



# **Estadística fundamental**

# **Distribuciones y simulación**

## **Máster en Data Science y Big Data (Finanzas)**

**Pilar Barrios**

Diciembre 2021

# Índice

---

1. Principales variables aleatorias y simulación
  - a. Variables aleatorias
  - b. Función de distribución y otros
  - c. Estadísticos
  - d. Simulación de variables aleatorias discretas
  - e. Variables aleatorias continuas
  - f. Simulación de variables aleatorias continuas
  - g. Modelos
  - h. Leyes de los grandes números y teorema del límite central
  - i. Intervalos de confianza

# Índice

---

## 2. Modelos multidimensionales y simulación

- a. Medidas de dependencia lineal
- b. El caso multidimensional: matriz de correlaciones
- c. Matrices de covarianza y de correlación muestrales
- d. La normal bidimensional (y multidimensional)

# 1 | Principales variables aleatorias y simulación

# 1.a Variables aleatorias

- Una **variable aleatoria** se define informalmente como un valor numérico que está afectado por el azar.
  - Es en definitiva una variable cuyos valores se obtienen de mediciones a partir de un experimento aleatorio.
- Formalmente una variable aleatoria es una función definida en el espacio muestral asociado a un experimento aleatorio, que asigna números reales a eventos de ese espacio muestral.
- Las variables aleatorias pueden ser:
  - Variables aleatorias **discretas**: Devuelve un conjunto de valores discreto.
  - Variables aleatorias **continuas**: El conjunto de posibles valores de la variable abarca un intervalo de números reales.

# 1.a Variables aleatorias

## Ejemplo:

- Se define el experimento de lanzar dos monedas al aire. El espacio muestral  $\Omega$  asociado a ese experimento es  $\Omega = \{cc, cx, xc, xx\}$  (“c” significa cara y “x” significa cruz).
- Sobre este espacio muestral  $\Omega$  se define la variable aleatoria  $K$  ( $K$  relaciona elementos de  $\Omega$  con números reales) consistente en asignar a cada suceso el número de caras obtenidas:
  - $K(cc)=2$
  - $K(cx)=K(xc)=1$
  - $K(xx)=0$

## 1.b Función de distribución y otros

- La **Distribución de Probabilidad** de una variable aleatoria es una función que asigna a cada suceso de la variable aleatoria la probabilidad de que dicho suceso ocurra.
  - Esta asignación se realiza mediante la **Función de Distribución**.
- La **Función de Distribución** es una función cuyo valor en  $x$  es la probabilidad de que la variable aleatoria sea menor que  $x$ .
  - Dada una variable aleatoria  $X$ , su función de distribución  $F_X$  es una función tal que:
$$F_X(x) = P(X \leq x)$$

En R las funciones de densidad, funciones de distribución, función cuantil y generación de números aleatorios de alguna distribución siguen la forma  $dxxx$ ,  $pxxx$ ,  $qxxx$  and  $rx$  respectivamente.

## 1.b Función de distribución y otros

Una descripción alternativa de una variable aleatoria  $X$ , más general y que será muy útil en simulación, viene dada por la función de distribución:

$$F_X(x) = P(X \leq x),$$

En general, una función de distribución cumple que

- $F_X(x) \leq F_X(y)$  si  $x \leq y$ ;
- $F_X(x) \rightarrow 0$  cuando  $x \rightarrow -\infty$ ;
- $F_X(x) \rightarrow 1$  cuando  $x \rightarrow +\infty$ ;

En el caso discreto,

$$F_X(x) = \sum_{j / x_j \leq x} p_j.$$

Nos referiremos a estas probabilidades como **probabilidades acumuladas**.

Obsérvese que no es más que una manera alternativa de escribir las probabilidades.

Las funciones de distribución, en este caso discreto, son funciones “escalonadas”.



# 1.b Función de distribución y otros

Variable X

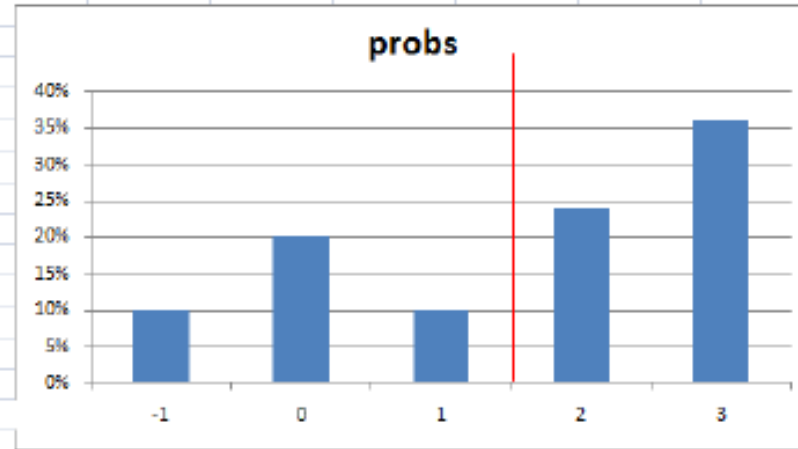
valores	probabilidades	probs acumuladas
-1	10%	10%
0	20%	30%
1	10%	40%
2	24%	64%
3	36%	100%

$E(X)$  1,56

$E(X^2)$  =SUMAPRODUCTO[valores^2;probs]

$V(X)$  1,9664

$\sigma(X)$  1,402283851



## 1.b Función de distribución y otros

- La función de distribución para **variables aleatorias discretas** se calcula como:

$$F(x) = P(X \leq x) = \sum_{k=-\infty}^x P(X = k)$$

- La función de distribución para **variables aleatorias continuas** se define como:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

- La función **f(t)** es la **función de densidad** de la probabilidad, la cual especifica la posibilidad relativa de que una variable aleatoria continua **X** tome un valor “cercano” a **x**.
  - El análogo a la función de densidad en distribuciones discretas es la **función de probabilidad**.

## 1.c. Estadísticos

- La manera de analizar y conocer una distribución es mediante el uso de **Estadísticos**. Un estadístico es una función que relaciona una distribución o muestra, con un número.
- Los estadísticos principales que permiten conocer una distribución son:
  - Media
  - Varianza y desviación estándar
  - Asimetría
  - Kurtosis
  - Moda
  - Percentiles y Mediana

# 1.c. Estadísticos

## Momentos de una variable aleatoria

Se trata de las cantidades

$E(X), E(X^2), E(X^3), E(X^4), \dots$

(el primero es la media), o bien de los llamados momentos centrados

$$\mu_2 = E([X - E(X)]^2), \mu_3 = E([X - E(X)]^3), \mu_4 = E([X - E(X)]^4), \dots$$

El momento  $\mu_2$  es la varianza. Son de uso habitual las dos siguientes cantidades, relacionadas con los momentos tercero y cuarto:

- Skewness o asimetría:  $\mu^3/\mu_2^{3/2}$
- (exceso de) kurtosis:  $\mu^4/\mu_2^2 - 3$

# 1.c. Estadísticos

## Momento de orden 1 (Media):

- La Media, Valor Esperado ó Esperanza Matemática de una variable aleatoria  $X$  se denota por  $E[X]$ .
- Representa una especie de “punto de equilibrio”.
- La esperanza matemática para una distribución discreta cuyos posibles valores son  $\{x_1, x_2, \dots, x_n\}$  se calcula como:

$$E[X] = \sum_{k=-\infty}^n x_i P(X = x_i)$$

- La esperanza matemática para una distribución continua cuya función de densidad es  $f(x)$  se calcula como:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

# 1.c. Estadísticos

## Momento de orden 1 (*Media*):

- La esperanza matemática es un operador lineal, y por tanto tiene las siguientes propiedades. Para  $X$  e  $Y$  variables aleatorias y  $c$  una constante:
  - $E[X+c]=E[X]+c$
  - $E[X+Y]=E[X]+E[Y]$
  - $E[cX]=cE[X]$

# 1.c. Estadísticos

## Momento de orden 2 (Varianza):

- La varianza de una variable aleatoria es una medida de dispersión que se define como la media aritmética del cuadrado de las desviaciones respecto a su media.

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

- Las unidades de medida de la varianza son distintas a las de la variable aleatoria, ya que la varianza es del orden de la variable elevada al cuadrado. Una medida de dispersión alternativa expresada en las mismas unidades que la variable aleatoria es la **Desviación Estándar (o desviación típica)**, la cual se define como la raíz cuadrada de la varianza:

$$STD[X] = \sqrt{Var(X)}$$

# 1.c. Estadísticos

## Momento de orden 2 (Varianza):

- Las propiedades de la Varianza son, para  $X$ ,  $Y$  variables aleatorias y  $a$  y  $b$  constantes:
  - $V(X) \geq 0$
  - $V(aX+b) = a^2 V(X)$
  - $V(X+Y) = V(X) + V(Y) + 2\text{Cov}(X, Y)$
  - $V(X-Y) = V(X) + V(Y) - 2\text{Cov}(X, Y)$
- $\text{Cov}(X, Y)$  es la **Covarianza** de  $X$  e  $Y$ . Es una medida de **dispersión conjunta entre dos variables aleatorias que se define como:**

