

# Statistique bayésienne avec R

Julien JACQUES

# Motivations



## Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine

Fernando P. Polack, M.D., Stephen J. Thomas, M.D., Nicholas Kitchin, M.D., Judith Abulafia, M.D., Alexandra Gurtman, M.D., Stephen Lambart, D.M., John L. Pene, M.D., Gonzalo Pérez-Marc, M.D., Gibson D. Moniz, M.D., Christian Zerbini, M.D., Ruth Bailey, B.Sc., Kara A. Swanson, Ph.D., Satrajit Bhattacharya, Ph.D., Kenneth Koury, Ph.D., Peg Li, Ph.D., Warren V. Kalish, Ph.D., David Casper, Ph.D., Robert W. French, Jr., M.D., Lucas L. Hunslett, M.D., Orlan Tsiang, M.D., Hagene Hart, M.D., Axel Schaefer, M.D., Serhaty David, M.D., Dora B. Treisman, D.V.M., Ph.D., Susan Mather, M.D., Philip R. Durrleman, M.D., Ph.D., Vijay Gargis, M.D., Adriane G. Jensen, Ph.D., and William C. Gubler, M.D., for the COV00100 Clinical Trial Group\*

**Table 2. Vaccine Efficacy against Covid-19 at Least 7 days after the Second Dose.\***

Efficacy End Point	BNT162b2		Placebo		Vaccine Efficacy, % (95% Credible Interval)‡	Posterior Probability (Vaccine Efficacy >30%)§
	No. of Cases	Surveillance Time (n)†	No. of Cases	Surveillance Time (n)†		
Covid-19 occurrence at least 7 days after the second dose in participants with- out evidence of infection	(N=18,198)		(N=18,325)		95.0 (90.3–97.6)	>0.9999
	8	2,214 (17,411)	162	2,222 (17,511)		
Covid-19 occurrence at least 7 days after the second dose in participants with and those without evidence of infection	(N=19,965)		(N=20,172)		94.6 (89.9–97.3)	>0.9999
	9	2,332 (18,559)	169	2,345 (18,708)		

\* The total population without baseline infection was 36,523; total population including those with and those without prior evidence of infection was 40,137.

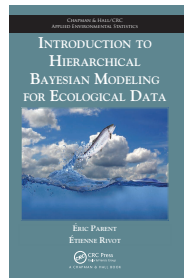
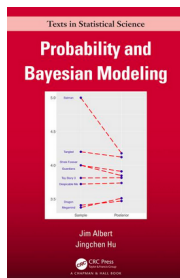
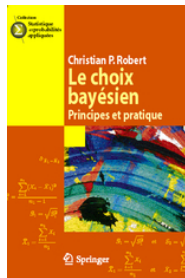
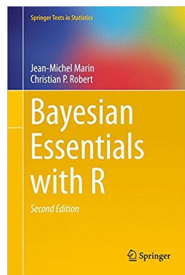
† The surveillance time is the total time in 1000 person-years for the given end point across all participants within each group at risk for the end point. The time period for Covid-19 case accrual is from 7 days after the second dose to the end of the surveillance period.

‡ The credible interval for vaccine efficacy was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

§ Posterior probability was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

**Figure 1: Pfizer-BioNTech**

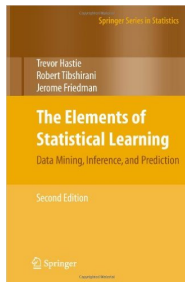
# Références bibliographiques



dont certains sont disponibles en ligne :

- ▶ Albert & Hu : <https://bayesball.github.io/BOOK/probability-a-measurement-of-uncertainty.html>

# Références bibliographiques



- ▶ Hastie et al. :  
<https://web.stanford.edu/~hastie/Papers/ESLII.pdf>

# Logiciels

Nous travaillerons avec :

- ▶ le logiciel R et les packages : bayess, BayesFactor, BMA, BAS, rjags
- ▶ le logiciel JAGS (via R)

Installation de JAGS et rjags :

1. installer JAGS : <http://sourceforge.net/projects/mcmc-jags/>
2. installer le package R rjags

Guide complet de JAGS :

<http://sourceforge.net/projects/mcmc-jags/files/Manuals/>

# Sommaire

## Les outils bayesiens

Inférence bayésienne

Rappel sur quelques lois de probabilité

Les lois a priori

Méthodes de Monte-Carlo

## Tests bayésiens

## Régression bayésienne

## Modèles graphiques (réseaux bayesiens hiérarchiques)

Les réseaux bayésiens sous JAGS

Utilisation de la modélisation hiérarchique en médecine

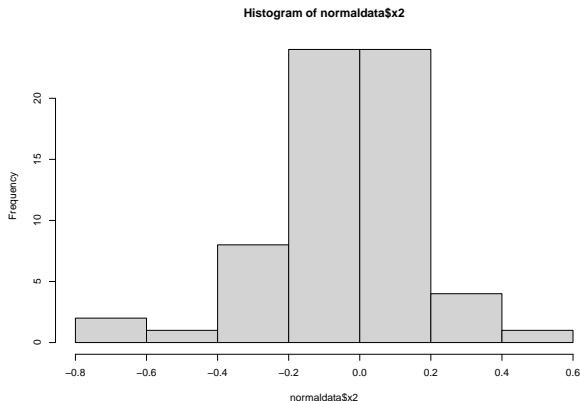
Utilisation de la modélisation hiérarchique en écologie

## Les outils bayesiens

# Les données normaldata

- mesures de différences de vitesses de deux faisceaux de lumière parcourant la même distance dans deux directions orthogonales (expérience d'Illingworth en 1927).

```
library(bayess)
data(normaldata)
hist(normaldata$x2)
```





# Inférence bayésienne

# Approche fréquentiste vs bayésienne

## Approche fréquentiste

- ▶ la variable d'intérêt  $x$  est supposée suivre une loi paramétrique :

$$x \sim f(\cdot, \theta)$$

- ▶ où  $\theta \in \Theta$  est un paramètre fixe, inconnu
- ▶ que l'on estime à partir un échantillon i.i.d.  $\underline{x} = (x_1, \dots, x_n)$
- ▶ en maximisant la **vraisemblance** des données

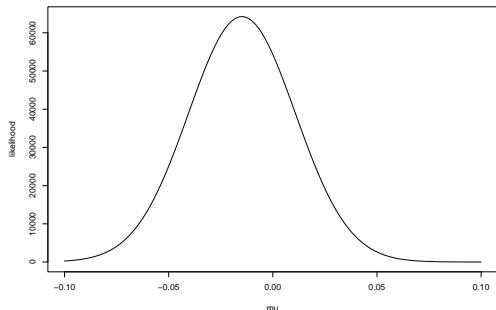
$$\ell(\underline{x}|\theta) = \prod_{i=1}^n f(x_i, \theta)$$

## Exemple 1 - normaldata

**Vraisemblance** en fonction de  $\mu$  pour un modèle  $\mathcal{N}(\mu, \sigma^2)$  :

$$\begin{aligned}\ell(\underline{x}|\theta) &= \prod_{i=1}^n \exp\{-(x_i - \mu)^2/2\sigma^2\}/\sqrt{2\pi}\sigma \\ &\propto \exp\{-[n(\mu - \bar{x})^2 + s^2]/2\sigma^2\}/\sigma^n\end{aligned}$$

```
lik=function(mu,x) prod(dnorm(x,mu,sd(x)))  
mu=seq(-.1,.1,.001); likelihood=NULL  
for (m in mu) likelihood=c(likelihood,lik(m,normaldata$x2))  
plot(mu,likelihood,type='l')
```



# Approche fréquentiste vs bayésienne

## Approche bayésienne

- ▶  $\theta$  n'est plus fixe mais considéré comme une variable aléatoire
- ▶ permet d'introduire l'information a priori dont on dispose sur  $\theta$ , i.e. l'information disponible sur  $\theta$  avant d'observer  $\underline{x}$ : pour Illingworth, il ne doit pas y avoir de différence de vitesse
- ▶ on note  $p(\theta)$  la loi a priori (*prior*) de  $\theta$

# Approche fréquentiste vs bayésienne

## Approche bayésienne

- ▶  $\theta$  n'est plus fixe mais considéré comme une variable aléatoire
- ▶ permet d'introduire l'information a priori dont on dispose sur  $\theta$ , i.e. l'information disponible sur  $\theta$  avant d'observer  $\underline{x}$ : pour Illingworth, il ne doit pas y avoir de différence de vitesse
- ▶ on note  $p(\theta)$  la loi a priori (*prior*) de  $\theta$
- ▶ on ne s'intéresse plus alors à la vraisemblance  $\ell(\underline{x}|\theta)$  mais à la loi a posteriori (*posterior*)

$$p(\theta|\underline{x}) = \frac{\ell(\underline{x}|\theta)p(\theta)}{p(\underline{x})} \propto \ell(\underline{x}|\theta)p(\theta)$$

(on reconnaît le théorème de Bayes)

Rq : contrairement à la vraisemblance, la loi a posteriori  $p(\theta|\underline{x})$  est une distribution de probabilité sur  $\Theta$

## Exemple 1 - normaldata

- ▶  $x \sim \mathcal{N}(\mu, \sigma^2)$  avec variance  $\sigma^2$  supposée connue
- ▶ supposons la loi a priori sur  $\mu$  :  $\mathcal{N}(0, \sigma^2)$
- ▶ on peut calculer la loi a posteriori

$$\begin{aligned} p(\mu|\underline{x}) &\propto p(\mu)\ell(\underline{x}|\theta) \\ &\propto \exp\{-\mu^2/2\sigma^2\} \exp\{-n(\mu - \bar{x})^2/2\sigma^2\} \\ &\propto \exp\{-(n+1)[\mu - n\bar{x}/(n+1)]^2/2\sigma^2\} \\ \Rightarrow \quad \mu|\underline{x} &\sim \mathcal{N}\left(\frac{n}{n+1}\bar{x}, \frac{\sigma^2}{n+1}\right) \end{aligned}$$

## Exemple 1 - normaldata

- ▶  $x \sim \mathcal{N}(\mu, \sigma^2)$  avec variance  $\sigma^2$  supposée connue
- ▶ supposons la loi a priori sur  $\mu$  :  $\mathcal{N}(0, \sigma^2)$
- ▶ on peut calculer la loi a posteriori

$$\begin{aligned} p(\mu|\underline{x}) &\propto p(\mu)\ell(\underline{x}|\theta) \\ &\propto \exp\{-\mu^2/2\sigma^2\} \exp\{-n(\mu - \bar{x})^2/2\sigma^2\} \\ &\propto \exp\{-(n+1)[\mu - n\bar{x}/(n+1)]^2/2\sigma^2\} \\ \Rightarrow \mu|\underline{x} &\sim \mathcal{N}\left(\frac{n}{n+1}\bar{x}, \frac{\sigma^2}{n+1}\right) \end{aligned}$$

- ▶ Rq 1 : la moyenne de la loi a posteriori est plus proche de 0 que l'estimateur classique  $\bar{x}$  de  $\mu$  : **cela est dû à la loi a priori centrée en 0, qui suppose donc que  $\mu$  est assez proche de 0.**
- ▶ Rq 2 : l'influence de la loi a priori s'estompe quand la taille d'échantillon grandit

# Estimation bayésienne

Inférence bayésienne :

- ▶ estimation ponctuelle  $\hat{\theta}$  de  $\theta$  ?
  - ▶ on peut choisir l'espérance de  $p(\theta|\underline{x})$
  - ▶ on peut aussi choisir le maximum a posteriori (MAP) : la valeur  $\hat{\theta}$  qui maximise  $p(\theta|\underline{x})$  en  $\theta$  (*mais pas toujours simple à trouver. . .*)
- ▶ l'approche bayésienne est une **approche inférentielle complète** qui permet de tout faire : estimation, intervalle de confiance, test d'hypothèse, vérification de modèle, . . .



# Intervalle de crédibilité

- ▶ dans une approche bayésienne on parle d'**intervalle de crédibilité** plutôt que d'intervalle de confiance :

$$p(\theta \in IC(\theta, \underline{x}) | \underline{x}) = 1 - \alpha$$

où  $1 - \alpha$  est la probabilité qu'un  $\theta \sim p(\theta | \underline{x})$  appartienne à l'intervalle  $IC(\theta, \underline{x})$ .

- ▶ le calcul de cet intervalle est trivial dès lors que l'on dispose de  $p(\theta | \underline{x})$ , en utilisant les quantiles de cette dernière
- ▶ attention : la loi a posteriori n'est pas toujours simple à obtenir (calcul explicite, approximation par simulation. . .)

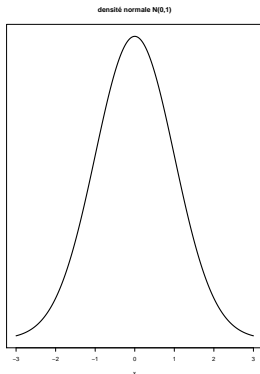
## Exercice 1 - normaldata

- ▶ récupérer les données `normaldata` dans le package `bayess`. Dans ces données, la seconde variable (que l'on appellera `shift`) mesure de différences de vitesses de deux faisceaux de lumière parcourant la même distance dans deux directions orthogonales (expérience d'Illingworth en 1927).
- ▶ estimer l'espérance  $\mu$  de la variable `shift` par maximum de vraisemblance. Tracer la vraisemblance.
- ▶ en supposant une loi a priori  $\mathcal{N}(0, \sigma^2)$  sur  $\mu$ , avec  $\sigma^2$  la variance de la variable `shift`, estimer  $\mu$  par maximum a posteriori. Tracer la loi a posteriori.

## Rappel sur quelques lois de probabilité

# Loi normale $\mathcal{N}(\mu, \sigma^2)$

- ▶ densité :  $p(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-(x - \mu)^2/2\sigma^2)$
- ▶ espérance :  $\mu$
- ▶ variance :  $\sigma^2$
- ▶ sous R : `dnorm`



# Loi de Bernoulli $\mathcal{B}(p)$

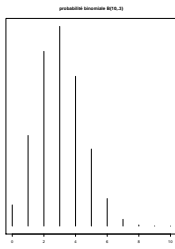
Modélise une variable aléatoire pouvant prendre 2 valeurs  $\{0, 1\}$ , 1 en cas de succès et 0 en cas d'échec, où  $p$  est la probabilité du succès.

- ▶ probabilité :  $p(x = 1|p) = p$
- ▶ espérance :  $p$
- ▶ variance :  $p(1 - p)$
- ▶ sous R : `dbinom` avec `size=1`

# Loi binomiale $\mathcal{B}(n, p)$

Modélise le nombre de succès lorsqu'on répète  $n$  expériences indépendante de Bernoulli.

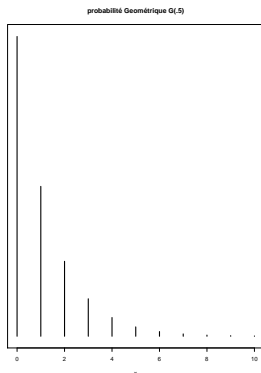
- ▶ probabilité :  $p(x = k|p) = C_n^k p^k (1 - p)^{n-k}$  pour tout  $k \in \{0, \dots, n\}$
- ▶ espérance :  $np$
- ▶ variance :  $np(1 - p)$
- ▶ propriétés asymptotiques :
  - ▶  $\mathcal{B}(n, p) \rightarrow \mathcal{P}(np)$  si  $n$  grand et  $np$  petit
  - ▶  $\mathcal{B}(n, p) \rightarrow \mathcal{N}(np, np(1 - p))$  si  $np$  et  $np(1 - p)$  grands
- ▶ sous R : `dbinom`



## loi géométrique $\mathcal{G}(p)$

Modélise le nombre d'échecs d'une suite d'expériences de Bernoulli avant l'obtention du premier succès.

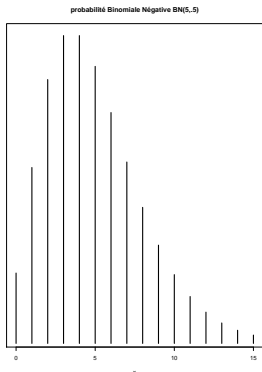
- ▶ probabilité :  $p(x = k|p) = p(1 - p)^k$  pour tout  $k \in \mathbb{N}$
- ▶ espérance :  $(1 - p)/p$
- ▶ variance :  $(1 - p)/p^2$
- ▶ sous R : `dgeom`



## loi binomiale négative $\mathcal{BN}(s, p)$

Modélise le nombre d'échecs d'une suite d'expériences de Bernoulli avant l'obtention de  $s$  succès.

- ▶ probabilité :  $p(x = k|p) = C_{s+k-1}^k p^s (1-p)^k$  pour tout  $k \in \mathbb{N}$
- ▶ espérance :  $s(1-p)/p$
- ▶ variance :  $s(1-p)/p^2$
- ▶ sous R : `dnbinom`

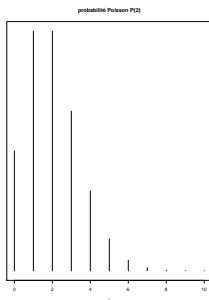




# loi de Poisson $\mathcal{P}(\lambda)$

Modélise un nombre d'événements indépendants dans un intervalle de temps donné.

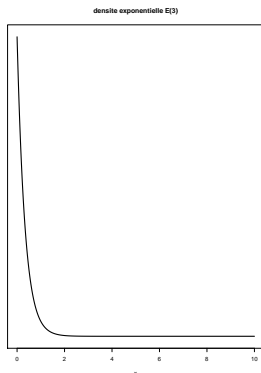
- ▶ probabilité :  $p(x = k|\lambda) = \lambda^k e^{-\lambda} / k!$  pour tout  $k \in \mathbb{N}$
- ▶ espérance :  $\lambda$
- ▶ variance :  $\lambda$
- ▶ propriété asymptotique :
  - ▶  $\mathcal{P}(\lambda) \rightarrow \mathcal{N}(\lambda, \lambda)$  si  $\lambda$  grand
- ▶ sous R : `dpois`



# loi exponentielle $\mathcal{E}(\lambda)$

Modélise un temps entre deux événements indépendants.

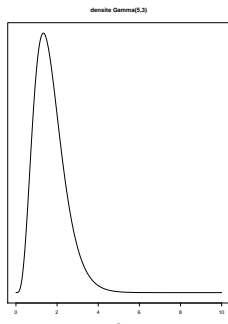
- ▶ probabilité :  $p(x|\lambda) = \lambda e^{-\lambda x}$  pour tout  $x \geq 0$
- ▶ espérance :  $1/\lambda$
- ▶ variance :  $1/\lambda^2$
- ▶ sous R : dexp



## loi gamma $\text{Gamma}(\alpha, \lambda)$

Modélise le temps nécessaire pour l'observation de  $\alpha$  événements indépendants. Cette définition est étendue au cas où  $\alpha$  n'est pas nécessairement entier (mais positif).

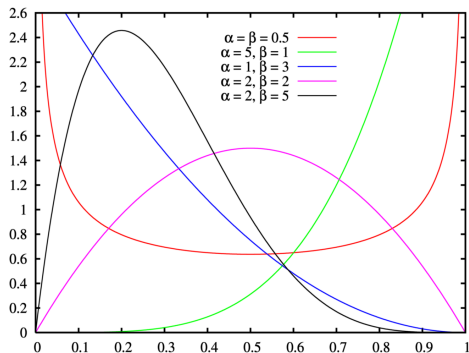
- ▶ probabilité :  $p(x|\alpha, \lambda) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$  pour tout  $x \geq 0$
- ▶ espérance :  $\alpha/\lambda$
- ▶ variance :  $\alpha/\lambda^2$
- ▶ sous R : `dgamma`



## loi beta $Beta(\alpha, \beta)$

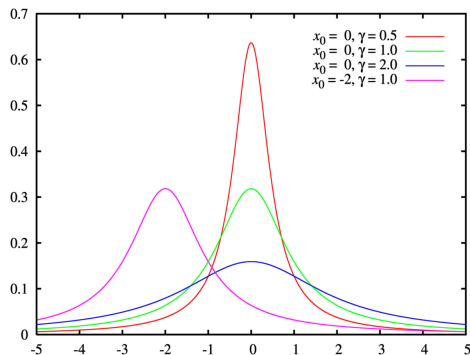
Loi très flexible sur  $]0, 1[$ , symétrique lorsque  $\alpha = \beta$ .

- ▶ probabilité :  $p(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$
- ▶ espérance :  $\alpha/(\alpha + \beta)$
- ▶ mode :  $(\alpha - 1)/(\alpha + \beta - 2)$  si  $\alpha, \beta > 1$
- ▶ variance :  $\alpha\beta/[(\alpha + \beta)^2(\alpha + \beta + 1)]$
- ▶  $Beta(1, 1)$  est la loi uniforme sur  $]0, 1[$
- ▶ sous R : `dbeta`



# loi de Cauchy centrée $\text{Cauchy}(0, a)$

- ▶ probabilité :  $p(x|a) = \frac{1}{\pi} \frac{a}{x^2 + a^2}$
- ▶ espérance, variance, ... : non définie
- ▶ sous R : dcauchy



Les lois a priori

# Différents types de loi a priori

Le choix de la loi a priori peut être fait :

- ▶ idéalement, en fonction du réel a priori dont on dispose
  - ▶ issu d'études précédentes,
  - ▶ issu d'une population statistique différente, ...  
(*mais attention, la loi a posteriori n'est pas toujours calculable analytiquement...*)
  - ▶ c'est le cadre typique du *transfer learning*
- ▶ de sorte à permettre le calcul analytique de la loi a posteriori : on parle de **loi a priori conjuguée**  
(*mais attention à l'influence des hyper-paramètres de la loi conjuguée...*)
- ▶ de sorte à atténuer l'effet de l'a priori : on parle de **loi a priori non informative**  
(*dans ce cas la loi a posteriori se rapproche de la vraisemblance...*)

## Loi a priori conjuguées

Quand on ne dispose pas d'information sur la forme de la loi a priori, on opte souvent pour le choix de la simplicité, c'est-à-dire le choix d'une **loi a priori conjuguée** qui facilite le calcul de la loi a posteriori :



# Loi a priori conjuguées

Exemple gaussien/gaussien :

- ▶  $\underline{x} = (x_1, \dots, x_n)$  avec  $x_i \sim \mathcal{N}(\theta, \sigma^2)$
- ▶  $\theta \sim \mathcal{N}(\mu, \tau^2)$
- ▶ alors

$$\theta|\underline{x} \sim \mathcal{N}\left(\frac{\tau^2 \bar{x} + \mu \frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}}, \frac{\tau^2 \frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}}\right)$$

où :

$$\begin{aligned} E[\theta|\underline{x}] &= \frac{\tau^2}{\tau^2 + \frac{\sigma^2}{n}} \bar{x} + \frac{\frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}} \mu = \frac{1}{\tau^2 + \frac{\sigma^2}{n}} \left( \tau^2 \bar{x} + \frac{\sigma^2}{n} \mu \right) \\ &= \frac{\tau^2 \frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}} \left( \frac{\bar{x}}{\frac{\sigma^2}{n}} + \frac{\mu}{\tau^2} \right) = \frac{\tau^2 \frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}} \left( \frac{\sum_i x_i}{\sigma^2} + \frac{\mu}{\tau^2} \right) \end{aligned}$$

# Loi a priori conjuguées

Exemple binomiale / Beta :

- ▶  $\underline{x} = (x_1, \dots, x_n)$  avec  $x_i \sim \mathcal{B}(n, \theta)$
- ▶  $\theta \sim \text{Beta}(\alpha, \beta)$
- ▶ alors

$$\theta | \underline{x} \sim \text{Beta}\left(\alpha + \underbrace{\sum_i x_i}_{\text{nb. succès}}, \beta + \underbrace{n - \sum_i x_i}_{\text{nb. echecs}}\right)$$

où

$$E[\theta | \underline{x}] = \frac{\alpha + \sum_i x_i}{\alpha + \beta + n}$$

et le mode (si  $\alpha > 1$  et  $\beta > 1$ ):

$$\frac{\alpha + \sum_i x_i - 1}{\alpha + \beta + n - 2}$$

# Loi a priori conjuguées

Exemple Poisson / Gamma :

- ▶  $\underline{x} = (x_1, \dots, x_n)$  avec  $x_i \sim \mathcal{P}(\lambda)$
- ▶  $\lambda \sim \mathcal{G}(\alpha, \beta)$
- ▶ alors

$$\lambda|\underline{x} \sim \mathcal{G}(\alpha + \sum_i x_i, \beta + n)$$

où

$$E[\lambda|\underline{x}] = \frac{\alpha + \sum_i x_i}{\beta + n}$$

# Loi a priori conjuguées

Exemple Exponentiel / Gamma :

- ▶  $\underline{x} = (x_1, \dots, x_n)$  avec  $x_i \sim \mathcal{E}(\lambda)$
- ▶  $\lambda \sim \mathcal{G}(\alpha, \beta)$
- ▶ alors

$$\lambda|\underline{x} \sim \mathcal{G}(\alpha + n, \beta + \sum_i x_i)$$

où

$$E[\lambda|\underline{x}] = \frac{\alpha + n}{\beta + \sum_i x_i}$$

# Loi a priori non informative

Une **loi non informative** est une loi qui porte une information sur le paramètre à estimer dont le poids dans l'inférence est réduit.

