Análisis Estadístico Multivariante

Francisco Javier Mercader Martínez

Índice

Tema 1: Vectores aleatorios

1.1) Introducción

Objetivo: estudiar k variables sobre una población de individuos (objetos).

Algunos ejemplos:

→ Las variables meteorológicas como temperatura, humedad y velocidad del viento.

→ La intensidad y la fase de una señal aleatoria que se miden en los canales de comunicación.

→ Los parámetros clínicos de los pacientes (como presión arterial, niveles de glucosa, etc.)

Habitualmente estas variables cualitativas o discretas que nos indicarán grupos de individuos.

Estas variables se representarán mediante vectores aleatorios sobre un espacio de probabilidad.

1) Definiciones

Un vector aleatorio (v.a.) k-dimensional sobre un espacio de probabilidad $(\Omega, \mathcal{S}, \mathcal{P})$ es $X = (X_1, \ldots, X_k)$ tal que

$$X_i^{-1}(-\infty, x] \in \mathcal{S}$$

para todo $x \in \mathbb{R}$, $i = 1, \ldots, k$

- Función de distribución conjunta

$F : \mathbb{R}^k \longrightarrow [0, 1]$,

$$F(x_1, \ldots, x_k) \coloneqq P[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k],$$

para todo $x_1, \ldots, x_k \in \mathbb{R}$.

1.2) Independencia de las variables aleatorias

- Definición

Las variables aleatorias $X_1, \ldots, X_k$ son independientes si los sucesos

$$\{x_1 \leq x_1\}, \{X_2 \leq x_2\}, \ldots, \{X_k \leq x_k\}$$

son independientes para todo $x_1, \ldots, x_k \in \mathbb{R}$.

Esto es equivalente a que

$$F(x_1, \ldots, x_k) = P[X_1 \leq x_1] \cdot P[X_2 \leq x_2] \cdots P[X_k \leq x_k]$$

para todo $x_1, \ldots, x_k \in \mathbb{R}$.

1.3) Distribuciones marginales

La función $F_{X_i}(x_i) = P[X_i \leq x_i]$ se denomina función de distribución marginal i-ésima y corresponde con la función de distribución de la variable aleatoria X_i

Las distribuciones marginales pueden obtenerse a partir de la distribución conjunta:

$$F_{X_i}(x_I) = F(+\infty, \ldots, +\infty, x_i, +\infty, \ldots, +\infty)$$

Análogamente, la función de distribución marginal del subvector aleatorio $(X_{i_1}, \ldots, X_{i_m})$ vendrá dada por

$$F_{X_{i_1},\ldots,X_{i_m}}(x_{i_1}, \ldots, x_{i_m}) = F(+\infty, \ldots, +\infty, x_{i_1}, +\infty, \ldots, +\infty, x_{i_m}, +\infty, \ldots, +\infty).$$

1.4) Vector aleatorio absolutamente continuo

Un vector aleatorio X es absolutamente continuo si existe una función $f : \mathbb{R}^k \longrightarrow \mathbb{R}$ no negativa (llamada función de densidad) tal que

$$F(x) = F(x_1, \ldots, x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(z_1, \ldots, z_k) \mathrm{d}z_k, \ldots, \mathrm{d}z_1,$$

para todo $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$

Usando el teorema fundamental del cálculo, se tiene que en cada punto de continuidad $(x_1, \ldots, x_k)$ de f:

$$\frac{\partial^k F(x_1, \ldots, x_k)}{\partial x_1, \ldots, \partial x_k} = f(x_1, \ldots, x_k).$$

Existen variables aleatorias cuya función de distribución es continua pero que no son absolutamente continuas (tienen una parte singular) y puede ocurrir que $X_1, \ldots, X_k$ sean absolutamente continuas y que $(X_1, \ldots, X_k)$ no lo sea.

→ Ejemplo: Si X_1 es una variable aleatoria absolutamente continua, entonces el vector aleatorio $X = (X_1, X_2)$ es continuo pero no absolutamente continuo.

→ De hecho, es completamente singular ya que está contenido en la recta $y = x$ que tiene medida cero en $\mathbb{R}^2$.

Esto ocurre si consideramos las notas de unos alumnos y sus medidas. En estos casos deberemos eliminar estas variables dependientes del vector.

1.5) Vector aleatorio discreto

Un vector aleatorio X se dice que es discreto si existe un conjunto numerable $\mathcal{S} \in \mathbb{R}^k$ tal que $P(X \in \mathcal{S}) = 1$.

Función masa de probabilidad de una vector aleatorio discreto:

$$P[X = x] = P[X_1 = x_1, \ldots, X_k = x_k]$$

para todo $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$, satisfaciendo:

→ $P[X = x] \geq 0, \ \forall x \in \mathcal{S}$

→ $\displaystyle\sum_{x \in \mathcal{S}} P[X = x] = 1$

Función de distribución de un vector aleatorio discreto:

$$F(x) = P[X \leq x] = \sum_{\substack{z \in \mathcal{S} \\ z \leq x}} P[X = z],$$

para todo $x \in \mathbb{R}^k$.

1.6) Distribuciones marginales

1.6.1) Caso continuo

- Distribución marginal de la variable aleatoria X_i

Sea $X = (X_1, \dots, X_k)$ un vector aleatorio continuo con función de densidad f entonces cada componente X_i es de tipo continuo y su función de distribución es;

$$F_{X_i}(x_i) = P[X_i \leq x_i] = \int_{-\infty}^{x_i} f_{X_i}(z_i)\mathrm{d}z_i,$$

con

$$f_{X_i} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(z_1, \dots, z_k)\mathrm{d}z_1, \dots, \mathrm{d}z_{i-1} \cdot \mathrm{d}z_{i+1} \dots, \mathrm{d}z_k,$$

para todo $z_i \in \mathbb{R}$.

La función de densidad marginal de cualquier subvector se calcularía de igual forma.

$X_1, \dots, X_k$ son independientes $\longleftrightarrow f(x_1, \dots, x_k) = f_{X_1}(x_1) \cdots f_{X_k}(x_k)$.

1.6.2) Caso discreto

- Distribución marginal de la variable aleatoria X_i

Sea $X = (X_1, \dots, X_l)$ un vector aleatorio discreto con $P[X \in \mathcal{S}] = 1$ y función masa de probabilidad $P[X = x]$, para todo $x \in \mathcal{S}$.

Si X_i es una componente arbitraria y por tanto discreta con valores en $\mathcal{S}_i$, entonces su función masa de probabilidad puede obtenerse a partir de la conjunta:

$$P[X_i = x_i] = \sum_{\substack{x_1 \dots, x_{i-1}, x_{i+1}, \dots, x_k \\ (x_1, \dots, x_i, \dots, x_n) \in \mathcal{S}}} P[X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x_i, X_{i+1} = x_{i+1}, \dots, X_k = x_k].$$

La función masa de probabilidad marginal de cualquier subvector se calcularía de igual forma.

