

Probabilidad y Estadísticas (C) - Resumen Apunte Bianco

Zamboni, Gianfranco

16 de febrero de 2023

Índice

I	Probabilidad	1
1.	Probabilidad	1
1.1.	Relación con Teoría de conjunto	1
1.2.	Definición de probabilidad	2
1.2.1.	Interpretación intuitiva	2
1.3.	Axiomas y propiedades de la probabilidad	2
1.4.	Espacios muestrales	3
1.5.	Probabilidad condicional	4
1.5.1.	Propiedades	4
1.5.2.	Teorema de la probabilidad total	4
1.5.3.	Teorema de Bayes	5
1.6.	Independencia	5
1.6.1.	Propiedades	5
1.6.2.	Independencia entre mas de dos elementos:	5
2.	Variables aleatorias discretas	5
2.1.	Función de probabilidad puntual y distribución acumulada	6
2.1.1.	Función de probabilidad puntual	6
2.1.2.	Función de distribución acumulada	6
2.1.3.	Propiedades	6
2.2.	Esperanza de una variable aleatoria	7
2.2.1.	Propiedades	7
2.3.	Varianza y desvío estandar de una v.a.	8
2.3.1.	Propiedades	8
2.4.	Distribución Binomial	8
2.5.	Distribución Geometrica	9
2.6.	Distrbución Binomial Negativa	9

2.7. Distribución Hipergeométrica	10
2.8. Distribución Poisson	10
2.8.1. Proceso de Poisson	11
3. Variables aleatorias continuas	11
3.1. Función de distribución acumulada	12
3.2. Esperanza y varianza de variables aleatorias continuas	13
3.3. Varianza de una variable aleatoria continua	13
3.4. Distribución Uniforme	13
3.5. Distribución Normal	14
3.5.1. Distribución normal estándar	14
3.6. Distribución Gamma	14
3.7. Distribución Exponencial	15
4. Distribución conjunta de variables aleatorias	16
4.1. Independencia de variables aleatorias	18
4.1.1. Esperanza, covarianza y correlación	18
4.2. Extensión a más de dos dimensiones	20
4.2.1. Distribución Multinomial	20
4.2.2. Vector aleatorio continuo	20
4.3. Sumas y promedios de variables aleatorias	21
4.4. Función Generadora de Momentos y sus Propiedades	21
4.4.1. Generación de números aleatorios	23
4.5. Desigualdad de Tchebycheft	24
4.5.1. Formas equivalentes de la desigualdad de Chebyshev	24
4.6. Ley de los Grandes Números	24
4.6.1. Corrección por continuidad	26
II Estadística	26
5. Introducción y estadística descriptiva	26
5.1. Estadística descriptiva	26
5.2. Datos cuantitativos	27
5.2.1. Esquema Tallo y Hoja	27
5.2.2. Gráficos de tallo-hojas espalda con espalda. Comparación de grupos	28
5.2.3. Histograma	28
5.3. Medidas de Resumen	29
5.3.1. Medidas de posición o centrado	29
5.3.2. Medidas de dispersión y variabilidad	30
5.3.3. Box-Plots	31

6. Estimación puntual	32
6.1. Métodos de estimación puntual	33
6.1.1. Método de momentos	33
6.1.2. Método de máxima verosimilitud	33
6.1.3. Propiedades de los estimadores y criterios de selección	33
7. Intervalos de confianza	35
7.1. Intervalos de confianza para los parámetros de una distribución normal	35
7.2. Intervalos de confianza para la media de la distribución normal con varianza conocida	36
7.3. Intervalos de confianza para la media de la distribución normal con varianza desconocida	36
7.4. Intervalos de confianza para la varianza de la distribución normal con media conocida	36
7.5. Intervalos de confianza para la varianza de la distribución normal con media desconocida	37
7.6. Método general para obtener intervalos de confianza	37
7.6.1. Determinación del tamaño de muestra	37
7.6.2. Metodo general	37
7.7. Intervalos de confianza de nivel asintótico $1-\alpha$	37
8. Test de Hipótesis	38
8.1. Tipos de hipotesis a testear	38
8.2. Test de hipótesis de nivel α para los parámetros de la distribución normal	39
8.3. Test de hipotesis de nivel aproximado α para la media de una distribución cualquiera	40
8.4. Relación entre tests de hipótesis bilaterales e intervalos de confianza	41

Warning: Este resumen no tiene demostraciones.

Parte I

Probabilidad

1. Probabilidad

Experimento: Es cualquier proceso o acción que genera observaciones y que puede ser repetible.

Espacio muestral asociado a un experimento: Es el conjunto de todos los resultados posibles del experimento. Lo notaremos S . Un espacio muestral puede ser *finito*, *infinito numerable* o *infinito no numerable*.

Eventos o sucesos: Se denomina suceso o evento a cualquier subconjunto del espacio muestral. Hay dos clases de eventos:

- **Evento simple o elemental:** Consiste de un único resultado individual.
- **Evento compuesto:** Consiste de más de un evento elemental.

1.1. Relación con Teoría de conjunto

Como un evento o suceso es un conjunto valen las siguientes propiedades:

- S es un subconjunto de S denominado *suceso cierto o seguro*.
- \emptyset es un subconjunto de S denominado *suceso imposible*.
- $A \cup B$ es el suceso *unión*.
- $A \cap B$ es el suceso *intersección*.
- A^c ó \overline{A} es el *opuesto* o *complemento* de A .
- $A - B = A \cap B^c$ es el suceso *diferencia*. Ocurre cuando ocurre A y no ocurre B .

Asociatividad:

$$A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C)$$

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

Conmutatividad

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

Distributividad:

$$\left(\bigcup_{i=1}^{\infty} A_i \right)^c = \bigcap_{i=1}^{\infty} A_i^c$$

$$\left(\bigcap_{i=1}^{\infty} A_i \right)^c = \bigcup_{i=1}^{\infty} A_i^c$$

1.2. Definición de probabilidad**1.2.1. Interpretación intuitiva**

Supongamos que se repite n veces un mismo experimento aleatorio en forma independiente y bajo las mismas condiciones. Sea n_A el número de veces que ocurre el suceso A en las n repeticiones. Se denomina *frecuencia relativa* de A en la secuencia de n repeticiones a

$$fr(A) = \frac{n_A}{n}$$

y cumple que:

1. $fr(A) = \frac{n_A}{n} \geq 0$
2. $fr(S) = \frac{n_S}{n} = \frac{n}{n} = 1$
3. Si $A \cap B = \emptyset \Rightarrow fr(A \cup B) = \frac{n_{A \cup B}}{n} = \frac{n_A + n_B}{n} = \frac{n_A}{n} + \frac{n_B}{n} = fr(A) + fr(B)$

1.3. Axiomas y propiedades de la probabilidad

Dado un experimento aleatorio y un espacio muestral asociado S , a cada evento A se le asociará un número que notaremos $P(A)$ y que llamaremos probabilidad del evento A . Esta asignación debe satisfacer los siguientes axiomas:

A1. $P(A) \geq 0$ para todo evento A .

A2. $P(S) = 1$

A3a. Si A_1, \dots, A_n es una colección finita de sucesos mutuamente excluyentes, es decir que $A_i \cap A_j = \emptyset \forall i \neq j$, entonces

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

A3b. Si A_1, \dots, A_n es una colección infinita numerable de sucesos mutuamente excluyentes, es decir que $A_i \cap A_j = \emptyset \forall i \neq j$, entonces

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

Algunas de las propiedades de P son:

1. $P(A^c) = 1 - P(A)$ para todo suceso A .
2. $P(\emptyset) = 0$
3. Si $A \subseteq B \Rightarrow P(A) \leq P(B)$ y $P(B - A) = P(B) - P(A)$
4. Dados sucesos cualesquiera A y B , $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5. Dados sucesos cualesquiera A y B , $P(A \cup B) \leq P(A) + P(B)$
6. Sean A_1, \dots, A_n eventos disjuntos, entonces:

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

7. Sean A y B dos eventos de un espacio muestral S , la probabilidad de que ocurra solo uno de los eventos es $P(A \text{ xor } B) = P(A \cup B) - P(A \cap B)$
8. Sean $\{A_i\}$ una sucesión creciente de eventos ($A_1 \subseteq A_2 \subseteq \dots$) tal que $\bigcup_{i=1}^n A_i = A_n$ entonces

$$P\left(\bigcup_{i=1}^n A_i\right) = P(A_n)$$

y

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n)$$

9. Sea $\{B_i\}_{i \leq 1}$ una sucesión decreciente de eventos ($B_{i+1} \subseteq B_i$) entonces

$$P\left(\bigcap_{i=1}^{\infty} B_i\right) = \lim_{n \rightarrow \infty} P(B_n)$$

10. Sea $(A_i)_{i \leq 1}$ un conjunto de eventos entonces

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

1.4. Espacios muestrales

Supongamos que el espacio muestral S asociado con cierto experimento es finito o infinito numerable. En este caso, una manera simple de trabajar es asignar probabilidades a los sucesos elementales, ya que cualquier suceso A será unión de sucesos elementales y estos son obviamente mutuamente excluyentes.

