_1 \in \omega$)

En este contexto $H_0: \theta = \theta_0$ vs $H_1: \theta \in \omega$ y $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$ ($\theta_1 \in \omega$) son equivalentes y podemos aplicar el T. de Neyman-Pearson al juego anterior, donde θ_1 es un valor genérico de ω . Sea $C(\theta_1)$ la región crítica obtenida y supongamos que su construcción es uniformemente válida para todo $\theta_1 \in \omega$. En este caso podemos garantizar que C es una región uniformemente óptima (óptima en todas las circunstancias). Este tipo de región crítica se llama Región Crítica Uniformemente Óptima o Uniformemente Más Poderosa (UMP)

Región Crítica Uniformemente Óptima (UMP)

Sea $X \sim f_X(x; \theta)$ y sea el juego de hipótesis $H_0: \theta = \theta_0$ vs $H_1: \theta \in \omega$ donde θ_0 y ω son conocidos, siendo ω un conjunto de valores alternativos para θ .

Una región crítica C es Uniformemente Óptima (o Uniformemente Más Poderosa) si dado un Nivel α predeterminado y dado un tamaño de muestra n , la Región Crítica C satisface el Teorema de Neyman-Pearson para todo $\theta \in \omega$, no dependiendo su forma del valor elegido de θ en ω .

Nota

- Si C es UMP, entonces $P(\text{Error I}) = P(C | H_0 \text{ es verdadera}) \leq \alpha$ y $\beta = P(\text{Error II}) = P(C^c | H_0 \text{ es falsa})$ es mínima $\forall \theta_1 \in \omega$
- No siempre existe la R.C. UMP

7.4 Hipótesis Compuestas vs Hipótesis Compuestas: Test de Razón de Verosimilitud

Este sistema de generar regiones críticas es bastante general y se aplica cuando no podemos generar una región crítica UMP.

Sea $X \sim f_X(x; \theta)$ sobre la cual tenemos la hipótesis nula $H_0: \theta \in \omega_0$ vs la hipótesis alterna $H_1: \theta \in \omega_0^c$ (donde ω_0 es un conjunto bien definido y ω_0^c es su complemento).

Sea $\hat{\theta}_0$ el estimador MV de θ bajo $H_0: \theta \in \omega_0$ o sea $\hat{\theta}_0$ maximiza $L(\theta)$ sobre $\theta \in \omega_0$. La probabilidad de la m.a. bajo H_0 es $L(\hat{\theta}_0) = \prod_{j=1}^n f_X(x_j; \hat{\theta}_0)$

Sea $\hat{\theta}$ el estimador MV de θ en general, o sea sobre $\theta \in \omega_0 \cup \omega_0^c$. La probabilidad de la m.a. es $L(\hat{\theta}) = \prod_{j=1}^n f_X(x_j; \hat{\theta})$

Definición (Razón de verosimilitud)

El cociente o "razón de verosimilitud" es $\lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$. Se cumple que

$$0 \leq \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} \leq 1.$$

Nota: Si $H_0: \theta \in \omega_0$ es cierta, esperamos que $\lambda \approx 1$

Definición (Región crítica de Razón de verosimilitud)

La región crítica de Razón de Verosimilitud C se define como el evento

$$C = \{(X_1, X_2, \dots, X_n) \mid \lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} \leq \lambda_0\} \text{ tal que } \lambda_0 \text{ satisface } P\left(\frac{L(\hat{\theta}_0)}{L(\hat{\theta})} \leq \lambda_0 \mid \theta \in \omega_0\right) = \alpha$$

Este test se llama "Test de Razón de Verosimilitud"

Propiedad

Si n es "grande" y $H_0: \theta \in \omega_0$ es verdadera, entonces $-2\ln(\lambda) \sim \chi^2(r)$ donde r es el número de parámetros especificado por $H_0: \theta \in \omega_0$. Esta propiedad puede usarse para hallar el valor λ_0 de determina la zona de rechazo en este test.

Fórmulas sobre Modelo de regresión lineal simple

8.1 Uso del modelo de regresión lineal simple

Cuando salvo variaciones aleatorias, una determinada variable cuantitativa X condiciona a otra variable cuantitativa Y de modo que cambios en X inducen cambios proporcionales en Y .

Geoméricamente lo anterior equivale a que en un plano cartesiano XY , las parejas de valores (X, Y) siguen una trayectoria rectilínea.

Algebraicamente la proporcionalidad equivale a que X e Y satisfacen la ecuación $Y = \alpha + \beta X + \varepsilon$ donde α y β son constantes características o sea son "parámetros" y ε es una variación aleatoria debida a que los agentes económicos no siempre tienen el mismo comportamiento.

