Introducción a la Inferencia Estadística

Precurso de

XXV Escuela de Otoño y XIX Encuentro Nacional de Biología Matemática

L. Leticia Ramírez Ramírez

CIMAT

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.

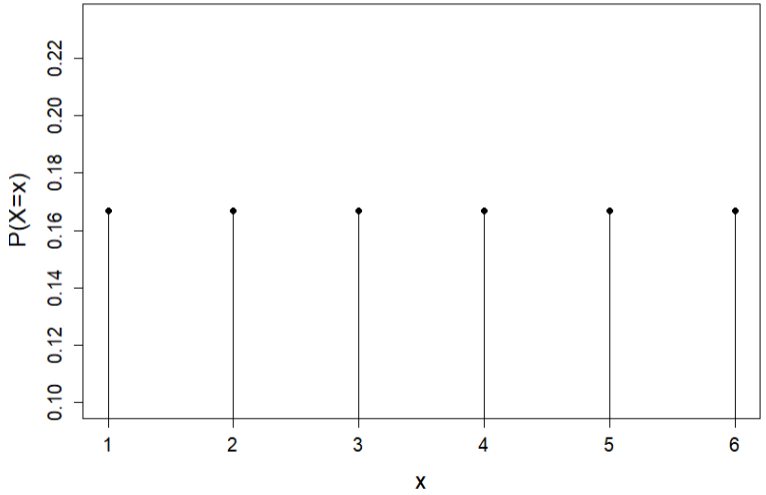
Este tipo de modelo aparece típicamente en la selección de **muestras al azar**.

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** $f(x)$.

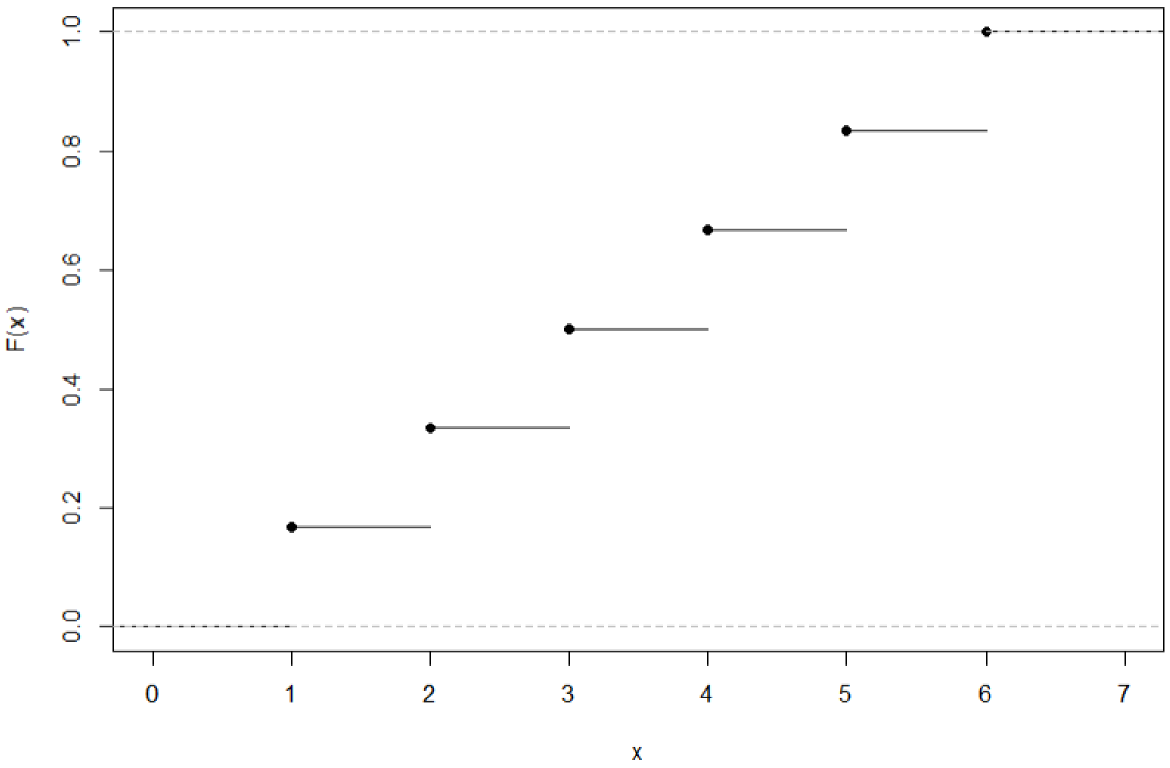
Bajo esta definición, claramente $p(x) \geq 0$ para toda x, y $\sum_{j=1}^{n} p(j) = 1$.