- ▶ par exemple, si  $\theta$  peut prendre  $q$  valeurs, alors on choisit  $p(\theta) = 1/q$  pour tout  $\theta \in \Theta$
- ▶ loi de **Jeffreys** :
  - ▶ c'est une méthode pour construire des loi donnant un poids plus important aux valeurs  $\theta$  où l'information de Fisher  $I(\theta)$  (quantité d'info. sur  $\theta$  apporté par  $\underline{x}$ ) est grande :  $p(\theta) = C\sqrt{I(\theta)}$
  - ▶ les lois obtenus sont invariantes par transformation (elles restent non informatives pour une transformation du paramètre)
  - ▶ cette méthode conduit souvent à des lois *impropres* ( $\int_{\Theta} p(\theta)d\theta = +\infty$ ), qui n'a donc plus d'interprétation et n'est utile que d'un point de vue calculatoire.

## Exemple de calcul avec une loi a priori conjuguée

- ▶ on s'intéresse au poids de naissance des enfants en Allemagne  
 $x \sim \mathcal{N}(\theta, \sigma^2)$
- ▶ on prélève un échantillon  $\underline{x} = (x_1, \dots, x_n)$  de taille  $n = 20$  et on calcule  $\bar{x} = 3.4$
- ▶ le *fréquentiste* estimerait  $\hat{\theta} = 3.4$

## Exemple de calcul avec une loi a priori conjuguée

- ▶ on s'intéresse au poids de naissance des enfants en Allemagne  
 $x \sim \mathcal{N}(\theta, \sigma^2)$
- ▶ on prélève un échantillon  $\underline{x} = (x_1, \dots, x_n)$  de taille  $n = 20$  et on calcule  $\bar{x} = 3.4$
- ▶ le *fréquentiste* estimerait  $\hat{\theta} = 3.4$
- ▶ par analogie avec ce que l'on sait sur la France, le *bayésien* pourrait supposer un a priori  $\theta \sim \mathcal{N}(3.2, 0.25)$
- ▶ supposons de plus que  $\sigma^2$  est :  $\sigma^2 = 1$  (en pratique il faudrait l'estimer...)

Quelle est alors l'espérance a posteriori de  $\theta$  sachant  $\underline{x}$  ?

## Exemple de calcul avec une loi a priori conjuguée

- ▶ on s'intéresse au poids de naissance des enfants en Allemagne  
 $x \sim \mathcal{N}(\theta, \sigma^2)$
- ▶ on prélève un échantillon  $\underline{x} = (x_1, \dots, x_n)$  de taille  $n = 20$  et on calcule  $\bar{x} = 3.4$
- ▶ le *fréquentiste* estimerait  $\hat{\theta} = 3.4$
- ▶ par analogie avec ce que l'on sait sur la France, le *bayésien* pourrait supposer un a priori  $\theta \sim \mathcal{N}(3.2, 0.25)$
- ▶ supposons de plus que  $\sigma^2$  est :  $\sigma^2 = 1$  (en pratique il faudrait l'estimer...)

Quelle est alors l'espérance a posteriori de  $\theta$  sachant  $\underline{x}$  ?

$$\begin{aligned}\hat{\theta} = E[\theta|\underline{x}] &= \frac{\tau^2 \frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}} \left( \frac{\bar{x}}{\frac{\sigma^2}{n}} + \frac{\mu}{\tau^2} \right) \\ &= \frac{0.25 \frac{1}{20}}{0.25 + \frac{1}{20}} \left( \frac{3.4}{\frac{1}{20}} + \frac{3.2}{0.25} \right) \\ &= 3.36667\end{aligned}$$



## Exemple 2 - pile ou face

On veut connaître la probabilité  $\theta$  de tomber sur face pour une pièce donnée. Nous considérons plusieurs a priori :

1. loi  $\beta(1, 1)$  : toutes les valeurs de  $\theta$  sont équiprobables (a priori non informatif) :  $p(\theta) \propto 1, \forall \theta$
2. loi  $\beta(2, 2)$ ,  $\beta(4, 4)$ ,  $\beta(8, 8)$

Nous observons l'échantillon suivant :  $\underline{x} = \{1, 1, 1, 0, 1, 1, 0, 0, 1, 1\}$ , soit  $y = 7$  "face" et 3 "pile".

### Questions

1. Quelle est la vraisemblance de  $\underline{x}$  ?
2. Pour chaque loi a priori :
  - ▶ calculer et représenter la loi a posteriori de  $\theta$ ,
  - ▶ estimer  $\theta$  ponctuellement,
  - ▶ donner un intervalle de crédibilité de  $\theta$ .

## Exemple 2 - pile ou face

- vraisemblance : 7 “face” de proba  $\theta$  et 3 pile de proba  $1 - \theta$

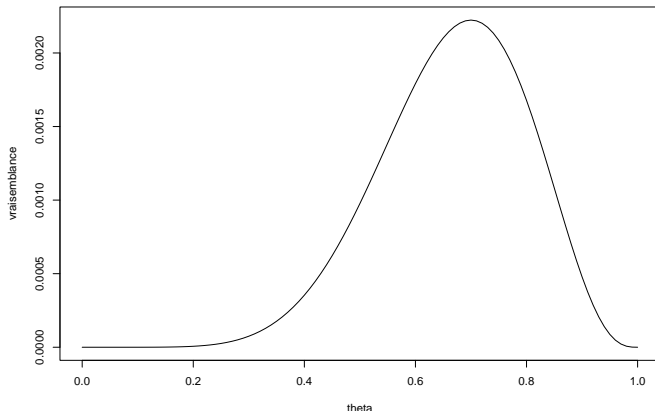
$$p(\underline{x}|\theta) = \theta^7(1 - \theta)^3$$

- la loi bêta( $a, b$ ) étant conjuguée pour la loi binomiale, on peut calculer la loi a posteriori :
  - loi a priori :  $p(\theta) = \text{bêta}(a, b) \propto \theta^{a-1}(1 - \theta)^{b-1}$
  - loi a posteriori :
$$p(\theta|\underline{x}) = \text{bêta}(a + 7, b + 3) \propto \theta^{a+7-1}(1 - \theta)^{b+3-1}$$

## Exemple 2 - pile ou face

Calcul de la vraisemblance (en fonction de  $\theta$ )

```
theta=seq(0,1,0.01)  
vraisemblance=theta^7*(1-theta)^3  
plot(theta,vraisemblance,type='l')
```



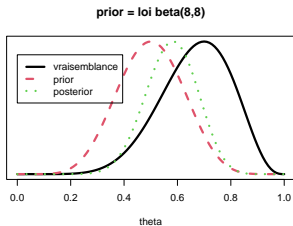
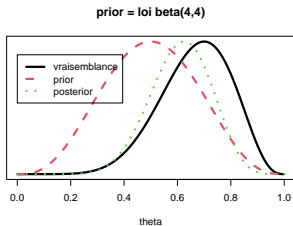
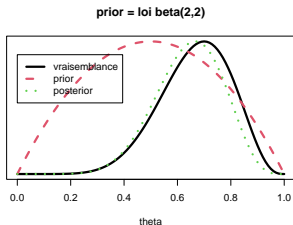
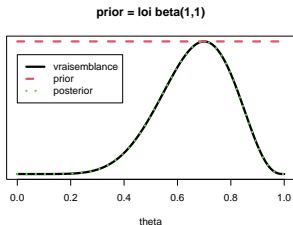
## Exemple 2 - pile ou face

On va représenter pour différentes valeurs des paramètres de la loi a priori :

- ▶ la vraisemblance,
- ▶ la loi a priori,
- ▶ la loi a posteriori (calculable ici car on a choisi un a priori conjugué)

```
par(mfrow=c(2,2))
for (i in c(1,2,4,8)){
  a=i;b=i;
  plot(theta,vraisemblance/max(vraisemblance),type='l',ylab="",yaxt="n",lwd=3)
  title(paste("prior = loi beta(",a,",",b,")",sep=""))
  prior=dbeta(theta,a,b)
  lines(theta,prior/max(prior),type='l',col=2,lty=2,lwd=3,ylab="",yaxt="n")
  posterior=dbeta(theta,a+7,b+3)
  lines(theta,posterior/max(posterior),type='l',col=3,lty=3,lwd=3,ylab="",
        yaxt="n")
  legend(0,.9,legend=c("vraisemblance", "prior","posterior"),col=1:3,lty=1:3,
        cex=1,lwd=3)
}
```

## Exemple 2 - pile ou face



Interprétation de l'influence de la loi a priori : plus sa variance est faible, plus il “tire à lui” la loi a posteriori.

## Exemple 2 - pile ou face

Sachant que la loi a posteriori est une loi Beta, on peut estimer  $\theta$  par le mode de la distribution a posteriori (Maximum A Posteriori, MAP).

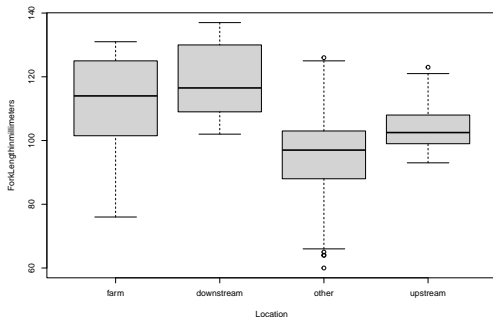
On peut aussi calculer facilement un intervalle de crédibilité à l'aide des quantiles de la loi Beta.

```
for (i in c(1,2,4,8)){  
  a=i;b=i;  
  cat('Estimation MAP :',(a+6)/(a+b+8),'\n')  
  cat('Intervalle de crédibilité : [',qbeta(0.025,a+7,b+3),',',  
      qbeta(0.975,a+7,b+3),']','\n')  
}
```

```
## Estimation MAP : 0.7  
## Intervalle de crédibilité : [ 0.3902574 , 0.8907366 ]  
## Estimation MAP : 0.6666667  
## Intervalle de crédibilité : [ 0.3857383 , 0.8614207 ]  
## Estimation MAP : 0.625  
## Intervalle de crédibilité : [ 0.3832837 , 0.815563 ]  
## Estimation MAP : 0.5833333  
## Intervalle de crédibilité : [ 0.3866535 , 0.7559763 ]
```

## Exercice : impact de la présence d'une pisciculture

- ▶ Les données `ForklengthData.csv` contiennent les longueurs de saumons juvénils le long d'une rivière en Bretagne (source [http://sirs.agrocampus-ouest.fr/bayes\\_V2/index.html](http://sirs.agrocampus-ouest.fr/bayes_V2/index.html)).
- ▶ On se pose la question de l'impact de la présence d'un ferme d'élevage installée sur cette rivière.
- ▶ Pour répondre à cette question, comparer la taille des poissons en amont et en aval de la ferme. Y-a-t'il une différence ?
- ▶ On utilisera une approche fréquentiste, puis une approche bayésienne en comparant les distributions a posteriori



# Exercice poids de naissance

```
data=read.table("Rcode/poidsnaissance.txt",header = T,sep=',',row.names=1)
head(data)
```

##	OBS	AGEGEST	SEXE	POIDNAIS	CIGJOUR	TAILMERE	POIDAVG	POIDFING
## 1	1	36	0	3300	0	160.0	67.3	82.7
## 2	2	38	0	3300	60	167.6	52.7	76.0
## 3	3	38	0	4100	20	167.6	64.2	79.6
## 4	4	38	1	2900	10	163.9	72.7	95.8
## 5	5	39	0	2820	0	161.3	50.0	63.3
## 6	6	39	0	3040	0	158.8	49.1	61.5

On s'intéresse au poids de naissance des enfants.

