

Programme

Lundi 06/09	10h30-12h00	Rappels en probabilité
Lundi 20/09	10h30-12h00	Principes de la démarche bayésienne en statistique
Lundi 04/10	10h30-12h00	Initiation aux modèles graphiques
Lundi 18/10	10h30-12h00	Distributions a priori Présentation des logiciels WinBUGS / JAGS / STAN https://inrae-fr.zoom.us/j/7675335297?pwd=dIBtZlpFSkZPU2RocnY4QXg3WUJmZz09 Mdp : Bayes2021!
Lundi 08/11	10h30-12h00	Vérification et évaluation de modèles https://inrae-fr.zoom.us/j/9141600517
Lundi 22/11	10h30-12h00	Estimation des distributions a posteriori avec des méthodes numériques https://inrae-fr.zoom.us/j/9141600517
Lundi 29/11	14h30-15h30	Séminaire 1 d'ouverture : Estimation de la distribution spatio-temporelle de mammifères marins
Lundi 06/12	14h30-15h30	Séminaire 2 d'ouverture : Approche bayésienne pour la prévision probabiliste de températures à partir de post-traitement de données d'ensemble
Lundi 13/12	14h30-15h30	Séminaire 3 d'ouverture : Reconstruction de l'histoire évolutive de langues humaines

Quelques outils pour la construction de DAG (graphs bayésiens orientés -> cours Modèles graphiques)

- TikZ : pour créer des diagrammes sous Overleaf/LaTeX

https://www.overleaf.com/learn/latex/TikZ_package

- Bayesian Analysis Reporting Guidelines : <https://doi.org/10.1038/s41562-021-01177-7>

Kruschke, J.K. Bayesian Analysis Reporting Guidelines. *Nat Hum Behav* (2021).

BiDAG: Bayesian Inference for Directed Acyclic Graphs – CRAN <https://cran.r-project.org/web/packages/BiDAG/>

- Option DoodleBUGS dans WinBUGS

<https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-bugs-resources-online/>

<https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>

- Graphical diagram in *Doing Bayesian Data Analysis*. <https://sites.google.com/site/doingbayesiandataanalysis/software-installation>

- Rgraphviz

- Application Shiny comparaison Bayésien-fréquentiste

<https://iupbsapps.shinyapps.io/KruschkeFreqAndBayesApp/>

Distributions a priori

Présentation des logiciels WinBUGS / JAGS / STAN

1^{ère} partie

- Distributions a priori (*cours Olivier David*)
- Elicitation de dires d'expert (pour définir des lois a priori) (*cours David Makowski*)

2^{ème} partie

- Logiciels d'inférence bayésienne (*cours avec Matthieu Authier et Sophie Ancelet*)
- exemple de code
- Un outil de couplage mécanistico-statistique (EDP/bayésien) : MSE (si on a le temps !)

Distributions *a priori*

Olivier David

MaIAGE, INRA, Jouy-en-Josas

BioBayes, Pornic, mars 2019

1. Exemple

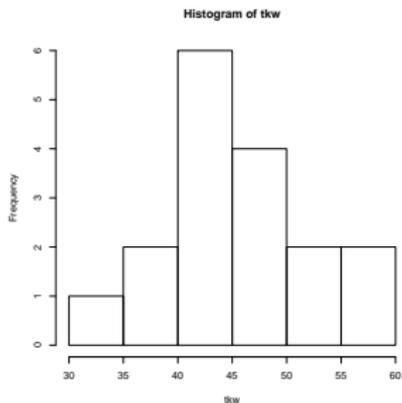
Essai à la ferme réalisé en 2015 pour comparer 14 populations de blé, comprenant 17 parcelles.

Essai réalisé en collaboration avec l'équipe "Diversité, Evolution et Adaptation des Populations" de l'UMR "Génétique Quantitative et Évolution-Le Moulon" dans le cadre d'un projet sur le développement de l'agriculture biologique pour le blé.

Plan de l'essai :

10	13	8	3	1	12	14	5	4
9	11	4	12	2	7	13	6	

Variable poids de mille grains (tkw) en g.



L'observation y_{ij} pour la population i et la répétition j est modélisée par :

$$y_{ij} = \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2).$$

Il y a 15 paramètres et 17 données.

Anova

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
germplasm	14	36691	2620.82	223.73	0.0004325 ***
Residuals	3	35	11.71		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Estimate	Std. Error	t value	Pr(> t)
germplasm 10	58.410	3.423	17.066	0.000438 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.423 on 3 degrees of freedom
Multiple R-squared: 0.999, Adjusted R-squared: 0.9946
F-statistic: 223.7 on 14 and 3 DF, p-value: 0.0004325

	2.5 %	97.5 %
germplasm 10	47.51768	69.30232

Quelles lois *a priori* utiliser pour α_i et σ_ϵ ?

2. Loi *a priori* informative

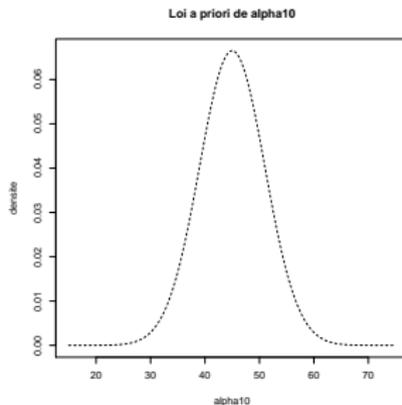
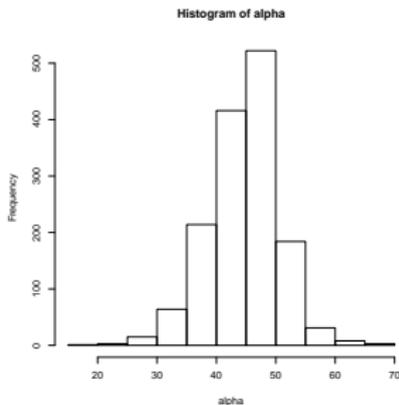
Loi *a priori informative* : loi qui cherche à prendre en compte l'information *a priori* dont on dispose sur les paramètres.

On utilise les résultats de l'anova : $\alpha_{10} \sim \mathcal{N}(58.410, 3.423^2)$.

	Mean	SD
alpha[10]	58.398365	2.0805987
sigma	2.649136	0.5259157

On peut construire une loi *a priori* informative en utilisant des données venant d'autres expériences, de la littérature etc.

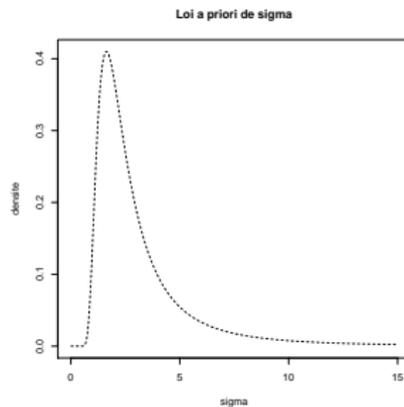
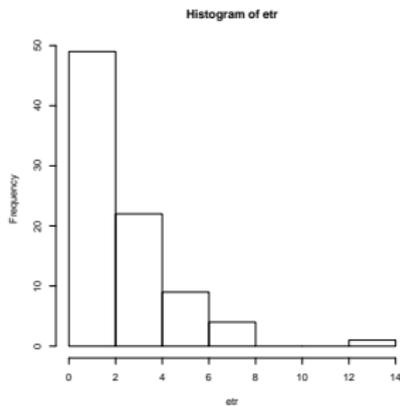
Histogramme des $\hat{\alpha}_i$ dans d'autres essais et loi *a priori* de α_{10} :



Loi *a priori* normale pour α : $\alpha_i \sim \mathcal{N}(\mu_0, \sigma_{\alpha 0}^2)$, $i = 1, \dots, 14$.

Moyenne des $\hat{\alpha}_i \approx 45 \Rightarrow \mu_0 = 45$. Ecart-type des $\hat{\alpha}_i \approx 6 \Rightarrow \sigma_{\alpha 0} = 6$.

Histogramme des $\hat{\sigma}_\epsilon$ dans d'autres essais et loi *a priori* de σ_ϵ :

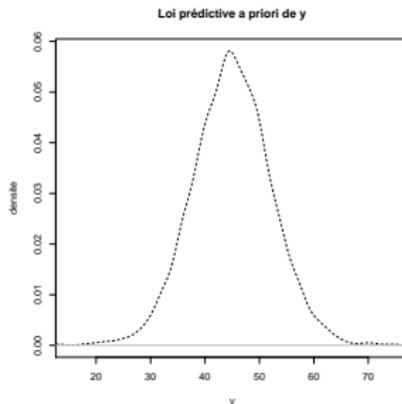
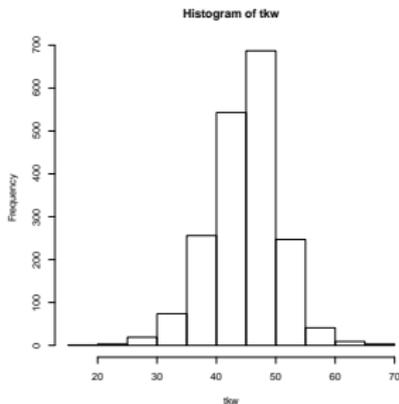


Loi *a priori* gamma pour σ_ϵ^{-2} : $\sigma_\epsilon^{-2} \sim \mathcal{G}(1, 4)$.

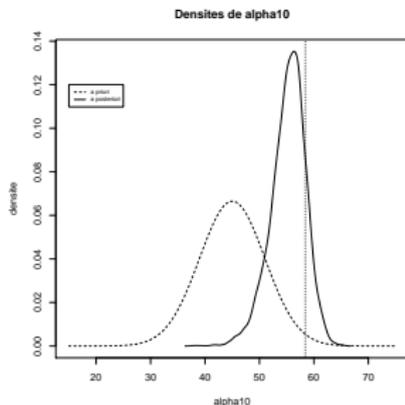
Loi prédictive *a priori* de y_{101} :

$$p(y_{101}) = \int p(y_{101}|\alpha_{10}, \sigma_\varepsilon) p(\alpha_{10}) p(\sigma_\varepsilon) d\alpha_{10} d\sigma_\varepsilon.$$

Histogramme de tkw dans d'autres essais et loi prédictive *a priori* de y_{101} :



Lois de α_{10} :



Résumés de la loi *a posteriori* :

	2.5%	25%	50%	75%	97.5%
alpha[10]	48.152562	53.36665	55.55667	57.411518	60.893645
sigma	1.860142	2.55137	3.10284	3.841427	5.898445

L'intervalle de crédibilité de α_{10} est moins grand que celui de l'anova.

3. Loi *a priori* “non informative”

Loi *a priori* “non informative” : loi qui cherche à apporter le moins d'information *a priori* possible sur les paramètres.

Loi *a priori* normale pour $\alpha|\sigma_\varepsilon$: $\alpha_i|\sigma_\varepsilon \sim \mathcal{N}\left(\mu_0, \frac{\sigma_\varepsilon^2}{r_0}\right)$, $i = 1, \dots, 14$.

