Modelos Estadísticos para la Regresión y la Clasificación
Estadística Descriptiva
Clase 2: Probabilidad

Mathias Bourel

Instituto de Matemática y Estadística Prof. Rafael Laguardia (IMERL)
Facultad de Ingeniería, Universidad de la República, Uruguay

7 de agosto de 2024

Plan

- Sea Ω un espacio muestral, es el conjunto de todos los sucesos elementales de un proceso aleatorio. Por ejemplo:
  - Si lanzo una moneda al aire $\Omega = \{C, X\}$
  - Si tiro un dado $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Si lanzo un dardo en el segmento $[0, 1]$, $\Omega = [0, 1]$.
- Una σ-álgebra $\mathcal{A}$ sobre Ω es el conjunto de todos los subconjuntos de Ω que serán "probabilizables".
- Una variable aleatoria $X : \Omega \to \mathbb{R}$ es una función que verifica que

$$\forall x \in \mathbb{R} \quad \{w : X(w) \leq x\} \in \mathcal{A}$$

Ejemplos

1. si se tiran dos dados, el espacio muestral es

$$\Omega = \{(n_1, n_2) : n_i \in \{1, \ldots, 6\}\}$$

   y una variable aleatoria podría ser la suma $X(n_1, n_2) = n_1 + n_2$.
2. si $\Omega = \{\text{población Montevideo}\}$ una variable aleatoria podría ser la áltura $X(w) = \text{áltura}$.

Notación

$$[X \leq x] = \{w \in \Omega : X(w) \leq x\}$$

- Un evento o suceso A contiene varios sucesos elementales.
- Propiedades de una probabilidad:
  1. $\mathbb{P}(\emptyset) = 0$
  2. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
  3. Si $A \subset B$ entonces $\mathbb{P}(A) \leq \mathbb{P}(B)$
  4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
  5. $\mathbb{P}\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} \mathbb{P}(A_n)$ para conjuntos A_n disjuntos dos a dos.

  Ejemplo: Lanzamiento de dos dados equilibrados e independientes.
  1. $\Omega = \{w_1 = (1,1), w_2 = (1,2), \ldots, w_{36} = (6,6)\}$
  2. Si $A =$la suma de dos valores es igual a 4 entonces

$$\mathbb{P}(A) = \mathbb{P}(\{(1,3),(2,2),(3,1)\}) = \mathbb{P}(1,1) + \mathbb{P}(2,2) + \mathbb{P}(3,1) = \frac{3}{36}$$

En el ejemplo anterior, si definimos como X a la variable aleatoria que devuelve la suma de los valores observados tenemos que $\mathbb{P}(A) = \mathbb{P}(X = 4)$.
Si la variable aleatoria es discreta, la ley de probabilidad de X consiste en dar los valores de las probabilidades $\mathbb{P}(X = x)$ para todo valor posible x de X.

Si $(\Omega, \mathcal{A}, \mathbb{P})$ es un espacio de probabilidad, X una variable aleatoria, la función de distribución de X es $F_X : \mathbb{R} \to \mathbb{R}$ definida por

$$F_X(x) = \mathbb{P}(X \leq x) \;\; \forall\, x \in \mathbb{R}$$

Propiedades:

1. $F_X(x) \in [0,1] \;\; \forall\, x \in \mathbb{R}$
2. Si $x_1 \leq x_2$ entonces $F_X(x_1) \leq F_X(x_2)$.
3. F_X es continua a la derecha, es decir $\lim\limits_{x \to a^+} F_X(x) = F_X(a)$.
4. $\lim\limits_{x \to \infty} F_X(x) = 1, \qquad \lim\limits_{x \to -\infty} F_X(x) = 0.$

Función de distribución empírica

Sea $X_1, \ldots, X_n$ una sucesión de variables aleatorias independientes con igual distribución F. La función de distribución empírica es

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{(-\infty,t]\}}(X_i) = \frac{1}{n}\#\{\text{cantidad de observaciones} \leq t\}$$

El teorema de Glivenko-Cantelli asegura que con probabilidad 1

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0$$

o sea la función de distribución empírica converge uniformemente a la función de distribución F.

Dos tipos de variables aleatorias

Sea $R_X = \{a \in \mathbb{R} : \mathbb{P}(X = a) > 0\}$ el conjunto de los puntos de discontinuidad de F_X. Se prueba que R_X es numerable.

- X es variable aleatoria discreta si y sólo si $\begin{cases} R_X \text{ es discreto} \\ \mathbb{P}(X \in R_X) = 1 \end{cases}$

  Toda variable aleatoria discreta tiene asociada una función de cuantía $p_X : \mathbb{R} \to \mathbb{R}$ definida por $p_X(x) = \mathbb{P}(X = x) \, \forall \, x \in \mathbb{R}$.

  Ejemplo: Bernoulli, Binomial, Geométrica, Hipergeométrica, Poisson.
- X es variable aleatoria continua si y sólo si $R_X = \emptyset$.
- X es una variable aleatoria absolutamente continua si y sólo si existe una función $f_X : \mathbb{R} \to [0, +\infty)$ tal que

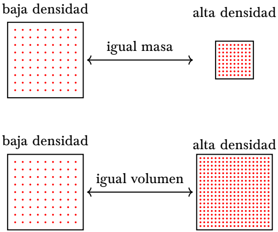
$$F_X(x) = \int_{-\infty}^{x} f_X(s)\, ds$$

  Observar que en este caso:
  - $\displaystyle\int_{-\infty}^{+\infty} f_X(s)\, ds = 1$
  - $\mathbb{P}(a \leq X \leq b) = \displaystyle\int_{a}^{b} f_X(s)\, ds$

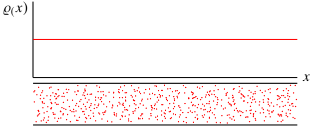
  Ejemplo: Uniforme, Normal, LogNormal.

Densidad

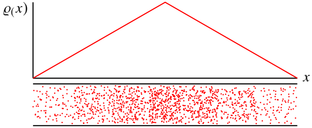
densidad = $\frac{\text{masa}}{\text{volumen}}$



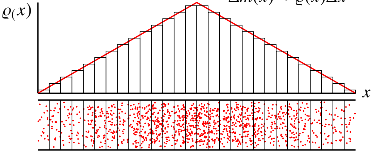
Uniforme



No uniforme



$\Delta m(x) \approx \varrho(x)\Delta x$

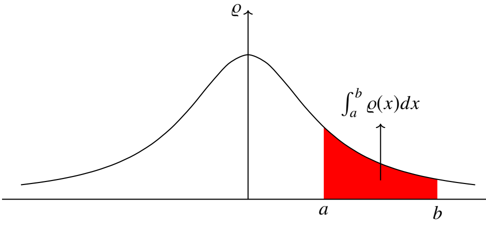


Definición

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

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

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



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

Esperanza y Varianza

- La *esperanza* de una variable aleatoria X es

$$\mathbb{E}(X) = \sum_{x \in R_X} x p_X(x) \quad \text{si } X \text{ es discreta}$$

$$\mathbb{E}(X) = \int x f_X(x)\, dx \quad \text{si } X \text{ es absolutamente continua}$$

  Un estimador de la esperanza de una variable aleatoria X es

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

  donde $X_1, \ldots, X_n$ son independientes y todos con la misma distribución que X

- La *varianza* de una variable aleatoria X es

$$\text{Var}(X) = \sigma^2 = \mathbb{E}\big(X - \mathbb{E}(X)\big)^2 = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2$$

  Estimadores de la varianza de una variable aleatoria X son

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \qquad s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

  donde $X_1, \ldots, X_n$ son independientes y todos con la misma distribución que X

1. Para la esperanza
   - $\mathbb{E}(X)$ no siempre existe
   - Si a es constante entonces $\mathbb{E}(a) = a$
   - $\mathbb{E}(X + a) = \mathbb{E}(X) + a$
   - $\mathbb{E}(aX) = a\mathbb{E}(X)$
   - $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
   - Si X e Y son independientes entonces $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. El recíproco es falso en general.
2. Para la varianza
   - $\text{Var}(X)$ minimiza la función $\mathbb{E}\left[(X - a)^2\right]$ ya que $\mathbb{E}\left[(X - a)^2\right] = \text{Var}(X) + (\mathbb{E}(X) - a)^2$
   - $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
   - $\text{Var}(X - a) = \text{Var}(X)$
   - $\text{Var}(aX) = a^2\text{Var}(X)$
   - $\text{Var}(X) = 0 \Leftrightarrow X = a$ c.s
   - $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$ donde
     $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
     $\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$
   - Si X e Y son independientes entonces $\text{Cov}(X, Y) = 0$ y
     $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$. El recíproco es en general falso.

1. **Distribución de Bernoulli** $Ber(p)$. Por ejemplo "resultado del lanzamiento de una moneda" que notamos por 1 o 0 según cara o cruz. Si $\mathbb{P}(X=1)=p$ entonces
   - $\mathbb{E}(X)=\sum_{i=0}^{1} i\mathbb{P}(X=i)=0\times(1-p)+1\times p=p$
   - $\mathbb{E}(X^2)=\sum_{i=0}^{1} i^2\mathbb{P}(X=i)=0\times(1-p)+1\times p=p$
   - $\text{Var}(X)=\mathbb{E}(X^2)-\mathbb{E}(X)^2=p-p^2=p(1-p)$
2. **Distribución binomial** $B(n,p)$. La variable aleatoria $X\in\{0,1\ldots,n\}$ cuenta la cantidad de caras obtenidas después de lanzar n veces la moneda, siendo p la probabilidad de obtener cara en cada lanzamiento
   - $\mathbb{P}(X=k)=C_k^n p^k(1-p)^{n-k}$.
   - $\mathbb{E}(X)=np;\quad \text{Var}(X)=np(1-p)$
3. **Distribución uniforme** $U(\{1,\ldots,n\})$. Si la variable aleatoria X toma valores en $\{1,\ldots,n\}$
   - $\mathbb{P}(X=i)=1/n$
   - $\mathbb{E}(X)=\sum_{i=0}^{1} i\mathbb{P}(X=i)=\frac{n+1}{2}$
   - $\text{Var}(X)=\mathbb{E}(X^2)-\mathbb{E}(X)^2=\frac{n^2-1}{12}$
4. **Distribución de Poisson** $\mathcal{P}(\lambda)$. $X\in\mathbb{N},\ \ \lambda>0$
   - $\mathbb{P}(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}$
   - $\mathbb{E}(X)=\text{Var}(X)=\lambda$.

Distribución Uniforme

Una variable aleatoria $X \ in[a,b]$ absolutamente continua tiene distribución uniforme en $[a,b]$ si

$$f_X(x) = \begin{cases} 0 & x \notin [a,b] \\ \frac{1}{b-a} & x \in [a,b] \end{cases}$$

Notación: $X \sim \mathcal{U}[a,b]$.

La función de distribución de $X \sim \mathcal{U}[a,b]$ es:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b \end{cases}$$

$\mathbb{E}(X) = \frac{a+b}{2}$.
$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$.
La suma de dos variables aleatorias uniformes no es uniforme.

Distribución Normal

Recordamos:

1. la densidad de una normal típica

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

   y escribimos $X \sim N(0, 1)$

Distribución Normal

Recordamos:

1. la densidad de una normal típica
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$
y escribimos $X \sim N(0,1)$
2. la densidad de una normal
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$
y escribimos $X \sim N(\mu, \sigma^2)$

Distribución Normal

Recordamos:

1. la densidad de una normal típica
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$
y escribimos $X \sim N(0,1)$
2. la densidad de una normal
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$
y escribimos $X \sim N(\mu, \sigma^2)$

Observaciones:

- Si $X \sim B(n, 0.5)$ y n tiene a infinito, entonces el gráfico de la ley de probabilidad que se obtiene tiende a la curva (simétrica) gaussiana.
- $\mathbb{E}(X) = \mu$ y $\mathrm{Var}(X) = \sigma^2$.
- La gaussiana tipíca es simétrica al rededor de 0 (alrededor de μ si no es típica). Tiene dos puntos de inflexión en -1 y en 1 ($\pm\sigma$ o $\mu \pm \sigma$ según el caso)
- Si σ es grande el pico de la gaussiana es chico (mucha dispersi'on a la media) y si σ es chico el pico de la gaussiana es grande (poca dispersión a la media).
- Si $X \sim N(\mu, \sigma^2)$ entonces $z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

