

1 Introducción

Se define el **experimento aleatorio** como un ejemplo no determinístico o aleatorio, ya que al realizar un experimento de esta índole, me limito al resultado del azar. Por ej. *Se lanza un dado y se observa el número que aparece en la cara superior.*

Características

- Es posible repetir el experimento indefinidamente sin cambiar esencialmente las condiciones en que se realiza.
- En general no podemos identificar un resultado particular, pero si se puede indicar el conjunto de todos los resultados posibles, denominado **espacio muestral**.
- Al repetir el experimento un gran número de veces, es posible el lograr distinguir un patrón particular de regularidad. Este patrón hace posible modelizar el experimento mediante un modelo matemático.

Probabilidad: Consiste en poder calcular que tan factible es algo. Para esto debo de conocer los datos de mis variables y con ello calculo la probabilidad de un evento.

Estadística: Mediante el resultado de un experimento aleatorio (obtengo una muestra), quiero estimar valores de probabilidad.

2 Estadística descriptiva

Parámetros

1. Media muestral: $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$. Es el promedio de los datos.
2. Mediana: m . En un ordenamiento de menor a mayor es el término central, a diferencia de la media no es sensible a los valores alejados de la media.
3. Cuartiles primero y tercero: q_1, q_3 . En un ordenamiento de menor a mayor son mediana de la primera mitad y de la segunda mitad de los datos.
4. Rango muestral: $R = M\acute{A}X - M\acute{I}N$. Distancia entre el mínimo y el máximo de los datos muestrales.
5. Rango intercuartil: $IQR = q_3 - q_1$. La distancia entre los cuartiles primero y tercero.
6. Desvío estándar muestral: $s = \sqrt{\frac{1}{n-1} \sum_{j=1}^n [x_j - \bar{x}]^2}$. Medida típica de variación de estos datos, es nula si y sólo si todos los datos son iguales y es tanto mayor cuanto más dispersos están los datos.

7. Coeficiente de simetría: $\gamma = \frac{1}{ns^3} \sum_{j=1}^n [x_j - \bar{x}]^3$.Medida típica de simetría de los datos respecto de la media. Es nulo si hay simetría. Si es positivo se dice que los datos tienen sesgo positivo y negativo en caso contrario.
8. Coeficiente de curtosis: $k = \frac{1}{ns^4} \sum_{j=1}^n [x_j - \bar{x}]^4 - 3$.Mide la forma de la distribución de datos en torno del promedio. Es positivo si los datos tienen alta concentración en torno de la media y negativo en caso contrario.
9. Coeficiente de variación: $CV = \frac{S}{|\bar{x}|}$. Es una medida adimensional que indica qué proporción representa la desviación estándar respecto de la media aritmética. Se utiliza con frecuencia en la comparación de la variabilidad de dos o más conjuntos de datos que difieren en unidades y/o magnitudes.

Si las frecuencias de las variables en la muestra son distintas:

$$\bar{x} = \frac{1}{n} \sum_{j=1}^m x_j f_j$$

$$s = \sqrt{\frac{1}{n-1} \sum_{j=1}^n [x_j - \bar{x}]^2 f_j} \rightarrow s = \sqrt{s^2} \quad \therefore s > 0$$

$$\gamma = \frac{1}{ns^3} \sum_{j=1}^n [x_j - \bar{x}]^3 f_j$$

$$k = \frac{1}{ns^4} \sum_{j=1}^n [x_j - \bar{x}]^4 f_j - 3$$

Según el coeficiente de Kurtosis:

- $k = 0$: El peso de las colas es similar al de una distribución normal.
- $k > 0$: El peso de las colas es menor al de una distribución normal.
- $k < 0$: El peso de las colas es mayor al de una distribución normal.

3 Axiomas de la probabilidad

3.1 Álgebra de sucesos

Se define un evento o suceso como un caso que quier evaluado, el cual puede ser una posibilidad o no dentro del espacio muestral, pero si que es requisito poder definirlo con una probabilidad de que suceda, aun así si esta probabilidad es nula (no pertenece al espacio muestral).

Sea S (espacio muestral) un conjunto y \sum una colección de subconjuntos (sucesos/eventos) de S . \sum es un σ - álgebra de \sum si:

- $S \in \sum$
- $A \in \sum \Rightarrow A^c \in \sum$
- $A_1, A_2, \dots \in \sum \Rightarrow A_1 \cup A_2 \cup \dots \in \sum$

Dos eventos A, B son **mutuamente excluyentes** $\leftrightarrow A \cap B = \phi$.

El triplete (S, \sum, P) es un **espacio de probabilidad** si $S \neq \phi$, \sum es un σ - álgebra de S y $P : \sum \rightarrow \mathbb{R}$ cumple las siguientes condiciones:

1. $P(A) \geq 0 \quad \forall A \in \sum$
2. $P(S) = 1$
3. E_1, E_2, \dots son mutuamente excluyentes en $\sum \Rightarrow P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$

Obs: Sea un evento $A \in \sum \Rightarrow A, A^c$ son mutuamente excluyentes $\rightarrow P(A^c) = 1 - P(A)$.

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

4 Métodos de enumeración

La regla de Laplace es muy útil para el cálculo de probabilidades cuando **todos los casos son igualmente probables**:

$$P(A) = \frac{\# \text{ casos favorables}}{\# \text{ total de casos}}$$

La combinatoria es una buena herramienta para calcular la cantidad de casos posibles como totales donde quiero distribuir m cosas en n espacios ($n > m$), sin importarme las permutaciones.

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

4.1 Permutaciones

- Si tenemos n objetos diferentes, se pueden agrupar (permutar) en $nP_n = n!$ formas distintas.
- Si tenemos n objetos diferentes y queremos escoger r de esos objetos, $0 \leq r \leq n$, y permutamos el r elegido \rightarrow Hay $nP_r = V_{n,r} = \frac{n!}{(n-r)!}$

4.2 Combinaciones

Si tenemos n objetos diferentes, y queremos contar el número de maneras como podemos escoger r de esos objetos, sin considerar el orden \rightarrow Hay $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ posibilidades.

