

Repaso probabilidad

Álvaro José Álvarez Arranz

02/01/2024

Espacios de probabilidad

Para empezar, tenemos que saber qué es un espacio de probabilidad.

El espacio muestral es el conjunto de todos los posibles resultados individuales de un experimento aleatorio. Cada elemento en el espacio se denomina “punto muestral” y representa un posible resultado experimental.

Por ejemplo, consideremos un dado. El espacio muestral de este caso será $\Omega = \{1, 2, 3, 4, 5, 6\}$ ya que esos son los posibles resultados individuales que podrían ocurrir. Otro ejemplo sería el de lanzar dos monedas, cuyo espacio muestral sería $\Omega = ((cara, cara), (cara, cruz), (cruz, cara), (cruz, cruz))$

Independencia de sucesos

Se dice que dos sucesos son **independientes** si y sólo si:

$$P(A \cap B) = P(A)P(B)$$

Ejemplo:

Consideremos el lanzamiento de dos dados justos. Sea A el suceso de obtener un 4 en el primer dado y B el suceso de obtener un 3 en el segundo dado. Estos sucesos son independientes porque la probabilidad de obtener un 4 en el primer dado no afecta la probabilidad de obtener un 3 en el segundo, y viceversa.

La probabilidad de obtener un 4 en el primer dado es $P(A) = \frac{1}{6}$, dado que hay un 4 en los 6 posibles resultados. Lo mismo pasa con el otro dado que ha obtenido un 3: $P(B) = \frac{1}{6}$.

La probabilidad conjunta de que ambos sucesos ocurran es el producto de las probabilidades individuales.

$$P(A \cap B) = P(A)P(B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Probabilidad condicionada

Asumiendo que B tiene probabilidad positiva, la **probabilidad condicionada** de A con respecto a B se define como:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Con esta definición, A y B son independientes si y solo si:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \stackrel{\text{ind}}{=} \frac{P(A)P(B)}{P(B)} = P(A)$$

Regla de la probabilidad total

Sea A_1, \dots, A_k una partición de Ω . Para cualquier suceso B se cumple:

$$P(B) = \sum_{i=1}^k P(B|A_i)P(A_i)$$

Teorema de Bayes

Sea A_1, \dots, A_k una partición de Ω tal que $P(A_i) > 0$ para todo i . Si $P(B) > 0$ entonces,

$$P(A_i|B) = \frac{P(B|A_i) P(A_i)}{\sum_{i=1}^k P(B|A_i) P(A_i)}$$

$P(A_i)$ recibe el nombre de probabilidad **a priori** y $P(A_i|B)$ **a posteriori**.

Variables aleatorias

Una **variable aleatoria** (v.a.) X es una función (medible) $X : \Omega \rightarrow \mathbb{R}$ que asigna un número real $X(\omega)$ a cada resultado ω de un experimento aleatorio.

La **función de distribución** $F : \mathbb{R} \rightarrow [0, 1]$ de una v.a. X es:

$$F(x) = P(X \leq x)$$

- Si X es el número de caras al tirar una moneda dos veces, determinar su función de distribución.

Solución:

Para determinar la función de distribución de X , primero necesitamos enumerar todas las posibles combinaciones de resultados.

Cuando se tira una moneda dos veces, hay cuatro posibles resultados para cada lanzamiento:

1. cara-cara (CC)
2. cara-cruz (CS)
3. cruz-cara (SC)
4. cruz-cruz (SS)

Ahora se determina el valor de X para cada una de esas combinaciones.

1. CC: $X=2$
2. CS: $X=1$
3. SC: $X=1$
4. SS: $X=0$

La función de distribución de X asigna probabilidades a estos valores. Dado que cada lanzamiento de la moneda es independiente y tiene una probabilidad de 0.5 para cara y 0.5 para cruz, podemos calcular las probabilidades para cada valor de X :

1. $P(X = 2) = P(CC) = 0.5 * 0.5 = 0.25$
2. $P(X = 1) = P(CS) + P(SC) = 0.5 * 0.5 + 0.5 * 0.5 = 0.5$
3. $P(X = 0) = P(SS) = 0.5 * 0.5 = 0.25$

- La función de distribución determina completamente la distribución de una v.a.
- F es monótona no decreciente, continua por la derecha, $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$

Funciones	Prefijos
Función de distribución	‘p’
Función cuantílica	‘q’
Función de densidad (continuas) o de probabilidad (discretas)	‘d’
Generación de números aleatorios	‘r’

Table 1: Ejemplo de tabla en LaTeX

Variables discretas y continuas

Una v.a. es **discreta** si toma un número numerable de valores x_1, x_2, \dots . Su función de probabilidad es: $p(x_i) = p_i = P(X = x_i)$, $i = 1, 2, \dots$.

Una v.a. es continua si existe una función $f(x) \geq 0$ con $\int f(x)dx = 1$ tal que

$$p(a < X < b) = \int_a^b f(x)dx$$

- Se dice que f es la **función de densidad** de X .
- Se verifica $F(x) = \int_{-\infty}^x f(t)dt$ y $F'(x) = f(x)$ si f es continua en x .

Distribuciones de probabilidad con R

- Cada distribución se nombra mediante una palabra clave o *alias*.
- Para una lista completa de las distribuciones se usa `help("Distributions")`
- A cada distribución se le antepone un prefijo que determina una función relacionada con ella

Ejemplos:

- Calcular la probabilidad de que una variable normal estandar sea mayor que 1.

```
library(tidyverse)
```

```
## Warning: package 'readr' was built under R version 4.3.2
```

```
## Warning: package 'dplyr' was built under R version 4.3.2
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.4      v readr      2.1.4
```

```
## v forcats   1.0.0      v stringr   1.5.0
```

```
## v ggplot2   3.4.3      v tibble    3.2.1
```

```
## v lubridate 1.9.2      v tidyr     1.3.0
```

```
## v purrr     1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
1 - pnorm(1)
```

```
## [1] 0.1586553
```

- Determinar el valor x tal que $P(Z > x) = 0.3$, donde Z es una v.a. con distribución normal estandar (es decir, el percentil 70 de la distribución)

```
qnorm(0.7)
```

```
## [1] 0.5244005
```

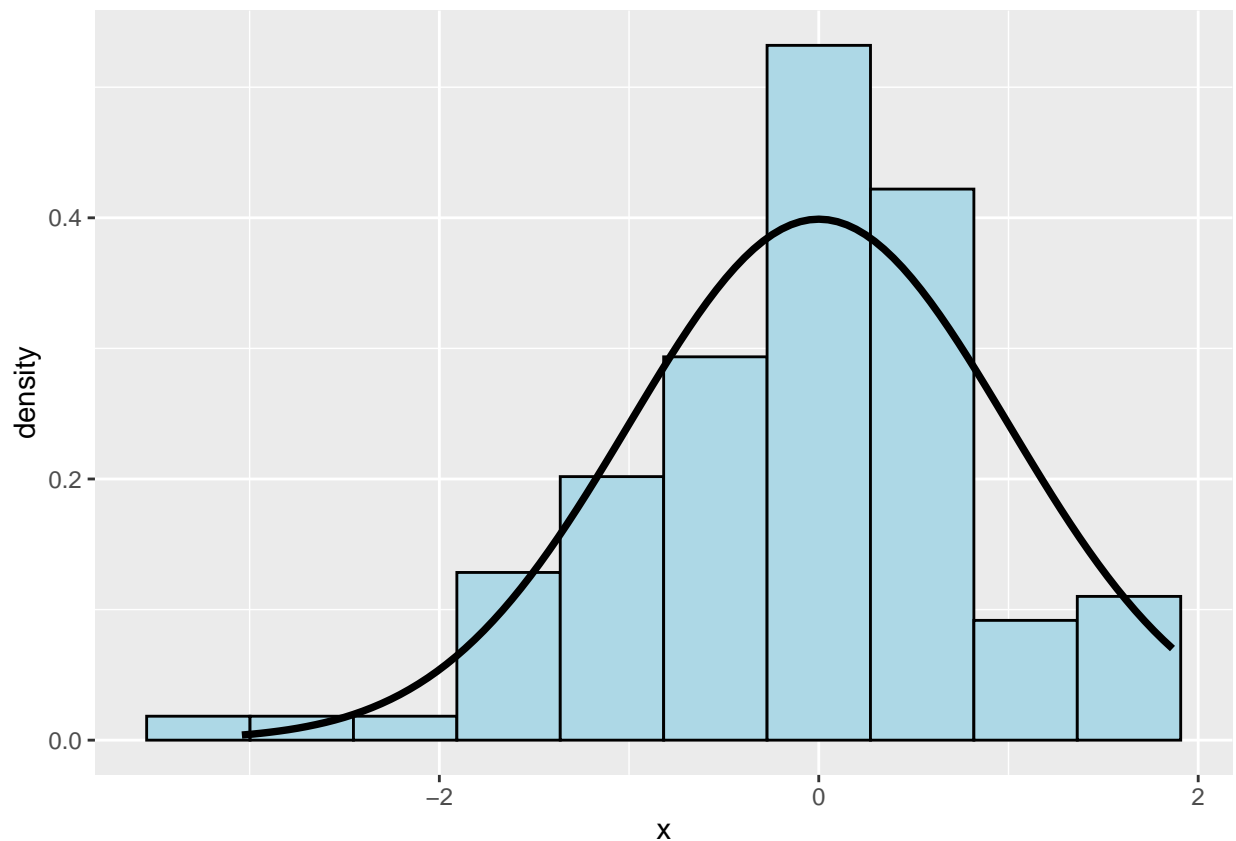
- Calcular la mediana de una variable exponencial de media 2.

```
qexp(0.5, rate=0.5)
```

```
## [1] 1.386294
```

