Getting into Bayesian Wizardry...
(with the eyes of a muggle actuary)

Arthur Charpentier

charpentier.arthur@uqam.ca

http://freakonometrics.hypotheses.org/

R in Insurance, London, July 2014

Getting into Bayesian Wizardry...
(with the eyes of a frequentist actuary)

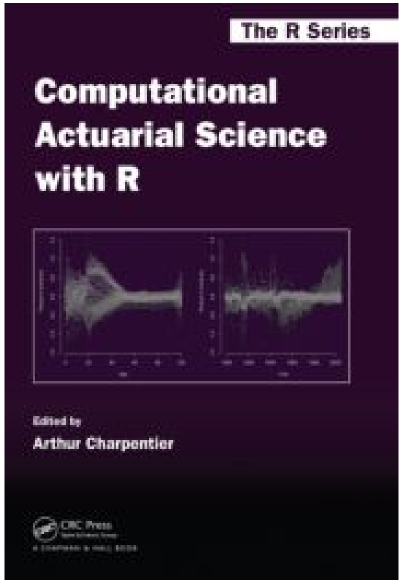
Arthur Charpentier

charpentier.arthur@uqam.ca

http://freakonometrics.hypotheses.org/

R in Insurance, London, July 2014

Professor of Actuarial Sciences, Mathematics Department, UQàM
(previously Economics Department, Univ. Rennes 1 & ENSAE Paristech
actuary in Hong Kong, IT & Stats FFSA)
PhD in Statistics (KU Leuven), Fellow Institute of Actuaries
MSc in Financial Mathematics (Paris Dauphine) & ENSAE
Editor of the freakonometrics.hypotheses.org's blog
Editor of (forthcoming) Computational Actuarial Science, CRC



"it's time to adopt modern Bayesian data analysis as standard procedure in our scientific practice and in our educational curriculum. Three reasons:

1. Scientific disciplines from astronomy to zoology are moving to Bayesian analysis. We should be leaders of the move, not followers.
2. Modern Bayesian methods provide richer information, with greater flexibility and broader applicability than 20th century methods. Bayesian methods are intellectually coherent and intuitive. Bayesian analyses are readily computed with modern software and hardware.
3. Null-hypothesis significance testing (NHST), with its reliance on p values, has many problems. There is little reason to persist with NHST now that Bayesian methods are accessible to everyone.

My conclusion from those points is that we should do whatever we can to encourage the move to Bayesian data analysis." John Kruschke,

(quoted in Meyers & Guszcza (2013))

Bayes vs. Frequentist, inference on heads/tails

Consider some Bernoulli sample $\boldsymbol{x} = \{x_1, x_2, \cdots, x_n\}$, where $x_i \in \{0, 1\}$.

X_i's are i.i.d. $\mathcal{B}(p)$ variables, $f_X(x) = p^x[1-p]^{1-x}$, $x \in \{0, 1\}$.

Standard frequentist approach

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^n x_i = \text{argmin}\Big\{\underbrace{\prod_{i=1}^n f_X(x_i)}_{\mathcal{L}(p;\boldsymbol{x})}\Big\}$$

From the central limit theorem

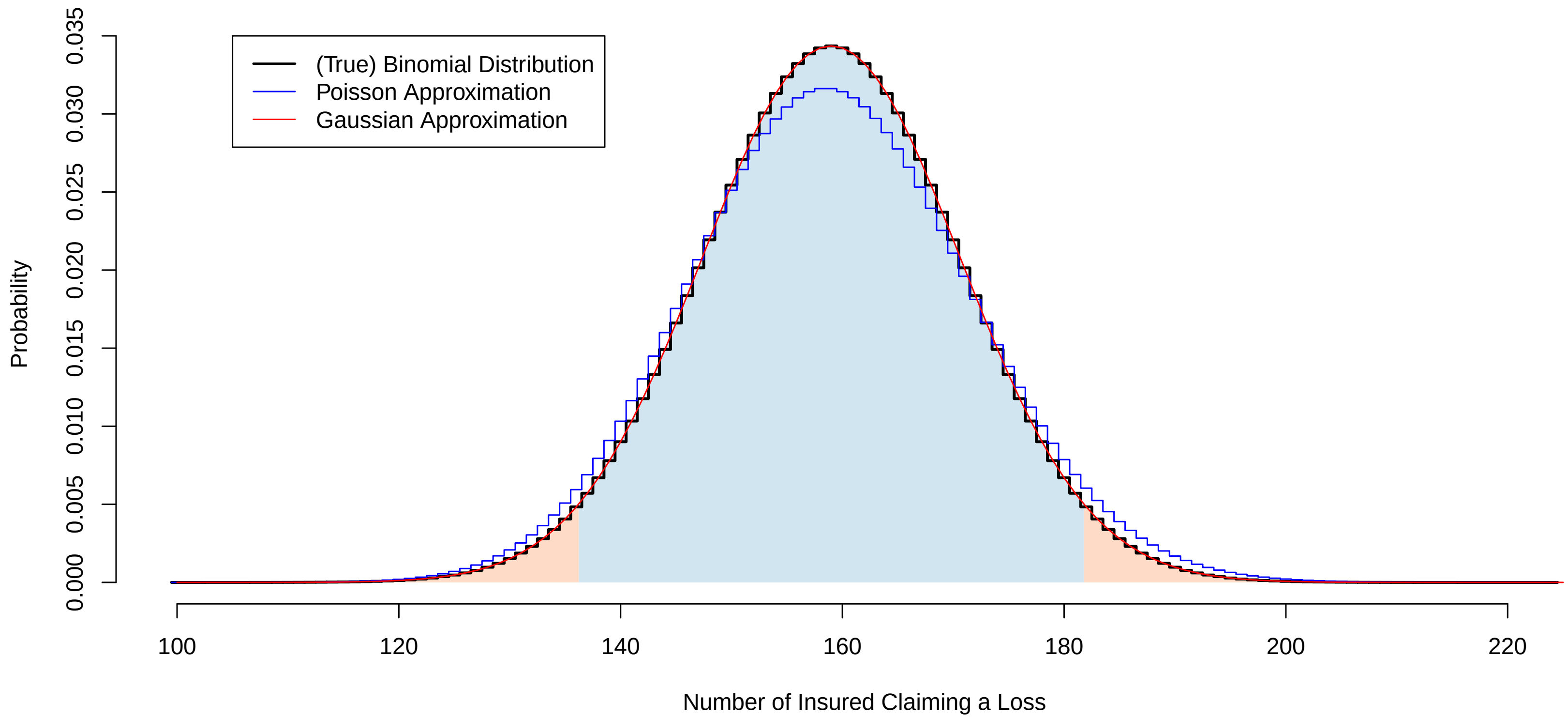
$$\sqrt{n}\frac{\widehat{p}-p}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ as } n \to \infty$$

we can derive an approximated 95% confidence interval

$$\left[\widehat{p} \pm \frac{1.96}{\sqrt{n}}\sqrt{\widehat{p}(1-\widehat{p})}\right]$$

Bayes vs. Frequentist, inference on heads/tails

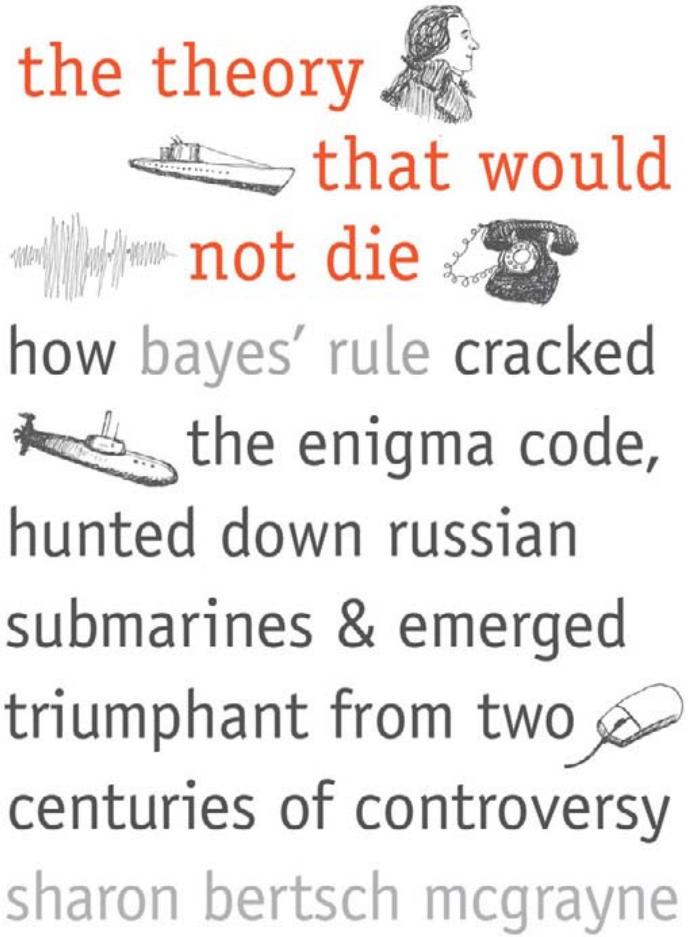
Example out of 1,047 contracts, 159 claimed a loss



Small Data and Black Swans

Example [Operational risk] What if our sample is $\boldsymbol{x} = \{0, 0, 0, 0, 0\}$?
How would we derive a confidence interval for p ?

"*INA's chief executive officer, dressed as Santa Claus, asked an unthinkable question: Could anyone predict the probability of two planes colliding in midair? Santa was asking his chief actuary, L. H. Longley-Cook, to make a prediction based on no experience at all. There had never been a serious midair collision of commercial planes. Without any past experience or repetitive experimentation, any orthodox statistician had to answer Santa's question with a resounding no.*"



Bayes, *the theory that would not die*

Liu et al. (1996) claim that “*Statistical methods with a Bayesian flavor* [...] *have long been used in the insurance industry*”.

