

## Módulo 3

# INTRODUCCIÓN AL ENFOQUE BAYESIANO

Curso de Posgrado: “Modelado y estimación de ocupación para poblaciones y comunidades de especies bajo enfoque Bayesiano”

CCT CONICET Mendoza  
24 - 28 Abril 2023



Instituto Nacional de  
Tecnología Agropecuaria  
Argentina



**GTBA**

Grupo Transdisciplinario de  
Biodiversidad y Agroecosistemas



**CONICET**



# ORIGEN DE INFERENCIA BAYESIANA

## Reverendo Thomas Bayes

- Nacido en 1701 (o 1702?) - 1761, Londres (o Hertfordshire?) Inglaterra
- “An essay towards solving a Problem in the doctrine of chances” (1763)



*“Dado el número de veces en que un evento desconocido ocurrió y falló: se requiere la posibilidad de que la probabilidad de que ocurra en un solo evento esté en un punto entre dos grados de probabilidad que pueden ser nombrados”*

# ORIGEN DE INFERENCIA BAYESIANA

*“Dado el número de veces en que un evento desconocido ocurrió y falló: se requiere la posibilidad de que la probabilidad de que ocurra en un solo evento esté en un punto entre dos grados de probabilidad que pueden ser nombrados”*

- $X \sim \text{Binomial}(1, p)$  (i.e. Bernoulli)  $\rightarrow$  evento aleatorio con dos resultados posibles
- “ $p$ ” = probabilidad de éxito en la prueba Bernoulli
- Teniendo en cuenta la incertidumbre acerca de  $p$

...pero es diferente al intervalo de confianza frecuentista!

# INFERENCIA BAYESIANA vs. FRECUENTISTA

- El uso del término “probabilidad”
  - ↗ Sets de datos hipotéticos (F)
  - ↘ Cantidades desconocidas; medida de la incertidumbre (B)
- ¿Qué preguntan?
  - *¿Cuál es la probabilidad de observar éstos datos, dado que dichas hipótesis son verdaderas?* (F)
  - *¿Cuál es la probabilidad que las hipótesis sean verdaderas dados los datos observados?* (B)
- Intervalos de Confianza vs. Intervalos de Credibilidad
  - I. Confianza: no contiene el parámetro con una probabilidad del 95%, sino que está basado en un número infinito de muestras **hipotético** (F)
  - I. Credibilidad: Probabilidad que el valor verdadero del parámetro se encuentre dentro de cierto rango, dado éstos mismos datos, modelo y previas (B)

# INFERENCIA BAYESIANA vs. FRECUENTISTA

Resumiendo podríamos decir que la gran diferencia es que con la inferencia Bayesiana, podemos obtener distribuciones de probabilidades para cantidades no observadas:

parámetros, valores perdidos, o datos nuevos que todavía no hemos colectado.

Podemos cuantificar incertidumbre y armar modelos realistas que tienen en cuenta observaciones imperfectas.



# INFERENCIA BAYESIANA vs. FRECUENTISTA

## *Frecuentista*

- Parámetros de modelos son fijos y desconocidos
- Incertidumbre evaluada sobre réplicas de set de datos hipotéticas (IC basado en muestra infinita)
- IMM

## *Bayesiano*

- Parámetros de modelos son variables aleatorias
- La probabilidad es una medida de incertidumbre sobre cantidades desconocidas (IC verdadero)
- Conocimiento previo
- Set de datos pequeños
- Modelos jerárquicos
- IMM difícil
- Computacionalmente difícil (MCMC)

# INFERENCIA BAYESIANA

- Reconoce y combina cuatro componentes del conocimiento:



conocimiento +  
previo

datos  
nuevos

modelo



conocimiento  
posterior

- Incorpora información previa para actualizar el conocimiento
- Subjetivo
  - Especificar “prior”; no todos los priors dan la misma inferencia (B)
  - Especificar estimador; no todos los estimadores dan la misma inferencia (F)
- Parámetros no son fijos, son resultados aleatorios
  - Representan la variación del parámetro
  - O la incertidumbre en el valor del parámetro

# ¿POR QUÉ USAR INFERENCIA BAYESIANA?

- Toda inferencia se basa en el **Teorema de Bayes**
  - No hay que preocuparse por la elección del estimador
- Estructura del modelo mas rica; relativamente fácil de implementar
- Focaliza en los modelos, no en el método
- Es objetivo: inferencia basada en la distribución posterior
- Relevante para situaciones de muestras pequeñas
- En muchos casos los MLE son suficientes. En otros, los métodos Bayesianos pueden ser muy útiles (o la única opción)
  - Modelos jerárquicos y efectos aleatorios
  - Interacciones complejas entre parámetros y estructuras de datos
- Múltiples fuentes de incertidumbre



# TEOREMA DE BAYES

- Probabilidades Condicionales

*La probabilidad de un evento dado que sabemos que el otro ejemplo ocurrió*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

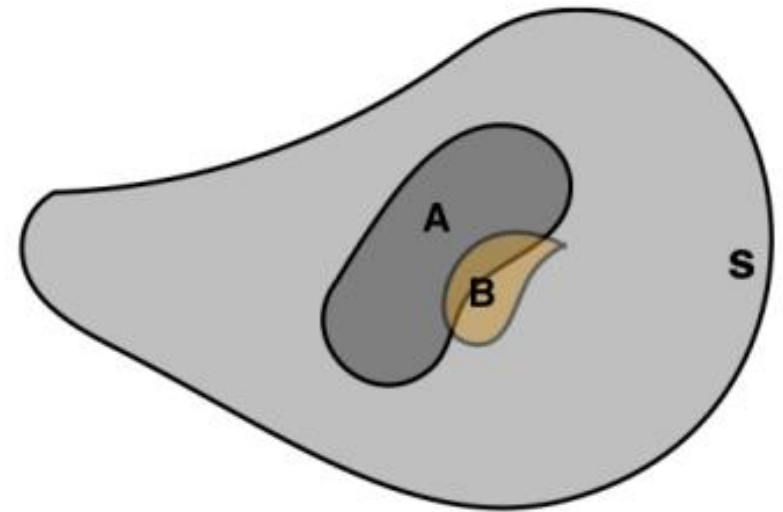
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$P(A|B) \times P(B) = P(A \cap B)$$

$$P(B|A) \times P(A) = P(A \cap B)$$

$$P(B|A) \times P(A) = P(A|B) \times P(B)$$

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A)}$$



# TEOREMA DE BAYES

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}$$

Ejemplo...