$X_1, \dots, X_k$ son independientes $\longleftrightarrow$ para todo $(x_1, \dots, x_k) \in \mathcal{S}$,

$$P[X_1 = x_1, \dots, X_k = x_k] = P[X_1 = x_1] \cdots P[X_k = x_k].$$

Nota:
A y B independientes $\longleftrightarrow P(A|B) = P(A)$
$P(A|B) = \dfrac{P(A \cap B)}{P(B)} = P(A) \longrightarrow P(A \cap B) = P(A) \cdot P(B)$

1.7) Distribuciones condicionadas

1.7.1) Caso continuo

- Distribución condicionada al valor de una variable

Sea $X = (X_1, \dots, X_k)$ un vector aleatorio continuo con función de densidad f.
Sea X_i una componente arbitraria y $x_i^* \in \mathbb{R}$ tal que $f_{X_i}(x_i^*) > 0$.
Se define la distribución condicionada de $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k)$ a $(X_i = x_i^*)$ como la determinada por la función de densidad:

$$f_{X_1, \dots, X_{i-1}, \dots, X_k | X_i = x_i^*}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k | x_i^*) = \frac{f(x_1, \dots, x^*, \dots, x_k)}{f_{X_i}(x_i^*)}.$$

- Distribución condicionada a valores de varias variables

Sea $X = (X_1, \dots, X_k)$ un vector aleatorio continuo con función de densidad f.
Sea $(X_{i_1}, \dots, X_{i_m})$ un subvector arbitrario y $(x_{i_1}^*, \dots, x_{i_m}^*) \in \mathbb{R}^m$ tal que:

$$f_{X_{i_1},\dots,X_{i_m}}(x_{i_1}^*, \dots, x_{i_m}^*) > 0.$$

Se define la distribución condicionada de $(X_1, \dots, X_{i_1-1}, X_{i_1+1}, \dots, X_{i_m-1}; X_{i_m+1}, \dots, X_k)$ a $(X_{i_1} = x_{i_1}^*, \dots, X_{i_m} = x_{i_m^*})$ como la determinada por la función de densidad:

$$f_{X_1,\dots,X_{i_1-1},X_{i_1+1},\dots,X_{i_m-1},\dots,X_k|X_{i_1}=x_{i_1}^*,\dots,X_{i_m}=x_{i_m^*}}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{i_m-1}, x_{i_m+1}, \dots, x_k|x_i^*) = \frac{f(x_1, \dots, x_{i_1}^*, \dots, x_{i_m}^* \dots, x_k)}{f_{X_{i_1},\dots,X_{i_m}}(x_{i_1}^*, \dots, x_{i_m}^*)}$$

1.7.2) Caso discreto

- Distribución condicionada al valor de una variable

Sea $X = (X_1, \dots, X_k)$ un vector aleatorio discreto.
Sea X_i una componente arbitraria y $x_i^* \in \mathbb{R}$ tal que

$$P[X_i = x_i^*] > 0.$$

Se define la distribución condicionada de $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k)$ a $(X_i = x_i^*)$ como la determinada por la función masa de probabilidad:

$$P[X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \dots, X_k = x_k | X_i = x_i^*] = \frac{P[X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x_i^*, X_{i+1} = x_{i+1}, \dots, X_k = x_k]}{P[X_i = x_i^*]}$$

para todo $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$ tal que $x_1, \dots, x_{i-1}, x_i^*, x_{i+1}, \dots, x_k \in \mathcal{S}$.

- Distribución condicionada a valores de varias variables

Sea $X = (X_1, \dots, X_k)$ un vector aleatorio discreto.

Sea $X_{i_1}, \dots, X_{i_m}$ un subvector arbitrario y $(x_{i_1}^*, \dots, x_{i_m}^*) \in \mathbb{R}^m$ tal que

$$P[X_{i_1} = x_{i_1}^*, \dots, X_{i_m} = x_{i_m}^*] > 0.$$

Se define la distribución condicionada de $(X_1, \dots, X_{i_1-1}, X_{i_1+1}, \dots, X_{i_m-1}, X_{i_m+1}, \dots, X_k)$ a $(X_{i_1} = x_{i_1}^*, \dots, X_{i_m} = x_{i_m}^*)$ como la determinada por la función masa de probabilidad:

$$P[X_1 = x_1, \dots, X_{i_1-1} = x_{i_1-1}, X_{i_1+1} = x_{i_1+1}, \dots, X_{i_m-1} = x_{i_m-1}, X_{i_m+1} = x_{i_m+1}, \dots, X_k = x_k | X_{i_1} = x_{i_1}^*, \dots, X_{i_m} = x_{i_m^*}] = \frac{P[X_1 = x_1, \dots, X_{i_1} = x_{i_1}^*, \dots, X_{i_m} = x_{i_m}^*, \dots, X_k = x_k]}{P[X_{i_1} = x_{i_1}^*, \dots, X_{i_m} = x_{i_m}^*]}$$

para todo $(x_1, \dots, x_{i_1}, x_{i_1+1}, \dots, x_{i_m-1}, x_{i_m+1}, \dots, x_k)$, tal que $(x_1, \dots, x_{i_1}^*, \dots, x_{i_m}^*, \dots, x_k) \in \mathcal{S}$

1.8) Distribución normal multivariante $\mathcal{N}_k(\mu, V)$

1) Función de densidad

$$f(x) = \frac{1}{\sqrt{|V|(2\pi)^k}} \exp\left(-\frac{1}{2}(x-\mu)'V^{-1}(x-\mu)\right),$$

para $x \in \mathbb{R}^k$, donde μ es un vector k-dimensional y V es una matriz $k \times k$ simétrica y definida positiva.

- Definiciones

Una matriz simétrica A, de dimensión $k \times k$, se dice que es definida positiva si se verifica que $x'Ax > 0$ para cualquier vector no nulo $x \in \mathbb{R}^k$.

Una matriz simétrica A, de dimensión $k \times k$, se dice que es semidefinida positiva si se verifica que $x'Ax \geq 0$ para cualquier vector $x \in \mathbb{R}^k$.

¿Cómo calcular la inversa de $V = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$ con R?

```
V <- matrix(c(1, 1/2,
            1/2, 1), nrow = 2, ncol = 2, byrow = TRUE)
solve(V)
```

```
##            [,1]       [,2]
## [1,]  1.3333333 -0.6666667
## [2,] -0.6666667  1.3333333
```

1.8.1) Normal bivariante

- Función de densidad

Caso bivariante, $k = 2$, para $\mu = (0, 0)$ y $V = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$.

