MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster

A review of the paper by S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White [1].

Liam de Búrca [1]

Although traditional parametric statistics are used in most scientific applications, they rely upon initial assumptions (e.g. normality) which may not always be true, potentially leading to false conclusions. In the spirit of combating these issues, nonparametric statistics rely on very few assumptions, to bolster objectivity and correctness. Applying nonparametric statistics to modelling using Markov chain Monte Carlo (MCMC) methods, leads to an expansion of the parameter or function space proportional to the amount of data. To ensure efficient mixing properties and convergence rates of MCMC algorithms, already well-established algorithms, such as the Metropolis-Hastings, can be slightly altered.

1 Introduction

In cases where a sample size is sufficiently small, such that the underlying probability distribution can not be identified, or that the sample is simply skewed towards one side, traditional parametric tests, which assume normality (e.g. t-tests), become unreliable [2]. Nonparametric tests avoid the assumption of normality, but suffer a loss of information as the values of the data are replaced by signs and ranks (Wilcoxon's sign test). Although nonparametric statistics no longer assume normality, they are still sensitive to subjectivity. One such test is the Kolmogorov-Smirnov (KS) test, which histograms two samples, and integrates to create their cumulative distributions functions (CDFs). The maximum difference between these CDFs is then the test-statistic for the KS test. Figure 1 shows two samples' histograms and their CDFs, using a bin width of 4, resulting in a test-statistic of 0.467. Using different bin-widths may lead to different test-statistics, ergo different conclusions. Nonparametric tests can therefore also suffer from user subjectivity.

Nonparametric MCMC takes this notion of objectivity one step further. Instead of starting with some equation for the forward problem (generating an observable from the system's state), we assume no knowledge of the inherent function. We therefore switch from sampling over parameter space to sampling over function space, which for complete objectivity and correctness, must be allowed to become ambiguously large (ideally infinite). The paper by Cotter et al. [1] discusses how well-established algorithms, most notably the Metropolis-Hastings, may be altered to ensure non-vanishing acceptance probabilities with increases in dimensionality, allowing for efficient mixing properties.

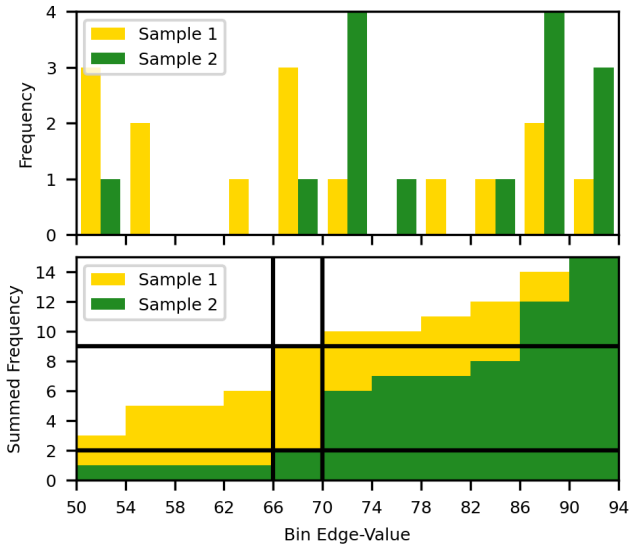

Figure 1: (Left) Histogram with a bin width of 4 showing two different samples. (Right) Cumulative probability curves. The black square highlight the maximum difference between the two samples. With this bin width, the test statistic for the KS test is $7/15 = 0.467$. Data was adapted from [2]

2 Review

In order to ensure consistent acceptance probabilities (probability that a proposed state is accepted) under mesh refinement (increase in the number of dimensions), standard MCMC algorithms can not be used. [1] compares the Metropolis-Hastings to the pCN (pre-conditioned Crank Nickolson) algorithms, with the main difference being the generation of new states ($v^{(k)}$):

$$\text{MH: } v^{(k)} = u^{(k)} + \beta \cdot N(0, C) \tag{1}$$

$$\text{pCN: } v^{(k)} = u^{(k)}\sqrt{1-\beta^2} + \beta \cdot N(0, C) \tag{2}$$

Affiliation University of Copenhagen
Correspondence wbm491@alumni.ku.dk
Version March 12, 2024

where $u^{(k)}$ is the current state, $v^{(k)}$ is the proposed state, β is a step-size, and $N(0,C)$ is a normal distribution with mean 0 and covariance C. As displayed in 2 from [1], the new pCN algorithm is able to maintain high acceptance probabilities through mesh refinement (i.e. increase in the number of dimensions), while maintaining the same step size.

