Ingeniería Matemática
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE
MA3403-4 Probabilidades y Estadística
Otoño 2016

Roberto Cortez
Cristóbal Parraguez
Bryan Sagredo

Pauta Control 3

P1. *a)* Sea $X_1, \dots, X_n$ una m.a.s. con distribución común Poisson(λ).

1) (2,0 ptos.) Encuentre $\hat{\lambda}$, el estimador de máxima verosimilitud de λ.

Solución. Recordemos que la función de distribución discreta de la variable Poisson es $p(k;\lambda) = e^{-\lambda}\lambda^k/k!$. Luego, la función de verosimilitud es

$$L = L(x_1, \dots, x_n; \lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\frac{\lambda^{n\bar{x}}}{\prod_{i=1}^{n} x_i!},$$

y entonces $\log L = -n\lambda + n\bar{x}\log\lambda - \log\prod_i x_i!$. Debemos maximizar con respecto a λ, para lo cual derivamos e igualamos a 0:

$$0 = \frac{\partial \log L}{\partial\lambda} = -n + \frac{n\bar{x}}{\lambda} \qquad \Rightarrow \qquad \lambda = \bar{x}.$$

Reemplazando los x_i por la muestra, obtenemos que $\hat{\lambda} = \bar{X}$.

2) (0,5 ptos.) ¿Es $\hat{\lambda}$ insesgado? Muestre que es consistente.

Solución. Sabemos que $\bar{X}$ es estimador insesgado de $\mu = \mathbb{E}(X_1)$. Pero como $X_1 \sim$ Poisson(λ), se tiene que $\mu = \lambda$. Luego, $\mathbb{E}(\hat{\lambda}) = \mathbb{E}(\bar{X}) = \lambda$, es decir, $\hat{\lambda}$ es insesgado. Además, por independencia, es claro que $\mathrm{var}(\hat{\lambda}) = \mathrm{var}(\bar{X}) = \mathrm{var}(X_1)/n = \lambda/n$, donde hemos usado que la varianza de una Poisson(λ) es λ. Luego, $\hat{\lambda}$ es insesgado y su varianza converge a 0 cuando $n \to \infty$, por lo tanto es consistente por resultado visto en cátedra (*observación:* también es válido argumentar que la convergencia casi segura de $\hat{\lambda}$ a λ (ítem siguiente) implica que $\hat{\lambda}$ es consistente, pues la convergencia casi segura implica la convergencia en probabilidad).

3) (0,5 ptos.) Muestre que $\hat{\lambda}$ converge casi seguramente a λ cuando $n \to \infty$.

Solución. El resultado buscado es directo: por la LFGN, sabemos que $\hat{\lambda} = \bar{X}$ converge casi seguramente a $\mu = \mathbb{E}(X_1) = \lambda$ cuando $n \to \infty$.

b) (3,0 ptos.) El gerente comercial de una marca de detergentes afirma que el 40 % del público prefiere dicha marca. Usted trabaja para la competencia, y cree que dicha información es exagerada. Para verificar quién tiene razón usted encarga realizar un estudio de mercado, el cual indica que 54 de 150 personas encuestadas prefiere la marca. Plantee hipótesis H_0 y H_1 adecuadas e indique su conclusión para $\alpha = 5\,\%$.

Solución. Anotemos $X_1, \dots, X_n$ la muestra de $n = 150$ personas encuestadas, donde X_i vale 1 si la persona i-ésima prefiere el detergente, y 0 en otro caso. Claramente $X_i \sim$ Bernoulli(p) para todo i, donde $p \in (0,1)$ es la probabilidad de que un cliente prefiera el detergente, la cual es desconocida. Planteamos las hipótesis

$$H_0 : p = 40\,\%$$
$$H_1 : p < 40\,\%.$$

Para decidir el test, trabajamos con el estadístico

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}},$$

donde $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$ y $p_0 = 40\,\%$. Dada la forma de H_1, rechazaremos H_0 si $Z < c$, y para determinar la constante c, imponemos que la probabilidad de equivocarse sea la deseada:

$$\alpha = 5\,\% = \mathbb{P}(\text{rechazar } H_0 \text{ cuando es cierta}) = \mathbb{P}(Z < c \mid H_0) \approx \mathbb{P}(\mathcal{N}(0,1) < c),$$

donde en el último paso hemos usado el hecho que Z posee distribución aproximada $\mathcal{N}(0,1)$ cuando $p = p_0$, gracias al TCL. Lo anterior equivale a $5\,\% = \mathbb{P}(\mathcal{N}(0,1) > -c)$, y de la tabla de la normal se obtiene $c = -1{,}65$ (ó $-1{,}64$). El valor observado de Z en la muestra es

$$Z_{\text{obs}} = \frac{\hat{p}_{\text{obs}} - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}} = \frac{\frac{54}{150} - \frac{40}{100}}{\sqrt{\frac{40}{100}\cdot\frac{60}{100}}/\sqrt{150}} = \frac{\frac{9}{25} - \frac{10}{25}}{\sqrt{\frac{4}{10}\cdot\frac{6}{10}\cdot\frac{1}{150}}} = \frac{-\frac{1}{25}}{\frac{1}{10}\sqrt{\frac{4}{25}}} = -1.$$

Como $Z_{\text{obs}} = -1$ no es menor que $c = -1{,}65$, no corresponde rechazar H_0. Es decir, no hay suficiente evidencia para rechazar la afirmación del gerente de la competencia.

P2. *a*) (2,0 ptos.) Un gatito quiere subir a la mesa, para lo cual intentará saltar a la silla primero, y desde la silla a la mesa. Cada salto puede resultar en éxito o fracaso con igual probabilidad e independiente del resto, y cada vez que un salto fracasa el gatito cae al suelo y sigue intentando. Calcule la cantidad esperada de saltos hasta que el gatito alcanza la mesa. *Indicación:* condicione en los resultados de los primeros 2 saltos; con esto obtenga una ecuación para la esperanza buscada y despeje.

Solución. Sea X_0 la variable aleatoria de la cantidad de saltos hasta que el gatito sube a la mesa. Sea Y la variable aleatoria dada por

$$Y = \begin{cases} 0 & \text{si el primer salto es fracaso,} \\ 1 & \text{si el primer salto es éxito y el segundo es fracaso,} \\ 2 & \text{si el primer y segundo salto son éxitos.} \end{cases}$$

Utilizando esperanzas condicionales, tenemos:

$$\begin{aligned} \mathbb{E}(X_0) &= \mathbb{E}(X_0 \mid Y=0)\mathbb{P}(Y=0) + \mathbb{E}(X_0 \mid Y=1)\mathbb{P}(Y=1) + \mathbb{E}(X_0 \mid Y=2)\mathbb{P}(Y=2) \\ &= [1 + \mathbb{E}(X_1 \mid Y=0)]\frac{1}{2} + [2 + \mathbb{E}(X_2 \mid Y=1)]\frac{1}{4} + [2 + \mathbb{E}(X_2 \mid Y=2)]\frac{1}{4}. \end{aligned}$$

donde X_i es la cantidad de saltos que faltan una vez realizados los primeros i saltos, para $i = 1, 2$. Es claro que $\mathbb{E}(X_1 \mid Y=0) = \mathbb{E}(X_2 \mid Y=1) = \mathbb{E}(X_0)$, pues cuando un salto es fracaso es como si el gatito volviese a comenzar. Similarmente, es claro que $\mathbb{E}(X_2 \mid Y=2) = 0$, pues si el gatito tuvo éxito en los primeros 2 saltos significa que ya alcanzó la mesa. Por lo tanto, la incógnita $x = \mathbb{E}(X_0)$ satisface la ecuación

$$x = \frac{1}{2}(1+x) + \frac{1}{4}(2+x) + \frac{2}{4},$$

la cual tiene por solución a $\mathbb{E}(X_0) = x = 6$.

b) Un restaurante puede servir 80 comidas por noche, y sólo acepta clientes con reservación. La experiencia muestra que cada cliente que reserva tiene un 20 % de probabilidad de no llegar al restaurante, independiente del resto.

1) (3,0 ptos.) Usando el TCL, muestre que la máxima cantidad de reservaciones que puede aceptar el restaurante para que la probabilidad de poder servir a todos los clientes que vendrán sea mayor o igual a 97,72 %, es aproximadamente $n = 90$. *Indicación:* defina X_i como la v.a. que vale 1 si el cliente de la i-ésima reserva efectivamente llega, y 0 si no.

Solución. Con la notación de la indicación, la cantidad de clientes que llegan al restaurant es $\sum_{i=1}^n X_i$. La probabilidad del evento deseado es:

$$\mathbb{P}\left(\sum_{i=1}^n X_i \leq 80\right) = \mathbb{P}\left(\frac{\frac{1}{n}\sum_{i=1}^n X_i - p}{\sigma/\sqrt{n}} \leq \frac{\frac{1}{n}80 - p}{\sigma/\sqrt{n}}\right),$$

donde $p = 0{,}8$ y $\sigma = \sqrt{p(1-p)} = \sqrt{0{,}8 \times 0{,}2} = 0{,}4$ son la esperanza y raíz de la varianza de cada X_i. Por el TCL, la variable a la izquierda de la desigualdad anterior es aproximadamente normal estándar, por lo tanto la probabilidad buscada es aproximadamente igual a

$$\mathbb{P}(\mathcal{N}(0,1) \leq (80 - 0{,}8n)/(0{,}4\sqrt{n})).$$

Ahora imponemos que esta probabilidad sea igual a 97,72 %, o equivalentemente, que $2{,}28\,\% = \mathbb{P}(\mathcal{N}(0,1) > (80-0{,}8n)/(0{,}4\sqrt{n}))$. Como $\mathbb{P}(\mathcal{N}(0,1) > 2) = 2{,}28\,\%$, esto implica que $(80-0{,}8n)/(0{,}4\sqrt{n}) = 2$, es decir, $x = \sqrt{n}$ cumple la ecuación cuadrática $x^2 + x - 100 = 0$, cuya solución positiva es

$$x = \frac{-1 + \sqrt{1+400}}{2} \approx 9{,}5.$$

Por lo tanto, $n \approx 9{,}5^2 \approx 90$. Luego, el restaurant puede aceptar hasta 90 reservaciones.

2) (1,0 pto.) Suponga que una noche se acepta el máximo de $n = 90$ reservaciones. Sin utilizar el TCL, obtenga una cota inferior para la probabilidad de que lleguen entre 63 y 81 clientes.

Solución. Anotemos $Y = \sum_{i=1}^n X_i$ la cantidad de clientes que efectivamente llega, y notemos que $\mu = \mathbb{E}(Y) = \sum_{i=1}^n \mathbb{E}(X_i) = 90 \times 80/100 = 72$ y que $\mathrm{var}(Y) = \sum_{i=1}^n \mathrm{var}(X_i) = 90 \times \frac{4}{5} \times \frac{1}{5} = \frac{360}{25}$, pues cada X_i tiene distribución Bernoulli(p) con $p = 80\,\%$. Notemos además que $Y \in [63, 81]$ equivale a que $|Y - 72| \leq 9$. Usando la desigualdad de Chebyshev, tenemos entonces:

$$\mathbb{P}(Y \in [63,81]) = \mathbb{P}(|Y-\mu| \leq 9) = 1 - \mathbb{P}(|Y-\mu| \geq 10) \geq 1 - \frac{\mathrm{var}(Y)}{10^2}.$$

Es decir, $\mathbb{P}(Y \in [63,81]) \geq 1 - 36/250 = 214/250 = 856/1000 = 85{,}6\,\%$.

P3. A usted se le encarga realizar un estudio preliminar sobre la concentración de mineral en un posible futuro yacimiento minero. La variable de interés se modela como una $\mathcal{N}(\mu, \sigma^2)$ con ambos parámetros desconocidos.

a) (3,0 ptos.) Usted toma 16 muestras de roca, obteniendo un promedio de 105,0 y una desviación estándar (es decir, la raíz del estimador insesgado de la varianza) de 9,0. Obtenga intervalos de confianza para μ y σ^2 al nivel 95 %.