Cálculo de la función de densidad en $x = (1, 1)$ utilizando la función `dmvnorm` de la librería `mvtnorm` de R:

```
library("mvtnorm")
V <- matrix(c(1, 1/2,
              1/2, 1), nrow = 2, ncol = 2, byrow = TRUE)
mu <- c(0, 0)
x <- c(1, 1)
dmvnorm(x, mean = mu, sigma = V)
```

```
## [1] 0.0943539
```

- Función de distribución

Cálculo (aproximado) de la función de distribución en $x = (1, 1)$ con la función:

`pmvnorm(lower = -Inf, upper = x, mean = mu, sigma = V)`

```
library("mvtnorm")
V <- matrix(c(1, 1/2,
              1/2, 1), nrow = 2, ncol = 2, byrow = TRUE)
mu <- c(0, 0)
x <- c(1, 1)
pmvnorm(lower = -Inf, upper = x, mean = mu, sigma = V)
```

```
## [1] 0.7452036
```

- Probabilidad en rectángulos

Cálculo (aproximado) de las probabilidades en rectángulos dando los límites inferiores y superiores del rectángulo. Por ejemplo, para calcular

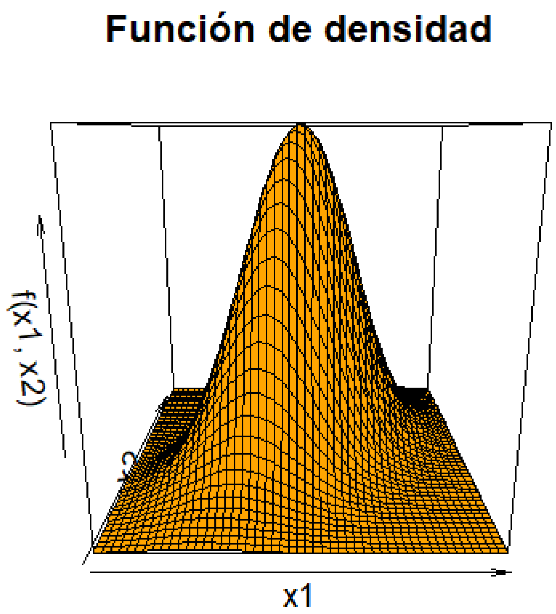
$$P(-1 < X_1 < 1, \ -1 < X_2 < 1)$$

```r
library("mvtnorm")
V <- matrix(c(1, 1/2,
              1/2, 1), nrow = 2, ncol = 2, byrow = TRUE)
mu <- c(0, 0)
x1 <- c(- 1, -1)
x2 <- c(1, 1)
pmvnorm(lower = x1, upper = x2, mean = mu, sigma = V)
```

```
## [1] 0.499718
```

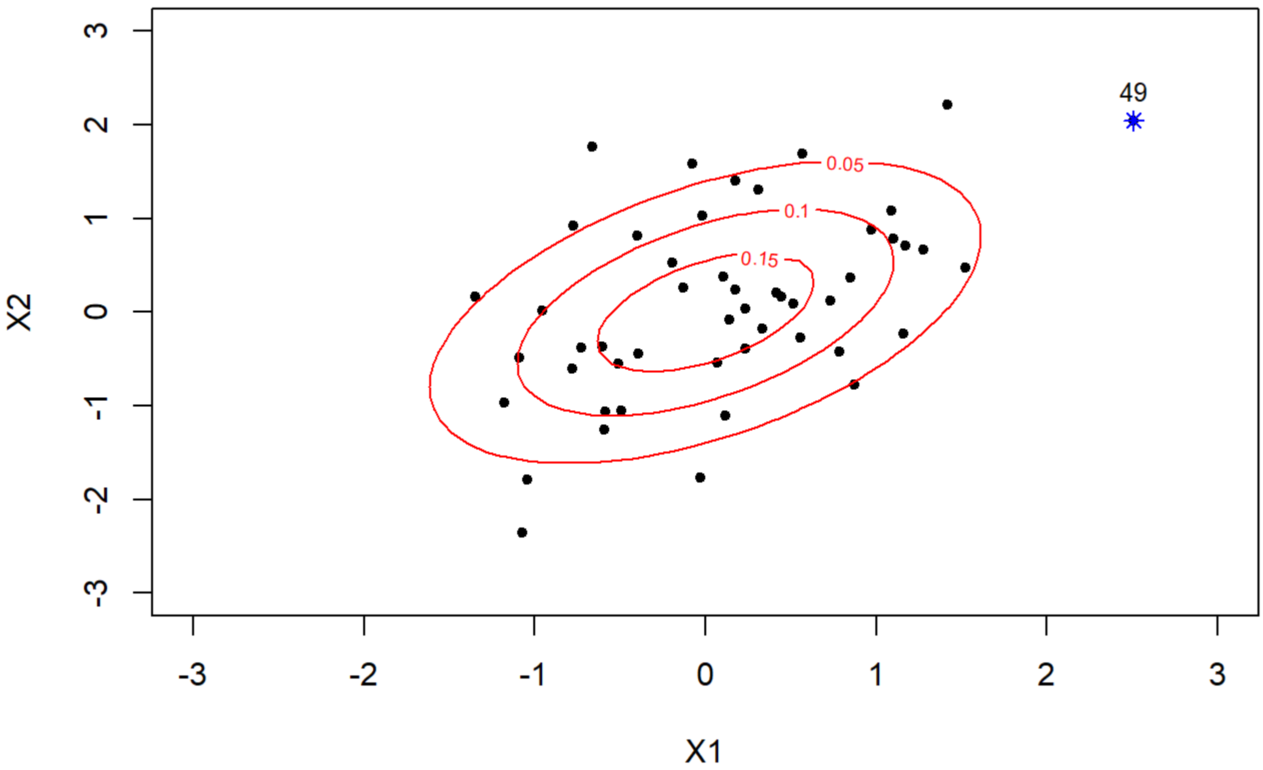
Su representación gráfica:

```r
f <- function(x1, x2) dmvnorm(data.frame(x1, x2), mu, V)
x <- seq(-3, 3, length = 50)
y <- seq(-3, 3, length = 50)
z <- outer(x, y, f)
persp(x, y, z, xlab = 'x1', ylab = 'x2', zlab = 'f(x1, x2)', col = 'orange', main = "Función de densidad")
```



Su representación gráfica ($f(x_1, x_2) = c$) y 50 datos simulados de este modelo

```r
#Se fija la semilla para la generación aleatoria
set.seed(123)
#Generación aleatoria del modelo
d <- rmvnorm(50, mu, V)
plot(d, xlab = "X1", ylab = "X2", pch = 20, xlim = c(-3, 3), ylim = c(-3, 3))
contour(x, y, z, nlevels = 4, add = T, col = 'red')
```



- Distancia de Mahalanobis

La distancia de Mahalanobis del vector x al vector μ basada en la matriz V:

$$D = \sqrt{(x-\mu)'V^{-1}(x-\mu)}$$

Tiene en cuenta la diferentes escalas de los datos y sus correlaciones.