History of Bayesian statistics, the theory that would not die by Sharon Bertsch McGrayne

“*[Arthur] Bailey spent his first year in New York [in 1918] trying to prove to himself that ‘all of the fancy actuarial [Bayesian] procedures of the casualty business were mathematically unsound.’ After a year of intense mental struggle, however, realized to his consternation that actuarial sledgehammering worked*” [...]

the theory
that would
not die
how bayes’ rule cracked
the enigma code,
hunted down russian
submarines, & emerged
triumphant from two
centuries of controversy
sharon bertsch mcgrayne

Yale university press new haven & london

Bayes, *the theory that would not die*

[...] "*He even preferred it to the elegance of frequentism. He positively liked formulae that described 'actual data . . . I realized that the hard-shelled underwriters were recognizing certain facts of life neglected by the statistical theorists.' He wanted to give more weight to a large volume of data than to the frequentists small sample; doing so felt surprisingly 'logical and reasonable'. He concluded that only a 'suicidal' actuary would use Fishers method of maximum likelihood, which assigned a zero probability to nonevents. Since many businesses file no insurance claims at all, Fishers method would produce premiums too low to cover future losses.*"

6.

arthur bailey

After the Second World War the first public challenge to the anti-Bayesian status quo came not from the military or university mathematicians and statisticians but from a Bible-quoting business executive named Arthur L. Bailey.

Bailey was an insurance actuary whose father had been fired and blackballed by every bank in Boston for telling his employers they should not be lending large sums of money to local politicians. So ostracized was the family that even Arthur's schoolmates stopped inviting him and his sister to parties. Turning his back on the New England establishment, Bailey enrolled at the University of Michigan in Ann Arbor. There he studied statistics in the mathematics department's actuarial program, earned a bachelor of science degree in 1928, and met his wife, Helen, who became an actuary for John Hancock Mutual Life before their children were born.[1]

Bailey's first job was, he liked to say, "in bananas," that is, in the statistics department of the United Fruit Company headquarters in Boston. When the department was eliminated during the Depression, Bailey wound up driving a fruit truck and chasing escaped tarantulas down Boston streets. He was lucky to have the job, and his family never lacked for bananas and oranges.

In 1937, after nine years in bananas, Bailey got a job in an unrelated field in New York City. There he was in charge of setting premium rates to cover risks involving automobiles, aircraft, manufacturing, burglary, and theft for the American Mutual Alliance, a consortium of mutual insurance companies.

Preferring church and community connections to the fair-weather friends of his youth, Bailey hid his growing professional success by living quietly in unpretentious New York suburbs. He relaxed by gardening, hiking

91

Bayes's theorem

Consider some hypothesis H and some evidence E, then

$$\mathbb{P}_E(H) = \mathbb{P}(H|E) = \frac{\mathbb{P}(H \cap E)}{\mathbb{P}(E)} = \frac{\mathbb{P}(H) \cdot \mathbb{P}(E|H)}{\mathbb{P}(E)}$$

Bayes rule,

$$\begin{cases} \text{prior probability } \mathbb{P}(H) \\ \text{versus posterior probability after receiving evidence } E,\ \mathbb{P}_E(H) = \mathbb{P}(H|E). \end{cases}$$

In Bayesian (parametric) statistics, $H = \{\theta \in \Theta\}$ and $E = \{\boldsymbol{X} = \boldsymbol{x}\}$.

Bayes' Theorem,

$$\pi(\theta|\boldsymbol{x}) = \frac{\pi(\theta) \cdot f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x})} = \frac{\pi(\theta) \cdot f(\boldsymbol{x}|\theta)}{\int f(\boldsymbol{x}|\theta)\pi(\theta)d\theta} \propto \pi(\theta) \cdot f(\boldsymbol{x}|\theta)$$

Small Data and Black Swans

Consider sample $\boldsymbol{x} = \{0, 0, 0, 0, 0\}$.

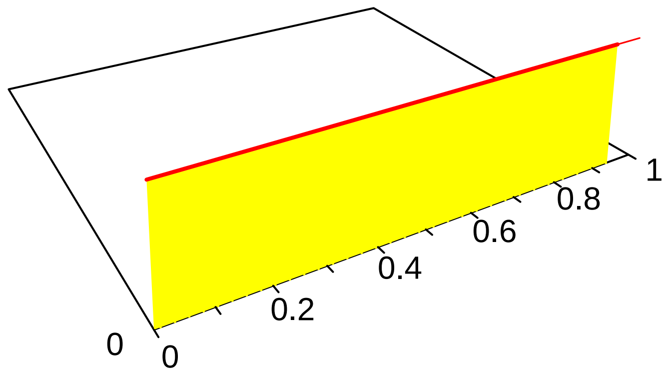
Here the likelihood is

$$\begin{cases} (x_i|\theta) = \theta^{x_i}[1-\theta]^{1-x_i} \\ f(\boldsymbol{x}|\theta) = \theta^{\boldsymbol{x}^\mathsf{T}\mathbf{1}}[1-\theta]^{n-\boldsymbol{x}^\mathsf{T}\mathbf{1}} \end{cases}$$

and we need a priori distribution $\pi(\cdot)$ e.g. a beta distribution

$$\pi(\theta) = \frac{\theta^\alpha[1-\theta]^\beta}{B(\alpha,\beta)}$$

$$\pi(\theta|\boldsymbol{x}) = \frac{\theta^{\alpha+\boldsymbol{x}^\mathsf{T}\mathbf{1}}[1-\theta]^{\beta+n-\boldsymbol{x}^\mathsf{T}\mathbf{1}}}{B(\alpha+\boldsymbol{x}^\mathsf{T}\mathbf{1},\beta+n-\boldsymbol{x}^\mathsf{T}\mathbf{1})}$$



On Bayesian Philosophy, Confidence vs. Credibility

for frequentists, a probability is a measure of the the frequency of repeated events

$\rightarrow$ parameters are fixed (but unknown), and data are random

for Bayesians, a probability is a measure of the degree of certainty about values

$\rightarrow$ parameters are random and data are fixed

"Bayesians : *Given our observed data, there is a 95% probability that the true value of θ falls within the credible region*

vs. Frequentists : *There is a 95% probability that when I compute a confidence interval from data of this sort, the true value of θ will fall within it.*" in Vanderplas (2014)

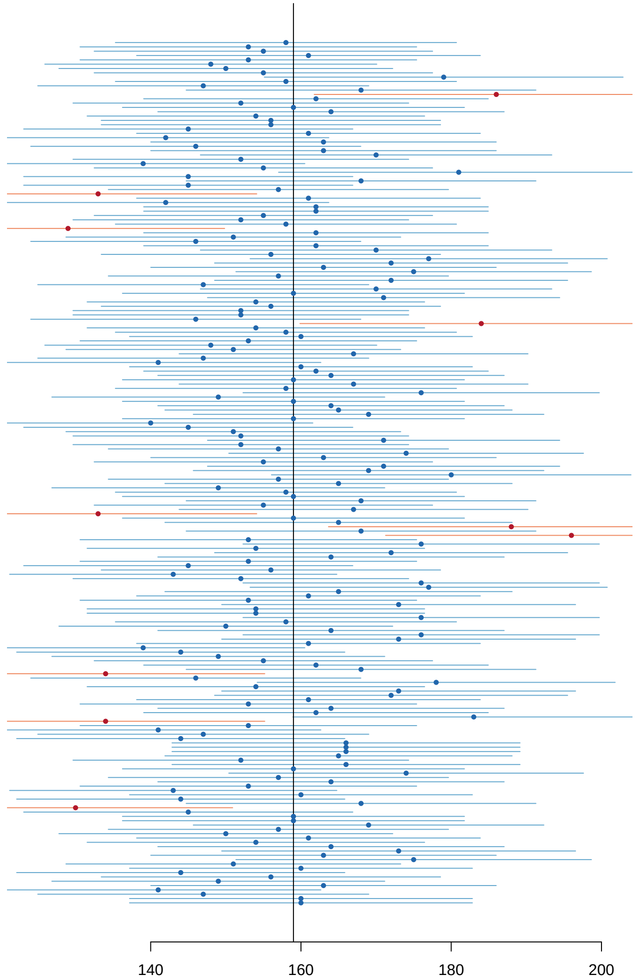
Example see Jaynes (1976), e.g. the truncated exponential

On Bayesian Philosophy, Confidence vs. Credibility

Example What is a 95% confidence interval of a proportion ? Here $\overline{x} = 159$ and $n = 1047$.

1. draw sets $(\tilde{x}_1, \cdots, \tilde{x}_n)_k$ with $X_i \sim \mathcal{B}(\overline{x}/n)$
2. compute for each set of values confidence intervals
3. determine the fraction of these confidence interval that contain $\overline{x}$

$\rightarrow$ the parameter is fixed, and we guarantee that 95% of the confidence intervals will contain it.

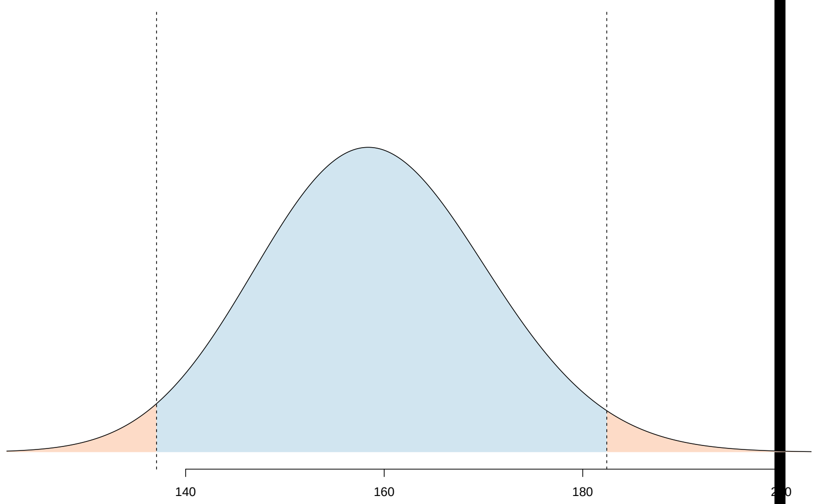


On Bayesian Philosophy, Confidence vs. Credibility

Example What is 95% credible region of a proportion ? Here $\overline{x} = 159$ and $n = 1047$.

