Soluciones del segundo parcial

1. En un pueblo andino en el que nieva todos los días de julio, se mide la cantidad de nieve caída cada día del mes. Un día del mes se considera un *récord* si la cantidad de nieve caída ese día es la máxima entre los días medidos hasta ese momento del mes. De esta forma, por ejemplo, el primer día de julio siempre es un *récord*.

   Si se sabe que la cantidad de nieve caída (en centímetros) en un día sigue una distribución uniforme $[0, 50]$,

   a) ¿Cuántos *récord* se espera tener a lo largo del mes de julio?

   b) ¿Cuál es el valor esperado de nieve caída en el último *récord* del mes?

   Aclaración: el mes de julio tiene 31 días.

   Solución: a) Sea X la cantidad de *récord* que se registran en el mes de julio. Queremos calcular $\mathbb{E}(X)$. Para eso, escribamos

$$X = \sum_{k=1}^{31} X_k$$

   , donde X_k vale 1 si el día k es un *récord* y 0 si no.

   Por linealidad de la esperanza, basta calcular $\mathbb{E}(X_k)$ para cada k. Notemos que, al ser Bernoulli, $\mathbb{E}(X_k) = \mathbb{P}($el día k es un *récord*$)$. Por lo tanto, calculemos esa probabilidad.

   Sabemos que $\mathbb{P}($el día 1 es un *récord*$) = 1$. Ahora calculemos $\mathbb{P}($el día 2 es un *récord*$)$:

   Para eso, definamos U_j como la cantidad de nieve caída en el día j, $j = 1, ..., 31$ y sabemos que son independientes. Por lo tanto, $\mathbb{P}($el día 2 es un *récord*$) = \mathbb{P}(U_1 < U_2)$.

   Recordemos que la densidad de U_j es $f_{U_j}(x) = \frac{1}{50}\mathbf{1}_{[0,50]}(x)$ para todo j, y que al ser independientes, la densidad conjunta será $f_{U_1,U_2}(x, y) = \frac{1}{50}\mathbf{1}_{[0,50]}(x)\frac{1}{50}\mathbf{1}_{[0,50]}(y) = \frac{1}{50^2}\mathbf{1}_{[0,50]}(x)\mathbf{1}_{[0,50]}(y)$.

   Luego, $\mathbb{P}(U_1 < U_2) = \mathbb{P}((U_1, U_2) \in B)$ con $B = \{(x, y) \in \mathbb{R}^2 : x < y\}$, y dicha probabilidad es igual a

$$\int_{-\infty}^{+\infty}\int_{x}^{+\infty} \frac{1}{50^2}\mathbf{1}_{[0,50]}(x)\mathbf{1}_{[0,50]}(y)dydx = \int_{0}^{50}\int_{x}^{50} \frac{1}{50^2}dydx = \frac{1}{2}$$

   Entonces $\mathbb{P}($el día 2 es un *récord*$) = \frac{1}{2}$.

   En general, $\mathbb{P}($el día k es un *récord*$) = \mathbb{P}(U_1 < U_k, U_2 < U_k, ..., U_{k-1} < U_k) = \mathbb{P}((U_1, ..., U_k) \in B)$ con $B = \{(x_1, ..., x_k) \in \mathbb{R}^k : x_i < x_k \forall i\}$, y dicha probabilidad es igual a

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{x_k}\cdots\int_{-\infty}^{x_k} \frac{1}{50^k}\mathbf{1}_{[0,50]}(x_1)\cdots\mathbf{1}_{[0,50]}(x_k)dx_1\cdots dx_{k-1}dx_k =$$

$$\int_{0}^{50}\int_{0}^{x_k}\cdots\int_{0}^{x_k} \frac{1}{50^k}dx_1\cdots dx_{k-1}dx_k = \frac{1}{k}$$

   Por lo tanto,

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{k=1}^{31} X_k\right) = \sum_{k=1}^{31}\mathbb{E}(X_k) = \sum_{k=1}^{31}\frac{1}{k}$$

   b) Notemos que el último *récord* será el máximo de las U_k. Por lo tanto, queremos calcular $\mathbb{E}(\max U_k)$. Para eso, calculemos la distribución del máximo.