8.2 Supuestos y parámetros del modelo lineal simple

Dado el modelo $Y_j = \alpha + \beta X_j + \varepsilon_j$, evaluado en una muestra aleatoria de n parejas $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ asumimos que:

- $E(\varepsilon_j) = 0 \quad \forall j$
- $V(\varepsilon_j) = \sigma^2 = \text{constante} \quad \forall j$
- $\rho_{\varepsilon_j \varepsilon_{j'}} = 0 \quad \forall j \neq j'$
- X es de valores predeterminados, medidos antes de registrar los valores de Y .

Los supuestos implican que $E(Y) = \alpha + \beta X$ y $V(Y) = \sigma^2$

Parámetros del modelo

- La constante α , es el valor esperado o promedio de Y cuando X es cero.
- La constante β , llamada la "pendiente" de la recta mide en cuántas unidades se espera que varíe Y cuando X aumenta en 1 unidad.
- σ^2 es la varianza del azar representado por el residuo aleatorio ε ; σ es la variación promedio arriba o debajo de la recta $Y = \alpha + \underbrace{\beta X}_{\text{Efecto de } X} + \underbrace{\varepsilon}_{\text{Efecto de azar}}$

8.3 Los estimadores de mínimos cuadrados ordinarios (OLS)

Se obtienen aplicando el método de mínimos cuadrados a la función objetivo

$$Q(\alpha, \beta) = \sum_{j=1}^n [Y_j - \alpha - \beta X_j]^2 \text{ que generan las "ecuaciones normales":}$$

$$\begin{pmatrix} n & n\bar{X} \\ n\bar{X} & \sum_{j=1}^n X_j^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} n\bar{Y} \\ \sum_{j=1}^n X_j Y_j \end{pmatrix}$$

- Los estimadores satisfacen $\begin{pmatrix} n & n\bar{X} \\ n\bar{X} & \sum_{j=1}^n X_j^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} n\bar{Y} \\ \sum_{j=1}^n X_j Y_j \end{pmatrix} \Leftrightarrow$

$$\beta = \frac{\sum_{j=1}^n X_j Y_j - n\bar{X}\bar{Y}}{\sum_{j=1}^n X_j^2 - n\bar{X}^2}; \quad \alpha = \bar{Y} - \beta\bar{X}$$

Estos mismos estimadores se obtienen aplicando Máxima Verosimilitud, bajo el supuesto adicional $\varepsilon \sim N(0, \sigma^2)$.

- La predicción de $E(Y)$ es $\hat{Y} = \alpha + \beta X$ o $\hat{Y}_j = \alpha + \beta X_j$; $\varepsilon_j = Y_j - \hat{Y}_j$ y $\sigma^2 = S_\varepsilon^2 = \frac{\sum_{j=1}^n \varepsilon_j^2}{n-2}$,

- β es una estimación y su "error de estimación" cuadrático es

$$S_{\beta_1}^2 = \frac{\sigma^2}{\sum_{j=1}^n X_j^2 - n\bar{X}^2} \text{ y el Error Estándar de estimación de } \beta \text{ es}$$

$$e.e.(\beta) \equiv S_{\beta} = \sqrt{\frac{\sigma^2}{\sum_{j=1}^n X_j^2 - n\bar{X}^2}} = \sqrt{\frac{\sigma^2}{(n-1)S_X^2}}.$$

8.4 Ajuste del Modelo

Se mide qué tan bien son representa el modelo a los datos.

- Con la correlación $R = r_{YX}$ aunque no mide exactamente la coincidencia, pues la correlación de Pearson mide asociación, no necesariamente coincidencia.
- Con el Coeficiente R^2

$$R^2 = \frac{SCR}{SCT} = \frac{\sum_{j=1}^n (\hat{Y}_j - \bar{Y})^2}{\sum_{j=1}^n (Y_j - \bar{Y})^2} = \frac{\text{Variabilidad en Y generada por X}}{\text{Variabilidad total en Y}}$$

R^2 mide la proporción de la variabilidad total en Y que es "explicada" o atribuible a las diferencias en la variable independiente X a través de la regresión. Es la proporción de diferencias en Y que se deben a las diferencias en X .

Equivalentemente $100R^2\%$ es el porcentaje de variabilidad (por extensión, también se dice "% de la varianza") de Y explicada por el modelo.

