

# Estadística Bayesiana

## Parte III

Irving Gómez Méndez



# Validación del Modelo

## Simulación de la predictiva previa

También se puede simular una muestra de la predictiva previa, para ello basta con simular una muestra de los parámetros  $\tilde{\theta}_1, \dots, \tilde{\theta}_S$ , a partir de la distribución previa y luego simular  $\tilde{Y}_s \sim P(Y|\tilde{\theta}_s)$ .

**Poder simular de la predictiva previa puede ayudar a discriminar entre distintas distribuciones previas.** Pues al simular de la predictiva previa podemos observar las consecuencias de las distribución previa sobre la variable de interés  $Y$ . Muchas de las técnicas convencionales para decidir la distribución previa pueden dar lugar a resultados más bien absurdos.

## Validación del modelo

Una vez que hemos completado los dos primeros pasos del análisis bayesiano: construcción de un modelo de probabilidad y cómputo de la distribución posterior, no deberíamos ignorar el paso (relativamente sencillo) de validar qué tan bien se ajusta nuestro modelo a los datos y a nuestro conocimiento del fenómeno.

Dado que de antemano sabemos que nuestro modelo no puede incluir todos los aspectos de la realidad, podemos averiguar cuáles aspectos no son capturados. Y, sobre todo, sobre la plausibilidad de nuestro modelo para el propósito para el que fue construido. **No se trata de preguntarnos si nuestro modelo es falso o verdadero, sino de cuáles son las principales deficiencias de nuestro modelo.**

## Estadísticas de prueba y $p$ -valor frecuentista

Desde un enfoque frecuentista se define una estadística de prueba  $T(\mathbf{Y})$  como una estadística de los datos usada para comparar los datos observados contra una réplica generada bajo el modelo propuesto. De esta manera, el  $p$ -valor clásico se define como

$$p_C = \mathbb{P}(T(\mathbf{Y}^{rep}) \geq T(\mathbf{Y}) | \theta).$$

Esto es,  $p_C$  es la probabilidad de obtener una estadística más extremista que la observada, con  $\theta$  fijo. En donde  $\theta$  puede ser un valor ‘nulo’ en pruebas de hipótesis o alguna estimación como el estimador de máxima verosimilitud.

## Contexto bayesiano

En el contexto bayesiano, aprovechamos que contamos con una distribución generativa dada por la distribución predictiva. Si el modelo ajusta bien, entonces **datos replicados con nuestro modelo generativo deberían ser similares a los datos observados**. Dicho de otra manera, los datos observados deben de parecer plausibles al considerar la distribución predictiva posterior. Nuestra técnica básica para validar el ajuste de un modelo es simular datos de la distribución predictiva posterior y compararlos con los datos observados.

## Significancia estadística y significancia práctica

El objetivo no es responder la pregunta: ¿los datos provienen del modelo supuesto?, cuya respuesta casi siempre será no, sino cuantificar las discrepancias entre los datos y el modelo, y saber cuándo esas discrepancias podrían haber surgido por puro azar, bajo los supuestos del modelo.

Si la falla del modelo en un aspecto importante es grande, podemos pensar en cambiar el modelo, si no, podemos ignorarla si no afecta las conclusiones principales. **El  $p$ -valor mide ‘significancia estadística’ no ‘significancia práctica’.**

## Réplicas de los datos

Para que la comparación sea correcta los datos replicados,  $\mathbf{Y}^{rep}$  deben de ser (cómo su nombre indica) réplicas de los datos observados. Es decir,  $\mathbf{Y}^{rep}$  debe ser de la misma dimensión que  $\mathbf{Y}$ , y si nuestro modelo cuenta con variables predictoras  $\mathbf{X}$ , entonces debemos usar exactamente los mismos valores de las variables predictoras.



## Cantidad de prueba y $p$ -valor bayesiano

Para medir la discrepancia entre el modelo ajustado y los datos, definimos una cantidad de prueba  $T(\mathbf{Y}, \theta)$ . Al ser  $\theta$  una variable aleatoria, la cantidad de prueba puede depender no sólo de los datos sino también del valor de  $\theta$ .

