Modelos Estadísticos para la Regresión y la Clasificación

Presentación y breve repaso de PyE

17 de agosto de 2023

Matías Carrasco

Docentes - 2do Semestre 2023

- Matías Carrasco (Lunes y Jueves)
- Juliana Faux (Miércoles)

Bibliografía

- Nivel Básico
  - Machine Learning Simplified: A Gentle Introduction to Supervised Learning (2022)
    Andrew Wolf
- Nivel Introductorio (adecuados para este curso)
  - A first course in machine learning (2012)
    Rogers & Girolami
  - An introduction to statistical learning with applications in Python (2023)
    James, Witten, Hastie, Tibshirani & Taylor
- Nivel Avanzado
  - Machine learning: a probabilistic perspective (2012)
    Murphy
- Biblio adicional se indicará en cada clase

Código

- Python
- Colab
- Kaggle
- Deepnote

Evaluación

Entregas obligatorias

- Se puede hacer en grupo de hasta 3
- Una entrega de los ejercicios de clase
- Una entrega de código

Definición de modelo estadístico

Un **modelo estadístico** es una representación condensada de nuestras **suposiciones** sobre los **datos** y su posible **generación**.

Más precisamente, un modelo estadístico M es una representación matemática de un **proceso aleatorio** que tiene el potencial de haber generado datos observables que nos interesen.

Formalmente

Un modelo estadístico es un par $M = (\mathcal{P}, \mathcal{Z})$ donde:

- $\mathcal{Z}$ es el espacio de **observaciones**.
- $\mathcal{P}$ es un conjunto de **distribuciones de probabilidad** en $\mathcal{Z}$.

Usos principales de un modelo estadístico

Descripción

Se utilizan para describir los valores típicos de una cantidad, o la relación entre diferentes variables.

Ejemplos

- ¿Cuál es la presión arterial “normal” en adultos?
- ¿Cómo se relaciona el nivel de educación con la tasa de desempleo?
- ¿Cómo se distribuye el tiempo que las personas pasan en redes sociales a lo largo de un día típico?
- ¿Cómo se compara el rendimiento académico entre estudiantes que estudian de noche versus aquellos que estudian durante el día?

Usos principales de un modelo estadístico

Predicción

Permiten sacar conclusiones sobre variables no observables directamente a partir de características observables.

Ejemplos

- A partir de una base de datos de caras etiquetadas, predecir la presencia y ubicación de rostros en nuevas fotos para aplicaciones de reconocimiento facial.
- Basado en comportamientos pasados y preferencias de usuarios, predecir qué productos, películas o libros podrían interesarles.
- Con base en registros médicos y características del paciente, predecir la probabilidad de que una persona desarrolle cierta enfermedad.
- Utilizando patrones de consumo y datos de sensores, predecir la demanda energética de un hogar.

Usos principales de un modelo estadístico

Intervención

Se usan para evaluar las implicaciones de las acciones en un sistema, relación causa-efecto.

Ejemplos

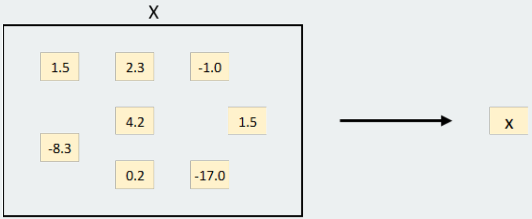
- ¿Cómo afectaría la implementación de carriles para bicicletas al uso del transporte público y privado?
- ¿Cómo impactaría una campaña masiva de vacunación en la propagación de una enfermedad infecciosa en una comunidad?
- ¿Qué impacto tendría la reducción del tamaño de las clases en los resultados académicos de los estudiantes?
- ¿Cómo afectaría a la biodiversidad local la introducción de una especie no nativa en un ecosistema?

Pensar una distribución como un generador de datos

Definición de **variable aleatoria** (informal)

- Una variable aleatoria X es una caja llena de tickets.
- En cada ticket está escrito un posible resultado numérico de un experimento.
- Todos los resultados posibles aparecen al menos una vez entre los tickets.
- Algunos pueden aparecer en muchos de ellos.

El resultado concreto x de un experimento se obtiene mezclando todos los tickets y seleccionando a ciegas solo uno.



Distribución de una variable aleatoria

- Una variable aleatoria tiene asociada una **distribución** de probabilidad $p(x)$.
- $p(x)$ describe el amontonamiento de los valores x que la variable puede tomar.
- Para representarla usamos **funciones de probabilidad** o **densidades** según el caso.

Variables discretas (informal)

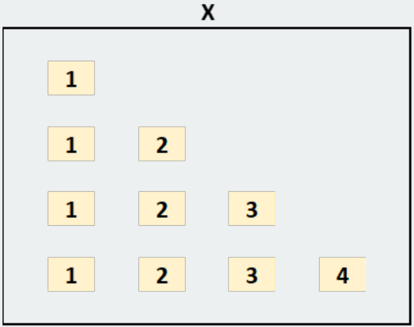
Las variables **discretas** se usan para contar ocurrencias o para representar categorías y sus valores son generalmente enteros.

Variables continuas (informal)

Las variables **continuas** se usan para mediciones y sus valores pueden ser números reales cualesquiera.

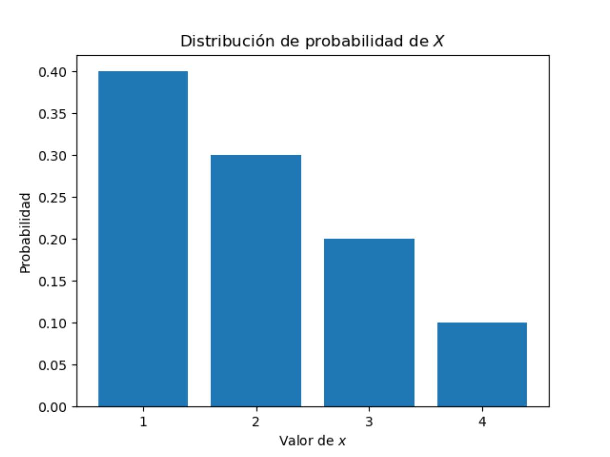
Ejemplo de variable discreta

Una urna tiene dos bolas rojas y tres blancas. Se extraen las bolas de la urna, una a la vez y sin reposición, hasta sacar una bola roja. Sea X el número de extracciones.



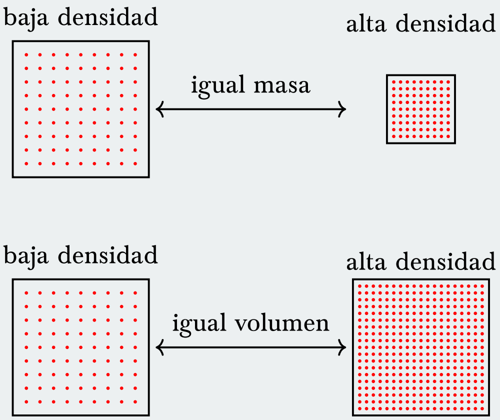
Ejemplo de variable discreta

Es apropiado un gráfico de barras para visualizar su distribución



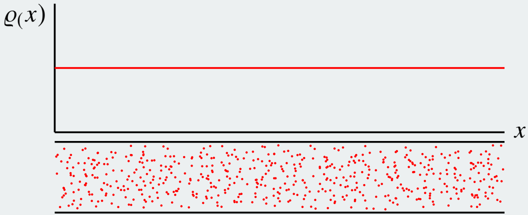
Variables continuas: el concepto de densidad

$$\text{densidad} = \frac{\text{masa}}{\text{volumen}}$$

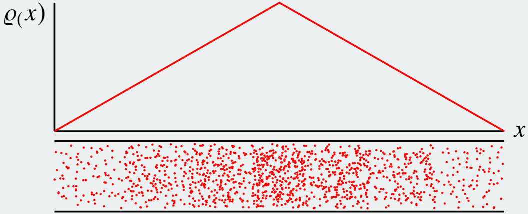


Variables continuas: el concepto de densidad

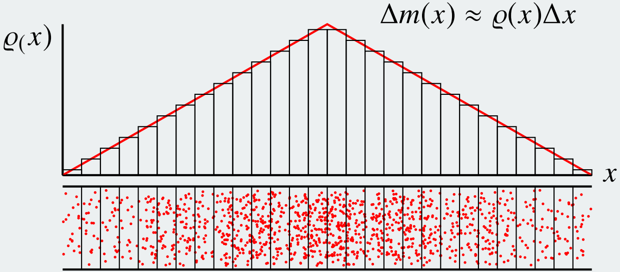
Uniforme



No uniforme



Variables continuas: el concepto de densidad



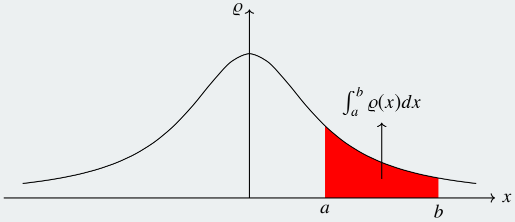
La noción de densidad de probabilidad

Definición

Una densidad de probabilidad mide el *amontonamiento* de casos.

