

Probabilidad y Estadística B

Resumen Probabilidad y Estadística B
Segundo cuatrimestre de 2021

Índice

1. Axiomas de Probabilidad	5
2. Experimentos con resultados equiprobables	5
2.1. Laplace	5
3. Conteo	5
3.1. Regla del producto	5
3.2. Permutaciones	5
3.3. Variaciones	5
3.4. Combinaciones	5
3.5. Bolas y urnas	6
4. Teoremas sobre conjuntos de eventos	6
4.1. Teorema 1	6
4.2. Teorema 2	6
4.3. Teorema σ -aditividad	6
5. Relaciones entre dos eventos: probabilidad condicional e independencia	6
5.1. Probabilidad condicional	6
5.2. Independencia de eventos	7
5.3. Teorema de Bayes	7
6. Variables aleatorias	7
6.1. Funcion de distribucion	7
6.2. Soporte de una V.A	8
6.3. Variables aleatorias discretas	8
6.4. Variables aleatorias continuas	8
6.5. Eventos equivalentes	8
7. Modelos continuos: distribuciones notables	9
7.1. Distribucion Uniforme	9
7.2. Distribucion exponencial	9
7.3. Funcion de Riesgo (para V.A.C.)	9
7.4. Distribucion Gamma	9
7.5. Distribucion normal estandar	9
7.6. Cuantil de una V.A	10
8. Funciones de variables aleatorias	10
8.1. Simulacion	10
9. Truncamiento	10
10. Vectores Aleatorios	10
10.1. Funcion de distribucion de un vector aleatorio	11
10.2. Propiedades del vector aleatorio ($\mathbb{X} = (X, Y)$)	11
10.3. Funcion de probabilidad de vectores aleatorios discretos (probabilidad conjunta)	11
10.3.1. Funciones de probabilidad marginales y de conjuntos	11
10.4. Funcion de densidad de un vector aleatorio continuo	11
10.4.1. Funciones de densidad marginales	11
10.5. Independencia de vectores aleatorios	12
10.5.1. Propiedades de la independencia de vectores aleatorios	12

11.Momentos	12
11.1. Esperanza	12
11.1.1. Propiedades	12
11.1.2. CASO GENERAL	12
11.2. Esperanza condicional	13
11.2.1. Propiedad	13
11.3. Varianza	13
11.3.1. Propiedad	13
11.3.2. Desvio estandar	13
11.3.3. Mediana	13
11.3.4. Moda	13
11.4. Esperanza de una funcion de un vector aleatorio	13
11.4.1. Propiedades de Orden	14
11.4.2. Propiedades importantes	14
11.5. Covarinza	14
11.5.1. Propiedades de la Covarinza	14
11.5.2. Coeficiente de correlacion	14
12.Prediccion	14
12.1. Los mejores predictores	15
13.Desigualdades	15
13.1. Desigualdad de Markov	15
13.2. Desigualdad de Tchevychev	15
14.Funcion de variable aleatoria (cambio de variable)	15
14.1. Metodo de eventos equivalentes	15
14.2. Metodo del Jacobiano (para vectores aleatorios)	15
15.Variables aleatorias condicionadas	16
15.1. Propiedades	16
16.Mezcla de variables aleatorias: Bayes para la mezcla	16
17.Esperanza condicional	16
17.1. Propiedades	17
18.Varianza Condicional	17
18.1. Propiedad (Pitagoras)	17
19.Proceso de Bernoulli	17
19.1. Condiciones	17
19.2. Variables definibles	18
20.Proceso de Poisson (puntual)	18
20.1. Variables definibles	18
20.2. Propiedades	18
20.2.1. Adelgazamiento	18
20.2.2. Superposicion de procesos de Poisson	18
20.3. Teorema	19
21.Distribucion Normal	19
21.1. Metodo de calculo de probabilidades	19
21.1.1. Propiedad	19

22.Distribucion Normal Multivariada	19
22.1. Propiedades	19
22.2. Teorema	20
23.Teoremas Limites	20
23.1. Ley de los grandes numeros	20
23.2. Ley debil de los grandes numeros	20
24.Teorema Central del Limite (TCL)	20
25.Resumen de sumas de V.A iid	20
26.Estadística	20
26.1. Concepto de muestra	21
27.Modelos parametricos	21
27.1. Familia parametrica de distribuciones	21
27.2. Ejemplos de familias parametricas	21
27.3. Familia Regular	21
27.4. Familias exponenciales	21
28.Funcion de Verosimilitud	22
29.Estadístico	22
29.1. Estadísticos Suficientes	22
29.2. Teorema de Factorizacion	22
29.3. Teorema	22
30.Estimadores	22
30.1. Metodo de maxima verosimilitud	23
30.2. Principio de invariancia	23
30.3. Bondad de los estimadores	23
30.4. Estimador Inssegado	23
30.4.1. Propiedad	23
30.4.2. Estimador asintoticamente inssegado	23
30.5. Estimador consistente	24
30.5.1. Teorema	24
30.5.2. Consistencia media cuadratica	24
30.6. Estimadores asintoticamente normales	24
30.6.1. Teorema	24
31.Distribuciones continuas importantes	24
31.1. Chi-Cuadrado	24
31.1.1. Teorema	24
31.1.2. Corolario	24
31.2. t de Student	25
31.3. Distribucion F de Fisher-Snedecor	25
31.4. Teorema	25
32.Test de hipotesis	25
32.1. Ensayo de hipotesis	25
32.2. Tipos de errores	25
32.3. Potencia del test	25
32.4. Test de hipotesis	26
32.5. Nivel de significacion del test	26

32.6. Test del cociente de verosimilitud	26
32.6.1. Test para hipótesis simple vs hipótesis simple	26

1. Axiomas de Probabilidad

Una probabilidad es una función de P que a cada evento A le hace corresponder un número real $P(A)$ con las siguientes propiedades:

1. $0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$
3. $A \cap B = \emptyset \implies P(A \cap B) = P(A) + P(B)$
4. $P(\bar{A}) = 1 - P(A)$

2. Experimentos con resultados equiprobables

2.1. Laplace

Evento A con M elementos y Ω espacio finito de N elementos:

$$P(A) = \frac{M}{N} = \frac{\text{card}(A)}{\text{card}(\Omega)}$$

3. Conteo

3.1. Regla del producto

Sirve para conjunto de pares ordenados entre dos conjuntos A y B : (cada uno de A con cada uno de B)

$$A \times B = \{(a, b) : a \in A, b \in B\} = \text{card}(A) \cdot \text{card}(B)$$

3.2. Permutaciones

Sirve para saber de cuantas formas se pueden ordenar n elementos de un conjunto:

$$n! = 1 \times 2 \times \dots \times n$$

3.3. Variaciones

Sirve para subconjuntos ordenados de k elementos, pertenecientes a un conjunto de n elementos, se representa como $(n)_k$:

$$(n)_k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

OBS: se hace con el botón nPr

3.4. Combinaciones

Sirve para subconjuntos sin ordenar de k elementos, pertenecientes a un conjunto de n elementos, se representa como $\binom{n}{k}$:

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

OBS: se hace con el botón nCr

3.5. Bolas y urnas

Sirve para bolas indistinguibles y urnas:

$$\#CP = \binom{B + (U - 1)}{B}$$

OBS: se hace con el boton nPr

4. Teoremas sobre conjuntos de eventos

4.1. Teorema 1

Sea $A(n)$ una sucesion de eventos tales que $A_n \subset A_{n+1} \forall n$ y $A = \bigcup_{i=0}^{\infty} A_i$:

$$P(A) = \lim_{n \rightarrow \infty} P(A_n)$$

4.2. Teorema 2

Sea $A(n)$ una sucesion de eventos tales que $A_{n+1} \subset A_n \forall n$ y $A = \bigcap_{i=0}^{\infty} A_i$:

$$P(A) = \lim_{n \rightarrow \infty} P(A_n)$$

4.3. Teorema σ -aditividad

Sea $A = \bigcup_{i=0}^{\infty} A_i \in \mathcal{A}$ con los eventos A_i mutuamente excluyentes 2 a 2, entonces:

$$P(A) = P\left(\bigcup_{i=0}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

5. Relaciones entre dos eventos: probabilidad condicional e independencia

5.1. Probabilidad condicional

Es la probabilidad que un evento A se de, sabiendo que ya se dio el evento B (A dado B):

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Propiedades de la probabilidad condicional

1. $0 \leq P(A|B) \leq 1, \forall A \in \mathcal{A}$
2. $P(\Omega|B) = 1$
3. Si $A \cap C = \emptyset \rightarrow P(A \cup C|B) = P(A|B) + P(C|B)$
4. Si $P(B) > 0$:
 - $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$
 - $P(A \cap B \cap C) = P(A|B \cap C) \cdot P(A|B) \cdot P(C) = P(A|B \cap C) \cdot P(B \cap C)$

Teorema de la probabilidad total

Dada una partición de Ω en B_1, B_2, \dots, B_n eventos, dado un evento superpuesto A , la probabilidad de A es:

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i)$$

5.2. Independencia de eventos

Dos eventos son independientes cuando:

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

Esto implica que hay la misma proporción de B en A que en todo Ω y viceversa.

Propiedades de la independencia de eventos

1. Si A y B son independientes, también lo son \bar{A} y B , A y \bar{B} , \bar{A} y \bar{B}
2. A_1, A_2, \dots, A_n son independientes si y solo si para cada subconjunto de más de dos elementos, la intersección de los sucesos coincide con el producto de las probabilidades.

5.3. Teorema de Bayes

Sean B_1, B_2, \dots, B_k una partición de Ω , A un evento de probabilidad positiva:

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{j=1}^k P(A|B_j) \cdot P(B_j)}$$

Se deduce de la definición de probabilidad condicional y el teorema de probabilidad total.

6. Variables aleatorias

Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y $X: \Omega \rightarrow \mathbb{R}$ una función, diremos que X es una variable aleatoria si $X^{-1}(B) \in \mathcal{A}$. Se puede calcular probabilidad como:

$$P(X^{-1}(B)) = P(X \in B)$$

6.1. Función de distribución

Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y X una V.A., definimos su función de distribución:

$$F_X(x) = P(X \leq x) \quad \forall x \in \mathbb{R}$$

Esta función se encarga de acumular probabilidad desde $-\infty$ hasta x .

OBS: $P(A < X \leq B) = F_X(B) - F_X(A)$

Propiedades de la función de distribución

1. $F_X \in [0, 1] \quad \forall x \in \mathbb{R}$.
2. F_X es monótona no decreciente.
3. F_X es continua a derecha.
4. $\lim_{x \rightarrow -\infty} F_X(x) = 0$ y $\lim_{x \rightarrow +\infty} F_X(x) = 1$

6.2. Soporte de una V.A

El soporte de X es:

$$S_X = \{x \in \mathbb{R} : F_X(x) - F_X(x^-) \neq 0 \vee \frac{dF_X(x)}{dx} \neq 0\}$$

6.3. Variables aleatorias discretas

La variable X tiene una distribución discreta si hay un conjunto $A \in \mathbb{R}$ finito o infinito numerable, tal que $P(X \in A) = 1$.

Sea para cada $x \in A : p_X(x) = P(X = x)$, se verifica que si $B \subset \mathbb{R}$:

$$P(X \in B) = \sum_{x \in B \cap A} p_X(x)$$

Y en particular:

$$\sum_{x \in A} p_X(x) = 1$$

Y la función de distribución es dado un $B = (-\infty, t]$ resulta:

$$P(X \in B) = P(X \leq t) = F_X(t) = \sum_{x \leq t} p_X(x)$$

6.4. Variables aleatorias continuas

Una variable aleatoria es continua si:

1.
 - a) El conjunto de valores posibles se compone de todos los números que hay en un solo intervalo o una unión excluyente de estos.
 - b) Ninguno de estos valores tiene un valor de probabilidad positivo $P(x = c) = 0 \quad \forall c \in \mathbb{R}$
2. Se dice que X es una variable continua si existe una función $f_X : \mathbb{R} \rightarrow \mathbb{R}$, llamada función de densidad de probabilidad, que satisface las siguientes condiciones:
 - a) $f_X \geq 0 \quad \forall x \in \mathbb{R}$
 - b) $\int_{-\infty}^{\infty} f_X(x) dx = 1$
 - c) Para cualquier a y b tales que $-\infty < a < b < +\infty$:

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

Teorema

Sea $F_X(x)$ una función de distribución de una V.A.C. (admite derivada), luego:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

OBS: La función de densidad solo existe para V.A.C.

6.5. Eventos equivalentes

Dos eventos son equivalentes si acumulan la misma probabilidad. Para V.A.D. significa que ambos eventos tienen la misma probabilidad.

7. Modelos continuos: distribuciones notables

7.1. Distribucion Uniforme

Supongamos una V.A.C. que toma todos los valores sobre un intervalo $[a, b]$. Si $f_X(x)$ esta dada por:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } a < X < b \\ 0 & \text{e.o.c} \end{cases}$$

Se denota como $X \sim \mathcal{U}(a, b)$

7.2. Distribucion exponencial

Una variable aleatoria tiene una distribucion exponencial de parametro $\lambda > 0$ si su funcion de densidad esta dada por:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{e.o.c} \end{cases}$$

Y su funcion de distribucion es:

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 0 \\ \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} & \text{e.o.c} \end{cases}$$

Propiedades de la exponencial

1. (PERDIDA DE MEMORIA) Si $X \sim \mathcal{E}(\lambda)$ entonces $P(X > t + s | X > t) = P(X > s)$ $\forall t, s \in \mathbb{R}$.
2. Si X es una V.A.C. y $P(X > t + s | X > t) = P(X > s) \forall t, s \in \mathbb{R}^+$ entonces existe $\lambda > 0$ tal que $X \sim \mathcal{E}(\lambda)$

7.3. Funcion de Riesgo (para V.A.C.)

Dada la funcion intensidad de fallas $\lambda(t)$, la funcion de distribucion es:

$$F(t) = 1 - e^{-\int_0^t \lambda(s) ds} \quad \text{si } t > 0$$

7.4. Distribucion Gamma

Se dice una V.A tiene distribucion Gamma de parametros λ y k si su funcion de densidad es:

$$f_X(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x} \quad \text{si } \{x > 0\}$$

7.5. Distribucion normal estandar

La V.A. X que toma los valores $-\infty < x < +\infty$ tiene una distribucion normal estandar si su funcion de densidad es:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Para calcular probabilidades de esta distribucion hay que mirar la tabla o integrar numericamente.

7.6. Cuantil de una V.A

Un cuantil α de X es cualquier numero x_α tal que :

$$P(X < x_\alpha) \leq \alpha \text{ y } P(X > x_\alpha) \leq 1 - \alpha$$

8. Funciones de variables aleatorias

Sea $Y = g(X)$ con X una variable aleatoria:
Si X es una V.A.D., Y sera discreta con:

$$p_Y(y) = P(Y = y) = \sum_{x \in B} p_X(x) \quad \text{con} \quad B = \{x \in \mathbb{R} : g(x) = y\}$$

Y en general:

$$F_Y(y) = P(Y \leq y) = P(g(x) \leq y)$$

Y con esta ultima se calcula la probabilidad $\forall y \in \mathbb{R}$

8.1. Simulacion

Sabiendo la distribucion de X y teniendo una variable aleatoria U para generar valores al azar, sabiendo su distribucion, entonces se busca una $F_U(u_i) = F_X(x_i)$, de donde se puede despejar x_i en funcion de u_i . Este despeje se puede hacer mediante la INVERSA GENERALIZADA:

$$F_X^{-1}(u) = \min\{x \in \mathbb{R} : F_X(x) \leq u\} \quad \text{con } u \in (0, 1)$$

Teorema

Si f es una funcion que cumple:

- Ser no decreciente.
- $\lim_{x \rightarrow +\infty} F(x) = 1$ y $\lim_{x \rightarrow -\infty} F(x) = 0$.
- Continua a derecha.

$\Rightarrow X = F^{-1}(U)$ con $U \sim \mathcal{U}(0, 1)$, se tiene que X es una V.A. cuya funcion de de distribucion es la funcion F dada.

9. Truncamiento

Sea X una V.A con $F_X(x) = P(X \leq x)$

$$F_{X|X \in A}(x) = P(X \leq x | X \in A) = \frac{P(X \leq x, X \in A)}{P(X \in A)}$$

Si X es continua, $f_X(x) = \frac{dF_X(x)}{dx}$

$$\Rightarrow f_{X|X \in A}(x) = \frac{d}{dx} F_{X|X \in A}(x) = \frac{f_X(x) \mathbb{1}\{X \in A\}}{P(X \in A)}$$

10. Vectores Aleatorios

$\mathbb{X} = (X_1, X_2, X_3, \dots, X_n)$ es un vector aleatorio de dimension n si para cada $j = 1, \dots, n$; $X_j : \Omega \rightarrow \mathbb{R}$ es una V.A.

10.1. Funcion de distribucion de un vector aleatorio

Sea \mathbb{X} un vector aleatorio de dimension n , definimos la funcion de distribucion de \mathbb{X} como:

$$F_{\mathbb{X}}(\bar{x}) = P(X_1 \leq x_1, X_2 \leq x_2, X_3 \leq x_3, \dots, X_n \leq x_n)$$

10.2. Propiedades del vector aleatorio ($\mathbb{X} = (X, Y)$)

1. $\lim_{x, y \rightarrow \infty} F_{\mathbb{X}}(x) = 1$,
2. $F_{\mathbb{X}}(x)$ es monotona no decreciente en cada variable.
3. $F_{\mathbb{X}}(x)$ es continua a derecha en cada variable.
4. $P((X, Y) \in (a_1, b_1) \times (a_2, b_2)) = F_{\mathbb{X}}(b_1, b_2) - F_{\mathbb{X}}(b_1, a_2) - F_{\mathbb{X}}(a_1, b_2) + F_{\mathbb{X}}(a_1, a_2)$

10.3. Funcion de probabilidad de vectores aleatorios discretos (probabilidad conjunta)

Sean X e Y dos V.A.D definidas en el espacio muestral Ω de un experimento. La funcion de probabilidad conjunta se define para cada par de numeros (x, y) como:

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

Debe cumplirse que:

1. $p_{X,Y}(x, y) \geq 0$
2. $\sum_x \sum_y p_{X,Y}(x, y) = 1$

10.3.1. Funciones de probabilidad marginales y de conjuntos

Para el caso de las variables aleatorias recién mencionadas, sus funciones de probabilidad marginales están dadas por:

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

$$p_Y(y) = \sum_x p_{X,Y}(x, y)$$

Para el caso de cualquier conjunto A compuesto por pares de valores (x, y) entonces:

$$P((X, Y) \in A) = \sum_{(x,y) \in A} \sum p_{X,Y}(x, y)$$

10.4. Funcion de densidad de un vector aleatorio continuo

Sean X e Y V.A.C una funcion de densidad de probabilidad conjunta $f_{X,Y}(x, y)$ de estas dos variables es una funcion que satisface:

1. $f_{X,Y}(x, y) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$

10.4.1. Funciones de densidad marginales

Para calcular las funciones de densidad marginales de X e Y :

1. $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$
2. $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$

10.5. Independencia de vectores aleatorios

Sea (X, Y) un vector aleatorio, las variables aleatorias X e Y son independientes si y solo si:

$$P((X \in A) \cap (Y \in B)) = P(X \in A) \cdot P(Y \in B) \quad \forall A, B$$

10.5.1. Propiedades de la independencia de vectores aleatorios

1. Se dice que $X_1 \dots X_n$ son V.A independientes si

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n)$$

2. Se dice que las V.A discretas X_1, \dots, X_n independientes si

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n)$$

3. Se dice que las V.A continuas X_1, \dots, X_n son independientes si

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$$

11. Momentos

11.1. Esperanza

Es el promedio ponderado de los valores que puede tomar una V.A. (centro de masa") Sea X una V.A.D con funcion de probabilidad $p_X(x)$, el valor esperado (o media) de X es:

1. Para discretas:

$$E(X) = \sum_{x \in R_x} x \cdot p_X(x)$$

2. Para continuas:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

11.1.1. Propiedades

1. El valor de la esperanza de cualquier funcion $h(x)$ (una V.A.) se calcula como:

- a) Para discretas:

$$E(h(x)) = \sum_{x \in R_x} h(x) \cdot p_x(x)$$

- b) Para continuas:

$$E(h(x)) = \int_{-\infty}^{\infty} h(x) \cdot f_x(x) dx$$

2. Sea X una V.A. con $E(X) = \mu$ si $h(x) = aX + b \rightarrow E(h(X)) = a\mu + b$

11.1.2. CASO GENERAL

Sea X una V.A. con funcion de distribucion $F_X(x) = P(X \leq x)$ si $h(X)$ es una funcion cualquiera de X , si definimos A como el conjunto de atomos (valores de X que concentren masa positiva), entonces:

$$E[h(X)] = \sum_{x \in A} h(x) \cdot P(X = x) + \int_{\mathbb{R}/A} h(x) \cdot F'_X(x) dx$$

11.2. Esperanza condicional

$$E[X|X \in A] = \frac{E[X \mathbb{1}\{X \in A\}]}{P(X \in A)}$$

Si despejo, y pienso en una particion tenemos:

$$E(X) = E[X|X \in A] \cdot P(X \in A) + E[X|X \in \bar{A}] \cdot P(X \in \bar{A})$$

11.2.1. Propiedad

Otra manera para calcular la esperanza que puede ser util:

$$E(X) = \int_0^\infty (1 - F_X(x))dx - \int_{-\infty}^0 F_X(x)dx$$

11.3. Varianza

Sea X una V.A, definimos la varianza X como:

$$Var(X) = E[(X - E(X))^2]$$

11.3.1. Propiedad

$$Var(X) = E[(X - E(X))^2] = E[X^2 - 2XE(X) + E(X)^2]$$

Si $E(X) = \mu$

$$\rightarrow Var(X) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2 = E(X^2) - E(X)^2$$

11.3.2. Desvio estandar

Se define como la raiz cuadrada de la varianza de una V.A.:

$$\sigma_x = \sqrt{Var(X)}$$

11.3.3. Mediana

Es el valor de X que acumula una probabilidad de 0.5 (es el cuantil 0.5) : $X/F_X(x) = 0,5$

11.3.4. Moda

Es el valor de X con mayor probabilidad.