$$P(H_p|D) = \frac{P(H_p) \times P(D|H_p)}{P(D)}$$



$$P(H_p|D) = \frac{P(H_p) \times P(D|H_p)}{P(H_p) \times P(D|H_p) + P(H_a) \times P(D|H_a)}$$

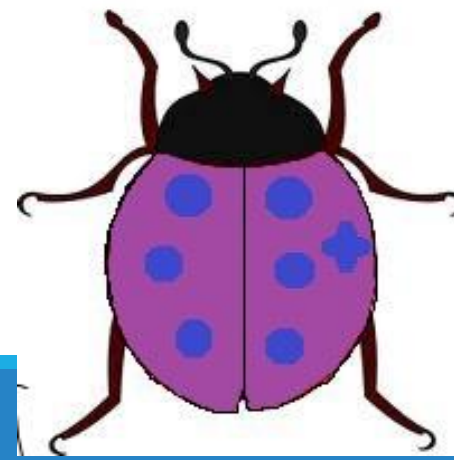
# TEOREMA DE BAYES

Un entomólogo encuentra lo que sería una subespecie rara de un coleóptero, dado a un patrón en las manchas de sus élitros. En la subespecie rara, el 98% tiene el patrón.

En la subespecie común, 5% tiene un patrón similar, pero el entomólogo no puede distinguirlos de memoria. La subespecie rara representa sólo el 0.1% de la población. ¿Cuál es la probabilidad que éste coleóptero sea raro?

$$P(\text{raro}|\text{patrón}) = \frac{P(\text{raro}) \times P(\text{patrón}|\text{raro})}{P(\text{raro}) \times P(\text{patrón}|\text{raro}) + P(\text{común}) \times P(\text{patrón}|\text{común})}$$

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$



# TEOREMA DE BAYES

98%  $P(\text{patrón}|\text{raro})$

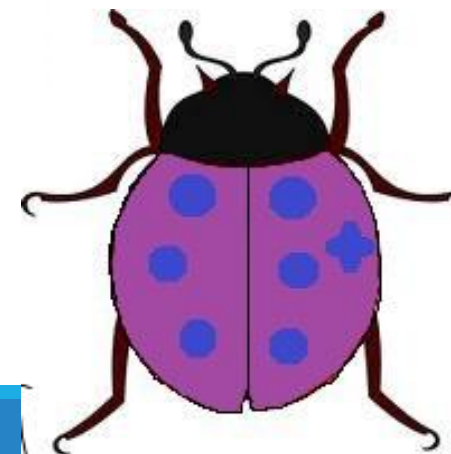
5%  $P(\text{patrón}|\text{común})$

0.1%  $P(\text{raro})$

$$P(\text{raro}|\text{patrón}) = \frac{P(\text{raro}) \times P(\text{patrón}|\text{raro})}{P(\text{raro}) \times P(\text{patrón}|\text{raro}) + P(\text{común}) \times P(\text{patrón}|\text{común})}$$

$$P(\text{raro}|\text{patrón}) = \frac{0.001 \times 0.98}{0.001 \times 0.98 + \underline{0.999 \times 0.05}}$$

$$P(\text{raro}|\text{patrón}) = 1.9\%$$



# TEOREMA DE BAYES

- Supongo  $x$  son los datos, y  $\{H_1, H_2, \dots, H_n\}$  son hipótesis que compiten mutuamente excluyentes y exhaustivas (modelos)\*:

$$P(H_i|x) = \frac{P(H_i) \times P(x|H_i)}{\sum_{j=1}^n P(H_j) \times P(x|H_j)}$$

- ...o parámetros

$$P(\theta_i|x) = \frac{P(\theta_i) \times P(x|\theta_i)}{\sum_{j=1}^n P(\theta_j) \times P(x|\theta_j)}$$

\* Podemos actualizar las creencias relativas en los modelos alternativos usando las verosimilitudes de cada modelo

# COMPONENTES EN INFERENCIA BAYESIANA

- Cantidades son conocidas o desconocidas
  - **Conocidas** (observadas)
    - Datos
    - Covariables
  - **Desconocidas** (no observadas) todas estas cantidades son tratadas de la misma manera
    - Parámetros
    - Predicciones
    - Datos faltantes

# COMPONENTES EN INFERENCIA BAYESIANA

Distribución a priori  
(previa)

Verosimilitud  
o “likelihood”

$$P(\theta_i | x) = \frac{P(\theta_i) \times P(x | \theta_i)}{\sum_j P(\theta_j) \times P(x | \theta_j)}$$

Probabilidad total de  
los datos  $P(x)$

Distribución Posterior  
de los parámetros,  
dado que observamos  
los datos  $x$

# COMPONENTES EN INFERENCIA BAYESIANA

$$P(\theta_i|x) = \frac{P(\theta_i) \times P(x|\theta_i)}{\sum_j P(\theta_j) \times P(x|\theta_j)}$$