4.3 Permutaciones cuando no todos los objetos son diferentes

Si tenemos n objetos tales que hay n_k de k clases distintas \rightarrow El número de permutaciones de estos n objetos está dado por:

$$P_n^{n_1 n_2 \dots n_k} = \frac{n!}{\prod_{i=1}^k n_i}$$

5 Independencia y probabilidad condicional

5.1 Sucesos independientes

Dos sucesos A, B son **independientes** $\leftrightarrow P(A \cap B) = P(A) \cdot P(B)$. Además, al ser independientes entre sí, también se cumple que: A, B^c ; A^c, B ; A^c, B^c serán independientes entre sí.

Prop: Sean los sucesos A_1, \dots, A_n independientes $\Rightarrow P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot \dots \cdot P(A_n)$

Prop: Si A y B son sucesos independientes $\Rightarrow P(A|B) = P(A) \wedge P(B|A) = P(B)$.

Se definen los sucesos A_1, A_2, \dots, A_k como una partición del espacio muestral S si:

- $A_i \neq \phi \quad i = 1, \dots, k$
- $A_i \cap A_j = \phi \quad i \neq j \quad i, j \in [1, k]$
- $S = \bigcup_{i=1}^k A_i$

5.2 Probabilidad condicional

Dados dos sucesos A, B con $P(B) > 0$, la probabilidad de que suceda A sabiendo que sucedió B , se conoce como **probabilidad condicional** y se define tal que:

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

Esta última relación dado el caso inverso, y como $P(A \cap B) = P(B \cap A) \rightarrow$ Llegamos a esa expresión.

Prop: $P(A) = P(B) \cdot P(A|B) + P(B^c) \cdot P(A|B^c)$.

Obs: $P(A \cap B^c) = P(A) - P(A \cap B)$.

5.3 Teorema de la probabilidad total

Sea A_1, A_2, \dots, A_k una partición del espacio muestral S , y B un suceso cualquiera, entonces:

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

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

6 Variables aleatorias discretas

Las variables aleatorias son funciones que transforman el espacio muestral S en valores reales, es decir: $X : S \rightarrow \mathbb{R}$. Por lo tanto, la probabilidad de cada valor de X está asociada a la probabilidad del espacio muestral S .

Para cada variable aleatoria X , se puede buscar cuáles son los valores que tienen asociada una probabilidad positiva. Este conjunto de valores se denomina **recorrido** y se nota R_X . Dada una variable aleatoria discreta X , se denomina la función de probabilidad puntual p_x a la función que a cada valor real le asigna su probabilidad:

$$p_x : \mathbb{R} \rightarrow [0, 1] / p_x(k) = P(X = k)$$

Obs: Dada una variable aleatoria discreta X y su recorrido R_x , se puede deducir que los eventos asociados a valores diferentes de X son mutuamente excluyentes, ya que X no puede tomar dos valores distintos simultáneamente.

Def: Las probabilidades de todos los valores del recorrido deben sumar 1: $\sum_{k \in R_x} p_x(k) = 1$.

Dada una variable aleatoria discreta X , su **valor esperado** se nota tal que: $E(X) = \mu_x = \sum_{k \in R_x} k \cdot p_x(k)$

6.1 Propiedades

- Sea una cte $c \rightarrow E(c) = c$.
- Sea una cte a y X una variable aleatoria discreta $\rightarrow E(a \cdot x) = a \cdot E(x)$.
- Sean X, Y dos variables aleatorias discretas $\rightarrow E(X + Y) = E(X) + E(Y)$.
- $E(X^2) = \sum_{k \in R_x} k^2 p_x(k)$.

Se define la **varianza** como $V(X) = E(X^2) - E^2(X)$.

Se define el **desvío** tal que $\sigma(X) = \sqrt{V(X)}$.

De la definición de tanto la varianza como el desvío, nacen otras propiedades:

- Sea c una cte $\rightarrow V(c) = \sigma(c) = 0$.
- Sea a una cte y X una variable aleatoria discreta $\rightarrow V(a \cdot X) = a^2 \cdot V(X) \Rightarrow \sigma(a \cdot X) = |a| \cdot \sigma(X)$.

- Sea c una cte y X una variable aleatoria discreta $\rightarrow V(X+c) = V(X) \Rightarrow \sigma(X+c) = \sigma(X)$
- Sean X, Y dos variables aleatorias discretas, por lo general vale que $V(X+Y) \neq V(X) + V(Y)$.

Sea X una variable aleatoria discreta, se define la **función de distribución acumulada** tal que:

$$F_X : \mathbb{R} \rightarrow [0, 1] / F_X(k) = P(X \leq k)$$

Algunas propiedades de la función de distribución acumulada son:

- Es monótona creciente $\rightarrow k_1 < k_2 \Rightarrow F_X(k_1) < F_X(k_2)$.
- $\lim_{k \rightarrow -\infty} F_X(k) = 0$.
- $\lim_{k \rightarrow +\infty} F_X(k) = 1$.
- Es continua a derecha $\rightarrow \lim_{a \rightarrow k^+} F_X(a) = F_X(k)$.
- $P(a < X \leq b) = F_X(b) - F_X(a)$.

Sea una variable aleatoria X (K valores posibles x_k)

$$\begin{aligned} E(X) &= \sum_{k=1}^K x_k \cdot P(X = k) \\ V(X) &= \sum_{k=1}^K (x_k - E(X))^2 \cdot P(X = k) \\ \sigma(X) &= \sqrt{V(X)} \\ \gamma(X) &= \frac{\sum_{k=1}^K (x_k - E(X))^3 \cdot P(X = k)}{\sigma^3(X)} \\ k(X) &= \frac{\sum_{k=1}^K (x_k - E(X))^4 \cdot P(X = k)}{\sigma^4(X)} - 3 \end{aligned}$$

6.2 V.A.D Notables

1. $X \sim \text{Bernoulli}(p)$: Sea X la variable aleatoria discreta correspondiente al éxito (= 1) o fracaso (= 0) en un experimento con probabilidad p de éxito.
2. $X \sim \text{Binomial}(n, p)$: Sea X la variable aleatoria discreta correspondiente al número de éxitos en una secuencia de n experimentos independientes cuyos únicos resultados posibles son éxito o fracaso. Cada uno de los experimentos tiene una probabilidad p de éxito.

