

MA34B-Estadística - Auxiliar 7
Profesora: Nancy Lacourly
Auxiliares: Jerónimo Escribano - Daniela Beltrán

Problema 1: Test de Hipótesis simple.

En general consideramos 2 hipótesis sobre algún parámetro y queremos saber cual de estas 2 hipótesis explica mejor los datos tomados

$$H_0 : \lambda = \lambda_0$$

$$H_1 : \lambda = \lambda_1$$

Para ello buscamos la mejor regla de decisión de manera que nos equivoquemos lo menos posible al aceptar o rechazar una de las dos hipótesis. Nos apoyamos en resultados:

Teorema: Si δ^* es una regla de decisión tal que:

se rechaza H_0 cuando $af_0(\underline{x}) < bf_1(\underline{x})$

se acepta H_0 cuando $af_0(\underline{x}) > bf_1(\underline{x})$

entonces $a\alpha(\delta^*) + b\beta(\delta^*) \leq a\alpha(\delta) + b\beta(\delta) \quad \forall \delta$ regla de decisión

Donde $f_0(\underline{x})$ y $f_1(\underline{x})$ son las funciones de verosimilitud bajo las hipótesis H_0 y H_1 respectivamente.

Recordemos que $\alpha(\delta) = \mathbb{P}(\text{rechazar } H_0 | H_0)$ y $\beta(\delta) = \mathbb{P}(\text{no rechazar } H_0 | H_1)$ bajo la regla δ . α y β se llaman error tipo 1 y error tipo 2 respectivamente.

Consideremos $X \sim \text{Poiss}(\lambda)$ y las hipótesis $H_0 : \lambda = \lambda_0 \quad H_1 : \lambda = \lambda_1$

tenemos que

$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

por lo que si obtenemos una m.a.s.,

$$f_n(\underline{x}|\lambda) = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!}$$

y la razón de verosimilitud queda

$$\Lambda = \frac{f_n(\underline{x}|\lambda_1)}{f_n(\underline{x}|\lambda_0)} = \frac{\lambda_1^{\sum x_i} e^{-n\lambda_1}}{\lambda_0^{\sum x_i} e^{-n\lambda_0}} = \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum x_i} e^{-n(\lambda_1 - \lambda_0)}$$

En consecuencia con el teorema anterior, si rechazamos H_0 cuando $\Lambda > \frac{a}{b}$ y aceptamos H_0 cuando $\Lambda < \frac{a}{b}$ estamos escogiendo la mejor regla, en el sentido que minimizamos el error ponderado $a\alpha(\delta) + b\beta(\delta)$

Como determinamos la región de rechazo a partir de la muestra?

Consideremos a y b tales que $a + b = 1$. Sigue que

$$\Lambda = \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum x_i} e^{-n(\lambda_1 - \lambda_0)} > \frac{a}{b}$$

y como $\ln(\cdot)$ es estrictamente creciente, y suponiendo que $a \neq 0$

$$\Rightarrow \ln(\Lambda) > \ln\left(\frac{a}{b}\right)$$

$$\Rightarrow \sum x_i \ln\left(\frac{\lambda_1}{\lambda_0}\right) > n(\lambda_1 - \lambda_0) + \ln\left(\frac{a}{b}\right)$$

Donde reconocemos 2 casos:

Si $\lambda_1 > \lambda_0$ entonces rechazamos H_0 si $\frac{\ln(\frac{a}{b}) + n(\lambda_1 - \lambda_0)}{\ln(\frac{\lambda_1}{\lambda_0})} < \sum x_i$

Si $\lambda_1 < \lambda_0$ entonces rechazamos H_0 si $\frac{\ln(\frac{a}{b}) + n(\lambda_1 - \lambda_0)}{\ln(\frac{\lambda_1}{\lambda_0})} > \sum x_i$

Notemos en el resultado anterior que el caso particular en que $a = b$, obtenemos $\ln(\frac{a}{b}) = 0$ y como consecuencia de lo anterior, las regiones de rechazo son del tipo $\sum x_i \gtrless c$ con c una constante que depende de $n, \lambda_0, \lambda_1, a, b$. (En el fondo esto es una consecuencia directa del teorema enunciado al principio y el lema de Neyman-Pearson).

Ahora supongamos que toleramos un error máximo tipo 1 α_0 fijo. O sea, que $\alpha(\delta) \leq \alpha_0$, ie,

$$\mathbb{P}(\text{rechazar } H_0 | H_0 \text{ es cierto}) = \mathbb{P}\left(\sum x_i > c | \lambda = \lambda_0\right) \leq \alpha_0 \text{ si } \lambda_1 > \lambda_0$$

$$\mathbb{P}(\text{rechazar } H_0 | H_0 \text{ es cierto}) = \mathbb{P}\left(\sum x_i < c' | \lambda = \lambda_0\right) \leq \alpha_0 \text{ si } \lambda_1 < \lambda_0$$

Para algun c o c' (dependiendo del caso). Cual es ese c ?