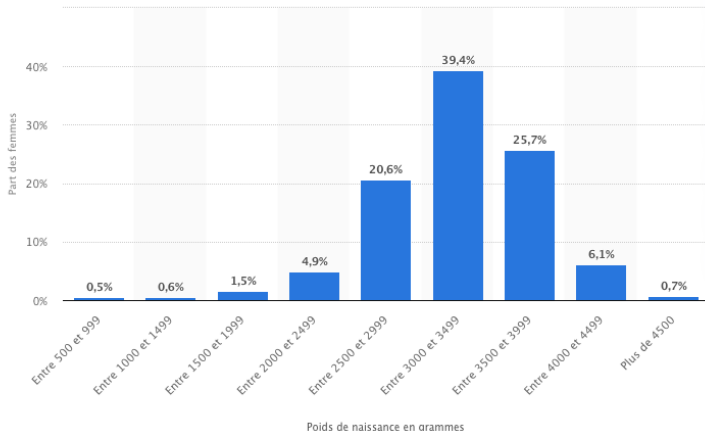
Les questions que l'on peut se poser :

1. quel le poids de naissance moyen ?
2. les garçons sont-ils plus lourds que les filles à la naissance ?
3. le poids de naissance dépend-il de la prise du poids de la mère ? Et du père ?
4. le poids de naissance dépend-il de l'âge gestationnel ?
5. le poids de naissance dépend-il du nombre de cigarettes fumées par jour ?
6. y-a-t'il un impact du nombre de cigarettes fumées par jour sur l'âge gestationnel et donc sur le poids de naissance ?
7. ...



## Exercice poids de naissance

Nous répondrons ici à la première question (estimer le poids de naissance moyen), de façon fréquentiste et de façon bayésienne en utilisant l'information suivante :



## Méthodes de Monte-Carlo

# Méthodes de Monte Carlo

Le calcul de la loi a posteriori fait intervenir le calcul de :

$$\int_{\Theta} \ell(\underline{x}|\theta) p(\theta) d\theta$$

qui est souvent non calculable analytiquement (en dehors du cas des lois conjuguées).

## **Méthode de Monte-Carlo** (simple)

Pour estimer l'intégrale

$$\mathcal{I} = \int h(\theta) p(\theta) d\theta$$

- ▶ simuler  $\theta_1, \dots, \theta_n \sim p(\theta)$
- ▶ approcher  $\mathcal{I}$  par

$$\hat{\mathcal{I}}_n = \frac{1}{n} \sum_{i=1}^n h(\theta_i)$$

# Monte Carlo en pratique

On cherche à approcher l'intégrale  $I = \int_0^1 \frac{1}{\sqrt{2*\pi}} e^{-\frac{x^2}{2}} dx$ .

Pour cela nous utilisons une méthode de Monte-Carlo :

- ▶ simuler un échantillon de taille  $n = 100$  de variables aléatoires  $x_i \sim \mathcal{U}[0, 1]$
- ▶ calculer la moyenne des  $\frac{1}{\sqrt{2*\pi}} e^{-\frac{x_i^2}{2}}$ , pour  $1 \leq i \leq n$
- ▶ répéter cela 100 fois et représenter la dispersion des résultats obtenus à l'aide d'un boxplot
- ▶ répéter cela avec  $n = 100, 10^4, 10^6$ .

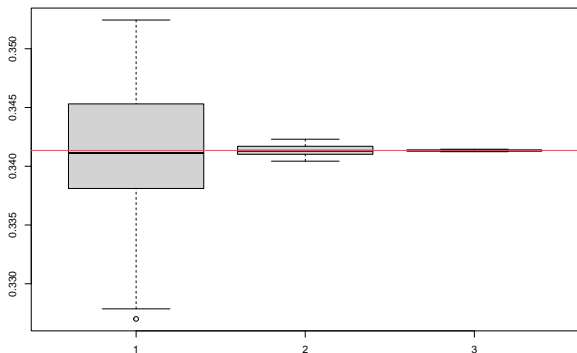
Pour information, cette intégrale vaut :

```
pnorm(1)-pnorm(0)
```

```
## [1] 0.3413447
```

# Monte Carlo en pratique

```
M=matrix(0,100,3)
for (i in 1:100){
  M[i,1]=mean(exp(-runif(100,0,1)^2/2)/sqrt(2*pi))
  M[i,2]=mean(exp(-runif(10^4,0,1)^2/2)/sqrt(2*pi))
  M[i,3]=mean(exp(-runif(10^6,0,1)^2/2)/sqrt(2*pi))
}
boxplot(M)
abline(h=(pnorm(1)-pnorm(0)),col=2)
```



## Simulation suivant $p(\theta)$

On veut simuler  $x \sim p(\theta)$ .

Soient :

- ▶  $f_\theta(\cdot)$  la densité de probabilité de la loi  $p(\theta)$
- ▶  $F_\theta(t) = \int_{-\infty}^t p(x; \theta) dx$  la fonction de répartition

Une façon classique de simuler  $x \sim p(\theta)$  est la suivante :

- ▶ simuler  $u \sim \mathcal{U}[0, 1]$
- ▶ calculer  $x = F_\theta^{-1}(u)$

Ce qui requiert le calcul de  $F_\theta^{-1}$  (et de  $F_\theta$ )...

# Méthodes de Monte Carlo

Lorsque l'on ne peut pas calculer facilement  $F_\theta$  et/ou  $F_\theta^{-1}$ , de nombreuses variantes existent :

- ▶ **Rejection Sampling**
- ▶ **Importance Sampling**
- ▶ **Monte Carlo par Chaîne de Markov (MCMC)**

une référence intéressante pour aller plus loin :

[https://www.mit.edu/~9.520/spring10/Classes/class21\\_mcmc\\_2010.pdf](https://www.mit.edu/~9.520/spring10/Classes/class21_mcmc_2010.pdf)

# Monte Carlo par Chaîne de Markov

- ▶ particulièrement utile lorsque  $\theta$  est de grande dimension
- ▶ On génère une suite de  $x_{i+1}|x_i$ , dont la génération est supposée simple, et dont la loi stationnaire (loi dont proviendront les  $x_i$  au bout d'un certain temps) est  $p(\theta)$
- ▶ deux algorithmes célèbres :
  - ▶ Metropolis-Hastings
  - ▶ Gibbs



# MCMC Metropolis-Hastings

- ▶ génère  $x_{i+1}|x_i$  suivant une loi *proposal*  $q(x_{i+1}|x_i)$ , par ex.  $\mathcal{N}(x_i, \sigma^2)$
- ▶ accepte  $x_{i+1}$  avec probabilité :

$$\min \left( 1, \frac{p_\theta(x_{i+1})q(x_i|x_{i+1})}{p_\theta(x_i)q(x_{i+1}|x_i)} \right)$$

- ▶ au bout d'un *certain temps* les  $x_i$  vont provenir de  $p(\theta) = p_\theta$

Remarques :

- ▶ cela fonctionne même si  $p(\theta)$  n'est connue qu'à une constante près (cas des posteriors. . .)
- ▶ le paramètre  $\sigma^2$  est à régler :
  - ▶ trop petit les  $x_i$  bougent trop peu et la chaîne converge trop lentement,
  - ▶ trop grand on accepte trop peu souvent et la chaîne converge trop lentement,

# MCMC Gibbs

- ▶  $x_i = (x_i^1, \dots, x_i^p)$  est en dimension  $p$
- ▶ génère  $x_{i+1}|x_i$  dimension par dimension :
  - ▶ initialise  $z_1 = x_i$
  - ▶ pour  $t$  de 1 à  $T$  :
    - ▶  $z_{t+1}^1 | z_t^2, \dots, z_t^p$
    - ▶  $z_{t+1}^2 | z_{t+1}^1, \dots, z_t^p$
    - ▶ ...
    - ▶  $z_{t+1}^p | z_{t+1}^1, \dots, z_{t+1}^{p-1}$
  - ▶ au bout d'un certain nombre d'itérations, le  $z_T = (z_T^1, \dots, z_T^p)$  obtenu provient de  $x_{i+1}|x_i$  : on peut donc choisir  $x_{i+1} = z_T$

## Remarques :

- ▶ très utile quand on ne sait pas calculer la loi jointe du vecteur  $(x_i^1, \dots, x_i^p)$
- ▶  $T$  n'a pas forcément à être très grand en pratique, car on répète ces  $T$  itérations pour chaque simulation de  $x_{i+1}|x_i$

## Exercice - Monte Carlo

On s'intéresse au nombre  $Y$  de face obtenus sur 10 lancers d'une pièce.

Considérons le modèle suivant :

$$Y|\pi \sim \mathcal{B}(10, \pi)$$

$$\pi \sim \text{Beta}(2, 2)$$

- ▶ quelle hypothèse est faite par cette loi a priori ?
- ▶ calculer la loi a posteriori théoriquement
- ▶ approcher cette loi a posteriori par Monte-Carlo simple
- ▶ superposer la loi théorique et la loi approchée. Tester différents nombre de simulations.

## Tests bayésiens

# Tests bayésiens

Question : est-ce que  $\mu = 0$  pour les normaldata ?

**Approche fréquentiste** On met en place un test de Student pour décider entre

- ▶  $H_0 : \mu = 0$
- ▶  $H_1 : \mu \neq 0$

**Approche bayésienne** On compare les deux modèles

- ▶  $M_1 : x \sim \mathcal{N}(0, \sigma^2)$
- ▶  $M_2 : x \sim \mathcal{N}(\mu, \sigma^2)$

# Facteur de Bayes

## Choix entre deux modèles

On veut comparer  $p(M_1|\underline{x})$  et  $p(M_2|\underline{x})$ .

Pour cela on introduit le **Bayes factor** :

$$B_{21}(\underline{x}) = \frac{p(\underline{x}|M_2)}{p(\underline{x}|M_1)} = \frac{p(M_2|\underline{x})/p(M_1|\underline{x})}{p(M_2)/p(M_1)}$$

qui correspond au odds ratio ou rapport de vraisemblance classique<sup>1</sup>, sauf qu'ici le paramètre est intégré et non maximisé sous chaque modèle

---

<sup>1</sup>déviance  $D = -2 \ln(p(\underline{x}|M_1, \hat{\theta})/p(\underline{x}|M_2, \hat{\theta}))$

# Facteur de bayes

## Echelle d'évidence de Jeffrey's

Bayes factor :

$$B_{21}(\underline{x}) = \frac{p(\underline{x}|M_2)}{p(\underline{x}|M_1)} = \frac{p(M_2|\underline{x})/p(M_1|\underline{x})}{p(M_2)/p(M_1)}$$

- ▶ si  $\log_{10}(B_{21}(\underline{x})) \in [0, 0.5]$ , la certitude que  $M_1$  est fausse est *faible*
- ▶ si  $\log_{10}(B_{21}(\underline{x})) \in [0.5, 1]$ , la certitude que  $M_1$  est fausse est *substantielle*
- ▶ si  $\log_{10}(B_{21}(\underline{x})) \in [1, 2]$ , la certitude que  $M_1$  est fausse est *forte*
- ▶ si  $\log_{10}(B_{21}(\underline{x})) > 2$ , la certitude que  $M_1$  est fausse est *décisive*

et idem en faveur de  $M_1$  pour les valeurs négatives.

Rq :  $\log_{10}(B_{21}(\underline{x})) > 1 \Leftrightarrow B_{21} > 10$ ,  $\log_{10}(B_{21}(\underline{x})) > 2 \Leftrightarrow B_{21} > 100, \dots$

# Facteur de bayes

Liens avec les tests classiques :

- ▶ globalement :
  - ▶  $\log_{10}(B_{21}(\underline{x})) > .5 \Leftrightarrow B_{21} > 1$  équivalent à une p-value  $< 0.1$
  - ▶  $\log_{10}(B_{21}(\underline{x})) > 1 \Leftrightarrow B_{21} > 10$  équivalent à une p-value  $< 0.05$
  - ▶  $\log_{10}(B_{21}(\underline{x})) > 2 \Leftrightarrow B_{21} > 100$  équivalent à une p-value  $< 0.01$
- ▶ l'approche bayésienne est symétrique entre  $M_1$  et  $M_2$ , alors qu'un test classique ne l'est pas entre  $H_0$  et  $H_1$  (risques  $\alpha \neq \beta$ )



# Facteur de bayes

Calcul du facteur de bayes :

$$B_{21}(\underline{x}) = \frac{p(M_2|\underline{x})/p(M_1|\underline{x})}{p(M_2)/p(M_1)} = \frac{\int_{\Theta_2} \ell_2(\theta_2|\underline{x})p_2(\theta_2)d\theta_2}{\int_{\Theta_1} \ell_1(\theta_1|\underline{x})p_1(\theta_1)d\theta_1}$$

qui peut parfois se calculer analytiquement.

