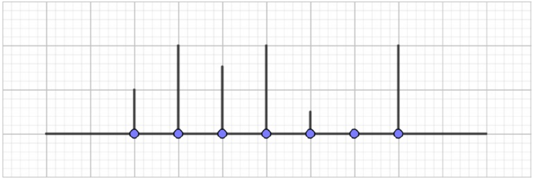
Variables aleatorias discretas

Alan Reyes-Figueroa

Introducción a la Ciencia de Datos (Aula 05) 25.enero.2021

Estadísticos



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).

Estadísticos

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

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

en caso de que la suma exista.

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,

Definición

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

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

en caso de que la suma exista.

Proposición

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* $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\,\mathbb{E}(X_2)$.

Prueba:

$$\begin{aligned}\mathbb{E}(aX_1+bX_2) &= \sum_{\omega}(aX_1+bX_2)(\omega)\mathbb{P}(\omega) = a\sum_{\omega}X_1(\omega)\mathbb{P}(\omega) + b\sum_{\omega}X_2(\omega)\mathbb{P}(\omega)\\ &= a\mathbb{E}(X_1)+b\mathbb{E}(X_2).\end{aligned}$$

Como $X_1 \perp X_2$, entonces $\mathbb{P}(X_1 = x_1, X_2 = x_2) = \mathbb{P}(X_1 = x_1)\,\mathbb{P}(X_2 = x_2)$. Luego

$$\begin{aligned}\mathbb{E}(X_1X_2) &= \sum_{(x_1,x_2)}(X_1X_2)(\omega)\mathbb{P}(X_1 = x_1, X_2 = x_2)\\ &= \sum_{(x_1,x_2)}X_1(x_1)X_2(x_2)\mathbb{P}(X_1 = x_2)\,\mathbb{P}(X_2 = x_2)\\ &= \Big(\sum_{x_1}X_1(x_1)\,\mathbb{P}(x_1)\Big)\Big(\sum_{x_2}X_2(x_2)\,\mathbb{P}(x_2)\Big) = \mathbb{E}(X_1)\,\mathbb{E}(X_2).\end{aligned}$$

Ejemplo

a) ¿Cuál es la esperanza de una v.a. constante?

b) Calcular $\mathbb{E}(3X + 2\mathbb{E}X)$.

<u>Solución</u>:

a) $\mathbb{E}(X) = \sum_x x\,\mathbb{P}(X = x) = c\,\mathbb{P}(X = c) = c(1) = c.$

b) $\mathbb{E}(3X + 2\mathbb{E}X) = 3\mathbb{E}(X) + 2\mathbb{E}(\mathbb{E}X) = 3\mathbb{E}(X) + 2\mathbb{E}(X) = 5\mathbb{E}(X).$

Esperanza

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
$J'(c) = 2\sum_{i=1}^n p_i(x_i - c) = 0.$
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 $\sum_{i=1}^n |x_i - c|_1$ es: la *mediana* de X.
- El valor que minimiza $\sum_{i=1}^n |x_i - c|_0$ es: la *moda* de X.

UNIVERSIDAD DEL VALLE DE GUATEMALA

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

Proposición

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

Proposición

1. *Para cualesquiera v.a. X y Y se cumple que* $\mathbb{E}X = \mathbb{E}(\mathbb{E}(X \mid Y = y))$.
2. *Para Para cualquier partición* $\{A_i\}$ *de* Ω, *vale* $\mathbb{E}X = \sum_i \mathbb{E}(X \mid A_i)\,\mathbb{P}(A_i)$.

Prueba: Ejercicio!

Ejemplo

Ejemplo

Definición

*Sea X una v.a. en $\mathbb{R}$. Definimos su **varianza** como:*

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

en caso de que este valor esperado exista.

Propiedades:

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

$$Var(aX_2 + bX_2) = a^2 Var(X_1) + b^2 Var(X_2).$$

UNIVERSIDAD DEL VALLE DE GUATEMALA

Prueba:

- $$Var(X) = \mathbb{E}\big((X - \mathbb{E}X)^2\big) = \sum_x (\cdot)^2 \mathbb{P}(\cdot) \geq 0,$$

  por ser suma de términos no-negativos.

