

Contenidos

- **Qué es** la simulación de eventos discretos
- Alimentación de un modelo de simulación. **Datos de entrada**
- Explotación de un modelo de simulación. **Datos de salida**
- Los modelos tienen que ser útiles: **verificación y validación**

Datos de entrada

- Qué hay que hacer en relación con los datos de entrada de un modelo de simulación
 - **Caracterizar** dichas variables:
 - ¿Cuánto tiempo transcurre entre la llegada de dos clientes consecutivos?
 - ¿Cuál es la demanda en un determinado periodo?
 - **Generar valores** de las variables de entrada:
 - Generar los tiempos específicos en los que llega cada cliente
 - Generar la demanda específica de cada periodo

Caracterización de las variables de entrada Alternativas

- Si existen datos históricos, tres alternativas
 - Alimentar con datos históricos
 - Obtener una función empírica
 - Obtener una función analítica
- Si no existen datos históricos
 - Información aproximada

Caracterización de las variables de entrada

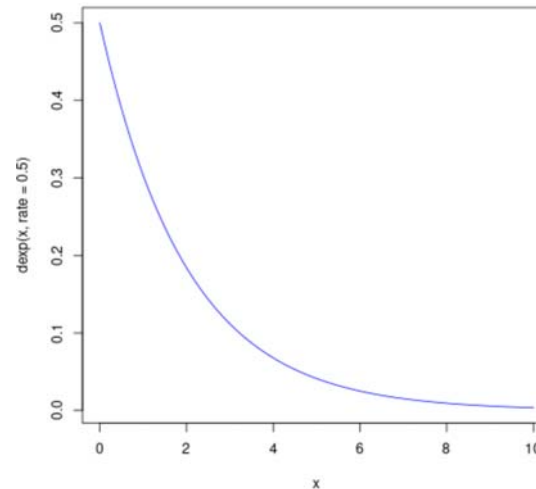
Datos históricos

- Existen **tres alternativas** para alimentar un modelo de simulación
 - Ejemplo: conocemos el tiempo entre cada dos averías de una máquina

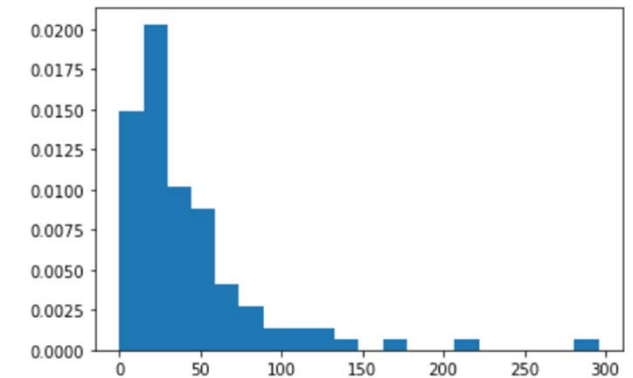
Datos históricos

21.6
85.1
105.1
84.5
46.1
12.0
58.8
108.6
74.0
2.3
3.2
47.5
7.9
3.5
15.3
9.8
71.7
7.9
33.4
14.0

Distribución teórica



Distribución empírica



Caracterización de las variables de entrada

Datos históricos

Datos históricos

21.6
85.1
105.1
84.5
46.1
12.0
58.8
108.6
74.0
2.3
3.2
47.5
7.9
3.5
15.3
9.8
71.7
7.9
33.4
14.0

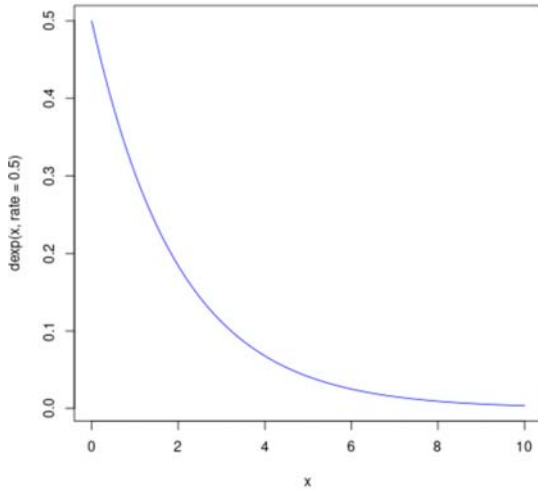
...

- Se utilizan los datos tal y como se recogieron del sistema.
- En el ejemplo:
 - La primera avería ocurre en $t=21.6$
 - La segunda avería ocurre en $t=21.6+85.1=106.3$
 - La tercera avería ocurre en $t=106.3+105.1=211.4$
- Los valores son los mismos cada vez que ejecutamos el modelo.

Caracterización de las variables de entrada

Datos históricos

Distribución teórica



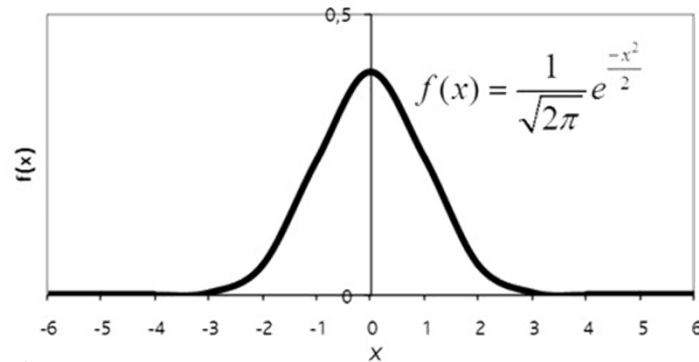
- Se dispone de una función (densidad/probabilidad o distribución).
- Podemos generar tantos números como queramos.
- Generamos un número aleatorio $U(0,1)$.
- Con ese número se obtiene un valor de la variable correspondiente.
- Los valores que se obtienen no son siempre los mismos*.