- Generar 100 valores de una distribución normal y representar el histograma con `ggplot2`. Añadir la curva de densidad teórica.

```
library(ggplot2)
set.seed(99)
x = rnorm(100)
df = data.frame(x=x)
ggplot(df)+
  geom_histogram(aes(x=x, y=after_stat(density)),
                 col='black', fill='lightblue', bins=10)+
  geom_function(fun=dnorm, col='black', linewidth=1.3)
```



Esperanza

La **esperanza** (o media) de una v.a. X es $E(X) := \mu = \int x dF(x)$. Esta notación significa:

- $E(X) = \sum_i x_i P(X = x_i)$, si X es discreta.
- $E(X) = \int x f(x) dx$, si X es continua.
- Es un resumen del centro en torno al cual toma valores X (\approx de muchas observaciones).
- Si $Y = g(X)$, entonces $E(Y) = \int g(x) dF(x)$, lo que significa que no es necesario calcular la distribución de Y , si no que basta con conocer X .
- La esperanza es lineal $E(aX + bY) = aE(X) + bE(Y)$
- Decimos que $E(X)$ existe si $\int |x| dF(x) < \infty$. Entonces, ¿qué ocurre con la distribución de Cauchy que tiene como $f(x) = \frac{1}{\pi(1+x^2)}$?

Solución:

La distribución de Cauchy es conocida por no tener momentos finitos. En el contexto de la integral del enunciado, si $F(x)$ es la función de distribución acumulativa asociada a la distribución de Cauchy, entonces la integral no converge, por lo que $E(X)$ no existe para la distribución de Cauchy.

Varianza y covarianza

Sea X una v.a. con esperanza μ .

La **varianza** de X es $Var(X) := \sigma^2 = E[(X - \mu)^2]$

- La varianza mide la dispersión de los valores que toma X en torno a μ .
- **No** es lineal: $Var(aX + b) = a^2 Var(X)$

La **desviación típica** de X es σ .

Sean X e Y v.a. con esperanzas μ_x y μ_y , y varianzas σ_x^2 y σ_y^2 .

Se define la **covarianza** entre X e Y como $Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] \equiv \sigma_{x,y}$

El **coeficiente de correlación** entre X e Y es

$$\rho(X, Y) = \frac{\sigma_{x,y}}{\sigma_x \sigma_y}$$

- La correlación mide el grado de relación lineal entre X e Y .
- Siempre se cumple que $-1 \leq \rho(X, Y) \leq 1$.

Algunas desigualdades

1. **Desigualdad de Markov:** Sea X una v.a. no negativa con esperanza $E(X) < \infty$, entonces, para todo $\epsilon > 0$ se cumple

$$P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}$$

2. **Desigualdad de Chebychev:** Sea X una v.a. con esperanza μ y varianza σ^2 , entonces, para todo $\epsilon > 0$ se cumple

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

3. **Desigualdad de Cauchy-Swarz:** Dadas dos v.a. X e Y con varianzas finitas, entonces

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$$

4. **Desigualdad de Jensen:** Si g es una función convexa,

$$E[g(X)] \geq g[E(X)]$$

Vectores aleatorios

Se dice que $X = (X_1, \dots, X_p)$ es un **vector aleatorio** si X_i es una v.a. para todo $i = 1, \dots, p$. Las distribuciones de las coordenadas X_i se llaman **distribuciones marginales**.

- El vector aleatorio X es continuo si existe una función de densidad conjunta $f : \mathbb{R}^p \rightarrow [0, \infty]$ tal que

$$P(X \in A) = \int_A f(x) dx$$

para $A \subset \mathbb{R}^p$ (medible).

- Las marginales son independientes si

$$P(X_1 \in A_1, \dots, X_p \in A_p) = \prod_{i=1}^p P(X_i \in A_i)$$

- En el caso continuo, la independencia equivale a que la densidad conjunta es el producto de las p densidades marginales.
- Bajo independencia, $E(X_1, \dots, X_p) = \prod_{i=1}^p E(X_i)$
- Si X e Y son independientes y tienen varianzas finitas, $Cov(X, Y) = 0$. El recíproco **no** es cierto.

