PROBABILIDAD Y ESTADÍSTICA
EXAMEN FEBRERO 2004

DATOS DEL ESTUDIANTE

No. de Parcial	Apellidos y Nombres	Cédula

- La duración total de la prueba es de **4 horas**.
- El puntaje mínimo necesario para aprobar es **50 puntos**.
- **Publicación de resultados:** Jueves 4 de marzo - 18:00hs.
- **Muestra de exámenes:** Viernes 5 de marzo - 18:00hs.

Problema 1 (30 puntos)

De una caja que contiene 3 bolas rojas y 2 azules se extrae una bola al azar y se la coloca en una segunda caja que contiene 4 bolas azules y 2 rojas. **A continuación** se extrae una bola al azar de la segunda caja.

1. **(10 puntos)** ¿Cuál es la probabilidad de que se extraiga la misma bola que se extrajo de la primera caja?
2. **(10 puntos)** ¿Cuál es la probabilidad de que la bola extraída de la segunda caja sea roja?
3. **(10 puntos)** Si la bola extraída de la segunda caja es roja, ¿cuál es la probabilidad de que sea la misma bola que se extrajo de la primera caja?

Problema 2 (30 puntos)

Se dice que una variable aleatoria X es una variable de *spin* de parámetro h, y se denota $X \sim \mathcal{S}(h)$, si X toma los valores $X = -1$, $X = 1$ con la siguiente distribución de probabilidad:

$$\mathbf{P}(X = x) = \frac{e^{hx}}{e^h + e^{-h}}, \quad x \in \{-1, 1\},$$

donde h es un parámetro real.

1. **(8 puntos)** Calcule $\mathbf{E}(X)$ y $\mathbf{Var}(X)$.
2. **(8 puntos)** Considere una sucesión de variables aleatorias $X_1, X_2, \ldots, X_n$ *i.i.d* $\sim \mathcal{S}(h)$. Halle un estimador consistente para el parámetro h (esto es, un estimador que converja casi seguramente a h cuando $n \to \infty$).

   Sugerencia: Use el valor obtenido para $\mathbf{E}(X_i)$.

3. El spin total de un conjunto de n partículas está definido por:

$$M = \sum_{i=1}^{n} X_i$$

donde $X_1, X_2, \ldots, X_n$ son variables aleatorias *i.i.d* y la variable $X_i \sim \mathcal{S}(h)$ representa el spin de la partícula i para cada $i = 1, 2, \ldots, n$.

Se dice que un sistema de n partículas presenta magnetización cuando $M \geq \frac{2}{3}n$.

   (a) **(7 puntos)** Se supone que el número total de partículas es $n = 9000$ y que $h = \log 2$. Calcule la probabilidad de que el sistema presente magnetización.

   (b) **(7 puntos)** Suponiendo como antes $n = 9000$, calcule el mínimo valor de h para el cual se cumple que la probabilidad de que el sistema presente magnetización es superior a $\frac{1}{2}$.

Sugerencia: Use el Teorema Central del Límite.

Problema 3 (40 puntos)

Los siguientes datos corresponden a los tiempos de vida (medidos en semanas) de colonias de bacterias criadas en un laboratorio bajo ciertas condiciones de temperatura y humedad:

3.31	5.48	1.37	7.44	2.73	4.19	3.18	2.23

Nota: En todos los test use el siguiente criterio de decisión: se acepta la hipótesis nula si el p-valor es superior a 0.10.

1. **(8 puntos)** Estudie la aleatoriedad de la muestra (realice **sólo un** test e indique cuál otro test podría ser implementado).

2. **(16 puntos)** ¿Es razonable suponer que los datos corresponden a una distribución exponencial?

3. **(16 puntos)** Se dispone ahora de una nueva muestra **i.i.d. e independiente de la anterior** (no hay que verificar estos supuestos) correspondiente a los tiempos de vida de colonias de bacterias criadas en condiciones diferentes de temperatura y humedad:

5.54	5.61	5.71	4.22	4.42	2.68	4.28	5.93

   Implemente un test de Kolmogorov-Smirnov de comparación de muestras para concluir si es razonable o no suponer que los nuevos datos tienen la misma distribución que los anteriores.

Soluciones
EXAMEN DE PROBABILIDAD Y ESTADÍSTICA
Febrero 2004

Solución problema 1

Se consideran los sucesos:

$$\begin{aligned} A &= \text{“La bola extraída de la primera caja es roja”} \\ B &= \text{“La bola extraída de la segunda caja es roja”} \\ C &= \text{“La bola extraída de la segunda caja es la misma que de la primera”} \end{aligned}$$

Entonces:

1. Se pide $\mathbf{P}(C) = \frac{1}{7}$ ya que cualquiera de las bolas de la segunda caja tiene igual probabilidad de salir (independientemente de cuál sea la bola agregada), hay $6+1=7$ bolas y la única "favorable" a C es la que se agregó.

2. Se pide $\mathbf{P}(B)$. Para ello hacemos:

$$\mathbf{P}(B) = \mathbf{P}(B \cap A) + \mathbf{P}(B \cap A^c) = \mathbf{P}(B \mid A)\,\mathbf{P}(A) + \mathbf{P}(B \mid A^c)\,\mathbf{P}(A^c)$$

   $\mathbf{P}(A)$ y $\mathbf{P}(A^c)$ dependen de la distribución de bolas de la primera caja y $\mathbf{P}(B \mid A)$ y $\mathbf{P}(B \mid A^c)$ dependen de cómo queda la segunda caja dado que se agregó una bola roja o azul. Los valores son:

$$\begin{aligned} \mathbf{P}(A) &= \frac{3}{5} \qquad \mathbf{P}(A^c) = \frac{2}{5} \\ \mathbf{P}(B \mid A) &= \frac{3}{7} \qquad \mathbf{P}(B \mid A) = \frac{2}{7} \end{aligned}$$

   Y por lo tanto:

$$\mathbf{P}(B) = \frac{3}{7}\frac{3}{5} + \frac{2}{7}\frac{2}{5} = \frac{13}{35}$$

3. Se pide $\mathbf{P}(C \mid B)$. Aplicando la definición de probabilidad condicional:

$$\mathbf{P}(C \mid B) = \frac{\mathbf{P}(B \cap C)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B \mid C)\,\mathbf{P}(C)}{\mathbf{P}(B)}$$

   $\mathbf{P}(B)$ y $\mathbf{P}(C)$ ya fueron calculadas. $\mathbf{P}(B \mid C)$ es la probabilidad de que se extraiga una roja dado que se extrae la misma. Este suceso tiene la misma probabilidad que A, ya que para que sea roja la segunda (dado que es la misma), debe ser roja la primera extraída. Por lo tanto, $\mathbf{P}(B \mid C) = \frac{3}{5}$ (puede hacerse también contando casos). Entonces:

$$\mathbf{P}(C \mid B) = \frac{\frac{3}{5}\frac{1}{7}}{\frac{13}{35}} = \frac{3}{13}$$

Solución problema 2

1. Se tiene que:

$$\mathbf{E}(X) = \mathbf{P}(X=1) - \mathbf{P}(X=-1) = \frac{e^h - e^{-h}}{e^h + e^{-h}} \qquad (= \tanh(h))$$

Además, $X^2 = 1$ con probabilidad 1, $\mathbf{E}(X^2) = 1$. Por lo tanto:

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = 1 - \left(\frac{e^h - e^{-h}}{e^h + e^{-h}}\right)^2 = \frac{4}{(e^h + e^{-h})^2} \qquad \left(= \frac{1}{\cosh^2(h)}\right)$$

2. Por la Ley Fuerte de los Grandes Números, se cumple que:

$$\bar{X}_n \underset{c.s.}{\longrightarrow} \mu = \mathbf{E}(X) = \frac{e^h - e^{-h}}{e^h + e^{-h}} \qquad (= \tanh(h))$$

Invertimos esta función escribiendo h en función de μ. Para ello despejamos h obteniendo:

$$h = g(\mu) = \frac{1}{2}\log\left(\frac{1+\mu}{1-\mu}\right) \qquad (= \operatorname{atanh}(\mu))$$

Como esta función g es continua para $\mu \in (-1,1)$ se tiene que:

$$\frac{1}{2}\log\left(\frac{1+\bar{X}_n}{1-\bar{X}_n}\right) = g(\bar{X}_n) \underset{c.s.}{\longrightarrow} g(\mu) = h$$

por lo que $\frac{1}{2}\log\left(\frac{1+\bar{X}_n}{1-\bar{X}_n}\right) = \operatorname{atanh}(\bar{X}_n)$ es un estimador consistente de h.

3. (a) Como $\frac{2}{3}n = 6000$ se pide $\mathbf{P}(M > 6000)$. Como M es suma de variables aleatorias vale el TCL y podemos escribir:

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} = \frac{\sqrt{n}\left(\frac{M}{n} - \mu\right)}{\sigma} \approx N(0,1)$$

Para $h = \log 2$, $\mu = \mathbf{E}(X_1) = 0.6$ y $\sigma^2 = \mathbf{Var}(X_1) = 0.64$ y entonces $\sigma = 0.8$.

$$\begin{aligned} \mathbf{P}(M > 6000) &= \mathbf{P}\left(\frac{\sqrt{9000}\left(\frac{M}{9000} - \mu\right)}{\sigma} > \frac{\sqrt{9000}\left(\frac{6000}{9000} - \mu\right)}{\sigma}\right) \approx \\ &\approx 1 - \Phi\left(\frac{\sqrt{9000}\,(0.667 - 0.6)}{0.8}\right) = 1 - \Phi(7.9) \approx 0 \end{aligned}$$

(b) Aplicando también TCL se tiene que:

$$\mathbf{P}(M > 6000) \approx 1 - \Phi\left(\frac{\sqrt{9000}\left(\frac{2}{3} - \mu\right)}{\sigma}\right)$$

por lo que la condición resulta que

$$\sqrt{9000}\left(\frac{2}{3} - \mu\right) < 0$$

o bien:

$$\mu = \frac{e^h - e^{-h}}{e^h + e^{-h}} > \frac{2}{3}$$

de donde

$$h > \frac{1}{2}\log\left(\frac{1+\frac{2}{3}}{1-\frac{2}{3}}\right) = \frac{1}{2}\log(5) = 0.805 \qquad \left(= \operatorname{atanh}\left(\frac{2}{3}\right)\right)$$

Solución problema 3

1. Se pueden aplicar el test de Rachas de Subidas y Bajadas y el test de Correlación de Rangos de Spearman. Los resultados son:

   Rachas: $R = 6$, $\alpha^* = 0.325$.

   Spearman: $r_s = -0.310$, $\alpha^* = 0.231$.

   en ambos casos se acepta la hipótesis nula de que los datos son *iid*.

2. Se aplica el Test de exponencialidad de Lilliefors obteniéndose un estadístico $D_n = 0.324$, y un p-valor $0.1 < \alpha^* < 0.15$ por lo que se acepta la hipótesis nula de que los datos son exponenciales.

3. El estadístico del Test de Kolmogorov-Smirnov de comparación de dos muestras es $D_m n = 0.625$ y por lo tanto $mnD = 40$ y $\alpha^* = 0.087$. En este caso se rechaza la hipótesis nula de que los datos provienen de la misma distribución.