*más sobre este punto más adelante

Cómo se caracteriza una función teórica

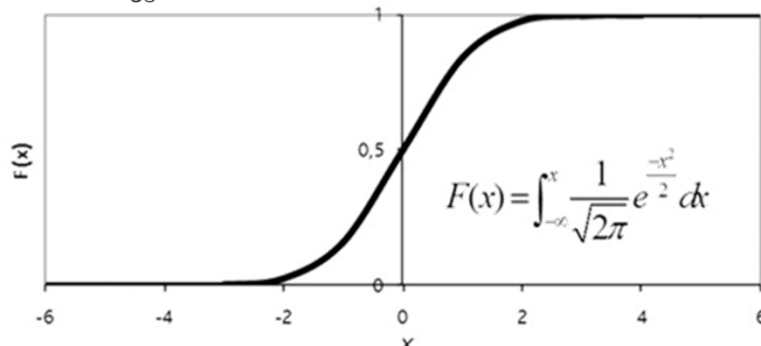
Funciones continuas

- Función de densidad
 $f(x)$



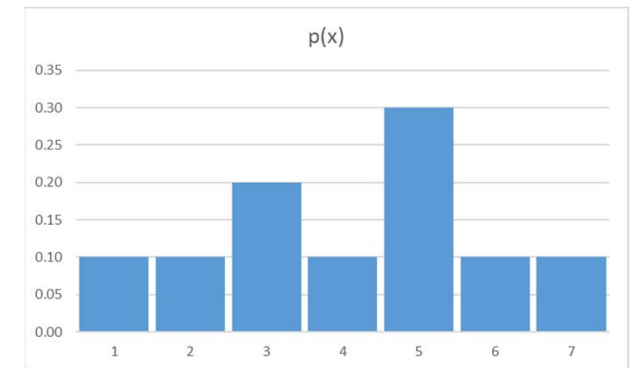
- Función de distribución

$$F(x) = p(X \leq x) = \int_{-\infty}^x f(z) dz$$



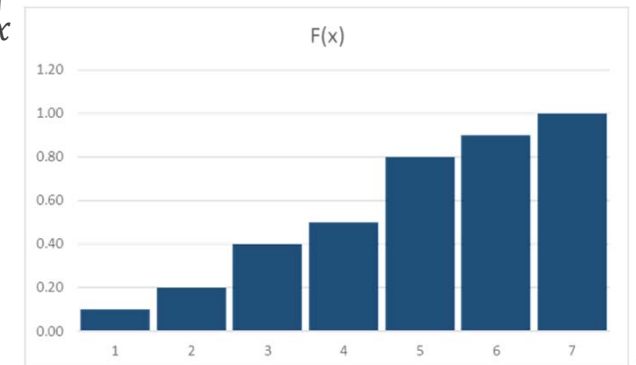
Funciones discretas

- Función de probabilidad
 $f(x) = p(X = x)$

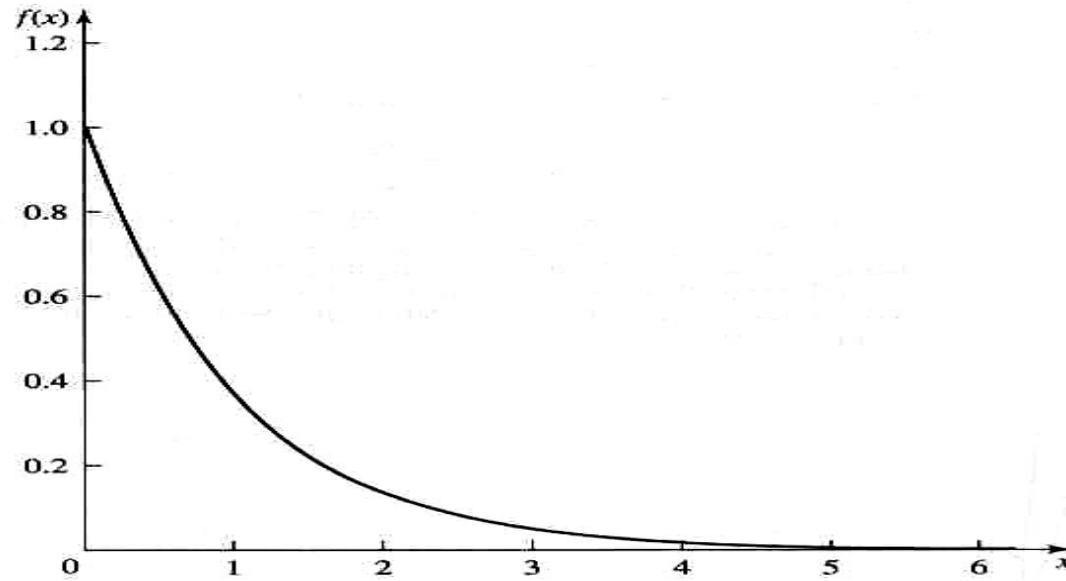


- Función de distribución

$$F(x) = p(X \leq x) = \sum_{x_i \leq x} p(x_i)$$



Ejemplos. Exponencial



Ej: $f(x)$ con $\beta=1$

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{si } x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

