Markov Chain Monte Carlo Methods

Eduardo Gomez

CSU, Chico

3/20/2019

Bayesian Inference

Bayes' Theorem:

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

Bayesian inference looks at our prior belief of a parameter θ and updates it with new data Y.

$$P(\theta|Y) = \frac{P(Y|\theta) * P(\theta)}{P(Y)}$$

$$\alpha_\theta \;\; P(Y|\theta) * P(\theta)$$

$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ \text{Posterior} & \text{Likelihood} & \text{Prior} \end{array}$$

Bayesian Inference

Example 1: Traffic stops in Connecticut

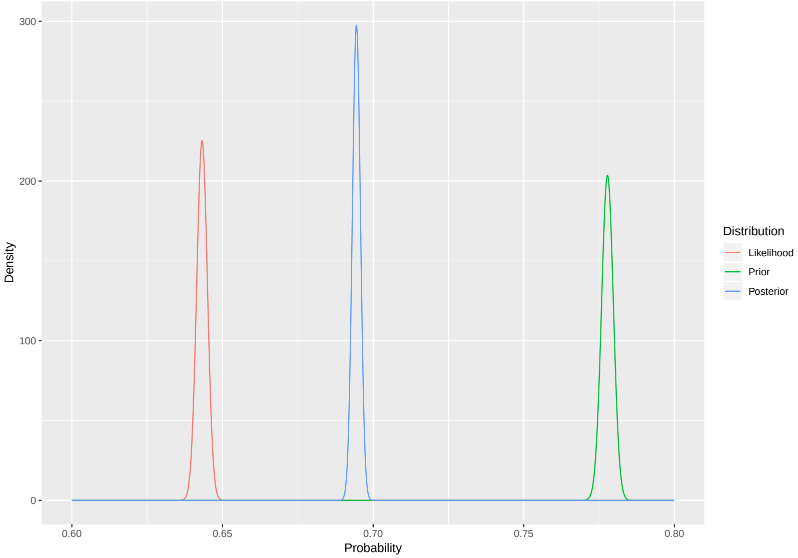
We're looking at the proportion of men that are stopped. Say we are pretty sure that the proportion is about 78%, which corresponds to a strong prior distribution. Let's give it a beta distribution. Because there are only two options for the officer to write down for their sex, the process, the likelihood, is a bernoulli distribution.

The posterior distribution is

$$P(\theta|Y) \;\; \alpha_\theta \;\; \theta^{35000+\Sigma(Y)-1} * (1-\theta)^{10000+N-\Sigma(Y)-1}$$

This is an example of a proper posterior distribution.

Bayesian Inference



Bayesian Inference

Example 2:

Maybe just talk about a Cauchy distribution. We use MCMC methods for this posterior distribution because the Cauchy distribution's expected value doesn't exist. So we can use MCMC methods to estimate the mean.

We have an integral for the posterior distribution that we can't find a normalizing constant for it to be a proper posterior.

That's all. Ask for help looking for an example?

MCMC

Using Markov chain Monte Carlo methods is a way to solve the previous example. They are primarily used to obtain a sample from a probability distribution when random sampling is hard, to approximate a distribution, or to compute an integral such as the expected value.

There are two aspects to these methods:

1. Markov Chains - describe a sequence of possible events/values where the probability of the next event/value is solely dependent on the current one (memoryless).
2. Monte Carlo methods - algorithms used to genereate random samples from the distribution in question to estimate its properties.

Explain how to come together for MCMC methods.

Metropolis-Hastings Algorithm

We have a target distribution f that we are trying to find some information about. We are going to create a chain of randomly generated values from a proposal distribution $g(y)$ (usually the normal distribution).

1 Generate the initial value x_0 from $g(y)$

2 Start a loop for the rest of the chain

   a Generate another value y from $g(y)$

   b We look at the probability of accepting the proposed value

$$\rho(x^t, y) = \frac{f(y) * g(x^t)}{f(x^t) * g(y)}$$

   where x^t is the value in the previous step in the chain and y is the proposed value

   c If some randomly generated number from the uniform distribution is less than the minimum of ρ and 1 then we accept the proposed value as the current value in the chain x^{t+1}, otherwise reject the proposed value and set $x^{t+1} = x^t$

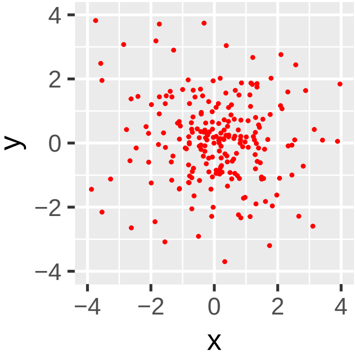
Hamiltonian Monte Carlo

Try to explain the physical process through the example of a bowl (?) State that we need a momentum variable as well as the position variable, which translates to our parameter of interest.

1. Generate a new value for the momentum variable, independent of the current value of the position variable.
2. Perform a Metropolis update.

Example 1: Generating a Cauchy Random Variable

Let's go back to the non-proper posterior distribution. Our target distribution is a two-dimension cauchy distribution. Here we have what a sample from the cauchy distribution looks like, what we expect.



Example 1: Generating a Cauchy Random Variable

Here is what out MCMC methods produce.