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

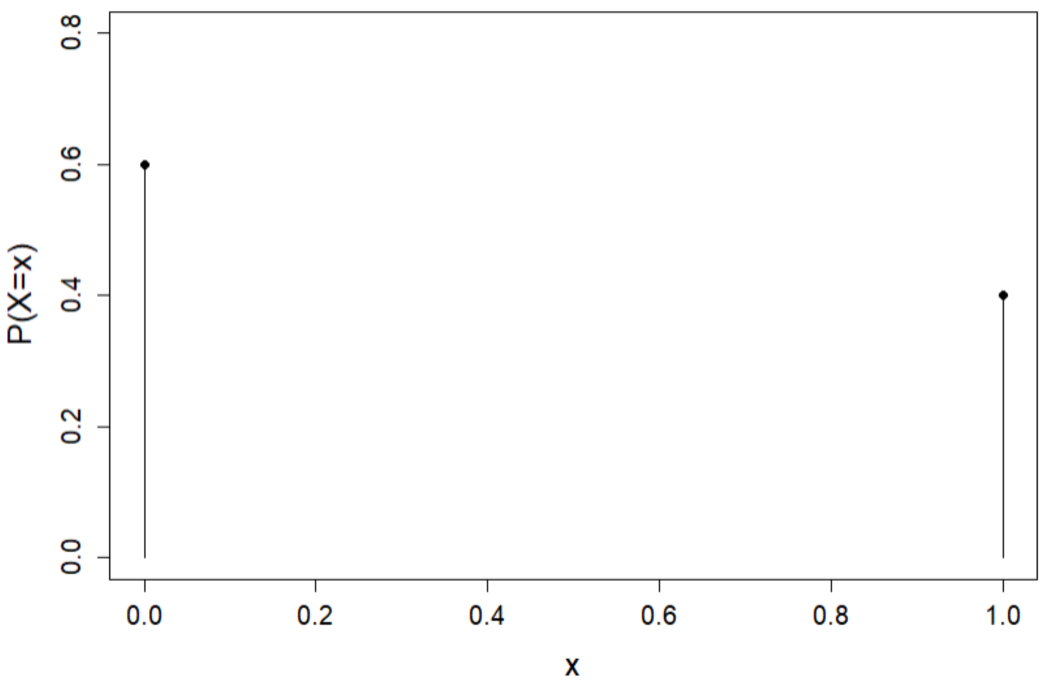
- Definamos un experimento en el cual hay únicamente dos posibles resultados: a uno de ellos le llamamos “éxito”, al otro “fracaso”.
- Generalmente le asignamos X un valor de 1 cuando ocurre un “éxito” y 0 cuando ocurre un “fracaso”.
- Si la probabilidad de observar un éxito es p entonces denotamos $X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

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

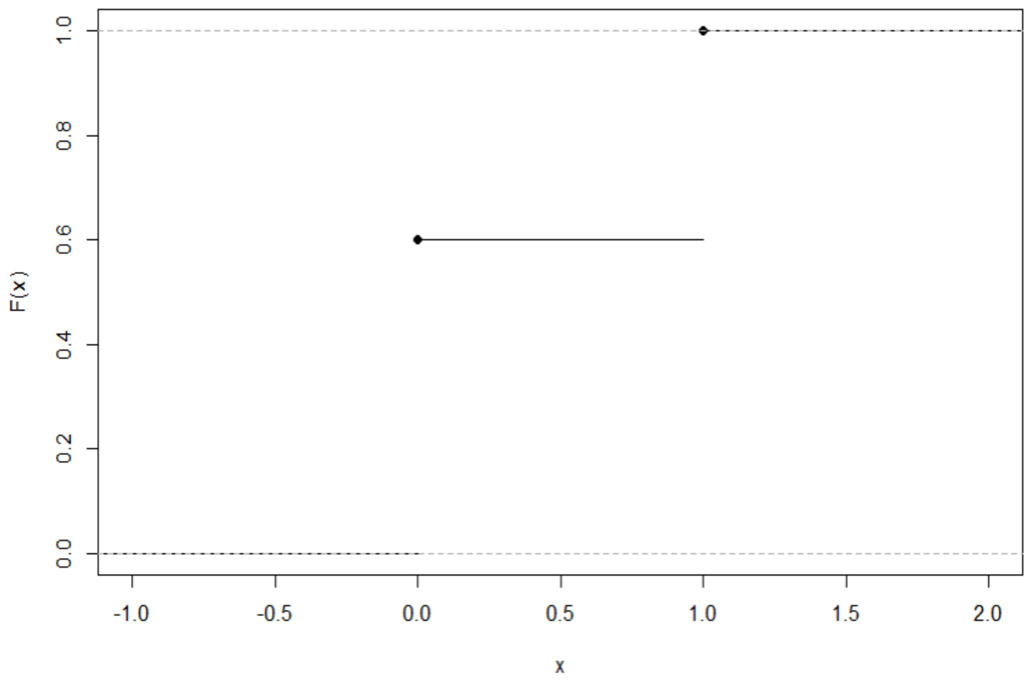
VAs Discretas: Distribución Bernoulli

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

p(x)



