Modelos de Regresión

– Enfoque Bayesiano –

Juan Carlos Martínez-Ovando

Maestría en Ciencia de Datos, ITAM

Fundamentos de Estadística
10 de noviembre de 2015

Contents

Motivación

Modelo de regresión

Recordemos, que tenemos el modelo de regresión lineal

$$y_i|\boldsymbol{x}_i, \beta, \sigma^2 \sim \mathrm{N}(y_i|\boldsymbol{x}_i'\beta, \sigma^2) \tag{1}$$

para $i = 1, \ldots, n$.

- Supuesto de simetría estocástica alrededor de las y_i's se define, formalmente, en términos de la distribución conjunta de las $(y_i, \boldsymbol{x}_i)$'s.
- la distribución de $\boldsymbol{x}_i$'s es indiferente para el análisis, **ya que nos interesa la dependencia de y_i respecto a $\boldsymbol{x}_i$**.
- Los parámetros del modelo son (β, σ^2).

Problema inferencial

- Desde el punto de vista bayesiano, debemos caracterizar la distribución final de (β, σ^2).

Motivación

Modelo de regresión

Recordemos, que tenemos el modelo de regresión lineal

$$y_i|\boldsymbol{x}_i, \beta, \sigma^2 \sim \mathrm{N}(y_i|\boldsymbol{x}_i'\beta, \sigma^2) \tag{1}$$

para $i = 1, \ldots, n$.

- Supuesto de simetría estocástica alrededor de las y_i's se define, formalmente, en términos de la distribución conjunta de las $(y_i, \boldsymbol{x}_i)$'s.
- la distribución de $\boldsymbol{x}_i$'s es indiferente para el análisis, **ya que nos interesa la dependencia de y_i respecto a $\boldsymbol{x}_i$.**
- Los parámetros del modelo son (β, σ^2).

Problema inferencial

- Desde el punto de vista bayesiano, debemos caracterizar la distribución final de (β, σ^2).

Motivación

Verosimilitud

La función de verosimilitud del modelo es

$$\mathbb{P}((y_i, \boldsymbol{x}_i) : i = 1, \ldots, n) = \mathrm{N}_n(\boldsymbol{y}|\boldsymbol{X}'\beta, \sigma^2 \boldsymbol{I}), \tag{2}$$

- $\boldsymbol{y}$ es un vector de dimensión $(n \times 1)$
- $\boldsymbol{X}$ es una matriz de dimensión $(n \times p)$ (p es el número de covariables). Cada renglón de $\boldsymbol{X}$ es $\boldsymbol{x}_i'$.
- $\boldsymbol{I}$ es una matriz de covarianza de dimensión $(n \times n)$

Parámetros:

- β, un vector de dimensión $(p \times 1)$
- σ^2, un valor escalar positivo.
- Esto es porque que se supone que la relaci n entre las parejas $(y_i, \boldsymbol{x}_i)$ es homogénea entre observaciones. **Caso de regresión homoscedástica.**

Motivación

Condición

Para estimar el caso homoscedástico (enfoque bayesiano o frecuentista) es fundamental mantener la siguientes condiciones:

- El número de observaciones n sea mayor que el número de covariables p.
- Las covariables que definen a $\boldsymbol{x}_i$ sean linealmente independientes entre sí.

Inferencia bayesiana

La inferencia bayesiana sobre (β, σ^2) se basa en la siguiente distribución conjunta,

$$\begin{aligned} \mathbb{P}((y_1, \boldsymbol{x_1}), \ldots, (y_n, \boldsymbol{x_n}), \beta, \sigma^2) &= \mathbb{P}((y_1, \boldsymbol{x_1}), \ldots, (y_n, \boldsymbol{x_n}) | \beta, \sigma^2) \Pi(\beta, \sigma^2) \\ &= \mathrm{N}_n(\boldsymbol{y} | \boldsymbol{X}'\beta, \sigma^2 \boldsymbol{I}) \Pi(\beta, \sigma^2) \end{aligned}$$

Pregunta

Cómo elegir $\Pi(\beta, \sigma^2)$?

- Opción 1: Distribución inicial conjugada.
- Opción 2: Distribución inicial no informativa.

Motivación

Condición

Para estimar el caso homoscedástico (enfoque bayesiano o frecuentista) es fundamental mantener la siguientes condiciones:

- El número de observaciones n sea mayor que el número de covariables p.
- Las covariables que definen a $\boldsymbol{x}_i$ sean linealmente independientes entre sí.

Inferencia bayesiana

La inferencia bayesiana sobre (β, σ^2) se basa en la siguiente distribución conjunta,

$$\begin{aligned}\mathbb{P}((y_1, \boldsymbol{x_1}), \ldots, (y_n, \boldsymbol{x_n}), \beta, \sigma^2) &= \mathbb{P}((y_1, \boldsymbol{x_1}), \ldots, (y_n, \boldsymbol{x_n}) | \beta, \sigma^2) \Pi(\beta, \sigma^2) \\ &= \mathrm{N}_n(\boldsymbol{y} | \boldsymbol{X}' \beta, \sigma^2 \boldsymbol{I}) \Pi(\beta, \sigma^2)\end{aligned}$$

Pregunta

Cómo elegir $\Pi(\beta, \sigma^2)$?

- Opción 1: Distribución inicial conjugada.
- Opción 2: Distribución inicial no informativa.

Motivación

Condición

Para estimar el caso homoscedástico (enfoque bayesiano o frecuentista) es fundamental mantener la siguientes condiciones:

- El número de observaciones n sea mayor que el número de covariables p.
- Las covariables que definen a $\boldsymbol{x}_i$ sean linealmente independientes entre sí.

Inferencia bayesiana

La inferencia bayesiana sobre (β, σ^2) se basa en la siguiente distribución conjunta,

$$
\begin{aligned}
\mathbb{P}((y_1, \boldsymbol{x_1}), \ldots, (y_n, \boldsymbol{x_n}), \beta, \sigma^2) &= \mathbb{P}((y_1, \boldsymbol{x_1}), \ldots, (y_n, \boldsymbol{x_n}) | \beta, \sigma^2) \Pi(\beta, \sigma^2) \\
&= \mathrm{N}_n(\boldsymbol{y} | \boldsymbol{X}' \beta, \sigma^2 \boldsymbol{I}) \Pi(\beta, \sigma^2)
\end{aligned}
$$

Pregunta

Cómo elegir $\Pi(\beta, \sigma^2)$?

- Opción 1: Distribución inicial conjugada.
- Opción 2: Distribución inicial no informativa.

Análisis bayesiano conjugado

Conjugacidad

La idea detr s del an lisis conjugado para un modelo general intercambiable $\mathbb{P}(y|\theta)$ y $\Pi(\theta)$ consiste en:

- Estudiar la estructura matem tica de $\mathbb{P}(y|\theta)$ en función de θ,
- Elegir la estructura de $\Pi(\theta)$ con base en dicha estrutura a manera que $\Pi(\theta|y)$ y $\Pi(\theta)$ tengan la "misma" forma funcional, salvo sus parámetros.

Teorema de Bayes

Recordemos que el Teorema de Bayes nos indica que

$$\Pi(\theta|y) \propto \mathbb{P}(y|\theta) \times \Pi(\theta). \tag{3}$$

Análisis bayesiano conjugado

Conjugacidad

La idea detr s del an lisis conjugado para un modelo general intercambiable $\mathbb{P}(y|\theta)$ y $\Pi(\theta)$ consiste en:

- Estudiar la estructura matem tica de $\mathbb{P}(y|\theta)$ en función de θ,
- Elegir la estructura de $\Pi(\theta)$ con base en dicha estrutura a manera que $\Pi(\theta|y)$ y $\Pi(\theta)$ tengan la "misma" forma funcional, salvo sus parámetros.

Teorema de Bayes

Recordemos que el Teorema de Bayes nos indica que

$$\Pi(\theta|y) \propto \mathbb{P}(y|\theta) \times \Pi(\theta). \tag{3}$$

Análisis bayesiano conjugado

Caso normal multivariado

Por simplicidad, consideraremos el parámetro de precisión, λ, en lugar de la varianza, σ^2, en el modelo normal. Así,

$$\boldsymbol{y}|\theta,\lambda \sim \mathrm{N}_n(y|\theta,\lambda^2\boldsymbol{I}) \tag{4}$$

La función de densidad de $\boldsymbol{y}$ estará dada por

$$f(\boldsymbol{y}|\theta,\lambda) \propto |\boldsymbol{\lambda}|^{1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{y}-\theta)'\boldsymbol{\lambda}(\boldsymbol{y}-\theta)\right\} \tag{5}$$

donde $\boldsymbol{\lambda} = \lambda\boldsymbol{I}$.

Enfoques

- Viendo a $f(\boldsymbol{y}|\theta,\lambda)$ como función de θ, condicional en λ, se asemeja al kernel de la distriución normal.
- Viendo a $f(\boldsymbol{y}|\theta,\lambda)$ como función de λ se asemeja al kernel de una distribución gamma.

Análisis bayesiano conjugado

Caso normal multivariado

Por simplicidad, consideraremos el parámetro de precisión, λ, en lugar de la varianza, σ^2, en el modelo normal. Así,

$$\boldsymbol{y}|\theta, \lambda \sim \mathrm{N}_n(y|\theta, \lambda^2 \boldsymbol{I}) \tag{4}$$

La función de densidad de $\boldsymbol{y}$ estará dada por

$$f(\boldsymbol{y}|\theta, \lambda) \propto |\boldsymbol{\lambda}|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y}-\theta)'\boldsymbol{\lambda}(\boldsymbol{y}-\theta)\right\} \tag{5}$$

donde $\boldsymbol{\lambda} = \lambda \boldsymbol{I}$.

Enfoques

- Viendo a $f(\boldsymbol{y}|\theta, \lambda)$ como función de θ, condicional en λ, se asemeja al kernel de la distriución normal.
- Viendo a $f(\boldsymbol{y}|\theta, \lambda)$ como función de λ se asemeja al kernel de una distribución gamma.

Análisis bayesiano conjugado

Distribución inicial

Los enfoques inducen la siguiente estructura

$$\Pi(\theta, \lambda) = \Pi(\theta|\lambda)\Pi(\lambda) \tag{6}$$

donde

$$\Pi(\theta|\lambda) = \mathrm{N}(\theta|m_0, \lambda S_0) \tag{7}$$

y

$$\Pi(\lambda) = \mathrm{Ga}(\lambda|a_0, b_0). \tag{8}$$

Esta distribución se conoce como normal-gamma.
La distribución normal-gamma es conjugada para el modelo normal multivariado con media y varianza desconocidos.

Análisis bayesiano conjugado

Distribución final

Aplicando el teorema de Bayes, se obtiene que la distribución final para (β, λ) es normal-gamma también, dada por

$$\Pi(\theta, \lambda|\text{datos}) \propto \mathrm{N}(\theta|m_1, \lambda S_1)\,\mathrm{Ga}(\lambda|a_1, b_1), \tag{9}$$

con

$$\begin{aligned} S_1 &= S_0 + (\boldsymbol{y} - \bar{y})(\boldsymbol{y} - \bar{y})' \\ m_1 &= S_1^{-1}(m_0 S_0 + \bar{y}) \\ a_1 &= a_0 + 1/2n \\ b_1 &= b_0 + 1/2(\boldsymbol{y} - \bar{y})'(\boldsymbol{y} - \bar{y}). \end{aligned}$$