1. draw random parameters p_k with from the posterior distribution, $\pi(\cdot|\boldsymbol{x})$
2. sample sets $(\tilde{x}_1, \cdots, \tilde{x}_n)_k$ with $X_{i,k} \sim \mathcal{B}(p_k)$
3. compute for each set of values means $\overline{x}_k$
4. look at the proportion of those $\overline{x}_k$ that are within this credible region $[\Pi^{-1}(.025|\boldsymbol{x}); \Pi^{-1}(.975|\boldsymbol{x})]$

$\rightarrow$ the credible region is fixed, and we guarantee that 95% of possible values of $\overline{x}$ will fall within it it.



Difficult concepts ? Difficult computations ?

We have a sample $\boldsymbol{x} = \{x_1, \cdots, x_d)$ i.i.d. from distribution $f_\theta(\cdot)$.

In predictive modeling, we need $\mathbb{E}(g(X)|\boldsymbol{x}) = \int x f_{\theta|\boldsymbol{x}}(x)dx$ where

$$f_{\theta|\boldsymbol{x}}(x) = f(x|\boldsymbol{x}) = \int f(x|\theta) \cdot \pi(\theta|\boldsymbol{x})d\theta$$

How can we derive $\pi(\theta|\boldsymbol{x})$?

Can we sample from $\pi(\theta|\boldsymbol{x})$ (use monte carlo technique to approximate the integral) ?

Computations not that simple... until the 90's : MCMC

Markov Chain

Stochastic process, $(X_t)_{t\in\mathbb{N}_\star}$, on some discrete space Ω

$$\mathbb{P}(X_{t+1}=y|X_t=x,\underline{\boldsymbol{X}}_{t-1}=\underline{\boldsymbol{x}}_{t-1})=\mathbb{P}(X_{t+1}=y|X_t=x)=P(x,y)$$

where P is a transition probability, that can be stored in a transition matrix, $\boldsymbol{P}=[P_{x,y}]=[P(x,y)]$.

Observe that $\mathbb{P}(X_{t+k}=y|X_t=x)=P_k(x,y)$ where $\boldsymbol{P}^k=[P_k(x,y)]$.

Under some condition, $\lim_{n\rightarrow\infty}\boldsymbol{P}^n=\boldsymbol{\Lambda}=[\boldsymbol{\lambda}^{\mathsf{T}}]$,

Problem given a distribution $\boldsymbol{\lambda}$, is it possible to generate a Markov Chain that converges to this distribution ?

Bonus Malus and Markov Chains

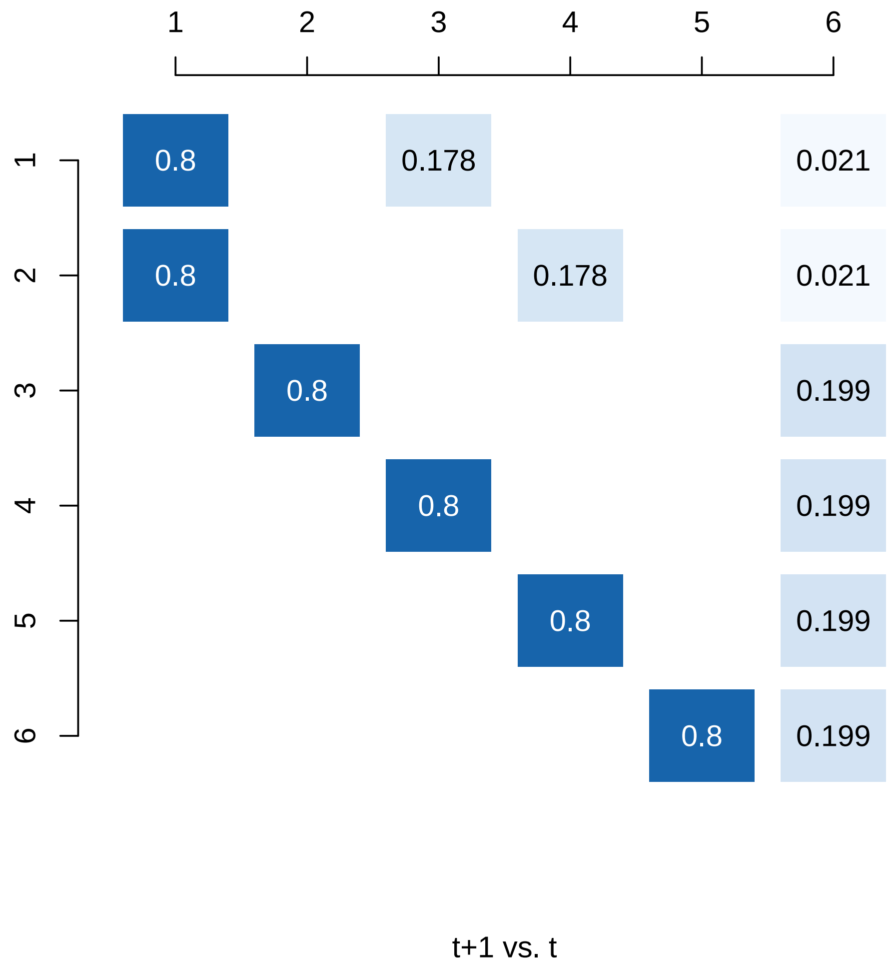
Ex no-claim bonus, see Lemaire (1995).

HONG KONG
Table B-9. Hong Kong System

Class	Premium	Class After 0 Claims	Class After 1 Claims	Class After ≥2 Claims
6	100	5	6	6
5	80	4	6	6
4	70	3	6	6
3	60	2	6	6
2	50	1	4	6
1	40	1	3	6

Starting class: 6.

Assume that the number of claims is $N \sim \mathcal{P}(10.536)$, so that $\mathbb{P}(N=0) = 10\%$.



Hastings-Metropolis

Back to our problem, we want to sample from $\pi(\theta|\boldsymbol{x})$

i.e. generate $\theta_1, \cdots, \theta_n, \cdots$ from $\pi(\theta|\boldsymbol{x})$.

Hastings-Metropolis sampler will generate a Markov Chain (θ_t) as follows,

- generate θ_1
- generate $\theta^\star$ and $U \sim \mathcal{U}([0,1])$,

  compute $R = \dfrac{\pi(\theta^\star|\boldsymbol{x})}{\pi(\theta_t|\boldsymbol{x})} \dfrac{P(\theta_t|\theta^\star)}{P(\theta^\star|\theta_{t-1})}$

  if $U < R$ set $\theta_{t+1} = \theta^\star$

  if $U \geq R$ set $\theta_{t+1} = \theta_t$

R is the acceptance ratio, we accept the new state $\theta^\star$ with probability $\min\{1, R\}$.

Hastings-Metropolis

Observe that

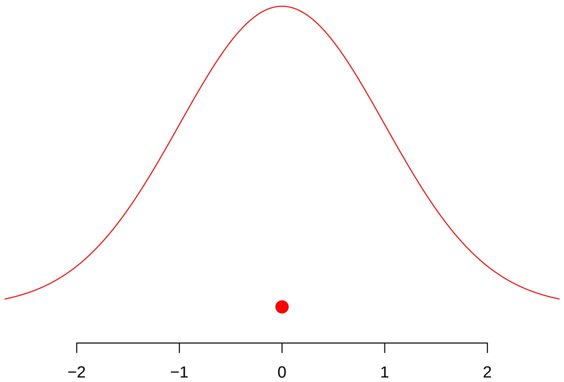
$$R = \frac{\pi(\theta^\star)\cdot f(\boldsymbol{x}|\theta^\star)}{\pi(\theta_t)\cdot f(\boldsymbol{x}|\theta_t)}\frac{P(\theta_t|\theta^\star)}{P(\theta^\star|\theta_{t-1})}$$

In a more general case, we can have a Markov process, not a Markov chain.

