Introducción a la Inferencia Estadística

Precurso de

XXV Escuela de Otoño y XIX Encuentro Nacional de Biología Matemática

CIMAT

L. Leticia Ramírez Ramírez

Contenido

Precursos en línea

	29 de julio	30 de julio	31 de julio	1 de agosto	2 de agosto
9:00-10:30	Biol Molecular ¿Que es el DNA? Dogma central	Estadística	Biol Molecular: Secuenciación	Estadística	Biol Molecular: Evolucion
10:30-11:00	Receso				
11:00-12:30	Estadística	Estadística	Estadística	Programación: librerías y gráficos	Programación: ciclos y funciones
12:30-14:30	Comida				
14:30-15:30	Programación: R Introducción a variables y R studio	Programacion: Python Introducción a variables y Jupyter notebook	Ecología	Ecología	Ecología
15:30-16:30	Bioinformática	TDA	Cálculo	Ecuaciones	Ecuaciones

Contenido

- Sesión 1: Variables aleatorias, FP, FDP, F(x) y momentos
- Sesión 2: Distribuciones Discretas y Continuas
- Sesión 3: Inferencia Estadística
- Sesión 4: Probabilidad y Esperanza Condicional
- Sesión 5: Inferencia Bayesiana

Variables Aleatorias

Una variable aleatoria (VA) es una función que adjudica eventos posibles a números reales.

Discreta. Toma un número contable de valores (finito o infinito).

- El resultado de un tiro de un dado
- El número de plantas con daños visibles por una plaga
- El número de animales contados en una zona

Continua. Toma valores en intervalos de los reales

- El peso de una persona seleccionada al azar
- La temperatura máxima de mañana en su ciudad

Mixtas. Toma valores en regiones y valores específicos de $\mathbb{R}$.

VAs Discretas: Distribución Uniforme

Cuando se tiene un número **finito** de resultados de un experimento, $x_1, x_2, \ldots, x_n$, cada uno de ellos **igualmente probable**, se dice que se tiene una variable aleatoria que puede ser modelada por una distribución uniforme discreta.

Notación:

$$X \sim U(\, x_1, x_2, \ldots, x_n \,).$$

Se lee: X se distribuye como una variable aleatoria uniforme.

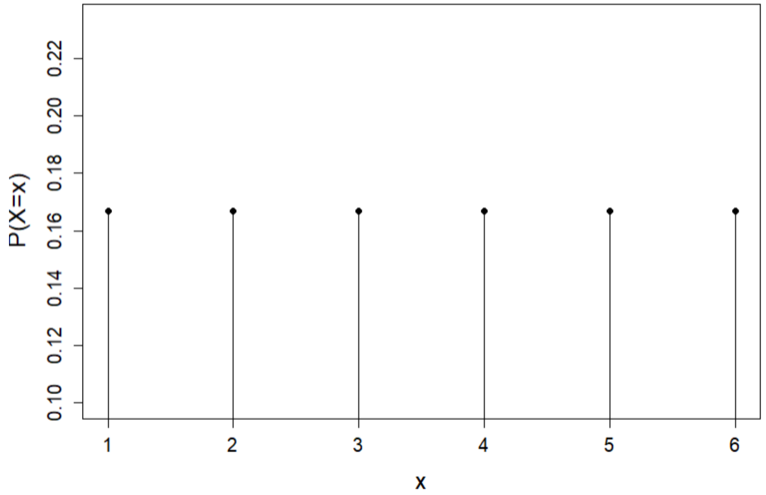
Este tipo de modelo aparece típicamente en la selección de **muestras al azar**.

El único parámetro de la distribución es n, el número posible de resultados.

VAs Discretas: Distribución Uniforme

Como la suma de todas las probabilidades debe ser uno y todos los valores son equiprobables, entonces la función de masa está dada por

$$P(X = x) = \begin{cases} 1/n & \text{para } x = x_1, x_2, \ldots, x_n \\ 0 & \text{en otro caso} \end{cases}$$



A $P(X = x)$ también se denota como $p(x)$ **o** $f(x)$.

