Resumen - Probabilidad y Estadística

Nicolás Margenat

1Q 2022

Contents

1 Concepto de Probabilidad

1.1 Definiciones

Definición 1. Espacio Muestral

Conjunto de todos los resultados posibles de un determinado experimento.

Definición 2. Suceso/Evento

Conjunto de resultados posibles de un experimento.

1.2 Álgebra de Sucesos

$$E : \text{Experimento aleatorio - Resultados inciertos}$$
$$S : \text{Espacio muestral}$$
$$A : \text{Suceso o Evento}$$

Cabe mencionar que:

$$A \subset S$$

Dos sucesos son **mutuamente excluyentes** sii

$$A \cap B = \varnothing$$

es decir, es *imposible* que sucedan juntos.

1.3 Frecuencia Relativa de un suceso

Si un determinado experimento se repite n veces y se cuenta el numero de veces que ocurre n_A, entonces la $frecuencia\ relativa$ de A viene dada por:

$$f(A) = \frac{n_A}{n} \in [0;1]$$

1.3.1 Propiedades

1. $f(A) \in [0;1]$
2. $f(S) = 1$
3. Si $A \cap B = \varnothing$, entonces:

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

1.4 Probabilidad de un suceso

1.4.1 Axiomática de Kolmogorov

- $P(A) \geq 0$ (La probabilidad de un suceso es un numero real no negativo)
- $P(S) = 1$
- Si $A \cap B = \varnothing$, entonces:

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

1.4.2 Teoremas

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

Con estos axiomas podemos derivar que:

- $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$
- Si $A \cap B = A \Rightarrow P(A) \leq P(B)$

1.4.3 Equivalencia de Expresiones

1. $B = (\bar{A} \cap B) \cup (A \cap B)$
2. $A \cup B = A \cup (\bar{A} \cap B)$

1.4.4 Desigualdad de Bonferroni

$$P(\bigcup_{j=1}^{n} E_j) \leq \sum_{j=1}^{n} P(E_j)$$

1.5 Regla de Laplace

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

Importante: Solo sirve en un espacio muestral finito y equiprobable.

2 Estadística Descriptiva

Estadística Descriptiva No Agrupada

$$\{x_i\}_{i=1}^{n}$$

2.1 Definiciones (ED No Agrupada)

2.1.1 Parámetros de Tendencia Central

Definición 1. Media Muestral o Promedio

Es un *parámetro de tendencia central.*
Se calcula:

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

Observación: El promedio de los desvíos de los datos respecto de la media SIEMPRE es 0, entonces habrá datos menores que la media y otro mayores que ella.

Definición 2. Mediana

Es un *parámetro de tendencia central posición.*
En un ordenamiento de menor a mayor, si el numero de datos es *impar* es el termino central. Si el numero de datos es *par* la mediana se calcula como el promedio de los dos términos centrales.
Notación: m

Definición 3. Cuartiles Primero y Tercero

Es un *parámetro de tendencia central posición.*
En un ordenamiento de menor a mayor son mediana de la primera mitad y de la segunda mitad de los datos.
Notación: q_1, q_3

2.1.2 Parámetros de Variabilidad

Definición 4. Rango muestral

La distancia entre el mínimo y el máximo de los datos muestrales. Es decir, longitud del intervalo al que pertenece el 100
Notación: R

Definición 5. Rango Intercuartílico

Distancia entre q_1 y q_3. Es decir, longitud del intervalo al que pertenecen el 50% central de los datos.
Notación: iqr

Definición 6. Desvío Estándar Muestral

Medida típica de variación de cierto conjunto de datos.
Se calcula como:

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

Definición 6. Varianza Muestral

Es el menor de los promedios de cuadrados de desvíos.

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

2.1.3 Parámetros de Forma

Definición 7. Coeficiente de Simetria

Medida típica de simetría de los datos respecto de la media, y es nulo si hay simetría. Si es positivo se dicen que los datos tienen sesgo positivo y negativo en caso contrario
Se calcula como:

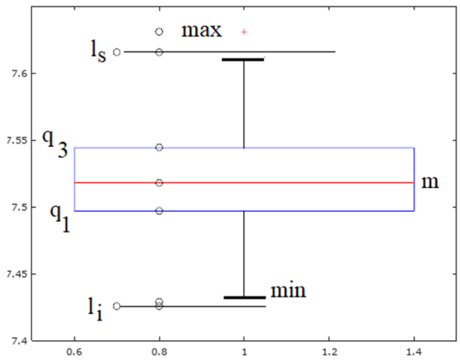
$$\gamma = \frac{1}{ns^3}\sum_{i=1}^{n}(x_i - \bar{x})^3$$

Definición 8. Coeficiente de Kurtosis

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.
Se calcula como:

$$\kappa = \frac{1}{ns^4}\sum_{i=1}^{n}(x_i - \bar{x})^4 - 3$$

2.2 Boxplot/Diagrama de caja

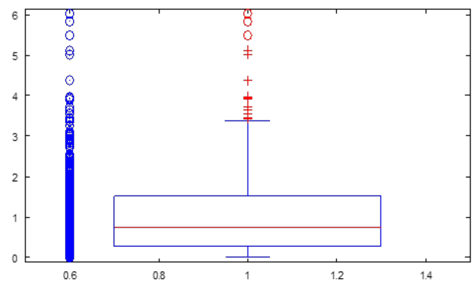


- La caja central se determina con q_1 y q_3

- Se representan con dos rayas los datos inmediatos a los limites de Tuckey, que se calculan de la siguiente manera:

$$\begin{aligned} l_s &= q_1 + 1,5 * iqr \\ l_i &= q_1 - 1,5 * iqr \end{aligned}$$

- Hay *datos outliers* de dos tipos: **moderados** (tienen un valor menor que $q_1 - 3 * iqr$) y **severos** (tienen valores mayores que $q_3 + 3 * iqr$)



+ moderados, O severos

- Si el máximo o mínimo están antes que el bigote calculado, entonces ahí va el bigote

Estadística Descriptiva Agrupada

$$\{\tilde{x}_r, y_r\}_{r=1}^{m}$$

2.3 Definiciones (ED Agrupada)

Defunción 1. Marca de clase

Es el punto medio del intervalo de clase r.

$$\tilde{x}_r = \frac{a_r + b_r}{2}$$

Definición 2. Frecuencia Absoluta de la clase r

Cantidad de datos del conjunto original que pertenecen al intervalo r.

2.3.1 Parámetros Característicos

Definición 3. Media Muestral o Promedio

Es un *parámetro de tendencia central.*
Se calcula:

$$\bar{x} = \frac{1}{n} \sum_{r=1}^{m} \tilde{x}_r y_r$$

Definición 4. Desvío Estándar Muestral

Medida típica de variación de cierto conjunto de datos.
Se calcula como:

$$s = \sqrt{\frac{1}{n-1}\sum_{r=1}^{m}(\tilde{x}_r - \bar{x})^2 y_r}$$

Definición 5. Coeficiente de Simetria

Medida típica de simetría de los datos respecto de la media, y es nulo si hay simetría. Si es positivo se dicen que los datos tienen sesgo positivo y negativo en caso contrario
Se calcula como:

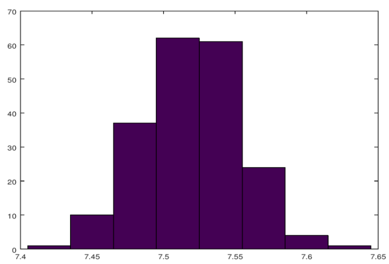
$$\gamma = \frac{1}{ns^3}\sum_{r=1}^{m}(\tilde{x}_r - \bar{x})^3 y_r$$

Definición 8. Coeficiente de Kurtosis

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.
Se calcula como:

$$\kappa = \frac{1}{ns^4}\sum_{r=1}^{m}(\tilde{x}_r - \bar{x})^4 y_r - 3$$

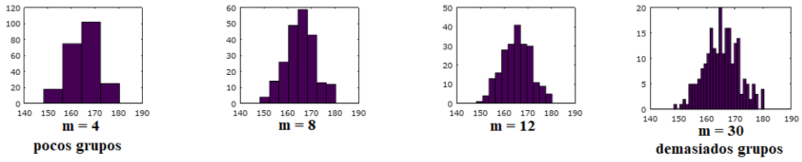
2.4 Histogramas



- Cada rectángulo tiene base en un intervalo de clase y la altura mide la frecuencia absoluta del intervalo.
- El área del rectángulo es proporcional a n (frecuencia absoluta) y L (longitud de intervalos)
- Para elegir el numero de rectángulos, que denominamos m, utilizamos la siguiente regla:

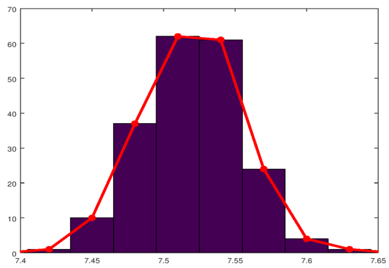
  $$m = \lceil log_2(n) \rceil$$

  a partir de este numero podemos sumar o restar un poco para ver si hay un mejor m.