Espacios finitos o numerables Sea $S = \{a_i : i \geq 1\} = \bigcup_{i \geq 1} \{a_i\}$ un espacio muestral, $p_i = P(\{a_i\})$ se dice función de probabilidad puntual y cumple con las siguientes características:

1. $P(S) = \sum_{i \geq 1} P(\{a_i\}) = \sum_{i \geq 1} p_i$
2. Sea A un evento tal que $A = \bigcup_{a_i \in A} \{a_i\}$ entonces $P(A) = \sum_{a_i \in A} p_i$

Espacios equiprobables Sea S un espacio muestral finito tal que $\#S = N$ y $p_i = c$ (c un valor constante) entonces S es un espacio equiprobable que cumple las siguientes propiedades:

1. $p_i = 1/N$, es decir los N sucesos elementales tienen la misma probabilidad.
2. $P(A) = \#A/\#S$

Partición de un espacio muestral: Una colección de eventos A_1, \dots, A_k constituye una partición del espacio muestral S si:

1. $A_i \cap A_j = \emptyset \forall i \neq j$
2. $P(A_i) > 0 \forall i$
3. $\bigcup_{i=1}^k A_i = S$

1.5. Probabilidad condicional

Sean A y B dos eventos tales que $P(B) > 0$, la probabilidad del evento A dado que ocurrió B es

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

1.5.1. Propiedades

1. Dado un suceso B fijo tal que $P(B) > 0$, $P(\bullet|B)$ es una probabilidad, en el sentido que satisface los axioma de probabilidad y, por lo tanto, todas las propiedades que se deducen a partir de ello.
2. **Regla del producto:** Dados dos sucesos A y B tales que $P(B) > 0$, $P(A \cap B) = P(B)P(A|B)$ y, además, si $P(A) > 0$, $P(A \cap B) = P(B|A)P(A)$
3. **Regla multiplicativa:** $P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$

1.5.2. Teorema de la probabilidad total

Teorema: Dada una partición $(A_i)_{i \geq 1}$ de S , $P(A_i) > 0 \forall i$ y un evento B de S ,

$$P(B) = \sum P(B|A_i)P(A_i)$$

1.5.3. Teorema de Bayes

Sea $(A_i)_{i \geq 1}$ una partición de S y B un evento con $P(B) > 0$, entonces

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^k P(B|A_i)P(A_i)}$$

El Teorema de Bayes describe cómo es posible “revisar” la probabilidad inicial de un evento o *probabilidad a priori* para reflejar la información adicional que nos provee la ocurrencia de un evento relacionado. La probabilidad revisada se denomina *probabilidad a posteriori*.

1.6. Independencia

Dos eventos A y B son independientes si $P(A \cap B) = P(A)P(B)$ si y solo si $P(A|B) = P(A)$. Si la igualdad no se cumple, decimos que A y B son dependientes.

1.6.1. Propiedades

1. Sean dos A y B dos sucesos, si $P(B) > 0$, A y B son independientes si y solo si $P(A|B) = P(A)$.
2. Si $A \cap B = \emptyset$ tal que $P(A) > 0$ y $P(B) > 0$, entonces, A y B no son independientes.
3. Si $P(B) = 0$ entonces B es independiente de cualquier suceso A tal que $P(A) > 0$.
4. Si $A \subseteq B$, $P(A) > 0$ y $P(B) < 1$, A y B no son independientes.
5. Si A y B son independientes, entonces A y B^c también lo son.

1.6.2. Independencia entre mas de dos elementos:

A_1, \dots, A_n son eventos independientes si y solo si $\forall k = 2 \dots n \ \forall i_1 < \dots < i_k$ tales que $1 \leq i_1 < \dots < i_k \leq n$, se verifica

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$$

Es decir que es necesario verificar $\binom{n}{2} + \binom{n}{3} + \dots + \binom{n}{n} = 2^n - n - 1$ condiciones.

Observación: Si los sucesos A_1, \dots, A_n son independientes, entonces son independientes de a pares pero la recíproca no es cierta.

2. Variables aleatorias discretas

Una **variable aleatoria** es una función $X : S \rightarrow \mathbb{R}$ que asocia a cada elemento de S un número real x .

R_x será el rango de la variable aleatoria X , es decir el conjunto de valores posibles de la variable aleatoria X .

Variable aleatoria discreta: Una variable aleatoria es discreta si toma un número finito o infinito numerable de valores.

Propiedad: Si X es una variable aleatoria discreta y toma valores x_1, x_2, \dots , entonces $g(X)$ es discreta con valores y_1, y_2, \dots , siendo $y_j = g(x_i)$ para, al menos, un valor de i .

2.1. Función de probabilidad puntual y distribución acumulada

2.1.1. Función de probabilidad puntual

La **función de probabilidad puntual o de masa** de una variable aleatoria discreta X esta definida para todo x como:

$$p_X(x) = P(X = x) = P(\{w \in S / X(w) = x\})$$

y satisface las siguientes propiedades:

1. $0 \leq p_X(x) \leq 1 \quad \forall x$
2. $\sum_{x \in R_X} p_X(x) = 1$

2.1.2. Función de distribución acumulada

La **función de distribución acumulada** de una variable aleatoria X con función de probabilidad puntual p_X se define para todo $x \in \mathbb{R}$, como:

$$F_X(x) = P(X \leq x) = \sum_{y \leq x, y \in R_X} p_X(y) \quad F_X : \mathbb{R} \rightarrow [0, 1]$$

Es decir que $F_X(x)$ es la probabilidad de que la variable aleatoria X tome valores menores o iguales que x .

Para v.a. discretas, se tiene que F_X es una función escalera no decreciente que toma valores entre cero y uno:

$$F_X(x) = P(X \leq x) = \sum_{y \leq x \in \mathbb{R}} p_X(y)$$

2.1.3. Propiedades

1. $0 \leq F_X(x) \leq 1 \quad \forall x \in \mathbb{R}$
2. Si $x_1 \leq x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$ (F_X es monotona no decreciente)
3. $\lim_{h \rightarrow 0^+} F_X(x + h) = F_X(x)$ (F_X es continua por derecha)
4. En cada punto x , el valor del salto es la probabilidad puntual, es decir:

$$p_X(x) = F_X(x) - F_X(x^-)$$

donde $x^- = \lim_{h \rightarrow 0^+} (x - h)$

Conociendo F_X se puede obtener cualquier probabilidad que involucre a X . Sean a y b tales que $a \leq b$, entonces:

$$\begin{aligned}P(a < X \leq b) &= F_X(b) - F_X(a) \\P(a \leq X \leq b) &= F_X(b) - F_X(a^-) \\P(a < X < b) &= F_X(b^-) - F_X(a) \\P(a \leq X < b) &= F_X(b^-) - F_X(a^-)\end{aligned}$$

2.2. Esperanza de una variable aleatoria

Sea X una variable aleatoria discreta que toma valores en R_X con función de probabilidad puntual $p_X(x)$, la **esperanza o valor esperado** de X se define como:

$$E(X) = \mu_X = \sum_{x \in R_X} x_i p(x_i)$$

siempre que $E(X) = \sum_{x \in R_X} |x_i| p(x_i) < \infty$. Si la serie de los valores absolutos diverge, la esperanza no puede definirse y decimos que no existe.

Ensayo Bernoulli: Sea X una variable aleatoria que toma solo dos valores, que designaremos 1 y 0 con la siguiente función de probabilidad puntual:

$$p_X(x) = \begin{cases} \alpha & \text{si } x = 1 \\ 1 - \alpha & \text{sino} \end{cases}$$

siendo $0 < \alpha < 1$, entonces su esperanza es $E(X) = 1 \cdot \alpha + 0 \cdot (1 - \alpha) = \alpha$

Interpretación: $E(X)$ es el centro de gravedad de la función de probabilidad puntual. Se ha demostrado que el promedio de los resultados obtenidos tiende a estabilizarse en un número que es $E(X)$, si es que ésta existe.

2.2.1. Propiedades

1. Si la variable aleatoria X tiene función de probabilidad puntual $p_X(x)$ para todo $x \in R_X$, entonces la esperanza de cualquier función real $h(x)$, está dada por

$$E(h(x)) = \sum_{x \in R_X} h(x) p_X(x)$$

si la serie es absolutamente convergente, o sea si $\sum_{x \in R_X} |h(x)| p_X(x) < \infty$

2. **Linealidad:** Si a y b son constante reales, $E(aX + b) = aE(X) + b$
3. Si X es una v.a. tal que $P(X = c) = 1$, entonces $E(X) = c$

2.3. Varianza y desvío estandar de una v.a.

Sea X una v.a. con función de probabilidad puntual $p_X(x)$ y esperanza μ_X , definimos la varianza de X como:

$$V(X) = \sigma_x^2 = E[(X - \mu_x)^2]$$

y la desvío estandar como:

$$\sigma_x = +\sqrt{V(X)}$$

Varianza de Ensayo Bernoulli: Sea X una variable aleatoria Bernoulli con función de probabilidad puntual:

$$p_X(x) = \begin{cases} \alpha & \text{si } x = 1 \\ 1 - \alpha & \text{sino} \end{cases}$$

siendo $0 < \alpha < 1$, entonces su varianza es

$$V(X) = (1 - \alpha)^2 \alpha + (0 - \alpha)^2 (1 - \alpha) = \alpha(1 - \alpha) [(1 - \alpha) + \alpha] = \alpha(1 - \alpha)$$

2.3.1. Propiedades

1. $V(X) = E(X^2) - (E(X))^2$
2. $V(aX + b) = a^2 V(X)$ y $\sigma_{aX+b} = |a| \sigma_X$
3. $P(X = c) = 1 \Rightarrow V(X) = 0$

2.4. Distribución Binomial

Un **ensayo bernoulli** es un experimento con dos posibles resultados: **Éxito = 1** ó **Fracaso = 0**.

Dada una sucesión de n ensayos Bernoulli con probabilidad de éxito p independientes entres si, se denomina a variable binomial a la v.a. X tal que $X =$ “número de éxitos en n repeticiones” y notamos $X \sim Bi(n, p)$.

■ **Rango:** $R_X = \{0..n\}$

■ **Función de probabilidad puntual**

$$P(X = k) = p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

■ **Función de distribución**

$$F_X(x) = \begin{cases} 0 & x < 0 \\ p_X(0) & 0 \leq x \leq 1 \\ \sum_{j=0}^{[x]} p_X(j) & k \leq x < k + 1 < n \end{cases}$$

donde $[x]$ denota la parte entera de x .

- **Esperanza y varianza**

$$E(X) = np \quad \text{y} \quad V(X) = np(1-p)$$

2.5. Distribución Geométrica

Supongamos que se repite un ensayo de Bernoulli de manera independiente hasta obtener el primer éxito y definamos la v.a. X = “numero de repeticiones hasta obtener el primer éxito”. Entonces si $P(\text{Éxito}) = p$, se dice que X tiene distribución Geométrica y se nota: $X \sim G(p)$

- **Rango:** $R_X = \mathbb{N}$

- **Función de probabilidad puntual:** $p_X(k) = (1-p)^{k-1}p$

- **Función de densidad:**

$$F_X(k) = \begin{cases} 0 & \text{si } x < 1 \\ 1 - (1-p)^k & \text{si } x \leq 1 \end{cases}$$

- **Esperanza y varianza**

$$E(X) = \frac{1}{p} \quad \text{y} \quad V(X) = \frac{1-p}{p^2}$$

Falta de memoria Sea $X \sim G(p)$ y sean n y m números naturales cualesquiera,

$$P(X > n+m | X > n) = P(X > m)$$

2.6. Distribución Binomial Negativa

Supongamos que se repite, de manera independiente, un ensayo Bernoulli con probabilidad de éxito $P(\text{Éxito}) = p$ y definamos la v.a. X = “número de repeticiones hasta obtener el r -ésimo éxito”, entonces X es una v.a. con distribución binomial negativa y se nota $X \sim BN(r, p)$.

