PROBABILIDAD Y ESTADÍSTICA
SEGUNDO PARCIAL 2004: SOLUCIÓN

Problema 1

a) i) Como f_X es una función de densidad debe cumplir $\int_{\mathbf{R}} f_X(x)dx = 1$. Por lo tanto:

$$1 = \int_a^\infty Ce^{-\lambda x} = \frac{C}{\lambda}e^{-\lambda a} \Rightarrow C = \lambda e^{\lambda a}.$$

ii) La mediana m_X queda definida por la condición $\int_{-\infty}^{m_X} f_X(x)dx = \frac{1}{2}$. De manera que, usando el valor de C calculado:

$$\frac{1}{2} = \int_a^{m_X} Ce^{-\lambda x} = 1 - e^{-\lambda(m_X - a)} \Rightarrow e^{-\lambda(m_X - a)} = \frac{1}{2}$$

y despejando m_X, se obtiene:

$$m_X = a + \frac{1}{\lambda}\log 2.$$

iii) El valor esperado está definido por $\mathbf{E}(X) = \int_{\mathbf{R}} x f_X(x)dx$. Calculando la integral resulta:

$$\mathbf{E}(X) = a + \frac{1}{\lambda}.$$

Otra manera de resolver: La densidad f_X corresponde a una variable exponencial desplazada en a. En efecto:

$$f_X(x) = \begin{cases} 0 & \text{si } x < a \\ \lambda e^{-\lambda(x-a)} & \text{si } a \leq x \end{cases}$$

y haciendo el cambio de variables $Y = X - a$ se obtiene una variable exponencial de parámetro λ. Usando la aditividad del valor esperado y el resultado conocido para el valor esperado de una exponencial se obtiene:

$$\mathbf{E}(X) = a + \mathbf{E}(Y) = a + \frac{1}{\lambda}.$$

iv) Usamos $\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X)$. Resulta:

$$\mathbf{E}(X^2) = \int_{\mathbf{R}} x^2 f_X(x)dx = a^2 + \frac{2}{\lambda^2} + \frac{2a}{\lambda}.$$

Con lo cual se obtiene:

$$\mathbf{Var}(X) = \frac{1}{\lambda^2}.$$

Otra manera de resolver: Usando propiedades de la varianza y la misma observación que en la parte anterior:

$$\mathbf{Var}(X) = \mathbf{Var}(a + Y) = \mathbf{Var}(Y) = \frac{1}{\lambda^2}.$$

b) i) El método de los momentos para la construcción de estimadores es equivalente a la resolución del siguiente conjunto de ecuaciones:

$$\begin{aligned} \overline{X_n} &= \mathbf{E}(X) = a + \frac{1}{\lambda} \\ \sigma_n^2 &= \mathbf{Var}(X) = \frac{1}{\lambda^2}, \end{aligned}$$

donde $\overline{X_n}$ y σ_n^2 son el promedio empírico y la varianza empírica respectivamente. La solución de ese sistema de ecuaciones nos proporciona los estimadores:

$$\hat{\lambda} = \frac{1}{\sigma_n} \quad \text{y} \quad \hat{a} = \overline{X_n} - \sigma_n.$$

ii) Estimadores consistentes para a y λ también pueden obtenerse resolviendo el siguiente conjunto de ecuaciones:

$$\begin{aligned} \overline{X_n} &= \mathbf{E}(X) = a + \frac{1}{\lambda} \\ m_n &= m = a + \frac{1}{\lambda}\log 2. \end{aligned}$$

Resulta:

$$\hat{\lambda} = \frac{1 - \log 2}{\overline{X_n} - m_n} \quad \text{y} \quad \hat{a} = \frac{m_n - \overline{X_n}\log 2}{1 - \log 2}.$$

c) El tiempo de recambio de la pieza número 144 es $Y_{144} = X_1 + X_2 + \ldots + X_{144}$ donde las variables X_i son independientes con distribución $X_i \sim PE(2, \lambda)$. Nos interesa estimar el máximo valor de λ para el cual se cumple

$$\mathbf{P}(Y_{144} \geq 576) \geq \frac{1}{2}.$$

Esta desigualdad es equivalente a la siguiente:

$$\mathbf{P}\left(\sqrt{\frac{144}{\mathbf{Var}(X)}}\left(\frac{Y_{144}}{144} - \mathbf{E}(X)\right) \geq \sqrt{\frac{144}{\mathbf{Var}(X)}}\,(576 - \mathbf{E}(X))\right) \geq \frac{1}{2}.$$

Ahora, usando

$$\mathbf{E}(X) = 2 + \frac{1}{\lambda} \quad \text{y} \quad \mathbf{Var}(X) = \frac{1}{\lambda^2}$$

y la aproximación del Teorema Central del Límite, el problema se reduce a calcular el máximo valor de λ para el cual se cumple

$$\mathbf{P}\left(N(0,1) \geq 12\lambda\left(2 - \frac{1}{\lambda}\right)\right) \geq \frac{1}{2},$$

de donde se obtiene la condición:

$$12\lambda\left(2 - \frac{1}{\lambda}\right) \leq 0,$$

que es equivalente a $\lambda \leq \frac{1}{2}$. De manera que

$$\lambda_{max} = \frac{1}{2}.$$

Problema 2

a) Para estudiar si los datos pueden suponerse aleatorios realizamos el *test de rachas de subidas y bajadas* y el *test de Spearman.* En caso de que la muestra pase ambos test consideraremos la hipótesis de aleatoriedad como cierta.

Test de Rachas

Recordamos que se definen las variables auxiliares U_i, que toman valores 0 ó 1, de la siguiente manera: $U_i = 1$ si $X_i \leq X_{i+1}$ y $U_i = 0$ en caso contrario.

X	U
1.90	
1.30	0
1.60	1
1.57	0
1.20	0
1.67	1
1.34	0
1.96	1
1.40	0
1.74	1
1.66	0
1.37	0

Hay 9 rachas, de manera que el estadístico del test es $R = 9$. De la tabla del test se obtiene el p-valor $\alpha^* = 0.2720$. Por lo tanto la muestra pasa el test.

Test de rangos de Spearman

Se ordenan los datos de menor a mayor. La tabla muestra los datos ordenados X_i^*, los índices i (esto es, el lugar que ocupaban los datos originalmente) y los rangos $R(X_i)$.

X_i^*	i	$R(X_i)$
1.20	5	1
1.30	2	2
1.34	7	3
1.37	12	4
1.40	9	5
1.57	4	6
1.60	3	7
1.66	11	8
1.67	6	9
1.74	10	10
1.90	1	11
1.96	8	12

De la tabla se obtiene $\sum_{i=1}^{12}(R(X_i) - i)^2 = 266$, y por lo tanto el el estadístico del test resulta

$$r_s = 1 - 6\frac{\sum_{i=1}^{12}(R(X_i) - i)^2}{12(12^2 - 1)} = 0.0699.$$

La tabla del test para $n = 12$ muestra que $\alpha^* \geq 0.1$. Por lo tanto la muestra pasa el test.

Podemos entonces asumir que los datos son aleatorios.

b) Estudiaremos ahora si la muestra tiene distribución normal. Se pide realizar un solo test, de manera que puede elegirse indistintamente entre el *test de normalidad de D'Agostino* y el *test de normalidad de Shapiro-Wilks.*

Test de normalidad de D'Agostino

El estadístico del test es

$$DA = \frac{1}{n^2\sigma_n}\sum_{i=1}^{n}\left(i - \frac{(n+1)}{2}\right)X_i^*,$$

donde $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X_n})^2$ es el estimador de la varianza.

A partir de los datos se obtiene $\overline{X_n} = 1.559$ y $\sigma_n^2 = 0.0537$.

