

# Meta-análisis para pocos estudios y presencia de ceros: un enfoque bayesiano.

Francisco-José Vázquez-Polo, Elías Moreno, Miguel A. Negrín y María Martel

<sup>1</sup>Departamento de Métodos Cuantitativos en Economía y Gestión.  
Universidad de Las Palmas de Gran Canaria.

<sup>2</sup>Departamento de Estadística. Universidad de Granada.

**Workshop Métodos Bayesianos'17.Madrid**



## Guión

### Introducción

### Modelo

- Meta-análisis para datos binarios

- Clustering

- Test de igualdad entre tratamientos

### Ejercicio de simulación

### Ejemplos con datos reales

- Canner (1987)

- Sutton et al. (2002)

### Discusión

### Líneas abiertas

## Introducción

- La comparación de  $k$  muestras independientes es un problema estadístico muy frecuente en salud: meta-análisis, análisis multicéntricos, variación en la práctica clínica, etc.
- Nosotros centraremos esta presentación en su aplicación en meta-análisis.
- El objetivo debe ser por tanto doble:
  - Analizar la heterogeneidad entre muestras ( $\theta_j$ )
  - Estimar el parámetro poblacional no condicionado al centro (meta-parámetro:  $\theta$ )

## Análisis convencional. Caso normal.

$$x_{ij} \sim N(x_{ij}|\theta_j, \sigma_j^2), j = 1, \dots, k \quad k \geq 2$$
$$N(\theta_j|\theta, \tau^2)$$

- DerSimonian y Laird (1986).
- La heterogeneidad entre muestras se mide a través del parámetro  $\tau^2$ .
- Es habitual contrastar si  $\tau^2 = 0$ , frente a la alternativa  $\tau^2 > 0$ . Esto equivale a:

$$H_0 = \theta_1 = \theta_2 = \dots = \theta_k$$

$$H_1 = \theta_1 \neq \theta_2 \neq \dots \neq \theta_k$$

- Índice  $I^2$  (Julian et al, 2003)

## Análisis convencional. Caso binomial.

$$x_j \sim \text{Bin}(x_j | \theta_j, n_j)$$

- Transformación: log odds ratio.

$$y_j = \log \frac{x_j}{n_j - x_j} \quad \text{with} \quad \eta_j = \log \frac{\theta_j}{1 - \theta_j}$$

- Aplicar el modelo normal a la transformación anterior.

$$y_j \sim \mathcal{N}(y_j | \eta_j, \sigma_j^2) \quad \text{Links:} \quad \eta_j | \eta \sim \mathcal{N}(\eta_j | \eta, \tau^2), \quad i = 1, \dots, k,$$

- Cuando  $x_j = 0$  se obvia dicho centro o se aplica alguna corrección de continuidad (0.5, 0.1, ...)



## Análisis convencional.

- Esta formulación sólo considera casos de heterogeneidad extrema, y existen otras situaciones intermedias posibles que son ignoradas.
- En este trabajo proponemos modelizar estas situaciones intermedias como un problema de selección de modelos y analizar el impacto de incorporar esta incertidumbre sobre la heterogeneidad muestral en los resultados finales.
- Consideraremos el análisis para datos binarios (respuesta 0 y 1) sin necesidad de la utilización de transformaciones.
- Malec y Sendrask (1992), Consonni and Veronese (1995) y Evans and Sendrask (2001) han tratado este problema aunque centrados en la estimación de  $\theta_j$ .

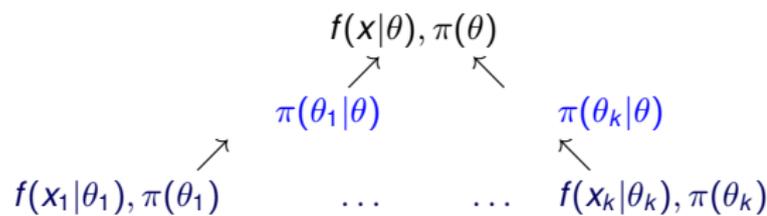


## Esquema del Meta-modelo

Meta-modelo

Distribución Link

Modelos experimentales



$\theta_j$  indica la efectividad del tratamiento, condicionada al centro  $j$  (i.i.d).

$\theta$  indica la efectividad no condicionada al centro (meta-parámetro).



## Meta-modelo datos binarios

- El modelo experimental

$$\text{Bin}(x_j | n_j, \theta_j) = \binom{n_j}{x_j} \theta_j^{x_j} (1 - \theta_j)^{n_j - x_j}, \quad \pi(\theta_j) = \text{Beta}(\theta_j | 1/2, 1/2).$$

Análisis Bayesiano objetivo con distribución a priori de Jeffreys.

- El meta-modelo

$$\text{Ber}(x | \theta) = \theta^x (1 - \theta)^{1-x}, \quad \pi(\theta) = \text{Beta}(\theta | 1/2, 1/2)$$

- ¿Qué distribución Link utilizar?

$$\pi(\theta_j | \theta)$$



## Meta-modelo datos binarios

### Condiciones de la distribución Link

- Asumimos que las distribuciones de referencia  $\pi(\theta_j)$  y  $\pi(\theta)$  son distribuciones a priori de Jeffreys. La distribución Link  $\pi(\theta_j|\theta)$  debe cumplir por tanto las siguientes condiciones:

1.

$$\int_0^1 \pi(\theta_j, \theta) d\theta_j = \pi(\theta), \quad \int_0^1 \pi(\theta_j, \theta) d\theta = \pi(\theta_j)$$

2.  $\pi(\theta_j|\theta)$  debe permitir distintos grados de concentración de los parámetros  $\theta_j$  alrededor de  $\theta$ .

## Clase de distribuciones Intrínsecas

- Clase de distribuciones **Intrínsecas** con distribuciones marginales uniformes (Berger y Pericchi, 1996; Moreno, 1997; Moreno et al., 1998)

$$\pi^I(\theta_j|\theta, t) = \pi^R(\theta_j) E_{z_1, \dots, z_t|\theta_j} \frac{f(z_1, \dots, z_t|\theta)}{\int f(z_1, \dots, z_t|\theta_j) \pi^R(\theta_j) d\theta_j},$$

$$\pi^I(\theta_j|\theta, t) = \sum_{z=0}^t \text{Beta}(\theta_j|z+1/2, t-z+1/2) \text{Bin}(z|t, \theta), \quad t = 1, 2, \dots$$

## Propiedades Link

- A medida que aumenta  $t$ , la concentración de  $\theta_j$  alrededor de  $\theta$  incrementa.
- Se puede observar que

$$\rho(\theta_j, \theta|t) = \frac{t}{2+t}, \quad \text{and} \quad \frac{1}{3} \leq \rho(\theta_j, \theta|t) \leq 1 \text{ for } t \in \{1, 2, \dots\}$$

