Estadística Bayesiana

Clase 2: Modelos uniparamétricos

Isabel Cristina Ramírez Guevara

Escuela de Estadística
Universidad Nacional de Colombia, Sede Medellín

Medellín, 6 de agosto de 2020

Modelos uniparamétricos

Vamos a ver el análisis de datos desde un enfoque Bayesiano a los modelos estadísticos que están determinados únicamente por un parámetro. En particular, vamos a estudiar los siguientes modelos:

a) Modelo Binomial
b) Modelo Poisson.
c) Modelo Normal con varianza conocida.
d) Modelo Normal con media conocida.

Modelo Binomial

El objetivo es estimar la proporción desconocida de un conjunto de pruebas Bernoulli, donde los datos $y_1, \cdots, y_n$ son una secuencia de ceros y unos, donde 1 implica un éxito y 0 un fracaso. Esta es secuencia de n intentos donde cada ocurrencia tiene dos posibilidades: éxito o fracaso.
Tenemos: $X=$ número de éxitos en n intentos, $X \sim$ Binomial (n, θ). Por lo tanto la verosimilitud es:

$$p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Modelo Binomial

El objetivo es estimar la proporción desconocida de un conjunto de pruebas Bernoulli, donde los datos $y_1, \cdots, y_n$ son una secuencia de ceros y unos, donde 1 implica un éxito y 0 un fracaso. Esta es secuencia de n intentos donde cada ocurrencia tiene dos posibilidades: éxito o fracaso.
Tenemos: $X=$ número de éxitos en n intentos, $X \sim$ Binomial (n, θ). Por lo tanto la verosimilitud es:

$$p(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

Si la distribución a priori es $\theta \sim$ U(0,1), la distribución posterior de θ es

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

Modelo Binomial

El objetivo es estimar la proporción desconocida de un conjunto de pruebas Bernoulli, donde los datos $y_1, \cdots, y_n$ son una secuencia de ceros y unos, donde 1 implica un éxito y 0 un fracaso. Esta es secuencia de n intentos donde cada ocurrencia tiene dos posibilidades: éxito o fracaso.
Tenemos: $X=$ número de éxitos en n intentos, $X \sim$ Binomial (n, θ). Por lo tanto la verosimilitud es:

$$p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Si la distribución a priori es $\theta \sim$ U(0,1), la distribución posterior de θ es

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

$$p(\theta|x) \propto \theta^x (1-\theta)^{n-x}(1)$$

Modelo Binomial

El objetivo es estimar la proporción desconocida de un conjunto de pruebas Bernoulli, donde los datos $y_1, \cdots, y_n$ son una secuencia de ceros y unos, donde 1 implica un éxito y 0 un fracaso. Esta es secuencia de n intentos donde cada ocurrencia tiene dos posibilidades: éxito o fracaso.
Tenemos: $X=$ número de éxitos en n intentos, $X \sim$ Binomial (n, θ). Por lo tanto la verosimilitud es:

$$p(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

Si la distribución a priori es $\theta \sim$ U(0,1), la distribución posterior de θ es

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

$$p(\theta|x) \propto \theta^x(1-\theta)^{n-x}(1)$$

$$p(\theta|x) \propto \theta^{x+1-1}(1-\theta)^{n-x+1-1}$$

Modelo Binomial

El objetivo es estimar la proporción desconocida de un conjunto de pruebas Bernoulli, donde los datos $y_1, \cdots, y_n$ son una secuencia de ceros y unos, donde 1 implica un éxito y 0 un fracaso. Esta es secuencia de n intentos donde cada ocurrencia tiene dos posibilidades: éxito o fracaso.
Tenemos: $X=$ número de éxitos en n intentos, $X \sim$ Binomial (n, θ). Por lo tanto la verosimilitud es:

$$p(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

Si la distribución a priori es $\theta \sim$ U(0,1), la distribución posterior de θ es

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

$$p(\theta|x) \propto \theta^x(1-\theta)^{n-x}(1)$$

$$p(\theta|x) \propto \theta^{x+1-1}(1-\theta)^{n-x+1-1}$$

$$\theta|x \sim \text{Beta}(x+1, n-x+1)$$

Modelo Binomial

Ahora suponga que $\theta \sim \text{Beta}(\alpha, \beta)$ entonces:

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Modelo Binomial

Ahora suponga que $\theta \sim \text{Beta}(\alpha, \beta)$ entonces:

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

La distribución posterior de θ está dada por:

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

Modelo Binomial

Ahora suponga que $\theta \sim \text{Beta}(\alpha, \beta)$ entonces:

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

La distribución posterior de θ está dada por:

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

$$p(\theta|x) \propto \theta^{x}(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Modelo Binomial

Ahora suponga que $\theta \sim \text{Beta}(\alpha, \beta)$ entonces:

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

La distribución posterior de θ está dada por:

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

$$p(\theta|x) \propto \theta^{x}(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$p(\theta|x) \propto \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

Modelo Binomial

Ahora suponga que $\theta \sim \text{Beta}(\alpha, \beta)$ entonces:

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

La distribución posterior de θ está dada por:

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

$$p(\theta|x) \propto \theta^{x}(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$p(\theta|x) \propto \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

$$\theta|x \sim \text{Beta}(x+\alpha, n-x+\beta),$$

por lo tanto la distribución Beta es una familia conjugada para la verosimilitud Binomial.

Modelo Binomial

Definición

Si $\mathcal{F}$ es una clase de disribuciones muestrales $p(y|\theta)$ y $\mathcal{P}$ es una clase de distribuciones a priori para θ, entonces la clase $\mathcal{P}$ es conjugada para $\mathcal{F}$ si $p(\theta|y) \in \mathcal{P}$ para todo $p(\cdot|\theta) \in \mathcal{F}$ y $p(\cdot) \in \mathcal{P}$.

De la distribución posterior de θ cuando la distribución a priori de $\theta \sim$ Beta(α, β) podemos decir que:

$$E(\theta|x) = \frac{x+\alpha}{\alpha+\beta+n} = \frac{\alpha}{\alpha+\beta}\left(\frac{\alpha+\beta}{\alpha+\beta+n}\right) + \frac{x}{n}\left(\frac{n}{\alpha+\beta+n}\right)$$

$$= \begin{matrix}\text{Media}\\ \text{Apriori}\end{matrix} \quad (W_1) \quad + \quad \begin{matrix}\text{Media}\\ \text{muestral}\end{matrix} \quad (W_2)$$

Modelo Binomial

Si $X \sim$ *Binomial* (n, θ) *y* $\theta \sim$*Beta*(α, β)*, entonces*
$\theta | x \sim$*Beta*$(x + \alpha, n - x + \beta)$

Modelo Binomial

Si X$\sim$ Binomial (n, θ) y θ $\sim$Beta(α, β), entonces
$\theta|x$ $\sim$Beta$(x + \alpha, n - x + \beta)$

Ejemplo

Se considera la estimación de la probabilidad de que nazca una niña dada la condición de placenta previa. En un estudio inicial llevado a cabo en Alemania, se encontró que de un total de 980 nacimientos con la condición de placenta previa, 437 eran niñas. ¿Cuánta evidencia proporcionan estos datos sobre la hipótesis de que la proporción de nacimientos hembras en la población de placenta previa es menor que la proporción 0.485 de niñas en la población general?.

Modelo Poisson

Es utilizado en datos de conteo. El parámetro θ representa la tasa promedio de ocurrencia del evento. Para un vector de n observaciones, $\mathbf{y}= (y_1, \cdots, y_n)$, independientes, la verosimilitud es:

$$p(\mathbf{y}|\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{y_i}}{y_i!} = \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!}$$

Modelo Poisson

Es utilizado en datos de conteo. El parámetro θ representa la tasa promedio de ocurrencia del evento. Para un vector de n observaciones, $\mathbf{y}= (y_1, \cdots, y_n)$, independientes, la verosimilitud es:

$$p(\mathbf{y}|\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{y_i}}{y_i!} = \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!}$$

Si $\theta \sim \text{Gamma}(\alpha, \beta)$ entonces:

$$p(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$$

Modelo Poisson

La distribución posterior es:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta)$$

Modelo Poisson

La distribución posterior es:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta)$$

$$p(\theta|\mathbf{y}) \propto \theta^{\sum_{i=1}^{n} y_i} e^{-n\theta} \theta^{\alpha-1} e^{-\beta\theta}$$

Modelo Poisson

La distribución posterior es:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta)$$

$$p(\theta|\mathbf{y}) \propto \theta^{\sum_{i=1}^{n} y_i} e^{-n\theta} \theta^{\alpha-1} e^{-\beta\theta}$$

$$p(\theta|\mathbf{y}) \propto \theta^{\sum_{i=1}^{n} y_i + \alpha - 1} e^{-\theta(n+\beta)}$$

Modelo Poisson

La distribución posterior es:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta)$$

$$p(\theta|\mathbf{y}) \propto \theta^{\sum_{i=1}^{n} y_i} e^{-n\theta} \theta^{\alpha-1} e^{-\beta\theta}$$

$$p(\theta|\mathbf{y}) \propto \theta^{\sum_{i=1}^{n} y_i+\alpha-1} e^{-\theta(n+\beta)}$$

$$\theta|\mathbf{y} \sim \text{Gamma}\left(\sum_{i=1}^{n} y_i + \alpha, n+\beta\right),$$

por lo tanto la Gamma es la familia conjugada para el parámetro θ de la distribución Poisson. Se tiene:

Si $Y \sim Poisson(\theta)$ y $\theta \sim Gamma(\alpha, \beta)$ entonces
$\theta|\mathbf{y} \sim Gamma\left(\sum_{i=1}^{n} y_i + \alpha, n+\beta\right)$

Modelo Poisson

$$E(\theta|\mathbf{y}) = \frac{\sum_{i=1}^{n} y_i + \alpha}{n+\beta} = \frac{\sum_{i=1}^{n} y_i}{n}\left(\frac{n}{n+\beta}\right) + \frac{\alpha}{\beta}\left(\frac{\beta}{n+\beta}\right)$$
$$= \begin{array}{c}\text{Media}\\ \text{muestral}\end{array}(W_1) + \begin{array}{c}\text{Media}\\ \text{Apriori}\end{array}(W_2)$$

Modelo Poisson

$$E(\theta|\mathbf{y}) = \frac{\sum_{i=1}^{n} y_i + \alpha}{n+\beta} = \frac{\sum_{i=1}^{n} y_i}{n}\left(\frac{n}{n+\beta}\right) + \frac{\alpha}{\beta}\left(\frac{\beta}{n+\beta}\right)$$

$$= \begin{array}{c}\text{Media}\\ \text{muestral}\end{array} (W_1) + \begin{array}{c}\text{Media}\\ \text{Apriori}\end{array} (W_2)$$

Ejemplo

Suponga que $X_i|\theta \sim$ Poisson(θ) y $\theta \sim$ Gamma(α, β). Se tiene que el número de accidentes aéreos fatales durante los años 1976 y 1985 son: 24, 25, 31, 31, 22, 21, 26, 20, 16, 22. Si $\tilde{X}$ es el número de accidentes fatales en 1986 y $\alpha = 101$ y $\beta = 5$ encuentre:

a) *La distribución posterior de θ.*

b) *El valor esperado y la varianza de $\tilde{X}$.*

Modelo Poisson en función de la tasa y la exposición

En algunas situaciones el número de casos esperados es el producto de la tasa θ y un valor ε_i (conocido) que equivale a la exposición, se pueden modelar como:

$$y_i \sim \text{Poisson}(\mu_i) \qquad \mu_i = \varepsilon_i \theta$$

θ: parámetro de interés.
ε_i: cantidad conocida.

Utilizando una distribución a priori Gamma (α, β) para θ, la distribución posterior es proporcional a:

$$\begin{aligned} p(\theta|\mathbf{y}) &\propto \prod_{i=1}^{n} \left(\theta^{y_i} e^{-\varepsilon_i \theta}\right) \theta^{\alpha-1} e^{-\beta\theta} \\ &= \theta^{\sum_{i=1}^{n} y_i + \alpha - 1} e^{-\theta\left(\sum_{i=1}^{n} \varepsilon_i + \beta\right)} \end{aligned}$$

$$\theta|\mathbf{y} \sim \text{Gamma}\left(\sum_{i=1}^{n} y_i + \alpha, \sum_{i=1}^{n} \varepsilon_i + \beta\right)$$

Ejemplo

Tres personas mueren de asma cada año de una población de 200000 habitantes. Estudios sobre mortalidad de asma alrededor del mundo muestran que las razones de mortalidad mayores de 1.5 por 100000 habitantes son raras en países occidentales, con razones de mortalidad por asma alrededor de 0.6 por 100000 habitantes. Encuentre la distribución posterior de θ: razón de mortalidad por asma en la población por cada 100000 habitantes por año y concluya.

*(**Distribución posterior con datos adicionales.**) Suponga que se tiene información de 10 años y se encuentra que la razón de mortalidad sigue siendo 1.5 por 100000 habitantes en la población de estudio. Se tiene además que el número de habitantes es constante: 200000. Encuentra la nuevamente la distribución posterior de θ y concluya.*