Esta variable aleatoria es una generalización de la variable aleatoria Geométrica, la cual corresponde al caso $r = 1$.

Sea $k \geq r$, $k \in \mathbb{N}$ entonces:

- **Rango:** $R_X = \{r, r+1, r+2, \dots\}$

- **Función de probabilidad puntual:**

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-1}$$

- **Función de densidad:**

$$F_X(k) = \begin{cases} 0 & x < r \\ \sum_{k=r}^{[x]} \binom{k-1}{r-1} p^r (1-p)^{k-1} & x \leq k < x+1 < n \end{cases}$$

- **Esperanza y Varianza**

$$E(X) = \frac{r}{p} \quad \text{y} \quad V(X) = \frac{r(1-p)}{p^2}$$

Observación: Esta variable aleatoria suele también definirse como el número de fracasos ante de obtener el r -ésimo éxito. Si la denotamos X^* , entonces:

■ **Rango:** $R_X = \mathbb{N} \cup \{0\}$

■ **Función de probabilidad puntual:**

$$p_{X^*}(k) = \binom{r+k-1}{k} p^r (1-p)^k$$

■ **Esperanza y Varianza**

$$E(X^*) = \frac{r(1-p)}{p} \quad y \quad V(X^*) = \frac{r(1-p)}{p^2}$$

2.7. Distribución Hipergeométrica

Supongamos que se tiene una población de N elementos que debe ser muestreada y cada elemento puede ser clasificado como éxito o fracaso. Sabiendo que hay D éxitos en la población, se extrae una muestra de n elementos de forma tal que cualquier subconjunto del mismo tamaño tiene la misma probabilidad de ser elegido, definimos $X =$ “número de éxitos en la muestra de tamaño n ”. Entonces X tiene distribución hipergeométrica y se nota $X \sim H(n, N, D)$

■ **Función de probabilidad puntual**

$$p_X(k) = \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}} \quad \max(0, n - (N - D)) \leq k \leq \min(n, D)$$

■ **Esperanza y Varianza**

$$E(X) = n \frac{D}{N} \quad y \quad V(X) = \frac{N-n}{N-1} n \frac{D}{N} \left(1 - \frac{D}{N}\right)$$

Observación 1 El factor $\left(\frac{N-n}{N-1}\right)$ que aparece en la expresión de la varianza se denomina factor de corrección por población finita.

Observación 2 Si n es pequeño en relación a N , la hipergeométrica puede ser aproximada por la distribución Binomial de parámetros n y $p = D/N$.

2.8. Distribución Poisson

Sea $X \sim Bi(n, p)$, con $n \rightarrow \infty$ y $p \rightarrow 0$, de manera que $np = \lambda$, entonces X tiene distribución Poisson y se note $X \sim \mathcal{P}(\lambda)$

1. **Rango:** $R_X = \{0, 1, 2, \dots\}$

2. **Función de probabilidad puntual**

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \rightarrow \frac{e^{-\lambda} \lambda^k}{k!} \quad \forall k \in \mathbb{N}_0$$

3. Esperanza y Varianza

$$E(X) = \lambda \quad \text{y} \quad V(X) = \lambda$$

2.8.1. Proceso de Poisson

Supongamos que se observa la ocurrencia de un evento a lo largo del tiempo y que existe una cantidad positiva $\theta > 0$, tal que:

1. La probabilidad de que ocurra exactamente un evento en un intervalo pequeño de longitud Δt es aproximadamente $\theta \Delta t$, es decir que:

$$P(\text{ocurra un evento en } \Delta t) = \theta \Delta t + o(\Delta t)$$

siendo $o(h)$ una función $g(h)$ tal que $\lim_{h \rightarrow 0} \frac{g(h)}{h} = 0$.

2. La probabilidad de que ocurra más de un evento en un intervalo pequeño de longitud Δt es despreciable cuando se la compara con la probabilidad de que ocurra un evento, es decir:

$$P(\text{ocurra un evento en } \Delta t) = o(\Delta t)$$

3. El número de eventos que ocurren en un intervalo es independiente del número de eventos que ocurren en otro intervalo disjunto.

Entonces, el número de ocurrencias de evento en un periodo de longitud t tiene distribución de Poisson de parámetro (θt) , es decir que la variable aleatoria $X_t =$ “número de ocurrencias del evento en el intervalo de longitud t ” satisface $X_t \sim \mathcal{P}(\theta t)$

Observación A θ se le suele llamar **tasa media de ocurrencia** o **intensidad** del Proceso de Poisson.

3. Variables aleatorias continuas

Una variable aleatoria X es continua si existe una función $f : \mathbb{R} \rightarrow \mathbb{R}^+$ llamada **función de densidad** de la variable aleatoria X tal que

$$P(x \in A) = \int_A f(x) dx \quad \forall A \subseteq \mathbb{R}$$

En particular, si $A = [a, b]$, entonces:

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

y $P(X = a) = P(a \leq X \leq a) = 0$

Propiedad: Para que una función $f(x)$ sea una función de densidad, debe satisfacer:

$$f(x) \geq 0 \quad \forall x \in \mathbb{R}$$
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Observación: $f(x)$ no es una probabilidad.

3.1. Función de distribución acumulada

La **función de distribución acumulada** de una variable aleatoria continua X con función de densidad $f(x)$ se define para todo $x \in \mathbb{R}$, como

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

Propiedad: Sea X una variable aleatoria continua,

1. $\forall x \in \mathbb{R}, F_X(x) \in [0, 1]$
2. Si $x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$, es decir, $F_X(x)$ es monótona no decreciente.
3. $F_X(x)$ es continua en todo punto.
4. $\lim_{x \rightarrow \infty} F_X(x) = 1$ y $\lim_{x \rightarrow -\infty} F_X(x) = 0$

Proposición: Sean a y b tales que $a \leq b$, entonces:

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b) = F(b) - F(a)$$

Proposición: Si X es una variable aleatoria continua con función de densidad $f(x)$ y función de distribución acumulada $F(x)$, entonces todo punto donde $F(x)$ es derivable vale que:

$$F'(x) = \frac{\partial F(x)}{\partial x} = f(x)$$

Percentiles de una distribución continua: Sea X una variable aleatoria continua con función de densidad $f(x)$ y función de distribución acumulada $F(x)$ y sea $0 < p < 1$. El percentil $(100p)$ -ésimo de la distribución de X es el valor x_p tal que

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

3.2. Esperanza y varianza de variables aleatorias continuas

Sea X una variable aleatoria continua con función de densidad $f(x)$, la **esperanza** o el **valor esperado** de X se define como

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

Proposición: Si la variable continua tiene función de densidad $f(x)$, entonces la esperanza de cualquier función real $h(x)$ está dada por:

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

Linealidad: Si a y b son constante reales, $E(aX + b) = aE(x) + b$.

3.3. Varianza de una variable aleatoria continua

Sea X una variable aleatoria continua con esperanza μ_X y densidad $f(x)$, la **varianza** de X es

$$V(X) = \sigma_X^2 = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx$$

Proposición: $V(X) = E(X^2) - (E(X))^2$

Propiedad: Sea X una variable aleatoria continua con densidad $f(x)$,

$$V(aX + b) = a^2 V(X) \quad y \quad \sigma_{aX+b} = |a| \sigma_X$$

3.4. Distribución Uniforme

Se dice que X tiene distribución uniforme en el intervalo $[A, B]$, si su función de densidad es:

$$f(x) = \frac{1}{B - A} I_{[A, B]}(x)$$

Notación: $X \sim U(A, B)$

Función de distribución acumulada

$$F(x) = \begin{cases} 0 & \text{si } x < A \\ \frac{x-A}{B-A} & \text{si } A \leq x \leq B \\ 1 & \text{si } x > B \end{cases}$$

Esperanza y Varianza

$$E(X) = \frac{A + B}{2} \quad y \quad V(X) = \frac{(B - A)^2}{12}$$

3.5. Distribución Normal

Se dice que X tiene distribución Normal de parámetros μ y σ^2 ($\mu \in \mathbb{R}$, $\sigma > 0$) $[A, B]$, si su función de densidad es:

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

Notación: $X \sim N(\mu, \sigma^2)$

El gráfico de la función de densidad normal tiene forma de campana con eje de simetría en $x = \mu$ y puntos de inflexión en $x = \mu + \sigma$ y $x = \mu - \sigma$.

3.5.1. Distribución normal estándar

Se dice que Z tiene distribución normal estándar si sus parámetros son $\mu = 0$ y $\sigma^2 = 1$, es decir $Z \sim N(0, 1)$. Su función de densidad estará dada por:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

y su función de distribución $\Phi(z)$ es:

$$\Phi(z) = F(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Esta integral no tiene una expresión analítica conocida.

Esperanza y Varianza Sea $Z \sim N(0, 1)$, entonces:

$$E(Z) = 0 \quad \text{y} \quad V(Z) = 1$$

Propiedades

1. Si $X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$
2. Si $Z \sim N(0, 1)$ y $\sigma > 0 \Rightarrow X = \sigma Z + \mu \sim N(\mu, \sigma^2)$
3. Si $X \sim N(\mu, \sigma^2) \Rightarrow E(X) = \mu$ y $V(X) = \sigma^2$

3.6. Distribución Gamma

Dado $\alpha > 0$, se define la función Gamma o función factorial como:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

Propiedades

1. Si $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
2. Si $\alpha \in \mathbb{N}$, $\Gamma(\alpha) = (\alpha - 1)!$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Distribución Gamma Se dice que X tiene distribución Gamma de parámetros α y λ ($\alpha > 0$, $\lambda > 0$) si su función de densidad está dada por

$$f(x) = \frac{e^{-\lambda x} x^{\alpha-1} \lambda^\alpha}{\Gamma(\alpha)} I_{(0,\infty)}(x)$$

Notación: $\Gamma X \alpha \lambda$

Si $\lambda = 1$, la distribución se denomina Gamma estándar y su densidad está dada por:

$$f(x) = \frac{e^{-x} x^{\alpha-1}}{\Gamma(\alpha)} I_{(0,\infty)}(x)$$

Esta función de densidad es estrictamente decreciente si $\alpha \leq 1$, y si $\alpha > 1$ alcanza un máximo y después decrece.

