



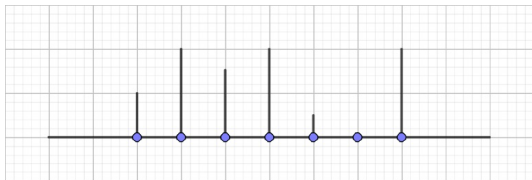
**FACULTAD de
CIENCIAS ECONÓMICAS**

ESTADÍSTICOS

ALAN REYES-FIGUEROA

ELEMENTS OF MACHINE LEARNING

(AULA 05) 24.ENERO.2023



Resúmenes de distribuciones:

- localización (promedio, rango, soporte o dominio);
- variabilidad (desviación estándar, varianza, entropía);
- forma de la distribución (kurtosis, histogramas, diagramas de probabilidad PP o QQ);
- simetría (sesgo, coeficiente de asimetría);
- En el caso de más variables: nos interesa algo que mida el grado de relación entre ellas (covarianza, correlación, información mutua).

Valores numéricos (o vectoriales) en términos de la variable aleatoria.
Resumen de una distribución.

Existen estadísticos con varios propósitos: localización, variabilidad, ...

Promedio: El **promedio** o **esperanza** (*expectativa, valor esperado*) de una variable aleatoria discreta X , $\mathbb{E}(X)$, se define como

$$\mathbb{E}(X) = \sum_x x \mathbb{P}(X = x).$$

Comentario: En la vida cotidiana usamos como promedio de $\{x_i\}_{i=1}^n$ a

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

En general,

Dada una función $g(\cdot)$, se define la **esperanza** $\mathbb{E}(g(X))$ como:

$$\mathbb{E}(g(X)) = \sum_x g(x) \mathbb{P}(X = x).$$

Propiedades:

1. (Linealidad) $\mathbb{E}(aX_1 + bX_2) = a\mathbb{E}(X_1) + b\mathbb{E}(X_2)$.
2. (Independencia) Si X_1, X_2 son independientes, entonces $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1) \mathbb{E}(X_2)$.

Media, mediana y moda

Media: Sea una variable aleatoria discreta X con probabilidad \mathbb{P} . La **esperanza** (*expectativa, valor esperado*) de X se define como

$$\mathbb{E}(X) = \sum_x x \mathbb{P}(X = x).$$

Mediana: Una **mediana** de X es cualquier valor $t \in \mathbb{R}$ que satisface $F_X(t) = \frac{1}{2}$. Dicho de otra manera, son las preimágenes $F_X^{-1}(1/2)$.

Obs! F_X en general no es invertible!! Denotamos $Q_X : [0, 1] \rightarrow \overline{\mathbb{R}}$ a la *función de cuantiles*, la inversa generalizada de F_X :

$$Q_X(\alpha) = \inf\{x \in \mathbb{R} : \alpha \leq F_X(x)\}, \quad \text{para } 0 < \alpha < 1.$$

Moda: Una **moda** de la distribución de X es cualquier máximo local de f_X . (unimodal, bimodal, ..., multimodal)

El valor esperado $\mathbb{E}(X)$ tiene otra propiedad importante: es el valor constante que minimiza la suma de errores cuadrados. Dado $\{x_i\}_{i=1}^n$ la imagen de la v.a. X , sea $p_i = \mathbb{P}(X = x_i)$. Queremos

$$\text{minimizar } J(c) = \text{minimizar } \sum_{i=1}^n p_i (x_i - c)^2.$$

Solución: Derivando con respecto de c , obtenemos

$$J'(c) = 2 \sum_{i=1}^n p_i (x_i - c) = 0.$$

$$\text{Luego } \sum_{i=1}^n p_i x_i = c \sum_{i=1}^n p_i = c \Rightarrow c = \sum_{i=1}^n p_i x_i = \sum_{i=1}^n x_i \mathbb{P}(X = x_i) = \mathbb{E}(X).$$

- El valor que minimiza $\sum_{i=1}^n p_i |x_i - c|_1$ es: la *mediana* de X .
- El valor que minimiza $\sum_{i=1}^n p_i |x_i - c|_0$ es: la *moda* de X .