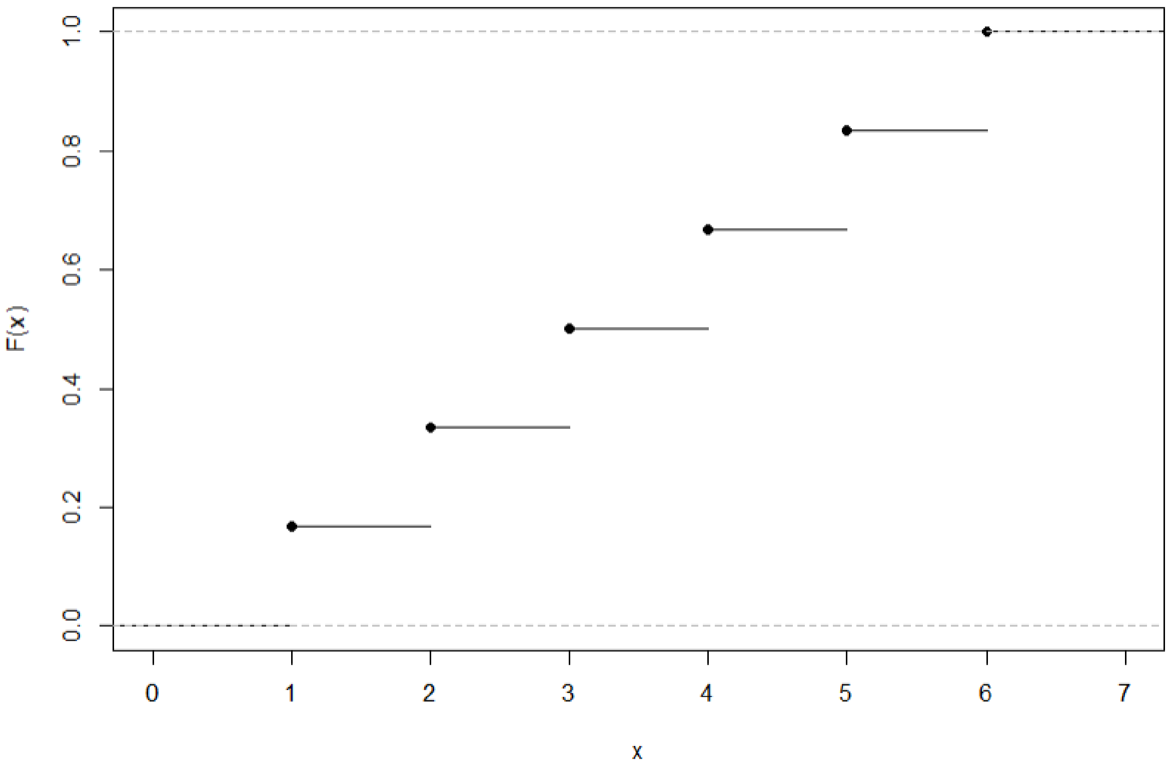
Bajo esta definición, claramente $p(x) \geq 0$ para toda x, y $\sum_{j=1}^{n} p(j) = 1$.

El caso más común de esta distribución es cuando la variable toma los valores enteros $\{1, 2, \ldots, n\}$.

Definiciones y propiedades

Def. $P(X = x) = p(x) = f(x)$ de denomina Función de Densidad de Probabilidad (FDP).

Def. La Función de Distribución se define como $F(x) = P(X \leq x) = \sum_{j \leq x} P(X = j)$



Propiedades

- $F(x)$ es continua por la derecha
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$

