Estadística Bayesiana

Clase 16: Distribuciones a priori no informativas

Isabel Cristina Ramírez Guevara

Escuela de Estadística
Universidad Nacional de Colombia, Sede Medellín

Medellín, 21 de octubre de 2020

Análisis Bayesiano Objetivo

Una de las principales críticas que muchos estadísticos hacen al paradigma Bayesiano es la subjetividad asociada a la selección de la distribución a priori. Los defensores del enfoque Bayesiano afirman que, en realidad, la objetividad de cualquier análisis estadístico puede estar en entredicho, ya que cualquier análisis estadístico involucra decisiones subjetivas en la selección del modelo y en el análisis de los resultados. Sin embargo, es claro que, usando el enfoque Bayesiano, dos investigadores que tengan los mismos datos pueden obtener resultados diferentes a partir de distribuciones a priori distintas.

Análisis Bayesiano Objetivo

Existen, sin embargo, situaciones en las cuales no se dispone de información a priori, o no se desea usar con el fin de garantizar la reproducibilidad de los resultados. Bayes y Laplace han propuesto mecanísmos para obtener distribuciones a priori de este tipo. Utilizar distribuciones a priori no informativas es la estrategia más empleada en la mayor parte de las aplicaciones, pese a las críticas de muchos Bayesianos.

Análisis Bayesiano Objetivo

Existen, sin embargo, situaciones en las cuales no se dispone de información a priori, o no se desea usar con el fin de garantizar la reproducibilidad de los resultados. Bayes y Laplace han propuesto mecanísmos para obtener distribuciones a priori de este tipo. Utilizar distribuciones a priori no informativas es la estrategia más empleada en la mayor parte de las aplicaciones, pese a las críticas de muchos Bayesianos.

Históricamente, estas a priori se han denominado no informativas. Sin embargo, algunos opinan que esta denominación no es conveniente, ya que distribución a priori siempre contendrá información de algún tipo sobre los parámetros. Por lo tanto, recientemente muchos autores han denominado a este tipo de a prioris "objetivas". También se conocen como vagas, difusas, planas o de referencia.

Análisis Bayesiano Objetivo

En cualquier caso, la característica común a todos los procedimientos que serán usados a continuación es que el cálculo o la selección de la distribución a priori no requiere de conocimiento previo sobre los parámetros, evitando el paso (frecuentemente difícil) de traducir dicho conocimiento en una distribución de probabilidad.

Análisis Bayesiano Objetivo

En cualquier caso, la característica común a todos los procedimientos que serán usados a continuación es que el cálculo o la selección de la distribución a priori no requiere de conocimiento previo sobre los parámetros, evitando el paso (frecuentemente difícil) de traducir dicho conocimiento en una distribución de probabilidad.

El uso de distribuciones a priori no informativas buscan que ellas tengan un impacto mínimo sobre la distribución posterior del parámetro de interés y que sea relativamente plana con relación a la verosimilitud. Esto busca que sean los datos los que tengan un claro dominio en la distribución posterior, y por lo tanto en todas las inferencias que de ellas se obtengan.

Análisis Bayesiano Objetivo

Estas distribuciones no informativas se reúnen en dos grupos:

i) Propias: cuando la distribución de probabilidad integra a una constante finita, se dice que es propia.

ii) Impropias: una distribución a priori $p(\theta)$ es impropia si

$$\int_{\Theta} p(\theta) d\theta = \infty$$

i. *Una distribución a priori impropia puede terminar en una posterior impropia y por lo tanto no se podrán hacer inferencias.*

ii. *Una distribución a priori impropia puede llevar a una posterior propia.*

Análisis Bayesiano Objetivo

Ejemplo

Suponga que $y_1, \cdots, y_n|\theta$ *son variables distribuidas normal e independientemente con media* θ *y varianza* σ^2 *conocida. Suponga que* $p(\theta) \propto 1$ *es la distribución a priori uniforme (impropia) sobre los números reales. Encuentre la distribución posterior de* θ*.*

Análisis Bayesiano Objetivo

El postulado original de Bayes y Laplace fue tomar como medida a priori "no informativa"la uniforme $p(\theta) = 1$. Sin embargo este postulado tiene dos inconvenientes:

1. Si un investigador supone que $p(\theta) = 1$ otro parametriza su problema en términos del parámetro $\eta = \exp(\theta)$ y supone $p(\eta) = 1$, los resultados a que llegan ambos con los mismos datos y verosimilitud son inconsistentes a pesar de que la transformación es un uno-uno. La razón es por la fórmula de cambio de variable, si $p(\theta) = 1$

$$p^*(\eta) = p(\log \eta)|\eta = \frac{1}{\eta}$$

   Entonces, según el postulado de Bayes-Laplace el primer investigador es no informativo respecto a θ pero no de $\eta = \exp(\theta)$.

2. Si el espacio de parámetros es no acotado, entonces $p(\theta) = 1$ no integra uno, es decir, es impropia.

Para resolver el primer inconveniente, formulado por Fisher, comenzó Jeffreys una línea de trabajo para obtener medidas invariantes.

A priori de Jeffreys.

Esta a priori está basada en la matriz de información de Fisher.

Definición

Sea $p(x|\theta)$ la densidad de x dado θ. La información de Fisher es definida como:

$$\mathbf{I}(\theta) = -E\left[\frac{\partial^2 \log(p(x|\theta))}{\partial\theta^2}\right]$$

Si $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$, entonces:

$$\mathbf{I}(\boldsymbol{\theta}) = -E\left[\frac{[\partial^2(\log(p(x|\boldsymbol{\theta})))}{\partial\theta_i\partial\theta_j}\right]_{pxp},$$

en este caso $\mathbf{I}(\boldsymbol{\theta})$ es una matriz de dimensión pxp.

A priori de Jeffreys.

La distribución a priori de Jeffreys se define como

$$p(\boldsymbol{\theta}) \propto |\mathbf{I}(\boldsymbol{\theta})|^{1/2}$$

Esta distribución es no informativa y en muchos casos es impropia.

Ejemplo

Se tienen v.a. independientes $y_1, \cdots, y_n$ Bernoulli con parámetro θ. Encontraremos la distribución a priori de Jeffreys para θ.

Ejemplo

Se tiene $X \sim N(\mu, \sigma^2)$. Calculemos la distribución a priori de Jeffreys para (μ, σ).

A priori de Jeffreys.

La distribución a priori de Jeffreys tiene la propiedad de invarianza, ya que para cualquier otra transformación uno a uno sigue siendo no informativa. Esto surge de la relación.

$$\mathbf{I}(\theta) = \mathbf{I}(\psi(\theta)) \left(\frac{\partial \psi(\theta)}{\partial \theta} \right)^2$$

donde $\psi(\theta)$ es una transformación uno a uno de θ. Así

$$(\mathbf{I}(\theta))^{1/2} = (\mathbf{I}(\psi(\theta)))^{1/2} \left| \frac{\partial \psi \theta}{\partial \theta} \right|$$

$$p(\theta) = p(\psi(\theta)) \left| \frac{\partial \psi(\theta)}{\partial \theta} \right|$$

Note que $\left| \frac{\partial \psi(\theta)}{\partial \theta} \right|$ es el valor absoluto del jacobiano de la transformación de θ a $\psi(\theta)$. Así

$$(\mathbf{I}(\theta))^{1/2} d\theta = (\mathbf{I}(\psi))^{1/2} d\psi$$

A priori de Jeffreys.

Ejemplo

Suponga $X \sim N(\mu, 1)$ *y* $\psi(\mu) = e^{\mu}$. *Encuentre la apriori de Jeffreys para* $\psi(\mu)$.

Ejemplo

Suponga $x_1, \cdots, x_n$ *son variables aleatorias i.i.d. Bernoulli*(θ) *y sea* $y = \sum x_i$. *Considere la transformación* $\tau = \frac{\theta}{(1-\theta)}$, $\tau =$ *odds ratio.*

a) *Determine la distribución de probabilidad de* y. *Escribala en función de* τ.

b) *Encuentre la distribución a priori de Jeffreys para* τ.

c) *Muestre que la distribución a priori de Jeffreys para* θ *es invariante con respecto a la transformación* $\tau = \frac{\theta}{1-\theta}$.

Cantidad pivotal.

Para modelos donde los parámetros son de localización y escala se utiliza el criterio de la cantidad pivotal para establecer distribuciones a priori no informativas.

- **Parámetros de localización:** Si la densidad de y es tal que $p(y-\theta|\theta)$ no depende de θ, digamos $f(u)$, donde $u = y - \theta$, diremos que $u = y - \theta$ es una cantidad pivotal y que θ es un parámetro de localización. En este caso, es razonable que una a priori objetiva para θ diera como resultado $f(u)$ para la distribución posterior $p(y-\theta|y)$. Esto implica que para la distribución posterior, $u = y - \theta$ también es una cantidad pivotal. Por lo tanto, $p(y-\theta|y) \propto p(\theta)p(y-\theta|\theta)p(\theta) \propto$ cte en el intervalo $(-\infty, \infty)$. Esto implica que la densidad a priori noinformativa es uniforme en θ.

Cantidad pivotal.

- **Parámetros de escala:** Si la densidad de y es tal que $p(\frac{y}{\theta}|\theta)$ no depende de θ, digamos $g(u)$, donde $u = \frac{y}{\theta}$, diremos $u = \frac{y}{\theta}$ es una cantidad pivotal y θ es un parámetro de escala. En tal caso, es razonable que una a priori no informativa diera como resultado de la posterior $p(\frac{y}{\theta}|y)$ una función $g(u)$ con $u = \frac{y}{\theta}$. Transformando variables,

$$p(y|\theta) = p(u|\theta)\left|\frac{du}{dy}\right|$$

$$p(y|\theta) = \frac{1}{\theta}p(u|\theta)$$

y en forma similar,

$$p(\theta|y) = p(u|y)\left|\frac{du}{d\theta}\right|$$

$$p(\theta|y) = \frac{y}{\theta^2}p(u|y)$$

Cantidad pivotal.

Haciendo $p(u|\theta) = p(u|y) = g(u)$, llegamos a que la a priori de referencia debe ser $p(\theta) \propto \frac{1}{\theta}$ o en forma equivalente $p(\log\theta) \propto 1$.