Esperanza condicional

Definición

Para la v.a. X y para un evento $A \in \mathcal{F}$, se define el **promedio condicional** (o **esperanza condicional**) de X dado A como

$$\mathbb{E}(X \mid A) = \sum_x x \mathbb{P}(X = x \mid A).$$

Definición

Para las v.a. X y Y , se define la **esperanza condicional** de X dado que Y es igual a un valor y , como

$$\mathbb{E}(X \mid Y = y) = \sum_x x \mathbb{P}(X = x \mid Y = y).$$

Esperanza condicional

En general, definimos $\mathbb{E}(g(X) \mid Y = y) = \sum_x g(x) \mathbb{P}(X = x \mid Y = y)$.

Propiedades:

Sean X, Y, Z v.a., $a, b \in \mathbb{R}$ y $g : \mathbb{R} \rightarrow \mathbb{R}$.

1. $\mathbb{E}(a \mid Y) = a$.
2. $\mathbb{E}(aX + bZ \mid Y) = a\mathbb{E}(X \mid Y) + b\mathbb{E}(Z \mid Y)$.
3. $\mathbb{E}(X \mid Y) \geq 0$ si $X \geq 0$.
4. $\mathbb{E}(X \mid Y) = \mathbb{E}(X)$ si X, Y son independientes.
5. $\mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}(X)$.
6. $\mathbb{E}(Xg(Y) \mid Y) = g(Y)\mathbb{E}(X \mid Y)$. En particular, $\mathbb{E}(g(Y) \mid Y) = g(Y)$.
7. $\mathbb{E}(X \mid Y, g(Y)) = \mathbb{E}(X \mid Y)$.
8. $\mathbb{E}(\mathbb{E}(X \mid Y, Z) \mid Y) = \mathbb{E}(X \mid Y)$.

Proposición (Ley de la probabilidad total para esperanzas)

Sean X, Y v.a. discretas, entonces

$$\mathbb{E}(X) = \sum_y \mathbb{E}(X \mid Y = y) \mathbb{P}(Y = y).$$

Ejemplo: Alguien anda perdido en la subterránea de Guanajuato. Está en un cruce con 3 opciones. Un camino le lleva a la salida en 10 minutos en promedio, un segundo camino le regresa a su lugar en promedio 15 minutos y un tercer camino le regresa a su lugar en promedio 5 minutos. Siempre elige alguna opción, independiente del pasado. ¿Cuánto va a tardar en promedio para salir?

Esperanza condicional

Solución:

- Opción 1: salida del subterráneo en 10m promedio.
- Opción 2: regresa al mismo lugar en 15m promedio.
- Opción 3: regresa al mismo lugar en 5m promedio.

Definimos E_i = elegir opción i , $i = 1, 2, 3$. T la v.a. = tiempo de salida.

$$\begin{aligned}\mathbb{E}(T) &= \sum_{i=1}^3 \mathbb{E}(T \mid E_i) \mathbb{P}(E_i) \\&= \mathbb{E}(T \mid E_1) \cdot \frac{1}{3} + \mathbb{E}(T \mid E_2) \cdot \frac{1}{3} + \mathbb{E}(T \mid E_3) \cdot \frac{1}{3} \\&= (10m) \cdot \frac{1}{3} + (15m + \mathbb{E}(T)) \cdot \frac{1}{3} + \mathbb{E}(5m + \mathbb{E}(T)) \cdot \frac{1}{3} \\&= \frac{10}{3}m + \frac{15}{3}m + \frac{5}{3}m + \frac{2}{3}\mathbb{E}(T).\end{aligned}$$

Luego, $\frac{1}{3}\mathbb{E}(T) = 10m \Rightarrow \mathbb{E}(T) = 30m$.

Sea X una variable aleatoria en \mathbb{R} . Definimos su **varianza** como:

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Propiedades:

- $\text{Var}(X) \geq 0$.
- $\text{Var}(aX) = a^2 \text{Var}(X)$.
- Si X_1, X_2 son independientes, entonces

$$\text{Var}(aX_1 + bX_2) = a^2 \text{Var}(X_1) + b^2 \text{Var}(X_2).$$

Covarianza

Consideramos ahora cómo relacionar dos variables aleatorias.