Notemos que la suma de n de v.a. Poisson de parámetro λ independientes es una Poisson de parámetro $n\lambda$. Con esto la condición anterior queda escrita como

$$\mathbb{P}(\text{Poiiss}(n\lambda_0) > c) = \alpha_0$$

Para poder calcular c supongamos que tenemos una muestra de tamaño $n = 50$, $\lambda_0 = 3$, $\lambda_1 = 5$ y $\alpha_0 = 0,05$. Con esto, tenemos que

$$\alpha_0 = 0,05 = \mathbb{P}\left(\sum x_i > c | H_0\right) = 1 - \mathbb{P}\left(\sum x_i \leq c | H_0\right)$$

De donde se concluye (por tablas estadísticas) que $c = 170$.

Ahora si suponemos que $\lambda_1 < \lambda_0$, por ejemplo, $n = 50$, $\lambda_1 = 3$, $\lambda_0 = 5$ y $\alpha_0 = 0,05$ obtenemos que $\alpha_0 = 0,05 = \mathbb{P}\left(\sum x_i < c | H_0\right)$ y consecuentemente que $c = 224$

Por lo tanto, con los datos anteriores, si $\lambda_1 = 5$ y $\lambda_0 = 3$, rechazamos con un nivel de significación de un 5% si $\sum x_i > 170$ y en el segundo caso si $\sum x_i < 224$

Supongamos ahora que sabemos que $\sum x_i = 180$, $\lambda_1 = 5$, $\lambda_0 = 3$ (estamos en el caso ($\lambda_1 > \lambda_0$)) y tomemos $c = \sum x_i = 180$.

Calculamos el p-valor $= \mathbb{P}(\text{Poiiss}(n\lambda) > c | H_0) = \mathbb{P}(\text{Poiiss}(n\lambda_0) > c) = 1 - \mathbb{P}(\text{Poiiss}(n\lambda_0) \leq c)$, que por tablas estadísticas resulta ser $p\text{-valor} = 0,0076 < 0,05$. Por lo tanto, si tenemos un nivel de significación del 5% debemos rechazar la hipótesis nula, puesto que la probabilidad de equivocarse si rechazamos H_0 es pequeña. Ojo que si rechazamos H_0 no necesariamente debemos aceptar H_1 . En caso que el p-valor sea mayor que nuestro nivel de significación debemos aceptar la hipótesis nula.

Problema 2 (Ejemplo)

Sea $X \sim \text{Gamma}(r, \lambda)$. X tiene función distribución $f(x) = \frac{x^{r-1} e^{-\frac{x}{\lambda}}}{\lambda^r \Gamma(r)}$. Sea $\{x_i\}$ una m.a.s. de tamaño n . Sean las hipótesis

$$H_0 : \lambda = \lambda_0$$

$$H_1 : \lambda = \lambda_1$$

Tenemos que

$$f_n(\underline{x}) = \frac{(\prod x_i)^{r-1} e^{-\frac{\sum x_i}{\lambda}}}{(\lambda^r \Gamma(r))^n}$$

y luego

$$\Lambda = \frac{(\prod x_i)^{r-1} e^{-\frac{\sum x_i}{\lambda_1}}}{(\prod x_i)^{r-1} e^{-\frac{\sum x_i}{\lambda_0}}} = \exp\left(-\sum x_i \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_0}\right)\right)$$

Supongamos que $\lambda_1 < \lambda_0$. (*Queda propuesto el análisis si $\lambda_1 > \lambda_0$*)

$$\Lambda > k \Rightarrow \sum x_i > \ln(k) \frac{\lambda_0 \lambda_1}{\lambda_0 - \lambda_1}$$

Sabemos que si $Z \sim \text{Gamma}(s, \lambda)$ y $Y \sim \text{Gamma}(t, \lambda)$ entonces $Z + Y \sim \text{Gamma}(s + t, \lambda)$ Por lo tanto, $\sum(x_i) \sim \text{Gamma}(nr, \lambda)$.

Supongamos que $\alpha_0 = 0,10$, $n = 60$, $\lambda_1 = 0,9$, $\lambda_0 = 1$ $r = 3$ y calculemos la región de rechazo:

$$0,10 = 1 - \mathbb{P}(\sum x_i < c) \Rightarrow c = 197,37$$

Por lo tanto, la hipótesis nula se rechaza si $\sum x_i > 197,37$