Bayesian Modelling
Numerical Inference
satRday

Bruno HERAULT

Department Forest & Environment

INPHB - Cirad

- your dataset is cleaned
- data stored in a secure database, copied in 3 (or more) places

you are now in front of your computer screen
you open R
you import your dataset
you check it
and you are happy with that

- now you are remembering the hypotheses you want to test
- and you are beginning to realize how complex can be your hypotheses and how complex may be science
- so that you start drawing some convoluted relationships between variables, twisted processes and wawering distorted curves

but..

..
after a while, you type on the keyboard *lm()* or perhaps *glm()*...

..
already resigned to think *linearly*

work program

1. bayesian vs frequentist inference
   - elements of theory
   - application to DBH distribution
2. MCMC numerical methods
   - principles
   - the algorythm
   - Application
3. some comments
4. why bayesian modeling in R

probabilist roots of bayesian statistics

Bayes Theorem

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

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

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

in terms of density

let X and Y be 2 random variables with densities π_X et π_Y

$$\pi_{X|Y}(x|y) = \frac{\pi_{Y|X}(y|x)\pi_X(x)}{\pi_Y(y)}$$

Rev. Thomas Bayes
1702-1761

formalisation for model inference

- your dataset $y_1, \ldots, y_n$
- your dataset are observations from a parametric model

$$Y_1, \ldots, Y_n \sim \pi_{Y|\theta} \quad i.i.d.$$

- the likelihood of this model is $L(y_1, \ldots, y_n; \theta) = \prod_{i=1}^{n} \pi_{Y|\theta}(y_i|\theta)$

frequentist estimation of θ

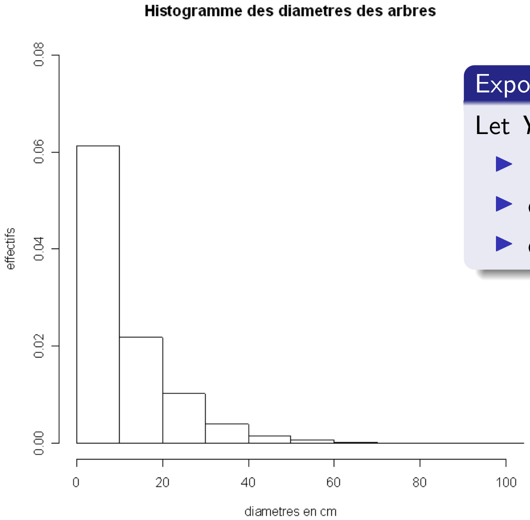
$\hat{\theta}$ maximize $L(y_1, \ldots, y_n; \theta)$

bayesian estimation of θ

- *a priori* law on θ with density π_θ
- calculate the *a posteriori* law

$$\pi_{\theta|\mathbf{Y}}(\theta|y_1, \ldots, y_n) = \frac{L(y_1, \ldots, y_n; \theta)\pi_\theta(\theta)}{\int L(y_1, \ldots, y_n; \theta)\pi_\theta(\theta) d\theta}$$

Diameter distribution in natural forests



Exponential model

Let Y be the DBH of a tree

- law : $Y \sim \exp(\theta)$
- density : $\pi_{Y|\theta}(y|\theta) = \theta \exp(-\theta y)$
- data : $\mathbf{y} = (y_1, \ldots, y_n)$

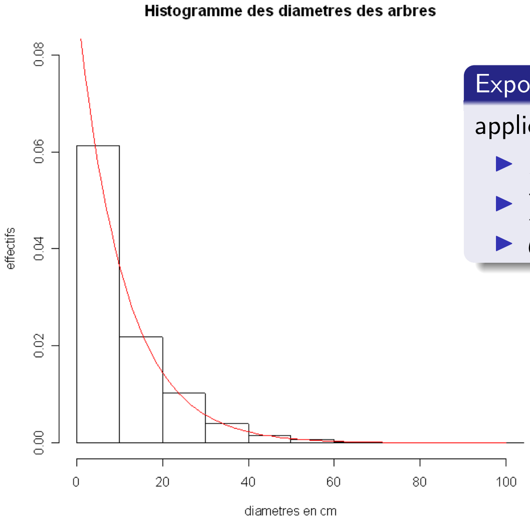
Diameter distribution in natural forests

frequentist inference

- the likelihood is $L(\mathbf{y};\theta) = \prod_{i=1}^{n} \theta \exp(-\theta y_i) = \theta^n \exp(-\theta \sum_{i=1}^{n} y_i)$
- the log-likelihood is $\ell(\theta) = n \log\theta - \theta \sum_{i=1}^{n} y_i$
- θ estimate at maximum likelihood is given by

$$\frac{\partial \ell(\hat{\theta})}{\partial \theta} = 0 \Longleftrightarrow \frac{n}{\theta} - \sum_{i=1}^{n} y_i = 0 \Longrightarrow \hat{\theta} = \frac{n}{\sum_{i=1}^{n} y_i}$$