11.4. Esperanza de una funcion de un vector aleatorio

La esperanza de una funcion $h(X, Y)$ esta dada por:

1. Si (X, Y) es un vector aleatorio discreto:

$$E(h(X, Y)) = \sum_x \sum_y h(x, y)p_{X,Y}(x, y)$$

2. Si (X, Y) es un vector aleatorio continuo:

$$E(h(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)f_{X,Y}(x, y)dxdy$$

11.4.1. Propiedades de Orden

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio, $g : \mathbb{R}^k \rightarrow \mathbb{R}$ una función, tenemos que:

1. Si $X > 0 \rightarrow E(X) > 0$
2. Si $g(x) > 0 \rightarrow E(g(X)) > 0$
3. Sea $h(x) > g(x) \rightarrow E(h(X)) > E(g(X))$
4. $E(|X|) \geq E(X)$
5. $E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}$

11.4.2. Propiedades importantes

1.

$$E\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i E(X_i)$$

2. Si X_1, \dots, X_n son independientes entonces:

$$E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i)$$

11.5. Covarinza

Sean X e Y dos V.A.:

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

11.5.1. Propiedades de la Covarinza

1. $Cov(X, Y) = E(X \cdot Y) - E(X)E(Y)$
2. Si X e Y son independientes entonces $E(X \cdot Y) = E(X)E(Y) \rightarrow Cov(X, Y) = 0$
3. $Cov(a + bX, c + dY) = b \cdot d \cdot Cov(X, Y)$
4. $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
5. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

11.5.2. Coeficiente de correlacion

Entre las V.A. X e Y esta dado por:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

$$\rightarrow -1 \leq \rho_{XY} \leq 1$$

Propiedad:

$$|\rho_{XY}| = 1 \Leftrightarrow P(aX + b = Y) = 1$$

12. Prediccion

Sea Y una V.A., $\mathbb{X} = (X_1, \dots, X_n)$ un vector aleatorio, existirá alguna función $g(\mathbb{X})$ que nos sirva para predecir a Y . Para encontrar dicha función se calcula el error cuadrático medio:

$$ECM = E[(Y - g(\mathbb{X}))^2]$$

12.1. Los mejores predictores

1. Constante: $E(X)$
2. Lineal: Recta de regresión de Y basada en X

$$g(X) = \hat{Y} = \frac{Cov(X, Y)}{Var(X)}(X - E(X)) + E(Y)$$

13. Desigualdades

13.1. Desigualdad de Markov

Sea $h : \mathbb{R} \rightarrow \mathbb{R}^+$ tal que h es par, y restringida a \mathbb{R}^+ es creciente, y sea X una V.A. tal que $E(h(X))$ existe, entonces $\forall t \in \mathbb{R}$,

$$P(|X| \geq t) \leq \frac{E[h(X)]}{h(t)}$$

Si además X es no negativa, $\forall a > 0$

$$P(X \geq a) \leq \frac{E(X)}{a}$$

13.2. Desigualdad de Tchevychev

Sea X una V.A. con varianza finita, $\forall k > 0$:

$$P(|X - E(X)| \geq k) \leq \frac{Var(X)}{k^2}$$

(desigualdad de Markov si $Y = X - E(X)$ y $h(t) = t^2$)

14. Función de variable aleatoria (cambio de variable)

Sea X una variable aleatoria, sea $Y = g(X)$ una función de la variable X , busquemos encontrar la distribución de la variable Y .

14.1. Método de eventos equivalentes

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$

14.2. Método del Jacobiano (para vectores aleatorios)

Suponga que Y_1 y Y_2 son V.A. continuas con función de densidad conjunta $f_{Y_1, Y_2}(y_1, y_2)$ y que para todo (y_1, y_2) tal que $f_{Y_1, Y_2}(y_1, y_2) > 0$, $u_1 = h_1(y_1, y_2)$, $u_2 = h_2(y_1, y_2)$ es una transformación 1 a 1 de (y_1, y_2) y (u_1, u_2) con inversa $y_1 = h_1^{-1}(u_1, u_2)$, $y_2 = h_2^{-1}(u_1, u_2)$. Si las inversas tienen derivadas parciales continuas respecto de u_1 y u_2 con jacobiano J , entonces:

$$f_{U_1, U_2}(u_1, u_2) = \frac{f_{Y_1, Y_2}(y_1, y_2)}{|J|} \Big|_{h_1^{-1}, h_2^{-1}}$$

15. Variables aleatorias condicionadas

Sea X e Y variables aleatorias discretas con $p_X(x) > 0$, la función de probabilidad de Y dada que $X = x$ es:

$$p_{Y|X=x}(y) = P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

$$\rightarrow p_{Y|X=x}(y) = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

Se define como $p_{Y|X=x}(y) = 0$ cuando $p_X(x) = 0$

Recordar:

$$p_{X,Y}(x, y) = p_{Y|X=x}(y)p_X(x)$$

$$p_{X,Y}(x, y) = p_{X|Y=y}(x)p_Y(y)$$

Sea (X, Y) un vector aleatorio continuo con densidad conjunta $f_{X,Y}(x, y)$ y densidad marginal $f_X(x)$, entonces para cualquier valor de X con el cual $f_X(x)$, la función de densidad de la variable condicionada Y dada $X = x$ es:

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Si $f_X(x) = 0$ entonces se define como $f_{Y|X=x} = 0$

15.1. Propiedades

1. Sea X e Y V.A. discretas tales que $p_{Y|X=x}(y) = p_Y(y) \quad \forall x \in \mathbb{R}$
 $\rightarrow X$ e Y son independientes.
2. $f_{X,Y}(x, y) = f_{Y|X=x}(y) \cdot f_X(x)$

16. Mezcla de variables aleatorias: Bayes para la mezcla

Sea M una V.A.D. con soporte finito y X una V.A.C., de manera que conozco la función de probabilidad de M y la función de densidad de las variables aleatorias $X|M = m$, $\forall m \in R_M$. La función de probabilidad de M dado que $X = x$ será:

$$p_{M|X=x}(m) = \frac{f_{X|M=m}(x) \cdot P(M = m)}{\sum_{i \in R_M} f_{X|M=i}(x) \cdot P(M = i)}$$

17. Esperanza condicional

Si $Y|X = x$ es una V.A.D.

$$\rightarrow E[Y|X = x] = \sum_{y \in R_{Y|X=x}} y \cdot p_{Y|X=x}(y)$$

Si $Y|X = x$ es una V.A.C.

$$\rightarrow E[Y|X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X=x}(y)$$

Ambas son función de x y se las llama **Función de Regresión ($\phi(x)$)**

De acá desprende la definición de Esperanza Condicional:

Si llamamos $\phi(x) = E[Y|X = x]$ a la esperanza de la variable condicionada Y dado que $X = x$, luego $\phi : \text{Sup}(X) \rightarrow \mathbb{R}$

Vamos a definir una variable aleatoria llamada **Esperanza Condicional** de Y dada X , denotada por $E[Y|X]$, como $\phi(X) = E[Y|X]$ (cuidado!!!! es una V.A, no una esperanza)

Otra definicion:

La V.A. Esperanza Condicional de Y dada X se define como $\phi = E[Y|X]$, con ϕ una funcion medible tal que $E[(Y - \phi(x)) \cdot t(X)] = 0$ para toda funcion t medible $t : \text{Sup}(x) \rightarrow \mathbb{R}$ tal que $Y \cdot t(X)$ tiene esperanza finita. Definimos $\phi(x) = E[Y|X = x]$

17.1. Propiedades

1. $E[E[Y|X]] = E[Y]$
2. Sea X e Y variables aleatorias, s y r funciones medibles tales que las variables aleatorias $r(X) \cdot s(Y)$; $r(X)$ y $s(Y)$ tiene esperanza finita, entonces

$$E[r(X) \cdot s(Y)|X] = r(X)E[s(Y)|X]$$

3. Sean Y_1 e Y_2 V.A con esperanza finita

$$E[aY_1 + bY_2|X] = aE[Y_1|X] + bE[Y_2|X]$$

4. $E[Y|X] = E[Y]$ si X e Y son independientes
5. $E[r(X)|X] = r(X)$

OBS: La esperanza condicional es de Y dado X es la funcion de la V.A. X que mejor predice a Y .

