Curso de Nivelación:
Probabilidad y estadística

Clase 1

Cronograma

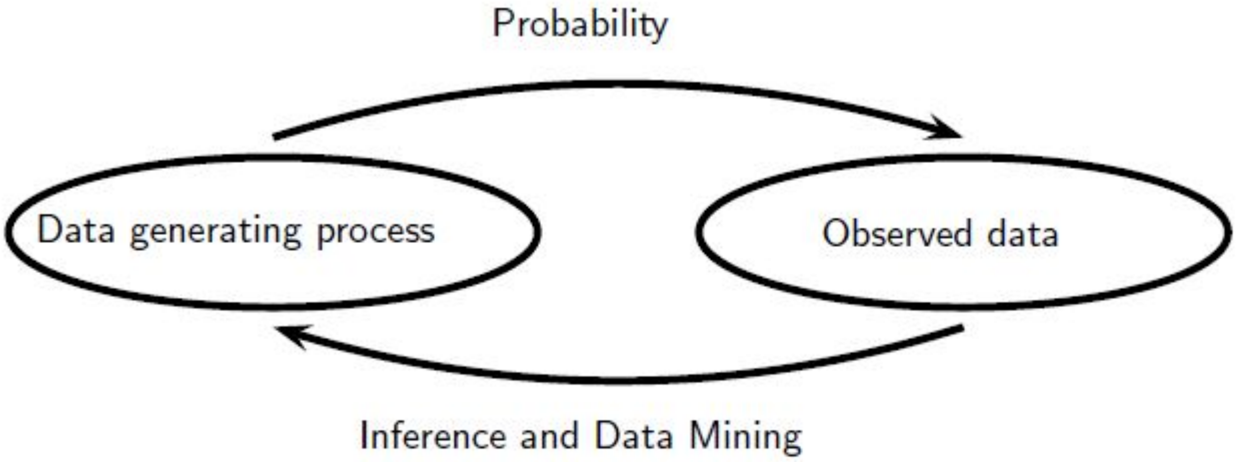
- **Clase 1:**
  - - Noción de probabilidad
  - - Espacio de probabilidad
  - - Espacios equiprobables, conteo (ideas básicas)
  - - Probabilidad condicional, teoremas de probabilidad total y Bayes
- **Clase 2:**
  - - Definición de variables aleatorias: variables continuas y discretas, función de densidad, de probabilidad y de distribución.
  - - Vectores aleatorios: distribución conjunta, distribución condicional, independencia.
  - - Momentos de variables aleatorias.

¿Qué es la probabilidad y la estadística?

¿Qué es la probabilidad?

El término **probabilidad** se refiere al estudio del azar y la incertidumbre en cualquier situación en la que varios escenarios pueden ocurrir.

Probabilidad vs. estadística



“All of Statistics: A Concise Course in Statistical Inference”, Larry Wasserman

Población y muestra

Una **población** es el conjunto total de los sujetos o unidades de análisis de interés

Una **muestra** es cualquier subconjunto de sujetos o unidades de aná;lisis de la población bajo estudio

Algunos elementos de la probabilidad

Un **experimento aleatorio:** acciones o procesos en los cuales conocemos todos los resultados posibles pero no sabemos con certeza cuál va a ocurrir.

Llamamos **espacio muestral** (Ω) al conjunto de resultados posibles de mi experimento aleatorio.

Un **evento o suceso** es cualquier subconjunto de resultados en el espacio muestral

Ejercicio 1

Proponer los espacios muestrales asociados a los siguientes experimentos:

a. Arrojo una moneda y observo el resultado
b. Tomo un caño de un depósito y mido su longitud
c. Arrojo un dado seis veces y observo cada uno de los resultados
d. Arrojo un dado seis veces y cuanto la cantidad de 1 observados
e. Registro la cantidad de llamadas en un call center entre las 9 y las 11hs

Una primera aproximación a las probabilidades

La **probabilidad** de un evento A es un número positivo (o nulo) que se le asigna a cada suceso o evento del espacio muestral que nos habla de la certeza que se tiene sobre ese evento.