Loi *a posteriori* de $\alpha|\sigma_\varepsilon$:

$$\alpha_i|\sigma_\varepsilon, y \sim \mathcal{N}\left(\mu'_i, \frac{\sigma_\varepsilon^2}{r'_i}\right), \quad i = 1, \dots, 14,$$
$$\mu'_i = \frac{r_0\mu_0 + r_i y_{i\bullet}}{r_0 + r_i}, \quad r'_i = r_0 + r_i,$$

r_i est le nombre de répétitions de la population i ,

$y_{i\bullet}$ est la moyenne des observations pour la population i .

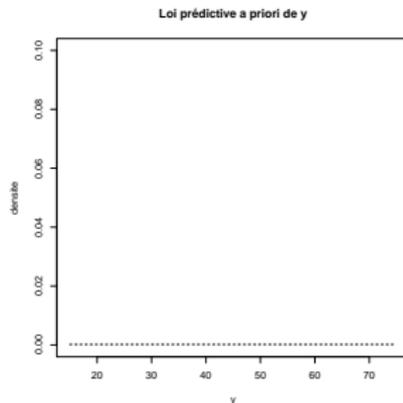
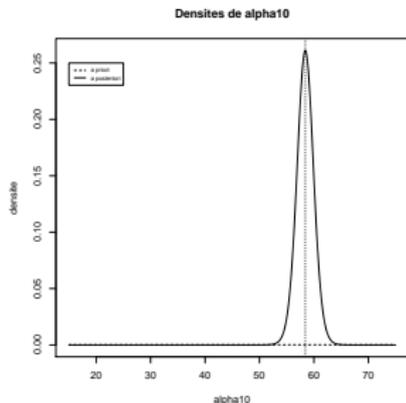
Loi *a priori conjuguée* avec la vraisemblance : la loi *a posteriori* appartient à la même famille de lois que la loi *a priori*.

La loi *a priori* apporte la même information qu'un jeu de données comprenant r_0 répétitions de la population i et tel que $\hat{\alpha}_i = \mu_0$.

La loi *a priori* a peu d'influence sur les résultats si les répétitions sont nombreuses.

Loi "non informative" pour α : $r_0 = 10^{-6}$.

Lois de α_{10} et loi prédictive *a priori* de y_{101} :



La loi *a priori* de α_{10} a une grande variance, elle est vague.

Résumés de la loi *a posteriori* :

	2.5%	25%	50%	75%	97.5%
alpha[10]	55.235266	57.380575	58.400574	59.413355	61.524943
sigma	1.144239	1.377826	1.536445	1.726912	2.240271

Il est en général plus difficile de construire une loi *a priori* “non informative” qu’une loi informative.

Dans cette école, on se limite à des lois vagues.

Une analyse de sensibilité à la loi *a priori* peut aider à détecter des problèmes.

Quand $r_0 \rightarrow 0$, la loi *a priori* de $\alpha|\sigma$ tend vers la loi uniforme sur \mathbb{R} .

Loi *a priori impropre* : loi *a priori* dont l'intégrale ne vaut pas 1 mais $+\infty$; elle n'est pas une véritable loi de probabilité.

Dans ce cas, il faut vérifier si la loi *a posteriori* est propre.

Quand $r_0 \rightarrow 0$, la loi *a posteriori* de $\alpha|\sigma$ est propre :

$$\alpha_i|\sigma_\varepsilon, y \sim \mathcal{N}\left(\alpha'_i, \frac{\sigma_\varepsilon^2}{r'_i}\right), \quad i = 1, \dots, 14,$$
$$\alpha'_i = y_{i\bullet}, \quad r'_i = r_i.$$

On peut donc utiliser la loi *a priori* uniforme pour $\alpha|\sigma_\varepsilon$.

4. Loi *a priori* hiérarchique

Loi *a priori* hiérarchique : loi *a priori* qui a plusieurs niveaux.

Loi *a priori* de α à 2 niveaux :

$$\text{Niveau 1} : y_{ij} \sim \mathcal{N}(\alpha_i, \sigma_\varepsilon^2),$$

$$\text{Niveau 2} : \alpha_i \sim \mathcal{N}(\mu, \sigma_\alpha^2), \quad \sigma_\varepsilon^{-2} \sim \mathcal{G}(1, 4),$$

$$\text{Niveau 3} : \mu \sim \mathcal{N}(0, \sigma_\mu^2), \quad \sigma_\alpha^{-2} \sim \mathcal{G}(\gamma, \gamma), \\ \sigma_\mu \rightarrow +\infty, \quad \gamma \rightarrow 0.$$

μ et σ_α sont des paramètres inconnus du niveau 2, appelés *hyperparamètres*.

On ne peut pas utiliser cette loi car la loi *a posteriori* est impropre.

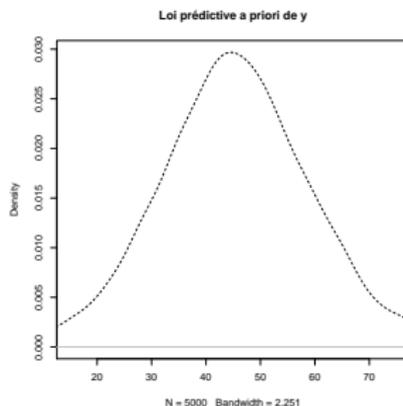
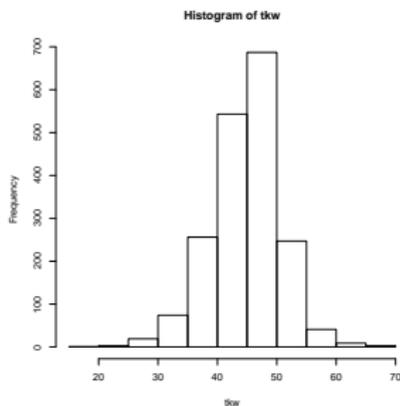
Loi *a priori* de α à 2 niveaux :

$$\text{Niveau 1} : y_{ij} \sim \mathcal{N}(\alpha_i, \sigma_\varepsilon^2),$$

$$\text{Niveau 2} : \alpha_i \sim \mathcal{N}(\mu, \sigma_\alpha^2), \quad \sigma_\varepsilon^{-2} \sim \mathcal{G}(1, 4),$$

$$\text{Niveau 3} : \mu \sim \mathcal{N}(45, 10^2), \quad \sigma_\alpha \sim \mathcal{N}^+(0, 10^2).$$

Histogramme de tkw dans d'autres essais et loi prédictive *a priori* de y_{101} :



Résumés de la loi *a posteriori* :

	2.5%	25%	50%	75%	97.5%
alpha[10]	45.097221	52.442597	55.387277	57.598797	61.534847
mu	42.097967	44.703527	45.897282	47.163301	49.796280
sigma	1.839705	2.610050	3.278263	4.293542	7.188869
sigma.alpha	1.723310	5.089808	6.239840	7.462387	10.520161

L'estimation de α_{10} s'est rapprochée de μ : on dit qu'elle est *rétrécie*.

L'intervalle de crédibilité de α_{10} est moins grand que celui de l'anova.

5. Références bibliographiques

- ▶ George E. P. Box and George C. Tiao. Bayesian inference in statistical analysis. Addison-Wesley, 1973.
- ▶ Collectif BioBayes. Initiation à la Statistique Bayésienne. Chapitre 7. Ellipses, 2015.
- ▶ Jean-Jacques Dreesbeke, Jeanne Fine, and Gilbert Saporta, editors. Méthodes bayésiennes en statistique. Editions Technip, 2002.
- ▶ A. Gelman (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian analysis, 1(3), 515-534.
- ▶ Jean-Michel Marin and Christian Robert. Bayesian core : a practical approach to computational Bayesian statistics. Springer Science & Business Media, 2007.
- ▶ Christian P. Robert. Le choix bayésien : Principes et pratique. Springer, 2006.

Elicitation de dires d'expert

Cours de David Makowski – Ecole chercheurs Biobayes 2019

Définition :

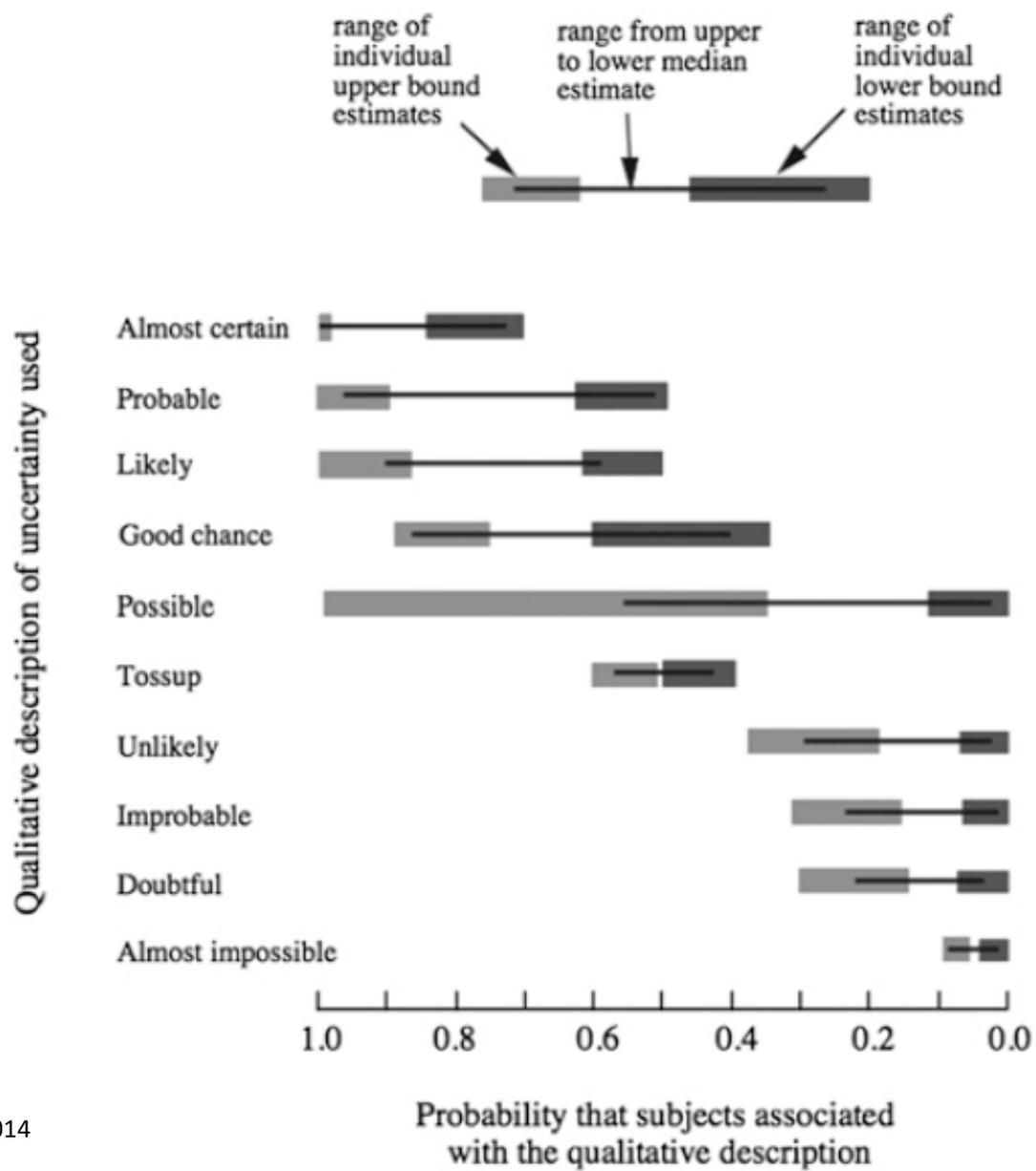
L'élicitation probabiliste est une méthode pour représenter et capitaliser les connaissances des experts sous la forme de distributions de probabilité

Intérêt

- Synthèse de connaissances
- Analyse d'incertitude
- Statistique bayésienne
- Nombreux développements méthodologiques dans les années 1990

« Qualitative Uncertainty Words Are Not Sufficient »

Morgan (2014)



D'après Morgan, 2014



Contents lists available at [ScienceDirect](#)

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft



Short communication

A web-based tool for eliciting probability distributions from experts

David E. Morris^{a,*}, Jeremy E. Oakley^b, John A. Crowe^a



Objectif

L'élicitation consiste à réaliser une synthèse quantitative des connaissances d'un expert (ou d'un groupe d'experts) sur une quantité d'intérêt Q pour laquelle il existe une incertitude due à un manque de données disponibles.

Le résultat d'une élicitation se présente sous la forme d'une distribution de probabilité reflétant les connaissances et le niveau d'incertitude de l'expert.

Définir des profils d'experts avant l'élicitation

- Définir le type d'expertises requises
 - Identifier les compétences recherchées
 - Définir les profils d'experts adaptés
- Questionnaire pour vérifier l'adéquation entre le profil des experts et les besoins

Les principales étapes de l'élicitation

- i. Définition de la quantité (ou des quantités)
- ii. Identification de l'expert ou des experts
- iii. Choix d'une méthode d'élicitation
- iv. Elicitation de l'expert (ou des experts)
- v. Détermination d'une distribution de probabilité

Si besoin

- vi. Combinaison des distributions obtenues avec différents experts

Méthodes d'élicitation

- Roulette
- Quartiles
- Tertiles

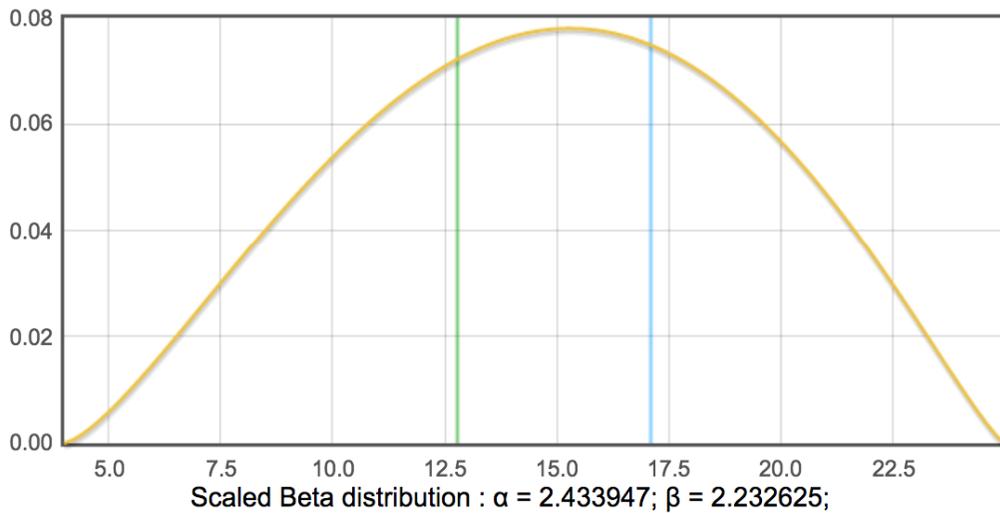
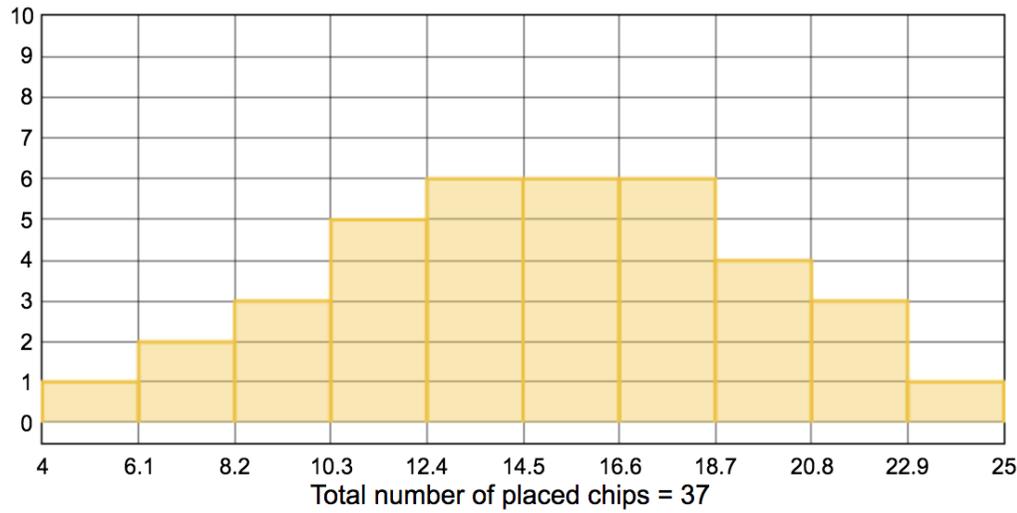
Deux outils informatiques gratuits

- Match tool

<http://optics.eee.nottingham.ac.uk/match/uncertainty.php>

- Package R « SHELF »

<https://cran.r-project.org/web/packages/SHELF/index.html>



Elicitation Options

Ranges

Lower Limit

Upper Limit

Input Mode

Roulette
 Quartile
 Tertile
 Probability
 Hybrid

Roulette Options

Number of Bins

Height of Grid

Fitting & Feedback

Distribution

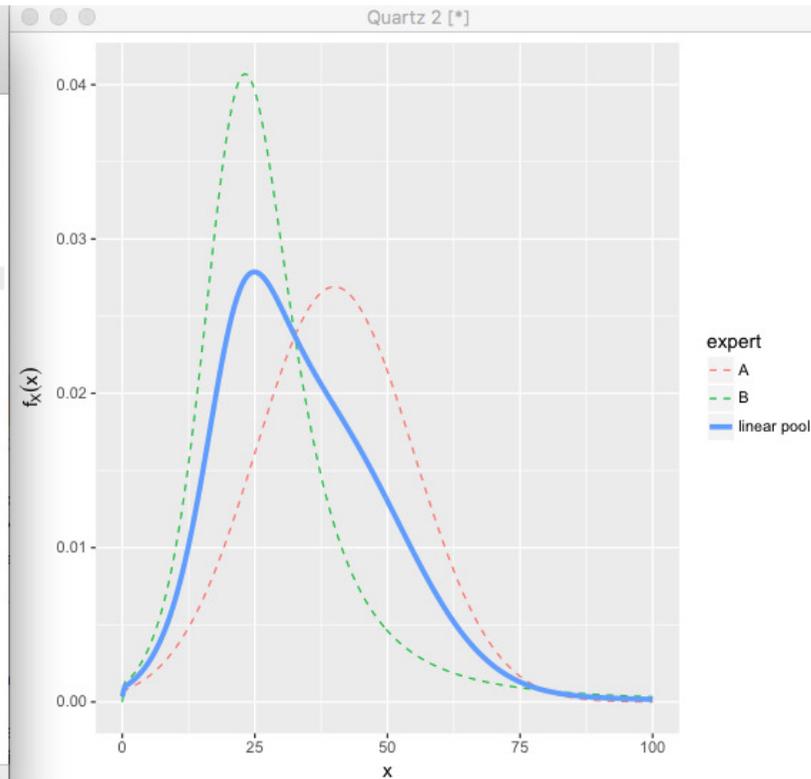
Normal
 Student-t
 Scaled Beta
 Gamma
 Log Normal
 Log Student-t
 Auto-select best fit

Feedback Percentiles

rd percentile =
 th percentile =

```
Sans titre
<fonctions> Recherche d'aide

1 library(SHELF)
2
3
4 ## 1) Elicit judgements from two experts individually
5 # Expert A states P(X<30)=0.25, P(X<40)=0.5, P(X<50)=0.75
6 # Expert B states P(X<20)=0.25, P(X<25)=0.5, P(X<35)=0.75
7 # Both experts state 0<X<100.
8
9 ## 2) Fit distributions to each expert's judgements
10 v <- matrix(c(30, 40, 50, 20, 25, 35), 3, 2)
11 p <- c(0.25, 0.5, 0.75)
12 myfit <- fitdist(vals = v, probs = p, lower = 0, upper = 100)
13
14 ## 3) Plot the fitted distributions, including a linear pool
15 plotfit(myfit, lp = T)
```



Méthode « roulette »

- i. Définir des valeurs min et max
- ii. Diviser la gamme de variation possible de la quantité Q en m cases (ex : $m=10$)
- iii. Demander à l'expert d'allouer n « jetons » à chaque case (ex: $n=20$).
 - L'ensemble des « jetons » décrit alors un histogramme.
 - La proportion de « jetons » allouée à chaque case reflète la probabilité que la valeur de Q soit dans cette case.
- iv. Ajuster une loi de probabilité à l'histogramme
- v. Faire valider le résultat de l'ajustement par l'expert.

Exemple

Elicitation du rendement de l'oignon « bio » dans une région du Sénégal ($t\ ha^{-1}$)

```
library(SHELF)
```

```
x <- roulette(lower = 4, upper = 25, gridheight = 10, nbins = 10)
```