Obs: $X \sim \text{Bernoulli}(p) \leftrightarrow X \sim \text{Binomial}(1, p)$

3. $X \sim \text{Hipergeométrica}(N, M, n)$: Considere una población con N elementos, M de los cuales son de un tipo dado. Se toma una muestra de tamaño n al azar y sin reposición y sea X la variable aleatoria discreta correspondiente al número de elementos en la muestra que son del tipo especial.

Obs: Si $N, N - M, M \gg n \rightarrow$ Es una binomial con $p = \frac{M}{N} \rightarrow X \sim \text{Binomial}(n, p)$

4. $X \sim \text{Geométrica}(p)$: Se repiten experimentos independientes y con igual probabilidad de éxito hasta que se obtenga un éxito.
5. $X \sim \text{Binomial negativa}(r, p)$: Se repiten experimentos independientes y con igual probabilidad de éxito hasta que se obtengan r éxitos.

| Tipos | R_x | $P(X = k)$ | μ_x | σ_x^2 |
|-----------------------------------|----------------|--|-----------------------|---|
| $\text{Bernoulli}(p)$ | $\{0, 1\}$ | p | p | $p \cdot q$ |
| $\text{Binomial}(n, p)$ | \mathbb{N}_0 | $\binom{n}{k} p^k q^{n-k}$ | $n \cdot p$ | $n \cdot p \cdot q$ |
| $\text{Hipergeométrica}(N, M, n)$ | - | $\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$ | $n \cdot \frac{M}{N}$ | $n \cdot \frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{N-n}{N-1}$ |
| $\text{Geométrica}(p)$ | \mathbb{N}_0 | $q^k p$ | $\frac{q}{p}$ | $\frac{q}{p^2}$ |
| $\text{Binomial negativa}(r, p)$ | \mathbb{N}_0 | $\binom{k+r-1}{k} q^k p^r$ | $r \frac{q}{p}$ | $r \frac{q}{p^2}$ |
| $\text{Poisson}(\lambda)$ | \mathbb{N}_0 | $\frac{\lambda^k}{k!} e^{-\lambda}$ | λ | λ |

7 Variable aleatoria continua

Sea X una variable aleatoria continua $\leftrightarrow F_X$ es continua. Además, F_x cumple las mismas propiedades que para el dominio discreto:

- $\alpha \leq \beta \Rightarrow F_X(\alpha) \leq F_X(\beta)$.
- $\lim_{\alpha \rightarrow -\infty} F_X(\alpha) = 0$.
- $\lim_{\alpha \rightarrow +\infty} F_X(\alpha) = 1$.

Se define la densidad de probabilidad como $f_x(x) = \frac{dF_X(x)}{dx} \forall x \in \mathbb{R}$ donde F_X sea derivable.

7.1 Propiedades

- $\int_{-\infty}^{\infty} f_x(x) dx = 1$.
- $f_x(x) \geq 0$.
- $E(X) = \mu_x = \int_{-\infty}^{\infty} x \cdot f_x(x) dx$.
- $Var(X) = \sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 \cdot f_x(x) dx$.
- $P(X \leq x) = \int_{-\infty}^x f_x(x) dx = F_X(X = x)$.

7.2 V.A.C Notables

7.2.1 $X \sim Uniforme(a, b)$

$$\bullet F_X(x): \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases} \quad \bullet f_x(x): \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & x \notin (a, b) \end{cases} \quad \bullet \mu_x: \frac{a+b}{2} \quad \bullet \sigma_x^2: \frac{(b-a)^2}{12}$$

7.2.2 $X \sim Exponencial(\lambda)$

$$\bullet F_X(x): \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases} \quad \bullet f_x(x): \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x > 0 \end{cases} \quad \bullet \mu_x: \frac{1}{\lambda} \quad \bullet \sigma_x^2: \frac{1}{\lambda^2}$$

7.2.3 $Z \sim N(\mu, \sigma)$

$$\bullet F_X(x): \phi\left(\frac{x-\mu}{\sigma}\right) = \Phi(z) \quad \bullet f_x(x): \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} \quad \bullet \mu_x: \mu \quad \bullet \sigma_x^2: \sigma^2$$

Notar que la normal estándar surge de la gaussiana mediante el cambio de variable $z = \frac{x-\mu}{\sigma}$.

Obs: Utilizamos la normal estándar, ya que la función de densidad de probabilidad

$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ de la normal o gaussiana, no tiene primitiva \rightarrow Es muy difícil calcular las probabilidades, ya que para eso utilizamos la función de acumulación, que es la integral de la densidad en el intervalo que queremos evaluar nuestra probabilidad.

7.3 Propiedades de la normal estándar

- $\phi(z) = 1 - \phi(-z).$
- $z = \phi^{-1}(\alpha).$
- $\phi(-z_\alpha) = 1 - \alpha \rightarrow z_{1-\alpha} = -z_\alpha.$

1 Función de variable aleatoria

Una función de una variable aleatoria es una variable aleatoria (Y) compuesta de otra variable aleatoria (X). Por ejemplo, tengo una variable aleatoria $V \sim N(0, 1)$, y otra variable aleatoria $W = \frac{V^2}{R}$, siendo R una cte. Si quiero hallar la función de densidad de probabilidad de V , puedo arrancar por la función de distribución acumulada:

$$F_W(w) = P(W \leq w) = P\left(\frac{V^2}{R} \leq w\right) = P(V \leq \sqrt{w \cdot R}) = F_Y(\sqrt{w \cdot R})$$

De esta forma ya conocemos $F_W(w)$, teniendo en cuenta los intervalos para cada w , y además podemos conocer la función de densidad de probabilidad, tal que $f_W(w) = \frac{dF_W(w)}{dw}$.