Diameter distribution in natural forests



Exponential model

application to real data

- $n = 39858$
- $\sum y_i = 434131.1$
- $\hat{\theta} = 0.0918$

Diameter distribution in natural forests

bayesian inference

- the likelihood does not change $L(\mathbf{y};\theta) = \theta^n \exp(-\theta \sum_{i=1}^{n} y_i)$
- we choose an *a priori* law on θ : $\pi(\theta)$
- θ estimate is given by the *a posteriori* law

$$\pi_{\theta|\mathbf{y}}(\theta|y_1,\ldots,y_n) = \frac{L(\mathbf{y};\theta)\pi(\theta)}{\int L(\mathbf{y};\theta)\pi(\theta)\mathrm{d}\theta} \propto L(\mathbf{y};\theta)\pi(\theta)$$

- the choice of the *a priori* law is fundamental

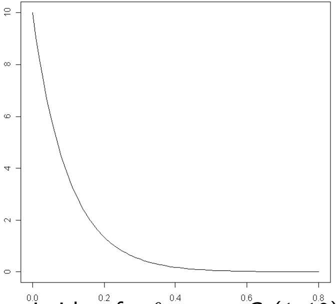
Diameter distribution in natural forests

you have to choose an *a priori* law for θ :

- common laws on $\mathbb{R}^+$ $\rightarrow$ no explicit *a posteriori* law
- conjugated law $\rightarrow$ known *a posteriori* law
- the Gamma law is conjugated to the exponential law

Exponential model

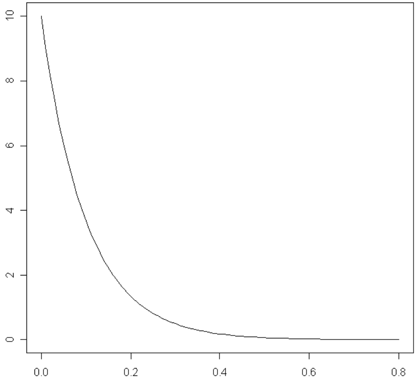
- prior of θ : $\pi_\theta(\theta) = Ga(\theta|\alpha, \beta)$
- posterior of θ :
  $\pi_{\theta|\mathbf{y}} = Ga(\alpha + n, \beta + \sum y_i)$
- posterior of Y :
  $\pi_{Y|\mathbf{y}} = Gg(\alpha + n, \beta + \sum y_i, 1)$



a priori law for θ : $\pi_\theta = Ga(1, 10)$

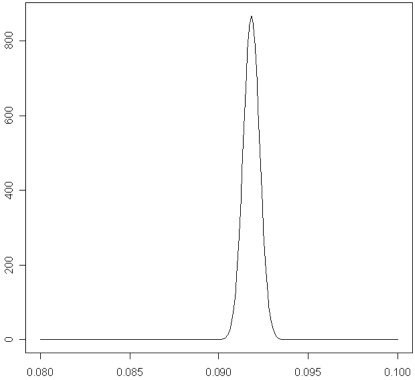
Diameter distribution in natural forests

with real data : $n = 39858$ et $\sum y_i = 434131.1$



prior law for θ :

$\pi_\theta = Ga(1, 10)$



posterior law for θ :

$\pi_{\theta|\mathbf{y}} = Ga(1 + 39858, 10 + 434131.1)$

Diameter distribution in natural forests

with real data : $n = 39858$ et $\sum y_i = 434131.1$

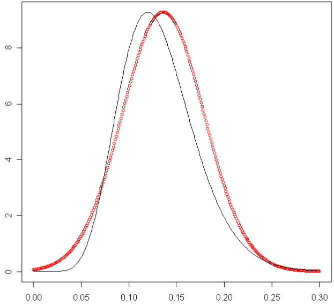
a posteriori law for D : $\pi_{Y|\mathbf{y}} = Gg(1 + 39858, 10 + 434131.1, 1)$

$\rightarrow$ result very close to the frequentist estimate

Diemeter distribution : bayesian ⇔ frequentist

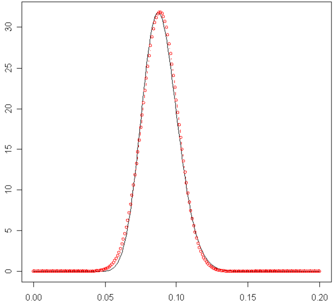
Impact of the sample size n

Sample size $n = 10$



$\pi_{\theta|\mathbf{y}}(\theta|\mathbf{y})$ (–) ⇔ $\mathcal{N}(\hat{\theta}, \frac{\hat{\theta}^2}{n})$ (∘∘)

Sample size $n = 50$



$\pi_{\theta|\mathbf{y}}(\theta|\mathbf{y})$ (–) ⇔ $\mathcal{N}(\hat{\theta}, \frac{\hat{\theta}^2}{n})$ (∘∘)

Comments

- the added values of bayesian inference is greater when the sample size is low
  - supplementary information from the *a priori* law
  - the *a priori* greatly helps inference but may also biased
  - the choice of the *a priori* law is crucial

- practical scheme

$$\left.\begin{array}{c} a\ priori \text{ law} \quad \pi_\theta(\theta) \\ + \\ \text{data} \quad \mathbf{y} \end{array}\right\} \longrightarrow a\ posteriori \text{ law} \quad \pi_{\theta|\mathbf{y}}(\theta|\mathbf{y})$$

- moving from «a priori» $\rightarrow$ «a posteriori» is nearly impossible analytically (excepted for conjugated laws)

$\hookrightarrow$ numerical methods MCMC to infer *a posteriori* laws

work program

Markov Chain Monte Carlo methods

Monte Carlo methods

- inference technique based on the simulation of a high number of random variables

advantages

- may be applied to a wide range of problems
- a few underlying hypothese
- easy to implement

constraints

- a good random generator
- computational power
- likelihood-explicit

Markov Chain Monte Carlo methods

- objective
  - sample a target law of density $\pi(z)$ where Z is a real variable with n-dimensions
- principle
  - we build a Markov $(Z^{(k)})_{k\geq 0}$ chain on $\mathbb{R}^n$ for which the asymptotic law is π
  - when the chain is stationary, we extract a random sample from the chain $(Z^{(k_1)}, \ldots, Z^{(k_N)})$ in order to estimate the parameter *a posteriori* distribution

Metropolis-Hastings (MH) sampler

- objective : build a Markov with π invariant
- ingredients :
  1. a proposal law/kernel π^{prop}
  2. a measurable function $\alpha \mapsto [0,1]$.
- principle :
  1. starting from $z \in E$, π^{prop} give a new candidate $\tilde{z} \in E$.
  2. this new candidate is accepted with probability $\alpha(z, \tilde{z})$, if not we conserve z.
- in practice
  - the proposal law/kernel π^{prop} is chosen by you
  - the MH algorythm determine α

the algorythm

Given π the density of the target law, π^{prop} the density of the proposal law/kernel

- initialisation : set $z^{(0)}$
- iteration t
  1. Given $z^{(t-1)}$, sample $\tilde{z} \sim \pi^{prop}(z^{(t-1)})$
  2. $\alpha(z^{(t-1)}, \tilde{z}) = \min\left(\dfrac{\pi(\tilde{z})/\pi^{prop}(z^{(t-1)}, \tilde{z})}{\pi(z^{(t-1)})/\pi^{prop}(\tilde{z}, z^{(t-1)})}, 1\right)$
  3. $z^{(t)} = \begin{cases} \tilde{z} & \text{with probability } \alpha \\ z^{(t-1)} & \text{with probability } 1-\alpha \end{cases}$