Distribución Normal

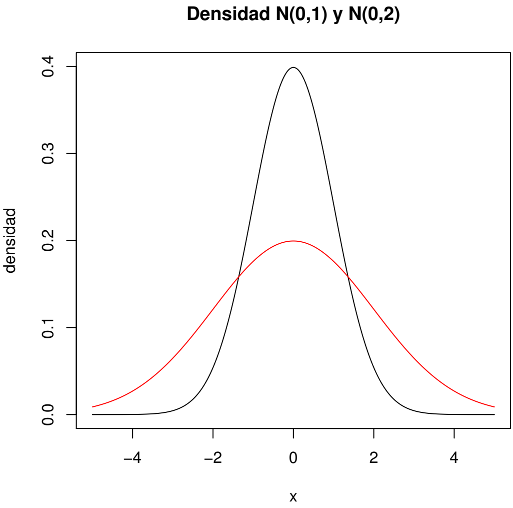
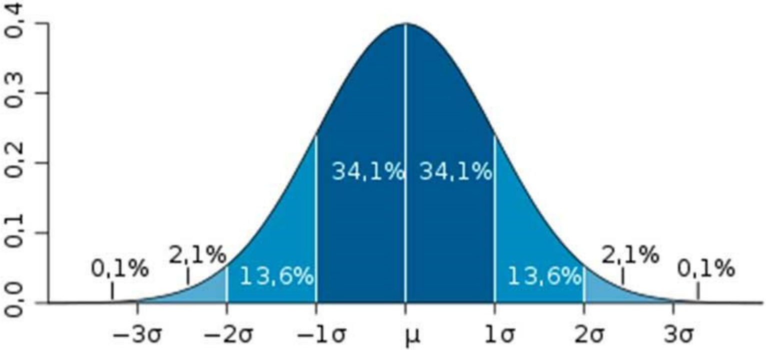


Figura: En negro $\mathcal{N}(0,2)$ y en rojo $\mathcal{N}(0,1)$



$\mathbb{P}(\mu - \sigma < X < \mu + \sigma) = \mathbb{P}\left(-1 < \frac{X-\mu}{\sigma} < 1\right) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \approx$ $2 \times 0{,}84 - 1 \approx 0,682$ Esto significa que para una distribución normal, hay un $31,7\,\%$ de chance de observar un desvío a la media mayor que σ.

Una variable aleatoria X tiene distribución lognormal si su logarítmo $Y = \ln(X)$ tiene distribución normal $\mathcal{N}(\mu_Y, \sigma_Y^2)$. Su función de densidad es

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_Y x} e^{\frac{-(\ln(x)-\mu_Y)^2}{2\sigma_Y^2}} \qquad \text{si } x > 0$$

Si X tiene distribución lognormal entonces

$$\mathbb{E}(X) = e^{\mu_Y + \sigma_Y^2/2}$$

$$\text{Var}(X) = (e^{\sigma_Y^2} - 1)e^{2\mu_Y + \sigma_Y^2}$$

En este caso la variable $Z = \frac{\ln(X)-\mu_Y}{\sigma_Y} \sim \mathcal{N}(0,1)$

1. - En general se hace la hipótesis de que una muestra proviene de realizaciones de una variable aleatoria con una cierta distribución en función de la naturaleza de la característica que se observa.
   - La distribución depende de uno o dos parámetros, en general desconocidos
   - Para ver si una muestra proviene una distribución determinada se puede comparar el histograma obtenido a partir de la muestra con la distribución de la ley. Pero también se puede trazar el QQ-plot, comparar los quantile empíricos con los teóricos, hacer un tesis de hipótesis, etc.
2. Sea $X_1, \ldots, X_n$ una muestra aleatoria simple ($X_1, \ldots, X_n$ independientes).
   - La media $\overline{X}_n$ es una variable aleatoria
   - Si suponemos que la variable aleatoria es normal con parámetros (μ, σ) entonces la media empírica tiene distribución normal con parámetros $\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
   - Si la variable aleatoria es cualquiera, con esperanza μ y varianza σ entonces para valores de n grandes la media empírica tiene distribución normal con parámetros $\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ (Teorema Central del Límite).

Plan

Sea $\mathbf{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$ una VA vectorial.
Función de distribución:

$$F_{\mathbf{x}}(x^0) = \mathbb{P}(\mathbf{x} \leq x^0) = \mathbb{P}(x_1 \leq x_1^0, x_2 \leq x_2^0, \ldots, x_p \leq x_p^0)$$

- Si $\mathbf{x}$ es discreta, entonces $p(x^0) = \mathbb{P}(\mathbf{x} = x^0) = \mathbb{P}(x_1 = x_1^0, x_2 = x_2^0, \ldots, x_p = x_p^0)$
- Decimos que $\mathbf{x}$ es absolutamente continua, si existe una función $f : \mathbb{R}^P \to \mathbb{R}$ no negativa con $\int_{\mathbb{R}^p} f(\mathbf{x})\, d\mathbf{x} = 1$ tal que

$$F_{\mathbf{x}}(x^0) = \int_{-\infty}^{x^0} f(\mathbf{x})\, d\mathbf{x} = \int_{-\infty}^{x_1^0} \int_{-\infty}^{x_2^0} \cdots \int_{-\infty}^{x_p^0} f(x_1, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p$$

Si x es escalar y absolutamente continua, entonces
$p(x^0) = \mathbb{P}\left(x \in [x^0 - \frac{\Delta x}{2}, x^0 + \frac{\Delta x}{2}]\right) = \int_{x^0 - \frac{\Delta x}{2}}^{x^0 + \frac{\Delta x}{2}} f(t)\, dt \approx f(x^0)\Delta x$
En general si $\mathbf{x}$ es vectorial $p(x^0) = f(x^0)\Delta\mathbf{x}$, siendo Δx el elemento de volumen.

