PROBABILIDAD Y ESTADÍSTICA
SEGUNDO PARCIAL, 28 de Junio de 2010

Problema 1 (20 puntos)

1. Se sabe que $\mathrm{E}(X) = a$ y $\mathrm{Var}(X) = 2b^2$, de donde $\mathrm{E}(X^2) = \mathrm{Var}(X) + \mathrm{E}(X)^2 = 2b^2 + a^2$.

   Sea $\overline{X}_n = \sum_{i=1}^{n} X_i$. Por el método de los momentos se obtienen estimadores $\widehat{a}$ y $\widehat{b}$ de a y b respectivamente, resolviendo el siguiente sistema:

$$\overline{X}_n = \widehat{a}$$
$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = 2\widehat{b}^2 + \widehat{a}^2$$

   es decir:

$$\overline{X}_n = \widehat{a}$$
$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2 = 2\widehat{b}^2$$

   Por lo tanto $\widehat{a} = \overline{X}_n$ es un estimador consistente de a y σ_n^2 es un estimador de $2b^2$. Luego, usando que la función $\sqrt{x}/2$ es continua se obtiene que $\widehat{b} = \frac{\sigma_n}{\sqrt{2}}$ es un estimador consistente de b.

2. La mediana m_X verifica que $F_X(m_X) = \int_{-\infty}^{m_X} f_X(t)dt = 1/2$. Observando que la densidad es simétrica respecto al parámetro a, es fácil deducir que $m_X = a$. En efecto:

$$F_X(a) = \int_{-\infty}^{a} \frac{1}{2b} e^{-\frac{t-a}{b}} dt = \frac{1}{2} e^{\frac{t-a}{b}} |_{-\infty}^{a} = \frac{1}{2} e^{\frac{a-a}{b}} = \frac{1}{2}$$

   Por lo tanto se tiene que la mediana empírica m_n es también un estimador consistente de a, donde:

$$m_n = \begin{cases} \frac{1}{2}(X^*_{\frac{n}{2}} + X^*_{\frac{n}{2}+1}) & \text{si } n \text{ es par} \\ X^*_{\frac{n+1}{2}} & \text{si } n \text{ es impar} \end{cases}$$

   siendo X_i^* los estadísticos de orden (muestra ordenada creciente).

3. Para $a = 0$, la densidad resulta $f_X(x) = \frac{1}{2b} e^{-|x|/b}$. La función de verosimilitud resulta entonces:

$$L(b) = \prod_{i=1}^{n} f_X(X_i) = \prod_{i=1}^{n} \frac{1}{2b} e^{-|X_i|/b}$$

   Maximizar $L(b)$ es equivalente a maximizar $\log(L(b)) = \sum_{i=1}^{n} -\log(2b) - \frac{|X_i|}{b}$. Derivando con respecto a b, se obtiene que:

$$\frac{\partial \log L(b)}{\partial b} = \frac{-n}{b} - \sum_{i=1}^{n} \frac{|X_i|}{b^2} = 0 \Leftrightarrow \frac{1}{b}\left(-n + \frac{1}{b}\sum_{i=1}^{n} |X_i|\right) \Leftrightarrow b = \frac{1}{n}\sum_{i=1}^{n} |X_i|$$

De estudiar el signo de la derivada, se deduce que $\widehat{b} = \frac{1}{n}\sum_{i=1}^{n} |X_i|$ es el estimador de máxima verosimilitud de b.

4. Se quiere calcular la función de distribución de $Y = |X|$ para $a = 0$. Observar que la variable Y sólo toma valores positivos, por lo tanto si $y < 0$, $F_Y(y) = 0$. Para $y > 0$,

$$F_Y(y) = P(Y \leq y) = P(|X| \leq y) = P(-y \leq X \leq y) = F_X(y) - F_X(-y)$$

donde para $x > 0$ se tiene que:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{0} \frac{1}{2b} e^{-\frac{t}{b}} dt + \int_{0}^{x} \frac{1}{2b} e^{-\frac{t}{b}} dt = \frac{1}{2} + \left(\frac{1}{2} - \frac{1}{2} e^{-\frac{x}{b}}\right) = 1 - \frac{1}{2} e^{-\frac{x}{b}}$$
$$F_X(-x) = \int_{-\infty}^{-x} \frac{1}{2b} e^{\frac{t}{b}} dt = \frac{1}{2} e^{\frac{-x}{b}}$$

Por lo tanto:

$$F_Y(y) = 1 - \frac{1}{2} e^{-\frac{y}{b}} - \frac{1}{2} e^{\frac{y}{b}} = 1 - e^{-y/b}$$

Se concluye entonces que Y tiene distribución exponencial de parámetro $1/b$.

Problema 2 (20 puntos)

1. Por la LFGN, se tiene que $\overline{X}_n$ es un estimador consistente de $\mu = \mathrm{E}(X_1) = \theta/2$. Por lo tanto $2\overline{X}_n$ es un estimador consistente de θ.