Application to DBH distribution

the exponential model

model law	$\pi_{Y\|\theta}(y) = \exp(\theta)$
likelihood	$L(\mathbf{y};\theta) = \theta^n \exp(-\theta \sum_{i=1}^n y_i)$
a priori law	$\pi_\theta(\theta) = Ga(\theta, 1, 10) = 10 \exp(-10\,\theta)$
a posteriori law	$\pi_{\theta\|\mathbf{y}} = Ga(1+n, 10+\sum y_i)$

imagine we do not know that $\pi_{\theta|\mathbf{y}} = Ga(1+n, 10+\sum y_i)$

$$\pi_{\theta|\mathbf{y}}(\theta|y_1,\ldots,y_n) = \frac{L(y_1,\ldots,y_n;\theta)\pi_\theta(\theta)}{\int L(y_1,\ldots,y_n;\theta)\pi_\theta(\theta)\mathrm{d}\theta} = \frac{\theta^n \exp(-\theta\sum_{i=1}^n y_i)\times 10\exp(-10\,\theta)}{\int \theta^n \exp(-\theta\sum_{i=1}^n y_i)\times 10\exp(-10\,\theta)\mathrm{d}\theta}$$

$$\propto \theta^n \exp\left(-\theta(10+\textstyle\sum_{i=1}^n y_i)\right)$$

however you know that

$$\pi_{\theta|\mathbf{y}}(\theta|y_1,\ldots,y_n) \propto \theta^n \exp\left(-\theta(10+\textstyle\sum_{i=1}^n y_i)\right)$$

here we are : infering the model with MCMC

building the Markov chain $\theta^{(1)}, \dots, \theta^{(t)}, \dots$ with asymptotic law $\pi_{\theta|\mathbf{y}}$

- the target law : $\pi_{\theta|\mathbf{y}}(\theta|y_1, \dots, y_n) \propto \theta^n \exp\{-\theta(10 + \sum_{i=1}^n y_i)\}$
- the proposal law : gaussian random walk $\tilde{\theta} \sim \mathcal{N}(\theta^{(t-1)}, \sigma^2)$

$$\pi^{prop}(\theta^{(t-1)}, \tilde{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\frac{1}{2\sigma^2}(\theta^{(t-1)} - \tilde{\theta})^2\}$$

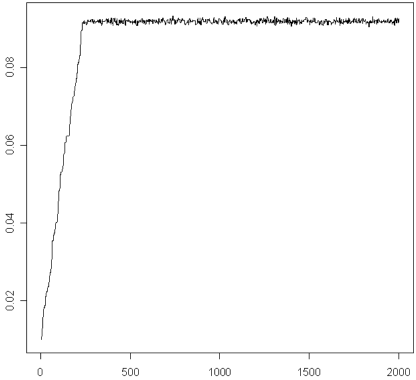
- the Metropolis-Hastings ratio is

$$r = \frac{\tilde{\theta}^n \exp\{-\tilde{\theta}(10 + \sum_{i=1}^n y_i)\}}{(\theta^{(t-1)})^n \exp\{-\theta^{(t-1)}(10 + \sum_{i=1}^n y_i)\}}$$

- updating the chain
  - $u \sim \mathcal{U}[0,\ 1]$
  - $\theta^{(t)} = \begin{cases} \tilde{\theta} & \text{if} \quad u \leq r \\ \theta^{(t-1)} & \text{if not} \end{cases}$

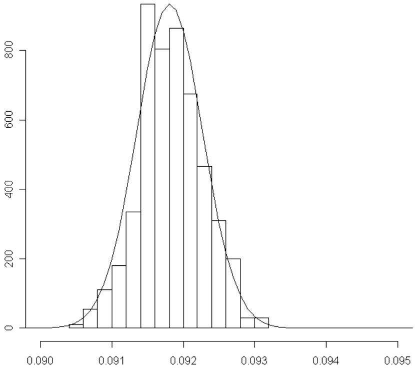
Application to DBH distribution

Markov chain for $\theta^{(t)}$



$\sigma = 0.001$ et $\theta^{(0)} = 0.01$

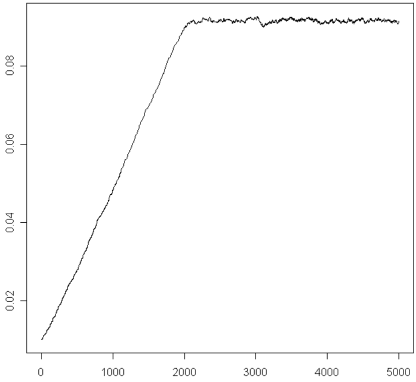
Estimate of $\pi_{\theta|\mathbf{y}}$



Impact of the proposal law

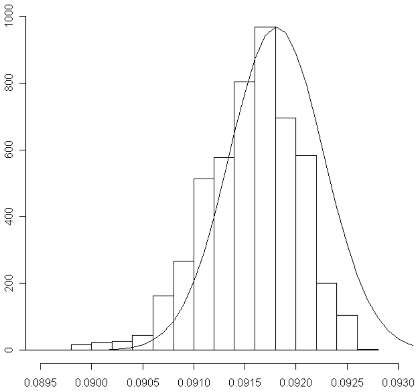
results with $\sigma = 0.0001$, *i.e.* a smaller step

