MCMC CONVERGENCE DIAGNOSTICS VIA REGENERATION WITH APPLICATION TO THE BAYESIAN LASSO

Yi-Lung Chen
Z. I. Botev

University of New South Wales
Sydney, 1466, AUSTRALIA

ABSTRACT

MCMC techniques are popular solutions to approximate quantities that are difficult to compute exactly. Unfortunately, despite its wide use across various fields, most Markov chain samplers lack theoretically justified methods to analyze their output. That is to say, most MCMC samplers lack of: a) a consistent variance estimator for a given ergodic average; and b) an estimator for the total variation distance between the distribution of a draw from the Markov chain sampler and its target distribution. In this study we demonstrate how one can systematically address a) and b) by exploiting the underlying regenerative structure of the simulation output. Roughly speaking, regenerative structure are the times when a stochastic process (in this case a Markov chain) scholastically `restarts' itself. Intuitively, if a Markov chain frequently `restarts' itself, it should have a fast convergence hence, one may examine the convergence of a Markov chain sampler by identifying these events.

1   INTRODUCTION

Let $\{X_k, k = 0,1,2,...\}$ be a discrete time stochastic process such as a Markov chain. Recall that the process is said to be a *zero-delayed discrete-time regenerative* process if there exists time $0 = T_0 \le T_1 \le T_2 \le ...$, such that for any r, $\{X_{T_r+k}, k \ge 0\}$ has the same distribution as $\{X_k, k \ge 0\}$ and is independent of $\{X_k, 0 \le k \le T_r - 1\}$. Consequently, the *tour lengths*, $M_r := T_{r+1} - T_r$, are iid. Intuitively, this means for each r, $X_k, T_r \le k \le T_{r+1} - 1$ is an iid segments of the original process, and each segment has a length of M_r. For example, with $(M_1, M_2, M_3, M_4) = (2, 3, 1, 4)$:

$$\underbrace{X_1, X_2}_{M_1=2}, \underbrace{X_3, X_4, X_5}_{M_2=3}, \underbrace{X_6}_{M_3=1}, \underbrace{X_7, X_8, X_9, X_{10}}_{M_4=4}, ...$$

In the context of a Markov chain sampler, one can potentially identify the regeneration times, $0 = T_0 \le T_1 \le T_2 \le ...$, from the simulation output (Mykland et al. 1995). Moreover we have the following theorems which provide powerful diagnostics to the MCMC sampler.

Theorem (TV bound estimator). Let $\kappa_t(\cdot | X_0)$ with $X_0 \sim \pi_0$ be the t-step transition kernel of a geometrically ergodic Markov chain with invariant density π. Suppose we can identify regenerative times of the Markov chain and assume that $X_0 \sim \pi_0$ initialized a new regenerative cycle. Then, we have (for some constant $\epsilon > 0$)

$$\| \mathbb{E}[\kappa_t(\cdot | X_0)] - \pi \|_{TV} \le \frac{\eta}{t+1} + O(\exp(-\epsilon t)),$$

where $\eta = \frac{\mathbb{E}M_1^2 - \mathbb{E}M_1}{2\mathbb{E}M_1}$ with $M_1, M_2, ...$ denoting the iid regenerative cycles.

Theorem(Elapsed time convergence). Suppose that $M_1, M_2, ...$ are the iid regenerative cycles of the Markov chain. Define the elapsed time process $E(t) = t - T_{N(t)}, t \geq 0$, where $Tn = M_1 + \cdots + M_n$ and $N(t) = \max\{n: T_n \leq t\}$. Then, the elapsed time is a discrete state-space Markov chain and such that:

$$\| \mathbb{E}[\kappa_t(\cdot | \boldsymbol{X}_0)] - \pi \|_{TV} \leq 2 \sup_A |\mathbb{P}(E(t) \in A) - \mathbb{P}(E(\infty) \in A)|,$$

where $E(\infty)$ is a stationary version of $E(t)$. In the other words, twice the total variation error of $E(t)$ bounds the total variation error of $\boldsymbol{X}_t$.

2 WHY SHOULD WE IDENTIFY REGENERATION TIMES IN MARKOV CHAIN OUTPUT?

Identifying the underlying regeneration times of a given Markov chain do require some algebraic manipulations and computations. However there are important benefits for output analysis, as follows:

- **A strongly consistent variance estimator.** Mykland et al. (1995) and Jones and Hobert (2001) provide a formula for a strongly consistent variance estimator for ergodic averages of a stochastic process. The formula requires one to identify regeneration times

- **TV distance bound and burn-in estimator.** The constant η in the first theorem can be estimated from the observed $M_1, M_2, ...$, consequently, we have a TV bound between the distribution of $\boldsymbol{X}_k$, the k-th draw of a Markov chain, and the target distribution.

  A corollary to this is that a burn-in size of at least $\left\lceil \frac{\sum_{k=1}^{N(t)} M_k^2 - \sum_{k=1}^{N(t)} M_k}{2\epsilon \sum_{k=1}^{N(t)} M\_k} \right\rceil$ guarantees the initial distribution of the Markov chain is within ϵ TV distance to the target distribution for any $\epsilon > 0$.

- **A diagnostic plot.** A popular practice for the convergence diagnostics of a Markov chain sampler is to examine autocorrelation plots of the processes $\{p_j \circ \boldsymbol{X}_k, k \geq 0\}$, where p_j is the $j-$the co-ordinate projection. The second theorem suggests the convergence of the process $\{\boldsymbol{X}_k, k \geq 0\}$ can be summarized by a single one-dimensional process $\{E(t), t \geq 0\}$. Consequently a single auto-correlation plot of the elapsed time process $\{E(t), t \geq 0\}$ is a sufficient visual aid.

Several figures can be obtained by applying regeneration to the Gibbs sampler (Park and Casella 2008), which simulates from the posterior density of the Bayesian Lasso linear regression model. The dataset we use consists of 442 observations of a medical measurement for the level of diabetes, each with 10 predictor variables.

REFERENCES

Jones, G. L., and J. P. Hobert. 2001. "Honest exploration of intractable probability distributions via Markov chain Monte Carlo". *Statistical Science* 312–334.

Mykland, P., L. Tierney, and B. Yu. 1995. "Regeneration in Markov chain samplers". *Journal of the American Statistical Association* 90(429):233–241.

Park, T. and G. Casella. 2008. "The Bayesian lasso". *Journal of American Statistical Association* 103(482):681–686.