Universidad de Chile

Facultad de Ciencias Físicas y Matemáticas

Departamento de Ingeniería Matemática

Auxiliar MA3403: Estimación e Intervalos de Confianza. Test de Hipótesis.

Profesor: Servet Martínez, Auxiliares: Gonzalo Contador, Francisco Unda.

P1. Sea $X_1 ... X_n$ una muestra aleatoria simple de una v.a. $X \sim Poisson(\lambda)$, con λ parámetro desconocido. Muestre que el estadístico $\sum\limits_{i=1}^{n} X_i$ es suficiente para λ.

Solución:
Como vimos al principio del semestre, si se tienen $(Y_i)_{i=1}^n$ variables aleatorias independientes con $Y_i \sim Poisson(\lambda_i)$, entonces se tiene que $\sum\limits_{i=1}^{n} Y_i \sim Poisson(\sum \lambda_i)$. Como la muestra es aleatoria y simple (los X_i son i.i.d.) se tiene que $\sum\limits_{i=1}^{n} X_i \sim Poisson(n\lambda)$. Para verificar que el estadístico es suficiente, calculemos la fdp condicional de la muestra dado el estadístico.

$$\mathbb{P}(X_1 = k_1, ..., X_n = k_n \mid \sum_{i=1}^{n} X_i = k) = \frac{\mathbb{P}(X_1 = k_1, ..., X_n = k_n, \sum\limits_{i=1}^{n} X_i = k)}{\mathbb{P}(\sum\limits_{i=1}^{n} X_i = k)}$$

Notemos que esta probabilidad será 0 si no se cumple $\sum\limits_{i=1}^{n} k_i = k$, pues los eventos serán disjuntos. Si se cumple $\sum\limits_{i=1}^{n} k_i = k$, entonces $\mathbb{P}(X_1 = k_1, ..., X_n = k_n, \sum\limits_{i=1}^{n} X_i = k) = \mathbb{P}(X_1 = k_1, ..., X_n = k_n) = \prod\limits_{i=1}^{n} \mathbb{P}(X_i = k_i)$ por la independencia de los valores muestrales. En tal caso:

$$\mathbb{P}(X_1 = k_1, ..., X_n = k_n \mid \sum_{i=1}^{n} X_i = k) = \frac{\prod\limits_{i=1}^{n} \mathbb{P}(X_i = k_i)}{\mathbb{P}(\sum\limits_{i=1}^{n} X_i = k)} = \frac{\prod\limits_{i=1}^{n} \frac{e^{-\lambda}\lambda^{k_i}}{k_i!}}{\frac{e^{-n\lambda}(n\lambda)^k}{k!}} = \frac{k!}{n^k \prod\limits_{i=1}^{n} k_i!}$$

La última igualdad se obtiene usando que $\sum\limits_{i=1}^{n} k_i = k$. Con esto, se observa que la distribución de la muestra dado $\sum\limits_{i=1}^{n} X_i$ no depende del parámetro desconocido λ, lo que prueba que el estadístico es suficiente para la estimación de este parámetro.■

P2. Considere $\theta \in \mathbb{R}^+$desconocido y $X_1...X_n$ una muestra aleatoria simple de una v.a. $X \sim Normal(\theta, \theta)$.

a) Construya un pivote (esto es, un estadístico cuya distribución no dependa del parámetro desconocido) que siga una distribución $Normal(0,1)$

b) Imponga condiciones sobre la muetra de manera que exista un intervalo de confianza de nivel $1-\alpha$ para el parámetro θ.

Solución:

a) Tenemos que $\mathbb{E}(X_i) = \theta = \mathbb{V}ar(X_i)\ \forall i$. Luego, utilizando el teorema central del límite, tenemos que

$$\frac{\sqrt{n}(\overline{X}-\theta)}{\sqrt{\theta}} \sim Normal(0,1)$$

cuya distribución no depende del parámetro θ.■

b) Debemos encontrar un intervalo $I=[a,b]$ de largo mínimo tal que $\mathbb{P}(\theta \in I) = 1-\alpha$. Observemos que

$$\theta \in I \Leftrightarrow a < \theta < b \Leftrightarrow \frac{\sqrt{n}(\overline{X}-b)}{\sqrt{\theta}} < \frac{\sqrt{n}(\overline{X}-\theta)}{\sqrt{\theta}} < \frac{\sqrt{n}(\overline{X}-a)}{\sqrt{\theta}}$$

Luego $\mathbb{P}(\theta \in I) = \mathbb{P}(\frac{\sqrt{n}(\overline{X}-b)}{\sqrt{\theta}} < \frac{\sqrt{n}(\overline{X}-\theta)}{\sqrt{\theta}} < \frac{\sqrt{n}(\overline{X}-a)}{\sqrt{\theta}}) = \mathbb{P}(-c < \frac{\sqrt{n}(\overline{X}-\theta)}{\sqrt{\theta}} < c) = 1-\alpha$. (Imponemos que el intervalo sea de la forma $[-c,c]$ para que tenga largo mínimo, puesto que la fdp de $\frac{\sqrt{n}(\overline{X}-\theta)}{\sqrt{\theta}}$ alcanza su máximo en 0, es estrìctamente monótona y no negativa en todo su dominio. Encontrado c, podemos despejar a y b para encontrar I de largo mínimo)

