1	2	3	4	Calificación

Probabilidad y Estadística (C)

Segundo parcial - 2C 2018 - 05/12/2018

Complete esta hoja y entréguela con el resto del examen. ***Realizar cada ejercicio en hoja separada****. Escribir el nombre en cada una. Al retirarse debe firmar una hoja de asistencia.*

Apellido y nombres:.................................. Nº de libreta: ....

mail:...........................@.................... Firma:.....................

Turno: | Tarde: 14 a 17 hs | Noche: 19 a 22 hs | Nº de hojas entregadas:

Criterio de aprobación: Para aprobar este examen es necesario sumar al menos 60 puntos.

Recuerde definir con palabras los eventos y/o las variables aleatorias involucradas, nombres y parámetros de las distribuciones. ***Justifique claramente sus afirmaciones****.*

1. Se sabe que un famoso jugador de basket hace siempre 15 tiros de doble y 5 tiros de triple por partido. La probabilidad de que enceste en un tiro de doble es 0.8 y la probabilidad de que enceste en un tiro de triple es 0.6. Las tiradas al aro son independientes. Una temporada consiste de 30 partidos para su equipo. Tenga presente que cada pelota encestada por tiro doble marca dos puntos, mientras que cada pelota encestada por tiro triple marca tres puntos.

a) (6 p) Sea X el número de puntos que hace el jugador en un partido. Verificar que $\mathbb{E}(X) = 33$ y $\text{Var}(X) = 20.4$.

b) (7 p) Dar una cota inferior para la probabilidad de que en una temporada el jugador consiga anotar entre 950 y 1030 puntos.

c) (6 p) Aproximar la probabilidad de que en una temporada el jugador anote entre 950 y 1000 puntos.

Otro famoso jugador de básket es tal que si D el es número de dobles que encesta en un partido y T es el número de triples que encesta en un partido, entonces $\mathbb{E}(D) = 12, \text{Var}(D) = 4$, $\mathbb{E}(T) = 4$, $\text{Var}(T) = 1$ y su correlación es $\rho_{D,T} = \frac{1}{2}$.

d) (7 p) Hallar la esperanza y varianza de la variable aleatoria Y, dada por el número de puntos que hace en un partido. Aproximar la probabilidad de que en una temporada supere los 1050 puntos.

2. Denotemos por X a la proporción de tiempo por día que un estudiante elegido al azar dedica a estudiar. Supongamos que la función de densidad de X es

$$f(x, \theta) = (\theta + 1)x^{\theta} \mathbb{1}_{(0,1)}(x),$$

donde θ es un parámetro desconocido, tomando valores en el conjunto de reales mayores a -1 ($\theta \in \Theta = (-1, \infty)$).

a) (5 p) Hallar $\widehat{\theta}_n$, el estimador de máxima verosimilitud de θ.

b) (4 p) Hallar la distribución de $Y = -\ln(X)$. Indique a que familia pertenece y cuánto vale $\mathbb{E}(Y)$.

c) (6 p) Hallar el sesgo $\widehat{\theta}_n$. Es el estimador insesgado? Es asintóticamente insesgado?

d) (6 p) Hallar el error cuadrático medio de $\widehat{\theta}_n$.

e) (4 p) Probar que $\widehat{\theta}_n$ es un estimador consistente de θ.

3. El rendimiento anual de almendros (en kg.) en parcelas cultivadas en el noroeste de Mendoza es una variable aleatoria con distribucón normal de media μ y varianza σ^2. Se observa el rendimiento anual de $n = 5$ parcelas elegidas al azar y se obtiene una media muestral de 525 kilogramos y un desvío muestral de 10 kilogramos.

a) (8 puntos) En base a los datos obtenidos, obtenga la estimación por intervalos para μ utilizando un procedimiento de nivel 0.95. El procedimiento utilizado para construir el intervalo, ¿es exacto o asintótico? Justifique su respuesta.

b) (8 puntos) El rendimiento anual medio de las parcelas de la provincia de La Rioja es de 500 kg. ¿Hay razones para sospechar que el rendimiento medio en Mendoza es mayor?

c) (9 puntos) Decidir si son verdaderas o falsas las siguientes afirmaciones:

  (a) Aproximadamente el 95% de las parcelas del noroeste de Mendoza tiene un rendimiento que pertenece al intervalo hallado en a), en base a la muestra observada.

  (b) La probabilidad de que el intervalo anterior contenga a μ es aproximadamente 0.95.

  (c) Si se extraen muchas muestras de tamaño 5 de manera independiente y para cada una se calcula el intervalo de confianza como en a), aproximadamente el 95% de estos intervalos contendrá a μ.