Supongamos que realizamos un experimento n veces, y estamos interesados en un evento particular A. Luego podemos definir:

- **Frecuencia absoluta:** Cantidad de veces que ocurre A en mis n experimentos
- **Frecuencia relativa (f_A):** la proporción de veces que ocurrió A entre mis n experimentos.

Si la cantidad de ensayos es lo suficientemente grande, esperamos que $f_A \approx \mathbb{P}(A)$

Más rigurosamente, $f_A \overset{n\to\infty}{\to} \mathbb{P}(A)$

Espacios equiprobables

Si estamos en presencia de un espacio **equiprobable**, es decir donde todos los elementos tienen las mismas chances de ocurrir, las probabilidades pueden calcularse como la proporción entre la cantidad de casos donde ocurre el experimento y la cantidad de elementos que existen en el espacio muestral. A esto se lo conoce como regla de **Laplace**

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{\#\text{"casos favorables"}}{\#\text{casos totales}}$$

Ejercicio 2

Vamos con un ejemplo donde conocemos la totalidad de la población:

Una empresa con 100 empleados divide a los mismos según el área de trabajo y la cantidad de años en la empresa. Se sabe que

#años\area	Contable y RRHH	Instalaciones	Atención al público
0	0	30	15
1	1	20	10
2	3	0	6
3	1	4	3
4+	5	1	1

Se elige una persona al azar de esa población. Hallar la probabilidad de que:

a. un empleado del área de Instalaciones esté trabajando hace exactamente 3 años
b. un empleado haya trabajado más de 5 años en la empresa
c. un empelado haya trabajado más de 3 años en la empresa y no sea de recursos humanos
d. un empleado esté trabajando hace más de 5 años o sea del área instalaciones

Ejercicio 3

Se arroja un dado equilibrado 2 veces. Calcular la probabilidad de

a. observar un 1 en el primer tiro y un 2 en el segundo
b. Observar un 1 y un 2
c. Observar dos resultados pares
d. La suma de los resultados no supere 7

Álgebra de eventos

Dado un Ω, definimos al **álgebra de eventos** ($\mathcal{A}$) como una familia de subconjuntos de Ω que satisface:

1. $\Omega \in \mathcal{A}$
2. $B \in \mathcal{A} \Rightarrow B^c \in \mathcal{A}$
3. $B, C \in \mathcal{A} \Rightarrow B \cup C \in \mathcal{A}$

Probabilidad: definición formal

Def: una **probabilidad** es una función $\mathbb{P} : \Omega \to [0,1]$ que satisface:

- $0 \leq \mathbb{P}(A) \leq 1, \ \forall A \in \mathcal{A}$
- $\mathbb{P}(\Omega) = 1$
- Si $A \cap B = \emptyset, \ \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

El conjunto $(\Omega, \mathcal{A}, \mathbb{P})$ define lo que se conoce como **espacio de probabilidad**

El álgebra representa los resultados “medibles” del experimento.

Corolario:

- $\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) = 1 \Rightarrow \mathbb{P}(A) = 1 - \mathbb{P}(A^c)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Ejercicio 4

En Argentina, el 80 % de los programadores usa Java, C o ambos; el 50 % usa Java y el 40 % usa C. ¿Cuál es la probabilidad de que al elegir un programador al azar use:

a) Java y C?

b) sólo Java?

c) sólo C?

d) ninguno de los dos lenguajes?

Probabilidades condicionales y proba. total

Def: Se llama **probabilidad condicional** de A dado B ($\mathbb{P}(A|B)$) a la probabilidad de que ocurra el evento A sabiendo que B ha ocurrido, y está definida por

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

Def: Diremos que los eventos $B_1, \ldots B_n$ forman una **partición** si $B_i \cap B_j = \emptyset \, \forall i, j$ y $\bigcup_{i=1}^{n} B_j = \Omega$.