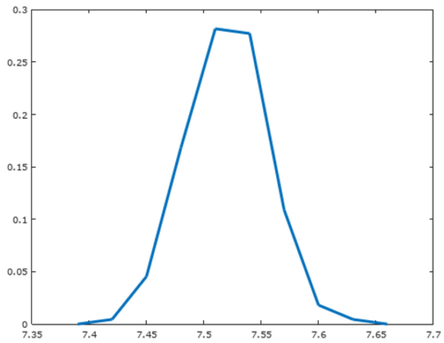
2.5 Polígonos

2.5.1 Polígono de Frecuencias Absolutas



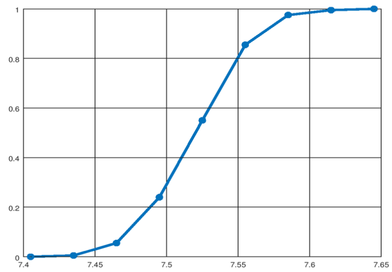
Para construirlo se unen las marcas de clase con lineas rectas, y luego se unen los extremos en donde la ordenada es 0 con las marcas de clase de los extremos.

2.5.2 Polígono de Frecuencias Relativas



Se construye de igual manera que el de frecuencias absolutas, pero usando las frecuencias relativas.
<u>Observación</u>: El área bajo este gráfico es L, pues la suma de las frecuencias relativas es 1.

2.5.3 Polígono de Frecuencias Relativas Acumuladas



Se representa la frecuencia acumulada relativa suponiendo que en cada intervalo la frecuencia aumenta linealmente.
Observación: La diferencia de ordenadas en los extremos de un intervalo cualquiera mide la fracción de los datos en ese intervalo.

3 Probabilidad Condicional

3.1 Definiciones

Definición 1. $P(A/B)$

Si A y B son sucesos, y $P(B) > 0$ entonces se introduce la probabilidad condicional de A dado que ocurrió B.
Notación: $P(A/B)$ (”*probabilidad de* A *dado* B”)

Definición 2. Sucesos independientes

Dos sucesos son independientes si la probabilidad de que ocurra uno no te influye sobre la probabilidad de otro. Además:

$$A \text{ y } B \text{ son independientes} \Leftrightarrow P(A \cap B) = P(A)P(B)$$

Por ejemplo, si tengo un tarro de bolitas con reposición y saco 2 bolitas, después como las bolitas las vuelvo a poner vuelvo a estar en el caso inicial cuando haga la segunda extracción (entonces son sucesos independientes).

3.2 Propiedades $P(A/B)$

1. $P(A/B) = \frac{P(A \cap B)}{P(B)}$ con $P(B) > 0$
2. $P(A \cap B) = P(A/B)P(B) = P(B/A)P(A)$ (puedo elegir la que mas me convenga)
3. Si $A \cap C = \varnothing$
$$\Rightarrow P((A \cup C)/B) = P(A/B) + P(C/B) \text{ con } P(B) > 0$$
4. $P(\bar{A}/B) = 1 - P(A/B)$ pero $P(A/\bar{B}) \neq 1 - P(A/B)$
5. $P((A \cup C)/B) = P(A/B) + P(C/B) - P((A \cap C)/B)$
6. Si $P(A \cap B) = 0 \Rightarrow P(A/B) = 0$
7. $P(H/E) = \frac{P(H)P(E/H)}{P(E)}$

3.3 Propiedades Sucesos Independientes

1. Si A y B son independientes, entonces también lo son: $\bar{A}$ y B; A y $\bar{B}$; y $\bar{A}$ y $\bar{B}$
2. *Condición de independencia para una familia de sucesos*: Sean $\{A_k\}_{k \in K}$ una familia de sucesos independientes, entonces:
$$P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j) \quad \forall J \subseteq K$$

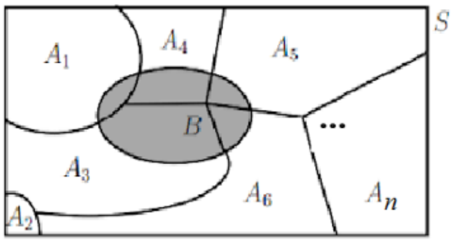
   Por ejemplo, para el caso de 3 sucesos:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$$
$$P(A_1 \cap A_3) = P(A_1) \cdot P(A_3)$$
$$P(A_2 \cap A_3) = P(A_2) \cdot P(A_3)$$
$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$$

3.4 Teoremas

Sea S un espacio muestral /

$$
\begin{aligned}
& S = \bigcup_{k=1}^{n} A_k \\
& A_k \neq \varnothing \\
& A_k \cap A_j = \varnothing \text{ si } k \neq j \\
& B = \bigcup_{k=1}^{n} (B \cap A_k) = \sum_{k=1}^{n} (B \cap A_k)
\end{aligned}
$$



Teorema de la Probabilidad Total

$$P(B) = \sum_{k=1}^{n} P(B/A_k)P(A_k)$$

Teorema de Bayes

$$P(A_j/B) = \frac{P(B/A_j)P(A_j)}{\sum_{k=1}^{n} P(B/A_k)P(A_k)}$$

4 Variables Aleatorias Discretas

4.1 Definiciones

Definición 1. Variable Aleatoria (va)

Una *variable aleatoria* es la consecuencia numérica de un experimento aleatorio.
Notación: X
Ademas, el conjunto de valores posibles de una variable aleatoria es:

$$X : S \to R_X \text{ donde } R_X \text{ es el recorrido o rango de } X$$

y S es el espacio muestral de un experimento aleatorio.

Definición 2. Valor Esperado de una vad

El *valor esperado de una variable aleatoria* es una medida de centralidad. Nos dice donde se centra la variable considerando las probabilidades.
Notación: $E(X)$

$$E(X) = \sum_x x p_X(x)$$

Definición 3. Varianza de una vad

La varianza de una variable aleatoria una medida de dispersión. Evalúa usando los valores de la probabilidad cuanto se tienden a alejar los valores del recorrido respecto de la media.
Notación: $V(X)$

$$V(X) = \sum_{x \in R_X} (x - E(X))^2 p_X(x)$$

Definición 4. Desvío Estándar de una vad

El desvío estándar de una variable aleatoria es una medida típica de los desvíos de sus valores respecto del valor esperado.
Notación: $\sigma(X)$

$$\sigma(X) = \sqrt{V(X)}$$

Definición 5. Coeficiente de Simetria de una vad

$$\gamma = \frac{1}{\sigma^3} \sum_{x \in R_X} (x - E(X))^3 p_X(X)$$

Definición 5. Coeficiente de Kurtosis de una vad

$$\gamma = \frac{1}{\sigma^4} \sum_{x \in R_X} (x - E(X))^4 p_X(X)$$

4.2 Propiedades

4.2.1 Propiedades de $E(X)$

1. $E(a) = a$ si a es una constante
2. $E(bX) = bE(X)$ si b es una constante
3. $E(X - E(X)) = 0 \ \ \forall X$
4. $E(X + Y) = E(X) + E(Y)$
5. $E(G(X)) = \sum_{x \in R_X} G(X) p_X(x)$

4.2.2 Propiedades de $V(X)$

1. $V(X) \geq 0$
2. $V(a) = 0$ si a es una constante
3. $V(bX) = b^2 V(X)$ si b es una constante
4. $V(X) = E(X^2) - (E(X))^2$ donde $E(X^2) = \sum_{x \in R_X} x^2 p_X(x)$

4.3 VADs Notables

4.3.1 VAD Bernoulli

Caracteriza la ocurrencia o no de un suceso al realizar un experimento aleatorio.
Características:

1. $X \sim Bernoulli(p)$

   $R_X = \{0, 1\}$ donde 0 es fracaso y 1 es éxito
2. $P(X = 1) = p$, $P(X = 0) = 1 - p = q$
3. $E(X) = p$
4. $V(X) = pq$

4.3.2 VAD Binomial

(**con reposición**) Cuenta el número de ocurrencias de un evento de interés en n de repeticiones *independientes* de un experimento aleatorio con igual probabilidad p de éxito.
Características:

1. $X \sim Binomial(n, p)$

   $R_X = \{0, 1, ..., n\}$
2. $P(X = k) = \binom{n}{k} p^k q^{n-k} \ \ \forall k \in R_X$ donde

   $\binom{n}{k}$: formas de elegir las posiciones de los éxitos

   p^k: probabilidad de k éxitos

   q^{n-k}: probabilidad de $n - k$ fracasos

3. $E(X) = np$
4. $V(X) = np(1-p)$

donde X = número de éxitos en los n experimentos.