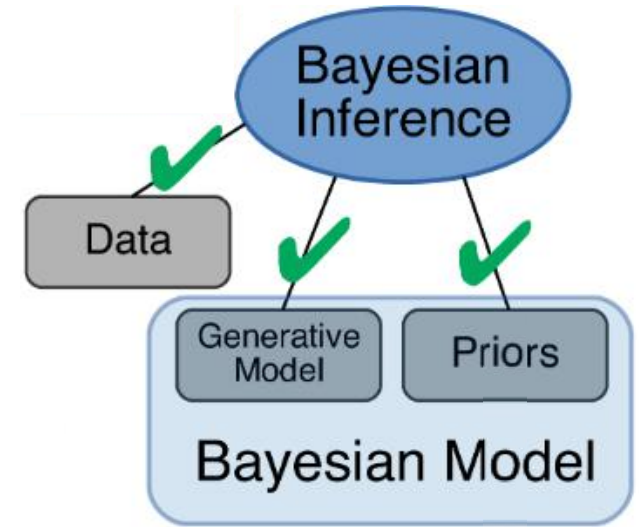
Eventos  $j$  son mutuamente exclusivos y exhaustivos; entonces,  $P(x)$  es un estado de la naturaleza, con explicaciones alternativas de cómo las cosas funcionan, y una constante

$$P(\theta_i|x) \propto P(\theta_i) \times P(x|\theta_i)$$

*Distribución posterior  $\propto$  Distribución a priori  $\times$  Verosimilitud*



# COMPONENTES EN INFERENCIA BAYESIANA



- $[x]$  Datos
  - $[x|\theta]$  Verosimilitud ("likelihood") de los datos condicional a  $\theta$
  - $[\theta]$  Distribución a priori del parámetro  $\theta$ 
    - Informativo
    - No-informativo
- $[\theta|x]$  Probabilidad posterior de  $\theta$ , dado que observamos los datos  $x$
- Casos simples pueden ser encontrados fácilmente, pero en la mayoría de los casos tiene que ser aproximado

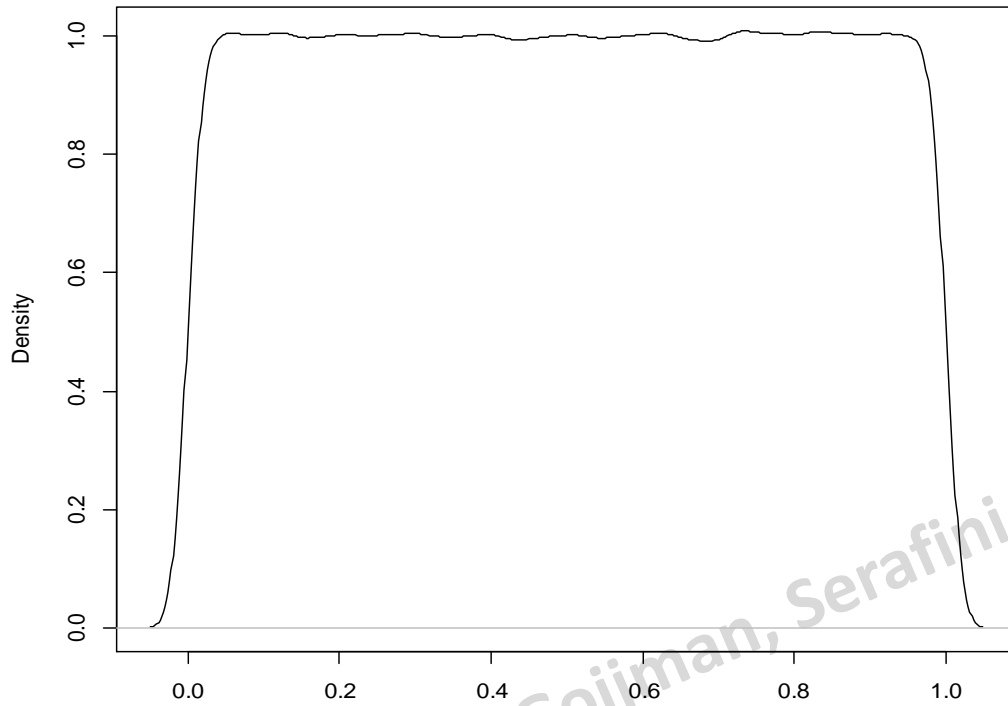
# COMPONENTES EN INFERENCIA BAYESIANA

- A priori no-informativas (“vagos”)
  - Generalmente, los “a priori” que tienen un mínimo rol en la inferencia
  - Dejan a los datos hablar por si mismos
  - Inferencia basada en los datos (los datos dominan la distribución posterior)
  - Ejemplos:
    - $\theta \sim \text{Beta}(1,1) = U(0,1)$
    - $\theta \sim \text{Normal}(\mu, \sigma)$   $\sigma = \text{grande}$  (baja precisión)

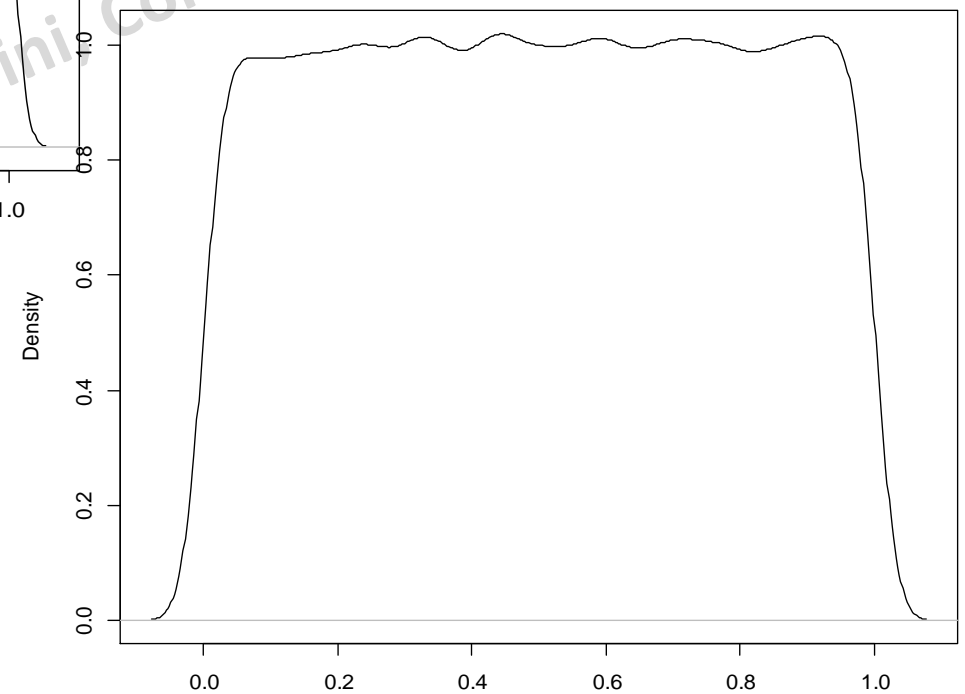
**BE CAREFUL**

**THIS MACHINE  
HAS NO BRAIN  
USE YOUR OWN**

Densidad  $x \sim \text{Beta}(1,1)$



Densidad  $x \sim \text{Uniforme}(0,1)$

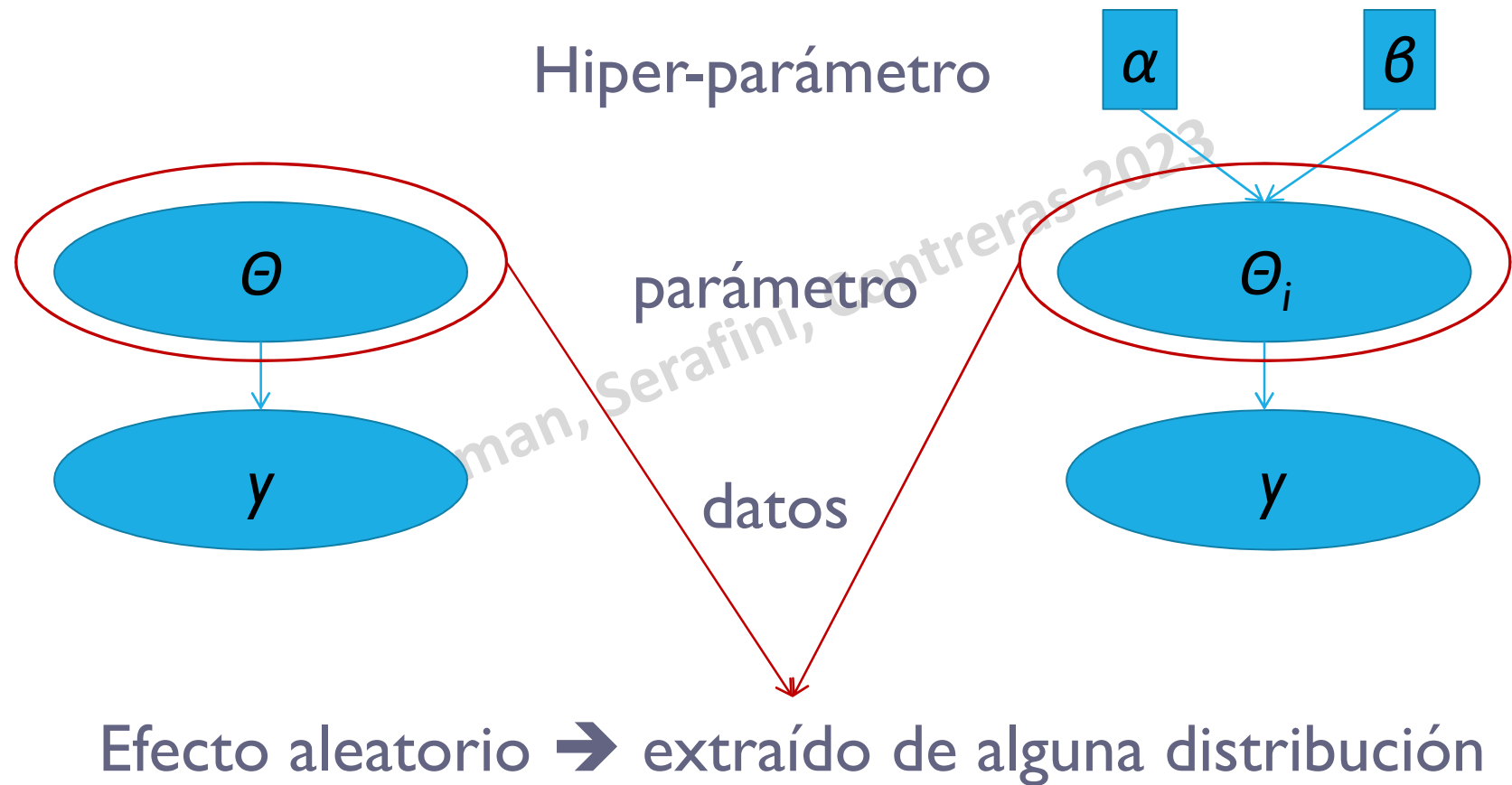


# MODELOS ALEATORIOS Y JERÁRQUICOS

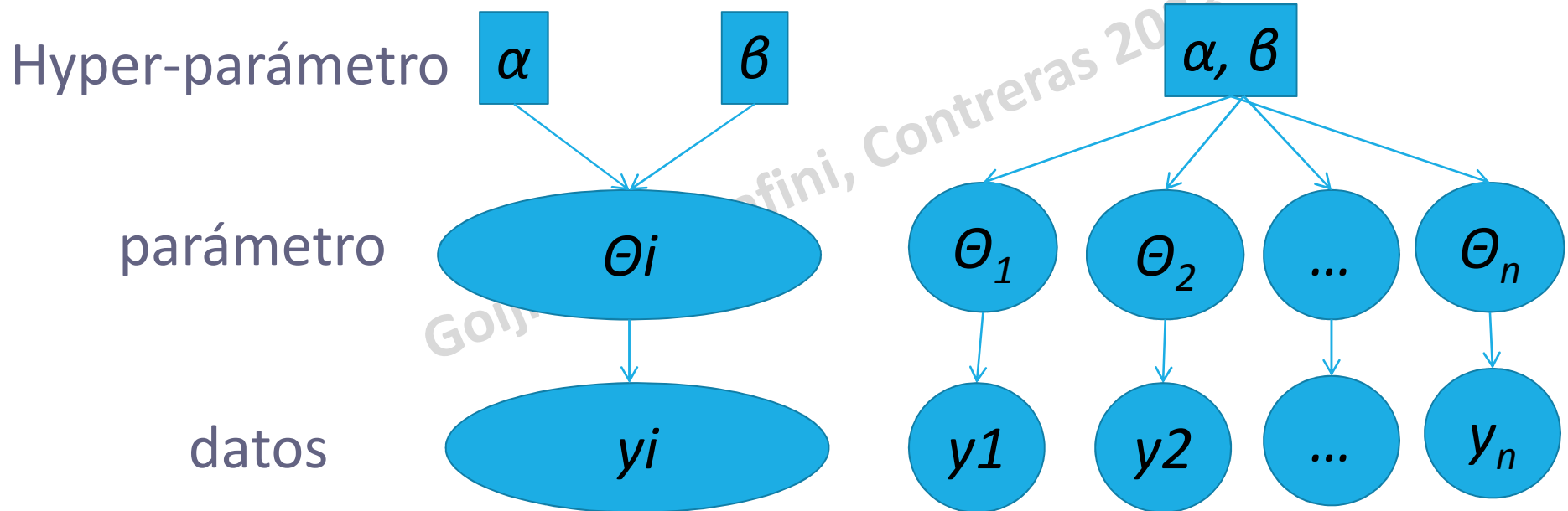
El término ‘modelos jerárquicos’ se ha convertido prácticamente en sinónimo de ‘Análisis Bayesiano’

- El tipo de análisis e inferencia es independiente de la formulación del modelo
- Los enfoques Bayesianos son naturales, convenientes y más efectivos
- A veces es la única manera para modelos complejos

# MODELOS ALEATORIOS Y JERÁRQUICOS



# MODELOS ALEATORIOS Y JERÁRQUICOS



# ¡UN MINUTO!


- Si los enfoques Bayesianos
  - Son más intuitivos
  - Incorporan el conocimiento previo
  - Permiten fácilmente la construcción de modelos de efectos aleatorios y jerárquicos
  - Etc...,



... ¿por qué no han sido más utilizados?

... ¿por qué no todo el mundo es Bayesiano?

# Ah....

- Resistencia a utilizar información a priori
  - Lo enfoques Bayesianos solían ser aplicados sólo a modelos complejos
  - Libros son difíciles de entender
  - No son enseñados en las universidades
  - Falta de programas
- 



# $[\theta|x]$ DISTRIBUCIÓN POSTERIOR DE $\theta$

- En casos sencillos puede ser encontrada fácilmente, en un paso
- La mayoría de los casos puede ser aproximada usando MCMC para generar muestras que converjan en la distribución posterior

$$P(\theta_i|x) = \frac{P(\theta_i) \times P(x|\theta_i)}{\sum_j P(\theta_j) \times P(x|\theta_j)}$$

- El denominador es difícil de calcular analíticamente! Pero métodos numéricos como MCMC hacen que los métodos Bayesianos sean accesibles a todos los científicos