18. Varianza Condicional

Sean X e Y variables aleatorias con $\text{Var}(Y)$ finita, la varianza de $Y|X = x$ sera:

$$\text{Var}(Y|X = x) = E[(Y - E(Y|X = x))^2|X = x]$$

Llamaremos $T(x) = \text{Var}(Y|X = x)$, $T : \text{Sup}(X) \rightarrow \mathbb{R}$ llamaremos varianza condicional de Y dada X a la V.A:

$$T(X) = \text{Var}(Y|X) = E[(Y - E(Y|X))^2|X]$$

Desarrollando el cuadrado y aplicando propiedades de esperanza condicional:

$$\text{Var}(Y|X) = E[Y^2|X] - E[Y|X]^2$$

18.1. Propiedad (Pitagoras)

$$\text{Var}(Y) = E[\text{Var}(Y|X)] + \text{Var}(E[Y|X])$$

19. Proceso de Bernoulli

19.1. Condiciones

1. Dicotomia (Sucede un evento, o no sucede)
2. Probabilidad constante
3. Los experimentos son independientes

Entonces tendremos una V.A. $X \sim \text{Ber}(p)$ y X_1, X_2, \dots una sucesion de V.A. tales que $X_1, X_2, \dots \sim X$

19.2. Variables definibles

Las variables definibles en este tipo de proceso son:

1. Y : 'Cantidad de exitos en n ensayos de Bernoulli', donde $Y \sim B(n, p)$
2. W : 'Cantidad de ensayos hasta lograr k exitos', donde $W \sim Pas(k, p)$
3. N : 'Cantidad de ensayos hasta el 1º exito', donde $N \sim \mathcal{G}(p)$

20. Proceso de Poisson (puntual)

Un proceso puntual aleatorio es un conjunto enumerable de puntos aleatorios ubicado sobre la recta real. En la mayoría de las aplicaciones un punto de un proceso puntual es el instante en que ocurre algún evento, motivo por el cual los puntos también se llaman eventos o arribos.

Llamemos $N(t)$ al número de eventos durante un intervalo específico $[0, t]$

1. El número de eventos durante intervalos de tiempo no superpuestos son variables aleatorias independientes (N_1 y N_2 son independientes).
2. La probabilidad de cada evento particular es la misma para todos los intervalos de longitud t , independientemente de la ubicación de cada intervalo y de la historia pasada del sistema (N_1 y N_2 tienen la misma distribución de probabilidades).
3. La probabilidad de obtener 2 o más eventos en un intervalo lo suficientemente chico es despreciable.

La variable $N(t)$ toma los valores posibles $0, 1, 2, \dots = \mathbb{N} \cup 0$ y su función de probabilidad está dada por

$$p_N(n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

donde $\lambda > 0$. Entonces $N(t)$ tiene distribución de Poisson de parámetro $\mu = \lambda t$

20.1. Variables definibles

En un Proceso de Poisson de intensidad o tasa de ocurrencia λ (PP(λ)):

1. $N(t)$: 'Cantidad de eventos en el intervalo de longitud t ', donde $N(t) \sim Poi(\lambda t)$
2. T : 'Tiempo entre dos eventos consecutivos', donde $T \sim Exp(\lambda)$
3. G : 'Tiempo hasta el k -ésimo evento de Poisson', donde $G \sim \Gamma(k, \lambda)$

20.2. Propiedades

20.2.1. Adelgazamiento

Sea $\{N(t); t > 0\}$ un proceso de Poisson de tasa λ . Cada vez que ocurre un evento se lo clasifica como de tipo I o de tipo II . Mas aun se clasifica como de tipo I con probabilidad p o como de tipo II con probabilidad $(1 - p)$, independientemente de todos los demás arribos. Sean $N_I(t)$ y $N_{II}(t)$ la cantidad de arribos de tipo I y II que ocurren en $[0, t]$. Es claro que $N(t) = N_I(t) + N_{II}(t)$. Estos dos nuevos procesos son independientes.

20.2.2. Superposicion de procesos de Poisson

Sean $\{N_1(t), t \geq 0\}$ y $\{N_2(t), t \geq 0\}$ procesos de Poisson independientes de tasas λ_1 y λ_2 respectivamente. Entonces $N(t) = N_1(t) + N_2(t)$ define un proceso de Poisson de tasa $\lambda = \lambda_1 + \lambda_2$.

20.3. Teorema

Bajo la condicion de que ocurrieron exactamente n arribos en el intervalo $[0, t]$ los tiempos de los n arribos S_1, S_2, \dots, S_n , considerados como variables aleatorias desordenadas, son independientes y estan distribuidas uniformemente sobre $[0, t]$

21. Distribucion Normal

La V.A. X que toma todos los valores reales $-\infty < x < \infty$ tiene una distribucion normal de parametros $\mu \in \mathbb{R}$ y $\sigma^2 > 0$ si su funcion de densidad es de la forma:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$\rightarrow X \sim \mathcal{N}(\mu, \sigma^2)$

Con $E[X] = \mu$ y $Var(X) = \sigma^2$

21.1. Metodo de calculo de probabilidades

Sabemos que $Z \sim \mathcal{N}(0, 1) \rightarrow P(Z \leq z) = \Phi(z)$

Si $X \sim \mathcal{N}(\mu, \sigma^2)$ busco una transformacion.

Se propone $Z = \frac{X-\mu}{\sigma}$ de manera tal que si $X \sim \mathcal{N}(\mu, \sigma^2) \rightarrow Z \sim \mathcal{N}(0, 1)$

21.1.1. Propiedad

Si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces:

$$P(\mu - \sigma < X < \mu + \sigma) = P(-1 < Z < 1) = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 2\Phi(2) - 1$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 2\Phi(3) - 1$$

22. Distribucion Normal Multivariada

Se dice que el vector aleatorio $X = (x_1, \dots, x_p)$ tiene distribucion normal multivariada de dimension p , de parametros $\mu \in \mathbb{R}^p$ y $\Sigma \in \mathbb{R}^{p \times p}$ (simetrica y definida positiva) $X \sim \mathcal{N}_p(\mu, \Sigma)$, si su funcion de densidad esta dada por:

$$f_X(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

Con X y x vectores en \mathbb{R}^p .

Definiendo μ y Σ (Matriz de Covarizanzas) como:

$$\mu = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} \quad \Sigma = \begin{bmatrix} Var(X_1) = Cov(X_1, X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_p) \\ \vdots & \dots & \dots & \vdots \\ Cov(X_p, X_1) & \dots & \dots & Cov(X_p, X_p) = Var(X_p) \end{bmatrix}$$

22.1. Propiedades

1. Si $X \sim \mathcal{N}_p(0, diag(\lambda_1, \dots, \lambda_p))$ entonces X_1, \dots, X_p son independientes y $X_i \sim \mathcal{N}(0, \lambda_i)$.
2. Si $X \sim \mathcal{N}_p(\mu, \Sigma)$ y $A \in \mathbb{R}^{p \times p}$ no singular entonces $AX + b \sim \mathcal{N}_p(A\mu + b, A\Sigma A^T)$.

