Bayesian Inference

Chapter 6: Regression and hierarchical models

M. Concepción Ausín
Department of Statistics
Universidad Carlos III de Madrid

Master in Business Administration and Quantitative Methods
Master in Mathematical Engineering

Introducción

- Abordaremos desde el punto de vista Bayesiano los principales modelos lineales normales y algunos modelos lineales generalizados.
- Introduciremos los modelos jerárquicos Bayesianos y analizaremos algunos ejemplos habituales.
- Los modelos más sencillos permiten utilizar distribuciones conjugadas, pero en la mayoría de los casos tendremos que hacer uso de los métodos MCMC para aproximar la distribución a posteriori de los parámetros del modelo.
- Incluso en modelos sencillos se puede hacer uso de los métodos MCMC si se quiere evitar hacer cálculos para obtener las distribuciones a posteriori analíticamente.

Contenidos

Modelos lineales normales

Un modelo lineal normal tiene la siguiente estructura:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon},$$

donde $\mathbf{y} = (y_1, \ldots, y_n)'$ es un conjunto de datos observados, $\mathbf{X}$ es una matriz conocida de dimensión $n \times k$, denominada matriz de diseño, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ es el conjunto de parámetros y

$$\epsilon \sim \mathcal{N}\left(\mathbf{0}_k, \frac{1}{\phi}\mathbf{I}_k\right).$$

Son ejemplos de modelos lineales normales el modelo de regresión lineal simple, de regresión multivariante y el modelo ANOVA.

Para este tipo de modelos existen distribuciones conjugadas a priori basadas en la normal-gamma multivariante.

Regresión lineal simple

Consideramos el caso más simple de modelos lineal normal que es el modelo de regresión lineal simple:

$$y_i = \alpha + \beta x_i + \epsilon_i, \qquad \text{para } i = 1, \ldots, n,$$

donde $\epsilon_i \sim \mathcal{N}(0, 1/\phi)$.

Para este modelo existe una distribución normal-gamma a priori conjugada:

$$(\alpha, \beta) \sim \mathcal{N}\left(0, \frac{1}{\phi} V\right)$$
$$\phi \sim \mathcal{G}\left(\frac{a}{2}, \frac{b}{2}\right)$$

Regresión lineal simple

En el límite, la distribución a priori impropia es

$$\pi(\alpha,\beta,\phi) \propto \frac{1}{\phi},$$

que da lugar a las siguientes distribuciones a posteriori:

$$\left.\begin{array}{c}\alpha\\ \beta\end{array}\right| \mathbf{y},\phi \sim \mathcal{N}\left(\left(\begin{array}{c}\hat{\alpha}\\ \hat{\beta}\end{array}\right), \frac{1}{\phi n s_{xx}}\left(\begin{array}{cc}\sum_{i=1}^{n} x_i^2 & -n\bar{x}\\ -n\bar{x} & n\end{array}\right)\right)$$

$$\phi \mid \mathbf{y} \sim \mathcal{G}\left(\frac{n-2}{2}, \frac{s_{yy}\left(1-r^2\right)}{2}\right)$$

donde,

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \quad \hat{\beta} = \frac{s_{xy}}{s_{xx}}, \quad s_{xx} = \sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2, \qquad s_{yy} = \sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2$$

$$s_{xy} = \sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right), \qquad r = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}, \qquad \hat{\sigma}^2 = \frac{s_{yy}\left(1-r^2\right)}{n-2}.$$

Regresión lineal simple

De aquí, se obtiene que las distrbuciones marginales a posteriori de α y β son:

$$\alpha \mid \mathbf{y} \sim \mathcal{T}\left(\hat{\alpha}, \hat{\sigma}^2 \frac{\sum_{i=1}^n x_i^2}{n s_{xx}}, n-2\right),$$
$$\beta \mid \mathbf{y} \sim \mathcal{T}\left(\hat{\beta}, \frac{\hat{\sigma}^2}{s_{xx}}, n-2\right).$$

Y así, por ejemplo un intervalo creíble al $(1-\alpha)\%$ para β es:

$$\hat{\beta} \pm \frac{\hat{\sigma}}{\sqrt{s_{xx}}} t_{n-2,\alpha/2},$$

que coincide con el intervalo clásico.

Regresión lineal simple

Ejemplo: El archivo de datos `cats` en la librería `MASS` de `R` incluye, en las variables `Bwt` y `Hwt`, el peso del cuerpo en kilos y del corazón en gramos de 144 gatos domésticos. Suponer un modelo de regresión lineal simple para explicar el peso del corazón en función del peso del cuerpo.

1. Mediante simulación Monte Carlo, obtener una muestra de la distribución a posteriori de los parámetros y aproximar los intervalos creíbles al 95%. Compararlos con los intervalos teóricos.
2. Realizar una análisis equivalete al anterior con un algoritmo MCMC.

Modelos lineales generalizados

Los modelos lineales generalizados extienden los modelos lineales normales permitiendo errores no normales y relaciones no lineales entre $\mathbf{x}$ e $\mathbf{y}$.

En este tipo de modelos, la $\mu = E[Y \mid \mathbf{x}]$ no es una función lineal de $\mathbf{x}$, sino que existe una relación a través de una función biyectiva, $g(\cdot)$, de modo que $g(\mu) = \mathbf{x}\boldsymbol{\theta}$.

Además, la densidad $f(y \mid \mathbf{x})$ no es necesariamente normal, sino que puede ser cualquier familia exponencial (Poisson, binomial, gamma, etc.)

Para estos modelos no existen distribuciones conjugadas y se hace necesario el uso de métodos MCMC.

Modelos lineales generalizados

Regresión logística

Se usa para estimar la probabilidad de que ocurra un suceso dadas unas covariables:

$$Y_i \mid p_i \sim \mathcal{B}in(n_i, p_i)$$
$$\log \frac{p_i}{1-p_i} = \mathbf{x}_i \boldsymbol{\theta}$$

Regresión Poisson

Se usa para estimar el número de ocurrencias de un suceso dadas unas covariables:

$$Y_i \mid p_i \sim \mathcal{P}(\lambda_i)$$
$$\log \lambda_i = \mathbf{x}_i \boldsymbol{\theta}$$

Regresión logística

Ejemplo: Se desea estimar la probabilidad de que un bebé prematuro se pueda alimentar con leche materna al ser dado de alta. Se tienen datos de 64 bebés prematuros de la edad gestacional (en semanas) al nacer (x), del número de bebés que pudieron ser alimentados con leche materna al salir del hospital ($\# \{y = 1\}$) y de los que no ($\# \{y = 1\}$):

x	28	29	30	31	32	33
$\# \{y = 0\}$	4	3	2	2	4	1
$\# \{y = 1\}$	2	2	7	7	16	14

Modelos jerárquicos

Supongamos que tenemos datos, $\mathbf{x}$, con función de densidad $f(\mathbf{x} \mid \boldsymbol{\theta})$ donde los valores de $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ son intercambiables, es decir, cualquier permutación de los mismos tiene la misma distribución.

En este tipo de situaciones, a menudo, se desea imponer a priori hipótesis adicionales sobre las relaciones entre los elementos de $\boldsymbol{\theta}$. Esto puede hacerse construyendo una densidad a priori, $\pi(\boldsymbol{\theta} \mid \phi)$, en la que $\boldsymbol{\theta}$ depende de un hiperparámetro desconocido.

De este modo, se construye lo que se denomina un modelo jerárquico. Lo más natural es asumir una distribución a priori, $\pi(\phi)$, sobre el hiperparámetro, ϕ.

Modelos jerárquicos

Ejemplo de un modelo normal jerárquico

Supongamos que n individuos se someten a m distintos tests de inteligencia propuestos por m psicólogos diferentes. Se asume que cada uno de los resultados, X_{ij}, sigue una distribución normal:

$$X_{ij} \mid \theta_i, \phi \sim \mathcal{N}\left(\theta_i, \frac{1}{\phi}\right), \qquad i = 1, \ldots, n, \quad j = 1, \ldots, m,$$

de modo que cada individuo tiene su propia media, θ_i. Suponiendo que estas medias son intercambiables, se puede asumir a priori:

$$\theta_i \mid \mu, \psi \sim \mathcal{N}\left(\mu, \frac{1}{\psi}\right),$$

donde los hiperparámetros son μ y ψ, que representan la media y precisión poblacional del resultado del test.

Modelos jerárquicos

Ejemplo: Realizar inferencia Bayesiana para el modelo anterior a partir de la siguiente muestra de resultados sobre 5 individuos sometidos cada uno de ellos a 3 test diferentes:

	1	2	3	4	5
Test 1	106	121	159	95	78
Test 2	108	113	158	91	80
Test 3	98	115	169	93	77

Modelos jerárquicos

Ejemplo de un modelo Poisson jerárquico

Se observan datos de fallos de n bombas en una central hidroeléctrica. Se supone que el número de fallos, X_i, obsevados en el período, t_i, sigue una distribución de Poisson:

$$X_i \mid \lambda_i \sim \mathcal{P}(\lambda_i t_i), \qquad \text{para } i = 1, \ldots, n,$$

Parece natural asumir que las tasas de fallos, λ_i, son intercambiables y asumir:

$$\lambda_i \mid \gamma \sim \mathcal{E}(\gamma),$$

donde γ es el hipérparámetro a priori.

Modelos jerárquicos

Ejemplo: Realizar inferencia Bayesiana para el modelo de Poisson anterior usando los datos del número de fallos, x_i, observados en 10 bombas en el período, t_i, medido en miles de horas, $i = 1, \ldots, 10$.

Bomba	1	2	3	4	5	6	7	8	9	10
t_i	94.5	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
x_i	5	1	5	14	3	19	1	1	4	22

Modelos jerárquicos

Ejemplo de un modelo de regresión jerárquico

Se miden 6 indicadores económicos, Y_i, $i = 1, \ldots, 6$, en 44 instantes de tiempo $t_1, \ldots, t_{44}$. Se modeliza la distribución de cada indicador mediante:

$$Y_{ij} \mid \beta_{0i}, \beta_{1i}, \tau \sim \mathcal{N}(\beta_{0i} + \beta_{1i}(t_j - \bar{t}), \tau)$$

y se asume que los parámetros de cada una de estas regresiones vienen modelizados por una distribución común:

$$\beta_{0i} \mid \mu_0, \tau_0 \sim \mathcal{N}(\mu_0, \tau_0)$$
$$\beta_{1i} \mid \mu_1, \tau_1 \sim \mathcal{N}(\mu_1, \tau_1)$$

Parece natural asumir que los indicadores son intercambiables.