# Cadena de Markov Monte Carlo (MCMC)

- ‘Monte Carlo’ → muestreo aleatorio (independiente).
- ‘Cadena de Markov’ → método de generación de muestras aleatorias. Series de números al azar donde el valor de cada uno es condicional al número inmediato anterior.
- Algoritmos MCMC se construyen de manera que las muestras son equivalentes a **muestras de la distribución posterior**.

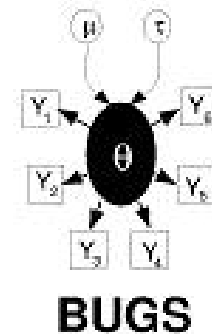
# Cadena de Markov Monte Carlo (MCMC)

- La ventaja de utilizar MCMC para muestrear de la distribución posterior es que no es necesario calcular el valor del denominador:
  - El cálculo es evitado porque cada muestra sucesiva depende de la proporción de dos distribuciones posteriores que comparten el mismo denominador, que se cancela (constante).

$$P(\theta_i | x) \propto P(\theta_i) \times P(x | \theta_i)$$

# Cadena de Markov Monte Carlo (MCMC)

- Escribí tu propio algoritmo
  - Metropolis-Hastings
  - Gibbs sampling (caso especial de M-H)
  - Combinación
- Usá Software (“enlatado”)
  - WinBUGS/BUGS
  - JAGS



```
JAGS 4.2.0 (32-bit)

C:\Users\andrea>jags-terminal
Welcome to JAGS 4.2.0 on Mon Mar 13 10:27:42 2023
JAGS is free software and comes with ABSOLUTELY NO WARRANTY
Loading module: basemod: ok
Loading module: bugs: ok
.
```

# GIBBS SAMPLER (Muestreo de Gibbs)

- Fijá un parámetro y sorteas una muestra al azar de la distribución posterior:

1.  $\text{Sortea}[\theta_1^{(1)} \sim \theta_1 | \theta_2^{(0)}, \theta_3^{(0)}, y]$

2.  $\text{Sortea}[\theta_2^{(1)} \sim \theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, y]$

3.  $\text{Sortea}[\theta_3^{(1)} \sim \theta_3 | \theta_1^{(1)}, \theta_2^{(1)}, y]$

4. Iterar... hasta llegar a una distribución estable (\*)

(\*) variedad de diagnósticos... o correr una cadena larga!

