Introduction to MCMC

Alberto Lumbreras

February 4, 2016

Overview

Introduction to Bayesian inference

Basic Monte Carlo methods

Markov Chain Monte Carlo methods

Introduction to Bayesian inference

Basic Monte Carlo methods

Markov Chain Monte Carlo methods

Intro to Bayesian Inference ("inverse probability")

A controversial formula

"The theory of inverse probability is founded upon an error, and must be wholly rejected!" Fisher, 1925

"Reduces all probability to a subjective judgement." Fisher, 1956

"Fallacious rubbish!" Fisher about Laplace, 1958

"The more I consider it, the more clearly it would appear that I have been doing almost exactly what Bayes had done in the 18th century." Fisher, 1959

Intro to Bayesian Inference ("inverse probability")

Thomas Bayes' formula (one more time)

$$\underbrace{p(\theta|y)}_{\textit{posterior}} = \frac{\overbrace{p(y,\theta)}^{\text{joint probability}}}{\int_\theta p(y,\theta)} = \frac{\overbrace{p(y|\theta)}^{\textit{likelihood}}\ \overbrace{p(\theta)}^{\textit{prior}}}{\int_\theta p(y|\theta)p(\theta)} \propto \overbrace{p(y|\theta)}^{\textit{likelihood}}\ \overbrace{p(\theta)}^{\textit{prior}} \qquad (1)$$

- Imagine we want to use Bayes in a non-trivial problem...
- ...where the denominator is not tractable (**usually**)
- ...but we can **always** evaluate the likelihood and the prior.
- How can I get access to the posterior?

Intro to Bayesian Inference

Computers to the rescue

$$\underbrace{p(\theta|y)}_{posterior} \propto \overbrace{p(y|\theta)}^{likelihood}\ \overbrace{p(\theta)}^{prior} \tag{2}$$

- The denominator was just a normalizing factor.
- Therefore drawing samples from $\overbrace{p(y|\theta)}^{likelihood}\ \overbrace{p(\theta)}^{prior}$ is like drawing samples directly from the posterior.
- "*I wish I had an Intel Core i7 :_(*"
  Thomas Bayes, circa 1750.

Introduction to Bayesian inference

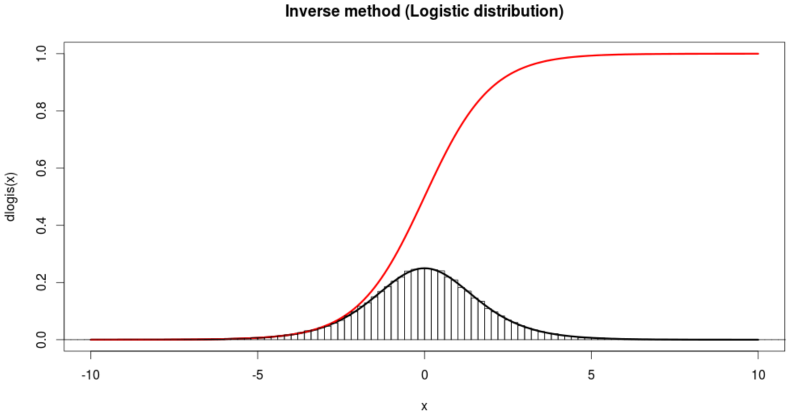
Basic Monte Carlo methods

Markov Chain Monte Carlo methods

Inverse method

Exploiting the Cumulative Distribution Function

1. Get the CDF: $\mathbb{R} \rightarrow [0,1]$
2. Draw samples from uniform distribution.
3. Map uniform samples to final samples through $CDF^{-1} : [0,1] \rightarrow \mathbb{R}$



Inverse method

Pros and cons

Pros:

- Faster possible method.

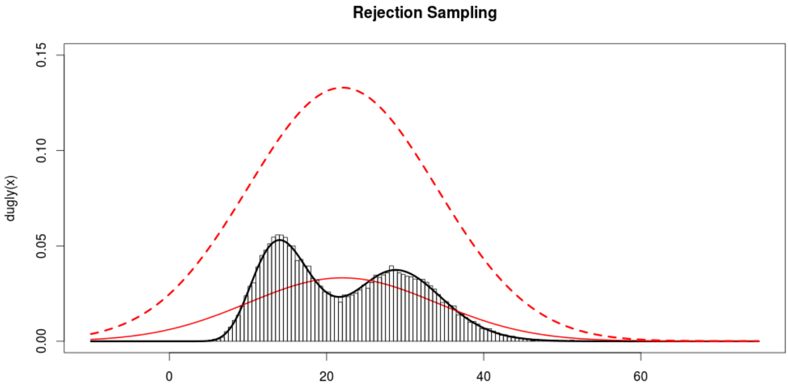
Cons:

- We don' have the CDF of most distributions.
- Only for unidimensional distributions.

Accept-Reject sampling

Sample from a comfort zone

1. Look for some pretty density q that you know how to can sample from.
2. Make it bigger than your ugly density p. (Kq)
3. Draw sample x_i from the pretty one.
4. Accept samples with probability $p(x_i)/(Kq(x_i))$



Accept-Reject sampling

Pros and cons

Pros:

- We can usually find a pretty envelope Kq.

Cons:

- If Kq is too wide or too big, lots of wasted (rejected) samples.
- Not always easy to find a good Kq.

Introduction to Bayesian inference

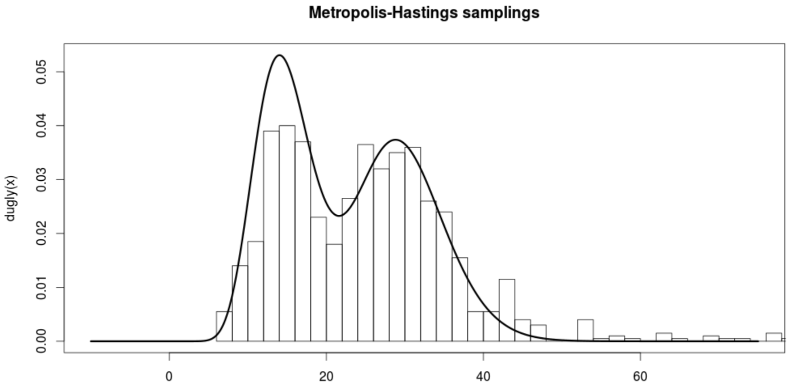
Basic Monte Carlo methods

Markov Chain Monte Carlo methods

Metropolis-Hastings

Jumping from last sample

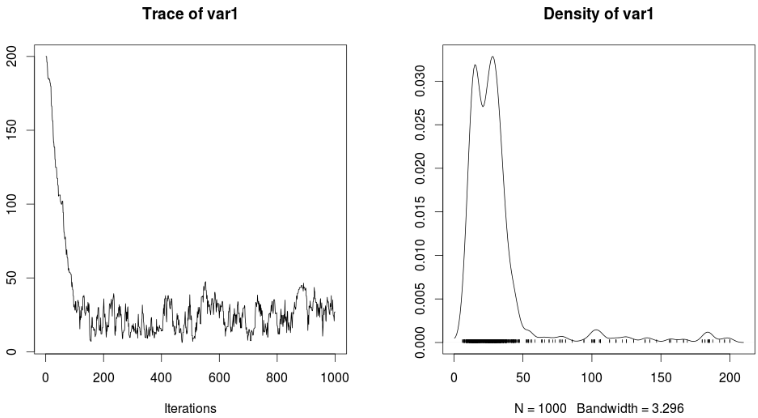
1. Start at some initial sample x_0
2. Random jump $x_{i+1} \sim x_i + \mathcal{N}(0, step)$
3. Accept with probability $p(x_{i+1})/p(x_i)$. If rejected, $x_{i+1} = x_i$
4. Repeat 2 and 3 N times (N is the length of your Markov Chain)



Metropolis-Hastings

About the Markov Chain

- First samples depend on x_0. We drop them (burn-in).
- After burn-in, we are sampling from the true distribution.
- Statistical convergence checks (see `coda` package in R).



Metropolis-Hastings

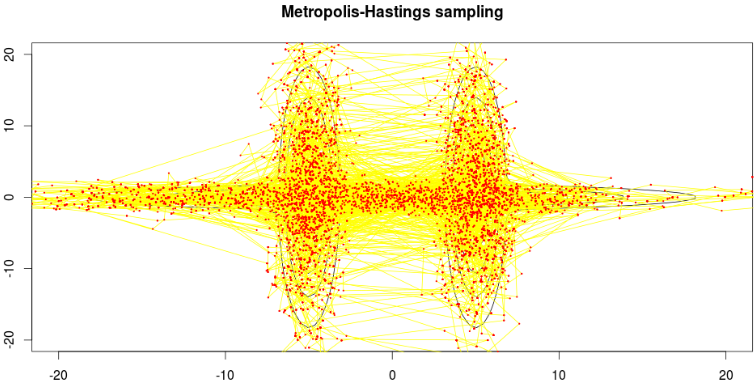
Pros and cons

Pros:

- Always works! (see ugly density below)
- Step size needs to be carefully tuned (for efficiency).

Cons:

- Rejects $\sim 70 - 80\%$ of samples (recommended).

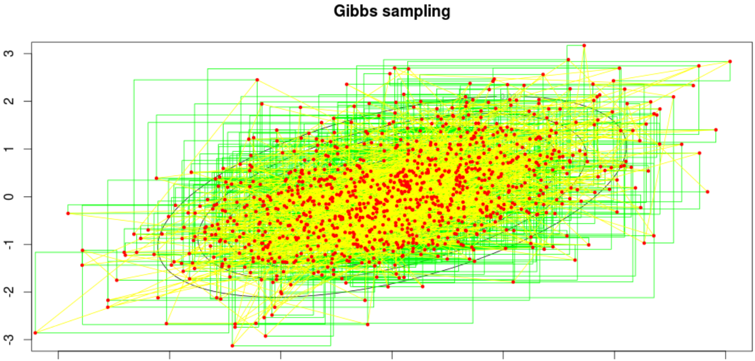


Gibbs sampling

Sample from pretty individual conditional probabilities.

Ugly $p(x1, x2, x3)$

1. Choose a initial x_1, x_2, x_3
2. Draw $x_1 \sim p(x_1|x_2, x_3)$ (easy!)
3. Draw $x_2 \sim p(x_2|x_1, x_3)$ (easy!)
4. Draw $x_3 \sim p(x_2|x_2, x_1)$ (easy!)
5. Repeat 1,2 and 3 N times.



Gibbs sampling

Prons and cons

Pros:

- Just a smart Metropolis-Hastings where all samples are accepted.

Cons:

- Only possible if conditional probabilities are pretty.
- (... and for that we need to use *conjugate priors*)

Wrap-up

- We sample because we can't do the maths analytically.
- Gibbs and Metropolis are a must in a Bayesian toolbox.

Caveat:

- burn-in can take toooo much.

Alternatives:

- Laplace approximations.
- **Variational Inference (approximation)** (very popular, very scalable)
- Expectation Propagation (approximation).
- ...

Readings

Historical debate:

- Aldrich, *R. A. Fisher on Bayes and Bayes' theorem*
- Fisher. *The Design of Experiments* (page 6)

MC Books and tutorials

- Kruschke. *Doing Bayesian Data Analysis.*
- Bishop. *Pattern Recognition and Machine Learning* (ch.11)
- Robert et Casella. *Méthodes de Monte Carlo avec R.*
- Gellman. *Bayesian Data Analysis.*
- Cam Davidson-Pilon. *Bayesian Methods for Hackers.*

Divulgation:

- Salsburg, D. *The lady tasting tea.*
- Silver, N. *The Signal and the Noise: Why So Many Predictions Fail –but Some Don't.*