

6. Modelos jerárquicos

- Los modelos jerárquicos Bayesianos simplifican la estimación simultánea de varios parámetros θ_i del mismo tipo con dos fines principales: (1) combinar la fuerza de los datos para mejorar la precisión de las estimaciones de los parámetros y (2) permitir incorporar incertidumbre en las estimaciones.
- Una ventaja de los métodos de simulación Bayesianos, como el MCMC, es que permiten realizar inferencias relativamente fáciles para modelos jerárquicos complejos.
- La idea de *combinar fuerzas* está basada en un supuesto de intercambiabilidad (simetría) entre los parámetros θ_i del modelo.
- INTECAMBIABILIDAD: La idea de intercambiabilidad se puede explicar de la siguiente manera: Consideremos un conjunto de k poblaciones, donde cada población está determinada por un parámetro θ_i , $i=1,\dots,k$. Sea Y_i una observación de la i -ésima población, i.e., cada población se puede representar mediante la función de densidad $f(y_i|\theta_i)$. Si no se cuenta con información que nos permita distinguir las θ_i 's unas de otras (además de la información de los datos Y_1,\dots,Y_k) y si no es razonable establecer algún orden o agrupación de los parámetros, entonces debemos suponer una especie de simetría entre los parámetros y ésta debe reflejarse en su distribución inicial. Dicha simetría se representa a través del concepto de intercambiabilidad.

- ¿CÓMO representamos intercambiabilidad en las θ_i 's?

La intercambiabilidad entre el conjunto $(\theta_1, \dots, \theta_k)$ se logra suponiendo que las θ_i 's son condicionalmente independientes dado el valor de un hiperparámetro, digamos φ , i.e.,

$$f(\theta_1, \dots, \theta_k) = \int f(\theta_1, \dots, \theta_k | \varphi) f(\varphi) d\varphi = \int \prod_{i=1}^k f(\theta_i | \varphi) f(\varphi) d\varphi,$$

donde $f(\varphi)$ es la distribución inicial del hiperparámetro.

- **MODELO:** El modelo jerárquico general toma la siguiente forma:

1) Condicional en φ y $(\theta_1, \dots, \theta_k)$, los datos Y_i son independientes, con densidades $f(y_i | \theta_i)$, i.e.,

$$f(y_1, \dots, y_k | \theta_1, \dots, \theta_k) = \prod_{i=1}^k f(y_i | \theta_i)$$

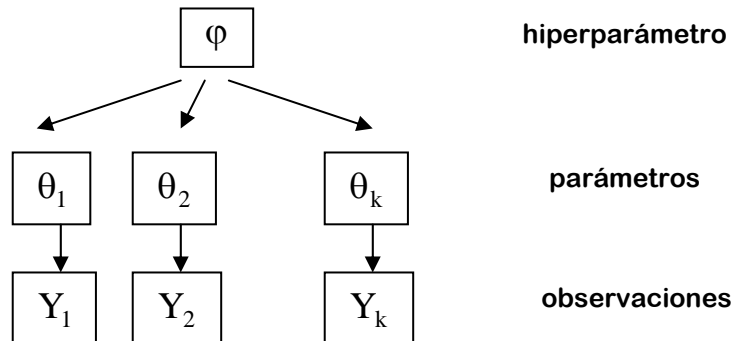
2) Condicional en φ , los parámetros θ_i son observaciones de la misma densidad $f(\theta | \varphi)$, i.e.,

$$f(\theta_1, \dots, \theta_k | \varphi) = \prod_{i=1}^k f(\theta_i | \varphi)$$

3) El hiperparámetro φ tiene función de densidad $f(\varphi)$.

La parte (1) representa el modelo paramétrico de los datos y las partes (2) y (3) representan la distribución inicial de los parámetros del modelo.

- *Representación gráfica* del modelo jerárquico. El modelo jerárquico se puede representar mediante el siguiente esquema:



- APLICACIONES / EXPLICACIONES.
 - Este tipo de modelos se utilizan para realizar inferencias dentro de un conjunto de unidades, por ejemplo, en estudios comparativos entre instituciones, como universidades, hospitales, bancos etc., los parámetros de interés podrían ser la calificación promedio en un examen por universidad, tasas de mortalidad por hospital, tasas de aceptación de créditos, etc.
 - Por lo general en este tipo de estudios comparativos se tienen varias unidades (universidades, hospitales, bancos, etc.) pero pocos datos de cada uno de ellos por separado. La idea de los modelos jerárquicos es combinar la información de distintas fuentes de datos (universidades, hospitales, bancos etc.), explotando la similitud entre los parámetros en términos de su génesis.

- ANÁLISIS DEL MODELO. Las componentes del modelo son:

Verosimilitud: $f(y_1, \dots, y_k | \theta_1, \dots, \theta_k) = \prod_{i=1}^k f(y_i | \theta_i)$

Inicial: $f(\theta_1, \dots, \theta_k, \varphi) = f(\theta_1, \dots, \theta_k | \varphi) f(\varphi)$,

donde, $f(\theta_1, \dots, \theta_k | \varphi) = \prod_{i=1}^k f(\theta_i | \varphi)$

La idea es hacer inferencias sobre los parámetros individuales $(\theta_1, \dots, \theta_k)$ y el parámetro común φ . Entonces,

Final: $f(\theta_1, \dots, \theta_k, \varphi | y) = f(\theta_1, \dots, \theta_k | \varphi, y) f(\varphi | y)$,

donde, $f(\theta_1, \dots, \theta_k | \varphi, y) \propto f(y_1, \dots, y_k | \theta_1, \dots, \theta_k) f(\theta_1, \dots, \theta_k | \varphi)$, y

$f(\varphi | y) \propto f(y | \varphi) f(\varphi)$ con $f(y | \varphi) = \int f(y | \theta_1, \dots, \theta_k) f(\theta_1, \dots, \theta_k | \varphi) d\theta$

- CASOS PARTICULARES: Dependiendo de la selección de la distribución que genera los datos, se pueden construir distintos modelos jerárquicos. Una clase de modelos jerárquicos general se obtiene al considerar $f(y_i | \theta_i)$ como un miembro de la familia exponencial y las distribuciones iniciales $f(\theta_i | \varphi)$ y $f(\varphi)$ miembros de la familia conjugada correspondiente. Ejemplos:

- Normal:

1) $Y_i | \mu_i \sim N(\mu_i, \tau_i)$, con τ_i conocida

2) $\mu_i | \varphi \sim N(\varphi, \lambda)$, con λ conocida

3) $\varphi \sim N(\varphi_0, \lambda_0)$, con φ_0 y λ_0 conocidas

Nota: Es posible que en la primera etapa se incluyan covariables, e.g.,

$Y_i | \beta_i \sim N(x_i' \beta_i, \tau_i)$, con $\beta_i | \varphi \sim N(\varphi, \lambda)$, y $\varphi \sim N(\varphi_0, \lambda_0)$.

○ Poisson:

- 1) $Y_i | \mu_i \sim \text{Po}(\mu_i)$
- 2) $\mu_i | \beta \sim \text{Ga}(\alpha, \beta)$, con α conocida
- 3) $\beta \sim \text{Ga}(\alpha_0, \beta_0)$, con α_0 y β_0 conocidas

Alternativamente,

- 2a) $\eta_i = \log(\mu_i)$, $\eta_i | \varphi \sim N(\varphi, \lambda)$, con λ conocida
- 3a) $\varphi \sim N(\varphi_0, \lambda_0)$, con φ_0 y λ_0 conocidas

○ Bernoulli:

- 1) $Y_i | \pi_i \sim \text{Bin}(n_i, \pi_i)$
- 2) $\pi_i | \alpha, \beta \sim \text{Beta}(\alpha, \beta)$
- 3) $\alpha \sim \text{Ga}(a_0, b_0)$, $\beta \sim \text{Ga}(c_0, d_0)$, con a_0 , b_0 , c_0 y d_0 conocidos

Alternativamente,

- 2a) $\eta_i = \text{logit}(\pi_i)$, $\eta_i | \varphi \sim N(\varphi, \lambda)$, con λ conocida
- 3a) $\varphi \sim N(\varphi_0, \lambda_0)$, con φ_0 y λ_0 conocidas

- EJERCICIO 9. Incidencia y tasas de mortalidad (Congdon, 2001). Considera un caso de estudio de niños con leucemia. Se cuenta con la información del número de muertes en dos regiones del Reino Unido en el año de 1950. Las tasas de muerte están clasificadas por edad (1-0:5, 2-6:14), el tipo de residencia del niño (1-rural, 2-urbano) y el tipo de cancer (1-Lymphoblastic, 2-Myeloblastic). Sea O_i el número de muertes observadas, P_i es el número total de niños (en millones) y θ_i la tasa de mortalidad en la clase i . Se proponen tres modelos para estos datos:

a) Modelo de efectos constantes:

$$Y_i|\theta \sim \text{Po}(\theta P_i), \theta \sim \text{Ga}(0.001, 0.001)$$

b) Modelo de efectos independientes:

$$Y_i|\theta_i \sim \text{Po}(\theta_i P_i), \theta_i \sim \text{Ga}(0.001, 0.001)$$

c) Modelo de efectos intercambiables:

$$Y_i|\theta_i \sim \text{Po}(\theta_i P_i), \theta_i|\alpha, \beta \sim \text{Ga}(\alpha, \beta), \alpha \sim \text{Ga}(1, 1), \beta \sim \text{Ga}(0.1, 1)$$

➤ EJERCICIO 10. Probabilidades de reclamación. Considera una compañía de seguros que cuenta con 10 tipos de seguros. Por el momento la compañía tiene un total de n_i asegurados en cada categoría y únicamente r_i de ellos le han hecho una reclamación, $i=1, \dots, 10$. Sea π_i la probabilidad de que un asegurado en la categoría i le haga una reclamación. Se proponen tres modelos:

a) Modelo con efectos constantes:

$$Y_i|\pi \sim \text{Bin}(n_i, \pi), \pi \sim \text{Beta}(1, 1)$$

b) Modelo con efectos independientes:

$$Y_i|\pi_i \sim \text{Bin}(n_i, \pi_i), \pi_i \sim \text{Beta}(1, 1)$$

c) Modelo con efectos intercambiables:

$$Y_i|\pi_i \sim \text{Bin}(n_i, \pi_i), \pi_i|\alpha, \beta \sim \text{Beta}(\alpha, \beta), \alpha \sim \text{Ga}(1, 1), \beta \sim \text{Ga}(1, 1)$$