Esperanza y Varianza: $X \sim \Gamma(\alpha, \lambda)$, entonces: $E(X) = \frac{\alpha}{\lambda}$ y $V(X) = \frac{\alpha}{\lambda^2}$

Propiedad: Si $X \sim \Gamma(\alpha, \lambda)$ y $a > 0$, entonces $aX \sim \Gamma(\alpha, \lambda/a)$

Observación: Algunos autores utilizan otra parametrización de la distribución Gamma, definiendo como segundo parámetro de la distribución a $\frac{1}{\lambda}$. Es decir $X \sim \Gamma(\alpha, \beta)$ si su función de densidad es:

$$f(x) = \frac{e^{-\frac{x}{\beta}} x^{\alpha-1}}{\lambda^\beta \Gamma(\alpha)} I_{(0,\infty)}(x)$$

En este caso, $E(X) = \alpha\beta$ y $V(X) = \alpha\beta^2$

3.7. Distribución Exponencial

Una variable aleatoria exponencial es una variable aleatoria Gamma con parámetro $\alpha = 1$. Entonces decimos que X tiene distribución exponencial de parámetro λ si su función de densidad es:

$$f(x) = e^{-\lambda x} \lambda I_{(0,\infty)}(x)$$

Notación: $X \sim \varepsilon(\lambda)$

Función de distribución acumulada

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

Esperanza y Varianza Si $X \sim \varepsilon(\lambda)$ entonces $E(X) = \frac{1}{\lambda}$ y $V(X) = \frac{1}{\lambda^2}$

Falta de memoria: Sea $X \sim \varepsilon(\lambda)$ y sean s y t números reales positivos cualesquiera, entonces:

$$P(X > s + t \mid X > s) = P(X > t)$$

Proposición: Dado un proceso de Poisson de intensidad v , si se define la variable aleatoria $T =$ “tiempo hasta la ocurrencia del primer evento”, entonces $T \sim \varepsilon(v)$

4. Distribución conjunta de variables aleatorias

Sean X e Y variables aleatorias discretas definidas sobre un espacio muestral S . La **función de probabilidad conjunta** del par (X, Y) , p_{XY} se define como:

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

El conjunto $R_{XY} = \{(x, y) \mid x \in R_X, y \in R_Y\}$ es el recorrido o rango del vector aleatorio (X, Y) .

Dado cualquier conjunto $A \subseteq \mathbb{R}^2$:

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

Una función de probabilidad conjunta satisface:

- $p_{XY}(x, y) \geq 0 \quad \forall (x, y)$
- $\sum_x \sum_y p_{XY}(x, y) = 1$

Definición: Sea (X, Y) un vector aleatorio discreto con función de probabilidad conjunta $p_{XY}(x, y)$, las **funciones de probabilidad marginal** de X e Y están dadas por

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

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

Definición: Sea (X, Y) un vector aleatorio discreto con función de probabilidad conjunta $p_{XY}(x, y)$, la **función de distribución acumulada conjunta** de X e Y está dada por

$$F_{XY}(x, y) = \sum_{s \leq x} \sum_{t \leq y} p_{XY}(s, t) \quad \forall (x, y) \in \mathbb{R}^2$$

Vectores continuos Sean X e Y variables aleatorias continuas definidas sobre un espacio muestral S . El vector (X, Y) es continuo si existe una función, denominada **función de densidad conjunta** $f_{XY}(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}_{\leq 0}$, tal que:

$$P((X, Y) \in A) = \int_A \int f_{XY}(x, y) dx dy \quad \forall A \subseteq \mathbb{R}^2$$

En particular, si $A = [a, b] \times [c, d]$,

$$P((X, Y) \in A) = \int_a^b \int_c^d f_{XY}(x, y) dy dx$$

Una función de probabilidad conjunta satisface:

- $f_{XY}(x, y) \geq 0 \quad \forall (x, y)$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dy dx = 1$

Definición: Diremos que el vector aleatorio tiene **distribución uniforme** sobre una región $A \subset \mathbb{R}^2$ si su densidad es constante sobre la región y 0 fuera de ella, es decir:

$$(X, Y) \sim U(A) \iff f_{XY}(x, y) = \begin{cases} k & \text{si } (x, y) \in A \\ 0 & \text{si } (x, y) \notin A \end{cases}$$

Definición: Sea (X, Y) un vector aleatorio continuo con función de densidad conjunta $p_{XY}(x, y)$, las **funciones de probabilidad marginal** de X e Y están dadas por

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$$

Definición: Sea (X, Y) un vector aleatorio continuo con función de densidad conjunta $p_{XY}(x, y)$, la **función de distribución acumulada conjunta** de (X, Y) está dada por

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(s, t) dt ds \quad \forall (x, y) \in \mathbb{R}^2$$

Definición: Sea (X, Y) un vector aleatorio discreto con función de probabilidad conjunta $p_{XY}(x, y)$, marginales $p_X(x)$ y $p_Y(y)$ y sea x tal que $p_X(x) > 0$, la **función de probabilidad condicional** de Y dado $X = x$ está dada por

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

Del mismo modo, sea y tal $p_Y(y) > 0$, la **función de probabilidad condicional** de X dado $Y = y$ está dada por:

$$P_{X|Y=y}(x) = \frac{p_{XY}(x, y)}{p_Y(y)}$$

Definición: Sea (X, Y) un vector aleatorio continuo con función de densidad conjunta $p_{XY}(x, y)$, marginales $f_X(x)$ y $f_Y(y)$ y sea x tal que $p_X(x) > 0$, la **función de probabilidad condicional** de Y dado $X = x$ está dada por

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

Del mismo modo, sea y tal $f_Y(y) > 0$, la **función de probabilidad condicional** de X dado $Y = y$ está dada por:

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

Se denomina **soporte de una densidad** al conjunto de valores en los cuales la densidad es positiva..

4.1. Independencia de variables aleatorias

Las variables aleatorias X e Y son independientes si y solo si para todo $a < b$ y $c < d$ se satisface:

$$P(\{a < X < b\} \cup \{c < Y < d\}) = P(a < X < b)P(c < Y < d)$$

Si esta condición no se satisface, diremos que X e Y son dependientes.

4.1.1. Esperanza, covarianza y correlación

Sean X e Y dos variables aleatorias discretas con función de probabilidad conjunta $p_{XY}(x, y)$ y sea $h(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$, entonces $h(X, Y)$ es una variable aleatorio y

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

siempre que esta esperanza exista.

Proposición: Sean X e Y dos variables aleatorias continuas con función de densidad conjunta $f_{XY}(x, y)$ y sea $h(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$, entonces $h(X, Y)$ es una variable aleatoria y

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

siempre que esta esperanza exista.

Proposición: Sean X e Y dos variables aleatorias discretas o continuas con función de probabilidad conjunta o densidad $p_{XY}(x, y)$ ó $f_{XY}(x, y)$ respectivamente y sea a y b números reales, entonces:

$$E(aX + bY) = aE(X) + bE(Y)$$

Proposición: Si X e Y son variables aleatorias independientes entonces $E(XY) = E(X)E(Y)$

Definición: Sean X e Y dos variables aleatorias con esperanzas μ_x y μ_y respectivamente, la **covarianza** entre X e Y se define como:

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

Propiedades

- $\text{Cov}(X, X) = V(X)$
- $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$
- Si X e Y son variables aleatorias independientes, $\text{Cov}(X, Y) = 0$. La reciproca no es cierta en general.

Definición: Sean X e Y dos variables aleatorias con esperanzas μ_x y μ_y respectivamente y varianzas positivas, el **coeficiente de correlación** entre X e Y se define como:

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

siendo σ_X y σ_Y , los desvíos standard de X e Y respectivamente..

Propiedades:

- Sean a, b, c y d números reales, $a \neq 0$, $c \neq 0$ y X e Y dos variables aleatorias cualesquiera con varianzas positivas, entonces

$$\rho(aX + b, cY + d) = \text{sg}(ac)\rho(X, Y)$$

donde sg denota la función signo.

- $-1 \leq \rho(X, Y) \leq 1$
- $|\rho(X, Y)| = 1 \iff Y = aX + b$ con probabilidad 1, para ciertos valores reales de a y b y $a \neq 0$.

4.2. Extensión a más de dos dimensiones

Sean X_1, \dots, X_k variables aleatorias discretas, la **función de probabilidad conjunta** del vector aleatorio (X_1, \dots, X_k) se define como:

$$p_{X_1, \dots, X_k}(x_1, \dots, x_k) = P(X_1 = x_1, \dots, X_k = x_k)$$

y dado cualquier conjunto $A \subseteq \mathbb{R}^k$:

$$P((X_1, \dots, X_k) \in A) = \sum_{(x_1, \dots, x_k) \in A} \dots \sum p_{X_1, \dots, X_k}(x_1, \dots, x_k)$$

4.2.1. Distribución Multinomial

Supongamos que se repite n veces en forma independiente una experiencia, que en cada repetición hay k resultados posibles ($k \leq 2$), cada uno de los cuales ocurre con probabilidad p_i ($1 \leq i \leq k$) y que estas probabilidades se mantienen constantes en todas las repeticiones.

Si definimos X_i como el número de veces que ocurre el resultado i ($1 \leq i \leq k$), la distribución conjunta de (X_1, \dots, X_k) se denomina distribución multinomial de parámetros n, p_1, \dots, p_k y notamos $(X_1, \dots, X_k) \sim M(n, p_1, \dots, p_k)$.