4.3.3 VAD Hipergeométrica

(**sin reposición**) Cuenta el número de elementos de cierta clase en una muestra aleatoria extraída sin reposición de una población que contiene elementos de esa clase.
Ejemplo: Población de N individuos divididos en 2 clases A y B. La A tiene M individuos y la B tiene $N - M$. Se toma una muestra de n individuos sin reposición.
Características:

1. $X \sim Hipergeométrica(N, M, n)$
   $R_X = max(0, n - (N - M)) \leq k \leq min(M, n)$

En grál: Si $Y_n \sim \mathcal{H}(N, B, n) \rightarrow$

si $n > N-B$, en algún momento se acaban los "fracasos"

si $B < n$, no puedo extraer más de B bolitas blancas

$R_{Y_n} = \{ i \in \mathbb{N}_0 \,/\, \max\{0, n-(N-B)\} \leq i \leq \min\{n, B\} \}$

si $N-B > n$ → puedo no extraer bolitas blancas porque nunca se acaban las que no lo son

si $n < B$ hay como máximo n bolitas blancas extraídas

2. $P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$ donde $max(0, n - (N - M)) \leq k \leq min(M, n)$
3. $E(X) = n\frac{M}{N}$
4. $V(X) = n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$

Observación: Si hay mas de 2 clases estas reglas ya no se cumplen.
Relación con la binomial:

$$X \sim Hipergeométrica(N, M, n),\ Y \sim Binomial(n, \frac{M}{N}) \ \Rightarrow \ P(X \leq k) \approx P(Y \leq k)$$

4.3.4 VAD Geométrica

(**con reposición**) Cuenta el número de repeticiones *independientes* de un experimento aleatorio hasta la primera ocurrencia de un evento de interés.

1. $X \sim Geom(p)$
   $R_X = \{1, 2, ...\} = \mathbb{N}$
2. $P(X = k) = (1-p)^{k-1}p \ \ \forall k \in R_X$
3. $E(X) = \frac{1}{p}$
4. $V(X) = \frac{1-p}{p^2}$

donde X = numero de intentos hasta el primer éxito (**incluido**).

Propiedades

- $P(X > M) = 1 - P(X \leq M) = (1-p)^m$
- $P(X > M + N | X > M) = P(X > N)$ ("*falta de memoria*")

4.3.5 VAD Poisson

Asociada a un especial proceso denominado de Poisson. Acá la vamos a usar como aproximación a la distribución binomial.
Se cuenta la cantidad de veces que ocurre algo, pero a diferencia de la binomial, estos eventos no tienen cantidad máxima. La ventaja de Poisson es que tiene un parámetro, mientras que la binomial tiene 2. Entonces nos puede servir cuando no se sabe n, ni p, pero se sabe el valor esperado λ (y n es grande y p es pequeño). Características:

1. $X \sim Poisson(\lambda)$ con $\lambda > 0$

   $R_X = \{0, 1, ...\} = \mathbb{N}_0$
2. $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}\ \forall k \in R_X$
3. $E(X) = \lambda$
4. $V(X) = \lambda$

Observación: Si $\lambda \in \mathbb{N}$ entonces tiene 2 valores máximos (λ y $\lambda - 1$). En cambio, si $\lambda \in \mathbb{R} - \mathbb{N}$ entonces tiene 1 valor máximo ($\lfloor \lambda \rfloor$).

4.4 Función Distribución de una vad X

Llamamos Función Distribución a $F : \mathbb{R} \to [0, 1]/$

$$F_X(x) = P(X \leq x) = \sum_{t \leq x0} p_X(t)$$

Propiedades

1. $\lim_{x \to \infty} F(x) = 1$
2. $\lim_{x \to -\infty} F(x) = 0$
3. $x_1 < x_2 \Rightarrow F(x_1) < F(x_2)$ (creciente)
4. $P(a < X \leq b) = F_X(b) - F_X(a)$

5 Variables Aleatorias Continuas

Son variables aleatorias cuyo rango es un intervalo de números reales o una unión de intervalos reales.

$$X : S \to R_X \subseteq \mathbb{R}$$

Aquí, $P(X = a) = 0$.

5.1 Definiciones

Definición 1. Valor Esperado de una vac

Se define como:

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

Definición 2. Varianza de una vac

Se define como:

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

donde $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$

Definición 3. Desvío estándar de una vac

Se define como:

$$\sigma(X) = \sqrt{V(X)}$$

Observación: Si el valor esperado y el desvío de X son finitos con ambos se puede construir un intervalo en el que la probabilidad de que X tome valores en él es significativamente grande.

Definición 4. Momentos de una va

Se definen dos momentos:

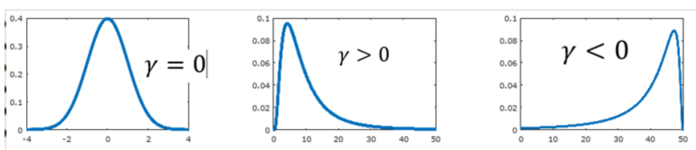
1. **Momento Absoluto de orden** n: $M_n = E(X^n)$
2. **Momento Centrado de orden** n: $\mu_n = E((X - E(X))^n$

con $n \in \mathbb{N}_0$

Definición 5. Coeficiente de Simetría de una vac

Se define como:

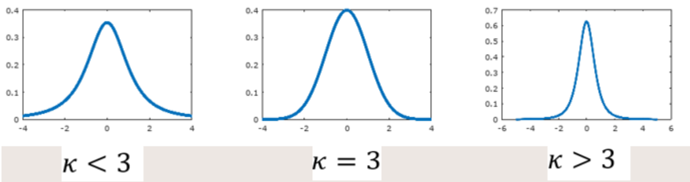
$$\gamma = \frac{\mu_3}{\sigma(X)^3}$$



Definición 6. Coeficiente de Kurtosis de una vac

Se define como:

$$\gamma = \frac{\mu_4}{\sigma(X)^4}$$



5.2 Función de Densidad de Probabilidad de X

Se define como:

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

Propiedades

1. $\int_{-\infty}^{\infty} f_X(x)dx = 1$
2. $\lim_{x\to+\infty} F(X) = 1$
3. $\lim_{x\to-\infty} F(X) = 0$

5.3 Función Distribución de una vac X

Llamamos Función Distribución a $F : \mathbb{R} \to [0, 1]/$

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

Propiedades

1. $\lim_{x\to\infty} F(x) = 1$
2. $\lim_{x\to-\infty} F(x) = 0$
3. $x_1 < x_2 \Rightarrow F(x_1) < F(x_2)$ (creciente)
4. $P(a < X < b) = F_X(b) - F_X(a)$
5. $F'_X(x) = f_X(x)$ (continua)

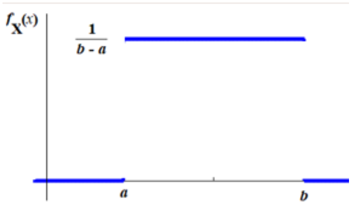
5.4 VACs Notables

5.4.1 VAC con Distribución Uniforme

$$X \sim U(a, b)$$

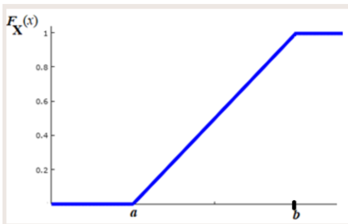
Una vac tiene *distribución uniforme* en el intervalo (a, b) si:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in (a, b) \\ \\ 0 & \text{si } x \notin (a, b) \end{cases}$$



Luego, si una vac tiene *distribución uniforme*, entonces:

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq a \\ \\ \frac{x-a}{b-a} & \text{si } x \in (a, b) \\ \\ 1 & \text{si } x \geq b \end{cases}$$



Propiedades

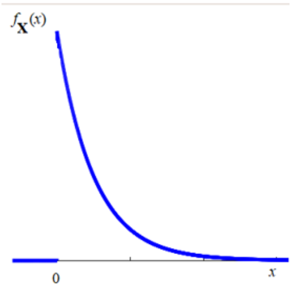
1. $E(X) = \frac{a+b}{2}$
2. $V(X) = \frac{(b-a)^2}{12}$

