Introducción a las variables aleatorias.

Mathieu Kessler

Departamento de Matemática Aplicada y Estadística
Universidad Politécnica de Cartagena

Cartagena, Enero 2010

Guión

Guión

Introducción

En problemas concretos

- Estamos interesados en funciones definidas sobre el espacio de resultados posibles de un experimento aleatorio.
- Los sucesos que nos interesan los expresaremos a través de estas funciones.

Damos un paso más en la construcción de modelos matemáticos de la incertidumbre

Asociados a experimentos aleatorios, definiremos el concepto de variable aleatoria.

Guión

Concepto de variable aleatoria

Definición

Una variable aleatoria (v.a.)- asocia un número o más generalmente una característica a todo resultado posible del experimento.

Ejemplos

- Experimento: *"Realizar una medición de la concentración de un producto"*
  v.a.: X: *"valor medido de la concentración"*.
- Experimento: *"Escoger al azar un dispositivo producido"*
  v.a. 'X: *"duración hasta el fallo"*
- Experimento: *"Lanzar tres veces una moneda"*
  v.a. X: *"Número de veces que ha salido cruz"*

Distribución de una v.a.

Un concepto fundamental

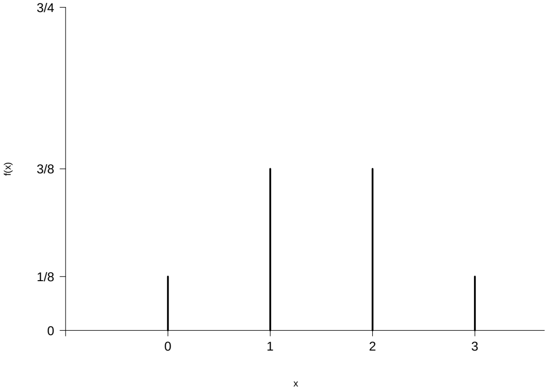
- Conocer la distribución de los valores de una v.a. X consiste en saber asignar a cualquier suceso relacionado con X una probabilidad.
- Ante un fenómeno aleatorio, construir un modelo para una v.a. de interés es decidir de una distribución para esta v.a.

Ejemplo de los tres lanzamiento de una moneda

- X número de veces que ha salido "cruz". Valores: 0,1,2,3
- La distribución está descrita por:

Valor	Probabilidad
0	1/8
1	3/8
2	3/8
3	1/8

Representación gráfica



Antes de realizar el experimento, ésta es nuestra expectación en cuanto al valor de X

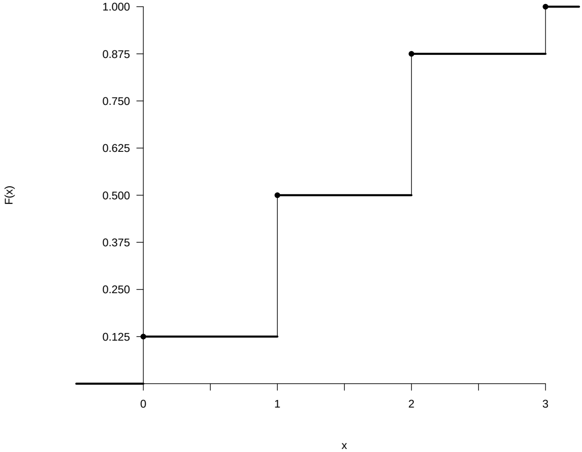
Función de distribución acumulada

Una manera de describir la distribución de una variable X

La función de distribución acumulada de una v.a. X es la función F_X que asocia a cualquier número real t la probabilidad de que X sea menor o igual a t, i.e.

$$F_X(t) = \mathbb{P}(X \leq t).$$

Ejemplo de los tres lanzamientos



Una propiedad fundamental de la función de distribución acumulada

Para todos números reales $a \leq b$,

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

Los dos tipos de variables aleatorias

Dos tipos de variables aleatorias

- V.a. discreta: sólo puede tomar un número finito o infinito numerable de valores.
- V.a. continua: puede tomar cualquier número en un intervalo.