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

- La covarianza de una variable consigo misma es la varianza:

$$\text{Cov}(X, X) = \text{Var}(X)$$



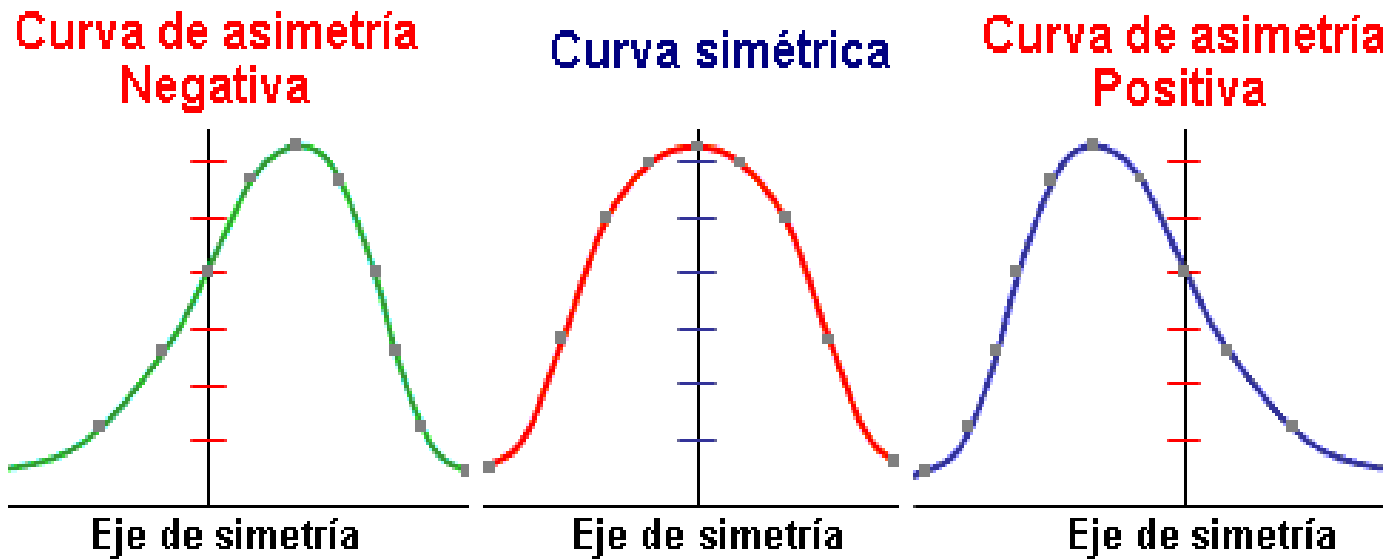
# 1.c. Estadísticos

## Relacionado con el momento de orden 3 (Asimetría):

- El coeficiente de asimetría es un estadístico que permite medir la asimetría de una distribución.
- Se dice que una distribución es simétrica si tiene la misma masa de probabilidad a ambos lados de la media.
- La asimetría es positiva si la distribución de probabilidad asigna más probabilidad a los datos que hay a la derecha de la media.
- Se dice que la asimetría es negativa si, al contrario, la distribución de probabilidad asigna más probabilidad a los datos que se encuentran a la izquierda de la media.

$$\text{Asimetria}[X] = E[(X - E[X])^3]/\sigma^3$$

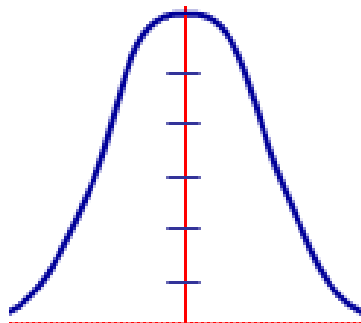
# 1.c. Estadísticos



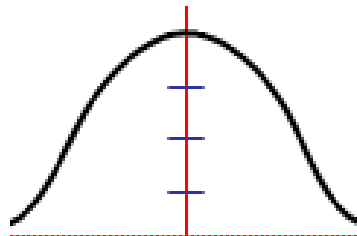
## 1.c. Estadísticos

### Relacionado con el momento de orden 4 (Kurtosis):

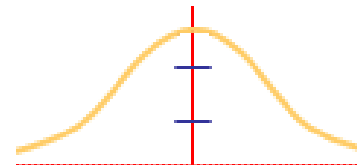
- La kurtosis mide cómo están concentrados los valores de la distribución en torno a la región central.
- Se dirá que la concentración es *Leptocúrtica* si el grado de concentración de valores en la región central es elevado.
- Se dirá que la distribución es *Mesocúrtica* si es normal.
- Y por último, si el grado de concentración de valores en la región central es bajo se dirá que la distribución es *Platicúrtica*



**Leptocúrtica**



**Mesocúrtica**



**Platicúrtica**

# 1.c. Estadísticos

## Moda:

- Se define como Moda estadística al valor con mayor frecuencia de la distribución.
- En una distribución puede haber más de una moda.
- Si existen dos modas, se dirá que la distribución es *bimodal*. Si existen 3 modas, se dirá que la distribución es *Trimodal*.
- En caso de que todos los valores tengan la misma frecuencia se dirá que la distribución no tiene Moda.

## Percentil:

- Se define como Percentil  $i$ -ésimo (donde  $i$  toma valores del 1 al 99) al número tal que los números menores que él suman un  $i\%$  de probabilidad (y los mayores un  $(100-i)\%$ ). Valdría también la definición para porcentajes no enteros.
- El percentil 50 es la **Mediana** estadística.

# 1.d. Simulación de variables aleatorias discretas

Una variable aleatoria, una especificación de valores y probabilidades, es un modelo probabilístico que pretende recoger la esencia de un cierto experimento aleatorio.

Simular consiste en obtener (muchas) muestras de la variable de manera que las frecuencias relativas con las que se obtiene cada uno de los valores posibles se aproximen a las probabilidades del modelo. Es decir, que el **histograma** muestral se asemeje a la distribución de probabilidades del modelo.

# 1.d. Simulación de variables aleatorias discretas

Muestras aleatorias con probabilidad discreta preelegida. Función R:  
`sample(x, tamaño, replace = FALSE, prob = NULL)`.

## Argumentos:

- > `x`: vector de más de un elemento (real, complejo, carácter o lógico) del que elegir las ocurrencias. O un entero positivo, en cuyo caso se elige del conjunto `1:x`
- > `tamaño`: entero no negativo que es el número de ocurrencias o extracciones a realizar.
- > `replace` si la extracción se hace o no con reemplazamiento.
- > **`prob`= vector de pesos a asignar a cada uno de los posibles valores que se extraen del conjunto especificado por `x`. Por defecto, todos los valores resultantes de `x` tienen la misma probabilidad.**

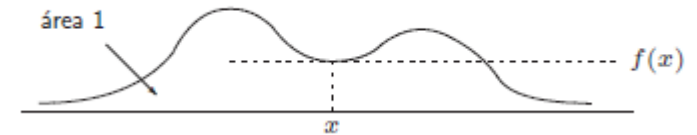
**Ejemplo:** Simulación de un dado

```
dadoBueno=sample(c('1','2','3','4','5','6'), 100, replace = TRUE);  
table(dadoBueno);  
dadoBueno  
barplot(table(dadoBueno))  
dadoTrucoNum=sample(c(1:6), 1000, replace = TRUE,prob = c(2,3,1,9,8,14))
```

# 1.e. Variables aleatorias continuas

Una variable continua  $X$  toma valores en  $\mathbb{R}$ .

Esto requiere algo más de técnica: en lugar de “probabilidades”, consideramos la función de densidad  $f(x)$ , una función no negativa ( $f(x) \geq 0$ ) y que encierra área 1,

$$\int_{-\infty}^{\infty} f(x)dx = 1$$


Ahora  $f(x)$  no representa la probabilidad de obtener el valor  $x$  en el experimento aleatorio que estamos modelando. Podemos interpretar el significado de  $f(x)$  de la siguiente manera: si  $h$  es muy pequeño, entonces

$f(x)h$  es, aproximadamente,  $P(x \leq X \leq x+h)$ .

Las probabilidades se calculan integrando: para todo  $a < b$ ,

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

# 1.e. Variables aleatorias continuas

## Cálculo de medias, varianzas y otros momentos

Las definiciones son las mismas que antes, pero ahora se calculan integrando:

Media:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

Media de una transformación de X:

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx.$$

Varianza:

$$\begin{aligned} V(X) &= E[(X - E(X))^2] = E(X^2) - E(X)^2 \\ &= \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \left( \int_{-\infty}^{\infty} xf(x)dx \right)^2. \end{aligned}$$



# 1.e. Variables aleatorias continuas

## Función de distribución

La definición es la misma,  $F_X(x) = P(X \leq x)$ , pero sus valores se obtienen integrando la función de densidad (hasta  $x$ ):

$$F_X(x) = \int_{-\infty}^x f(y)dy.$$

NOTA. La función de densidad y la función de distribución describen completamente el modelo. Son descripciones alternativas. En el caso discreto, se pasa de una a otra sumando y restando.

En el caso continuo, si disponemos de la función de densidad  $f(x)$ , obtenemos la de distribución integrando, como hemos visto arriba.

Si disponemos de  $F(x)$ , entonces recuperamos la de densidad derivando:

$$F'(x) = f(x)$$

# 1.f. Simulación de variables aleatorias continuas

El mecanismo universal de simulación requiere saber calcular la inversa de la función de distribución  $F(x)$ .

NOTA. La "función"  $F^{-1}$  es a veces llamada la **función cuantil**:

- si  $F(x)$  es continua y estrictamente creciente, entonces la función cuantil es la inversa de  $F(x)$ ;
- en general, se define como

$$F^{-1}(p) = \inf\{x \in \mathbb{R} : p \leq F(x)\}$$

(en el caso de distribuciones discretas, utilizamos la función `buscarv`).

Consta, de nuevo, de dos pasos:

1. Generamos una lista de números  $u_1, \dots, u_n$  con arreglo a la uniforme  $[0,1]$ .
2. Con la función inversa de  $F$ , que llamaremos  $F^{-1}$ , calculamos la lista de números  $x_1 = F^{-1}(u_1)$ ,  $x_2 = F^{-1}(u_2)$ ,  $\dots$ ,  $x_n = F^{-1}(u_n)$ .

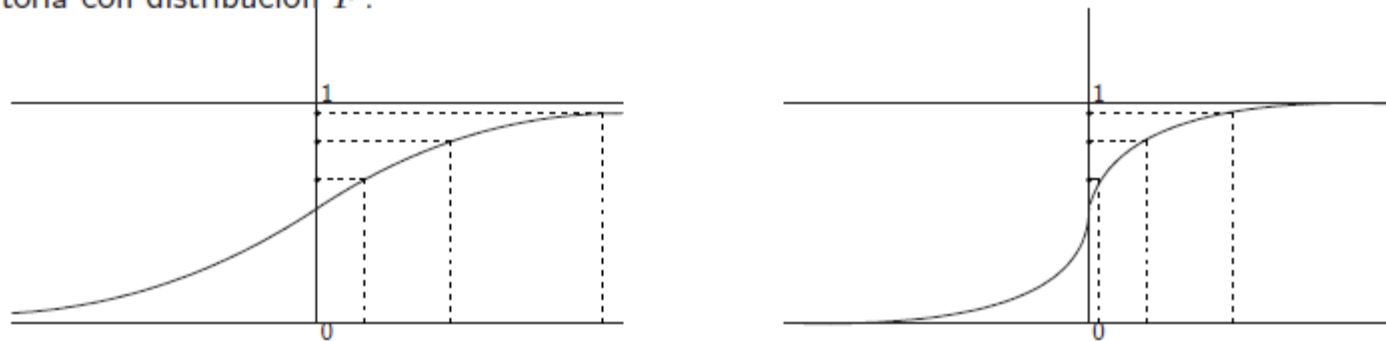
Estos números  $x_1, \dots, x_n$  ya son muestras de la variable  $X$ .

# 1.f. Simulación de variables aleatorias continuas

## Nota

**JUSTIFICACIÓN.** Se basa en el segundo de estos resultados:

- (a) Si  $X$  es una variable aleatoria con función de distribución  $F$ , entonces  $U = F(X)$  es una variable aleatoria uniforme.
- (b) O al revés: si  $U$  es una variable aleatoria uniforme, entonces  $X = F^{-1}(U)$  es una variable aleatoria con distribución  $F$ .