5.4.2 VAC con Distribución Exponencial

$$X \sim E(\lambda) \text{ con } \lambda > 0$$

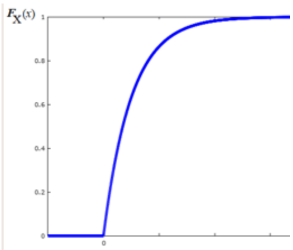
Una vac tiene *distribución exponencial* de parámetro $\lambda > 0$ si:

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



Luego, si una vac tiene *distribución exponencial*, entonces:

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



de donde podemos obtener que:

$$f_X(x) = P(X \geq x) = 1 - F_X(x) = \begin{cases} 1 & \text{si } x \leq 0 \\ \\ e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

Propiedades

1. $E(X) = \frac{1}{\lambda}$
2. $E(X^n) = \frac{n!}{\lambda^n}$ con $n \in \mathbb{N}$
3. $V(X) = \frac{1}{\lambda^2}$
4. $\sigma(X) = E(X) = \frac{1}{\lambda}$
5. $P(X > a + d | X > a) = P(X > d)$ (”*falta de memoria*”), es decir tiene que ver con la longitud del intervalo

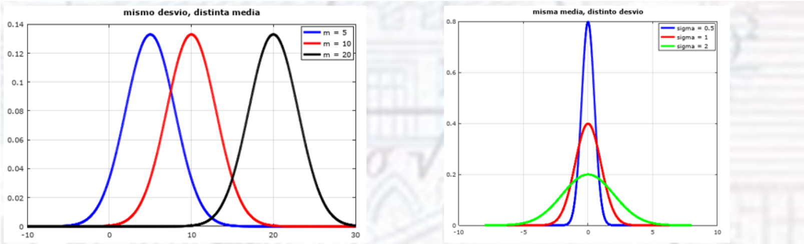
5.4.3 VAC con Distribución Normal

$$X \sim N(\mu, \sigma)$$

Una vac tiene *distribución normal* si:

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

donde μ es *media, mediana y modo* y σ es el *desvío estándar.*



Como el parámetro μ es la media, mediana y modo al aumentar se traslada "la campana". Este parámetro es de **posición**.

Al variar el parámetro σ se modifican las escalas horizontal y vertical. El intervalo de valores de interés de la variable tiene ancho $6\,\sigma$. Este parámetro es de **escala o forma**.

Propiedades A

1. $E(X) = \mu$
2. $\gamma(X) = 0$
3. $V(X) = \sigma^2$
4. $\sigma(X) = \sigma$
5. $\kappa(X) = 3$ (esto es sin restarle 3)

Propiedades B

1. $f(x) > 0 \;\; \forall x \in \mathbb{R}$
2. $\lim_{|x|\to\infty} f_X(x) = 1$
3. $\int_{-\infty}^{\infty} f_X(x) = 1$
4. $f_X(\mu + d) = f_X(\mu - d) \;\; \forall d$
5. $\int_{-\infty}^{\mu} f_X(x) = \int_{\mu}^{\infty} f_X(x) = 0,5$
6. $f'_X(x) = f''_X(x) = 0$
7. $P(a < X < b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$

Cálculo de probabilidad en un intervalo

1. Estandarizamos la variable X, para ello tomamos $Z = \frac{X-\mu}{\sigma}$:

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}dx \;\Rightarrow\; F_X(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{z^2}{2}}dz = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

2. $P(a < X < b) = P(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$ y usando la tabla o una calculadora calculamos.

Propiedades C

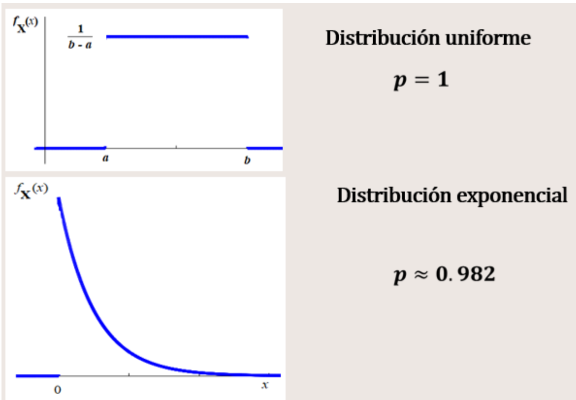
1. $P(Z < -z) = P(Z > z) = 1 - P(Z < z)$
2. $\Phi(-z) = 1 - \Phi(z)$
3. $P(-a < Z < a) = \Phi(a) - \Phi(-a) = 2\Phi(a) - 1$

No resumí fractiles porque no entendí nada.

5.5 Intervalo de Interés

Tomamos el intervalo:

$$p = P(E(X) - 3\sigma(X) < X < E(X) + 3\sigma(X))$$



Es decir, con este intervalo podemos encapsular prácticamente todo el intervalo de la función probabilidad. De hecho, la probabilidad de que la variable aleatoria tome valores en dicho intervalo es siempre $\geq \frac{8}{9}$.

6 Variables Aleatorias Bidimensionales (VA2D)

Sean

$$X : S \to R_X$$
$$Y : S \to R_Y$$

variables aleatorias definidas en el espacio muestral de un experimento aleatorio.
Entonces decimos que el par (X, Y) es una VA2D. Además, el *recorrido* de (X, Y) es:

$$R_{X,Y} = \{(x,\, y) \in R_X \times R_Y \,/\, \exists s \in S : x = X(s) \ \wedge\ y = Y(s)\} \subseteq R_X \times R_Y$$

Propiedades

1. $E(X + Y) = E(X) + E(Y)$
2. $Cov(X,\, Y) = E((X - E(X))(Y - E(Y)))$
3. $V(X + Y) = V(X) + V(Y) + 2Cov(X,\, Y)$

Propiedades si son INDEPENDIENTES

1. $E(XY) = E(X)E(Y)$
2. $Cov(X,\, Y) = 0$
3. $V(X + Y) = V(X) + V(Y)$

6.0.1 Covarianza

1. $Cov(X,\, Y) = Cov(Y,\, X)$
2. $Cov(X,\, C) = 0$ si C es constante
3. $Cov(X,\, X) = V(X)$
4. $Cov(X,\, Y) = E(XY) - E(X)E(Y)$ (”*fórmula reducida de cálculo de la covarianza*”)
5. $Cov(aX,\, bY) = ab\, E(XY) - E(X)E(Y)$
6. $Cov(X,\, Y) \neq 0 \ \Rightarrow\ X$ e Y son *dependientes*

6.0.2 Coeficiente de Correlación (lineal) de dos va2d

Se define como:

$$\rho(X,\, Y) = \frac{Cov(X,\, Y)}{\sigma(X)\sigma(Y)}$$

donde $\rho(X,\, Y) \in [-1,\, 1]$

Propiedades

1. $\rho(X,\, Y) = \rho(Y,\, X)$
2. $\rho(X,\, X) = 1$
3. $\rho(X,\, C) = 0$ si C es constante
4. $ab > 0 \;\Rightarrow\; \rho(aX,\, bY) = \rho(X,\, Y)$
5. Si existen constantes $a \neq 0$, b y $Y = aX + b \;\Leftrightarrow\; \rho(X,\, Y) = sign(a) = \begin{cases} 1 & \text{si } a > 0 \\ -1 & \text{si } a < 0 \end{cases}$

6.1 VA2D Discretas

Cuando los recorridos de ambas variables es *discreto* entonces el recorrido de la va2d es discreto.

6.1.1 Función de Probabilidad Conjunta

Se define como:

$$p_{XY} : R_{XY} \to [0,\, 1] / \, p_{XY}(x,\, y) = P(X = x,\, Y = y) = \begin{cases} 0 & \text{si } (x,\, y) \notin R_{XY} \\ p_{XY}(x,\, y) > 0 & \text{si } (x,\, y) \in R_{XY} \end{cases}$$

6.1.2 Cálculo de Probabilidades de Sucesos

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

6.1.3 Valores Esperados de Funciones de X e Y

$$E(G(X,\, Y)) = \sum_{(x,\, y)\in R_{XY}} \sum G(x,\, y) p_{XY}(x,\, y)$$

6.1.4 Distribuciones de Probabilidad de Funciones de X e Y

Tomo $Z = G(x,\, y)$. Entonces:

$$F_Z(z) = P(Z \leq z) = \sum_{G(x,y)\leq z} \sum p_{XY}(x,\, y)$$

6.1.5 Funciones de Probabilidad Marginales

$$p_X(x) = \textstyle\sum_{y\in R_Y} p_{XY}(x,\, y) \quad (1)$$

$$p_Y(y) = \textstyle\sum_{x\in R_X} p_{XY}(x,\, y) \quad (2)$$

En (1) la variable x está fija, y solo varía la y. En (2) la variable y está fija, y solo varía la x.