$F_{\text{máx}\, U_k}(t) = \mathbb{P}(\text{máx}\, U_k \leq t) = \mathbb{P}(U_1 \leq t, ..., U_{31} \leq t) \overset{iid}{=} (\mathbb{P}(U_1 \leq t))^{31} = \left(\frac{t}{50}\right)^{31}$ si $0 \leq t \leq 50$ (vale 0 si $t \leq 0$ y vale 1 si $50 \leq t$). Por lo tanto la densidad será $f_{\text{máx}\, U_k}(t) = F_{\text{máx}\, U_k}(t)' = \frac{31}{50^{31}} t^{30} \mathbf{1}_{[0,50]}(t)$. Luego,

$$\mathbb{E}(\text{máx}\, U_k) = \int_{-\infty}^{+\infty} x \frac{31}{50^{31}} x^{30} \mathbf{1}_{[0,50]}(x) dx = \int_0^{50} \frac{31}{50^{31}} x^{31} dx = \frac{31}{50^{31}} \frac{50^{32}}{32} = \frac{31}{32} 50 \approx 48{,}4375$$

2. Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias independientes tales que $P(X_1 = 0) = 1$ y para $k \geq 2$:

$$P(X_k = k) = \frac{1}{2k \log(k)}, \quad P(X_k = -k) = \frac{1}{2k \log(k)}, \quad P(X_k = 0) = 1 - \frac{1}{k \log(k)},$$

Sea $S_n = \sum_{k=1}^n X_k$.

a) Probar que $\frac{S_n}{n} \xrightarrow{p} 0$.

b) Sea $A_k = \{|X_k| > \frac{k}{2}\}$. Probar que

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \right) = 1$$

c) Probar que $P(\frac{S_n}{n} \to 0) = 0$.

Solución: a) Sea $\varepsilon > 0$. Queremos probar que $\mathbb{P}(\left|\frac{S_n}{n}\right| > \varepsilon) \underset{n}{\to} 0$. Acotemos con Tchebychev y, para eso, calculemos $\mathbb{E}(\frac{S_n}{n})$ y $\text{Var}(\frac{S_n}{n})$, notando que $\mathbb{E}(X_1) = 0 = \text{Var}(X_1)$:

$$\mathbb{E}\left(\frac{S_n}{n}\right) = \frac{1}{n}\mathbb{E}(S_n) = \frac{1}{n}\sum_{k=2}^n \mathbb{E}(X_k)$$

$$\text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\text{Var}(S_n) \overset{indep}{=} \frac{1}{n^2}\sum_{k=2}^n \text{Var}(X_k)$$

Además,

$$\mathbb{E}(X_k) = k\frac{1}{2k\log(k)} - k\frac{1}{2k\log(k)} = 0$$

$$\text{Var}(X_k) = \mathbb{E}(X_k^2) - 0^2 = \mathbb{E}(X_k^2) = k^2\frac{1}{k\log(k)} = \frac{k}{\log(k)}$$

Por lo tanto,

$$\mathbb{E}\left(\frac{S_n}{n}\right) = 0$$

$$\text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{k=2}^n \frac{k}{\log(k)} \leq \frac{1}{\log(n)}$$

donde la última desigualdad vale porque la función $x/log(x)$ es creciente para $x > e$, cada sumando es menor igual que $n/log(n)$.

Luego, por Tchebychev, $\mathbb{P}(\left|\frac{S_n}{n}\right| > \varepsilon) \leq \frac{\text{Var}(\frac{S_n}{n})}{\varepsilon^2} \leq \frac{1}{\log(n)\varepsilon^2} \underset{n}{\to} 0$

b) Primero notemos que los eventos A_k son independientes ya que dependen de las X_k lo son.

Luego, podemos usar Borel Cantelli 2 y, si probamos que $\sum \mathbb{P}(A_k) = +\infty$, tendremos lo que queremos.

Notemos que $\mathbb{P}(A_k) = \frac{1}{k\log(k)}$ y por lo tanto $\sum \mathbb{P}(A_k) = \sum \frac{1}{k\log(k)}$.

Una forma de ver que esa serie diverge es usar el criterio de la integral y probar que $\int \frac{1}{xlog(x)}dx$ diverge.

c) Observemos que por el inciso anterior, vale que $|X_n/n| > \frac{1}{2}$ infinitas veces con probabilidad 1. En particular, se deduce que $P(X_n/n \to 0) = 0$ ya que si $\omega \in \Omega$ es tal que $|X_n(\omega)/n| > \frac{1}{2}$ infinitas veces entonces $X_n(\omega) \nrightarrow 0$.

Por último, afirmamos que si para un $\omega \in \Omega$ vale que $S_n(\omega)/n \to 0$ entonces necesariamente $X_n(\omega)/n \to 0$. Esto va a implicar que $P(S_n/n \to 0) \leq P(X_n/n \to 0) = 0$. Para probar la afirmación, observemos que si $S_n(\omega)/n \to 0$ entonces:

$$\frac{X_n(\omega)}{n} = \frac{S_n(\omega) - S_{n-1}(\omega)}{n} = \frac{S_n(\omega)}{n} - \frac{n-1}{n} \cdot \frac{S_{n-1}(\omega)}{n-1} \to 0.$$

Otra solución: Notemos que el ítem b) sigue valiendo aún sin módulo (la serie quedaría $\sum \mathbb{P}(A_k) = \sum \frac{1}{2k\log(k)}$ que también diverge). Eso quiere decir que para cada $\omega \in \Omega$ existen infinitos n para los cuales $X_n/n > 1/2$, es decir $X_n = n$. Fijemos un $\omega \in \Omega$ y consideremos esos subíndices que dijimos. Para esos n,

$$\frac{S_n}{n} - \frac{S_{n-1}}{n-1} = \frac{S_{n-1}+X_n}{n} - \frac{S_{n-1}}{n-1} = \frac{S_{n-1}+n}{n} - \frac{S_{n-1}}{n-1} = 1 + \frac{(n-1)S_{n-1} - nS_{n-1}}{n(n-1)} = 1 - \frac{S_{n-1}}{n(n-1)} \geq 1/2$$

donde en la última desigualdad se usó que $S_{n-1} = X_1 + ... + X_{n-1} \leq 1 + ... + (n-1) = n(n-1)/2$.

Luego, eso dice que para todo $\omega \in \Omega$, S_n/n no converge. Por lo tanto $\mathbb{P}(S_n/n \to 0) = 0$.

3. Sea $(X_n)_{n\in\mathbb{N}}$ una sucesión de variables aleatorias iid con media $\mu_X \neq 0$, varianza σ_X^2, y sea $(Y_n)_{n\in\mathbb{N}}$ una sucesión de variables aleatorias iid con media μ_Y, varianza σ_Y^2, tal que Y_j, X_k son independientes para cualquier elección de j, k.

   a) Hallar el límite en distribución de $Z_n = \sqrt{n}\left(\frac{\overline{Y_n}}{\overline{X_n}} - \frac{\mu_Y}{\mu_X}\right)$

   b) Si g es una función continua tal que $g(\mu_Y) = \sigma_X$, probar que $W_n = \sqrt{n}\left(\frac{\overline{X_n} - \mu_X}{g(\overline{Y_n})}\right) \xrightarrow{\mathcal{D}} Z \sim N(0,1)$

Solución: a) Escribamos a Z_n de otra forma, para poder usar los teoremas que conocemos.

$$Z_n = \sqrt{n}\left(\frac{\overline{Y_n}}{\overline{X_n}} - \frac{\mu_Y}{\mu_X}\right) = \frac{1}{\overline{X_n}\mu_X}\sqrt{n}\left(\overline{Y_n}\mu_X - \overline{X_n}\mu_Y\right)$$

Ahora observemos lo siguiente: definiendo $W_n = Y_n\mu_X - X_n\mu_Y$, tenemos que $\overline{W_n} = \overline{Y_n}\mu_X - \overline{X_n}\mu_Y$.

Además, $\mathbb{E}(W_n) = \mathbb{E}(Y_n)\mu_X - \mathbb{E}(X_n)\mu_Y = \mu_Y\mu_X - \mu_X\mu_Y = 0$ y $\mathrm{Var}(W_n) = \sigma_Y^2\mu_X^2 + \sigma_X^2\mu_Y^2$ (aquí usamos la independencia).

Luego:

- Por LGN, $\overline{X_n} \xrightarrow{p} \mu_X$ y, por Slutzky, $\dfrac{1}{\overline{X_n}\mu_X} \xrightarrow{p} \dfrac{1}{\mu_X^2}$
- Por TCL, $\sqrt{n}\overline{W_n} \xrightarrow{\mathcal{D}} N(0, \sigma_Y^2\mu_X^2 + \sigma_X^2\mu_Y^2)$

Por lo tanto, por Slutzky de nuevo,

$$Z_n \xrightarrow{\mathcal{D}} \frac{1}{\mu_X^2} N(0, \sigma_Y^2\mu_X^2 + \sigma_X^2\mu_Y^2) \sim N(0, \frac{\sigma_Y^2\mu_X^2 + \sigma_X^2\mu_Y^2}{\mu_X^4})$$

b) Observemos que por LGN, $\overline{Y_n} \xrightarrow{p} \mu_Y$ y, al ser g continua, $g(\overline{Y_n}) \xrightarrow{p} g(\mu_Y) = \sigma_X$. Además, por el TCL sabemos que $\sqrt{n}\left(\overline{X_n} - \mu_X\right) \xrightarrow{\mathcal{D}} N(0, \sigma_X^2)$. Luego, por Slutzky, $W_n \xrightarrow{\mathcal{D}} \frac{1}{\sigma_X} N(0, \sigma_X^2) \sim N(0,1)$

4. Martín tiene una gata color gris de 2 años. Su vecino Maxi, también tiene una gata gris pero de 1,5 años. Un día Martín llega a su casa y se encuentra con las dos gatas adentro. Para decidir cuál es la suya lo que hace es observar a una gata y contar la cantidad de minutos que pasan hasta que maulla por primera vez.

