Ingeniería Matemática
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE
MA3403 Probabilidades y Estadística
Primavera 2013

Roberto Cortez
Mauricio Duarte
Joaquín Fontbona

Pauta Examen

20 de diciembre de 2013
Tiempo: 3 horas

P1. (a) (3,0 ptos.) Sea M_n el número de caras obtenidas por Michelle luego de lanzar n monedas y S_n el número de caras obtenidas por Sebastián luego de lanzar n monedas. Usando probabilidades totales

$$\mathbb{P}(M_{n+1} > S_n) = \mathbb{P}(M_{n+1} > S_n \mid M_n > S_n)\mathbb{P}(M_n > S_n) + \mathbb{P}(M_{n+1} > S_n \mid M_n = S_n)\mathbb{P}(M_n = S_n),$$

puesto que $\mathbb{P}(M_{n+1} > S_n \mid M_n < S_n) = 0$.
Es claro que $\mathbb{P}(M_{n+1} > S_n \mid M_n = S_n) = \frac{1}{2}$, y además $\mathbb{P}(M_{n+1} > S_n \mid M_n > S_n) = 1$. Luego,

$$\mathbb{P}(M_{n+1} > S_n) = \mathbb{P}(M_n > S_n) + \frac{1}{2}\mathbb{P}(M_n = S_n).$$

Pero $\mathbb{P}(M_n = S_n) = 1 - \mathbb{P}(M_n > S_n) - \mathbb{P}(M_n < S_n)$, lo que por simetría además es igual a $1 - 2\mathbb{P}(M_n > S_n)$. Finalmente,

$$\mathbb{P}(M_{n+1} > S_n) = \mathbb{P}(M_n > S_n) + \frac{1}{2}\left(1 - 2\mathbb{P}(M_n > S_n)\right) = \frac{1}{2}.$$

(b) 1) (1,5 ptos.) Si T_A y T_B son los tiempos que tardan en pasar los buses de las líneas A y B respectivamente, es claro que $T = \text{mín}(T_A, T_B)$. Calculemos la distribución acumulada de T, utilizando el hecho que $T > x$ si y sólo si $T_A > x$ y $T_B > x$:

$$F_(T)(x) = \mathbb{P}(T \leq x) = 1 - \mathbb{P}(T > x) = 1 - \mathbb{P}(T_A > x)\mathbb{P}(T_B > x),$$

donde en el último paso hemos utilizado la independencia entre T_A y T_B. Como T_A es variable exponencial de parámetro λ_A, es directo que $\mathbb{P}(T_A > x) = e{-}\lambda_A x$, ídem para T_B. Luego:

$$F_T(x) = 1 - e^{-\lambda_A x}e^{-\lambda_B x} = 1 - e^{-(\lambda_A+\lambda_B)x},$$

y derivando obtenemos $f_T(x) = (\lambda_A + \lambda_B)e^{-(\lambda_A+\lambda_B)x}$, lo cual vale para $x \geq 0$, mientras que para $x \leq 0$ es claro que $f_T(x) = 0$ (pues T sólo toma valores positivos). Es decir, T es una variable exponencial de parámetro $\lambda_A + \lambda_B = 1/10 + 1/20 = 3/20$.

2) (1,5 ptos.) Fijemos las 8:50 como hora 0, con lo cual la hora J en que llegua el jefe es una variable uniforme en el intervalo $[0, 20]$, y la hora a la que usted llega es T. Deseamos calcular la probabilidad de que usted llegue tarde y que el jefe llegue antes que usted, es decir, $T > 10$ y $J < T$. Condicionando en los posibles resultados de J y utilizando la independencia entre T y J, tenemos:

$$\begin{aligned}\mathbb{P}(T > 10, J < T) &= \int_0^{20} \mathbb{P}(T > 10, J < T \mid J = t) f_J(t)dt \\ &= \int_0^{20} \mathbb{P}(T > 10, t < T)\frac{dt}{20}.\end{aligned}$$

Para $t > 20$, la probabilidad dentro de la integral corresponde a $\mathbb{P}(T > t) = e^{-\lambda t}$, mientras que si $t \leq 10$ dicha probabilidad es $\mathbb{P}(T > 10) = e^{-\lambda\times 10}$, donde $\lambda = 3/20$ es el parámetro de T. Con esto:

$$\begin{aligned}\mathbb{P}(T > 10, J > T) &= \int_0^{10} e^{-10\lambda}\frac{dt}{20} + \int_{10}^{20} e^{-\lambda t}\frac{dt}{20} \\ &= \frac{e^{-10\lambda}}{2} + \frac{e^{-10\lambda} - e^{-20\lambda}}{20\lambda} \\ &= \frac{e^{-3/2}}{2} + \frac{e^{-3/2} - e^{-3}}{3}.\end{aligned}$$