La correspondiente función de probabilidad conjunta está dada por:

$$p_{X_1, \dots, X_k}(x_1, \dots, x_k) = \begin{cases} \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} & \text{si } 0 \leq x_i \leq n \ \forall i, \ \sum_{i=1}^n x_i = n \\ 0 & \text{en otro caso} \end{cases}$$

En forma similar a lo hecho para el caso bidimensional se pueden definir las **funciones de probabilidad marginal**.

La **distribución marginal** de X_i es binomial de parámetros n y p , para todo $1 \leq i \leq k$

4.2.2. Vector aleatorio continuo

El vector aleatorio (X_1, \dots, X_k) es continuo si existe una función $f_{X_1 \dots X_k} : \mathbb{R}^k \rightarrow \mathbb{R}_{\geq 0}$, denominada **función de densidad conjunta**, tal que

$$P((X_1, \dots, X_k) \in A) = \int \dots \int f_{X_1 \dots X_k}(x_1, \dots, x_k) dx_1 \dots dx_k \ \forall A \subseteq \mathbb{R}^k$$

En forma similar a lo hecho para el caso bidimensional se pueden definir las **funciones de densidad marginal**.

Definición: X_1, \dots, X_k son variables aleatorias **independientes** si y solo si:

$p_{X_1, \dots, X_k}(x_1, \dots, x_k) = p_{X_1}(x_1) \dots p_{X_k}(x_k)$ en el caso discreto,

$f_{X_1, \dots, X_k}(x_1, \dots, x_k) = f_{X_1}(x_1) \dots p_{X_k}(x_k)$ en el caso continuo.

4.3. Sumas y promedios de variables aleatorias

- Sean $X \sim \mathcal{P}(\lambda)$ e $Y \sim \mathcal{P}(\mu)$ variables aleatorias independientes y sea $V = X + Y$ entonces $V \sim \mathcal{P}(\lambda + \mu)$
- Sean $X \sim \varepsilon(\lambda)$ e $Y \sim \mathcal{P}(\lambda)$ variables aleatorias independientes y sea $V = X + Y$ entonces $V \sim \Gamma(2, \lambda)$
- Sean $X \sim \Gamma(\alpha, \lambda)$ e $Y \sim \Gamma(\beta, \lambda)$ variables aleatorias independientes y sea $V = X + Y$ entonces $V \sim \Gamma(\alpha + \beta, \lambda)$

Propiedad: Sean X_1, \dots, X_n variables aleatorias cualesquiera con $E(X_i) = \mu_i$ y $V(X_i) = \sigma^2$ y a_1, \dots, a_n números reales, entonces:

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

$$V\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \sigma_i^2 + 2 \sum_{i < j} a_i a_j \text{cov}(X_i, X_j)$$

Corolario: Sean X_1, \dots, X_n variables aleatorias idénticamente distribuidas con $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ y a_1, \dots, a_n números reales, entonces:

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

$$V\left(\sum_{i=1}^n X_i\right) = n\sigma^2$$

y

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

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

4.4. Función Generadora de Momentos y sus Propiedades

Si X es una variable aleatoria, el momento de orden k de X se define como $E(X^k)$ siempre que la esperanza exista.

Función generadora de momentos: La función generadora de momentos de una variable aleatoria X es una función $M_X(t)$ a valores reales definida como:

$$M_X(t) = E(e^{tX}) \begin{cases} \sum_{x \in R_X} e^{tx} p_X(x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x) dx & \text{si } X \text{ es continua} \end{cases}$$

siempre que el valor esperado exista para todo $t \in (-h, h)$, $h > 0$.

Teorema: Sea X una v.a. para la cual existe la función generadora de momentos $M_X(t)$ entonces:

$$E(X^n) = \left. \frac{\partial^n}{\partial t^n} M_X(t) \right|_{t=0}$$

Lema: Si la función $g(t)$ definida por

$$g(t) = \sum_x e^{tx} p(x) \quad \text{ó} \quad g(t) = \int_{-\infty}^{\infty} f(x) dx$$

converge para todo $t \in (-h, h)$ para algún $h > 0$, entonces existen las derivadas de orden n de $g(t)$ para todo $t \in (-h, h)$ y para todo n entero positivo y se obtienen como

$$\frac{\partial^n g(t)}{\partial t^n} = \sum_x \frac{\partial^n e^{tx} p(x)}{\partial t^n} \quad \text{ó} \quad \frac{\partial^n g(t)}{\partial t^n} = \int_{-\infty}^{\infty} \frac{\partial^n e^{tx}}{\partial t^n} f(x) dx$$

Propiedad: Sea X una variable aleatoria con función generadora de momentos $M_X(t)$, entonces si $Y = aX + b$, entonces $M_Y(t) = e^{bt} M_X(at)$

Teorema de unicidad: Si existe la función generadora de momentos de una variable aleatoria, es única y determina a la función de densidad o probabilidad de la variable aleatoria salvo a lo

sumo en un conjunto de probabilidades 0.

Distribución	$M_X(t)$
$Bi(n, p)$	$(e^t + 1 - p)^n$
$P(\lambda)$	$e^{\lambda(e^t - 1)}$
$N(\mu, \sigma^2)$	$e^{\frac{\sigma^2 t^2}{2} + \mu t}$
$E(\lambda)$	$\frac{\lambda}{\lambda - t}$
$G(\alpha, \lambda)$	$\left(\frac{\lambda}{\lambda - t}\right)^\alpha$
$U(a, b)$	$\frac{e^{tb} - e^{ta}}{t(b - a)}$
$G(p)$	$\frac{pe^t}{1 - (1 - p)e^t}$
$BN(r, p)$	$\left(\frac{pe^t}{1 - (1 - p)e^t}\right)^r$

Función generadora de momentos en la suma de variables aleatorias independientes:
Sean, en principio X e Y dos variables aleatorias independientes, entonces la función generadora de la suma $X + Y$ es el producto de las funciones generadoras, es decir:

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

4.4.1. Generación de números aleatorios

El método más conocido es el **método de congruencias**. Para generar números con distribución uniforme en el intervalo $(0, 1)$ debemos generar X_n enteros entre 0 y un número natural m y luego tomar la fracción:

$$U_n = \frac{X_n}{m}$$

Dados cuatro números m, a, c y X_0 , formamos la secuencia de números aleatorios X_n de la siguiente forma:

$$X_{n+1} \equiv (aX_n + c) \pmod{m}, \quad n \geq 0$$

es decir que X_{n+1} es el resto entero de dividir $aX_n + c$ por m . Sea U una variable aleatoria con distribución $U(0, 1)$ y G una función de distribución acumulada continua y estrictamente creciente.

Si $X = G^{-1}(U)$, entonces la función de distribución acumulada de X es G , es decir $F_X = G$.

Teorema: Sea U una variable aleatoria con distribución $U(0, 1)$ y G una función de distribución acumulada. Existe una función H tal que $H(U)$ tiene distribución acumulada G .

4.5. Desigualdad de Tchebycheft

Sea X una variable aleatoria con $E(X) = \mu$ y $V(X) = \sigma^2 < \infty$, entonces $\forall \epsilon > 0$

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

Observación: La cota que provee la desigualdad de Chebyshev puede ser grosera o, peor aún, no informativa.

4.5.1. Formas equivalentes de la desigualdad de Chebyshev

■ $\forall \epsilon > 0$

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

■ $\forall k > 1$

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

■ $\forall k > 1$

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

4.6. Ley de los Grandes Números

Sea (X_n) ($n \geq 1$) una sucesión de variables aleatorias, diremos que X_n **converge en probabilidad** a la variable aleatoria X y lo notaremos $X_n \xrightarrow{p} X$, si:

$$\lim_{n \rightarrow \infty} P(|X_n - \mu| > \epsilon) = 0$$

Ley de los Grandes Números: Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas con $E(X) = \mu$ y $V(X) = \sigma^2 < \infty$, entonces

$$\bar{X}_n \xrightarrow{p} \mu$$

siendo $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ el denominado promedio muestral.

Propiedades

- Si X_1, \dots, X_n son variables independientes tales que $X \sim Bi(n_i, p)$, entonces:

$$\sum_{i=1}^n X_i \sim Bi\left(\sum_{i=1}^n n_i, p\right)$$

- Si X_1, \dots, X_n son variables independientes tales que $X \sim \mathcal{P}(\lambda_i)$, entonces:

$$\sum_{i=1}^n X_i \sim \mathcal{P}\left(\sum_{i=1}^n \lambda_i\right)$$

- Si X_1, \dots, X_n son variables independientes tales que $X \sim G(p)$, entonces:

$$\sum_{i=1}^n X_i \sim BN(n, p)$$

- Si X_1, \dots, X_n son variables independientes tales que $X \sim \varepsilon(\lambda)$, entonces:

$$\sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$$

- Si X_1, \dots, X_n son variables independientes tales que $X \sim \Gamma(n_i, \lambda)$, entonces:

$$\sum_{i=1}^n X_i \sim \Gamma\left(\sum_{i=1}^n n_i, \lambda\right)$$

- Si X_1, \dots, X_n son variables independientes tales que $dnom X \mu_i \sigma^2$ y a_1, \dots, a_n son números reales, entonces:

$$\sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma^2\right)$$

Teorema del Central Límite Sean X_1, \dots, X_n variables aleatorias e idénticamente distribuidas con $E(X_i) = \mu$ y $V(X_i) = \sigma^2 < \infty$, sea $T = \sum_{i=1}^n X_i$, entonces si n es suficientemente grande vale que:

$$\frac{T - n\mu}{\sqrt{n}\sigma} \stackrel{(a)}{\sim} N(0, 1) \quad \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \stackrel{(a)}{\sim} N(0, 1)$$

o dicho de otro modo,

$$\frac{T - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} Z \sim N(0, 1) \quad \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} Z \sim N(0, 1)$$

donde la convergencia en distribución (\xrightarrow{d}) se interpreta del siguiente modo:

$$P\left(\frac{T - n\mu}{\sqrt{n}\sigma} \leq a\right) \cong \Phi(a) \quad P\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq a\right) \cong \Phi(a)$$

es decir, que las funciones de distribución convergen a la función de distribución normal standard.

4.6.1. Corrección por continuidad

Cuando se aproxima una distribución discreta por una continua, como es el caso de la aproximación de la distribución binomial por la normal, es necesario efectuar una corrección.