Servirá para detectar las observaciones más alejadas del vector de medias que podrían ser observaciones atípicas (`outliers`) que no provengan de nuestra población o contengan errores.

→ Cuando se pueda, se deberán chequear y, si es posible, corregir o eliminar.

→ En otros casos, se deberán mantener por ser observaciones correctas que hay que tener en cuenta.

- Cálculo de la distancia de Mahalanobis

Para calcular las distancias de Mahalanobis al cuadrado de los datos al vector de medias (teóricas o muestrales) podemos utilizar la función `mahalanobis`.

```
V <- matrix(c(1, 1/2,
              1/2, 1), nrow = 2, ncol = 2, byrow = TRUE)
mu <- c(0, 0)
dM1 <- mahalanobis(d, mu, V)
dM2 <- mahalanobis(d, colMeans(d), cov(d))
```

- Distancias de los datos simulados al vector de medias teóricas μ con respecto a V

```
summary(dM1)
```

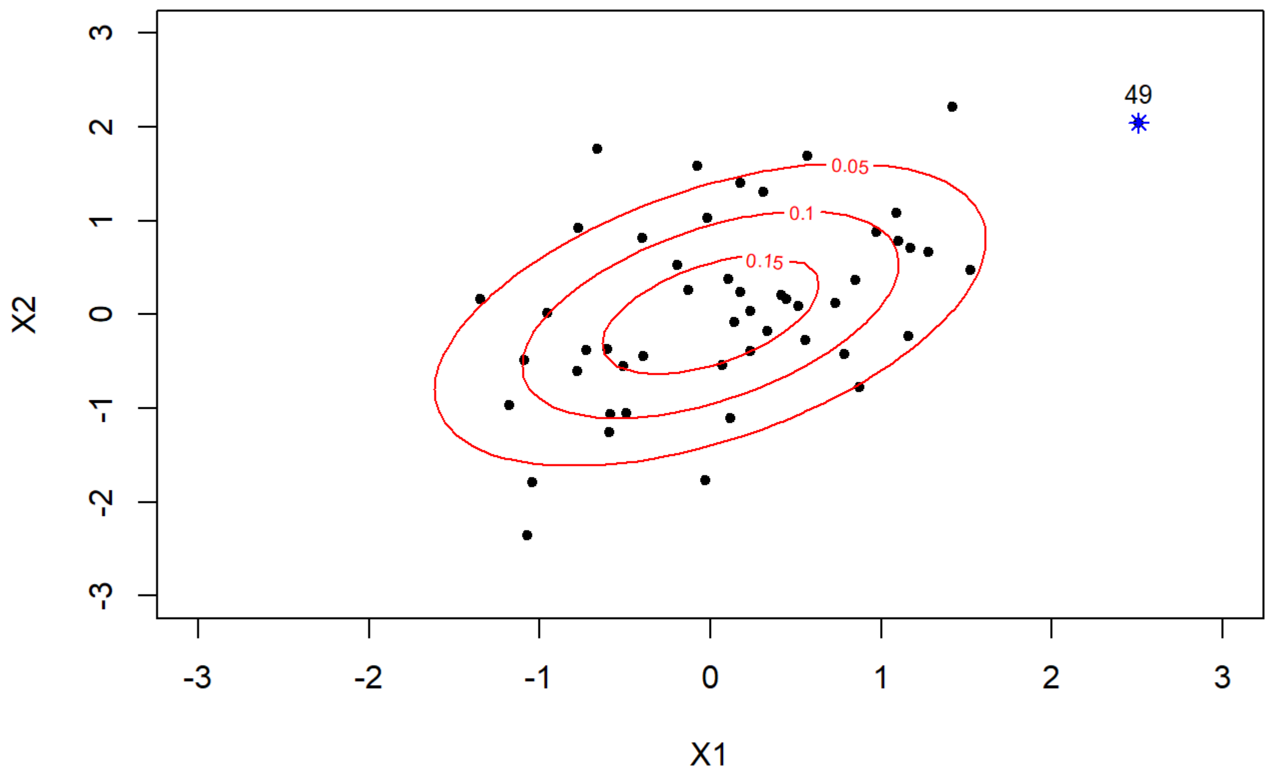
```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.05216 0.41016 1.26433 1.66615 2.31591 7.13332
```

- ¿Dónde se encuentra la observación más alejada del vector de medias?

```r
d[which.max(dM1), ]
```

```
## [1] 2.509470 2.046512
```

```r
f <- function(x1, x2) dmvnorm(data.frame(x1, x2), mu, V)
x <- seq(-3, 3, length = 50)
y <- seq(-3, 3, length = 50)
z <- outer(x, y, f)
plot(d, xlab = "X1", ylab = "X2", pch = 20, xlim = c(-3, 3), ylim = c(-3, 3))
points(d[which.max(dM1), 1], d[which.max(dM1), 2], col = "blue", pch = 8)
text(d[which.max(dM1), 1], d[which.max(dM1), 2], which.max(dM1), cex = 0.8, pos = 3)
contour(x, y, z, nlevels = 4, add = T, col = 'red')
```



- Distancias de los datos simulados al vector de medias muestrales $\overline{x}$ con respecto a $\mathcal{S}$

```r
summary(dM2)
```