22.2. Teorema

Sean X_1, \dots, X_n V.A. aleatorias independientes con $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$ y sea $Y = \sum_{i=1}^n a_i X_i$, entonces Y tendrá una distribución normal de parámetros $\mu_Y = \sum_{i=1}^n a_i \mu_i$ y $\sigma_Y^2 = \sum_{i=1}^n a_i^2 \sigma_i^2$.

23. Teoremas Límites

23.1. Ley de los grandes números

Si se tiene una sucesión de V.A. $(X_n)_{n \geq 1}$ independientes con $E[X_i] = \mu < \infty$ y $Var(X_i) = \sigma^2 < \infty$, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ entonces $\bar{X} \rightarrow \mu$.

23.2. Ley débil de los grandes números

$$P(|\bar{X}_n - \mu| > \varepsilon) \rightarrow_{n \rightarrow \infty} 0$$

Entonces $E(\bar{X}_n) = \mu$, $Var(\bar{X}_n) = \frac{\sigma^2}{n}$

24. Teorema Central del Límite (TCL)

Sean $(X_n)_{n \geq 1}$ una sucesión de V.A. indep. e idénticamente distribuidas con $E(X_i) = \mu < \infty$ y $Var(X_i) = \sigma^2 < \infty$, $i = 1, \dots, n$, entonces (bajo ciertas condiciones generales):

$$\frac{\sum_{i=1}^n X_i - n \cdot \mu}{\sqrt{n \cdot \sigma^2}} \rightarrow Z \sim \mathcal{N}(0, 1)$$

Es decir:

$$P\left(\frac{\sum_{i=1}^n X_i - n \cdot \mu}{\sqrt{n \cdot \sigma^2}} \leq t\right) \rightarrow_{n \rightarrow \infty} \Phi(t)$$

OBS: Saber siempre que esto es una aproximación!

25. Resumen de sumas de V.A. iid

1. Si $X_1, \dots, X_n \sim \text{Ber}(p) \rightarrow \sum_{i=1}^n X_i \sim B(n, p)$
2. Si $X_1, \dots, X_n \sim G(p) \rightarrow \sum_{i=1}^n X_i \sim \text{Pas}(n, p)$
3. Si $X_1, \dots, X_n \sim \text{Exp}(\lambda) \rightarrow \sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$
4. Si $X_1, \dots, X_n \sim (p) \rightarrow \sum_{i=1}^n X_i \sim B(n, p)$

26. Estadística

Inferencia: Construir un modelo sobre la base de datos experimentales y extraer conclusiones.
 Totalidad de los resultados experimentales posibles \rightarrow POBLACIÓN.
 Conjunto de datos que se obtiene de realizar el experimento una cierta cantidad de veces \rightarrow MUESTRA.

26.1. Concepto de muestra

Variable aleatoria X , definida sobre (Ω, \mathcal{A}, P) con distribución $F_X(x) = P(X \leq x)$ que se desconoce (al menos parcialmente)

$X \rightarrow$ "observable" de experimento aleatorio.

Quiero saber como se comporta la POBLACION.

Muestra aleatoria: Sucesión de variables independientes X_1, X_2, \dots todas iid a X .

Tendremos que $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n F_X(x_i), \forall n \in \mathbb{N}$

Notación:

$\underline{X} = (X_1, \dots, X_n)$ muestra aleatoria de tamaño n con X_1, \dots, X_n .

$\underline{x} = (x_1, \dots, x_n)$ n observaciones obtenidas al realizar n repeticiones independientes de tamaño n .

27. Modelos paramétricos

La distribución de X pertenece a una familia de distribuciones \mathcal{F} que dependen de un parámetro desconocido.

27.1. Familia paramétrica de distribuciones

$\mathcal{F} = \{F_\theta(x) : \theta \in \Theta\}$ será una familia de distribuciones de probabilidad parametrizadas por un subconjunto no vacío $\Theta \in \mathbb{R}^p$ llamada espacio paramétrico.

Correspondencia uno a uno :

$$F_{\theta_1}(x) \neq F_{\theta_2}(x) \leftrightarrow \theta_1 \neq \theta_2$$

27.2. Ejemplos de familias paramétricas

$$1. X \sim \text{Ber}(p) \rightarrow f_p(x) = p^x(1-p)^{1-x}, 0 < p < 1 \text{ con } p \in \Theta = (0, 1)$$

$$2. T \sim \text{Exp}(\lambda) \rightarrow f_\lambda(t) = \lambda e^{-\lambda t}, \lambda > 0$$

$$3. Y \sim \mathcal{N}(\mu, \sigma^2) \rightarrow f_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \mu \in \mathbb{R}, \sigma^2 > 0$$

27.3. Familia Regular

1. El soporte de $f_\theta(x)$ no depende de θ .
2. f_θ es derivable con respecto a $\theta \quad \forall x$.
3. El conjunto paramétrico $\Theta \in \mathbb{R}^p$ es discreto.

27.4. Familias exponenciales

Se dice que una familia de distribuciones (continuas o discretas) en \mathbb{R}^q con distribución F_θ , $\theta \in \Theta \subset \mathbb{R}^k$ es una familia exponencial a k parámetro, si una función de densidad (o probabilidad) se puede escribir como:

$$f_\theta(x) = A(\theta) e^{\sum_{i=1}^k C_i(\theta) r_i(x)} \cdot h(x)$$

Donde:

$$\begin{array}{ll} C(\theta) & \Theta \rightarrow \mathbb{R} \\ A(\theta) & \Theta \rightarrow \mathbb{R}^+ \\ r_i(x) & \mathbb{R}^q \rightarrow \mathbb{R} \\ h(x) & \mathbb{R}^q \rightarrow \mathbb{R}^+ \end{array}$$

28. Funcion de Verosimilitud

Es la funcion conjunta (de densidad o probabilidad) vista como funcion del parametro desconocido θ :

$$\mathcal{L}(\theta) = \prod_{i=1}^n f_{\theta}(x_i) \text{ si } \underline{x} \text{ es continuo.}$$
$$\mathcal{L}(\theta) = \prod_{i=1}^n p_{\theta}(x_i) \text{ si } \underline{x} \text{ es discreto.}$$

29. Estadistico

Un estadistico es cualquier funcion medible $T_n = T(\underline{X})$ con valores en un espacio euclideo de dimension finita.

Dada una m.a. \underline{X} , un estadistico es una funcion de la m.a. que evaluada en los valores observados, debe poder resultar en un valor numerico. Esta no puede depender de los parametros desconocidos. El estadistico es una funcion de la m.a. $\underline{X} \rightarrow$ es una V.A.

29.1. Estadisticos Suficientes

Sea $\underline{X} = (X_1, \dots, X_n)$ un vector aleatorio de dimension n cuya distribucion es $F_{\theta}(\underline{x})$, $\theta \in \Theta$ se dice que un estadistico $T = r(\underline{X})$ es suficiente para θ si la distribucion de \underline{X} condicionada a que $T = t$ es independiente de $\theta \quad \forall \quad t$.

Esto significa que si conozco a T y la distribucion de $\underline{X}|T = t$, entonces puedo reconstruir una muestra con la misma distribucion que la muestra original.