En general, cuando la variable es discreta y $x_i - x_{i-1} = 1$, la corrección se realiza en la forma $P(X \leq a) = P(X \leq a + 0.5)$ y $P(X \geq a) = P(X \geq a - 0.5)$

Proposición: Sean W_1, \dots, W_k variables aleatorias independientes con distribución $E(1)$. Consideremos el siguiente proceso. Comenzamos a medir el tiempo en $t = 0$ y consideramos que ocurre el primer evento en el instante W_1 , el segundo en el instante $W_1 + W_2$, y en general el k -ésimo evento en el instante $W_1 + \dots + W_k$. Si para $t > 0$, definimos la variable aleatoria X_t = cantidad de eventos que ocurren en el intervalo $[0, t]$, entonces X_t es una variable discreta y su distribución es $P(t)$.

Parte II

Estadística

5. Introducción y estadística descriptiva

Los métodos de **Análisis Exploratorio** o **Estadística Descriptiva** ayudan a comprender la estructura de los datos, de manera de detectar tanto un patrón de comportamiento general como apartamientos del mismo.

La **inferencia estadística** nos permite tanto hacer predicciones y estimaciones como decidir entre dos hipótesis opuestas relativas a la población de la cual provienen los datos (test de hipótesis).

Los métodos estadísticos aplicados sobre datos obtenidos a partir de muestras aleatorias, permiten cuantificar el error que podemos cometer en una estimación o calcular la probabilidad de cometer un error al tomar una decisión en un test de hipótesis.

5.1. Estadística descriptiva

Examinaremos los datos en forma descriptiva con el fin de:

- Organizar la información
- Sintetizar la información
- Ver sus características más relevantes
- Presentar la información..

Población: Conjunto total de los sujetos o unidades de análisis de interés en el estudio.

Muestra: Cuaquier subconjunto de sujetos o unidades de análisis de la población en estudio.

Unidad de análisis o observación: Objeto bajo estudio.

Variable: Cualquier caracterísitca de la unidad de observación que interese registrar y que en el momento de ser registrada puede ser transformada en un número.

Valor de una variable, Dato, Observación o Medición: número que describe a la característica de interés en una unidad de observación particular.

Caso o Registro: Conjunto de mediciones realizadas sobre una unidad de observación.

5.2. Datos cuantitativos

5.2.1. Esquema Tallo y Hoja

Nos da una primera aproximación rápida al a distribución de los datos sin perder de vistas las observaciones.

1. Ordenamos los datos de menor a mayor.
2. Separamos cada observación en dos partes: **tallo** y **hoja**.
3. Listamos en forma vertical y creciente los tallos y agregamos las hojas a la derecha del tallo correspondiente

Ejemplo: La siguiente tabla contiene 45 observaciones:

96	93	88	117	127	95	113	96
108	94	148	156	139	142	94	107
125	155	155	103	112	127	117	120
112	135	132	111	125	104	106	139
134	119	97	89	118	136	125	143
120	103	113	124	138			

A continuación se muestra el diagrama tallo hoja:

9	3445667
10	334678
11	122337789
12	00455577
13	2456899
14	238
15	556

¿Que se puede ver en este tipo de diagramas?

- Rango de las observaciones, valores máximos y mínimos.
- Forma de la distribución: simétrica, asimetría a derecha, asimetría a izquierda y cuántos picos tiene la distribución.
- Posición del centro de la distribución y concentración de los datos.
- Desviaciones marcadas respecto al comportamiento general: outliers o valores atípicos.

¿Como elegimos los valores de los tallos? En general se recomienda utilizar entre 8 y 20 tallos. El número de tallos debe ser tal que permita mostrar una imagen general de la estructura del conjunto de datos. Demasiados detalles en general serán poco informativos, demasiado agrupamiento puede distorsionar la imagen del conjunto.

Cuando el volumen de dato es muy grande conviene usar otro tipo de gráficos que tambien son de fácil interpretación.

5.2.2. Gráficos de tallo-hojas espalda con espalda. Comparación de grupos

Los gráficos de tallo-hojas son útiles para comparar la distribución de una variable en dos condiciones o grupos. EL gráfico se denomina tallo-hojas espalda con espalda porque ambos grupos comparte los tallos.

T1		T2
	5	47
	6	2
74	7	37
963	8	778999
660	9	0358
9662	10	222
821	11	7
70	12	
2	13	
	14	
	15	
	16	

5.2.3. Histograma

1. Se divide el rango de los datos en **intervalos o clases** que no se superpongan. Las clases deben ser **excluyentes** y **exhaustivas**.
2. Se cuenta la cantidad de datos en cada intervalo o clase, es decir la **frecuencia**.

3. Se gráfica el histograma en un par de ejes coordenados representando en las abscisas los intervalos y sobre cada uno de ellos un rectángulo cuya área sea proporcional a la frecuencia relativa de dicho intervalo.

No es necesario que todos los intervalos tengan la misma longitud.

Un conjunto de datos que no se distribuye simétricamente, se dice que es **asimétrico**.

5.3. Medidas de Resumen

5.3.1. Medidas de posición o centrado

Un modo de resumir un conjunto de datos numéricos es a través de un número que represente a todos, en el sentido de ser un valor típico para el conjunto. Supongamos que tenemos un conjunto de n datos que genéricamente representaremos por: x_1, \dots, x_n :

Promedio o Media Muestral: Es el punto de equilibrio del conjunto de datos. Es una medida muy sensible a la presencia de datos anómalos (outliers).

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

Mediana muestral: Es una medida el centro de los datos en tanto divide a la muestra ordeada en dos partes de igual tamaño. Deja la mitad de los datos cada lado. La mediana es resistente a la presencia de datos atípicos. También puede ser útil cuando algunos datos han sido censurados.

$$\tilde{x} = \begin{cases} x_{(k+1)} & \text{si } n = 2k + 1 \\ \frac{x_{(k)} + x_{(k+1)}}{2} & \text{si } n = 2k \end{cases}$$

Media α -podada: Es un promedio calculado sobre los datos una vez que se han eliminado $\alpha \cdot 100\%$ de los datos más pequeños y $\alpha \cdot 100\%$ de los datos más grandes. Es una medida intermedia entre la media y la mediana.

$$\bar{x}_\alpha = \frac{x_{([n\alpha]+1)} + \dots + x_{(n-[n\alpha])}}{n - 2[n\alpha]}$$

Otra posible manera de definirla es eliminando $(n \cdot \alpha)$ datos en cada extremo. Si $(n \cdot \alpha)$ es entero y, cuando no lo es, interpolando entre dos medias α -podadas, una en la cual se podan $[n \cdot \alpha]$ en cada extremo y otra en la que se podan $[n \cdot \alpha] + 1$ datos en cada extremo.

La media es una media α -podada con $\alpha = 0$ y la mediana una media podada con α tan próximo a 0.5 como sea posible.

Se elegirá el α dependiendo de cuantos outliers se pretende excluir y de cuán robusta queremos que sea la medida de posición. Una elección bastante común es $\alpha = 0.10$, que excluye un 20% de los datos.

Si la distribución es simétrica, la mediana y la media identifican al mismo punto. Si es asimétrica derecha (cola larga hacia la derecha), entonces $\bar{x} > \tilde{x}$ y si es asimétrica a izquierda (cola larga hacia la izquierda), entonces $\bar{x} < \tilde{x}$.

5.3.2. Medidas de dispersión y variabilidad

Rango muestral: Es la diferencia entre el valor más grande y el más pequeño de los datos:

$$\text{Rango} = \max(X_i) - \min(X_i)$$

Esta medida es muy sensible a la presencia de outliers. Además no capta la dispersión interna del conjunto de datos.

Varianza muestral: Mide la variabilidad de los datos alrededor de la media muestral.

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

Desvío Estándar Muestral: $S = \sqrt{S^2}$

El desvío estándar tiene las mismas unidades que los datos. Y estas medidas son sensibles a la presencia de datos atípicos.

Coefficient de Variación: Es una medida que relaciona el desvío estándar con la media de una muestra:

$$CV = \frac{S}{\bar{x}}$$

No tiene propiedades estadísticas muy interesantes.

Distancia intercuartil: El percentil $\alpha \cdot 100\%$ de la muestra ($0 < \alpha < 1$) es el valor por debajo del cual se encuentra el $\alpha \cdot 100\%$ de los datos en la muestra ordenada.

Para calcularlo:

1. Ordenamos la muestra de menor a mayor.
2. Buscamos el dato que ocupa la posición $\alpha \cdot (n + 1)$. Si este número no es entero se interpolan los dos adyacentes.

La mediana coincide con el percentil 50 %. Llamaremos **cuartil inferior** al percentil 25 % y cuartil superior al percentil 75 %.

Entre los cuartiles se halla aproximadamente el 50 % central de los datos y el rango de estos es:

$$d_i = \text{distancia intercuartil} = \text{cuartil superior} - \text{cuartil inferior}$$

Cuartos y distancias entre cuartos:

- Se ordena la muestra y se calcula la mediana de los datos.
- Dividimos a la muestra ordenada en dos partes: la **primera** corresponde a los datos más pequeños que la mediana y la **segunda** corresponde a los datos más grandes que la mediana.
- Si el tamaño de la muestra es *par*, el **cuarto inferior** es la mediana de la primera mitad, mientras que el **cuarto superior** es la mediana de la segunda mitad.
- Si el tamaño de la muestra es *impar*, a la primera y a la segunda parte se las expande agregándose a cada una de ellas la mediana de todos los datos. El **cuarto inferior** es la mediana de la primera parte expandida y el **cuarto superior** es la mediana de la segunda parte expandida.

Definimos la **distancia entre cuartos** como:

$$d_c = \text{cuarto superior} - \text{cuarto inferior}$$

Desvío Absoluto Mediano (Desviación absoluta respecto de la mediana):

$$MAD = \text{mediana}(|x_i - \tilde{x}|)$$

1. Ordenamos los datos de menor a mayor.
2. Calculamos la mediana.
3. Calculamos la distancia de cada dato a la mediana.
4. Despreciamos el signo de las distancias y las ordenamos de menor a mayor.
5. Buscamos la mediana de las distancias sin signo.

Si deseamos comparar la distancia intercuartil y la MAD , se compara a S con $\frac{MAD}{0.675}$ ó $\frac{d_I}{1.35}$.