$P(X \in (a, b]) = F(b) - F(a)$ para $a, b \in \mathbb{R}$

VAs Discretas: Distribución Bernoulli

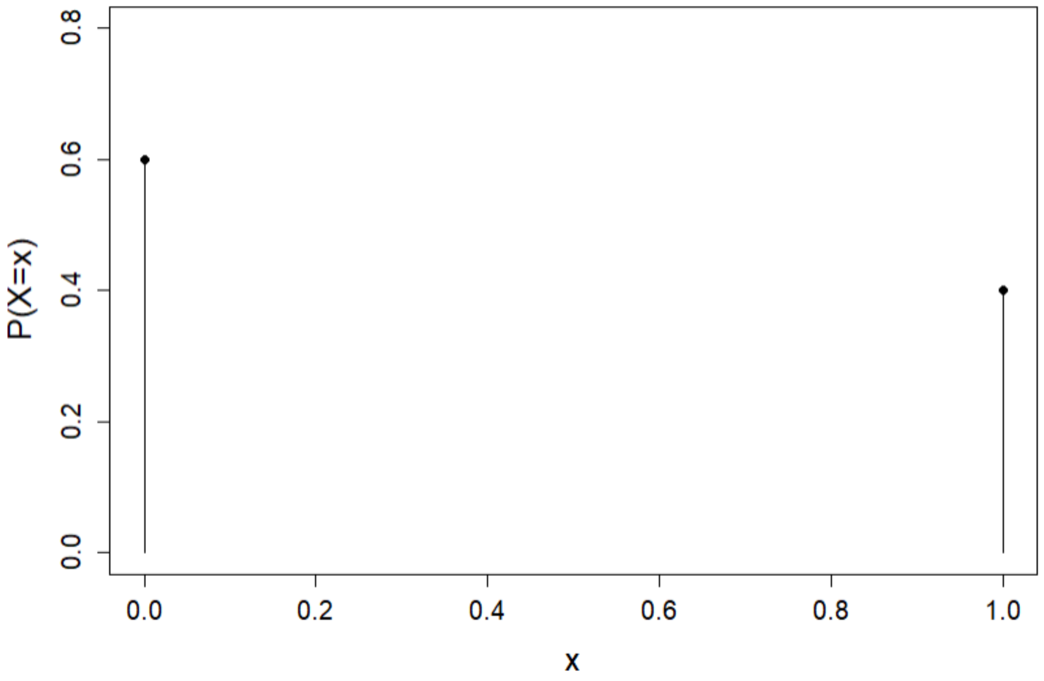
- Definamos un experimento en el cual hay únicamente dos posibles resultados: a uno de ellos le llamamos "éxito", al otro "fracaso".
- Generalmente le asignamos X un valor de 1 cuando ocurre un "éxito" y 0 cuando ocurre un "fracaso".
- Si la probabilidad de observar un éxito es p entonces denotamos $X \sim \text{Bernoulli}(p)$ y su FDP es

$$p(x) = p^x(1-p)^{1-x}, \quad \text{para} \quad x = 0, 1.$$

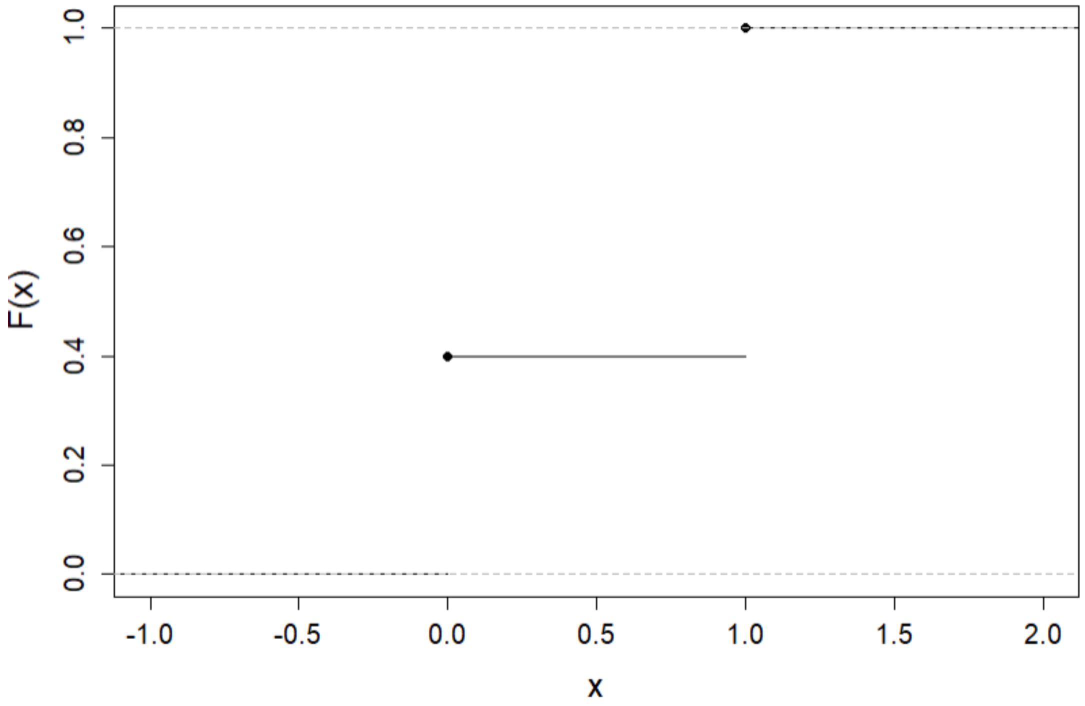
VAs Discretas: Distribución Bernoulli

$$X \sim \mathrm{Ber}(0.6)$$

p(x)



