Bayesian Statistics: A Brief Introduction

Hedibert F. Lopes[1]
hedibert.org

[1]Professor of Statistics and Econometrics at Insper, São Paulo.

Outline

Bayesian paradigm

- Combination of different sources/levels of information
- Sequential update of beliefs
- A single, coherent framework for
  - Statistical inference/learning
  - Model comparison/selection/criticism
  - Predictive analysis and decision making
- Drawback: Computationally challenging

Example 1: Is Diego ill?

- Diego claims some discomfort and goes to his doctor.
- His doctor believes he might be ill (he may have the flu).
- $\theta = 1$: Diego is ill.
- $\theta = 0$: Diego is not ill.
- θ is the “state of nature” or “proposition”

Adding some modeling

The doctor can take a binary and imperfect "test" X in order to learn about θ:

$$\begin{cases} P(X=1|\theta=0)=0.40, & \text{false positive} \\ P(X=0|\theta=1)=0.05, & \text{false negative} \end{cases}$$

These numbers might be based, say, on observed frequencies over the years and over several hospital in a given region.

$X = 1$ is observed

Data collection

The doctor performs the test and observes $X = 1$.

$X = 1$ is observed

Data collection
The doctor performs the test and observes $X = 1$.

Decision making
How should the doctor proceed?

$X = 1$ is observed

Data collection
The doctor performs the test and observes $X = 1$.

Decision making
How should the doctor proceed?

Maximum likelihood estimation
Since

$$0.95 = P(X = 1|\theta = 1) >> P(X = 1|\theta = 0) = 0.40$$

a maximum likelihood argument estimates that $\hat{\theta} = 1$.

$X = 1$ is observed

Data collection
The doctor performs the test and observes $X = 1$.

Decision making
How should the doctor proceed?

Maximum likelihood estimation
Since

$$0.95 = P(X = 1|\theta = 1) >> P(X = 1|\theta = 0) = 0.40$$

a maximum likelihood argument estimates that $\hat{\theta} = 1$.

$\Rightarrow$ Diego is believed to have disease A.

Bayesian learning

Suppose the doctor claims that

$$P(\theta = 1) = 0.70$$

Bayesian learning

Suppose the doctor claims that

$$P(\theta = 1) = 0.70$$

This information can be based on the doctor's sole experience or based on existing health department summaries or any other piece of existing historical information.

Bayesian learning

Suppose the doctor claims that

$$P(\theta = 1) = 0.70$$

This information can be based on the doctor's sole experience or based on existing health department summaries or any other piece of existing historical information.

Overall rate of positives
The doctor can anticipate the overall rate of positive tests:

$$\begin{aligned} P(X = 1) &= P(X = 1|\theta = 0)P(\theta = 0) \\ &+ P(X = 1|\theta = 1)P(\theta = 1) \\ &= (0.4)(0.3) + (0.95)(0.7) = 0.785 \end{aligned}$$

Turning the Bayesian crank

Once $X = 1$ is observed, i.e. once Diego is submitted to the test X and the outcome is $X = 1$, what is the probability that Diego is ill?

Turning the Bayesian crank

Once $X = 1$ is observed, i.e. once Diego is submitted to the test X and the outcome is $X = 1$, what is the probability that Diego is ill?

Common (and wrong!) answer: $P(X = 1|\theta = 1) = 0.95$

Turning the Bayesian crank

Once $X = 1$ is observed, i.e. once Diego is submitted to the test X and the outcome is $X = 1$, what is the probability that Diego is ill?

Common (and wrong!) answer: $P(X = 1|\theta = 1) = 0.95$

Correct answer: $P(\theta = 1|X = 1)$

Turning the Bayesian crank

Once $X = 1$ is observed, i.e. once Diego is submitted to the test X and the outcome is $X = 1$, what is the probability that Diego is ill?

Common (and wrong!) answer: $P(X = 1|\theta = 1) = 0.95$

Correct answer: $P(\theta = 1|X = 1)$

Simple probability identity (Bayes' rule):

$$\begin{aligned} P(\theta = 1|X = 1) &= \frac{P(\theta = 1)P(X = 1|\theta = 1)}{P(X = 1)} \\ &= \frac{0.70 \times 0.95}{0.785} \\ &= 0.8471338 \end{aligned}$$

Combining both pieces of information

By combining

doctor's existing information $\quad + \quad$ data information

the updated probability that Diego is ill is 85%.

Combining both pieces of information

By combining

doctor's existing information $\quad + \quad$ data information

the updated probability that Diego is ill is 85%.

More generally,

$$\texttt{Posterior} = \frac{\texttt{Prior} \times \texttt{Likelihood}}{\texttt{Predictive}}$$

Posterior predictive

The doctor is still not convinced and decides to perform a second more reliable test (Y):

$$P(Y = 0|\theta = 1) = 0.01 \quad \text{versus} \quad P(X = 0|\theta = 1) = 0.05$$
$$P(Y = 1|\theta = 0) = 0.04 \quad \text{versus} \quad P(X = 1|\theta = 0) = 0.40$$

Posterior predictive

The doctor is still not convinced and decides to perform a second more reliable test (Y):

$$\begin{array}{lll} P(Y=0|\theta=1)=0.01 & \text{versus} & P(X=0|\theta=1)=0.05 \\ P(Y=1|\theta=0)=0.04 & \text{versus} & P(X=1|\theta=0)=0.40 \end{array}$$

Overall rate of positives
Once again, the doctor can anticipate the overall rate of positive tests, but now conditioning on $X=1$:

$$\begin{aligned} P(Y=1|X=1) &= P(Y=1|\theta=0)P(\theta=0|X=1) \\ &+ P(Y=1|\theta=1)P(\theta=1|X=1) \\ &= (0.04)(0.1528662)+(0.99)(0.8471338) \\ &= 0.8447771 \end{aligned}$$

$Y = 1$ is observed

Once again, Bayes rule leads to

$$
\begin{aligned}
P(\theta = 1|X = 1, Y = 1) &= \frac{P(Y = 1|\theta = 1)P(\theta = 1|X = 1)}{P(Y = 1|X = 1)} \\
&= \frac{(0.99)(0.8471338)}{0.8447771} \\
&= 99.2762\%
\end{aligned}
$$

$Y = 1$ is observed

Once again, Bayes rule leads to

$$\begin{aligned} P(\theta = 1|X = 1, Y = 1) &= \frac{P(Y = 1|\theta = 1)P(\theta = 1|X = 1)}{P(Y = 1|X = 1)} \\ &= \frac{(0.99)(0.8471338)}{0.8447771} \\ &= 99.2762\% \end{aligned}$$

$$P(\theta = 1|H) = \begin{cases} 70\% & , H\text{: before } X \text{ and } Y \\ 85\% & , H\text{: after } X = 1 \text{ and before } Y \\ 99\% & , H\text{: after } X = 1 \text{ and } Y = 1 \end{cases}$$

$Y = 1$ is observed

Once again, Bayes rule leads to

$$
\begin{aligned}
P(\theta = 1 | X = 1, Y = 1) &= \frac{P(Y = 1|\theta = 1)P(\theta = 1|X = 1)}{P(Y = 1|X = 1)} \\
&= \frac{(0.99)(0.8471338)}{0.8447771} \\
&= 99.2762\%
\end{aligned}
$$

$$
P(\theta = 1|H) = \begin{cases} 70\% & , H\text{: before } X \text{ and } Y \\ 85\% & , H\text{: after } X = 1 \text{ and before } Y \\ 99\% & , H\text{: after } X = 1 \text{ and } Y = 1 \end{cases}
$$

It is easy to see that $Pr(\theta = 1|Y = 1) = 98.2979\%$.