Análisis bayesiano conjugado

Caso normal multivariado

Por simplicidad, consideraremos el parámetro de precisión, λ, en lugar de la varianza, σ^2, en el modelo normal. Así,

$$\boldsymbol{y}|\boldsymbol{X}|\theta, \lambda \sim \mathrm{N}_n(\boldsymbol{y}|\boldsymbol{X}'\theta, \lambda^2 \boldsymbol{I}) \tag{10}$$

La función de densidad de $\boldsymbol{y}$ estará dada por

$$f(\boldsymbol{y}|\theta, \lambda) \propto |\boldsymbol{\lambda}|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}'\theta)'\boldsymbol{\lambda}(\boldsymbol{y} - \boldsymbol{X}'\theta)\right\} \tag{11}$$

donde $\boldsymbol{\lambda} = \lambda \boldsymbol{I}$.

Enfoques

- Viendo a $f(\boldsymbol{y}|\theta, \lambda)$ como función de θ, condicional en λ, se asemeja al kernel de la distriución normal p-variada.
- Viendo a $f(\boldsymbol{y}|\theta, \lambda)$ como función de λ se asemeja al kernel de una distribución gamma.

Análisis bayesiano conjugado

Caso normal multivariado

Por simplicidad, consideraremos el parámetro de precisión, λ, en lugar de la varianza, σ^2, en el modelo normal. Así,

$$\boldsymbol{y}|\boldsymbol{X}|\theta, \lambda \sim \mathrm{N}_n(\boldsymbol{y}|\boldsymbol{X}'\theta, \lambda^2 \boldsymbol{I}) \tag{10}$$

La función de densidad de $\boldsymbol{y}$ estará dada por

$$f(\boldsymbol{y}|\theta, \lambda) \propto |\boldsymbol{\lambda}|^{1/2} \exp\left\{ -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}'\theta)'\boldsymbol{\lambda}(\boldsymbol{y} - \boldsymbol{X}'\theta) \right\} \tag{11}$$

donde $\boldsymbol{\lambda} = \lambda \boldsymbol{I}$.

Enfoques

- Viendo a $f(\boldsymbol{y}|\theta, \lambda)$ como función de θ, condicional en λ, se asemeja al kernel de la distriución normal p-variada.
- Viendo a $f(\boldsymbol{y}|\theta, \lambda)$ como función de λ se asemeja al kernel de una distribución gamma.

Análisis bayesiano conjugado

Distribución inicial

Los enfoques inducen la siguiente estructura

$$\Pi(\theta, \lambda) = \Pi(\theta|\lambda)\Pi(\lambda) \tag{12}$$

donde

$$\Pi(\theta|\lambda) = \mathrm{N}(\theta|m_0, \lambda S_0) \tag{13}$$

y

$$\Pi(\lambda) = \mathrm{Ga}(\lambda|a_0, b_0). \tag{14}$$

En este caso,

- m_0 es un vector p-dimensional, y
- S_0 es una matriz de covarianzas p-dimensional.

Análisis bayesiano conjugado

Distribución final

Aplicando el teorema de Bayes, se obtiene que la distribución final para (β, λ) es normal-gamma también, dada por

$$\Pi(\theta, \lambda|\text{datos}) \propto \mathrm{N}(\theta|m_1, \lambda S_1)\,\mathrm{Ga}(\lambda|a_1, b_1), \tag{15}$$

con

$$\begin{aligned}
S_1 &= S_0 + \boldsymbol{X}'\boldsymbol{X} \\
m_1 &= S_1^{-1}(S_0 m_0 + \boldsymbol{X}'\boldsymbol{y}) \\
a_1 &= a_0 + 1/2n \\
b_1 &= b_0 + 1/2(\boldsymbol{y} - \boldsymbol{X}'m_1)'\boldsymbol{y} + 1/2(m_0 - m_1)'S_0 m_0.
\end{aligned}$$