4. Históricamente, el 20% del mercado prefiere el jabón de la marca A. Para incrementar las ventas, la empresa A realiza una intensa campaña de publicidad. Al finalizar la misma se entrevistan $n = 400$ individuos y se les pregunta si prefieren o no la marca A, procurando demostrar que la campaña fue exitosa.

a) (5 p) Exprese H_0 y H_1 en términos de p, la probabilidad de que un cliente prefiera el jabón de la marca A al finalizar la campaña.

b) (5 p) El gerente de la empresa decide concluir que la campaña de publicidad es exitosa si al menos 92 (es decir, si 92 o más) de los 400 clientes entrevistados prefiere la marca A. Especifique cuál es la región de reachazo de la hipótesis nula H_0 e indique, de manera aproximada, cual es el nivel del criterio propuesto.

c) (9 p) El dueño de la empresa quiere establecer otro criterio de modo que, con probabilidad 0.05, la campaña se declare exitosa cuando en realidad no lo fue. Construya una región de rechazo para satisfacer al dueño. Si al realizar la encuesta se observa que 92 de los 400 encuestados preferirían la marca A, calcule de manera aproximada el p-valor correspondiente a los datos obtenidos. Con estos datos, se rechaza la hipótesis nula a nivel 0.05?

d) (6 p) Hallar de manera aproximada la probabilidad de cometer un error de tipo II con el criterio propuesto por el dueño si en realidad $p = 0.24$.

Resolución

1. a) $D \sim \text{Bi}(15, 0.8)$, $T \sim \text{Bi}(5, 0.6)$

$$\begin{aligned} X &= 2D + 3T \\ \mathbb{E}(X) &= 2\mathbb{E}(D) + 3\mathbb{E}(T) = 30 \cdot 0.8 + 15 \cdot 0.6 = 33 \\ \text{Var}(X) &= 4\text{Var}(D) + 9\text{Var}(T) = 60 \cdot 0.8 \cdot 0.2 + 45 \cdot 0.6 \cdot 0.4 = 20.4 \end{aligned}$$

b)

$$\begin{aligned} P(950 \leq 30\overline{X}_{30} \leq 1030) &= P(-40 \leq 30\overline{X}_{30} - 990 \leq 40) = 1 - P(|30\overline{X}_{30} - 990| > 40) \\ &\geq 1 - \frac{30 \cdot 20.4}{40^2} = 0.6175 \end{aligned}$$

c)

$$\begin{aligned} P(950 \leq 30\overline{X}_{30} \leq 1000) &= P\left(-\frac{40}{\sqrt{612}} \leq \frac{30\overline{X}_{30} - 990}{\sqrt{612}} \leq \frac{10}{\sqrt{612}}\right) \\ &\approx \Phi\left(\frac{10}{\sqrt{612}}\right) + \Phi\left(\frac{40}{\sqrt{612}}\right) - 1 \approx 0.6554 + 0.9474 - 1 = 0.6028 \end{aligned}$$

d)

$$\begin{aligned} \mathbb{E}(Y) &= 2\mathbb{E}(D) + 3\mathbb{E}(T) = 36 \\ \text{Var}(Y) &= 4\text{Var}(D) + 12\text{Cov}(D, T) + 9\text{Var}(T) \\ &= 4\text{Var}(D) + 12\rho_{D,T}\sqrt{\text{Var}(D)\text{Var}(T)} + 9\text{Var}(T) = 37 \\ P(30\overline{Y}_{30} > 1050) = P(\overline{Y}_{30} > 35) &= P\left(\frac{\overline{Y}_{30} - 36}{\sqrt{37/30}} > -\sqrt{\frac{30}{37}}\right) = \Phi\left(\sqrt{\frac{30}{37}}\right) \approx 0.8159 \end{aligned}$$

2. a)

$$f_{X_1 \dots X_n}(x_1, \dots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

Como $f(x, \theta) = 0$ para $x \notin (0, 1)$, asumo que $x_i \in (0, 1)$ para todo i, pues $P(X_i \in (0, 1)) = 1$. Luego, $f(x_i, \theta) > 0$ para todo i, y puedo aplicar ln.

$$\ln f_{X_1 \dots X_n}(x_1, \dots, x_n, \theta) = \sum_{i=1}^{n} \ln f(x_i, \theta)$$

$$\frac{\partial \ln f_{X_1 \dots X_n}(x_1, \dots, x_n, \theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln f(x_i, \theta)}{\partial \theta}$$

$$\ln f(x_i, \theta) = \ln(\theta + 1) + \theta \ln(x_i) \quad \frac{\partial \ln f(x_i, \theta)}{\partial \theta} = \frac{1}{\theta + 1} + \ln(x_i)$$

$$0 = \frac{n}{\theta + 1} + \sum_{i=1}^{n} \ln(x_i) \Rightarrow \hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \ln(X_i)} - 1$$

b)

$$\begin{aligned} F_{-\ln(X)}(y) &= P(X \geq e^{-y}) = 1 - F_X(e^{-y}) \\ f_{-\ln(X)}(y) &= f_X(e^{-y})e^{-y} = (\theta + 1)e^{-(\theta+1)y}\mathbb{1}_{(0,1)}(e^{-y}) = (\theta + 1)e^{-(\theta+1)y}\mathbb{1}_{(0,+\infty)}(y) \end{aligned}$$