1.1 Transformaciones útiles

Si tengo dos variables aleatorias Y, X , tal que $Y = a \cdot X + b$:

- $\mu_Y = a \cdot \mu_X + b$
- $\sigma_Y^2 = a^2 \cdot \sigma_X^2 \rightarrow \sigma_Y = |a| \cdot \sigma_X$
- $\gamma_Y = \text{sign}(a) \cdot \gamma_X$
- $K_Y = K_X$

2 Variable aleatoria bidimensional

2.1 Discretas

Sean X, Y dos variables aleatorias discretas. Su comportamiento puede ser descrito por la **función de probabilidad conjunta**.

$$p_{X,Y}(x, y) = P(X = x, Y = y) \quad x \in R_X, y \in R_Y$$

Las funciones de probabilidad marginales pueden ser obtenidas como:

$$p_X(x) = P(X = x) = \sum_{y \in R_Y} p_{X,Y}(x, y) \quad x \in R_X$$

$$p_Y(y) = P(Y = y) = \sum_{x \in R_X} p_{X,Y}(x, y) \quad y \in R_Y$$

Obs: Las variables aleatorias son independientes si:

$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \quad x \in R_X, y \in R_Y$$

El valor esperado de una función h de ambas variables aleatorias ($h : \mathbb{R}^2 \rightarrow \mathbb{R}$) viene dado por:

$$E[h(X, Y)] = \sum_{x \in R_X, y \in R_Y} h(x, y)p_{X,Y}(x, y) = \sum_{x \in R_X} \sum_{y \in R_Y} h(x, y)p_{X,Y}(x, y)$$

Sean dos funciones $k, m : \mathbb{R} \rightarrow \mathbb{R}$. Luego, si X e Y son independientes:

$$E[k(X) \cdot m(Y)] = E[k(X)] \cdot E[m(Y)]$$

2.2 Continuas

El comportamiento de dos variables aleatorias continuas X e Y es descrito por una **densidad de probabilidad conjunta** $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}$. A partir de ésta, se pueden calcular todas las posibilidades relacionadas a ambas variables como:

$$P((X, Y) \in B) = \iint_{(x,y) \in B} f_{X,Y}(x, y) \, dx dy \quad B \subseteq \mathbb{R}^2$$

La densidad de las probabilidades marginales es tal que:

$$f_X = \int_{\mathbb{R}} f_{X,Y}(x, y) \, dy \quad f_Y = \int_{\mathbb{R}} f_{X,Y}(x, y) \, dx$$

Si éstas son independientes $\rightarrow f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$.

El valor esperado de una función h de ambas variables aleatorias ($h : \mathbb{R}^2 \rightarrow \mathbb{R}$) viene dado por:

$$E[h(X, Y)] = \iint_{\mathbb{R}^2} h(x, y) f_{X,Y}(x, y)$$

Sean dos funciones $k, m : \mathbb{R} \rightarrow \mathbb{R}$. Luego, si X e Y son independientes:

$$E[k(X) \cdot m(Y)] = E[k(X)] \cdot E[m(Y)]$$

La covarianza de dos variables aleatorias está definida por:

$$Cov[X, Y] = E[X \cdot Y] - E[X] \cdot E[Y]$$

El coeficiente de correlación de dos variables aleatorias está definido como:

$$\rho(X, Y) = \frac{Cov[X, Y]}{\sqrt{Var[X] \cdot Var[Y]}}$$

Sean X e Y dos variables aleatorias, se cumple que:

$$E[X + Y] = E[X] + E[Y] \quad E[(X + Y)^2] = E[X^2] + 2 \cdot E[X \cdot Y] + E[Y^2]$$

La varianza está dada por: $Var[X + Y] = Var[X] + 2 \cdot Cov[X, Y] + Var[Y]$. Pero si X e Y son independientes $\rightarrow Var[X + Y] = Var[X] + Var[Y]$.

Caso práctico:

$$\begin{aligned} P(X + Y = z) &= \sum_{x \in R_X, y \in R_Y : x+y=z} p_{X,Y}(x, y) \\ \text{(por independencia)} \quad &= \sum_{x \in R_X} p_X(x) p_Y(z - x) = \sum_{x \in R_Y} p_X(z - y) p_Y(y) \end{aligned}$$

Para variables aleatorias continuas:

$$P(X + Y = z) = \int_{\mathbb{R}} f_X(x) f_Y(z - x) dx = \int_{\mathbb{R}} f_X(z - y) f_Y(y) dy$$

2.3 Distribuciones condicionales

1. Discreta:

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

2. Continua:

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

3 Procesos estocásticos

Un proceso estocástico es una familia de variables aleatorias $\{X(t)\}_{t \in \mathbb{T}}$, donde \mathbb{T} es el conjunto de índices o espacio de parámetros del proceso.

1. Proceso discreto: $p(x_1, t_1, \dots, x_n, t_n) = P(X(t_1) = x_1, \dots, X(t_n) = x_n)$
2. Proceso continuo: $f(x_1, t_1, \dots, x_n, t_n)$. En general no es sencilla esta forma de describir los procesos, por lo que se utilizan simplificaciones, como los procesos estacionarios, procesos de Markov, etc.

3.1 Proceso estocástico estacionario

Un proceso estocástico es **estacionario** si:

1. Discreto:

$$p(x_1, t_1 + \Delta t, \dots, x_n, t_n + \Delta t) = p(x_1, t_1, \dots, x_n, t_n)$$

2. Continuo:

$$f(x_1, t_1 + \Delta t, \dots, x_n, t_n + \Delta t) = f(x_1, t_1, \dots, x_n, t_n)$$

En resumen, $\{X(t)\}_t$ es estacionario si **es igual** a $\{X(t + \Delta t)\}_t$.