> **Solución.** Veamos el intervalo para μ: usamos el estadístico $T = (\bar{X} - \mu)/(S/\sqrt{n})$, el cual posee distribución $t_{n-1} = t_{15}$. Imponemos el nivel deseado a un intervalo simétrico para T:
>
> $$95\,\% = \mathbb{P}(T \in [-c, c]) = 1 - 2\mathbb{P}(T > c),$$
>
> luego $\mathbb{P}(T > c) = 2{,}5\,\%$. De la tabla de una t-student obtenemos que $c = 2{,}131$. Despejando μ, tenemos:
>
> $$95\,\% = \mathbb{P}\left(-c \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq c\right) = \mathbb{P}\left(\bar{X} - \frac{cS}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{cS}{\sqrt{n}}\right).$$
>
> Reemplazando los valores de $\bar{X}$ y S obtenidos en la muestra, tenemos que el intervalo de confianza para μ es
>
> $$\left[105 - \frac{2{,}131 \times 9}{4}, 105 + \frac{2{,}131 \times 9}{4}\right].$$
>
> Veamos el intervalo para σ^2: usamos el estadístico $U = (n-1)S^2/\sigma^2$, el cual posee distribución $\chi^2_{n-1} = \chi^2_{15}$. Imponemos el nivel deseado a U usando un intervalo $[a, b]$ que deja igual probabilidad debajo de a que por encima de b, es decir:
>
> $$\begin{aligned} 95\,\% = \mathbb{P}(U \in [a, b]) \quad &\Rightarrow \quad 2{,}5\,\% = \mathbb{P}(U < a) = \mathbb{P}(U > b) \\ &\Rightarrow \quad 97{,}5\,\% = \mathbb{P}(U > a) \text{ y } 2{,}5\,\% = \mathbb{P}(U > b). \end{aligned}$$
>
> De una tabla de la chi-cuadrado, esto entrega $a = 6{,}26$ y $b = 27{,}5$. Despejando σ^2, obtenemos:
>
> $$95\,\% = \mathbb{P}\left(a \leq \frac{(n-1)S^2}{\sigma^2} \leq b\right) = \mathbb{P}\left(\frac{(n-1)S^2}{b} \leq \sigma^2 \leq \frac{(n-1)S^2}{a}\right),$$
>
> y reemplazando los valores de la muestra obtenemos el intervalo deseado para σ^2:
>
> $$\left[\frac{15 \times 81}{27{,}5}, \frac{15 \times 81}{6{,}26}\right].$$

Los inversionistas están preocupados por estos datos, sobre todo por la variabilidad de la concentración del mineral, pues podría afectar negativamente los procesos productivos. Para tener mayor información al respecto, se le encarga a usted planificar un *nuevo* estudio sobre σ^2.

b) (1,5 pto.) Usted planea obtener un nuevo intervalo de confianza para σ^2 con una nueva muestra de 101 datos y el mismo nivel de confianza de 95 %. Calcule el *largo esperado* del intervalo a obtener, en función de σ^2.

> **Solución.** Por lo hecho en el ítem anterior sabemos que el intervalo será $[(n-1)S^2/b, (n-1)S^2/a]$, donde ahora $n = 101$, S^2 es el estimador insesgado de la varianza obtenido en esta nueva muestra, y a, b se calculan usando una $\chi_1 00^2$, es decir, $a = 74{,}2$ y $b = 129{,}6$.

El largo de este intervalo es la diferencia de los extremos, luego el largo esperado es

$$\mathbb{E}\left(\frac{(n-1)S^2}{a} - \frac{(n-1)S^2}{b}\right) = (n-1)\left(\frac{1}{a} - \frac{1}{b}\right)\mathbb{E}(S^2).$$

Pero como S^2 es estimador insesgado de σ^2, tenemos que $\mathbb{E}(S^2) = \sigma^2$. Luego, el largo esperado es $100\left(\frac{1}{74{,}2} - \frac{1}{129{,}6}\right)\sigma^2$.

c) (1,5 pto.) Se lleva a cabo el estudio planificado por usted, obteniendo un promedio de 108,0 y desviación estándar de 10,0. Por simplicidad y dado que el tamaño de la muestra es bastante grande, se trabajará con σ^2 igual al estimador insesgado obtenido en la muestra. Con estos datos, obtenga un nuevo intervalo de confianza para μ al 99 %.

Solución. En este caso usamos el estadístico $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ con $\sigma = 10^2 = 100$, el cual posee distribución $\mathcal{N}(0,1)$. Imponemos el nivel deseado a un intervalo simétrico para Z:

$$99\,\% = \mathbb{P}(Z \in [-c, c]) = 1 - 2\mathbb{P}(Z > c),$$

luego $\mathbb{P}(Z > c) = 0{,}5\,\%$. De la tabla de la normal obtenemos $c = 2{,}57$ (ó 2,58). Despejando μ, obtenemos:

$$99\,\% = \mathbb{P}\left(-c \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq c\right) = \mathbb{P}\left(\bar{X} - \frac{c\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{c\sigma}{\sqrt{n}}\right).$$

Reemplazando los valores correspondientes, el intervalo buscado es:

$$\left[108 - \frac{25{,}7}{\sqrt{101}}, 108 + \frac{25{,}7}{\sqrt{101}}\right].$$