El  $p$ -valor bayesiano se define entonces como la probabilidad de que la cantidad de prueba evaluada en los datos replicados,  $T(\mathbf{Y}^{rep}, \theta)$ , sea más extrema que la cantidad de prueba evaluada en los datos observados.

$$\begin{aligned} p_B &= \mathbb{P}(T(\mathbf{Y}^{rep}, \theta) \geq T(\mathbf{Y}, \theta) | \mathbf{Y}) \\ &= \mathbb{E} \left[ \mathbb{1}_{T(\mathbf{Y}^{rep}, \theta) \geq T(\mathbf{Y}, \theta)} | \mathbf{Y} \right] \end{aligned}$$

## Estimación del $p$ -valor bayesiano

Como  $p_B = \mathbb{E} \left[ \mathbb{1}_{T(\mathbf{Y}^{rep}, \theta) \geq T(\mathbf{Y}, \theta)} | \mathbf{Y} \right]$ , entonces una manera de estimar el  $p$ -valor es simulando  $S$  valores de la distribución posterior  $\tilde{\theta}_1, \dots, \tilde{\theta}_S$  y  $S$  réplicas,  $\mathbf{Y}_1^{rep}, \dots, \mathbf{Y}_S^{rep}$  de la distribución predictiva posterior y calcular

$$\hat{p}_B = \frac{1}{S} \sum_{s=1}^S \mathbb{1}_{T(\mathbf{Y}_s^{rep}, \tilde{\theta}_s) \geq T(\mathbf{Y}, \tilde{\theta}_s)}.$$

## Validación marginal del modelo

Además, podemos estimar la distribución marginal predictiva  $p(Y_i|\mathbf{Y})$  y compararla con los datos para buscar datos atípicos o validar el modelo de manera individual, es decir, podemos estimar

$$p_i = \mathbb{P}(T(Y_i^{rep}, \theta) \geq T(Y_i, \theta) | \mathbf{Y}).$$

Si  $Y_i$  es continua, podemos hacer  $T(Y_i, \theta) = Y_i$ . Si  $Y_i$  es discreta, una opción es

$$p_i = \mathbb{P}(Y_i^{rep} > Y_i | \mathbf{Y}) + \frac{1}{2} \mathbb{P}(Y_i^{rep} = Y_i).$$

Si los  $p$ -valores marginales se concentran cerca de 0 y 1, entonces los datos están sobredispersos comparados con el modelo, mientras que si los  $p$ -valores se concentran en 0.5 los datos tienen menor dispersión que la que estima el modelo.

- ▶ Mostrar ejemplo de la estatura de la población !Kung.
- ▶ Mostrar ejemplo con los datos de Newcomb.
- ▶ Mostrar ejemplo de independencia en lanzamientos Bernoulli.

# Regresión Bayesiana

## Regresión de la estatura en el peso

Para estos ejercicios vamos a usar los datos de la comunidad !Kung. Consideraremos sólo a los adultos, esto porque la estatura está fuertemente relacionada con la edad hasta antes de la adultez.

Sea  $h_i$  la estatura del  $i$ -ésimo individuo,  $w_i$  su peso y  $\bar{w}$  el peso promedio de los adultos. Vamos a considerar el modelo:

$$h_i | \mu_i, \sigma^2 \sim \mathcal{N}(\mu_i, \sigma^2)$$

$$\mu_i = \alpha + \beta(w_i - \bar{w})$$

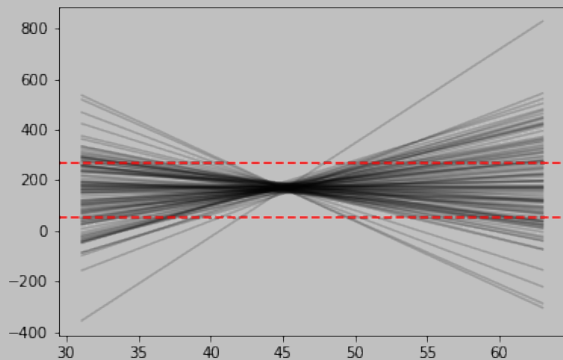
$$\alpha \sim \mathcal{N}(170, 10^2)$$

$$\beta \sim \mathcal{N}(0, 10^2)$$

$$\sigma \sim \mathcal{U}(0, 33)$$

Note cuando  $w_i = \bar{w}$ ,  $\mu_i = \alpha$ . Es decir cuando el peso es igual al promedio de la población la estatura promedio será  $\alpha$ , así que hace mucho sentido modelar  $\alpha \sim \mathcal{N}(170, 10^2)$ . Para  $\sigma$  simplemente usamos una distribución con poca información. Pero **¿por qué hemos puesto esa previa para  $\beta$ ?**

Una manera de verificar qué tan sensatas son estas previa es simulando valores de  $\alpha$  y  $\beta$ ,  $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_m$  y luego graficar  $\alpha_j + \beta_j(w - \bar{w})$ ,  $j = 1, \dots, m$ .



¡Evidentemente esta previa para  $\beta$  es absurda!



Cualquiera sabe que el peso y la estatura guardan (hasta cierto punto) una correlación positiva, por lo que es sensato considerar una distribución previa estrictamente positiva para  $\beta$ . Por lo que consideramos el siguiente modelo:

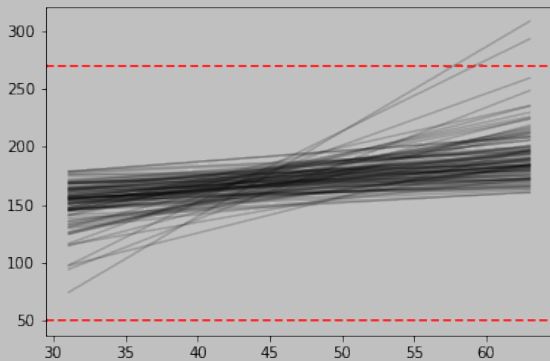
$$h_i | \mu_i, \sigma \sim \mathcal{N}(\mu_i, \sigma^2)$$

$$\mu_i = \alpha + \beta(w_i - \bar{w})$$

$$\alpha \sim \mathcal{N}(170, 10^2)$$

$$\beta \sim \text{lognormal}(0, 1)$$

$$\sigma \sim \mathcal{U}(0, 33)$$



## Estimador MAP

Una vez establecido el modelo podemos crear una rendija de valores para  $\alpha$ ,  $\beta$  y  $\sigma$ . Y como antes, simular una muestra de la posterior. Con esta muestra podemos estimar puntualmente los parámetros y graficar la recta con mayor probabilidad a posteriori  $\hat{\alpha} + \hat{\beta}(w - \bar{w})$ .

## Intervalos de confianza

Como  $\mu$  depende de los parámetros y estos tienen una distribución posterior, entonces  $\mu$  también cuenta con una distribución posterior. En particular **fijando un peso**  $w$ , y contando con una muestra a posteriori  $\tilde{\alpha}_1, \dots, \tilde{\alpha}_S, \tilde{\beta}_1, \dots, \tilde{\beta}_S$  podemos obtener una muestra a posteriori  $\tilde{\mu}_1, \dots, \tilde{\mu}_S$  usando la relación:

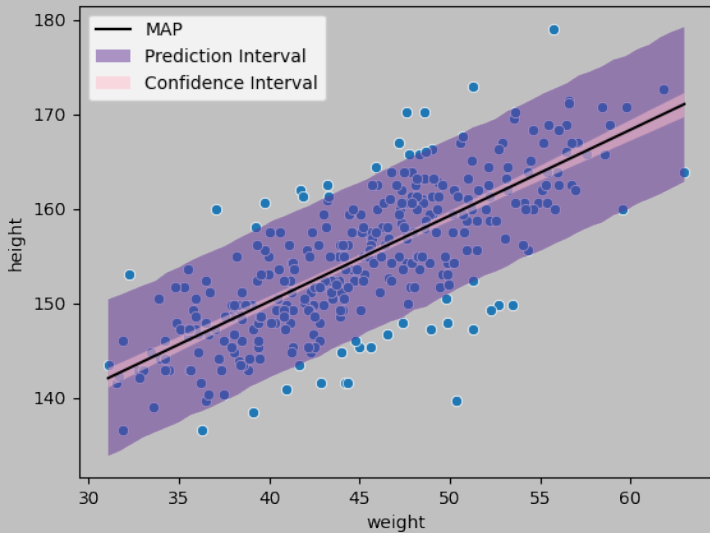
$$\tilde{\mu}_s = \tilde{\alpha}_s + \tilde{\beta}_s(w - \bar{w}).$$

A partir de esta muestra podemos obtener un intervalo de confianza para la estatura media dado el peso  $w$ . Variando  $w$  desde el valor más bajo hasta el valor más alto podemos obtener un intervalo de confianza para la función de regresión.

## Intervalos de predicción

Finalmente, para agregar intervalos de predicción hay que recordar que  $h|\mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2)$ . Así **para un peso fijo**  $w$  podemos obtener una muestra de la posterior de  $\mu$  como se explicó antes,  $\tilde{\mu}_1, \dots, \tilde{\mu}_S$ . Además, podemos simular una muestra de la posterior de  $\sigma$ ,  $\tilde{\sigma}_1, \dots, \tilde{\sigma}_S$ . Por lo tanto, podemos obtener una **muestra de la posterior de  $h$** ,  $\tilde{h}_1, \dots, \tilde{h}_S$ , donde  $\tilde{h}_s \sim \mathcal{N}(\tilde{\mu}_s, \tilde{\sigma}_s^2)$ . A partir de la muestra  $\tilde{h}_1, \dots, \tilde{h}_S$  podemos obtener un intervalo de predicción dado el peso  $w$ . Variando  $w$  desde el valor más bajo hasta el valor más alto podemos obtener un intervalo de predicción para la función de regresión.

- ▶ Mostrar ejemplo de regresión de la estatura de los adultos !Kung.



# Aproximaciones Computacionales

## Aproximación por cuadratura

Es claro que una de las limitantes del método de aproximación usando una rendija es que escala pobremente cuando incrementamos el número de parámetros en el modelo (maldición de la alta dimensionalidad). De los ejemplos anteriores también hemos visto que las posteriores de los parámetros tienden a tomar una forma acampanada cuando el tamaño de la muestra es grande (teorema del límite central).



Esta observación nos permite introducir el método de aproximación por cuadratura. Hay que recordar que la densidad normal tiene esta forma de campana y que el logaritmo es proporcional a  $\lambda(Y - \mu)^2$ , i.e. es una función cuadrática (una parábola). Usando este hecho podemos aproximar la posterior de los parámetros mediante una distribución normal.

## *Markov chain Monte Carlo, MCMC*

Hay varios modelos, como los modelos jerárquicos en los que ni la aproximación por rendija ni la aproximación cuadrática son satisfactorias. Como resultado, se han desarrollado varias técnicas nada intuitivas para implementar dichos modelos, como *Markov chain Monte Carlo* (MCMC) o inferencia variacional. Siendo MCMC la más popular. De hecho, es ampliamente reconocido que gracias a MCMC es que la estadística bayesiana resurgió en la década de los 1990's. Y, aunque MCMC es más antigua que 1990, el poder de cómputo necesario para ejecutar la técnica no lo es, por lo que el resurgimiento de la estadística bayesiana es también un logro de la ingeniería computacional.

El beneficio en el uso de técnicas avanzadas de aproximación de la posterior, como MCMC o inferencia variacional, es escapar del extraño supuesto de una normal multivariada para la distribución posterior conjunta. Esto se vuelve extremadamente útil en modelos más complejos como los modelos jerárquicos, en los que la distribución posterior es lejana a una gaussiana. Sin embargo, existe un enorme reto conceptual ya que estas técnicas siguen una estrategia nada obvia.

La simulación mediante cadenas de Markov engloba varios métodos, en los que los valores muestreados para los parámetros se obtienen de distribuciones que se aproximan cada vez más a la distribución posterior, y que dependen de la última muestra obtenida (de allí que sean cadenas de Markov). Los detalles de por qué estos métodos funcionan corresponden a un curso de cómputo y simulación estocástica.

- ▶ Mostrar ejemplo de muestreador Gibbs.
- ▶ Mostrar aproximación por cuadratura y MCMC para el ejemplo binomial.
- ▶ Mostrar aproximación por cuadratura y MCMC para la estatura de adultos !Kung.

## Regresión considerando toda la población !Kung

Consideremos una vez más el ejemplo de predecir la estatura de la población !Kung según su peso, pero ahora tomemos todos los individuos de la muestra. Ahora estamos dispuestos a aceptar que pueda haber mayor variabilidad en la estatura así que cambiamos la distribución previa de  $\sigma$  por  $\sigma \sim \mathcal{U}(0, 50)$ . Antes de ajustar el modelo estandarizamos la altura y el peso (restamos la media y dividimos por el desvío padrón), más adelante veremos por qué hemos hecho esto.

- ▶ Mostrar modelo lineal para la regresión de la estatura de los adultos !Kung.
- ▶ Mostrar modelo lineal para la regresión de la estatura de toda la población !Kung.

El modelo lineal no es satisfactorio al considerar toda la población !Kung. Por lo que en su lugar proponemos el siguiente modelo cuadrático:

$$h_i | \mu_i, \sigma^2 \sim \mathcal{N}(\mu_i, \sigma^2)$$

$$\mu_i = \alpha + \beta_1 y_i + \beta_2 y_i^2$$

$$\alpha \sim \mathcal{N}(170, 10^2)$$

$$\beta_1 \sim \text{lognormal}(0, 1)$$

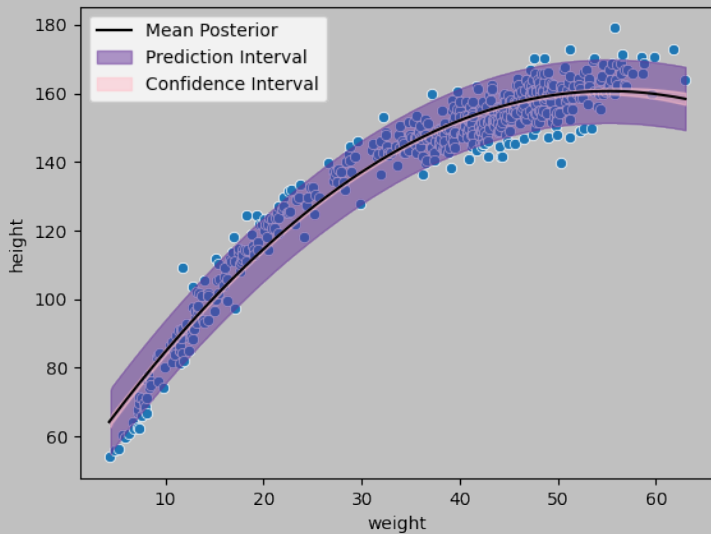
$$\beta_2 \sim \mathcal{N}(0, 1)$$

$$\sigma \sim \mathcal{U}(0, 50),$$

donde  $y_i = (w_i - \bar{w})/s_w$ .

Como estamos considerando una potencia de la variable predictora conviene estandarizarla antes para evitar posibles problemas numéricos.

- Mostrar modelo cuadrático para toda la población !Kung



# Tarea 5

- ▶ Dejar tarea 5.



# Modelos Jerárquicos

## Ensayo biológico

*Ejemplo tomado de la sección 3.7 de BDA3.*

Durante el desarrollo de nuevos medicamentos o sustancias químicas, es común realizar estudios sobre su toxicidad en animales. Dichos estudios se realizan administrando varias dosis a distintos grupos de animales.

La respuesta es típicamente binaria: por ejemplo, vive o muere, hay tumor o no hay tumor. Un experimento de este tipo genera datos de la forma

$$(X_i, n_i, Y_i); \quad i = 1, \dots, k,$$

donde  $X_i$  representa la  $i$ -ésima de  $k$  dosis (generalmente medidas en una escala logarítmica) administrada a  $n_i$  animales, de los cuales,  $Y_i$  tuvieron una respuesta positiva.

Un ejemplo con datos reales se encuentra en la siguiente tabla, donde se realizaron pruebas en 20 animales, 5 por cada una de las 4 dosis.

Dosis, $X_i$ (log g/ml)	Animales, $n_i$	Muertes, $Y_i$
-0.86	5	0
-0.3	5	1
-0.05	5	3
0.73	5	5

Datos de Racine et al (1986). *Bayesian methods in practice: experiences in the pharmaceutical industry (with discussion)*. Applied Statistics.

Para este tipo de datos, podemos considerar el siguiente modelo:

$$Y_i | \theta_i \sim \text{Binomial}(n_i, \theta_i),$$

donde  $\theta_i$  es la probabilidad de muerte de los animales dada la dosis  $X_i$ . El modelo más simple para la relación dosis-respuesta, es decir la relación entre  $\theta_i$  y  $X_i$ , es lineal:  $\theta_i = \alpha + \beta X_i$ . Sin embargo, este modelo presenta serios problemas cuando  $\theta_i \rightarrow \infty$ , **¿por qué?**

La solución estándar es transformar  $\theta_i$ , en la relación dosis-respuesta:

$$\text{logit}(\theta_i) = \alpha + \beta X_i,$$

donde  $\text{logit}(\theta_i) = \log(\theta_i / (1 - \theta_i))$ . A este modelo se le llama **regresión logística**.

Podemos usar el siguiente modelo para estos datos:

$$Y_i | \theta_i \sim \text{Binomial}(n_i, \theta_i)$$

$$\text{logit}(\theta_i) = \alpha + \beta X_i$$

$$p(\alpha) \propto \mathbb{1}_{\mathbb{R}}(\alpha)$$

$$p(\beta) \propto \mathbb{1}_{\mathbb{R}}(\beta)$$

Un parámetro de interés en los estudios biológicos es la dosis letal mediana  $\text{LD}_{50}$ , que corresponde a la dosis cuya probabilidad de muerte es del 50%.

$$\theta = \text{logit}^{-1}(\alpha + \beta \text{LD}_{50}) = 0.5$$

entonces,  $\alpha + \beta \text{LD}_{50} = \text{logit}(0.5) = 0$ , y  $\text{LD}_{50} = -\alpha/\beta$ .

## Modelos jerárquicos

En su versión más simple, en los modelos jerárquicos los parámetros  $\theta_j$  forman una muestra simple de la distribución previa, que es determinada por un vector de (hiper)parámetros  $\phi$ , es decir

$$p(\theta|\phi) = \prod_{j=1}^J p(\theta_j|\phi).$$

En general  $\phi$  es desconocido y, por lo tanto, tiene su propia distribución (hiper)previa  $p(\phi)$ .

La distribución previa conjunta es

$$p(\theta, \phi) = p(\theta|\phi)p(\phi),$$

y la distribución posterior conjunta es

$$\begin{aligned} p(\theta, \phi|\mathbf{Y}) &\propto p(\theta, \phi)p(\mathbf{Y}|\theta, \phi) \\ &= p(\theta, \phi)p(\mathbf{Y}|\theta) \\ &= p(\theta|\phi)p(\phi)p(\mathbf{Y}|\theta). \end{aligned}$$

Al asignar una distribución previa para  $\phi$  tenemos que tener cuidado al usar una distribución impropia y checar que la distribución posterior exista.

Note que la condicional posterior de  $\theta$  se puede obtener como

$$\begin{aligned} p(\theta|\phi, \mathbf{Y}) &\propto p(\theta|\phi)p(\mathbf{Y}|\theta) \\ &= f(\phi, \mathbf{Y})p(\theta|\phi)p(\mathbf{Y}|\theta). \end{aligned}$$

donde  $f(\phi, \mathbf{Y})$  es la ‘constante’ de normalización que depende de  $\phi$  y  $\mathbf{Y}$ .



Podemos calcular la posterior de  $\phi$  integrando sobre  $\theta$  la distribución posterior conjunta:

$$p(\phi|\mathbf{Y}) = \int p(\theta, \phi|\mathbf{Y})d\theta.$$

O, usando la fórmula de probabilidad condicional:

$$\begin{aligned} p(\phi|\mathbf{Y}) &= \frac{p(\theta, \phi|\mathbf{Y})}{p(\theta|\phi, \mathbf{Y})} \\ &\propto \frac{p(\theta|\phi)p(\phi)p(\mathbf{Y}|\theta)}{f(\phi, \mathbf{Y})p(\theta|\phi)p(\mathbf{Y}|\theta)} \\ &\propto \frac{p(\phi)}{f(\phi, \mathbf{Y})} \end{aligned}$$

## Tumores en ratas

*Ejemplo tomado de la sección 5.1 y 5.3 de BDA3.*

Como mencionamos anteriormente, al evaluar nuevos medicamentos, es común realizar estudios en roedores. Para este ejemplo suponga que 4 de un total de 14 ratas desarrollaron pólipos endometriales (un tipo de tumor). Podemos suponer un modelo binomial para el número de ratas con tumor dado  $\theta$ , donde  $\theta$  es la probabilidad de desarrollar dicho tumor. Por conveniencia, podemos usar para  $\theta$  la previa conjugada  $\theta \sim \text{Beta}(\alpha, \beta)$

Para asignar valores para  $\alpha$  y  $\beta$  recurramos a nuestro conocimiento previo sobre  $\theta$ , por ejemplo, mediante su su promedio y desviación. Sin embargo, esta información no es conocida típicamente. En vez de ello, contamos con datos históricos.

Supongamos que contamos con  $J$  datos históricos, y que modelamos

$$Y_j | \theta_j \sim \text{Binomial}(n_j, \theta_j),$$

con el número de ratas,  $n_j$ , conocido. Y

$$\theta_j \sim \text{Beta}(\alpha, \beta).$$

Sólo falta asignar una distribución previa para  $(\alpha, \beta)$ , pero debemos checar que la posterior exista.

$$p(\theta, \alpha, \beta | \mathbf{Y}) \propto p(\alpha, \beta) p(\theta | \alpha, \beta) p(\mathbf{Y} | \theta)$$

$$\propto p(\alpha, \beta) \prod_{j=1}^J \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1} (1 - \theta_j)^{\beta-1} \prod_{j=1}^J \theta_j^{Y_j} (1 - \theta_j)^{1-Y_j}$$

Luego,

$$p(\theta|\alpha, \beta, \mathbf{Y}) \propto \prod_{j=1}^J \theta_j^{\alpha+Y_j-1} (1-\theta_j)^{\beta+n_j-Y_j-1}.$$

Es decir, dado  $(\alpha, \beta)$ , los elementos de  $\theta$  son independientes y siguen una distribución beta *a posteriori*,

$$\theta_j|\alpha, \beta, \mathbf{Y} \sim \text{Beta}(\alpha + Y_j, \beta + n_j - Y_j),$$

y

$$p(\theta|\alpha, \beta, \mathbf{Y}) = \prod_{j=1}^J \frac{\Gamma(\alpha + \beta + n_j)}{\Gamma(\alpha + Y_j)\Gamma(\beta + n_j - Y_j)} \theta_j^{\alpha+Y_j-1} (1-\theta_j)^{\beta+n_j-Y_j-1}$$

Reescribiendo  $p(\theta|\alpha, \beta, \mathbf{Y})$  de la forma:

$$p(\theta|\alpha, \beta, \mathbf{Y}) = f(\alpha, \beta, \mathbf{Y})p(\theta|\alpha, \beta)p(\mathbf{Y}|\theta),$$

obtenemos que

$$f(\alpha, \beta, \mathbf{Y}) = \prod_{j=1}^J \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \frac{\Gamma(\alpha + \beta + n_j)}{\Gamma(\alpha + Y_j)\Gamma(\beta + n_j - Y_j)} \left[ \binom{n_j}{Y_j} \right]^{-1}.$$

Entonces,

$$p(\alpha, \beta|\mathbf{Y}) \propto p(\alpha, \beta) \prod_{j=1}^J \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + Y_j)\Gamma(\beta + n_j - Y_j)}{\Gamma(\alpha + \beta + n_j)}.$$

En el libro de BDA3, los autores proponen como una previa no informativa  $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$ , la cual genera una distribución posterior propia.

Note que **podemos evitar el enorme esfuerzo matemático** de checar la existencia de la distribución posterior **si nos restringimos a proponer distribuciones propias** para los hiperparámetros.

## Modelo normal jerárquico

- ▶ Checar notas sobre modelo normal jerárquico.
- ▶ Checar artículo sobre estimación de la pobreza en Tailandia.
- ▶ Checar ejemplo sobre estudio en 8 escuelas.



## Meta-análisis

*Ejemplo tomado de la sección 5.6 de BDA3.*

Meta-análisis es el proceso para resumir e integrar los resultados de distintos estudios.

Los datos del siguiente ejemplo resumen los infartos de miocardio en 22 ensayos clínicos, cada uno consiste de dos grupos de pacientes que fueron asignados aleatoriamente para recibir beta-bloqueadores (medicamentos que afectan al sistema nervioso y que reducen la presión arterial).

El ensayo clínico  $j$  involucra a  $n_{0j}$  individuos en el grupo control y  $n_{1j}$  en el grupo tratado, generando  $y_{0j}$  y  $y_{1j}$  muertes en el grupo control y el grupo tratado, respectivamente, entonces el modelo involucra dos binomiales independientes con probabilidades  $p_{0j}$  y  $p_{1j}$ , respectivamente.

Una cantidad de interés es la razón de momios:

$$\rho_j = \frac{\frac{p_{1j}}{1-p_{1j}}}{\frac{p_{0j}}{1-p_{0j}}}.$$

Sin embargo, por su interpretabilidad y su cercanía con una distribución gaussiana, nos centraremos en la inferencia del logaritmo de la tasa de momios,  $\theta_j = \log \rho_j$ .

Calculamos el logaritmo de los momios observados:

$$y_j = \log \left( \frac{y_{1j}}{n_{1j} - y_{1j}} \right) - \log \left( \frac{y_{0j}}{n_{0j} - y_{0j}} \right),$$

con una varianzaz aproximada

$$\sigma_j^2 = \frac{1}{y_{1j}} + \frac{1}{n_{1j} - y_{1j}} + \frac{1}{y_{0j}} + \frac{1}{n_{0j} - y_{0j}}$$

Este es nuestro modelo propuesto:

$$y_j | \theta_j, \sigma_j^2 \sim \text{Normal}(\theta_j, \sigma_j^2)$$

$$\theta_j \sim \text{Normal}(\mu, \tau^2)$$

$$p(\mu) \propto \mathbb{1}_{\mathbb{R}}(\mu)$$

$$p(\tau) \propto \mathbb{1}_{\mathbb{R}}(\tau)$$

