Bayesian Data Analysis Session 2

Rick Rijkse

10/2/2017

Recap

With Bayesian stats we put a distribution on the model parameters. They get updated by observed data.

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

It is almost always infeasable to obtain the posterior directly, because the marginal $P(D)$ is too complex to obtain.

We have seen that we can use sampling to normalize the posterior.

Many θs

Acceptance-rejection sampling impractical with many parameters. Why?

Alternative MCMC.

General idea:

$f(x)$ is proportional to posterior

θ_1 is some initialisation value.

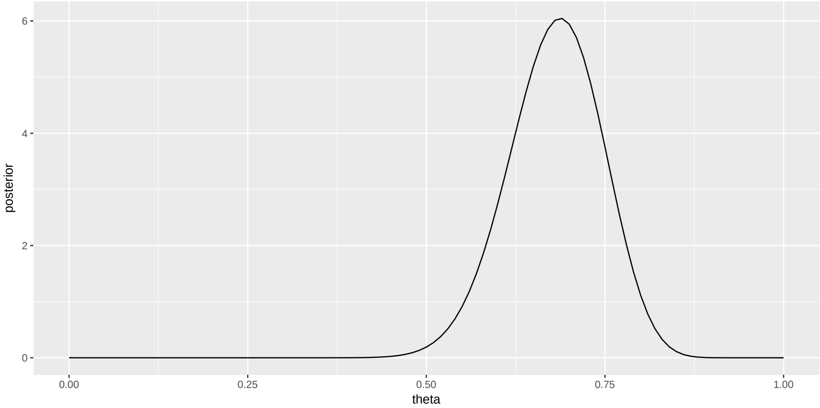
For values 2 to T:

1. sample $\theta_{t+1} \sim N(\theta_t, \sigma)$
2. accecpt θ_{t+1} with $P = \min\left(\frac{f(\theta_{t+1})}{f(\theta_t)}, 1\right)$
3. if accepted next iteration with $\theta_t = \theta_{t+1}$, else redo with θ_t

Example beta-binomial

We know that beta-binomial is conjugate, so we check with the analytical solution.

Lets do a $Beta(3,3)$ posterior and a 31 successes out of 44 likelihood. Posterior is $Beta(34,16)$.



Integration

So the function $P(D|\theta)P(\theta)$ is proportional to the posterior, but does not integrate to 1.

```
posterior_prop <- function(theta) {
  dbeta(theta, 3, 3) * dbinom(31, 44, theta)
}

integrate(posterior_prop, 0, 1)
```

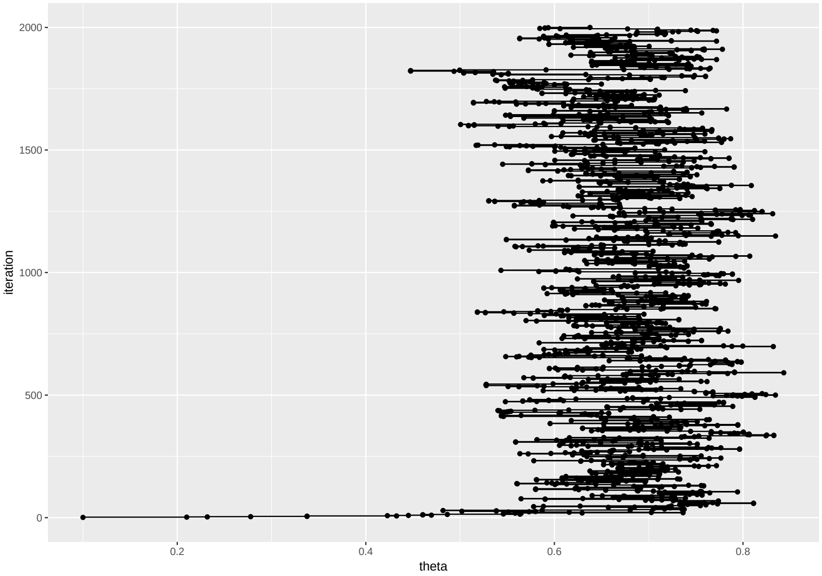
```
## 0.02907361 with absolute error < 3.9e-06
```

Metropolis

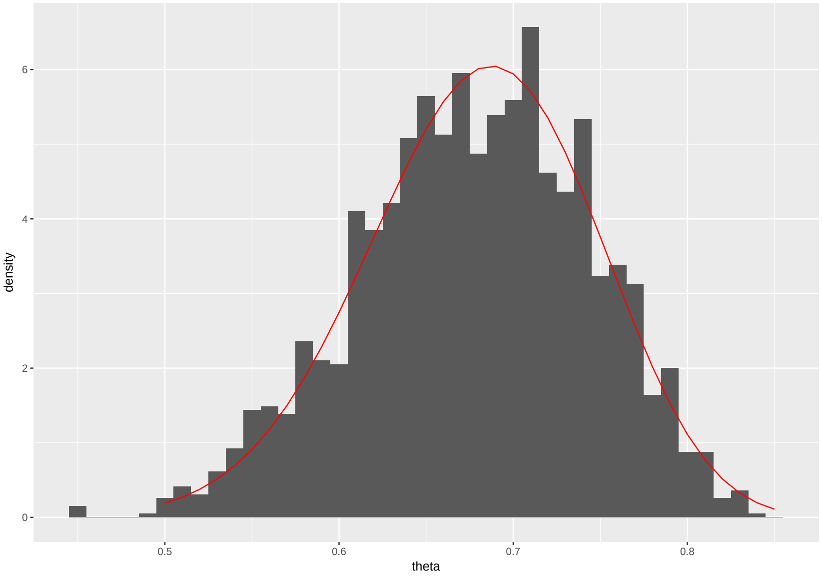
Start with a value of $\theta = 0.1$, proposal $\sigma = 0.05$.

```
set.seed(4242)
current_theta <- 0.1
theta_frame <- data_frame(iteration = 1:2000, theta = 0)
for(i in 1:nrow(theta_frame)) {
  theta_frame$theta[i] <- current_theta
  proposed <- rnorm(1, current_theta, 0.05)
  accept <- (posterior_prop(proposed) /
               posterior_prop(current_theta)) >
    runif(1)
  if (accept) current_theta <- proposed
}
```

Metropolis



Compare with analytical



Metropolis

The MCMC needs to "find" the posterior first. First samples are always discarded (burn-in).

What would happen when we increased σ, and when we decrease it?

Stationary = MCMC reached the posterior. Usually start multiple chains from different initial points and see if they all reach the same distribution.

With many θ's

Vanilla Metropolis theoratically always works, but can be very impractical.

The proposal distribution must be fit for the problem at hand, so desires manual specification.

Default settings easily under- or overshoot. Both yielding efficiency loss.

Long waiting times and convergence issues.

Gibbs sampling (JAGS)

Often we cannot derive the joint posterior of all θ but we can get conditional posterior of individual θ values.

In each iteration we update the θ values one-by-one, by conditioning on the current value of the other parameters. This allows to sample directly from the distribution.

Pro: sampling guaranteed from posterior (no rejections)

Con: one-by-one, conditionals must be known.

Hamiltonian MC (Stan)

Before sampling from the proposal, calculate the gradient of the function at the current (multivariate) point.

Determine a trajectory in the direction of the highest gradient, sample from this trajectory.

Pro: very effective especially with correlated parameters, can be used when posterior is unknow.

Con: gradient calculations