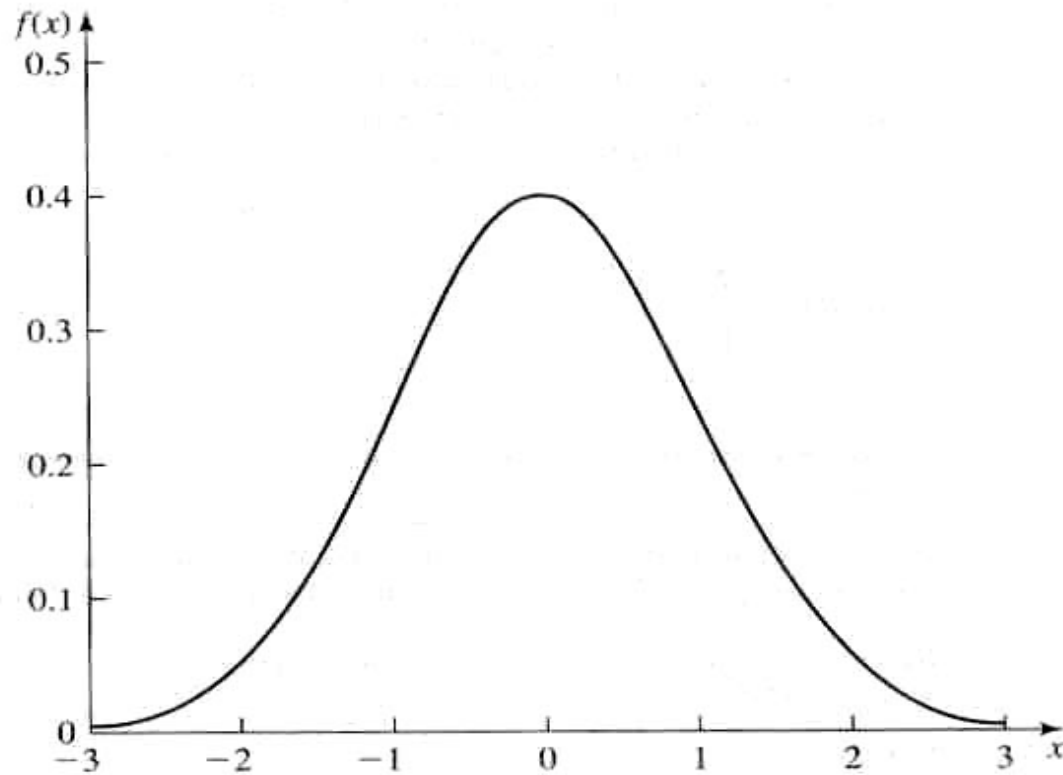
$$F(x) = \begin{cases} 1 - e^{-x/\beta} & \text{si } x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

- Rango: $[0, \infty)$

- Media: β

- Varianza: β^2

Ejemplos. Normal

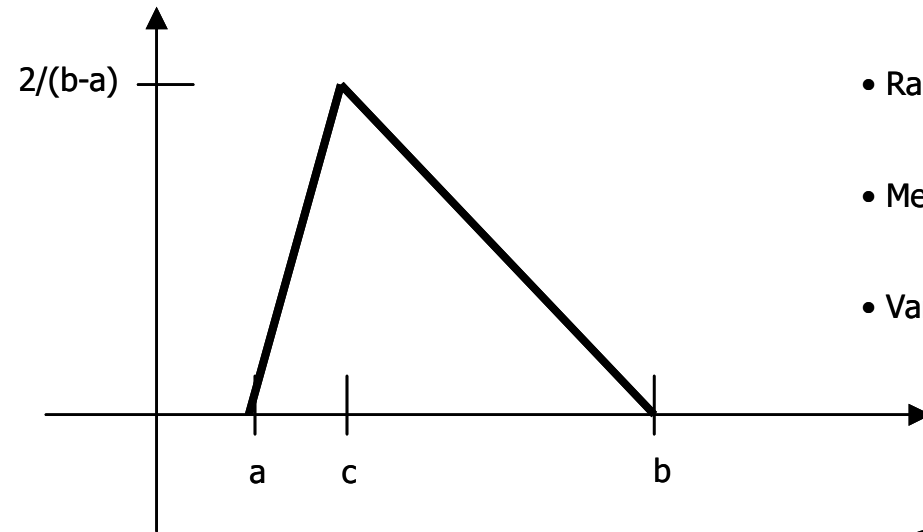


Normal (0,1)

- Rango: $(-\infty, \infty)$
- μ no acotada y σ positiva
- Media: μ
- Varianza: σ^2

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

Ejemplos. Triangular



- Rango: $[a, b]$

- Media: $(a+b+c)/3$

- Varianza: $(a^2 + b^2 + c^2 - ab - ac - bc)/18$

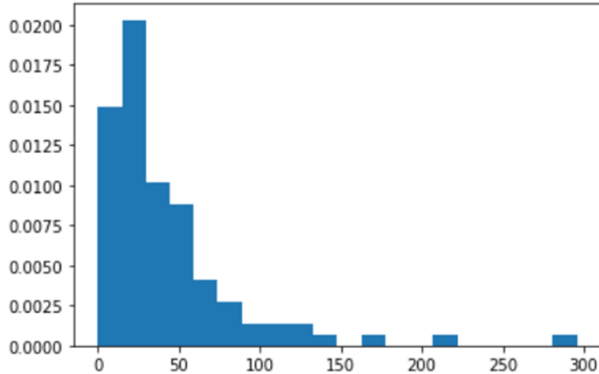
$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{si } a \leq x \leq c \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{si } c < x \leq b \\ 0 & \text{en otro caso} \end{cases}$$

$$F(x) = \begin{cases} 0 & x < a \\ \frac{(x-a)^2}{(b-a)(c-a)} & \text{si } a \leq x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{si } c < x \leq b \\ 1 & \text{si } b < x \end{cases}$$

Caracterización de las variables de entrada

Datos históricos

Distribución empírica



- Convertimos las frecuencias observadas en probabilidades
- Por ejemplo:
 - Si el 10% de las veces el tiempo entre averías estuvo en el intervalo $[0,20)$
 - La probabilidad de tener un tiempo entre averías en el intervalo $[0,20)$ es 0.1
- Generación de valores
 - Se selecciona un intervalo
 - Se elige uno valor dentro del intervalo
- Los valores que se obtienen no son siempre los mismos*