Quand ce n'est pas le cas, on utilise des approximations numériques  
(*Méthodes de Monte Carlo*)

# Critère BIC

## BIC - Bayesian Information Criterion

Le critère **BIC** est une approximation de la vraisemblance intégrée :

$$BIC_{M_1} = -2 \ln \ell(\underline{x}|\hat{\theta}) + k \ln n \simeq -2 \ln \int_{\Theta} \ell(\underline{x}|\theta) p(\theta) d\theta$$

qui a l'avantage de se calculer très facilement dès lors qu'on dispose du maximum de vraisemblance  $\hat{\theta}$ .

Ce critère est largement utilisé, que l'on soit dans une approche bayésienne ou non.

## Critères AIC et DIC

D'autres critères sont proches du critères BIC, avec le même objectif mais des fondations théoriques différentes.

Le critère **AIC** est un critère *fréquentiste* de choix de modèle, qui approxime la **déviante** du modèle (écart de Kullback-Leibler entre le *vrai* modèle et le modèle estimé)

$$AIC_{M_1} = -2 \ln \ell(\underline{x}|\hat{\theta}) + 2k$$

Le critère **DIC - Deviance Information Criterion** est une généralisation du critère AIC pour les modèles hiérarchiques. Nous l'utiliserons plus tard.

# Test bayésiens sous R

Pour comparer deux populations, on utilisera :

- ▶ fréquentiste : `t.test` ou `wilcox.test`
- ▶ bayésien : `ttestBF`

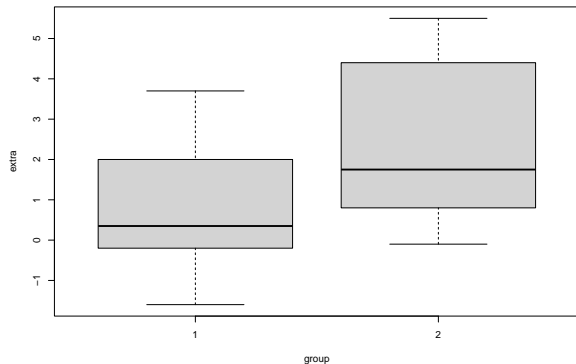
Pour comparer plus de deux populations, on utilisera :

- ▶ fréquentiste : `aov` ou `kruskal.test`
- ▶ bayésien : `anovaBF`

## Exemple 3 - données sleep

Les données sleep contiennent l'effet de 2 somnifères (group) sur la durée de sommeil, pour 10 patients.

```
boxplot(extra~group,data=sleep)
```



Tester la différence entre les groupes avec une approche classique

## Exemple 3 - données sleep

Nous allons maintenant évaluer la différence entre les groupes avec une approche bayésienne :

- ▶  $\delta = \frac{\mu_1 - \mu_2}{\sigma}$
- ▶ on compare les modèles
  - ▶  $M_1 : (\mu_1 - \mu_2)/\sigma \sim \mathcal{N}(0, 1)$
  - ▶  $M_2 : (\mu_1 - \mu_2)/\sigma \sim \mathcal{N}(\delta, 1)$  avec  $\delta \sim \text{Cauchy}(rscale)$

```
library(BayesFactor)
ttestBF(x=sleep$extra[sleep$group==1] -
        sleep$extra[sleep$group==2], rscale=sqrt(2)/2)
```

```
## Bayes factor analysis
## -----
## [1] Alt., r=0.707 : 17.25888 ±0%
##
## Against denominator:
##   Null, mu = 0
## ---
## Bayes factor type: BFoneSample, JZS
```

## Exemple 3 - données sleep

En calculant le  $\log_{10}$  de BayesFactor, on conclut à une forte certitude que le modèle M1 (supposant l'absence de différence) est faux. Le résultat est le même qu'avec l'approche inférentielle classique.

```
log10(17.25888)
```

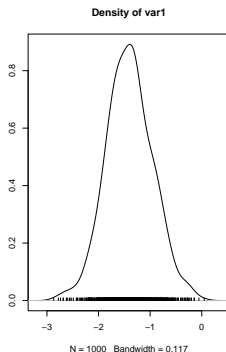
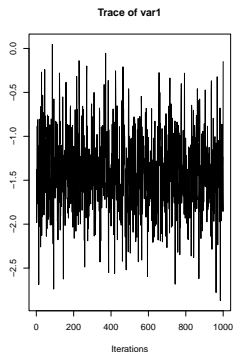
```
## [1] 1.237013
```

## Exemple 3 - données sleep

On peut aussi simuler l'esperance de la différence suivant la loi a posteriori

```
samples = ttestBF(x = sleep$extra[sleep$group==1]
                  -sleep$extra[sleep$group==2],
                  rscale=sqrt(2)/2,posterior = TRUE,
                  iterations = 1000)

plot(samples[, "mu"])
```

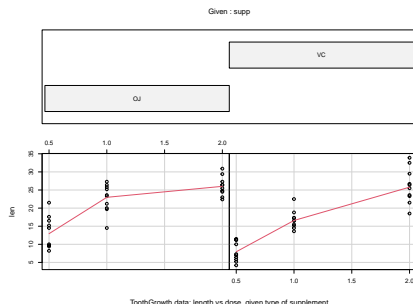




## Exercice 2 - données ToothGrowth

ToothGrowth contient la longueur des dents de 60 cochons d'Inde en fonction de deux sources de vitamine C (*jus d'orange et acide ascorbique*)

```
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```



Tester les effets sources et doses avec :

- ▶ une approche classique
- ▶ une approche bayésienne

# Exercice poids de naissance

```
data=read.table("Rcode/poidsnaissance.txt",header = T,sep=',',row.names=1)
head(data)
```

##	OBS	AGEGEST	SEXE	POIDNAIS	CIGJOUR	TAILMERE	POIDAVG	POIDFING
## 1	1	36	0	3300	0	160.0	67.3	82.7
## 2	2	38	0	3300	60	167.6	52.7	76.0
## 3	3	38	0	4100	20	167.6	64.2	79.6
## 4	4	38	1	2900	10	163.9	72.7	95.8
## 5	5	39	0	2820	0	161.3	50.0	63.3
## 6	6	39	0	3040	0	158.8	49.1	61.5

Répondre de façon fréquentiste et bayésienne :

1. les garçons sont-ils plus lourds que les filles à la naissance ?
2. le poids de naissance dépend-il du fait que la mère soit fumeuse ?
3. l'âge gestationnel dépend-il du fait que la mère soit fumeuse ?

## Régression bayésienne

## Régression linéaire *fréquentiste*

Le modèle linéaire ordinaire est

$$y|\alpha, \beta, \sigma^2 \sim \mathcal{N}(\alpha \mathbf{1}_n + \mathbf{x}\beta, \sigma^2 \mathbf{I}_n)$$

avec

$$\begin{aligned} E[y_i|\alpha, \beta, \sigma^2] &= \alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \\ V(y_i|\alpha, \beta, \sigma^2) &= \sigma^2 \end{aligned}$$

L'estimation par maximum de vraisemblance (débiaisée pour  $\hat{\sigma}^2$ ) donne

$$\begin{aligned} \hat{\alpha} &= \bar{y} \\ \hat{\beta} &= (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'(\mathbf{y} - \bar{y}) \\ \hat{\sigma}^2 &= \frac{(\mathbf{y} - \alpha \mathbf{1}_n - \mathbf{x}\beta)'(\mathbf{y} - \alpha \mathbf{1}_n - \mathbf{x}\beta)}{n - p - 1} = \frac{s^2}{n - p - 1} \end{aligned}$$

# Régression bayésienne

Avec **a priori non informatif de Jeffreys**

$$\begin{aligned}y|\alpha, \beta, \sigma^2 &\sim \mathcal{N}(\alpha \mathbf{1}_n + \mathbf{X}\beta, \sigma^2 \mathbf{I}_n) \\(\alpha, \beta, \sigma^2) &\propto \sigma^{-2}\end{aligned}$$

Les estimateurs bayésiens (espérance  $E^\pi$  de la loi a posteriori) sont alors :

$$\begin{aligned}E^\pi[\alpha|y] &= \hat{\alpha} \\E^\pi[\beta|y] &= \hat{\beta} \\E^\pi[\sigma^2|y] &= s^2/(n - p - 3)\end{aligned}$$

Notons que les estimations de  $\alpha$  et  $\beta$  sont les mêmes que dans le cas ordinaires, et que l'estimation de  $\sigma^2$  est un peu plus grande (plus pessimiste).

# Régression bayésienne

## Avec **avec a priori informatif de Zellner**

- ▶ idée de Zellner : proposer un a priori conjugué mais en réduisant le nombre d'hyper-paramètres
- ▶  $\Rightarrow$  a priori gaussien sur  $\beta$  dont on doit fixer l'espérance  $\tilde{\beta}$  et le facteur d'importance  $g$   
( $g = n \Leftrightarrow$  a priori a autant d'importance qu'une observation,  
 $g = 1 \Leftrightarrow$  et il a autant d'importance que toutes les données ensemble)
- ▶  $\Rightarrow$  a priori non informatif de Jeffreys (impropre) sur  $(\alpha, \sigma^2)$

$$\begin{aligned}y|\alpha, \beta, \sigma^2 &\sim \mathcal{N}(\alpha \mathbf{1}_n + \mathbf{x}\beta, \sigma^2 \mathbf{I}_n) \\ \beta|\alpha, \sigma^2 &\sim \mathcal{N}(\tilde{\beta}, g\sigma^2(\mathbf{x}'\mathbf{x})^{-1}) \\ (\alpha, \sigma^2) &\propto \sigma^{-2}\end{aligned}$$

L'estimateur bayésien pour  $\beta$  est alors  $E^\pi[\beta|y] = \frac{g}{g+1}(\hat{\beta} + \tilde{\beta}/g)$

# Régression bayésienne sous R : BayesReg

La fonction BayesReg (package bayess) implémente une régression bayésienne avec a priori de Zellner

```
data("caterpillar")  
y=log(caterpillar$y)  
x=as.matrix(caterpillar[,1:8])
```

on normalise les données pour rendre les valeurs des coefficients de régression comparable

```
x=scale(x)
```

## Régression bayésienne sous R : BayesReg

On réalise une de régression *classique*, où l'on voit que seules les variables 1, 2 et 7 semblent significatives

```
summary(lm(y~x))
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4710 -0.4474 -0.1769  0.6121  1.5602
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.81328    0.15356  -5.296 1.97e-05 ***
## xx1         -0.52722    0.21186  -2.489  0.0202 *
## xx2         -0.39286    0.16974  -2.315  0.0295 *
## xx3          0.65133    0.38670   1.684  0.1051
## xx4         -0.29048    0.31551  -0.921  0.3664
## xx5         -0.21645    0.16865  -1.283  0.2116
## xx6          0.29361    0.53562   0.548  0.5886
## xx7         -1.09027    0.47020  -2.319  0.0292 *
## xx8         -0.02312    0.17225  -0.134  0.8944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```



# Régression bayésienne sous R : BayesReg

Effectuons une regression Bayésienne avec un a priori de Zellner nul et une faible confiance dans cet a priori ( $g = n$ )

```
res1=BayesReg(y,x,betatilde = c(0,0,0,0,0,0,0,0),g=length(y))
```

```
##
##          PostMean PostStError Log10bf EvidAgaH0
## Intercept  -0.8133      0.1407
## x1         -0.5039      0.1883  0.7224      (**)
## x2         -0.3755      0.1508  0.5392      (**)
## x3          0.6225      0.3436 -0.0443
## x4         -0.2776      0.2804 -0.5422
## x5         -0.2069      0.1499 -0.3378
## x6          0.2806      0.4760 -0.6857
## x7         -1.0420      0.4178  0.5435      (**)
## x8         -0.0221      0.1531 -0.7609
##
##
## Posterior Mean of Sigma2: 0.6528
## Posterior StError of Sigma2: 0.939
```

Même résultat que pour l'approche classique.

## Régression bayésienne sous R : BayesReg

On introduit un cette fois un a priori sur le fait que le coefficient de  $x_3$  est égal à 1, et ce avec une confiance forte ( $g = 1$ ). Pour ne pas avoir d'impact sur les autres variables, on fixe leur effet à l'aide de la régression moindres carrés

```
res1=BayesReg(y,x,betatilde = c(-.52,-.39,1,-.29,-.21,.29,-1.09,-0.02),g=1)
```

```
##
##           PostMean PostStError Log10bf EvidAgaH0
## Intercept   -0.8133      0.1443
## x1          -0.5196      0.1386  1.3091      (***)
## x2          -0.3884      0.1111  1.129      (***)
## x3           0.8207      0.2530  0.4505      (*)
## x4          -0.2880      0.2065  0.0683      (*)
## x5          -0.2116      0.1104  0.2686      (*)
## x6           0.2896      0.3505 -0.0722
## x7          -1.0818      0.3077  1.1332      (***)
## x8          -0.0214      0.1127 -0.1458
##
##
## Posterior Mean of Sigma2: 0.6872
## Posterior StError of Sigma2: 0.9885
```

Sans surprise,  $x_3$  devient alors significative, tout comme d'autres variables (car on a insisté sur leur effet constaté dans les données)

## Exercice données longley

- ▶ charger les données longley, les étudier puis réaliser une régression ordinaire
- ▶ comparer à une régression bayésienne en jouant sur les hyper-paramètres de l'a priori de Zellner

## Exercice données IFPEN

- ▶ charger les données source et établir un modèle de régression
- ▶ charger les données target. Tirer aléatoirement un échantillon de petite taille, 12 observations, et comparer en qualité de prédiction sur le reste des données :
  - ▶ le modèle de régression classique estimé sur ces 12 observations
  - ▶ un modèle de régression bayésienne avec comme a priori les coefficients du modèle source

# Exercice poids de naissance

```
data=read.table("Rcode/poidsnaissance.txt",header = T,sep=',',row.names=1)
head(data)
```

##	OBS	AGEGEST	SEXE	POIDNAIS	CIGJOUR	TAILMERE	POIDAVG	POIDFING
## 1	1	36	0	3300	0	160.0	67.3	82.7
## 2	2	38	0	3300	60	167.6	52.7	76.0
## 3	3	38	0	4100	20	167.6	64.2	79.6
## 4	4	38	1	2900	10	163.9	72.7	95.8
## 5	5	39	0	2820	0	161.3	50.0	63.3
## 6	6	39	0	3040	0	158.8	49.1	61.5

Effectuer une régression du poids de naissance en fonction des autres variables disponibles :

- ▶ de façon fréquentiste
- ▶ de façon bayésienne en cherchant à traduire et introduire votre propre a priori

# Choix de modèle en régression bayésienne

- ▶ avec  $p$  variables, le nombre de modèles possibles est  $2^p$
- ▶ il est impossible de comparer tous les modèles dans une approche fréquentiste classique
- ▶ en régression bayésienne, **le choix du modèle** (i.e. des variables à inclure dans le modèle) **est considéré comme un paramètre**, avec un a priori généralement uniforme (ou plus généralement *beta-binomial*).
- ▶ la fonction ModChoBayesReg implémente ce modèle de régression bayésienne intégrant le choix de modèle.
  - ▶ si  $p < 15$  les probabilités a posteriori de tous les modèles sont calculées
  - ▶ si  $p > 15$ , une approximation MCMC est utilisée

## Régression bayésienne sous R : ModChoBayesReg

La fonction ModChoBayesReg implémente un modèle de régression bayésienne avec a priori de Zellner intégrant le choix de modèle.

```
data("caterpillar")
y=log(caterpillar$y)
x=as.matrix(caterpillar[,1:8])
x=scale(x)
ModChoBayesReg(y,x)
```

```
##
## Number of variables less than 15
## Model posterior probabilities are calculated exactly
##
##      Top10Models PostProb
## 1          1 2 7    0.0767
## 2           1 7    0.0689
## 3          1 2 3 7    0.0686
## 4           1 3 7    0.0376
## 5           1 2 6    0.0369
## 6          1 2 3 5 7    0.0326
## 7           1 2 5 7    0.0294
```

## Exercice - données longley

- choisissez le meilleur modèle de régression bayésienne, en jouant sur le paramètre  $g$  de Zellner



# Bayesian Model Averaging

- ▶ choisir un unique modèle  $\mathcal{M}_m$  parmi  $2^p$  modèles possibles peut-être restrictif, sachant que les différences entre leur probabilité a posteriori sont parfois faibles
- ▶ le **Bayesian Model Averaging** préconise de moyenner les modèles en fonction de leur probabilité a posteriori. Ainsi :

$$E^\pi[\beta|y] = \sum_{m=1}^M E^\pi[\beta|y, \mathcal{M}_m] p(\mathcal{M}_m|y)$$

où  $M < 2^p$  est le nombre de modèles conservés (*on ne conserve que les modèles dont la probabilité a posteriori est supérieure à  $1/OR$  le maximum des probabilités a posteriori*).

- ▶ cette technique est utilisable dès lors que l'on fait de la prédiction (en classification également donc), et quelque soit le type de modèle que l'on utilise (réseaux de neurones, arbres de régression ...)

# Bayesian Model Averaging

- ▶ la fonction `bicreg` du package `BMA` implémente cette technique, mais sans grand contrôle sur le choix des a priori.
- ▶ la fonction `bas.lm` du package `BAS` permet de contrôler la loi a priori sur les modèles.

## Bayesian Model Averaging - bicreg(BMA)

```
library('BMA',quietly = T)
bma=bicreg(x,y)
summary(bma)
```

```
##
```

```
## Call:
```

```
## bicreg(x = x, y = y)
```

```
##
```

```
##
```

```
## 49 models were selected
```

```
## Best 5 models (cumulative posterior probability = 0.3884
```

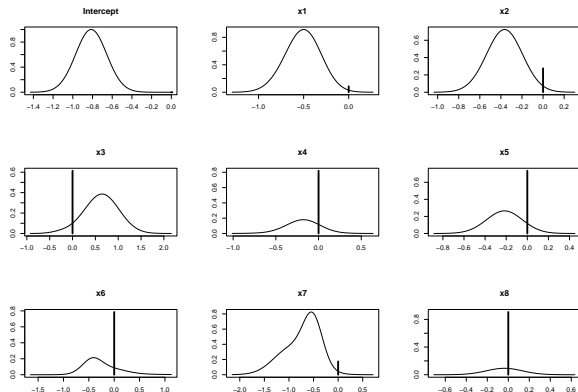
```
##
```

```
##
```

	p!=0	EV	SD	model 1	model 2	mod
## Intercept	100.0	-0.813281	0.15861	-0.8133	-0.8133	-0
## x1	91.4	-0.467197	0.23971	-0.6209	-0.4548	-0
## x2	72.4	-0.265331	0.21969	-0.3541	-0.3276	
## x3	38.6	0.239301	0.38959	0.6590	.	
## x4	18.0	-0.033645	0.11514	.	.	
## x5	26.5	-0.057497	0.13029	.	.	
## x6	21.4	-0.056307	0.19728	.	.	
## x7	82.5	-0.595837	0.43263	-0.9762	-0.4674	-0

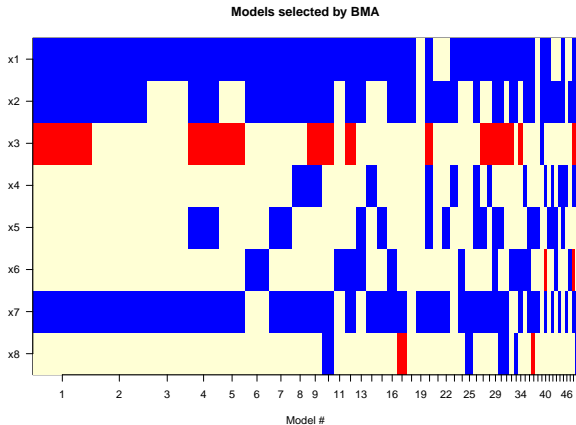
# Bayesian Model Averaging - bicreg(BMA)

```
plot(bma)
```



## Bayesian Model Averaging - bicreg(BMA)

```
imageplot.bma(bma)
```



## Exercice - données longley

- ▶ Tester le Bayesian Model Averaging avec les fonctions `bicreg`.

## Exercice - données IFPEN

Récupérer les données dans le matériel supplémentaire de la publication :

*Iapteff L., Jacques J., Rolland M. and Celse B. Reducing the number of experiments required for modelling the hydrocracking process with kriging through Bayesian transfer learning. Journal of the Royal Statistical Society, 2021*

disponible ici :

<https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12516>

- ▶ Effectuer une régression linéaire sur les données sources.
- ▶ Effectuer ensuite une régression bayésienne sur les données cibles, en ne prenant que 20 données sur l'ensemble de celles disponibles, et en utilisant la régression sur les données sources comme a priori.
- ▶ Comparer la qualité du modèle fréquentiste appris sur ces 20 données

# Régression bayésienne

Quelques a priori particuliers bien connus :

- ridge regression :  $\beta \sim \mathcal{N}(0, \frac{\sigma^2}{\lambda} I)$ :

$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - \alpha - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}$$

- LASSO regression :  $\beta \sim \operatorname{Laplace}(\frac{2\sigma^2}{\lambda})$ :

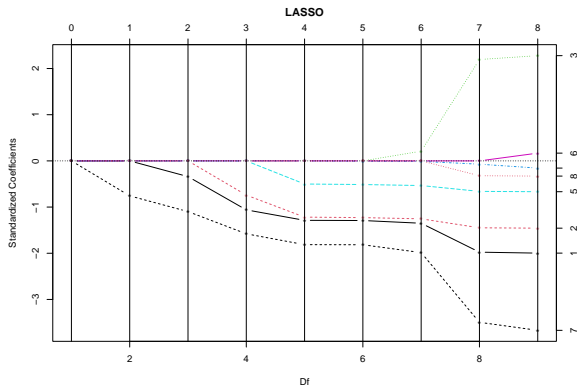
$$\hat{\beta}^{LASSO} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - \alpha - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$



# Exemple LASSO - données caterpillar

L'algorithme LARS permet de calculer les solutions LASSO pour toutes les valeurs de  $\lambda$  en une seule passe

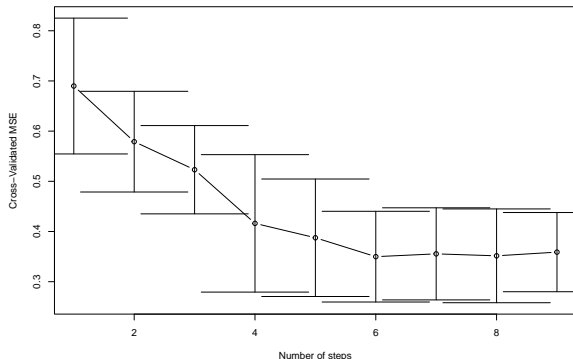
```
library('lars')  
model_lasso=lars(as.matrix(caterpillar[, -9]), caterpillar$y,  
                 type="lasso", trace=F, normalize=TRUE)  
plot(model_lasso, xvar='df', plottype='coeff')
```



## Exemple LASSO - données caterpillar

Pour choisir le  $\lambda$  optimal, on calcule la MSE par validation croisée (CV) à chaque étape de l'algorithme LARS. Le choix de  $\lambda$  revient à choisir l'étape de l'algorithme conduisant à la meilleure CV-MSE.

```
cv=cv.lars(as.matrix(caterpillar[, -9]), caterpillar$y, K=3,  
           trace=F, plot.it=T, se=T, type=c("lasso"),  
           mode='step', normalize=TRUE)
```



## Exemple LASSO - données caterpillar

Plutôt que de choisir l'étape conduisant CV-MSE la plus faible, on choisit l'étape la plus petite (en numéro d'étape) telle que la CV-MSE moyenne soit dans l'intervalle de confiance de la plus petite CV-MSE moyenne. Ainsi on choisit un modèle un peu plus parcimonieux (moins de variables actives) avec une CV-MSE moyenne non significativement différente de la plus faible observée.

Ici, la plus faible CVMSE moyenne est à l'étape 5. Mais la CVMSE moyenne de l'étape 4 est dans la bande de confiance; On va sélectionner la solution de l'étape 4 (*attention la première étape dans les sorties est l'étape 0, sans covariable*).

```
print(model_lasso$lambda[4])
```

```
## [1] 0.6939303
```

```
print(model_lasso$beta[4,])
```

```
##           x1           x2           x3           x4
## -0.001453016 -0.017941130  0.000000000  0.000000000  0.0000000
##           x7           x8
## -0.492714400  0.000000000
```

## Exemple LASSO - données caterpillar

- ▶ Les coefficients de régression non nuls sont ceux des variables significatives. La valeur de ces coefficients est néanmoins pénalisés par la pénalité LASSO qui a permis de mettre les autres coefficients non significatifs à 0.
- ▶ En pratique, il est souvent plus efficace de ne se servir de la régression LASSO que comme une méthode de sélection de variables.
- ▶ Une fois les variables pertinentes identifiées, il est donc préférable de ré-estimer un modèle de régression linéaire classique

## Exercice - données longley

- ▶ Tester une régression LASSO sur les données longley.



## Exercice - données Prostate

- ▶ Récupérer le jeu de données Prostate dans le package `lasso2`.
- ▶ Essayer de construire le meilleur modèle de régression linéaire possible pour prédire la variable `lcavol` (logarithme du volume du cancer)

Modèles graphiques (réseaux bayesiens  
hiérarchiques)



## Les réseaux bayésiens sous JAGS

# Les graphes acycliques dirigés

## Les **graphes acycliques dirigés** (DAG)

- ▶ un outil de modélisation général.
- ▶ graphique qui montre les relations (stochastiques et déterministes) entre les paramètres et les données.
- ▶ idée : retranscrire toutes les étapes qui, à partir des paramètres, permettent de générer les données.

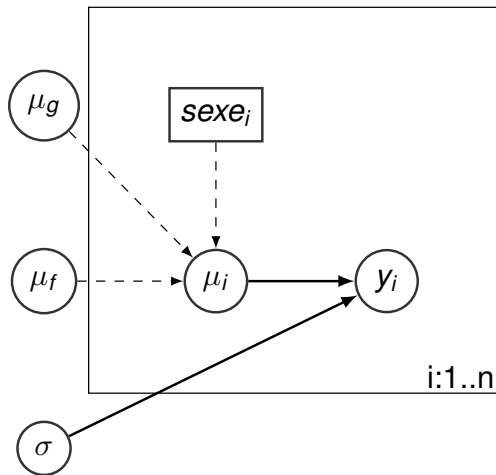
## Propriétés des DAG :

- ▶ acyclique : pas de cycle
- ▶ dirigés : les liens ont un sens (direction)

# Formalisme des DAG

- ▶ Noeuds du graphe
  - ▶ covariables : donnée fixe, supposée sans erreur : rectangle ex: *âge*
  - ▶ variable aléatoire (donnée observée à modéliser, paramètres) : ellipse ex: *poids à la naissance, intercept, pente*
- ▶ Liens du graphe
  - ▶ lien déterministe : flèche en pointillés ex:  $\mu_i = \alpha + \beta \times \text{age}_i$
  - ▶ lien stochastiques : flèche en trait plein ex:  $y_i \sim \mathcal{N}(\mu_i, \sigma^2)$
- ▶ Tous les noeuds du graphe à un même niveau sont contenus dans un cadre

## Exemple poids de naissance



# Exemple poids de naissance

Partie déterministe :

- ▶  $\mu_i = \mu_f$  si l'enfant est une fille ( $\text{sexe}_i = f$ )
- ▶  $\mu_i = \mu_g$  si l'enfant est un garçon ( $\text{sexe}_i = g$ )

Partie stochastique :

- ▶  $y_i \sim \mathcal{N}(\mu_i, \sigma^2)$

Information a priori

- ▶  $\mu_f$  et  $\mu_g$  sont uniformes entre 2.5kg et 5kg
- ▶  $\sigma$  est uniforme entre 0.2kg et 0.8kg

# Projet BUGS: Bayesian inference Using Gibbs Sampling

Développement et mise à disposition de logiciels flexibles permettant de mettre en oeuvre l'inférence bayésienne sur des modèles complexes, en utilisant les MCMC. Quelques outils disponibles :

- ▶ Winbugs
- ▶ Openbugs
- ▶ JAGS (Just Another Gibbs sampler - Martyn Plummer)
- ▶ Stan (nouveaux algorithmes - Andrew Gelman)
- ▶ RevBayes (pour la phylogénie)
- ▶ ...

# Installation de JAGS et rjags

- ▶ installer JAGS :

<http://sourceforge.net/projects/mcmc-jags/>

- ▶ installer le package R rjags

```
install.packages(rjags)
```

- ▶ guide complet de JAGS :

<http://sourceforge.net/projects/mcmc-jags/files/Manuals/>

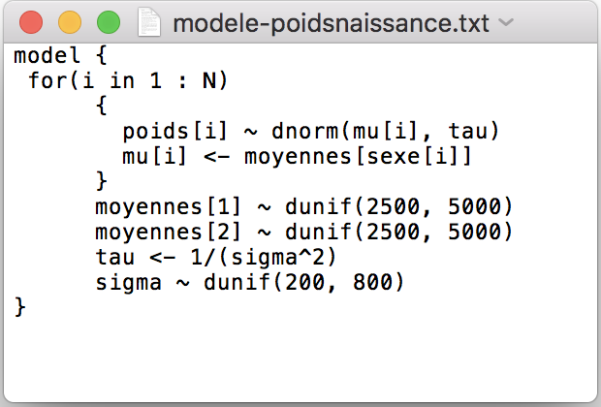
# Langage de type BUGS

- ▶ c'est un langage déclaratif qui ressemble à R
- ▶ déclaration des liens logiques :  
`noeud <- fonction(certains autres noeuds)`
- ▶ déclaration des liens stochastiques :  
`noeud ~ distribution(certains autres noeuds ou hyper-paramètres)`
- ▶ Attention : une sortie du modèle (données cibles que l'on modélise) doit toujours être définie de façon stochastique.
- ▶ un noeud est univarié (utiliser des boucles pour définir des vecteurs, matrices, array)



## Code pour le poids des enfants à la naissance

A écrire dans un fichier texte comme ci-dessous.



```
model {  
  for(i in 1 : N)  
  {  
    poids[i] ~ dnorm(mu[i], tau)  
    mu[i] <- moyennes[sexe[i]]  
  }  
  moyennes[1] ~ dunif(2500, 5000)  
  moyennes[2] ~ dunif(2500, 5000)  
  tau <- 1/(sigma^2)  
  sigma ~ dunif(200, 800)  
}
```

# Langage de type BUGS

- ▶ fonctions de base
  - ▶ `exp(x)` `log(x)` `sqrt(x)` `min(x1,x2)`
  - ▶ `step(x)` (1 si  $x \geq 0$ , 0 sinon)
  - ▶ `ifelse(x,a,b)` (si  $x$  alors  $a$  sinon  $b$ )
- ▶ fonctions vectorielles
  - ▶ `mean(v[])` `sd(v[])` `inverse(M[])`
  - ▶ `ranked(v[],k)` ( $k$ -ème élément de  $v$  dans l'ordre croissant)
- ▶ fonction de lien pouvant être utilisé à gauche de `<-`
  - ▶ `logit(y)` `probit(y)` `log(y)`
- ▶ exemples de distribution
  - ▶ `x~dbern(p)`
  - ▶ `x~dbin(p,n)`
  - ▶ `x~dnorm(mu,tau)` (attention tau est l'inverse de la variance)
  - ▶ `x~dbeta(alpha,beta)`
  - ▶ `x~dmulti(p[],n)` (multinomiale)
  - ▶ `x~dmnorm(mu[],T[,])` (T inverse de la matrice de variance)

# Utiliser rjags

- ▶ il faut commencer par définir les données (*attention à la cohérence des noms dans le modèle et dans les données*)

```
data=read.table('Rcode/poidsnaissance.txt',header = T,sep=',')
sexe=data$SEXE+1
poids=data$POIDNAIS
library(rjags)
data <- list(poids=poids, sexe=sexe, N=length(poids))
```

- ▶ il faut définir l'initialisation (ici 3 différentes):

```
inits <- list(list(moyennes=c(2600,4000), sigma=500),
               list(moyennes=c(4500,2700), sigma=700),
               list(moyennes=c(4000,4000), sigma=300))
```

# Utiliser rjags

- ▶ on définit ensuite le modèle, les données, les initialisations

```
m1 <- jags.model('Rcode/modelepoidsnaissance.txt', data=data,  
                 inits=inits, n.chains=3)
```

```
## Compiling model graph  
##   Resolving undeclared variables  
##   Allocating nodes  
## Graph information:  
##   Observed stochastic nodes: 48  
##   Unobserved stochastic nodes: 3  
##   Total graph size: 108  
##  
## Initializing model
```

# Utiliser rjags

- ▶ on itère l'algorithme MCMC sur une période de chauffe

```
update(m1, 3000)
```

- ▶ puis de nouvelles itérations permettant de générer les paramètres selon leur loi a posteriori

```
mcmc1 <- coda.samples(m1, variable.names=c("moyennes", "sigma"),  
                      n.iter=2000)
```

- ▶ l'objet mcmc1 ainsi obtenu est une liste de 3 tableaux (1 par init.). Le tableau  $i$  s'obtient par `mcmc1[[i]]`. Les colonnes de chaque tableau correspondent aux paramètres, les lignes aux itérations.
- ▶ on peut par exemple calculer la moyenne a posteriori de  $\mu_1$  :

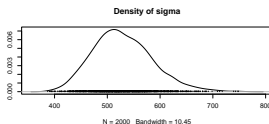
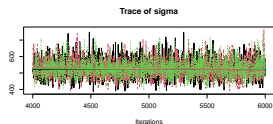
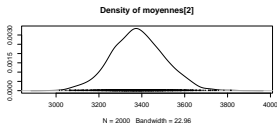
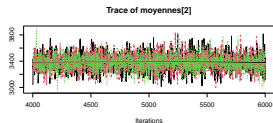
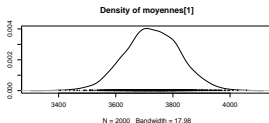
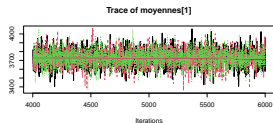
```
mean(mcmc1[[1]][, "moyennes[1]"])
```

```
## [1] 3728.765
```

# Utiliser rjags

- ▶ on peut aussi représenter les simulations des paramètres selon leur lois a posteriori par l'algo. MCMC

```
plot(mcmc1)
```



## Utiliser rjags

- Pour savoir si on a atteint l'état stationnaire de la chaîne MCMC, on peut calculer l'indice de réduction de la variance de Gelman :

$$\sqrt{\frac{\text{variance totale}}{\text{variance intra-chaînes}}}$$

```
gelman.diag(mcmc1)
```

```
## Potential scale reduction factors:
```

```
##
```

```
##           Point est. Upper C.I.
```

```
## moyennes[1]           1           1
```

```
## moyennes[2]           1           1
```

```
## sigma                1           1
```

```
##
```

```
## Multivariate psrf
```

```
##
```

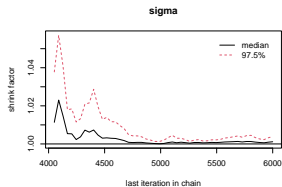
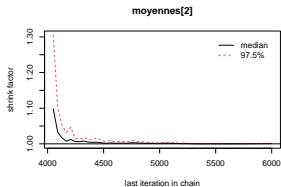
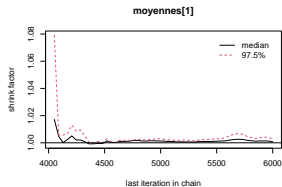
```
## 1
```

Ici l'indice est de 1, ce qui signifie que les 3 chaînes échantillonnent suivant la même distribution (a posteriori), la convergence peut être supposée

# Utiliser rjags

On peut aussi représenter graphiquement l'indice de Gelman :

```
gelman.plot(mcmc1)
```

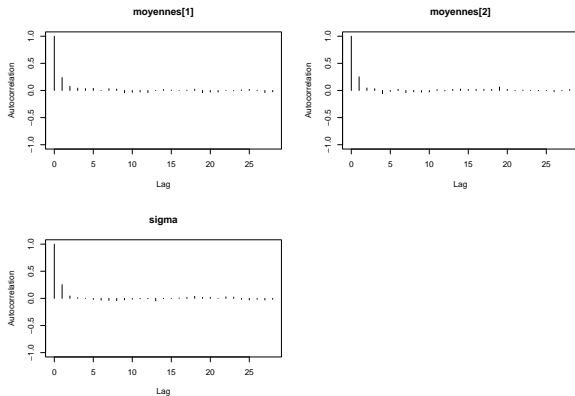




# Utiliser rjags

- Une chaîne MCMC ne doit pas être auto-corrélée (sinon il faut plus d'itérations). On peut vérifier cela à l'aide de la commande suivante :

```
autocorr.plot(mcmc1[[1]])
```



Ici pas de soucis particulier.