DISTRIBUCIONES MARGINALES:

- $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ con distribución conjunta f_{x_1,x_2} entonces

$$f_{x_1}(x_1) = \int_{\mathbb{R}} f_{x_1,x_2}(x_1, x_2)\, dx_2 \qquad f_{x_2}(x_2) = \int_{\mathbb{R}} f_{x_1,x_2}(x_1, x_2)\, dx_1$$

  y con abuso de notación:

$$f(x_1) = \int_{\mathbb{R}} f(x_1, x_2)\, dx_2 \qquad f(x_2) = \int_{\mathbb{R}} f(x_1, x_2)\, dx_1$$

- $\int_{\mathbb{R}} f_{x_1}(x_1)\, dx_1 = \int_{\mathbb{R}} f_{x_2}(x_2)\, dx_2 = 1$

DISTRIBUCIONES CONDICIONADAS:

- Sea el vector aleatorio $X = (\mathbf{x_1}, \mathbf{x_2}) \in \mathbb{R}^{p\times 2}$.
  Definimos la distribución condicionada de $\mathbf{x_1}$ para un valor de $\mathbf{x_2} = \mathbf{x_2^0}$ como

$$f(\mathbf{x_1}|\mathbf{x_2} = \mathbf{x_2^0}) = \frac{f(\mathbf{x_1}, \mathbf{x_2^0})}{f(\mathbf{x_2^0})} \quad \text{suponiendo que } f(\mathbf{x_2^0}) \neq 0$$

  Esto es consistente con el concepto de probabilidad condicionada, pues, suponiendo que las variables $\mathbf{x_1}$ y $\mathbf{x_2}$ son escalares

$$\underbrace{f(x_1|x_2^0)\Delta x_1}_{\mathbb{P}(x_1|x_2^0)} = \underbrace{\frac{f(x_1, x_2^0)\Delta x_1 \Delta x_2}{f(x_2^0)\Delta x_2}}_{\frac{\mathbb{P}(x_1, x_2^0)}{\mathbb{P}(x_2^0)}}$$

  Entonces

$$f(\mathbf{x_1}, \mathbf{x_2}) = f(\mathbf{x_1}|\mathbf{x_2})f(\mathbf{x_2}) \qquad f(\mathbf{x_1}, \mathbf{x_2}) = f(\mathbf{x_2}|\mathbf{x_1})f(\mathbf{x_1})$$

La distribución marginal de $\mathbf{x}_2$ se puede calcular como

$$f(\mathbf{x}_2) = \int f(\mathbf{x}_2|\mathbf{x}_1) f(\mathbf{x}_1) d\,\mathbf{x}_1$$

Observar que si multiplicamos por $\Delta \mathbf{x}_2$ esto se puede interpretar como

$$f(\mathbf{x}_2)\Delta \mathbf{x}_2 = \int f(\mathbf{x}_2|\mathbf{x}_1) f(\mathbf{x}_1) d\,\mathbf{x}_1 \Delta \mathbf{x}_2$$

$$p(\mathbf{x}_2) = \sum \underbrace{f(\mathbf{x}_2|\mathbf{x}_1)\Delta \mathbf{x}_2}_{p(\mathbf{x}_2|\mathbf{x}_1)} \underbrace{f(\mathbf{x}_1)\Delta \mathbf{x}_1}_{p(\mathbf{x}_1)}$$

FORMULA DE BAYES:

$$f(\mathbf{x}_1|\mathbf{x}_2) = \frac{f(\mathbf{x}_2|\mathbf{x}_1) f(\mathbf{x}_1)}{f(\mathbf{x}_2)} = \frac{f(\mathbf{x}_2|\mathbf{x}_1) f(\mathbf{x}_1)}{\int f(\mathbf{x}_2|\mathbf{x}_1) f(\mathbf{x}_1) d\,\mathbf{x}_1}$$

Ejemplo

x_1 votar a algun candidato c_1, c_2, c_3, c_4 y x_2 nivel de ingreso A (alto), M (medio), B (bajo). Presentamos la distribución conjunta de votos:

	A	M	B
c_1	.1	.05	.01
c_2	.05	.20	.04
c_3	.04	.25	.07
c_4	.01	.1	.08

Distribuciones marginales:

	A	M	B	votos
c_1	.1	.05	.01	0.16
c_2	.05	.20	.04	0.29
c_3	.04	.25	.07	0.36
c_4	.01	.1	.08	0.19
ingresos	0.2	0.6	0.2	

Ejemplo

Distribución condicionada de los votos por personas con nivel de ingreso B:

	c_1	c_2	c_3	c_4
B	$\frac{0{,}01}{0{,}2} = 0{,}05$	$\frac{0{,}04}{0{,}2} = 0{,}2$	$\frac{0{,}07}{0{,}2} = 0{,}35$	$\frac{0{,}08}{0{,}2} = 0{,}4$

Distribución condicionada de los ingresos por votantes del candidato c_4:

	A	M	B
c_4	$\frac{0{,}01}{0{,}19} = 0{,}0526$	$\frac{0{,}1}{0{,}19} = 0{,}5263$	$\frac{0{,}08}{0{,}19} = 0,4211$

Independencia de vectores aleatorios

$\mathbf{x_1}$ y $\mathbf{x_2}$ son **independientes** si

$$f(\mathbf{x_2}|\mathbf{x_1}) = f(\mathbf{x_2})$$

lo cual equivale a

$$f(\mathbf{x_1}, \mathbf{x_2}) = f(\mathbf{x_1})f(\mathbf{x_2})$$

Notación: $\mathbf{x_1} \perp \mathbf{x_2}$

Observación: si $\mathbf{x_1} \perp \mathbf{x_2}$ entonces $g_1(\mathbf{x_1}) = \mathbf{y_1} \perp \mathbf{y_2} = g_2(\mathbf{x_2})$

Sea $\mathbf{x} = (x_1, \ldots, x_p)$ una VA en $\mathbb{R}^p$. El vector de medias es

$$\mathbb{E}(\mathbf{x}) = \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} = \begin{pmatrix} \mathbb{E}(x_1) \\ \vdots \\ \mathbb{E}(x_p) \end{pmatrix}$$

Si $\mathbf{x}$ es continua entonces notamos $\mathbb{E}(\mathbf{x}) = \int \mathbf{x} f(\mathbf{x}), d\mathbf{x}$ donde $\mathbb{E}(x_i) = \int x f_{x_i}(x)\, dx \,\forall\, i$

Proposición 1

- $\mathbb{E}(A\mathbf{x} + \mathbf{b}) = A\mathbb{E}(\mathbf{x}) + \mathbf{b}$ *siendo* A *una matriz y* $\mathbf{b}$ *un vector.*
- *Si* $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2})$ *y* $a, b \in \mathbb{R}$ *entonces* $\mathbb{E}(a\mathbf{x_1} + b\mathbf{x_2}) = a\mathbb{E}(\mathbf{x_1}) + b\mathbb{E}(\mathbf{x_2})$