Lo que no está claro es cómo calcular  $F^{-1}$ . En ocasiones, dispondremos de una fórmula explícita, en otras necesitaremos recurrir a métodos numéricos de cálculo.

# 1.g. Modelos

Nos disponemos a analizar las características (función de densidad, de distribución, medias y varianzas) de los modelos aleatorios más habituales en el contexto financiero. Avanzaremos también algunos de los contextos en los que tienen utilidad. Además, diseñaremos métodos de simulación para cada uno de ellos.

- Variable Bernoulli  $B(p)$
- Variable discreta general
- Variable uniforme  $U[a, b]$
- Variable normal  $N(\mu, \sigma)$
- Variable binomial  $\text{Bin}(n, p)$
- Variable geométrica  $\text{Geom}(p)$
- Variable Poisson *Poisson*  $(\lambda)$
- Variable hipergeométrica
- Variable multinomial
- Variable  $\chi^2_v$
- Variable t de Student  $t_v$
- Variable exponencial

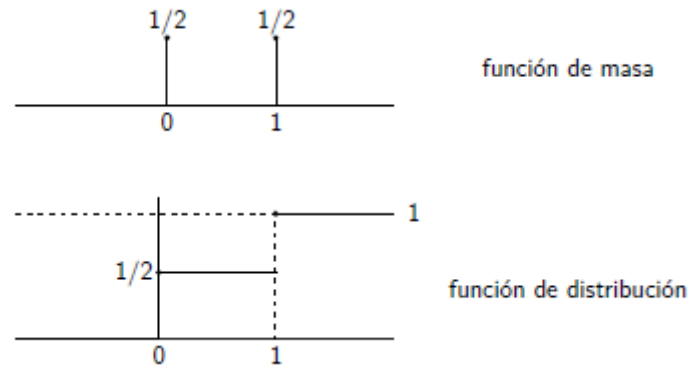
# 1.g. Modelos

## Cara y cruz: variable Bernoulli

La moneda regular,  $BER(1/2)$ :

$$X \rightarrow \begin{cases} 0 & \text{probabilidad } 1/2 \\ 1 & \text{probabilidad } 1/2 \end{cases}$$

X sigue una  $BER(p)$  si toma el valor 1 con probabilidad  $p$  y el 0 con  $1 - p$ . Su media y varianza son  $E(X) = p$  y  $V(X) = p(1 - p)$ .



# 1.g. Modelos

## Variable discreta general

Más generalmente, podemos considerar una variable aleatoria  $X$  que tome una lista de valores  $x_1, \dots, x_n$  con probabilidades  $p_1, \dots, p_n$ .

La media y la varianza son:

$$E(X) = \sum_{j=1}^n x_j p_j \qquad V(X) = \sum_{j=1}^n x_j^2 p_j - \left( \sum_{j=1}^n x_j p_j \right)^2$$

Para simular valores de  $X$ , sólo tenemos que definir los “cortes” en el intervalo  $[0, 1]$  dados por  $p_1, p_1 + p_2, p_1 + p_2 + p_3, \dots$

(esto es, las probabilidades acumuladas), junto con el generador de la uniforme aleatorio().

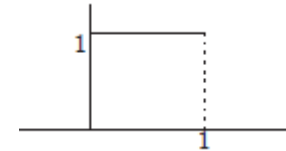
# 1.g. Modelos

## La distribución uniforme

```
dunif(x, min = 0, max = 1, log = FALSE)
punif(q, min = 0, max = 1, lower.tail = TRUE, log.p = FALSE)
qunif(p, min = 0, max = 1, lower.tail = TRUE, log.p = FALSE)
runif(n, min = 0, max = 1)
```

Función de densidad

$$f_X(x) = \begin{cases} 1 & \text{si } x \in [0,1] \\ 0 & \text{resto} \end{cases}$$



Función de distribución

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$



Media y varianza

$$E(X) = \int_0^1 x dx = 1/2$$

$$V(X) = \int_0^1 x^2 dx - 1/4 = 1/3 - 1/4 = 1/12$$

# 1.g. Modelos

## La distribución normal

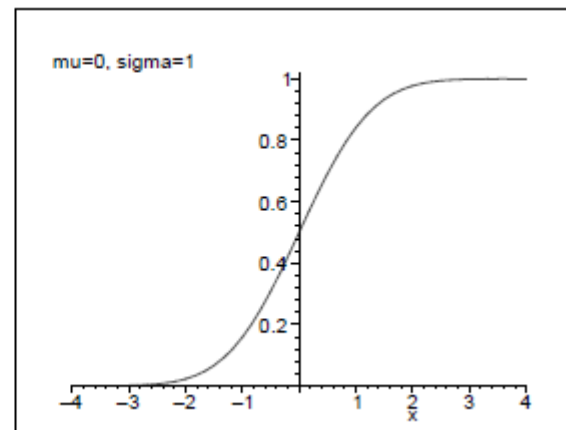
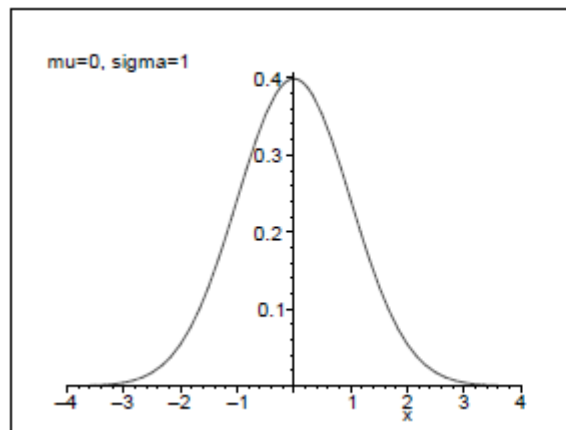
```
dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
rnorm(n, mean = 0, sd = 1)
```

Decimos que una variable aleatoria  $X$  sigue una distribución normal estándar, lo que denotaremos como  $N(0,1)$ , si tiene como funciones de densidad y distribución a

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

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$$

Una tal variable tiene media 0 y varianza (y desviación típica) 1.





## La distribución normal

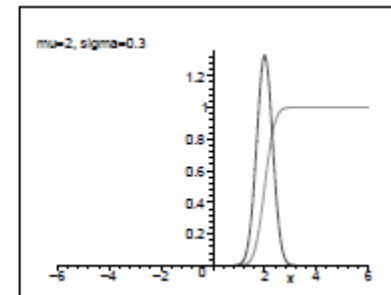
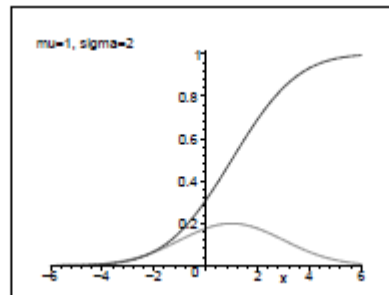
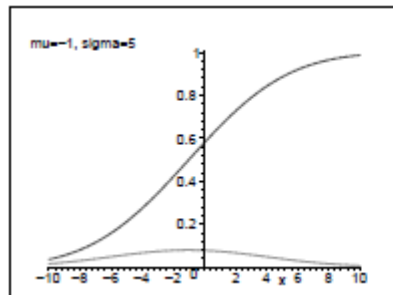
En general, tenemos dos parámetros a nuestra disposición, que denotamos habitualmente por  $\mu$  y  $\sigma > 0$ . Las funciones de densidad y distribución de una normal  $N(\mu, \sigma)$  son

$$\phi_{\mu, \sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad \Phi_{\mu, \sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right) dy$$

Los parámetros  $\mu$  y  $\sigma$  resultan ser la media y la desviación típica de la variable  $X$ :

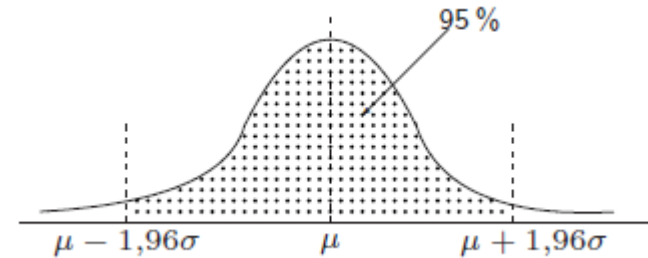
$$E(X) = \int_{-\infty}^{\infty} x\phi_{\mu, \sigma}(x)dx = \mu; \quad V(X) = \int_{-\infty}^{\infty} x^2\phi_{\mu, \sigma}(x)dx - \mu^2 = \sigma^2.$$

**Nota:** Una variable  $X$  normal  $N(\mu, \sigma)$  se puede escribir como  $X = \mu + \sigma Y$ , donde  $Y$  es  $N(0, 1)$ . El parámetro  $\mu$  desplaza el centro de la distribución, el parámetro  $\sigma$  la “ensancha” o la “estrecha”:



# La distribución normal

De hecho, el parámetro  $\sigma$  nos da una idea precisa de “dónde” pueden estar, con muy alta probabilidad, los números sorteados de una normal  $N(\mu, \sigma)$ . Véase, por ejemplo, la figura de la derecha.



