Inferencia Bayesiana

Proyecto Final

1. Se realizó un pequeño experimento con el fin de medir el riesgo de cierto tipo de tumor en un grupo de ratas, dadas diferentes dosis de una droga. El propósito del estudio es estudiar la relación entre la dosis y la respuesta, i.e. la tasa a la que el riesgo de tumor crece o decrece como función de la dosis. Los datos del experimento se presentan en la Tabla 1, donde x representa el nivel de la dosis, mientras que n_x y y_x denotan, respectivamente, el número de ratas tratadas y el número de ratas que presentan tumor en cada nivel ($x = 0, 1, 2$).

Tabla 1

x	n_x	y_x
0	14	4
1	34	4
2	34	2

Sea π_x la probabilidad de que una rata en el grupo x desarrolle un tumor. Consideren el modelo

$$Y_x \sim \text{Bin}(\pi_x, n_x) \qquad (x = 0, 1, 2).$$

Dado que las investigadoras están interesadas en la forma como varía π_x en función de la dosis x, propusieron el modelo

$$\text{logit}(\pi_x) = \alpha + \beta\, x \qquad (x = 0, 1, 2).$$

El parámetro de interés para las investigadoras es la pendiente β, pero no cuentan con información inicial sobre su valor.

Proporcionen un resumen (lo más completo posible) de la distribución final de β suponiendo una distribución inicial no informativa en la que α y β se asumen independientes, con $\alpha \sim N(0, 1000)$ y $\beta \sim N(0, 1000)$; esto es, con media 0 y varianza 1000.

2. Dado que el tamaño de las muestras en el problema anterior es muy pequeño, y en vista de la falta de información inicial, las investigadoras se dieron a la tarea de buscar información relevante en la literatura. Como producto de esta labor, encontraron datos de 10 estudios similares con ratas de la misma cepa, Desafortunadamente, todos estos datos correspondían a *controles*; es decir, ratas a las que no se les aplicó la droga. Los datos se presentan en la Tabla 2a. Aquí $n_{0,i}$ y $y_{0,i}$ denotan, respectivamente, el número total de ratas y el número de ratas que presentaron un tumor en el i-ésimo estudio ($i = 1, 2, \ldots, 10$).

Tabla 2a

Estudio i	$n_{0,i}$	$y_{0,i}$
1	10	1
2	13	2
3	48	10
4	19	5
5	20	0
6	18	0
7	25	2
8	49	5
9	48	9
10	19	4

No satisfechas con estos datos, las investigadoras siguieron buscando trabajos recientes (no publicados). Finalmente encontraron dos reportes muy relevantes, de donde extrajeron los siguientes datos:

Tabla 2b

x	$n_{x,11}$	$y_{x,11}$
0	7	3
1	16	5
2	18	2

Tabla 2c

x	$n_{x,12}$	$y_{x,12}$
0	5	2
1	11	1
2	9	0

En vista de que para los datos de la Tabla 2a sólo se recabó información de controles, el nivel de la dosis es $x = 0$ en todos esos casos. Por lo tanto el modelo que propusieron para esos datos es

$$Y_i \sim \text{Bin}(\pi_{0,i}, n_{0.i}), \qquad i = 1, 2, \ldots, 10,$$

donde

$$\text{logit}(\pi_{0,i}) = \alpha_i, \qquad i = 1, 2, \ldots, 10.$$

Por otra parte, para los datos de las Tablas 2b y 2c (Estudios 11 y 12), las investigadoras supusieron un modelo de la misma forma que el del Problema 1, es decir:

$$Y_{x,i} \sim \text{Bin}(\pi_{x,i}, n_{x,i}) \qquad (x = 0, 1, 2); \;\; i = 11, 12,$$

con

$$\text{logit}(\pi_{x,i}) = \alpha_i + \beta_i\, x \qquad (x = 0, 1, 2); \;\; i = 11, 12.$$

Para simplificar el análisis en esta etapa, las investigadoras decidieron considerar todos estos estudios suficientemente similares como para suponer que los datos de las Tablas 1, 2a, 2b y 2c *provienen de un solo experimento*, de manera que $\alpha_1 = \alpha_2 = \cdots = \alpha_{12} = \alpha$ y $\beta_{11} = \beta_{12} = \beta$.

Utilizando *la misma distribución inicial* que en el Ejercicio 1, proporcionen un resumen (lo más completo posible) de la distribución final de β.

3. Poco tiempo después, una de las investigadoras tuvo la oportunidad de asistir a un curso de Análisis Bayesiano de Modelos Jerárquicos y convenció al resto del equipo de que ésa es la manera más apropiada de analizar los datos con los que contaban. Específicamente, dado que todos los estudios eran similares, consideraron que podían utilizar los 12 estudios que encontraron en la literatura para complementar la información de su experimento original (ver Tabla 1).

Las investigadoras supusieron entonces que los parámetros $\{\alpha, \alpha_1, \ldots, \alpha_{12}\}$ eran intercambiables, con distribución común $N(\alpha^*, \sigma^2_\alpha)$, y también que los parámetros $\{\beta, \beta_{11}, \beta_{12}\}$ eran intercambiables, con distribución común $N(\beta^*, \sigma^2_\beta)$. Finalmente, tanto para α^* como para β^* supusieron una distribución $N(0, 100)$, mientras que para $\tau_\alpha = 1/\sigma^2_\alpha$ y $\tau_\beta = 1/\sigma^2_\beta$ consideraron una distribución Gamma$(0.01, 0.01)$.

Proporcionen un resumen (lo más completo posible) de la distribución final de β (la correspondiente al Ejercicio 1) bajo estas condiciones.

4. Comparen y discutan los resultados de estos tres análisis.

¡Diviértanse!