2. Se tiene que $\sigma^2 = \mathrm{Var}(X_1) = \frac{\theta^2}{12}$, por lo tanto $\sigma = \frac{\theta}{\sqrt{12}} = \frac{\mu}{\sqrt{3}}$. Como antes, se tiene que $\overline{X}_n$ es un estimador consistente de μ, de donde $\widehat{\sigma} = \frac{\overline{X}_n}{\sqrt{3}}$ es un estimador consistente de σ. Además, se tiene que:

$$\mathrm{E}(\widehat{\sigma}) = \frac{1}{\sqrt{3}} \mathrm{E}(\overline{X}_n) = \frac{\mu}{\sqrt{3}} = \sigma$$
$$\mathrm{Var}(\widehat{\sigma}) = \frac{1}{3} \mathrm{Var}(\overline{X}_n) = \frac{1}{3n} \mathrm{Var}(X_1) = \frac{\sigma^2}{3n}$$

3.

$$P\left(\theta \in \left[2\overline{X}_n - 2\frac{\overline{X}_n z_{\alpha/2}}{\sqrt{3n}}, 2\overline{X}_n + 2\frac{\overline{X}_n z_{\alpha/2}}{\sqrt{3n}}\right]\right) = P\left(\frac{\theta}{2} \in \left[\overline{X}_n - \frac{\overline{X}_n z_{\alpha/2}}{\sqrt{3n}}, \overline{X}_n + \frac{\overline{X}_n z_{\alpha/2}}{\sqrt{3n}}\right]\right)$$
$$= P\left(-\frac{\overline{X}_n z_{\alpha/2}}{\sqrt{3n}} \leq \overline{X}_n - \frac{\theta}{2} \leq \frac{\overline{X}_n z_{\alpha/2}}{\sqrt{3n}}\right) = P\left(-z_{\alpha/2} \leq \sqrt{3n}\left(\frac{\overline{X}_n - \theta/2}{\overline{X}_n}\right) \leq z_{\alpha/2}\right)$$
$$P\left(-z_{\alpha/2} \leq \sqrt{n}\left(\frac{\overline{X}_n - \theta/2}{\overline{X}_n/\sqrt{3}}\right) \leq z_{\alpha/2}\right)$$

Observando que $\theta/2 = E(X_1)$ y que $\overline{X}_n/\sqrt{3}$ es un estimador consistente de σ, se tiene que $\sqrt{n}\left(\frac{\overline{X}_n-\theta/2}{\overline{X}_n/\sqrt{3}}\right)$ tiene distribución aproximadamente normal. Por lo tanto, si $Z \sim \mathcal{N}(0,1)$, se tiene que:

$$P\left(\theta \in \left[2\overline{X}_n - 2\frac{\overline{X}_n z_{\alpha/2}}{\sqrt{3n}}, 2\overline{X}_n + 2\frac{\overline{X}_n z_{\alpha/2}}{\sqrt{3n}}\right]\right) \approx P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right)$$
$$= \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.$$

De donde se concluye que el intervalo propuesto, es un intervalo de confianza a nivel α para θ.

4. (a) Hay que probar que $P_{H_0}(\mathcal{R}_\alpha) \approx \alpha$, donde $P_{H_0}(\mathcal{R}_\alpha)$ es la probabilidad de rechazar H_0 asumiendo que la hipótesis nula es verdadera (error tipo I).

$$P_{H_0}(\mathcal{R}_\alpha) = P_{H_0}\left(\frac{\theta_0}{2\overline{X}_n} \leq 1 - \frac{z_\alpha}{\sqrt{3n}}\right) = P_{H_0}\left(\overline{X}_n - \theta_0/2 \geq \frac{z_\alpha \overline{X}_n}{\sqrt{3n}}\right)$$
$$P_{H_0}\left(\sqrt{n}\left(\frac{\overline{X}_n - \theta_0/2}{\overline{X}_n/\sqrt{3}}\right) \geq z_\alpha\right) \approx P(Z \geq z_\alpha) = \alpha$$

donde usamos que bajo H_0, $\mathrm{E}(X_1) = \theta_0/2$ y por lo tanto $\sqrt{n}\left(\frac{\overline{X}_n-\theta_0/2}{\overline{X}_n/\sqrt{3}}\right)$ tiene distribución aproximadamente normal.
Observar que en este caso también se podría usar que $\sigma = \theta_0/\sqrt{12}$, y en ese caso se tiene que $\sqrt{n}\left(\frac{\overline{X}_n-\theta_0/2}{\theta_0/\sqrt{12}}\right)$ tiene distribución normal, lo que conduce al mismo resultado.

