Estadística
CURE

Práctico 2

Curso 2018

EJERCICIO 1. La distribución gamma es una familia de distribuciones continuas, de dos parámetros α y β, cuya pdf es

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad 0 < x < \infty, \alpha > 0, \beta > 0. \tag{1}$$

El parámetro α se llama *parámetro de forma*, porque tiene influencia sobre el máximo de la distribución, mientras que el parámetro β se llama *paràmeto de escala*, ya que su mayor influencia recae en la extensión de la distribución.

(a) Demuestre si $X \sim \Gamma(\alpha,\beta)$, el valor esperado es $E(X) = \alpha\beta$ y la varianza es $Var(X) = \alpha\beta^2$.

(b) Demuestre que el estimador máximo verosímil (MLE) de β para la distribución gamma es $\hat{\beta} = \bar{X}/\alpha$. Se asumirá que α es conocido.

(c) Si se asume que tanto α como β son desconocidos, no hay una fórmula explícita para los MLE de α y β. En este caso el máximo se puede reducir a un problema univariado utilizando el resultado de la parte (b) y luego resolver numéricamente. Escriba un programa para calcular los MLEs de α y β dados los siguientes datos:

$$22.0, 23.9, 20.9, 23.8, 25.0, 24.0, 21.7, 23.8, 22.8, 23.1, 23.1, 23.5, 23.0, 23.0$$

Si su resultado numérico es correcto, debería obtener los valores numéricos $(\hat{\alpha}, \hat{\beta}) = (514.219, 0.0450)$.

Nota. Estos datos corresponden al largo (en milímetros) de una muestra de huevos de aves de la familia Cuculidae.

EJERCICIO 2. En este ejercicio vemos algunas propiedades elementales de la *distribución chi-cuadrado* con p grados de libertad. Esta distribución es una sub-familia de la distribución gamma, que se obtiene de la Ec. (1) haciendo $\alpha = p/2$ con p entero y $\beta = 2$. Se denota χ^2_p y su pdf está dada por:

$$f(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{\frac{p}{2}-1} e^{-x/2}, \quad 0 < x < \infty.$$

(a) Calcule el valor esperado y la varianza de la distribución chi cuadrado.

(b) Si $Z \sim \mathcal{N}(0,1)$, entonces $Z^2 \sim \chi^2_1$. Es decir, el cuadrado de una distribución normal estándar tiene distribución chi-cuadrado.

(c) (*Opcional*) Si $X_1, \ldots, X_n$ son v.a. iid tales que $X_i \sim \chi^2_{p_i}$, entonces

$$X_1 + \ldots + X_n \sim \chi^2_{p_1+\ldots p_n}.$$

Es decir, v.a. chi-cuadrado independientes se distribuyen según una chi-cuadrado, y los grados de libertado también son aditivos.

EJERCICIO 3. Si $X_1, \ldots, X_n$ es una muestra aleatoria de una distribución normal $\mathcal{N}(\mu, \sigma^2)$, y si $\bar{X} = (1/n)\sum_{i=1}^n X_i$ y $S^2 = [1/(n-1)]\sum_{i=1}^n (X_i - \bar{X})^2$, la cantidad

$$(X - \mu)/(S/\sqrt{n})$$

tiene una *distribución t de Student* con $n-1$ grados de libertad.

De manera equivalente, una v.a. X tiene una distribución t de Student con p grados de libertad, y escribimos $X \sim t_p$, si su pdf es

$$f(t|p) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1+t^2/p)^{(p+1)/2}}, \quad -\infty < t < \infty.$$

Sea X una v.a. distribución t de Student con p grados de libertad.

(a) Calcule la media y la varianza de X.

(b) Demuestre que X^2 tiene una distribución F con 1 y p grados de libertad.

(c) Si $f(x|p)$ es la pdf de X, demuestre que

$$\lim_{p\to\infty} f(x|p) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty$$

Este resultado sugiere –correctamente– que a medida que $p \to \infty$, X converge a una v.a. $\mathcal{N}(0,1)$.

Sugerencia: utilice la fórmula de Stirling.

(d) Utilice los resultados de las partes (a) y (b) para justificar que, a medida que $p \to \infty$, X^2 converge a una v.a. con distribución chi-cuadrado χ^2_1.

EJERCICIO 4. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución binomial $\mathrm{Bin}(k,p)$, donde p es conocido y k es desconocido. Por ejemplo, lanzamos una moneda que sabemos que no está cargada y observamos x_i caras, pero desconocemos cuántas veces la moneda fue lanzada. Recuerde que la función de verosimilitud en este caso es

$$L(k|x,p) = \prod_{i=1}^n \binom{k}{x_i} p^{x_i}(1-p)^{k-x_i}.$$

(a) Demuestre que el MLE es un número $\hat{k} \in \mathbb{Z}$ que satisface las desigualdades

$$(k(1-p))^n \geq \prod_{i=1}^n (k - x_i), \qquad ((k+1)(1-p))^n < \prod_{i=1}^n (k+1-x_i),$$

y además es el mayor entero menor o igual a $1/\hat{z}$, donde $\hat{z}$ se define como aquél valor de z que es solución de la ecuación

$$(1-p)^n = \prod_{i=1}^{n}(1 - x_i z).$$

(b) Calcule $\hat{k}$ utilizando la parte anterior si $p = 1/2$, $n = 4$, y $X_1 = 0$, $X_2 = 20$, $X_3 = 1$ y $X_4 = 19$.

EJERCICIO 5. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias iid con distribución normal $\mathcal{N}(\theta, \sigma^2)$, con θ y σ^2 desconocidos. Utilizando el método de maximizaciones sucesivas demuestre que los estimadores de máxima verosimilitud correspondientes son $\bar{X}$ y $n^{-1}(X_i - \bar{X})^2$.

Sugerencia. El siguiente resultado puede resultar útil. Sean $x_1, \ldots, x_n$ un conjunto de números y sea $\bar{x}$ su media. Entonces $\text{mín}_{a\in\mathbb{R}}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$.

EJERCICIO 6. En este ejercicio estudiamos los estimadores Bayesianos de la distribución normal. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $\mathcal{N}(\theta, \sigma^2)$, y supongamos que la distribución a priori para θ es una normal $\mathcal{N}(\mu, \tau^2)$. Se asumirá que σ^2, μ y τ^2 son todos conocidos.

(a) Encuentre la pdf conjunta de $\bar{X}$ y de θ.

(b) Pruebe que $m(\bar{x}|\sigma^2, \mu, \tau^2)$, la distribución marginal de $\bar{X}$, es $\mathcal{N}(\mu, (\sigma^2/n) + \tau^2)$.

(c) Pruebe que $\pi(\theta|\bar{x}, \sigma^2, \mu, \tau^2)$, la distribución a posteriori de θ, es normal de media y varianza

$$\mathrm{E}(\theta|x) = \frac{\tau^2}{\tau^2 + \sigma^2}x + \frac{\sigma^2}{\sigma^2 + \tau^2}\mu, \qquad \mathrm{Var}(\theta|x) = \frac{\sigma^2\tau^2}{\tau^2 + \sigma^2}.$$

(d) Interprete estos resultados.

EJERCICIO 7. En este ejercicio estudiamos el error cuadrático medio o *mean squared error* (MSE) del estimador Bayesiano de una binomial. Sean $X_1, \ldots, X_n$ iid con distribución Bernoulli(p).

(a) Demuestre que el estimador Bayesiano de la probabilidad de suceso es

$$\hat{p}_B = \frac{Y + \alpha}{\alpha + \beta + n}.$$

(b) Calcule el MSE de dicho estimador Bayesiano.

(c) Demuestre que la elección $\alpha = \beta = \sqrt{n/4}$ da un MSE constante para $\hat{p}_B$.