F(x)



Distribuciones conjuntas

Tiramos un dado "justo" dos veces. El espacio de resultados tiene 36 elementos:

$$(1,1),(1,2),(1,3),\ldots,(4,6),(5,6),(6,6).$$

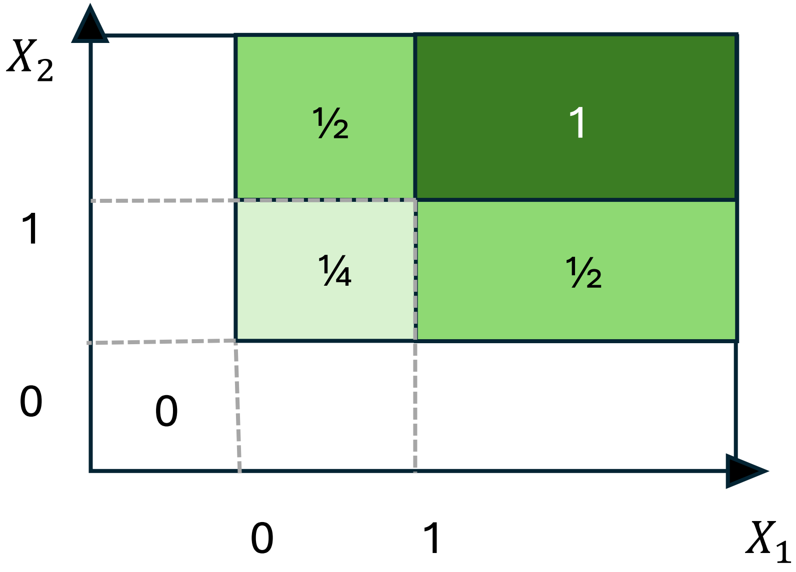
Cada uno con la misma probabilidad. Entonces la distribución conjunta es

$$P(X_1 = x_1 \text{ y } X_2 = x_2) = P(X_1 = x_1, X_2 = x_2) = \begin{cases} 1/36 & \text{si } x_1, x_2 = 1,2,3,4,5,6 \\ 0 & \text{en otro caso} \end{cases}$$

Distribuciones conjuntas

La Función de Distribución conjunta es

$$F_{X_1,X_2}(x_1,x_2) = P(X_1 \leq x_1, X_2 \leq x_2) = \begin{cases} 0 & \text{si } x_1 < 0 \text{ o } x_2 < 0 \\ 1/4 & \text{si } x_1, x_2 \in [0,1) \\ 1/2 & \text{si } x_1 \in [0,1), x_2 > 1 \text{ o } x_1 > 1, x_2 \in [0,1) \\ 1 & \text{eoc} \end{cases}$$



VAs Discretas: Distribución Binomial

Consideramos que ahora tenemos repeticiones **independientes** de experimentos Bernoulli(p) y definimos a X como $X =$ Núm. de **"éxitos"** en n observaciones.

Entonces X tiene distribución Binomial con parámetros n **y** p. Su FDP es

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

donde

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

Independencia Estocástica

Def. Decimos que dos VAs son independientes ssi

$$P(X \in A \text{ y } Y \in B) = P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

para todo $A, B \in \mathcal{B}$.

Def. Decimos que las variables aleatorias $X_1, X_2, \ldots, X_n$ son independientes ssi para cualquier subconjunto de las n variables tenemos

$$P(X_{i_1} \in A_1 \cap X_{i_2} \in A_2 \cap \cdots \cap X_{i_m} \in A_m) = P(X_{i_1} \in A_1)P(X_{i_2} \in A_2) \cdots P(X_{i_m} \in A_m)$$

para todo $A_1, \ldots, A_m$ en $\mathcal{B}$.

Independencia Estocástica

Para checar idependencia, es más sencillo usar

Teo. Las variables aleatorias $X_1, X_2, \ldots, X_n$ son independientes ssi

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \times \cdots \times F_{X_n}(x_n) = \prod_{j=1}^{n} F_{X_j}(x_j)$$

para $(x_1, \ldots, x_n)$ en $\mathbb{R}^n$.