- Ejemplo: experimento: "Tirar tres veces una moneda". X: número de cruces.
- Ejemplo: experimento: "Tirar una pelota", X: distancia a la que cae.

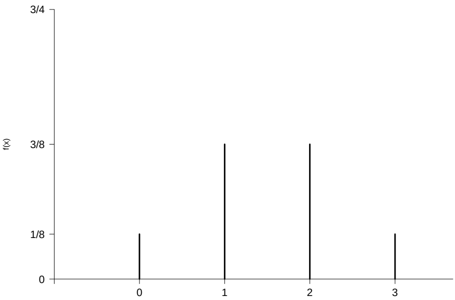
Guión

Descripción de la distribución de una v.a. discreta

Si X es una variable discreta, $x_1, \ldots, x_n, \ldots$ sus valores posibles.
Para describir la distribución de X, basta con especificar

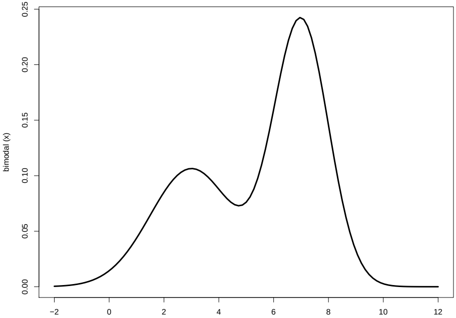
$$\mathbb{P}(X = x_i) = f_X(x_i) \qquad = f_X(x_i),$$

Función



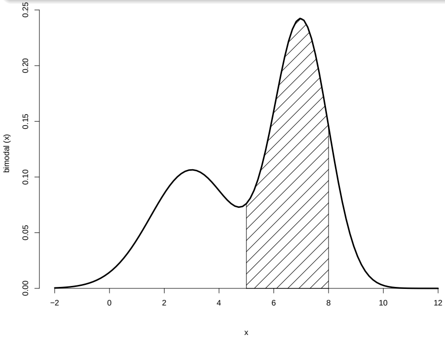
Descripción de la distribución de una v.a. continua

Si X es una variable continua, usaremos una función para indicar en qué regiones de su intervalo de valores posibles, creemos más probable que tome valores.
Lo haremos a través de una función de densidad



Descripción de la distribución de una v.a. continua

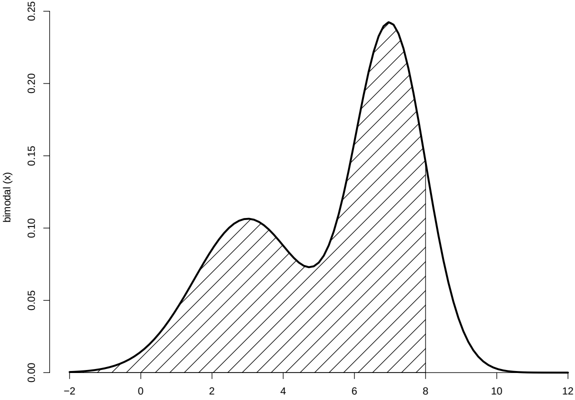
La probabilidad de que X pertenezca a un intervalo dado ($a < X \leq b$, por ejemplo) es el área encerrada debajo de la curva en el intervalo (a, b).



$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx.$$

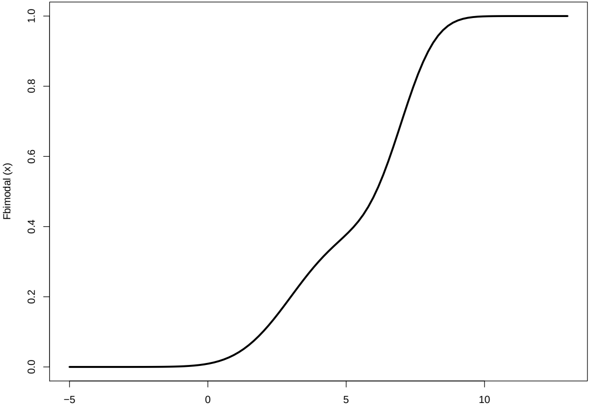
Función de distribución acumulada

Necesitamos calcular el área acumulada, variando el límite derecho:



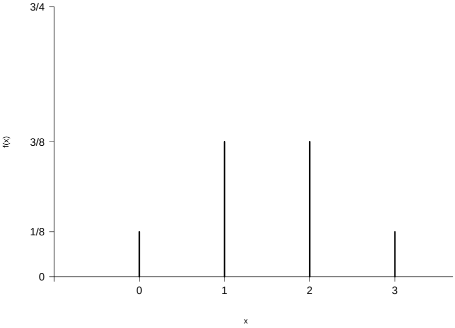
Función de distribución acumulada

Obtenemos

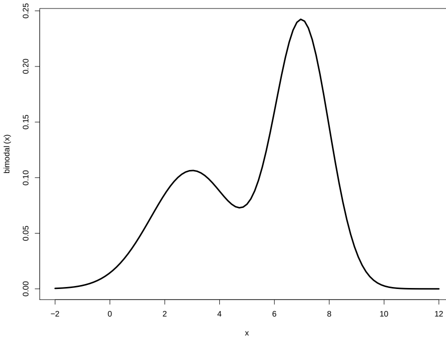


En resumen

Variable discreta: función de probabilidad



Variable continua: función de densidad



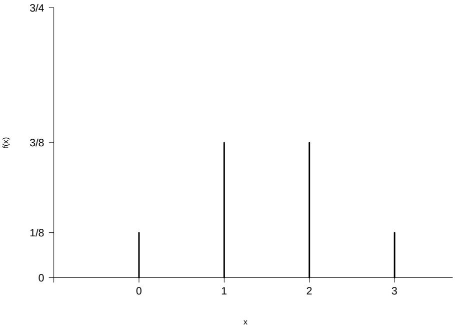
Indicadores numéricos para resumir las distribuciones

Al igual que para conjuntos de datos, buscamos resumir con indicadores numéricos la distribución de una variable:

- ¿Donde está el centro?
- ¿Presenta mucha variabilidad?
- ¿Es simétrica?
- ¿Es unimodal?

Sólo veremos indicadores para las dos primeras preguntas.

Valor promedio de una distribución



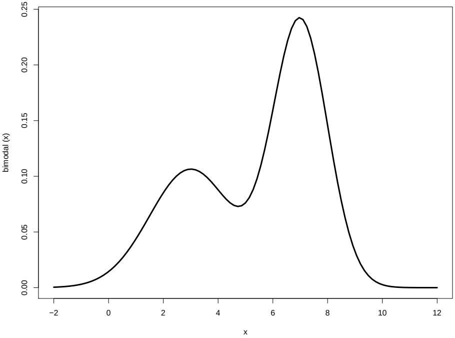
¿Qué valor escogemos como el centro de esta distribución?
El centro de gravedad:

$$1/8\times 0+1\times 3/8+2\times 3/8+3\times 1/8$$

$$= 3/2 = \mathbb{E}[X]$$

$$\boxed{\mathbb{E}[X] = \sum_{x_i} x_i \mathbb{P}(X = x_i).}$$

Valor promedio de una distribución: variable continua



¿Qué valor escogemos como el centro de esta distribución?
El centro de gravedad:

$$\boxed{\mu_X = \mathbb{E}[X] = \int_{-\infty}^{+\infty} x \cdot f(x)dx.}$$

Esperanza

La cantidad $\mathbb{E}[X]$ se llama:

- valor promedio de X
- valor esperado de X
- media de X
- esperanza de X.

Será útil considerar también el valor esperado de una función de X, por ejemplo

$$\mathbb{E}[X^2], \qquad \mathbb{E}[X^3], \qquad \mathbb{E}[e^X]\ldots$$

En general:

$$\mathbb{E}[g(X)]$$

Esperanza de una función $g(X)$

Para una función g,

consideramos la esperanza de $g(X)$ como

- Para una variable discreta:

$$\boxed{\mathbb{E}[g(X)] = \sum_{x_i} g(x_i)\mathbb{P}(X = x_i).}$$

- Para una variable continua:

$$\boxed{\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot f(x)dx.}$$

Varianza y desviación típica

Para medir la dispersión de la distribución

- Varianza

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

  Es el valor promedio de la distancia entre X y su valor promedio...