F(x)



Distribuciones conjuntas

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

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

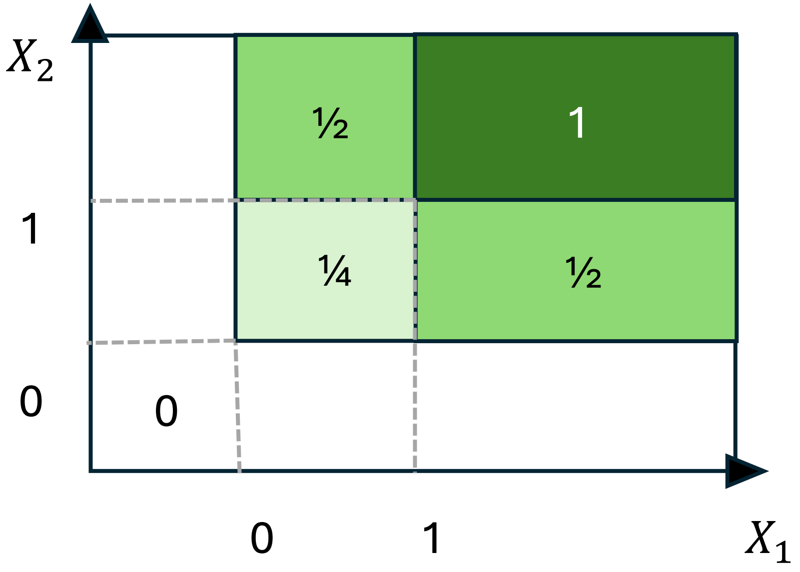
Cada uno con la misma probabilidad. Entonces la distribución conjunta es

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

Distribuciones conjuntas

En el caso de dos tiros de monedas justas, X_1 y X_2 (con distribución Bernoulli(1/2)) 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

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

A manera de ejercicio, puedes verificar que:

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

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

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

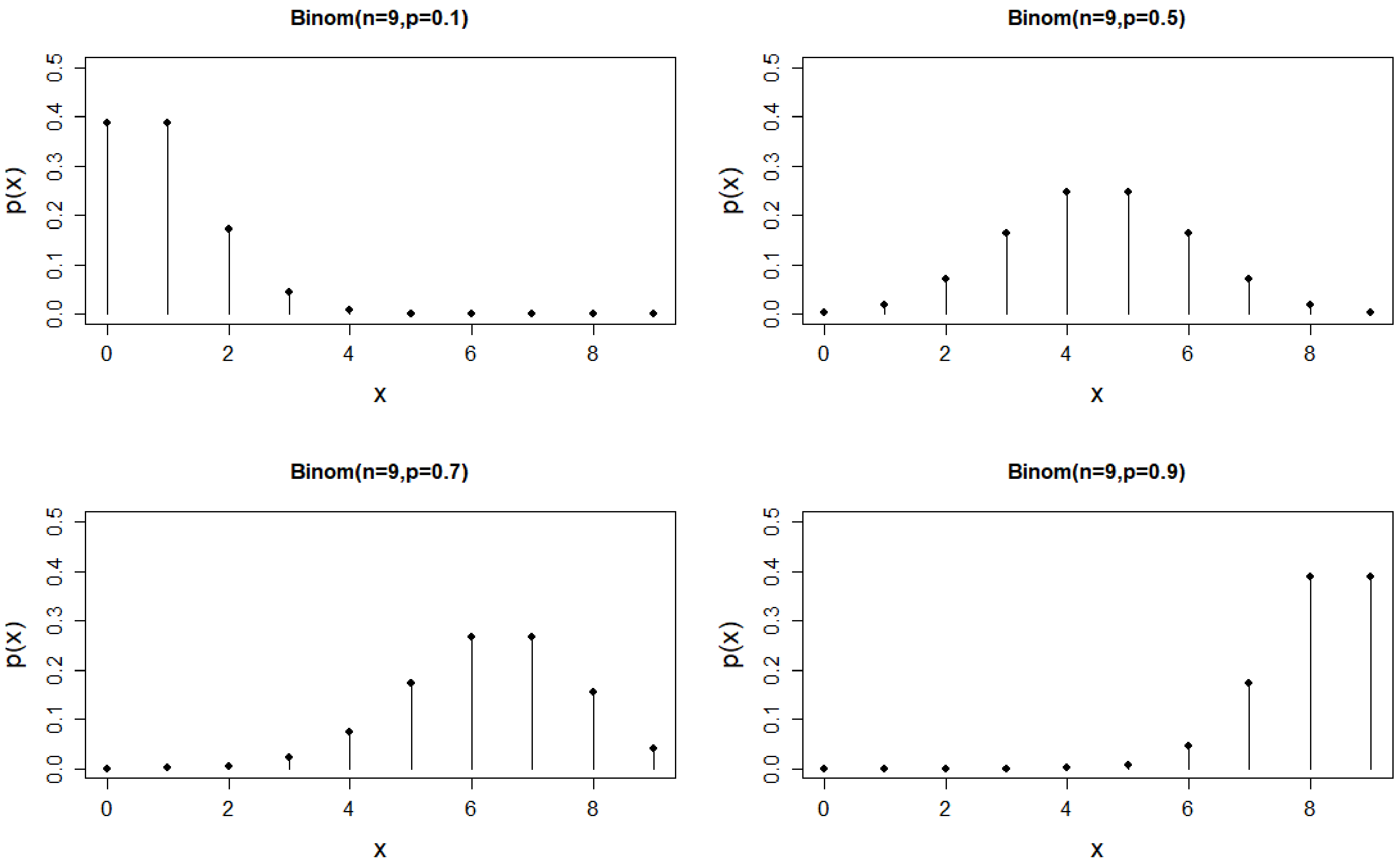
con FDP

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

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



Momentos

Los momentos de (una distribución de probabilidad de) una variable aleatoria X son medidas que describen las características de (la distribución de) la variable aleatoria X.

Distinguimos principalmente entre los momentos (alrededor del cero) o alrededor del primer momento (éstos se denominan momentos centrales).

El estudio de los momentos tiene mucha utilidad cuando no se conoce la distribución de probabilidad de una variable aleatoria o se quiere resumir algunas de sus características.

Momentos

Def. El r**-ésimo momento** (alrededor del cero) se define como:

$$E\left[X^r\right] = \sum_x x^r P(x).$$

Def. El **primer momento** es la **media**, o el valor esperado, es decir

$$\mu = E[X].$$

La media indica dónde la FDP tienden a agruparse, por eso es una medida de tendencia central.

Momento Centrales

Def. El ***r*-ésimo momento central** es el r-ésimo momento alrededor de la media μ y se define como:

$$E\left[(X-\mu)^r\right] = \sum_x (x-\mu)^r p(x).$$

El primer momento central es cero:

$$\mu_1 = E\left[(X-\mu)^1\right] = E[X-\mu] = E[X] - \mu = 0.$$

Def. El **segundo momento central** se denomina como **la varianza**:

$$E\left[(X-\mu)^2\right] = \mathrm{Var}(X) = E\left[X^2\right] - E^2[X] = E\left[X^2\right] - \mu^2.$$

Propiedades

1. Las FDP marginales se pueden obtener de las FDP conjuntas. Ej.

$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2), \qquad P(X_2 = x_2) = \sum_{x_1} P(X_1 = x_1, X_2 = x_2)$$

2. La esperanza de una función de una VA X, $g(X)$ (donde g es $\mathcal{B}$ medible) es

$$E(g(X)) = \sum_{x} g(x)p(x)$$

3. La Esperanza es lineal: Si X y Y son VAs, a, b escalares entonces $E(aX + bY) = aE(X) + bE(Y)$.

$$\begin{aligned} E(aX + bY) &= \sum_{x}\sum_{y}(ax + by)p_{X,Y}(x, y) = \sum_{x}\sum_{y}(ax)p_{X,Y}(x, y) + \sum_{x}\sum_{y}(by)p_{X,Y}(x, y) \\ &= a\sum_{x} x \sum_{y} p_{X,Y}(x, y) + b\sum_{y} y \sum_{x} p_{X,Y}(x, y) = a\sum_{x} xp_X(x) + b\sum_{y} yp_Y(y) = aE(X) + bE(Y) \end{aligned}$$

$\blacksquare$

Propiedades

4. Si X y Y son independientes, entonces $E(XY) = E(X)E(Y)$

$$E(XY) = \sum_y \sum_x p_{X,Y}(x,y) = \sum_x \sum_y p_X(x) p_Y(y) = \sum_x x p_X(x) \sum_y y p_Y(y) = E(X)E(Y) \qquad \blacksquare$$

5. Si X y Y son independientes, entonces $g(X)$ y $h(Y)$, con g y h son funcientes $\mathcal{B}$-medibles, son también independientes.