Markov chain of $\theta^{(t)}$



σ = **0.0001 et** $\theta^{(\mathbf{0})}$ = **0.01**

Estimate of $\pi_{\theta|\mathbf{y}}$

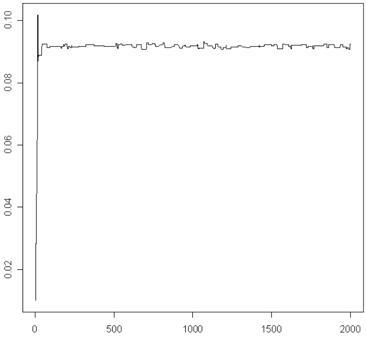


Application to DBH distribution

impact of the proposal law

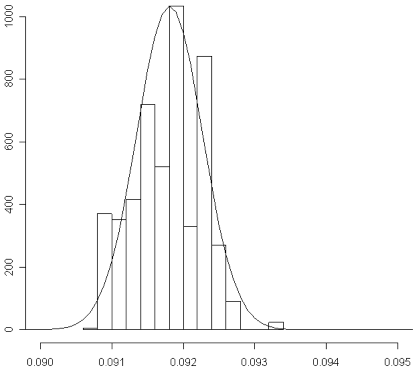
results for $\sigma = 0.01$, *i.e.* a bigger step

Markov chain of $\theta^{(t)}$



σ = **0.01 et** $\theta^{(0)}$ = **0.01**

Estimation de $\pi_{\theta|\mathbf{y}}$



work program

in practice some critical points

- initial values may be critical
- uneasy to optimize the proposal law / to diagnose convergence
- burning time maybe **too** long

other MC algorithms

- Gibbs : sequential updates of model parameters

  iteration $Z^{(k)} \to Z^{(k+1)}$ is $\begin{cases} Z_\ell^{(k+1)} \sim \pi_{\ell|\neg\ell} \\ \text{for } \ell \in \{1,\dots,n\} \end{cases}$

  with $\pi_{\ell|\neg\ell}$ the marginal law of Z_ℓ given $Z_1,\dots,Z_{\ell-1},Z_{\ell+1},\dots,Z_n$
- Metropolis within Gibbs : mixing MH and Gibbs
- Hamiltonian MC : oriented walk
  Implied in Rstan http ://mc-stan.org/

work program

the bayesian future

ECOLOGY LETTERS

Explore this journal >

Why environmental scientists are becoming Bayesians

James S. Clark

First published: 15 December 2004 Full publication history

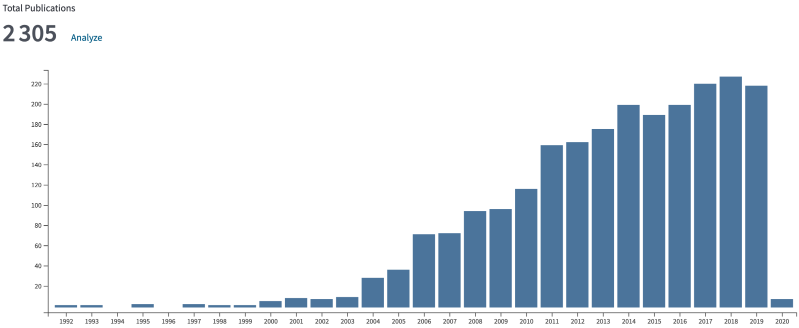
DOI: 10.1111/j.1461-0248.2004.00702.x View/save citation