Notemos que la condición $-c < \frac{\sqrt{n}(\overline{X}-\theta)}{\sqrt{\theta}} < c$ es equivalente a $\frac{n(\overline{X}-\theta)^2}{\theta} < c^2$, o bien a que $n\theta^2 - (2n\overline{X}+c^2)\theta + n\overline{X} < 0$. Despejando los valores críticos de la ecuación cuadrática para θ obtenemos

$$\theta_{crit} = \frac{2n\overline{X}+c^2 \pm \sqrt{4n^2\overline{X}^2 + 4nc^2\overline{X} - c^4 - 4n^2\overline{X}^2}}{2n}$$

Así, para que exista el intervalo de confianza debe cumplirse que el discriminante sea positivo, de lo contrario no habrán valores que satisfagan la condición $\theta \in I$. Esto es, $4nc^2\overline{X} - c^4 > 0$, o bien, $\overline{X} > \frac{c^2}{4n}$. (Notemos que la condición no sólo depende del promedio muestral y del tamaño de la muestra, sino también del nivel de confianza requerido para el intervalo)

P3. En este problema, considere $X \sim Uniforme(0,\theta)$, donde $\theta > 0$ es un parámetro desconocido.

a) Considere $X_1...X_n$ muestra aleatoria simple de X. Demuestre que el estimador para θ, $\hat{\theta} = \underset{i}{max} X_i$ es consistente.

b) Suponga ahora una muestra de tamaño 1, X_1, de X. Encuentre una cota inferior para θ de un 95% de confianza.

Solución:

a) Calculemos primero la función de densidad f de $\hat{\theta}$. Tenemos, para $x \in (0, \theta)$

$$\mathbb{P}(\hat{\theta} < x) = \mathbb{P}(\underset{i}{max} X_i < x) = \mathbb{P}(X_1 < x, ..., X_n < x) = \prod_{i=1}^{n} \mathbb{P}(X_i < x) = (\frac{x}{\theta})^n$$

Luego $f(x) = \dfrac{nx^{n-1}}{\theta^n}$. Verifiquemos la consistencia del estimador. Sea $\varepsilon > 0$. Notando que $\hat{\theta}$ no puede tomar valores mayores que θ

$$\mathbb{P}(\mid \hat{\theta} - \theta \mid < \varepsilon) = \int_{\theta-\varepsilon}^{\theta+\varepsilon} f(x)dx = \int_{\theta-\varepsilon}^{\theta} f(x)dx = 1 - (\frac{\theta - \varepsilon}{\theta})^n$$

De donde obtenemos $\mathbb{P}(\mid \hat{\theta} - \theta \mid > \varepsilon) = (\frac{\theta-\varepsilon}{\theta})^n$. Luego, se cumple

$$\underset{n \to \infty}{lim} \mathbb{P}(\mid \hat{\theta} - \theta \mid > \varepsilon) = 0$$

Así, se demuestra que $\hat{\theta}$ converge en probabilidad hacia θ. Luego, $\hat{\theta}$ es consistente.$\blacksquare$

★Propuesto: Verificar que $\tilde{\theta} = \frac{n+1}{n}\hat{\theta}$ también es consistente. Para esto, verifiquen que el estimador es insesgado en este caso, y recuerden que cuando el estimador es insesgado la consistencia equivale a que la varianza tiende a 0 al aumentar el tamaño de la muestra (¿por qué?).

b) Calculemos la distribución de $U = \frac{X_1}{\theta}$. Utilizando el teo. de cambio de variables en una dimensión $f_U(u) = f_X(u\theta) = 1$ si $u \in (0,1)$. Luego, $U \sim Uniforme(0,1)$. Calculando c tal que $\mathbb{P}(U < c) = 0.95$, se tiene $c = 0.95$. Así

$$\mathbb{P}(\frac{X_1}{\theta} < 0.95) = 0.95 = \mathbb{P}(\frac{X_1}{0.95} < \theta)$$

con lo que se obtiene que la cota buscada es $\frac{X_1}{0.95}$

P4. Considere $X_1...X_n$ muestra aleatoria simple de $X \sim exponencial(\lambda)$, con $\lambda > 0$ desconocido. Encuentre el estimador de máxima verosimilitud de λ.

Solución:

Escribimos la función de verosimilitud de la muestra. Por independencia de la muestra, esta será el producto de las densidades de cada valor muestral, osea

$$f(x) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i} = \lambda^n e^{-\lambda \sum X_i}$$

Para encontrar el estimador, debemos encontrar el valor $\hat{\lambda}$ que maximiza la función de verosimilitud. Por ser $ln(x)$ una función estrictamente creciente, tenemos que el valor que maximice f también maximizará $ln(f)$. Luego

$$ln(f(x)) = nln(\lambda) - \lambda \sum X_i$$

derivando con respecto al parámetro a estimar e imponiendo condición de máximo, se obtiene

$$\frac{n}{\lambda} - \sum X_i = 0 \Leftrightarrow \lambda = \frac{n}{\sum X_i} = \frac{1}{\overline{X}}$$

Luego, el estimador de máxima verosimilitud es el inverso del promedio muestral.