

# Estadística Bayesiana

## Modelos multiparámétricos

Ms Carlos López de Castilla Vásquez

Universidad Nacional Agraria La Molina

2017-1



# Introducción

- Aunque un problema puede incluir varios parámetros de interés las conclusiones se presentan solo para algunos pocos.
- En este caso el objetivo final del análisis bayesiano será obtener la distribución posterior marginal de estos parámetros.
- Se empieza por calcular la distribución posterior conjunta y luego las distribuciones marginales de interés por integración o a través del proceso de simulación.
- En muchos problemas no existe un real interés en algunos de los parámetros involucrados pero deben ser considerados para poder construir un modelo realista. Parámetros de este tipo son llamados *parámetros de ruido*.

# Introducción

- Suponga que  $\theta = (\theta_1, \theta_2)$  y que se está interesado solo en hacer inferencias sobre  $\theta_1$  por lo que  $\theta_2$  es un parámetro de ruido.
- Para obtener la densidad posterior de  $\theta_1$  primero se debe calcular la densidad posterior conjunta:

$$f(\theta_1, \theta_2 | y) \propto f(y | \theta_1, \theta_2) f(\theta_1, \theta_2)$$

y luego la densidad posterior marginal integrando sobre  $\theta_2$ :

$$f(\theta_1 | y) = \int f(\theta_1, \theta_2 | y) d\theta_2$$

# Introducción

- La densidad posterior conjunta puede factorizarse de la siguiente manera:

$$f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | y) = f(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2, y) f(\boldsymbol{\theta}_2 | y)$$

con lo que se tiene:

$$f(\boldsymbol{\theta}_1 | y) = \int f(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2, y) f(\boldsymbol{\theta}_2 | y) d\boldsymbol{\theta}_2$$

- La distribución posterior puede ser obtenida simulando valores de  $\boldsymbol{\theta}_2$  desde  $f(\boldsymbol{\theta}_2 | y)$  y luego valores de  $\boldsymbol{\theta}_1$  desde  $f(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2, y)$  usando el valor de  $\boldsymbol{\theta}_2$  ya extraído.

## Distribución a priori de Jeffreys

- La versión multivariada para la distribución a priori no informativa de Jeffreys es:

$$f(\boldsymbol{\theta}) \propto [J(\boldsymbol{\theta})]^{1/2}$$

donde  $J(\boldsymbol{\theta})$  es el determinante de la *matriz de información de Fisher* cuyos elementos son:

$$J_{ii} = E \left[ \left. \frac{d^2 \log f(y|\boldsymbol{\theta})}{d\theta_i^2} \right| \boldsymbol{\theta} \right] \quad J_{ij} = E \left[ \left. \frac{d^2 \log f(y|\boldsymbol{\theta})}{d\theta_i d\theta_j} \right| \boldsymbol{\theta} \right]$$

- Una alternativa es usar el producto de las distribuciones a priori de Jeffreys asumiendo *independencia*.

## Distribución a priori de Jeffreys

- Sea la distribución fundamental:

$$y|\boldsymbol{\theta} \sim \mathcal{N}(\mu, \sigma^2)$$

Hallar la distribución a priori no informativa de Jeffreys para  $\boldsymbol{\theta} = (\mu, \sigma^2)$ .

- Sea la distribución fundamental:

$$\mathbf{y}|\boldsymbol{\pi} \sim \mathcal{M}(n, \pi_1, \pi_2, \pi_3)$$

Hallar la distribución a priori no informativa de Jeffreys para  $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$ .

## Distribución a priori no informativa

- Considere un vector de  $n$  observaciones independientes desde la distribución:

$$y_1, \dots, y_n \sim \mathcal{N}(\mu, \sigma^2)$$

- Este modelo será analizado a partir de una distribución a priori *no informativa*. Sin embargo los resultados pueden ser fácilmente extendidos al caso de distribuciones *informativas*.
- La distribución a priori no informativa para  $\mu$  y  $\sigma^2$  asumiendo independencia entre ambos parámetros es:

$$f(\mu, \sigma^2) \propto \sigma^{-2}$$

# Distribución a priori no informativa

- La distribución posterior conjunta es:

$$\begin{aligned} f(\mu, \sigma^2 | \mathbf{y}) &\propto \sigma^{-n-2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right\} \\ &\propto \sigma^{-n-2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2 \right] \right\} \\ &\propto \sigma^{-n-2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (n-1)s^2 + n(\bar{y} - \mu)^2 \right] \right\} \end{aligned}$$

donde las *estadísticas suficientes* son  $\bar{y}$  y  $s^2$ .



## Distribución a priori no informativa

- La distribución posterior conjunta se puede factorizar de la siguiente manera:

$$f(\mu, \sigma^2 | \mathbf{y}) = f(\mu | \sigma^2, \mathbf{y}) f(\sigma^2 | \mathbf{y})$$

- La *distribución posterior condicional de  $\mu$  dado  $\sigma^2$*  es:

$$\mu | \sigma^2, \mathbf{y} \sim \mathcal{N}\left(\bar{y}, \frac{\sigma^2}{n}\right)$$

- La *distribución posterior marginal de  $\sigma^2$*  es:

$$\sigma^2 | \mathbf{y} \sim \chi^2\text{-inversa de escala}(n - 1, s^2)$$

## Distribución a priori no informativa

- La *distribución posterior marginal* para  $\mu$  se obtiene de la siguiente manera:

$$\begin{aligned} f(\mu|\mathbf{y}) &= \int_0^\infty f(\mu, \sigma^2|\mathbf{y}) d\sigma^2 \\ &\propto \left[ 1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2} \right]^{-n/2} \end{aligned}$$

- La *distribución posterior marginal* para  $\mu$  es:

$$\mu|\mathbf{y} \sim t_{n-1}(\bar{y}, s^2/n)$$

## Distribución a priori no informativa

- Un intervalo de credibilidad al  $(1 - \alpha) \times 100\%$  para  $\mu$  basado en su distribución marginal es:

$$\bar{y} - t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{y} + t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}$$

- La *distribución predictiva posterior* para  $\tilde{y}$  es:

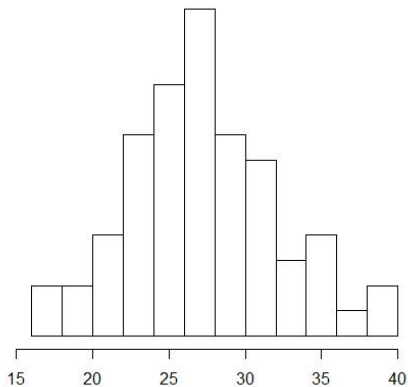
$$\begin{aligned} f(\tilde{y}|\mathbf{y}) &= \int \int f(\tilde{y}|\mu, \sigma^2) f(\mu, \sigma^2|\mathbf{y}) d\mu d\sigma^2 \\ &\sim t_{n-1} \left( \bar{y}, \left( 1 + \frac{1}{n} \right) s^2 \right) \end{aligned}$$

## Ejemplo: Estimación de la velocidad de la luz

- Se realizó un experimento en 1882 para medir la velocidad de la luz registrando el tiempo que requiere para viajar una distancia de 7442 metros.
- Las medidas tomadas como desviaciones desde 24 800 nanosegundos muestran una distribución aproximadamente simétrica.
- El objetivo es realizar el proceso de inferencia posterior para la media asumiendo una distribución a priori no informativa.
- Hallar un intervalo de credibilidad del 95 % para  $\mu$  usando su distribución posterior marginal.