Luego podemos describir al evento A como $A = (A \cap B_1) \cup \ldots \cup (A \cap B_n)$ de forma que

$$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \ldots + \mathbb{P}(A \cap B_n) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \ldots + \mathbb{P}(A|B_n)\mathbb{P}(B_n)$$

Fórmula de probabilidad total

Ejercicio 5

Usando la tabla del ejercicio 2:

a. Calcular la probabilidad de que el empleado haya estado más de 3 años en la empresa sabiendo que pertenece al área contable
b. Si el empleado ha estado al menos 2 años en la empresa, hallar la probabilidad de que pertenezca al área de instalaciones

Ejercicio 6

Se arroja un dado equilibrado dos veces. Hallar la probabilidad de haber observado un 4 sabiendo que la suma de los resultados fue par

Ejercicio 7

En una urna hay una bola verde y dos bolas rojas. En cada paso se extrae una bola al azar y se la repone junto con otra del mismo color

1. Calcular la probabilidad de que al finalizar el segundo paso la urna contenga dos bolas verdes y tres rojas.

Teorema de Bayes e independencia

Teorema de Bayes: Sean $B_1, \dots B_n$ una partición de Ω, y A un evento con probabilidad positiva:

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

Def: Diremos que dos eventos A y B son **independientes** si y sólo sí vale que

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Ejercicio 7 (cont)

En una urna hay una bola verde y dos bolas rojas. En cada paso se extrae una bola al azar y se la repone junto con otra del mismo color

1. Si al finalizar el segundo paso la urna contiene dos bolas verdes y tres rojas, ¿cuál es la probabilidad de que en el primer paso se haya extra ído una bola roja?

Bonus: algunas ideas de conteo

El chiste ahora es aprender a contar de forma sencilla, para ello nos podemos usar técnicas de conteo:

- **Regla del producto:** Si tengo m elementos en el conj. A y n elementos en el conjunto B, entonces puedo formar mxn pares (a,b).

  Ejemplo: Se arrojan 3 dados equilibrados. De cuántas formas puedo observar 3 números pares?

- **Permutaciones:** # de formas en que puedo ordenar n elementos.

  $$n! = n(n-1)(n-2)\ldots 1$$

  Ejemplo: ¿De cuántas formas puedo ordenar 5 libros en un estante?

Bonus: algunas ideas de conteo

- **Variaciones:** : # de formas combinaciones ordenadas que puedo hacer con n elementos tomados de a k.

  nPk := $(n)_k = n(n-1)\ldots(n-k+1) = \frac{n(n-1)\ldots 1}{(n-k)(n-k-1)\ldots 1} = \frac{n!}{(n-1)!}$

  Ejemplo: Cuántas formas tengo de colocar 3 libros en un estante de 5 que tengo.

- **Combinaciones:** # de combinaciones distintas (sin importar el orden) que puedo hacer con n elementos tomados de a k.

  nCk:= $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\ldots 1}{[(k(k-1)\ldots 1][(n-k)(n-k-1)\ldots 1]}$

  Ejemplo: (Control de calidad) ¿Cuántas muestras diferentes de 10 piezas puedo elegir de un lote de 100?

Ejercicio 8

1. Se tiene un naipe español de 40 cartas. Se extraen al azar sin reposición tres cartas. Hallar
    a. la probabilidad de que las tres sean de oro.
    b. la probabilidad de que las tres sean del mismo palo.
    c. la probabilidad de que las tres sean iguales.
    d. la probabilidad de que las tres sean de palos diferentes.
2. Repetir para 3 extracciones con reposición

Curso de Nivelación:
Probabilidad y estadística

Clase 2

Variables aleatorias

Variables aleatorias

Una **variable aleatoria** (v.a.) es una función que mapea cada elemento ω del espacio muestral Ω a los números reales a un número real $X(\omega)$

Def: Dado un espacio muestral $(\Omega, \mathcal{A}, \mathbb{P})$ y una función $X : \Omega \to \mathbb{R}$ diremos que X es una variable aleatoria si $X^{-1}(B) \in \mathcal{A}$

Def: Dado un $(\Omega, \mathcal{A}, \mathbb{P})$ y X una v.a., se define su **función de distribución** como

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

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) \; \forall a \leq b$$

Propiedades de la función de distribución

Sea $F_X(x)$ una función de distribución asociada a la v.a. X.