Obs: Se dice que un proceso estocástico es estacionario en sentido amplio si su *esperanza* y *varianza* son constantes.

3.2 Incrementos independientes

Un proceso estocástico tiene incrementos independientes si para $t_1, t_2, t_3, t_4 \in \mathbb{T}$ se cumple que:

$$[t_1, t_2) \cap [t_3, t_4) = \emptyset \Rightarrow (X(t_2) - X(t_1)) \text{ y } (X(t_4) - X(t_3)) \text{ son v.a independientes}$$

Se dice que un proceso estocástico tiene incrementos estacionarios si las variables aleatorias $(X(t_2) - X(t_1))$ y $(X(t_2 + \Delta t) - X(t_1 + \Delta t))$ tienen la misma distribución de probabilidades.

3.3 Procesos de Markov

Se dice que un proceso estocástico es un proceso de Markov si para todo conjunto de valores $t_1 \leq t_2 \leq \dots \leq t_n$, $t_i \in \mathbb{T}$, y para todo $x_1, x_2, \dots, x_n \in \mathbb{E}$ se cumple:

1. Discreto:

$$p(x_n, t_n | x_{n-1}, t_{n-1}, \dots, x_1, t_1) = p(x_n, t_n | x_{n-1}, t_{n-1})$$

2. Continuo:

$$f(x_n, t_n | x_{n-1}, t_{n-1}, \dots, x_1, t_1) = f(x_n, t_n | x_{n-1}, t_{n-1})$$

Obs: El proceso depende del estado más reciente y no de toda la historia.

3.4 Proceso de conteo

Considero un proceso estocástico $\{T(k)\}_{k \in \mathbb{N}_0}$ tal que $T_0 = 0$ y $k \leq m \Rightarrow T_k \leq T_m$. Este tipo de proceso puede utilizarse para modelar los instantes de ocurrencia de ciertos eventos. Puede tener un espacio de estados discreto o continuo. En general, se le asocia un proceso de conteo $\{N(t)\}_{t \geq 0}$ definido como $N(t) = \max\{k : T_k \leq t\}$. Por lo tanto, $N(t)$ cuenta la cantidad de eventos hasta el instante t , y cuenta con las siguientes proposiciones/características:

- $N(0) = 0$.
- $N(t) \in \mathbb{N}_0$.
- $s \leq t \Rightarrow N(s) \leq N(t)$.
- $N(t) - N(s)$ es la cantidad de eventos en el intervalo $(s, t]$.

3.5 Proceso de Poisson

Un proceso de conteo puede considerarse un proceso de Poisson con tasa $\lambda > 0$ si satisface las siguientes condiciones:

- Tiene incrementos independientes.
- Los incrementos son estacionarios.
- La probabilidad de que exactamente un evento ocurra en un intervalo de longitud Δt es $\lambda \Delta t + o(\Delta t)$.
- La probabilidad de que más de un evento ocurra en un intervalo de longitud Δt es $o(\Delta t)$.

Si cumple estas condiciones: $N(t) \sim \text{Poisson}(\lambda t) \Rightarrow P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$ con $k = 0, 1, \dots$

Obs: Sea $T(k)_{k \in \mathbb{N}_0}$ el proceso estocástico que modela los instantes de ocurrencia de eventos asociados al proceso de Poisson $N(t)_{t \in \mathbb{R}_{\geq 0}}$. También, se puede demostrar que, para todo $k \in \mathbb{N}_0$, $(T(k+1) - T(k))$ son variables aleatorias i.i.d con distribución exponencial, de parámetro λ . Es decir, los tiempos entre eventos son variables exponenciales independientes.

3.6 cadenas de Markov

Las cadenas de Markov son procesos de Markov en los cuales el espacio de estados es discreto: $E = \{s_1, s_2, s_3, \dots\}$. Nosotros nos concentraremos en cadenas cuyo espacio de parámetro o conjunto de índices es también discreto: $\mathbb{T} = \mathbb{N}_0$. Una cadena de Markov $\{X(n)\}_{n \in \mathbb{N}_0}$ queda completamente descrita por:

- 1. Distribución de probabilidades del estado inicial:

$$p_j(0) = P(X(0) = s_j) \text{ con } \sum_j p_j(0) = 1.$$

- 2. Probabilidades de transición entre cada par de estados en cada instante de tiempo:

$$p_{ij}(n) = P(X(n+1) = s_j | X(n) = s_i) \text{ con } \sum_j p_{ij}(n) = 1 \ \forall \ i.$$

Por la ec. de Chapman-Kolmogorov: $p_j(n+1) = \sum_i p_{ij}(n)p_i(n).$

La matriz de probabilidades será:

$$\mathbb{P} = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \cdots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix}$$

Notar que los elementos en cada filan deben sumar 1. Y si quiero conocer la distribución de probabilidades en el futuro, es decir, en n pasos, entonces es \mathbb{P}^n .

3.7 Estados en cadenas homogéneas

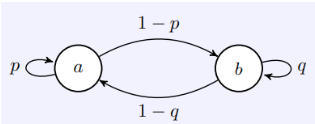
Si $P(X(n+m) = s_j | X(n) = s_i) > 0$ para algún m , entonces se dice que el estado s_j es accesible desde el estado s_i . Si todos los estados de una cadena de Markov son accesibles de todos los otros estados, se dice que la cadena de Markov es irreducible. Un estado s_i es absorbente si $p_{ii} = 1$. Un estado s_j es periódico con período T_0 si $P(X(n) = s_j | X(0) = s_j) = 0$ salvo para $n = k$.

3.8 Diagrama de estados de Markov

Se trata de un grafo dirigido y con aristas etiquetadas. Cada nodo del grafo se corresponde a un estado posible de la cadena de Markov. La etiqueta de cada arista dirigida indica la probabilidad de transición de un estado hacia otro.

$$\begin{aligned} P(X(n+1) = a | X(n) = a) &= p, & P(X(n+1) = b | X(n) = a) &= 1 - p \\ P(X(n+1) = a | X(n) = b) &= q, & P(X(n+1) = b | X(n) = b) &= 1 - q \end{aligned}$$

Obs: La suma de las flecha(probabilidades) que salen de cada nodo debe ser 1.



Teorema: Una cadena de Markov finita irreducible tiene una distribución estacionaria de probabilidades $\rightarrow \exists \vec{\pi} = \lim_{n \rightarrow \infty} \vec{p}(n)$, y ese límite es independiente de la distribución de probabilidades del estado original $\vec{p}(0)$.

Teorema: Si $\exists k$ / todos los elementos $\mathbb{P}_{ij}^k > 0$, la cadena de Markov es regular \rightarrow Existe una distribución de probabilidades estacionaria.

Si la cadena de Markov tiene una distribución estacionaria de probabilidades $\rightarrow \vec{\pi}$ es la distribución de probabilidades a largo plazo, y la puedo calcular tal que:

$$\vec{\pi} = \vec{\pi} \mathbb{P}$$

4 Suma de variables aleatorias

Sean $\{X_k\}_{k=1}^n$ variables aleatorias independientes, llamamos a la sumatoria de éstas S_n :

$$S_n = \sum_{k=1}^n X_k$$

$$E[S_n] = \sum_{k=1}^n E[X_k] \quad Var[S_n] = \sum_{k=1}^n Var[X_k]$$

1. **Suma de v.a. Bernoulli es binomial:** Sean $X_1, X_2, \dots, X_n \sim \text{Bernoulli}(p)$ independientes. Luego, $\sum_{k=1}^n X_k \sim \text{Binomial}(n, p)$.
2. **Suma de v.a. binomiales es binomial:** Sean $N_1 \sim \text{Binomial}(n, p)$ y $N_2 \sim \text{Binomial}(m, p)$ independientes. Luego $N_1 + N_2 \sim \text{Binomial}(n + m, p)$. El ejemplo anterior es un caso especial.
3. **Suma de v.a. Poisson es Poisson:** Sean $N_1 \sim \text{Poisson}(\lambda_1)$ y $N_2 \sim \text{Poisson}(\lambda_2)$ independientes. Luego $N_1 + N_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.
4. **Suma de v.a. normales es normal:** Sean $X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ y $X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$ independientes. Luego $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.
5. **Suma de v.a. geométricas es binomial negativa:** Sean $X_1 \sim \text{Geo}(p)$ y $X_2 \sim \text{Geo}(p)$ independientes. Luego $X_1 + X_2 \sim \text{BinoNeg}(2, p)$.
6. **Suma de v.a. binomiales negativas es binomial negativa:** Sean $X_1 \sim \text{BinoNeg}(n, p)$ y $X_2 \sim \text{BinoNeg}(m, p)$ independientes. Luego $X_1 + X_2 \sim \text{BinoNeg}(n + m, p)$.
7. **Suma de v.a. exponenciales es gama:** Sean $X_1, X_2, \dots, X_n \sim \text{Expo}(\lambda)$ independientes. Luego, $\sum_{k=1}^n X_k \sim \Gamma(n, \lambda)$, donde

$$f_{\Gamma(n, \lambda)}(x) = \frac{(\lambda x)^{n-1}}{(n-1)!} \lambda e^{-\lambda x} \quad \text{para } x > 0.$$

8. **Suma de los cuadrados de v.a. normales es chi(ji)-cuadrado:** Sean $X_1, X_2, \dots, X_n \sim \mathcal{N}(0, 1)$ independientes. Luego, $\sum_{k=1}^n X_k^2 \sim \chi_n^2$ (chi(ji)-cuadrado con n grados de libertad), donde

$$f_{\chi_n^2}(x) = \left(\frac{x}{2}\right)^{\frac{n}{2}} \frac{e^{-\frac{x}{2}}}{x \Gamma\left(\frac{n}{2}\right)} \quad \text{para } x > 0.$$

9. **Suma de v.a. chi-cuadrado es chi-cuadrado:** Sean $X_1 \sim \chi_n^2$ y $X_2 \sim \chi_m^2$ independientes. Luego $X_1 + X_2 \sim \chi_{n+m}^2$.
10. **Suma de v.a. Cauchy:** Sean $X_1 \sim \text{Cauchy}(x_1, \gamma_1)$ y $X_2 \sim \text{Cauchy}(x_2, \gamma_2)$ independientes. Luego $X_1 + X_2 \sim \text{Cauchy}(x_1 + x_2, \gamma_1 + \gamma_2)$. Recordemos que

$$f_{X_1}(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x - x_1)^2}.$$

4.1 Desigualdad de Markov

$$P(|X| \geq \varepsilon) \leq \frac{E[|X|]}{\varepsilon} \quad \varepsilon > 0$$

4.2 Desigualdad de Chebychev

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{\text{Var}[X]}{\varepsilon^2} \quad \varepsilon > 0$$

La desigualdad de Chebychev se puede usar para la suma S_n variables aleatorias i.i.d $\{X_k\}_{k=1}^n$ para todo $\varepsilon > 0$:

$$P\left(\left|\frac{S_n}{n} - E[X_1]\right| \geq \varepsilon\right) \leq \frac{Var[X_1]}{n\varepsilon^2}$$

4.3 Ley de los grandes números

Por la **ley débil de los grandes números**, se dice que para todo $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0$$

La probabilidad de que el promedio y el valor medio difieran en algo se hace cada vez más pequeña a medida que se promedian más datos. Por lo que la **ley fuerte de los grandes números** afirma que:

$$P(\lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1$$

Sea $A \subseteq \Omega$ un evento y sea $p = P(A)$. Supongamos que el experimento se repite varias veces en forma independiente, de modo que definimos la variable que define el evento como:

$$l_k = \begin{cases} 1 & \text{Ocurrió A en el k-ésimo experimento} \\ 0 & \text{Caso contrario} \end{cases}$$

Las variables $l_k(A)$ con i.i.d con $E[l_k(A)] = p$, y $Var[l_k(A)] = p(1 - p)$, definimos la frecuencia relativa de ocurrencia del evento A en los primeros n experimentos como:

$$\hat{P}_n = \frac{1}{n} \sum_{k=1}^n l_k(A)$$

4.4 Teorema central del límite

Sean $\{X_k\}_{k=1}^n$ i.i.d con $\mu = E[X_1]$ y $\sigma^2 = Var[X_1]$, definimos $Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow$ se define el teorema central del límite como:

$$\lim_{n \rightarrow \infty} F_{Z_n}(z) = \lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z)$$

Obs: A medida que n se hace más grande, Z_n se parece más a una variable $Z \sim N(0, 1)$, por lo que para $n > 20$ se cumplen las siguientes aproximaciones:

$$F_{Z_n}(z) = P(Z_n \leq z) \simeq \Phi(z)$$

$$F_{\bar{X}_n}(x) = P(\bar{X}_n \leq x) \simeq \Phi\left(\frac{x - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$F_{S_n}(s) = P(S_n \leq s) \simeq \Phi\left(\frac{s - n\mu}{\sqrt{n} \sigma}\right)$$

Sea $A \subseteq \Omega$ un evento y sea $p = P(A)$, de modo que definimos por la ley fuerte de los grandes números que $\lim_{n \rightarrow \infty} \hat{P}_n = p$ (con probabilidad 1). Por el teorema central del límite, para un n muy grande:

$$F_{\hat{P}_n}(q) = P(\hat{P}_n \leq q) \simeq \phi\left(\frac{q - p}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

4.5 Corrección ppor continuidad

El teorema central del límite es válido tanto para variables aleatorias continuas, como para variables aleatorias discretas. Para un n muy grande (pero finito), estamos aproximando la distribución de una variable aleatoria discreta por la de una continua. Para que la aproximación sea correcta, hay que ajustar la distribución continua de la normal a la distribución discreta.

$$P(S_n = s) = P\left(s - \frac{1}{2} < S_n \leq s + \frac{1}{2}\right) \simeq \phi\left(\frac{s + \frac{1}{2} - n\mu}{\sqrt{n} \sigma}\right) - \phi\left(\frac{s - \frac{1}{2} - n\mu}{\sqrt{n} \sigma}\right)$$
$$P(\bar{X}_n = x) = P\left(x - \frac{1}{2n} < \bar{X}_n \leq x + \frac{1}{2n}\right) \simeq \phi\left(\frac{x + \frac{1}{2n} - \mu}{\frac{\sigma}{\sqrt{n}}}\right) - \phi\left(\frac{x - \frac{1}{2n} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

5 Intervalos de confianza

Se define \bar{X} como un estimador puntual, ya que devuelve un solo valor como aproximación del parámetro.

Se mide la bondad de un estimador puntual mediante el error cuadrático medio:

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

El estimador es insesgado si $E[\hat{\theta}] = \theta$. Un estimador es ideal si es insesgado de mínima varianza.

Obs: Si tomo un n grande \rightarrow Se dice que $\hat{\theta}$ es asintóticamente insesgado si:

$$\lim_{n \rightarrow \infty} E[\hat{\theta}(X_1, \dots, X_n)] - \theta = 0$$

Sean X_1, \dots, X_n variables aleatorias tales que su función de densidad de probabilidad/función de probabilidad acumulada se puede escribir con θ como parámetro a estimar \rightarrow Ésta función se la denomina verosimilitud. Y el estimador de máxima verosimilitud es el valor de θ que maximiza la probabilidad.

Obs: Los estimadores de máxima verosimilitud son asintóticamente consistentes.

5.1 Estimación puntual de la media

Sean X_1, \dots, X_n variables aleatorias i.i.d con media $\mu = E[X_1]$. Un estimador de la media es el promedio:

$$\hat{\mu} = \hat{\mu}(X_1, \dots, X_n) = \frac{1}{n} \sum_{k=1}^n X_k$$

Observaciones:

1. Es un estimador insesgado.

2. Por la ley de los grandes números \rightarrow Es un estimador consistente.
3. Si $\sigma^2 = Var[X_1] \rightarrow ECM(\hat{\mu}) = Var[\bar{X}] = \frac{\sigma^2}{n}$.
4. Si $X_1 \sim \mathcal{N}(\mu, \sigma) \Rightarrow \hat{\mu} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$.
5. Si $\sigma^2 = Var[X_1] < \infty$, por TCL, $\hat{\mu}$ es asintóticamente normal. Es decir que, para n grande, la distribución del estimador $\hat{\mu}$ se puede aproximar por la de una variable aleatoria $\sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$.

5.2 Estimación puntual de la varianza

Sean X_1, \dots, X_n variables aleatorias i.i.d con media $\mu = E[X_1]$ y varianza $\sigma^2 = Var[X_1]$. Un estimador de la varianza es la varianza muestral:

$$\hat{\sigma}^2 = \hat{\sigma}^2(X_1, \dots, X_n) = S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2 = \frac{1}{n-1} \sum_{k=1}^n X_k^2 - \frac{n}{n-1} \bar{X}^2$$

Observaciones:

1. Es un estimador insesgado.
2. La varianza del estimador es: $ECM(\hat{\sigma}^2) = Var[S^2] = \frac{\sigma^4}{n} (k + 2 + \frac{2}{n-1})$, con $k = \frac{E[(X_1 - E[X_1])^4]}{\sigma^4} - 3$. Si X_1 es normal $\rightarrow Var[S^2] = \frac{2\sigma^4}{n-1}$.
3. Si $X_1 \sim \mathcal{N}(\mu, \sigma) \Rightarrow \frac{(n-1)S^2}{\sigma^2} \sim X_{n-1}^2$.
4. Si $k < \infty$, Chebychev nos permite probar una versión débil de los grandes números y que se trata de un estimador consistente.
5. Si $k < \infty$, por una versión del TCL, S^2 es asintóticamente normal. Por lo que, para un n grande, la distribución de estimador de S^2 se puede aproximar por la de una variable aleatoria $\sim \mathcal{N}(\sigma^2, \sqrt{Var[S^2]})$.