Independencia: Ejemplo 1

Dos lanzamientos de una moneda donde la probabilidad de sol o águila es $1/2$. Los resultados posibles son (s,s), (s,a),(a,s),(a,a). Que por ser equiprobables, cada uno tiene probabilidad de $1/4$. Ahora se puede verificar que

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$$

Por ejemplo,

$$F_{X_1,X_2}(x_1, x_2) = 0 = 0 \times F_{X_2}(x_2) = F_{X_1}(x_1)F_{X_2}(x_2) \text{ si } x_1 < 0$$
$$F_{X_1,X_2}(x_1, x_2) = 1/2 = 1/2 \times 1 = F_{X_1}(x_1)F_{X_2}(x_2) \text{ si } x_1 \in [0, 1), x_2 > 1$$
$$F_{X_1,X_2}(x_1, x_2) = 1 = 1 \times 1 = F_{X_1}(x_1)F_{X_2}(x_2) \text{ si } x_1 > 1, x_2 > 1$$

Independencia: Ejemplo 1

Por ser X_1 y X_2 independientes, entonces

$$P(X_1 = x_1, X_2 = x_2) = 1/4 = 1/2 \times 1/2 = P(X_1 = x_1)P(X_2 = x_2)$$

para $x_1, x_2 \in \{0, 1\}$.

Independencia: Ejemplo 2

El concepto de independencia en la práctica se puede argumentar cuando los resultados de una variable aleatoria no influyen en los resultados de la otra. Así en el tiro consecutivos de una moneda donde la probabilidad de éxito es p podemos calcular las probabilidades conjuntas como

$$\begin{aligned}
P(X_1 = 1, X_2 = 1, x_3 = 1) &= p \times p \times p = p^3 \\
P(X_1 = 1, X_2 = 1, x_3 = 0) &= p \times p \times (1-p) = p^2(1-p) \\
P(X_1 = 1, X_2 = 0, x_3 = 1) &= p \times (1-p) \times p = p^2(1-p) \\
P(X_1 = 0, X_2 = 1, x_3 = 1) &= (1-p) \times p \times p = p^2(1-p) \\
&\vdots
\end{aligned}$$

VAs Discretas: Distribución Binomial

Considerando el número de exitos en tres ensayos Bernoulli(p) vemos que

$$\mathrm{P}(X_1 = 0, X_2 = 0, X_3 = 0) = \mathrm{P}(X_1 = 0)\mathrm{P}(X_2 = 0)\mathrm{P}(X_3 = 0) = qqq = q^3 = \binom{3}{0} q^3$$

A manera de ejercicio, puedes verificar que:

$$\begin{aligned}
\mathrm{P}(X = 1) &= \mathrm{P}(\{X_1 = 1, X_2 = 0, X_3 = 0\} \text{ ó } \{X_1 = 0, X_2 = 1, X_3 = 0\} \text{ ó } \{X_1 = 0, X_2 = 0, X_3 = 1\}) \\
&= 3pqq = 3pq^2 = \binom{3}{1} pq^2, \qquad \textbf{1} \quad \textbf{“éxito”}; \\
\mathrm{P}(X = 2) &= \mathrm{P}(\{X_1 = 1, X_2 = 1, X_3 = 0\} \text{ ó } \{X_1 = 1, X_2 = 0, X_3 = 1\} \text{ ó } \{X_1 = 0, X_2 = 1, X_3 = 1\}) \\
&= 3ppq = 3p^2q = \binom{3}{2} p^2q, \qquad \textbf{2} \quad \textbf{“éxitos”}; \\
\mathrm{P}(X = 3) &= \mathrm{P}(\{X_1 = 1, X_2 = 1, X_3 = 1\}) \\
&= ppp = 3p^3 = \binom{3}{3} p^3, \qquad \textbf{3} \quad \textbf{“éxitos”}.
\end{aligned}$$

VAs Discretas: Distribución Binomial

El ejemplo y argumento anterior se puede formalizar para demostrar que si $X_1, X_2, \ldots, X_n$ son VAs independientes y con la misma distribución Bernoulli(p) entonces la VA definida como

$$Y = \sum_{j=1}^{n} X_j \sim \text{Binom}(n, p)$$

con FDP

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

Se puede verificar que la suma, sobre todos los valores de Y, es 1, haciendo uso del **Teorema del Binomio de Newton**:

$$(a+b)^n = \sum_{y=0}^{n} \binom{n}{y} a^{n-y} b^y.$$

VAs Discretas: Distribución Binomial