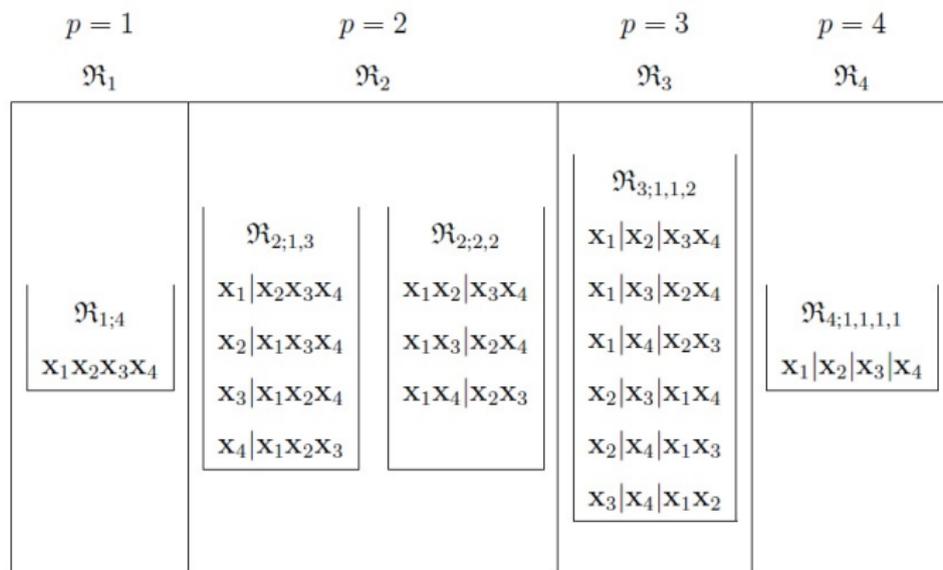
## Clustering

- Definición de cluster: dos muestras  $i$  y  $j$ , para  $i \neq j$ , están en el mismo cluster si  $\theta_i = \theta_j$ .
- El clustering puede plantearse como un problema de selección de modelos desde una perspectiva Bayesiana (Casella et al., 2014)
- Lo plantearemos como un problema de particiones (Hartigan, 1990).
- Cada cluster vendrá definido por los parámetros  $p$  y  $\mathbf{r}_p = (r_1, \dots, r_p)$ , donde  $p$  señala el número de clusters y  $r_i$  indica el cluster en el que se sitúa el centro  $i$ . Por ejemplo, la partición  $\mathbf{r}_2 = (1, 2, 1, 2)$  indicaría el cluster  $\{x_1, x_3\}$ ,  $\{x_2, x_4\}$ .



## Clustering

Ejemplo para 4 muestras:





## Clustering

- Se utilizará la estructura a priori uniforme jerárquica (Casella et al., 2014) para: clase, subclase y partición.

$$\pi^{HU}(p, \mathbf{r}_p | k) = \pi(p, \mathbf{r}_p | \mathfrak{R}_{p;k_1, \dots, k_p}, k) \pi(\mathfrak{R}_{p;k_1, \dots, k_p} | \mathfrak{R}_p, k) \pi(p | k),$$

- Por ejemplo, la distribución a priori de la partición

$$x_1, x_2 | x_3, x_4 \text{ será: } \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{24}$$

$p = 1$	$p = 2$		$p = 3$	$p = 4$
$\mathfrak{R}_1$	$\mathfrak{R}_2$		$\mathfrak{R}_3$	$\mathfrak{R}_4$
$\mathfrak{R}_{1;4}$ $x_1 x_2 x_3 x_4$ <b>1/4</b>	$\mathfrak{R}_{2;1,3}$ $x_1   x_2 x_3 x_4$ $x_2   x_1 x_3 x_4$ $x_3   x_1 x_2 x_4$ $x_4   x_1 x_2 x_3$ <b>1/32</b>	$\mathfrak{R}_{2;2,2}$ $x_1 x_2   x_3 x_4$ $x_1 x_3   x_2 x_4$ $x_1 x_4   x_2 x_3$ <b>1/24</b>	$\mathfrak{R}_{3;1,1,2}$ $x_1   x_2   x_3 x_4$ $x_1   x_3   x_2 x_4$ $x_1   x_4   x_2 x_3$ $x_2   x_3   x_1 x_4$ $x_2   x_4   x_1 x_3$ $x_3   x_4   x_1 x_2$ <b>1/24</b>	$\mathfrak{R}_{4;1,1,1,1}$ $x_1   x_2   x_3   x_4$ <b>1/4</b>

## Verosimilitud del meta-parámetro

- La verosimilitud del meta-parámetro  $\theta$ , condicionada a un determinado cluster:

$$f(\mathbf{x}|\rho, \mathbf{r}_\rho, \theta, k, t) = \prod_{j=1}^p \int \left( \prod_{i:r_i=j} f(\mathbf{x}_i|\theta_j) \right) \pi^l(\theta_j|\theta, t) d\theta_j =$$

$$= \prod_{j=1}^p \left\{ \frac{\Gamma(t+1)}{\Gamma(t+1/2)\Gamma(1/2)} \frac{\Gamma(\mathbf{x}_j+1/2)\Gamma(m_j+t-\mathbf{x}_j+1/2)}{\Gamma(m_j+t+1)} {}_3F_2 \left( \mathbf{a}, \mathbf{b}, \frac{\theta}{\theta-1} \right) (1-\theta)^t, \right.$$

donde  $\mathbf{x}_j = \sum_{i:r_i=j} \mathbf{x}_i$ ,  $m_j = \sum_{i:r_i=j} n_i$ ,  
 $\mathbf{a} = (1/2 - t, -t, 1/2 + \mathbf{x}_j)$  y  $\mathbf{b} = (1/2, 1/2 - m_j - t + \mathbf{x}_j)$ ,  
 para  $j = 1, \dots, p$ .

## Cantidades de interés

- Una vez definida la distribución a priori uniforme jerárquica para los cluster, la verosimilitud del meta-parámetro  $\theta$  vendrá dada por:

$$f(\mathbf{x}|\theta, k, t) = \sum_{\rho=1}^k \sum_{\mathbf{r}_\rho \in \mathfrak{R}_\rho} f(\mathbf{x}|\rho, \mathbf{r}_\rho, \theta, k, t) \pi^{HU}(\rho, \mathbf{r}_\rho|k).$$

- La distribución a posteriori del meta-parámetro  $\theta$  se obtiene a partir de la expresión:

$$\pi(\theta|\mathbf{x}, k, t) = \sum_{\rho=1}^k \sum_{\mathbf{r}_\rho \in \mathfrak{R}_\rho} \pi(\theta|\mathbf{x}, \rho, \mathbf{r}_\rho, k, t) \Pr(\rho, \mathbf{r}_\rho|\mathbf{x}, k).$$



## Cantidades de interés

- La distribución predictiva de la meta-variable  $x$  para  $x = 1$  es:

$$E(\theta|\mathbf{x}, t) = \sum_{p=1}^k \sum_{\mathbf{r}_p \in \mathfrak{R}_p} E(\theta|\mathbf{x}, p, \mathbf{r}_p, k, t) \Pr(p, \mathbf{r}_p|\mathbf{x}, t, k).$$

## Comparación de tratamientos: Test de igualdad

- Test de igualdad sobre el meta-parámetro para distintos tratamientos

$$M_0 : \left\{ f(\mathbf{x}_1|\xi, k, t)f(\mathbf{x}_2|\xi, k, t), \pi^R(\xi) \right\}$$

vs.

$$M_1 : \left\{ f(\mathbf{x}_1|\theta, k, t)f(\mathbf{x}_2|\mu, k, t), \pi^R(\theta)\pi^R(\mu) \right\}.$$

- Siendo el factor Bayes:

$$B_{10}(\mathbf{x}_1, \mathbf{x}_2, k, t) = \frac{\int f(\mathbf{x}_1|\theta, k, t)\pi^R(\theta)d\theta \int f(\mathbf{x}_2|\mu, k, t)\pi^R(\mu)d\mu}{\int f(\mathbf{x}_1|\xi, k, t)f(\mathbf{x}_2|\xi, k, t)\pi^R(\xi)d\xi},$$

## Simulación

- Consideramos seis muestras independientes  $\mathbf{x} = (x_1, \dots, x_6)$ , con tamaño muestral  $n_j = 100$  para  $j = 1, \dots, 5$  y  $n_6 = 500$ . Las muestras se simulan a partir de seis distribuciones binomiales con parámetros  $\theta_j = 0.1$  for  $j = 1, \dots, 5$  y  $\theta_6 = 0.7$ . Las primeras 5 muestras contienen cierta cantidad de ceros.
- Para  $k = 6$  hay un total de 203 posibles agrupaciones (cluster).
- Repetimos las simulaciones 500 veces y estimamos la probabilidad a posteriori de cada posible cluster, así como la estimación del meta-parámetro  $\theta$ .

## Simulación. Resultados.

En 485 de las 500 repeticiones el modelo más probable fue el verdadero  $\{X_1, X_2, X_3, X_4, X_5\}$ ,  $\{X_6\}$ .

**Table 3.** Mean of the posterior probabilities of the top cluster models across the 500 simulations.

Top cluster model	Mean of the posterior probability
$\{1, 2, 3, 4, 5\}, \{6\}$ (true model)	0.87
$\{1\}, \{2, 3, 4, 5\}, \{6\}$	0.015
the rest	< 0.01
Other models	
$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$ (full heterogeneity)	0.005
$\{1, 2, 3, 4, 5, 6\}$ (homogeneity)	< $10^{-4}$

## Simulación. Resultados.

**Table 4.** Mean of the posterior expectation of the meta-parameter under three selected models for 500 simulations.

Cluster model	Mean of the posterior expectation of $\theta$
True model: {1, 2, 3, 4, 5}, {6}	0.37
Mixture model	0.36
Full heterogeneity: {1}, {2}, {3}, {4}, {5}, {6}	0.18

## Simulación.

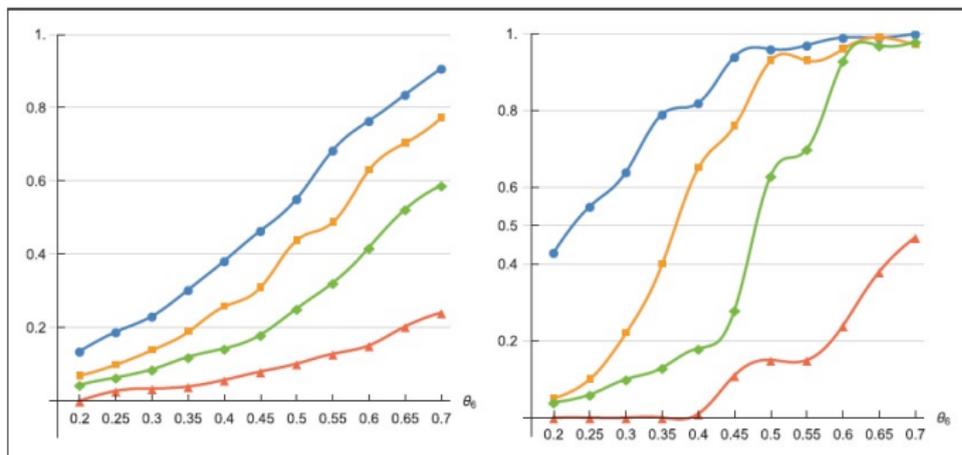
- Modificamos la simulación anterior para distintos valores de  $\theta_6$  desde 0.2 hasta 0.7.

$$\theta_1 = \dots = \theta_5 = 0.1, \theta_6 \in (0.2, 0.7)$$

- Modificamos los tamaños muestrales, considerando 4 casos:

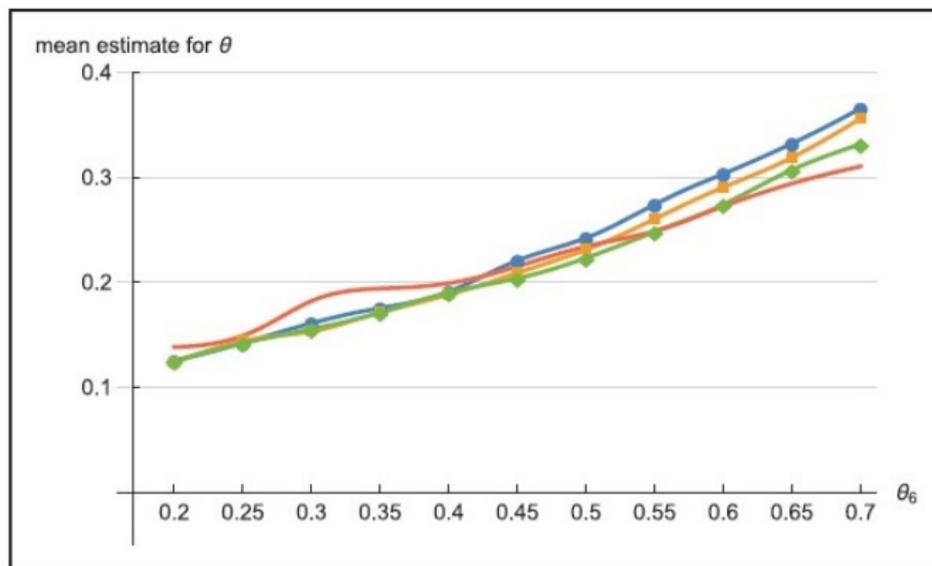
1.  $n_1 = \dots = n_5 = 10, n_6 = 50$
2.  $n_1 = \dots = n_5 = 25, n_6 = 125$
3.  $n_1 = \dots = n_5 = 50, n_6 = 250$
4.  $n_1 = \dots = n_5 = 100, n_6 = 500$

## Simulación. Resultados.



**Figure 2.** **Left panel:** Behaviour of the mean of the posterior probabilities of the true model across the 500 simulations for different values of  $\theta_0$  from 0.2 to 0.7 by step of 0.05. **Right panel:** Average rate of success of the true model across the 500 simulations. In both pictures, symbols  $\Delta$ ,  $\diamond$ ,  $\square$  and  $\circ$  refer to  $n_0 = 50, 125, 250$ , and 500, respectively.

## Simulación. Resultados.



**Figure 3.** Behaviour of the mean estimate for  $\theta$  of the true model across the 500 simulations for different values of  $\theta_6$  from 0.2 to 0.7 by step of 0.05. Symbols  $\Delta$ ,  $\diamond$ ,  $\square$  and  $\circ$  refer to  $n_6 = 50, 125, 250$ , and 500, respectively.

## Ejemplo 1. Canner, 1987

- Analizar si la tasa de mortalidad por infarto de miocardio se reduce con una dosis diaria de aspirina.

Trial	Aspirin		Placebo	
	$n_i$	$x_i$	$n_i$	$x_i$
UK-1	615	49	624	67
CDPA	758	44	771	64
GAMS	317	27	309	32
UK-2	832	102	850	126
PRIS	810	85	406	52
AMIS	2267	246	2257	219

## Ejemplo 1. Canner, 1987

- Clusters más probables para cada uno de los tratamientos

Aspirin		Placebo	
Cluster Model	Post. Prob.	Cluster Model	Post. Prob.
{1,2,3},{4,5,6}	0.25	{1,2,3,6},{4,5}	0.21
{1,2},{3,4,5,6}	0.11	{1,2,3,5,6},{4}	0.14
{2},{1,3,4,5,6}	0.08	{1,2,6},{3,4,5}	0.06
the rest	< 0.05	{1,2,3,6},{4,5}	0.05
		the rest	< 0.05

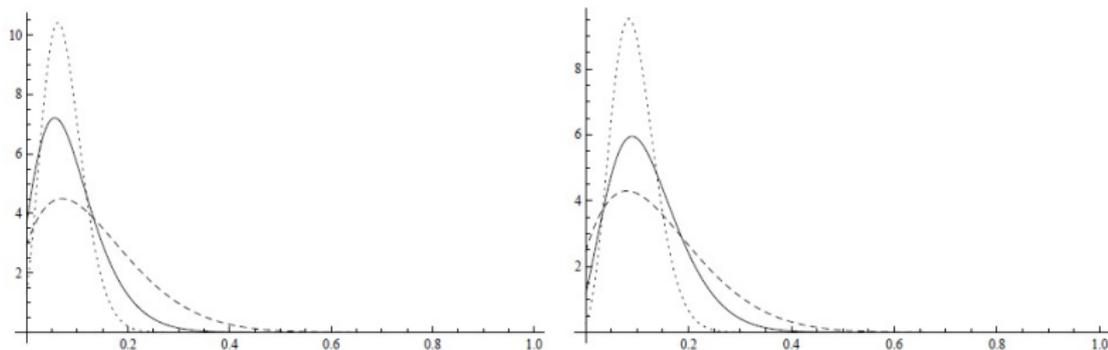
## Ejemplo 1. Canner, 1987

- Estimación del meta-parámetro para cada cluster.

Aspirin		Placebo	
Cluster	$E(\theta \text{Cluster})$	Cluster	$E(\theta \text{Cluster})$
{1,2,3},{4,5,6}	0.098	{1,2,3,6},{4,5}	0.126
{1,2},{3,4,5,6}	0.095	{1,2,3,5,6},{4}	0.130
{2},{1,3,4,5,6}	0.089	{1,2,6},{3,4,5}	0.122
		{1,4,5},{2,3,6}	0.120

## Ejemplo 1. Canner, 1987

- Distribución a posteriori del meta-parámetro para el clustering y los casos extremos de homogeneidad y heterogeneidad perfecta



- Las estimaciones del meta-parámetro:  
 $E[\theta_{aspirin} | \text{datos}] = 0.09, E[\theta_{placebo} | \text{datos}] = 0.122$
- Test de igualdad: 0.814

## Ejemplo 2. Sutton et al, 2002

- Infecciones en cesáreas. Uso de antibióticos vs. placebo.

Trial	Antibiotics		Placebo	
	$n_i$	$x_i$	$n_i$	$x_i$
1	100	3	33	0
2	11	1	17	5
3	42	0	40	0
4	167	4	140	5
5	36	1	25	1
6	115	12	117	15
7	16	0	16	1

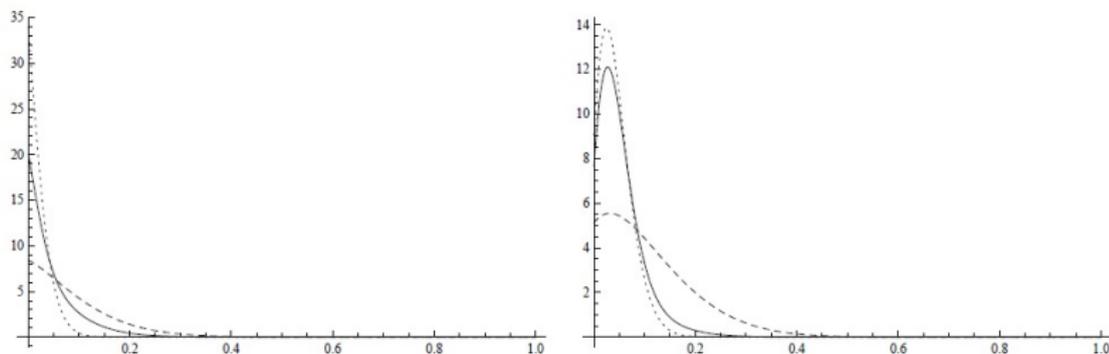
## Ejemplo 2. Sutton et al, 2002

- Clusters más probables para cada uno de los tratamientos

Antibiotics		Placebo	
Cluster Model	Post. Prob.	Cluster Model	Post. Prob.
{1},{2},{3},{4},{5},{6},{7}	0.13	{1},{2},{3},{4},{5},{6},{7}	0.26
{1, 2, 3, 4, 5, 7},{6}	0.08	{1, 3},{2},{4},{5},{6},{7}	0.04
{1, 3, 4, 5,7},{2, 6}	0.04	{1, 3, 4, 5, 7}, {2, 6}	0.03
{1, 2, 3, 4, 5, 6, 7}	0.02	{1, 3, 4, 5, 7},{2},{6}	0.03
the rest	< 0.02	the rest	< 0.03

## Ejemplo 2. Sutton et al, 2002

- Distribución a posteriori del meta-parámetro para el clustering y los casos extremos de homogeneidad y heterogeneidad perfecta



- Las estimaciones del meta-parámetro:  
 $E[\theta_{antibiotics} | \text{datos}] = 0.047, E[\theta_{placebo} | \text{datos}] = 0.055$
- Test de igualdad: 0.90



## Discusión

- El clustering puede interpretarse como un problema de selección de modelos bayesiano
- Los resultados principales que obtendríamos son:
  1. La probabilidad a posteriori de cada posible cluster.
  2. La estimación del meta-parámetro condicionado a cada cluster.
  3. La estimación del meta-parámetro no-condicionado.
  4. Test de hipótesis de igualdad para el meta-parámetro de 2 tratamientos.



## Líneas abiertas

- Estudiar la sensibilidad de los resultados al uso de la distribución a priori jerárquica para los clustering.
- Resolver el problema para muchas muestras ( $k > 10$ )