Una variable aleatoria X tiene densidad de probabilidad $p(x)$ si cumple

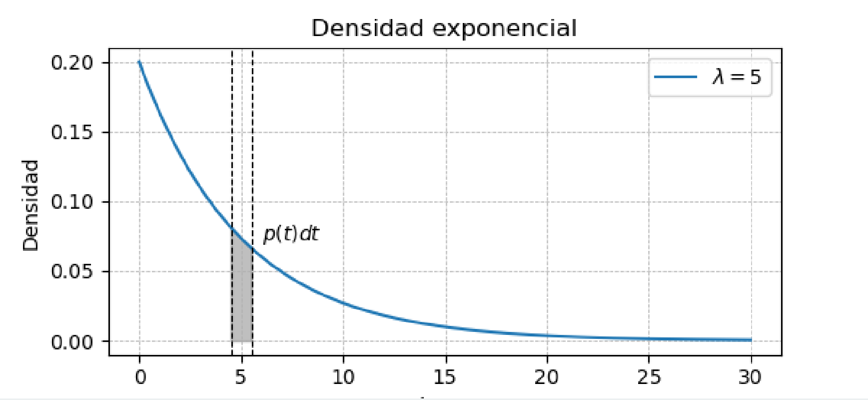
$$\mathbf{P}(a < X \leq b) = \int_a^b p(x)dx$$



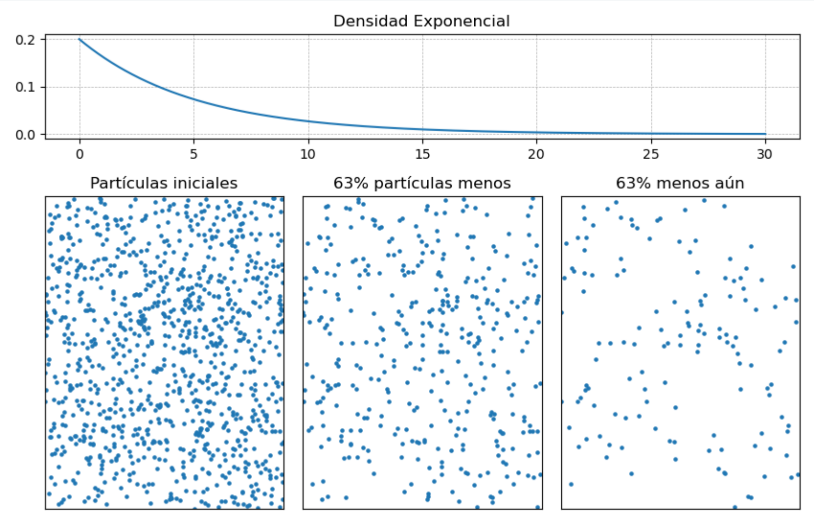
$$\mathbf{P}(x \leq X \leq x + \Delta x) \approx p(x)\Delta x, \quad \Delta x \approx 0.$$

Ejemplo de variable continua

- X es el tiempo de vida de una partícula (hasta descomponerse o mutar)
- Un ejemplo clásico es la densidad exponencial $p(t) = \frac{1}{\lambda} e^{-t/\lambda}, \quad t > 0$

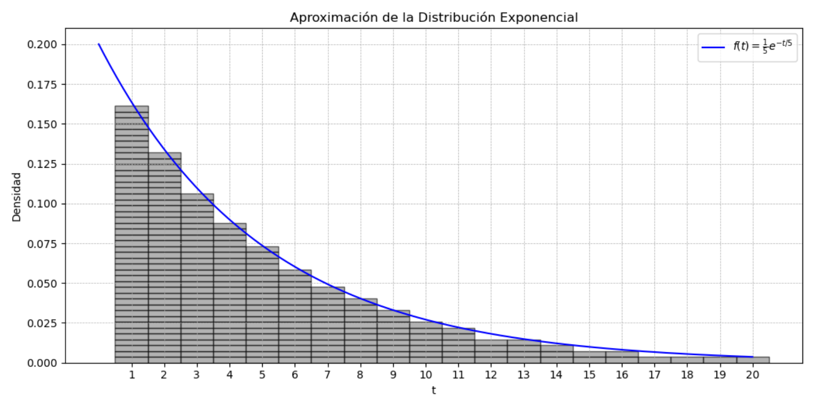


Ejemplo de variable continua



Limitantes de la analogía con las cajas de tickets

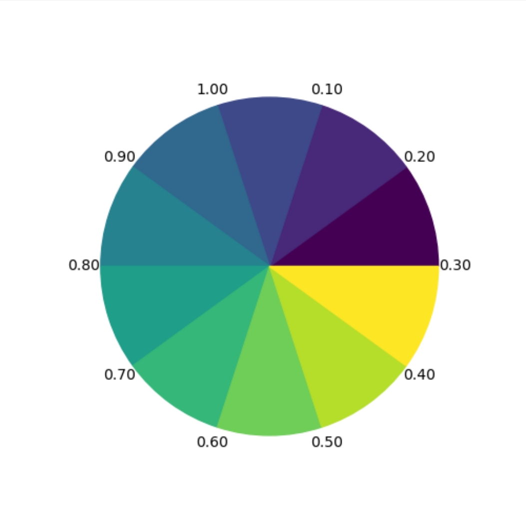
- No es posible representar una variable continua con una caja de tickets.
- Tampoco representar una variable discreta con probabilidades irracionales.
- Aunque sí se puede aproximar su distribución tan bien como una quiera.



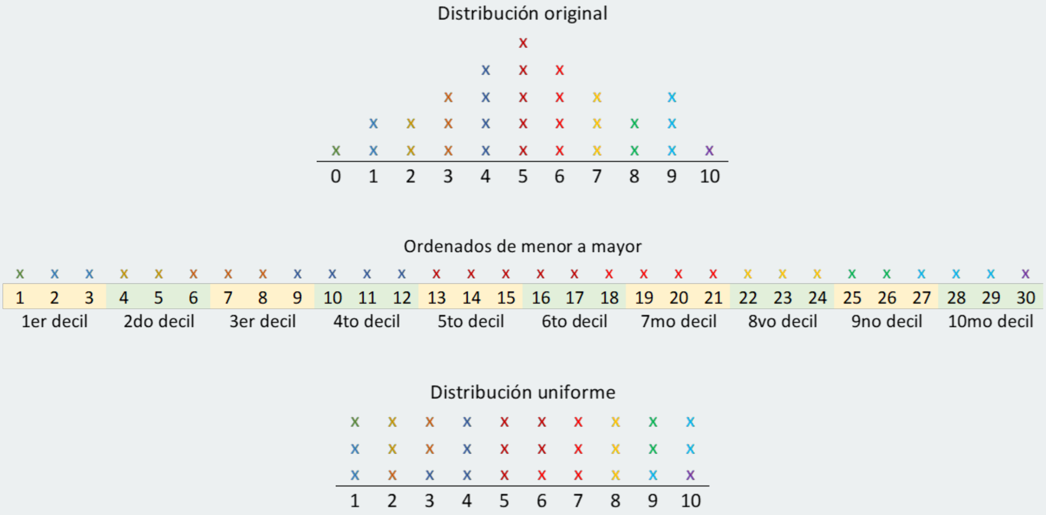
Una analogía alternativa: rueda de la fortuna

Distribución uniforme

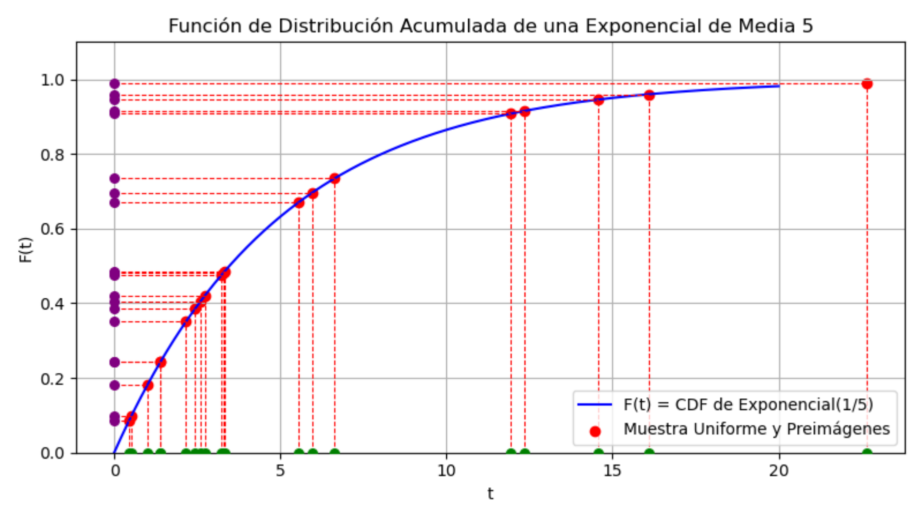
- La uniforme en el intervalo [0,1] puede representarse como una rueda de la fortuna.
- Al girar la rueda, todos los segmentos tienen la misma probabilidad de ser seleccionados.
- Con la uniforme podemos generar cualquier otra distribución.