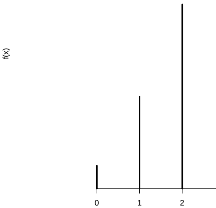
- Desviación típica:

$$\boxed{\sigma_X = \sqrt{Var(X)}}.$$

Guión

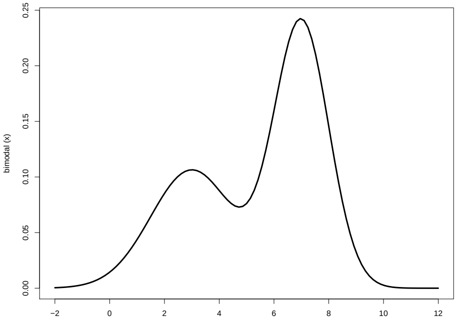
Escoger un modelo

Ante un fenómeno aleatorio, construir un modelo para la distribución de valores de $X \Leftrightarrow$ escoger una función de probabilidad o una función de densidad.



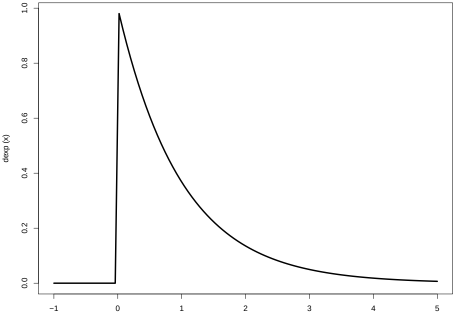
Escoger un modelo

Ante un fenómeno aleatorio, construir un modelo para la distribución de valores de $X \Leftrightarrow$ escoger una función de probabilidad o una función de densidad.



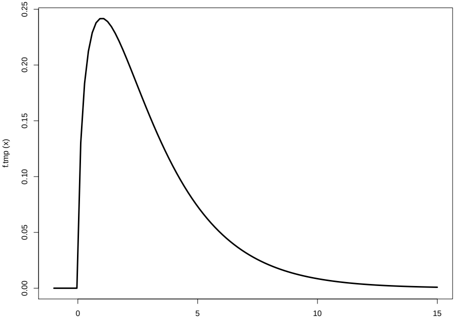
Escoger un modelo

Ante un fenómeno aleatorio, construir un modelo para la distribución de valores de $X \Leftrightarrow$ escoger una función de probabilidad o una función de densidad.



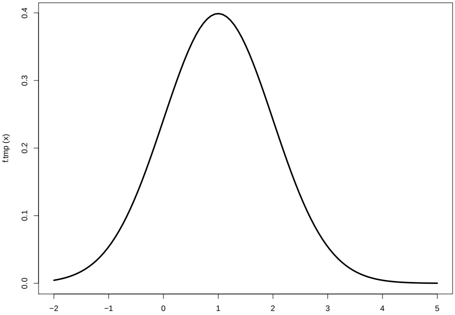
Escoger un modelo

Ante un fenómeno aleatorio, construir un modelo para la distribución de valores de X $\Leftrightarrow$ escoger una función de probabilidad o una función de densidad.



Escoger un modelo

Ante un fenómeno aleatorio, construir un modelo para la distribución de valores de $X \Leftrightarrow$ escoger una función de probabilidad o una función de densidad.

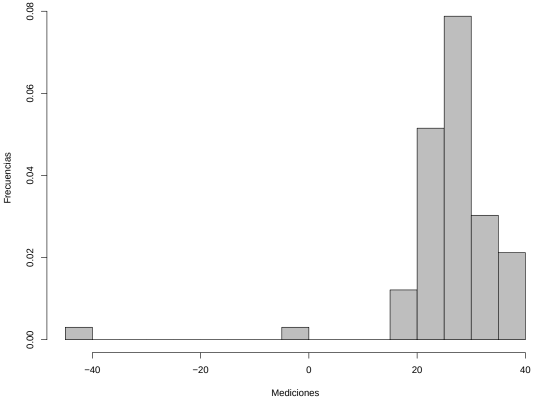


¿Cómo escoger un modelo?