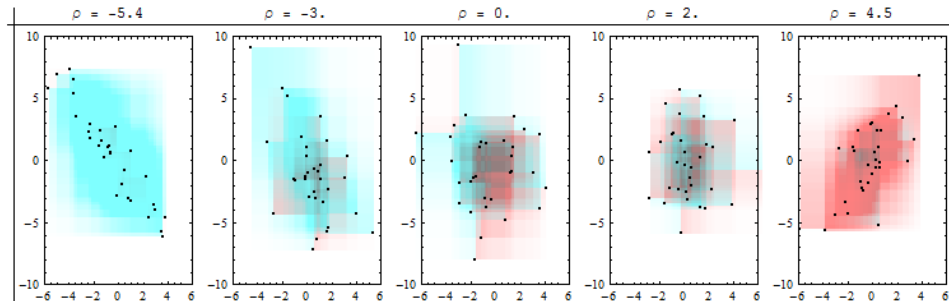
Dada dos variables aleatorias X_1, X_2 (definidas sobre el mismo espacio). Definimos su **covarianza** como:

$$\text{Cov}(X_1, X_2) = \mathbb{E}[(X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)].$$

Propiedades:

- $\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1)$.
- $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$.
- $\text{Cov}(aX, X) = a\text{Var}(X)$.
- Si X_1, X_2 son independientes, entonces $\text{Cov}(X_1, X_2) = 0$.

Covarianza



Dadas dos variables aleatorias X, Y , definimos su **correlación** (o **coeficiente de correlación**) como:

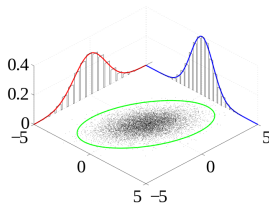
$$\rho(X, Y) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

Propiedades:

- $\rho(X, Y) = \rho(Y, X)$.
- $-1 \leq \rho(X, Y) \leq 1$.
- $\rho(aX, bY) = \rho(X, Y)$.
- $\rho(aX, X) = \text{sign}(a)$.
- Si X, Y son independientes, entonces $\rho(X, Y) = 0$.

Correlación

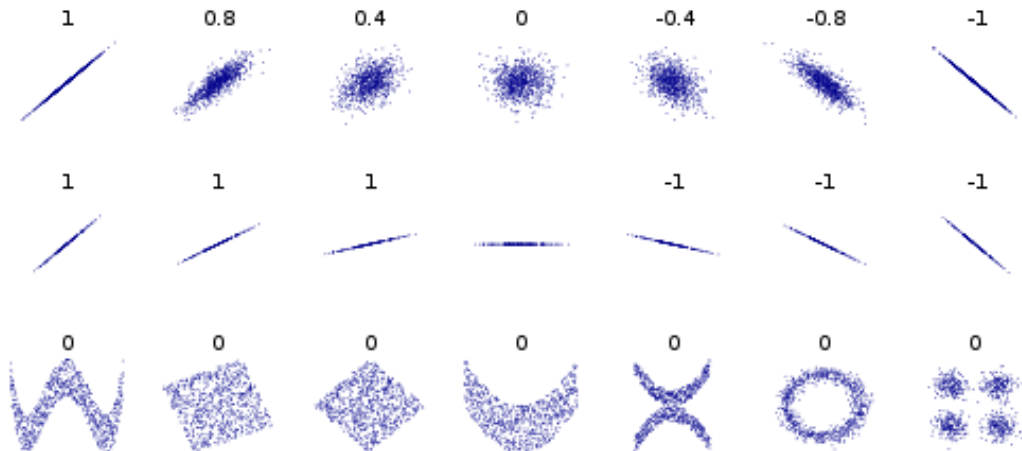
Por ejemplo, para el caso de dos v.a. normales X y Y :



tenemos



Correlación



Ya vimos que la varianza presenta limitaciones (igual que la covarianza).

Punto de partida: medir la sorpresa asociada el evento $X = x$, $I(x)$. La entropía es el valor esperado de esta sorpresa $\mathbb{E}(I(x))$.

¿Cómo medimos esta sorpresa o incerteza?

- Un evento que ocurre con alta probabilidad no genera sorpresa.
- Un evento que ocurre con baja probabilidad genera mayor sorpresa (más entre menor es \mathbb{P}).

¿Cómo definir $I(x)$? Tenemos varias alternativas simples

$$I(x) = \frac{1}{\mathbb{P}(X = x)}, \quad I(x) = 1 - \mathbb{P}(X = x), \quad I(x) = -\log \mathbb{P}(X = x).$$

Definición

Sea X una v.a. discreta. Definimos su **entropía de Shannon** como:

$$H(X) = - \sum_x \mathbb{P}(X = x) \log \mathbb{P}(X = x).$$

Comentario: Shannon definió la entropía en un contexto de teoría de la información (bits), usa \log_2 . Si $p = 0$, usualmente se define $p \log p = 0$.

Definición

Sea X una v.a. discreta. Definimos su **entropía de Gini o coeficiente de Gini** por:

$$G(X) = \sum_x \mathbb{P}(X = x) (1 - \mathbb{P}(X = x)) = 1 - \sum_x \mathbb{P}(X = x)^2.$$

1. Dibuja dos distribuciones o variables aleatorias (discretas) distintas, con mismo promedio y entropía, pero varianzas diferentes.
2. Toma una v.a. $X \in \{0, 1\}$. Calcular la varianza y la entropía de Shannon y de Gini en función de $p = \mathbb{P}(X = 1)$.
Compara las gráficas de $H(X)$ y $2G(X)$.
3. ¿Cuáles son los valores mínimo y máximo para $H(X)$ y $G(X)$?

Entropía condicional

Definición

Sean X, Y dos variables aleatorias, la **entropía condicional** de X dado Y es

$$H_Y(X) = \mathbb{E}H(X | Y) = - \sum_y \left(\sum_x \mathbb{P}(X = x | Y = y) \log \mathbb{P}(X = x | Y = y) \right) \mathbb{P}(Y = y).$$

Obs. No es simétrica: $H_Y(X) \neq H_X(Y)$.

Definición

Sean X, Y dos variables aleatorias, la **información mutua** de X y Y está dada por

$$I(X, Y) = H(X) - H_Y(X).$$

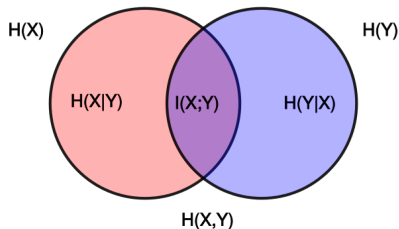
Proposición

$$I(X, Y) = I(Y, X).$$

Definición

La **entropía conjunta** de X y Y es

$$H(X, Y) = - \sum_x \sum_y \mathbb{P}(x, y) \log \mathbb{P}(x, y).$$



Un diagrama de Venn que muestra relaciones aditivas y sustractivas entre varias medidas de información asociadas con las variables X y Y . El área contenida por ambos círculos es la entropía conjunta $H(X, Y)$. El círculo de la izquierda (rojo y violeta) es la entropía individual $H(X)$, siendo el rojo la entropía condicional $H_Y(X)$. El círculo de la derecha (azul y violeta) es $H(Y)$, y el azul es $H_X(Y)$. El violeta es la información mutua $I(X, Y)$.

Entropía

Sean P una distribución discreta de probabilidad, la **entropía** de P es

$$H(P) = - \sum_x P(x) \log P(x).$$

Sean P, Q dos distribuciones discretas de probabilidad, la **entropía cruzada** (*cross-entropy*) de P y Q es

$$H(P, Q) = - \sum_x P(x) \log Q(x).$$

Además, la **divergencia de Kullback-Leibler** de P y Q se define como

$$\begin{aligned} D_{KL}(P \parallel Q) &= - \sum_x P(x) \log \frac{Q(x)}{P(x)} \\ &= - \sum_x P(x) \log Q(x) + \sum_x P(x) \log P(x) = H(P, Q) - H(P). \end{aligned}$$