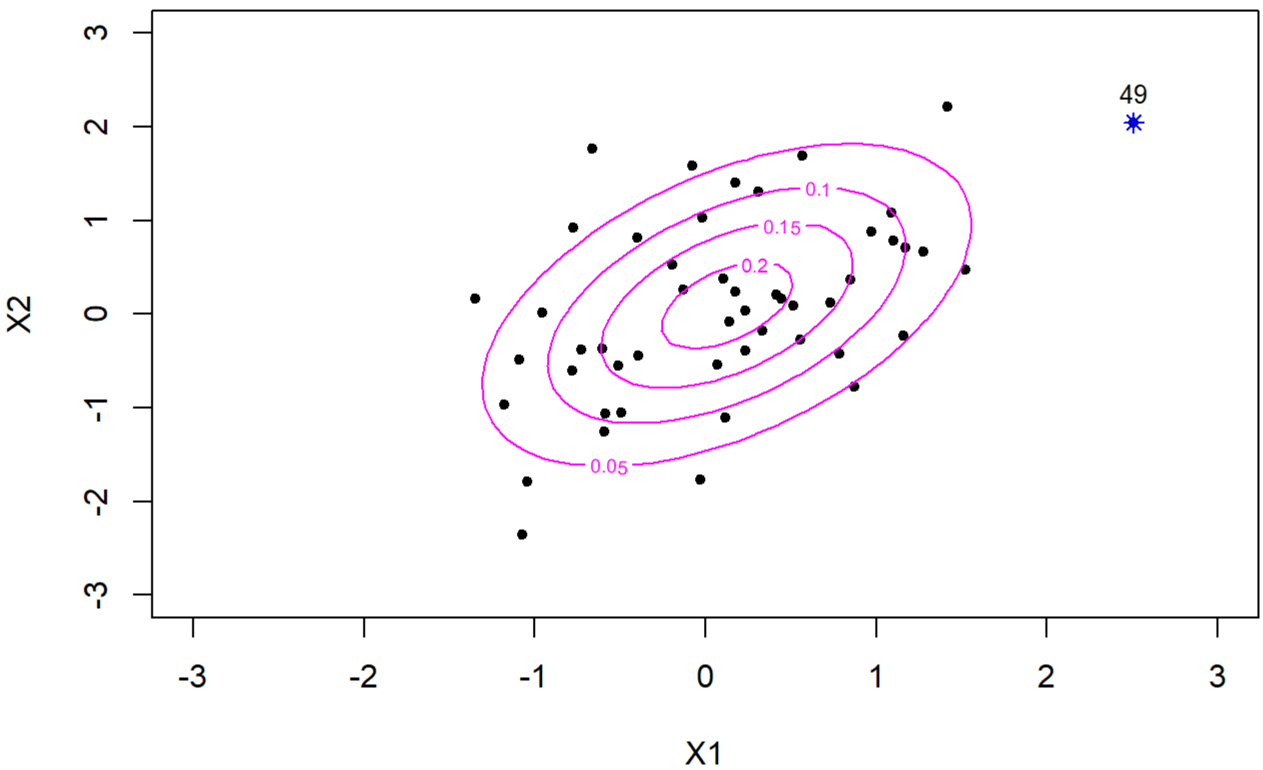
```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.02114 0.67111 1.52636 1.96000 2.64131 8.65906
```

- ¿Dónde se encuentra la observación más alejada del vector de medias?

```r
d[which.max(dM2), ]
```

```
## [1] 2.509470 2.046512
```

```
f <- function(x1, x2) dmvnorm(data.frame(x1, x2), colMeans(d), cov(d))
x <- seq(-3, 3, length = 50)
y <- seq(-3, 3, length = 50)
z <- outer(x, y, f)
plot(d, xlab = "X1", ylab = "X2", pch = 20, xlim = c(-3, 3), ylim = c(-3, 3))
points(d[which.max(dM2), 1], d[which.max(dM2), 2], col = "blue", pch = 8)
text(d[which.max(dM2), 1], d[which.max(dM2), 2], which.max(dM2), cex = 0.8, pos = 3)
contour(x, y, z, nlevels = 4, add = T, col = 'magenta')
```



1.9) Distribución multinomial $\mathcal{M}_k(n, p_1, \ldots, p_k)$

- Modelo multinomial

$(X_1, \ldots, X_k)$: variables aleatorias que representan el número de veces que ocurre el suceso A_i en un experimento aleatorio repetido n veces con k opciones dadas por $\{A_1, \ldots, A_k\}$ y con probabilidades constantes $p_i = P(A_i)$, para $i = 1, \ldots, k$.
Función masa de probabilidad conjunta:

$$p(x_1, \ldots, x_k) = P[X_1 = x_1, \ldots, X_k = x_k] = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k},$$

para enteros no negativos tales que $x_1 + \cdots + x_k = n$ y donde $p_i \in [0, 1]$ satisface $p_1 + \cdots + p_k = 1$.

Distribuciones marginales: X_i sigue una distribución binomial $B(n, p_i)$, con $E(X_i) = np_i$.

1.10) Estadístico de Pearson

- Discrepancias entre lo observado y lo esperado

Contexto: Lanzamos un dado $n =$ veces, $p_i = \dfrac{1}{6}$ para todo i, y los valores esperados son $np_i = 10$, para $i = 1, \ldots, 6$.

Objetivo: Medir las discrepancias entre valores observados y esperados.

Sea $X = (X_1, \ldots, X_k)$ una variable aleatoria con distribución multinomial, entonces el estadístico

$$T = \sum_{i=1}^{k} \frac{X_i - np_i}{np_i}$$

sigue una distribución Chi-cuadrado χ^2_{k-1} de Pearson con $k-1$ grados de libertad, cuando $n \longrightarrow \infty$.

1.11) Medias y covarianzas

- Definiciones

Dado el vector aleatorio.

→ El vector de medias (o esperanza matemática de X) se define como:

$$\mu := E[X] = (E[X_1], \ldots, E[X_k])' = (\mu_1, \ldots, \mu_k)'$$

(note que es un vector columna).

→ La matriz de covarianzas (o varianzas-covarianzas) se define como:

$$V = (\sigma_{i,j}),$$

donde $\sigma_{i,j}$ es la covarianza entre X_i y X_j, definida como:

$$\sigma_{i,j} = \mathrm{Cov}(X_i, X_j) = E\left[(X_i - \mu_i)(X_j - \mu_j)\right] = E[X_i X_j] - \mu_i \mu_j$$

Notemos que $\sigma_{i,i} = E\left[(X_i - \mu_i)^2\right] = \mathrm{Var}(X_i) = \sigma_i^2$.

- Cálculo de la esperanza matemática

La media de cada componente X_i del vector puede calcularse a partir de la distribución conjunta o a partir de la marginal.

→ Caso discreto:

$$\begin{aligned} E[X_i] &= \sum_{x_i} x_i P[X_i = x_i] \\ &= \sum_{x_1, \ldots, x_k} x_i P[X_1 = x_1, \ldots, X_k = x_k] \end{aligned}$$

→ Caso continuo:

$$\begin{aligned} E[X_i] &= \int_{\mathbb{R}} x_i f_{X_i}(x_i)\, \mathrm{d}x_i \\ &= \int_{\mathbb{R}^k} x_i f(x_1, \ldots, x_k)\, \mathrm{d}x_1 \cdots \mathrm{d}x_k \end{aligned}$$

1.11.1) Esperanza de la transformación $g : \mathbb{R}^k \longrightarrow \mathbb{R}$

- Caso discreto

Sea $g : \mathbb{R}^k \longrightarrow \mathbb{R}$ una función medible $\longrightarrow Y = g(X)$ es una variable aleatoria .