$Y = 1$ is observed

Once again, Bayes rule leads to

$$
\begin{aligned}
P(\theta = 1|X = 1, Y = 1) &= \frac{P(Y = 1|\theta = 1)P(\theta = 1|X = 1)}{P(Y = 1|X = 1)} \\
&= \frac{(0.99)(0.8471338)}{0.8447771} \\
&= 99.2762\%
\end{aligned}
$$

$$
P(\theta = 1|H) = \begin{cases} 70\% & , H\text{: before } X \text{ and } Y \\ 85\% & , H\text{: after } X = 1 \text{ and before } Y \\ 99\% & , H\text{: after } X = 1 \text{ and } Y = 1 \end{cases}
$$

It is easy to see that $Pr(\theta = 1|Y = 1) = 98.2979\%$.

Conclusion: Don't consider test X, unless it is "cost" free.

Example 2: Gaussian measurement error

Goal: Learn θ, a physical quantity.

Example 2: Gaussian measurement error

Goal: Learn θ, a physical quantity.

Measurement: X

Example 2: Gaussian measurement error

Goal: Learn θ, a physical quantity.

Measurement: X

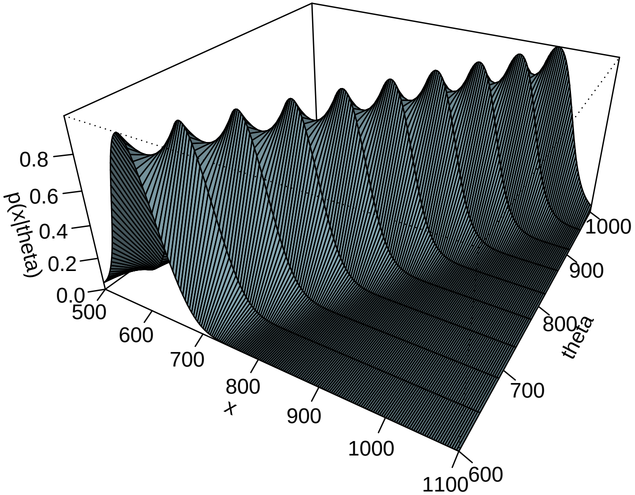
Model: $(X|\theta) \sim N(\theta, (40)^2)$

Example 2: Gaussian measurement error

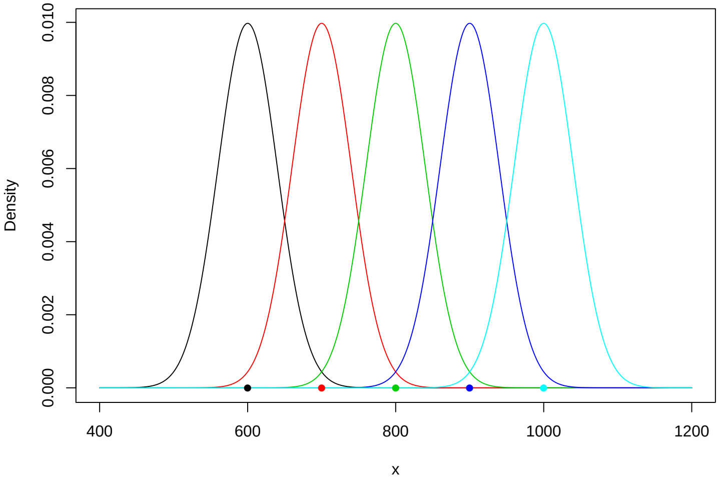
Goal: Learn θ, a physical quantity.

Measurement: X

Model: $(X|\theta) \sim N(\theta, (40)^2)$



$p(x|\theta)$ for $\theta \in \{600, 700, \ldots, 1000\}$



Large and small prior experience

Prior A: Physicist A (large experience): $\theta \sim N(900, (20)^2)$

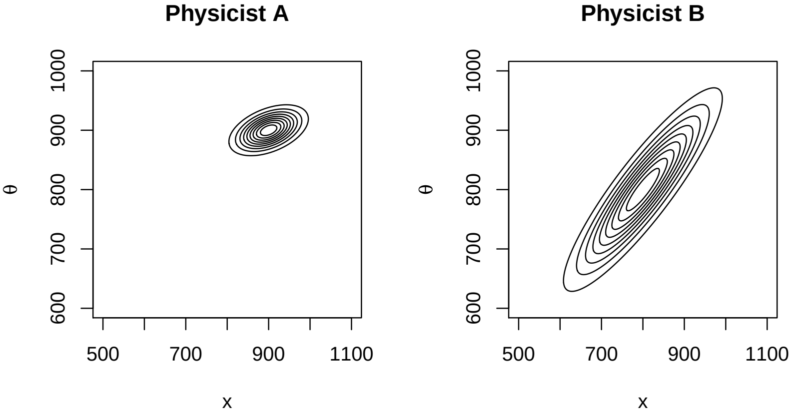
Large and small prior experience

Prior A: Physicist A (large experience): $\theta \sim N(900, (20)^2)$

Prior B: Physicist B (not so experienced): $\theta \sim N(800, (80)^2)$

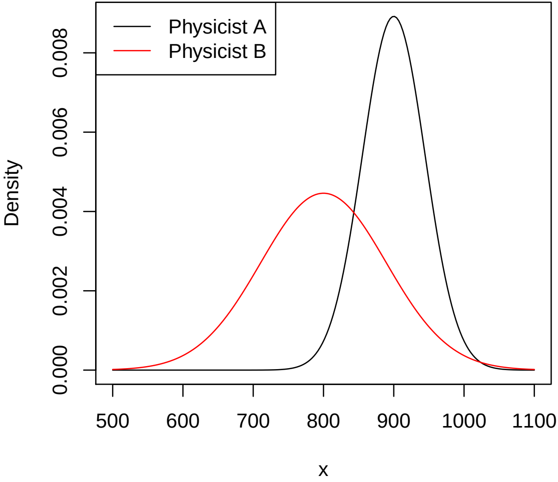
Large and small prior experience

Prior A: Physicist A (large experience): $\theta \sim N(900, (20)^2)$
Prior B: Physicist B (not so experienced): $\theta \sim N(800, (80)^2)$
Joint density: $p(x, \theta) = p(x|\theta)p(\theta)$

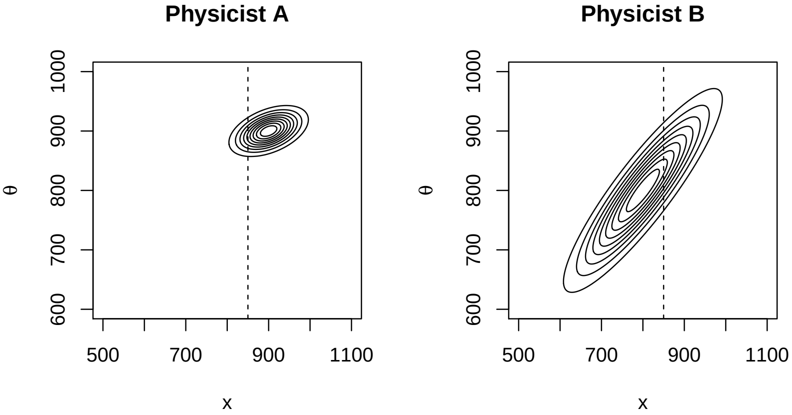


Predictive densities

$$p(x) = \int_{-\infty}^{\infty} p(x|\theta)p(\theta)d\theta$$



Observation: $X = 850$



Posterior (updated) densities

Physicist A
After observing $x = 850$, if follows that

$$(\theta|X = 850) \sim N(890, (17.9)^2)$$

against the prior $\theta \sim N(900, (20)^2)$

Posterior (updated) densities

Physicist A
After observing $x = 850$, if follows that

$$(\theta | X = 850) \sim N(890, (17.9)^2)$$

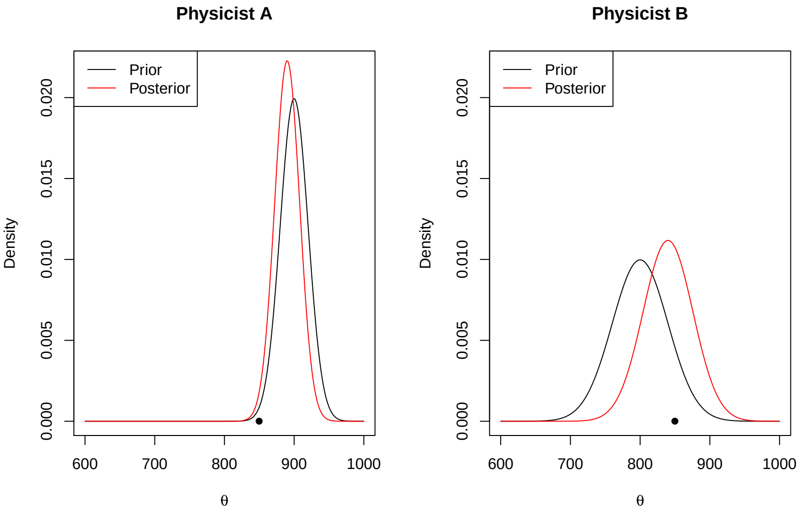
against the prior $\theta \sim N(900, (20)^2)$

Physicist B
After observing $x = 850$, if follows that

$$(\theta | X = 850) \sim N(840, (35.7)^2)$$

against the prior $\theta \sim N(800, (40)^2)$

Priors and posteriors



Bayesian computation: predictive

Prior: $\theta \sim N(\theta_0, \tau_0^2)$
Model: $x|\theta \sim N(\theta, \sigma^2)$

Bayesian computation: predictive

Prior: $\theta \sim N(\theta_0, \tau_0^2)$
Model: $x|\theta \sim N(\theta, \sigma^2)$

$$
\begin{aligned}
p(x) &= \int_{-\infty}^{\infty} p(x|\theta)p(\theta)d\theta \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\tau_0^2}} e^{-\frac{(\theta-\theta_0)^2}{2\tau_0^2}} d\theta \\
&= \frac{1}{\sqrt{2\pi(\sigma^2+\tau_0^2)}} e^{-\frac{(x-\theta)^2}{2(\sigma^2+\tau_0^2)}}
\end{aligned}
$$

or

$$x \sim N(\theta_0, \sigma^2 + \tau_0^2)$$

Bayesian computation: posterior

$$
\begin{aligned}
p(\theta|x) &= \frac{p(x|\theta)p(\theta)}{p(x)} \propto p(x|\theta)p(\theta) \\
&= (2\pi\sigma^2)^{-1/2} e^{-\frac{(x-\theta)^2}{2\sigma^2}} (2\pi\tau_0^2)^{-1/2} e^{-\frac{(\theta-\theta_0)^2}{2\tau_0^2}} \\
&\propto \exp\left\{-\frac{1}{2}\left[(\theta^2 - 2\theta x)/\sigma^2 + (\theta^2 - 2\theta\theta_0)/\tau_0^2)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\tau_1^2}(\theta-\theta_1)^2\right\}
\end{aligned}
$$

or

$$\theta|x \sim N(\theta_1, \tau_1^2)$$

where

$$\theta_1 = \left(\frac{\sigma^2}{\sigma^2+\tau_0^2}\right)\theta_0 + \left(\frac{\tau_0^2}{\sigma^2+\tau_0^2}\right)x \quad \text{and} \quad \tau_1^2 = \tau_0^2\left(\frac{\sigma^2}{\sigma^2+\tau_0^2}\right)$$

Combination of information

Let

$$\pi = \frac{\sigma^2}{\sigma^2 + \tau_0^2} \in (0, 1)$$

Therefore,

$$E(\theta|x) = \pi E(\theta) + (1 - \pi)x$$

and

$$V(\theta|x) = \pi V(\theta)$$

When τ_0^2 is much larger than σ^2, $\pi \approx 0$ and the posterior collapses at the observed value x!

Bayesian computational statistics

Deriving the posterior (via Bayes rule)

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

and computing the predictive

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

can become very challenging!

Bayesian computational statistics

Deriving the posterior (via Bayes rule)

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

and computing the predictive

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

can become very challenging!

Bayesian computation was done on limited, unrealistic models until the Monte Carlo revolution (and the computing revolution) of the late 1980's and early 1990's.

A more conservative physicist

Prior A: Physicist A (large experience): $\theta \sim N(900, 400)$

Prior B: Physicist B (not so experienced): $\theta \sim N(800, 1600)$

A more conservative physicist

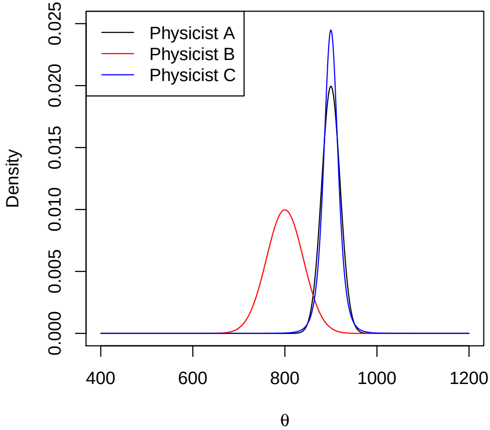
Prior A: Physicist A (large experience): $\theta \sim N(900, 400)$
Prior B: Physicist B (not so experienced): $\theta \sim N(800, 1600)$
Prior C: Physicist C (largeR experience): $\theta \sim t_5(900, 240)$

$$V(\text{Prior C}) = \frac{5}{5-2} 240 = 400 = V(\text{Prior A})$$

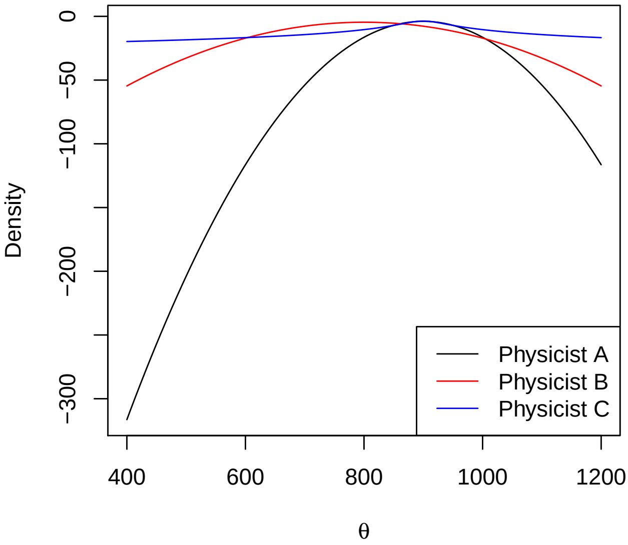
A more conservative physicist

Prior A: Physicist A (large experience): $\theta \sim N(900, 400)$
Prior B: Physicist B (not so experienced): $\theta \sim N(800, 1600)$
Prior C: Physicist C (largeR experience): $\theta \sim t_5(900, 240)$