# JAGS

## What is JAGS?

JAGS is Just Another Gibbs Sampler. It is a program for analysis of Bayesian hierarchical models using Markov Chain Monte Carlo (MCMC) simulation not wholly unlike [BUGS](#). JAGS was written with three aims in mind:

- To have a cross-platform engine for the BUGS language
- To be extensible, allowing users to write their own functions, distributions and samplers.
- To be a platform for experimentation with ideas in Bayesian modelling

JAGS is licensed under the GNU General Public License version 2. You may freely modify and redistribute it under certain conditions (see the file COPYING for details).

## News

[No Title]

See the [JAGS NEWS](#) blog for news about the project. If you want to be kept informed of updates to JAGS, then subscribe to the RSS news feed.

## Latest version

The latest release is JAGS 4.3.1. It was released on 12 April 2022.

## Downloads

To download JAGS, please visit the [files page](#) of the mcmc-jags project at sourceforge. You will find the source for JAGS there as well as binary packages for Mac OS X (Thanks to Matt Denwood and the pioneering work of Bill Northcott) and Windows.

Binaries for Linux are distributed separately. There are packages for [various RPM-based Linux distributions](#) (RHEL 5, 6, 7; OpenSUSE 13.1, 13.2, TumbleWeed; CentOS 5, 6, 7; Fedora 20) (Thanks to Lars Vilhuber), [Debian](#) and the development version of [Ubuntu](#) (Thanks to Dirk Eddelbuettel). You can track the current status of the Debian packages [here](#). Another [Ubuntu repository](#) for older versions of Ubuntu is provided by Michael Rutter.

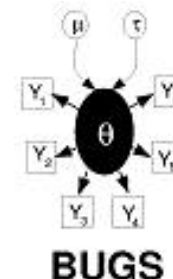
Daniel Meliza also maintains a port of JAGS on [MacPorts](#).





# The BUGS Project

## welcome



Welcome  
Page

Latest News

Contact  
us/BUGS list

WinBUGS

New  
WinBUGS  
examples

The BUGS  
Book

FAQs

DIC

GeoBUGS

## Background to BUGS

The BUGS (**B**ayesian inference **U**sing **G**ibbs **S**ampling) project is concerned with flexible software for the Bayes methods. The project began in 1989 in the MRC Biostatistics Unit, Cambridge, and led initially to the 'Classic' Bi College School of Medicine at St Mary's, London.

Development is now focussed on the [OpenBUGS](#) project.

## WinBUGS 1.4.3

This site at the MRC Biostatistics Unit is primarily concerned with the stand-alone [WinBUGS 1.4.3](#) package.

- Features a graphical user interface and on-line monitoring and convergence diagnostics.
- Over 30000 downloads, and a huge number of [applications](#) and links.
- [WinBUGS development site](#) includes facilities to add distributions, functions, and includes add-ons for pha
- Can be called from R with [R2WinBUGS](#).

WinBUGS 1.4.3 is a stable version which is recommended for standard use. However many developments are n

**Note: The preferred reference for citing WinBUGS in scientific papers is:** Lunn, D.J., Thomas, A., Best, N. structure, and extensibility. *Statistics and Computing*, **10**:325--337.

## OpenBUGS

# Ejemplo – Dist a priori no informativa

WinBUGS14

File Tools Edit Attributes Info Model Inference Options Doodle Map Text Window Help

WinBUGS Licence

zorzal\_tamano\_lecture

```
#Promedio tamaño Zorzales (n=10)
```

```
model
```

```
{  
  mean ~ dnorm (0, 1.0E-6) # tamaño medio de Zorzales  
                        # (uninformative prior)  
  var ~ dlnorm (0, 1.0E-6) # varianza tamaño Zorzales  
                        # (uninformative prior)  
  
  prec <- 1/var #varianza a presicion
```

```
for (i in 1:10) #para cada uno de los 10 zorzales
```

```
{  
  Size[i] ~ dnorm (mean, prec) # Tamaño Zorzal trazado de  
                        #normal distribution  
}
```

```
#Tamaño 10 Zorzales
```

```
Data
```

```
list(Size=c(7.9, 8.1, 11, 10.6, 9.2, 8, 9.8, 10.1, 10.9, 9))
```

```
Initial Values
```

```
list(var=100, mean=10)
```

```
#Average size of American Robins (n=10), with prior information (mean=9.4  
(0.81 0.52))
```

Specification Tool

check model load data

compile num of chains 3

load inits for chain 1

gen inits

Sample Monitor Tool

node mean chains 1 to 3 percentiles

beg 1 end 1000000 thin 1

clear set trace history density

stats coda quantiles bgr diag auto cor

2.5
5
10
25
median
75
90
95
97.5

Update Tool

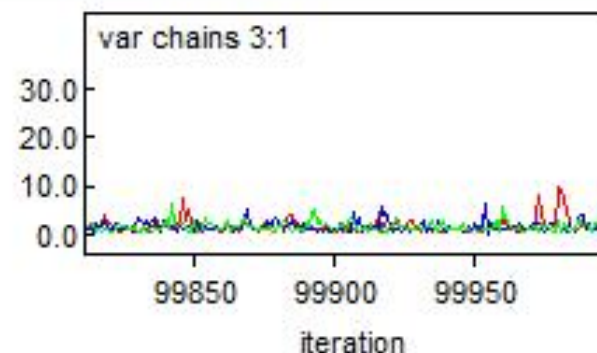
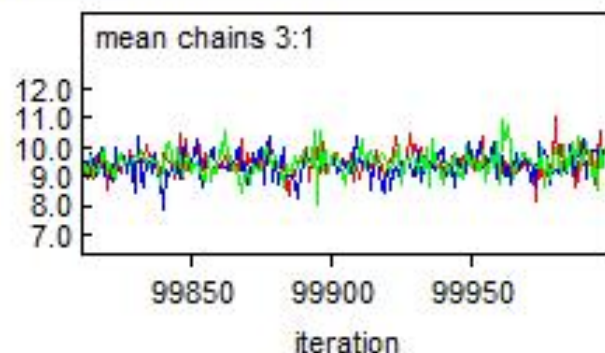
updates 100000 refresh 100