$Y \sim \mathcal{E}(\theta + 1) \Rightarrow \mathbb{E}(Y) = (\theta + 1)^{-1}$

c) $-n\overline{\ln(X)} \sim \Gamma(n, \theta+1)$

$$E\left(\hat{\theta}\right) = n\int_0^{+\infty} \frac{e^{-(\theta+1)x}x^{n-2}(\theta+1)^n}{\Gamma(n)}\,dx - 1 = \frac{n}{n-1}(\theta+1) - 1$$
$$b\left(\hat{\theta}\right) = \frac{\theta+1}{n-1}$$

No es insesgado, pero es asintóticamente insesgado.

d)

$$V\left(\hat{\theta}\right) = n^2\int_0^{+\infty} \frac{e^{-(\theta+1)x}x^{n-3}(\theta+1)^n}{\Gamma(n)}\,dx - n^2\frac{(\theta+1)^2}{(n-1)^2} = n^2\frac{(\theta+1)^2}{(n-1)(n-2)} - n^2\frac{(\theta+1)^2}{(n-1)^2}$$
$$= \frac{n^2(\theta+1)^2}{(n-1)^2(n-2)}$$
$$\text{ECM}\left(\hat{\theta}\right) = \frac{n^2+n-2}{n-2}\frac{(\theta+1)^2}{(n-1)^2} = \frac{(n+2)(\theta+1)^2}{(n-2)(n-1)}$$

e) Es asintóticamente insesgado y su varianza tiende a 0. Alternativamente, $-\overline{\ln(X)} \xrightarrow{p} (\theta+1)^{-1}$ por LGN, y por continuidad $\hat{\theta} \xrightarrow{p} \theta$.

3. a)

$$1-\alpha = P\left(\frac{|\overline{X}-\mu|}{S/\sqrt{n}} \leq t_{n-1,\alpha/2}\right) \Rightarrow \frac{|525-\mu|}{10/\sqrt{5}} \leq t_{4,0.025} = 2.7765$$
$$[512.583114521, 537.416885479]$$

Es exacto, pues $\sqrt{n}(\overline{X}-\mu)/S \sim t_{n-1}$.

b) Sí. Si se plantea el test de hipótesis $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$, con μ_0 menor al extremo inferior del intervalo obtenido, se tiene que

$$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} > t_{n-1,\alpha/2}$$

Por lo que la hipótesis nula se rechazaría para el nivel de significación $\alpha/2$. En este caso particular, $500 < 512.583114521$ y se rechaza la hipótesis nula para el nivel 0.025.

c) i. Falso. El porcentaje aproximado de parcelas con rendimiento en el intervalo $[a, b]$ es

$$P\left(a \leq X \leq b\right) = P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Como puede observarse, esto depende de μ y σ por lo que no se puede asegurar cuánto vale sin conocer tales parámetros. Por ejemplo, si fijamos σ y $\mu \to +\infty$ la expresión tiende a 0, mientras que si fijamos $\mu = a$ y $\sigma \to 0^+$ la expresión tiende a 1/2.

ii. Falso. μ es un valor desconocido pero determinístico por lo que, o bien $\mu \in [512.58, 537.42]$, o bien $\mu \notin [512.58, 537.42]$.

iii. Verdadero. Como se dijo en el ítem anterior, μ es determinístico, pero tanto $\overline{X} - 2.7765S/\sqrt{5}$ como $\overline{X} + 2.7765S/\sqrt{5}$ son variables aleatorias, cuyo valor depende del resultado del experimento aleatorio considerado: elegir al azar 5 parcelas del noroeste de Mendoza y observar su rendimiento anual. En consecuencia, se puede hablar de la probabilidad

$$P\left(\mu \in \left[\overline{X} - \frac{2.7765}{\sqrt{5}}S, \overline{X} + \frac{2.7765}{\sqrt{5}}S\right]\right)$$

y por la contrucción realizada en el ítem a, esta probabilidad es 0.95, por lo que si se repite muchas veces el experimento aleatorio mencionado, aproximadamente el 95% de los intervalos así calculados contendrá a μ.

4. a) $H_0 : p = 0.2$, $H_1 : p > 0.2$

b)

$$P\left(\frac{\overline{X} - 0.2}{0.02} \geq 1.5\right) \approx 1 - 0.9332 = 0.0668$$

c)

$$0.05 \approx P\left(\frac{\overline{X} - 0.2}{0.02} \geq 1.64\right)$$

El p-valor fue calculado en el ítem anterior. La hipótesis nula no se rechaza pues $0.0668 > 0.05$.

d)

$$P\left(\frac{\overline{X} - 0.2}{0.02} < 1.64\right) = P\left(\frac{\overline{X} - 0.24}{\sqrt{0.1824/20}} < -\frac{0.144}{\sqrt{0.1824}}\right) \approx 1 - \Phi(0.34) = 0.3669$$