

SC1 Assessment 1

Sam Bowyer

2023-01-06

Gibbs Sampling

Intro

In the final week's content of this module we saw that we can use the Metropolis-Hastings algorithm to sample from some distribution $P(x)$ given a function f proportional to its pdf and a proposal/jumping distribution $Q(x'|x)$ whose pdf $q(x'|x)$ we know and from which we can generate samples. The key motivation behind this MCMC method is that P may be a distribution for which sampling is generally difficult, but using this algorithm we only need to sample from Q , which we may choose to be something simple (often a Gaussian) from which we can easily sample.

One problem that can arise, particularly in very high dimensions, is that of finding a suitable jumping distribution Q —choosing the wrong Q (which can be especially difficult if the different dimensions behave differently to one another) can lead to slow mixing times, meaning you have to generate a large number of samples before they start to accurately approximate P . A related MCMC technique known as Gibbs sampling is very commonly used in this situation, it works by sampling each dimension individually from its conditional distribution (the conditioning occurs based on the most recent samples from every other dimension), rather than attempting to sample the entire joint distribution of Q . Before we go into the details and implementation of Gibbs sampling, we will first provide a brief recap of the Metropolis-Hastings algorithm.

Metropolis-Hastings MCMC

(can probably reuse a lot of the course website code) To simulate samples from target dist P using a proposal dist Q you need: 1. a function f proportional to P 's pdf 2. $Q(x'|x)$'s pdf $q(x')$ 3. The ability to sample $x' \sim Q(x'|x)$.

MCMC considerations

Burn in to discard early samples which may have been generated before the Markov chain stabilised to the stationary distribution P .

Nearby samples are correlated—often solved by only actually taking every n th sample (so if we generate $k = 10000000$ samples (after burn in), then we get 1000 (hopefully fairly uncorrelated) samples with $n = 1000$).

when in high-dims, choosing the right jumping/proposal dist Q can be v. difficult, so it may be better to consider how we'd tackle this problem using the conditional dists. ...->->-> Gibbs sampling (thus instead of having to sample Q from a v. high dimension once every time step, instead we sample many simpler conditional distributions every time step—this is particularly useful if each dim/variable is conditioned only on very few other dims/variables (e.g. in most hierarchical models))

Gibbs Sampler Details

Algo in pseudocode/words w/ maths explanations of what's going on
code

Example 1 (2D Guassian?)

explain the two r.v.s and their conditional dists

plot samples over time (+ $\text{cumsum}(x - \text{mean}(xs))$) to see how it takes some time to stabilise -> implement burn in too

check autocorrelation -> only sample every n th sample

check mean/sd of marginals from the generated samples

Example 2 (higher dim?)

i dunno, just more dims i guess