6. Si X y Y son independientes, entonces $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

$$\begin{aligned} \mathrm{V}(X+Y) &= E((X+Y)^2) - [E(X+Y)]^2 = E(X^2 + 2XY + Y^2) - [\mu_x + \mu_Y]^2 \\ &= E(X^2) + \cancel{2\mu_X \mu_Y} + E(Y^2) - \mu_X^2 - \cancel{2\mu_X \mu_Y} - \mu_Y^2 = \mathrm{V}(X) + \mathrm{V}(Y) \qquad \blacksquare \end{aligned}$$

7. Si X es una VA y c una constante, entonces $\mathrm{V}(cX) = c^2 \mathrm{V}(X)$.

$$\mathrm{V}(cX) = E((cX)^2) - [E(cX)]^2 = E(c^2 X^2) - [E(c)E(X)]^2 = c^2 E(X^2) - c^2 E(X)^2 = c^2 \mathrm{V}(X) \qquad \blacksquare$$

Media y Varianza: Distribución Uniforme$(\{1, \ldots, n\})$

$$\mathrm{E}(X) = \sum_{x=1}^{n} x\frac{1}{n} = \frac{1+2+\cdots+n}{n} = \frac{n+1}{2},$$

y

$$\begin{aligned}
\mathrm{V}(X) = \; & \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = \sum_{x=1}^{n} x^2\frac{1}{n} - \left(\frac{n+1}{2}\right)^2 \\
= \; & \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{2n^2+3n+1}{6} - \frac{n^2+2n+1}{4} \\
= \; & \frac{2(2n^2+3n+1) - 3(n^2+2n+1)}{12} = \frac{4n^2+6n+2-3n^2-6n-3}{12} \\
= \; & \frac{n^2-1}{12}.
\end{aligned}$$

Media y Varianza: Distribución Bernoulli(p)

$$\mathrm{E}(X) = \sum_{x=0}^{1} x \cdot f(x) = 0 \cdot (1-p) + 1 \cdot p = \mathbf{p}$$

$$\mathrm{E}(X^2) = \sum_{x=0}^{1} x^2 \cdot f(x) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$$

$$\mathrm{V}(X) = \mathrm{E}(X^2) + [\mathrm{E}(X)]^2 = p - p^2 = p(1-p) = \mathbf{pq}$$

Media y Varianza: Distribución Binomial(n, p)

Recordando las propiedades del valor esperado y el valor que toma la esperanza de una v.a. $Bernoulli(p)$, tenemos que el valor esperado de Y es

$$\mathrm{E}(Y) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} p = np$$

y su varianza, por ser las X_i's independientes, es

$$\mathrm{V}(Y) = \mathrm{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{V}(X_i) = \sum_{i=1}^{n} pq = npq.$$

VAs Discretas: Distribución Poisson

Supóngase ahora que estamos interesado en el número de **"éxitos"**, pero ahora preguntándonos por la posibilidad de que ocurran en un cierto **intervalo de tiempo, espacio o volumen**, por ejemplo

- Número de defectos en una plancha de acero de $1\,m^2$,
- Número de bacterias en $1\,cm^3$ de agua potable,
- Número de llamadas que llegan a un conmutador en un minuto.

En estos casos interesa la variable aleatoria

$X =$ # de **"éxitos"** obtenidos en un cierto **intervalo (temporal o espacial)**.

VAs Discretas: Distribución Poisson

Nuestros supuestos son:

- El número promedio de **"éxitos"**, λ (la intensidad), sobre el intervalo dado se mantiene **constante**.

  Se considera que el comportamiento es estable. Incluso, si se estuviera interesado en un múltiplo t (no necesariamente entero) del intervalo original, el valor λ_1 *correspondiente será* $\lambda_1 = \lambda t$.