6.1.6 Función de Probabilidad Condicional

$$p_{X|Y=y}(x,y) = \frac{p_{XY}(x,y)}{p_Y(y)} \quad \text{con } p_Y(y) \neq 0$$

$$p_{Y|X=x}(x,y) = \frac{p_{XY}(x,y)}{p_X(x)} \quad \text{con } p_X(x) \neq 0$$

6.1.7 Generalización de Probabilidad de Intersección

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

6.1.8 Independencia de VA2D Discretas

Dos va2d discretas son independientes sii:

$$p_{XY}(x,\, y) = p_X(x)p_Y(y) \;\; \forall x,\, y$$

6.2 VA2D Continuas

Cuando el recorrido de ambas variables son intervalos reales, o unión de ellos, entonces la va es *continua*.

6.2.1 Función de Densidad de Probabilidad Conjunta

$$P(A) = \int\int_A f_{X,Y}(x,\, y)dxdy$$

Propiedades

1. $f_{X,Y}(x,\, y) \geq 0 \;\; \forall (x,\, y) \in \mathbb{R}^2$
2. $\int\int_{\mathbb{R}^2} f_{X,Y}(x,\, y)dxdy = 1$

6.2.2 Cálculo de Probabilidades de Sucesos

$$P(A) = P((X,\, Y) \in A) = \int\int_{(x,y)\in A} f_{X,Y}(x,\, y)dxdy$$

6.2.3 Valores Esperados de Funciones de X e Y

$$E(G(X,\, Y)) = \int\int_{\mathbb{R}^2} G(x,y) f_{X,Y}(x,\, y)dxdy$$

6.2.4 Distribuciones de Probabilidad de Funciones de X e Y

Tomo $Z = G(x,\, y)$. Entonces:

$$F_Z(z) = P(Z \leq z) = \int\int_{A_Z} f_{X,Y}(x,\, y)dxdy$$

donde $A_Z = \{(x,\, y)/\, G(x,\, y) \leq z\}$

6.2.5 Funciones de Densidad de Probabilidad Marginales

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

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

En (1) la variable x está fija, y solo varía la y. En (2) la variable y está fija, y solo varía la x.

6.2.6 Función de Densidad de Probabilidad Condicional

$$f_{X|Y=y}(x, y) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad \text{con } f_Y(y) > 0$$

$$f_{Y|X=x}(x, y) = \frac{f_{XY}(x,y)}{f_X(x)} \quad \text{con } f_X(x) > 0$$

6.2.7 Generalización de Probabilidad de Intersección

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

6.2.8 Independencia de VA2D Continuas

Dos va2d continuas son independientes sii:

$$f_{XY}(x,\, y) = f_X(x) f_Y(y) \;\; \forall (x,\, y) \in \mathbb{R}^2$$

7 Suma de Variables Independientes Idénticamente Distribuidas

Sean X, Y va iid con $E(X) = \mu$ y $V(X) = \sigma^2$. Entonces:

1. $E(X + Y) = 2\mu$
2. $V(X + Y) = 2\sigma^2$
3. $E(\sum_{i=1}^{n} X_i) = n\mu$
4. $V(\sum_{i=1}^{n} X_i) = n\sigma^2$
5. $E(\bar{X}_n) = \mu$ donde $\bar{X}_n = \frac{1}{n}\sum_{n}^{i=1} X_i$ (promedio)
6. $E(\bar{X}_n) = \frac{\sigma^2}{n}$ donde $\bar{X}_n$ es el promedio

Casos Especiales

1. $X_1 \sim Bin(n_1,\, p)$, $X_2 \sim Bin(n_1,\, p)$

$$\Rightarrow X_1 + X_2 \sim Bin(n_1 + n_2,\, p)$$

2. $X_1 \sim Poiss(\lambda_1)$, $X_2 \sim Poiss(\lambda_2)$

$$\Rightarrow X_1 + X_2 \sim Poiss(\lambda_1 + \lambda_2)$$

3. $X_1 \sim N(\mu_1,\, \sigma_1)$, $X_2 \sim N(\mu_2,\, \sigma_2)$

$$\Rightarrow X_1 + X_2 \sim N\left(\mu_1 + \mu_2,\, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

4. $X_1,\, X_2,\, ...,\, X_n$ va iid con $X_i \sim Bernoulli(p)$

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

5. $X_1,\, X_2,\, ...,\, X_n$ va iid con $X_i \sim Exp(\lambda)$

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

donde $f_{S_n}(t) = \frac{\lambda(\lambda t)^{n-1}e^{-\lambda t}}{(n-1)!}$

7.1 Teorema Central del Límite

Sean $X_1,\, X_2,\, ...X_n$ va iid con $E(X_1) = \mu$ y $V(X_1) = \sigma^2$. Luego para n grande se tiene que:

$$\bar{X}_n \approx N\left(\mu,\, \frac{\sigma}{\sqrt{n}}\right)$$

$$\sum_{i=1}^{n} X_i \approx N\left(n\mu,\, \sqrt{n}\sigma\right)$$

Ejemplos

1. $X_i \sim U(0,\,1)$
$$\Rightarrow\ \bar{X}_n \approx N\left(0.5,\ \frac{1}{\sqrt{12n}}\right)$$

2. $X_i \sim \mathit{Bernoulli}(p)$, $S_n = \sum_{i=1}^{n} X_i$
$$\Rightarrow\ S_n \sim \mathit{Bin}(n,\,p) \approx N(np,\ \sqrt{np(1-p)})$$

3. $X_i \sim \mathit{Exp}(\lambda)$, $S_n = \sum_{i=1}^{n} X_i \sim \mathit{Gamma}(n,\,\lambda)$
$$\Rightarrow\ S_n \approx N\left(\frac{n}{\lambda},\ \frac{\sqrt{n}}{\lambda}\right)$$

7.2 Ley Débil de los Grandes Números

Sean X_1, X_2, ... va iid con $E(X_1) = \mu$ y $V(X_1) = \sigma^2$. Entonces:

$$lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0\ \ \forall\epsilon$$

Observación: Con esta ley, podemos estimar valores medio de va y probabilidades realizando repeticiones independientes de un experimento un gran número de veces.

7.3 Desigualdad de Markov

Sea X una va solo toma valores positivos. Entonces:

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

7.4 Desigualdad de Chebychev

Sea X una va con $E(X) = \mu$ y $V(X) = \sigma^2$. Entonces:

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

con $\epsilon > 0$

8 Inferencia Estadística

8.1 Definiciones

Definición 1. Parámetros poblacionales

Sea X una variable aleatoria definida en la población. Entonces, los *parámetros de la distribución de probabilidades de* X se denominan parámetros poblacionales.
Toma el valor 1 si sucede A, y 0 si no sucede $\Rightarrow$ $P(A) = p$

Definición 2. Muestra

Subconjunto de la población. El muestreo es aleatorio, es decir, todos los miembros de la población tienen la *misma probabilidad* de pertenecer a la muestra.
Hay 2 tipos:

1. Muestreo con Reemplazo: Cada observación se realiza sobre un miembro de la población y se retorna.
2. Muestreo sin Reemplazo: Cada observación se realiza sobre un miembro de la población y no se retorna.

Definición 3. Estimador/Estadístico

Sea $\{X_1, X_2, ..., X_n\}$ una *muestra aleatoria* de tamaño n.
Toda función real de la muestra aleatoria define una va que se denomina estimador/estadístico.

$$T = H(X_1, X_2, ..., X_n)$$

Los estimadores son **va, son función de varias va iid**.
Notación: Decimos que $\hat{\theta}$ es estimador del parámetro poblacional θ.

Propiedades deseables

1. Estimador Insesgado: $\hat{\theta}$ es un estimador insesgado del parámetro poblacional θ si

$$E(\hat{\theta}) = \theta$$

   y quiere decir que las va están centradas en lo que se busca estimar.
   Para ver si un estimador es o no insesgado, es básicamente sacar el valor estimado y ver si coincide con lo que quiero estimar.