De una distribución a la uniforme vía cuantiles



De la uniforme a una distribución vía la FDA

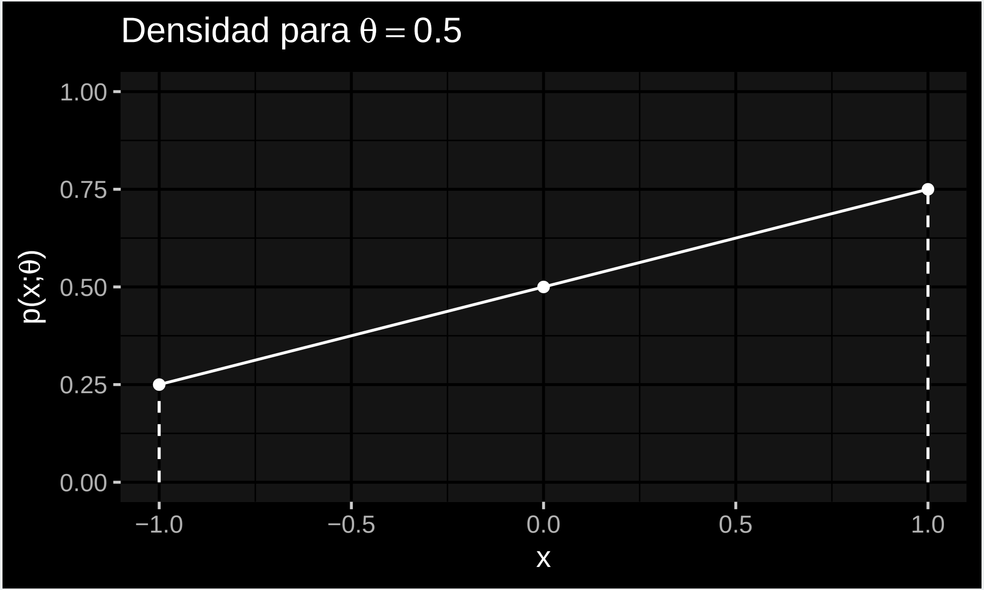


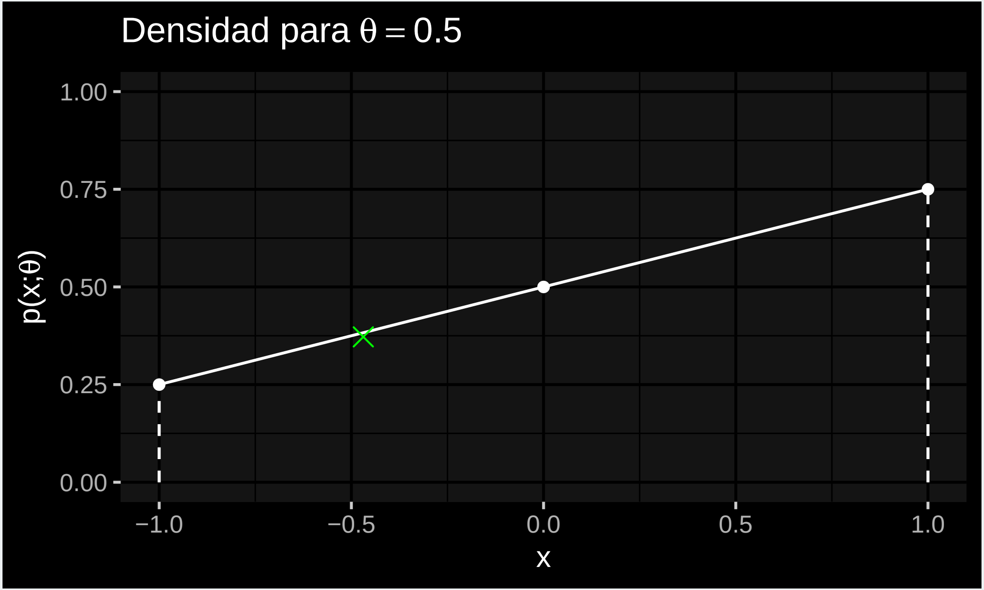
Otra forma de generar: aceptación-rechazo

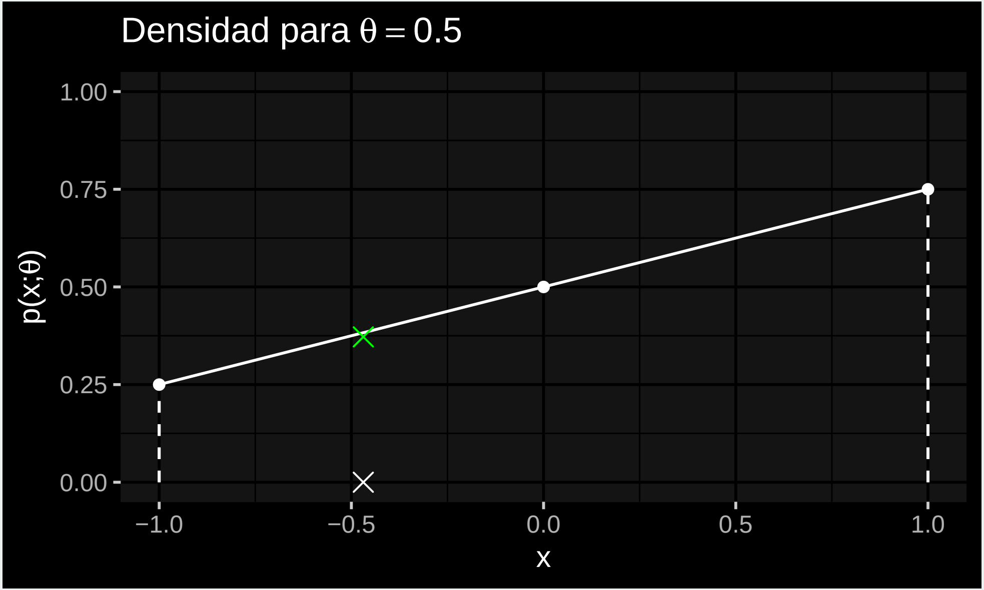
Consideremos la densidad

$$p(x;\theta) = \frac{1}{2}(1+\theta x) \quad -1 \leq x \leq 1$$

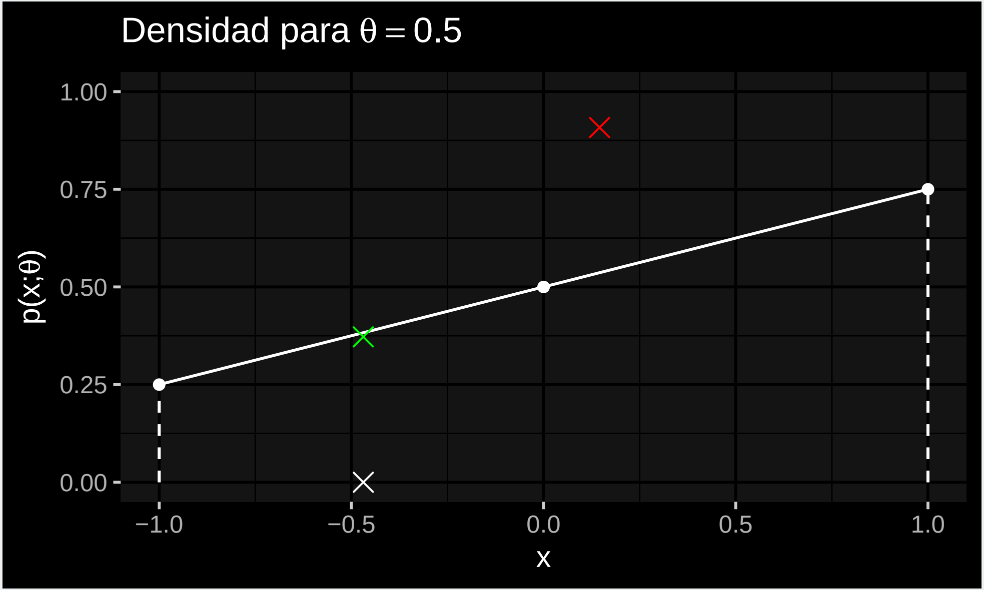
El parámetro θ también varía entre -1 y 1.

