Ingeniería Matemática
FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE
MA3403-3 Probabilidades y Estadística, Otoño 2012

Roberto Cortez
Francisco Castro
Alfredo Torrico

Pauta Control 3

P1. (a) Calculemos primero el estimador de máxima verosimilitud. La función de verosimilitud corresponde a:

$$L(x_1,\dots,x_n)=\prod_{i=1}^n f(x_i)=\prod_{i=1}^n r x_i^{r-1}\mathbf{1}_{[0,1]}(x_i)=r^n\left[\prod_{i=1}^n x_i\right]^{r-1}\left[\prod_{i=1}^n \mathbf{1}_{[0,1]}(x_i)\right].$$

Para maximizar lo anterior con respecto a r, basta trabajar con la parte que depende de r, por lo cual ignoramos el producto de las indicatrices. Tomamos $\ln(\cdot)$ al resto, derivamos e igualamos a 0:

$$0=\frac{d}{dr}\left(n\ln(r)+(r-1)\sum_{i=1}^n \ln(x_i)\right)=\frac{n}{r}+\sum_{i=1}^n \ln(x_i).$$

Despejando r, se obtiene entonces $\hat{r}_{\mathrm{MV}}=-n/\sum_{i=1}^n \ln(X_i)$. Veamos ahora el estimador del método de los momentos, para lo cual calculamos la esperanza de la variable en consideración:

$$\mathbb{E}(X_i)=\int_{-\infty}^{\infty} x f(x)dx=\int_0^1 x r x^{r-1}dx=r\left.\frac{x^{r+1}}{r+1}\right|_0^1=\frac{r}{r+1}=\bar{X},$$

donde la última igualdad la hemos impuesto, para aplicar el método de los momentos. Despejando r, obtenemos

$$\hat{r}_{\mathrm{mom}}=\frac{\bar{X}}{1-\bar{X}}.$$

(b) Sean $X_1,\dots,X_n$ los resultados de las duraciones de las 16 baterías probadas.

1) Trabajamos con el estadístico

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

el cual sabemos que se distribuye como una t_{n-1}, es decir, como una t_{15}. Imponemos que T esté en un intervalo simétrico con probabilidad $1-\alpha$ con $\alpha=0{,}05$:

$$1-\alpha=\mathbb{P}(T\in[-c,c])=1-2\mathbb{P}(T>c),\quad \text{es decir}\quad \mathbb{P}(T>c)=\alpha/2=0{,}025.$$

Mirando la tabla de la distribución t-student, obtenemos $c=2{,}131$. Despejando μ en la inclusión $T\in[-c,c]$, se tiene que

$$\mu\in\left[\bar{X}-c\frac{s}{\sqrt{n}},\bar{X}+c\frac{s}{\sqrt{n}}\right]=\left[7-2{,}131\frac{\sqrt{0{,}9}}{4},7+2{,}131\frac{\sqrt{0{,}9}}{4}\right].$$

2) Trabajamos con el estadístico

$$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}},$$

el cual sabemos que se distribuye como una normal estándar. Imponemos:

$$1-\alpha=\mathbb{P}(Z\in[-c,c])=1-2\mathbb{P}(Z>c),\quad \text{es decir}\quad \mathbb{P}(Z>c)=\alpha/2=0{,}025.$$

Mirando la tabla de una normal estándar obtenemos $c = 1{,}96$. Por lo tanto, despejando μ en la inclusión $Z \in [-c, c]$, se obtiene

$$\mu \in \left[\bar{X} - c\frac{\sigma}{\sqrt{n}}, \bar{X} + c\frac{\sigma}{\sqrt{n}}\right] = \left[7 - \frac{1{,}96}{4}, 7 + \frac{1{,}96}{4}\right].$$

3) El nivel de confianza es el mismo si mantenemos el mismo $c = 1{,}96$. Por otro lado, el largo del intervalo anterior es la diferencia de los extremos, es decir

$$2\frac{1{,}96}{4} = \frac{1{,}96}{2}.$$

Queremos encontrar un n^* tal que que este intervalo se reduzca un 20 %, para lo cual imponemos entonces que el largo del nuevo intervalo sea un 80 % veces la cantidad anterior, es decir:

$$2\frac{1{,}96}{\sqrt{n^*}} = 0{,}8\frac{1{,}96}{2}, \quad \text{es decir} \quad \sqrt{n^*} = 5.$$

Por lo tanto $n^* = 25$, lo que significa que deben probarse 9 baterías adicionales para reducir el largo del intervalo un 20 %.

P2. (a) Por el lema de Neyman-Pearson, sabemos que el test más potente tiene región de rechazo de la forma

$$R = \left\{\vec{x} \in \mathbb{R}^n : \frac{L(\vec{x}; \lambda_0)}{L(\vec{x}; \lambda_1)} \leq \eta\right\},$$

donde η es una constante, y $L(\vec{x}; \lambda)$ es la verosimilitud de la muestra. Se tiene que:

$$L(\vec{x}; \lambda) = \prod_{i=1}^{n} f(x_i; \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}.$$

Por lo tanto, la desigualdad que define la región R corresponde a:

$$\eta \geq \frac{\lambda_0^n e^{-\lambda_0 \sum_{i=1}^{n} x_i}}{\lambda_1^n e^{-\lambda_1 \sum_{i=1}^{n} x_i}} = \frac{\lambda_0^n}{\lambda_1^n} e^{n(\lambda_1 - \lambda_0)\bar{x}},$$

lo cual equivale a

$$\text{Cte} = \frac{\log(\eta) + n\log(\lambda_1/\lambda_0)}{n(\lambda_1 - \lambda_0)} \geq \bar{x},$$

como deseábamos. Notemos que mientras $\lambda_1 > \lambda_0$, no cambia el sentido de la desigualdad al dividir por $(\lambda_1 - \lambda_0)$ en el último paso. Esto significa que para cualquier $\lambda_1 > \lambda_0$, la forma de la región de rechazo es siempre la misma. Por lo tanto, el test sí es uniformemente más potente.

(b) Restando $1/\lambda_0$ y dividiendo por $1/(\lambda_0\sqrt{n})$ en la desigualdad $\bar{x} \leq \text{Cte}$, obtenemos

$$\frac{\bar{x} - 1/\lambda_0}{1/(\lambda_0\sqrt{n})} \leq \frac{\text{Cte} - 1/\lambda_0}{1/(\lambda_0\sqrt{n})} = c,$$

como requeríamos. Ahora, notemos que la media μ y varianza σ^2 de la variable $\exp(\lambda)$ en estudio corresponden a $1/\lambda$ y $1/\lambda^2$, respectivamente. Por lo tanto, al imponer que el error de tipo I sea el α especificado, tenemos que:

$$5\,\% = \mathbb{P}(\vec{X} \in R \mid H_0) = \mathbb{P}\left(\frac{\bar{X} - 1/\lambda_0}{1/(\lambda_0\sqrt{n})} \leq c \;\middle|\; \lambda = \lambda_0\right) = \mathbb{P}\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq c\right) \approx \mathbb{P}(\mathcal{N}(0,1) \leq c).$$

donde en último paso hemos utilizado el TCL para aproximar la variable por una normal estándar. Por simetría de la normal, la probabilidad anterior es igual a $\mathbb{P}(\mathcal{N}(0,1) \geq -c)$, y de la tabla normal, obtenemos $c = -1{,}65$, como queríamos probar.

(c) Sabemos que la potencia del test corresponde a la probabilidad, dado H_1, de rechazar H_0. Es decir, la potencia corresponde a

$$\begin{aligned}
\mathbb{P}(\vec{X} \in R \mid H_1) &= \mathbb{P}\left(\frac{\bar{X} - 1/\lambda_0}{1/(\lambda_0\sqrt{n})} \leq c \,\middle|\, \lambda = \lambda_1 \right) \\
&= \mathbb{P}(\bar{X} \leq c/(\lambda_0\sqrt{n}) + 1/\lambda_0 \mid \lambda = \lambda_1) \\
&= \mathbb{P}\left(\frac{\bar{X} - 1/\lambda_1}{1/(\lambda_1\sqrt{n})} \leq \frac{c/(\lambda_0\sqrt{n}) + 1/\lambda_0 - 1/\lambda_1}{1/(\lambda_1\sqrt{n})} \,\middle|\, \lambda = \lambda_1 \right) \\
&= \mathbb{P}\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{-1{,}65/\sqrt{25} + 1 - 1/2}{1/(2\sqrt{25})} \right) \\
&\approx \mathbb{P}(\mathcal{N}(0,1) \leq 1{,}7) \\
&= 1 - \mathbb{P}(\mathcal{N}(0,1) \geq 1{,}7),
\end{aligned}$$

donde nuevamente hemos utilizado el TCL para aproximar la variable a la izquierda de la desigualdad por una normal estándar. Mirando la tabla normal, la probabilidad de la última línea corresponde a 4,46 %. Es decir, la potencia es de un 95,54 %.

(d) El p-valor corresponde a la probabilidad, dado H_0, de obtener un valor al menos tan extremo como el obtenido en la muestra. Como la forma de la región de rechazo es $\bar{X} \leq \textsc{Cte}$, esto quiere decir:

$$\begin{aligned}
p\text{-valor} &= \mathbb{P}(\bar{X} \leq \bar{X}_{\text{obs}} \mid H_0) \\
&= \mathbb{P}\left(\frac{\bar{X} - 1/\lambda_0}{1/(\lambda_0\sqrt{n})} \leq \frac{0{,}6 - 1/\lambda_0}{1/(\lambda_0\sqrt{n})} \,\middle|\, \lambda = \lambda_0 \right) \\
&= \mathbb{P}\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{0{,}6 - 1}{1/\sqrt{25}} \right) \\
&\approx \mathbb{P}(\mathcal{N}(0,1) \leq -2),
\end{aligned}$$

donde hemos utilizado el TCL igual que antes. Por simetría, la última probabilidad es igual a $\mathbb{P}(\mathcal{N}(0,1) \geq 2)$, y mirando una tabla normal obtenemos que el p-valor corresponde a 2,28 %. Como este valor es menor que $\alpha = 5\,\%$, corresponde rechazar H_0.

P3. (a) Sean $X_1, \ldots, X_n$ las duraciones de las ampolletas, con $n = 100$, de modo que $Y = \sum_{i=1}^n X_i$. Sabemos que la esperanza de cada X_i corresponde a $\mu = 5$, de manera que el valor esperado de Y es

$$\mathbb{E}(Y) = \mathbb{E}\left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \mathbb{E}(X_i) = n\mu = 500.$$

El evento en que después de 525 horas aún hay una ampolleta funcionando corresponde a que $Y \geq 525$. Utilizando la desigualdad de Markov, obtenemos la cota buscada:

$$\mathbb{P}(Y \geq 525) \leq \frac{\mathbb{E}(Y)}{525} = \frac{500}{525} = \frac{20}{21}.$$

(b) Sabemos que la varianza de cada X_i es $\sigma^2 = 25$. Además, como los X_i son independientes, sus covarianzas son 0, lo que significa que la varianza de su suma es igual a la suma de sus varianzas. Por lo tanto,

$$\operatorname{var}(Y) = \operatorname{var}\left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \operatorname{var}(X_i) = n\sigma^2 = 100 \times 25,$$

luego, la raíz de la varianza de Y es $10 \times 5 = 50$. El evento en que se acaban las ampolletas entre las horas 475 y 525 corresponde a $400 < Y < 600$, o equivalentemente, $|Y-500| < 100$. Utilizando la desigualdad de Chebyshev, obtenemos:

$$\mathbb{P}(|Y-500| < 100) = 1 - \mathbb{P}(|Y - \mathbb{E}(Y)| \geq 100) \geq 1 - \frac{\mathrm{var}(Y)}{100^2} = 1 - \frac{50^2}{100^2} = 75\,\%.$$

(c) Usando el TCL, tenemos:

$$\begin{aligned}
\mathbb{P}(Y \geq 525) &= \mathbb{P}(n\bar{X} \geq 525) \\
&= \mathbb{P}\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq \frac{\frac{525}{n}-\mu}{\sigma/\sqrt{n}}\right) \\
&\approx \mathbb{P}\left(\mathcal{N}(0,1) \geq \frac{\frac{525}{100}-5}{5/10}\right) \\
&= \mathbb{P}(\mathcal{N}(0,1) \geq 0{,}5),
\end{aligned}$$

y mirando una tabla normal, obtenemos que la probabilidad anterior es de 30,85 %.

(d) Sean $Y_1, \ldots, Y_m$ las duraciones de las ampolletas adicionales, con $m = 50$. Sabemos que cada Y_i tiene media $\nu = 3$ y varianza $\tau^2 = 22$. El tiempo total que están prendidas todas las ampolletas es $Z = \sum_{i=1}^n X_i + \sum_{i=1}^m Y_i$. Utilizando el TCL, notemos que

$$\sum_{i=1}^n X_i = n\left(\frac{\sigma}{\sqrt{n}}\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right] + \mu\right) \approx n\left(\frac{\sigma}{\sqrt{n}}\mathcal{N}(0,1) + \mu\right) = \mathcal{N}(n\mu, n\sigma^2),$$

es decir, la suma de los X_i es una variable aproximadamente normal con media $n\mu$ y varianza $n\sigma^2$. Análogamente, la suma de los Y_i es aproximadamente normal con media $m\nu$ y varianza $m\tau^2$. Como los X_i y los Y_i son independientes, estas normales son independientes entre sí. Como al sumar normales independientes los parámetros se suman, tenemos entonces que Z tiene distribución aproximadamente normal con media $n\mu + m\nu$ y varianza $n\sigma^2 + m\tau^2$. Luego:

$$\begin{aligned}
\mathbb{P}(Z \geq 700) &\approx \mathbb{P}(\mathcal{N}(n\mu + m\nu, n\sigma^2 + m\tau^2) \geq 700) \\
&= \mathbb{P}\left(\mathcal{N}(0,1) \geq \frac{700 - n\mu - m\nu}{\sqrt{n\sigma^2 + m\tau^2}}\right) \\
&= \mathbb{P}\left(\mathcal{N}(0,1) \geq \frac{700 - 100 \times 5 - 50 \times 3}{\sqrt{100 \times 25 + 50 \times 22}}\right) \\
&= \mathbb{P}(\mathcal{N}(0,1) \geq 0{,}833).
\end{aligned}$$

Mirando una tabla normal, obtenemos que la probabilidad buscada es aproximadamente igual a 20,33 %.