**Simulación:** para evaluar  $\Phi^{-1}$  hay que recurrir a aproximaciones numéricas.

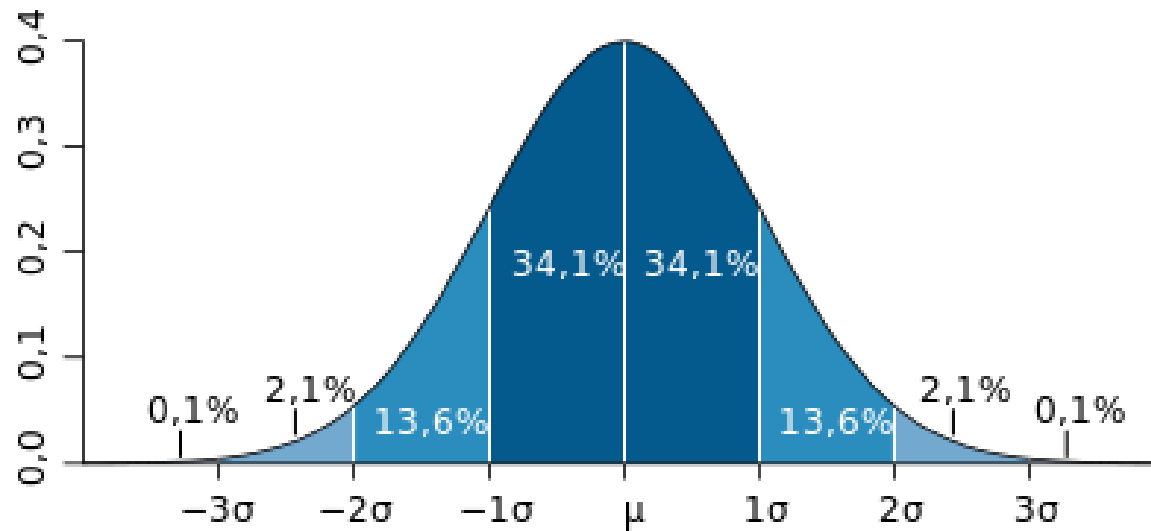
# La distribución normal

## Propiedades

- Es simétrica con respecto a su media
- Su rango de valores va desde  $-\infty$  a  $\infty$
- Es **reescalable y trasladable**: Si  $X$  es una variable aleatoria de media  $\mu$  y desviación estándar  $\sigma$   $N(\mu, \sigma)$  y  $a$  y  $b$  son constantes, entonces:
  - $aX + b \sim N(a\mu + b, a\sigma)$
- Es **sumable**: Si  $X$  es una variable aleatoria  $N(\mu_1, \sigma_1)$  e  $Y$  es una variable aleatoria  $N(\mu_2, \sigma_2)$  entonces su suma:
  - $X + Y \sim N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$
- Estandarización de variables normales: si  $X$  es una variable aleatoria  $N(\mu, \sigma)$ , entonces:
  - $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$

# La distribución normal

La distribución de la probabilidad de la distribución normal a lo largo de su media es la siguiente:



# La distribución normal

Es posible tabular el número de desviaciones típicas a partir de la media que devuelve cada percentil:

Percentil	Número de desviaciones
80%	1.28155
90%	1.64485
95%	1.95996
98%	2.32635
99%	2.57583
99.5%	2.80703
99.8%	3.09023

# 1.g. Modelos

## Distribución Log-Normal

```
dlnorm(x, meanlog = 0, sdlog = 1, log = FALSE)
plnorm(q, meanlog = 0, sdlog = 1, lower.tail = TRUE, log.p = FALSE)
qlnorm(p, meanlog = 0, sdlog = 1, lower.tail = TRUE, log.p = FALSE)
rlnorm(n, meanlog = 0, sdlog = 1)
```

La distribución Log-normal es una distribución continua aplicable a cualquier variable aleatoria cuyo logaritmo se distribuya mediante una normal.

$$X \sim \text{LogNormal}(\mu, \sigma) \Leftrightarrow \text{Ln}(X) \sim N(\mu, \sigma)$$

### Función de densidad:

La función de densidad de una distribución log normal viene definida por:

$$f_{\mu, \sigma}(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$

### Función de Distribución:

La función de Distribución viene dada por la función de distribución de la normal. Si X es lognormal y Z su normal asociada, entonces:

$$P(X < x) = P(Z < \text{Ln}(x))$$

# Distribución Log-Normal

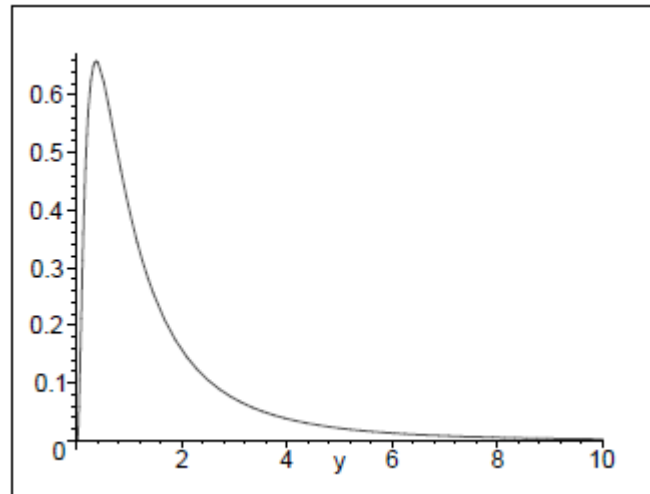
## Estadísticos:

Media:  $E[X] = e^{\mu + \frac{\sigma^2}{2}}$

Varianza:  $Var[X] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$

## Propiedades

Al contrario que la distribución normal, la distribución lognormal no es sumable, ni trasladable ni reescalable ni simétrica.



# 1.g. Modelos

Sumamos los resultados de cara y cruz:

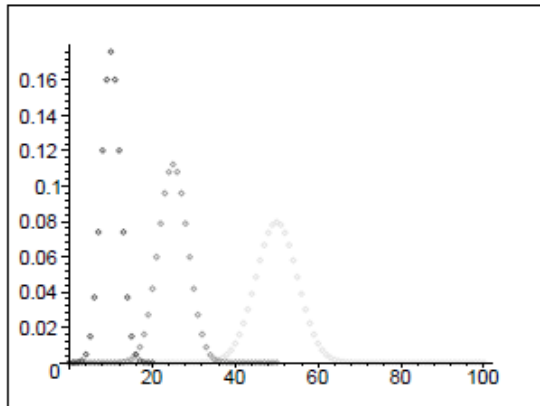
la distribución binomial

Si  $X$  sigue una distribución  $\text{Bin}(n, p)$ , entonces, para cada  $j = 0, \dots, n$ ,

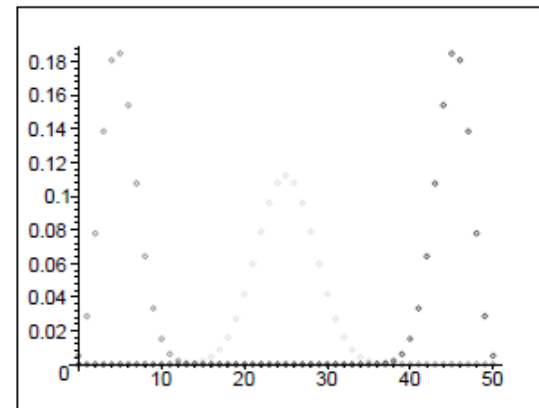
$$P(X = j) = \binom{n}{j} p^j (1-p)^{n-j}.$$

Al parámetro  $n$  nos referiremos como “número de repeticiones”; al parámetro  $p$ , como “probabilidad de éxito”.  $X$  cuenta número de “éxitos”. La media y varianza son  $E(X) = np$  y  $V(X) = np(1-p)$ .

Cómo cambia cuando  $n$  crece:



Cómo cambia cuando  $p$  se mueve en  $[0, 1]$ :



```
dbinom(x, size, prob, log = FALSE)
pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)
qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)
rbinom(n, size, prob)
```



# 1.g. Modelos

## Esperamos hasta el primer éxito: la distribución geométrica

```
dgeom(x, prob, log = FALSE)
pgeom(q, prob, lower.tail = TRUE, log.p = FALSE)
qgeom(p, prob, lower.tail = TRUE, log.p = FALSE)
rgeom(n, prob)
```

Un solo parámetro  $p$ . Una variable  $X$  sigue una  $GEOM(p)$  si toma valores  $j=1, 2, 3, 4, \dots$  con probabilidades

$$P(X = j) = p(1 - p)^{j-1}.$$

La media y varianza son  $E(X) = 1/p$  y  $V(X) = (1 - p)/p^2$ .

$X$  es un tiempo de espera (por ejemplo hasta que sale la primera cara, hasta que una empresa con una probabilidad de default,  $p$ , entra en quiebra)

# 1.g. Modelos

## La distribución de Poisson

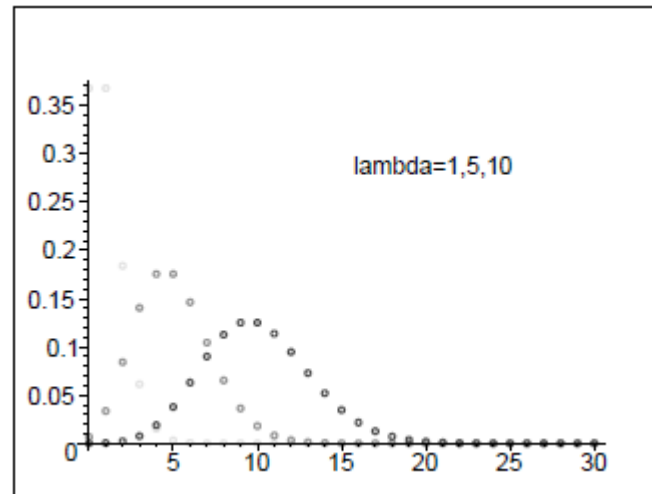
```
dpois(x, lambda, log = FALSE)
ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)
rpois(n, lambda)
```

Fijamos un parámetro  $\lambda > 0$ . La variable aleatoria  $X$  toma valores  $\{0, 1, 2, 3, \dots\}$  con probabilidades

$$P(X = j) = \frac{\lambda^j}{j!} e^{-\lambda}.$$

La media y varianza son  $E(X) = V(X) = \lambda$ .

Cómo cambia con  $\lambda$ :



## La distribución de Poisson

- Simular 1000 números extraídos de una Poisson, para distintos valores de  $\lambda$  (con la tabla de valores y probabilidades, convenientemente truncada). Ver histogramas.
- Fijado  $\lambda$ , generar muestras de una binomial de parámetros  $n$  y  $p$ . Manteniendo fijo el producto  $n \times p = \lambda$ , hacer  $n$  cada vez más grande. Comparar el resultado con los histogramas obtenidos en el ejercicio anterior.

La razón del comportamiento observado en el último ejercicio (y una de las utilidades básicas de la Poisson) es que si  $n$  es muy grande y  $np = \lambda$ , entonces

$$P(BIN(n, p) = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} = P(Poisson(\lambda) = k)$$

Así que, si  $n$  es grande,  $p$  pequeño, y su producto es  $\lambda$ , la Binomial de parámetros  $(n, p)$  es “muy parecida” a la Poisson de parámetro  $\lambda$ .

# 1.g. Modelos

## La distribución hipergeométrica

```

dhyper(x, m, n, k, log = FALSE)
phyper(q, m, n, k, lower.tail = TRUE, log.p = FALSE)
qhyper(p, m, n, k, lower.tail = TRUE, log.p = FALSE)
rhyper(nn, m, n, k)

```

Es una distribución discreta relacionada con muestreos aleatorios y sin reemplazamiento. Supóngase que se tiene una población de  $N$  elementos de los cuales,  $d$  pertenecen a la categoría A y  $N-d$  a la B. La distribución hipergeométrica mide la probabilidad de obtener  $x$  ( $0 \leq x \leq d$ ) elementos de la categoría A en una muestra sin reemplazamiento de  $n$  elementos de la población original.

$$P(X = x) = \frac{\binom{d}{x} \binom{N-d}{n-x}}{\binom{N}{n}}$$

donde  $N$  es el tamaño de la población y  $n$  el de la muestra extraída,  $d$  es el número de elementos en la población original que pertenecen a la categoría deseada y  $x$  es el número de elementos en la muestra que pertenecen a dicha categoría.

