

brms

Estadística bayesiana al alcance de todos

David Hervás Marín y Victoria Fornés Ferrer
Plataforma de Data Science, Bioestadística y
Bioinformática. IIS La Fe
bioestadistica@iislafe.es

IX Jornadas de Usuarios de R, Granada 2017

Paradigma bayesiano vs. frecuentista

Inferencia frecuentista

Inferencia bayesiana

Asunción: familia de la función de distribución de las variables

+

Datos

Estimación *puntual* de los parámetros del modelo mediante el MLE

$$\hat{\theta} = \arg \max \{l(\theta)\}$$

+

Información previa: $P(\theta)$

+

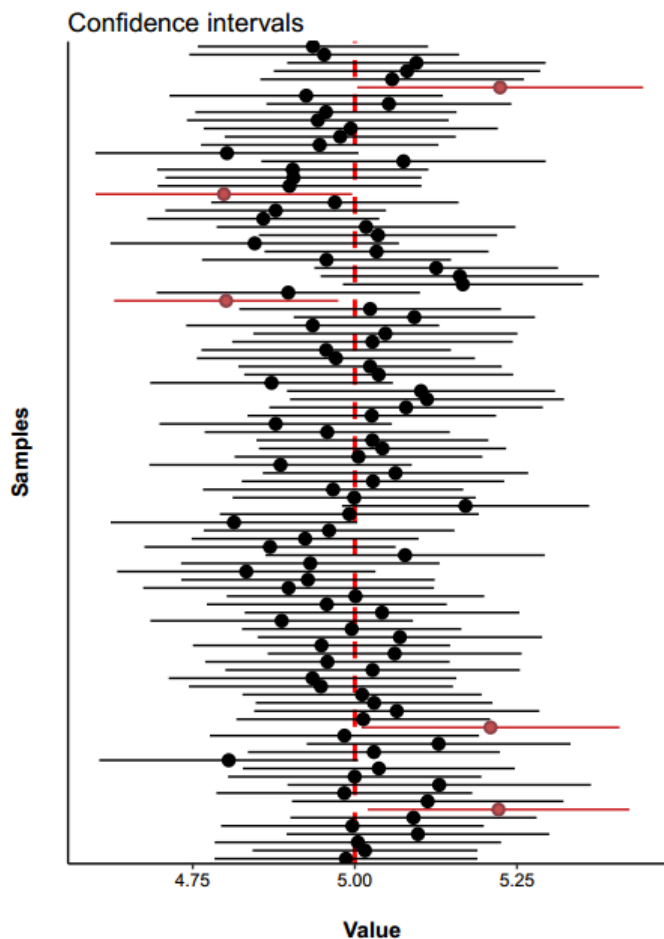
Datos: Función de verosimilitud: $P(X | \theta)$

Estimación de la *distribución* a posteriori de los parámetros del modelo mediante el Teorema de Bayes