Sea $\mathbf{x} = (x_1, \ldots, x_p)$ una VA en $\mathbb{R}^p$. Su matriz de varianzas y covarianzas es la matriz cuadarada

$$Var(\mathbf{x}) = V_x = \mathbb{E}[(\mathbf{x} - \mu)(\mathbf{x} - \mu)'] \in \mathcal{M}_{p\times p}$$

$$V_x = \begin{pmatrix} \sigma_1^2 & s_{12} & \ldots & s_{1p} \\ s_{12} & \sigma_2^2 & \ldots & s_{2p} \\ \vdots & & \ddots & \vdots \\ s_{1p} & s_{2p} & \ldots & \sigma_p^2 \end{pmatrix}$$

siendo

$$\sigma_i^2 = Var(x_i), \;\; s_{ij} = Cov(x_i, x_j) = \mathbb{E}(x_i x_j) - \mathbb{E}(x_i)\mathbb{E}(x_j)$$

Propiedades:

1. V_x es simétrica (es claro)
2. V_x es semidefinida positiva, es decir para todo $\mathbf{w} \in \mathbb{R}^p$ se tiene que $\mathbf{w}' V_x \mathbf{w} \geq 0$.
   En efecto, sea $\mathbf{w} \in \mathbb{R}^p$ y defino $y = (\mathbf{x} - \mu)'\mathbf{w} \in \mathbb{R}$. Entonces

$$\mathbb{E}(y) = \mathbb{E}((\mathbf{x} - \mu)'\mathbf{w}) = \mathbb{E}((\mathbf{x} - \mu))'\mathbf{w} = 0$$

   Por lo tanto

$$Var(y) = \mathbb{E}(y^2) = \mathbf{w}'\mathbb{E}((\mathbf{x} - \mu)(\mathbf{x} - \mu)')\mathbf{w} = \mathbf{w}' V_x \mathbf{w} \geq 0$$

Sea $\mathbf{x} = (x_1, \ldots, x_p)$ un vector aleatorio de $\mathbb{R}^p$ con densidad $f_x(\mathbf{x})$ y sea otro vector aleatorio $\mathbf{y} = (y_1, \ldots, y_p) \in \mathbb{R}^p$ definido por

$$\begin{cases} y_1 = g_1(x_1, \ldots, x_p) \\ y_2 = g_2(x_1, \ldots, x_p) \\ \qquad \vdots \\ y_p = g_p(x_1, \ldots, x_p) \end{cases}$$

donde suponemos que existen las funciones inversas
$x_1 = h_1(y_1, \ldots, y_p), \ldots, x_p = h_p(y_1, \ldots, y_p)$, siendo $g_1, \ldots, g_p, h_1, \ldots, h_p$ diferenciables.

Entonces puede demostrarse (regla del Jacobiano) que:

$$f_y(\mathbf{y}) = f_x(\mathbf{x}) \left| \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \cdots & \frac{\partial x_1}{\partial y_p} \\ \vdots & & & \vdots \\ \frac{\partial x_p}{\partial y_1} & \cdots & \cdots & \frac{\partial x_p}{\partial y_p} \end{pmatrix} \right|$$

Caso particular. Supongamos que $\mathbf{y} = A\mathbf{x}$ siendo $A \in \mathcal{M}_{p \times p}$ invertible. Entonces

$$f_y(\mathbf{y}) = f_x(A^{-1}\mathbf{y}) |\det(A^{-1})|$$

Supongamos que $\mathbf{y} = A\mathbf{x}$ siendo $A \in \mathcal{M}_{p\times p}$. Entonces

Proposición 2

1. $\mu_y = A\mu_x$
2. $V_y = AV_xA'$

Demostración.

1. Ya lo vimos.
2. $V_y = \mathbb{E}((\mathbf{y}-\mu_y)(\mathbf{y}-\mu_y)')$
$= \mathbb{E}(A(\mathbf{x}-\mu_x)(\mathbf{x}-\mu_x)'A') = A\mathbb{E}((\mathbf{x}-\mu_x)(\mathbf{x}-\mu_x)')A' = AV_xA'$

$\square$

La esperanza de $\mathbf{x}_1$ condicionada a $\mathbf{x}_2$ es:

$$\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2) = \int \mathbf{x}_1 f(\mathbf{x}_1|\mathbf{x}_2)\, d\mathbf{x}_1$$

y es una función de $\mathbf{x}_2$.
Si $\mathbf{x}_2$ es un valor fijo entonces $\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2)$ es constante.
Si $\mathbf{x}_2$ es un variable aleatoria entonces $\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2)$ es una variable aleatoria.

Proposición 3

Se cumple que

$$\mathbb{E}(\mathbf{x}_1) = \mathbb{E}\big(\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2)\big)$$

y la esperanza de la media condicionada es la esperanza marginal

Demostración:

$$\mathbb{E}(\mathbf{x}_1) = \int \mathbf{x}_1 f(\mathbf{x}_1) d\mathbf{x}_1 = \int \mathbf{x}_1 \left(\int f(\mathbf{x}_1|\mathbf{x}_2) f(\mathbf{x}_2) d\mathbf{x}_2 \right) d\mathbf{x}_1$$

$$= \int \int \mathbf{x}_1 f(\mathbf{x}_1|\mathbf{x}_2) f(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 = \int f(\mathbf{x}_2) \left(\int \mathbf{x}_1 f(\mathbf{x}_1|\mathbf{x}_2) d\mathbf{x}_1 \right) d\mathbf{x}_2$$

$$= \int f(\mathbf{x}_2) \mathbb{E}(\mathbf{x}_1|\mathbf{x}_2) d\mathbf{x}_2 = \mathbb{E}\big(\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2)\big)$$

Definición

$$\mathrm{Var}(\mathbf{x}_1|\mathbf{x}_2) = \mathbb{E}[(\mathbf{x}_1 - \mathbb{E}(\mathbf{x}_1|\mathbf{x}_2))^2|\mathbf{x}_2]$$

Propiedad

$$\mathrm{Var}(\mathbf{x}_1|\mathbf{x}_2) = \mathbb{E}({\mathbf{x}_1}^2|\mathbf{x}_2) - (\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2))^2$$

Lo verificamos para cada y:

$$\begin{aligned}\mathrm{Var}(\mathbf{x}_1|\mathbf{x}_2 = y) =& \mathbb{E}[(\mathbf{x}_1 - \mathbb{E}(\mathbf{x}_1|\mathbf{x}_2 = y))^2|\mathbf{x}_2 = y) \\ =& \mathbb{E}[{\mathbf{x}_1}^2 + (\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2 = y))^2 - 2\mathbf{x}_1\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2 = y)|\mathbf{x}_2 = y)\end{aligned}$$

Como $\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2 = y)$ es una constante cuando se considera la distribución condicionada a $\mathbf{x}_2 = y$, el resultado lo deducimos de la linealidad de la esperanza condicionada.