La media y la varianza son

$$E(X) = \frac{nd}{N} \quad V(X) = \left(\frac{N-n}{N-1}\right) \left(\frac{nd}{N}\right) \left(1 - \frac{d}{N}\right)$$

Definiendo  $p=d/N$  y  $q=1-p$ , se tiene que

$$E(X) = np \quad V(X) = npq \left(\frac{N-n}{N-1}\right)$$

## La distribución hipergeométrica

$$P(X = x) = \frac{\binom{d}{x} \binom{N-d}{n-x}}{\binom{N}{n}}$$

$$E(X) = \frac{nd}{N} \quad V(X) = \left(\frac{N-n}{N-1}\right) \left(\frac{nd}{N}\right) \left(1 - \frac{d}{N}\right)$$

$$E(X) = np \quad V(X) = npq \left(\frac{N-n}{N-1}\right)$$

La distribución hipergeométrica es aplicable a muestreos sin reemplazamiento y la binomial a muestreos con reemplazamiento.

En situaciones en las que el número esperado de repeticiones en el muestreo es presumiblemente bajo, puede aproximarse la primera por la segunda.

Esto es así cuando  $N$  es grande y el tamaño relativo de la muestra extraída,  $n/N$ , es pequeño.

# 1.g. Modelos

## La distribución multinomial

```
rmultinom(n, size, prob)  
dmultinom(x, size = NULL, prob, log = FALSE)
```

Es una generalización de la distribución binomial.

La distribución binomial es la probabilidad de un número de éxitos en  $N$  sucesos de Bernoulli independientes, con la misma probabilidad de éxito en cada suceso.

En una distribución multinomial, el análogo a la distribución de **Bernoulli** es la distribución **categorica**, donde cada suceso concluye en únicamente un resultado de un número finito  $K$  de los posibles, con probabilidades  $p_1, \dots, p_K$  (tal que  $p_i \geq 0$  para  $i=1, \dots, K$  y  $p_1 + \dots + p_K = 1$ ); y con  $n$  sucesos independientes.

Entonces sea la variable aleatoria  $X_i$ , que indica el número de veces que se ha dado el resultado  $i$  sobre los  $n$  sucesos. El vector  $X=(X_1, \dots, X_K)$  sigue una distribución multinomial con parámetros  $n$  y  $p$ , donde  $p = (p_1, \dots, p_K)$ .

## La distribución multinomial

La función de probabilidad de la distribución multinomial es:

$$P(X_1 = x_1, \dots, X_k = x_k) = \begin{cases} \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} & \text{cuando } \sum_{i=1}^k x_i = n \\ 0 & \text{resto} \end{cases}$$

Para enteros no negativos  $x_1, \dots, x_k$ .

La esperanza matemática del suceso  $i$  observado en  $n$  pruebas es:

$$E(X_i) = np_i.$$

Y su varianza es:

$$V(X_i) = np_i(1 - p_i).$$

# 1.g. Modelos

## La distribución $\chi^2_v$

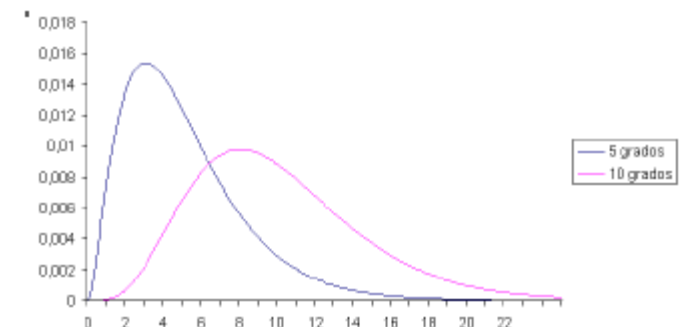
```
dchisq(x, df, ncp = 0, log = FALSE)
pchisq(q, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)
qchisq(p, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)
rchisq(n, df, ncp = 0)
```

Fijamos un entero positivo  $v$ . Una variable sigue una distribución  $\chi^2_v$  (se lee “ji” o “chi” cuadrado) con  $v$  grados de libertad si se escribe como la suma

$$\sum_{j=1}^v Z_j^2,$$

donde las variables  $Z_1, \dots, Z_v$  son normales  $N(0, 1)$  (independientes). Esta definición permite simular directamente valores de la  $\chi^2_v$  (simulando los de las correspondientes  $Z_j$ ). Obsérvese que una  $\chi^2_v$  solo toma valores positivos. Su media y varianza son:

$$E(\chi^2_v) = v \qquad V(\chi^2_v) = 2v$$



Una distribución  $\chi^2_v$  centrada es igual a una distribución Gamma con parámetro de forma  $a = v/2$  y parámetro de escala  $s = 2$ . Una distribución  $\chi^2_2$  centrada es un caso especial de distribución exponencial.



# 1.g. Modelos

## La distribución $t_v$ de Student

```
dt(x, df, ncp, log = FALSE)
pt(q, df, ncp, lower.tail = TRUE, log.p = FALSE)
qt(p, df, ncp, lower.tail = TRUE, log.p = FALSE)
rt(n, df, ncp)
```

Si  $Y$  es  $N(0, 1)$  y  $X$  es una  $\chi^2_v$  con  $v$  grados de libertad (independientes entre sí),

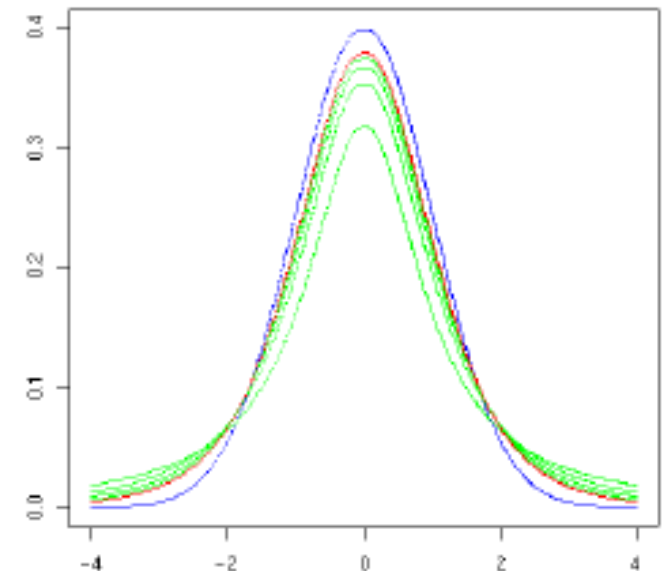
$$Z = \sqrt{v} \frac{Y}{\sqrt{X}}$$

sigue una  $t_v$  de Student.

Su media y varianza son

$$E(Z) = 0 \text{ y } V(Z) = v/(v-2) \text{ (para } v > 2).$$

Es simétrica con respecto al origen (skewness 0) y tiene colas más gruesas que la normal (exceso de curtosis  $6/(n - 4)$  para  $n > 4$ ). Cuando  $n$  es grande, la  $t_n$  se asemeja a una  $N(0, 1)$ . En el gráfico de la derecha la normal es la línea azul, mientras que las demás representan  $t$  de Student con parámetros crecientes (en rojo, la de  $n = 10$ ).



# 1.g. Modelos

## Tiempo de supervivencia: la exponencial

Para  $\lambda > 0$ , una variable  $T$  es una  $EXP(\lambda)$  si

$$(f. \text{ densidad}) \quad f_T(t) = \begin{cases} 0 & t \leq 0 \\ \lambda e^{-\lambda t} & t > 0 \end{cases}$$

$$(f. \text{ distribucion}) \quad F_T(t) = \begin{cases} 0 & t \leq 0 \\ 1 - e^{-\lambda t} & t > 0 \end{cases}$$

La media es  $E(T) = 1/\lambda$  y la varianza,  $V(T) = 1/\lambda^2$

Si  $T$  modeliza el momento de default de una empresa,

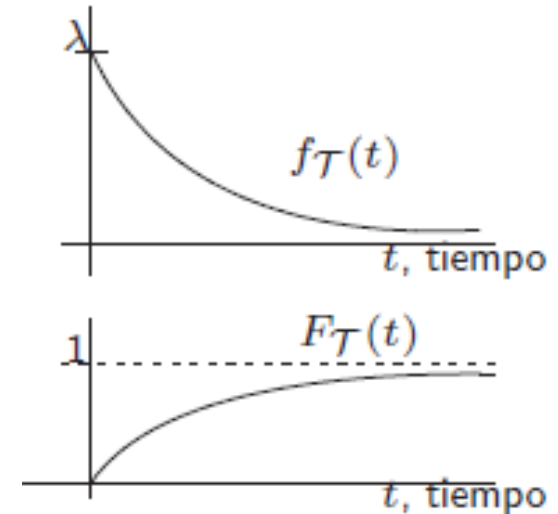
$$P(T > t) = e^{-\lambda t}$$

es la probabilidad de que no haya default hasta  $t > 0$ . En las aplicaciones, el parámetro  $\lambda$  es la intensidad. Si  $\lambda$  es grande, esperamos que el default se produzca pronto.

**Simulación:** generamos  $u$  (uniforme en  $[0, 1]$ ) y entonces

$$t = \frac{1}{\lambda} \ln \left( \frac{1}{1-u} \right)$$

```
dexp(x, rate = rate, log = FALSE)
pexp(q, rate = rate, lower.tail = TRUE, log.p = FALSE)
qexp(p, rate = rate, lower.tail = TRUE, log.p = FALSE)
rexp(n, rate = rate)
```



# 1.h. Concentración de promedios. Leyes de los grandes números y TLC

Si  $X_1, \dots, X_n$  son variables aleatorias i.i.d. con media común  $\mu$  y desviación típica común  $\sigma$ , entonces

$$\text{(suma)} \quad S_n = \sum_{j=1}^n X_j \quad \longrightarrow \quad \begin{cases} \mathbf{E}(S_n) = n\mu \\ \mathbf{V}(S_n) = n\sigma^2 \end{cases} \quad y \quad \sigma(S_n) = \sqrt{n} \sigma$$

$$\text{(promedio)} \quad Y_n = \frac{1}{n} \sum_{j=1}^n X_j \quad \longrightarrow \quad \begin{cases} \mathbf{E}(Y_n) = \mu \\ \mathbf{V}(Y_n) = \frac{\sigma^2}{n} \end{cases} \quad y \quad \sigma(Y_n) = \frac{\sigma}{\sqrt{n}}$$

(En realidad, las expresiones de las varianzas son válidas sin más que las variables  $X_i$  sean incorreladas).

