Introducción a la Estadística y Ciencia de Datos

GUÍA DE ACTIVIDADES - CLASE DE TEST DE HIPÓTESIS (CONTINUACIÓN)

1. El tiempo (en horas) que un operario demora en realizar una tarea es una variable aleatoria X con densidad
$$f_\theta(x) = \frac{\theta}{2}\left(\frac{x}{2}\right)^{\theta-1} \mathbf{1}\{0 < x < 2\}, \quad \theta > 0$$
Diseñar un test asintótico de nivel $\alpha = 0{,}05$ para $H_0 : \theta \leq 2$ vs. $H_1 : \theta > 2$. Se midió el tiempo de 36 operarios y se obtuvo que
$$-\sum_{i=1}^{36} \ln(x_i/2) = 19{,}498\,.$$
¿Qué concluye?

2. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población con distribución $\mathcal{P}(\lambda)$. Consideramos las hipótesis $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$, con $\lambda_0 > 0$.

   a) Probar que el estadístico del test de cociente de máxima verosimilitud para testear las hipótesis planteadas puede escribirse como
   $$\Lambda = 2n(\lambda_0 - \overline{X} - \overline{X}\ln(\lambda_0/\overline{X}))$$

   b) Diseñar un test asintótico de nivel α para las hipótesis planteadas usando la distribución asintótica del cociente de máxima verosimilitud.

   c) La cantidad de errores de tipeo por página que comete un tipográfo sigue una distribución de Poisson. Se quiere chequear si el valor medio de dicha distribución sigue siendo el valor histórico, que es de 3 errores por página. Para ello se ha tomado una muestra de 30 observaciones y se obtienen los siguientes valores:

   $$\begin{array}{ccccccccccccccc} 7, & 3, & 5, & 6, & 6, & 4, & 1, & 5, & 1, & 3, & 2, & 2, & 4, & 3, & 5, \\ 3, & 4, & 1, & 3, & 2, & 7, & 1, & 4, & 1, & 4, & 5, & 2, & 3, & 4, & 9. \end{array}$$

   ¿Qué decisión se toma a nivel 5 %? Plantear las hipótesis, el estadístico, la zona de rechazo y su decisión a partir de los datos.

3. Las variables aleatorias normalmente distribuídas $X_1, \ldots, X_n$ se denominan *serialmente correlacionadas* o siguen un *modelo autorregresivo* si se escriben como
$$X_i = \theta X_{i-1} + \epsilon_i \quad i = 1, ..., n$$
donde $X_0 = 0$, θ el parámetro desconocido y $\epsilon_1, ..., \epsilon_n$ es una muestra aleatoria con distribución $\mathcal{N}(0, \sigma_0^2)$ con σ_0^2 una constante conocida. El vector aleatorio $\mathbf{X} = (X_1, ..., X_n)$ tiene función de densidad conjunta de la forma
$$f_{\mathbf{X}}(\mathbf{x}, \theta) = (2\,\pi\sigma_0^2)^{-\frac{1}{2}n} e^{-\frac{1}{2\,\sigma_0^2}\sum_{i=1}^{n}(x_i - \theta x_{i-1})^2}$$

donde $x_0 = 0$. El estimador de máxima verosimilitud de θ, expresado en función del vector aleatorio $\mathbf{X}$, resulta

$$\hat{\theta}(\mathbf{X}) = \frac{\displaystyle\sum_{i=1}^{n} X_i X_{i-1}}{\displaystyle\sum_{i=1}^{n} X_{i-1}^2}$$

Dadas las hipótesis

$$H_0 : \theta = 0 \quad \text{contra} \quad H_1 : \theta \neq 0\,.$$

a) Probar que el estadístico del test de cociente de verosimilitud resulta

$$E = \frac{\left(\displaystyle\sum_{i=1}^{n} X_i X_{i-1}\right)^2}{\displaystyle\sum_{i=1}^{n} X_i^2}\,.$$

b) Hallar el test de cociente de verosimilitud de nivel asintótico α para las hipótesis planteadas. *(Notar que H_0 indica independencia de la muestra $\mathbf{X}$ y H_1 confirma la correlación serial.)*