2. Estimador Consistente: $\hat{\theta}$ es un estimador consistente del parámetro poblacional θ si

$$\lim_{n\to\infty} E((\hat{\theta} - \theta)^2) = 0$$

8.2 Estimadores/Estadísticos Importantes

Media Muestral

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Sea $Z = \frac{\bar{X}-\mu}{\sigma}\sqrt{n}$. Entonces,

- Si X tiene distribución normal $\Rightarrow Z \sim N(0,\,1)$
- Si X no tiene distribución normal $\Rightarrow Z$ tiene distribución asintótica $(n \to \infty)$

Varianza Muestral

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

Sea $Z = \frac{\hat{p}-p}{\sqrt{p(1-p)}}\sqrt{n} \Rightarrow Z \sim N(0,\,1)$ en forma asintótica $(n \to \infty)$.

Proporción Muestral

$$\hat{p} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Si $X \sim N(\mu,\,\sigma) \Rightarrow D^2 = \frac{S^2(n-1)}{\sigma^2} \sim \chi^2$ con $n-1$ grados de libertad.

8.3 Estimadores

Como $\{X_1,\,...,\,X_n\}$ son va iid. Entonces, $E(X_k) = \mu$ y $V(X_k) = \sigma^2\ \forall k \leq n$. Luego:

$$E(X) = \mu \text{ y } V(X) = \sigma^2$$

De esta manera,

1. $E(\bar{X}) = \mu$
2. $V(\bar{X}) = \frac{\sigma^2}{n}$
3. $E(S^2) = \sigma^2$
4. $V(S^2) = \frac{2\sigma^4}{n-1}$ si $X \sim N(\mu,\,\sigma)$
5. $E(\hat{p}) = p$
6. $V(\hat{p}) = \frac{p(1-p)}{n}$

8.4 Estimación de la proporción

Sean $X_1,\,...,\,X_n \sim Be(p)$ va iid. Podemos estimar $\hat{p}$ como

$$\hat{p} \approx N\left(p,\,\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right)$$

Hay dos maneras:

1. **Puntual**: Evaluando la proporción muestral para una muestra y resulta un número (que será la estimación puntual p).

2. **Intervalo de Confianza**: Intervalo con centro en la estimación puntual y semi-amplitud dependiente de la distribución de probabilidades de la proporción muestral, y el tamaño de la muestra.

$$IC_\gamma = \left(\mu - \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}} ; \mu + \frac{\sigma}{\sqrt{n}} z_{\frac{1+\gamma}{2}}\right)$$

donde $z_{\frac{1+\gamma}{2}} = \phi^{-1}(\frac{1+\gamma}{2})$ y γ es el nivel de confianza.

8.5 Distribución t-Student

Sean $X_1, ..., X_n \sim N(\mu, \sigma)$ va iid. Entnonces se define la *distribución t-Student con* $n-1$ *grados de libertad* como:

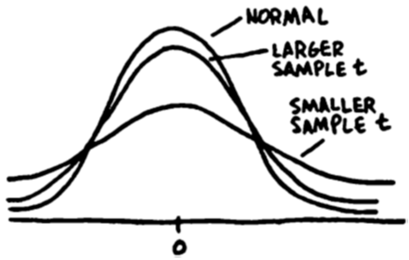
$$T = \frac{\bar{X}_n - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

donde $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ y $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$.

Propiedades

1. $E(T) = 0$
2. $V(T) = \begin{cases} \frac{n-1}{n-3} & \text{si } n > 3 \\ \infty & \text{si } 2 < n \leq 3 \end{cases}$
3. $\kappa(T) = \begin{cases} \frac{6}{n-5} & \text{si } n > 5 \\ \infty & \text{si } 3 < n \leq 5 \end{cases}$

Notemos que $lim_{n\to\infty} V(T) = 1$ y $lim_{n\to\infty}\kappa(T) = 0$. Además, $T \approx N(0, 1)$.



8.6 Intervalos de Confianza para la media

Tenemos dos maneras de hacerlo:

1. Si *n es grande* $\Rightarrow$ $(\mu - z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}};\ \mu + z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}})$
2. Si *n no es muy grande* $\Rightarrow$ $(\bar{X}_n - \frac{S}{\sqrt{n}} t_{n-1,\, \frac{1+\gamma}{2}};\ \bar{X}_n + \frac{S}{\sqrt{n}} t_{n-1,\, \frac{1+\gamma}{2}})$

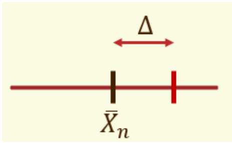
donde $t_{\nu,\,\alpha}$ se define como $P(T \leq t_{\nu,\,\alpha}) = \alpha$ si $T \sim t_{\nu,\,\alpha}$ (esta en tablas).
Observación: $t_{\nu,\,\alpha} = -t_{\nu,\,1-\alpha}$

Tenemos dos tipos de intervalos: *unilaterales* y *bilaterales*

Unilaterales

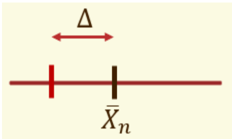
Puede ser el *intervalo a derecha*:

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



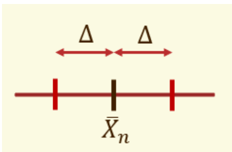
o el *intervalo a izquierda*:

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



Bilaterales

$$\left(\bar{X}_n - \frac{S}{\sqrt{n}} t_{n-1,\,\frac{1+\gamma}{2}};\ \bar{X}_n + \frac{S}{\sqrt{n}} t_{n-1,\,\frac{1+\gamma}{2}}\right)$$



9 Prueba de Hipótesis

Una *prueba de hipótesis* es un procedimiento con el que se busca tomar una decisión sobre el *valor de verdad de una hipótesis estadística*. Al hacerlo decidimos si rechazar o no esa hipótesis estadística. Siempre tenemos dos hipótesis:

- Hipótesis nula (H_0): Es sostener lo que se dice o lo que se viene haciendo hasta ahora. Es decir, mantener el status quo.
- Hipótesis alternativa (H_1) : Es sostener lo opuesto a lo que se dice o cambiar lo que se viene haciendo hasta ahora.

Ejemplo

Una agencia de protección al consumidor desea poner a prueba la afirmación de un fabricante de pinturas según la cual el tiempo medio de secado de su nueva pintura de secado rápido es de no más de 20 min. Entonces:

- H_0: El fabricante dice la verdad, por lo que el tiempo de secado es ≤ 20.
- H_1: El fabricante miente, por lo que el tiempo de secado es > 20.

Pero no podemos estar 100% seguros de esto ya que el azar puede hacer que cometamos un error. Por lo que tenemos que tener en cuenta distintos casos:

<table>
<tr><td></td><td></td><td colspan="2">Toma de decisión basada en H_0</td></tr>
<tr><td></td><td></td><td>Aceptamos H_0</td><td>Rechazamos H_0</td></tr>
<tr><td rowspan="2">Valor de verdad de H_0</td><td>H_0 verdadera</td><td>:D</td><td>ERROR TIPO I (α)</td></tr>
<tr><td>H_0 falsa</td><td>ERROR TIPO II (β)</td><td>:D ("potencia de la prueba")</td></tr>
</table>

Observación: La *potencia de la prueba* se puede calcular como: $1 - \beta$
Para esto, tenemos dos **estrategias**:

1. Fijar, de antemano, un valor α_m como cota superior al ERROR TIPO I.
2. Reportar qué tan probable es que una muestra haya sido tan contraria a H_0. Es decir, "asumiendo que H_0 es verdadera, ¿cuán probable es no observar lo observado?". Entonces, si es muy poco probable es "malo". A esto lo llamamos *p-value* (es una medida de qué tan biased está nuestra H_0).

9.1 Definiciones

Definición 1. Nivel de significación de la prueba

Es la máxima probabilidad de cometer un error de tipo I que es posible admitir. Normalmente se fija en un valor α_m.

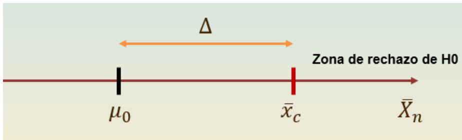
Definición 2. Valor p/p-value

El valor p se define como la probabilidad de que *si la hipótesis nula es verdadera el estadístico de prueba dé tan mal o peor de lo que dio.*

9.2 Pruebas de Hipótesis en $\bar{X}_n$ con σ conocido

Dado $X \sim N(\mu, \sigma)$, con σ conocido.
Además, como *estadístico* se utilizará el valor medio: $\bar{X}_n \sim N(\mu, \frac{\sigma}{\sqrt{n}})$

9.2.1 Prueba de cola derecha



Es cuando las hipótesis cumplen lo siguiente:

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$

Estrategia 1

Fijamos en α_m la *máxima probabilidad de cometer un error de tipo I* que es posible admitir.
Entonces, rechazaremos H_0 si $\bar{X}_n$ es mas grande que $\bar{x}_c = \mu_0 + z_{1-\alpha_m} \frac{\sigma}{\sqrt{n}}$. O lo que es lo mismo:

$$P_{\mu_0}(\bar{X}_n > \bar{x}_c) = 1 - \Phi\left(\frac{\bar{x}_c - \mu_0}{\sigma}\sqrt{n}\right) > \alpha_m$$

Si esto se cumple, entonces *rechazaremos* H_0.