Propiedades:

- $F_X(x) \in [0,1] \; \forall \, x \in \mathbb{R}$
- $F_X(x)$ es monótona no decreciente
- $F_X(x)$ es continua a derecha
- $\lim_{x \to -\infty} F_X(x) = 0$ y $\lim_{x \to \infty} F_X(x) = 1$

Variables aleatorias discretas

Si X es una variable aleatoria **discreta**, entonces su función de distribución $F_X(x)$ es monótona no decreciente de a saltos (es escalonada). Llamamos **átomos** (A) al conjunto de puntos donde la función de distribución tiene saltos

Además tiene asociada una **función (de masa) de probabilidad**

$$p_X(x) = \mathbb{P}(X = x)$$

Propiedades:

- $p_X(x) > 0$
- $\sum_{x \in A} p_X(x) = 1$
- $\mathbb{P}(X \in B) = \sum_{x \in B} p_X(x)$

Algunas distribuciones conocidas

- Bernoulli: diremos que $X \sim Ber(p)$ si X toma valores 1 (*éxito*) o 0 (*fracaso*) y

$$p_X(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$$

- Binomial: cuenta la cantidad de éxitos en n ensayos independientes. Diremos que $X \sim Bin(n,p)$ y

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \ldots, n$$

- Geométrica: cuenta la cantidad de ensayos hasta el primer éxito. Diremos que $X \sim \mathcal{G}(p)$ y

$$p_X(x) = (1-p)^{x-1} p, x = 1, 2, \ldots$$

Variables aleatorias continuas

Si X es una variable aleatoria **continua**, entonces su función de distribución $F_X(x)$ es monótona estrictamente creciente (no presenta saltos).

Además tiene asociada una **función de densidad** $f_X(x)$

Propiedades:

- $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$
- $\int_{\mathbb{R}} f_X(x)) = 1$
- $\mathbb{P}(X \in A) = \int_{x \in A} f_X(x)dx$

Algunas variables importantes

- Uniforme: todos los puntos son equiprobables. $X \sim \mathcal{U}(a,b)$

$$f_X(x) = \frac{1}{b-a}\mathbf{I}\{a < x < b\}$$

- Exponencial: sirve para modelar tiempos hasta eventos que no tienen memoria. Por ejemplo fallas casuales. $X \sim \mathcal{E}(\lambda)$

$$f_X(x) = \lambda e^{-\lambda x}\mathbf{I}\{x > 0\}$$

- Normal (gaussiana). $X \sim \mathcal{N}(\mu, \sigma^2)$

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

Propiedades: $X \sim \mathcal{N}(\mu_X, \sigma_X^2),\ Y \sim \mathcal{N}\mu_Y, \sigma_Y^2) \rightarrow X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

$X \sim \mathcal{N}(\mu, \sigma^2) \rightarrow \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ (estandarización)

Distribución conjunta y marginales

Si tenemos dos variables X e Y se define su función de distribución conjunta como

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y)$$

En este caso, vale la regla del rectángulo

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c)$$

Caso continuo: $f_{X,Y}(x,y)$ es la función de densidad conjunta y se definen las funciones de densidad marginales como $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy$ y

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

Caso discreto: $p_{X,Y}(x,y)$ es función de probabilidad conjunta y se definen las funciones de probabilidad marginal como $p_X(x) = \sum_y p_{X,Y}(x,y)$

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

Independencia de v.a.

Diremos que dos v.a. X e Y son independientes si vale que

$$F_{X.Y}(x, y) = F_X(x)F_Y(y) \quad \forall\, x, y \in \mathbb{R}$$

Caso discreto:

$$p_{X.Y}(x, y) = p_X(x)p_Y(y) \quad \forall\, x, y \in \mathbb{R}$$

Caso continuo:

$$f_{X.Y}(x, y) = f_X(x)f_Y(y) \quad \forall\, x, y \in \mathbb{R}$$

Distribuciones condicionales

Al igual que como definimos probabilidades condicionales para eventos podemos definir distribuciones condicionales para variables aleatorias.

Caso discreto:

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

Caso continuo:

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

Qué ocurre si X e Y son independientes?