Roulette elicitation

Show fit

Spread end probs over empty bins

Distribution

Normal

Student t

Gamma

Log normal

Log Student t

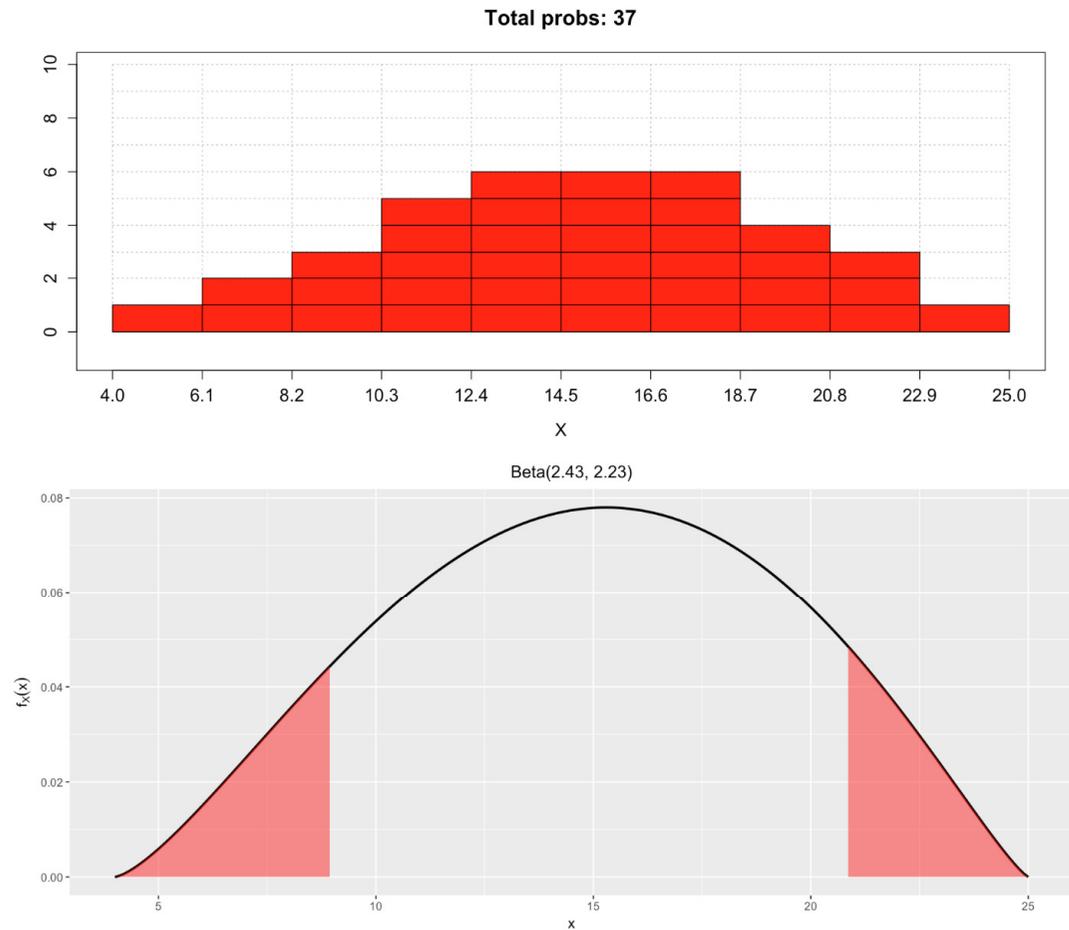
Beta

Best fitting

Student-t degrees of freedom

lower feedback quantile

upper feedback quantile



```
myfit <- fitdist(vals = x$v, probs = x$p, lower = 4, upper = 25)  
myfit
```

```
plotfit(myfit, ql=0.1,qu=0.9, xlab="Rendement bio au Senegal")
```

```
feedback(myfit)
```

```
feedback(myfit, quantiles=c(0.33,0.5,0.66))
```

```
> myfit
$Normal
  mean    sd
1 14.97745 4.705922

$Student.t
 location scale df
1 14.98137 3.8291 3

$Gamma
  shape    rate
1 5.647559 0.4990778

$Log.normal
 mean.log.X sd.log.X
1 2.354798 0.4182309

$Log.Student.t
 location.log.X scale.log.X df.log.X
1 2.361113 0.3454594 3

$Beta
  shape1 shape2
1 2.433733 2.232446

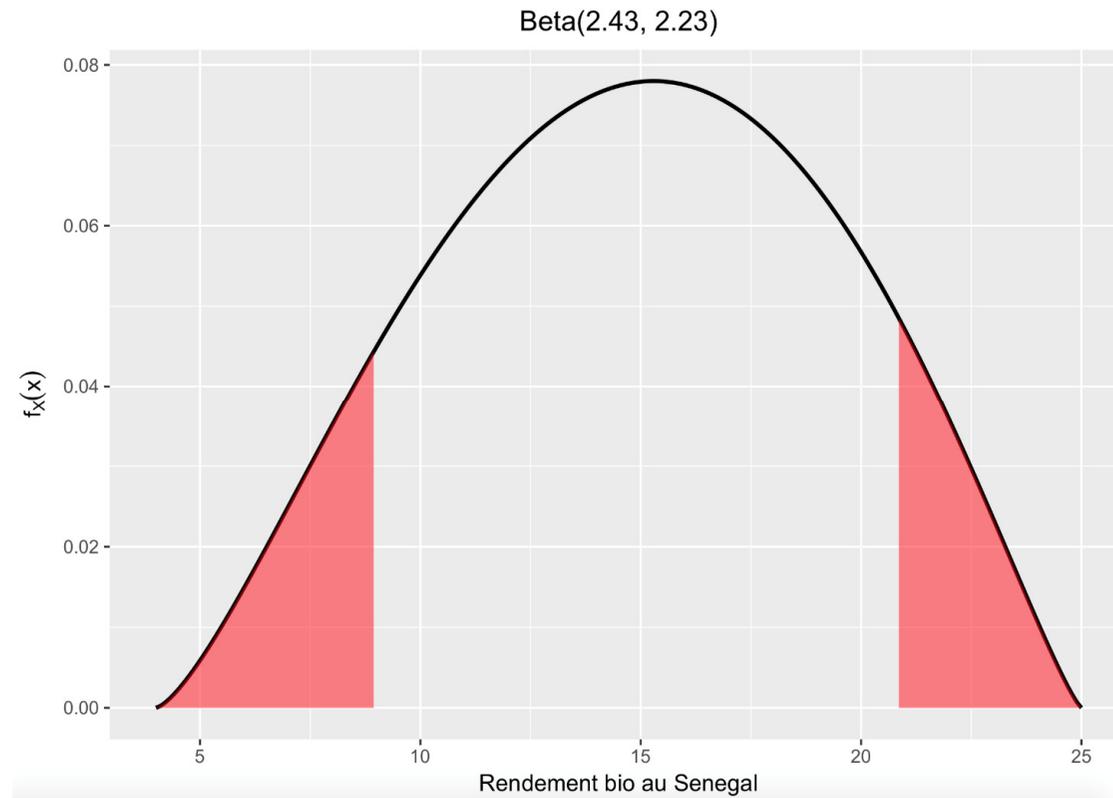
$ssq
  Normal Student-t Gamma Log normal Log Student-t Beta
1 0.0008969088 0.005310216 0.009809939 0.01733336 0.01951732 0.000490555

$best.fitting
 best.fit
1 Beta
```

```
> feedback(myfit, quantiles=c(0.33,0.5,0.66))
```

```
$fitted.quantiles
```

	Normal	Student-t	Gamma	Log normal	Log Student-t	Beta
0.33	12.9	13.1	12.8	12.8	13.0	12.8
0.5	15.0	15.0	14.7	14.5	14.6	15.0
0.66	16.9	16.7	16.7	16.5	16.4	17.1



Méthodes d'élicitation

- Roulette
- **Quartiles**
- **Tertiles**

Méthode « quartiles/tertiles »

- i. Définir des valeurs min et max
- ii. Demander à l'expert de définir les valeurs des
 - Quartiles (0.25, 0.75) et la médiane
 - Tertiles (0.33, 0.66) et la médiane.
- iii. Ajuster une loi de probabilité aux valeurs
- iv. Faire valider le résultat de l'ajustement par l'expert.

Elicitation

Parameter limits

4, 25

Parameter values

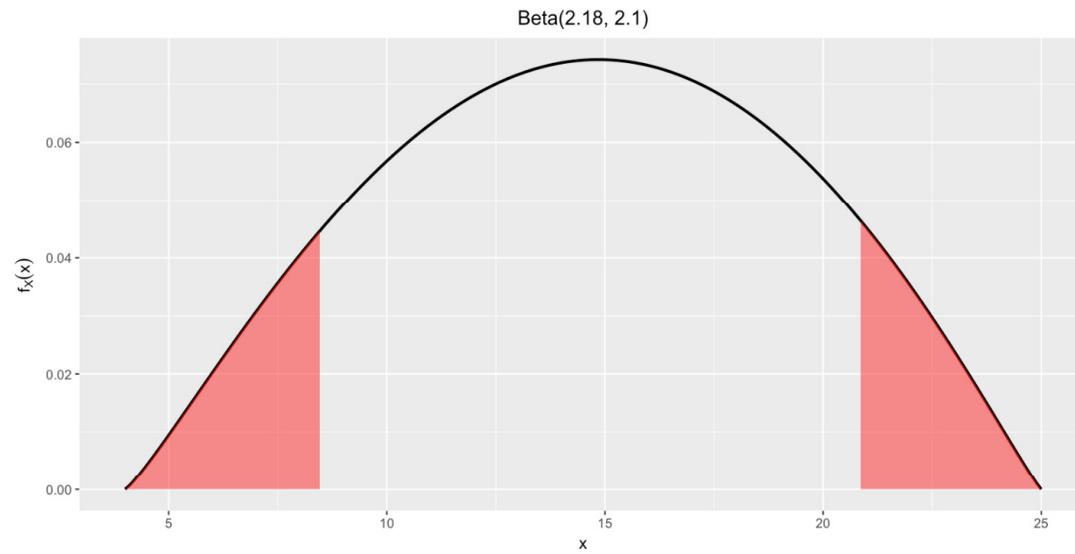
11, 15, 18

Cumulative probabilities

0.25, 0.5, 0.75

Distribution

- Histogram
- Normal
- Student t
- Gamma
- Log normal
- Log Student t
- Beta
- Best fitting



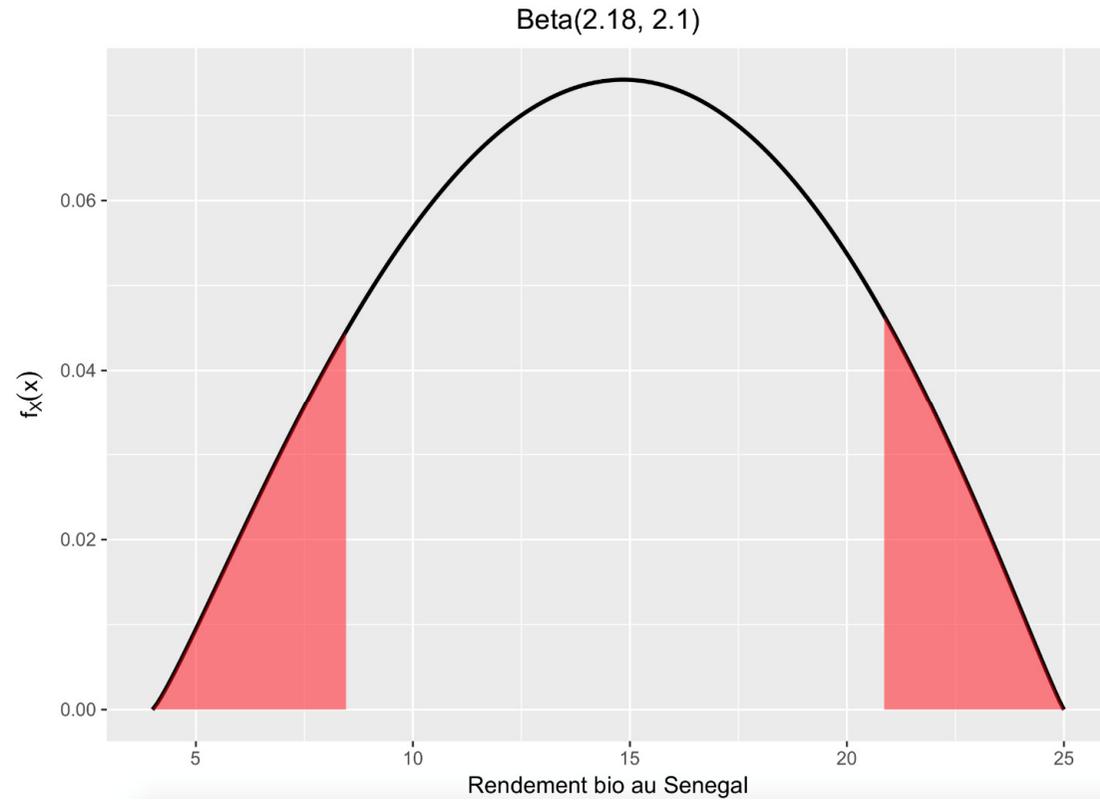
quantiles	values
0.10	8.46
0.90	20.90

```
myfit<- fitdist(vals=c(11,15,18), probs=c(0.25,0.5,0.75), lower=4, upper=25)
myfit
plotfit(myfit, ql=0.1,qu=0.9, xlab="Rendement bio au Senegal")
feedback(myfit, quantiles=c(0.33,0.5,0.66))
```

```
> feedback(myfit, quantiles=c(0.33,0.5,0.66))
```

```
$fitted.quantiles
```

	Normal	Student-t	Gamma	Log normal	Log Student-t	Beta
0.33	12.4	12.5		12.3	12.2	12.3
	12.4					
0.5	14.7	14.8		14.5	14.4	14.4
	14.7					
0.66	16.9	16.8				
	16.9					



Que faire avec plusieurs experts
(pas d'accord entre eux) ?