Si H_0 fuera falsa, estaríamos cometiendo un error que depende de "qué tan falsa sea":

$$\beta(\mu) = P_\mu(\bar{X}_n \leq \bar{x}_c) = \Phi\left(\frac{\bar{x}_c - \mu}{\sigma}\sqrt{n}\right)$$

Algunos *valores conocidos* de β:

1. $\beta(\mu_0) = 1 - \alpha_m$
2. $1 - \beta(\mu_0)$ se llama potencia de la prueba
3. $\beta(\bar{x}_c) = 0.5$
4. $\lim_{\mu \to \infty} \beta(\mu) = 0$

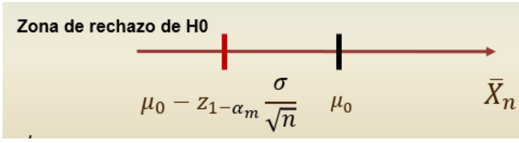
Estrategia 2

Reportamos que tan probable es que una muestra haya sido tan contraria a H_0. O lo que es lo mismo:

$$\underbrace{P_{\mu_0}(\bar{X}_n > \bar{x}_{obs})}_{\text{p-value}} = 1 - \Phi\left(\frac{\bar{x}_{obs} - \mu_0}{\sigma}\sqrt{n}\right) < \alpha_m$$

Si esto se cumple, entonces *rechazaremos* H_0.

9.2.2 Prueba de cola izquierda



Es cuando las hipótesis cumplen lo siguiente:

$$H_0 : \mu \geq \mu_0$$
$$H_1 : \mu < \mu_0$$

Estrategia 1

Fijamos en α_m la *máxima probabilidad de cometer un error de tipo I* que es posible admitir. Entonces, rechazaremos H_0 si $\bar{X}_n$ es mas chico que $\bar{x}_c = \mu_0 - z_{1-\alpha_m}\frac{\sigma}{\sqrt{n}}$. O lo que es lo mismo:

$$P_{\mu_0}(\bar{X}_n \leq \bar{x}_c) = \Phi\left(\frac{\bar{x}_c - \mu_0}{\sigma}\sqrt{n}\right) > \alpha_m$$

Si esto se cumple, entonces *rechazaremos* H_0.

Si H_0 fuera falsa, estaríamos cometiendo un error que depende de "qué tan falsa sea":

$$\beta(\mu) = P_\mu(\bar{X}_n > \bar{x}_c) = 1 - \Phi\left(\frac{\bar{x}_c - \mu}{\sigma}\sqrt{n}\right)$$

Algunos *valores conocidos* de β:

1. $\beta(\mu_0) = 1 - \alpha_m$
2. $1 - \beta(\mu_0)$ se llama potencia de la prueba
3. $\beta(\bar{x}_c) = 0.5$
4. $\lim_{\mu\to-\infty} \beta(\mu) = 0$

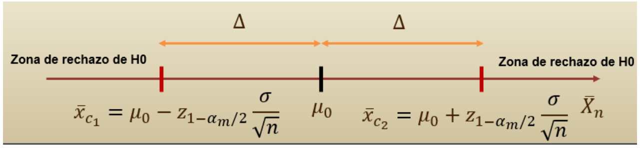
Estrategia 2

Reportamos que tan probable es que una muestra haya sido tan contraria a H_0. O lo que es lo mismo:

$$\underbrace{P_{\mu_0}(\bar{X}_n < \bar{x}_{obs})}_{\text{p-value}} = \Phi\left(\frac{\bar{x}_{obs} - \mu_0}{\sigma}\sqrt{n}\right) < \alpha_m$$

Si esto se cumple, entonces *rechazaremos* H_0.

9.2.3 Prueba de dos colas



Es cuando las hipótesis cumplen lo siguiente:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

Estrategia 1

Fijamos en α_m la *máxima probabilidad de cometer un error de tipo I* que es posible admitir. Entonces, rechazaremos H_0 si $\bar{X}_n$ es mas chica que $\bar{x}_{c_1} = \mu_0 - z_{1-\alpha_m}\frac{\sigma}{\sqrt{n}}$ o mas grande $\bar{x}_{c_2} = \mu_0 + z_{1-\alpha_m}\frac{\sigma}{\sqrt{n}}$. O lo que es lo mismo:

$$P_\mu(\bar{X}_n < \bar{x}_{c_1} \cup \bar{X}_n > \bar{x}_{c_2}) = 1 - \left[\Phi\left(\frac{\bar{x}_{c_2} - \mu}{\sigma}\sqrt{n}\right) - \Phi\left(\frac{\bar{x}_{c_1} - \mu}{\sigma}\sqrt{n}\right)\right] > \alpha_m$$

Si esto se cumple, entonces *rechazaremos* H_0.

Si H_0 fuera falsa, estaríamos cometiendo un error que depende de "qué tan falsa sea":

$$\beta(\mu) = P_\mu(\bar{x}_{c_1} \leq \bar{X}_n \leq \bar{x}_{c_2}) = \Phi\left(\frac{\bar{x}_{c_2} - \mu}{\sigma}\sqrt{n}\right) - \Phi\left(\frac{\bar{x}_{c_1} - \mu}{\sigma}\sqrt{n}\right)$$

Algunos *valores conocidos* de β:

1. $\beta(\mu_0) = 1 - \alpha_m$ (potencia de la prueba)
2. $1 - \beta(\mu_0)$ se llama potencia de la prueba
3. $\beta(\bar{x}_{c_1}) = \beta(\bar{x}_{c_1}) \approx 0.5$
4. $\lim_{\mu\to\pm\infty} \beta(\mu) = 0$

Estrategia 2

Reportamos que tan probable es que una muestra haya sido tan contraria a H_0. O lo que es lo mismo:

$$\underbrace{P_{\mu_0}(|\bar{X}_n - \mu_0| > \Delta_{obs})}_{\text{p-value}} = 2\left[1 - \Phi\left(\frac{\Delta_{obs}\sqrt{n}}{\sigma}\right)\right] < \alpha_m$$

donde $\Delta_{obs} = |\bar{x}_{obs} - \mu_0|$.
Si esto se cumple, entonces *rechazaremos* H_0.

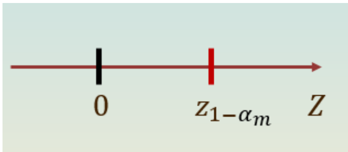
9.3 Pruebas de Hipótesis en Z con σ conocido

Usando $Z = \frac{(\bar{X}_n - \mu_0)\sqrt{n}}{\sigma}$ obtenemos un *estadístico de muestra adimensional.*
Además sabemos que para n grande se cumple:

$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Luego, podemos hacer lo mismo que hicimos para $\bar{X}_n$ en la proporción.