Si X es de tipo discreto,

$$\exists E[g(X)] \longleftrightarrow \sum_{x_1,\ldots,x_k} |g(x_1,\ldots,x_k)|\, P[X_1 = x_1,\ldots,X_k = x_k] < \infty$$

Y en caso de existir:

$$E[g(X_1,\ldots,X_k)] = \sum_{x_1,\ldots,x_k} g(x_1,\ldots,x_k) P[X_1 = x_1,\ldots,X_k = x_k]$$

- Caso continuo

Sea $g : \mathbb{R}^k \longrightarrow \mathbb{R}$ una función medible $\longrightarrow Y = g(X)$ es una variable aleatoria .

Si X es de tipo continuo,

$$\exists E[g(X)] \longleftrightarrow \int_{\mathbb{R}^k} |g(x_1,\ldots,x_k)|\, f_X(x_1,\ldots,x_k)\, \mathrm{d}x_1 \cdots \mathrm{d}x_k < \infty$$

Y en caso de existir:

$$E[g(X_1,\ldots,X_k)] = \int_{\mathbb{R}^k} g(x_1,\ldots,x_k) f_X(x_1,\ldots,x_k)\, \mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

- Propiedades

V es una matriz simétrica y semidefinida positiva ($x'Vx \geq 0$, para todo $x \in \mathbb{R}^k$).

En forma matricial,

$$V = E\left[(X-\mu)(X-\mu)'\right] = E[XX'] - \mu\mu'.$$

donde la esperanza de una matriz aleatoria se define como la matriz de las esperanzas de cada variable.

Si X_i y X_j son independientes, entonces

$$E[X_i X_j] = E[X_i]E[X_j]$$

y, por lo tanto, $\mathrm{Cov}(X_i, X_j) = 0$. El recíproco no es cierto.

Si $X \longrightarrow \mathcal{N}_k(\mu, V)$, se puede demostrar que μ es el vector de medias y V es la matriz de covarianzas.

1.12) Correlación

La correlación (lineal de Pearson) entre X_i y X_j se define como

$$\rho_{i,j} = \mathrm{Corr}(X_i, X_j) = \frac{\sigma_{i,j}}{\sigma_i \sigma_j}$$

siendo $\rho_{i,i} = \mathrm{Corr}(X_i, X_j) = 1$.

Mide el grado de relación lineal entre X_i y X_j.

Puede demostrarse que

$$-1 \leq \rho_{i,j} \leq 1.$$

Se dice que X_i y X_j son incorreladas si $\rho_{i,j} = 0$.

Si son independientes serán incorreladas, pero el recíproco no es cierto.

La matriz de correlaciones es $R = (\rho_{i,j})$.

1.13) Correlación entre vectores aleatorios

Análogamente, si X e Y son vectores aleatorios (de dimensiones cualesquiera), se define su matriz de covarianzas como

$$\operatorname{Cov}(X, Y) = (\operatorname{Cov}(X_i, Y_j))$$

y su matriz de correlaciones como

$$\operatorname{Corr}(X, Y) = (\operatorname{Corr}(X_i, Y_j)).$$

Puede demostrarse que

$$\operatorname{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])'].$$

Evidentemente, $\operatorname{Cov}(X) = \operatorname{Cov}(X, X)$.

- Propiedades

Si X, Y, Z son vectores (columna) aleatorios, se verifican las propiedades siguientes:

1) $E[a_1 g_1(X) + a_2 g_2(X)] = a_1 E[g_1(X)] + a_2 E[X_2]$, donde $a_1, a_2 \in \mathbb{R}$ y g_1 y g_2 son funciones medible de vectores aleatorios.

2) $X = (Y, Z), E_X[g(Y)] = E_Y[g(Y)]$, donde g es una función medible de un vector aleatorio, E_X denota la esperanza en la distribución conjunta y E_Y en la distribución marginal.

3) Si X e Y son independientes, entonces

$$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)],$$

donde g_1 y g_2 son funciones medibles cualesquiera de vectores aleatorios .

4) $E[AX + b] = AE[X] + b, A \in M_{m,k} b' \in \mathbb{R}^m$.

5) $\operatorname{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$.

6) Si $X_1, \ldots, X_k$ son independientes, $\operatorname{Cov}(X_i, X_j) = 0$.

7) $\operatorname{Var}(X_i + X_j) = \operatorname{Var}(X_i) + 2\operatorname{Cov}(X_i, X_j) + \operatorname{Var}(X_j)$.

8) $\operatorname{Cov}(aX_i + b, cX_j + d) = ac\operatorname{Cov}(X_i, X_j)$, donde $a, b, c, d \in \mathbb{R}$.

9) $\operatorname{Cov}(X) = E[(X - \mu)(X - \mu)'] = E[XX'] - \mu\mu'$.

10) $\operatorname{Var}(a'X) = a'\operatorname{Cov}(X)a = \sum_{i,j} a_i a_j \sigma_{i,j}$, donde $a \in \mathbb{R}^k$.

11) $\operatorname{Cov}(AX + b) = A\operatorname{Cov}(X)A'$, donde $A \in M_{m,k}$ y $b' \in \mathbb{R}^m$.

12) Si $X_1, \ldots, X_k$ son independientes, $\operatorname{Corr}(X_i, X_j) = 0$.

13) $\operatorname{Corr}(aX_i + b, cX_j + d) = \operatorname{Corr}(X_i, X_j)$, donde $a, b, c, d \in \mathbb{R}$.

14) $-1 \leq \operatorname{Corr}(X_i, X_j) \leq 1$.

15) $\operatorname{Corr}(X_i, aX_i + b) = \pm 1$, donde $a, b \in \mathbb{R}$ (según el signo de a).

16) $\operatorname{Corr}(X) = \delta^{-1}\operatorname{Cov}(X)\delta^{-1}$, donde δ es la matriz diagonal formada por las desviaciones típicas ($\delta = \operatorname{diag}(\sigma_1, \ldots, \sigma_k)$).

17) $\operatorname{Cov}(X, Y) = (\operatorname{Cov}(X_i, Y_j)) = \operatorname{Cov}(Y, X)'$.

18) $\operatorname{Cov}(X + Z, Y) = \operatorname{Cov}(X, Y) + \operatorname{Cov}(Z, Y)$

19) Si X e Y tienen la misma dimensión, entonces $\text{Cov}(X+Y) = \text{Cov}(X) + \text{Cov}(X,Y) + \text{Cov}(Y,X) + \text{Cov}(Y)$.

20) $\text{Cov}(AX, BY) = A\,\text{Cov}(X,Y)B'$, donde A y B son matrices (de dimensiones adecuadas).

21) Si X e Y independientes, entonces $\text{Cov}(X,Y) = 0$.

- Demostración apartado (10)

Directamente se tiene que:

$$\text{Var}(a'X) = \text{Cov}(a'X, a'X) = E[a'(X-\mu)(X-\mu)'a] = a\,\text{Cov}(X)a$$

Como consecuencia, se obtiene que la matriz de covarianzas $\text{Cov}(X)$ es semidefinida positiva ya que $\text{Var}(a'X) \geq 0$.