Si $X = (X_1, \dots, X_p)'$ un vector aleatorio p -dimensional:

- Su **vector de medias** es $E(X) = \mu = (\mu_1, \dots, \mu_p)'$, donde $\mu_i = E(X_i)$.
- Su **matriz de covarianzas** es $Var(X) := Cov(X) := \Sigma$, cuya posición (i, j) es $\sigma_{i,j} = Cov(X_i, X_j)$.
- Demostración $Var(X) = E(XX') - \mu\mu'$
 1. Expansión de $Var(X)$: $Var(X) = E[(X - \mu)(X - \mu)']$
 2. Expandimos $E[(X - \mu)(X - \mu)']$: $E[(X - \mu)(X - \mu)'] = E(XX') - E(X\mu') - E(\mu X') + E(\mu\mu')$
 3. Propiedades:
 - $E(a) = a$ para cualquier a constante.
 - $E(aX) = aE(X)$ para cualquier a constante.
 - $E(X + Y) = E(X) + E(Y)$ para variables aleatorias X e Y independientes.
 4. Sustitución: $E[(X - \mu)(X - \mu)'] = E(XX') - \mu'E(X) - \mu E(X') + \mu\mu'$
 5. Propiedades de la media: $E(X) = \mu$
 6. Sustitución de la media: $E[(X - \mu)(X - \mu)'] = E(XX') - \mu'\mu - \mu\mu' + \mu\mu'$
 7. Simplificación: $E[(X - \mu)(X - \mu)'] = E(XX') - \mu\mu'$
- Si A es una matriz $q \times p$ y $b \in \mathbb{R}^q$, entonces $E(AX + b) = A\mu + b$ y $Var(AX + b) = E[A(X - \mu)(X - \mu)'A'] = A\Sigma A'$.

Esperanza y varianza condicionadas

Sean X y Y dos v.a. La esperanza y la varianza de Y condicionada a X (es decir, conocemos X) se define:

$$E(Y|X = x) = \int y dF_{Y|X=x}(y)$$

$$Var(Y|X = x) = E[Y - E(Y|X = x)]^2|X = x]$$

- Las distribuciones condicionadas se calculan de manera similar a las probabilidades condicionadas. Por ejemplo, si f es la densidad conjunta de (X, Y) y f_X la densidad marginal de X ,

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}$$

- Tanto $E(Y|X)$ como $Var(X|Y)$ son variables aleatorias (funciones de X) cuando no se condiciona a un valor concreto.
- Si X e Y son independientes, se cumple $E(Y|X) = E(Y)$
- $E(g(X)Y|X) = g(X)E(Y|X)$
- $E(a_1Y_1 + a_2Y_2|X) = a_1E(Y_1|X) + a_2E(Y_2|X)$
- Ley de la esperanza iterada: $E(Y) = E[E(Y|X)]$
- El valor $E(Y|X)$ da la mejor predicción posible de Y a partir de X : Para cualquier g ,

$$E[(Y - g(x))^2] \geq E[(Y - E(Y|X))^2]$$

- Una identidad útil para la varianza:

$$Var(Y) = E[Var(Y|X)] + Var[E(Y|X)]$$

Distribución normal multivariante

El vector aleatorio X es **normal p -dimensional** con vector de medias μ y matriz de covarianzas Σ si su densidad es:

$$f(x) = |\Sigma|^{-1/2} (2\pi)^{-p/2} \exp\left\{-\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu)\right\}, x \in \mathbb{R}^p$$

- ¿Qué pasa si $\mu = 0$ y $\Sigma = \mathbb{I}$?

Solución:

- Sustituimos en la expresión y obtenemos:

$$f(x) = |\mathbb{I}_p|^{-1/2} (2\pi)^{-p/2} \exp\left\{-\frac{1}{2}x' \mathbb{I}_p^{-1}x\right\}$$

- Dado que la matriz identidad \mathbb{I}_p es invertible y su determinante es 1, la expresión se simplifica aún más:

$$f(x) = (2\pi)^{-p/2} \exp\left\{-\frac{1}{2}x'x\right\}$$

- Se usa la notación $X \equiv N_p(\mu, \Sigma)$.
- si $X \equiv N_p(\mu, \Sigma)$, A es matriz $q \times p$ y $b \in \mathbb{R}^q$, entonces

$$AX + b \equiv N_q(A\mu + b, A\Sigma A')$$

Sea $X \equiv N_p(\mu, \Sigma)$. Consideramos la partición $X = (X_a, X_b)$, con $X_a \in \mathbb{R}^q$ y $X_b \in \mathbb{R}^{p-q}$ y consideramos las particiones de $\mu = (\mu_a, \mu_b)$ y $\Sigma = [[\Sigma_{aa}, \Sigma_{ab}], [\Sigma_{ba}, \Sigma_{bb}]]$

- $X_a \equiv N_q(\mu_a, \Sigma_{aa})$
- X_a y X_b son independientes si y solo si $\Sigma_{ab} = 0$.
- Distribución condicionada:

$$X_a|(X_b = x) \equiv N_q(\mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x - \mu_b), \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})$$

- Si se aplica la propiedad anterior a un vector normal bidimensional (X, Y) , se obtiene la distribución condicionada $Y|X$.