

# Identificación del modelo

---

Chequeo de la distribución

# Chequeo de la Distribución

---

- Supongamos que se observa una m. a. simple  $X_1, X_2, \dots, X_n$  de una v. a.  $X$  que tiene una cierta distribución de probabilidad. Existen muchos casos en los que dicha distribución es desconocida, por ello, en base al razonamiento o a la experiencia suponemos una cierta distribución para los datos, y luego debemos verificar si los datos realmente responden a la distribución postulada (identificación del modelo).

# Teorema para verificar distribuciones

---

Sea  $X$  una v. a. continua con función de distribución acumulada  $F_X$ , sea  $X_1, \dots, X_n$  una muestra aleatoria de  $X$  y sea  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  la muestra ordenada, entonces:

$$E(F_X(X_{(j)})) = \frac{j}{n+1}$$

**Aplicación:** Sea  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$  una muestra observada y ordenada, suficientemente grande, luego

$$E(F_X(X_{(j)})) = \frac{j}{n+1} \Rightarrow F_X(x_{(j)}) \approx \frac{j}{n+1}$$

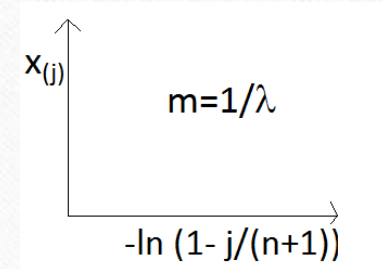
**Interpretación:** Los valores observados, se espera, que dividan el área debajo de la curva en  $n+1$  partes iguales, cada una de área  $1/(n+1)$

# Caso particular 1:

## Chequeo de la distribución Exponencial

Si los datos provienen de una distribución  $E(\lambda)$ , entonces, se debe cumplir que, para un  $n$  grande:

$$F_X(x_{(j)}) \approx \frac{j}{n+1} \Rightarrow 1 - e^{-\lambda x_{(j)}} \approx \frac{j}{n+1} \Rightarrow x_{(j)} \approx -\frac{1}{\lambda} \ln\left(1 - \frac{j}{n+1}\right)$$



Es decir que la relación entre los valores observados ordenados y los números  $-\ln\left(1 - \frac{j}{n+1}\right)$

debe ser aproximadamente lineal, si esto no ocurre se concluye que los datos no provienen de una distribución Exponencial.

**Nota:** No debe esperarse una relación lineal exacta

# Ejemplo: chequeo de la exponencial

Se desea verificar a partir de los datos:

0.53 0.03 1.12 0.53 0.23 0.16 1.39 3.71 1.61

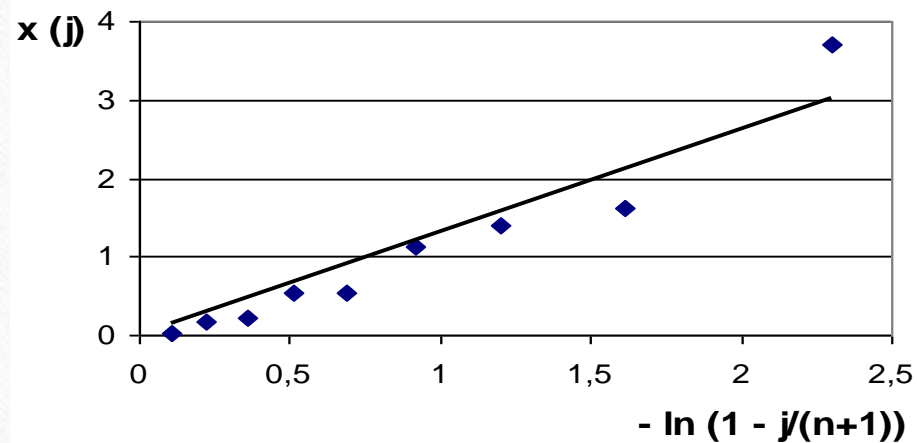
que el tiempo en horas que se demora en reparar una bomba, sigue una distribución Exponencial.