Lo mismo le ocurre a la matriz de correlaciones $\text{Corr}(X)$ ya que es la matriz de covarianzas de las variables aleatorias tipificadas $Z_i = \dfrac{X_i - \mu_i}{\sigma_i}$.

1.14) Resultados básicos de la inferencia

- Contexto

En la práctica, todas las medidas, varianzas y covarianzas serán desconocidas por lo que tenemos que estimarlas.
Para ello dispondremos de una muestra de individuos (objetos) en los que se han medido todas las variables.
Proporcionamos resultados básicos de inferencia para poder aplicar las técnicas multivariantes que desarrollaremos en temas posteriores.
Se ilustran estos procedimientos de inferencia con conjuntos de datos de `R`, accesibles con `data()`.

1.14.1) ¿Cómo se representan las muestras aleatorias?

- Matriz de la muestra aleatoria simple

En general, nuestra muestra aleatoria se representará como:

i	X_1	X_2	$\cdots$	X_k	Y
O_1	$X_{1,1}$	$X_{1,2}$	$\cdots$	$X_{1,k}$	Y_1
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
O_i	$X_{i,1}$	$X_{i,2}$	$\cdots$	$X_{i,k}$	Y_i
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
O_n	$X_{n,1}$	$X_{n,2}$	$\cdots$	$X_{n,k}$	Y_n

Si no hay grupos supondremos que los objetos

$$O_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,k})'$$

son una muestra aleatoria simple de X, es decir, serán vectores (columna) aleatorios independientes con la misma distribución que X.
Si no hay grupos supondremos lo mismo en cada grupo

La variable Y solo se usará para detonar la variable respuesta en regresión.

En algunos casos usaremos la matriz $M = (X_{i,j})$ que será una matriz aleatoria.

1.14.2) ¿Cómo se muestran los valores muestrales?

- Matriz de datos

En general, nuestra muestra se representará como:

i	x_1	x_2	$\cdots$	x_k	y
o_1	$x_{1,1}$	$x_{1,2}$	$\cdots$	$x_{1,k}$	y_1
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
o_i	$x_{i,1}$	$x_{i,2}$	$\cdots$	$x_{i,k}$	y_i
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
o_n	$x_{n,1}$	$x_{n,2}$	$\cdots$	$x_{n,k}$	y_n

La variable Y solo se usará para detonar la variable respuesta en regresión.

En algunos casos usaremos la matriz de datos $M = (x_{i,j})$

Si no hay grupos, supondremos que los vectores

$$o_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,k})'$$

son una realización de una muestra aleatoria simple de X, es decir, serán vectores (columna) con los datos muestrales.
Si hay grupos supondremos lo mismo en cada grupo.

RELACIÓN DE PROBLEMAS: VECTORES ALEATORIOS
ANÁLISIS ESTADÍSTICO MULTIVARIANTE
GRADO EN CIENCIA E INGENIERÍA DE DATOS

1. Sea (X, Y) un vector aleatorio con función de densidad conjunta
$$f(x,y) = \begin{cases} 1 & \text{si } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$
Hallar las distribuciones marginales y condicionadas.

2. Obtener las distribuciones marginales y condicionadas asociadas al vector aleatorio (X, Y) con función de densidad
$$f(x,y) = \begin{cases} 2 & \text{si } 0 < x < 1, 0 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

3. Sea (X, Y) un vector aleatorio con función de densidad
$$f(x,y) = \begin{cases} \frac{3}{4}\left[xy + \frac{x^2}{2}\right] & \text{si } 0 < x < 1, 0 < y < 2 \\ 0 & \text{en otro caso} \end{cases}$$
Hallar la distribución marginal de X y la distribución de Y condicionada a $X = \frac{1}{2}$.

4. Sea $\mathbf{X} = (X_1, X_2)$ un vector aleatorio con función masa de probabilidad
$$P[X_1 = x_1, X_2 = x_2] = \frac{k}{2^{x_1+x_2}}, x_1, x_2 \in \mathbb{N},$$
donde k es una constante. Obtener las distribuciones marginales y condicionadas.

5. Calcular la función de densidad de una distribución normal bidimensional en (1, 1) si las medias son cero, las varianzas 1 y 4, y la covarianza 1.

6. Sea (X, Y) un vector aleatorio con distribución uniforme en el cuadrado unidad, $[0,1] \times [0,1]$, con función de densidad conjunta
$$f(x,y) = \begin{cases} 1 & \text{si } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$
Calcular el valor esperado de $g(X, Y) = XY^2$, es decir, $E[XY^2]$.

7. (X, Y) vector aleatorio discreto con función masa de probabilidad conjunta:

$X \backslash Y$	1	2
1	1/9	2/9
2	2/9	4/9

   a) Calcular $E[X + Y]$, $E[2X + 3Y]$.
   b) Obtener el vector de medias, la matriz de covarianzas y la matriz de correlaciones del vector (X, Y).
   c) ¿Son independientes? ¿Están incorreladas?

8. Demostrar que el vector de medias muestral es el punto de $\mathbb{R}^k$ que minimiza la suma de las distancias al cuadrado (error cuadrático medio, MSE).

1) Sea (X, Y) un vector aleatorio con función de densidad conjunta

$$f(x,y) = \begin{cases} 1 & \text{si } 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Hallar las distribuciones marginales y condicionadas

$$\underset{0<x<1}{f_X(x)} = \int_{\infty}^{+\infty} f(x,y)\ \mathrm{d}y \ = \cancel{\int_{-\infty}^{0} 0\ \mathrm{d}y} + \int_0^1 1\ \mathrm{d}y + \cancel{\int_0^{+\infty} 0\ \mathrm{d}y} \ =$$

$$[y]_{y=0}^{y=1} = 1 \qquad f_X(x) = \begin{cases} 1 & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$\underset{0<y<1}{f_Y(y)} = \int_{-\infty}^{+\infty} f(x,y)\ \mathrm{d}x \ = \int_0^1 1\ \mathrm{d}x \ = \ [x]_{x=0}^{x=1} \ = \ 1 \qquad f_Y(y) \ =$$

$$\begin{cases} 1 & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$f(x,y) = \begin{cases} 1 & 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$\underset{f_X(x^*)>0}{y|x = x^*} \longrightarrow \underset{0<x^*<1}{f_{y|x=x^*}} \ = \frac{f(x^*, y)}{f_X(x^*)} = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$f(x,y) = f_X(x) \cdot f_Y(y) \quad X$ e Y independientes

2) Obtener las distribuciones marginales y condicionadas asociadas al vector aleatorio (X, Y) con función de densidad

$$f(x,y) = \begin{cases} 2 & \text{si } 0 < x < 1,\ 0 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

$$\underset{0<x<1}{f_X(x)} = \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y = \int_0^x 2\mathrm{d}y = [2y]_{y=0}^{y=x} = 2x \longrightarrow \begin{cases} 2x & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$\underset{0<y<1}{f_Y(y)} = \int_{-\infty}^{+\infty} f(x,y)\ \mathrm{d}x \ = \int_y^1 2\ \mathrm{d}x \ = \ [2x]_{x=y}^{x=1} \ = \ 2 - 2y \ \longrightarrow$$

$$\begin{cases} 2 - 2y & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Los recintos son dependientes.