Nota: Estimación de la correlación de Pearson ρ_{XY}

Se puede probar que una estimación consistente de la correlación de Pearson ρ_{XY} entre dos variables aleatorias X e Y es

$$\rho_{XY} \equiv r_{XY} = \frac{\sum_{j=1}^n (Y_j - \bar{Y})(X_j - \bar{X})}{(n-1)S_X S_Y} = \frac{\sum_{j=1}^n X_j Y_j - n\bar{X}\bar{Y}}{(n-1)S_X S_Y}.$$

Relación entre r_{XY} , β_X y R^2

- $\beta = \frac{r_{XY} S_Y}{S_X}$
- $SCR = \beta \sum_{j=1}^n (Y_j - \bar{Y})(X_j - \bar{X}) = \frac{r_{XY} S_Y}{S_X} \times r_{XY} (n-1) S_X S_Y = r_{XY}^2 (n-1) S_Y^2$
- $R^2 = \frac{SCR}{SCT} = \frac{r_{XY}^2 (n-1) S_Y^2}{(n-1) S_Y^2} = r_{XY}^2$
- $SCE = SCT - SCR = (n-1) S_Y^2 [1 - r_{XY}^2]$

8.5 Contrastes en el Análisis de Regresión.

Si asumimos $\varepsilon_j \sim N(0, \sigma^2)$ además de los supuestos clásicos, tenemos

Proposición 1

$$(a) \beta \sim N(\beta_1, \sigma_{\beta_1}^2) \text{ donde } \sigma_{\beta_1}^2 = \frac{\sigma^2}{\sum_{j=1}^n X_j^2 - n\bar{X}^2}$$

$$(b) W = (n-2)\sigma^2 / \sigma^2 \sim \chi^2(n-2). \text{ Es decir } SCE / \sigma^2 \sim \chi^2(n-2)$$

Proposición 2

$$t = \frac{(\beta_1 - \beta_1)}{S_{\beta_1}} \sim t(n-2) \text{ donde } S_{\beta_1} = \sqrt{\frac{\sigma^2}{\sum_{j=1}^n X_j^2 - n\bar{X}^2}}$$

8.5.1 Contraste general sobre β

Para contrastar $H_0: \beta = b$ donde b es un valor hipotético predeterminado, contra las distintas alternativas H_1 uni o bilaterales, aplicando variantes del Teorema de Neyman-Pearson se llega a:

Como en general $t = \frac{(\beta - b)}{S_{\beta_1}} \sim t(n-2)$, entonces si $H_0: \beta_1 = b$ es cierta, el estadístico

$t = \frac{(\beta_1 - b)}{S_{\beta_1}} \sim t(n-2)$ y $E(t) = 0$. Entonces un valor de t muy alejado de cero es buena razón para

rechazar $H_0: \beta = b$.

Considerando "muy alejados" de cero a los valores de t que tienen probabilidad menor que un nivel de significación α (**no** confundir con el intercepto del modelo!) predeterminado, llegamos a:

Hipótesis Nula	Hipótesis Alternativa	Rechazar H_0 si	Tipo de contraste
$H_0: \beta = b$	$H_1: \beta > b$	$t > t_{1-\alpha}$	Unilateral derecho
	$H_1: \beta < b$	$t < -t_{1-\alpha}$	Unilateral izquierdo
	$H_1: \beta \neq b$	$ t > t_{1-\alpha/2}$	Bilateral
$t_{1-\alpha}$ y $t_{1-\alpha/2}$ percentiles $1-\alpha$ y $1-\alpha/2$ de la tabla t de Student: $t(k=n-2)$			

8.5.2 Contraste de coeficiente nulo $H_0: \beta = 0$

$H_0: \beta = 0$ equivale a decir que Y no depende de X . El estadístico de contraste es

$$t = \beta / S_{\beta_1}$$

Hipótesis Nula	Hipótesis Alternativa	Rechazar H_0 si	Tipo de contraste
$H_0: \beta = 0$	$H_1: \beta > 0$	$t > t_{1-\alpha}$	Unilateral derecho
	$H_1: \beta < 0$	$t < -t_{1-\alpha}$	Unilateral izquierdo
	$H_1: \beta \neq 0$	$ t > t_{1-\alpha/2}$	Bilateral
$t_{1-\alpha}$ y $t_{1-\alpha/2}$ percentiles $1-\alpha$ y $1-\alpha/2$ de la tabla t de Student: $t(k=n-2)$			

Nota:

Como $\beta = \frac{r_{XY} S_Y}{S_X}$ y es posible pasar de un coeficiente a otro, entonces el con-

traste de $H_0: \beta = 0$ es equivalente al de $H_0: \rho_{XY} = 0$ y al de $H_0: R^2 = 0$

La consecuencia es que las hipótesis $H_0: \beta = 0$, $H_0: \rho_{XY} = 0$ y $H_0: R^2 = 0$ son todas equivalentes.