5.3 Estimación puntual de una proporción

Sean X_1, \dots, X_n variables aleatorias i.i.d con $X_1 \sim \text{Bernoulli}(p)$. cada X_i marca la ocurrencia o no de un dado evento en una serie de experimentos independientes.

$$\hat{p} = \hat{p}(X_1, \dots, X_n) = F = \frac{1}{n} \sum_{k=1}^n X_k$$

Observaciones:

1. Es un estimador insesgado.
2. Por la ley de los grandes números \rightarrow Es un estimador consistente.

3. El error cuadrático medio es: $ECM(\hat{p}) = Var[\hat{p}] = \frac{p(1-p)}{n}$.
4. $n\hat{p} \sim \text{Binomial}(n, p)$.
5. Por TCL, \hat{p} es asintóticamente normal. Para n grande, la distribución del estimador \hat{p} se puede aproximar por la de una variable aleatoria $\sim \mathcal{N}(p, \sqrt{\frac{p(1-p)}{n}})$.

5.4 Estimación de intervalos

En ciertos casos, en vez de dar una estimación puntual de un parámetro θ , se realiza la estimación de un intervalo de confianza $(\hat{\theta}_l^\alpha, \hat{\theta}_u^\alpha)$ tal que:

$$P(\hat{\theta}_l^\alpha(X_1, \dots, X_n) < \theta < \hat{\theta}_u^\alpha(X_1, \dots, X_n)) = 1 - \alpha$$

Donde $1 - \alpha$ es el nivel de confianza del estimador.

Obs: Si se fija $\hat{\theta}_l^\alpha = -\infty \rightarrow$ Intervalo unilateral a derecha. Si se fija $\hat{\theta}_u^\alpha = +\infty \rightarrow$ Intervalo unilateral a izquierda. Si ambos $|\hat{\theta}_l^\alpha|, |\hat{\theta}_u^\alpha| < \infty \rightarrow$ Intervalo bilateral.

5.5 Intervalo para la media con varianza conocida

Sean X_1, \dots, X_n variables aleatorias i.i.d con media $\mu = E[X_1]$ **desconocida** y varianza $\sigma^2 = Var[X_1]$ **conocida**. Los intervalos con un nivel de confianza $\gamma \cdot 100\%$, con $\gamma \in [0, 1]$:

1. Unilateral a derecha:

$$I_r^\gamma = \left(-\infty, \bar{X} + z_\gamma \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Donde $z_p = \Phi^{-1}(p)$.

2. Unilateral a izquierda:

$$I_l^\gamma = \left(\bar{X} - z_\gamma \cdot \frac{\sigma}{\sqrt{n}}, +\infty \right)$$

Donde $z_\alpha = -z_{1-\alpha}$.

3. Bilateral:

$$I_{lr}^\gamma = \left(\bar{X} - z_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Si las variables aleatorias no son normales, pero n es grande, el TCL nos permite utilizar estos mismos intervalos de confianza.

Obs: Sean o no normales las variables aleatorias, si la varianza(σ^2) es desconocida, para n muy grande, se pueden utilizar estos mismos intervalos de confianza reemplazando σ desconocido por el estimador S .

5.6 Intervalo para la proporción con muestras grandes

Sean p una probabilidad y \hat{p} un estimador resultante de n experimentos independientes. Si n es grande, se tienen los siguientes intervalos con un nivel de confianza de $\gamma \cdot 100\%$:

1. Unilateral a derecha:

$$I_r^\gamma = \left[0, \hat{p} + z_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

2. Unilateral a izquierda:

$$I_l^\gamma = \left(\hat{p} - z_\gamma \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, 1 \right]$$

3. Bilateral:

$$I_{lr}^\gamma = \left(\hat{p} - z_{\frac{1+\gamma}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{1+\gamma}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

Estos intervalos de confianza son muy parecidos a los de la media con varianza conocida. Esto es debido a que p es el valor medio de las variables aleatorias i.i.d con distribución Bernoulli, \hat{p} es la media muestral y al hecho que podemos aplicar al TCL. Sin embargo, hay algunas diferencias:

- $p \in [0, 1]$.
- Se desconoce la varianza real, y se la reemplaza por el estimador $\hat{p}(1-\hat{p})$. Se puede mostrar que esta aproximación es buena para n grande.

5.7 Intervalo para la media de variables aleatorias normales

Sean X_1, \dots, X_n variables aleatorias i.i.d normales con media $\mu = E[X_1]$ y varianza $\sigma^2 = \text{Var}[X_1]$ desconocidas. Luego, se tienen los siguientes intervalos con un nivel de confianza de $\gamma \cdot 100\%$:

1. Unilateral a derecha:

$$I_r^\gamma = \left(-\infty, \bar{X} + t_{n-1, \gamma} \frac{S}{\sqrt{n}} \right)$$

Donde $t_{k,p}$ es el $100p$ percentil de una variable aleatoria con distribución t-student con k grados de libertad.

2. Unilateral a izquierda:

$$I_l^\gamma = \left(\bar{X} - t_{n-1, \gamma} \frac{S}{\sqrt{n}}, +\infty \right)$$

Recordemos que $t_{n-1, \alpha} = -t_{n-1, 1-\alpha}$.

3. Bilateral:

$$I_{lr}^\gamma = \left(\bar{X} - t_{n-1, \frac{1+\gamma}{2}} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1, \frac{1+\gamma}{2}} \frac{S}{\sqrt{n}} \right)$$

Dado que para k muy grande la distribución t de student es muy parecida a la normal, para $n > 200$ no hace mucha diferencia si se utilizan los percentiles de distribución Gaussiana.