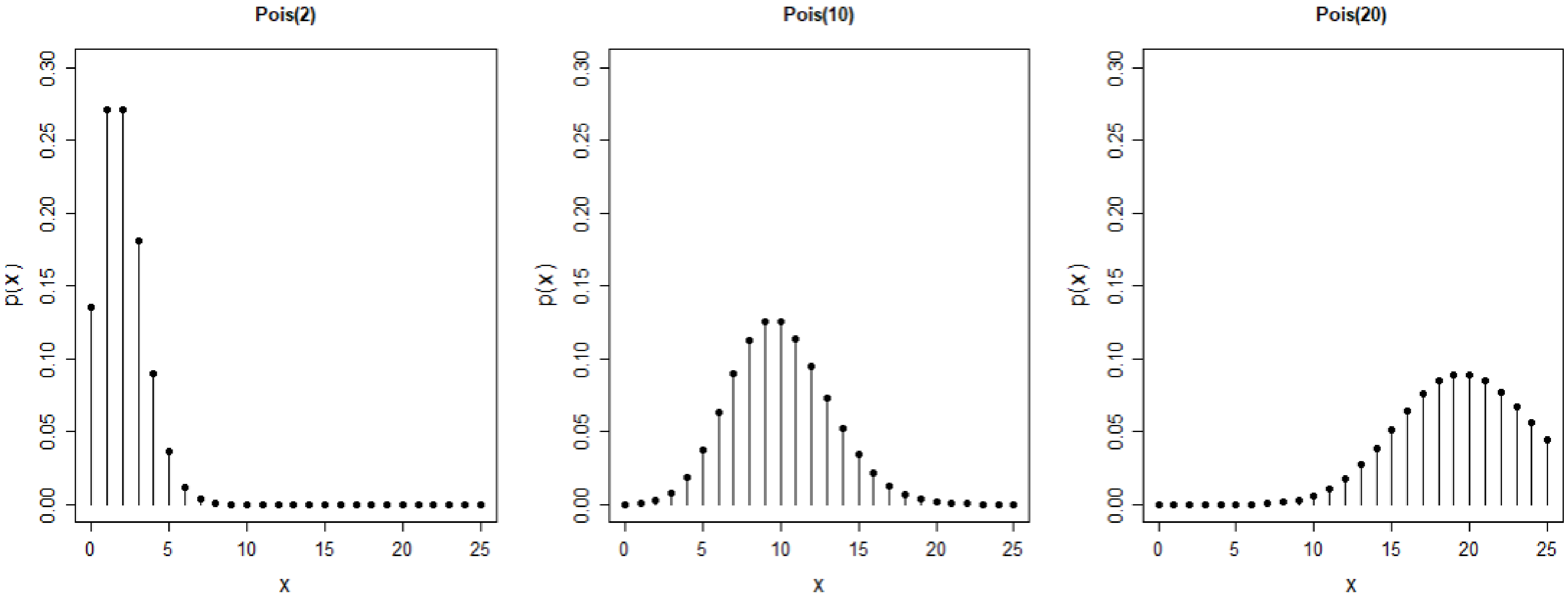
- Los **éxitos aparecen en forma aleatoria en intervalos de la misma magnitud**, los **intervalos no se traslapan** y son **independientes entre ellos**.

- La probabilidad de tener una observaciones en intervalos de longitud Δt es $\lambda \Delta t$, la probabilidad de no tener ninguna observación es $(1 - \lambda)\Delta t$ y la probabilidad de tener más de una observación es $o(\Delta t)$.

VAs Discretas: Distribución Poisson

Bajo los supuestos anterires, tenemos que la función de probabilidad de la **Distribución Poisson** que describe el número de observaciones por unidad (de tiempo, espacio o volumen) es

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \text{para } x = 0, 1, 2, \ldots$$



VAs Discretas: Distribución Poisson

$$\boxed{E(X) = \mathrm{V}(X) = \lambda}$$

$$E(X) = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

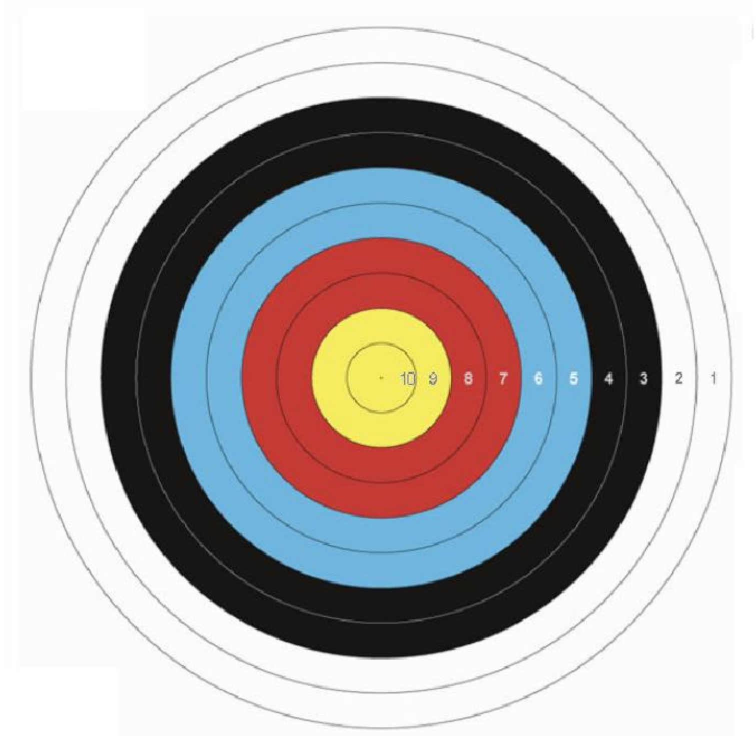
$$E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{x\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{d}{d\lambda} \left[\frac{\lambda^x}{(x-1)!} \right] = \lambda e^{-\lambda} \frac{d}{d\lambda} \left[\sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \right]$$