update thin 1 iteration 0

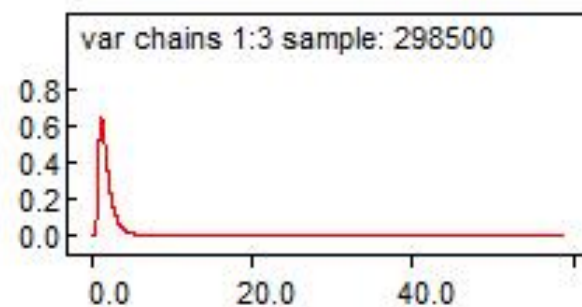
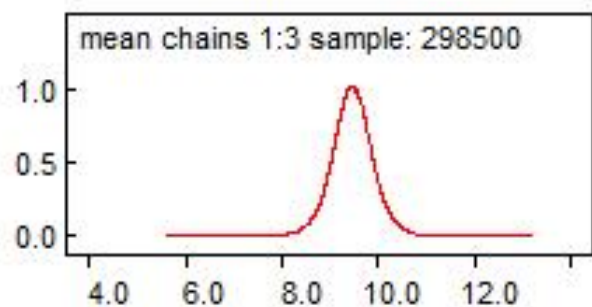
☐ over relax ☐ adapting



## Dynamic trace



## Kernel density



## Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
mean	9.46	0.4296	8.077E-4	8.601	9.46	10.32	501	298500
var	1.845	1.174	0.003894	0.682	1.55	4.788	501	298500

# Ejemplo – Dist a priori informativa

```
WinBUGS14
File Tools Edit Attributes Info Model Inference Options Doodle Map Text Window Help

zorzal_tamano_lecture

#Tamaño promedio de Zorzales (n=10), con información prior (mean=9.45, sd=0.8); var
(0.81, 0.52)

model
{
  mean ~ dnorm (9.45, 1.56) # tamaño medio de zorzales en el sitio
                             # (informative prior)
  var ~ dlnorm (0.81, 1.92) # variance de tamaño zorzal
                           # entre sitios (un sitio tiene mas
                           # diff. tamaños que otros
                           # (informative prior)

  prec <- 1/var #variance to precision

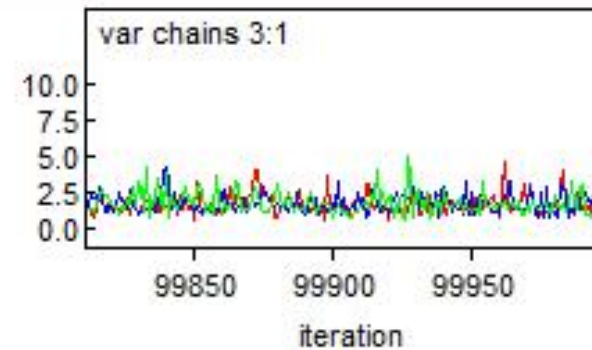
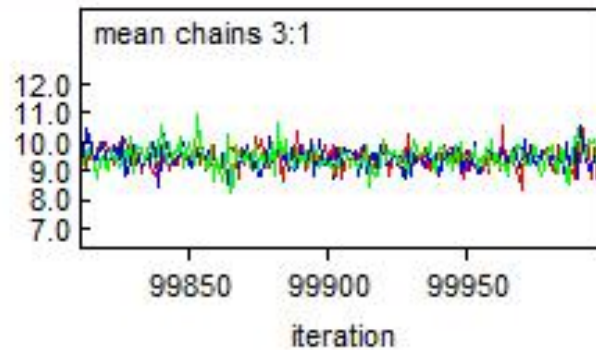
  for (i in 1:10) #para cada uno de los 10 zorzales
  {
    Size[i] ~ dnorm (mean, prec) # Tamaño Zorzal generado de
                                #normal distribution
  }
}

#Tamaño de 10 Zorzales
Data
list(Size=c(7.9, 8.1, 11, 10.6, 9.2, 8, 9.8, 10.1, 10.9, 9))

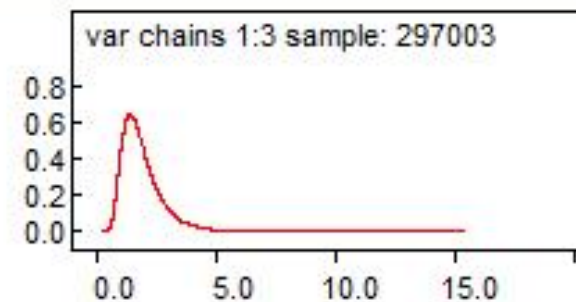
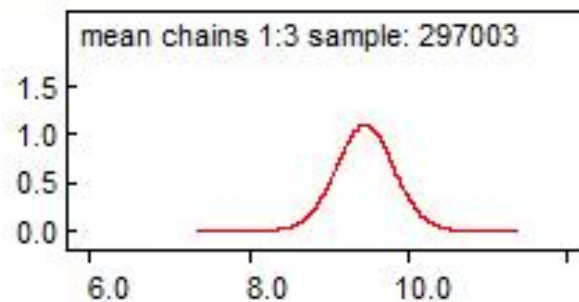
Initial Values
list(var=100, mean=10)
```

3

### Dynamic trace



### Kernel density



### Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
mean	9.458	0.3757	6.739E-4	8.711	9.458	10.2	1000	297003
var	1.88	0.8641	0.002346	0.8073	1.69	4.068	1000	297003

## Resultados: Distribución a priori Informativa vs. No-Informativa

Modelo	Nodo	Mean	SD	2.5%	97.5%
No-informativo	Mean	9.46	0.4296	8.601	10.32
	Var	1.845	1.174	0.682	4.788
Informativo	Mean	9.458	0.3757	8.711	10.2
	Var	1.88	0.8641	0.807	4.068