$y|x = x^*$

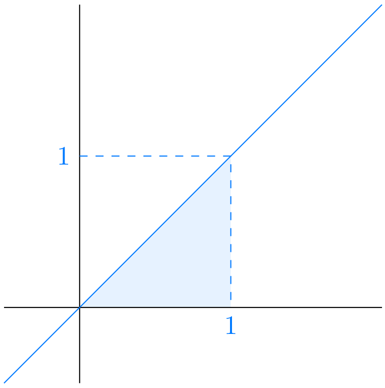
$f_X(x^*) > 0$

$$\underset{0<x^*<1}{f_{y|x=x^*}(y|x^*)} = \frac{f(x^*, y)}{f(x^*)} = \begin{cases} \dfrac{2}{2x^*} & 0 < y < x^* \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \dfrac{1}{x^*} & 0 < y < x^* \\ 0 & \text{en otro caso} \end{cases}$$

$x|y = y^*$

$f_Y(y^*) > 0$

$$\underset{0<y^*<1}{f_{x|y=y^*}(x|y^*)} = \frac{f(x, y^*)}{f_Y(y^*)} = \begin{cases} \dfrac{2}{2 - 2y^*} & \text{si } y^* < x < 1 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \dfrac{1}{1 - y^*} & \text{si } y^* < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$



3) Sea (X, Y) un vector aleatorio con función de densidad

$$f(x,y) = \begin{cases} \frac{3}{4}\left[xy + \frac{x^2}{2}\right] & \text{si } 0 < x < 1,\ 0 < y < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Hallar la distribución marginal de X y la distribución de Y condicionada a $X = \frac{1}{2}$.

$$\underset{0<x<1}{f_X(x)} = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y = \int_0^2 \frac{3}{4}\left[xy + \frac{x^2}{2}\right] = \frac{3}{4}\left[\frac{xy^2}{2} + \frac{x^2}{2}\right]_{y=0}^{y=2} =$$

$$\frac{3}{4}\left(2x + \frac{x^2}{2}\right) \longrightarrow \begin{cases} \frac{3}{4}\left(2x + \frac{x^2}{2}\right) & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$y|x = x^*$

$$\underset{0<x^*<1}{f_{y|x=x^*}(y|x^*)} = \frac{f(x^*, y)}{f(x^*)} = \frac{\frac{3}{4}\left(x^*y + \frac{(x^*)^2}{2}\right)}{\frac{3}{4}\left(2x^* + \frac{(x^*)^2}{2}\right)} = \frac{x^*y + \frac{(x^*)^2}{2}}{2x^* + \frac{(x^*)^2}{2}} = \frac{2x^*y + (x^*)^2}{4x^* + (x^*)^2} \xrightarrow{x^*=\frac{1}{2}} \frac{y + \frac{1}{4}}{2 + \frac{1}{4}} = \frac{4y+1}{9}$$

$$\begin{cases} \frac{4y+1}{9} & \text{si } 0 < y < 2 \\ 0 & \text{en otro caso} \end{cases}$$

4) Sea $X = (X_1, X_2)$ un vector aleatorio con función masa de probabilidad

$$P[X_1 = x_1, X_2 = x_2] = \frac{k}{2^{x_1+x_2}}, x_1, x_2 \in \mathbb{N}$$

donde k es una constante. Obtener las distribuciones marginales y condicionadas.

Primero hay que encontrar k para que la función masa de probabilidad sume 1. Para ello, sumamos sobre todos los posibles valores de x_1 y x_2:

$$\sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty} P[X_1 = x_1, X_2 = x_2] = \sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty} \frac{k}{2^{x_1+x_2}} = k\sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty}\left(\frac{1}{2}\right)^{x_1+x_2}$$

Dado que la serie $\sum_{n=0}^{\infty}\left(\frac{1}{2}\right)^n$ es una serie geométrica de suma a 2, la doble suma es 4. Por lo tanto, para que la función masa de probabilidad sume 1:

$$4k = 1 \longrightarrow k = \frac{1}{4}$$

La distribución marginal de X_1 se obtiene sumando la función masa de probabilidad conjunta de todos los posibles valores de X_2:

$$P[X_1 = x_1] = \sum_{x_2=0}^{\infty} P[X_1 = x_1, X_2 = x_2] = \sum_{x_2=0}^{\infty} \frac{1}{4}\left(\frac{1}{2}\right)^{x_1+x_2} = \frac{1}{4}\left(\frac{1}{2}\right)^{x_1} \sum_{x_2=0}^{\infty}\left(\frac{1}{2}\right)^{x_2} = \frac{1}{4}\cdot\left(\frac{1}{2}\right)^{x_1} \cdot 2 = \frac{1}{2^{x_1+1}}$$

$$P[X_2 = x_2] = \sum_{x_1=0}^{\infty} P[X_1 = x_1, X_2 = x_2] = \sum_{x_1=0}^{\infty} \frac{1}{4}\left(\frac{1}{2}\right)^{x_1+x_2} = \frac{1}{4}\left(\frac{1}{2}\right)^{x_2} \sum_{x_1=0}^{\infty}\left(\frac{1}{2}\right)^{x_1} = \frac{1}{4}\cdot\left(\frac{1}{2}\right)^{x_2} \cdot 2 = \frac{1}{2^{x_2+1}}$$

Ahora obtenemos las distribuciones condicionadas:

$$P[X_1 = x_1 | X_2 = x_2] = \frac{P[X_1 = x_1, X_2 = x_2]}{P[X_2 = x_2]} = \frac{\frac{1}{4}\left(\frac{1}{2}\right)^{x_1+x_2}}{\frac{1}{2^{x_2+1}}} = \frac{1}{2^{x_1+1}}$$

$$P[X_2 = x_2 | X_1 = x_1] = \frac{P[X_1 = x_1, X_2 = x_2]}{P[X_1 = x_1]} = \frac{\frac{1}{4}\left(\frac{1}{2}\right)^{x_1+x_2}}{\frac{1}{2^{x_1+1}}} = \frac{1}{2^{x_2+1}}$$

Por lo tanto, las distribuciones condicionadas son las mismas que las marginales, lo que indica que X_1 y X_2 son variables aleatorias independientes.