# Concentración de promedios. Leyes de los grandes números y TLC

**DETALLE MATEMÁTICO.** Supongamos que  $X_1, \dots, X_n$  son variables **incorreladas**. Esto es,  $E(X_i X_j) = E(X_i)E(X_j)$  si  $i \neq j$ . Entonces

$$V\left(\sum_{j=1}^n X_j\right) = E\left(\left[\sum_{j=1}^n X_j\right]^2\right) - \left[E\left(\sum_{j=1}^n X_j\right)\right]^2 = E\left(\left[\sum_{j=1}^n X_j\right]^2\right) - \left[\sum_{j=1}^n E(X_j)\right]^2.$$

Pero

$$\begin{aligned} E\left(\left[\sum_{j=1}^n X_j\right]^2\right) &= E\left(\sum_{j=1}^n X_j^2 + \sum_{i \neq j} X_i X_j\right) = \sum_{j=1}^n E(X_j^2) + \sum_{i \neq j} E(X_i X_j) = \sum_{j=1}^n E(X_j^2) + \sum_{i \neq j} E(X_i)E(X_j) \\ &= \sum_{j=1}^n [V(X_j) + E(X_j)^2] + \sum_{i \neq j} E(X_i)E(X_j) = \sum_{j=1}^n V(X_j) + \sum_{j=1}^n E(X_j)^2 + \sum_{i \neq j} E(X_i)E(X_j) \\ &= \sum_{j=1}^n V(X_j) + \left[\sum_{j=1}^n E(X_j)\right]^2. \end{aligned}$$

De manera que, simplemente,

$$V\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n V(X_j)$$

Si además las  $X_j$  tienen varianza común  $\sigma^2$ , entonces  $V(\sum_{j=1}^n X_j) = n\sigma^2$ .

La varianza de la variable promedio se obtiene sin más que aplicar que  $V(aX) = a^2 V(X)$ .

# Concentración de promedios. Leyes de los grandes números y TLC

Así que la variable promedio

tiene media  $\mu$ .

$$Y_n = \frac{1}{n} \sum_{j=1}^n X_j$$

Pero además tiene una desviación típica  $\sigma/\sqrt{n}$ , cantidad que, si  $n$  es muy grande, es muy pequeña.

De manera que los valores del promedio  $Y_n$  se concentran extraordinariamente en torno a la media.

La cuantificación de esta primera impresión es lo que se conoce como ley (débil) de los grandes números.

El ingrediente fundamental de este resultado es la propiedad general conocida como desigualdad de Chebyshev.

## Ley (débil) de los grandes números

$X_1, \dots, X_n$  son i.i.d. con media  $\mu$  y varianza  $\sigma^2$ . Consideramos la variable “promedio”

$$Y_n = \frac{1}{n} \sum_{j=1}^n X_j. \quad \longrightarrow \quad \begin{cases} E(Y_n) &= \mu; \\ \sigma(Y_n) &= \frac{\sigma}{\sqrt{n}} \quad (\text{por independencia}). \end{cases}$$

Si  $n$  es grande, los valores de  $Y_n$  se concentran en torno a la media común a las  $X_j$ , esto es,  $\mu$ . Más precisamente, y con la ayuda de la desigualdad de Chebyshev,

$$P\left(\left|\frac{1}{n} \sum_{j=1}^n X_j - \mu\right| > \lambda\right) \leq \frac{\sigma^2}{n \lambda^2}.$$

Con probabilidad tan próxima a 1 como queramos, la media aritmética se desvía de  $\mu$  tan poco como queramos (si el número  $n$  de “copias” de  $X_j$  es suficientemente grande).

# El Teorema del Límite Central

De unas variables  $X_1, X_2, \dots, X_n$  “solo” sabemos que son i.i.d. con media  $\mu$  y varianza  $\sigma^2$ . Sabemos que el “promedio”  $Y_n$  es, aproximadamente, como  $\mu$ .

Pero, ¿podemos estimar cantidades como  $P(Y_n \leq a)$ ?

Es decir, ¿podemos conocer (con cierta precisión) la función de distribución de  $Y_n$ ? Parecería que depende de la distribución de las  $X_j$ , pero. . .

Tenemos variables aleatorias  $X_1, X_2, \dots$ , i.i.d., todas con media  $\mu$  y varianza  $\sigma^2$ . La variable aleatoria promedio tipificada (es decir, con media 0 y desviación típica 1)

$$Z_n = \frac{1}{\sigma/\sqrt{n}} \left[ \frac{1}{n} \sum_{j=1}^n X_j - \mu \right]$$

cumple que

$$P(Z_n \leq x) \xrightarrow{n \rightarrow \infty} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy.$$

$$P(x_1 \leq Z_n \leq x_2) \xrightarrow{n \rightarrow \infty} \Phi(x_2) - \Phi(x_1) = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-y^2/2} dy.$$

Es decir, la variable  $Z_n$  se distribuye (aproximadamente, y si  $n$  es grande) como una  $N(0, 1)$ .

# El Teorema del Límite Central

**Versión para sumas (no promedios):** la variable  $\frac{\sum_{j=1}^n X_j - n\mu}{\sigma\sqrt{n}}$  se distribuye (aproximadamente, y si  $n$  es grande) como una  $\mathcal{N}(0, 1)$ .

Algunas observaciones sobre el TLC

Obsérvese que:

Ley débil de los grandes números: 
$$\frac{\sum_{j=1}^n X_j - n\mu}{n} \rightarrow 0$$

Teorema central del límite: 
$$\frac{\sum_{j=1}^n X_j - n\mu}{\sigma\sqrt{n}} \rightarrow \mathcal{N}(0, 1)$$



# Aplicación del TLC al cálculo de probabilidades

Dadas unas variables  $X_1, X_2, \dots, X_n$  i.i.d. consideramos la variable aleatoria promedio

$$Y_n = \frac{1}{n} \sum_{j=1}^n X_j$$

El Teorema Central del Límite nos dice que

$$\mathbf{P}(Y_n \leq a) = \mathbf{P}(Y_n - \mu \leq a - \mu) = \mathbf{P}\left(\frac{Y_n - \mu}{\sigma/\sqrt{n}} \leq \frac{a - \mu}{\sigma/\sqrt{n}}\right) \stackrel{n \text{ grande}}{\approx} \Phi\left(\frac{a - \mu}{\sigma/\sqrt{n}}\right).$$

Esta aproximación (cuando sea aplicable) nos permite entender la variable  $Y_n$  sin necesidad de conocer explícitamente la distribución de las  $X_j$ : de ellas basta conocer la media  $\mu$  y la desviación típica  $\sigma$  (además, claro, del número de sumandos  $n$ ).

# Aplicación del TLC al cálculo de probabilidades

- 1) Simular 100 muestras aleatorias simples de una variable aleatoria uniforme en  $[2,4]$  de 1000 datos cada una, en 100 columnas, sumarlas por componentes, almacenando en una variable SumaMuestra y observar cómo es el comportamiento de la muestra resultante en SumaMuestra formada por la suma de las 100 variables uniformes. Comparar la muestra de SumaMuestra con la distribución normal.

(La v.a. Uniforme  $[a,b]$  tiene:  $\text{Media}_U = (a+b)/2=3$  ;  $\text{Varianza}_U = (b-a)^2/12=0.333\dots$ ,  $\text{Desv. Típica} = 0.5773503$ . Y para la suma de 10 v.a. uniformes independientes:  
 $\text{Media} = 3*10=30$ ,  $\text{Varianza} = 10*0.33\dots=3.33\dots$ :  $\text{Desv. Típ.} = 1.826$ )

- 2) Realizar el análisis de los datos en SumaMuestra mediante un histograma y un diagrama cuantil-cuantil (qqplot) con referencia a la distribución normal.

# Aplicación del TLC al cálculo de probabilidades

Se genera una matriz en que cada columna es una muestra de `TamanoMuestra` elementos de una v.a uniforme continua entre 2 y 4. Hay `NumMuestras` columnas. Compárense con los 'poblacionales'  $\text{media}=30$ ,  $\text{sd}=1.826$

```
TamanoMuestra=1000;NumMuestras=100;  
muestra=array(0,c(TamanoMuestra,NumMuestras))  
SumaMuestra=rep(0,TamanoMuestra)# Definir vector con 0's TamanoMuestra veces
```

```
for (i in 1:NumMuestras) {  
  muestra[,i]=runif(TamanoMuestra,2,4) #llenar columna i con num. aleat. unif.  
  SumaMuestra=SumaMuestra+muestra[,i]  
}
```

```
hist(SumaMuestra)  
qqnorm(SumaMuestra)
```



## 1.i. Intervalos de confianza

Buena parte del análisis de cuestiones aleatorias que hemos visto hasta aquí se basa en un enfoque “experimental”, y como tal, está expuesto a “errores de medida”.

Por ejemplo, si disponemos de un mecanismo de simulación de una cierta variable aleatoria  $Y$ , para calcular su media  $E(Y)$  generamos una muestra grande de valores  $y_1, y_2, \dots, y_n$  para luego tomar la media aritmética

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j.$$

Nos gustaría disponer estimaciones de la posible discrepancia entre el valor “real”  $E(Y)$  y el promedio obtenido.

O mejor aún, estimar a priori cuántas muestras (el valor de  $n$ ) son necesarias para que esa discrepancia no supere un cierto umbral (un 1%, por ejemplo).

## 1.i. Intervalos de confianza

Los números  $y_1, \dots, y_n$  son muestras (independientes) de la variable aleatoria  $Y$ , cuyas características son en principio desconocidas. Consideremos la variable aleatoria promedio

$$\bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j.$$

de la que sabemos que

$$E(\bar{Y}) = E(Y), \quad V(\bar{Y}) = V(Y) / n$$

Quizás no conozcamos su distribución precisa, pero el teorema del límite central nos dice que

$$\frac{\bar{Y} - E(Y)}{\sigma(Y) / \sqrt{n}}$$

es aproximadamente como una  $N(0,1)$ .

Lo que nos permite establecer la siguiente estimación a priori.