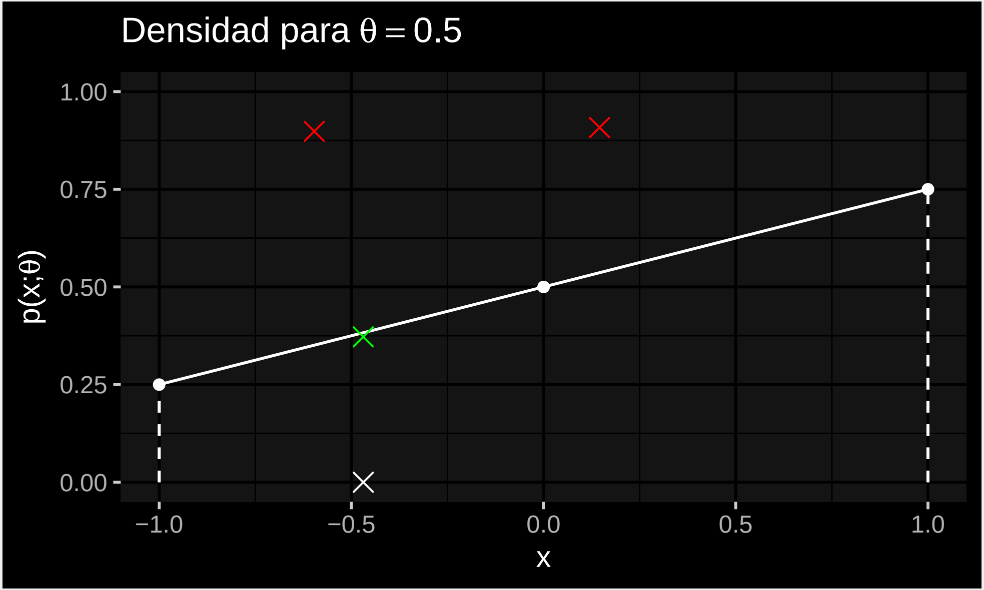

Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x


Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x

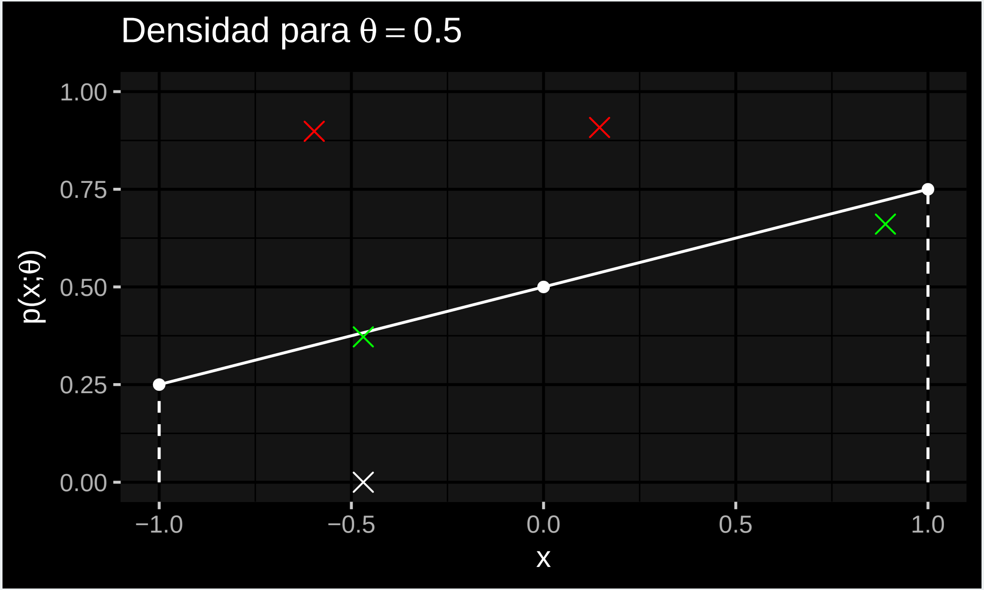


Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x

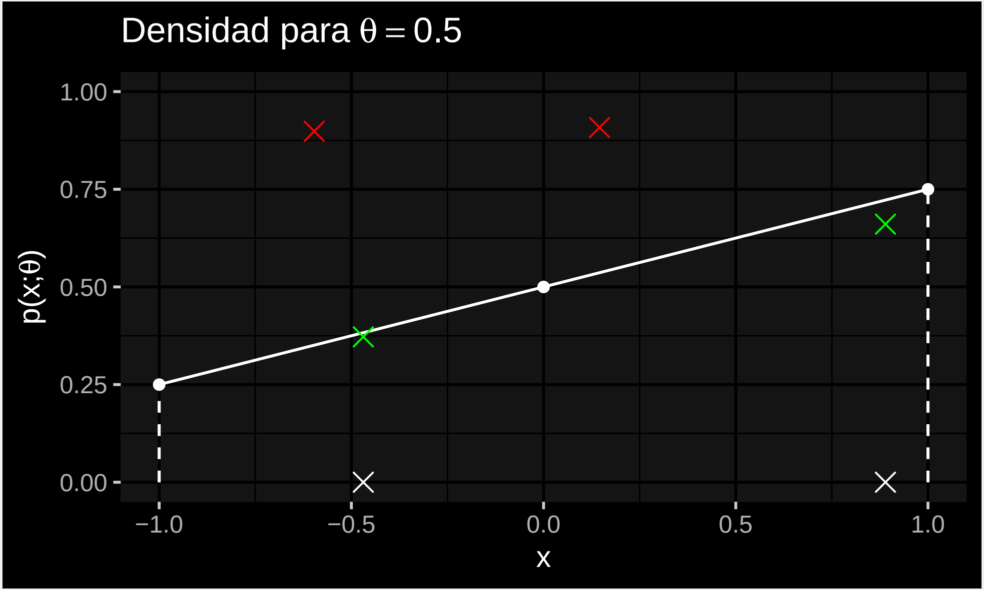

Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x



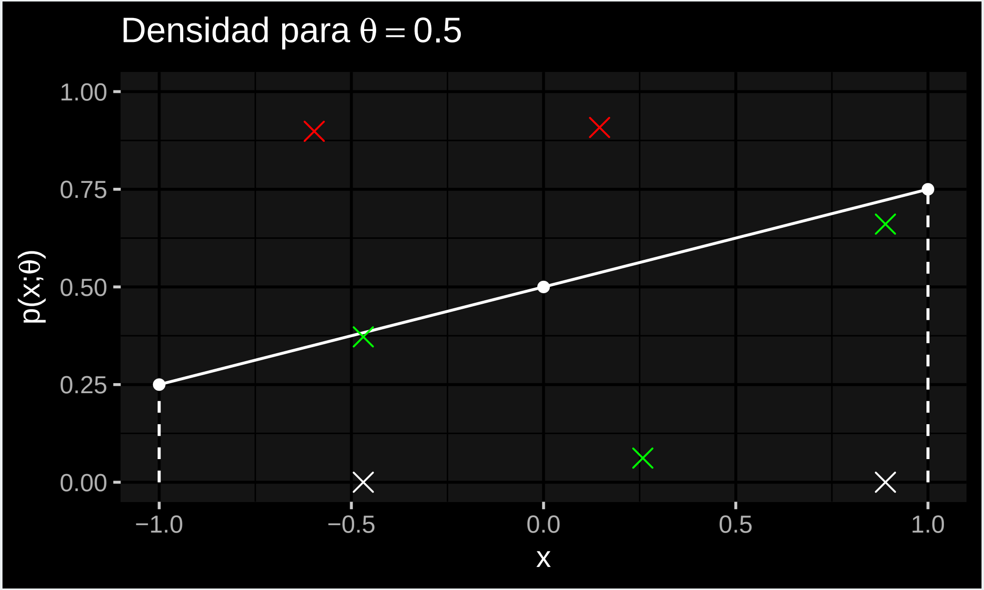
Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
−1.0
−0.5
0.0
0.5
1.0
x
p(x;θ)