29.2. Teorema de Factorizacion

Sea \underline{X} un vector aleatorio con funcion de densidad (o probabilidad) conjunta $f_{\theta}(\underline{x})$, $\theta \in \Theta$ entonces el estadistico $T = r(\underline{X})$ es suficiente para θ si y solo si existen dos funciones h y g tales que:

$$f_{\theta}(x) = g(r(\underline{x}), \theta) \cdot h(\underline{x})$$

29.3. Teorema

Una familia exponencial a k parametros tiene como estadistico suficiente para $\theta \in \mathbb{R}^k$ al vector:

$$T = (r_1(\underline{X}), \dots, r_k(\underline{X}))$$

Si X_1, \dots, X_n es una m.a. de una distribucion perteneciente a una familia exponencial a k parametros entonces al vector aleatorio \underline{X} tambien tiene una distribucion perteneciente a una familia exponencial a k parametros, y el estadistico suficiente para θ basado en la m.a. sera:

$$T = \left(\sum_{i=1}^n r_1(X_i), \dots, \sum_{i=1}^n r_k(X_i) \right)$$

30. Estimadores

1. Es un estadistico cuyos valores se consideran medidas experimentales de un parametro desconocido.
2. Un estimador es una funcion de la muestra (estadistico) que provee un valor aproximado de un parametro o caracteristica desconocida.

30.1. Metodo de maxima verosimilitud

Es un metodo para construir estimadores puntuales. Se basa en que, en los experimentos aleatorios, los resultados observados deben tener alta probabilidad de ocurrir.

Diremos que $\hat{\theta}(\underline{X})$ es un estimador de maxima verosimilitud de θ si se cumple que:

$$f_{\hat{\theta}}(\underline{X}) = \max_{\theta} f_{\theta}(\underline{x})$$

Es decir, buscamos el valor de θ que maximiza la funcion de verosimilitud.

$$\hat{\theta} = \operatorname{argmax}_{\theta} \mathcal{L}(\theta)$$

A partir de la funcion de verosimilitud $\mathcal{L}(\theta)$ vamos a buscar el valor de θ que maximiza dicha funcion, luego $\hat{\theta} = \hat{\theta}(\underline{X})$.

Si Θ es un subconjunto abierto tal que el soporte de $f_{\theta}(\underline{x})$ no depende de θ , como la funcion logaritmo es monotona creciente, maximizar $\mathcal{L}(\theta)$ es lo mismo que maximizar $\log(\mathcal{L}(\theta))$. Luego, el EMV verificara que:

$$\frac{d}{d\theta} \ln(\mathcal{L}(\theta)) = 0$$

30.2. Principio de invariancia

Supongamos que $\lambda = q(\theta)$ es una funcion biunivoca de θ . Si $\hat{\theta}$ es el EMV de θ entonces $\hat{\lambda} = q(\hat{\theta})$ sera el EMV de λ .

30.3. Bondad de los estimadores

Dada $X_1, \dots, X_n \sim^{iid} F_{\theta}(x)$, $\theta \in \Theta$ una m.a. Estimamos θ por $\hat{\theta}$. El riesgo de estimar θ con $\hat{\theta}$ se mide con el error cuadratico medio.

$$R(\theta, \hat{\theta}) = ECM(\hat{\theta}) = E[(\theta - \hat{\theta})^2]$$

Diremos que un estimador optimo para θ sera $\hat{\theta}$ tal que

$$ECM(\hat{\theta}^k) \leq ECM(\hat{\theta}), \forall \hat{\theta}$$

30.4. Estimador Inssegado

Un estimador es inssegado para θ si:

$$E_{\theta}(\hat{\theta}) = \theta, \quad \forall \theta \in \Theta$$

En caso contrario, diremos que el estimador es sesgado, y definiremos su sesgo como:

$$B(\hat{\theta}) = E_{\theta}(\hat{\theta}) - \theta$$

30.4.1. Propiedad

Dado un estimador de θ , se tiene que:

$$ECM(\hat{\theta}) = \operatorname{Var}_{\theta}(\hat{\theta}) + B(\hat{\theta})^2$$

Por lo tanto, si es inssegado, $B = 0$ y $ECM(\hat{\theta}) = \operatorname{Var}_{\theta}(\hat{\theta})$

30.4.2. Estimador asintoticamente inssegado

Un estimador es asintoticamente inssegado si:

$$\lim_{n \rightarrow \infty} E_{\theta}(\hat{\theta}) = \theta, \quad \forall \theta \in \Theta$$

30.5. Estimador consistente

Dada una sucesión de estimadores $\hat{\theta}_n$ de θ , decimos que $T = \hat{\theta}$ es (debilmente) consistente si $\forall \varepsilon > 0$:

$$P_{\theta}(|T - \theta| > \varepsilon) \rightarrow_{n \rightarrow \infty} 0$$

30.5.1. Teorema

Sea una sucesión de estimadores de $\hat{\theta}_n$ de θ . Si $\text{Var}_{\theta}(\hat{\theta}) \rightarrow 0$ y $E_{\theta}(\hat{\theta}) \rightarrow \theta$. entonces $\hat{\theta}_n$ es debilmente consistente.

30.5.2. Consistencia media cuadratica

$$\lim_{n \rightarrow \infty} ECM(\hat{\theta}) = 0, \quad \forall \theta \in \Theta$$

30.6. Estimadores asintoticamente normales

Se dice que $\hat{\theta}_n$ es una sucesión de estimadores asintoticamente normales si $\sqrt{n}(\hat{\theta}_n - q(\theta))$ converge en distribución a una normal con media cero y varianza $\frac{q'(\theta)^2}{I(\theta)}$. $I(\theta)$ se llama **Numero de informacion de fisher** y se calcula como:

$$I(\theta) = E\left[\left(\frac{d}{d\theta} \ln(f_{\theta}(X))\right)^2\right] = -E\left[\frac{d^2}{d\theta^2} \ln(f_{\theta}(X))\right]$$

(vale solo para familias regulares)

30.6.1. Teorema

Bajo ciertas condiciones muy generales, sea $\hat{\theta}_n(\underline{X})$ un EMV de θ consistente y sea $q(\theta)$ derivable con $q'(\theta) \neq 0, \quad \forall \theta$, entonces $q(\hat{\theta}_n)$ es asintoticamente normal para estimar $q(\theta)$. Si $\sqrt{n}\sqrt{I(\theta)} \cdot (\hat{\theta}_n - \theta) \rightsquigarrow^a \mathcal{N}(0, 1)$ y $\hat{\theta}$ es un estimador consistente para θ (todo esto ocurrirá con los EMV), entonces vale que:

$$\sqrt{n}\sqrt{I(\hat{\theta}_n)} \cdot (\hat{\theta}_n - \theta) \rightsquigarrow^a \mathcal{N}(0, 1)$$

31. Distribuciones continuas importantes

31.1. Chi-Cuadrado

La V.A. X tiene una distribución \mathcal{X}^2 de parametro v (grados de libertad) si su densidad esta dada por:

$$f_X(x) = \frac{1}{\Gamma(v/2)} \left(\frac{1}{2}\right)^{v/2} \cdot x^{\frac{v}{2}-1} e^{-x/2} \quad \text{si } x > 0$$

Se puede observar que coincide con una variable con distribución $\Gamma(\frac{v}{2}, \frac{1}{2})$. De ahí deducimos fácilmente que $E(X) = v$ y $\text{Var}(X) = 2v$.

31.1.1. Teorema

Si $x_1, \dots, X_n, i = 1, \dots, n$, entonces $Y = \sum_{i=1}^n X_i$ tendrá distribución \mathcal{X}_v^2 , con $v = \sum_{i=1}^n V_i$.

