A theme common in Mathematics appears in Statistics as well—the distinction between existence and construction. As folks began to write down Gibbs sampling algorithms to fit models during the early 1990’s, the focus was often on “existence” algorithms. The goal was to build an algorithm/chain that had the properties of irreducibility and aperiodicity, and that had the posterior distribution as the limiting distribution. From there, Markov chain theory justified empirical estimation based on a realization of the Markov chain.

In a fairly short span of time, work passed from existence to construction, as shortcomings of many algorithms were discovered. The goal became not only construction of an algorithm, but construction of an algorithm that exhibited decent convergence and mixing behavior. This required diagnosis of poor performance with an emphasis on describing the root cause of the poor performance. The next couple of pages show plots that contrast the behavior of two algorithms. Clues to the inadequacy of the first are (i) strips of repeated values for a parameter and (ii) an implausible posterior distribution. Both of these features are improved by the addition of the remixing step. Code is available on the course web site (to get these plots, you will need to load code from two R scripts).

Along with running samplers, early work on Gibbs sampling focused on improving estimation. The two early techniques that were suggested are Rao-Blackwellization of the estimator and subsampling. A closer look at the arguments for these methods is warranted.

The traditional form of the Rao-Blackwellized estimator replaces the right hand side of the empirical estimator with a conditional expectation.

\[
E[\theta_1 | X] = \frac{1}{N} \sum_{i=1}^{N} \theta_1^{(i)}
\]

becomes

\[
E[\theta_1 | X] = \frac{1}{N} \sum_{i=1}^{N} E[\theta_1^{(i)} | \theta_{<i}^{(i-1)}].
\]

The plots on the final page show some of the features of the estimators and the benefit to Rao-Blackwellization of the estimator of the \( \theta_i \). Note that the benefit of Rao-Blackwellization will typically differ for different parameters in a model, and that it varies substantially across different Gibbs samplers.

Note also that techniques can have use beyond improvement of estimators. Both Rao-Blackwellization and subsampling find use in improved convergence diagnostics. Running Rao-Blackwellized convergence diagnostics can often uncover substantial dependence that would otherwise be missed. Subsampling is, of course, useful for graphics with extraordinarily large sample sizes (as is often the case when working with Markov chain Monte Carlo algorithms).
Figure 1: Time series plot and kernel density estimate of the posterior for $\theta_{82}$. The mass parameter is $M = 1$ and there is no remixing step.
Figure 2: Time series plot and kernel density estimate of the posterior for $\theta_{82}$. The mass parameter is $M = 1$. A remixing step is used in the algorithm.
Figure 3: Autocorrelation plots for the two samplers. The top panel is for the no-remixing algorithm; the bottom panel for the remixing algorithm.
Figure 4: A plot of the variance of the generated values against their mean. Each point represents one of the $\theta_i$. The lower panel shows a plot of batch-means estimates of standard errors for the empirical estimator and the Rao-Blackwellized (based on the generation of the theta star) estimator.
Figure 5: Time series plot and kernel density estimate of the posterior for \( \theta_{82} \). The mass parameter is \( M = 1 \) and there is no remixing step. These plots are from a run of only 2000 iterates. Visibility of the repeated values for \( \theta_{82} \) is enhanced.