Sciences Economiques & Sociales de la Santé
& Traitement de l'Information Médicale

www.sesstim-orspaca.org

Dr. Eric-André SAULEAU
Laboratoire de Biostatistique et Informatique Médicale

Faculté de Médecine
Université de Strasbourg

Introduction à la culture Bayésienne - Élicitation des lois a priori

mars 2015

Introduction à la culture bayésienne
L'élicitation des lois *a priori*

Erik-André Sauleau
ea.sauleau@unistra.fr

Université de Strasbourg
Laboratoire de Biostatistique - Faculté de Médecine
EA 3430 « Progression tumorale et micro-environnement »
LabEx IRMIA « Institut de Recherche en Mathématiques, ses Interactions, ses Applications »

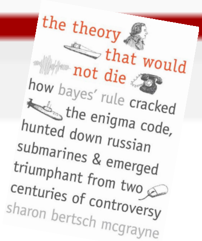
Hôpitaux Universitaires de Strasbourg
Pôle de Santé Publique - Santé au Travail - Hygiène hospitalière

Webinar QuanTIM - 13 mars 2015

Plan

Plan

Contexte général



- Ouvrage récent « grand public »

  McGrayne S. *The theory that would not die*. London: Yale University Press, 2011.

→ Sous-titre « how Bayes' rule cracked the Enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy »

- Réponse de statisticiens sur la pratique de l'inférence bayésienne :
  - « si c'est pour mettre des priors non informatifs, autant faire de l'inférence fréquentiste »
  - « Malheureusement, la science est influencée par certaines modes (ou Zeitgeist) »

La probabilité inverse

Notation : $\text{Pr}(\mathcal{A})$ est la probabilité de survenue d'un événement $\mathcal{A}$

- $\mathcal{A}_i$ est une cause de $\mathcal{B}$
- $\text{Pr}(\mathcal{B}|\mathcal{A}_i)$ est la probabilité que $\mathcal{B}$ se produise quand $\mathcal{A}_i$ est présent
- L'intérêt porte sur $\text{Pr}(\mathcal{A}_i|\mathcal{B})$: quand $\mathcal{B}$ s'est produit, quelle est la probabilité que $\mathcal{A}_i$ en soit la cause ?
- Problème du XVIII^{ième} siècle
- Exemple de Jacob Bernoulli (1713)

Bernoulli J. *Ars conjectandi*. Basel: Thurnisiorum, 1713.

Théorème de Bayes selon Bayes

- Notations
  - On note $\mathcal{A}$ et $\mathcal{B}$ deux événements
  - Probabilité de réalisation de $\mathcal{A}$: $\Pr(\mathcal{A})$
  - Probabilité de réalisation de $\mathcal{A}$ et $\mathcal{B}$: $\Pr(\mathcal{A} \cap \mathcal{B})$
  - Probabilité conditionnelle $\Pr(\mathcal{A}|\mathcal{B}) = \frac{\Pr(\mathcal{A} \cap \mathcal{B})}{\Pr(\mathcal{B})}$
-  $\leftarrow$ ceci n'est pas Thomas Bayes

$$\boxed{\Pr(\mathcal{A}|\mathcal{B}) = \frac{\Pr(\mathcal{B}|\mathcal{A}) \times \Pr(\mathcal{A})}{\Pr(\mathcal{B})}}$$

Théorème de Bayes selon Laplace

- Notations
  - Les causes possibles de $\mathcal{B}$ sont $\mathcal{A}_1, \ldots, \mathcal{A}_i, \ldots, \mathcal{A}_n$
-  $\leftarrow$ PS Laplace (1774)

$$\boxed{\Pr(\mathcal{A}_i|\mathcal{B}) = \frac{\Pr(\mathcal{B}|\mathcal{A}_i) \times \Pr(\mathcal{A}_i)}{\sum_n \Pr(\mathcal{B}|\mathcal{A}_i) \times \Pr(\mathcal{A}_i)}}$$

Inférence bayésienne

- Notations
  - $\mathcal{A}$: paramètre(s) $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$
  - $\mathcal{B}$: données D
- Application du théorème $\rightarrow$ $\mathrm{Pr}(\boldsymbol{\theta}|D) \propto \mathrm{Pr}(D|\boldsymbol{\theta}) \times \mathrm{Pr}(\boldsymbol{\theta})$
  - $\mathrm{Pr}(\boldsymbol{\theta}|D)$: loi des paramètres *a posteriori* avec les données
  - $\mathrm{Pr}(D|\boldsymbol{\theta})$: vraisemblance des données
  - $\mathrm{Pr}(\boldsymbol{\theta})$: loi des paramètres *a priori* avant les données

posterior	$\propto$	vraisemblance	$\times$	prior
$\pi(\boldsymbol{\theta}\|D)$	$\propto$	$\mathcal{L}_{\boldsymbol{\theta}}(D)$	$\times$	$\mathrm{p}(\boldsymbol{\theta})$

Bayésien vs fréquentiste : deux visions différentes

L'inférence bayésienne ne renie pas la vraisemblance

- Vraisemblance utilisée pour les estimations *a posteriori*
- Renier la vraisemblance $\Rightarrow$ renier les données
- *A posteriori*, balance entre prior et données

Exemple Estimation de la moyenne d'une loi normale

$$w\mu_0 + (1-w)\frac{\sum y}{n}$$

- Somme pondérée de la moyenne *a priori* μ_0 et de la moyenne des données $\bar{y}$
- Poids w dépend de la taille de l'échantillon et des variances σ_0^2 et σ^2

Bayésien vs fréquentiste : deux visions différentes

- Sur l'inférence scientifique donc sur l'incertitude
  - Fréquentiste : outils basés sur l'expérimentation (vraisemblance)
  - Bayésien : la seule réponse est probabiliste
- Sur la définition d'une probabilité
  - Fréquentiste : fréquence $\lim_{n \to +\infty} \left(\frac{m}{n} \right)$
  - Bayésien : degré raisonnable de croyance

$\rightarrow$ Bayes, Ramsay, de Finetti

Le terme de *culture*

- Difficile d'utiliser les termes tels que
  - École
  - Fréquentisme et bayésianisme
  - Théorie
  - Paradigme

    Kuhn T. *La structure des révolutions scientifiques*. Flammarion, Paris, 1983.
- Préférer culture $\rightarrow$ référence à un ensemble de travaux communs
- Pratique ?

Plan

Application du théorème de Bayes à l'inférence

$$\boxed{\pi(\boldsymbol{\theta}|D) \propto \mathcal{L}_{\boldsymbol{\theta}}(D) \times \mathrm{p}(\boldsymbol{\theta})}$$

- $\pi(\boldsymbol{\theta}|D)$: loi des paramètres *a posteriori* avec les données
- $\mathcal{L}_{\boldsymbol{\theta}}(D)$: vraisemblance des données
- $\mathrm{p}(\boldsymbol{\theta})$: loi des paramètres *a priori* avant les données

Toute la difficulté et la beauté de l'inférence bayésienne est la détermination de cette loi *a priori* $\Rightarrow$ travail conjoint clinicien (expert) - statisticien

= **élicitation**

Définition (Wikipedia) : incitation d'un locuteur à un autre à statuer sur différentes hypothèses

Retour sur la compréhension bayésienne des probabilités

degré raisonnable de croyance

- Propriété minimale : cohérence → loi de probabilité
- Deux courants
  1. Bayésien subjectif : la cohérence suffit
  2. Bayésien objectif : d'autres propriétés sont nécessaires → deux agents avec la même connaissance produiront la même élicitation

exemple : principe de la raison insuffisante

A partir de maintenant nous serons des bayésiens objectifs

Deux axes pour l'élicitation

1. Choix de la forme fonctionnelle = famille de la loi
   - La loi *a priori* est conjuguée pour la vraisemblance si la loi *a posteriori* est de la même classe que la loi *a priori*

Vraisemblance	Paramètre	Prior conjugué
Bernoulli		Beta, Uniforme
Binomiale		Beta, Uniforme
Normale	Moyenne	Normal
	Précision	Gamma
Poisson	Moyenne	Gamma

   - Contraignant, quelle réalité pour les données ?
   - Ce n'est plus une nécessité (McMC) mais accélère la convergence

Deux axes pour l'élicitation

1. Choix de la forme fonctionnelle = famille de la loi
2. Choix des paramètres de la loi (hyperparamètres)

Analyse de sensibilité : vérifier l'effet de différents priors sur les estimations

Estimation d'une fréquence : position du problème

- Estimer la fréquence de réponse clinique à un an après traitement d'une tumeur rare
- Critère RECIST (Response Evaluation Criteria in Solid Tumors Group)
  - Positif : réponse complète clinique ou réponse partielle clinique
  - Négatif : stabilisation et progression
- 52 patients sont suivis et 39 présentent une réponse clinique
- Réponse clinique observée : 75%
- Estimer la fréquence dans la population = inférence