P2. (a) 1) (1,0 ptos.) Note que X is a variable aleatoria no negativa. Entonces, para $z > 0$,

$$\mathbb{P}(X^k > z) = \mathbb{P}(X > \sqrt[k]{z}) = \int_{\sqrt[k]{z}}^{\infty} Cx^{k-1}e^{-(x/\lambda)^k}dx.$$

Usando el cambio de variable $y = x^k$, tenemos $dy = kx^{k-1}dx$, luego

$$\mathbb{P}(X^k > z) = \int_z^{\infty} Cke^{-y/\lambda^k}dy,$$

lo que muestra que la densidad de X^k es una constante por e^{-x/λ^k}, es decir, $X^k \sim \exp(1/\lambda^k)$.

2) (2,0 ptos.) Para $x_1, \ldots, x_n \geq 0$, la función de verosimilitud está dada por

$$L = L(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^n \frac{k}{\lambda^k} x_i^{k-1} e^{-(x_i/\lambda)^k} = \frac{k^n}{\lambda^{nk}} e^{-\frac{1}{\lambda^k}\sum_{i=1}^n x_i^k} \left[\prod_{i=1}^n x_i\right]^{k-1},$$

luego

$$\log L = n \log k - nk \log \lambda - \frac{1}{\lambda^k}\sum_{i=1}^n x_i^k + (k-1)\sum_{i=1}^n \log x_i.$$

Para maximizar lo anterior con respecto a λ, derivamos con respecto a λ e igualamos a 0:

$$0 = \frac{\partial \log L}{\partial \lambda} = -\frac{nk}{\lambda} + \frac{k}{\lambda^{k+1}}\sum_{i=1}^n x_i^k.$$

Despejando λ y reemplazando los $x_1, \ldots, x_n$ por la muestra $X_1, \ldots, X_n$, obtenemos lo deseado:

$$\hat{\lambda} = \left(\frac{1}{n}\sum_{i=1}^n X_i^k\right)^{1/k}.$$

Por lo realizado previamente, sabemos que las variables $Y_i = X_i^k$ son exponenciales de parámetro $1/\lambda^k$, y son independientes pues los X_i lo son. Luego, por la ley fuerte de los grandes números, se tiene que

$$\frac{1}{n}\sum_{i=1}^n X_i^k = \frac{1}{n}\sum_{i=1}^n Y_i = \overline{Y} \to \mathbb{E}(Y_1),$$

casi seguramente cuando $n \to \infty$. Pero sabemos que la esperanza de una variable exponencial es el recíproco del parámetro, luego $\mathbb{E}(Y_1) = \lambda^k$. Por lo tanto, se obtiene la convergencia buscada:

$$\hat{\lambda} = \left(\frac{1}{n}\sum_{i=1}^n X_i^k\right)^{1/k} \to (\lambda^k)^{1/k} = \lambda.$$

(b) (3,0 ptos.) Calculando directamente, usando la independencia y el hecho de que todas las variables tienen la misma distribución,

$$\begin{aligned}\mathbb{P}(Y < y) &= \mathbb{P}(\text{máx}(X_1, \ldots, X_n) < y) \\ &= \mathbb{P}(X_1 < y, X_2 < y, \ldots, X_n < y) \\ &= \mathbb{P}(X_1 < y)^n = y^n,\end{aligned}$$

para cada $y \in [0, 1]$. Derivando, obtenemos que

$$f_Y(y) = ny^{n-1}\mathbf{1}_{[0,1]}(y).$$

P3. (a) 1) (1,0 pto.) Utilizamos el estadístico

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}},$$

el cual sabemos que tiene distribución t-student con $n-1 = 24$ grados de libertad. Para encontrar el intervalo buscado, imponemos que T esté en un intervalo simétrico:

$$95\,\% = \mathbb{P}(T \in [-c, c]) = 1 - \mathbb{P}(|T| > c) = 1 - 2\mathbb{P}(T > c),$$

es decir, $\mathbb{P}(T > c) = 2{,}5\,\%$. De una tabla de una t-student, obtenemos $c = 2{,}064$. Despejando μ en la inclusión $T \in [-c, c]$, tenemos:

$$\begin{aligned} -c \leq \frac{\bar{X} - \mu}{s/\sqrt{n}} \leq c \\ \Leftrightarrow \qquad \bar{X} - \frac{cs}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{cs}{\sqrt{n}} \end{aligned}$$

Notando que $s^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2 = \frac{2400}{24} = 100$, se tiene que $cs/\sqrt{n} = 2{,}064 \times 10/\sqrt{25} = 4{,}128$. Con esto, el intervalo buscado es:

$$\left[\bar{X} - \frac{cs}{\sqrt{n}}, \bar{X} + \frac{cs}{\sqrt{n}}\right] = [65 - 4{,}128, \quad 65 + 4{,}128] = [60{,}872, \quad 69{,}128]$$

2) (1,0 pto.) Utilizamos el estadístico

$$V = (n-1)\frac{s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

el cual sabemos que posee distribución chi-cuadrado con $n-1 = 24$ grados de libertad. Imponemos $95\,\% = \mathbb{P}(V \in [a, b])$, donde a y b satisfacen simetría de probabilidad, es decir $\mathbb{P}(V < a) = \mathbb{P}(V > b)$. Por lo tanto $\mathbb{P}(V \geq a) = 97{,}5\,\%$ y $\mathbb{P}(V > b) = 2{,}5\,\%$, y de la tabla se obtiene $a = 12{,}4$ y $b = 39{,}4$. Despejando σ^2 en la inclusión $V \in [a, b]$, tenemos:

$$\begin{aligned} a \leq (n-1)\frac{s^2}{\sigma^2} \leq b \\ \Leftrightarrow \quad \frac{(n-1)s^2}{b} \leq \sigma^2 \leq \frac{(n-1)s^2}{a}. \end{aligned}$$

Notando que $(n-1)s^2 = \sum(X_i - \bar{X})^2 = 2400$, el intervalo de confianza para σ^2 es

$$\left[\frac{2400}{39{,}4}, \quad \frac{2400}{12{,}4}\right] \approx [60, \quad 200].$$

3) (1,0 pto.) El largo del intervalo obtenido para σ^2 es la diferencia de sus extremos, es decir, $(n-1)s^2[\frac{1}{a} - \frac{1}{b}]$. Luego, el valor esperado del largo es

$$\mathbb{E}\left((n-1)s^2\left[\frac{1}{a} - \frac{1}{b}\right]\right) = (n-1)\left[\frac{1}{a} - \frac{1}{b}\right]\mathbb{E}(s^2) = 24 \times \left[\frac{1}{12{,}4} - \frac{1}{39{,}4}\right]\sigma^2,$$

donde en el último paso hemos utilizado el hecho que s^2 es un estimador insesgado de σ^2.

(b) (3,0 ptos.) Las probabilidades estimadas en base a la muestra son $\hat{p}_j = n_j/n$, donde n_1, n_2 y n_3 son la cantidad de clientes que prefieren el envase A, B y C, respectivamente. Las probabilidades de la hipótesis nula son equiprobables entre las $k = 3$ posibilidades, es decir, $p_j^0 = 1/k = 1/3$, para $j = 1, 2, 3$. Con esto, calculemos el valor observado en la muestra del estadístico Δ:

$$\begin{aligned} \Delta_{\text{obs}} &= n\sum_{j=1}^{k}\frac{(\hat{p}_j - p_j^0)^2}{p_j^0} \\ &= n\sum_{j=1}^{k}\frac{\left(\frac{n_j}{n} - \frac{n}{nk}\right)^2}{\frac{1}{k}} \\ &= \frac{k}{n}\sum_{j=1}^{k}(n_j - n/k)^2 \\ &= \frac{1}{60}[(50-60)^2 + (73-60)^2 + (57-60)^2], \end{aligned}$$

es decir, $\Delta_{\text{obs}} = [100 + 169 + 9]/60 = 278/60 \approx 4{,}633$. El p-valor corresponde a la probabilidad, bajo H_0, de obtener algo al menos tan extremo como lo observado en la muestra. Sabemos que bajo H_0, el estadístico Δ tiene distribución aproximada de una χ^2_{k-1} (con $k - 1 = 2$), con lo cual obtenemos:

$$p\text{-valor} = \mathbb{P}(\Delta \geq \Delta_{\text{obs}} \mid H_0) \approx \mathbb{P}(\chi^2_2 \geq 4{,}633) \approx 0{,}1,$$

donde en el último paso hemos utilizado la tabla de la distribución chi-cuadrado. Notemos que $\alpha = 5\,\% < 10\,\% = p$-valor, por lo tanto no corresponde rechazar la hipótesis nula. Es decir, no hay suficiente evidencia para afirmar que los clientes poseen preferencia por alguno de los envases.