9.3.1 Prueba de cola derecha

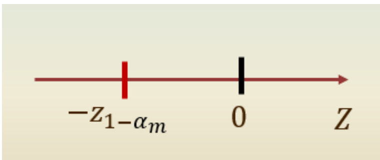


Es cuando las hipótesis cumplen lo siguiente:

$$H_0 : p \leq p_0$$
$$H_1 : p > p_0$$

Además, **p-value** $= 1 - \Phi(|z_{obs}|)$

9.3.2 Prueba de cola izquierda

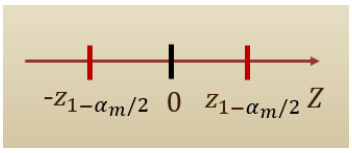


Es cuando las hipótesis cumplen lo siguiente:

$$H_0 : p \geq p_0$$
$$H_1 : p < p_0$$

Además, **p-value** $= \Phi(|z_{obs}|)$

9.3.3 Prueba de dos colas



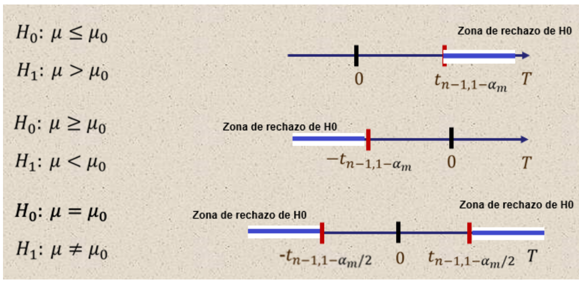
Es cuando las hipótesis cumplen lo siguiente:

$$H_0 : p = p_0$$
$$H_1 : p \neq p_0$$

Además, **p-value** $= 2[1 - \Phi(|z_{obs}|)]$

9.4 Pruebas de Hipótesis en $\bar{X}_n$ con σ desconocido

Dada $X \sim N(\mu, \sigma)$. Se considera una muestra aleatoria $\{X_1, ..., X_n\}$ de tamaño n de X.
Tomamos $T = \frac{\bar{X}-\mu}{S}\sqrt{n} \sim t_{n-1}$



Observación: Si comparamos los fractiles de t y la *normal estándar*, entonces se puede ver que:

$$t_{n-1,\,w} > z_w \; \forall w \in (0,\, 1)$$

10 Procesos Estocásticos

Un proceso estocástico es un proceso/sistema que se desarrolla/evoluciona en el tiempo (y/o espacio) mientras que pasa por fluctuaciones al azar.
En cada instante/posición se tiene una **va** cuyo valor es **estado del proceso** y cuya distribución de probabilidades **puede cambiar con el tiempo/posición**.

10.1 Definiciones

Definición 1. Proceso Estocástico

Conjunto de variables aleatorias $X = \{X(t) : t \in T\}$

Definición 2. Espacio del Parámetro

Es la T que aparece en $X = \{X(t) : t \in T\}$

Definición 3. Espacio de Estados

Es la unión de los recorridos de las variables $X(t)$.
Notación: E

Ejemplos

- $X(t)$: La temperatura corporal tomada en la axila izquierda en el instante t.
  $\Rightarrow T = \mathbb{R}_{\geq 0}$, $E \subset \mathbb{R}$
- $N(t)$: Número de personas que han arribado a la fila de un banco hasta el instante t.
  $\Rightarrow T = \mathbb{R}_{\geq 0}$, $E \subset \mathbb{N}_0$

Observación: Notemos que en el primer caso T y E son *continuos*, mientras que en el segundo T es continuo pero E es *discreto*.

Definición 4. Incremento Estacionario

Se dice que un proceso estocástico tiene *incrementos estacionarios* sii:

$$P(X(t_2 + \tau) - X(t_1 + \tau) \leq x) = P(X(t_2) - X(t_1)) \;\; \forall t_1 < t_2, \forall \tau, \forall x$$

Es decir, *la distribución de incrementos "no depende" del tiempo.*
Por ejemplo, la caminata aleatoria es de incrementos estacionarios.

Definición 5. Incremento Independientes

Se dice que un proceso estocástico tiene *incrementos independientes* sii:

$$\forall n, \forall (l_1, u_1), ..., (l_n, u_n) \text{ con } l_i < u_i \text{ y } (l_k, u_m) \cap (l_k, u_k) = \varnothing \text{ si } m \neq k$$

se cumple que las variables aleatorias:

$$\Delta_i = X(u_i) - X(l_i) \text{ son independientes entre sí}$$

Luego, podemos describir al proceso como *una suma de incrementos independientes* de la siguiente manera:

$$l_i = u_{i-1} \text{ y } X(u_n) - X(l_1) = \sum_{i=2}^{n} \Delta_i$$

Por ejemplo, la caminata aleatoria es de incrementos independientes.

Definición 5. Proceso de Conteo

Se dice que un proceso estocástico $N(t)$ *es de conteo* sii:

1. $N(0) = 0$
2. $N(t) \geq 0 \;\; \forall t \geq 0$
3. $s < t \;\Rightarrow\; N(s) \leq N(t)$
4. $E = \mathbb{N}_0$

Observación: $N(t) - N(s)$ son los eventos que ocurrieron *después de s pero antes de t.*

Definición 6. Cadenas homogénes

Es cuando en una *cadena de Markov*, las probabilidades de transición no dependen del tiempo:

$$p_{ij}(n) = p_{ij}$$

10.2 Procesos de Poisson

Se dice que un **proceso estocástico de conteo** $N(t)$ es un *proceso de Poisson de parámetro* λ sii:

1. Tiene **incrementos independientes**.
2. Los **incrementos son estacionarios**.
3. La probabilidad de que exactamente **un evento** ocurra en un intervalo de tiempo de longitud h es $\lambda h + o(h)$
4. La probabilidad de que **más de un evento** ocurra en un intervalo de tiempo de longitud h es $o(h)$

Características de los Procesos de Poisson

1. La probabilidad de que en un intervalo de tiempo $(0, t)$ hayan ocurrido n eventos del proceso de Poisson es:
$$p_n(t+h) = P(N(t+h) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$
entonces $N(t) \sim Poisson(\lambda t)$
2. Los tiempos entre eventos (τ) tienen distribución exponencial.
Los tiempos entre eventos (τ) son independientes entre sí.

   $\Rightarrow$ Los tiempos entre eventos (τ) son variables iid con distribución exponencial de parámetro λ

3. El tiempo S_k hasta la ocurrencia del evento k es una vac con distribución *Gamma* de parámetros k y λ. Luego, se verifica que:

$$P(S_k < t) = P(N_t \geq k) = 1 - P(N_t < k)$$

10.3 Cadenas de Markov

Se dice que un proceso estocástico es un proceso de Markov sii $\forall n$, $\forall t_1 < t_2 < ... < t_n$, $\forall x$, se cumple que:

$$P(X(t_n) \leq x | X(t_{n-1}), ..., X(t_1)) = P(X(t_n) \leq x | X(t_{n-1}))$$

es decir, *sólo importa el valor más reciente.*
De esta manera, se llama **cadena de Markov** a un proceso de Markov que tiene un *spacio de parámetro discreto*: $T = \mathbb{N}^0$, y (en nuestro caso) *un espacio de estados discreto* $E = \{e_1, e_2, ...\}$.

Observaciones

Una cadena de Markov $\{X(n)\}_{n\in\mathbb{N}^0}$ queda completamente descripta por:

1. La distribución de probabilidades del estado inicial.

$$p_j(0) = P(X(0) = e_j)$$
$$\sum_j p_j(0) = 1$$

que tambien se puede representar como *vector de probabilidades*:

$$\vec{p}(n) = (p_1(n) \;\; p_2(n) \;\; ...)$$

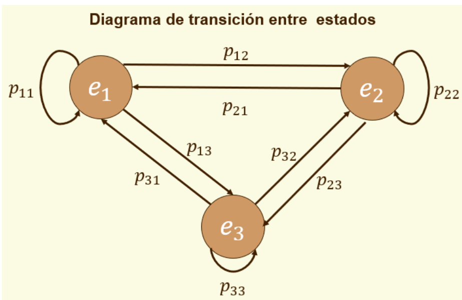
2. Las probabilidades de transición entre cada par de estados.

$$p_{ij}(n) = P(X(n+1) = e_j | X(n) = e_j)$$

que también se pueden representar como *matriz de probabilidades de transición de un sólo paso*:

$$\mathbb{P}(n) = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

Esto último se puede ver gráficamente como un grafo dirigido de la siguiente manera:



10.3.1 Ecuación de Chapman-Kolmogorov

Forma vectorial: $p_j(n+1) = \sum_i p_i(n) p_{ij}(n)$

Forma matricial: $\vec{p}_j(n+1) = \vec{p}(n)\mathbb{P}(n)$

Si la *cadena fuese homogénea*, entonces: $\vec{p}(n+1) = \vec{p}(0)\mathbb{P}^n$ por lo que la **matriz de probabilidades en k pasos** se obtendría como: $\mathbb{P}^{(k)} = \mathbb{P}^k$

10.3.2 Probabilidad a largo plazo/Distribución estacionaria

$$\vec{\pi} = \lim_{n\to\infty} \vec{p}(n) = \left[\lim_{n\to\infty} \vec{p}(n)\right] \mathbb{P} = \vec{\pi}\mathbb{P}$$

Observaciones

1. Si existe $\Rightarrow$ se corresponde con un autovector a izquierda correspondiente al autovalor 1 de $\mathbb{P}$.
2. No toda distribución que satisface la ecuación es una distribución estacionaria.

La **condición suficiente para la existencia de una distribución estacionaria** es:
"Dada una cadena de Markov, si $\exists k / \mathbb{P}^k$ tiene todos sus elementos positivos $\Rightarrow$ existe una distribución de probabilidad estacionaria".