USO DE LA PRUEBA DE CHI-CUADRADO Y K-S
CASO EXPONENCIAL

Prueba Chi-Cuadrado.

Consideremos los siguientes N ($N = 50$) datos que se sospechan provienen de una distribución exponencial:

4.8836	2.3710	5.4863	0.4128	0.5140	3.5525	0.0100	0.1258	0.6072	0.0385
1.8700	2.0893	0.3445	2.1685	2.1992	0.7800	3.5711	1.0869	1.6796	4.9266
3.3049	0.3198	0.8513	0.8234	0.3551	0.2849	1.9687	0.9654	0.8164	3.9926
6.6115	2.0679	0.7423	0.8220	7.6054	0.0406	6.2950	8.3504	2.0288	1.5100
2.2095	8.9253	0.4075	5.7358	0.1409	6.0335	0.2485	2.4816	0.6662	1.0702

Debemos establecer el número de intervalos k y las frecuencias observadas (O_i) y las esperadas (E_i) para cada intervalo. Cuando consideramos distribuciones uniformes, por simplicidad, cada intervalo tenia la misma amplitud que al mismo tiempo los hacia equiprobables. Si la distribución es exponencial, los intervalos de la forma $[a_o , a_1]$, $[a_1 , a_2]$,…, $[a_{k-1} , a_k]$, al hacerlos equiprobables, tendrán distintas amplitudes. Debemos seguir garantizando que $E_i \geq 5$. Esto implica que debemos tener, para los 50 números dados, no más de 50/5 = 10 intervalos. Fijando k tendremos que la probabilidad p de cada intervalo será $1/k$. Para los distintos intervalos tenemos que:

$$F(a_i) = 1 - e^{-\lambda a_i}$$

donde los a_i representan los puntos finales del i-esimo intervalo para $i = 1,2,\ldots,k$ con $a_o = 0$ y $a_k = \infty$. Como $F(a_i)$ es el acumulado de cero hasta a_i, podemos escribir:

$$ip = 1 - e^{-\lambda a_i} \quad \text{o} \quad e^{-\lambda a_i} = 1 - ip \quad \text{o} \quad a_i = -\frac{1}{\lambda}\ln(1 - ip) \quad \text{para } i = 0,1,\ldots,k$$

Para $k = 5$ ($p = 0.2$) y los datos anteriores tenemos que $\overline{X} = 2.3279$ y $\hat{\lambda} = 1/\overline{X} = 0.4296$ y obtenemos la siguiente tabla:

Intervalo	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
[0.000 , 0.519]	13	10	0.9
(0.519 , 1.189]	11	10	0.1
(1.189 , 2.133]	7	10	0.9
(2.133 , 3.747]	8	10	0.4
(3.747 , ∞]	11	10	0.1
	50	50	2.4

De las tablas obtenemos que $\chi^2_{[0.05,3]} = 7.81$ que es mayor que $\chi^2_0 = 2.4$ por lo que no hay evidencia de que los datos no sean exponenciales. En este caso los grados de libertad son $k - s - 1 = 5 - 1 - 1 = 3$ donde s es el numero de parámetros estimados de la muestra que en este caso es solo la media.

Prueba de Kolmogorov-Smirnov (K-S).

Consideremos los primero 10 números de la serie anterior y procedamos ahora con la prueba K-S:

```
4.8836  2.3710  5.4863  0.4128  0.5140  3.5525  0.0100  0.1258  0.6072  0.0385
```

i	x_i	$S_n(x_i) = i/10$	$F_X(x_i) = 1 - e^{-\hat{\lambda} x_i}$	$S_n(x_i) - F_X(x_i)$	$F_X(x_i) - S_n(x_{i-1})$
1	0.0100	0.1	0.0055	0.0945	0.0055
2	0.0385	0.2	0.0212	0.1788	-0.0788
3	0.1258	0.3	0.0675	0.2325	-0.1325
4	0.4128	0.4	0.2049	0.1951	-0.0951
5	0.5140	0.5	0.2484	0.2516	-0.1516
6	0.6072	0.6	0.2863	0.3137	-0.2137
7	2.3710	0.7	0.7321	-0.0321	0.1321
8	3.5525	0.8	0.8610	-0.0610	0.1610
9	4.8836	0.9	0.9337	-0.0337	0.1337
10	5.4863	1	0.9525	0.0475	0.0525
				D^+ = 0.3137	D^- = 0.1610
$\bar{X}$ =	1.8002			D_n = 0.3137	
$\hat{\lambda}$ =	0.5555				

Al igual que se hace un ajuste para la prueba de Chi-Cuadrado dado que estamos estimando parámetros de la muestra, en este caso debemos aplicar la siguiente tabla:

		$1-\alpha$				
Case	**Adjusted test statistic**	**0.850**	**0.900**	**0.950**	**0.975**	**0.990**
All parameters known	$\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right) D_n$	1.138	1.224	1.358	1.480	1.628
$N(\bar{X}(n), S^2(n))$	$\left(\sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}}\right) D_n$	0.775	0.819	0.895	0.955	1.035
$\text{expo}(\bar{X}(n))$	$\left(D_n - \frac{0.2}{n}\right)\left(\sqrt{n} + 0.26 + \frac{0.5}{\sqrt{n}}\right)$	0.926	0.990	1.094	1.190	1.308

Corrigiendo tenemos $\left(0.3137 - \frac{0.2}{10}\right)\left(\sqrt{10} + 0.26 + \frac{0.5}{\sqrt{10}}\right) = 1.0515$ que es menor que 1.094 por lo cual no rechazamos la hipótesis de la exponencialidad de los datos.