Estimation d'une fréquence : inférence bayésienne

Situation conjuguée

Données : y réponses cliniques sur n sujets

1. Vraisemblance : $y \sim \text{Bin}\,(\theta, n)$ où θ est la fréquence de réponse clinique
2. Prior sur θ : $\text{p}(\theta) \sim \beta\,(a, b)$
3. Posterior : $\pi(\theta|y) \sim \beta\,(a + y, b + n - y)$

Il reste à spécifier les paramètres a et b de la loi *a priori*

Élicitation : généralités

- Le clinicien a la connaissance, le biostatisticien doit la transformer pour en faire la distribution *a priori* (forme et hyperparamètres)
- Sources
  - Littérature : articles, méta-analyses
  - Jugement d'expert(s)
    - Expert unique
    - Conférence de consensus
- Deux difficultés au jugement d'un expert
  1. Donner des valeurs numériques à un jugement subjectif
  2. Nombre fini de jugements $\rightarrow$ distribution unique
- Deux résultats
  1. Élicitation non informative
  2. Élicitation informative

Élicitation non informative

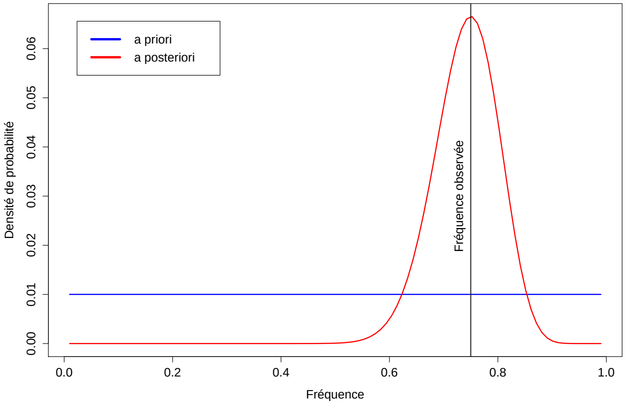
- Si on n'a pas d'information préalable $\Rightarrow$ prior non informatif
- Principe de la raison insuffisante : quand les données ne fournissent aucun argument pour préférer une valeur a_1 à une autre a_2, alors $\mathrm{p}(a_1) = \mathrm{p}(a_2)$

  Laplace PS. *Mémoire sur la probabilité des causes par les événements*. Imprimerie Royale, Paris, 1774.

- Toutes les valeurs sont équiprobables *a priori* $\rightarrow$ loi uniforme
- Fisher (1930), citant Boole : « equal distribution of ignorance »

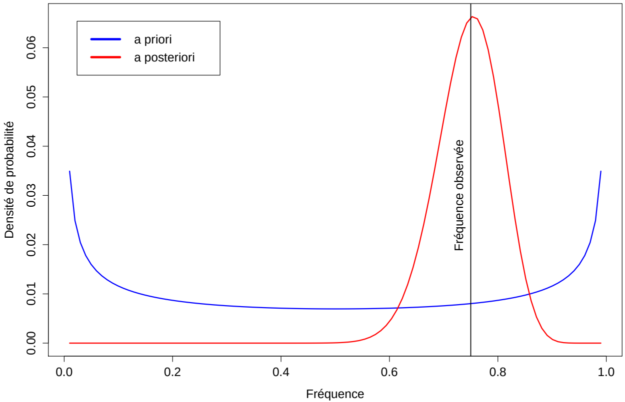
Élicitation non informative

$$\theta_{\text{obs}} = 0,75 = 39/52, \text{ prior plat } \mathsf{p}(\theta) \sim \beta(a = 1, b = 1)$$



Élicitation non informative

$\theta_{\text{obs}} = 0,75 = 39/52$, prior de Jeffreys $\mathsf{p}(\theta) \sim \beta(a = 0.5, b = 0.5)$



Élicitation informative

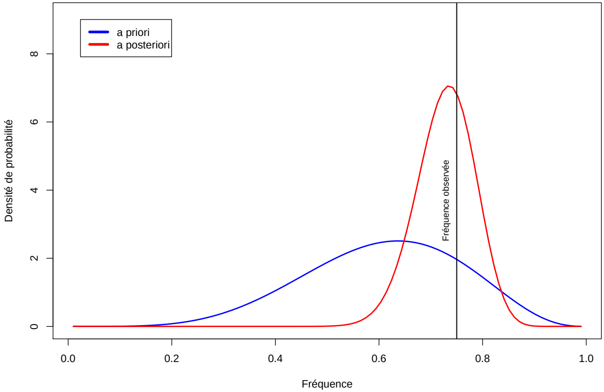
- Si on a de l'information préalable $\Rightarrow$ prior informatif

  Exemple de la fréquence et de la loi beta *a priori*
- Deux paramètres à déterminer (a et b) $\Leftrightarrow$ deux questions à poser
  - Quelle est la probabilité p_1 que la fréquence soit supérieure à une valeur K_1 ?
  - Quelle est la probabilité p_2 que la fréquence soit inférieure à une valeur K_2 ?

$\rightarrow$ Les réponses donnent deux percentiles de la loi et permettent de déterminer a et b

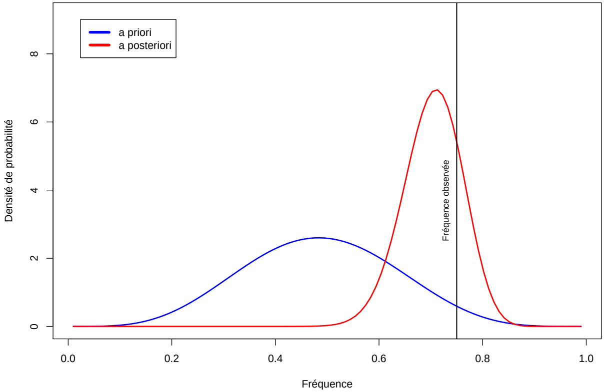
Élicitation différente $\Rightarrow$ posterior différent

$\theta_{\text{obs}} = 0,75 = 39/52$, prior $\mathrm{p}(\theta) \sim \boldsymbol{\beta}(a = 5.9, b = 3.8)$
10% en dessous de 0,40 et 10% au dessus de 0,80



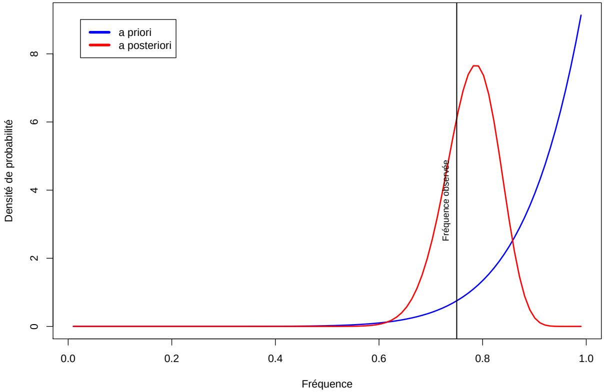
Élicitation différente ⇒ posterior différent

$\theta_{\text{obs}} = 0,75 = 39/52$, prior $\text{p}(\theta) \sim \beta(a = 5.4, b = 5.7)$
10% en dessous de 0,30 et 10% au dessus de 0,70



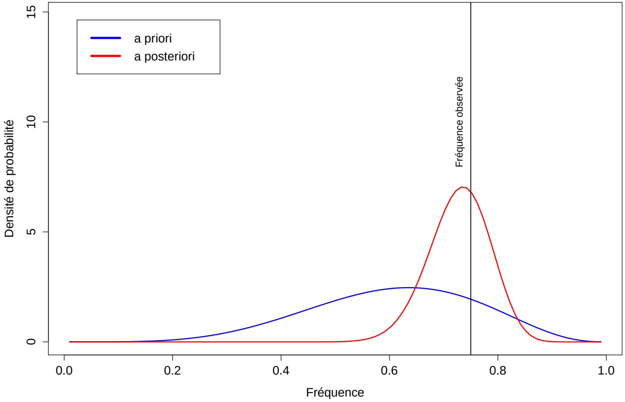
Élicitation différente $\Rightarrow$ posterior différent

$\theta_{\text{obs}} = 0,75 = 39/52$, prior $\text{p}(\theta) \sim \boldsymbol{\beta}(a = 10.0, b = 1.0)$
10% en dessous de 0,80 et 1% au dessus de 0,95