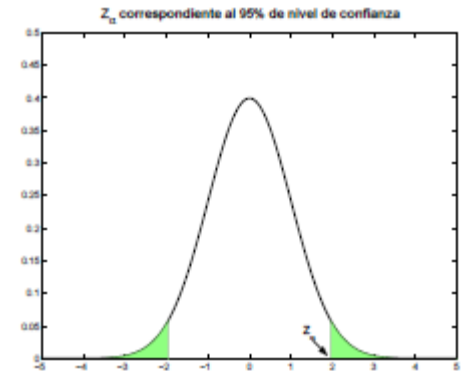
# 1.i. Intervalos de confianza

Dado un cierto  $\alpha$  (por ejemplo,  $\alpha = 5\%$ ), llamemos  $z_{\alpha/2}$  al valor para el que

$$P(-z_{\alpha/2} < N(0,1) < z_{\alpha/2}) = 1 - \alpha$$

Entonces,

$$\begin{aligned} 1 - \alpha &\approx P(-z_{\alpha/2} < \frac{\bar{Y} - E(Y)}{\sigma(Y)/\sqrt{n}} < z_{\alpha/2}) \\ &= P(\bar{Y} - z_{\alpha/2} \frac{\sigma(Y)}{\sqrt{n}} < E(Y) < \bar{Y} + z_{\alpha/2} \frac{\sigma(Y)}{\sqrt{n}}) \end{aligned}$$



Es decir, que hay una alta probabilidad ( $1 - \alpha$ ) de que el valor medio buscado,  $E(Y)$ , no difiera de  $\bar{Y}$  en más de  $z_{\alpha/2} \frac{\sigma(Y)}{\sqrt{n}}$ .

El problema es que, en general, no conocemos  $\sigma(Y)$ .

## 1.i. Intervalos de confianza

Pero, una vez que dispongamos de las muestras  $y_1, \dots, y_n$  de  $Y$ , tendremos que la muestra de la variable  $Y$  es

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j$$

y podemos estimar  $\sigma(Y)$  con la desviación típica muestral

$$s = \frac{1}{n-1} \sum_{j=1}^n (y_j - \bar{y})^2$$

Para, a posteriori, tener confianza  $1 - \alpha$  de que se tenga

$$\bar{y} - z_{\alpha/2} \frac{s}{\sqrt{n}} \leq E(Y) \leq \bar{y} + z_{\alpha/2} \frac{s}{\sqrt{n}}$$

El intervalo de confianza tiene tamaño

$$2z_{\alpha/2} \frac{s}{\sqrt{n}}$$

<https://seeing-theory.brown.edu/index.html#firstPage>

# 2 | Modelos multidimensionales y simulación



## 2.a. Medidas de dependencia lineal

Habitualmente manejamos un par de medidas de dependencia (lineal) entre dos variables:

- La covarianza:  $\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$ .
- Y la versión sin unidades, la correlación, que es un número entre  $-1$  y  $1$ :

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{V(X) V(Y)}}$$

La correlación es una medida adimensional, absoluta, entre  $-100\%$  y  $100\%$ , fácilmente interpretable. La covarianza, por su parte, tiene unidades (cuadráticas). Sin embargo, muchos cálculos se hacen con covarianzas (para pasar a correlaciones al final).

**Nota:** Dos variables incorreladas no tienen por qué ser independientes.

## 2.a. Medidas de dependencia lineal

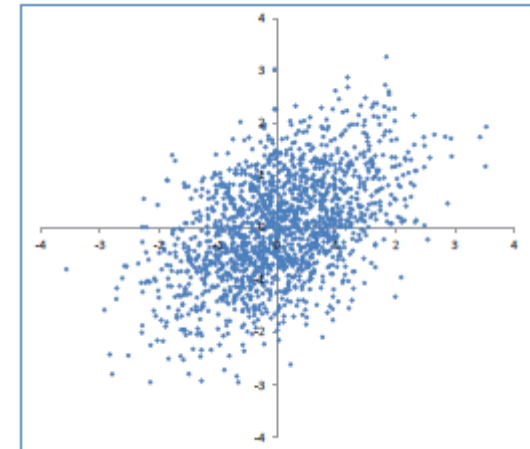
¿Qué significa la covarianza/correlación?

Observemos la definición de la covarianza:

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

Si esa cantidad es positiva, quiere decir que es más probable que  $X - E(X)$  e  $Y - E(Y)$  tengan el mismo signo. Es decir, que si  $X$  produce un valor “grande” por encima de la media, es altamente probable que  $Y$  produzca un valor también grande (por encima de la suya). O al revés, si  $X$  es pequeño, es bastante probable que  $Y$  sea pequeño.

Si simuláramos muestras de dos variables aleatorias con correlación, por ejemplo, del 50%, y las representáramos en un gráfico de dispersión, esperaríamos ver una situación como la de la figura, en la que (relativamente) pocos puntos caen en los cuadrantes superior/izquierda e inferior/derecha.



## 2.b. El caso multidimensional: matriz de correlaciones

Tenemos un vector  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  de variables aleatorias.

Conjuntamente, tienen una tabla (n-dimensional) de probabilidades conjuntas (en el caso discreto), o una función de densidad conjunta  $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$  en el caso continuo.

El vector aleatorio  $\mathbf{X}$  tiene

- un vector de medias  $\mu = (\mu_1, \mu_2, \dots, \mu_n) = (E(X_1), E(X_2), \dots, E(X_n))$ ;
- y un vector de varianzas:  $(V(X_1), V(X_2), \dots, V(X_n))$ , donde

$$V(X_j) = E((X_j - \mu_j)^2) = E(X_j^2) - \mu_j^2.$$

O también el correspondiente vector de desviaciones típicas,

$$\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n),$$

donde  $\sigma_j = \sqrt{V(X_j)}$ .

## 2.b. El caso multidimensional: matriz de correlaciones

Para registrar la información sobre la dependencia (lineal) entre (los pares de) variables utilizamos una tabla  $n \times n$ , la **matriz de varianzas/covarianzas**:

$$V = \begin{pmatrix} V(X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & V(X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & V(X_n) \end{pmatrix}$$

donde  $\text{cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = E(X_i X_j) - \mu_i \mu_j$

O la **matriz de correlaciones**:

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n} & \rho_{2n} & \cdots & 1 \end{pmatrix}$$

donde

$$\rho_{ij} = \frac{\text{cov}(X_i, X_j)}{\sigma_i \sigma_j}$$

Obsérvese que si disponemos de  $V$  podemos calcular  $\Sigma$ . Si tenemos  $\Sigma$ , necesitamos también el vector de desviaciones típicas  $\sigma$  para calcular  $V$ .

Tanto  $V$  como  $\Sigma$  son **simétricas y definidas positivas**.

## 2.b. El caso multidimensional: matriz de correlaciones

### Cálculo de matrices de covarianza y de correlaciones.

Partimos de  $\mathbf{Y} = (Y_1, \dots, Y_n)$ , con vector de medias  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ , y queremos calcular su matriz de correlaciones. Pasamos a  $\tilde{\mathbf{Y}} = \mathbf{Y} - \boldsymbol{\mu} = (Y_1 - \mu_1, Y_2 - \mu_2, \dots, Y_n - \mu_n)$ . Y ahora

$$\begin{aligned} \tilde{\mathbf{Y}}^T \tilde{\mathbf{Y}} &= \begin{pmatrix} Y_1 - \mu_1 \\ Y_2 - \mu_2 \\ \vdots \\ Y_n - \mu_n \end{pmatrix} (Y_1 - \mu_1, Y_2 - \mu_2, \dots, Y_n - \mu_n) = \\ &= \begin{pmatrix} (Y_1 - \mu_1)^2 & (Y_1 - \mu_1)(Y_2 - \mu_2) & \cdots & (Y_1 - \mu_1)(Y_n - \mu_n) \\ (Y_2 - \mu_2)(Y_1 - \mu_1) & (Y_2 - \mu_2)^2 & \cdots & (Y_2 - \mu_2)(Y_n - \mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (Y_n - \mu_n)(Y_1 - \mu_1) & (Y_n - \mu_n)(Y_2 - \mu_2) & \cdots & (Y_n - \mu_n)^2 \end{pmatrix} \end{aligned}$$

Luego

$$\mathbf{E}(\tilde{\mathbf{Y}}^T \tilde{\mathbf{Y}}) = \mathbf{V}$$

De manera análoga, si hacemos

$$\hat{\mathbf{Y}} = \left( \frac{Y_1 - \mu_1}{\sigma_1}, \frac{Y_2 - \mu_2}{\sigma_2}, \dots, \frac{Y_n - \mu_n}{\sigma_n} \right), \quad \text{entonces} \quad \mathbf{E}(\hat{\mathbf{Y}}^T \hat{\mathbf{Y}}) = \boldsymbol{\Sigma}$$

## 2.c. Matrices de covarianza y de correlación muestrales

Partimos de  $n$  series, cada una de las cuales consta de  $N$  datos. La media (muestral, estadística) de la serie  $j$  es

$$\mu^{(j)} = \frac{1}{N} \sum_{k=1}^N x_k^{(j)}$$

serie 1	serie 2	.....	serie $n$
$x_1^{(1)}$	$x_1^{(2)}$	$\dots$	$x_1^{(n)}$
$x_2^{(1)}$	$x_2^{(2)}$	$\dots$	$x_2^{(n)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_N^{(1)}$	$x_N^{(2)}$	$\dots$	$x_N^{(n)}$

La varianza y la desviación típica (muestrales, estadísticas) de cada serie  $j$  se definen mediante

$$V^{(j)} = \frac{1}{N} \sum_{k=1}^N (x_k^{(j)} - \mu^{(j)})^2 \quad y \quad \sigma^{(j)} = \sqrt{V^{(j)}}$$

La covarianza (estadística) entre las series  $i$  y  $j$  es

$$V^{(i,j)} = \frac{1}{N} \sum_{k=1}^N (x_k^{(i)} - \mu^{(i)})(x_k^{(j)} - \mu^{(j)})$$

## 2.c. Matrices de covarianza y de correlación muestrales

Y las correspondientes correlaciones (estadísticas)

$$\rho^{(i,j)} = \frac{\mathbf{V}^{(i,j)}}{\sigma^{(i)}\sigma^{(j)}}$$

Las matrices de covarianza y de correlación (muestrales, estadísticas) son las siguientes:

La covarianza (estadística) entre las series  $i$  y  $j$  es:

$$\mathbb{V}_{\text{muestral}} = \begin{pmatrix} \mathbf{V}^{(1)} & \mathbf{V}^{(1,2)} & \dots & \mathbf{V}^{(1,n)} \\ \mathbf{V}^{(2,1)} & \mathbf{V}^{(2)} & \dots & \mathbf{V}^{(2,n)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{V}^{(n,1)} & \mathbf{V}^{(n,2)} & \dots & \mathbf{V}^{(n)} \end{pmatrix} \quad \text{y} \quad \Sigma_{\text{muestral}} = \begin{pmatrix} 1 & \rho^{(1,2)} & \dots & \rho^{(1,n)} \\ \rho^{(2,1)} & 1 & \dots & \rho^{(2,n)} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{(n,1)} & \rho^{(n,2)} & \dots & 1 \end{pmatrix}$$

## 2.c. Matrices de covarianza y de correlación muestrales

Cálculo de matrices de covarianza y de correlación muestrales.

