Segundo PBL. 16 de Abril de 2009

Estadística I. GDO-ADE 73, 74, 75, 79 y 80

Alumnos (por orden alfabético): Grupo pequeño n^o:

1.

2.

3.

Puntuación sobre 7.5 puntos. Tiempo de realización: 1 hora y 15 min.
Todas las respuestas finales y los cálculos intermedios deben escribirse en esta misma hoja en los espacios reservados para ello. Expresad el resultado final con 4 decimales.

1. **(2 pts.)** Sean X e Y dos variables aleatorias discretas con función de probabilidad conjunta

Y \ X	0	1	2
-1	1/10	1/5	1/10
1	3/20	3/10	3/20

a) **(0.25 pts.)** Calculad la función de probabilidad marginal de X.

Solución.

Y \ X	*0*	*1*	*2*
-1	*1/10*	*1/5*	*1/10*
1	*3/20*	*3/10*	*3/20*
$P(X=x)$	*1/4*	*1/2*	*1/4*

b) **(0.25 pts.)** Calculad la función de probabilidad marginal de Y.

Solución.

Y \ X	*0*	*1*	*2*	$P(Y=y)$
-1	*1/10*	*1/5*	*1/10*	*2/5*
1	*3/20*	*3/10*	*3/20*	*3/5*

c) **(0.5 pts.)** ¿Son X e Y independientes? ¿Por qué?

Solución. *Sí, porque* $P(X=x, Y=y) = P(X=x)\cdot P(Y=y)$, $\forall x$, $\forall y$:

$$P(X=0,Y=-1) = \frac{1}{10} = \frac{1}{4}\cdot\frac{2}{5} = P(X=0)\cdot P(Y=-1)$$
$$P(X=1,Y=-1) = \frac{1}{5} = \frac{1}{2}\cdot\frac{2}{5} = P(X=1)\cdot P(Y=-1)$$
$$P(X=2,Y=-1) = \frac{1}{10} = \frac{1}{4}\cdot\frac{2}{5} = P(X=2)\cdot P(Y=-1)$$
$$P(X=0,Y=1) = \frac{3}{20} = \frac{1}{4}\cdot\frac{3}{5} = P(X=0)\cdot P(Y=1)$$
$$P(X=1,Y=1) = \frac{3}{10} = \frac{1}{2}\cdot\frac{3}{5} = P(X=1)\cdot P(Y=1)$$
$$P(X=2,Y=1) = \frac{3}{20} = \frac{1}{4}\cdot\frac{3}{5} = P(X=2)\cdot P(Y=1)$$

d) **(0.5 pts.)** Calculad la esperanza y la varianza de Y.

Solución.

$$\begin{aligned} E[Y] &= \sum_y y \cdot P(Y = y) = -1 \cdot \frac{2}{5} + 1 \cdot \frac{3}{5} = \frac{1}{5}. \\ Var[Y] &= E[Y^2] - E[Y]^2 = \sum_y y^2 \cdot P(Y = y) - E[Y]^2 = (-1)^2 \cdot \frac{2}{5} + 1^2 \cdot \frac{3}{5} - \left(\frac{1}{5}\right)^2 \\ &= 1 - \frac{1}{25} = \frac{24}{25}. \end{aligned}$$

e) **(0.5 pts.)** Sabiendo que $E[X] = 1$ y $Var[X] = \frac{1}{2}$, calculad la varianza de $X - Y$ justificando todos los pasos.

Solución. Puesto que X e Y son variables aleatorias independientes, *tenemos que* $Var[X - Y] = Var[X] + Var[Y]$, *por tanto:*

$$Var[X - Y] = Var[X] + Var[Y] = \frac{1}{2} + \frac{24}{25} = \frac{25 + 48}{50} = \frac{73}{50}.$$

2. **(3 pts.)** En una fábrica de embalaje de cereales para el desayuno, se empaquetan los cereales en cajas cuyo peso anunciado es de 500 gramos. Sin embargo en la práctica, la cantidad de cereal que se coloca en cada caja es una variable aleatoria con media 500 gramos y desviación típica 25 gramos. Frecuentemente se realizan controles de calidad, durante los cuales se eligen 50 cajas al azar de forma independiente para pesar su contenido.

a) **(0.5 pts.)** ¿Cuál es la distribución del peso medio de las 50 cajas? ¿Por qué?

Solución. *Sea X = peso del cereal contenido en cada caja, la v.a. de interés. Se toma una m.a.s., $X_1, \ldots, X_{50}$, de tamaño 50 y nos interesa la distribución del estadístico* $\overline{X} = \frac{1}{50} \sum_{i=1}^{50} X_i$.
Sabemos que

$$E[\overline{X}] = E[X] = 500 \quad y \quad DT[\overline{X}] = \frac{DT[X]}{\sqrt{n}} = \frac{25}{\sqrt{50}}.$$

Además, al ser el tamaño de la muestra aleatoria simple suficientemente grande ($50 > 30$) *podemos aplicar el teorema central del límite. Por lo tanto la distribución de $\overline{X}$ será aproximadamente normal, es decir*

$$\overline{X} \sim N(500, \frac{25}{\sqrt{50}}) \text{ aproximadamente.}$$

b) **(1.25 pts.)** Durante los controles de calidad, si el peso medio de las 50 cajas es superior a 510 gramos o inferior a 490 gramos se detiene la cadena de embalaje para proceder al equilibrado de las máquinas. ¿Cuál es la probabilidad de que se detenga la cadena durante uno de estos controles?