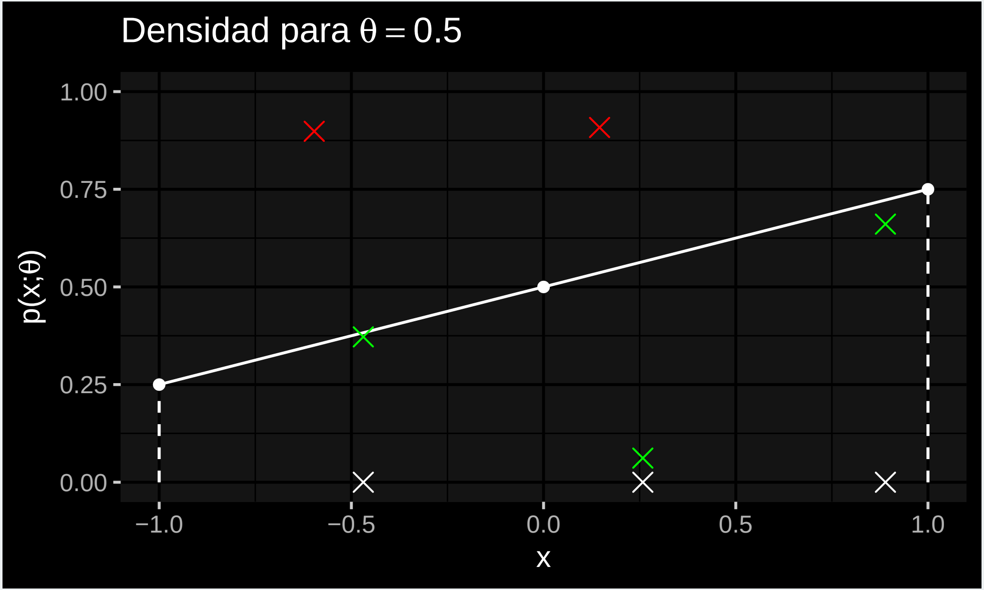


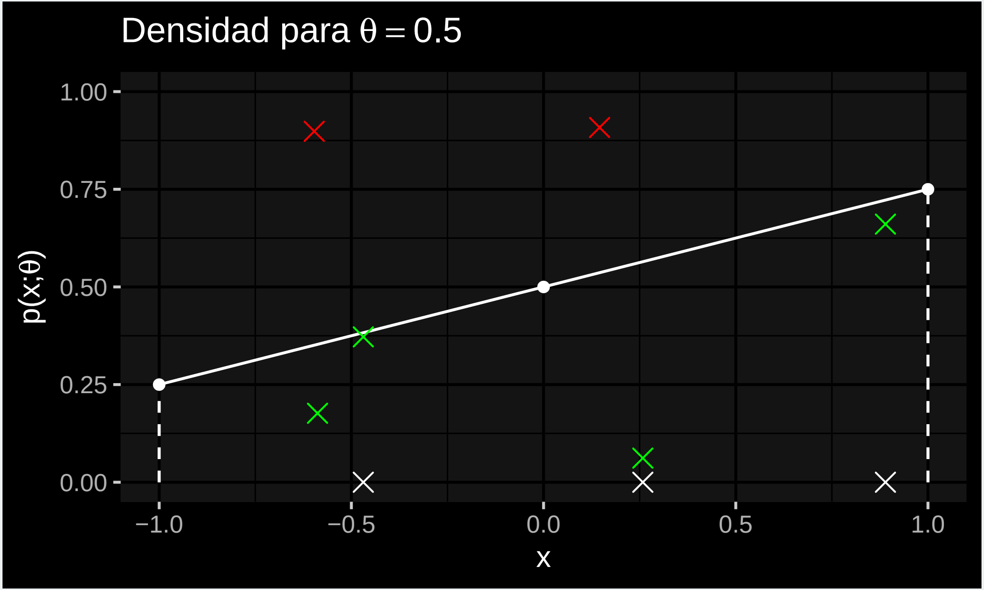
Densidad para θ = 0.5
p(x;θ)
1.00
0.75
0.50
0.25
0.00
−1.0
−0.5
0.0
0.5
1.0
x

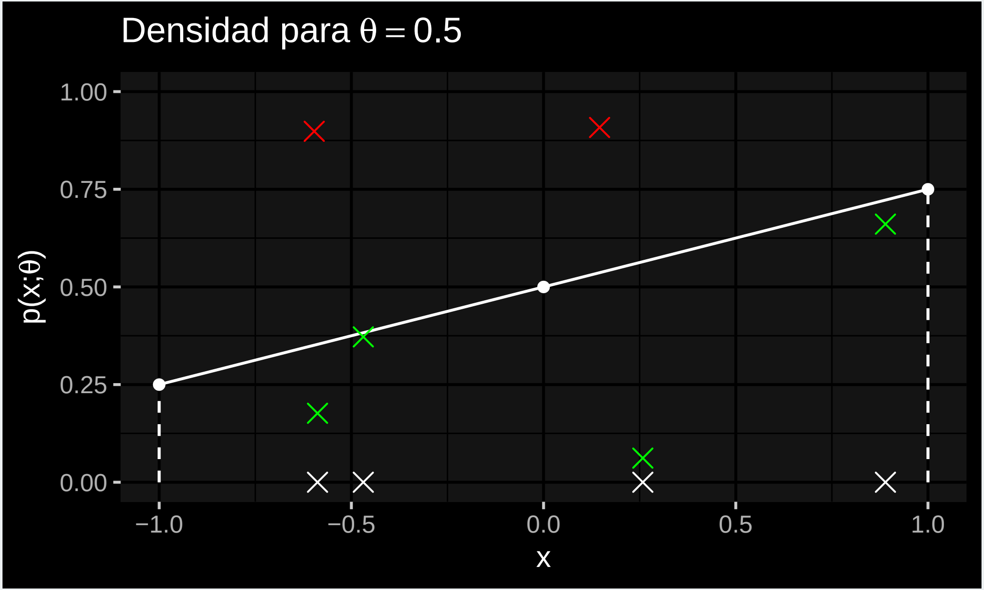


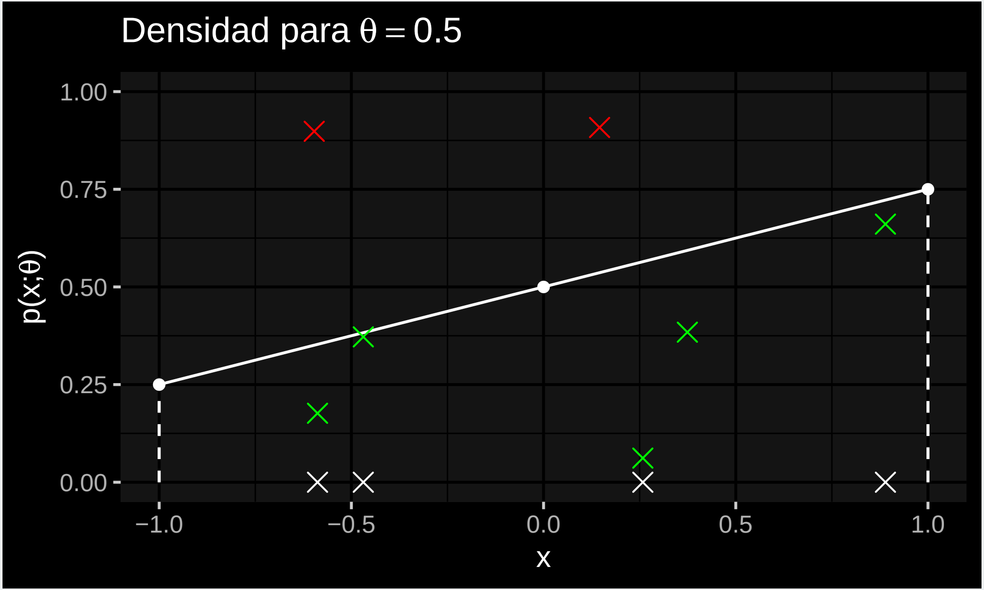
Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x

