

It is important to run separate MCMC chains (from different starting points) as it is hard to evaluate the convergence of the chains. If the chains individually appear to converge, but the inference results are different, it can suggest multi-modality in the posterior with the possibility that the chains got ‘stuck’ in a local mode. It is only possible to detect this with multiple chains all starting from different starting points.

Drawing samples from a MCMC chain has the guarantee that as the number of samples tends to  $\infty$ , the samples will be as if they were drawn i.i.d from the target distribution. For all practical purposes we are unable to draw infinitely many samples. We therefore initialize the chain at some unknown starting point and take a finite number of samples. This alone would lead to biased results as the starting point (due to random initialization) might have been in an area of low probability. For the finite number of samples, we have therefore drawn too many samples from this low probability area simply because the chain was initialized in this space. Rather, if we allow a ‘burn-in’ period where the sampler can find a region of high probability, we hope to not induce this bias in our inference.