## Utiliser rjags

- Enfin, on peut exploiter les chaînes MCMC pour obtenir les estimations bayésiennes des paramètres :

```
summary(mcmc1)
```

```
##
## Iterations = 4001:6000
## Thinning interval = 1
## Number of chains = 3
## Sample size per chain = 2000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##              Mean      SD Naive SE Time-series SE
## moyennes[1] 3727.8  96.62   1.2474           1.5748
## moyennes[2] 3382.4 123.37   1.5927           2.0795
## sigma       529.1  56.50   0.7294           0.9752
##
## 2. Quantiles for each variable:
##
```

## Utiliser rjags

Enfin, on pourra si l'on veut comparer différents modèles examiner la valeur du critère DIC, qui doit être la plus petite possible

```
dic.samples(m1,n.iter=1000)
```

```
## Mean deviance: 737
```

```
## penalty 3.12
```

```
## Penalized deviance: 740.1
```

## Exercice poids de naissance

Toujours sur les poids des enfants à la naissance, introduire d'autres variables (présentes dans le fichier).

Utilisation de la modélisation hiérarchique en médecine

# Hypothèse pour les données hiérarchiques

On veut faire de l'inférence sur des paramètres  $\theta_1, \dots, \theta_N$  liés à  $N$  unités : individus, sous-ensemble, essais cliniques (méta-analyses)

Hypothèses possibles :

- ▶ paramètres identiques :  $\theta_n = \theta$ , pas d'effet *unité*, les données proviennent toutes de la même population
- ▶ paramètres indépendants :  $\theta_n \neq \theta_m$  pour tout  $n \neq m$ , effet *unité* présent, les données de chaque unité peuvent être analysées indépendamment
- ▶ paramètres échangeables : les  $\theta_n$  sont supposés non identiques mais similaires (provenant d'une même distribution), leur index ne porte aucune information

## Exemple - chirurgie cardiaque infantile

```
print(chirurgie)
```

##	hopital	operation	deces
## 1	bristol	143	41
## 2	leicester	187	25
## 3	leeds	323	24
## 4	oxford	122	23
## 5	guys	164	25
## 6	liverpool	405	42
## 7	southampton	239	24
## 8	great_ormond_st	482	53
## 9	newcastle	195	26
## 10	harefield	177	25
## 11	birmingham	581	58
## 12	brompton	301	31

## Exemple - chirurgie cardiaque infantile

On suppose que le nombre de décès  $y_i$  au sein de l'hôpital  $i$  est :

$$y_i \sim \mathcal{B}(n_i, \theta_i)$$

Plusieurs hypothèses sont possibles :

1. paramètres identiques (*pas d'effet hôpital*) :

$$\theta_1 = \dots = \theta_N = \theta$$

2. paramètres indépendants (*effet fixe hôpital*) :

$$\theta_1 \neq \dots \neq \theta_N$$

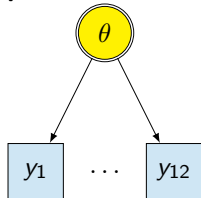
3. paramètres échangeables (*effet aléatoire hôpital*) :

$$\text{logit}(\theta_n) \sim \mathcal{N}(\mu, \sigma^2)$$

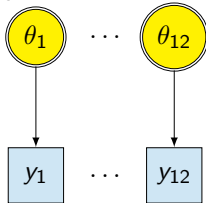


## Exemple - chirurgie cardiaque infantile

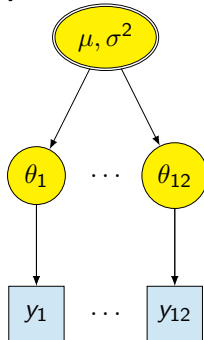
**Identical parameters**



**Independent parameters**

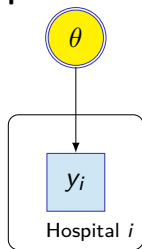


**Exchangeable parameters**

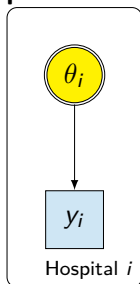


## Exemple - chirurgie cardiaque infantile

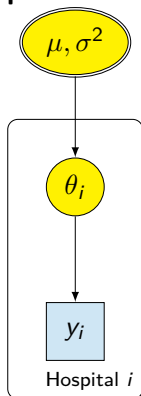
**Identical parameters**



**Independent parameters**



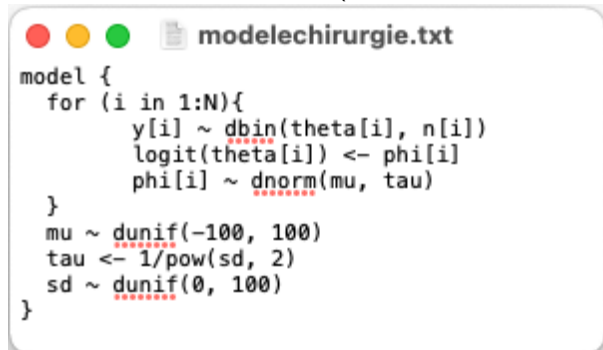
**Exchangeable parameters**



## Exemple - chirurgie cardiaque infantile

- ▶ la comparaison entre les différents types d'effets sera fait en comparant les 3 modèles
- ▶ Rq: *la comparaison entre effet fixe et effet aléatoire est habituellement un choix de modélisation pour un fréquentiste*

Le modèle le plus complet (effet hopital aléatoire) se code par :



```
modelechirurgie.txt
model {
  for (i in 1:N){
    y[i] ~ dbin(theta[i], n[i])
    logit(theta[i]) <- phi[i]
    phi[i] ~ dnorm(mu, tau)
  }
  mu ~ dunif(-100, 100)
  tau <- 1/pow(sd, 2)
  sd ~ dunif(0, 100)
}
```

## Exercice - chirurgie cardiaque infantile

- ▶ coder les trois modèles et comparer les critères DIC.
- ▶ conclure sur l'analyse statistique.

## Exemple - méta-analyse

On s'intéresse à la survenue d'infections chez les prématurés (septicémie), et l'effet d'injection intraveineuse d'immunoglobuline.

```
print(sepsis)
```

##	etude	annee	sepsis_t	total_t	sepsis_c	total_c
## 1	Bussel	1990	20	61	23	65
## 2	Chirico	1987	2	43	8	43
## 3	Clapp	1989	0	56	5	59
## 4	Conway	1990	8	34	14	32
## 5	Fanaroff	1994	186	1204	209	1212
## 6	Haque	1986	4	100	5	50
## 7	Ratrisawadi	1991	10	68	13	34
## 8	Sandberg	2000	19	40	13	41
## 9	Tanzer	1997	3	40	8	40
## 10	Weisman	1994	40	372	39	381

## Exemple - méta-analyse

Soient, pour chaque étude  $s$  :

- ▶  $y_{sT}$  : nb de succès (non infection) parmi les nouveaux-nés traités par immunoglobuline intraveineuse
- ▶  $n_{sT}$  : nb de nouveaux-nés traités par immunoglobuline intraveineuse
- ▶  $y_{sC}$  : nb de succès (non infection) parmi les nouveaux-nés non traités (contrôles)
- ▶  $n_{sC}$  : nb de nouveaux-nés non traités

# Exemple - méta-analyse

Hypothèses :

1. paramètres identiques (méta-analyse à effets fixes):

$$\begin{aligned}y_{sT} &\sim \mathcal{B}(n_{sT}, p_{sT}) \\ \text{logit}(p_{sT}) &= \alpha_s + \delta \\ y_{sC} &\sim \mathcal{B}(n_{sC}, p_{sC}) \\ \text{logit}(p_{sC}) &= \alpha_s\end{aligned}$$

où

- ▶  $\alpha_s$  log odds d'avoir un succès chez les contrôles dans l'étude  $s$
- ▶  $\alpha_s + \delta$  log odds d'avoir un succès chez les traités dans l'étude  $s$
- ▶  $\delta$  log odds ratio

# Exemple - méta-analyse

Hypothèses :

2. paramètres échangeables (méta-analyse à effets aléatoires):

$$\begin{aligned}y_{sT} &\sim \mathcal{B}(n_{sT}, p_{sT}) \\ \text{logit}(p_{sT}) &= \alpha_s + \mu_s \\ y_{sC} &\sim \mathcal{B}(n_{sC}, p_{sC}) \\ \text{logit}(p_{sC}) &= \alpha_s \\ \mu_s &\sim \mathcal{N}(\delta, \tau^2)\end{aligned}$$

où

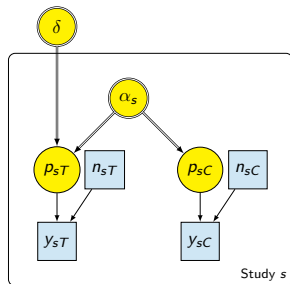
- ▶  $\alpha_s$  log odds d'avoir un succès chez les contrôles dans l'étude  $s$
- ▶  $\alpha_s + \mu_s$  log odds d'avoir un succès chez les traités dans l'étude  $s$
- ▶  $\delta$  log odds ratio global

Ce modèle permet de prendre en compte l'hétérogénéité des procédures, caractéristiques des patients, etc. . . entre les sites, dans la variabilité totale observée.



# Exemple - méta-analyse

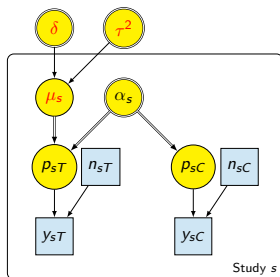
Modèle à paramètres indépendants (effets fixes)



```
for (s in 1:Ns){  
  for (a in 1:2){  
    y[s,a] ~ dbin(p[s,a], n[s,a])  
    logit(p[s,a]) <- alpha[s] +  
      delta[a]  
  }  
  alpha[s] ~ dnorm(0, 0.01)  
}  
delta[1] <- 0  
delta[2] ~ dnorm(0, 0.01)
```

# Exemple - méta-analyse

## Modèle à paramètres échangeable (effets aléatoires)



```
for (s in 1:Ns){  
  for (a in 1:2){  
    y[s,a] ~ dbin(p[s,a], n[s,a])  
    logit(p[s,a]) <- alpha[s] +  
      mu[s,a]  
  }  
  mu[s, 1] <- 0  
  mu[s, 2] ~ dnorm(delta, prec.mu)  
  
  alpha[s] ~ dnorm(0, 0.01)  
}  
delta ~ dnorm(0, 0.01)  
  
prec.mu <- 1/(sd.mu * sd.mu)  
sd.mu ~ dunif(0, 10)
```

## Exercice - méta-analyse

- ▶ comparer les deux modèles à l'aide des critères DIC
- ▶ conclure sur l'analyse statistique

## Pour aller plus loin

- ▶ d'autres exemples d'analyses statistiques :

[https://rstudio-pubs-static.s3.amazonaws.com/272658\\_ae4d482c86514674be17042c852ebbf.html](https://rstudio-pubs-static.s3.amazonaws.com/272658_ae4d482c86514674be17042c852ebbf.html)

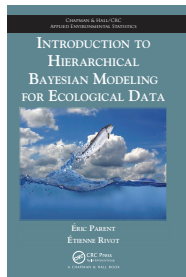
- ▶ gestion des données manquantes, données censurées :

[https://www.mrc-bsu.cam.ac.uk/wp-content/uploads/2019/06/handouts\\_master.pdf](https://www.mrc-bsu.cam.ac.uk/wp-content/uploads/2019/06/handouts_master.pdf)

## Utilisation de la modélisation hiérarchique en écologie

# Bibliographie

## Introduction to hierarchical bayesian modeling for ecological data



Données et codes R associés disponibles sur :  
[http://sirs.agrocampus-ouest.fr/bayes\\_V2/index.html](http://sirs.agrocampus-ouest.fr/bayes_V2/index.html)

# Capture / Marque / Recapture

On dispose des données suivantes (section 9.2 du livre) :

- ▶  $y_1$  : nombre de poissons piégés
- ▶  $y_2+y_3$  : poissons capturés non relachés
- ▶  $y_4=y_1-(y_2+y_3)$  : nombre de poissons marqués et relachés
- ▶  $y_5$  : nombre de poissons marqués recapturés
- ▶  $y_6$  : nombre de poissons non marqués recapturés

On suppose que

$$y_1 \sim \mathcal{B}(n, p)$$

où :

- ▶  $n$  est la taille de la population, que l'on cherche à connaître
- ▶  $p$  est l'efficacité de piègeage

Estimer la loi a posteriori de ces deux paramètres