Consideramos los datos como una matriz  $H$  (de dimensiones  $N \times n$ ):

$$H = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(n)} \\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ x_N^{(1)} & x_N^{(2)} & \cdots & x_N^{(n)} \end{pmatrix}$$

Obsérvese que

$$\begin{aligned} H^T H &= \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & \cdots & \cdots & x_N^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & \cdots & \cdots & x_N^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_1^{(n)} & x_2^{(n)} & \cdots & \cdots & \cdots & x_N^{(n)} \end{pmatrix} \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(n)} \\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_N^{(1)} & x_N^{(2)} & \cdots & x_N^{(n)} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{k=1}^N x_k^{(1)} x_k^{(1)} & \sum_{k=1}^N x_k^{(1)} x_k^{(2)} & \cdots & \sum_{k=1}^N x_k^{(1)} x_k^{(n)} \\ \sum_{k=1}^N x_k^{(2)} x_k^{(1)} & \sum_{k=1}^N x_k^{(2)} x_k^{(2)} & \cdots & \sum_{k=1}^N x_k^{(2)} x_k^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^N x_k^{(n)} x_k^{(1)} & \sum_{k=1}^N x_k^{(n)} x_k^{(2)} & \cdots & \sum_{k=1}^N x_k^{(n)} x_k^{(n)} \end{pmatrix} \end{aligned}$$



## 2.c. Matrices de covarianza y de correlación muestrales

Si hacemos que todas las series tengan media 0 y formamos la correspondiente matriz  $\tilde{H}$  (de dimensiones  $N \times n$ ):

$$\tilde{H} = \begin{pmatrix} x_1^{(1)} - \mu^{(1)} & x_1^{(2)} - \mu^{(2)} & \dots & x_1^{(n)} - \mu^{(n)} \\ x_2^{(1)} - \mu^{(1)} & x_2^{(2)} - \mu^{(2)} & \dots & x_2^{(n)} - \mu^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ x_N^{(1)} - \mu^{(1)} & x_N^{(2)} - \mu^{(2)} & \dots & x_N^{(n)} - \mu^{(n)} \end{pmatrix} \quad \text{entonces} \quad \boxed{V_{\text{muestral}} = \frac{1}{N} \tilde{H}^T \tilde{H}}$$

Y si lo que hacemos es normalizar cada una de las series (media 0 y desviación típica 1):

$$\tilde{\tilde{H}} = \begin{pmatrix} \frac{x_1^{(1)} - \mu^{(1)}}{\sigma^{(1)}} & \frac{x_1^{(2)} - \mu^{(2)}}{\sigma^{(2)}} & \dots & \frac{x_1^{(n)} - \mu^{(n)}}{\sigma^{(n)}} \\ \frac{x_2^{(1)} - \mu^{(1)}}{\sigma^{(1)}} & \frac{x_2^{(2)} - \mu^{(2)}}{\sigma^{(2)}} & \dots & \frac{x_2^{(n)} - \mu^{(n)}}{\sigma^{(n)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_N^{(1)} - \mu^{(1)}}{\sigma^{(1)}} & \frac{x_N^{(2)} - \mu^{(2)}}{\sigma^{(2)}} & \dots & \frac{x_N^{(n)} - \mu^{(n)}}{\sigma^{(n)}} \end{pmatrix}$$

entonces

$$\boxed{\Sigma_{\text{muestral}} = \frac{1}{N} (\tilde{\tilde{H}})^T \tilde{\tilde{H}}}$$

## 2.d. La normal bidimensional (y multidimensional)

El modelo multidimensional más importante, por su amplio uso en el contexto financiero (y porque permite cálculos explícitos y se presta a mecanismos sencillos de simulación), es, sin duda, la normal multidimensional.

Se trata de un modelo en el que cada variable por separado es una normal (con cierta media y varianza). Pero, además, toda la información conjunta queda (únicamente) codificada por las correlaciones.

Lo emplearemos, por ejemplo, para describir rendimientos conjuntos de una cartera de activos financieros: el rendimiento de cada activo es una normal, y conjuntamente siguen una normal multidimensional.

Para empezar, veamos el caso bidimensional.

```
library(MASS)
Sigma <- matrix(c(10,3,3,2),2,2)
Sigma
b<-mvrnorm(n = 1000, rep(0, 2), Sigma)
plot(b)
```

## 2.d. La normal bidimensional (y multidimensional)

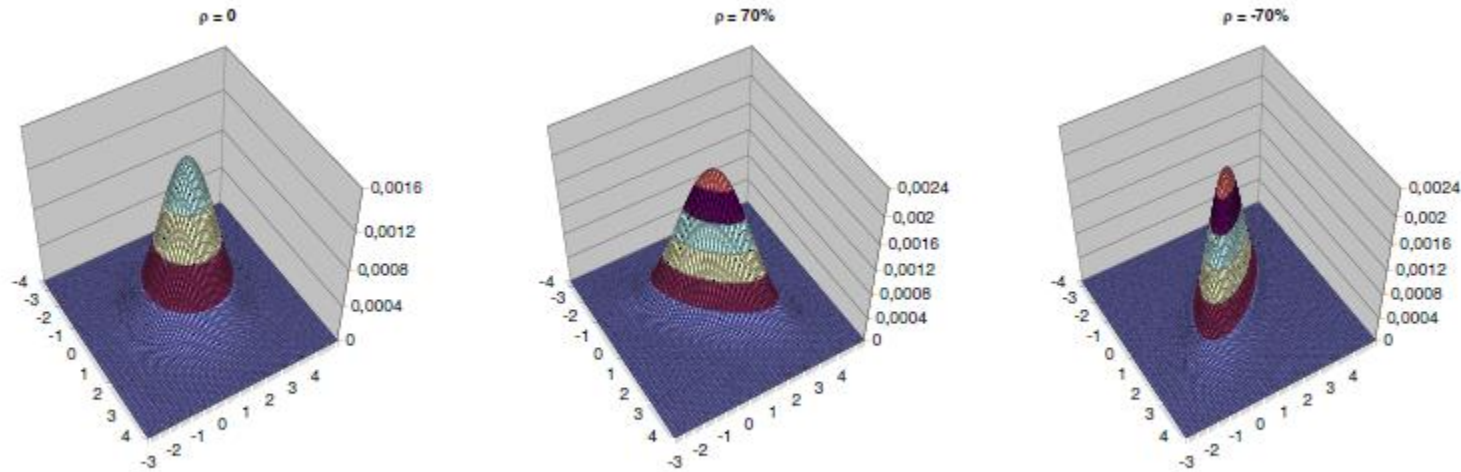
La normal bidimensional. **Primera versión.** El par  $(X, Y)$  sigue una normal bidimensional, y tanto  $X$  como  $Y$  son  $N(0,1)$ . Hay un único parámetro,  $-1 \leq \rho \leq 1$  (que resulta ser, justamente, la correlación entre  $X$  e  $Y$ ). La función de densidad conjunta es

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

Para cualquier  $\rho \in [-1, 1]$ , la superficie definida por la función anterior encierra volumen 1 ( $\iint f(x, y) dx dy = 1$ ) y sus marginales son normales  $N(0,1)$ .

## 2.d. La normal bidimensional (y multidimensional)

Las gráficas siguientes dan una idea del papel de  $\rho$ .



**Nota importante.** En la normal bidimensional, el que  $X$  e  $Y$  sean incorreladas ( $\rho = 0$ ) es equivalente a que sean independientes.

## 2.d. La normal bidimensional (y multidimensional)

**Segunda versión (más general).** El par de variables  $(X, Y)$  sigue una normal bidimensional, pero ahora

- $X$  es  $N(\mu_1, \sigma_1)$ ;
- mientras que  $Y$  es  $N(\mu_2, \sigma_2)$ ;
- el parámetro  $\rho$  es, de nuevo, la correlación entre  $X$  e  $Y$ .

La función de densidad conjunta, que tiene cinco parámetros,

$$-1 < \rho < 1, \mu_1, \mu_2 \in \mathbb{R} \text{ y } \sigma_1, \sigma_2 > 0,$$

viene dada por:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left( -\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} \right] \right)$$

De nuevo, correlación cero es lo mismo que independencia.

## 2.d. La normal bidimensional (y multidimensional)

**La normal multidimensional.** Las variables  $(X_1, \dots, X_n)$  siguen una normal multidimensional

- con vector de medias  $\mu = (\mu_1, \dots, \mu_n)$ ;
- y matriz de varianza-covarianza  $\mathbf{V}$

si la función de densidad conjunta viene dada por

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \det(\mathbf{V})^{1/2}} \exp \left( -\frac{1}{2} (\mathbf{x} - \mu)^T \mathbf{V}^{-1} (\mathbf{x} - \mu) \right)$$

donde  $\mathbf{x} = (x_1, \dots, x_n)$ .

Si la matriz  $\mathbf{V}$  es diagonal (es decir, si las  $X_j$  son incorreladas), entonces las variables  $X_j$  son independientes.

## 2.d. La normal bidimensional (y multidimensional)

Algunos comentarios sobre la normal bidimensional (o multidimensional).

1. Sobre sumas de normales. Si  $(X, Y)$  sigue una normal bidimensional, de parámetros  $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ , entonces cualquier combinación lineal  $Z = aX + bY$  sigue una distribución normal de parámetros

$$E(Z) = a\mu_1 + b\mu_2, \quad V(Z) = a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\sigma_1\sigma_2\rho.$$

2. Transformaciones lineales. Con más generalidad, si  $(X, Y)$  sigue una normal bidimensional, entonces el vector  $(X', Y')$  dado por

$$(X', Y') = (X, Y) \begin{pmatrix} a & c \\ b & d \end{pmatrix} \implies \begin{cases} X' = aX + bY \\ Y' = cX + dY \end{cases}$$

sigue una normal bidimensional.

Si  $\mathbf{X}$  es una normal  $n$ -dimensional  $\mathcal{N}(\boldsymbol{\mu}, \mathbb{V})$  e  $\mathbf{Y} = \mathbf{XA}$ , donde  $\mathbf{A}$  es una matriz  $n \times n$ , entonces  $\mathbf{Y}$  sigue una normal multidimensional  $\mathcal{N}(\boldsymbol{\mu}\mathbf{A}, \mathbf{A}^T\mathbb{V}\mathbf{A})$



Afi

Escuela  
de Finanzas

---

© 2021 Afi Escuela de Finanzas. Todos los derechos reservados.