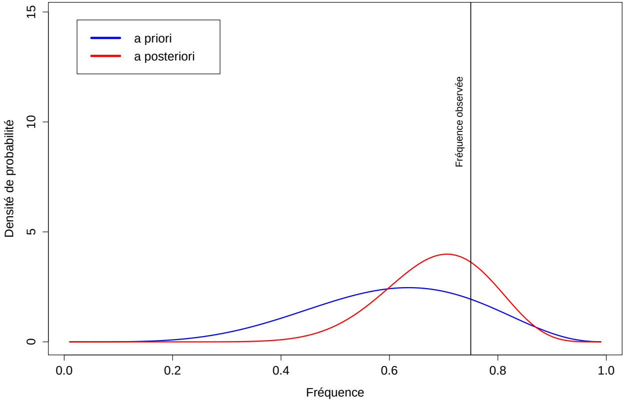
Prior vs données

$$\theta_{\text{obs}} = 0,75 = 39/52, \text{ prior } \mathrm{p}(\theta) \sim \beta(a = 5.9, b = 3.8)$$



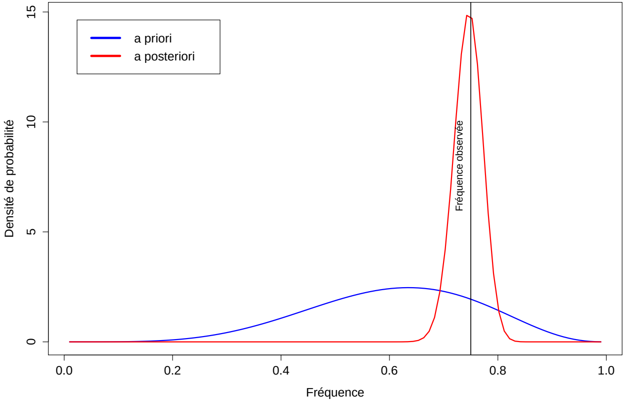
Prior vs données

$$\theta_{\text{obs}} = 0,75 = 9/12, \text{ prior p}(\theta) \sim \beta(a = 5.9, b = 3.8)$$



Prior vs données

$$\theta_{\text{obs}} = 0,75 = 390/520, \text{ prior } \mathrm{p}(\theta) \sim \beta(a = 5.9, b = 3.8)$$



Élicitation : conclusion

- Point clé de l'inférence bayésienne : élicitation
  - Pas de loi *a priori* totalement indiscutable
  - Inclure toute connaissance utile
- $posterior \propto vraisemblance \times prior$
  - $posterior$: biostatisticien et technique
  - $vraisemblance$: données
  - $prior$: expert (clinicien) et facilitateur (biostatisticien)
- Analyse de sensibilité
  - Confrontation des *a posteriori* avec différentes *a priori*
  - « Analyse ascendante » : trouver la loi *a priori* qui aboutit à une conclusion donnée et évaluer la crédibilité de cette loi

Plan

- Inférence bayésienne
- Inférence fréquentiste

Merci de votre attention

Avantages conceptuels du bayésien

1. « Épistémologie naturalisée et évolutionniste »
    - Mise à jour d'une connaissance *a priori*
    - Modèles complexes
2. Le bayésien répond aux questions du clinicien
    - $\theta | D$
    - $\Pr(H_0 | D)$
    - Formulation de H_0
3. En fréquentiste tout est basé sur grands nombres et/ou répétitions : distribution asymptotique, intervalle de confiance, probabilité

Avantages pratiques du bayésien

- Souplesse dans les modèles surtout complexes
- Présentation hiérarchique des modèles
- Données manquantes
- Comparaisons multiples
- Recherche clinique : le bayésien est conceptuellement séquentiel
- Petits échantillons $\rightarrow$ poids de la loi *a priori*

Inconvénients conceptuels du bayésien

-

Inconvénients pratiques du bayésien

- Lourdeur mathématique
- Pas de routine de calcul du nombre nécessaire de sujets
- Problèmes techniques
  - Logiciels
  - Résultats souvent par simulation (McMC) $\Rightarrow$ pas de résultat analytique
  - Lenteur des itérations
- Oblige à savoir exactement ce que l'on fait : prior, modèle, ...