31.1.2. Corolario

Se llama distribución \mathcal{X}^2 con n grados de libertad a la distribución de $U = \sum_{i=1}^n Z_i^2$ donde $Z_1, \dots, Z_n \rightsquigarrow^{iid} \mathcal{N}(0, 1)$.

31.2. t de Student

Sean $Z \sim \mathcal{N}(0, 1)$ y $U \sim \mathcal{X}^2$, entonces si Z y U son independientes, $T = \frac{Z}{\sqrt{U/n}} \sim t_n$.

31.3. Distribucion F de Fisher-Snedecor

Sean U y V dos variables aleatorias indep. con distribucion \mathcal{X}^2 de n_1 y n_2 g.l. respectivamente. Entonces:

$$F = \frac{U/n_1}{V/n_2} \sim \mathcal{F}_{n_1, n_2}$$

31.4. Teorema

Sean $X_1, \dots, X_n \sim^{iid} \mathcal{N}(\mu, \sigma^2)$

1. $Z = \sqrt{n}(\frac{\bar{X}-\mu}{\sigma}) \sim \mathcal{N}(0, 1)$.
2. $W = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \mathcal{X}_{n-1}^2$.
3. w y Z son independientes.
4. Si $S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1}$ entonces $T = \sqrt{n} \frac{(\bar{X} - \mu)}{S} \sim t_{n-1}$

32. Test de hipotesis

Una hipotesis es una suposicion o conjetura sobre la naturaleza, cuyo valor de verdad o falsedad no se conoce. Una hipotesis estadistica es una hipotesis sobre una o mas poblaciones estadisticas o sobre un fenomeno aleatorio.

1. Hipotesis estadistica \rightarrow Poblacion.
2. Analisis \rightarrow Muestra.

32.1. Ensayo de hipotesis

Procedimiento que se sigue para tratar de averiguar si una hipotesis es verdadera o falsa. Siempre se van a presentar dos hipotesis:

1. H_1 : (Hipotesis de investigador) Hipotesis alternativa.
2. H_0 : Hipotesis nula. Objeto de ensayo. (Se formula con la intension de ser rechazada)

32.2. Tipos de errores

1. Error de tipo I : Se comete cuando se rechaza una hipotesis nula que era verdadera. Debe tener MUY baja probabilidad.
2. Error de tipo II : Se comete cuando no se rechaza una hipotesis nula que era falsa.

[EI es mas grave que EII]

32.3. Potencia del test

Probabilidad de rechazar la hipotesis nula.

32.4. Test de hipotesis

Es una regla de decision entre H_0 y H_1 , y la expresamos como una funcion de la muestra aleatoria $\delta(\underline{X})$ que puede tomar los valores 0 o 1. Si $\delta(\underline{X}) = 1$ se rechaza H_0 , y en caso contrario no se rechaza.

Sea $\underline{X} = (X_1, \dots, X_n)$ una m.a. de una poblacion con distribucion $F_\theta(x)$, $\theta \in \Theta$. Sean Θ_1 y Θ_2 tales que $\Theta_1 \cup \Theta_2 = \Theta$ y $\Theta_1 \cap \Theta_2 = \emptyset$. Un test para este problema es una regla de decision basada en \underline{X} para decidir entre 2 hipotesis.

$$H_0 : \theta \in \Theta_1 \quad vs \quad H_1 : \theta \in \Theta_2$$

Entonces:

1. $P(\text{'Error I'}) = P_\theta(\text{'Rechazar } H_0'), \theta \in \Theta_1$
2. $P(\text{'Error II'}) = P_\theta(\text{'No rechazar } H_0'), \theta \in \Theta_2$

Y la potencia del test es $\Pi_\delta(\theta) = P_\theta(\text{'Rechazar } H_0') = P_\theta(\delta(\underline{X}) = 1) = E_\theta \delta(\underline{X})$.

OBS:

1. $P(\text{'Error I'}) = \Pi_\delta(\theta), \theta \in \Theta_1$
2. $P(\text{'Error II'}) = 1 - \Pi_\delta(\theta), \theta \in \Theta_2$

32.5. Nivel de significacion del test

Es la maxima probabilidad de cometer un error de tipo I.

$$\alpha = \sup_{\theta \in \Theta_1} \Pi_\delta(\theta)$$

Se llama **p-valor** de un test al menor nivel de significacion para el cual se rechaza H_0 , para una observacion dada. (Es la probabilidad de encontrar un valor tan o mas extremo que el que se encontro con la muestra observada).

Llamaremos curva caracteristica:

$$\beta(\theta) = P_\theta(\text{'Error II'}) = 1 - \Pi_\delta(\theta), \theta \in \Theta_2$$

32.6. Test del cociente de verosimilitud

32.6.1. Test para hipotesis simple vs hipotesis simple

Es el caso cuando tenemos Θ_1 y Θ_2 con un solo elemento cada una:

$$H_0 : \theta = \theta_1 \quad vs. \quad H_1 : \theta = \theta_2$$

Regla de decision (Test):

$$\delta(\underline{x}) = \begin{cases} 1 & \text{si } \frac{f_{\theta_2}(\underline{x})}{f_{\theta_1}(\underline{x})} > K_\alpha \\ 0 & \text{si } \text{no} \end{cases}$$

Dada un α fijo, debemos hallar K_α que cumpla:

$$\alpha = P_{\theta_1}((\delta(\underline{X}) = 1))$$

Supongamos que $X_1, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ nuestra m.a., y que por alguna razon conocemos σ^2 .

$$H_0 : \mu = \mu_1 \quad vs. \quad H_1 : \mu = \mu_2$$

Desarrollando el cociente de verosimilitud, aplicando logaritmo y desarrollando cuadrados:

$$\delta(\underline{x}) = \begin{cases} 1 & \text{si } \sum_{i=1}^n (X_i) > K'_\alpha \\ 0 & \text{si } \text{no} \end{cases}$$

Y para hallar el valor de K'_α usamos la 2^o condicion.

$$\alpha = P_{\mu_1}(\delta(\underline{X}) = 1) = P_{\mu_1}\left(\sum_{i=1}^n (X_i) > K'_\alpha\right) = P_{\mu_1}\left(\frac{\sum_{i=1}^n (X_i) - n \cdot \mu_1}{\sqrt{n\sigma^2}} < z_\alpha\right)$$

Entonces el test con nivel de significacion α .

$$\delta(\underline{x}) = \begin{cases} 1 & \text{si } \frac{\sum_{i=1}^n (X_i) - n \cdot \mu_1}{\sqrt{n\sigma^2}} < z_\alpha \\ 0 & \text{si } \text{no} \end{cases}$$

OBS: $\delta(\underline{X})$ depende de α, n la m.a. y el valor del parametro. Bajo H_0 la regla sera la misma si $H_1 : \mu < \mu_1$.

La funcion potencia para este caso:

$$\begin{aligned} \Pi_\delta(\mu) &= P_\mu(\delta(\underline{X}) = 1) \\ &= P_\mu\left(\frac{\sum_{i=1}^n (X_i) - n \cdot \mu_1}{\sqrt{n\sigma^2}} < z_\alpha\right) \\ &= P_\mu\left(\sum_{i=1}^n (X_i) < z_\alpha \cdot \sqrt{n\sigma^2} + n \cdot \mu_1\right) \\ &= \Phi\left(z_\alpha - \frac{\sqrt{n}}{\sigma}(\mu - \mu_1)\right) \end{aligned}$$