$$V(\text{Prior C}) = \frac{5}{5-2}240 = 400 = V(\text{Prior A})$$



Closer look at the tails



Predictive and posterior of physicist C

For model $x|\theta \sim N(\theta, \sigma^2)$ and prior of $\theta \sim t_\nu(\theta_0, \tau^2)$, the integral

$$p(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\tau_0^2}} \left(1 + \frac{1}{\nu}\left(\frac{\theta-\theta_0}{\tau_0}\right)^2\right)^{-\frac{\nu+1}{2}} d\theta$$

is not analytically available.

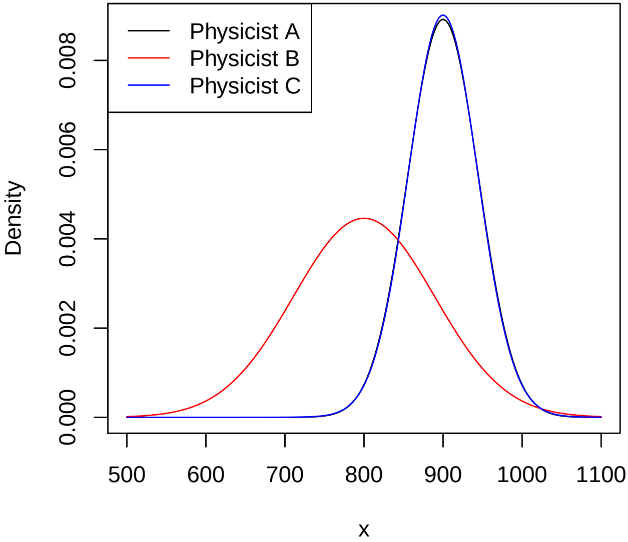
Similarly,

$$p(\theta|x) \propto \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}\left(1 + \frac{1}{\nu}\frac{(\theta-\theta_0)^2}{\tau_0^2}\right)^{-\frac{\nu+1}{2}}$$

is of no known form.

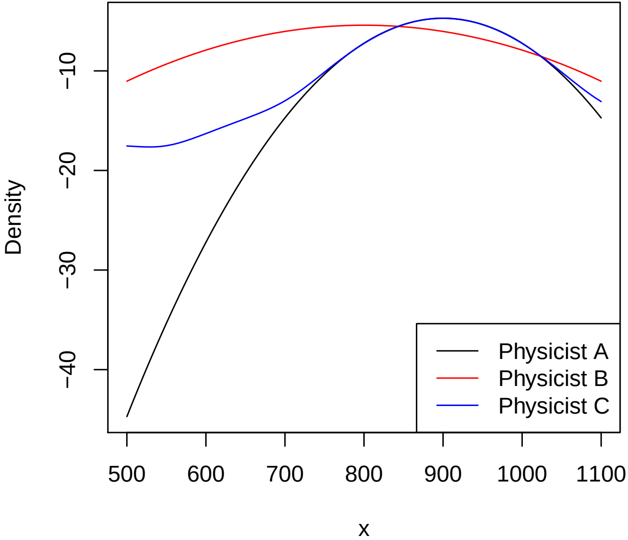
Predictives

Monte Carlo approximation to $p(x)$ for physicist C.



Log predictives

Physicist C has similar knowledge as physicist A, but does not rule out smaller values for x.



Monte Carlo integration

The integral

$$p(x) = \int p(x|\theta)p(\theta)d\theta = E_{p(\theta)}\{p(x|\theta)\}$$

can be approximated by Monte Carlo as

$$\hat{p}_{MC}(x) = \frac{1}{M}\sum_{i=1}^{M} p(x|\theta^{(i)})$$

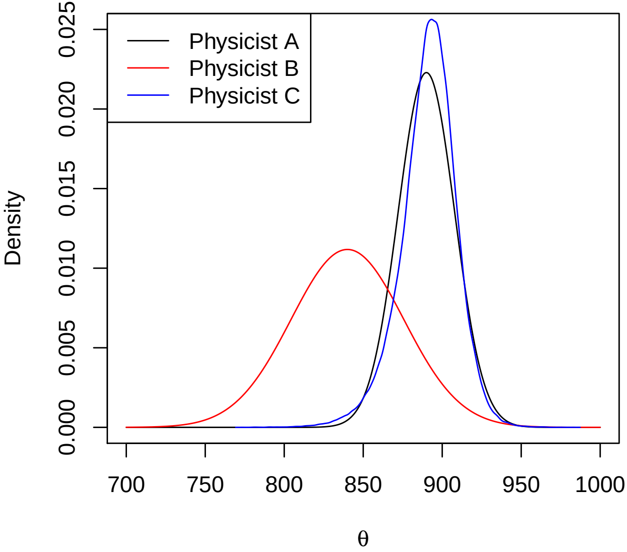
where

$$\{\theta^{(1)}, \ldots, \theta^{(M)}\} \sim p(\theta)$$

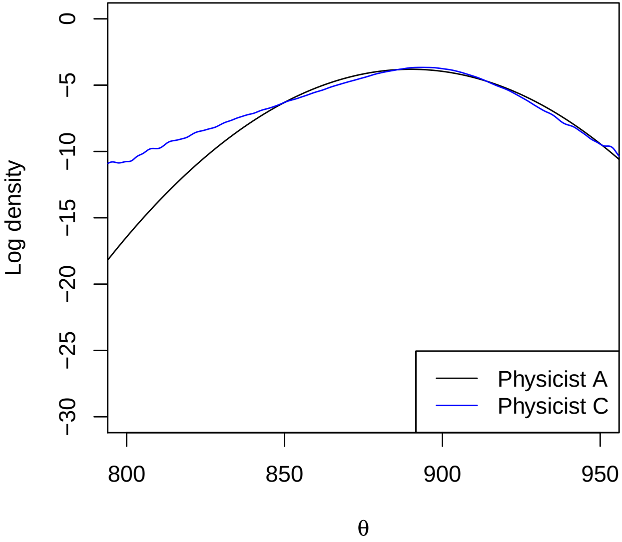
We used $M = 1,000,000$ draws in the previous two plots.

Posteriors for θ

Monte Carlo approximation to $p(\theta|x)$ for physicist C.



Log posteriors



Monte Carlo simulation via SIR

Sampling importance resampling (SIR) is a well-known MC tool that resamples draws from a candidate density $q(\cdot)$ to obtains draws from a target density $\pi(\cdot)$.

Monte Carlo simulation via SIR

Sampling importance resampling (SIR) is a well-known MC tool that resamples draws from a candidate density $q(\cdot)$ to obtains draws from a target density $\pi(\cdot)$.

SIR Algorithm:

1. Draws $\{\theta^{(i)}\}_{i=1}^{M}$ from candidate density $q(\cdot)$
2. Compute resampling weights: $w^{(i)} \propto \pi(\theta^{(i)})/q(\theta^{(i)})$
3. Sample $\{\tilde{\theta}^{(j)}\}_{j=1}^{N}$ from $\{\theta^{(i)}\}_{i=1}^{M}$ with weights $\{w^{(i)}\}_{i=1}^{M}$.

Monte Carlo simulation via SIR

Sampling importance resampling (SIR) is a well-known MC tool that resamples draws from a candidate density $q(\cdot)$ to obtains draws from a target density $\pi(\cdot)$.

SIR Algorithm:

1. Draws $\{\theta^{(i)}\}_{i=1}^{M}$ from candidate density $q(\cdot)$
2. Compute resampling weights: $w^{(i)} \propto \pi(\theta^{(i)})/q(\theta^{(i)})$
3. Sample $\{\tilde{\theta}^{(j)}\}_{j=1}^{N}$ from $\{\theta^{(i)}\}_{i=1}^{M}$ with weights $\{w^{(i)}\}_{i=1}^{M}$.

Result: $\{\tilde{\theta}^{(1)}, \ldots, \tilde{\theta}^{(N)}\} \sim \pi(\theta)$

Bayesian bootstrap

When ...

- the target density is the posterior $p(\theta|x)$, and
- the candidate density is the prior $p(\theta)$, then
- the weight is the likelihood $p(x|\theta)$:

$$w^{(i)} \propto \frac{p(\theta^{(i)})p(x|\theta^{(i)})}{p(\theta^{(i)})} = p(x|\theta^{(i)})$$

Note: We used $M = 10^6$ and $N = 0.1M$ in the previous two plots.

MC is expensive!

Exact solution

$$I = \int_{-\infty}^{\infty} \exp\{-0.5\theta^2\} d\theta = \sqrt{2\pi} = 2.506628275$$

MC is expensive!

Exact solution

$$I = \int_{-\infty}^{\infty} \exp\{-0.5\theta^2\} d\theta = \sqrt{2\pi} = 2.506628275$$

Let us assume that

$$I = \int_{-\infty}^{\infty} \exp\{-0.5\theta^2\} d\theta = \int_{-5}^{5} \exp\{-0.5\theta^2\} d\theta$$

MC is expensive!

Exact solution

$$I = \int_{-\infty}^{\infty} \exp\{-0.5\theta^2\} d\theta = \sqrt{2\pi} = 2.506628275$$

Let us assume that

$$I = \int_{-\infty}^{\infty} \exp\{-0.5\theta^2\} d\theta = \int_{-5}^{5} \exp\{-0.5\theta^2\} d\theta$$

Grid approximation (less than 0.01 seconds to run)

For $\theta_1 = -5$ $\theta_2 = -5 + \Delta$, $\ldots, \theta_{1001} = 5$ and $\Delta = 0.01$,

$$\hat{I}_{hist} = \sum_{i=1}^{1001} \exp\{-0.5\theta_i^2\}\Delta = 2.506626875$$

MC integration

It is easy to see that

$$\begin{aligned}\int_{-5}^{5} \exp\{-0.5\theta^2\} d\theta &= \int_{-5}^{5} 10 \exp\{-0.5\theta^2\} \frac{1}{10} d\theta \\ &= E_{U(-5,5)} \left[10 \exp\{-0.5\theta^2\} \right]\end{aligned}$$

MC integration

It is easy to see that

$$\int_{-5}^{5} \exp\{-0.5\theta^2\} d\theta = \int_{-5}^{5} 10 \exp\{-0.5\theta^2\} \frac{1}{10} d\theta$$
$$= E_{U(-5,5)} \left[10 \exp\{-0.5\theta^2\}\right]$$

Therefore, for $\{\theta^{(i)}\}_{i=1}^{M} \sim U(-5,5)$,

$$\hat{I}_{MC} = \frac{1}{M} \sum_{i=1}^{M} 10 \exp\{-0.5\theta^{(i)2}\}$$

M	$\hat{I}_{MC}$	MC error
1,000	2.505392026	0.10640840352
10,000	2.507470696	0.03380205878
100,000	2.506948869	0.01067906810

To improve on digital point, one needs M^2 draws!

It takes about 0.02 seconds to run.

Monte Carlo methods

- They are expensive.
- They are scalable.
- Readily available MC error bounds.

Why not simply use deterministic approximations?

Let us consider the bidimensional integral, for $\theta = (\theta_1, \theta_2, \theta_3)$,

$$I = \int \exp\{-0.5\theta'\theta\} d\theta = (2\pi)^{3/2} = 15.74960995$$

Grid approximation (20 seconds)

$$\hat{I}_{hist} = \sum_{i=1}^{1001} \sum_{j=1}^{1001} \sum_{k=1}^{1001} \exp\{-0.5(\theta_{1i}^2 + \theta_{2j}^2 + \theta_{3k}^2)\}\Delta^3 = 15.74958355$$

Monte Carlo approximation (0.02 seconds)

M	$\hat{I}_{MC}$	MC error
1,000	15.75223328	2.2768286659
10,000	15.72907660	0.7515860214
100,000	15.75368350	0.2236006764

Gibbs sampler

The Gibbs sampler is the most famous of the Markov chain Monte Carlo methods.

Roughly speaking, one can sample from the joint posterior of $(\theta_1, \theta_2, \theta_3)$

$$p(\theta_1, \theta_2, \theta_3 | y)$$

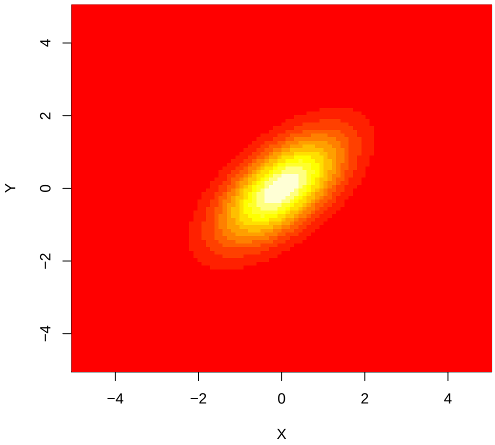
by iteratively sampling from the full conditional distributions

$$p(\theta_1 | \theta_2, \theta_3, y)$$
$$p(\theta_2 | \theta_1, \theta_3, y)$$
$$p(\theta_3 | \theta_1, \theta_1, y)$$

After a *warm up* phase, the draws will behave as coming from posterior distribution.

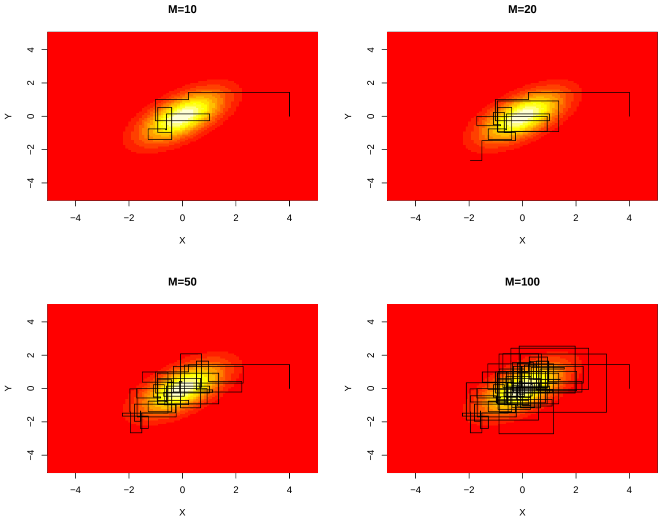
Taget distribution: bivariate normal with $\rho = 0.6$

$$p(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy - y^2}{2(1-\rho^2)}\right\}$$



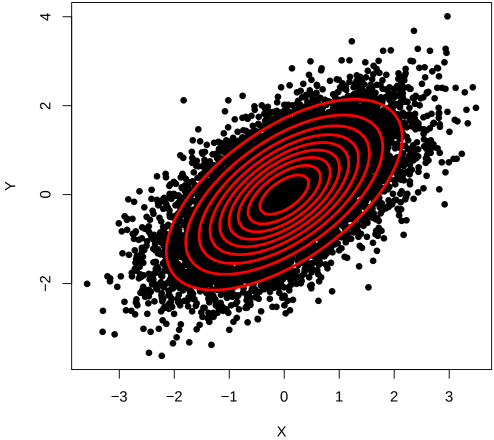
Full conditional distributions

Easy to see that $x|y \sim N(\rho y, 1-\rho^2)$ and $y|x \sim N(\rho x, 1-\rho^2)$.
Initial value: $x^{(0)} = 4$

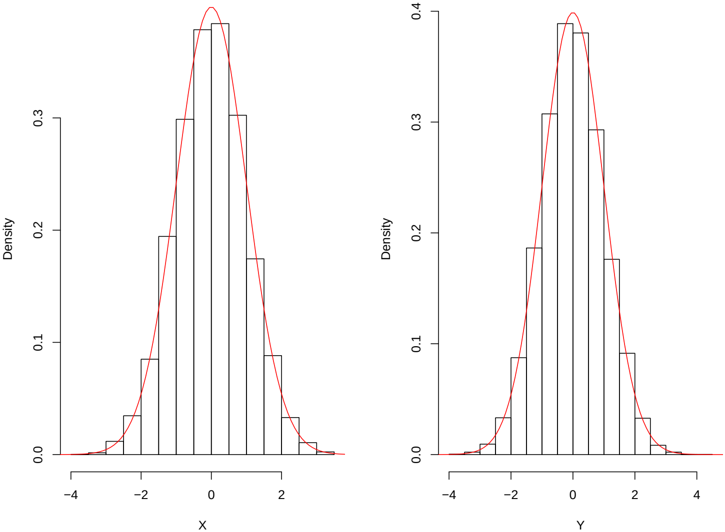


Posterior draws

Running the Gibbs sampler for 11,000 iterations and discarding the first 1,000 draws.



Marginal posterior distributions



Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm is, in fact, more general than the Gibbs sampler and older (1950's).

One can sample from the joint posterior $p(\theta_1, \theta_2, \theta_3|y)$ by iteratively sampling θ_1^* from a proposal density $q_1(\cdot)$ and accepting the draw with probability

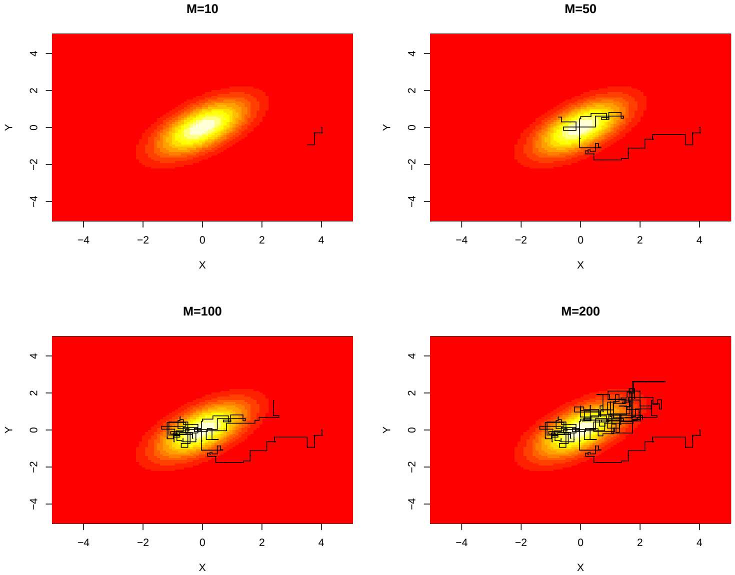
$$\min\left\{1, \frac{p(\theta_1^*, \theta_2, \theta_3|y)}{p(\theta_1, \theta_2, \theta_3|y)} \frac{q_1(\theta_1)}{q_1(\theta_1^*)}\right\},$$

with θ_2 and θ_3 fixed at the final draws from the previous iteration. The steps are repeated for θ_2^* and θ_3^*.

After a *warm up* phase, the draws will behave as coming from posterior distribution.

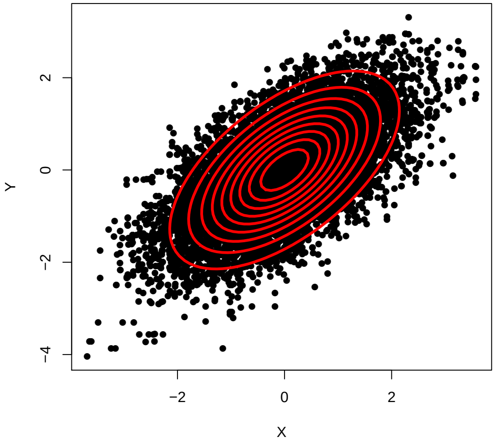
Random-walk Metropolis algorithm

The proposals are $x^* \sim N(x^{old}, 0.25)$ and $y^* \sim N(y^{old}, 0.25)$

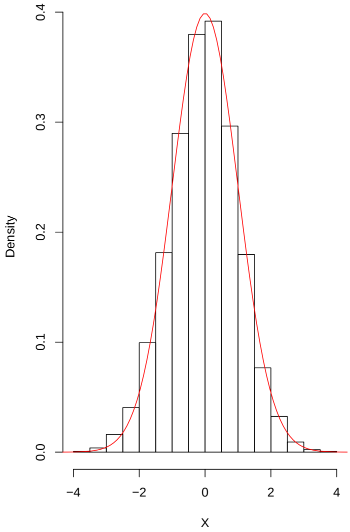


Posterior draws

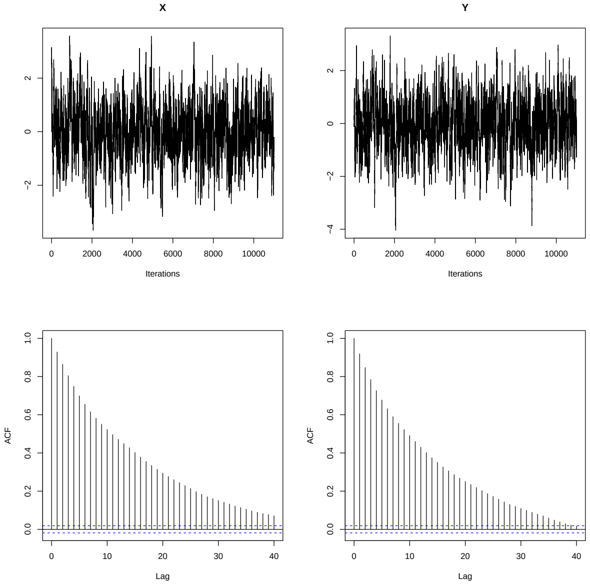
Running the Metropolis-Hastings algorithm for 11,000 iterations and discarding the first 1,000 draws.



Marginal posterior distributions

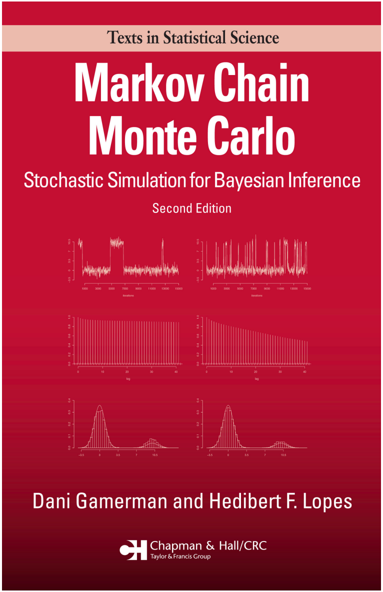


Markov chains and autocorrelation



Want to learn more?

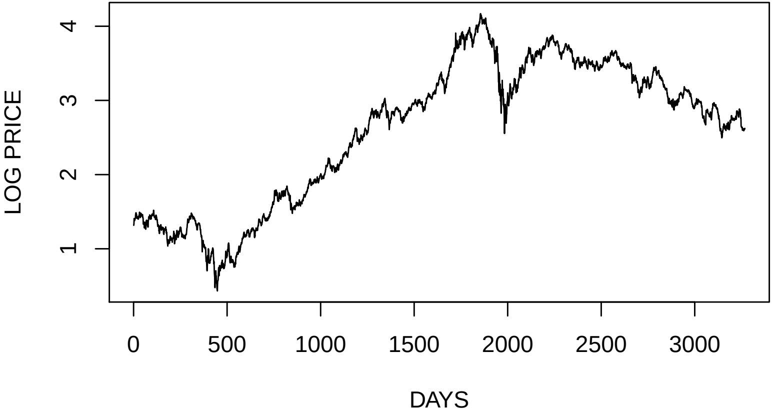
hedibert.org has a link to book webpage.



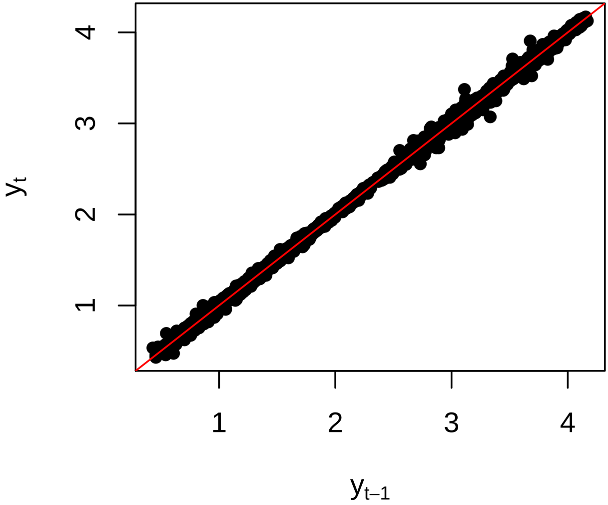
Example 3: Time-varying variance modeling

Modeling Petrobrás' log-returns

Time span: 12/29/2000 - 12/31/2013 ($n = 3268$ days)

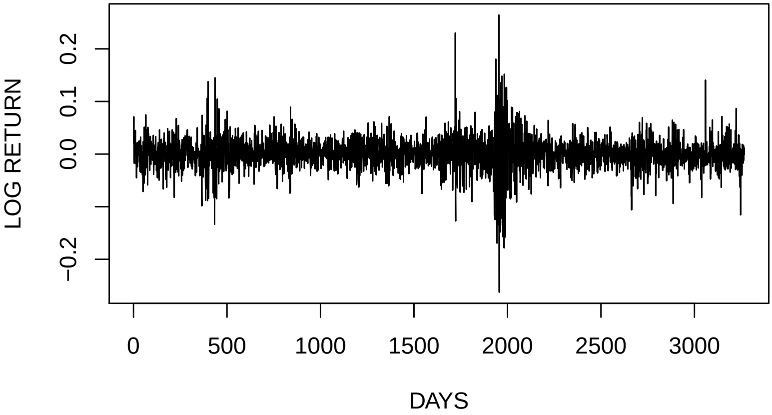


Scatterplot of y_{t-1} versus y_t

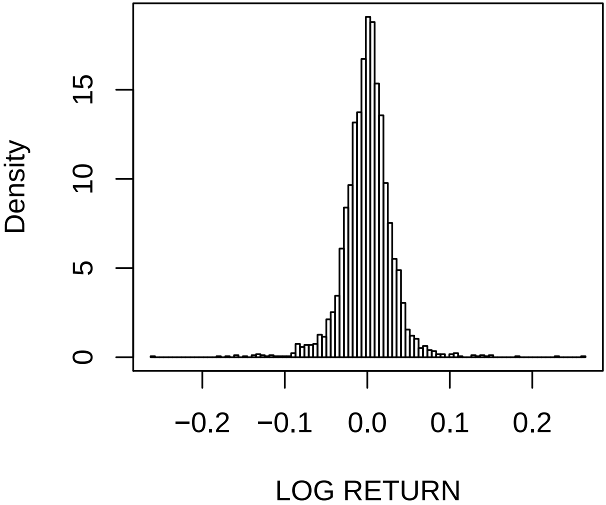


Log return: $r_t = y_t - y_{t-1} = \log(p_t/p_{t-1})$

Time span: 01/02/2001 - 12/31/2013 ($n = 3267$ days)



Histogram of r_t



Training and testing samples

Years 2001-2006:
The first $n_0 = 1506$ days are used for prior specification.

Years 2007-2013:
The last $n = 1760$ days are used for posterior inference.

GARCH(1,1) with t errors

The GARCH(1,1) model with Student-t innovations:

$$
\begin{aligned}
r_t &\sim t_\nu(0, \rho h_t) \\
h_t &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta h_{t-1},
\end{aligned}
$$

where $\alpha_0 > 0$, $\alpha_1 \geq 0$ and $\beta > 0$.

We set the initial variance to $h_0 = 0$ for convenience.

We let $\rho = (\nu - 2)/\nu$ so that

$$V(r_t | h_t) = \frac{\nu}{\nu - 2} \rho h_t = h_t.$$

Prior

Let $\psi = (\alpha', \beta, \nu)'$ and $\alpha = (\alpha_0, \alpha_1)'$.

The prior distribution of ψ is such that

$$p(\alpha, \beta, \mu) = p(\alpha)p(\beta)p(\nu)$$

where

$$\begin{aligned} \alpha &\sim N_2(\mu_\alpha, \Sigma_\alpha) I_{(\alpha>0)} \\ \beta &\sim N(\mu_\beta, \Sigma_\beta) I_{(\beta>0)} \end{aligned}$$

and

$$p(\nu) = \lambda \exp\{-\lambda(\nu - \delta)\} I_{(\lambda>\delta)}$$

for $\lambda > 0$ and $\delta \geq 2$, such that $E(\nu) = \delta + 1/\lambda$.

Normal case: $\lambda = 100$ and $\delta = 500$.

bayesGARCH

`bayesGARCH:` Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations

```
bayesGARCH(r,mu.alpha = c(0,0),Sigma.alpha=1000*diag(1,2),
           mu.beta=0,Sigma.beta=1000,
           lambda=0.01,delta=2,control=list())
```

Paper: Ardia and Hoogerheide (2010) Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations. *The R Journal*, 2,41-47.

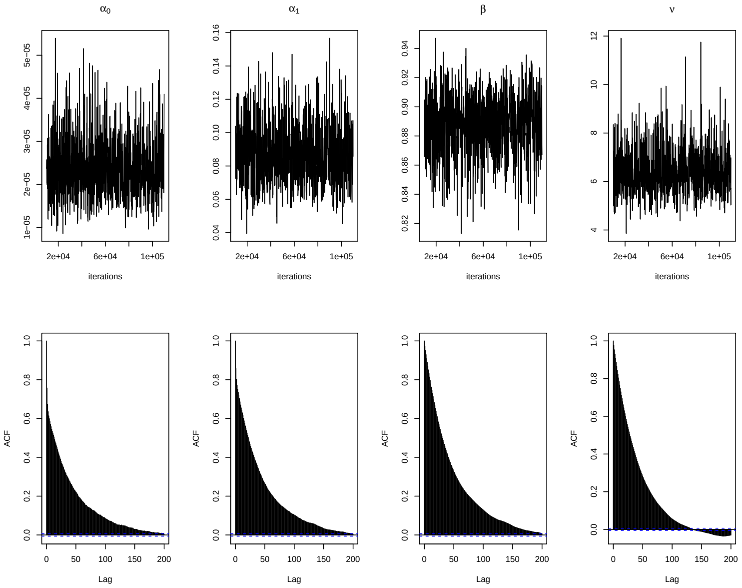
`http://cran.r-project.org/web/packages/bayesGARCH`

Example of R script

Recall that r_0 are Petrobras' returns for the first part of the data.

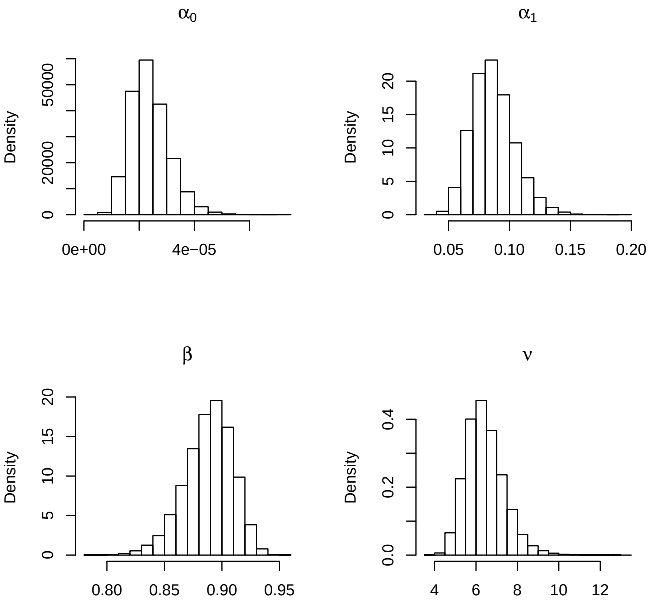
```
M0    = 10000       # to be discarded (burn-in)
M     = 10000       # kept for posterior inference
niter = M0+M

MCMC.initial = bayesGARCH(r0,mu.alpha=c(0,0),Sigma.alpha=1000*diag(1,2),
                          mu.beta=0,Sigma.beta=1000,lambda=0.01,delta=2,
                          control=list(n.chain=1,l.chain=niter,refresh=100))

draws = MCMC.initial$chain1

range = (M0+1):niter

par(mfrow=c(2,2))
ts.plot(draws[range,1],xlab="iterations",main=expression(alpha[0]),ylab="")
ts.plot(draws[range,2],xlab="iterations",main=expression(alpha[1]),ylab="")
ts.plot(draws[range,3],xlab="iterations",main=expression(beta),ylab="")
ts.plot(draws[range,4],xlab="iterations",main=expression(nu),ylab="")
```

MCMC output



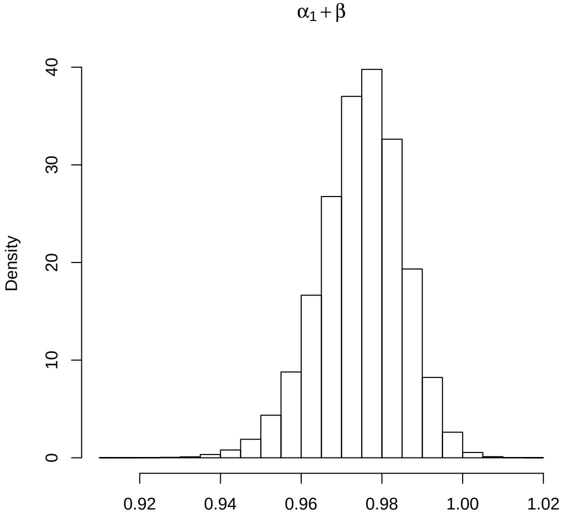
α_0
α_1
β
ν
iterations
iterations
iterations
iterations
ACF
ACF
ACF
ACF
Lag
Lag
Lag
Lag

Marginal posterior distributions



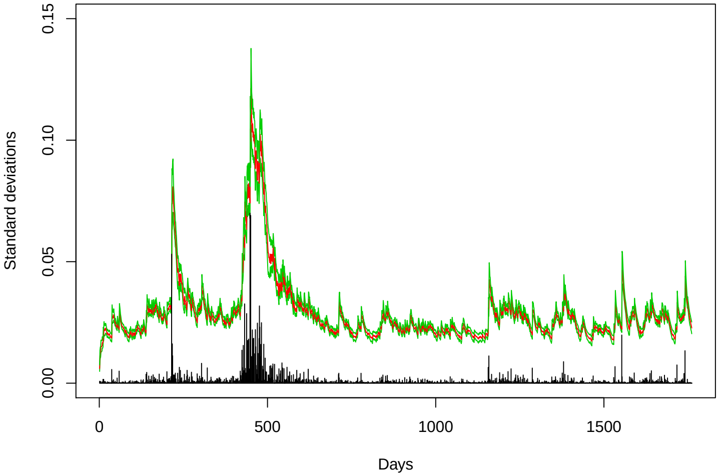
$p(\alpha_1 + \beta|\mathtt{data})$

$$Pr(\alpha_1 + \beta > 1|\mathtt{data}) = 0.0034$$



Quantiles from $p(h_t^{1/2}|\texttt{data})$

Percentiles 2.5%, 50% and 97.5% of $p(h_t^{1/2}|\texttt{data})$
Black vertical lines: r_t^2



Final remarks

- Model and prior are equally important
- Monte Carlo methods are here to stay
- Bayesian approach is the same across model complexity
- More flexibility to cycles between exploratory data analysis, modeling and inference

Final remarks

- Model and prior are equally important
- Monte Carlo methods are here to stay
- Bayesian approach is the same across model complexity
- More flexibility to cycles between exploratory data analysis, modeling and inference
- It pays to be Bayes!