Números de resumen: Los 5 números de resumen de la distribución de un conjunto de datos consisten en el **mínimo**, el **cuartil inferior**, la **mediana**, el **cuartil superior** y el **máximo**.

5.3.3. Box-Plots

1. Representamos una escala vertical u horizontal.
2. Dibujamos una caja cuyos extremos son los cuartiles y dentro de ella un segmento que corresponde a la mediana.
3. A partir de cada extremo dibujamos un segmento hasta el dato más alejado que está a lo sumo $1.5d_I$ del extremo de la caja. Estos segmentos se llaman bigotes.

4. Marcamos con un * a aquellos datos que están entre $1.5d_I$ y $3d_I$ de cada extremo y con un círculo a aquellos que están a mas de $3d_I$ de cada extremo.

A partir de un box-plot podemos apreciar los siguientes aspectos de la distribución e un conjunto dado:

- Posición
- Dispersión
- Asimetría
- Longitud de las colas
- Puntos anómalos o outliers.

Outliers: Su detección es importante pues pueden determinar o influenciar fuertemente los resultados de un análisis estadístico clásico. Si no hay evidencia de error y su valor es posible no deberían ser eliminados. Asimismo, la presencia de outliers puede indicar que la escala elegida no es la más adecuada.

Boxplots paralelos Una aplicación muy útil de los boxplots es la comparación de la distribución de dos o más conjuntos de datos graficando en una escala común los boxplots de cada una de las muestras.

QQ-Plot (Normal Probability Plot): El QQ-Plot es un gráfico que nos sirve para evaluar la cercanía de una distribución dada, en particular a la distribución normal.

Consideremos a la muestra aleatoria X_1, \dots, X_n y los correspondientes estadísticos de orden $X_{(1)}, \dots, X_{(n)}$. En el QQ-Plot se grafican en el eje de abscisas los percentiles de la distribución teórica (en nuestro caso normal) y en el eje de ordenadas las observaciones ordenadas, que pueden ser vistas como percentiles empíricos.

6. Estimación puntual

La mayoría de las distribuciones de probabilidad dependen de cierto número de parámetros.

El objetivo de la *estimación puntual* es usar una muestra para obtener números que, en algún sentido, sean los que mejor representan a los verdaderos valores de los parámetros de interés.

Un **estimador puntual** de un parámetro θ , es un valor que puede ser considerado representativo de θ y se indicará $\hat{\theta}$. Se obtiene a partir de alguna función de la muestra.

6.1. Métodos de estimación puntual

6.1.1. Método de momentos

Dada una muestra aleatoria X_1, \dots, X_n se denomina **momento muestral de orden k** a

$$\frac{\sum_{i=1}^n X_i^k}{n}$$

Sea X_1, \dots, X_n una muestra aleatoria de una distribución con función de probabilidad puntual o función de densidad que depende de m parámetros $\theta_1, \dots, \theta_m$. Los estimadores de momentos de $\theta_1, \dots, \theta_m$ son los valores $\hat{\theta}_1, \dots, \hat{\theta}_m$ que se obtienen igualando m momentos poblacionales con los correspondientes momentos muestrales. En general, se obtienen resolviendo el siguiente sistema de ecuaciones:

$$\frac{\sum_{i=1}^n X_i^k}{n} = E(X^k) \quad k = 1, 2, \dots, m$$

6.1.2. Método de máxima verosimilitud

Sean X_1, \dots, X_n variables aleatorias con función de probabilidad conjunta $p_{\bar{X}}(x_1, \dots, x_n)$ o función de densidad conjunta $f_{\bar{X}}(x_1, \dots, x_n)$ que depende de m parámetros $\theta_1, \dots, \theta_m$. Cuando (x_1, \dots, x_n) son los valores observados y la función de probabilidad o de densidad conjunta se considera función de los parámetros $\theta_1, \dots, \theta_m$, se denomina **función de verosimilitud** y se denota $L(\theta_1, \dots, \theta_m)$.

Los estimadores de máxima verosimilitud (EMV) de $\theta_1, \dots, \theta_m$ son los valores $\hat{\theta}_1, \dots, \hat{\theta}_m$ que maximizan la función de verosimilitud, o sea los valores tales que

$$L(\hat{\theta}_1, \dots, \hat{\theta}_m) \geq L(\tilde{\theta}_1, \dots, \tilde{\theta}_m)$$

La forma general de los EMV se obtiene reemplazando los valores observados x_i por las variables X_i

Propiedad de la invarianza de los EMV: Sea $\hat{\theta}$ el EMV de θ y sea h una función inyectiva con dominio en el rango de valores posibles de θ , entonces el EMV de $h(\theta)$ es $h(\hat{\theta})$

6.1.3. Propiedades de los estimadores y criterios de selección

Dada una muestra X_1, \dots, X_n , donde $X_i \sim F_\theta$, un estimador puntual del parámetro θ . La diferencia

$$\hat{\theta} - \theta$$

es el error de estimación y una estimación será más precisa cuanto menor sea este error.

Definición: Un estimador puntual $\hat{\theta}$ del parámetro θ es **insesgado** si

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

Si $\hat{\theta}$ no es insesgado, se denomina **sesgo** de $\hat{\theta}$ a $b(\hat{\theta}) = E_{\theta}(\hat{\theta}) - \theta$.

Definición: Un estimador puntual $\hat{\theta}$ del parámetro θ basado en una muestra X_1, \dots, X_n es **asintóticamente insesgado** si

$$E_{\theta}(\hat{\theta}) \xrightarrow{n \rightarrow \infty} \theta \quad \forall \theta$$

Principio de estimación insesgada de mínima varianza: Entre todos los estimadores insesgados de θ , elegir el de menor varianza. El estimador resultante se denomina **Inssegado de Mínima Varianza Uniformemente** (IMVU).

Teorema: Sea X_1, \dots, X_n una muestra aleatoria de una distribución $N(\mu, \sigma^2)$. Entonces \bar{X} es estimador IMVU de μ

Definición: El error estándar de un estimador $\hat{\theta}$ es su desviación estándar, es decir

$$\sigma = \sqrt{V_{\theta}(\hat{\theta})}$$

Si el error standard depende de parámetros desconocidos, éstos se reemplazan por un estimador y se obtiene el error standard estimado.

Definición: Sea $\hat{\theta}$ un estimador de θ , su error cuadrático medio es:

$$ECM_{\theta}(\hat{\theta}) = E_{\theta} \left[\left(\hat{\theta} - \theta \right)^2 \right]$$

Si el estimador $\hat{\theta}$ es insesgado el error cuadrático medio es igual a la varianza del estimador.

Proposición: $ECM_{\theta}(\hat{\theta}) = V_{\theta}(\hat{\theta}) + [b(\hat{\theta})]^2$, siendo $b(\hat{\theta})$ el sesgo del estimador.

Principio de estimación del menor error cuadrático medio: Dados dos o más estimadores del parámetro θ , elegir el menor ECM. Si el estimador sesgado tiene una varianza mucho menor que el insesgado, podría ser preferible su uso.

Definición: Sea X_1, \dots, X_n una muestra aleatoria de una distribución que depende de un parámetro θ y sea $\hat{\theta}_n$ un estimador puntual de θ basado en esa muestra. Diremos que $\{\hat{\theta}_n\}$ es una sucesión **consistente** (o más brevemente que $\hat{\theta}_n$ es un estimador consistente de θ) si

$$\hat{\theta}_n \xrightarrow{p} \theta$$

es decir, si $\forall \epsilon > 0, P(|\hat{\theta}_n - \theta| > \epsilon) \xrightarrow{n \rightarrow \infty} 0$

Proposición: Sean X_1, \dots, X_n una muestra aleatoria de una distribución que depende de un parámetro θ y sea $\hat{\theta}_n$ un estimador de θ basado en la muestra de tamaño n . Si

1. $E_{\theta}(\hat{\theta}_n) \xrightarrow{n \rightarrow \infty} \theta$
2. $V_{\theta}(\hat{\theta}_n) \xrightarrow{n \rightarrow \infty} 0$

entonces, $\hat{\theta}_n$ es consistente de θ .

7. Intervalos de confianza

Sea X_1, \dots, X_n una muestra aleatoria de una distribución que depende un parámetro θ . Dadas dos funciones de la muestra $a(X_1, \dots, X_n)$ y $b(X_1, \dots, X_n)$ tales que

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

con α pequeño, el intervalo $[a(X_1, \dots, X_n), b(X_1, \dots, X_n)]$ se denomina **intervalo de confianza de nivel $1 - \alpha$** para el parámetro θ .

Una vez construido el intervalo a partir de una muestra dada, tenemos *confianza* de que el intervalo contenga a θ .

7.1. Intervalos de confianza para los parámetros de una distribución normal

Sean dos variables aleatorias $Z \sim N(0, 1)$ y $U \sim \chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ independientes, entonces

$$T = \frac{Z}{\sqrt{\frac{U}{n}}} \sim t_n$$

Y se dice que T tiene distribución t de Student con n grados de libertad. Esta distribución está tabulada para diferentes valores de n . Su densidad es simétrica respecto al 0 y tiene forma de campana, pero tiene colas más pesadas que la distribución normal standard.

Proposición: Sean X_1, \dots, X_n una muestra aleatoria de una distribución $N(\mu, \sigma^2)$, entonces

■

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \iff \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1)$$

■

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2 \quad \text{con } S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

■ \bar{X} y S^2 son independientes.