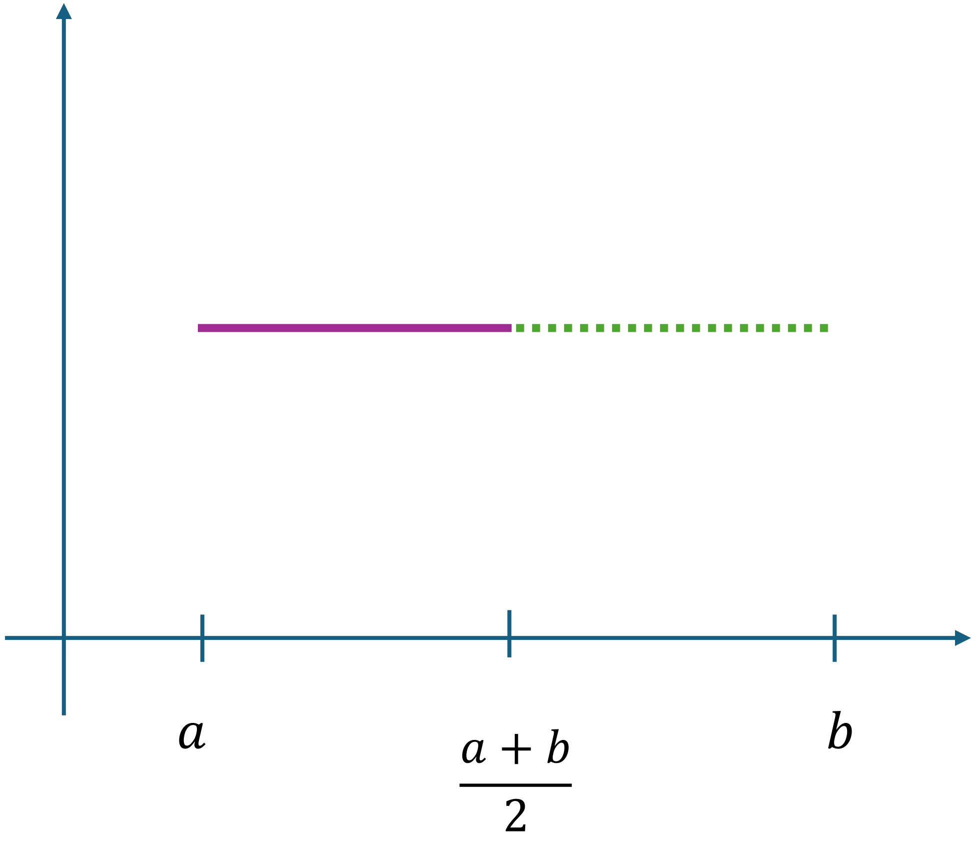
$$= \lambda e^{-\lambda} \frac{d}{d\lambda} \left[\lambda \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \right] = \lambda e^{-\lambda} \frac{d}{d\lambda} \left[\lambda e^{\lambda} \right] = \lambda e^{-\lambda} \left[e^{\lambda} + \lambda e^{\lambda} \right] = \lambda + \lambda^2.$$