E.g. $P(\theta^\star|\theta_t)\sim\mathcal{N}(\theta_t,1)$

Using MCMC to generate Gaussian values

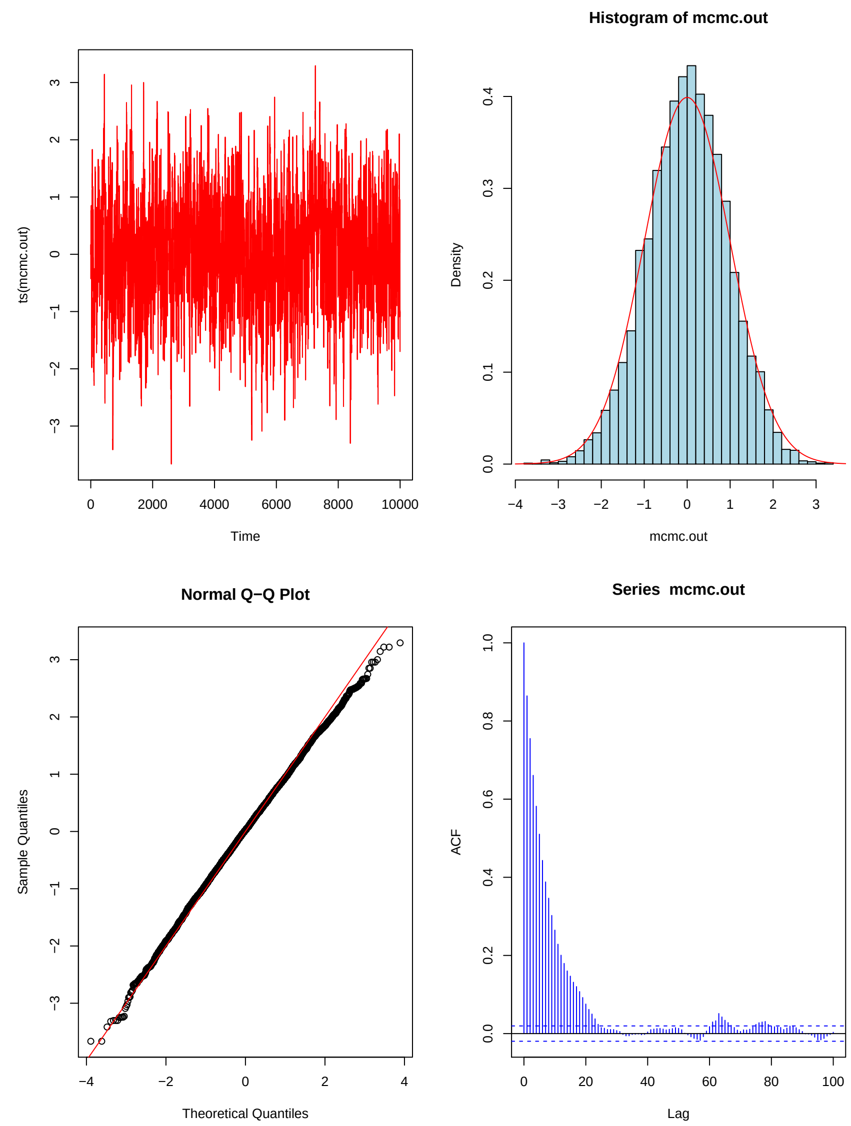
```
> metrop1 <- function(n=1000,eps=0.5){
+ vec <- vector("numeric", n)
+ x=0
+ vec[1] <- x
+ for (i in 2:n) {
+ innov <- runif(1,-eps,eps)
+ mov <- x+innov
+ aprob <- min(1,dnorm(mov)/dnorm(x))
+ u <- runif(1)
+ if (u < aprob)
+ x <- mov
+ vec[i] <- x
+ }
+ return(vec)}
```



Using MCMC to generate Gaussian values

```
> plot.mcmc <- function(mcmc.out) {
+ op <- par(mfrow=c(2,2))
+ plot(ts(mcmc.out),col="red")
+ hist(mcmc.out,30,probability=TRUE,
+ col="light blue")
+ lines(seq(-4,4,by=.01),dnorm(seq(-4,4,
+ by=.01)),col="red")
+ qqnorm(mcmc.out)
+ abline(a=mean(mcmc.out),b=sd(mcmc.out))
+ acf(mcmc.out,col="blue",lag.max=100)
+ par(op) }

> metrop.out<-metrop1(10000,1)
> plot.mcmc(metrop.out)
```



Heuristics on Hastings-Metropolis

In standard Monte Carlo, generate θ_i's i.i.d., then

$$\frac{1}{n}\sum_{i=1}^{n} g(\theta_i) \rightarrow \mathbb{E}[g(\theta)] = \int g(\theta)\pi(\theta)d\theta$$

(strong law of large numbers).

Well-behaved Markov Chains ($\boldsymbol{P}$ aperiodic, irreducible, positive recurrent) can satisfy some ergodic property, similar to that LLN. More precisely,

- $\boldsymbol{P}$ has a unique stationary distribution λ, i.e. $\lambda = \lambda \times \boldsymbol{P}$
- ergodic theorem

$$\frac{1}{n}\sum_{i=1}^{n} g(\theta_i) \rightarrow \int g(\theta)\lambda(\theta)d\theta$$

even if θ_i's are not independent.

Heuristics on Hastings-Metropolis

Remark The conditions mentioned above are

- aperiodic, the chain does not regularly return to any state in multiples of some k.
- irreducible, the state can go from any state to any other state in some finite number of steps
- positively recurrent, the chain will return to any particular state with probability 1, and finite expected return time

MCMC and Loss Models

Example A Tweedie model, $\mathbb{E}(X) = \mu$ and $\text{Var}(X) = \varphi \cdot \mu^p$. Here assume that φ and p are given, and μ is the unknown parameter.

$\rightarrow$ need a predictive distribution for μ given $\boldsymbol{x}$.

Consider the following transition kernel (a Gamma distribution)

$$\mu|\mu_t \sim \mathcal{G}\left(\frac{\mu_t}{\alpha}, \alpha\right)$$

with $\mathbb{E}(\mu|\mu_t) = \mu_t$ and $\text{CV}(\mu) = \dfrac{1}{\sqrt{\alpha}}$.

Use some a priori distribution, e.g. $\mathcal{G}(\alpha_0, \beta_0)$.

MCMC and Loss Models

- generate μ_1
- at step t : generate $\mu^\star \sim \mathcal{G}\left(\alpha^{-1}\mu_t, \alpha\right)$ and $U \sim \mathcal{U}([0,1])$,

  compute $R = \dfrac{\pi(\mu^\star)\cdot f(\boldsymbol{x}|\mu^\star)}{\pi(\mu_t)\cdot f(\boldsymbol{x}|\theta_t)}\dfrac{P_\alpha(\mu_t|\theta^\star)}{P_\alpha(\theta^\star|\theta_{t-1})}$

  if $U < R$ set $\theta_{t+1} = \theta^\star$

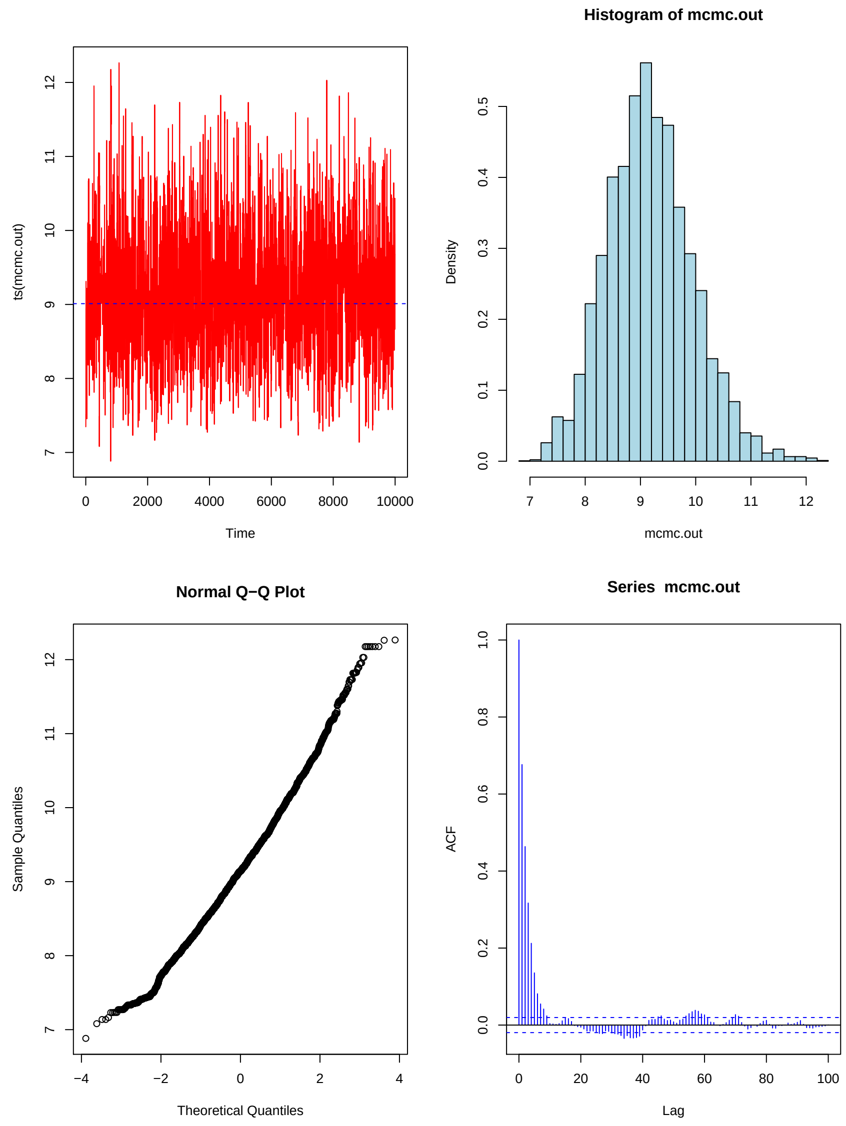
  if $U \geq R$ set $\theta_{t+1} = \theta_t$

where

$$f(\boldsymbol{x}|\mu) = \mathcal{L}(\mu) = \prod_{i=1}^{n} f(x_i|\mu, p, \varphi),$$

$f(x\cdot|\mu, p, \varphi)$ being the density of the Tweedie distribution, `dtweedie function (x, p, mu, phi)` from `library(tweedie)`.

```
> p=2 ; phi=2/5
> set.seed(1) ; X <- rtweedie(50,p,10,phi)
> metrop2 <- function(n=10000,a0=10,
+ b0=1,alpha=1){
+ vec <- vector("numeric", n)
+ mu <- rgamma(1,a0,b0)
+ vec[1] <- mu
+ for (i in 2:n) {
+ mustar <- rgamma(1,vec[i-1]/alpha,alpha)
+ R=prod(dtweedie(X,p,mustar,phi)/dtweedie
+ (X,p,vec[i-1],phi))*dgamma(mustar,a0,b0)/
+ dgamma(vec[i-1],a0,b0)* dgamma(vec[i-1],
+ mustar/alpha,alpha)/dgamma(mustar,
+ vec[i-1]/alpha,alpha)
+ aprob <- min(1,R)
+ u <- runif(1)
+ ifelse(u < aprob,vec[i]<-mustar,
+ vec[i]<-vec[i-1]) }
+ return(vec)}
> metrop.output<-metrop2(10000,alpha=1)
```



Gibbs Sampler

For a multivariate problem, it is possible to use Gibbs sampler.

