

Both rejection and importance sampling will work well when the target distribution P is low dimensional and when we have a known distribution Q that approximates (and majorizes) P well. The Markov Chain Monte Carlo methods rather use current samples to draw new **correlated** samples. These methods never reject samples and so while they may take *burn-in and thinning* to try achieve independent samples from the target distribution, they do not (always) depend on a potentially poorly approximating proposal function Q . Gibbs sampling requires a structured setup in the conditional dependence of the latent factors. In the Bayesian setting, we would need to write the posterior $P(\theta_{\{k\}}, \theta_{\{-k\}} | X)$ in terms of the factored conditional distributions $P(\theta_{\{k\}} | \theta_{\{-k\}}, X)$ and $P(\theta_{\{-k\}} | \theta_{\{k\}}, X)$ where $\theta_{\{k\}}$ denotes some subset of parameters that are disjoint from $\theta_{\{-k\}}$. If it is possible to compute these conditionals, then Gibbs sampling might be more useful than Random-Walk Metropolis. In fact it is simply a structured form of the MH algorithm where the acceptance probabilities are 1 due to the forms of these structured conditionals. Note Gibbs is likely to produce better results for posteriors where latent variables are correlated. Random-Walk Metropolis will be preferable when this factoring is intractible or expensive to calculate.

MCMC methods require that dependent draws from the incorrect proposal distribution will converge to independent draws from the correct target distribution if the Markov chain transition is specified correctly and as the number of draws tends to ∞ . We therefore strictly require convergence of the Markov chains for the method to work. The convergence of the chain implies that samples are independent and therefore we can evaluate the degree of auto-correlation to gain insight into the success of this method. Having a low degree of auto-correlation is a necessary but not sufficient condition to ensure good sampling. It is easy to compute but it has the disadvantage of not necessarily highlighting whether the chain converged **to the correct distribution**. In terms of evaluating the approximate Bayesian solution (after the sampling has converged), posterior predictive checks might give insight into the accuracy/oversight of a model.