  N'est-ce pas plutôt un avantage ? ☺

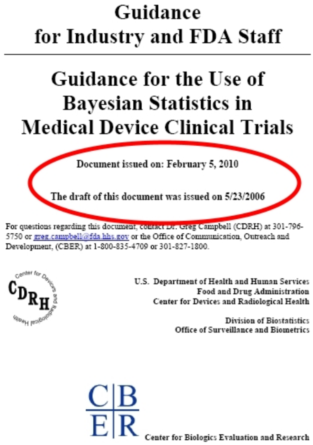
Freins au bayésien

- Lourdeur mathématique
- Réticences au changement de « culture »
- Sentiment de subjectivité
    - Implicite en fréquentiste (α, ...)
    - Explicite en bayésien
- Difficultés de publication (revues cliniques)

Freins au bayésien

Difficultés de publication $\Rightarrow$ il faut convaincre !

- Investigateurs
- Statisticiens
- Éditeurs et reviewers
- Agences de régulation du médicament



**Guidance
for Industry and FDA Staff**

**Guidance for the Use of
Bayesian Statistics in
Medical Device Clinical Trials**

Document issued on: February 5, 2010

The draft of this document was issued on 5/23/2006

For questions regarding this document, contact Dr. Greg Campbell (CDRH) at 301-796-5750 or greg.campbell@fda.hhs.gov or the Office of Communication, Outreach and Development, (CBER) at 1-800-835-4709 or 301-827-1800.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Division of Biostatistics
Office of Surveillance and Biometrics

Center for Biologics Evaluation and Research

Freins au bayésien

- La culture fréquentiste reste largement dominante ☹
- Prédiction de Bruno de Finetti : en 2020 l'inférence bayésienne prévaudra

Statistique mathématique

Accessible

- Held L, Bové DS. *Applied statistical inference : likelihood and Bayes*. Springer, 2014. (Porte bien son titre)
- Draper D. *Bayesian modeling : inference and prediction*, 2005. (Code R)

Statistique mathématique

Assez accessible

- Christensen R, Johson W, Branscum A, Hanson TE. *Bayesian ideas and data analysis : an introduction for scientists and statisticians*. CRC Press, 2011. (Code R et WinBUGS)
- Lesaffre E, Lawson AB. *Bayesian biostatistics* Wiley, 2012.
- Bolstad WM. *Introduction to Bayesian statistics*. Wiley, 2007. (Historique)
- Gelman A, Carlin JB, Stern AS, Rubin DB. *Bayesian data analysis*. Chapman & Hall/CRC, 2004.
- O'Hagan A, West M. *The Oxford handbook of applied bayesian analysis*. Oxford University Press, 2010. (Chapitre = application pratique)
- Boreux JJ, Parent E, Bernier J. *Pratique du calcul Bayesien*. Springer, 2010.

Statistique mathématique

Éventuellement

- Robert CP. *Le choix Bayesien*. Springer, 2006. (Difficile)
- Parent E, Bernier J. *Le raisonnement Bayesien*. Springer, 2007. (Difficile)
- Bolstad WM. *Understanding computational Bayesian statistics*. Wiley, 2010. (Difficile)
- Bernardo JM, Smith AFM. *Bayesian Theory*. Wiley, 2000. (Historique)

Applications

Accessible, code

- Albert J. *Bayesian computation with R*. Springer, 2007. (Pédagogique, code R)
- Kruschke JK. *Doing Bayesian data analysis : a tutorial with R and BUGS*. Chapman & Hall/CRC, 2010. (Situations pratiques, code R et WinBUGS)
- Ntzoufras I. *Bayesian modeling using WinBUGS*. Wiley, 2009. (Situations pratiques, code WinBUGS)
- Hoff P. *A first course in Bayesian statistical methods*. Springer 2009. (Pédagogique, code R mais limité)

Applications

Accessible, éventuellement ...

- Congdon P. *Applied Bayesian hierarchical methods*. CRC Press, 2010. (Situations pratiques, code R et WinBUGS mais pas pédagogique)
- Congdon P. *Applied Bayesian modelling*. Wiley, 2003. (Situations pratiques mais pas pédagogique)
- Congdon P. *Bayesian statistical modelling*. Wiley, 2006. (Situations pratiques, code WinBUGS mais pas pédagogique)
- Marin JM, Robert CP. *Bayesian essentials with R*. Springer, 2014. (Situations pratiques, code R mais complexe)

Applications

A suivre . . .

Kruschke JK. *Doing Bayesian data analysis : a tutorial with R, JAGS and Stan*. Academic Press/Elsevier, 2014. (Tout juste sorti !)

Livre de chevet

McGrayne SB. *The theory that would not die : how Bayes' rule cracked the enigma code, hunted down russian submarines and emerged triumphant from two centuries of controversy*. Yale UP, 2011.

Merci de votre attention

A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule.