PROBABILIDAD Y ESTADÍSTICA
EXAMEN DICIEMBRE 2003

DATOS DEL ESTUDIANTE

Apellidos y Nombres	Cédula

Nombre del docente de Teórico	
Nombre del docente de Práctico	

- Si concurre a más de un teórico o un práctico, incluya los nombres de todos sus docentes.
- Si no recuerda el nombre del docente ponga el horario.
- Si no concurre a clase, responda "NO", pero no deje su casillero en blanco.

- La duración total de la prueba es de **4 horas**.
- **Publicación de resultados:** Lunes 29 de diciembre - 18:00hs.
- **Muestra de exámenes:** Lunes 29 de diciembre - 18:30hs.

¡BUENA SUERTE!

Problema 1 (50 puntos)

Para modelar la cantidad X de productos defectuosos que se encuentran en una línea de producción en determinado período de tiempo se utiliza un modelo de dos parámetros llamado "Poisson con Ceros Forzados" (PCF).

Decimos que la variable X tiene distribución de Poisson con ceros Forzados, y se nota $X \sim PCF(\lambda, p)$, si X puede escribirse como $X = YZ$ con Y, Z independientes tales que:

Y tiene distribución de Poisson de parámetro λ ($Y \sim \mathcal{P}(\lambda)$)

Z tiene distribución de Bernoulli de parámetro p ($Z \sim \mathrm{Ber}(p)$)

El modelo representa que si la producción tiene una muy baja tasa de errores, la probabilidad de que no haya defectuosos es más alta que en un modelo de Poisson tradicional.

1. **(12 puntos)** Hallar $\mathbf{P}(X = k)\ \forall k \geq 0$.
2. **(10 puntos)** Decimos que $X = 0$ es un "cero forzado" cuando ocurre $Z = 0$. Dado que se observa $X = 0$ hallar la probabilidad de que no sea un "cero forzado" (esto es $Z = 1$).
3. **(11 puntos)** Hallar $\mathbf{E}(X)$, $\mathbf{E}\left(X^2\right)$ y $\mathbf{Var}(X)$.

4. **(12 puntos)** Dada una muestra $X_1, \ldots, X_n$ *iid* $\sim PCF(\lambda, p)$, halle estimadores $\hat{\lambda}$ y $\hat{p}$ de los parámetros por el método de los momentos.

5. **(5 puntos)** Para la siguiente lista de 10 observaciones:

0	0	0	2	0	0	4	1	0	0

estime λ y p por el método de los momentos.

Problema 2 (50 puntos)

Se dispone de una máquina para llenar latas con 10 c.c. de aceite (las latas llenadas al borde tiene una capacidad de 11 c.c.). Un operario afirma haberla ajustado de modo que la cantidad de aceite que envía el pico de la máquina envasadora es de $10 + \varepsilon$ c.c., donde ε es una variable aleatoria con la distribución $\mathbf{U}\left[-\frac{1}{10}, \frac{1}{10}\right]$. Para verificar la afirmación anterior se estudió el contenido de 12 latas obteniendo los siguientes valores de ε:

0.098	0.068	−0.004	−0.026	−0.012	0.096
0.05	0.062	0.036	0.04	0	−0.054

Nota: Para cada una de las partes, realice **un solo** test de hipótesis e indique cuál o cuáles otros son aplicables si los hay.

1. **(8 puntos)** Estudiar la aleatoriedad de la muestra.

2. **(15 puntos)** ¿Es razonable suponer que la distribución de la muestra es $\mathbf{U}\left[-\frac{1}{10}, \frac{1}{10}\right]$? Justifique su respuesta.

3. **(12 puntos)** Si la distribución del error fuese $\mathbf{U}\left[-\frac{1}{10}, \frac{1}{10}\right]$ el error "medio" sería 0 c.c. ¿Los datos experimentales confirman la afirmación anterior? Justifique su respuesta.

4. **(15 puntos)** Se dispone de una muestra de 8 latas llenadas antes del ajuste del operario.

−0.016	−0.122	−0.015	−0.104	0.005	0.001	−0.012	−0.105

Se supone que esta muestra es *iid* e independiente de la anterior. Se asume además que esta muestra tiene la misma dispersión que la anterior (no hay que verificar estos supuestos). Evalúe si existe corrimiento entre las muestras.

Soluciones
EXAMEN DE PROBABILIDAD Y ESTADÍSTICA
Diciembre 2003

Solución problema 1

1. Para $k = 0$ se tiene que

$$\begin{aligned}\mathbf{P}(X=0) &= \mathbf{P}(YZ=0) = \mathbf{P}(Z=0) + \mathbf{P}(Y=0, Z=1) = \\ &= \mathbf{P}(Z=0) + \mathbf{P}(Y=0)\,\mathbf{P}(Z=1) = 1 - p + pe^{-\lambda}\end{aligned}$$

   Si $k > 0$ entonces:

$$\mathbf{P}(X=k) = \mathbf{P}(YZ=k) = \mathbf{P}(Y=k, Z=1) = \mathbf{P}(Y=k)\,\mathbf{P}(Z=1) = pe^{-\lambda}\frac{\lambda^k}{k!}$$

2. Se tiene que:

$$\mathbf{P}(Z=1 \mid X=0) = \frac{\mathbf{P}(Z=1, X=0)}{\mathbf{P}(X=0)}$$

   y como $X = YZ$, se cumple que $\{X=0, Z=1\} = \{Y=0, Z=1\}$. Por lo tanto:

$$\mathbf{P}(Z=1 \mid X=0) = \frac{\mathbf{P}(Y=0, Z=1)}{\mathbf{P}(X=0)} = \frac{\mathbf{P}(Y=0)\,\mathbf{P}(Z=1)}{\mathbf{P}(X=0)} = \frac{pe^{-\lambda}}{1-p+pe^{-\lambda}}$$

3. Utilizando que Y y Z son independientes se tiene:

$$\mathbf{E}(X) = \mathbf{E}(YZ) = \mathbf{E}(Y)\,\mathbf{E}(Z) = \lambda p$$

   A su vez:

$$\mathbf{E}\left(X^2\right) = \mathbf{E}\left(Y^2Z^2\right) = \mathbf{E}\left(Y^2\right)\mathbf{E}\left(Z^2\right)$$

   Como $Y \sim \mathcal{P}(\lambda)$, $\mathbf{Var}(Y) = \lambda$ y por lo tanto $\mathbf{E}\left(Y^2\right) = \mathbf{Var}(Y) + [\mathbf{E}(Y)]^2 = \lambda + \lambda^2$. Además, $Z^2 = Z$ por lo que $\mathbf{E}\left(Z^2\right) = p$. Por lo tanto:

$$\mathbf{E}\left(X^2\right) = p\left(\lambda + \lambda^2\right)$$

   Utilizando las dos anteriores se tiene que:

$$\mathbf{Var}(Z) = \mathbf{E}\left(X^2\right) - [\mathbf{E}X]^2 = \lambda p + \lambda^2 p - \lambda^2 p^2$$

4. Se tiene el siguiente sistema:

$$\begin{cases}\mathbf{E}(X) = \mu_1 = \lambda p \\ \mathbf{E}\left(X^2\right) = \mu_2 = \lambda p + \lambda^2 p\end{cases}$$

   Por lo tanto, restando, $\mu_2 - \mu_1 = \lambda^2 p$. Dividiendo esta ecuación entre la primera se tiene que:

$$\lambda = \frac{\mu_2 - \mu_1}{\mu_1} \qquad \text{y} \qquad p = \frac{\mu_1^2}{\mu_2 - \mu_1}$$

   Y por lo tanto, sustituyendo μ_1 y μ_2 por sus estimadores:

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

   se llega a los estimadores deseados.

5. Para los valores allí presentados y con la notación anterior se tiene que:

$$\hat{\mu}_1 = \frac{7}{10} = 0.7 \quad \hat{\mu}_2 = \frac{21}{10} = 2.1$$

por lo que los estimadores quedan:

$$\hat{\lambda} = 2 \quad \hat{p} = 0.35$$

Solución problema 2

Nota: En todos los casos se detallan los resultados de todos los tests aplicables. En cada parte bastaba con hacer uno solo de los test.

1. **Test de rachas de subidas y bajadas.**

   Estadístico: $\mathbf{R} = 7$. Como $\mathbf{R} = 7 < \mathbf{E}(R) = \frac{23}{3}$ planteamos

$$\begin{cases} \mathbf{H}_0 : & X_1, ..., X_n \text{ son } iid \\ \mathbf{H}_1 : & \text{“hay pocas rachas”} \end{cases}$$

   Como $\alpha^* = 0.4453$, no se rechaza $\mathbf{H}_0$.

   Test de rangos de Spearman.

   Estadístico: $\mathbf{R}_s = -0.385$. Como el estadístico de correlación de Spearman dio negativo planteamos:

$$\begin{cases} \mathbf{H}_0 : & X_1, ..., X_n \text{ son } iid \\ \mathbf{H}_1 : & \text{“hay tendencia decreciente”} \end{cases}$$

   resultando de la tabla que: $\alpha^* > 0.1$ y por lo tanto no se rechaza $\mathbf{H}_0$.

2. Como la distribución está perfectamente determinada (es uniforme con parámetros conocidos) podemos aplicar el **Test de Kolmogorov- Smirnov** (no podemos aplicar un Test χ^2 pues es para n “grande”). Planteamos:

$$\begin{cases} \mathbf{H}_0 : & F = \mathbf{U}\left[-\frac{1}{10}, \frac{1}{10}\right] \\ \mathbf{H}_1 : & F \neq \mathbf{U}\left[-\frac{1}{10}, \frac{1}{10}\right] \end{cases}$$

   Estadístico: $D_n = 0.287$ y por lo tanto $\alpha^* > 0.2$. No se rechaza $\mathbf{H}_0$.

3. No podemos aplicar tests paramétricos para la media pues tenemos pocos datos (no es aplicable el Teorema Central del Límite). Pero si podemos aplicar tests no paramétricos para la mediana (pues si la distribución es uniforme coinciden).

   Test de signos.

   Estadístico: $\mathbf{U} = 8$. Como $\mathbf{U} > \mathbf{E}(\mathbf{U}) = \frac{n}{2} = 6$, planteamos:

$$\begin{cases} \mathbf{H}_0 & m = 0 \\ \mathbf{H}_1 & m > 0 \end{cases}$$

   Siendo $\alpha^* = 0.1938$, no se rechaza $\mathbf{H}_0$.

Test rangos signados de Wilcoxon.

Estadístico $\mathbf{W}_n^+ = 61$, como $\mathbf{W}^+ > \mathbf{E}\left(\mathbf{W}^+\right) = \frac{n(n+1)}{4} = 39$, planteamos:

$$\begin{cases} \mathbf{H}_0 & m = 0 \\ \mathbf{H}_1 & m > 0 \end{cases}$$

Resultando $\alpha^* = 0.046$, por lo que se rechaza $\mathbf{H}_0$

4. Aplicamos el **Test de Mann -Whitney-Wilcoxon** para alternativa de corrimiento.

   Llamando X a la segunda muestra de $m = 8$ datos e Y a la primera muestra de $n = 12$ datos el estadístico queda $\mathbf{T}_X = 52.5$ (hay un empate que se resuelve promediando). Como $\mathbf{T}_X < \mathbf{E}\left(\mathbf{T}_X\right) = \frac{m(n+m+1)}{2} = 84$ planteamos:

$$\begin{cases} \mathbf{H}_0 & X \sim Y \\ \mathbf{H}_1 & X \sim Y + \theta, \text{ con } \theta < 0 \end{cases}$$

   Como $n > 10$ usamos la aproximación normal resultando

$$z_L = \frac{T_X + 0.5 - \frac{m(m+n+1)}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} = -2.39$$

   por lo que el p-valor queda $\alpha^* = \Phi(-2.39) = 0.008 < 0.1$ por lo que se rechaza $\mathbf{H}_0$.