Cited by: 351 articles Refresh Citing literature

Am score 5

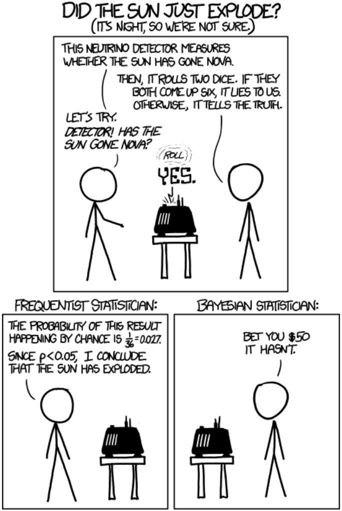
View issue TOC
Volume 8, Issue 1
January 2005
Pages 2–14

the bayesian future

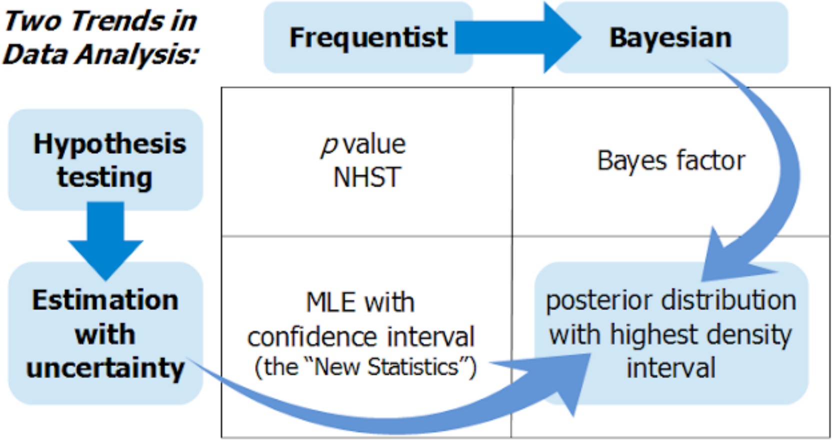


Keywords : Bayes + Africa

the bayesian way of thinking



the bayesian way of thinking



the bayesian advantages

You can

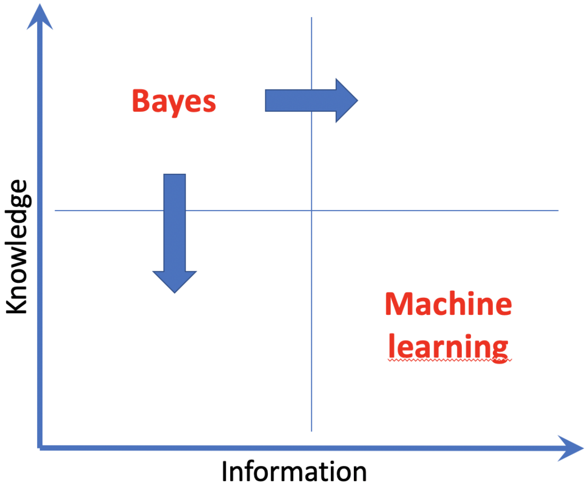
- incorporate prior knowledge
- design your own model
- inspect your posterior distribution
- easily propagate uncertainties, even for complex models
- update your models as soon as you want

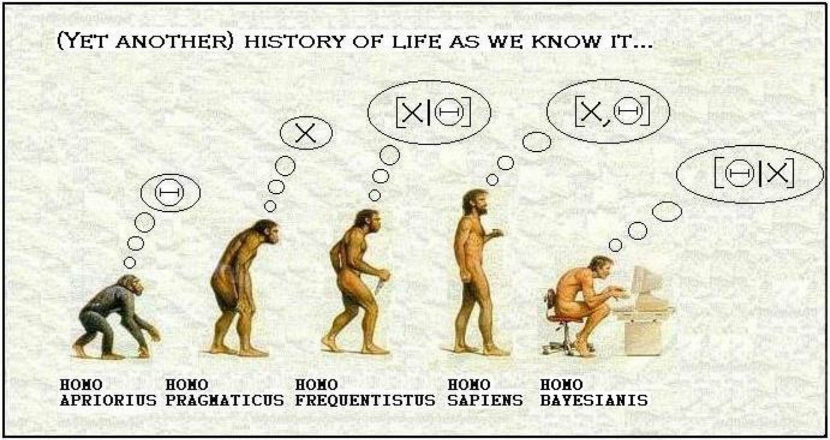
the bayesian advantages

You will

- separate creativity from inference
- build very complex models
- not forget your prior results each time you start modeling
- be useful for decision making
- have fun in modeling

the bayesian niche




(YET ANOTHER) HISTORY OF LIFE AS WE KNOW IT...
[X|Θ]
[X,Θ]
[Θ|X]
X
Θ
HOMO APRIORIUS
HOMO PRAGMATICUS
HOMO FREQUENTISTUS
HOMO SAPIENS
HOMO BAYESIANIS