## Ejemplo: Estimación de la velocidad de la luz

Figura 1: Histograma para los tiempos observados



## Ejemplo: Estimación de la velocidad de la luz

## Intervalo usando la distribución posterior marginal

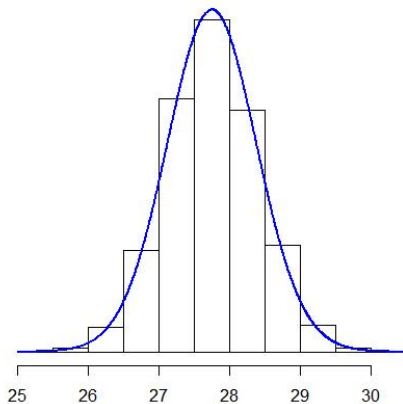
```
> mediaty <- mean(y)
> s2 <- var(y)
> n <- length(y)
> mediaty + c(-1, 1)*qt(p = 0.975,df = n - 1)* sqrt(s2/n)
[1] 26.4802 29.0198
```

## Intervalo usando el proceso de simulación

```
> library(geoR)
> sigma2 <- rinvchisq(n = 10000, df = n - 1, scale = s2)
> mu <- rnorm(n = 10000, mean = mediaty, sd = sqrt(sigma2/n))
> mean(mu)+c(-1, 1)*qt(p = 0.975,df = n - 1)*sd(mu)
[1] 26.45811 29.04037
```

## Ejemplo: Estimación de la velocidad de la luz

Figura 2: Histograma para los valores simulados



# Distribución a priori conjugada

- La distribución a priori conjugada esta dada por las siguientes especificaciones:

$$\mu|\sigma^2 \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{k_0}\right)$$

$$\sigma^2 \sim \chi^2 - \text{inversa de escala } (v_0, \sigma_0^2)$$

- La distribución anterior es llamada normal  $\chi^2$ -inversa de escala y se denota por:

$$\mu, \sigma^2 \sim \mathcal{N}\chi^2\left(\mu_0, \frac{\sigma_0^2}{k_0}, v_0, \sigma_0^2\right)$$



# Distribución a priori conjugada

- La distribución posterior es:

$$\mu, \sigma^2 | \mathbf{y} \sim \mathcal{N}\chi^2 \left( \mu_n, \frac{\sigma_n^2}{k_n}, v_n, \sigma_n^2 \right)$$

donde se puede demostrar que:

$$\begin{aligned}\mu_n &= \frac{k_0}{k_0 + n} \mu_0 + \frac{n}{k_0 + n} \bar{y} \\ k_n &= k_0 + n \\ v_n &= v_0 + n \\ v_n \sigma_n^2 &= v_0 \sigma_0^2 + (n - 1) s^2 + \frac{k_0 n}{k_0 + n} (\bar{y} - \mu_0)^2\end{aligned}$$

# Distribución a priori conjugada

- La *distribución posterior condicional* de  $\mu$  dado  $\sigma^2$  es:

$$\begin{aligned}\mu|\sigma^2, \mathbf{y} &\sim \mathcal{N}\left(\mu_n, \frac{\sigma^2}{k_n}\right) \\ &\sim \mathcal{N}\left(\frac{\frac{k_0}{\sigma^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{k_0}{\sigma^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{k_0}{\sigma^2} + \frac{n}{\sigma^2}}\right)\end{aligned}$$

- La *distribución posterior marginal* de  $\sigma^2$  es:

$$\sigma^2|\mathbf{y} \sim \chi^2 - \text{inversa de escala } (v_n, \sigma_n^2)$$

# Distribución a priori conjugada

- La *distribución posterior marginal* de  $\mu$  es:

$$f(\mu|\mathbf{y}) \propto \left[ 1 + \frac{k_n (\mu - \mu_n)^2}{v_n \sigma_n^2} \right]^{-(v_n+1)/2}$$
$$\sim t_{v_n} \left( \mu_n, \frac{\sigma_n^2}{k_n} \right)$$

- Un intervalo de credibilidad al  $(1 - \alpha) \times 100\%$  para  $\mu$  es:

$$\mu_n - t_{v_n, 1-\alpha/2} \frac{\sigma_n}{\sqrt{k_n}} \leq \mu \leq \mu_n + t_{v_n, 1-\alpha/2} \frac{\sigma_n}{\sqrt{k_n}}$$

## Ejemplo: Rendimiento de un proceso químico

- Se sabe que el rendimiento porcentual de un proceso químico es aproximadamente 12 %. Sin embargo se han observado rendimientos por debajo del valor mencionado.
- Para comprobar si el rendimiento ha disminuido se toma una muestra aleatoria de tamaño 20 de un lote de materia prima y se obtuvo una media igual a 11.5 y una desviación estándar de 0.3567.
- Suponga que el rendimiento porcentual tiene distribución normal y que la distribución a priori conjunta es:

$$\mu, \sigma^2 \sim \mathcal{N}\chi^2 \left( \mu_0 = 12, \frac{\sigma_0^2}{k_0} = \frac{4}{2}, v_0 = 3, \sigma_0^2 = 4 \right)$$

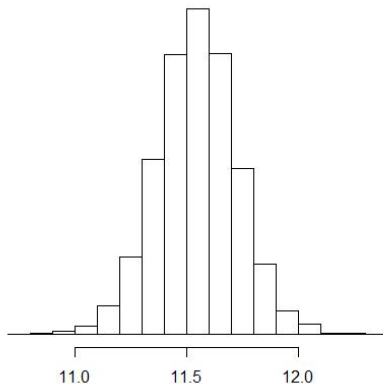
## Ejemplo: Rendimiento de un proceso químico

### Distribución posterior usando simulación

```
> library(geoR)
> sigma2 <- rinvchisq(n = 10000, df = 23, scale = 0.64660906)
> mu <- rnorm(n = 10000, mean = 11.545454, sd =
sqrt(sigma2/22))
> mean(mu)+c(-1, 1)*qt(p = 0.975,df = 23)*sd(mu)
[1] 11.17601 11.91572
> hist(mu, freq=F, yaxt="n", ylab="", xlab="", main="")
```

## Ejemplo: Rendimiento de un proceso químico

Figura 3: Histograma para los valores simulados



# Distribución a priori conjugada

- Si  $\mathbf{y} = (y_1, \dots, y_k)$  es el vector de observaciones entonces:

$$\mathbf{y}|\boldsymbol{\theta} \sim \mathcal{M}(n, \boldsymbol{\theta})$$

donde  $\sum_{j=1}^k \theta_j = 1$  y  $\sum_{j=1}^k y_j = n$ .

- La distribución a priori conjugada es Dirichlet:

$$\boldsymbol{\theta} \sim \mathcal{D}(\alpha_1, \dots, \alpha_k)$$

- La distribución posterior es:

$$\boldsymbol{\theta}|\mathbf{y} \sim \mathcal{D}(\alpha_1 + y_1, \dots, \alpha_k + y_k)$$

## Ejemplo: Intención de voto

- Se realizó una encuesta en abril del 2011 para analizar la intención de voto presidencial en la segunda vuelta de las últimas elecciones.
- De las 1570 personas encuestadas  $y_1 = 660$  apoyaron a Ollanta Humala,  $y_2 = 581$  a Keiko Fujimori y  $y_3 = 329$  votarían en blanco o anularían su voto.
- Si se asume que el muestreo es aleatorio entonces:

$$(y_1, y_2, y_3) \sim \mathcal{M}(n = 1570, \theta_1, \theta_2, \theta_3)$$

- Una estimación de interés sería  $\theta_1 - \theta_2$  la diferencia entre quienes apoyan a los dos principales candidatos.



## Ejemplo: Intención de voto

- Considerando una distribución a priori no informativa con  $\alpha_1 = \alpha_2 = \alpha_3 = 1$  la distribución posterior es:

$$\theta | \mathbf{y} \sim \mathcal{D}(y_1 + 1 = 661, y_2 + 1 = 582, y_3 + 1 = 330)$$

- Se podría obtener la distribución posterior de  $\theta_1 - \theta_2$  por simulación.

### Distribución posterior usando simulación

```
> library(mixAK)
> theta <- rDirichlet(n = 10000, alpha = c(661, 582, 330))
> hist(theta[,1] - theta[,2], freq = F, yaxt = "n", ylab = "", xlab = "", main = "")
```

## Ejemplo: Intención de voto

Figura 4: Histograma para los valores simulados de  $\theta_1 - \theta_2$