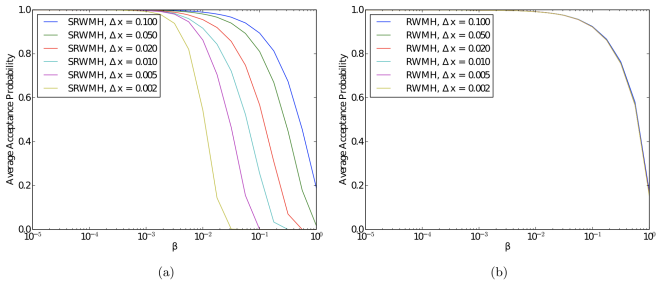

Figure 2: Screenshot of Figure 1 from [1] displaying the insensitivity of the pCN algorithm to mesh refinement. Left: Metropolis-Hastings, right: pre-conditioned Crank Nickolson.

[1] explains, that the pCN algorithm's independence to mesh refinement lies partially in the choice, that the covariance in $N(0,C)$ is equal to the covariance of the Gaussian prior, i.e. although nonparametric statistics tend to shy away from Gaussian statistics, they are still enforced for MCMC over function space.

To further test the efficiency of sampling over function space, the pCN method was compared to Gibb's sampling (Metropolis-within-Gibbs, MwG), see Figure 3. Analysis of the auto-correlation times showed that sampling over function space produced independent samples much faster, by almost an order of magnitude.

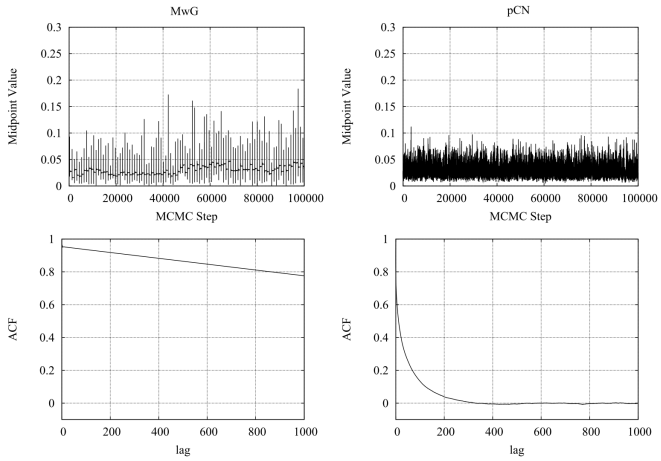

Figure 3: Partial screenshot of Figure 2 from [1]. Notice the much lower auto-correlation time of the pCN algorithm, allowing it to create independent samples much faster than a Metropolis-within-Gibbs algorithm.

In order to further optimize, the step size, β, must be chosen. [1] conducted two experiments, the first using a β-value tuned to reach an average acceptance probability of 0.5, the second using a uniform distribution to randomly determine the step size at each iteration. The results, shown in Figure 4, show almost identical marginal probability distributions (a), and that smaller step sizes cause larger acceptance probabilities (b). Although the two Markov chains are very similar, applying a uniform distribution to the step size may have a significant benefit, namely allowing the algorithm to make large jumps between areas of high probability density.

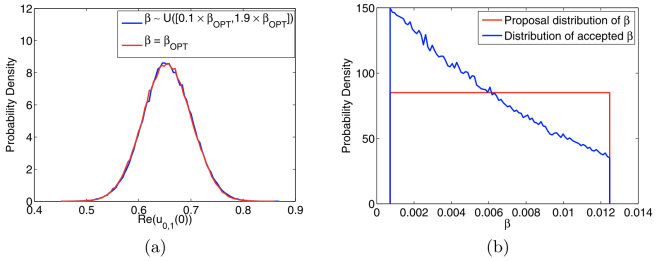

Figure 4: Screenshot of Figure 5 from [1] displaying the marginal distributions from two Markov chains. The first using an optimized constant step size, while the second uses a random step size.

3 Conclusions

Nonparametric statistics are a powerful tool when the assumption of normality is wrong. However, in order to apply the central ideas of nonparametric statistics to problems within physics, and not just medicine or finance, we must adapt our well-established algorithms such that they perform efficiently regardless of the numbers of parameters. [1] displays how a relatively simple modification to the Metropolis-Hastings algorithm can produce huge improvements in acceptance probabilities and auto-correlation times, allowing efficient probability density sampling. Lastly, it was shown, that the choosing a randomly generated step size, β, produced similar results in terms of sampling, while theoretically allowing the Markov chain to travel between regions of high probability density.

References

[1] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. Mcmc methods for functions: Modifying old algorithms to make them faster. *Statistical Science*, 28(3), August 2013.

[2] Nahm FS. Nonparametric statistical tests for the continuous data: the basic concept and the practical use. *Korean J Anesthesiol*, 2016.