*más sobre este punto más adelante

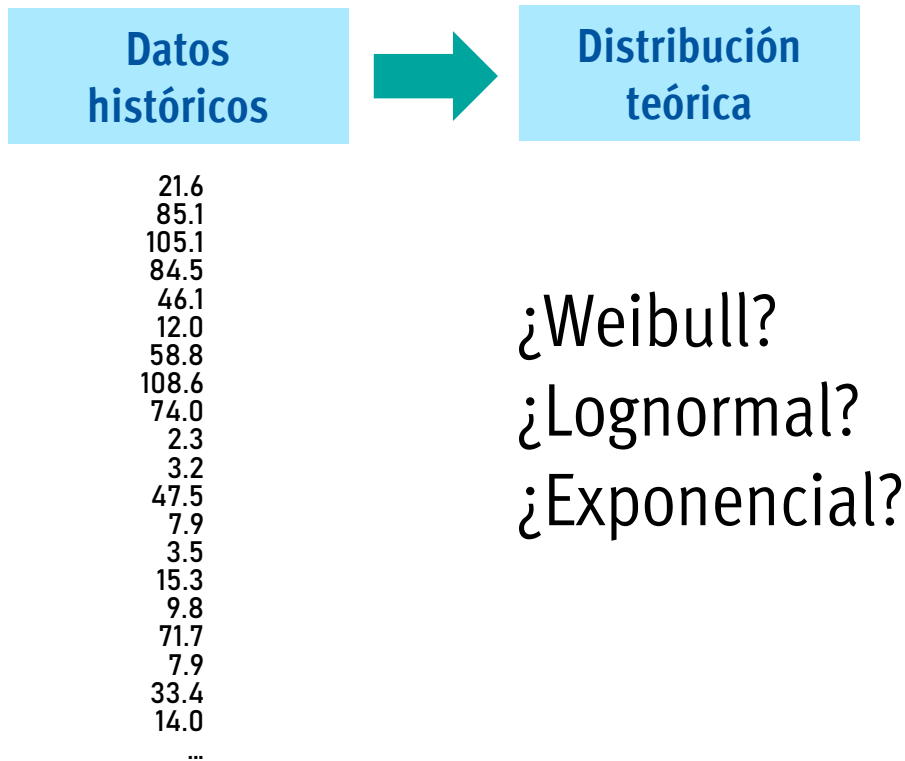
Caracterización de las variables de entrada

- Idoneidad de cada alternativa
 - Alimentar con datos históricos:
 - Suelen ser pocos.
 - Solo permiten simular lo que ocurrió y tal y como ocurrió en el pasado.
 - Adecuado: validación.
 - Obtener una función empírica:
 - Permiten obtener un conjunto “infinito” de valores.
 - Se puede general cualquier valor entre el máximo y el mínimo observados, pero no fuera de ese intervalo.
 - Obtener una función analítica:
 - Permiten obtener un conjunto “infinito” de valores.
 - Evita irregularidades.
 - Permite obtener valores fuera del rango de los observados.
 - Formulación más compacta.

Caracterización de las variables de entrada.

Tests de ajuste

- Lógica general:
 - Comprobar la independencia de los datos.
 - Plantear una hipótesis sobre el tipo de distribución.
 - Estimar los parámetros de la distribución.
 - Realizar un test de ajuste de los datos.
- Algunas técnicas:
 - Test χ^2 .
 - Test Kolmogorov-Smirnov.
- Los paquetes estadísticos incorporan estos tests.



Test de ajuste

χ^2

Distribuciones teóricas. Test de ajuste

Etapas

1. **Comprobar la independencia de los datos.**
2. Plantear una hipótesis sobre el tipo de distribución.
3. Estimar los parámetros de la distribución.
4. Realizar un test de ajuste de los datos.

Test de ajuste

Etapa 1. Independencia de los datos

- Muchas de las técnicas utilizadas en el ajuste de datos presuponen que las observaciones de la variable son independientes
- Ejemplo dependencia: esperas en cola de los clientes sucesivos en un puesto de servicio congestionado
- Test de independencia:
 - Representación de (x_i, x_{i+1})
 - Cálculo del coeficiente de correlación
$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

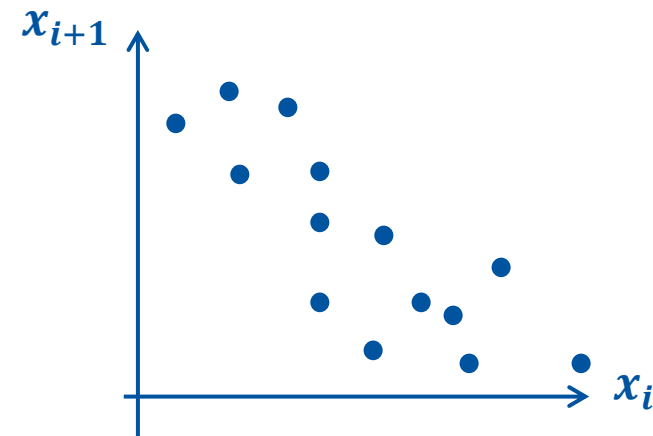
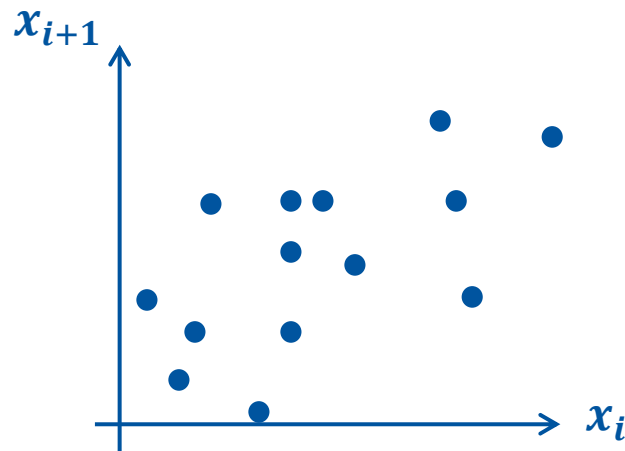
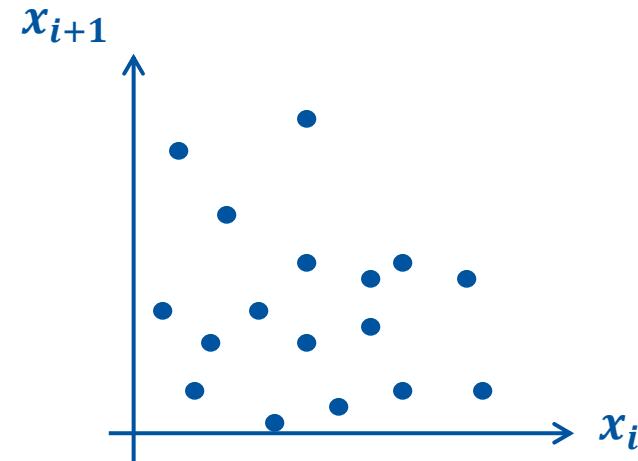
Test de ajuste