Solución. *Sabemos que* $\overline{X} \sim N(500, \dfrac{25}{\sqrt{50}})$ *aproximadamente. La probabilidad de que se detenga la cadena de embalaje es*

$$\begin{array}{lcl} P(\overline{X} < 490 \text{ ó } \overline{X} > 510) & = & P(\overline{X} < 490) + P(\overline{X} > 510) \\ & = & P\left(\dfrac{\overline{X}-500}{25/\sqrt{50}} < \dfrac{490-500}{25/\sqrt{50}}\right) + P\left(\dfrac{\overline{X}-500}{25/\sqrt{50}} > \dfrac{510-500}{25/\sqrt{50}}\right) \\ \left[Z = \dfrac{\overline{X}-500}{25/\sqrt{50}} \sim N(0,1)\right] & = & P\left(Z < \dfrac{490-500}{25/\sqrt{50}}\right) + P\left(Z > \dfrac{510-500}{25/\sqrt{50}}\right) \\ & = & P(Z < -2.8284) + P(Z > 2.8284) = 2 \cdot P(Z > 2.8284) \\ & = & 2 \cdot 0.0023 = 0.0046. \end{array}$$

c) **(1.25 pts.)** Si suponemos ahora que el peso de cereal en cada caja sigue una distribución normal (con los parámetros del enunciado), calculad, justificando todos los pasos, la probabilidad de que la varianza muestral en m.a.s. de tamaño 25 sea superior a 735.

Solución. *Si* X = *peso del cereal contenido en cada caja* $\sim N(500, 25)$, *entonces sabemos que*

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$$

(de forma exacta, independientemente de n*) donde* $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$ *es el estadístico varianza muestral.*
En nuestro caso tendríamos

$$\frac{24}{25^2} S^2 \sim \chi^2_{24}.$$

Por tanto

$$\begin{array}{lcl} P(S^2 > 735) & = & P(\dfrac{24}{25^2} S^2 > \dfrac{24}{25^2} 735) \\ \left[W = \dfrac{24}{25^2} S^2 \sim \chi^2_{24}\right] & = & P\left(W > 28.2240\right) \approx 0.25. \end{array}$$

3. **(2.5 pts.)** La temperatura de una cámara de refrigeración es una variable aleatoria X con función de densidad

$$f_X(x) = \frac{1}{b} \quad \text{si} \quad x \in [0, b]$$

donde **b** es un parámetro desconocido. Para estimar la temperatura máxima que se puede alcanzar en la cámara (**b**), se eligen al azar diez momentos distintos durante el día obteniéndose una m.a.s. de temperaturas $X_1, X_2, \ldots, X_{10}$. Definimos los estimadores:

$$T_1(X_1, X_2, \ldots, X_{10}) = 2 \cdot \overline{X}$$

$$T_2(X_1, X_2, \ldots, X_{10}) = \text{máx}\{X_1, X_2, \ldots, X_{10}\}$$

El primer estimador consiste en calcular la media de las diez temperaturas y multiplicarla por dos. El segundo estimador consiste en tomar el máximo de las diez temperaturas observadas.

a) **(1 pto.)** ¿Es T_1 un estimador insesgado de **b**?

Solución. *T_1 será un estimador insesgado de* **b** *si su esperanza es* **b**.

$$\begin{aligned} E[T_1] &= E[2\cdot \overline{X}] = 2\cdot E[\overline{X}] = 2\cdot E[X] \\ &= 2\cdot \int_{-\infty}^{\infty} x\cdot f_X(x)dx = 2\cdot \int_0^b x\frac{1}{b}\cdot dx = 2\cdot \frac{1}{b}\left[\frac{x^2}{2}\right]_0^b \\ &= 2\cdot \frac{1}{b}\frac{b^2}{2} = b. \end{aligned}$$

Por lo tanto T_1 es un estimador insesgado de **b**.

b) **(0.5 pts.)** Sabiendo que $E[T_2] = \frac{10}{11}b$, ¿en base a qué criterio compararíais T_1 y T_2? Justificad vuestra respuesta.

Solución. *Puesto que T_1 es un estimador insesgado de* **b** *y T_2 no, no podemos comparar directamente sus varianzas, sino que tenemos que utilizar el error cuadrático medio:*

$$ECM[T] = Sesgo(T)^2 + Var[T].$$

Será preferible el estimador con menor error cuadrático medio.

c) **(1 pto.)** Sabiendo que $Var[X] = \frac{b^2}{12}$ y $Var[T_2] = \frac{5}{726}b^2$, ¿qué estimador es preferible, en base al criterio establecido en el apartado anterior?

Solución. *Calculamos el E.C.M. de cada estimador, para ello hemos de calcular su sesgo y su varianza.*
T_1 es insesgado, por lo tanto $Sesgo(T_1) = 0$.

$$\begin{aligned} Var[T_1] &= Var[2\cdot \overline{X}] = 4\cdot Var[\overline{X}] = 4\cdot \frac{Var[X]}{10} \\ &= 4\cdot \frac{b^2/12}{10} = \frac{b^2}{30}. \end{aligned}$$

El error cuadrático medio de T_1 es

$$ECM[T_1] = Var[T_1] + Sesgo(T_1)^2 = \frac{b^2}{30} + 0 = \frac{b^2}{30} = 0.0333\, b^2.$$

Calculemos ahora el sesgo de T_2:

$$Sesgo(T_2) = E[T_2] - b = \frac{10}{11}b - b = \frac{-1}{11}b.$$

El error cuadrático medio de T_2 es

$$ECM[T_2] = Var[T_2] + Sesgo(T_2)^2 = \frac{5}{726}b^2 + \left(\frac{-1}{11}b\right)^2 = \frac{5}{726}b^2 + \frac{1}{121}b^2 = 0.0152\, b^2.$$

El estimador preferible es del de menor error cuadrático medio, es decir, T_2.