- $$\begin{aligned} Var(aX) &= \mathbb{E}\big((aX - \mathbb{E}(aX)^2\big) = \mathbb{E}\big((aX - a\mathbb{E}X)^2\big) \\ &= \mathbb{E}\big(a^2(X - \mathbb{E}X)^2\big) = a^2\mathbb{E}\big((X - \mathbb{E}X)^2\big) \\ &= a^2 Var(X). \end{aligned}$$

Prueba: Suponga que X_1, X_2 son independientes. Entonces, $\mathbb{E}(X_1 X_2) = \mathbb{E}(X_1)\,\mathbb{E}(X_2)$. Luego

$$
\begin{aligned}
Var(aX_1 + bX_2) &= \mathbb{E}\big([(aX_1 + bX_2) - \mathbb{E}(aX_1 + bX_2)]^2\big) \\
&= \mathbb{E}\big([a(X_1 - \mathbb{E}X_1) + b(X_2 - \mathbb{E}X_2)]^2\big) \\
&= \mathbb{E}\big(a^2(X_1 - \mathbb{E}X_1)^2 + b^2(X_2 - \mathbb{E}X_2)^2 + 2ab(X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)\big) \\
&= a^2\mathbb{E}\big((X_1 - \mathbb{E}X_1)^2\big) + b^2\mathbb{E}\big((X_2 - \mathbb{E}X_2)^2\big) + 2ab\mathbb{E}\big((X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)\big) \\
&= a^2 Var(X_1) + b^2 Var(X_2) + 2ab\mathbb{E}\big(X_1X_2 - X_1\mathbb{E}X_2 - (X_2\mathbb{E}X_1 + (\mathbb{E}x_1)(\mathbb{E}X_2)\big) \\
&= a^2 Var(X_1) + b^2 Var(X_2) + 2ab\big(\mathbb{E}(X_1X_2) - (\mathbb{E}x_1)(\mathbb{E}X_2) - (\mathbb{E}x_1)(\mathbb{E}X_2) + (\mathbb{E}x_1)(\mathbb{E}X_2)\big) \\
&= a^2 Var(X_1) + b^2 Var(X_2) + 2ab\big(\mathbb{E}(X_1X_2) - (\mathbb{E}x_1)(\mathbb{E}X_2)\big) = a^2 Var(X_1) + b^2 Var(X_2).
\end{aligned}
$$

Definición

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

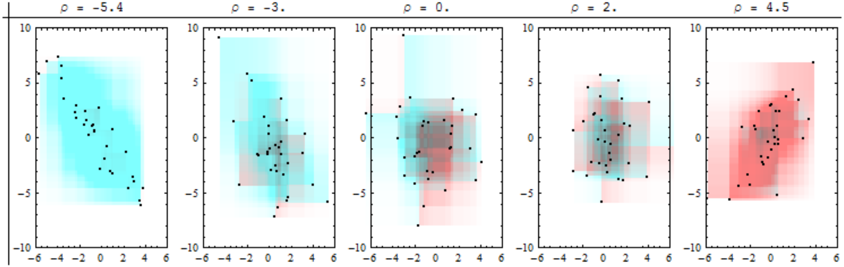
$$Cov(X_1, X_2) = \mathbb{E}\Big[(X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2)\Big],$$

en caso de que este valor esperado exista.

Propiedades:

- $Cov(X_1, X_2) = Cov(X_2, X_1)$.
- $Cov(aX, bY) = abCov(X, Y)$.
- $Cov(aX, X) = aVar(X)$.
- Si X_1, X_2 son independientes, entonces $Cov(X_1, X_2) = 0$.