Entonces

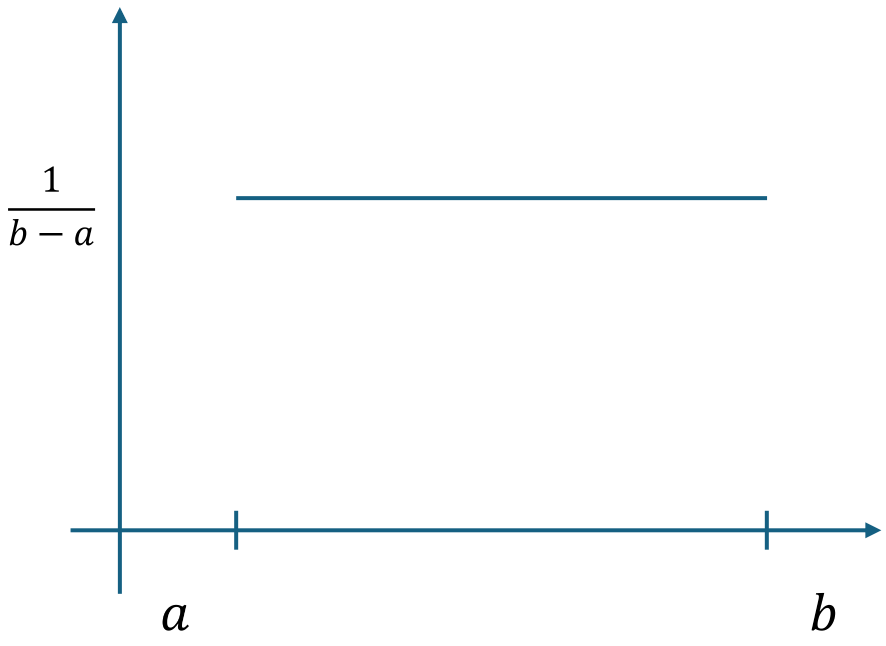
$$\mathrm{V}(X) = \lambda + \lambda^2 - \lambda^2 = \lambda \qquad \blacksquare$$

VAs Continuas





VAs Continuas: Distribución Uniforme



Una v.a. X tiene una **distribución uniforme continua**, denotada con $X \sim \text{Unif}(a, b)$ ssi su función de densidad (FD) está dada por

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b, \\ \\ 0 & \text{otra parte.} \end{cases}$$

En el caso continuo, calculamos las probabilidades de que X otme valor en una region $[c, d]$ como

$$P(X \in [c, d]) = \int_c^d f(x) dx.$$

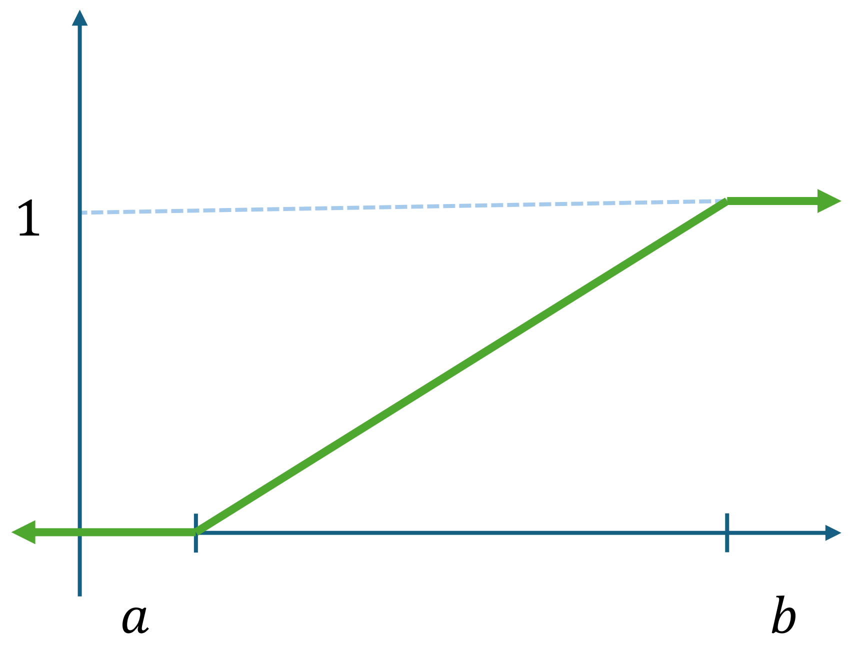
Definiciones en el caso de VAs continuas
Función de Distribución

La Función de Distribución se define como $F(x) = P(X \leq x)$. En el caso continuo ésta corresponde a

$$F(x) = \int_{-\infty}^{x} f(u)du.$$

Cuando $X \sim \text{Unif}(a, b)$ entonces tenemos que

$$F(x) = \begin{cases} 0 & \text{si } x < a, \\ \int_a^x f(u)du = \int_a^x \frac{1}{b-a}du = \frac{x-a}{b-a} & \text{si } x \in (a, b), \\ 1 & \text{si } x \geq b. \end{cases}$$



Definiciones en el caso de VAs continuas
Función de Distribución

Al igual que en el caso discreto, podemos calcular probabilidades a partir de la FD. Por ejemplo, para $c, d \in \mathbb{R}$, $c < d$:

$$P(X \in (c, d]) = F(d) - F(c)$$
$$P(X \in [c, d)) = F(d) - F(c)$$
$$P(X \in [c, d]) = F(d) - F(c)$$
$$P(X \in (c, d)) = F(d) - F(c)$$