# ALGUNAS CONSIDERACIONES PARA SU IMPLEMENTACIÓN

**Monitoreo de Convergencia** (distribución de equilibrio alcanzada)

- Más de una cadena, iniciada de un punto arbitrario. Luego de un tiempo, las cadenas son independientes a los valores iniciales seleccionados.
- Los valores iniciales de la cadena de Markov pueden influir los resultados hasta un número suficiente de muestras descartar algunas de las primeras muestras como “burn in”.
- Cada muestra adicional contiene sólo una fracción de la información acerca de la distribución posterior se requiere un gran número de muestras.
- R-hat

# EJEMPLO MODELO JERÁRQUICOS



Model

```
{  
lambda ~ dlnorm(0, 1.0E-6)
```

→ Distr. a priori no informativa para el número medio de ranas en un charco

```
for (i in 1:15)  
{
```

→ Cara cada uno de los 15 charcos

```
y[i] ~ dpois(lambda)  
}  
}
```

→ Número de ranas que siguen un distribución Poisson con media lambda

```
data  
list (y=c(6,3,1,2,1,7,1,5,2,8,2,3,5,9,11))
```

```
initial values  
list(lambda=5)
```

PERO sabemos q el n medio de ranas en un charco, puede ser diferente, y queremos agregar variación extra

# EJEMPLO MODELO JERÁRQUICO



Model

```
{  
lambda ~ dlnorm(0, 1.0E-6)
```

**mu ~ dnorm(0, 1.0E-6)** → Prior no Informativa para media de log # de ranas

**sd ~ dunif(0,10)** → Prior no Informativa para sd of de log # de ranas

**tau <- 1/(sd\*sd)** → Presición

```
for (i in 1:15)
```

```
{
```

**lambda[i] ~ dlnorm(mu, tau)** → Media de ranas con distribución log normal

**y[i] ~ dpois(lambda[i])** → Número de ranas que siguen un distribución

```
}
```

Poisson con media lambda

```
}
```

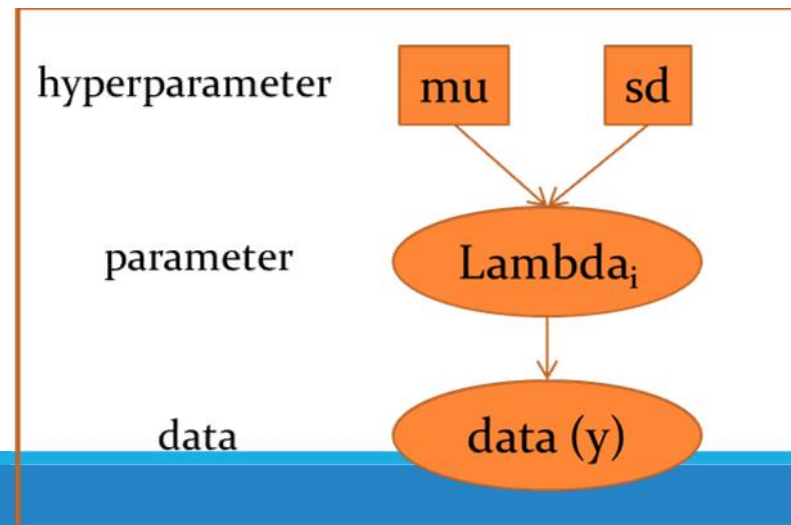
data

```
list (y=c(6,3,1,2,1,7,1,5,2,8,2,3,5,  
9,11))
```

initial values

```
list(lambda=c(5,2,1,3,4,2,5,2,1,3,4,2  
,5,2,1),
```

```
mu=2,sd=5)
```





# SELECCIÓN DE MODELOS

- Modelo más aceptable
- Muy sensible a los priors
- Dificultad para definir el número efectivo de parámetros ( $p_D$ )
- No hay una solución clara o consenso para enfoques Bayesianos
  - Mientras que la inferencia y predicción son llevados a cabo de mejor manera en los enfoques Bayesianos
  - La selección de modelos es mejor llevada a cabo desde una perspectiva frecuentista (ej. AIC)
- Inferencia multi-modelo es difícil!



# MÉTODOS DE SELECCIÓN DE MODELOS

- Bayes factors
  - Compara modelos de a pares (“likelihood ratio”)
  - No afectado por priors
- BIC (Bayesian information criteria)
  - No afectado por priors
  - Intenta identificar un modelo “verdadero”
  - Requiere especificación de pD
  - Promediado de modelos
- AIC (Akaike information criteria)
  - Dificultad en modelos de efectos mixtos, pD
  - Promediado de modelos
- DIC (Deviance information criteria)
  - Una generalización del AIC teniendo en cuenta pD en modelos mixtos (los estima)
  - No hay promediado de modelos

IF YOU'RE FEELING OVERWHELMED...  
YOU SIMPLY HAVEN'T FOUND THE

ANSWER.

KEEP LOOKING!



# REFERENCIAS

- Andrew Gelman, John B. Carlin, Hal S. Stern and Donald B. Rubin. 2004. Bayesian Data Analysis. Chapman & Hall.
- Marc Kery. 2009. Introduction to WinBUGS for Ecologists. Academic Press.
- Marc Kery and Michael Schaub. 2012. Bayesian population analysis using WinBUGS. A hierarchical perspective. Academic Press.
- William A. Link and Richard J. Barker. 2010. Bayesian Inference with Ecological Applications. Academic Press.
- Michael A. McCarthy. 2007. Bayesian Methods for Ecology. Cambridge University Press.
- J. Andrew Royle and Robert M. Dorazio. 2008. Hierarchical modeling and inference in ecology. The analysis of data from populations, metapopulations, and communities. Academic Press.

## FUENTE INTERESANTE:

<https://www.countbayesie.com/blog/2016/5/1/a-guide-to-bayesian-statistics>