$$F_X(x_{(j)}) \cong \frac{j}{n+1} \Rightarrow 1 - e^{-\lambda x_{(j)}} \cong \frac{j}{n+1}$$

$$\Rightarrow -\lambda x_{(j)} \cong \ln\left(1 - \frac{j}{n+1}\right) \Rightarrow x_{(j)} \cong -\frac{1}{\lambda} \ln\left(1 - \frac{j}{n+1}\right)$$

j	$x_{(j)}$	$j/(n+1)$	$-\ln(1 - j/(n+1))$
1	0.03	0.1	0.11
2	0.16	0.2	0.22
3	0.23	0.3	0.36
4	0.53	0.4	0.51
5	0.53	0.5	0.69
6	1.12	0.6	0.92
7	1.39	0.7	1.20
8	1.61	0.8	1.61
9	3.71	0.9	2.30

$$x_{(j)} \cong -\frac{1}{\lambda} \ln \left( 1 - \frac{j}{n+1} \right)$$



j	$x_{(j)}$	$j/(n+1)$	$-\ln(1 - j/(n+1))$
1	0.03	0.1	0.11
2	0.16	0.2	0.22
3	0.23	0.3	0.36
4	0.53	0.4	0.51
5	0.53	0.5	0.69
6	1.12	0.6	0.92
7	1.39	0.7	1.20
8	1.61	0.8	1.61
9	3.71	0.9	2.30

Conclusión:

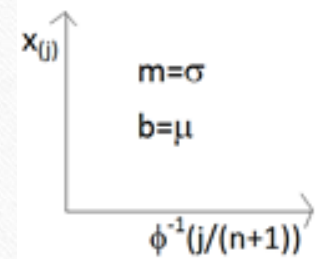
Como  $x(j)$  vs  $-\ln(1 - j/(n+1))$  ajusta a una recta, se concluye que los datos provienen de una distribución Exponencial.

# Caso particular 2:

## Chequeo de la distribución Normal

Si los datos provienen de una distribución  $N(\mu, \sigma)$ , entonces, aplicando el teorema anterior, para  $n$  grande:

$$\Phi\left(\frac{x_{(j)} - \mu}{\sigma}\right) \approx \frac{j}{n+1} \Rightarrow x_{(j)} \approx \mu + \sigma\Phi^{-1}\left(\frac{j}{n+1}\right)$$



Es decir que la relación entre los valores observados ordenados y los números  $\Phi^{-1}\left(\frac{j}{n+1}\right)$

debe ser aproximadamente lineal, si esto no ocurre se concluye que los datos no provienen de una distribución Normal.

**Nota:** No debe esperarse una relación lineal exacta

# Ejemplo: chequeo de una Normal

---

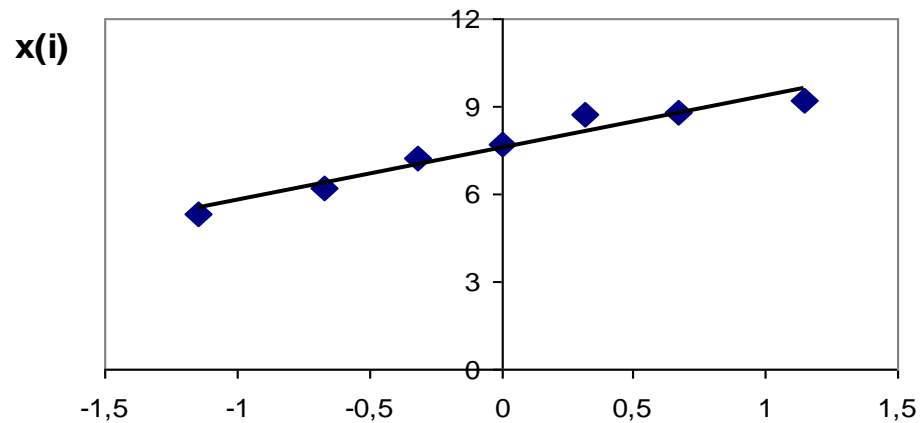
Los siguientes datos son las calificaciones obtenidas en una evaluación de niños de escuela primaria:

- 7.2   6.2   7.7   8.8   8.7   9.2   5.3

Se quiere verificar que los datos siguen efectivamente una distribución Normal

$$\Phi\left(\frac{x_{(i)} - \mu}{\sigma}\right) = \frac{i}{n+1} \Rightarrow \frac{x_{(i)} - \mu}{\sigma} = \Phi^{-1}\left(\frac{i}{n+1}\right) \Rightarrow x_{(i)} = \sigma \Phi^{-1}\left(\frac{i}{n+1}\right) + \mu$$

# Ejemplo: Chequeo de la Normal



$i$	$x_{(i)}$	$i/(n+1)$	$\phi^{-1}(i/(n+1))$
1	5.3	0.125	-1.15
2	6.2	0.250	-0.67
3	7.2	0.375	-0.32
4	7.7	0.500	0
5	8.7	0.625	0.32
6	8.8	0.750	0.67
7	9.2	0.875	1.15

Conclusión:

Como  $x_{(i)}$  vs  $\phi^{-1}\left(\frac{i}{n+1}\right)$  ajusta a una recta, se concluye que los datos provienen de una distribución Normal.

# Caso discreto: se puede aplicar con cualquier distribución discreta

(hacemos el Chequeo de una distribución de Poisson)

- Supongamos que los datos provienen de una distribución  $P(\lambda)$ .
- Se sabe que la frecuencia relativa de un cierto valor se aproxima a la probabilidad de ocurrencia de este, es decir

$$\frac{f_{\text{obs}}(x)}{n} \approx P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$
 por tratarse de una variable Poisson,  
luego aplicando logaritmo natural, para  $n$  grande

$$\ln(f_{\text{obs}}(x)) - \ln(n) \approx -\lambda + x \ln \lambda - \ln(x!)$$

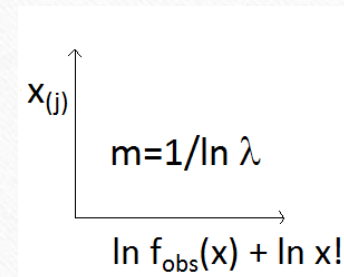
$$\ln(f_{\text{obs}}(x)) + \ln(x!) \approx x \ln \lambda + \ln(n) - \lambda$$

---

$$\ln(f_{\text{obs}}(x)) + \ln(x!) \approx x \ln \lambda + \ln(n) - \lambda$$

Entonces si graficamos  $x$  vs  $\ln(f_{\text{obs}}(x)) + \ln(x!)$   
debemos obtener aproximadamente una recta.

Si esto no ocurre se concluye que los datos no tienen distribución Poisson



# Ejemplo: Chequeo de una Poisson

---

La cantidad de errores de tipeo que comete una secretaria por página son: 2 2  
0 4 2 3 4 3 1 0.

Verificar que estos datos siguen efectivamente una distribución de Poisson.



x	$f_{\text{obs}}(x)$	$\ln f_{\text{obs}}(x) + \ln x!$
0	2	0.69
1	1	0
2	3	1.79
3	2	2.48
4	2	3.87
	10	

# Ejemplo: Chequeo de una Poisson

$$\frac{f_{\text{obs}}(x)}{n} \cong P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

$$\ln f_{\text{obs}}(x) \cong \ln n - \lambda + x \ln \lambda - \ln x!$$

$$\ln f_{\text{obs}}(x) + \ln x! = x \ln \lambda + (\ln n - \lambda)$$

$$x = \frac{1}{\ln \lambda} (\ln f_{\text{obs}}(x) + \ln x!) - \frac{1}{\ln \lambda} (\ln n - \lambda)$$

Conclusión:

Como  $x_{(i)}$  vs  $\ln f_{\text{obs}}(x) + \ln x!$  ajusta a una recta, se concluye que los datos provienen de una distribución Poisson.

x	$f_{\text{obs}}(x)$	$\ln f_{\text{obs}}(x) + \ln x!$
0	2	0.69
1	1	0
2	3	1.79
3	2	2.48
4	2	3.87
	10	