Momentos

Esperanza

La esperanza o media es el valor esperado de la variable aleatoria. Es un promedio ponderado de los valores que puede tomar la variable.

Caso discreto: $\mathbb{E}[X] = \sum_{x\in A} x\, p_X(x)$

Caso continuo: $\mathbb{E}[X] = \int_{\mathbb{R}} x\, f_X(x)\, dx$

Propiedades:

- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ (linealididad)
- Si X,Y son independientes $\mathbb{E}[X\,Y] = \mathbb{E}[X]\mathbb{E}[Y]$
- Si $Y = g(X)$ vale que $\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum_{x\in A} g(x) p_X(x)$ (si X es v.a.c)

  $\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x)\, df$ (si X es continua)

Varianza

La varianza de una v.a. X mide su dispersión respecto de la media, y se define como

$$var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Propiedades:

- $var(aX + b) = a^2 var(X)$
- Si X, Y son **independientes**, $var(X + Y) = var(X) + var(Y)$

Se define también el desvío estándar como

$$\sigma_X = \sqrt{var(X)}$$

Vectores aleatorios:

Para el caso del vector aleatorio *(X,Y)* definimos la esperanza como

$$\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y) \text{ (caso discreto)}$$

$$\mathbb{E}[g(X,Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y) f_{X,Y}(x,y) dx dy \text{ (caso continuo)}$$

Definimos la covarianza entre dos v.a. X, Y como:

$$cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

La covarianza mide la relación **lineal** entre las variables.

El coeficiente de correlación entre ambas variables se define como

$$\rho = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

Covarianza

Propiedades de la covarianza:

- $cov(aX + bY, cZ) = ac\, cov(X, Z) + bc\, cov(Y, Z)$
- $cov(X, X) = var(X)$
- $var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$
- Si $cov(X, Y) = 0$ diremos que las variables están descorrelacionadas
- Si X, Y son independientes, entonces $cov(X, Y) = 0$ (Cuidado! No vale para el otro lado)

Momentos

En general, definimos el ***n*-ésimo momento** de la v.a. X como $\mathbb{E}[X^n]$

Se define la **función generadora de momentos** como

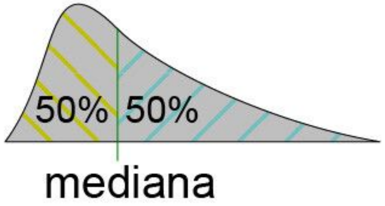
$$M_X(t) = \mathbb{E}[e^{tX}]$$

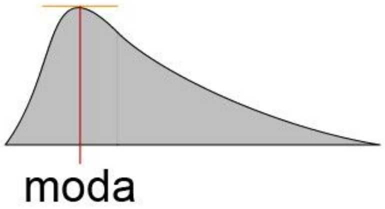
Si que $\mathbb{E}[e^{tX}]$ existe (es finita) para algún intervalo que contiene al cero, se puede calcular el n-ésimo momento de X como la derivada n-ésima de $M_X(t)$ evaluada en 0:

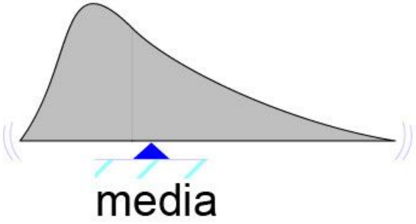
$$\mathbb{E}[X^n] = \frac{d}{dt} M_X(t)|_{t=0}$$

Algunos estadísticos importantes

- Media
- Mediana
- Moda







- Cuantiles

$$q_p : \mathbb{P}(X < q_p) = p$$

$$q_{0.25} : \mathbb{P}(X < q_{0.25}) = 0.25$$ Cuantil 0.25 o 1er cuartil

$$q_{0.75} : \mathbb{P}(X < q_{0.75}) = 0.75$$ Cuantil 0.75 o 3er cuartil

Bibliografía

Bibliografía sugerida

- “Mathematical Statistics with Applications”, Dennis D. Wackerly, William Mendenhall, Richard L. Scheaffer
- “All of Statistics: A Concise Course in Statistical Inference”, Larry Wasserman
- “Mathematics for Machine Learning”, Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong