

# Análisis bayesiano de datos experimentales

Diego Salmerón Martínez  
dsm@um.es

<http://webs.um.es/dsm/miwiki/doku.php>

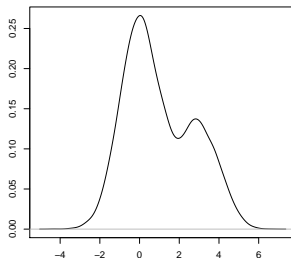
- ▶ Repaso de variables aleatorias
- ▶ Introducción
- ▶ Estimación a posteriori. Distribuciones a priori no informativas
- ▶ Análisis bayesiano de muestras con distribución normal
- ▶ Métodos de Monte Carlo (MC) y de Monte Carlo y Cadenas de Markov (MCMC)
- ▶ Aspectos computacionales con R y WinBUGS
- ▶ Inferencia bayesiana en modelos lineales y en modelos de regresión logística

# Repaso de variables aleatorias

# Variable aleatoria de tipo continuo

$X$  es una v.a. 1-dim ( $X \in \mathbb{R}$ ) con función de densidad  $f(x)$

$$X \sim f(x) \quad P(X < a) = \int_{x < a} f(x) dx$$



# Variable aleatoria de tipo discreto

$$X \in \{a_1, a_2, \dots\}$$

$$a_i \in \mathbb{R}, \forall i = 1, 2, \dots$$

v.a. con función puntual de probabilidad  $f(x)$

$$f(a_i) = P(X = a_i), \quad i = 1, 2, \dots$$

$$X \sim f(x) \qquad P(X \leq a) = \sum_{a_i \leq a} f(a_i)$$

# Dos características de las variables aleatorias

**Media** de  $X$

$$E(X) = \int xf(x)dx$$

$$E(X) = \sum_i a_i f(a_i)$$

**Varianza** de  $X$

$$Var(X) = E((X - E(X))^2) = \int (x - E(X))^2 f(x) dx$$

$$Var(X) = E((X - E(X))^2) = \sum_i (a_i - E(X))^2 f(a_i)$$

# Dos variables aleatorias

$(X, Y)$  es una v.a. con función de densidad  $f(x, y)$

$$(X, Y) \sim f(x, y)$$

$$f_X(x) = \int f(x, y) dy, \quad f_Y(y) = \int f(x, y) dx$$

$$f_{X|Y}(x | y) = \frac{f(x, y)}{f_Y(y)}, \quad f_{Y|X}(y | x) = \frac{f(x, y)}{f_X(x)}$$

$$f_X(x) = f(x), \quad f_Y(y) = f(y)$$

$$f_{X|Y}(x | y) = f(x | y), \quad f_{Y|X}(y | x) = f(y | x)$$

$$f(x | y) = \frac{f(x, y)}{f(y)}, \quad f(y | x) = \frac{f(x, y)}{f(x)}$$



# Multivariate

$$(X_1, \dots, X_k) \sim f(x_1, \dots, x_k)$$

$$f(x_1) = \int f(x_1, \dots, x_k) dx_2 \dots dx_k$$

$$f(x_1, x_2) = \int f(x_1, \dots, x_k) dx_3 \dots dx_k$$

$$f(x_1 \mid x_2, \dots, x_k) = \frac{f(x_1, \dots, x_k)}{f(x_2, \dots, x_k)}$$

►  $X \sim f(x)$

►  $h(x)$  una función tal que  $Y = h(X)$  es v.a. 1-dim

$$E(Y) = \int y f_Y(y) dy$$

$$E(Y) = \int h(x) f(x) dx$$

# Introducción

# Inferencia paramétrica. Modelo estadístico.

Los datos  $\mathbf{x} = (x_1, \dots, x_n)$  son la observación de una variable aleatoria

El punto de partida en inferencia paramétrica es un modelo de probabilidad para los datos  $\mathbf{x} = (x_1, \dots, x_n)$

Generalmente dicho modelo vendrá definido mediante ecuaciones con variables aleatorias, o por una función de densidad  $f(\mathbf{x}|\theta)$  parametrizada por el parámetro  $\theta \in \Theta \subseteq \mathbb{R}^k$

$$\mathbf{x}|\theta \sim f(\mathbf{x}|\theta), \quad \theta \in \Theta$$

# Objetivo

Los datos han sido generados según el modelo de probabilidad

$$\mathbf{x}|\theta \sim f(\mathbf{x}|\theta), \quad \theta \in \Theta$$

para algún valor desconocido de  $\theta$

**Dados los datos  $\mathbf{x}$** , el objetivo es estimar  $\theta$ , evaluar hipótesis sobre  $\theta$ , y hacer predicciones

# Algunos ejemplos

- ▶ Muestra de la  $N(\mu, \sigma^2)$
- ▶ Modelo de regresión lineal
- ▶ Modelos para Expresión diferencial de genes
- ▶ Cinética de la degradación de los alimentos

## Muestra de la $N(\mu, \sigma^2)$

Una muestra de una distribución normal  $N(\mu, \sigma^2)$

$$x_i = \mu + \sigma \varepsilon_i, \quad \varepsilon_i \sim N(0, 1), \quad i = 1, \dots, n$$

$$x_i \mid \mu, \sigma \sim N(\mu, \sigma^2), \quad i = 1, \dots, n$$

Función de densidad

$$\begin{aligned} f(x_1, \dots, x_n \mid \mu, \sigma) &= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \\ &= c\sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2\right) \end{aligned}$$

## Modelo de regresión lineal

$$y_i = \beta_0 + x_{i1}\beta_1 + \dots + x_{ik}\beta_k + \varepsilon_i, \quad i = 1, \dots, n$$

$$\varepsilon_1, \dots, \varepsilon_k \sim N(0, \sigma^2)$$

Función de densidad

$$f(\mathbf{y} \mid \mathbf{X}, \beta, \sigma) = c\sigma^{-n} \exp \left( -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \right)$$



## Expresión diferencial de genes

$G$  genes

Expresión de los genes: valores numéricos

Dos grupos (p. ej., enfermos y no enfermos)

¿Los niveles medios de expresión del gen  $g$  son diferentes entre los grupos?

$$\mu_{g1} \neq \mu_{g2}?$$

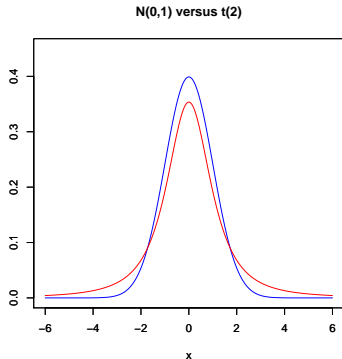
¿Que implicaciones tendría que  $\mu_{g1} \neq \mu_{g2}$ ?

Hay varios pasos para obtener la expresión genética usando microarrays: desde la hibridación hasta el análisis de imagen

Los datos suelen presentar **Outliers**

La distribución Normal puede no ser apropiada

La distribución  $t$  es una alternativa pues presenta colas más pesadas



$$y_{gij} = \mu_{gi} + \frac{\epsilon_{gij}}{\sqrt{w_{gij}}}$$

$$\epsilon_{gij} \sim N(0, \sigma_{gi}^2), \quad w_{gij} \sim \text{Gamma}(\nu_j/2, \nu_j/2)$$

$$g = 1, \dots, G, \quad i = 1, 2, \quad j = 1, \dots, J$$

Gottardo et al. (2006). Bayesian Robust Inference for Differential Gene Expression in Microarrays with Multiple Samples. *Biometrics* 62, 10–18.

## Cinética de la degradación de los alimentos

La cinética de la degradación ruminal de la materia, el nitrógeno y otros constituyentes de la pared celular, suele describirse con modelos no lineales

$$y_i = a + b(1 - e^{-ct_i}) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, N$$

$y_i$  es la proporción degradada hasta el instante de tiempo  $t_i$

$a$  representa el sustrato soluble y completamente degradable

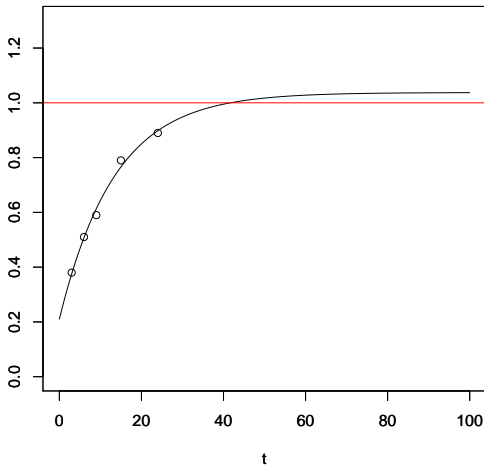
$b$  representa la fracción insoluble pero potencialmente degradable del sustrato

$c$  representa la velocidad de degradación

$$0 < a, b, a + b < 1, c > 0$$

$t = (3, 6, 9, 15, 24)$   $y = (0.38, 0.51, 0.59, 0.79, 0.89)$

Ajuste de mínimos cuadrados



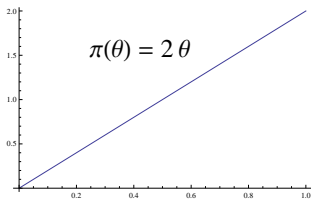
Ørskov ER, McDonald I. The estimation of protein degradability in the rumen from incubation measurements weighted according to rate of passage. J. Agric. Sci. Camb. 1979;92:499–503.

# Inferencia bayesiana

- ▶ El parámetro  $\theta$  es desconocido, tenemos incertidumbre sobre  $\theta$
- ▶ Modelamos  $\theta$  como una v.a. mediante una distribución de probabilidad  $\pi(\theta)$  llamada **distribución a priori**
- ▶ Las conclusiones sobre  $\theta$  se realizan en términos de probabilidades mediante la **distribución a posteriori**  $\pi(\theta \mid \mathbf{x})$

La distribución a priori  $\pi(\theta)$  refleja información sobre los valores que puede tomar  $\theta$

►  $\pi(\theta) = 2\theta, 0 < \theta < 1$



►  $\theta \sim N(0, \sigma_0^2)$

$\theta$  está alrededor de 0,  $P(-4\sigma_0 < \theta < 4\sigma_0) \approx 0.99$ , y es igual de fácil que sea positivo que negativo

# Algo en lo que pensar

¿Qué probabilidad aplicamos a una persona SI observamos que es fumadora?

$P(\text{Cáncer pulmón})$

$P(\text{Cáncer pulmón} \mid \text{fumador})$

$P(\text{Cáncer pulmón} \mid \text{No fumador})$





# Un primer ejemplo de inferencia bayesiana

- ▶ Lanzar un dado y anotar  $X = 1$  si ha salido un número par y  $X = 0$  si ha salido impar

$$p = P(X = 1)$$

- ▶ Dicho experimento lo realizamos  $n$  veces de manera independiente y bajo las mismas condiciones obteniendo los datos  $\mathbf{x} = (x_1, \dots, x_n)$ , donde  $x_i \in \{0, 1\}$ , y por tanto

$$f(\mathbf{x}|p) = p^{n\bar{\mathbf{x}}}(1 - p)^{n(1-\bar{\mathbf{x}})}$$

donde  $\bar{\mathbf{x}} = \sum_{i=1}^n x_i / n$ .

# Aplicamos el razonamiento bayesiano

- ▶ Parámetro:  $p \in (0, 1)$
- ▶ Modelamos  $p \sim U(0, 1)$ :  $\pi(p) = 1$
- ▶  $(\mathbf{x}, p)$  es una variable aleatoria

¿Qué probabilidad aplicamos a una persona SI observamos que es fumadora?

$P(\text{Cáncer pulmón})$

$P(\text{Cáncer pulmón} \mid \text{fumador})$

$P(\text{Cáncer pulmón} \mid \text{No fumador})$



Si hemos observado  $\mathbf{x}$ , la incertidumbre sobre el parámetro desconocido  $p$  la tenemos cuantificada mediante

$$\pi(p|\mathbf{x}) = \frac{f(\mathbf{x}, p)}{m(\mathbf{x})} \quad (1)$$

donde  $m(\mathbf{x})$  es la densidad marginal de  $\mathbf{x}$

La densidad

$$\pi(p|\mathbf{x}) = \frac{f(\mathbf{x}, p)}{m(\mathbf{x})} = \frac{f(\mathbf{x}|p)\pi(p)}{m(\mathbf{x})}$$

contiene información sobre el parámetro  $p$  después de observar los datos  $\mathbf{x}$ , es decir *a posteriori*

# Definición

Si el modelo paramétrico para los datos es

$$\mathbf{x}|\theta \sim f(\mathbf{x}|\theta), \quad \theta \in \Theta$$

y la distribución a priori es  $\pi(\theta)$ , entonces la distribución a posteriori de  $\theta$  viene dada por

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{m(\mathbf{x})}, \quad \theta \in \Theta \quad (2)$$

## Ejemplo

$$f(\mathbf{x}|p) = p^{n\bar{\mathbf{x}}}(1-p)^{n(1-\bar{\mathbf{x}})}$$

$$\pi(p) = \begin{cases} 1 & \text{if } p \in [0, 1] \\ 0 & \text{if } p \notin [0, 1] \end{cases}$$

$$\pi(p|\mathbf{x}) = \frac{f(\mathbf{x}|p)\pi(p)}{m(\mathbf{x})} \propto p^{n\bar{\mathbf{x}}}(1-p)^{n(1-\bar{\mathbf{x}})}$$

$$\mathcal{B}(p \mid a, b) = cp^{a-1}(1-p)^{b-1}$$

Distribución Beta de parámetros  $a = n\bar{\mathbf{x}} + 1$  y  $b = n(1 - \bar{\mathbf{x}}) + 1$

## Práctica

Usando R, vamos a simular una muestra (los datos)  $\mathbf{x}$  con  $n=10$ , 20, 50, y 200, tomando  $p = 0.4$ , y pintaremos la distribución a posteriori de  $p$

```
n=10;x=rbinom(n,1,0.4)
```

```
#Prior and Posterior distribution#
```

```
a=sum(x)+1;b=n*(1-mean(x))+1
```

```
p=seq(0,1,.005)
```

```
prior=rep(1,length(p))
```

```
y=dbeta(p,a,b,ncp=0,log=F)
```

```
plot(p,prior,ylim=c(0,max(y)),type="l",ylab="");par(new=T)
```

```
plot(p,y,type="l",ylab="")
```

```
abline(v=0.4,lty=2,col="RED",ylim=c(0,max(y)))
```

## Algunos detalles

- ▶ El parámetro  $\theta$  puede ser multidimensional  $\theta = (\theta_1, \theta_2)$

$$\pi(\theta_1, \theta_2 | \mathbf{x}) = \frac{f(\mathbf{x} | \theta_1, \theta_2)\pi(\theta_1, \theta_2)}{m(\mathbf{x})}$$

$$\pi(\theta_1 | \mathbf{x}) = \int \pi(\theta_1, \theta_2 | \mathbf{x}) d\theta_2$$

- ▶ Como  $\pi(\theta | \mathbf{x})$  define una distribución de probabilidad sobre  $\theta$ , entonces

$$\pi(\theta | \mathbf{x}) \propto f(\mathbf{x} | \theta)\pi(\theta)$$



# Estimación a posteriori

- ▶ Se basa en  $\pi(\theta \mid \mathbf{x})$
- ▶ Podemos describir  $\pi(\theta \mid \mathbf{x})$  mediante cuantiles, media, varianza ...
- ▶ Para parámetros unidimensionales

$$P(a < \theta < b \mid \mathbf{x}) = \int_a^b \pi(\theta \mid \mathbf{x}) d\theta = 0.95$$

Intervalo de credibilidad al 95 %

¿diferencia con el intervalo de confianza?

# Intervalo de credibilidad

Dado  $\alpha \in (0, 1)$ , diremos que un intervalo  $I \subseteq \mathbb{R}$  es un intervalo de credibilidad del  $(1 - \alpha)100\%$  para el parámetro unidimensional  $\nu = h(\theta)$  si

$$P(h(\theta) \in I | \mathbf{x}) = \int_{h(\theta) \in I} \pi(\theta | \mathbf{x}) d\theta = 1 - \alpha$$

# Media y varianza a posteriori

La media y varianza a posteriori del parámetro unidimensional  $h(\theta)$ , son respectivamente la media y la varianza de la distribución a posteriori de  $h(\theta)$ , es decir,

$$E(h(\theta)|\mathbf{x}) = \int h(\theta)\pi(\theta|\mathbf{x})d\theta$$

y

$$Var(\theta|\mathbf{x}) = \int (h(\theta) - E(h(\theta)|\mathbf{x}))^2 \pi(\theta|\mathbf{x}) d\theta$$

## Ejemplo

$$\pi(p|\mathbf{x}) = \frac{f(\mathbf{x}|p)\pi(p)}{m(\mathbf{x})} \propto p^{n\bar{x}}(1-p)^{n(1-\bar{x})}$$

La distribución a posteriori es una  $\mathcal{B}(a, b)$  con  $a = n\bar{x} + 1$  y  $b = n(1 - \bar{x}) + 1$

La media a posteriori de  $p$  es

$$E(p|\mathbf{x}) = \frac{a}{a+b} = \frac{n\bar{x} + 1}{n + 2}$$

Si  $n$  es suficiente grande, dicha media a posteriori se aproxima a  $\bar{x}$

La varianza a posteriori converge a 0 cuando  $n \rightarrow +\infty$ .

Dos priors diferentes pueden  
producir resultados diferentes

$$\mathcal{L} \pi(\theta)?$$

# Distribuciones a priori no informativas

La elección de  $\pi(\theta)$  es una de las principales dificultades para realizar inferencia bayesiana

- ▶ Algunas veces el investigador tiene información sobre los parámetros antes de la observación de los datos
- ▶ Usar procedimientos automáticos conocidos como *objetivos* o *no informativos* para producir  $\pi(\theta)$ , que tienen como objetivo atenuar el impacto sobre la inferencia resultante, [Bernardo y Smith, 1994]
- ▶ Afortunadamente, la prior suele ejercer poca influencia cuando hay datos suficientes

# Regla de Jeffreys

[Jeffreys, 1961]: tomar la raíz cuadrada de la medida de información de Fisher

$$\pi(\theta) \propto \left( E_{\mathbf{x}|\theta} \left( -\frac{d^2}{d\theta^2} \log f(\mathbf{x}|\theta) \right) \right)^{1/2}$$

cuando el parámetro  $\theta$  es unidimensional

El símbolo  $E_{\mathbf{x}|\theta}$  significa esperanza respecto de la distribución dada por  $f(\mathbf{x}|\theta)$

## Ejemplo

Supongamos que  $\mathbf{x} = (x_1, \dots, x_n)$  es una muestra de  $n$  observaciones independientes de la distribución de Poisson de media  $\lambda$

$$f(\mathbf{x}|\lambda) \propto \lambda^{n\bar{x}} \exp(-n\lambda)$$

$$\frac{d}{d\lambda} \log f(\mathbf{x}|\lambda) = n\bar{x}/\lambda - n \qquad \frac{d^2}{d\lambda^2} \log f(\mathbf{x}|\lambda) = -n\bar{x}/\lambda^2$$

Como  $E(n\bar{x}/\lambda^2) = n\lambda/\lambda^2 = n/\lambda$ , entonces  $\pi(\lambda) \propto 1/\sqrt{\lambda}$



## Ejemplo

La distribución  $\pi(\lambda)$  es impropia

$$\int_0^{+\infty} 1/\sqrt{\lambda} = +\infty$$

pero la distribución a posteriori que resulta de aplicar la definición es propia

$$\pi(\lambda \mid \mathbf{x}) \propto \lambda^{n\bar{x}-1/2} \exp(-n\lambda)$$

La distribución Gamma  $\mathcal{G}(n\bar{x} + 1/2, 1/n)$  con parámetro de forma  $n\bar{x} + 1/2$  y parámetro de escala  $1/n$

## Ejemplo

Si hemos observado

$\mathbf{x} = (1, 4, 6, 2, 4, 5, 3, 5, 3, 4, 3, 3, 1, 3, 4, 3, 3, 1, 1, 4, 8, 5, 3, 1, 0)$

dibuja la distribución a posteriori  $\pi(\lambda \mid \mathbf{x})$ , y calcula media y varianza a posteriori, y un IC95 % basado en los cuantiles de la distribución a posteriori. Utiliza las funciones `dgamma` y `qgamma`. Genera simulaciones (`rgamma`) de esta distribución gamma y describe esas simulaciones como si fuera una muestra: media, varianza, histograma ...

# Análisis bayesiano de muestras con distribución normal

# Una población normal $N(\mu, \sigma^2)$

Supongamos que los datos  $x_1, x_2, \dots, x_n$  son observaciones independientes con distribución  $N(\mu, \sigma^2)$

$$f(\mathbf{x}|\mu, \sigma) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

Consideramos dos casos

**Caso 1.** El parámetro  $\sigma$  es conocido  $\Rightarrow \theta = \mu$

**Caso 2.** Los parámetros  $\mu$  y  $\sigma$  son desconocido  $\Rightarrow \theta = (\mu, \sigma)$

## Caso 1. El parámetro $\sigma$ es conocido

- ▶ La distribución a priori de Jeffreys es  $\pi(\mu) \propto 1$
- ▶ Prior impropia ( $\int \pi(\mu) d\mu = +\infty$ ), pero la definición de distribución a posteriori (2) proporciona una distribución propia

$$\pi(\mu|\mathbf{x}) \propto f(\mathbf{x}|\mu, \sigma)\pi(\mu) \propto \exp\left(-\frac{(\bar{\mathbf{x}} - \mu)^2}{2\sigma^2/n}\right)$$

$$\pi(\mu|\mathbf{x}) = N(\mu \mid \bar{\mathbf{x}}, \sigma^2/n)$$

El intervalo

$$I = \bar{\mathbf{x}} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

es un intervalo de credibilidad al 95 % para  $\mu$

Coincide *numéricamente* con el intervalo de confianza al 95 % para  $\mu$  cuando  $\sigma^2$  es conocida

## Caso 2. Los parámetros $\mu$ y $\sigma$ son desconocidos.

Prior:  $\pi(\mu, \sigma) \propto 1/\sigma$

Impropia:  $\int \pi(\mu, \sigma) d\mu d\sigma = +\infty$

## Caso 2. Los parámetros $\mu$ y $\sigma$ son desconocidos.

Prior:  $\pi(\mu, \sigma) \propto 1/\sigma$

Impropia:  $\int \pi(\mu, \sigma) d\mu d\sigma = +\infty$

Para  $n \geq 2$ , la distribución a posteriori (2) es una distribución propia

$$\pi(\mu, \sigma | \mathbf{x}) \propto \sigma^{-n-1} \exp \left( -\frac{1}{2\sigma^2} (\nu s^2 + n(\mu - \bar{\mathbf{x}})^2) \right)$$



## Caso 2. Los parámetros $\mu$ y $\sigma$ son desconocidos.

Prior:  $\pi(\mu, \sigma) \propto 1/\sigma$

Impropia:  $\int \pi(\mu, \sigma) d\mu d\sigma = +\infty$

Para  $n \geq 2$ , la distribución a posteriori (2) es una distribución propia

$$\pi(\mu, \sigma | \mathbf{x}) \propto \sigma^{-n-1} \exp \left( -\frac{1}{2\sigma^2} (\nu s^2 + n(\mu - \bar{\mathbf{x}})^2) \right)$$

si  $s^2 = \nu^{-1} \sum (x_i - \bar{\mathbf{x}})^2 > 0$ , donde  $\nu = n - 1$

## Distribución a posteriori de $\mu$

$$\pi(\mu, \sigma \mid \mathbf{x}) \propto \sigma^{-n-1} \exp \left( -\frac{1}{2\sigma^2} (\nu s^2 + n(\mu - \bar{\mathbf{x}})^2) \right)$$

## Distribución a posteriori de $\mu$

$$\pi(\mu, \sigma \mid \mathbf{x}) \propto \sigma^{-n-1} \exp \left( -\frac{1}{2\sigma^2} (\nu s^2 + n(\mu - \bar{\mathbf{x}})^2) \right)$$

$$\pi(\mu \mid \mathbf{x}) = \int_0^{+\infty} \pi(\mu, \sigma \mid \mathbf{x}) d\sigma \propto \frac{1}{\left( 1 + \frac{n(\mu - \bar{\mathbf{x}})^2}{\nu s^2} \right)^{n/2}}$$

La distribución a posteriori de  $\mu$  es la distribución de la variable aleatoria  $\bar{\mathbf{x}} + \frac{s}{\sqrt{n}}t$ , donde  $t$  es una t-Student con  $\nu$  grados de libertad

La media a posteriori de  $\mu$  es  $\bar{\mathbf{x}}$  y un intervalo de credibilidad al  $(1 - \alpha)100\%$  viene dado por

$$\left( \bar{\mathbf{x}} - t_{\alpha/2, \nu} \frac{s}{\sqrt{n}}, \bar{\mathbf{x}} + t_{\alpha/2, \nu} \frac{s}{\sqrt{n}} \right)$$

donde  $t_{\alpha/2, \nu}$  es el cuantil  $1 - \alpha/2$  de la distribución Student-t con  $\nu$  grados de libertad.

## Distribución a posteriori de $\sigma$

$$\pi(\mu, \sigma \mid \mathbf{x}) \propto \sigma^{-n-1} \exp \left( -\frac{1}{2\sigma^2} (\nu s^2 + n(\mu - \bar{\mathbf{x}})^2) \right)$$

$$\pi(\sigma \mid \mathbf{x}) = \int \pi(\mu, \sigma \mid \mathbf{x}) d\mu \propto \sigma^{-n} \exp \left( -\frac{\nu s^2}{2\sigma^2} \right)$$

que está relacionada con la distribución Gamma!!!

# Diferencia de las medias de dos poblaciones normales

Cuantificar las diferencias entre las medias de dos variables

# Diferencia de las medias de dos poblaciones normales

Cuantificar las diferencias entre las medias de dos variables

- ▶ Comparar los niveles de colesterol para dos tratamientos:  
comparar las medias de colesterol según el tratamiento

# Diferencia de las medias de dos poblaciones normales

Cuantificar las diferencias entre las medias de dos variables

- ▶ Comparar los niveles de colesterol para dos tratamientos:  
comparar las medias de colesterol según el tratamiento
- ▶ Comparar los niveles medios de expresión de un gen entre dos grupos de interés



- ▶ Supongamos que tenemos dos muestras independientes
  - ▶  $\mathbf{x}_1 = (x_{11}, x_{12}, \dots, x_{1n_1})$  iid  $N(\mu_1, \sigma_1^2)$
  - ▶  $\mathbf{x}_2 = (x_{21}, x_{22}, \dots, x_{2n_2})$  iid  $N(\mu_2, \sigma_2^2)$
- ▶ Los parámetros desconocidos son

- ▶ Supongamos que tenemos dos muestras independientes
  - ▶  $\mathbf{x}_1 = (x_{11}, x_{12}, \dots, x_{1n_1})$  iid  $N(\mu_1, \sigma_1^2)$
  - ▶  $\mathbf{x}_2 = (x_{21}, x_{22}, \dots, x_{2n_2})$  iid  $N(\mu_2, \sigma_2^2)$
- ▶ Los parámetros desconocidos son

$$\mu_1, \mu_2, \sigma_1, \sigma_2$$

pero estamos interesados en el parámetro

- ▶ Supongamos que tenemos dos muestras independientes
  - ▶  $\mathbf{x}_1 = (x_{11}, x_{12}, \dots, x_{1n_1})$  iid  $N(\mu_1, \sigma_1^2)$
  - ▶  $\mathbf{x}_2 = (x_{21}, x_{22}, \dots, x_{2n_2})$  iid  $N(\mu_2, \sigma_2^2)$
- ▶ Los parámetros desconocidos son

$$\mu_1, \mu_2, \sigma_1, \sigma_2$$

pero estamos interesados en el parámetro

$$\alpha = \mu_2 - \mu_1$$

- ▶ Supongamos que tenemos dos muestras independientes
  - ▶  $\mathbf{x}_1 = (x_{11}, x_{12}, \dots, x_{1n_1})$  iid  $N(\mu_1, \sigma_1^2)$
  - ▶  $\mathbf{x}_2 = (x_{21}, x_{12}, \dots, x_{2n_2})$  iid  $N(\mu_2, \sigma_2^2)$
- ▶ Los parámetros desconocidos son

$$\mu_1, \mu_2, \sigma_1, \sigma_2$$

pero estamos interesados en el parámetro

$$\alpha = \mu_2 - \mu_1$$

- ▶ La inferencia sobre el parámetro  $\alpha$  se basa en su distribución a posteriori