■

$$\sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{n-1}$$

7.2. Intervalos de confianza para la media de la distribución normal con varianza conocida

Sean X_1, \dots, X_n una muestra aleatoria de una distribución $N(\mu, \sigma_0^2)$ con varianza σ^2 conocida, entonces

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma_0} \sim N(0, 1)$$

y

$$P\left(-z_{\alpha/2} \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma_0} \leq z_{\alpha/2}\right) = 1 - \alpha$$

de donde se deduce el siguiente intervalo de confianza de nivel $1 - \alpha$ para μ ,

$$\left[\bar{X} - z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}}\right]$$

7.3. Intervalos de confianza para la media de la distribución normal con varianza desconocida

Sean X_1, \dots, X_n una muestra aleatoria de una distribución $N(\mu, \sigma^2)$, entonces

$$\sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{n-1}$$

y el intervalo de confianza de nivel $1 - \alpha$ para μ es:

$$\left[\bar{X} - t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}\right]$$

7.4. Intervalos de confianza para la varianza de la distribución normal con media conocida

Sean X_1, \dots, X_n una muestra aleatoria de una distribución $N(\mu_0, \sigma^2)$ con media μ_0 conocida, entonces

$$\frac{X_i - \mu_0}{\sigma} \sim N(0, 1)$$

y el intervalo de confianza de nivel $1 - \alpha$ para μ es:

$$\left[\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n, \alpha/2}^2}, \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n, 1-\alpha/2}^2}\right]$$

7.5. Intervalos de confianza para la varianza de la distribución normal con media desconocida

Sean X_1, \dots, X_n una muestra aleatoria de una distribución $N(\mu, \sigma^2)$, entonces

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

y el intervalo de confianza de nivel $1 - \alpha$ para μ es:

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

7.6. Método general para obtener intervalos de confianza

7.6.1. Determinación del tamaño de muestra

Consideremos el intervalo de confianza para μ con varianza conocida en el caso de una muestra aleatoria normal. La longitud del intervalo obtenido es

$$L = 2z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}}$$

Si se desea un intervalo de longitud menor que igual que L_0 , entonces

$$n \geq \left(\frac{2z_{\alpha/2}\sigma_0}{L_0} \right)^2$$

7.6.2. Metodo general

Sea X_1, \dots, X_n una muestra aleatoria de una distribución que depende un parámetro θ . Supongamos que existe una función $T(X_1, \dots, X_n, \theta)$ cuya distribución no depende de θ ni de ningún otro parámetro desconocido. Entonces, existen dos variables a y b tales que

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

y, a partir de esta expresión, es posible obtener un intervalo de confianza para θ .

La función $T(X_1, \dots, X_n, \theta)$ se denomina **pivote**.

7.7. Intervalos de confianza de nivel asintótico $1-\alpha$

Sea X_1, \dots, X_n una muestra aleatoria de una distribución que depende un parámetro θ . Dadas dos sucesiones $\{a_n(X_1, \dots, X_n)\}$ y $\{b_n(X_1, \dots, X_n)\}$ tales que

$$\lim_{n \rightarrow \infty} P(a_n(X_1, \dots, X_n) \leq \theta \leq b_n(X_1, \dots, X_n)) = 1 - \alpha$$

la sucesión de intervalos $[a_n(X_1, \dots, X_n), b_n(X_1, \dots, X_n)]$ es una sucesión de **intervalos de confianza de nivel asintótico $1-\alpha$** para el parámetro θ . También, se dice que, si n es suficientemente grande, el intervalo $[a_n(X_1, \dots, X_n), b_n(X_1, \dots, X_n)]$ tiene nivel aproximado $1 - \alpha$

Propiedad

$$\left. \begin{array}{l} Y_n \xrightarrow{d} Y \\ U_n \xrightarrow{p} a \end{array} \right\} \Rightarrow U_n Y_n \xrightarrow{d} aY$$

8. Test de Hipótesis

Se proponen dos hipótesis: A la primera se la denomina **hipótesis nula** y se designa H_0 . Esta hipótesis implica que no hay efecto, es la hipótesis del status quo, o sea la de no cambio respecto de la situación inicial. La segunda hipótesis se denomina **hipótesis alternativa** y se designa H_1 . Se la puede llamar la hipótesis del investigador.

Un test es una regla de decisión basada en un **estadístico** o función de la muestra, en este caso \bar{X} , y en una **zona de rechazo**, es decir un conjunto de valores para los cuáles se rechaza la hipótesis nula H_0 .

Al tomar una decisión en base a una muestra, podemos cometer dos tipos de error:

	No se rechaza H_0	Se rechaza H_0
H_0 es cierta	OK	Error tipo I
H_0 no es cierta	Error tipo II	OK

Llamaremos **nivel de significación del test**, y lo designaremos α , a la *probabilidad de error tipo I* y designaremos β a la *probabilidad de error tipo II*.

Como el estadístico se contruye bajo la condición de que H_0 es verdadera, lo que podemos controlar es la probabilidad de error tipo I.

Definición: La **función de potencia** es un test, $\pi(\mu)$, es la probabilidad de rechazar la hipótesis nula cuando el valor verdadero del parámetro es μ .

$$\pi(\mu) = \begin{cases} \alpha(\mu) & \text{si } \mu \in H_0 \\ 1 - \beta(\mu) & \text{si } \mu \notin H_0 \end{cases}$$

donde $\alpha(\mu)$ y $\beta(\mu)$ denota las probabilidad de error tipo I y tipo II respectivamente cuando el verdadero valor del parámetro es μ .

8.1. Tipos de hipótesis a testear

Hipotesis unilaterales

- $H_0: \theta = \theta_0$ (ó $\theta \leq \theta_0$) vs $H_1: \theta > \theta_0$
- $H_0: \theta = \theta_0$ (ó $\theta \geq \theta_0$) vs $H_1: \theta < \theta_0$

Hipotesis bilaterales

- $H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$

8.2. Test de hipótesis de nivel α para los parámetros de la distribución normal

Sea X_1, \dots, X_n una muestra aleatoria de una distribución $N(\mu, \sigma^2)$

Test para la media cuando la varianza es conocida: Supongamos que $\sigma^2 = \sigma_0^2$ es conocida y consideremos las siguientes hipótesis:

1. $H_0: \mu = \mu_0$ (ó $\mu \leq \mu_0$) vs $H_1: \mu > \mu_0$
2. $H_0: \mu = \mu_0$ (ó $\mu \geq \mu_0$) vs $H_1: \mu < \mu_0$
3. $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

El estadístico del test es:

$$T = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma_0}$$

y $T \sim N(0, 1)$.

Región de rechazo:

1. $T \geq z_\alpha$
2. $T \leq -z_\alpha$
3. $|T| \geq z_{\alpha/2}$

Test para la media cuando la varianza es desconocida: Supongamos ahora que la varianza es desconocida y consideremos las mismas hipótesis sobre μ :

1. $H_0: \mu = \mu_0$ (ó $\mu \leq \mu_0$) vs $H_1: \mu > \mu_0$
2. $H_0: \mu = \mu_0$ (ó $\mu \geq \mu_0$) vs $H_1: \mu < \mu_0$
3. $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

El estadístico del test es:

$$T = \sqrt{n} \frac{\bar{X} - \mu_0}{s_0}$$

y $T \sim t_{n-1}$.

Región de rechazo:

1. $T \geq t_{n-1, \alpha}$
2. $T \leq -t_{n-1, \alpha}$
3. $|T| \geq t_{n-1, \alpha/2}$

Test para la varianza cuando la media es desconocida: Las hipótesis a testear son:

1. $H_0: \sigma^2 = \sigma_0^2$ (ó $\sigma^2 \leq \sigma_0^2$) vs $H_1: \sigma^2 > \sigma_0^2$
2. $H_0: \sigma^2 = \sigma_0^2$ (ó $\sigma^2 \geq \sigma_0^2$) vs $H_1: \sigma^2 < \sigma_0^2$
3. $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$

El estadístico del test es:

$$U = \frac{(n-1)S^2}{\sigma_0^2}$$

y $U \sim \chi_{n-1}^2$.

Región de rechazo:

1. $U \geq \chi_{n-1, \alpha}^2$
2. $U \leq -\chi_{n-1, \alpha}^2$
3. $|U| \geq \chi_{n-1, \alpha/2}^2$ ó $|U| \leq \chi_{n-1, 1-\alpha/2}^2$

8.3. Test de hipótesis de nivel aproximado α para la media de una distribución cualquiera

Sea X_1, \dots, X_n una muestra aleatoria de una distribución con media μ y varianza $\sigma^2 < \infty$. Aplicando el Teorema Central del Límite, sabemos que

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} Z \sim N(0, 1)$$

Además, utilizando la propiedad enunciada al construir intervalos de confianza de nivel asintótico $(1 - \alpha)$ para la media de una distribución cualquiera,

$$\left. \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \xrightarrow{d} N(0, 1) \right\} \Rightarrow \sqrt{n} \frac{\bar{X} - \mu}{S} \xrightarrow{d} N(0, 1)$$

Por lo tanto, si n es suficientemente grande,

$$\sqrt{n} \frac{\bar{X} - \mu}{S} \xrightarrow{d} N(0, 1)$$

Supongamos que se desea testear a nivel aproximado α alguna de la hipótesis siguientes:

1. $H_0: \mu = \mu_0$ (ó $\mu \leq \mu_0$) vs $H_1: \mu > \mu_0$
2. $H_0: \mu = \mu_0$ (ó $\mu \geq \mu_0$) vs $H_1: \mu < \mu_0$
3. $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

y que n es suficientemente grande. Utilizando como estadístico

$$T = \sqrt{n} \frac{\bar{X} - \mu_0}{S}$$

y las siguientes regiones de rechazo:

Región de rechazo:

1. $T \geq z_\alpha$
2. $T \leq -z_\alpha$
3. $|T| \geq z_{\alpha/2}$

8.4. Relación entre tests de hipótesis bilaterales e intervalos de confianza

Sea X_1, \dots, X_n una muestra aleatoria de una distribución $N(\mu, \sigma^2)$. Sabemos que el intervalo de confianza para μ de nivel $1 - \alpha$ está dado por

$$\left[\bar{X} - t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} \right]$$

Deseamos testear a nivel α las siguientes hipótesis:

$$H_0: \mu = \mu_0 \quad \text{vs} \quad H_1: \mu \neq \mu_0$$

Podríamos construir un test de nivel α rechazando H_0 si μ_0 no pertenece al intervalo de confianza.

Proposición: Sea $IC(X_1, \dots, X_n)$ un intervalo de confianza de nivel $1 - \alpha$ para un parámetro θ , obtenido a partir de una muestra aleatoria X_1, \dots, X_n . Consideremos el problema de testear las hipótesis:

$$H_0: \theta = \theta_0 \quad \text{vs} \quad H_1: \theta \neq \theta_0$$

El test que rechaza H_0 cuando $\theta_0 \notin IC(X_1, \dots, X_n)$, tiene nivel α .