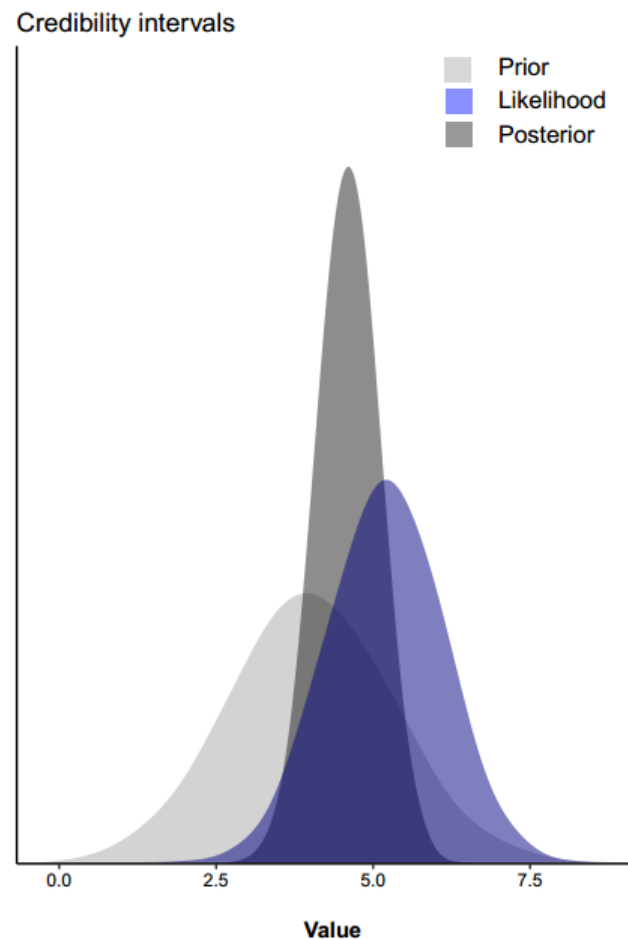
$$P(\theta|X) = P(X|\theta)*P(\theta)$$

Estimación de parámetros

Con un 95% de confianza, el intervalo contiene el verdadero valor del parámetro de la población



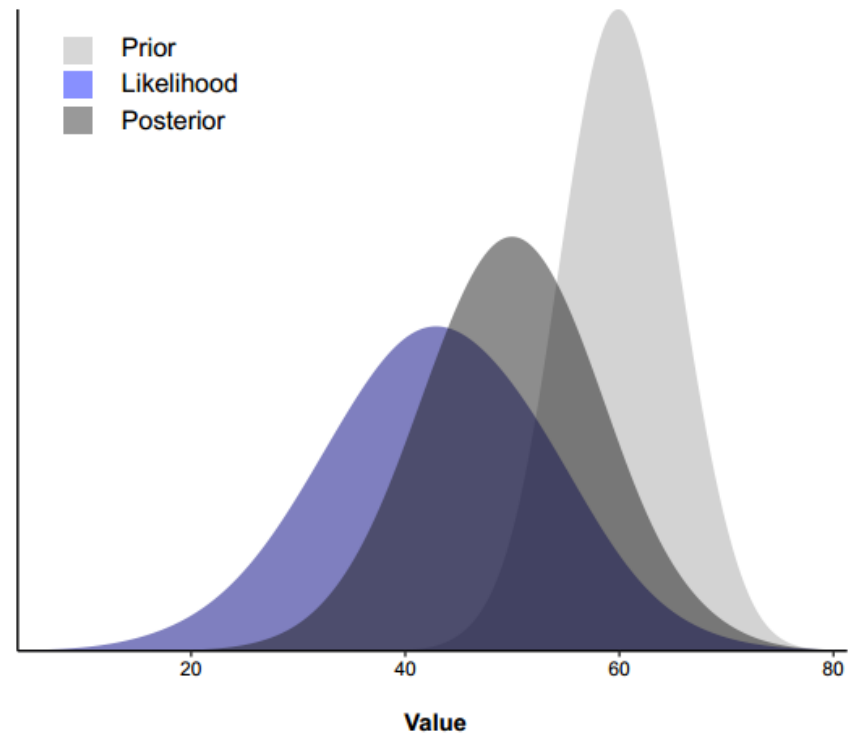
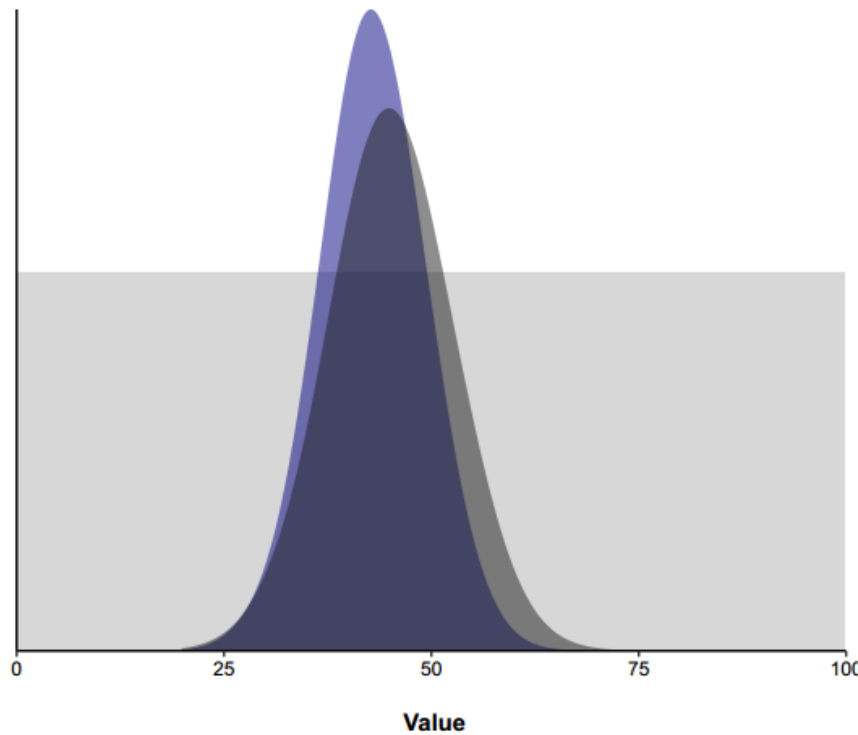
Existe una probabilidad de 0.95 de que el intervalo contenga el valor del parámetro de la población



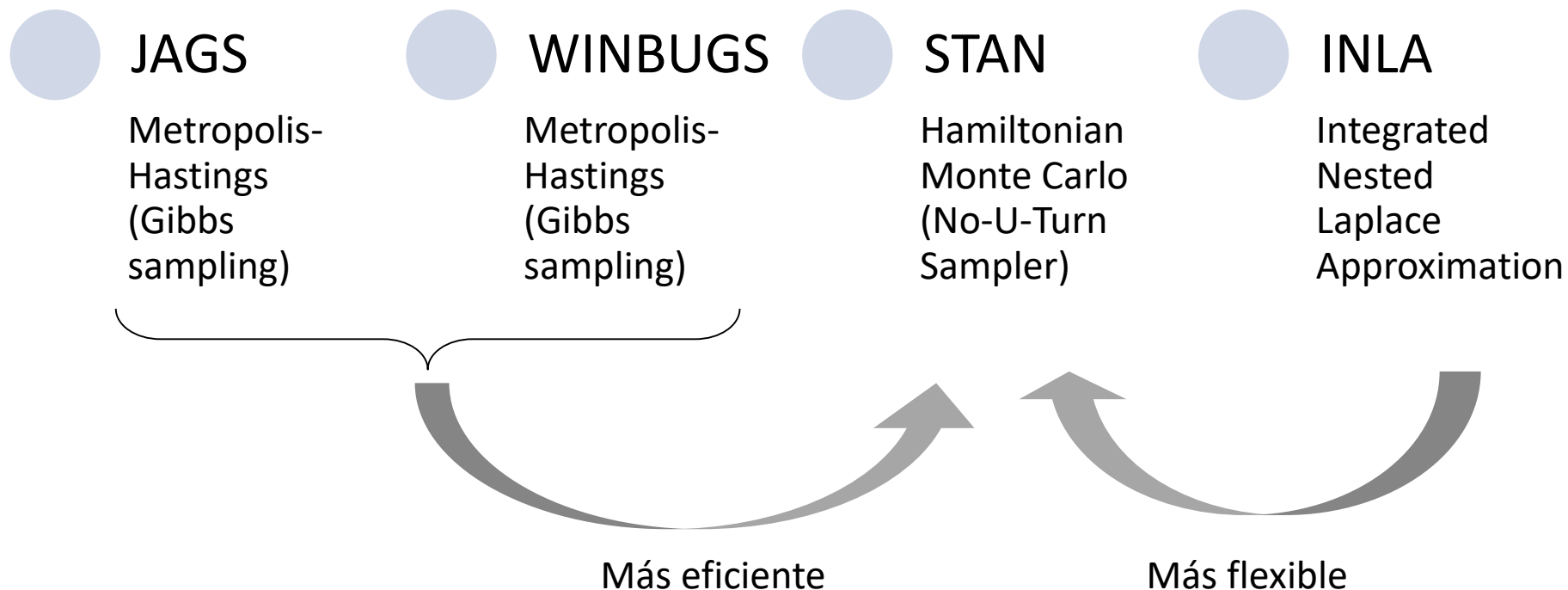
Importancia de las priors

Less informative

More informative →



Softwares disponibles



Código Stan

```
#Alternativa bayesiana a la función lm
sink("lmbayes.stan")
cat("
data {
  int N; //the number of observations
  int K; //the number of columns in the model matrix
  real y[N]; //the response
  matrix[N,K] X; //the model matrix
}
parameters {
  vector[K] beta; //the regression parameters
  real sigma; //the standard deviation
}
transformed parameters {
  vector[N] linpred;
  linpred = X*beta;
}
model {
  beta[1] ~ uniform(-500, 500);
  for(i in 2:K)
    //prior for the slopes following Gelman 2008
    beta[i] ~ uniform(-50, 50);
  y ~ normal(linpred,sigma);
}
")
sink()

#load libraries
library(rstan); library(coda)
#the model
X <- model.matrix(~ hp + disp + gear + am, data=mtcars)
y <- mtcars$mpg
m_norm<-stan(file="lmbayes.stan",data = list(N=dim(mtcars)[1],
      k=6,y=y,X=X),pars = c("beta","sigma"));m_norm
round(coef(mod1),2)
```



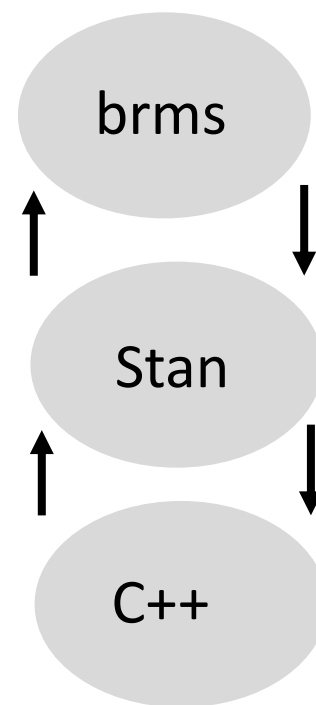
Código Stan vs brms

```
#Alternativa bayesiana a la función lm
sink("lmbayes.stan")
cat("
data {
  int N; //the number of observations
  int K; //the number of columns in the model matrix
  real y[N]; //the response
  matrix[N,K] X; //the model matrix
}
parameters {
  vector[K] beta; //the regression parameters
  real sigma; //the standard deviation
}
transformed parameters {
  vector[N] linpred;
  linpred = X*beta;
}
model {
  beta[1] ~ uniform(-500, 500);
  for(i in 2:K)
  //prior for the slopes following Gelman 2008
    beta[i] ~ uniform(-50, 50);
  y ~ normal(linpred,sigma);
}
")
sink()

#load libraries
library(rstan); library(coda)
#the model
X <- model.matrix(~ hp + disp + gear + am, data=mtcars)
y <- mtcars$mpg
m_norm<-stan(file="lmbayes.stan",data = list(N=dim(mtcars)[1],
      k=6,y=y,X=X),pars = c("beta","sigma"));m_norm
round(coef(mod1),2)
```

```
#bufffff!!!!
#Alternativa con brms
mod1b <- brm(mpg ~ hp + disp + gear + am, data=mtcars)
report(mod1b)

#Eso es otra cosa!
#Cambiar previas
mod1c <- brm(mpg ~ hp + disp + gear + am, data=mtcars,
  prior=c(set_prior("uniform(-50, 50)", class="b")))
report(mod1c)
```



brms es un paquete creado por Paul-Christian Bürkner (University of Münster) para el ajuste de modelos bayesianos que da soporte a cualquier tipo de modelo:

Linear models	Survival
Robusts linear models	Ordinal regression
Logistic regression	Zero-inflated and hurdle models
Categorical models	Generalized additive models
Count data models	Non-linear models

La sintaxis es la misma que la del paquete lme4:

```
brm(form = y ~ x1 + x2 + x3 + (1|id), data, family = gaussian(),  
prior = NULL, chains = 4, iter = 2000, ...)
```


Let's practice!!



David Hervás Marín

bioestadistica@iislafe.es

Victoria Fornés Ferrer

apoyo_bioestadistica@iislafe.es

Plataforma de Data Science, Bioestadística y
Bioinformática. IIS La Fe

IX Jornadas de Usuarios de R, Granada 2017



<https://github.com/David-Hervas/9JRes>

<https://cran.r-project.org/bin/windows/Rtools/>