(b) Se busca aproximar $\beta = P(\text{error tipo II}) = P_{H_1}(\mathcal{R}_\alpha^c)$, siendo $P_{H_1}(\mathcal{R}_\alpha^c)$ la probabilidad de no rechazar H_0 asumiendo que la hipótesis alternativa es verdadera. Con las mismas ideas que en el cáclulo de la parte anterior se tiene que:

$$\beta = P_{H_1}(\mathcal{R}_\alpha^c) = P_{H_1}\left(\frac{\theta_0}{2\overline{X}_n} \geq 1 - \frac{z_\alpha}{\sqrt{3n}}\right) = P\left(\overline{X}_n \leq \frac{\theta_0}{2} + \frac{\overline{X}_n z_\alpha}{\sqrt{3n}}\right)$$
$$= P\left(\sqrt{n}\left(\frac{\overline{X}_n - \theta_1/2}{\overline{X}_n/\sqrt{3}}\right) \leq \frac{\sqrt{n}}{\overline{X}_n/\sqrt{3}}\left(\frac{\theta_0}{2} - \frac{\theta_1}{2}\right) + z_\alpha\right) \approx \Phi\left(\frac{\sqrt{n}}{\overline{X}_n/\sqrt{3}}\left(\frac{\theta_0}{2} - \frac{\theta_1}{2}\right) + z_\alpha\right)$$

Observar que en este caso la media es $\theta_1/2$ pues estamos asumiendo que la hipótesis verdadera es H_1, de donde $\sqrt{n}\left(\frac{\overline{X}_n-\theta_1/2}{\overline{X}_n/\sqrt{3}}\right)$ tiene distribución aproximadamente normal. Para los valores indicados se tiene que $\beta = 0.0256$.
Al igual que en la parte anterior se podría usar $\theta_1/\sqrt{12}$ en lugar de $\overline{X}_n/\sqrt{3}$.

Problema 3 (20 puntos)

1. Para el test de Rachas, el estadístico es $R = 7$, y de la tabla con $n = 10$, se obtiene que el p-valor $\alpha^* = 0.4524$ en el caso:

$$\begin{cases} H_0 : \text{Los datos son iid} \\ H_1 : \text{Muchas rachas} \end{cases}$$

Dado que $\alpha^* > 0.1$, no se rechaza H_0.

Para el test de correlación de rangos de Spearman se tiene que $R_S = 0.2$ y de la tabla con $n = 10$ se obtiene que $\alpha^* = 0.292$ en el caso:

$$\begin{cases} H_0 : \text{Los datos son iid} \\ H_1 : \text{Tendencia creciente} \end{cases}$$

Por lo tanto, tampoco se rechaza H_0 y los datos pueden considerarse iid.

2. Se plantea el test de D'Agostino:

$$\begin{cases} H_0 : \text{Los datos son normales} \\ H_1 : \text{ no } H_0 \end{cases}$$

Se obtiene que el estadístico es $DA_n = 0.2817$. De la tabla con $n = 10$ se tiene que $\alpha^* > 0.2$. Por lo tanto no se rechaza la hipótesis nula de que los datos son normales.

Si el test elegido es el test de Lilliefors para normales, se tiene que $\mu_n = 0.495$ y $s_n = 0.1805$. En este caso:

$$\begin{cases} H_0 : \text{Los datos son normales} \\ H_1 : \text{ no } H_0 \end{cases}$$

El estadístico de prueba en este caso resulta $D_n = 0.13$ y de la tabla correspondiente se obtiene que $\alpha^* > 0.2$. Por lo tanto no se rechaza H_0.

3. Se plantea el test de hipótesis sobre la media de datos gaussianos con varianza desconocida:

$$\begin{cases} H_0 : \mu \geq 0.5 \\ H_1 : \mu < 0.5 \end{cases}$$

La región crítica en este caso es: $\mathcal{R}_\alpha = \left\{ \sqrt{n} \left(\frac{\overline{X}_n - 0.5}{s_n} \right) \leq -t_\alpha(n-1) \right\}$. Para $n = 10$ y $\alpha = 0.05$, de la tabla t-student se obtiene que $t_{0.05}(9) = 3.25$. A partir de los datos se obtiene que $\sqrt{n} \left(\frac{\overline{X}_n - 0.5}{s_n} \right) = -0.0876$. Por lo tanto, como el estadístico no pertenece a la región crítica, no se rechaza H_0. Entonces, no es posible afirmar que la media de la variable PV es inferior a 0.5.

4. Se plantea un test de hipótesis sobre la varianza para datos gaussianos:

$$\begin{cases} H_0 : \sigma^2 = 0.05 \\ H_1 : \sigma^2 \neq 0.05 \end{cases}$$

La región crítica en este caso es: $\mathcal{R}_\alpha = \left\{ \frac{(n-1)S_n^2}{\sigma^2} \notin \left(\chi_{1-\alpha/2}(n-1), \chi_{\alpha/2}(n-1) \right) \right\}$. Para $n = 10$ y $\alpha = 0.05$, se tiene que $\chi_{\alpha/2}(9) = 19.023$ y $\chi_{1-\alpha/2}(9) = 2.7$. A partir de los datos se tiene que $\frac{(n-1)S_n^2}{\sigma^2} = 5.865$, por lo tanto no se rechaza la hipótesis nula y se puede asumir que $\sigma^2 = 0.05$.