Covarianza



Definición

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

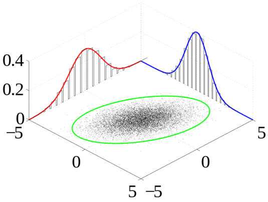
$$\rho(X,Y) = \frac{Cov(X_1,X_2)}{\sqrt{Var(X)\,Var(Y)}}.$$

Propiedades:

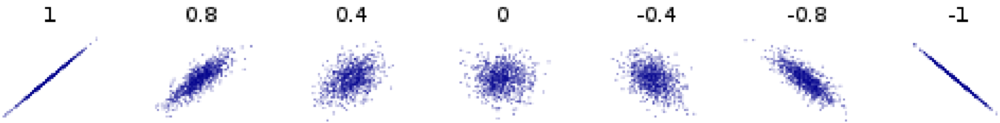
- $\rho(X,Y) = \rho(Y,X)$.
- $-1 \leq \rho(X,Y) \leq 1$.
- $\rho(aX,bY) = \rho(X,Y)$.
- $\rho(aX,X) = (a)$.
- Si X, Y son independientes, entonces $\rho(X,Y) = 0$.

Correlación

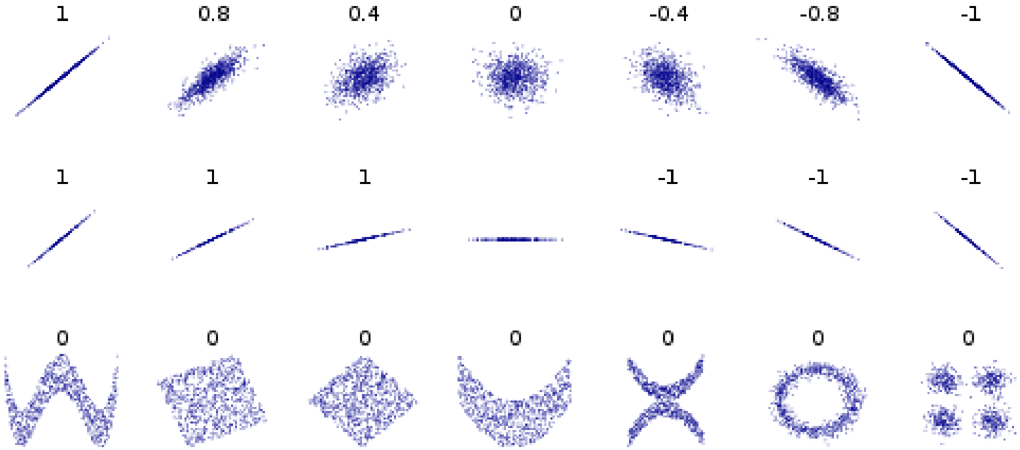
Por ejemplo, para el caso de dos v.a. normales X y Y:



tenemos



Correlación



Sorpresa!

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

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

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:*

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

Ejercicios

1. Dibuja dos distribuciones o variables aleatorias (discretas) distintas, con mismo promedio y entropía, pero varianza diferentes.

2. Toma una v.a. $X \in \{0, 1\}$. Calcular la varianza y la entropía de Shannnon y de Gini en función de $p = \mathbb{P}(X = 1)$.

   Compara la gráficas de $H(X)$ y $2G(X)$.

3. ¿Cuáles son los valores mínimo y máximo para $H(X)$ y $G(X)$?

Variables aleatorias discretas

- distribución Uniforme $U[a..b]$,
- distribución Bernoulli $Ber(p)$,
- distribución Binomial $Binom(n, p)$,
- distribución Geométrica $Geom(p)$,
- distribución Poisson $Poisson(\lambda)$,

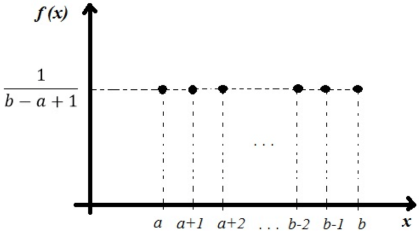
- distribución Rademacher $Rad(p)$,
- distribución Binomial Negativa $NB(r, p)$,
- distribución Hipergeométrica $Hypergeometric(N, K, n)$.

1. Distribución Uniforme

$$X \sim U[a..b] \;\Leftrightarrow\; \mathbb{P}(X=k) = \tfrac{1}{b-a+1}, \quad \text{para } k = a, a+1, \ldots, b.$$

Obs.

- Esta distribución depende de dos parámetros (de localización): a y b.
- El caso $a=b$, con $\mathbb{P}(X=a=b)=1$ se llama una v.a. *degenerada*.

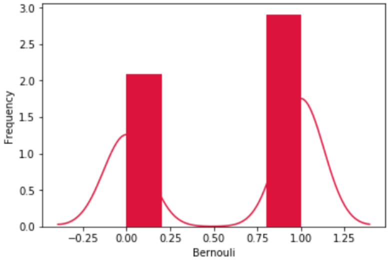


2. Distribución Bernoulli

$$X \sim Ber(p) \;\Leftrightarrow\; \mathbb{P}(X=1) = p,\; \mathbb{P}(X=0) = 1-p, \quad \text{para } 0 \leq p \leq 1.$$

Obs.

- La distribución es simétrica si, y sólo si, $p = 1/2$.
- $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$.



Variables aleatorias discretas

La distribución Bernoulli tiene una hermana gemela: la distribución de Rademacher.

$$X \sim Rad(p) \;\Leftrightarrow\; \mathbb{P}(X=1)=p,\; \mathbb{P}(X=-1)=1-p, \;\; \text{para } 0 \le p \le 1.$$

Preguntas:

- ¿Cuál es la media y varianza de la distribución $Rad(p)$.
- Sean X, Y v.a., con $X \sim Ber(p)$ y $Y \sim Rad(p)$. Escribir X en términos de Y, y Y en términos de X.

La distribución de Bernoulli es importante para escribir situaciones donde se cuenta la ocurrencia de eventos. La variable $X \sim Ber(p)$ cuenta o indica la ocurrencia del evento de interés.

3. Distribución Binomial

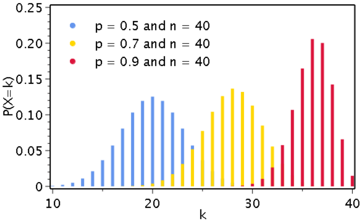
$$X \sim Binom(n,p) \;\Leftrightarrow\; \mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{para } k = 0, 1, \ldots, n.$$

Interpretación: Si $\{X_i\}_{i=1}$ son v.a. *i.i.d.* con $X_i \sim Ber(p)$, entonces

$$X = \sum_{i=1}^{n} X_i \sim Binom(n,p).$$

Obs.

- La distribución es simétrica si, y sólo si, $p = 1/2$.
- $\mathbb{E}(X) = np$, $Var(X) = np(1-p)$.



4. Distribución Geométrica

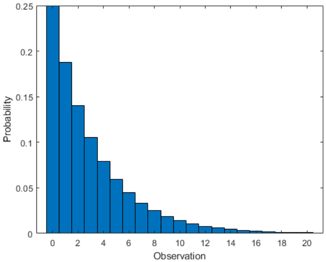
$$X \sim Geom(n,p) \;\Leftrightarrow\; \mathbb{P}(X=k) = p(1-p)^{k-1}, \quad \text{para } k = 1,2,3,\ldots$$

Interpretación: Si $\{X_i\}_{i=1}$ son v.a. *i.i.d.* con $X_i \sim Ber(p)$, entonces

$$X = \text{el momento del primer éxito en } \{X_i\} \sim Geom(p).$$

Obs.

- La probabilidad va decayendo en forma geométrica con k.
- $\mathbb{E}(X) = \frac{1}{p}$.



5. Distribución Poisson

$$X \sim Poisson(\lambda) \;\Leftrightarrow\; \mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \text{para } k = 0, 1, 2, \ldots$$

Interpretación: Si $\{X_i\}_{i=1}$ son v.a. *i.i.d.* con $X_i \sim Ber(p)$, entonces

$$X = \text{el momento del primer éxito en } \{X_i\} \sim Geom(p).$$

Obs.

- Cuenta el número de llegadas de un proceso con tiempos exponenciales $Exp(\lambda)$.
- $\mathbb{E}(X) = \lambda$. Representa el número esperado de veces que ocurra el fenómeno durante un intervalo dado.