Por otra parte:

$$\sum_{i=1}^{12}\left(i - \frac{13}{2}\right)X_i^* = 9.435$$

De donde resulta el valor del estadístico $DA = 0.2824$. La tabla correspondiente a $n = 12$ muestra que $\alpha^* > 0.2$, de manera que la muestra pasa el test.

Test de normalidad de Shapiro-Wilks

El estadístico del test es

$$W = \frac{b_n^2}{n\sigma_n^2},$$

donde $b_n = \sum_{i=1}^{[n/2]} a_i(X_{n-i+1}^* - X_i^*)$ y las constantes a_i se buscan en la tabla. En nuestro caso resulta:

$$\begin{aligned} b_n &= 0.5475(X_{12}^* - X_1^*) + 0.3325(X_{11}^* - X_2^*) + 0.2347(X_{10}^* - X_3^*) \\ &\quad +0.1586(X_9^* - X_4^*) + 0.0922(X_8^* - X_5^*) + 0.0303(X_7^* - X_6^*) \\ &= 0.7819. \end{aligned}$$

De donde se obtiene el valor del estadístico $W = 0.948$. La tabla correspondiente a $n = 12$ muestra que $\alpha^* > 0.5$ y por lo tanto la muestra pasa el test.

c) Para datos normales se pueden construir Intervalos de Confianza exactos para la media y la varianza.

Intervalo de Confianza exacto para la media

El Intervalo de Confianza para μ al nivel α cuando la muestra es gaussiana está dado por

$$I = \left[\overline{X_n} - t_{\frac{\alpha}{2}}(n-1)\frac{s_n}{\sqrt{n}}, \overline{X_n} - t_{\frac{\alpha}{2}}(n-1)\frac{s_n}{\sqrt{n}} \right],$$

donde $s_n^2 = \frac{n}{n-1}\sigma_n^2$ y $t_{\frac{\alpha}{2}}(n-1)$ se busca en la tabla de la distribución de Student con $n-1$ grados de libertad.

En nuestro caso, con $\alpha = 0.05$ y $n = 12$, obtenemos $t_{0.025}(11) = 2.201$. Por otra parte $\overline{X_n} = 1.559$ y $s_n = 0.2423$. De manera que el Intervalo de Confianza queda

$$I = [1.405, 1.713].$$

Intervalo de Confianza exacto para la varianza

El Intervalo de Confianza para σ^2 al nivel α cuando la muestra es gaussiana está dado por

$$I = \left[\frac{(n-1)s_n^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)s_n^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \right],$$

donde $\chi^2_{\frac{\alpha}{2}}(n-1)$ se busca en la tabla de la distribución χ^2 con $n-1$ grados de libertad.

En nuestro caso, con $\alpha = 0.05$ y $n = 12$, obtenemos $\chi^2_{0.025}(11) = 21.92$ y $\chi^2_{0.975}(11) = 3.82$. De manera que el Intervalo de Confianza queda

$$I = [0.0293, 0.1684].$$

d) Se asume que las muestras son independientes entre sí, por lo tanto podemos aplicar el *test de Kolmogorov-Smirnov para la comparación de 2 muestras.*

Test de Kolmogorov-Smirnov de comparación de 2 muestras

El estadístico del test es

$$D = \sup_{t \in \mathbf{R}} |F_m^X(t) - F_n^Y(t)|,$$

donde $F_m^X(t)$ y $F_n^Y(t)$ son las funciones de distribución empíricas para cada una de las muestras. En nuestro caso $m = n = 12$ y es fácil verificar que la máxima diferencia entre las funciones de distribución empíricas es

$$D = \frac{2}{12}.$$

Por lo tanto $mnD = 24$ y la tabla del test nos da $\alpha^* > 0.2$ y por lo tanto es razonable suponer que las muestras tienen la misma distribución.

Nota: No es válido aplicar el *test de Mann-Whitney-Wilcoxon* sin tener ninguna información adicional sobre la distribución de la Muestra 2, porque ese es un test para detectar corrimientos y en principio ni siquiera sabemos si las muestras tienen la misma forma para su distribución.