   Sea Y la edad de la gata observada, donde $\mathbb{P}(Y = 2) = \frac{1}{3}$ y $\mathbb{P}(Y = 1,5) = \frac{2}{3}$ y supongamos que la cantidad de minutos X que pasan hasta que maulla la gata, dado que $Y = y$, es una variable aleatoria exponencial de parámetro $\frac{1}{2y-1}$.

   a) Hallar la densidad de X y calcular $\mathbb{P}(X < 2)$.

   b) Hallar $\mathbb{E}(X \mid Y)$, $\mathbb{E}(X)$ y $\mathrm{Var}(X)$.

   c) Calcular la probabilidad de que Martín haya observado a su gata y haya tenido que esperar al menos 2 minutos.

Solución: a) Sabemos que $X \mid Y = y \sim \exp(\frac{1}{2y-1})$, es decir $F_{X|Y=y}(t) = 1 - e^{-\frac{t}{2y-1}}$. Por lo tanto, $F_X(t) = \mathbb{P}(X \leq t) \overset{probatotal}{=} \mathbb{P}(X \leq t \mid Y = 2)\mathbb{P}(Y = 2) + \mathbb{P}(X \leq t \mid Y = 1,5)\mathbb{P}(Y = 1,5) = (1 - e^{-\frac{t}{3}})\frac{1}{3} + (1 - e^{-\frac{t}{2}})\frac{2}{3}$. Entonces la densidad de X será

$$f_X(t) = F_X(t)' = \frac{1}{9}e^{-\frac{t}{3}} + \frac{1}{3}e^{-\frac{t}{2}}$$

si $t \geq 0$ y 0 si no.

Por lo que calculamos antes, $\mathbb{P}(X < 2) = \mathbb{P}(X \leq 2) = F_X(2) = (1 - e^{-\frac{2}{3}})\frac{1}{3} + (1 - e^{-\frac{2}{2}})\frac{2}{3} = 1 - e^{-\frac{2}{3}} - e^{-1}$

b) Como $X \mid Y = y \sim \exp(\frac{1}{2y-1})$, $\mathbb{E}(X \mid Y = y) = 2y - 1$. Por lo tanto, por el principio de sustitución, $\mathbb{E}(X \mid Y) = 2Y - 1$.

Para calcular $\mathbb{E}(X)$, recordemos que $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid Y)) \overset{a)}{=} \mathbb{E}(2Y - 1) \overset{linealidad}{=} 2\mathbb{E}(Y) - 1$. Ahora calculemos $\mathbb{E}(Y)$, que al ser discreta, la podemos calcular como $\mathbb{E}(Y) = 2 \cdot \frac{1}{3} + 1,5 \cdot \frac{2}{3} = \frac{5}{3}$. Finalmente, $\mathbb{E}(X) = 2 \cdot \frac{5}{3} - 1 = \frac{7}{3}$.

Para calcular $\mathrm{Var}(X)$, usaremos la fórmula $\mathrm{Var}(X) = \mathrm{Var}(\mathbb{E}(X \mid Y)) + \mathbb{E}(\mathrm{Var}(X \mid Y))$ (ej 5 de la práctica 10)

Por lo anterior, $\mathrm{Var}(\mathbb{E}(X \mid Y)) = \mathrm{Var}(2Y - 1) = 4\mathrm{Var}(Y)$ y $\mathbb{E}(\mathrm{Var}(X \mid Y)) = \mathbb{E}((2Y-1)^2) = 4\mathbb{E}(Y^2) - 4\mathbb{E}(Y) + 1$.

La esperanza de Y ya la calculamos, y $\mathbb{E}(Y^2) = 2^2 \cdot \frac{1}{3} + (1,5)^2\frac{2}{3} = \frac{17}{6}$. Por lo tanto $\mathrm{Var}(Y) = \frac{17}{6} - (\frac{5}{3})^2 = \frac{1}{18}$.