Asumiendo  $\sigma_1 = \sigma_2 = \sigma$ , los parámetros desconocidos son  $\mu_1$ ,  $\mu_2$  y  $\sigma$

Prior:  $\pi(\mu_1, \mu_2, \sigma) \propto 1/\sigma$

Asumiendo  $\sigma_1 = \sigma_2 = \sigma$ , los parámetros desconocidos son  $\mu_1$ ,  $\mu_2$  y  $\sigma$

Prior:  $\pi(\mu_1, \mu_2, \sigma) \propto 1/\sigma$

La distribución a posteriori de

$$\frac{\alpha - (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)}{s \sqrt{1/n_1 + 1/n_2}}$$

es la distribución t-Student con  $\nu = n_1 + n_2 - 2$  grados de libertad, donde  $s^2 = (\nu_1 s_1^2 + \nu_2 s_2^2)/\nu$ .

Asumiendo  $\sigma_1 = \sigma_2 = \sigma$ , los parámetros desconocidos son  $\mu_1$ ,  $\mu_2$  y  $\sigma$

Prior:  $\pi(\mu_1, \mu_2, \sigma) \propto 1/\sigma$

La distribución a posteriori de

$$\frac{\alpha - (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)}{s \sqrt{1/n_1 + 1/n_2}}$$

es la distribución t-Student con  $\nu = n_1 + n_2 - 2$  grados de libertad, donde  $s^2 = (\nu_1 s_1^2 + \nu_2 s_2^2)/\nu$ .

Cuando no es asumible  $\sigma_1 = \sigma_2$ , es necesario recurrir a métodos numéricos para hacer inferencias sobre  $\alpha = \mu_2 - \mu_1$

Datos muestrales del estudio

Sex differences in high density lipoprotein cholesterol in urban blacks

Fichero colesterol.R

Earl Ford, Richard Cooper, Brian Simmons, Sherry Katz, Rashmi Patel . Sex differences in high density lipoprotein cholesterol in urban blacks. American Journal of Epidemiology. 1988, 127, 753–761



# Métodos de Monte Carlo (MC) y de Monte Carlo y Cadenas de Markov (MCMC)

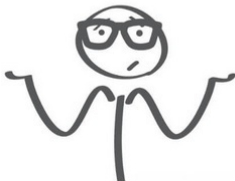
Necesitamos calcular integrales



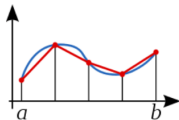
$$E(h(\theta)|\mathbf{x}) = \int h(\theta)\pi(\theta|\mathbf{x})d\theta \quad (3)$$



La media de h de tita es lo mismo que la media de montecarlo.



- ▶ Es necesario recurrir a métodos **numéricos** iterativos
- ▶ Cuando la integral es de dimensión elevada, los métodos numéricos clásicos



no son eficientes

- ▶ En estadística bayesiana se usan métodos de simulación de  $\pi(\theta \mid \mathbf{x})$ : **MC, MCMC, ABC, ...**

**Simular** una variable aleatoria es **Generar** la observación de dicha v.a.

$Ber(0.5)$ : Proceso mecánico basado en lanzar una moneda perfecta

Bernoulli



$Ber(0.5)$ : Ordenador usando secuencias de números *pseudoaleatorios*



# Método de Monte Carlo

Si tenemos una sucesión de simulaciones de la variable a posteriori podemos aproximar las cantidades.

Si  $\theta^{(1)}, \theta^{(2)}, \dots$  es una sucesión de v.a. iid  $\pi(\theta|\mathbf{x})$ , entonces

$$\lim_m \frac{1}{m} \sum_{i=1}^m h(\theta^{(i)}) = E(h(\theta)|\mathbf{x})$$

La media de  $h$  de tita es lo mismo que la media de montecarlo.

# Método de Monte Carlo

Si  $\theta^{(1)}, \theta^{(2)}, \dots$  es una sucesión de v.a. iid  $\pi(\theta|\mathbf{x})$ , entonces

$$\lim_m \frac{1}{m} \sum_{i=1}^m h(\theta^{(i)}) = E(h(\theta)|\mathbf{x})$$

Para aproximar la integral  $E(h(\theta)|\mathbf{x}) = \int h(\theta)\pi(\theta | \mathbf{x})d\theta$ , podemos

# Método de Monte Carlo

Si  $\theta^{(1)}, \theta^{(2)}, \dots$  es una sucesión de v.a. iid  $\pi(\theta|\mathbf{x})$ , entonces

$$\lim_m \frac{1}{m} \sum_{i=1}^m h(\theta^{(i)}) = E(h(\theta)|\mathbf{x})$$

Para aproximar la integral  $E(h(\theta)|\mathbf{x}) = \int h(\theta)\pi(\theta | \mathbf{x})d\theta$ , podemos

1. simular de  $\pi(\theta | \mathbf{x})$  un gran número  $m$  de veces

# Método de Monte Carlo

Si  $\theta^{(1)}, \theta^{(2)}, \dots$  es una sucesión de v.a. iid  $\pi(\theta|\mathbf{x})$ , entonces

$$\lim_m \frac{1}{m} \sum_{i=1}^m h(\theta^{(i)}) = E(h(\theta)|\mathbf{x})$$

Para aproximar la integral  $E(h(\theta)|\mathbf{x}) = \int h(\theta)\pi(\theta | \mathbf{x})d\theta$ , podemos

1. simular de  $\pi(\theta | \mathbf{x})$  un gran número  $m$  de veces
2. evaluar la función  $h(\theta)$  en cada simulación  $\theta^{(i)}$  y



# Método de Monte Carlo

Si  $\theta^{(1)}, \theta^{(2)}, \dots$  es una sucesión de v.a. iid  $\pi(\theta|\mathbf{x})$ , entonces

$$\lim_m \frac{1}{m} \sum_{i=1}^m h(\theta^{(i)}) = E(h(\theta)|\mathbf{x})$$

Para aproximar la integral  $E(h(\theta)|\mathbf{x}) = \int h(\theta)\pi(\theta | \mathbf{x})d\theta$ , podemos

1. simular de  $\pi(\theta | \mathbf{x})$  un gran número  $m$  de veces
2. evaluar la función  $h(\theta)$  en cada simulación  $\theta^{(i)}$  y
3. calcular  $\sum_{i=1}^m h(\theta^{(i)})/m$

## Detalle

Si  $\theta^{(1)}, \theta^{(2)}, \dots$  son simulaciones de  $\pi(\theta|\mathbf{x})$ , entonces  
 $\nu^{(1)} = h(\theta^{(1)}), \nu^{(2)} = h(\theta^{(2)}), \dots$

# Detalle

Si  $\theta^{(1)}, \theta^{(2)}, \dots$  son simulaciones de  $\pi(\theta|\mathbf{x})$ , entonces  
 $\nu^{(1)} = h(\theta^{(1)}), \nu^{(2)} = h(\theta^{(2)}), \dots$  son simulaciones de  $\pi(\nu | \mathbf{x})$

Con simulaciones lo tenemos  
todo !!!



Dos poblaciones normales

$$N(\mu_1, \sigma_1^2) \text{ y } N(\mu_2, \sigma_1^2)$$

Parámetros:

$$\mu_1, \mu_2, \sigma_1, \sigma_2$$

- ▶ Supongamos que tenemos dos muestras independientes

- ▶  $\mathbf{x}_1 = (x_{11}, x_{12}, \dots, x_{1n_1})$  iid  $N(\mu_1, \sigma_1^2)$

- ▶  $\mathbf{x}_2 = (x_{21}, x_{12}, \dots, x_{2n_2})$  iid  $N(\mu_2, \sigma_2^2)$

- ▶ Los parámetros desconocidos son

$$\mu_1, \mu_2, \sigma_1, \sigma_2$$

- ▶ Estamos interesados en el parámetro

$$\nu = \mu_2 - \mu_1$$

- ▶ La inferencia sobre el parámetro  $\nu$  se basa en su distribución a posteriori
- ▶ **No asumimos igualdad de varianzas**

- ▶ Simular de  $\pi(\nu|\mathbf{x})$  mediante simulaciones de  $\pi(\mu_1, \mu_2, \sigma_1, \sigma_2|\mathbf{x})$
- ▶ Prior:  $\pi(\mu_1, \mu_2, \sigma_1, \sigma_2) \propto \frac{1}{\sigma_1 \sigma_2}$
- ▶ A posteriori  $(\mu_1, \sigma_1)$  y  $(\mu_2, \sigma_2)$  son independientes:

$$\pi(\mu_1, \mu_2, \sigma_1, \sigma_2|\mathbf{x}) = \pi_1(\mu_1, \sigma_1|\mathbf{x}_1)\pi_2(\mu_2, \sigma_2|\mathbf{x}_2)$$

Se pueden simular por separado

- Para simular

$$\pi_1(\mu_1, \sigma_1 | \mathbf{x}_1) = \pi_1(\sigma_1 | \mathbf{x}_1) \pi_1(\mu_1 | \sigma_1, \mathbf{x}_1)$$

podemos simular  $\pi_1(\sigma_1 | \mathbf{x}_1)$  y luego  $\pi_1(\mu_1 | \sigma_1, \mathbf{x}_1)$ .

- La distribución

$$\pi_1(\sigma_1 | \mathbf{x}_1) \propto \sigma_1^{-n_1} \exp \left( -\frac{\nu_1 s_1^2}{2\sigma_1^2} \right)$$

está relacionada con la distribución Gamma.

- La distribución  $\pi_1(\mu_1 | \sigma_1, \mathbf{x}_1)$  es la normal de media  $\bar{\mathbf{x}}_1$  y varianza  $\sigma_1^2/n_1$ .

La siguiente función para R proporciona una simulación de la distribución a posteriori de  $\nu = \mu_2 - \mu_1$



```
simula.difmedias=function(x1,x2){  
  
n1=length(x1);n2=length(x2);  
  
m.x1=mean(x1);m.x2=mean(x2);  
  
q1=(n1-1)*var(x1);q2=(n2-1)*var(x2);  
  
u=rgamma(1,shape=(n1-1)/2,scale=2/q1);  
sigma1=1/sqrt(u);  
mu1=rnorm(1,m.x1,sigma1/sqrt(n1));  
  
u=rgamma(1,shape=(n2-1)/2,scale=2/q2);  
sigma2=1/sqrt(u);  
mu2=rnorm(1,m.x2,sigma2/sqrt(n2));  
  
dif=mu2-mu1  
return(dif)}
```

# Práctica

Simular datos del modelo  $N(\mu, \sigma^2)$  usando el comando `rnorm`, para

$$n_1 = 15, \mu_1 = 0, \sigma_1 = 0.1$$

$$n_2 = 20, \mu_2 = 1 \text{ y } \sigma_2 = 0.2$$

Aplicar a los datos obtenidos la función `simula.difmedias` para obtener la media a posteriori de  $\nu = \mu_2 - \mu_1$ , un IC 95 %, un histograma de la distribución a posteriori de  $\nu$ , y  $P(\nu > 1 \mid \text{Datos})$  mediante simulaciones de dicha distribución.

Comparar el IC95 % con un resultado frecuentista

# Muestreo de importancia

Si podemos evaluar  $\pi(\theta|\mathbf{x})$  <sup>distribución a posteriori</sup>

Por cada  $g$  podemos saber todo lo que hay en la derivada.

$$E(h(\theta)|\mathbf{x}) = \int \frac{h(\theta)\pi(\theta|\mathbf{x})}{g(\theta)} g(\theta) d\theta$$

puede aproximarse usando Monte Carlo simulando de  $g(\theta)$

¿Cómo?

La  $g$  puede ser una normal o lo que queramos.

- ▶ La densidad  $g(\theta)$  puede ser cualquiera
- ▶ Una elección que suele ser satisfactoria consiste en tomar la densidad de una normal
  - ▶ Centrada en  $\hat{\theta}$
  - ▶ Con varianza proporcional a  $\hat{V}$
- ▶ t-Student es otra opción
- ▶ Requiere poder evaluar  $\pi(\theta|\mathbf{x})$ : el problema será  $m(\mathbf{x})$

## Sampling Importance Resampling (SIR)

Para calcular la marginal tambien se utiliza el mismo método de multiplicar y dividir la g de tita°

$$m(\mathbf{x}) = \int f(\mathbf{x}|\theta)\pi(\theta)d\theta = \int \frac{f(\mathbf{x}|\theta)\pi(\theta)}{g(\theta)}g(\theta)d\theta$$

A. F. M. Smith and A. E. Gelfand. Bayesian Statistics without Tears: A Sampling-Resampling Perspective The American Statistician. 1992, 46, 84-88.

## Algoritmo

1. Simulamos independientemente  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(m)}$  de  $g(\theta)$
2. Calculamos

$$\omega_i \propto \frac{f(\mathbf{x}|\theta^{(i)})\pi(\theta^{(i)})}{g(\theta^{(i)})}, \quad i = 1, \dots, m$$

3. Obtenemos la aproximación

$$\int h(\theta)\pi(\theta|\mathbf{x})d\theta \approx \frac{\sum_{i=1}^m h(\theta^{(i)})\omega_i}{\sum_{i=1}^m \omega_i}$$

Las simulaciones  $\theta^{(i)}$  al pesarlas con  $\omega_i$ , pueden ser utilizadas para la inferencia bayesiana

# Práctica

Supongamos que los datos son una muestra de una distribución normal  $N(\mu, \sigma^2)$  y que sabemos que  $\mu \in (0, 5)$  y  $\sigma \in (0, 6)$ . Simular los datos ( $n=20$ ) tomando unos valores concretos de  $\mu$  y de  $\sigma$ , y realizar inferencia bayesiana sobre los parámetros.

La función de densidad de los datos es

$$f(\mathbf{x}|\mu, \sigma) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

Tomaremos  $\pi(\mu, \sigma) \propto \sigma^{-1} 1_{(0,5)}(\mu) 1_{(0,6)}(\sigma)$

Usaremos SIR y tomaremos  $g(\mu, \sigma)$  la densidad de la distribución uniforme en el espacio paramétrico

Fichero PracticaSIR.R



# MCMC ...

# Cadena de Markov

Son iteraciones de las simulaciones que dependen de la anterior.

Diremos que la sucesión de variables aleatorias  $\theta^{(1)}, \theta^{(2)}, \dots$  es una cadena de Markov si la distribución de probabilidad de  $\theta^{(k+1)} | \theta^{(1)}, \theta^{(2)}, \dots, \theta^{(k)}$  no depende de  $(\theta^{(1)}, \dots, \theta^{(k-1)})$

$$f(\theta^{(k+1)} | \theta^{(1)}, \theta^{(2)}, \dots, \theta^{(k)}) = f(\theta^{(k+1)} | \theta^{(k)})$$

## Ejemplo

$$\theta^{(t+1)} = \frac{1}{2}\theta^{(t)} + \varepsilon_t, \quad t = 1, \dots$$

$$\theta^{(1)} = 0, \quad \varepsilon_t \sim N(0, 1), \quad t \geq 1$$

```
theta={}

valor=0
for(i in 1:1000){
  valor=.5*valor+rnorm(1,0,1)
  theta=c(theta,valor)
}

plot(theta,type="l")
```

# Metropolis-Hasting

- ▶ Desarrollado por Metropolis et al. (1953) y posteriormente por Hasting (1970)
- ▶ Aplicado a  $\pi(\theta|\mathbf{x})$  produce una cadena de Markov  $(\theta^{(n)})$  cuya distribución límite es  $\pi(\theta|\mathbf{x})$

$$\lim_m \frac{1}{m} \sum_{i=1}^m h(\theta^{(i)}) = \int h(\theta) \pi(\theta|\mathbf{x}) d\theta = E(h(\theta) | \mathbf{x})$$

Necesita una densidad de transición  $q(\theta'|\theta)$  llamada densidad instrumental o candidata, y un valor inicial  $\theta^{(0)}$

1. Simular  $\theta' \sim q(\theta'|\theta^{(k)})$
2. Simular  $u \sim U(0, 1)$  y calcular

$$\rho_k = \min \left( 1, \frac{\pi(\theta')f(\mathbf{x}|\theta')}{\pi(\theta^{(k)})f(\mathbf{x}|\theta^{(k)})} \frac{q(\theta^{(k)}|\theta')}{q(\theta'|\theta^{(k)})} \right)$$

3. Si  $u < \rho_k$  entonces tomar  $\theta^{(k+1)} = \theta'$ , en otro caso tomar  $\theta^{(k+1)} = \theta^{(k)}$

No necesitamos calcular  $m(\mathbf{x})$  !!!

# Gibbs sampling

Si  $\theta = (\theta_1, \dots, \theta_p)$ , es complicado proponer una candidata adecuada para aplicar directamente el algoritmo de Metropolis-Hastings

El algoritmo Gibbs sampling produce una cadena de Markov  $(\theta^{(t)})$  con distribución límite la distribución  $\pi(\theta|\mathbf{x})$

Gibbs sampling: simulación de las full conditional distributions !!!

$$\pi(\theta_j \mid \theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_p, \mathbf{x})$$

Dimensión menor

La transición  $\theta^{(k)} \rightarrow \theta^{(k+1)}$  está definida por los siguientes pasos

1.  $\theta_1^{(k+1)} \sim \pi(\theta_1 | \theta_2^{(k)}, \dots, \theta_p^{(k)}, \mathbf{x})$

2.  $\theta_2^{(k+1)} \sim \pi(\theta_2 | \theta_1^{(k+1)}, \theta_3^{(k)}, \dots, \theta_p^{(k)}, \mathbf{x})$

$\vdots$

p.  $\theta_p^{(k+1)} \sim \pi(\theta_p | \theta_1^{(k+1)}, \theta_2^{(k+1)}, \dots, \theta_{p-1}^{(k+1)}, \mathbf{x})$

La transición  $\theta^{(k)} \rightarrow \theta^{(k+1)}$  está definida por los siguientes pasos

1.  $\theta_1^{(k+1)} \sim \pi(\theta_1 | \theta_2^{(k)}, \dots, \theta_p^{(k)}, \mathbf{x})$