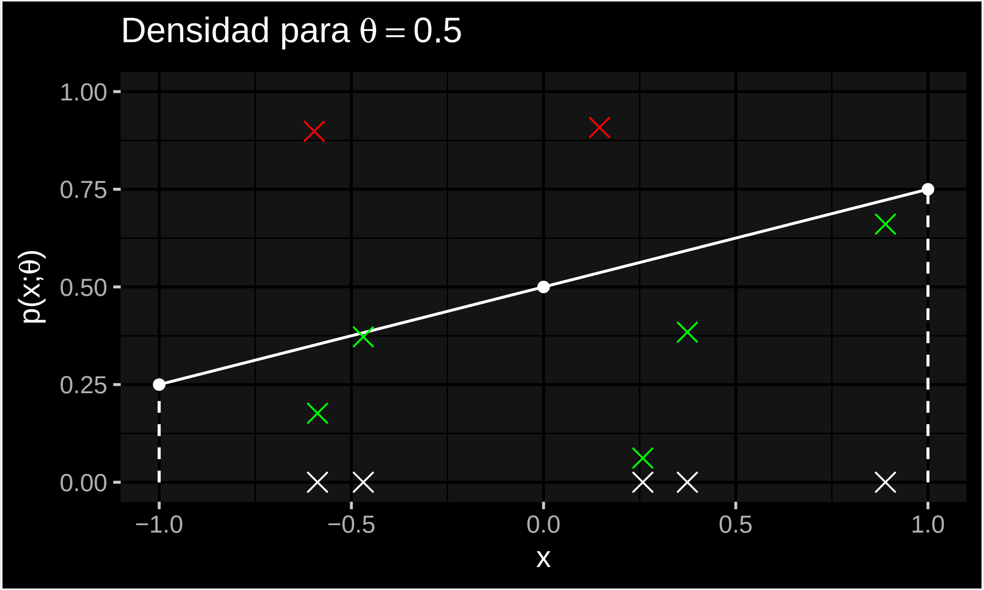

Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x

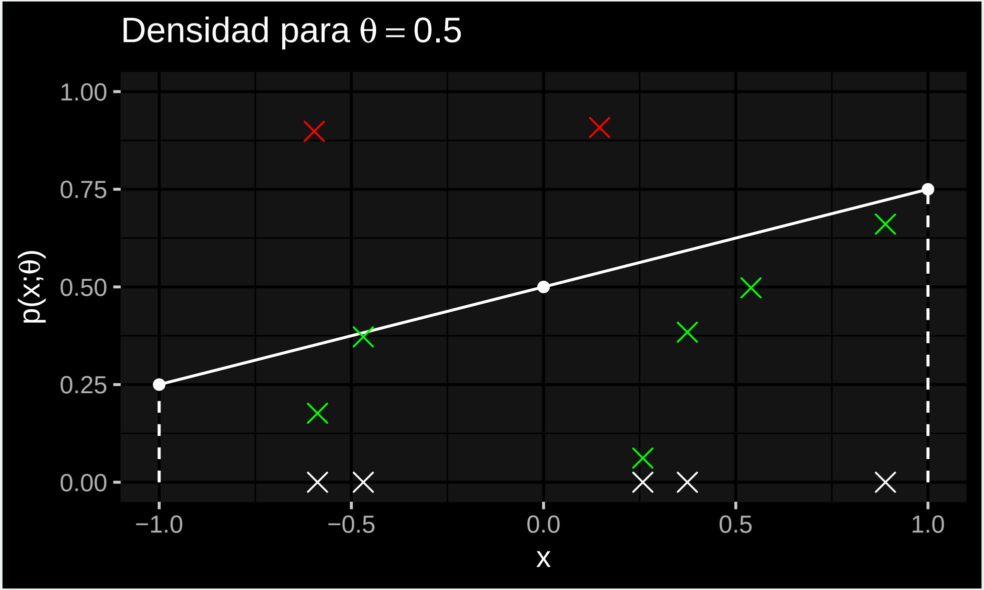

Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x

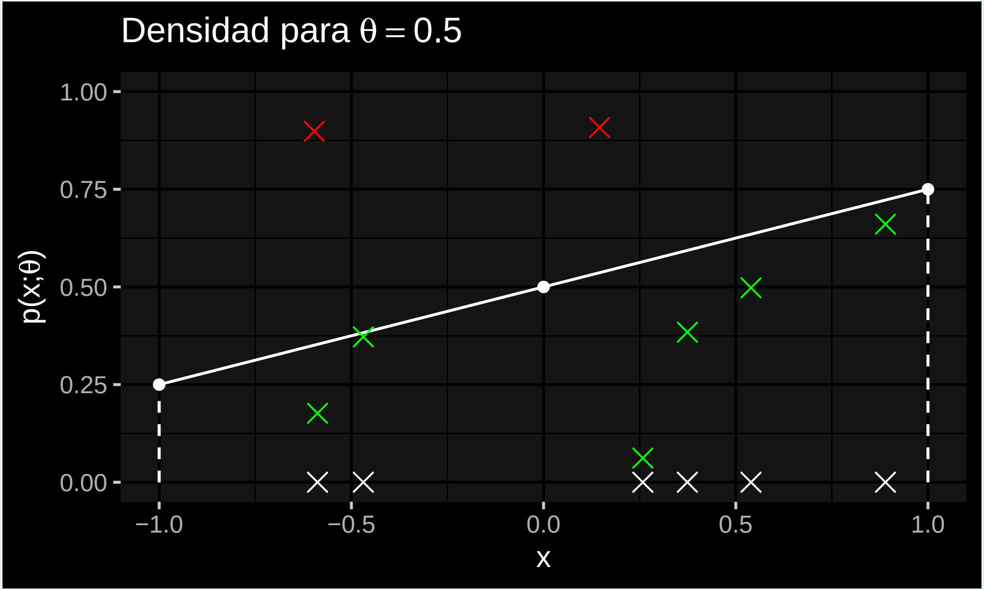

Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x

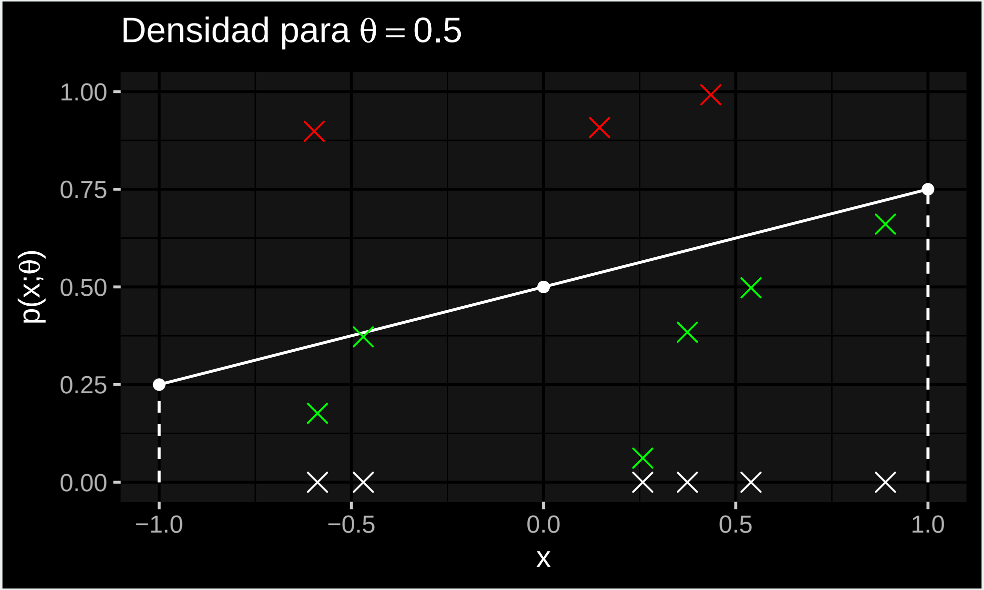

Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x

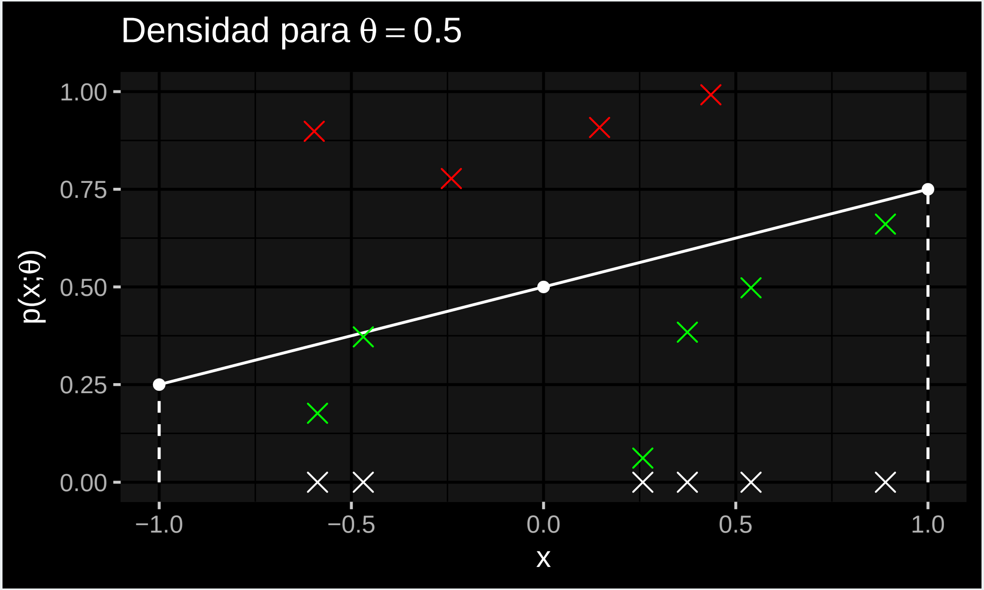

Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x

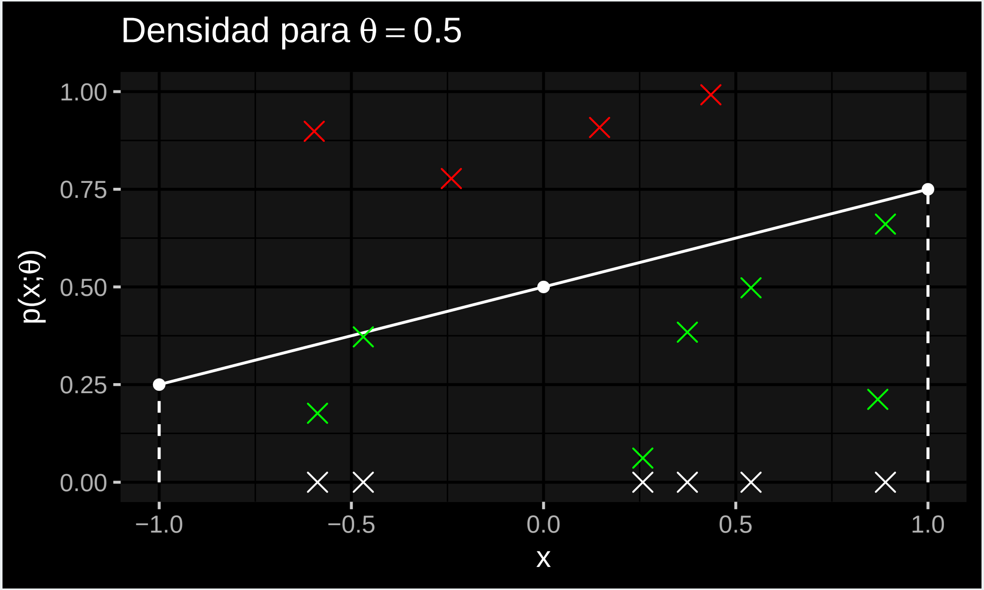

Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x

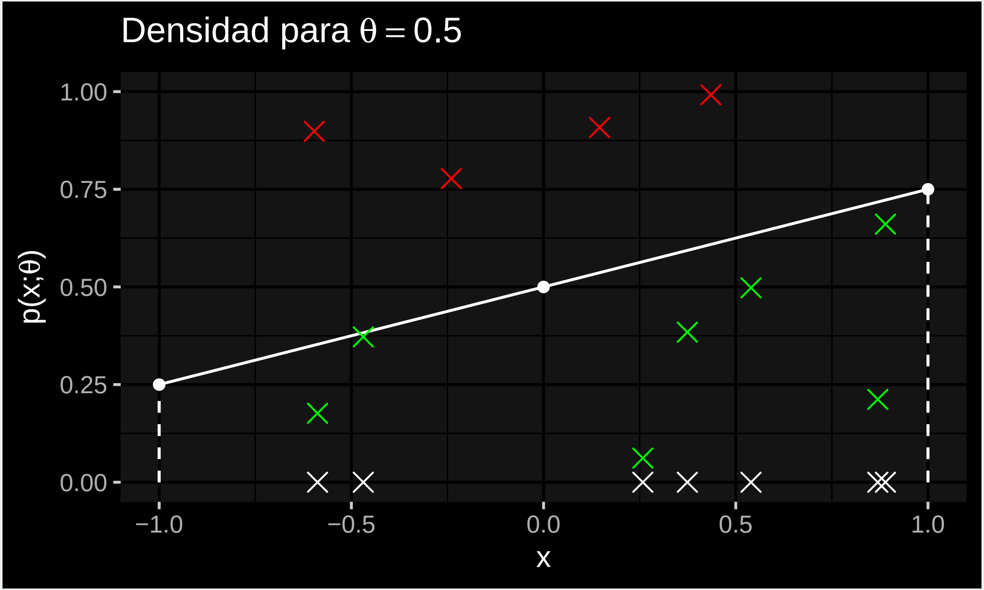

Densidad para θ = 0.5
p(x;θ)
1.00
0.75
0.50
0.25
0.00
−1.0
−0.5
0.0
0.5
1.0
x


Densidad para θ = 0.5
p(x;θ)
1.00
0.75
0.50
0.25
0.00
−1.0
−0.5
0.0
0.5
1.0
x


Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
−1.0
−0.5
0.0
0.5
1.0
x
p(x;θ)


Densidad para θ = 0.5
p(x;θ)
1.00
0.75
0.50
0.25
0.00
−1.0
−0.5
0.0
0.5
1.0
x


Densidad para θ = 0.5
p(x;θ)
1.00
0.75
0.50
0.25
0.00
−1.0
−0.5
0.0
0.5
1.0
x


Densidad para θ = 0.5
1.00
0.75
0.50
0.25
0.00
p(x;θ)
−1.0
−0.5
0.0
0.5
1.0
x

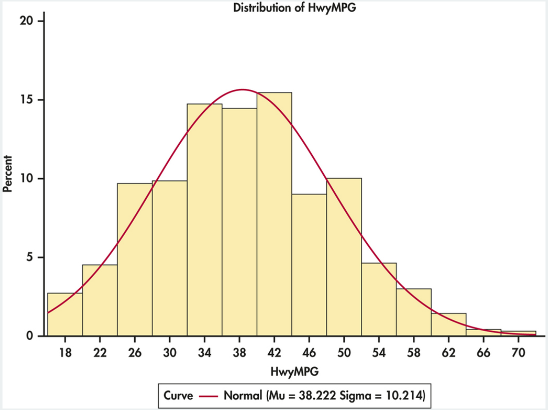
Resultado

El resultado de nuestra simulación es

$$-0.47 \quad 0.89 \quad 0.26 \quad -0.59 \quad 0.37 \quad 0.54 \quad 0.87$$

Estimar una densidad: histograma

El método más sencillo es mediante un histograma.



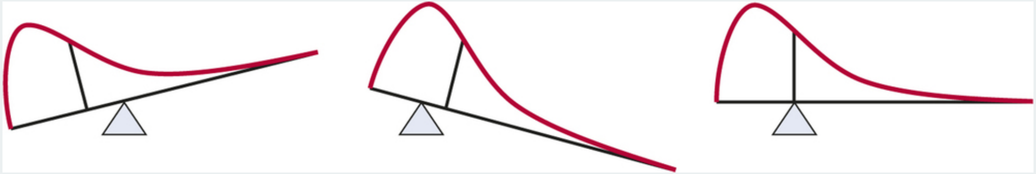
Volveremos sobre esto más adelante.

Características a observar: la media

Definición de Esperanza

La esperanza de una variable X es

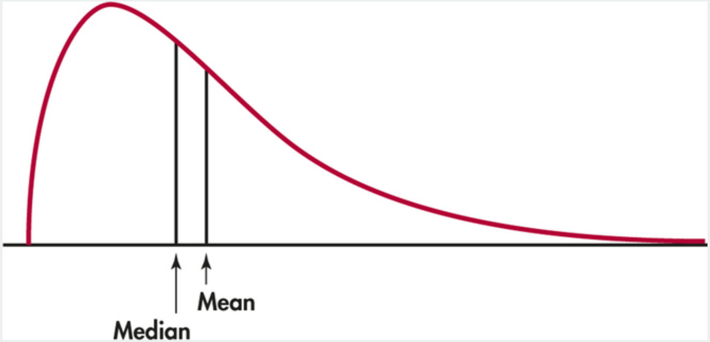
$$\mathbb{E}[X] = \begin{cases} \sum_x x p(x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{+\infty} x p(x) dx & \text{si } X \text{ es continua} \end{cases}$$



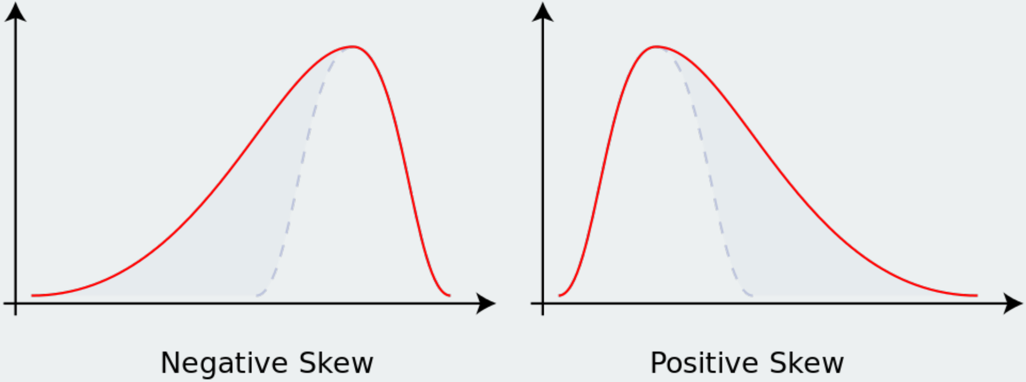
Características a observar: cuantiles

Definición de Cuantil

El cuantil x_p de orden p, para cualquier $p \in [0, 1]$, de una variable continua X se define como la preimagen por F_X de p. Para $p = 1/2$ se llama mediana.

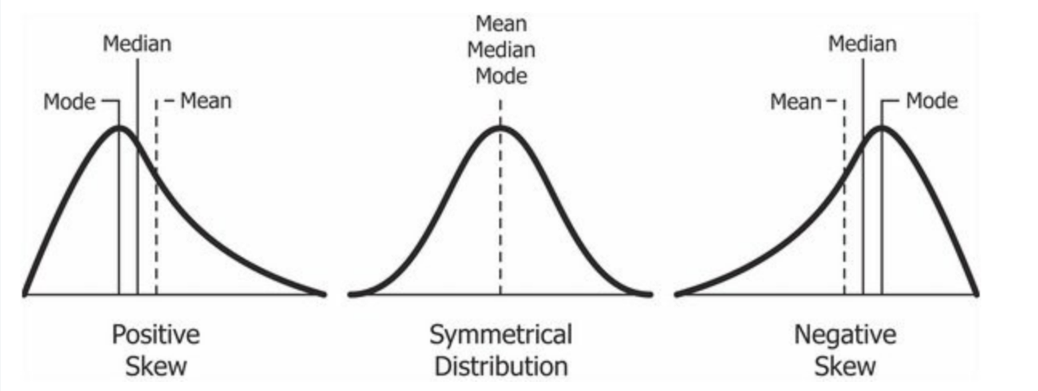


Características a observar: simetría



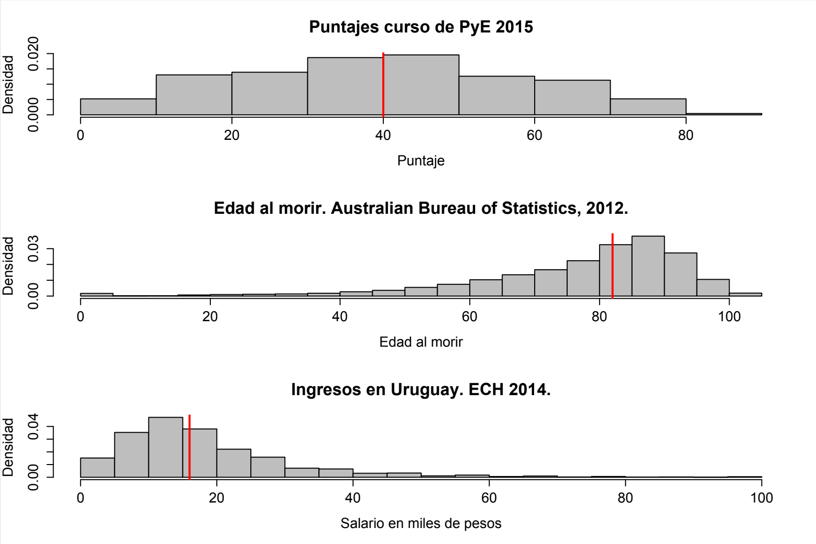
Fuente: Wikipedia

Características a observar: simetría



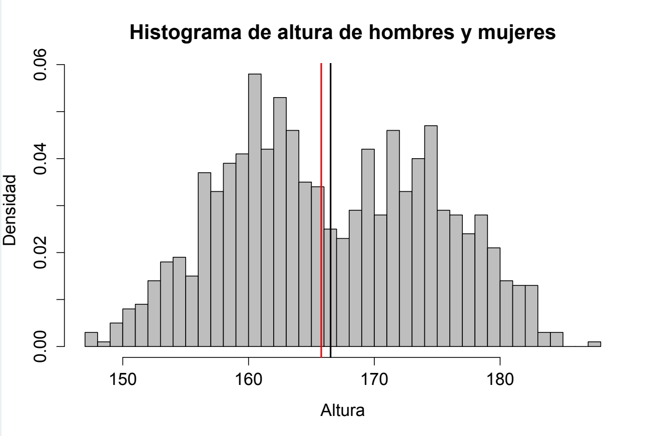
Fuente: Wikipedia

Características a observar: simetría



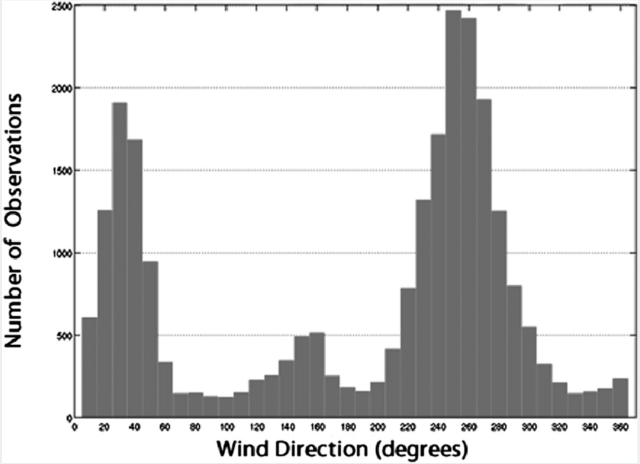
Características a observar: multimodalidad

La multimodalidad puede darse por muchas razones. Una muy común es la mezcla de subpoblaciones dentro de una misma población:



Características a observar: multimodalidad

Pero también puede ser una característica intrínseca de la distribución:



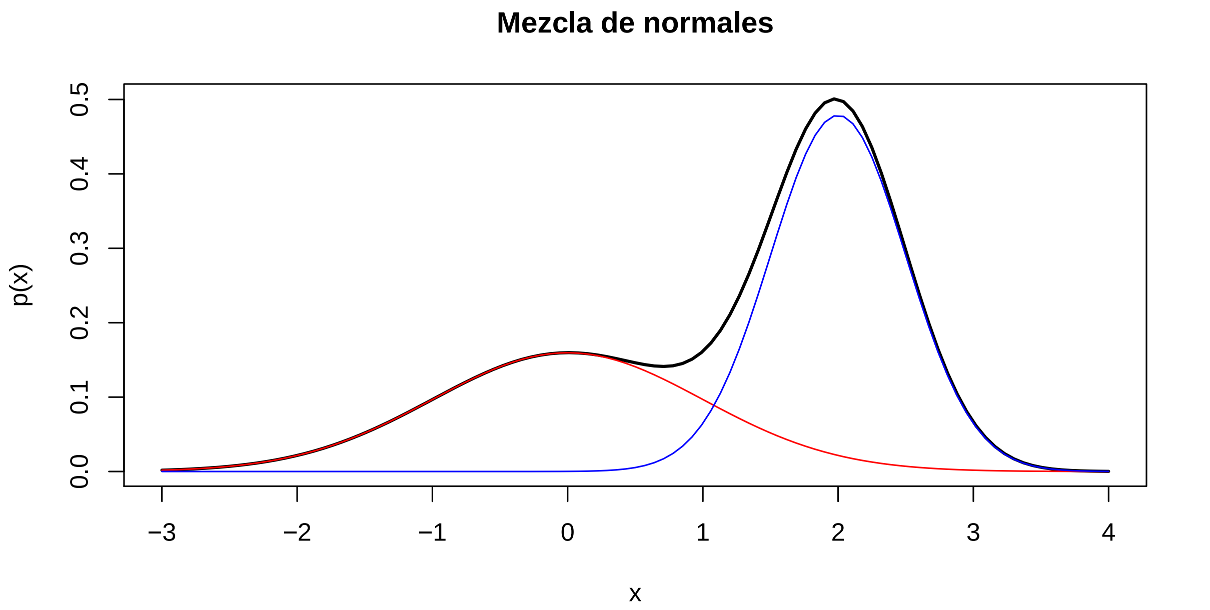
Fuente: Climatology of High Wind Events in the Owens Valley, California

Mezclas de distribuciones

Definición de mezcla

Dado un conjunto finito de funciones de densidad de probabilidad $p_1(x), \ldots, p_n(x)$, y pesos $w_1, \ldots, w_n$ tales que $w_i \geq 0$ y $\sum w_i = 1$, la distribución de la mezcla se puede representar por la densidad

$$p(x) = \sum_{i=1}^{n} w_i \, p_i(x).$$


Mezcla de normales
p(x)
0.0
0.1
0.2
0.3
0.4
0.5
x
−3
−2
−1
0
1
2
3
4

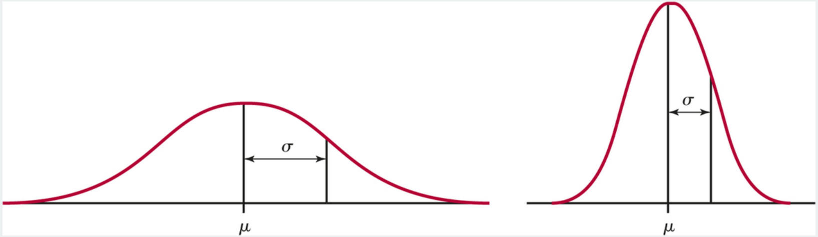
Características a observar: varianza y desvío

Definición de varianza y desvío

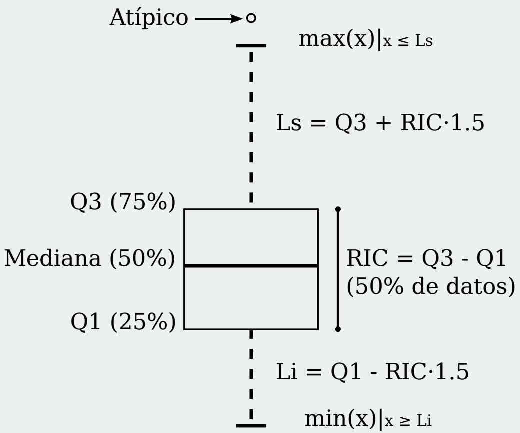
Sea X una variable con media $\mu = \mathbb{E}[X]$. La varianza de X se define como

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

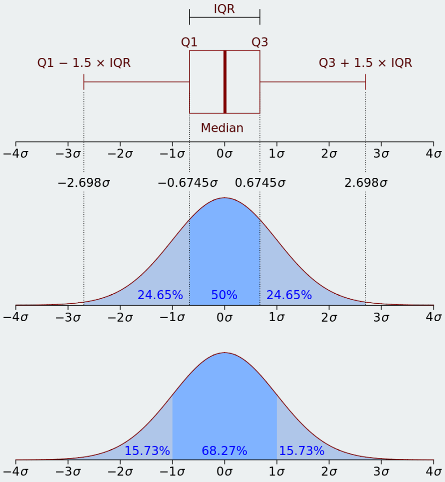
El desvío estándar es la raíz de la varianza $\sigma = \sqrt{\mathrm{Var}[X]}$



Resumen visual-numérico: el boxplot



Boxplot: comparación con la normal



Fuente: Wikipedia

En la distribución normal $N(\mu, \sigma^2)$, el 99.73 % del área debajo la campana está a menos de 3σ de distancia de μ.
El IQR es en este caso $1{,}349\sigma$. El valor $0{,}6745\sigma$ se conoce históricamente como *error probable*.

Ejercicio: interpretar un resumen numérico

Abajo se muestran indicadores que caracterizan la distribución de notas de dos clases paralelas de un curso de Inglés. El puntaje máximo es 100.

	Clase 1	Clase 2
Promedio	72	78
Mediana	73	65
Desvío estándar	6	16

1. Bosquejar el histograma de la distribución de notas de cada clase.
2. ¿En cuál de las dos clases es más probable encontrar un estudiante con nota alta?