Finalmente, $\mathrm{Var}(X) = 4\frac{1}{18} + 4\frac{17}{6} - 4\frac{5}{3} + 1 = \frac{2}{9} + \frac{102}{9} - \frac{60}{9} + 1 = \frac{53}{9}$

c) Queremos calcular $\mathbb{P}(Y = 2, X \geq 2)$. Pasando al condicional, $\mathbb{P}(Y = 2, X \geq 2) = \mathbb{P}(X \geq 2 \mid Y = 2)\mathbb{P}(Y = 2)$. Ahora, notando que $\mathbb{P}(X \geq 2 \mid Y = 2) + \mathbb{P}(X < 2 \mid Y = 2) = 1$, tenemos que $\mathbb{P}(Y = 2, X \geq 2) = (1 - \mathbb{P}(X < 2 \mid Y = 2))\mathbb{P}(Y = 2) = (1 - (1 - e^{-\frac{2}{3}}))\frac{1}{3} = \frac{e^{-\frac{2}{3}}}{3}$

5. Sea $X_1, ..., X_n$ una muestra aleatoria de una distribución con densidad

$$f_\theta(x) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}} \mathbf{1}_{[0,+\infty)}(x)$$

a) Hallar el estimador de máxima verosimilitud de θ.

b) Decidir si el estimador hallado es insesgado o asintóticamente insesgado.

c) Decidir si el estimador hallado es consistente.

Solución: a) La función de verosimilitud será

$$L(\theta\ x_1, ..., x_n) = \prod_{k=1}^{n} f_\theta(x_k) = \frac{\prod_{k=1}^{n} x_k}{\theta^n} e^{-\frac{\sum_{k=1}^{n}(x_k)^2}{2\theta}} \prod_{k=1}^{n} \mathbf{1}_{[0,+\infty)}(x_k)$$

Al ser $L(\theta\ \ x_1, ..., x_n)$ una función no negativa, y al ser log una función creciente, maximizar $L(\theta\ x_1, ..., x_n)$ es equivalente a maximizar

$$g(\theta) = \log(L(\theta\ x_1, ..., x_n)) = (\sum_{k=1}^{n} x_k) - n\log(\theta) - \frac{\sum_{k=1}^{n}(x_k)^2}{2\theta}$$

Calculemos

$$\partial g(\theta) = \frac{-n}{\theta} + \frac{\sum_{k=1}^{n}(x_k)^2}{2\theta^2}$$

Igualando a cero y despejando, obtenemos el estimador $\hat{\theta} = \frac{\sum_{k=1}^{n}(x_k)^2}{2n}$.

Para verificar que es un máximo, basta calcular la derivada segunda y chequear que sea negativa en dicho valor.

b) Tenemos que chequear si $\mathbb{E}(\hat{\theta}) = \theta$. Para eso, por la linealidad de la esperanza, debemos hallar $\mathbb{E}(X_k^2)$. Hay dos formas de hacerlo: una, haciendo la cuenta, teniendo en cuenta que X_k^2 es una función de X_k (y por lo tanto su esperanza será integrar $x^2 f(x)$).

Otra forma es la siguiente: probaremos que $X_k^2 \sim \exp(\frac{1}{2\theta})$, y por lo tanto $\mathbb{E}(X_k^2) = 2\theta$. Veamos eso: integrando la densidad dada, $F_X(x) = 1 - e^{-\frac{x^2}{2\theta}}$ y por lo tanto $F_{X_k^2}(x) = \mathbb{P}(X_k^2 \leq x) = \mathbb{P}(X_k \leq \sqrt{x}) - \mathbb{P}(X_k \leq -\sqrt{x}) = F_X(\sqrt{x}) - F_X(-\sqrt{x}) = 1 - e^{-\frac{x}{2\theta}} - 0 = 1 - e^{-\frac{x}{2\theta}}$, que es la acumulada de una exponencial de parámetro $\frac{1}{2\theta}$.

Por lo tanto, $\mathbb{E}(\hat{\theta}) = \frac{\sum_{k=1}^{n}\mathbb{E}(X_k^2)}{2n} = \frac{2\theta n}{2n} = \theta$ y el estimador resulta insesgado.

c) Tenemos que chequear si $\hat{\theta} \xrightarrow{p} \theta$. Apelando a la LGN (pues probamos que X_k^2 es una exponencial y por lo tanto tiene varianza finita), sabemos que $\frac{\sum_{k=1}^{n}(X_k)^2}{n} \xrightarrow{p} \mathbb{E}(X_k^2) = 2\theta$. Luego, como al multiplicar por constantes la convergencia en probabilidad se mantiene (por ejemplo por Slutzky), $\hat{\theta} = \frac{1}{2}\frac{\sum_{k=1}^{n}(X_k)^2}{n} \xrightarrow{p} \frac{1}{2}\mathbb{E}(X_k^2) = \theta$, y el estimador resulta consistente.