Para decidir de un modelo para la distribución de nuestra variable de interés, tenemos dos posibilidades:

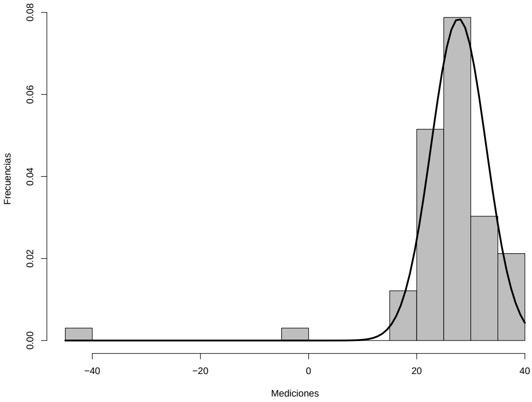
1. Deducirlo de la naturaleza y las propiedades del fenómeno de interés. Ejemplo de los tres lanzamientos.
2. Basarnos en un número suficiente de observaciones del fenómeno, realizar un histograma y identificar un modelo conocido en un "catálogo" de distribuciones.

Ejemplo de elección basada en el histograma..



¿Qué distribución escoger?

Ejemplo de elección basada en el histograma..



Veremos un mini catálogo de distribuciones

Para variables discretas

1. La distribución binomial
2. La distribución de Poisson

Para variables continuas

1. La distribución uniforme
2. La distribución exponencial
3. La distribución Normal

La distribución binomial

Contexto

La distribución binomial aparece cuando se dan las condiciones siguientes:

- Tenemos un primer experimento aleatorio simple, con una situación dicotómica, es decir una situación con dos sucesos posibles A y A^c (o ocurre A o no ocurre A).
- Repetimos este experimento simple n veces de manera independiente.
- Consideramos la variable X="Número de veces que ha ocurrido A en las n realizaciones del experimento simple.

$$\boxed{X \sim \mathcal{B}(n, p),}$$

i.e X sigue una distribución Binomial, de parámetros n (el número de veces que repetimos el experimento simple) y p (la probabilidad de que, en una realización del experimento simple, ocurra A)

La distribución binomial

Ejemplo

Una empresa produce piezas con 1% de defectuosas. Las piezas se empaquetan en cajas de 10 unidades. Si consideramos el experimento aleatorio que consiste en escoger al azar una caja entre la producción, ¿cuál es la distribución de la variable X="número de piezas defectuosas en la caja".

- Experimento: "llenar una caja con piezas de la producción"
- Equivalente a realizar 10 veces el experimento simple:

Escoger al azar UNA pieza en la producción.

- En el experimento simple, situación dicotómica: A="La pieza es defectuosa"(probabilidad 1%), A^C "la pieza es correcta".

$$\boxed{X \sim \mathcal{B}(10, 0.01)}$$

Propiedades de la distribución binomial $\mathcal{B}(n, p)$

Sólo para información...

- Valores posibles: $0, 1, 2, \ldots, n$.
- Función puntual de probabilidad. Para $i = 0, 1, \ldots, n$,

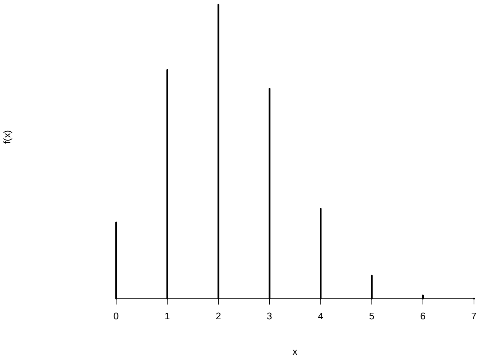
$$f_X(i) = \mathbb{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i},$$

donde $\binom{n}{i} = \frac{n!}{i! \cdot (n-i)!}$,

- $\mathbb{E}[X] = np \qquad Var(X) = np(1-p)$.

Ejemplo de gráfica de la función de probabilidad de una binomial

$$\boxed{X \sim \mathcal{B}(7, 0.3):}$$



La distribución de Poisson

Contexto

La distribución de Poisson aparece en situaciones en las que se cuenta el número de apariciones de un determinado suceso:

- o bien en un intervalo de tiempo dado (como el número de partículas emitidas en un segundo por un material radioactivo)
- bien en un recinto físico (como el número de fallos en un metro de alambre de hierro producido.)

Llamamos λ el número medio de apariciones del suceso de interés por intervalo de tiempo.

$$\boxed{X \sim \mathcal{P}(\lambda),}$$

i.e X sigue una distribución de Poisson de parámetro λ.

Propiedades de la distribución de Poisson $\mathcal{P}(n, p)$

Sólo para información...

- Valores posibles: $0, 1, 2, \ldots, ...$
- Función puntual de probabilidad. Para $i = 0, 1, \ldots,$

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

- $\mathbb{E}[X] = \lambda \qquad Var(X) = \lambda.$

Ejemplo de gráfica de la función de probabilidad de una Poisson

$$\boxed{X \sim \mathcal{P}(4) :}$$

La distribución uniforme

La más sencilla

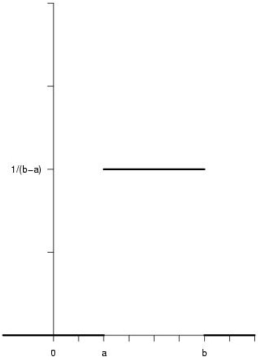
La distribución uniforme sobre el intervalo $[a, b]$:

$$f_X(x) = \begin{cases} \frac{1}{(b-a)} & \text{si } a \leq x \leq b, \\ 0 & \text{en otro caso.} \end{cases}$$

En cualquier calculadora científica

El comando `RANDOM` simula una variable uniforme sobre el intervalo $[0, 1]$

La distribución uniforme



¿Esperanza?

$$\mathbb{E}[X] = \frac{a+b}{2}.$$

La distribución exponencial

Contexto

La distribución exponencial aparece en las mismas situaciones que la distribución de Poisson pero mide el tiempo entre dos apariciones del suceso de interés, por ejemplo:

- el tiempo que discurre entre dos emisiones de partículas radioactivas.
- el tiempo entre dos llegadas de clientes en una cola.

Llamamos λ el número medio de apariciones del suceso de interés por intervalo de tiempo.

$$\boxed{X \sim \mathcal{E}xp(\lambda),}$$

i.e X sigue una distribución exponencial de parámetro λ.

Propiedades de la distribución exponencial $\mathcal{E}xp(\lambda)$

Sólo para información...

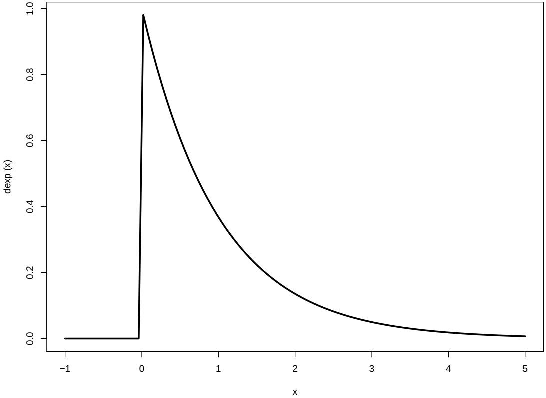
- Valores posibles: $x > 0$.
- Función de densidad. Para $x > 0$,

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en otro caso.} \end{cases}$$

- $\mathbb{E}[X] = 1/\lambda \qquad Var(X) = 1/\lambda^2.$

Ejemplo de gráfica de la densidad de una Exponencial

$$\boxed{X \sim \mathcal{E}xp(1):}$$



La distribución Normal

Definición

Sea μ un número real y σ^2 un número real positivo.
X sigue una distribución Normal de parámetros μ y σ^2 si su densidad es

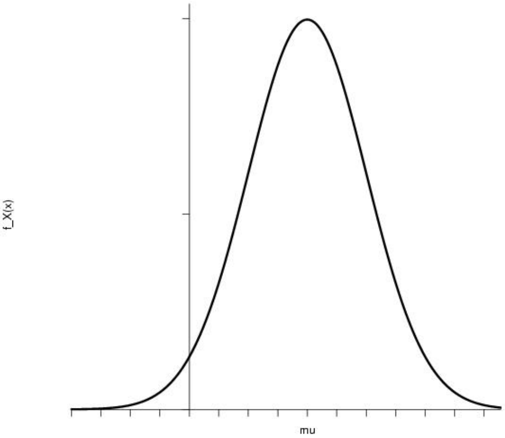
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$$\boxed{X \sim \mathcal{N}(\mu, \sigma^2),}$$

¡La estrella de las distribuciones!

- Aparece en la mayoría de los procedimientos estadísticos más usados.
- Tipicamente en contexto de mediciones (margen de error etc..)

La densidad de la distribución Normal



Propiedades de la distribución Normal $\mathcal{N}(\mu, \sigma^2)$

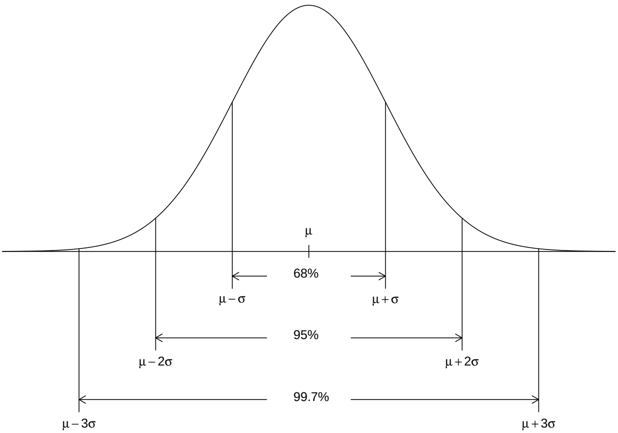
- Valores posibles: $x \in \mathbb{R}$.
- La función de densidad es simétrica respecto a $x = \mu$. ¿Qué vale $\mathbb{P}(X > \mu)$?
- La curva de densidad nunca se cruza con el eje Ox.
- $$\mathbb{E}[X] = \mu, \qquad var(X) = \sigma^2.$$
- La función de distribución acumulada requiere aproximaciones numéricas. (disponible en R por ejemplo)

Una regla que hay que saber

A pesar de que las áreas requieren aproximaciones numéricas, se cumple

$$
\begin{aligned}
\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) &\simeq 0.68 \\
\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &\simeq 0.95 \\
\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &\simeq 0.997,
\end{aligned}
$$

Una regla que hay que saber



Ejemplo de ilustración

Primer ejemplo

Decidimos modelizar la altura de los españoles varones adultos por una distribución Normal de media 175cm y desviación típica $\sigma = 5cm$.

- ¿Entre qué altura se encuentra comprendido el 95% de los varones españoles?
- Calcular el percentil 50 de la distribución de la altura.
- Calcular el percentil 84 y el percentil 16 de la distribución de la altura

Ejemplo de ilustración

Segundo ejemplo

Debemos modelizar la distribución de valores de una variable X, con media que estimamos igual a 3.2. Si pensamos razonable afirmar que el 95% de los valores de la distribución están entre 2.8 y 3.6, ¿qué valor tomaremos como σ?

Para el cálculo de probabilidades asociadas a una Normal

- Existen tablas para la Normal estándar ($Z \sim \mathcal{N}(0, 1)$)
- `R` incluye algoritmos de cálculo de probabilidades asociadas a muchas distribuciones. Para la normal:

  $\boxed{\texttt{pnorm(x,mean,sd)}}$

  calcula la función de distribución acumulada de la distribución Normal con media especificada por `mean` y desviación típica especificada por `sd`.
- Por ejemplo $X \sim \mathcal{N}(3, 2)$:
  - $P(X < 2.5)$=`pnorm(2.5,mean=3,sd=2)`
  - $P(X > 4)$=`1-pnorm(4,mean=3,sd=2)`
  - $P(-1 \leq X \leq 4) =$
    `pnorm(4,mean=3,sd=2)-pnorm(-1,mean=3,sd=2)`