Probabilidad y estadística

Clase 6

Test de hipótesis

Motivación

Un test de hipótesis es una manera formal de elegir entre dos opciones, o hipótesis.

Un ejemplo de la vida cotidiana: un juicio. Se cuenta con dos hipótesis: la persona es inocente o es culpable. Para decidir que es culpable (y enviarlo a la cárcel) el juez a tener que tener evidencia suficiente para tomar esa decisión.

La presuposición de inocencia ("inocente hasta que se demuestre lo contrario") es lo que se conoce como hipótesis nula. La otra es la que se conoce como hipótesis alternativa.

Formalicemos esta idea

Sea una m.a. $\underline{X} = X_1, \ldots, X_n$ de una población con distribución perteneciente a una familia $F_\theta(x)$ con $\theta \in \Theta$. Sean Θ_0 y Θ_1 tales que $\Theta_0 \cup \Theta_1 = \emptyset$ y $\Theta_0 \cup \Theta_1 = \Theta$. Un test para este problema será una regla basada en $\underline{X}$ para decidir entre las dos hipótesis

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_1$$

Definición: Se llama test a una función $\delta(\underline{X})$ que puede tomar valores 0 o 1.

Formalicemos esta idea

En un test de hipótesis se plantean dos hipótesis:

H_1: es la hipótesis del investigador (lo que se desea probar). Se la conoce como hipótesis alternativa

H_0: se formula con el objetivo de ser rechazada. Se la conoce como hipótesis nula.

Diremos que rechazamos la hipótesis nula cuando $\delta(\underline{X}) = 1$, en caso contrario decimos que no hay evidencia suficiente para rechazar H_0.

Tipos de error

Error de tipo I: Es el error que se comete al rechazar una hipótesis nula que era verdadera. Buscamos que esto tenga muy baja probabilidad.

$$\mathbb{P}(EI) = \mathbb{P}_{\theta\in\Theta_0}(\delta(\underline{X}) = 1)$$

Error de tipo II: Es el error que se comete al no rechazar una hipótesis nula que era falsa.

$$\mathbb{P}(EII) = \mathbb{P}_{\theta\in\Theta_1}(\delta(\underline{X}) = 0)$$

Potencia del test

Def: Se llama potencia del test a la probabilidad de rechazar una hipótesis nula en función del parámetro desconocido sobre el que se plantea la hipótesis.

$$\pi_\delta(\theta) = \mathbb{P}_\theta(\delta(\underline{X}) = 1).$$

Luego podemos reescribir

$$\mathbb{P}(EI) = \pi_\delta(\theta),\ \theta \in \theta_0 \quad \text{y} \quad \mathbb{P}(EII) = 1 - \pi_\delta(\theta),\ \theta \in \Theta_1$$

Nivel de significación y p-valor

Def: Se llama nivel de significación del test a la máxima probabilidad de cometer un error de tipo I:

$$\alpha = \sup_{\theta \in \theta_0} \pi_\delta(\theta)$$

Def: Se llama p-valor de un test al menor nivel de significación para el cual se rechaza H_0 para una observación dada

Sobre la construcción de reglas de decisión

En la práctica, las reglas de decisión se construyen basándose en una estadística de la m.a. $\underline{X}$, es decir que son de la forma $\delta(\underline{X}) = \boldsymbol{I}\{T(\underline{X}) \in \mathcal{R}\}$. A $\mathcal{R}$ la conoce como región crítica o de rechazo .

Resulta intuitivo que el estadístico $T(\underline{X})$ se corresponda con el estadístico suficiente (o una transformación del mismo)

En particular, podemos utilizar el método del pivote visto para intervalos de confianza para construir el test.

Tipos de hipótesis

En general, tenemos 3 tipos de hipótesis que deseamos testear:

$$H_0: \theta \geq \theta_0 \ vs. \ H_1: \theta < \theta_0$$

$$H_0: \theta \leq \theta_0 \ vs. \ H_1: \theta > \theta_0$$

Hipótesis unilaterales

$$H_0: \theta = \theta_0 \ vs. \ H_1: \theta \neq \theta_0$$

Hipótesis bilaterales

Si el pivote es decreciente en θ, un test para

$H_0: \theta \leq \theta_0 \ vs. \ H_1: \theta > \theta_0$ será $\delta(\underline{X}) = \boldsymbol{I}\{U_{\theta_0} > k_\alpha\}$