Deux approches

- Méthode 1 : Elicitation d'un consensus

Le groupe d'experts est élicité collectivement, et une distribution unique est obtenue.

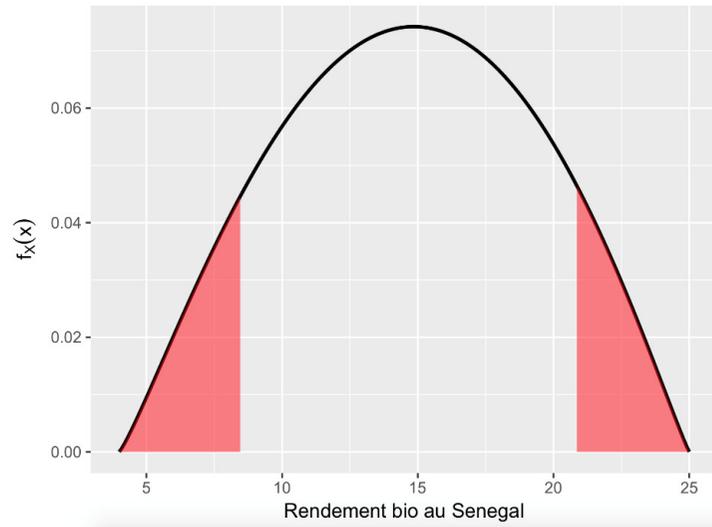
- Méthode 2 : Combiner des distributions individuelles

Une distribution est élicitée pour chaque expert, puis les distributions individuelles sont combinées.

Percentiles	Expert A	Expert B	Expert C
25%	11 kg ha ⁻¹	9 kg ha ⁻¹	13 kg ha ⁻¹
50%	15 kg ha ⁻¹	11 kg ha ⁻¹	19 kg ha ⁻¹
75%	18 kg ha ⁻¹	13 kg ha ⁻¹	20 kg ha ⁻¹

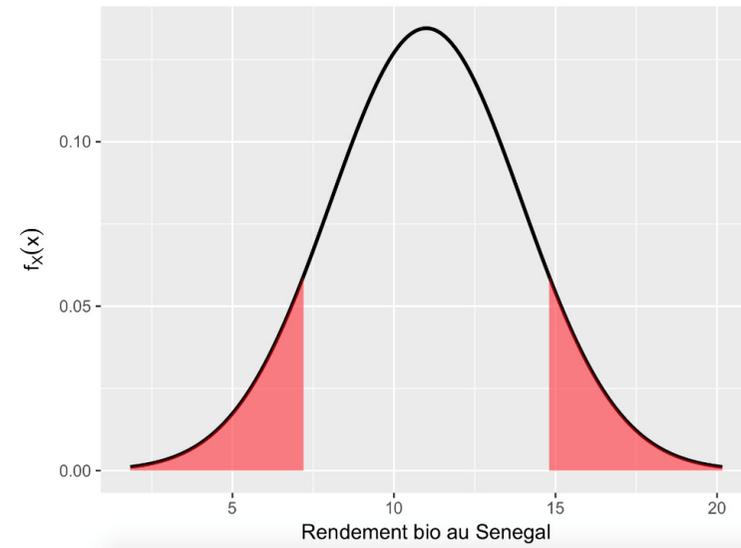
EXPERT 1

Beta(2.18, 2.1)



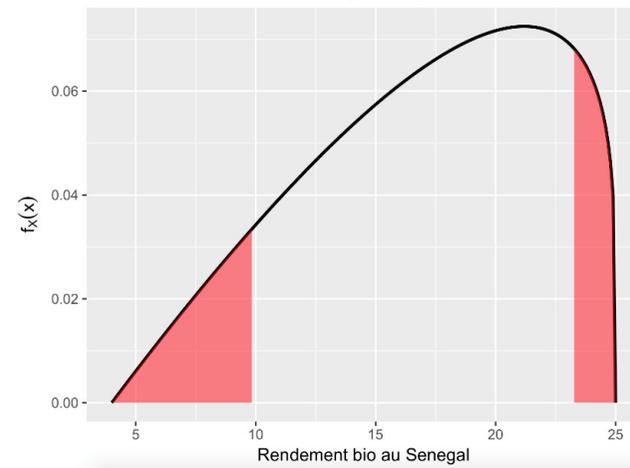
EXPERT 2

Normal (mean = 11, sd = 2.97)



EXPERT 3

Beta(2, 1.22)



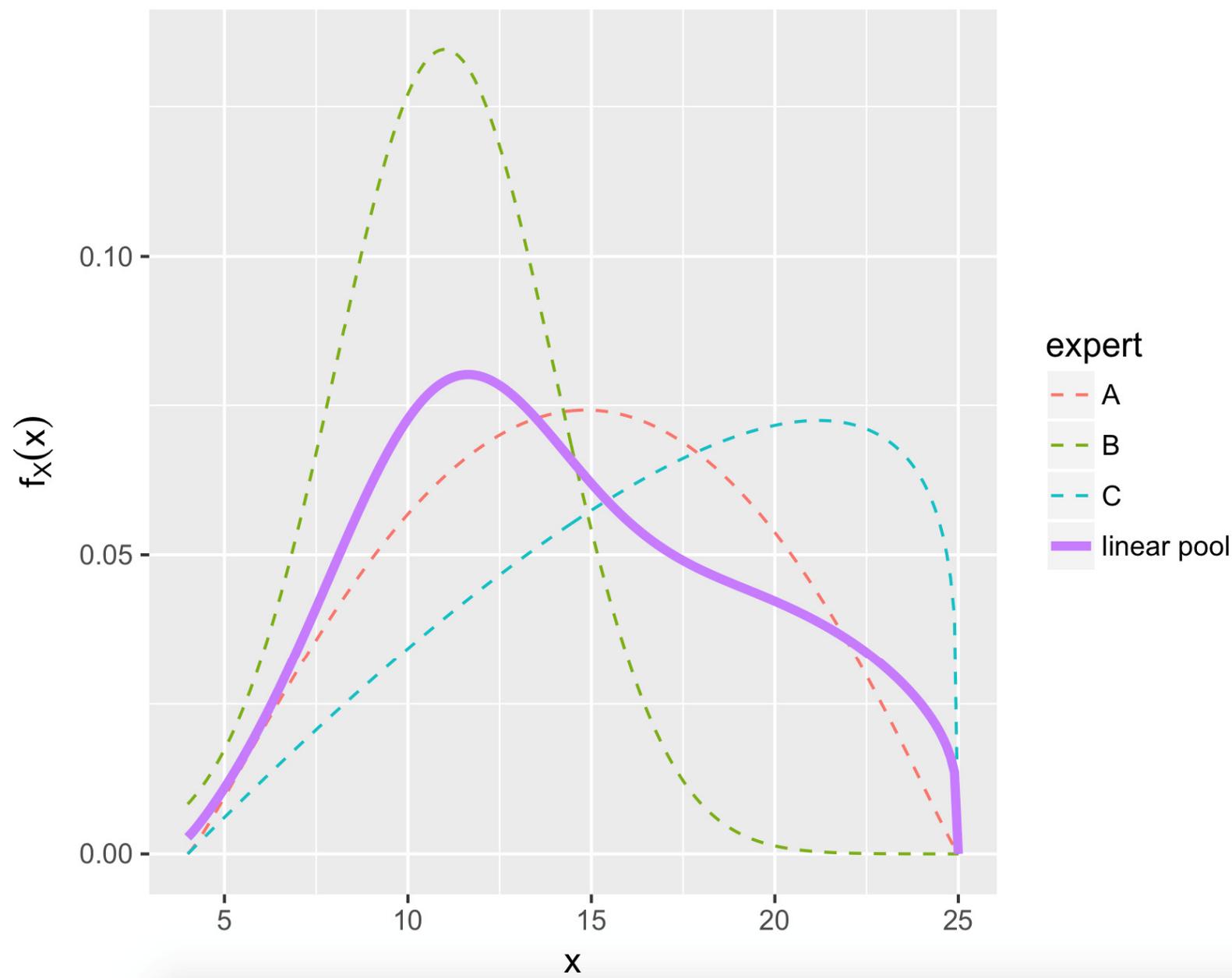
```
v <- matrix(  
c(11, 15, 18,  
  9, 11, 13,  
 13, 19, 20),  
3, 3)
```

```
p <- c(0.25, 0.5, 0.75)
```

```
myfit <- fitdist(vals = v, probs = p, lower = 4,  
upper = 25)  
myfit
```

```
plotfit(myfit, lp = T)
```

```
plinearpool(myfit, x=c(5, 10, 15))
```



A quoi ça sert ?

- Utilisation directe en cas de manque de données
- Analyse d'incertitude
- Définition de distributions a priori en statistique bayésienne

Deux principaux logiciels sur le « marché »

Logiciel	Avantage	Inconvénient
MATCHTOOL	<ul style="list-style-type: none">• Accessible depuis internet• Aucune installation de logiciel• Convivial• Gratuit• Sessions communes• Plusieurs méthodes• Bien documenté	<ul style="list-style-type: none">• Pas utilisable en local• Mono-expert• Une seule quantité élicitée à la fois• Réutilisation des résultats difficile
SHELF	<ul style="list-style-type: none">• Utilisable en local• Multi-experts• Peut éliciter plusieurs quantités• Gratuit• Plusieurs méthodes• Lien direct avec R• Bien documenté	<ul style="list-style-type: none">• Pas accessible depuis internet• Nécessite une installation de logiciel• Pas convivial (programmation)• Réutilisation des résultats difficile

Présentation des logiciels d'inférence bayésienne

Matthieu Authier, Emily Walker, Sophie Ancelet

19 mars 2019, Biobayes

Plan

- ▶ Introduction à l'inférence bayésienne et aux MCMC
- ▶ Outils et logiciels
- ▶ Prise en main WinBUGS / JAGS / Stan
- ▶ Comparaison WinBUGS / JAGS / Stan

Introduction à l'inférence bayésienne et aux MCMC

Loi a posteriori (cf. cours Eric) :

$$[\theta|y] \propto [y|\theta] \times [\theta]$$

- ▶ Ce que l'on cherche à quantifier en bayésien :
 - ▶ la loi a posteriori

$$f(\theta | Y) = \frac{f(Y | \theta)\pi(\theta)}{\int_{\Theta} f(Y | \alpha)\pi(\alpha)d\alpha}$$

- ▶ et ses caractéristiques : moments a posteriori, maximum a posteriori, quantiles a posteriori, intervalles de crédibilité...

Introduction à l'inférence bayésienne et aux MCMC

- ▶ Ce que l'on cherche à quantifier en bayésien :
 - ▶ la loi a posteriori

$$f(\theta | Y) = \frac{f(Y | \theta)\pi(\theta)}{\int_{\Theta} f(Y | \alpha)\pi(\alpha)d\alpha}$$

- ▶ et ses caractéristiques : moments a posteriori, maximum a posteriori, quantiles a posteriori, intervalles de crédibilité...
- ▶ Mais la loi a posteriori peut être difficilement calculable
 - ▶ Modèle avec nombreux paramètres et variables latentes :

$$f(\theta_1, \dots, \theta_K | Y) = \frac{f(Y | \theta_1, \dots, \theta_K)\pi(\theta_1, \dots, \theta_K)}{\int_{\Theta_1} \dots \int_{\Theta_K} f(Y | \theta_1, \dots, \theta_K)\pi(\alpha_1, \dots, \alpha_K)d\alpha_1 \dots d\alpha_K}$$

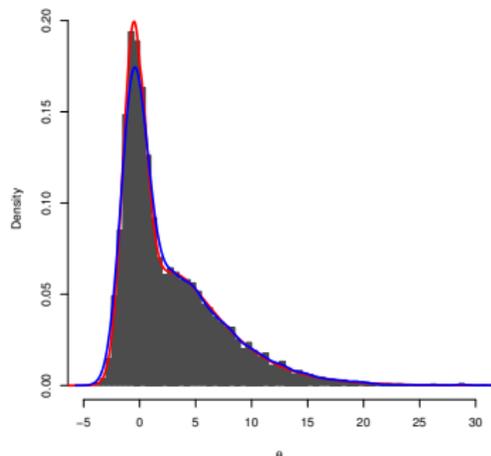
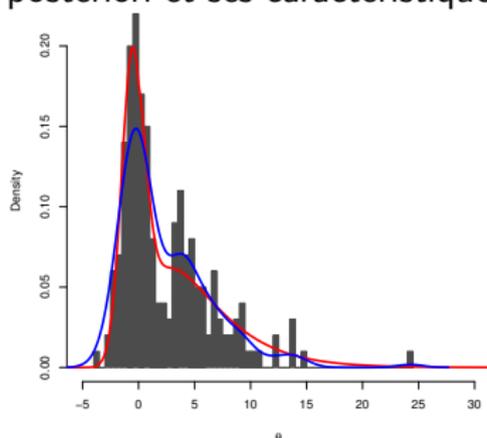
- ▶ Les intégrales multiples (grande dimension) rendent difficile le calcul de la posterior jointe $f(\theta_1, \dots, \theta_K | Y)$, des posteriors marginales $f(\theta_k | Y)$, des moments a posteriori $E(\theta_i^q | Y)$...

Introduction à l'inférence bayésienne et aux MCMC

- ▶ Méthodes numériques pour :
 - ▶ générer un échantillon issu de la loi a posteriori
 - ▶ sans passer par le calcul d'intégrales multiples

Algorithme de Monte Carlo par chaînes de Markov (MCMC) par exemple

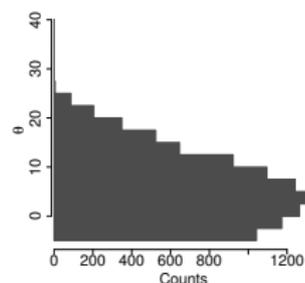
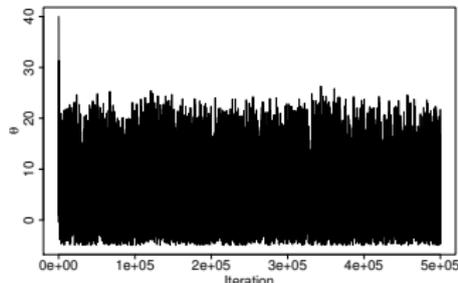
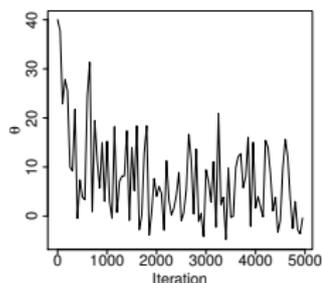
- ▶ Générer un grand échantillon pour correctement approcher la loi a posteriori et ses caractéristiques



Echantillons de taille 200 (histo gauche) et 10000 (histo droite)
Densités a posteriori vraies (rouge) et estimées (bleu)

MCMC : Présentation

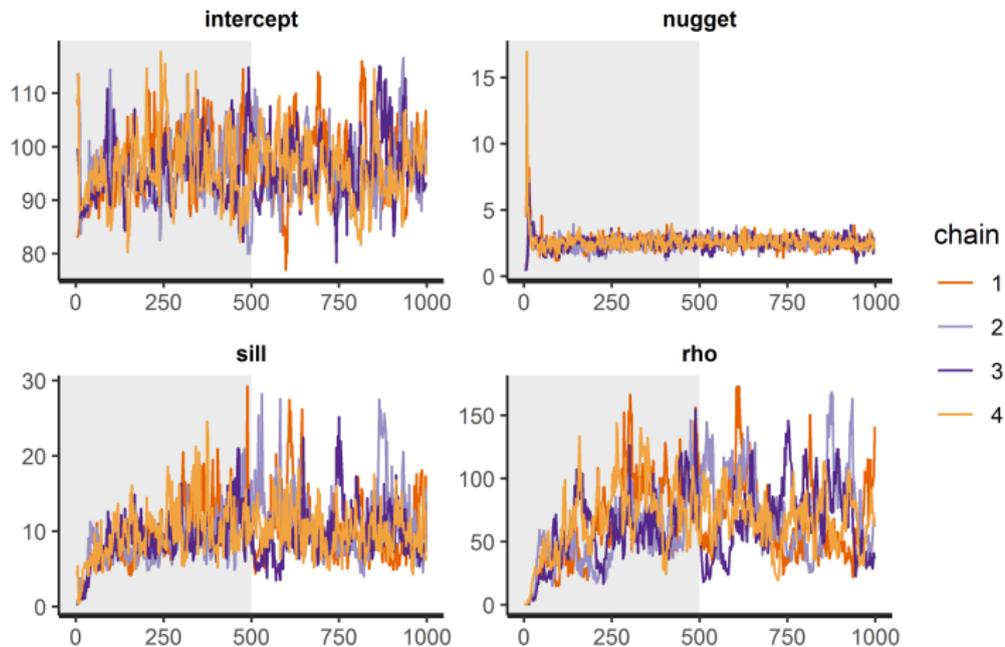
- ▶ Méthodes de Monte Carlo par Chaînes de Markov
- ▶ Algorithmes séquentiels : une séquence de réalisations dépendantes (i.e. une chaîne) de θ est générée
- ▶ Exploration ciblée de l'espace des paramètres (et des variables latentes)
- ▶ Qu'est-ce qu'une chaîne ?



cf. cours Samuel Soubeyrand jour 3

MCMC : Présentation

chaînes à problème...



Logiciels

- ▶ MCMC (Markov Chain Monte Carlo) : **WinBUGS**, OpenBUGS, **JAGS** (appelé depuis R avec package rjags) , Nimble
- ▶ HMC (Hamiltonian Monte Carlo) : **Stan** (appelé depuis R avec package rstan)
- ▶ INLA (Integrated Nested Laplace Approximations) : R-INLA
- ▶ PMC (Population Monte Carlo) : BIIPS, Nimble

Logiciels pendant l'école chercheurs

- ▶ WinBUGS : BUGS écrit en Pascal (sur Windows seulement), interface clique-bouton
- ▶ JAGS : Just Another Gibbs Sampler (Martyn Plummer), package R **rjags**
- ▶ Stan : Stanislas Ulaw, co-inventeur des méthodes de Monte Carlo, package R **rstan**

Logiciel	WinBUGS	JAGS	Stan	ABC
Package	R2winBUGS	rjags	rstan	rABC
Convivialité	R/interface +++	R ++	R +	R/interface ++
Algorithme	MCMC MHwGibbs	MCMC MHwGibbs	HamiltonienMC	
Activité	↘	→	↗	→

Prise en main des 3 logiciels

- ▶ WinBUGS (version interface + version R2WinBUGS)
- ▶ JAGS
- ▶ Stan

Trois étapes

1. modèle bayésien
2. simulation de données
3. faire tourner le modèle

Prise en main des 3 logiciels

1. modèle bayésien : script du modèle BUGS

```
model{  
  #Poisson model script  
  for (i in 1:N) {  
    y[i] ~ dpois(lambda)  
  }  
  log(lambda) <- theta  
  theta ~ dnorm(0, .0001)  
  ynew ~ dpois(lambda)  
}
```

- ▶ Twiddles symbol \sim for a stochastic relation
- ▶ Left arrow \leftarrow for a deterministic relation
- ▶ Hash key $\#$ for comments
- ▶ Arrays are indexed by terms within square brackets.

Prise en main des 3 logiciels

1. modèle bayésien
2. simulation de données

```
set.seed(1234)  
N <- 100  
y <- rpois(N, 10)
```

3. faire tourner le modèle

Comparaison WinBUGS / JAGS / Stan

Logiciel	WinBUGS	JAGS (rjags)	Stan (rstan)
Manuel	BUGS book	≤ 100 pages	≥ 600 pages
Rédaction du code	non structuré	non structuré	très structuré
Paramétrisation	$y \leftarrow \text{dnorm}(\text{mu}, \text{sd})$	$y \leftarrow \text{dnorm}(\text{mu}, 1/(\text{sd}*\text{sd}))$	$y = \text{normal}(\text{mu}, \text{sd});$

Comparaison WinBUGS / JAGS / Stan

Logiciel	WinBUGS	JAGS (rjags)	Stan (rstan)
Manuel Rédaction du code	BUGS book non structuré	≤ 100 pages non structuré	≥ 600 pages très structuré
Paramétrisation	$y \leftarrow \text{dnorm}(\text{mu}, 1/(\text{sd}*\text{sd}))$	$y \leftarrow \text{dnorm}(\text{mu}, 1/(\text{sd}*\text{sd}))$	$y = \text{normal}(\text{mu}, \text{sd});$
ALGO	Gibbs Metropolis HMC MLE ODE	oui oui non non oui	oui oui non non non
			non non oui oui mais ? oui mais ?

Comparaison WinBUGS / JAGS / Stan

Logiciel	WinBUGS	JAGS (rjags)	Stan (rstan)
Manuel Rédaction du code Paramétrisation	BUGS book non structuré $y \leftarrow \text{dnorm}(\mu, 1/(\text{sd}*\text{sd}))$	≤ 100 pages non structuré $y \leftarrow \text{dnorm}(\mu, 1/(\text{sd}*\text{sd}))$	≥ 600 pages très structuré $y = \text{normal}(\mu, \text{sd});$
ALGO	Gibbs Metropolis HMC MLE ODE	oui oui non non oui	non non oui oui mais ? oui mais ?
Message d'erreur Complexité du modèle Temps de calcul	(nombre d'effets aléatoires)	TRAP (debug = TRUE) lent burn-in	pas très précis plus ou moins rapide burn-in
			précis plus ou moins rapide warm-up

Comparaison WinBUGS / JAGS / Stan

Comparaison sur le modèle Bernoulli

Comparaison sur le modèle Normal

Exemple 1 : données binaires

- ▶ vraies valeurs : $\mu = 0$ & $sd = \frac{\pi^2}{3}$
- ▶ Modèle

$$\mu \sim \text{Normale}(0, \text{prior}_{\text{prec}})$$

$$\text{prior}_{\text{prec}} = 1/(1.5 * 1.5)$$

$$\text{prec} = \frac{1}{sd^2}$$

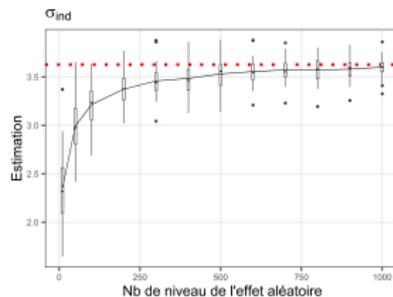
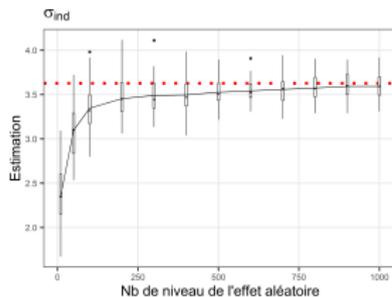
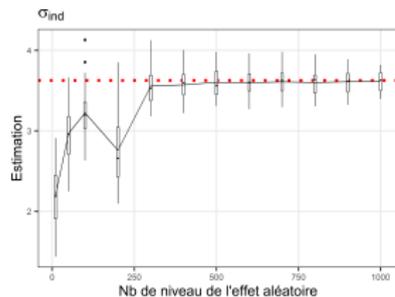
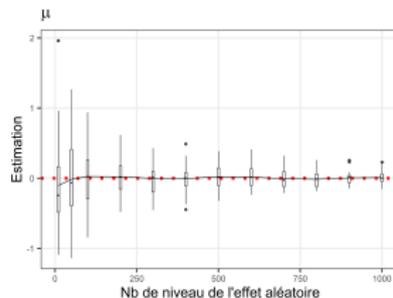
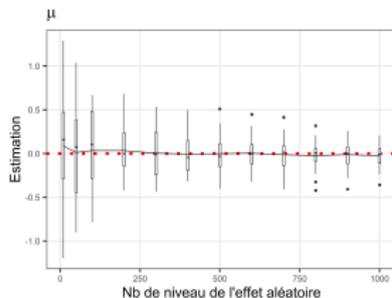
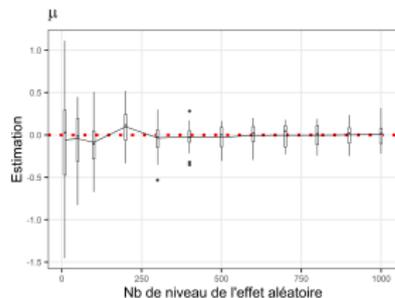
$$sd \sim \text{Normale}(0, 1)$$

$$\text{precision}_{\text{res}} = \frac{1}{sd_{\text{res}}^2}$$

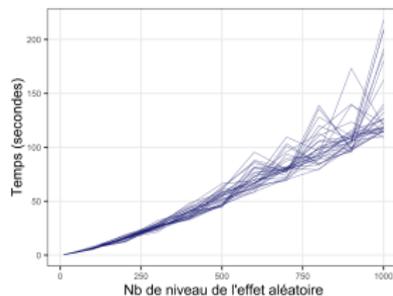
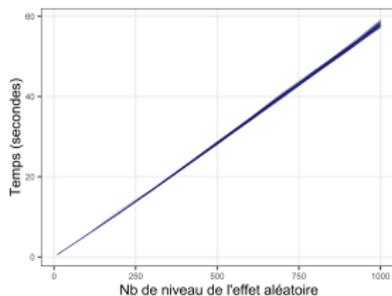
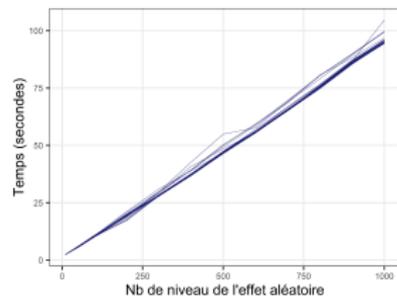
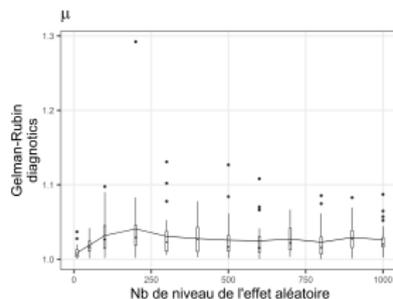
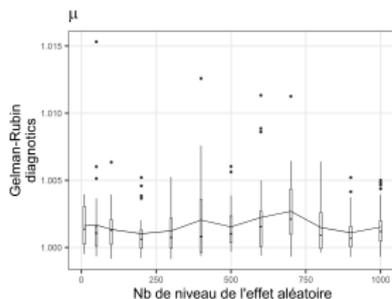
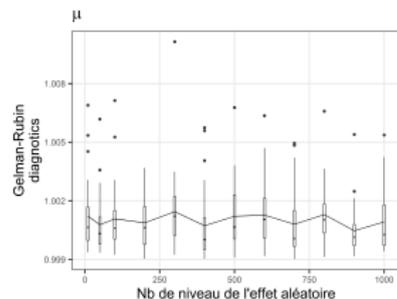
$$\text{for } i \text{ in } 1 : n_{\text{ind}} \alpha_j \sim \text{Normale}(\mu, \text{prec})$$

$$\text{for } j \text{ in } 1 : n_{\text{obs}} y_j \sim \text{Bernoulli}(\text{logit}(\alpha[\text{ind}[j]]))$$

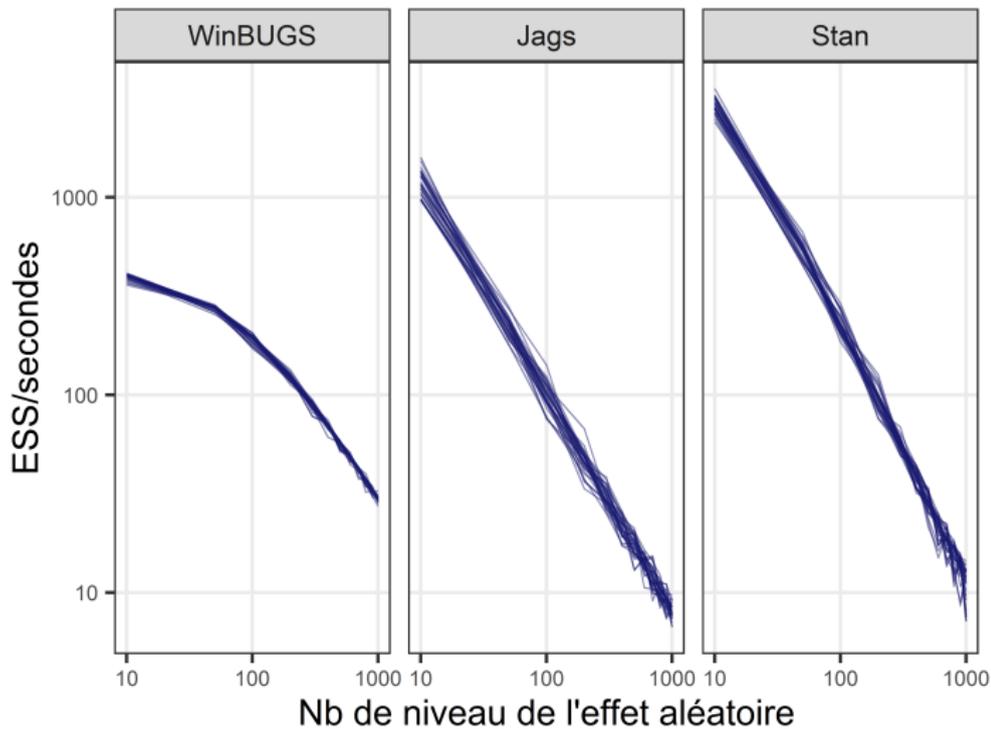
Comparison WinBUGS / JAGS / Stan (Bernoulli)



Comparaison WinBUGS / JAGS / Stan (Bernoulli)



Comparaison WinBUGS / JAGS / Stan (Bernoulli)



Exemple 2 : données Normales

- ▶ vraies valeurs : $\mu = 0$ & $sd = 2$
- ▶ Modèle

$$\mu \sim \text{Normale}(0, 1)$$

$$prec = \frac{1}{sd^2}$$

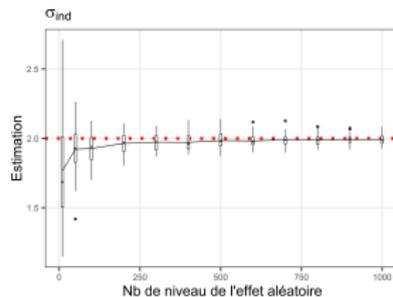
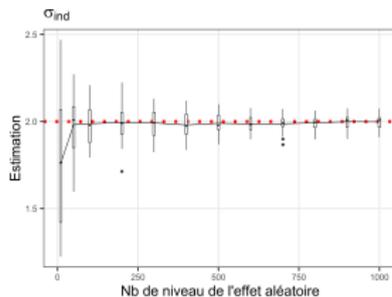
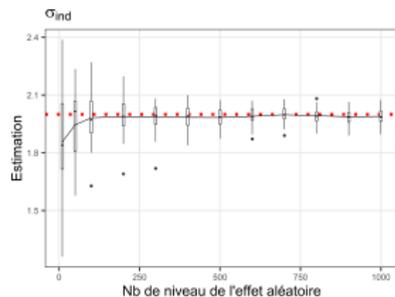
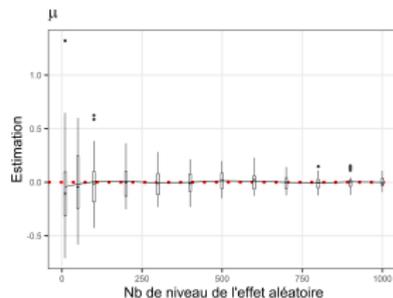
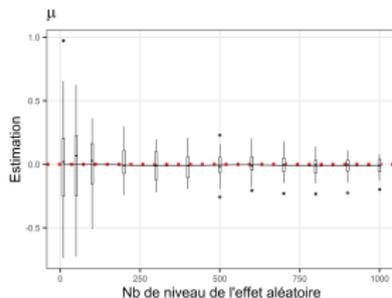
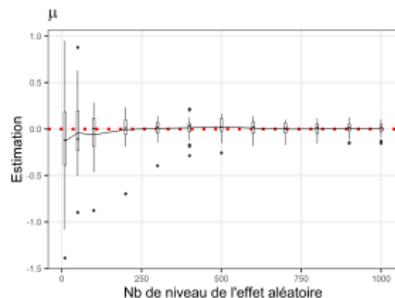
$$sd \sim \text{Normale}(0, 1)$$

$$precision_{res} = \frac{1}{sd_{res}^2}$$

$$\text{for } i \text{ in } 1 : n_{ind} \alpha_i \sim \text{Normale}(\mu, prec)$$

$$\text{for } j \text{ in } 1 : n_{obs} y_j \sim \text{Normale}(\alpha[ind[j]], prec_{res})$$

Comparison WinBUGS / JAGS / Stan (Normal)



Logiciels / packages R

<https://sites.google.com/site/doingbayesiandataanalysis/software-installation>

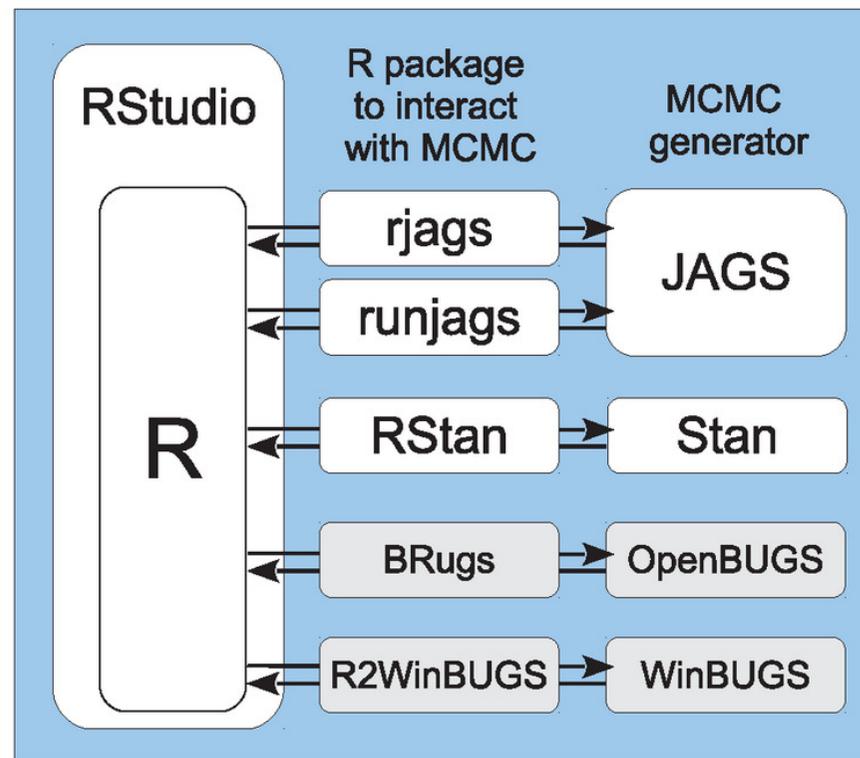
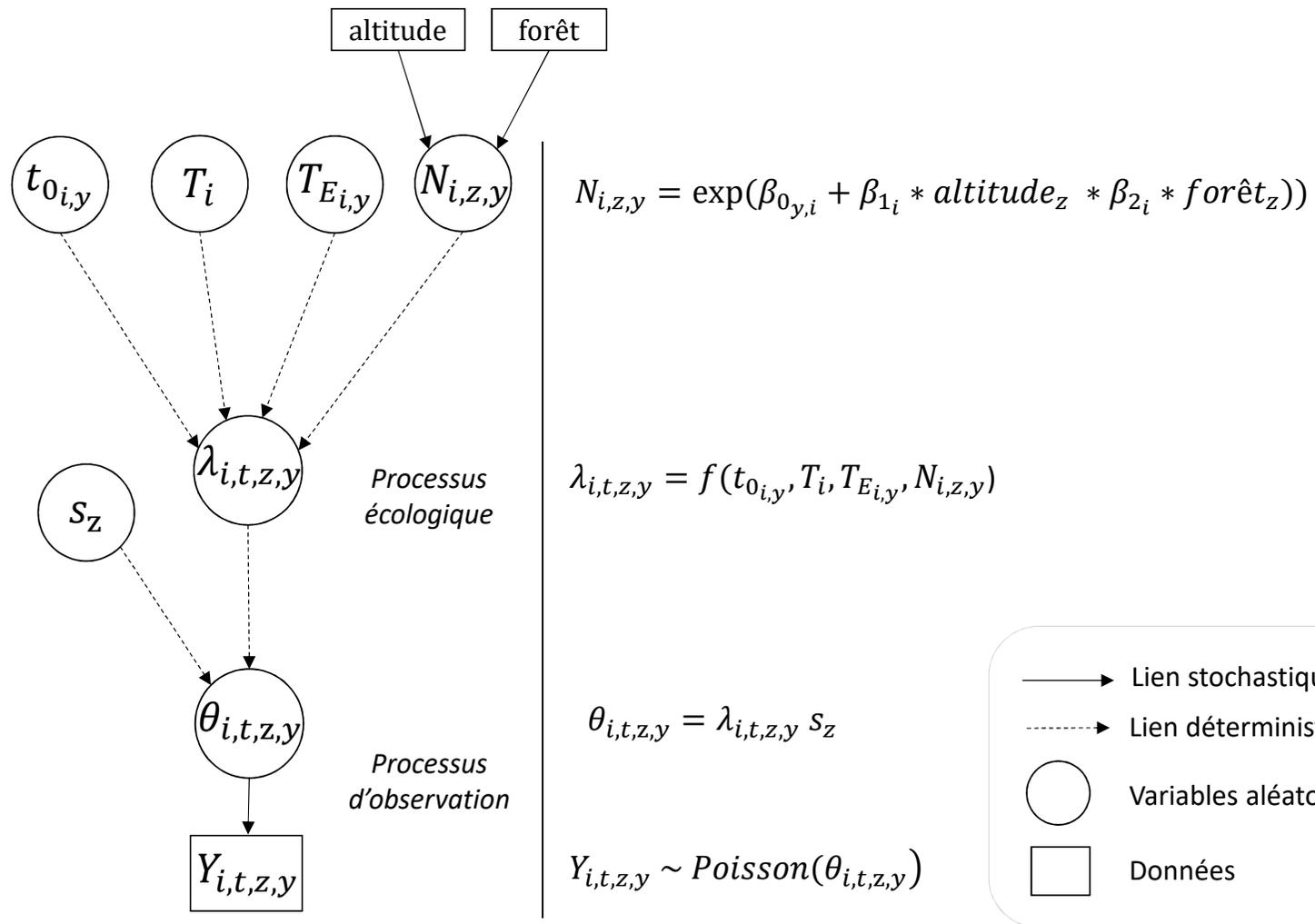


Figure 8.1: Relation of R programming language to other software tools. On the left, RStudio is an editor for interacting with R. The items on the right are various programs for generating MCMC samples of posterior distributions. The items in the middle are packages in R that interact with the MCMC generators. Copyright © Kruschke, J. M. (2014). Doing Bayesian Data Analysis: A Tutorial with R, Wiley.

Méca-stat

- Approche mécanistico-statistique : couplage entre un modèle mécaniste et l'estimation statistique des paramètres
- Outil « maison » : MSE (méca-stat environnement)
<https://mse.mathnum.inrae.fr/>



Bayesian Analysis Reporting Guidelines (BARG)

Table 1 | List of key reporting points for the BARG

Preamble

- A. *Why Bayesian.* If the audience requires it, explain what benefits will be gleaned by a Bayesian analysis (as opposed to a frequentist analysis).
- B. *Goals of analysis.* Explain the goals of the analysis. This prepares the audience for the type of models to expect and how the results will be described.

Step 1. Explain the model

- A. *Data variables.* Explain the dependent (predicted) variables and independent (predictor) variables.
- B. *Likelihood function and parameters.* For every model, explain the likelihood function and all the parameters, distinguishing clearly between parameters of primary theoretical interest and ancillary parameters. If the model is multilevel, be sure that the hierarchical structure is clearly explained, along with any covariance structure if multivariate parameter distributions are used.
- C. *Prior distribution.* For every model, explain and justify the prior distribution of the parameters in the model.
- D. *Formal specification.* Include a formal specification (mathematical or computer code) of the likelihood and prior, located either in the main text or in a publicly and persistently accessible online supplementary material.
- E. *Prior predictive check.* Especially when using informed priors but even with broad priors, it is valuable to report a prior predictive check to demonstrate that the prior really generates simulated data consistent with the assumed prior knowledge.

Step 2. Report details of the computation

- A. *Software.* Report the software used, including any specific added packages or plugins.
- B. *MCMC chain convergence.* Report evidence that the chains have converged, using a convergence statistic such as PSRF, for every parameter or derived value.
- C. *MCMC chain resolution.* Report evidence that the chains have high resolution, using the ESS, for every parameter or derived value.
- D. *If not MCMC.* If using some computational procedure other than MCMC, be aware of and report inherently inaccurate approximations, especially for the limits of credible intervals.

Step 3. Describe the posterior distribution

- A. *Posterior predictive check.* Provide a posterior predictive check to show that the model usefully mimics the data.
- B. *Summarize posterior of variables.* For continuous parameters, derived variables and predicted values, report the central tendency and limits of the credible interval. Explicitly state whether you are using density-based values (mode and HDI) or quantile-based values (median and ETI), and state the mass of the credible interval (for example, 95%).
- C. *BF and posterior model probabilities.* If conducting model comparison or hypothesis testing, report the BF and posterior probabilities of models for a range of prior model probabilities.

Step 4. Report decisions (if any) and their criteria

- A. *Why decisions?* Explain why the decisions are theoretically meaningful and which decision procedure is being used. Regardless of which decision procedure is used, if it addresses null values, it should be able to accept the null value not only reject it.
- B. *Loss function.* If utilities and a loss function for a decision rule are defined, these should be explained and reported.
- C. *ROPE limits.* If using a continuous-parameter posterior distribution as the basis for decision, state and justify the limits of the ROPE and the required probability mass.
- D. *BF, decision threshold and model probabilities.* If using model comparison or hypothesis testing as the basis for a decision, state and justify the decision threshold for the posterior model probability, and the minimum prior model probability that would make the posterior model probability exceed the decision threshold.
- E. *Estimated values too.* If deciding about null values, always also report the estimate of the parameter value (central tendency and credible interval).

Step 5. Report sensitivity analysis

- A. *For broad priors.* If the prior is intended to be vague or only mildly informed so that it has minimal influence on the posterior, show that other vague priors produce similar posterior results.
- B. *For informed priors.* If the prior is informed by previous research, show what posterior results from a vague prior or from a range of differently informed priors.
- C. *For default priors.* If using a default prior, show the effect of varying its settings. Be sure that the range of default priors constitutes theoretically meaningful priors, and consider whether they mimic plausible empirically informed priors.
- D. *BFs and model probabilities.* If the analysis involves model comparison or hypothesis testing, then for each prior report not only the BFs but also the posterior model probabilities for a