Etapa 1. Independencia de los datos

Nº obs.	Valor
1	x_1
2	x_2
...	...
i	x_i
...	...
n	x_n



(x_1, x_2)
 (x_2, x_3)
 (x_3, x_4)
...
 (x_{n-1}, x_n)



Test de ajuste

Etapa 2. Hipótesis sobre la distribución

- Observación de las medidas estadísticas de los datos

FUNCIÓN	ESTIMADOR	COMENTARIOS
Mínimo, máximo	$\min_i\{x_i\}, \max_i\{x_i\}$	Estimador del rango
Media μ	$\bar{X}(n)$	Medida de tendencia central
Mediana $X_{0.5}$	$\hat{X}_{0.5}(n) = \begin{cases} x_{((n+1)/2)} & \text{si } n \text{ impar,} \\ \frac{(x_{(n/2)} + x_{(n/2+1)})}{2} & \text{si } n \text{ par} \end{cases}$	Medida de tendencia central
Varianza σ^2	$S_{n-1}(n)$	Medida de variabilidad
Coficiente de variación σ/μ	$S_{n-1}(n)/\bar{X}(n)$	Medida de variabilidad

- Representación gráfica de los datos

Test de ajuste

Etapa 2. Hipótesis sobre la distribución

- División del rango de los datos en k intervalos de igual amplitud:

$$[b_0, b_1), [b_1, b_2), [b_1, b_2) \dots [b_{k-1}, b_k)$$

- Cálculo de la proporción de datos en cada intervalo (frecuencia): $h_1, h_2 \dots h_k$.

$$h_1 = \frac{n^{\circ} \text{ observaciones en } [b_0, b_1)}{n^{\circ} \text{ total de observaciones}}$$

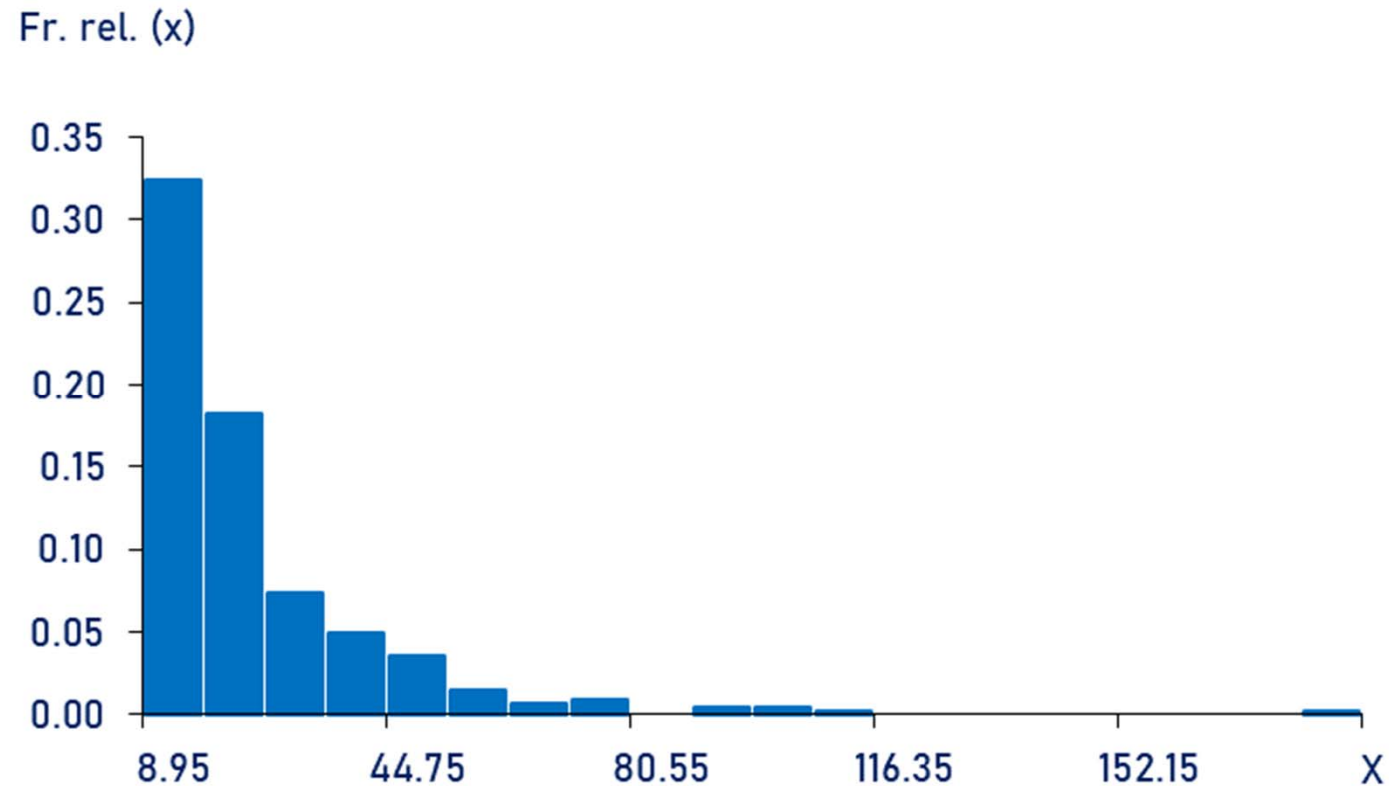
$$h_1 + h_2 + \dots + h_k = 1$$

- Representación de la frecuencia frente a los intervalos

Test de ajuste

Etapa 2. Hipótesis sobre la distribución

- Ejemplo de histograma



Test de ajuste

Etapa 3. Estimación de parámetros

- A la vista del histograma y el resto de consideraciones se formula una hipótesis sobre el tipo de distribución (ej.: exponencial).
- A continuación es necesario estimar los parámetros de la distribución.
- En general, utilizaremos el EMV (Estimador de Máxima Verosimilitud).

Test de ajuste

Etapa 3. Estimación de parámetros

- EMV: función de verosimilitud $L(\theta)$
 - Si x es discreta: $L(\theta) = p_{\theta}(x_1) \times p_{\theta}(x_2) \times \cdots \times p_{\theta}(x_n)$
 - Si x es continua: $L(\theta) = f_{\theta}(x_1) \times f_{\theta}(x_2) \times \cdots \times f_{\theta}(x_n)$
 - La estimación del parámetro $\hat{\theta}$, obtenida cumple:

$$L(\hat{\theta}) \geq L(\theta), \forall \theta$$

- Ejemplo: en el caso de una exponencial, el parámetro es la media de las observaciones.

Test de ajuste

Etapa 4. Ajuste

- Una vez caracterizada la distribución a la que “aparentemente” se ajustan los datos se puede realizar un test de ajuste
- Uno de los más conocidos es el test de la χ^2
- El test permite contrastar la hipótesis H_0 : los datos $x_1, x_2 \dots x_n$ son variables aleatorias IID con una función de densidad $f(x)$ de parámetros $\alpha, \beta \dots$

Test de ajuste

Etapa 4. Ajuste test χ^2 . Justificación

- Variable E_j : número de observaciones (de las n disponibles) pertenecen a un intervalo j :

- $E_j \sim \text{Binomial: } B(n, p_j)$

- $E_j \sim \text{Normal: } N\left(np_j, \sqrt{np_j(1-p_j)}\right)$ si $np_j > 5$ y $n(1-p_j) > 5$

- $\frac{E_j - np_j}{\sqrt{np_j(1-p_j)}} \sim N(0,1) = Z$

- $\frac{(E_j - np_j)^2}{np_j(1-p_j)} \sim Z^2$ y si $p_j \ll 1$: $\frac{(E_j - np_j)^2}{np_j} \sim Z^2$

Test de ajuste

Etapa 4. Ajuste test χ^2 . Justificación

- $\frac{(E_j - np_j)^2}{np_j} \sim Z^2$

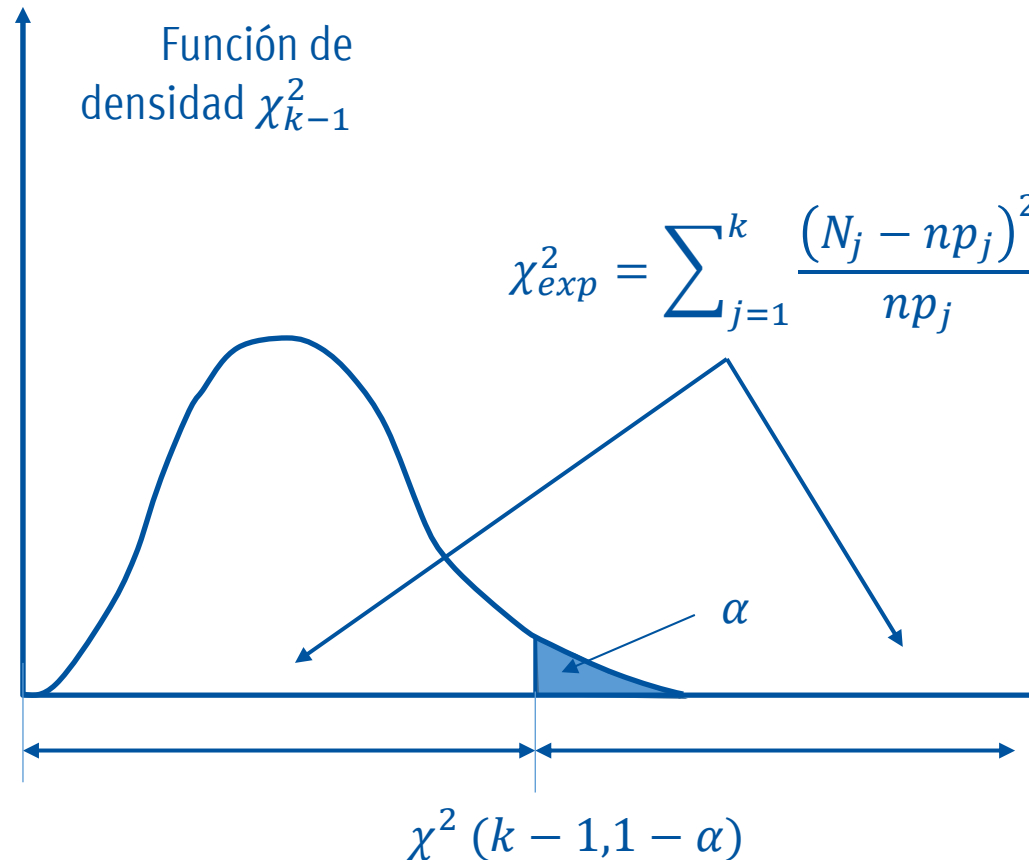
- $\sum_{j=1}^k Z^2 \sim \chi_{k-1}^2$

- $\sum_{j=1}^k \frac{(E_j - np_j)^2}{np_j} \sim \chi_{k-1}^2$

Test de ajuste

Etapa 4. Ajuste test χ^2 . Justificación

$$\sum_{j=1}^k \frac{(E_j - np_j)^2}{np_j} \sim \chi_{k-1}^2$$



Test de ajuste

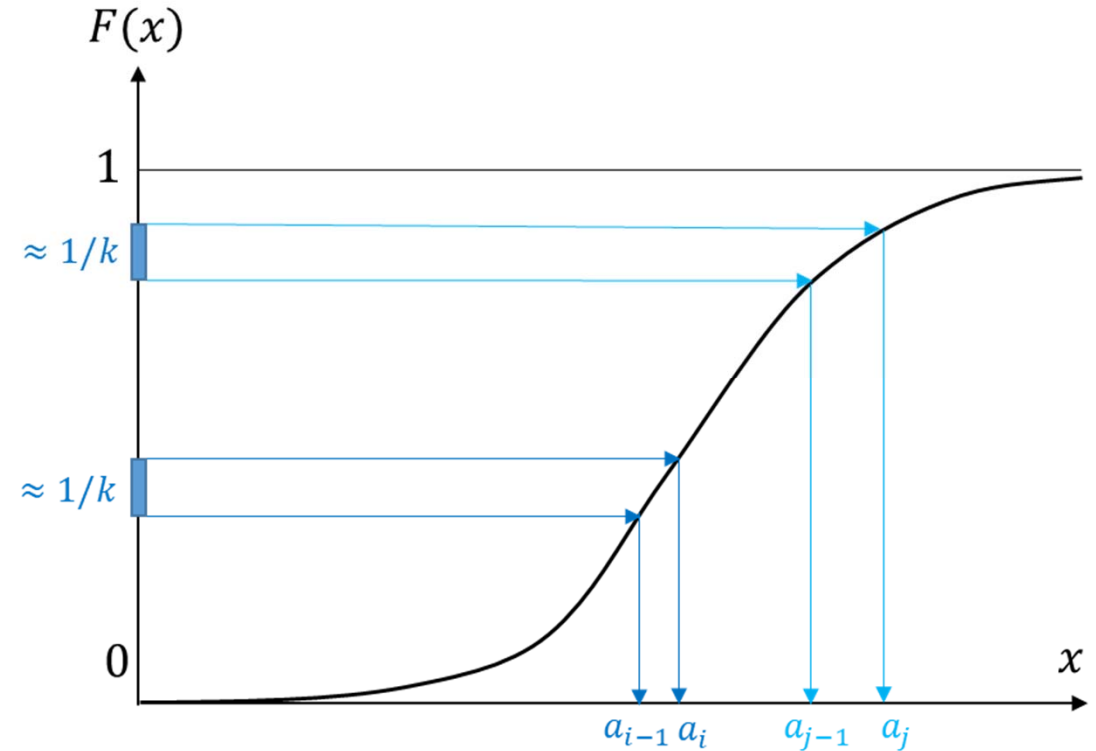
Etapa 4. Ajuste test χ^2

- Se divide el rango la variable X (variable teórica) en k intervalos:

$$[a_0, a_1), [a_1, a_2), [a_1, a_2) \dots [a_{k-1}, a_k)$$

(a_0 o a_k pueden alcanzar el valor $-\infty$ o $+\infty$)

- Con las siguientes condiciones
 - $k \geq 3$ $np_j \geq 5$
 - $p_j = p$ para todo j (la probabilidad de que una observación pertenezca a un intervalo debe de ser igual o muy parecida)
 - Donde $p_j = P(a_{j-1} \leq x < a_j), j = 1 \dots k$



Test de ajuste

Etapa 4. Ajuste test χ^2

- Dados los datos $x_1, x_2 \dots x_n$ se recuenta el número de ellos, N_j que pertenece a cada intervalo:

N_j = número de observaciones en el intervalo $[a_{j-1}, a_j)$

Naturalmente, $\sum_j N_j = n$

- Se calcula el valor del estadístico χ_{exp}^2 :

$$\chi_{exp}^2 = \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j}$$

Test de ajuste

Etapa 4. Ajuste test χ^2 . Resultado

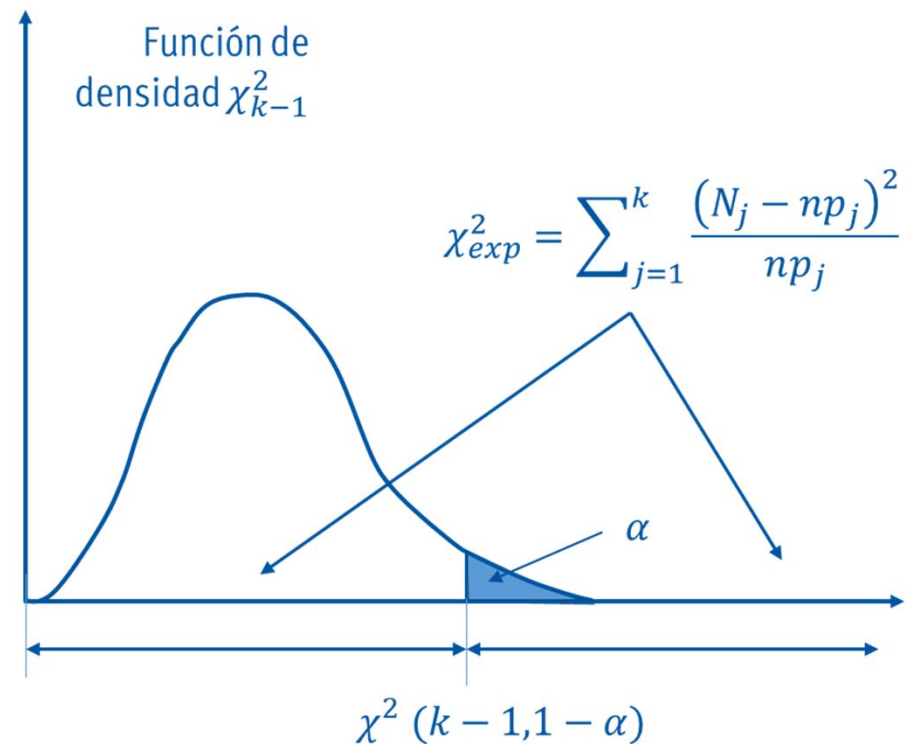
Se comparan χ_{exp}^2 y $\chi_{(k-1,1-\alpha)}^2$ con:

k : número de intervalos

α : nivel de significación ($1 - \alpha$: nivel de confianza)

Si $\chi_{exp}^2 > \chi_{(k-1,1-\alpha)}^2$ se rechaza la hipótesis nula

Si $\chi_{exp}^2 \leq \chi_{(k-1,1-\alpha)}^2$ no hay evidencia estadística para rechazar la hipótesis nula



Generar valores de las variables de entrada

- Dos aspectos esenciales:
 - Generación de números aleatorios $[0,1]$.
 - Generación de valores de las variables de entrada.
- Los entornos de simulación incorporan esta funcionalidad.

Generar valores de números aleatorios

Generador lineal congruente

- Se define una serie de números pseudoaleatorios U_i mediante una fórmula recurrente:

$$U_i = Z_i / m$$
$$Z_i = (a Z_{i-1} + c) \bmod m$$

Donde:

- Z_i : serie de números enteros
 - m : módulo
 - a : multiplicador
 - c : incremento
 - Z_0 : semilla
- m, a, c y Z_0 son números enteros no negativos, y deben satisfacer: $a < m$; $c < m$ y $Z_0 < m$

Generar valores de números aleatorios

Generador lineal congruente

- Los U_i pueden tomar alguno de los siguientes valores racionales: $0, 1/m, 2/m, \dots, (m-1)/m$
- El valor de m se suele tomar muy grande: 10^9 . De esa forma el conjunto es muy denso (aproximadamente, 1 billón de valores posibles)
- El generador tiene un ciclo a partir del cual se repiten sus valores. La longitud del ciclo se llama **periodo** del generador, de valor máximo: m
- Ejemplo de generador: $a = 16807, m = 2^{31} - 1, c = 0$

Generar valores de las variables de entrada

- Dos aspectos esenciales:
 - Generación de números aleatorios $[0,1]$.
 - **Generación de valores de las variables de entrada.**
- Los entornos de simulación incorporan esta funcionalidad.

Generar valores de valores de variables aleatorias

- Se requiere para proporcionar al modelo de simulación valores de las variables de entrada.
- Se generan a partir de números aleatorios.
- Existe una amplia variedad de técnicas, que varían en exactitud y eficiencia.
- La elección de la técnica depende del tipo de distribución de probabilidad.
- Ejemplos de técnicas: **transformada inversa**, composición, convolución...

Generar valores de valores de variables aleatorias. Transformada inversa

- Está basado en propiedades de la **función de distribución**.
- Procedimiento de aplicación:
 - Obtención de $F^{-1}(y)$
 - Generación de un número aleatorio $U(0,1): r$
 - Cálculo del valor de la variable x , como $x = F^{-1}(r)$

Generar valores de valores de variables aleatorias. Transformada inversa

La probabilidad de que

$$a \leq X \leq b$$

debe ser

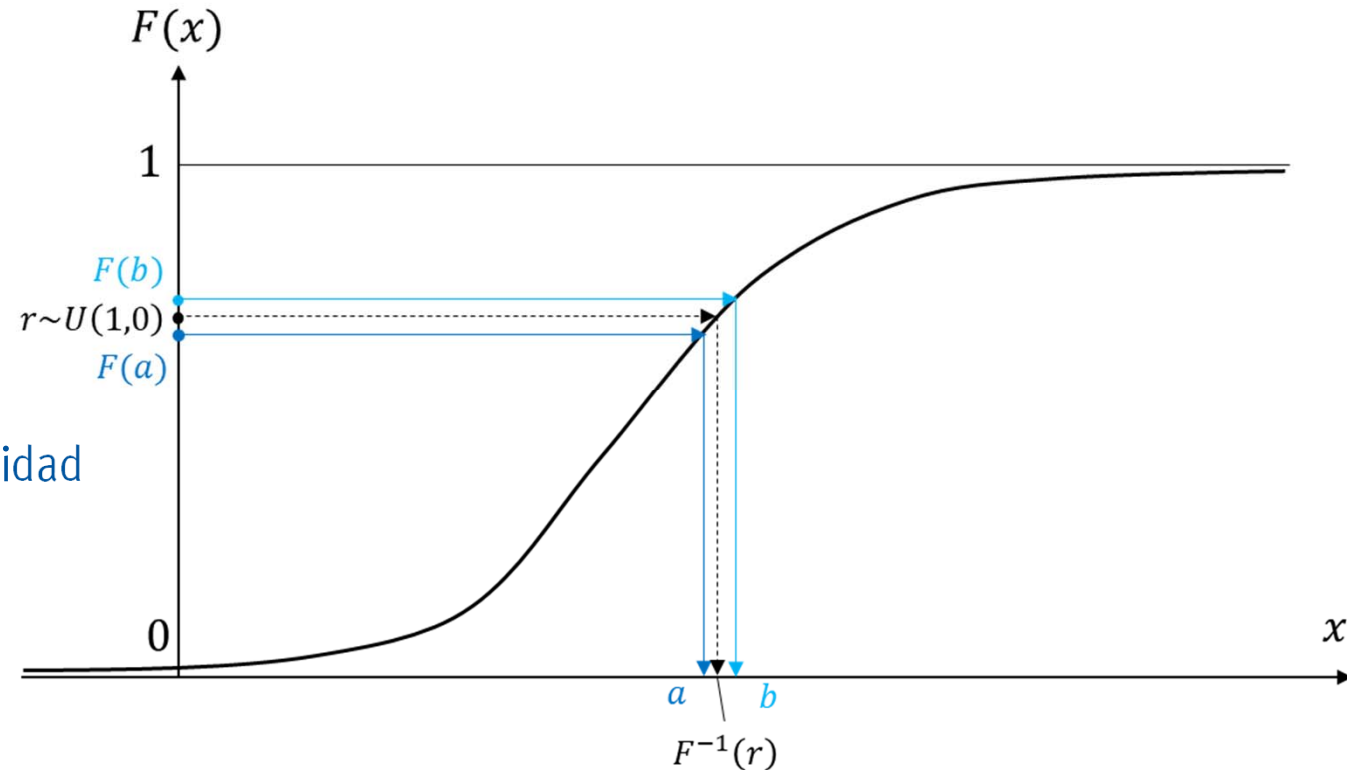
$$F(b) - F(a)$$

Al generar un número aleatorio $r \sim U(0,1)$, la probabilidad de que r esté entre $F(b) - F(a)$ es, precisamente:

$$F(b) - F(a)$$

Si genero el valor $x = F^{-1}(r)$ entonces:

$$p(a \leq x = F^{-1}(r) \leq b) = F(b) - F(a)$$



Generar valores de valores de variables aleatorias. Transformada inversa

- Generación de variables de una **exponencial** de media λ
- Función de distribución:

$$F(x) = \begin{cases} 1 - e^{-x/\lambda} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$y = 1 - e^{-x/\lambda} \Rightarrow y - 1 = e^{-x/\lambda} \Rightarrow \text{Log}(1 - y) = -\frac{x}{\lambda} \\ \Rightarrow x = -\lambda \text{Log}(1 - y)$$

$$F^{-1}(y) = -\lambda \text{Log}(1 - y)$$