$H_0: \theta \geq \theta_0 \ vs. \ H_1: \theta < \theta_0$ será $\delta(\underline{X}) = \boldsymbol{I}\{U_{\theta_0} < k_\alpha\}$

$H_0: \theta = \theta_0 \ vs. \ H_1: \theta \neq \theta_0$ será $\delta(\underline{X}) = \boldsymbol{I}\{k_{\alpha/2} < U_{\theta_0} < k_{1-\alpha/2}\}$

Test con nivel de significación asintótico

Def: Sea $\underline{X}_n = (X_1, \ldots, X_n)$ una m.a. de duna población con distribución $F_\theta(x),\ \theta \in \Theta$. Se desea testear

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_1$$

Se dirá que una sucesión de test tiene nivel de significación asintótico α si

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_0} \pi_\delta(\theta) = \alpha$$

Nuevamente, podemos basar el test en la distribución asintótica de EMV, o bien usando TCL.

Estimación Bayesiana

Enfoque frecuentista vs. enfoque Bayesiano

Hasta ahora vimos métodos de estimación paramétricas que siguen el enfoque frecuentista. Se asume que parámetro de interés, θ, es desconocido, pero constante. Es decir que no se hacen suposiciones previas respecto de θ.

El enfoque Bayesiano, en cambio, supone que se tiene alguna información previa sobre el parámetro. Esta información previa se expresa en forma de una distribución sobre θ, llamada distribución a priori

Enfoque frecuentista: θ es una constante (desconocida)
Enfoque Bayesiano: θ es una v.a.

Distribución a priori

Según el problema se pueden tener distintas interpretaciones acerca de la distribución a priori $\pi(\theta)$:

- La distribución a priori está basada en experiencias previas similares
- La distribución a priori expresa una creencia subjetiva.

A la variable aleatoria la llamaremos Θ, para distinguirla del valor que toma θ.

También cambia la interpretación de familia de distribución, ya que ahora $F_\theta(x)$ es la distribución condicional de X dado que $\Theta = \theta$

Distribución a posteriori

Una vez observada la m.a. $\underline{X} = (X_1, \ldots, X_n)$ se puede calcular la distribución de Θ dada $\underline{X} = \underline{x}$. A esta distribución se la conoce como distribución a posteriori y está dada por

$$f_{\Theta|\underline{X}=\underline{x}}(\theta) = \frac{f_{\underline{X}|\Theta=\theta}(\underline{x})\pi(\theta)}{\int_{-\infty}^{\infty} f_{\underline{X}|\Theta=\theta}(\underline{x})\pi(\theta)d\theta}.$$

Obs: Si alguna de las variables ($\underline{X}$ o Θ) son discretas reemplazaremos la función de densidad por una función de probabilidad, y si Θ reemplazaremos la integral por una sumatoria.