2.  $\theta_2^{(k+1)} \sim \pi(\theta_2 | \theta_1^{(k+1)}, \theta_3^{(k)}, \dots, \theta_p^{(k)}, \mathbf{x})$

$\vdots$

p.  $\theta_p^{(k+1)} \sim \pi(\theta_p | \theta_1^{(k+1)}, \theta_2^{(k+1)}, \dots, \theta_{p-1}^{(k+1)}, \mathbf{x})$

- En muchas situaciones las condicionadas pueden realizarse directamente



La transición  $\theta^{(k)} \rightarrow \theta^{(k+1)}$  está definida por los siguientes pasos

1.  $\theta_1^{(k+1)} \sim \pi(\theta_1 | \theta_2^{(k)}, \dots, \theta_p^{(k)}, \mathbf{x})$

2.  $\theta_2^{(k+1)} \sim \pi(\theta_2 | \theta_1^{(k+1)}, \theta_3^{(k)}, \dots, \theta_p^{(k)}, \mathbf{x})$

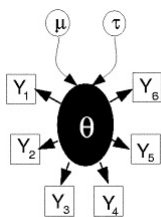
$\vdots$

p.  $\theta_p^{(k+1)} \sim \pi(\theta_p | \theta_1^{(k+1)}, \theta_2^{(k+1)}, \dots, \theta_{p-1}^{(k+1)}, \mathbf{x})$

- ▶ En muchas situaciones las condicionadas pueden realizarse directamente
- ▶ Si no es así, es posible aplicar el algoritmo de Metropolis-Hastings con **solo** una iteración para simular en los pasos del algoritmo Gibbs sampling

# Aspectos computacionales con R y WinBUGS

El programa WinBUGS permite simular la distribución a posteriori combinando varias técnicas MCMC (MH, Gibbs Sampling, ...) *produciendo* una cadena de Markov convergente a  $\pi(\theta \mid \mathbf{x})$



**BUGS**

WinBUGS  
with  
DoodleBUGS  
Version 1.4

# Componentes de un modelo para WinBUGS

- ▶ El modelo para los datos y la prior
- ▶ Los datos y las constantes
- ▶ El valor inicial de la cadena de Markov

# El modelo para los datos y la prior

```
model
{
  for(i in 1:n1){
    x1[i] ~ dnorm(mu1,tau1)
  }
  for(i in 1:n2){
    x2[i] ~ dnorm(mu2,tau2)
  }
  mu1 ~ dflat()
  mu2 ~ dflat()
  .
  .
  .
  dif<-mu2-mu1
}
```

Este primer for es para declarar la distribucion de los datos.  
Normal


Son iteraciones de las simulaciones que dependen de la anterior.

- ▶ La notación  $\sim$  significa *tiene por distribución*
- ▶  $x \sim \text{dnorm}(\mu, \tau)$  significa que  $x$  se modela como una variable con distribución normal de media  $\mu$  y **varianza  $1/\tau$**
- ▶ La notación  $< -$  hace el papel de  $=$  para definir parámetros en función de otros ( $\nu = \mu_2 - \mu_1$ ) o valores en función de los datos ( $z_1 = \sqrt{y_1}$ )

WinBUGS14

FileToolsEditAttributesInfoModelInferenceOptionsDoodleMapTextWindowHelp

WinBUGS Licence



**BUGS**

Introduction

This software and a (hereinafter called 'agreement and ma This is a legal agre College of Science provided in the follo

Users are required WinBUGS PACKA registration and acc There is no fee pay WinBUGS package package can upgra

The current fee is 2

By completing and the terms of this lic

It should be empha WinBUGS PACKA offer advice on inter Any assistance wil

WinBUGS User Manual

## Permission and Disclaimer

[please click here to read the legal bit](#)

More informally, potential users are reminded to be extremely careful using this program for serious statistical analysis. We have tested the program on quite a wide set of examples, but be particularly careful with types of model that are currently not featured. If there is a problem, WinBUGS might just crash, which is not very good, but it might well carry on and produce answers that are wrong, which is even worse. Please let us know of any successes or failures.

**Beware: MCMC sampling can be dangerous!**

## Contents

[Introduction](#) ⇌ ⇌

[Model Specification](#) ⇌ ⇌

[The Model Menu](#) ⇌ ⇌

[The Info Menu](#) ⇌ ⇌

[Batch-mode: Scripts](#)

[WinBUGS Graphics](#) ⇌ ⇌

[Tutorial](#) ⇌ ⇌

[Distributions](#) ⇌ ⇌

[Compound](#)

[DoodleB](#)

[The Infer](#)

[The Opti](#)

[Tricks: A](#)

[Tips and](#)

[Changin](#)

[Referenc](#)

User manual F1

Doodle help


Examples Vol I

Examples Vol II

Licence

About WinBUGS

Distributions



**BUGS**

## Distributions

### Contents

[Discrete Univariate](#) ⇌ ⇌

[Continuous Univariate](#) ⇌ ⇌

[Discrete Multivariate](#) ⇌ ⇌

[Continuous Multivariate](#) ⇌ ⇌

## Discrete Univariate

### Bernoulli

$p \sim \text{dbern}(p)$

# Los datos y las constantes

Se introducen con el comando **list**

```
list(x1=c(0.149,0.007,...),  
x2=c(0.922,1.053,...),n1=15,n2=20)
```



# El valor inicial de la cadena de Markov

Los valores iniciales se definen para lo que se declaró con  $\sim$  en el modelo, menos los datos observados

```
list(mu1=0,mu2=0,...)
```

# Práctica con WinBUGS

En esta práctica vamos a realizar inferencia sobre la diferencia de las medias de dos variables con distribución normal sin asumir igualdad de varianzas

El fichero necesario se llama Practica1-Winbugs-modelo-diferencia

# Pasos para ejecutar WinBUGS

- ▶ Abrimos WinBUGS, **File>New** y escribimos (pegamos) el modelo, los datos y los valores iniciales en la ventana que aparece

# Pasos para ejecutar WinBUGS

- ▶ Abrimos WinBUGS, **File>New** y escribimos (pegamos) el modelo, los datos y los valores iniciales en la ventana que aparece
- ▶ **Model>Specification** y aparece la ventana **Specification Tool**. Seleccionamos con el puntero del ratón el texto del modelo y seleccionamos **chek model** en la ventana **Specification Tool**. Si el modelo está bien escrito, aparecerá el mensaje **model is syntactically correct** en la esquina inferior izquierda

- ▶ Cargar los datos: seleccionamos con el puntero del ratón los datos y hacemos click en **load data** de la ventana **Specification Tool**. Aparecerá el mensaje **data loaded** en la esquina inferior izquierda

- ▶ Cargar los datos: seleccionamos con el puntero del ratón los datos y hacemos click en **load data** de la ventana **Specification Tool**. Aparecerá el mensaje **data loaded** en la esquina inferior izquierda
- ▶ Compilar: hacemos click en **compile** de la ventana **Specification Tool**

- ▶ Cargar los datos: seleccionamos con el puntero del ratón los datos y hacemos click en **load data** de la ventana **Specification Tool**. Aparecerá el mensaje **data loaded** en la esquina inferior izquierda
- ▶ Compilar: hacemos click en **compile** de la ventana **Specification Tool**
- ▶ Cargar los valores iniciales: seleccionamos con el puntero del ratón los valores iniciales y hacemos click en **load inits**; aparecerá **model is initialized** en la esquina inferior izquierda. En este momento el programa está preparado para simular la cadena de Markov

► **Inference** > **Samples** y aparece la ventana **Sample Monitor Tool**.

Tenemos que especificar los parámetros que deseamos simular, en principio todas las componentes de  $\theta$ . Como para este ejemplo solo estamos interesados en  $\mu_2 - \mu_1$ , escribiremos dif en **node** y hacemos click en **set**



- ▶ **Inference>Samples** y aparece la ventana **Sample Monitor Tool**.

Tenemos que especificar los parámetros que deseamos simular, en principio todas las componentes de  $\theta$ . Como para este ejemplo solo estamos interesados en  $\mu_2 - \mu_1$ , escribiremos dif en **node** y hacemos click en **set**

- ▶ **Model>Update** y aparece la ventana **Update Tool**.

Podemos ir lanzando trozos de la cadena haciendo click en **update**.

Cuando tenemos un gran número de iteraciones (simulaciones), volvemos a la ventana **Sample Monitor Tool**. Si hacemos click en **stats**, aparece un descriptivo de la distribución a posteriori del parámetro que aparece en **node**, basado en las simulaciones creadas hasta el momento.

Media a posteriori, IC95

Si usamos la función `simula.difmedias`

```
> lista={}  
> for(k in 1:30000){  
+ lista=c(lista,simula.difmedias(x1,x2))  
+ }  
>  
> mean(lista)  
[1] 1.056417  
> quantile(lista,c(0.025,0.975))  
2.5\% 97.5\  
0.9612464 1.1508548
```

Cambiar las priors para  $\mu_1$  y  $\mu_2$  por distribuciones normales centradas en 0 y varianza 100: `dnorm(0, 0.01)`

¿Qué ocurre?

Valorar la convergencia mirando las trazas y las autocorrelaciones

Cambiar el primer dato a 4 (outlier) y lanzar el modelo. ¿Cómo cambia la inferencia a posteriori?

Cambiar el primer dato a 4 (outlier) y lanzar el modelo. ¿Cómo cambia la inferencia a posteriori?

La distribución t-Student es una distribución con colas más pesadas que la distribución Normal y puede usarse si sospechamos que en nuestros datos hay outliers.

Utilizar la distribución t-Student con  $k = 2$  grados de libertad para modelar los datos y observar lo que pasa con las estimaciones a posteriori.

$$x \sim dt(\mu, \tau, k)$$

Valorar la convergencia mirando las trazas y las autocorrelaciones

# Práctica

A un grupo de 20 pacientes hemos suministrado el fármaco 1 y a otro grupo de 24 pacientes el fármaco 2. Tras un periodo de tiempo la enfermedad desaparece en 16 de los 20 pacientes con fármaco 1, y en 21 de los 24 pacientes con fármaco 2. Estimar cuanto es más probable sanar con el fármaco 2 que con el 1. Este parámetro es un cociente de dos probabilidades  $\theta = p_2/p_1$

Preparar un modelo en WinBUGS y lanzarlo para obtener la media a posteriori, un IC 95 % y la gráfica de la densidad a posteriori de  $\theta$ . Usar tres cadenas de Markov con 10000 simulaciones cada una. Graficar las trazas y las autocorrelaciones y discutir en base a ellas la convergencia de las cadenas.



# R y WinBUGS

- ▶ Desde R podemos usar WinBUGS y combinar la funcionalidad de ambos programas

# R y WinBUGS

- ▶ Desde R podemos usar WinBUGS y combinar la funcionalidad de ambos programas
- ▶ Necesitaremos cargar en R la librería R2WinBUGS



# R y WinBUGS

- ▶ Desde R podemos usar WinBUGS y combinar la funcionalidad de ambos programas
- ▶ Necesitaremos cargar en R la librería R2WinBUGS
- ▶ La función que llama a WinBUGS desde R es bugs

# R y WinBUGS

- ▶ Desde R podemos usar WinBUGS y combinar la funcionalidad de ambos programas
- ▶ Necesitaremos cargar en R la librería R2WinBUGS
- ▶ La función que llama a WinBUGS desde R es bugs
- ▶ Necesitamos los mismos ingredientes que son necesarios para usar WinBUGS: el modelo, los datos y los valores iniciales

# R y WinBUGS

- ▶ Desde R podemos usar WinBUGS y combinar la funcionalidad de ambos programas
- ▶ Necesitaremos cargar en R la librería R2WinBUGS
- ▶ La función que llama a WinBUGS desde R es bugs
- ▶ Necesitamos los mismos ingredientes que son necesarios para usar WinBUGS: el modelo, los datos y los valores iniciales

El modelo debe estar en un archivo de texto

La función bugs buscará dicho fichero en el directorio de trabajo actual de la sesión que tengamos abierta de R

- ▶ Los datos deben declararse de la siguiente forma  
`datos=list(x1=varx1,x2=varx2,...)`

x1 en el fichero del modelo

varx1 en R

- ▶ Los datos deben declararse de la siguiente forma

```
datos=list(x1=varx1,x2=varx2,...)
```

x1 en el fichero del modelo

varx1 en R

- ▶ Los valores iniciales

```
iniciales=function(){  
  list(theta=valor inicial para theta)  
}
```

por ejemplo

```
iniciales=function(){  
  list(mu1=0,mu2=0,xi1=0,xi2=0)  
}
```

- ▶ Es una buena práctica lanzar varias cadenas y empezar cada vez en valores diferentes.

- ▶ Usando

```
iniciales=function(){  
  list(mu1=rnorm(1,0,1),mu2=rnorm(1,0,1),  
  xi1=rnorm(1,0,1),xi2=rnorm(1,0,1))  
}
```

cada vez que la función iniciales sea llamada, los valores iniciales serán simulados según la distribución que pongamos

Las siguientes líneas de comandos llaman a WinBUGS para que simule 3 cadenas, 15000 iteraciones, eliminando las primeras 1000, y pidiendo solo las simulaciones correspondientes al parámetro dif:

```
directorio.winbugs="C:/Program Files/WinBUGS14"
```

```
res=bugs(data=datos,inits=iniciales,  
model.file="modelo.txt",  
parameters.to.save=c("dif"),n.iter=15000,n.burnin=1000,  
n.thin=1,n.chain=3,  
bugs.directory=directorio.winbugs)
```

todo estará en el objeto res

# Práctica

Usando los mismos datos que hemos usado para estimar la diferencia de las medias con WinBUGS, preparar una sintaxis en R para realizar las inferencias mediante R y WinBUGS

Descomprimir dif\_R\_WB





# Modelo de regresión lineal

El modelo de regresión lineal establece una relación entre una variable continua  $Y$ , y variables explicativas  $X_1, \dots, X_k$

# Modelo de regresión lineal

El modelo de regresión lineal establece una relación entre una variable continua  $Y$ , y variables explicativas  $X_1, \dots, X_k$

Las observaciones  $y_1, \dots, y_n$  de la variable  $y$  son independientes y tienen una distribución normal con media

$$E(y_i | \beta, \sigma, \mathbf{x}_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

El modelo en notación matricial es

$$\mathbf{y}|\beta, \sigma, X \sim N_n(X\beta, \sigma^2 I_n), \quad (4)$$

donde  $\mathbf{y} = (y_1, \dots, y_n)^T$ ,  $X$  es la matriz con filas  $(1, x_{i1}, \dots, x_{ik})$ ,  $i = 1, \dots, n$ , y  $\beta$  es el vector de los coeficientes de regresión  $(\beta_0, \beta_1, \dots, \beta_k)^T$

El modelo en notación matricial es

$$\mathbf{y}|\beta, \sigma, X \sim N_n(X\beta, \sigma^2 I_n), \quad (4)$$

donde  $\mathbf{y} = (y_1, \dots, y_n)^T$ ,  $X$  es la matriz con filas  $(1, x_{i1}, \dots, x_{ik})$ ,  $i = 1, \dots, n$ , y  $\beta$  es el vector de los coeficientes de regresión  $(\beta_0, \beta_1, \dots, \beta_k)^T$

Asumiremos que  $X^T X$  tiene inversa

El modelo en notación matricial es

$$\mathbf{y}|\beta, \sigma, X \sim N_n(X\beta, \sigma^2 I_n), \quad (4)$$

donde  $\mathbf{y} = (y_1, \dots, y_n)^T$ ,  $X$  es la matriz con filas  $(1, x_{i1}, \dots, x_{ik})$ ,  $i = 1, \dots, n$ , y  $\beta$  es el vector de los coeficientes de regresión  $(\beta_0, \beta_1, \dots, \beta_k)^T$

Asumiremos que  $X^T X$  tiene inversa

¿interpretación los coeficientes  $\beta_1, \dots, \beta_k$  del modelo (4)?

## Función de verosimilitud

$$f(\mathbf{y}|\beta, \sigma^2, X) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta)\right)$$

Priori de Jeffreys:  $\pi(\beta, \sigma^2|X) \propto 1/\sigma^2$

## Posterior

$$\beta|\sigma^2, \mathbf{y}, X \sim N_{k+1}(\hat{\beta}, \sigma^2(X^T X)^{-1})$$

$$\sigma^2|\mathbf{y}, X \sim \mathcal{IG}((n - k - 1)/2, s^2/2)$$

donde

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y} = MLE$$

$$s^2 = (\mathbf{y} - X\hat{\beta})^T(\mathbf{y} - X\hat{\beta})$$

## Práctica

Usando WinBUGS (aunque puede hacerse sin MCMC) y R, ajustar el modelo de regresión lineal con la base de datos caterpillar.txt, tomando la variable respuesta  $\log V11$ , y el resto de variables como variables de ajuste.

```
> Bdatos=read.table("caterpillar.txt")
```

```
> Bdatos[1:5,]
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
1	1200	22	1	4.0	14.8	1.0	1.1	5.9	1.4	1.4	2.37
2	1342	28	8	4.4	18.0	1.5	1.5	6.4	1.7	1.7	1.47
3	1231	28	5	2.4	7.8	1.3	1.6	4.3	1.5	1.4	1.13
4	1254	28	18	3.0	9.2	2.3	1.7	6.9	2.3	1.6	0.85
5	1357	32	7	3.7	10.7	1.4	1.7	6.6	1.8	1.3	0.24

Almacenamos el nombre de las variables en memoria

```
> attach(Bdatos)
```

```
> names(Bdatos)
```

```
[1] "V1" "V2" "V3" "V4" "V5" "V6" "V7"  
"V8" "V9" "V10" "V11"
```



La variable respuesta será  $\log V11$

```
> y=log(V11)
```

y ajustamos el modelo desde el punto de vista frecuentista

```
> ajuste=lm(y~V1+V2+V3+V4+V5+V6+V7+V8+V9+V10)
```

```
> summary(ajuste)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	10.998412	3.060272	3.594	0.00161	**
V1	-0.004431	0.001557	-2.846	0.00939	**
V2	-0.053830	0.021900	-2.458	0.02232	*
V3	0.067939	0.099472	0.683	0.50174	
V4	-1.293636	0.563811	-2.294	0.03168	*
V5	0.231637	0.104378	2.219	0.03709	*
V6	-0.356800	1.566464	-0.228	0.82193	
V7	-0.237469	1.006006	-0.236	0.81558	
V8	0.181060	0.236724	0.765	0.45248	
V9	-1.285316	0.864847	-1.486	0.15142	
V10	-0.433106	0.734869	-0.589	0.56162	

---

```
library(R2WinBUGS)
```

```
datos=list(y=y,V1=V1,V2=V2,V3=V3,V4=V4,V5=V5,V6=V6,  
V7=V7,V8=V8,V9=V9,V10=V10,n=length(y))
```

```
iniciales=function(){list(beta=rnorm(10,0,1),inter=0,xi=0)}
```

```
model="modelo-reglineal-catepillar.txt"
```

```
directorio.winbugs="C:/Program Files/WinBUGS14"
```

```
resultado=bugs(data=datos,inits=iniciales,model.file=model,  
parameters.to.save=c("inter","beta","sigma"),n.iter=10000,  
n.burnin=1000,n.thin=10,n.chain=3,  
bugs.directory=directorio.winbugs,DIC=F)  
}
```

```
> print(resultado,digits=3)
```

Inference for Bugs model at "modelo-reglineal-catepillar.txt", fit using WinBUG

3 chains, each with 10000 iterations (first 1000 discarded), n.thin = 10

n.sims = 2700 iterations saved

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
inter	10.987	3.166	4.877	8.874	11.000	13.082	17.241	1.001	2700
beta[1]	-0.004	0.002	-0.007	-0.005	-0.004	-0.003	-0.001	1.001	2200
beta[2]	-0.053	0.023	-0.097	-0.068	-0.054	-0.038	-0.008	1.001	2700
beta[3]	0.066	0.105	-0.143	-0.002	0.066	0.134	0.275	1.001	2700
beta[4]	-1.299	0.590	-2.482	-1.661	-1.293	-0.928	-0.127	1.001	2700
beta[5]	0.232	0.110	0.005	0.163	0.231	0.304	0.453	1.000	2700
beta[6]	-0.343	1.644	-3.594	-1.393	-0.323	0.729	2.903	1.001	2700
beta[7]	-0.231	1.032	-2.275	-0.920	-0.249	0.472	1.856	1.002	1700
beta[8]	0.183	0.249	-0.305	0.023	0.181	0.344	0.678	1.001	2700
beta[9]	-1.282	0.939	-3.183	-1.881	-1.304	-0.684	0.663	1.001	2700
beta[10]	-0.452	0.761	-1.970	-0.955	-0.454	0.044	1.047	1.001	2700
sigma	0.858	0.136	0.643	0.764	0.842	0.934	1.168	1.001	2700

For each parameter, n.eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

Usar la función `attach.bugs`

Cambiar la prior sobre sigma por la siguiente prior para tau

$$\text{tau} \sim \text{dgamma}(0.001, 0.001)$$

y repetir las estimaciones a posteriori sobre los parámetros

# Inferencia bayesiana en modelos de regresión logística

Variable respuesta es una variable que toma los valores 0 y 1, es decir, una variable de bernoulli

El modelo de regresión logística asume la siguiente relación entre los datos  $y_i$  y los valores  $x_{i1}, x_{i2}, \dots, x_{ik}$  de las covariables

$$\text{logit } p_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}, \quad i = 1, \dots, n$$

donde  $p_i = P(y_i = 1 | \beta, \mathbf{x}_i)$

Asumiendo que los datos  $y_i$  son independientes dados  $\beta$  y  $\mathbf{x}_i$ ,  $i = 1, \dots, n$ , la función de verosimilitud es

$$f(\mathbf{y}|\beta, \mathbf{x}_1, \dots, \mathbf{x}_n) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i}$$

Para este modelo no es posible simular la distribución a posteriori directamente

Usando  $\pi(\beta) = 1$ , la implementación general de este modelo en WinBUGS es

```
model{  
  
  for(i in 1:n){  
    y[i] ~ dbin(p.bound[i],1)  
    p.bound[i] <- max(0, min(1, p[i]))  
    logit(p[i]) <-beta0+beta[1]*x1[i]+...+beta[k]*xk[i]  
  }  
  
  beta0 ~ dflat()  
  for(i in 1:k){beta[i] ~ dflat()}  
  
}
```





Durante 1986 el Baystate Medical Center de Massachusetts recogió datos sobre el peso de recién nacidos para identificar factores que incrementan el riesgo de bajo peso al nacer

Algunos de esos factores, medidos en las madres, fueron tabaco, raza, partos previos y edad

Utilizaremos los datos de estas variables para ejemplificar el ajuste bayesiano de un modelo de regresión logística

El ajuste clásico en R es

```
summary(glm(LOW~factor(SMOKE)+factor(RACE)+factor(edad5c)+  
factor(PTL>0),family=binomial(logit),data=datos))
```

# Cinética de la degradación de los alimentos

La cinética de la degradación ruminal de la materia, el nitrógeno y otros constituyentes de la pared celular, suele describirse con modelos no lineales

$$y_i = a + b(1 - e^{-ct_i}) + \varepsilon_i, \quad i = 1, \dots, N$$

- ▶  $\varepsilon_i \sim N(0, \sigma^2)$
- ▶  $y_i$  es la proporción degradada hasta el instante de tiempo  $t_i$
- ▶  $a$  representa el sustrato soluble y completamente degradable
- ▶  $b$  representa la fracción insoluble pero potencialmente degradable del sustrato
- ▶  $c$  representa la velocidad de degradación
- ▶  $0 < a, b, a + b < 1, c > 0$

```
t = c(3,6,9,15,24)
y = c(0.38,0.51,0.59,0.79,0.89)
```

```
plot(t,y,ylim=c(0,1))
```

```
datos=data.frame(y=y,t=t)
sol=nls(y ~ a+b*(1-exp(-c*t)),
        data = datos,
        start = list(a=0.3,b=0.3,c=0.1))
```

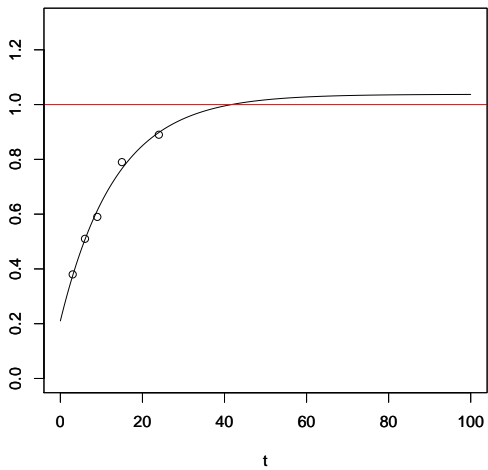
```
summary(sol)
```

```
estimacion=summary(sol)$coef
```

```
data.frame(MLE=estimacion[,1],  
inf=estimacion[,1]-1.96*estimacion[,2],  
sup=estimacion[,1]+1.96*estimacion[,2])
```

```
a=estimacion[1,1]  
b=estimacion[2,1]  
c=estimacion[3,1]
```

```
tiempos=seq(0,100,.1)  
plot(tiempos,a+b*(1-exp(-c*tiempos)),type="l",ylim=c(0,1.3)  
ylab="",xlab="t")  
abline(h=1,col="RED")  
par(new=T)  
plot(t,y,ylim=c(0,1.3),xlim=c(0,100),ylab="",xlab="t")
```



```

model
{
for( i in 1 : N ) {
y[i] ~ dnorm(mu[i],tau)
mu[i] <- a + b*(1-exp(-c*t[i]))
}
a ~ dunif(0.0, 1.0)
b ~ dunif(0.0, z)
z<-1-a
tau ~ dgamma(0.001,0.001)
sigma <- 1 / sqrt(tau)
c ~ dunif(0.0, 0.5)
}

list(t=c(3,6,9,15,24),y=c(0.38,0.51,0.59,0.79,0.89), N=5)
list(t=c(3,6,9,15,24,50,60,70),y=c(0.38,0.51,0.59,0.79,
0.89,NA,NA,NA),N=8)

list(a=0.3,b=0.3,c=0.1,tau=10)

```

untitled

model

```

{
  for(i in 1:N){
    y[i] ~ dnorm(mu[i],tau)
    mu[i] <- a + b*(1-exp(-c*t[i]))
  }

  a ~ dunif(0.0, 1.0)
  b ~ dunif(0.0, z)
  z<-1-a

  tau ~ dgamma(0.001,0.001)
  sigma <- 1 / sqrt(tau)
  c ~ dunif(0.0, 0.5)

}

list(t= c(3,6,9,15,24),y= c(0.38,0.51,0.59,0.79,0.89), N= 5)

list(t= c(3,6,9,15,24,50,60,70),y= c(0.38,0.51,0.59,0.79,0.89,NA,NA,NA), N= 8)

list(a=0.3,b=0.3,c=0.1,tau=10)

```

Update Tool

updates 100000

update thin 100

☐ over relax

Specification Tool

check model

load data

compile

num of chains 3

load inits

for chain 1

gen inits

Sample Monitor Tool

node c

beg 1 end 1000000

clear set trace

stats coda quantiles

Comparison Tool

node mu

beg 1

end 1000000

other y

box plot

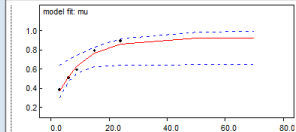
caterpillar

axis t






model fit

scatterplot

model fit





-  Berger, J. y Bernardo, J. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.*, 84:200-207.
-  Bernardo, J. M. (1979). Reference posterior distribution for bayesian inference. *J. R. Stat. Soc. Ser. B*, 41:113-147.
-  Jeffreys, H. (1961). Theory of Probability. Oxford University Press, London.
-  Bernardo, J. y Smith, A. (1994). Bayesian Theory. John Wiley, New York..
-  Box, G.E.P. y Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis. Addison-Wesley.