Example Assume that the loss ratio of a company has a lognormal distribution, $LN(\mu, \sigma^2)$, .e.g

```
> LR <- c(0.958, 0.614, 0.977, 0.921, 0.756)
```

Example Assume that we have a sample $\boldsymbol{x}$ from a $\mathcal{N}(\mu, \sigma^2)$. We want the posterior distribution of $\boldsymbol{\theta} = (\mu, \sigma^2)$ given $\boldsymbol{x}$. Observe here that if priors are Gaussian $\mathcal{N}\left(\mu_0, \tau^2\right)$ and the inverse Gamma distribution $IG(a, b)$, them

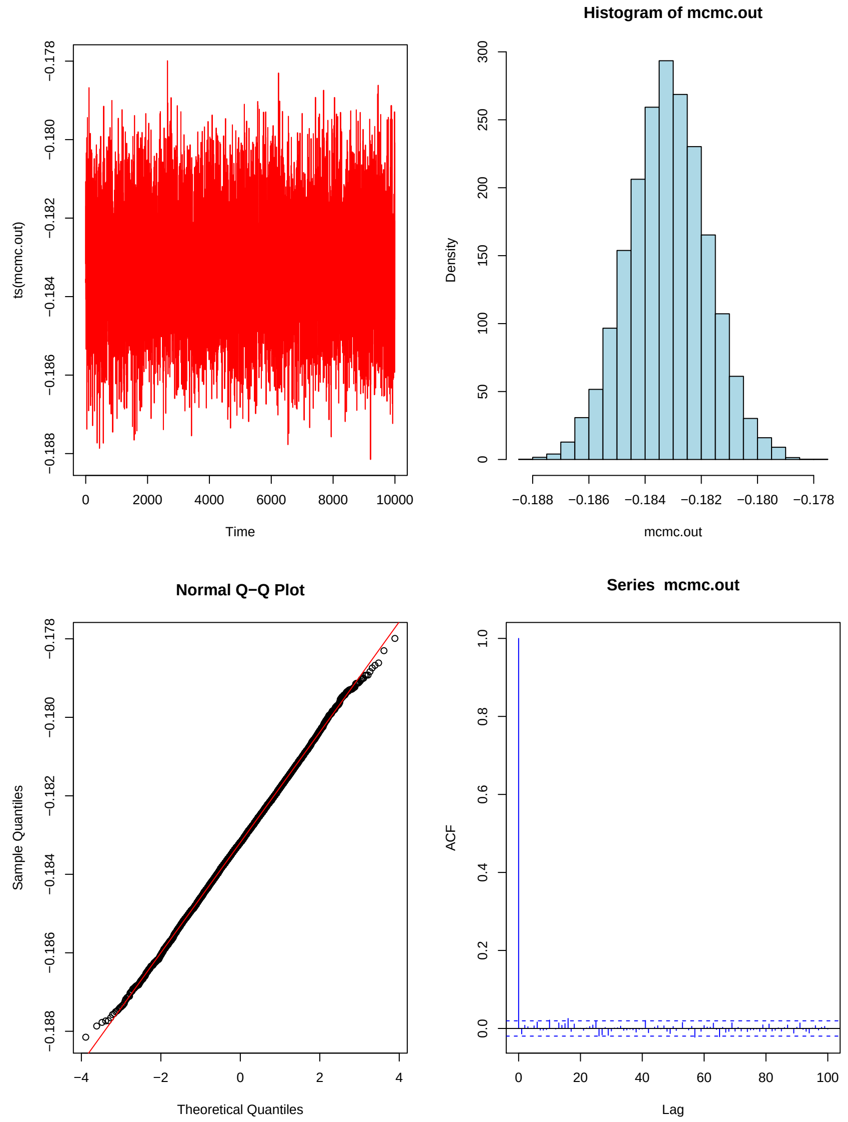
$$\begin{cases} \mu|\sigma^2, \boldsymbol{x} \sim \mathcal{N}\left(\dfrac{\sigma^2}{\sigma^2 + n\tau^2}\mu_0 + \dfrac{n\tau^2}{\sigma^2 + n\tau^2}\overline{x}, \dfrac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}\right) \\ \sigma^2|\mu, \boldsymbol{x} \sim IG\left(\dfrac{n}{2} + a, \dfrac{1}{2}\displaystyle\sum_{i=1}^{n}[x_i - \mu]^2 + b\right) \end{cases}$$

More generally, we need the conditional distribution of $\theta_k|\boldsymbol{\theta}_{-k}, \boldsymbol{x}$, for all k.

```
> x <- log(LR)
```

Gibbs Sampler

```
> xbar <- mean(x)
> mu <- sigma2=rep(0,10000)
> sigma2[1] <- 1/rgamma(1,shape=1,rate=1)
> Z <- sigma2[1]/(sigma2[1]+n*1)
> mu[1] <- rnorm(1,m=Z*0+(1-Z)*xbar,
+ sd=sqrt(1*Z))
> for (i in 2:10000){
+ Z <- sigma2[i-1]/(sigma2[i-1]+n*1)
+ mu[i] <- rnorm(1,m=Z*0+(1-Z)*xbar,
+ sd=sqrt(1*Z))
+ sigma2[i] <- 1/rgamma(1,shape=n/2+1,
+ rate <- (1/2)*(sum((x-mu[i])^2))+1)
+ }
```



Gibbs Sampler

Example Consider some vector $\boldsymbol{X} = (X_1, \cdots, X_d)$ with indépendent components, $X_i \sim \mathcal{E}(\lambda_i)$. We sample to sample from $\boldsymbol{X}$ given $\boldsymbol{X}^\top \mathbf{1} > s$ for some threshold $s > 0$.

- start with some starting point $\boldsymbol{x}_0$ such that $\boldsymbol{x}_0^\top \mathbf{1} > s$
- pick up (randomly) $i \in \{1, \cdots, d\}$

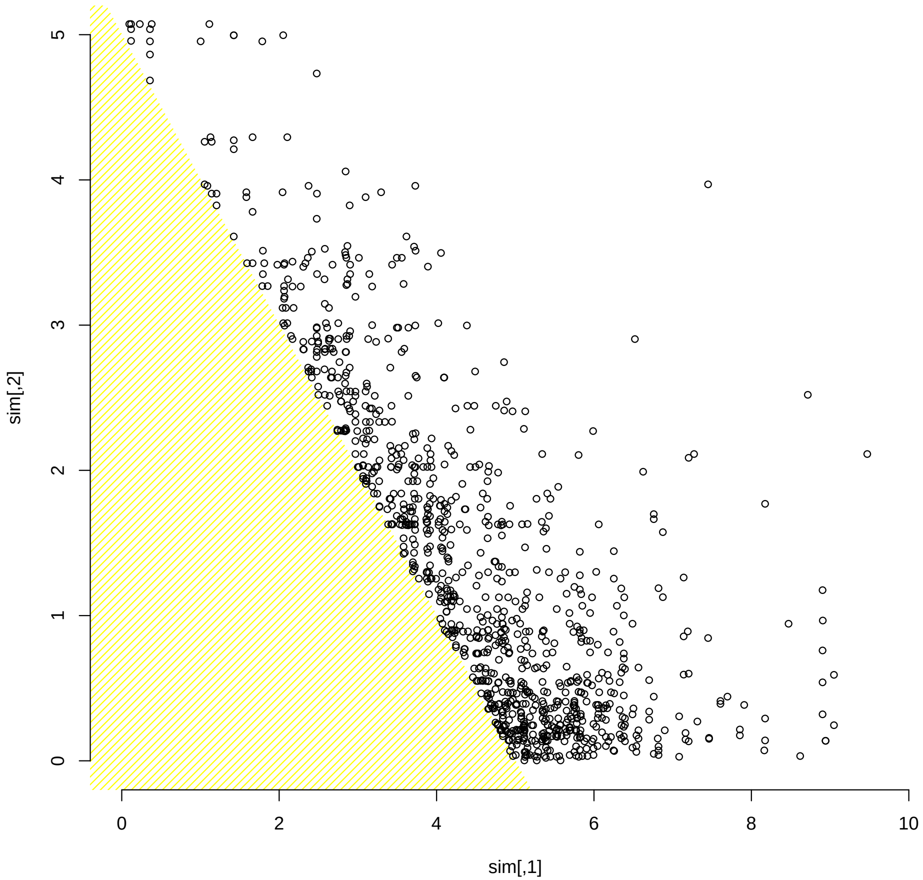
  X_i given $X_i > s - \boldsymbol{x}_{(-i)}^\top \mathbf{1}$ has an Exponential distribution $\mathcal{E}(\lambda_i)$

  draw $Y \sim \mathcal{E}(\lambda_i)$ and set $x_i = y + (s - \boldsymbol{x}_{(-i)}^\top \mathbf{1})_+$ until $\boldsymbol{x}_{(-i)}^\top \mathbf{1} + x_i > s$

E.g. losses and allocated expenses

Gibbs Sampler

```
> sim <- NULL
> lambda <- c(1,2)
> X <- c(3,3)
> s <- 5
> for(k in 1:1000){
+ i <- sample(1:2,1)
+ X[i] <- rexp(1,lambda[i])+
+ max(0,s-sum(X[-i]))
+ while(sum(X)<s){
+ X[i] <- rexp(1,lambda[i])+
+ max(0,s-sum(X[-i])) }
+ sim <- rbind(sim,X) }
```



JAGS and STAN

Martyn Plummer developed JAGS *Just another Gibbs sampler* in 2007 (stable since 2013) in `library(runjags)`. It is an open-source, enhanced, cross-platform version of an earlier engine BUGS (Bayesian inference Using Gibbs Sampling).

STAN `library(Rstan)` is a newer tool that uses the Hamiltonian Monte Carlo (HMC) sampler.

HMC uses information about the derivative of the posterior probability density to improve the algorithm. These derivatives are supplied by algorithm differentiation in C/C++ codes.

JAGS on the $\mathcal{N}(\mu, \sigma^2)$ distribution

```
> library(runjags)
> jags.model <- "
+ model {
+ mu ~ dnorm(mu0, 1/(sigma0^2))
+ g ~ dgamma(k0, theta0)
+ sigma <- 1 / g
+ for (i in 1:n) {
+ logLR[i] ~ dnorm(mu, g^2)
+ }
+ }"
```

```
> jags.data <- list(n=length(LR),
+ logLR=log(LR), mu0=-.2, sigma0=0.02,
+ k0=1, theta0=1)

> jags.init <- list(list(mu=log(1.2),
+ g=1/0.5^2),
+ list(mu=log(.8),
+ g=1/.2^2))

> model.out <- autorun.jags(jags.model,
+ data=jags.data, inits=jags.init,
+ monitor=c("mu", "sigma"), n.chains=2)
> traceplot(model.out$mcmc)
> summary(model.out)
```

STAN on the $\mathcal{N}(\mu, \sigma^2)$ distribution

```
> library(rstan)
> stan.model <- "
+ data {
+ int<lower=0> n;
+ vector[n] LR;
+ real mu0;
+ real<lower=0> sigma0;
+ real<lower=0> k0;
+ real<lower=0> theta0;
+ }
+ parameters {
+ real mu;
+ real<lower=0> sigma;
+ }
+ model {
+ mu ~ normal(mu0, sigma0);
+ sigma ~ inv_gamma(k0, theta0);
+ for (i in 1:n)
+ log(LR[i]) ~ normal(mu, sigma);
+ }"

> stan.data <- list(n=length(LR), r=LR, mu0=mu0,
+ sigma0=sigma0, k0=k0, theta0=theta0)
> stan.out <- stan(model_code=stan.model,
+ data=stan.data, seed=2)
> traceplot(stan.out)
> print(stan.out, digits_summary=2)
```

MCMC and Loss Models

Example Consider some simple time series of Loss Ratios,

$$LR_t \sim \mathcal{N}(\mu_t, \sigma^2) \text{ where } \mu_t = \phi\mu_{t-1} + \varepsilon_t$$

E.g. in JAGS we can define the vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_T)$ recursively

```
+ model {
+ mu[1] ~ dnorm(mu0, 1/(sigma0^2))
+ for (t in 2:T) { mu[t] ~ dnorm(mu[t-1], 1/(sigma0^2)) }
+ }
```

MCMC and Claims Reserving

Consider the following (cumulated) triangle, $\{C_{i,j}\}$,

	0	1	2	3	4	5
0	3209	4372	4411	4428	4435	4456
1	3367	4659	4696	4720	4730	4752.4
2	3871	5345	5398	5420	5430.1	5455.8
3	4239	5917	6020	6046.1	6057.4	6086.1
4	4929	6794	6871.7	6901.5	6914.3	6947.1
5	5217	7204.3	7286.7	7318.3	7331.9	7366.7

λ_j	1.3809	1.0114	1.0043	1.0018	1.0047
σ_j	0.7248	0.3203	0.04587	0.02570	0.02570

(from Markus' `library(ChainLadder)`).

A Bayesian version of Chain Ladder

	0	1	2	3	4	5
0		1.362418	1.008920	1.003854	1.001581	1.004735
1		1.383724	1.007942	1.005111	1.002119	
2		1.380780	1.009916	1.004076		
3		1.395848	1.017407			
4		1.378373				

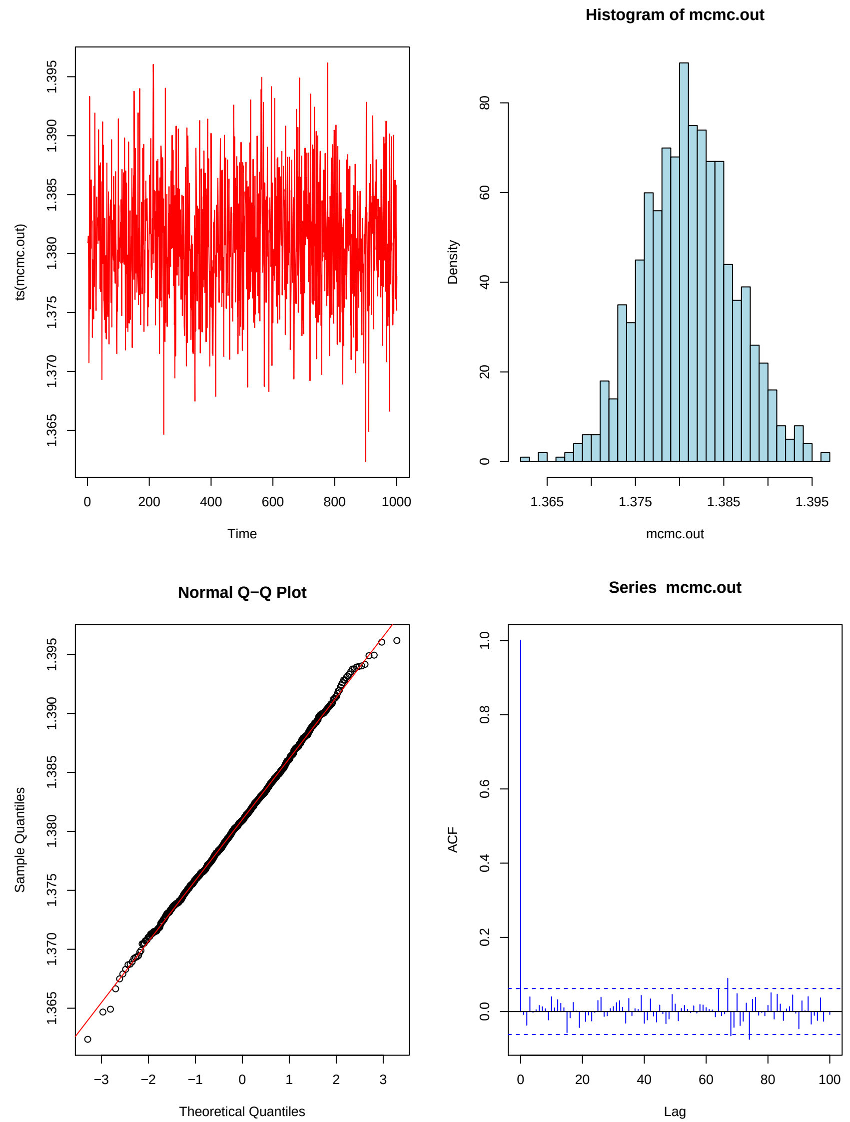
λ_j	1.3809	1.0114	1.0043	1.0018	1.0047
σ_j	0.7248	0.3203	0.04587	0.02570	0.02570

Assume that $\lambda_{i,j} \sim \mathcal{N}\left(\mu_j, \dfrac{\tau_j}{C_{i,j}}\right)$.

We can use Gibbs sampler to get the distribution of the transition factors, as well as a distribution for the reserves,

```
> source("http://freakonometrics.free.fr/
triangleCL.R")
> source("http://freakonometrics.free.fr/
bayesCL.R")
> mcmcCL<-bayesian.triangle(PAID)
> plot.mcmc(mcmcCL$Lambda[,1])
> plot.mcmc(mcmcCL$Lambda[,2])
> plot.mcmc(mcmcCL$reserves[,6])
> plot.mcmc(mcmcCL$reserves[,7])

> library(ChainLadder)
> MCL<-MackChainLadder(PAID)
> m<-sum(MCL$FullTriangle[,6]-
+ diag(MCL$FullTriangle[,6:1]))
> stdev<-MCL$Total.Mack.S.E
> hist(mcmcCL$reserves[,7],probability=TRUE,
> breaks=20,col="light blue")
> x=seq(2000,3000,by=10)
> y=dnorm(x,m,stdev)
> lines(x,y,col="red")
```



A Bayesian analysis of the Poisson Regression Model

In a Poisson regression model, we have a sample $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_i, y_i)\}$,

$$y_i \sim \mathcal{P}(\mu_i) \text{ with } \log \mu_i = \beta_0 + \beta_1 x_i.$$

In the Bayesian framework, β_0 and β_1 are random variables.

Example: for instance `library(arm)`, (see also `library(INLA)`)

The code is very simple : from

```
> reg<-glm(dist~speed,data=cars,family=poisson)
```

get used to

```
> regb <- bayesglm(dist~speed,data=cars,family=poisson)
```

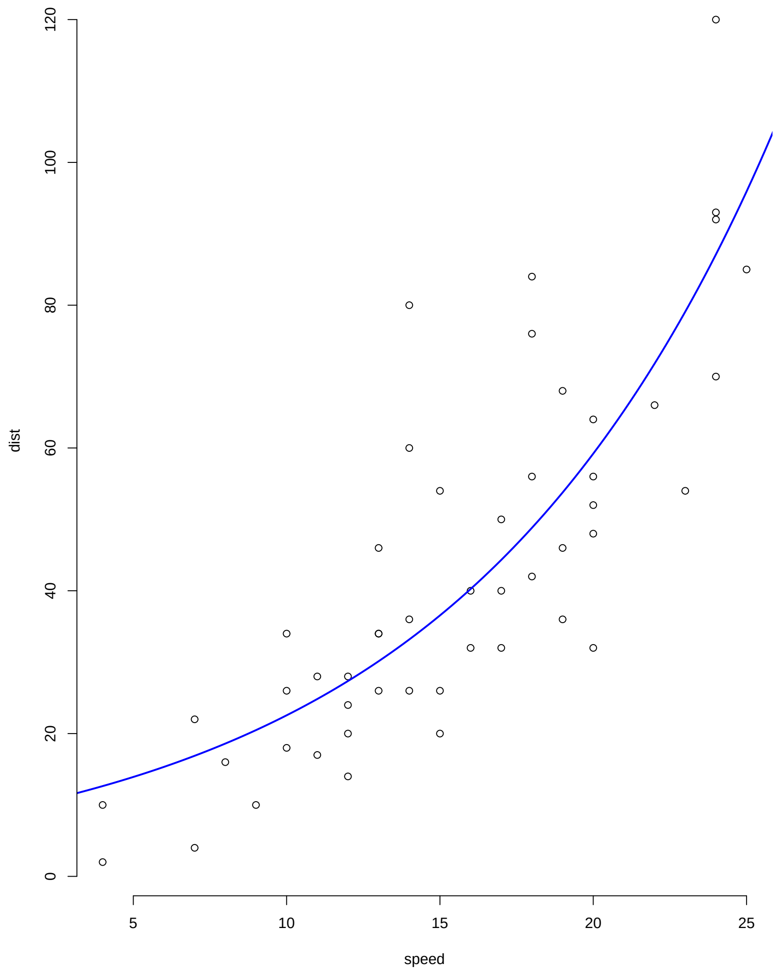
A Bayesian analysis of the Poisson Regression Model

```
> newd <- data.frame(speed=0:30)
> predreg <- predict(reg,newdata=
+ newd,type="response")
> plot(cars,axes)
> lines(newd$speed,predreg,lwd=2)

> library(arm)
> beta01<-coef(sim(regb))

> for(i in 1:100){
> lines(newd$speed,exp(beta01[i,1]+
> beta01[i,2]*newd$speed))}

> plot.mcmc(beta01[,1])
> plot.mcmc(beta01[,2])
```



Other alternatives to classical statistics

Consider a regression problem, $\mu(x) = \mathbb{E}(Y|X = x)$, and assume that smoothed splines are used,

$$\mu(x) = \sum_{i=1}^{k} \beta_j h_j(x)$$

Let $\boldsymbol{H}$ be the $n \times k$ matrix, $\boldsymbol{H} = [h_j(x_i)] = [\boldsymbol{h}(x_i)]$, then $\widehat{\boldsymbol{\beta}} = (\boldsymbol{H}^\mathsf{T}\boldsymbol{H})^{-1}\boldsymbol{H}^\mathsf{T}\boldsymbol{y}$, and

$$\widehat{\text{se}}(\widehat{\mu}(x)) = [\boldsymbol{h}(x)^\mathsf{T}(\boldsymbol{H}^\mathsf{T}\boldsymbol{H})^{-1}\boldsymbol{h}(x)]^{\frac{1}{2}}\widehat{\sigma}$$

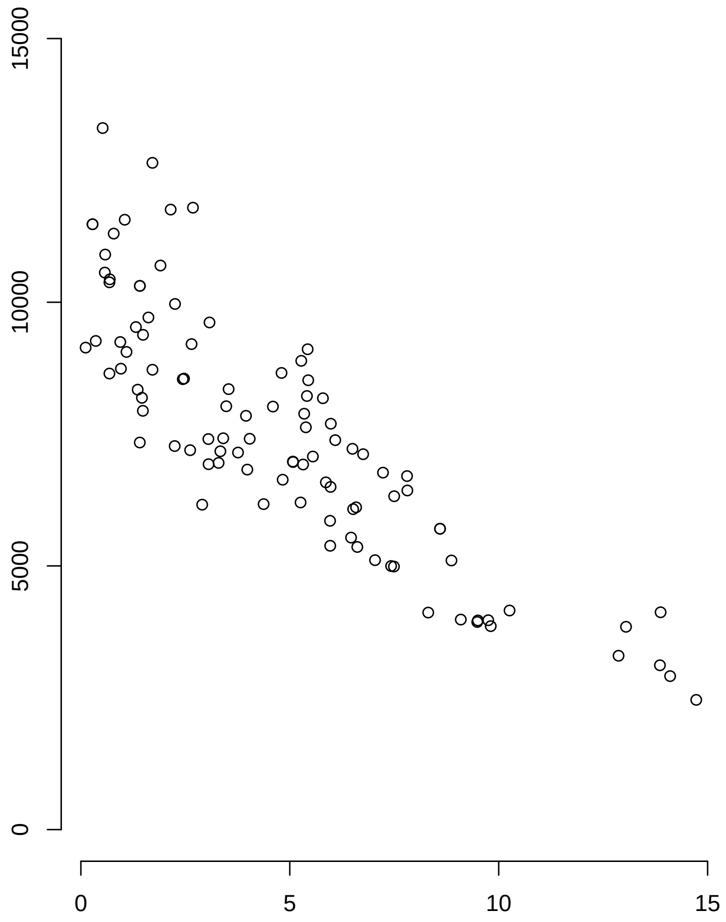
With a Gaussian assumption on the residuals, we can derive (approximated) confidence bands for predictions $\widehat{\mu}(x)$.

Smoothed regression with splines

```
> dtf <- read.table(
+ "http://freakonometrics.free.fr/
   theftinsurance.txt",sep=";",
+ header=TRUE)
> names(dtf)<-c("x","y")

> library(splines)
> reg=lm(y~bs(x,df=4),data=dtf)

> yp=predict(reg,type="response",
+ newdata=new,interval="confidence")
```



Bayesian interpretation of the regression problem

Assume here that $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{\Sigma})$ as the priori distribution for $\boldsymbol{\beta}$.

Then, if $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_i, y_i), i = 1, \cdots, n\}$, the posterior distribution of $\mu(x)$ will be Gaussian, with

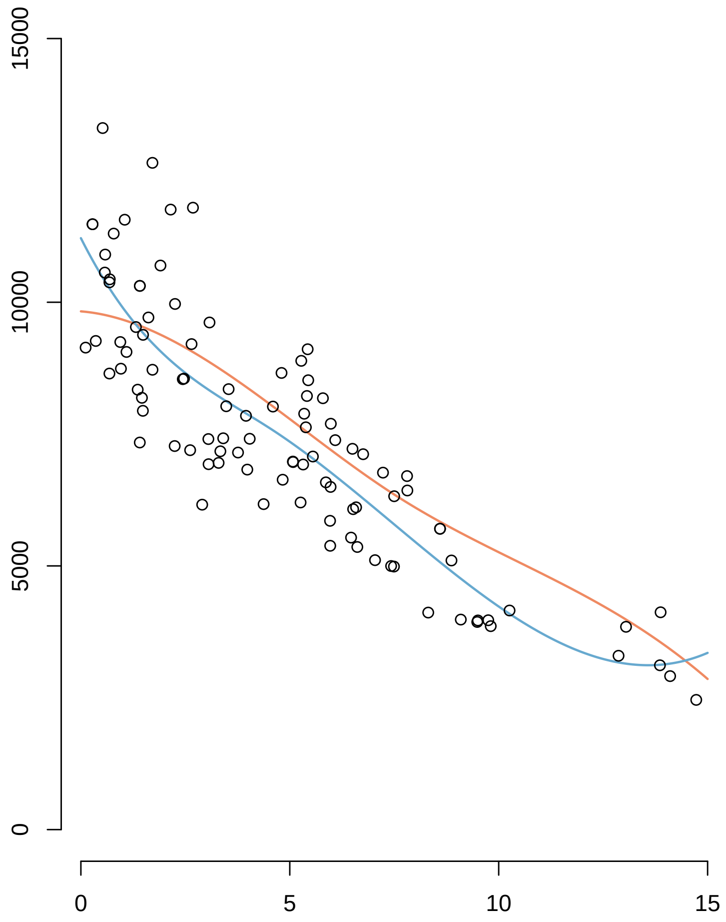
$$\mathbb{E}(\mu(x)|\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{h}(x)^{\mathsf{T}} \left(\boldsymbol{H}^{\mathsf{T}} \boldsymbol{H} + \frac{\sigma^2}{\tau} \boldsymbol{\Sigma}^{-1} \right)^{-1} \boldsymbol{H}^{\mathsf{T}} \boldsymbol{y}$$

$$\text{cov}(\mu(x), \mu(x')|\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{h}(x)^{\mathsf{T}} \left(\boldsymbol{H}^{\mathsf{T}} \boldsymbol{H} + \frac{\sigma^2}{\tau} \boldsymbol{\Sigma}^{-1} \right)^{-1} \boldsymbol{h}(x')\sigma^2$$

Example $\boldsymbol{\Sigma} = \mathbb{I}$

Bayesian interpretation of the regression problem

```
> tau <- 100
> sigma <- summary(reg)$sigma
> H=cbind(rep(1,nrow(dtf)),matrix(bs(b$x,
+ df=4),nrow=nrow(dtf)))
> h=cbind(rep(1,nrow(new)),matrix(bs(new$x,
+ df=4),nrow=nrow(new)))
> E=h%*%solve(t(H)%*%H + sigma^2/tau*
+ diag(1,ncol(H)))%*%t(H)%*%dtf$y
> V=h%*%solve(t(H)%*%H + sigma^2/tau*
+ diag(1,ncol(H)))%*% t(h) * sigma^2
> z=E+t(chol(V))%*%rnorm(length(E))
```



Bootstrap strategy

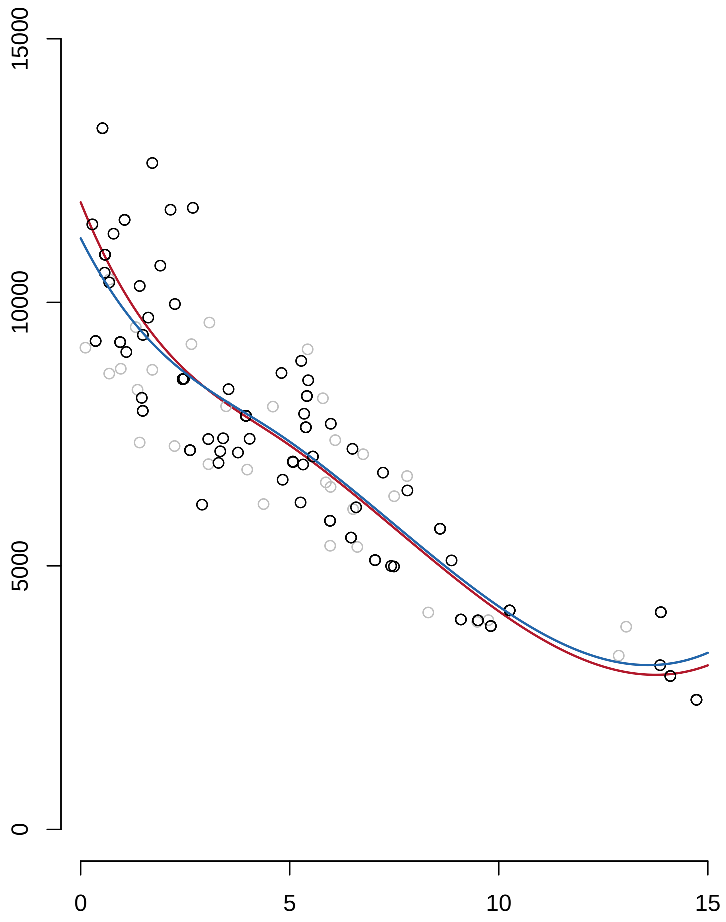
Assume that $Y = \mu(x) + \varepsilon$, and based on the estimated model, generate pseudo observations, $y_i^\star = \widehat{\mu}(x_i) + \widehat{\varepsilon}_i^\star$.

Based on $(\boldsymbol{x}, \boldsymbol{y}^\star) = \{(x_i, y_i^\star), i = 1, \cdots, n\}$, derive the estimator $\widehat{\mu}^\star(\star)$

(and repeat)

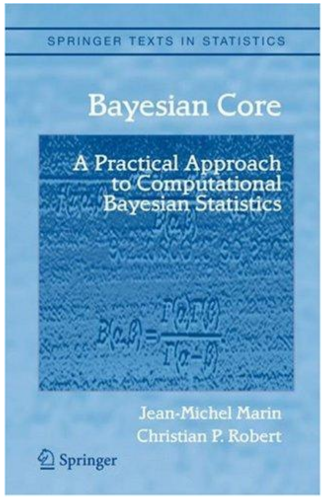
Bootstrap strategy

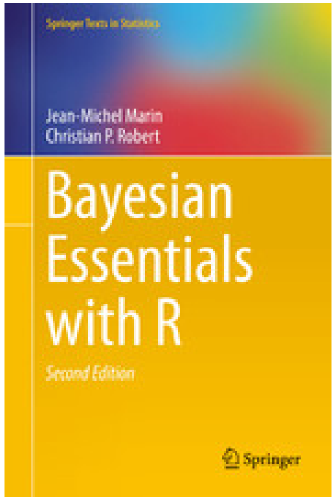
```
> for(b in 1:1000) {
+ i=sample(1:nrow(dtf),size=nrow(dtf),
+ replace=TRUE)
+ regb=lm(y~bs(x,df=4),data=dtf[i,])
+ ypb[,b]=predict(regb,type="response",
+ newdata=new))
+ }
```

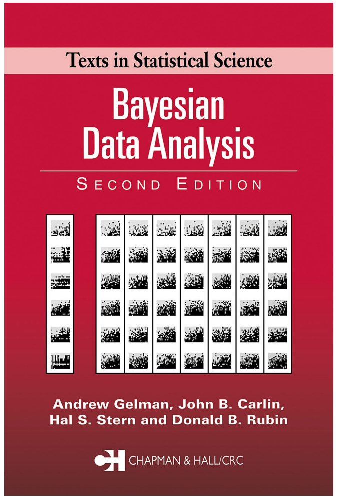


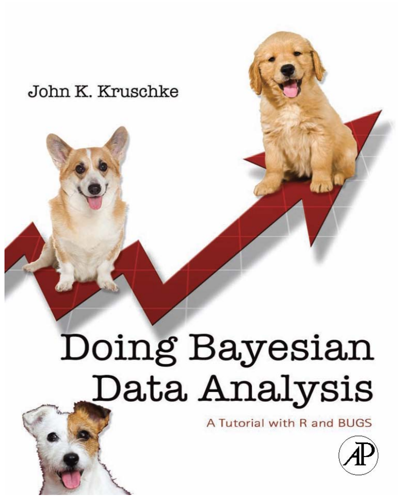
Observe that the bootstrap is the Bayesian case, when $\tau \to \infty$.

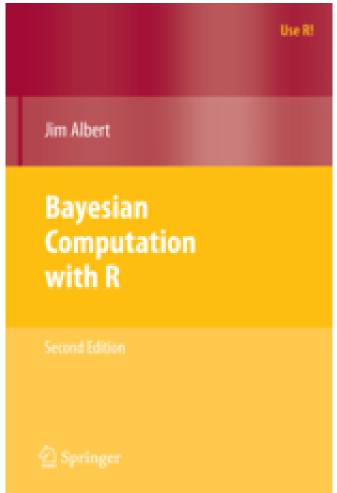
Some additional references (before a conclusion)

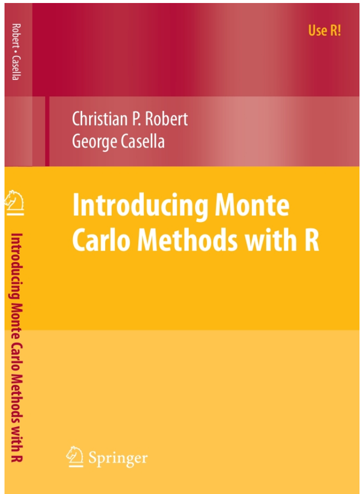


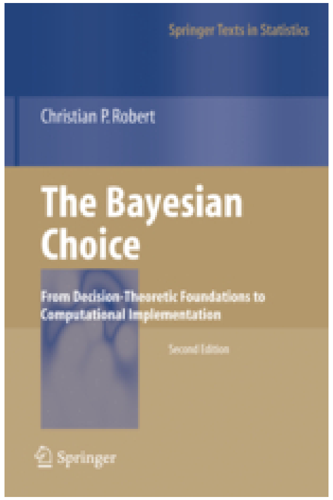


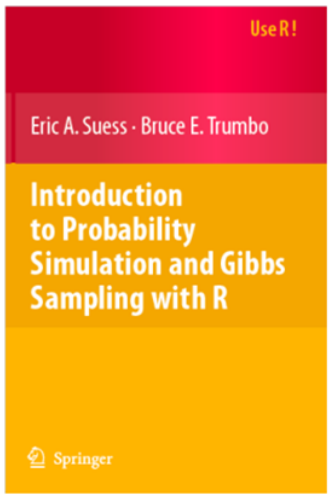


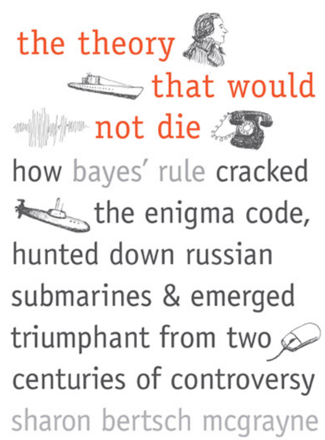




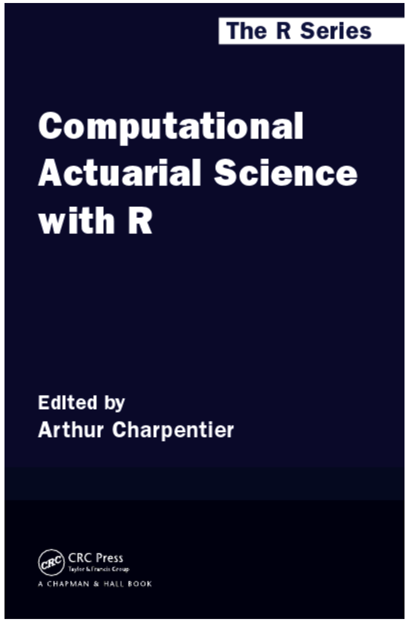








Take-Away Conclusion



Kendrick (2006), about computational economics: "*our thesis is that computational economics offers a way to improve this situation and to bring new life into the teaching of economics in colleges and universities* [...] *computational economics provides an opportunity for some students to move away from too much use of the lecture-exam paradigm and more use of a laboratory-paper paradigm in teaching under graduate economics. This opens the door for more creative activity on the part of the students by giving them models developed by previous generations and challenging them to modify those models.*"

It is probably the same about computational actuarial science, thanks to R...

Take-Away Conclusion

Efron (2004) claimed that "*Bayes rule is a very attractive way of reasoning, and fun to use, but using Bayes rule doesn't make one a Bayesian*".

Bayesian models offer an interesting alternative to standard statistical techniques, on small datasets as well as on large ones (see applications to hierarchical and longitudinal models).

Computational issues are not that complicated... once you get used to the bayesian way of seen a statistical model.