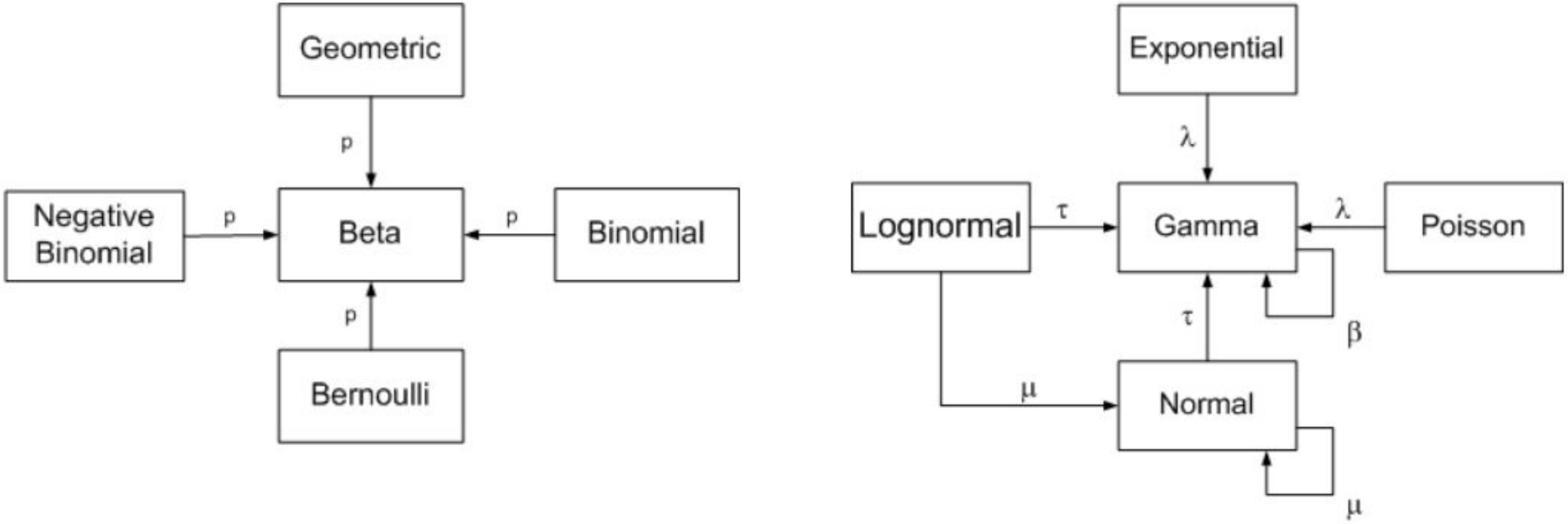
Cómo elegir la distribución a priori

- Cuando no tenemos información acerca del parámetro vamos a usar distribuciones a priori que no favorezcan ningún valor, como por ejemplo una distribución uniforme.

- Hay que prestar atención al soporte que tiene la distribución a priori elegida, ya que los valores de θ que no pertenezcan al soporte original quedan "eliminados".

- Por simplicidad, suelen elegirse **distribuciones conjugadas**. Las familias de distribuciones conjugadas son aquellas para las cuales la función de dist. a posteriori va a pertenecer a la misma familia de distribuciones que la dist. a priori.

Ejemplos de familias de dist. conjugadas



Fuente

Estimadores puntuales Bayesianos

Una ventaja de este método es que se pueden definir de manera natural estimadores óptimos. Según que se considere como función de pérdida se obtendrán distintos estimadores puntuales.

Si consideramos una función de pérdidas $l(\theta, d)$, que es el costo de estimar al parámetro θ por el valor d, si $\hat{\theta} = \varphi(\underline{X})$, entonces la pérdida será una v.a. $l(\Theta, \varphi(\underline{X}))$.

La pérdida esperada es lo que se conoce como riesgo de Bayes. Esto significa que dada una distribución a priori $\pi(\theta)$, un estimador de Bayes será el que minimice $r(\varphi, \pi) = \mathbb{E}[l(\Theta, \varphi(\underline{X}))]$

Distintos estimadores según la función de riesgo

- Si consideramos una función de pérdida cuadrática: $\ell(\theta, d) = (\theta - d)^2$, el estimador de Bayes será es que minimice el ECM. ¿Quién era este estimador? $\varphi(\underline{X}) = \mathbb{E}[\Theta|\underline{X}]$ (con la esperanza tomada respecto de la distribución a posteriori.)

- Si consideramos la pérdida ℓ_1 dada por $\ell(\theta, d) = |\theta - d|$, el estimador de Bayes será la mediana de la distribución a posteriori de Θ condicionada a $\underline{X} = \underline{x}$.

- Muchas veces se utiliza lo que se conoce como máximo a posteriori (MAP) que se corresponde con la moda de la distribución a posteriori de $\Theta|\underline{X} = \underline{x}$, es decir el de el valor de θ que maximiza la dist. a posteriori. En general no es un estimador Bayesiano.

Cálculo de probabilidades

Para poder estimar probabilidades a partir de la distribución a posteriori de los parámetros aleatorios, usaremos la fórmula de probabilidad total. Por ejemplo, si Θ y X son v.a. continuas

$$\mathbb{P}(X \in A) = \int_A \int_{-\infty}^{\infty} f_{X|\Theta=\theta}(x) f_{\Theta|\underline{X}=\underline{x}}(\theta) d\theta dx$$

Bibliografía

Bibliografía

- Notas de estadística. Graciela Boente y Víctor Yohai
- “All of Statistics : A Concise Course in Statistical Inference”, Larry Wasserman, Springer-Verlag New York Inc.
- “Mathematical statistics with applications”, Mendenhall, W., Scheaffer, R. L., & Wackerly, D. D. Boston: Duxbury Press.