Descomposición de la varianza

$$\mathrm{Var}(\mathbf{x}_1) = \mathrm{Var}(\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2)) + \mathbb{E}(\mathrm{Var}(\mathbf{x}_1|\mathbf{x}_2))$$

Se sabe que $\mathbb{E}(\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2)) = \mathbb{E}(\mathbf{x}_1)$ y $\mathbb{E}(\mathbb{E}({\mathbf{x}_1}^2|\mathbf{x}_2)) = \mathbb{E}({\mathbf{x}_1}^2)$. Entonces:

$$\begin{aligned}
\mathrm{Var}(\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2)) + \mathbb{E}(\mathrm{Var}(\mathbf{x}_1|\mathbf{x}_2)) =& \mathbb{E}[(\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2))^2] - [\mathbb{E}(\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2))]^2 \\
&+ \mathbb{E}(\mathbb{E}({\mathbf{x}_1}^2|\mathbf{x}_2)) - \mathbb{E}[(\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2))^2] \\
=& \mathbb{E}[(\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2))^2] - [\mathbb{E}(\mathbf{x}_1)]^2 \\
&+ \mathbb{E}({\mathbf{x}_1}^2) - \mathbb{E}[(\mathbb{E}(\mathbf{x}_1|\mathbf{x}_2))^2] \\
=& \mathrm{Var}(\mathbf{x}_1).
\end{aligned}$$

Descomposición de la varianza

En particular, si x_1 es una variable aleatoria real y $\mathbb{E}(x_1) = \mu_1$ entonces:

- $(A) : \mathbb{E}\left((x_1 - \mu_1)^2\right) = \mathrm{Var}(x_1)$
- (B) : como $\mu_1 = \mathbb{E}(\mathbb{E}(x_1|\mathbf{x_2}))$ entonces $\mathrm{Var}(\mathbb{E}(x_1|\mathbf{x_2})) = \mathbb{E}\left((\mathbb{E}(x_1|\mathbf{x_2}) - \mu_1)^2\right)$
- (C) : como $\mathrm{Var}(x_1|\mathbf{x_2}) = \mathbb{E}((x_1 - \mathbb{E}(x_1|\mathbf{x_2}))^2|\mathbf{x_2})$ entonces $\mathbb{E}(\mathrm{Var}(x_1|\mathbf{x_2})) = \mathbb{E}\left((x_1 - \mathbb{E}(x_1|\mathbf{x_2}))^2\right)$

Por lo tanto de

$$\mathrm{Var}(x_1) = \underbrace{\mathrm{Var}(\mathbb{E}(x_1|\mathbf{x_2}))}_{(B)} + \underbrace{\mathbb{E}(\mathrm{Var}(x_1|\mathbf{x_2}))}_{(C)}$$

se deduce que:

$$\underbrace{\mathbb{E}\left((x_1 - \mu_1)^2\right)}_{(A)} = \underbrace{\mathbb{E}\left((\mathbb{E}(x_1|\mathbf{x_2}) - \mu_1)^2\right)}_{(B)} + \underbrace{\mathbb{E}\left((x_1 - \mathbb{E}(x_1|\mathbf{x_2}))^2\right)}_{(C)}$$

Descomposición de la varianza

El segundo termino promedia las varianzas de las distribuciones condicionadas. El primer termino recoge las diferencias entre la media global μ_1 y las medias condicionadas.

- Si $x_1 \perp \mathbf{x_2}$ entonces

$$\mathbb{E}(x_1|\mathbf{x_2}) = \int x_1 f(x_1|\mathbf{x_2})\, dx_1 = \int x_1 f(x_1)\, dx_1 = \mathbb{E}(x_1) = \mu_1$$

  y por lo tanto el segundo miembro $var(\mathbb{E}(x_1|\mathbf{x_2})) = 0$

- En modelos lineales univariantes, si $\overline{x}$ es la media global:

$$\frac{1}{n}\sum(x_i - \overline{x})^2 = \underbrace{\frac{1}{n}\sum(\widehat{x}_i - \overline{x})^2}_{(B)} + \underbrace{\frac{1}{n}\sum(x_i - \widehat{x}_i)^2}_{(C)}$$

  se obtiene la descomposición de la varianza total como suma de una varianza explicada por el modelo (B) y de una varianza no explicada por el modelo (C)

Decimos que $\mathbf{x} = (x_1, \ldots, x_d)$ es **gaussiana típica** en $\mathbb{R}^d$ si tiene densidad conjunta

$$f_x(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{1}{2}||\mathbf{x}||^2}$$

Notación:

$$\mathbf{x} \sim N(0_{\mathbb{R}^d}, I_d)$$

Decimos que $\mathbf{x} = (x_1, \ldots, x_d)$ es **gaussiana típica** en $\mathbb{R}^d$ si tiene densidad conjunta

$$f_x(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{1}{2}||\mathbf{x}||^2}$$

Notación:

$$\mathbf{x} \sim N(0_{\mathbb{R}^d}, I_d)$$

Decimos que $\mathbf{y}$ es **gaussiana** en $\mathbb{R}^d$ o **normal multivariada** con media $\mu \in \mathbb{R}^d$ y matriz de covarianza $\Sigma = AA'$ siendo $A \in \mathcal{M}_{d\times d}$ si $\mathbf{y}$ tiene la misma distribución que $\mu + A\mathbf{x}$ con $\mathbf{x} \sim N(0_{\mathbb{R}^d}, I_d)$.

Notación:

$$\mathbf{y} \sim N(\mu, \Sigma)$$

Sea $\mathbf{x} = (x_1, x_2) \sim N(0_{\mathbb{R}^2}, I_2)$

Si defino

$$\mathbf{y} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Entonces

$$\mathbf{y} \sim N\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \right)$$

library(mvtnorm) library(MASS)

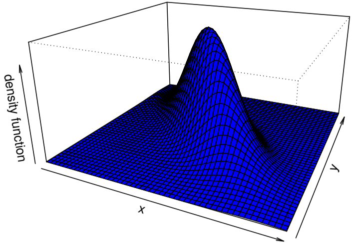


Figura: $N(\mu, \Sigma)$ con $\mu = (0,0)$ y $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Densidad de una normal $N(\mu, \Sigma)$

Si $\mathbf{y}$ es **gaussiana** en $\mathbb{R}^d$, $\mathbf{y} \sim N(\mu, \Sigma)$ con $\Sigma = AA'$ e $\mathbf{y} = \mu + A\mathbf{x}$ entonces

$$\mathbf{y} = Q(\mathbf{x})$$

con $Q : \mathbb{R}^d \to \mathbb{R}^d$ definida por $Q(\mathbf{x}) = \mu + A\mathbf{x}$.

[1] Si $y = Ax + \mu$ entonces $f_y(y) = f_x(A^{-1}(y - \mu))|det(A)^{-1}|$

Densidad de una normal $N(\mu, \Sigma)$

Si $\mathbf{y}$ es **gaussiana** en $\mathbb{R}^d$, $\mathbf{y} \sim N(\mu, \Sigma)$ con $\Sigma = AA'$ e $\mathbf{y} = \mu + A\mathbf{x}$ entonces

$$\mathbf{y} = Q(\mathbf{x})$$

con $Q : \mathbb{R}^d \to \mathbb{R}^d$ definida por $Q(\mathbf{x}) = \mu + A\mathbf{x}$.

Supongamos que A es invertible, entonces Q es biyectiva, $J_Q = \det(A)$ y por lo tanto, por el teorema del Jacobiano [1], Y es absolutamente continua con densidad:

$$f_y(\mathbf{y}) = f_{Q(x)}(\mathbf{y}) = f_X(A^{-1}(\mathbf{y} - \mu))\frac{1}{\sqrt{\det(\Sigma)}} = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}||A^{-1}(\mathbf{y}-\mu)||^2}$$

$$= \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(A^{-1}(\mathbf{y}-\mu))'A^{-1}(\mathbf{y}-\mu)} = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\mathbf{y}-\mu)'\Sigma^{-1}(\mathbf{y}-\mu)}$$

y por lo tanto

$$f_y(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}$$

[1] Si $y = Ax + \mu$ entonces $f_y(y) = f_x(A^{-1}(y - \mu))|det(A)^{-1}|$

Normal multivariada (teoremas)

Se puede probar:

1. Si $\mathbf{x} = (x_1, \ldots, x_d) \sim N(\mu, \Sigma)$ entonces $x_1, \ldots, x_d$ independientes si y sólo si son incorrelacionadas (es decir Σ es una matriz diagonal).

Se puede probar:

1. Si $\mathbf{x} = (x_1, \ldots, x_d) \sim N(\mu, \Sigma)$ entonces $x_1, \ldots, x_d$ independientes si y sólo si son incorrelacionadas (es decir Σ es una matriz diagonal).
   - Es bien conocido que si $x_1, \ldots, x_d$ son independientes entonces $x_1, \ldots, x_d$ están incorrelacionadas y esto implica que la matriz de varianzas del vector $\mathbf{x} = (x_1, \ldots, x_d)$ es diagonal.

Se puede probar:

1. Si $\mathbf{x} = (x_1, \ldots, x_d) \sim N(\mu, \Sigma)$ entonces $x_1, \ldots, x_d$ independientes si y sólo si son incorrelacionadas (es decir Σ es una matriz diagonal).

    - Es bien conocido que si $x_1, \ldots, x_d$ son independientes entonces $x_1, \ldots, x_d$ están incorrelacionadas y esto implica que la matriz de varianzas del vector $\mathbf{x} = (x_1, \ldots, x_d)$ es diagonal.
    - El recíproco no es siempre cierto en general, pero si x es gaussiano se cumple, pues

$$f_{\mathbf{x}}(x) = \prod_{i=1}^{d} f_{x_i}(x_i)$$

Se puede probar:

1. Si $\mathbf{x} = (x_1, \ldots, x_d) \sim N(\mu, \Sigma)$ entonces $x_1, \ldots, x_d$ independientes si y sólo si son incorrelacionadas (es decir Σ es una matriz diagonal).

    - Es bien conocido que si $x_1, \ldots, x_d$ son independientes entonces $x_1, \ldots, x_d$ están incorrelacionadas y esto implica que la matriz de varianzas del vector $\mathbf{x} = (x_1, \ldots, x_d)$ es diagonal.
    - El recíproco no es siempre cierto en general, pero si x es gaussiano se cumple, pues

$$f_{\mathbf{x}}(x) = \prod_{i=1}^{d} f_{x_i}(x_i)$$

2. Si $\mathbf{x}_1, \ldots, \mathbf{x}_n$ es una sucesión de vectores aleatorios i.i.d, con $E(\mathbf{x}_1) = \mu$ y matriz de covarianzas Σ entonces

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{x}_i - \mu) \xrightarrow[n]{D} N_d(0_{\mathbb{R}^d}, \Sigma)$$

La distribución normal de un vector (x_1, x_2) de media $\mu = (\mu_1, \mu_2)$ y matriz de covarianza

$$\Sigma = \begin{pmatrix} \sigma_1^2 & cov(x_1, x_2) \\ cov(x_2, x_1) & \sigma_2^2 \end{pmatrix}$$

tiene como densidad

$$f_{(x_1,x_2)}(x_1, x_2) = \frac{1}{(\sqrt{2\pi})^2\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x_1-\mu_1 \;\; x_2-\mu_2)\Sigma^{-1}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}\right)$$

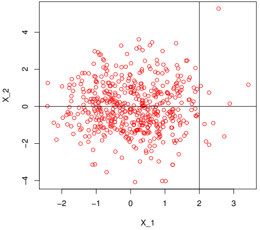


Figura: $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$$\mathbb{P}(X_2 > 0 | X_1 > 2) = \mathbb{P}(X_2 > 0)$$

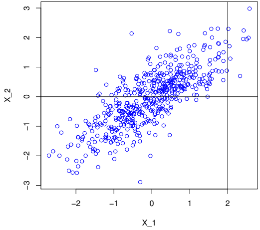


Figura: $\Sigma = \begin{pmatrix} 1 & 0{,}8 \\ 0{,}8 & 1 \end{pmatrix}$

$$\mathbb{P}(X_2 > 0 | X_1 > 2) > \mathbb{P}(X_2 > 0)$$

Sea $\mathbf{x} \sim N(\mu, \Sigma)$ entonces tenemos las siguientes propiedades:

1. La distribución es simétrica alrededor de μ.
   Esto es porque $f(\mu + a) = f(\mu - a)$.
2. La distribución tiene un único máximo en μ.
   Al ser Σ definida positiva, el término del exponente $(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)$ es siempre positivo, y $f(\mathbf{x})$ es máxima cuando este término es nulo, o sea si $\mathbf{x} = \mu$.
3. La media es μ y la matriz de varianzas-covarianzas es Σ.
4. Las distribuciones marginales son normales.
5. Si $\mathbf{x} \in \mathbb{R}^d$ es normal y $A \in \mathcal{M}_{k \times d}$ es una matriz entonces $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^k$ es normal.

Si cortamos con hiperplanos paralelos a las p variables, se obtienen curvas de nivel [2] cuya ecuación es

$$(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) = cte \quad \text{(elipsoides)}$$

y define entonces una medida de la distancia de $\mathbf{x}$ al centro μ.
Esta medida se llama distancia de Mahalanobis y

$$D^2 = (\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)$$

Supongamos que tenemos dos normales $N(0,1)$ y $N(10,10^2)$. Sea $x=3$, entonces:

1. La distancia euclidea de $x=3$ a 0 es más corta que la distancia de 3 a 10.
2. La distancia de Mahalanobis de $x=3$ a la distribución que tiene desviación típica 1 es $(3-0)1(3-0)=9$
3. La distancia de Mahalanobis de $x=3$ a la distribución que tiene desviación típica 10 es $(3-10)\frac{1}{100}(3-10)=0{,}49$

Con la distancia de Mahalanobis, el punto $x=3$ está más cerca de la segunda distribución. Es más probable que provenga de ella (lo veremos en breve).

[2] La curva de nivel α de una funcón $f:\mathbb{R}^d\to\mathbb{R}$ se define como $\mathcal{C}_\alpha=\{\mathbf{x}\in\mathbb{R}^d : f(\mathbf{x})=\alpha\}$

Distribución χ^2

Si $X_1, \ldots, X_d$ son variables aleatorias independientes con distribución normal $\mathcal{N}(0,1)$ entonces decimos que la variable $Z = X_1^2 + \cdots + X_d^2$ tiene distribución chi cuadrado con d grados de libertad.

Notamos $X \sim \chi^2(d)$

La función de densidad de una variable aleatoria con distribución $\chi^2(d)$ es

$$f(x) = \frac{2^{-\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} x^{\frac{d}{2}-1} e^{-\frac{x}{2}} \quad x > 0, d > 0$$

siendo $\Gamma(d) = \int_0^{+\infty} x^{d-1} e^{-x}\, dx$

Propiedades

Si $X \sim \chi^2(d)$ entonces:

1. $\mathbb{E}(X) = d$
2. $\mathrm{Var}(X) = 2d$

Propiedad

Si $\Sigma \in \mathcal{M}_{p\times p}$, la distancia de Mahalanobis asociada tiene distribución $\chi^2(p)$

Plan

Distribuciones mezcladas

Muchas veces los datos multivariantes provienen de poblaciones distintas. Si suponemos que tenemos G poblaciones, entonces

$$\mathbb{P}(\mathbf{x} \in A) = \sum_{i=1}^{G} \mathbb{P}(\mathbf{x} \in A, \mathbf{x} \in i) = \sum_{i=1}^{G} \mathbb{P}(\mathbf{x} \in A | \mathbf{x} \in i)\mathbb{P}(\mathbf{x} \in i) = \sum_{i=1}^{G} \pi_i \mathbb{P}(\mathbf{x} \in A | \mathbf{x} \in i)$$

y por lo tanto la función de densidad de la población es

$$f(\mathbf{x}) = \sum_{i=1}^{G} \pi_i f_i(\mathbf{x})$$

Distribuciones mezcladas

Muchas veces los datos multivariantes provienen de poblaciones distintas. Si suponemos que tenemos G poblaciones, entonces

$$\mathbb{P}(\mathbf{x} \in A) = \sum_{i=1}^{G} \mathbb{P}(\mathbf{x} \in A, \mathbf{x} \in i) = \sum_{i=1}^{G} \mathbb{P}(\mathbf{x} \in A|\mathbf{x} \in i)\mathbb{P}(\mathbf{x} \in i) = \sum_{i=1}^{G} \pi_i \mathbb{P}(\mathbf{x} \in A|\mathbf{x} \in i)$$

y por lo tanto la función de densidad de la población es

$$f(\mathbf{x}) = \sum_{i=1}^{G} \pi_i f_i(\mathbf{x})$$

Se prueba que

1. La media de una mezcla es

$$\mu = \sum_{i=1}^{G} \pi_i \mu_i$$

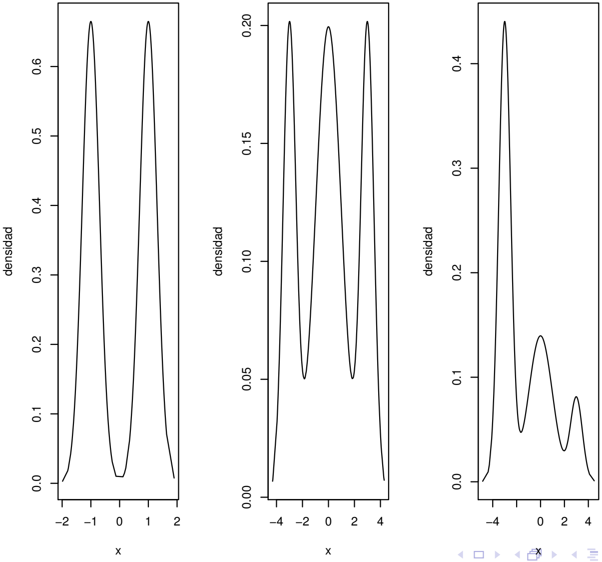
.

2. La matriz de varianzas y covarianzas de una mezcla es

$$V = \sum_{i=1}^{G} \pi_i V_i + \sum_{i=1}^{G} \pi_i (\mu_i - \mu)(\mu_i - \mu)'$$

donde V_i es la matriz de varianzas y covarianzas de la distribución i.

Mezcla de normales multivariadas



Mezcla de tres normales