HYDRA: A Java library for Markov Chain Monte Carlo

Gregory R. Warnes *

Abstract

HYDRA is an open-source, platform-neutral library for performing Markov Chain Monte Carlo. It implements the logic of standard MCMC samplers within a framework designed to be easy to use and to extend while allowing integration with other software tools. In this paper, we describe the problem that motivated our work, outline our goals for the HYDRA project, and describe the current features of the HYDRA library. We then provide a step-by-step example of using HYDRA to simulate from a mixture model drawn from cancer genetics, first using a variable-at-a-time Metropolis sampler and then a Normal Kernel Coupler. We conclude with a discussion of future directions for HYDRA.

Keywords: Markov Chain Monte-Carlo, Gibbs Sampling, Software Library

1 Introduction

Markov Chain Monte Carlo (MCMC) is a method of performing numerical integration of analytically intractable functions that can expressed as distributions (12; 9). One strength of MCMC is that it can simulate from distributions without requiring that the densities be properly normalized. This makes it particularly useful for Bayesian statistical models, where properly normalizing the densities often requires integration of analytically intractable functions over high-dimensional spaces.

After an initial burn-in period, a properly constructed MCMC sampler will generate a sequence of (non-independent) samples, \(X_0, X_1, \ldots, X_N\) from a specified probability distribution, \(\Pi\). Using these samples, the expectations under \(\Pi\) of any function \(g\) can be estimated by computing the sample path

*Gregory R. Warnes is a Coordinator, Biostatistics and Reporting, Pfizer Global Research and Development, MS 8260-114, Eastern Point Road, Groton, CT 06340. E-mail: gregory.r.warnes@groton.pfizer.com. The author would like to thank Adrian E. Raftery, Thomas Lumley, and Anthony Rossini of the Department of Statistics and Department of Biostatistics of the University of Washington for helpful discussions and guidance. This research was supported in part by NIH/NIAID Grant no. 5 T32 AI07450-09, NIH Grant no. 1 PO1 CA76466, NIH Grant no. 1 PO1 CA76466, and ONR Grant no. N00014-96-1-0192.
average $\hat{E}(g) = \frac{1}{N} \sum_{t=0}^{N} g(X_t)$. As of this writing, two software packages are available that implement Markov Chain Monte Carlo for statistical applications, WinBUGS and FBM.

WinBUGS (6; 7) is a software package for performing Bayesian inference using Gibbs sampling. It provides tools for specifying the model, running the Gibbs sampler, and monitoring convergence using a “point-and-click” graphical interface. A noteworthy feature is that it allows models to be specified using either a text-based notation or via a graphical model created with the DoodleBUGS interface (16).

While WinBUGS is mature and is available free of charge from the MRC Biostatistics Unit web site (17), it has several drawbacks. First, WinBUGS is designed to perform only Gibbs and componentwise Metropolis sampling and does not allow specification of alternative sampling methods. As a consequence, WinBUGS cannot be used when Gibbs sampling or Metropolis-within-Gibbs are inappropriate. Second, the source code to WinBUGS is not available to the user. This makes it impossible for users to add features to WinBUGS. Thus, users who require features (such as statistical distributions or sampling methods) not provided by WinBUGS are forced to abandon the package entirely. In addition the inability to access the source code prohibits the use of WinBUGS for experimentation with or customization of sampling algorithms. This prevents WinBUGS from being used as a tool for research on MCMC methods.

Radford Neal’s FBM (“Flexible Bayesian Modeling”) software (13), first released in 1995, is a less well known package that implements a variety of MCMC methods and includes the C source code. While FBM is more flexible than WinBUGS, the FBM documentation and interface are considerably more difficult to understand.

Both WinBUGS and FBM are restricted to specific operating systems. While older versions of BUGS were available for Unix systems, the current version is available only for systems running versions of Microsoft Windows. FBM, on the other hand, is available only for Unix systems. In addition, neither package is integrated with standard statistical tools. This requires the user to learn the interface of an additional software package in order to use MCMC.

These drawbacks appear to have discouraged or prevented many users from taking advantage of the considerable effort and expertise represented by WinBUGS and FBM. As evidence of the general dissatisfaction with the available tools, all of the statisticians we have observed using or researching MCMC write their own custom software. This results in considerable duplicated effort. Worse, since properly debugging and verifying software algorithms is a difficult and time-consuming task, it is likely that many of the hand-written software programs contain undetected errors. This can lead to the presentation of faulty analyses. Finally, lack of integrated software support for MCMC has led many applied researchers to avoid Bayesian statistical methods entirely.

2 Our Approach

Clearly, there is a need for better MCMC software. Our goal is is to produce a software tool that

1. implements standard MCMC techniques,
2. is easy to use,
3. is reliable,
4. is applicable to a wide variety of problems,
5. allows access to the underlying algorithms,
6. can be easily customized and extended,
7. is integrated with traditional statistical packages, and
8. is available on all common computer platforms.

The Hydra MCMC library is a first step toward providing software that achieves these goals. Hydra is a object-oriented library that implements the logic of standard MCMC methods. Although Hydra can be used directly in custom MCMC programs, is designed to be used as a basis for MCMC software which provides a more user-friendly interface and which is integrated with standard statistical packages.

Hydra is implemented using Java, a platform independent object-oriented language designed for general programming tasks. We selected Java (10) because it enabled the library to meet a number of our stated goals. First, Java’s support of formal interfaces facilitated the construction of a library that is easy to use while being flexible and easy to extended. In particular, the use of interfaces permits users to provide components of the MCMC algorithm that are tuned to their specific problem. This allows the user to extend the Hydra package to support new problems or MCMC techniques without changing the existing code. Second, Java provides features that reduce common programming errors and is supported by a wealth of standard libraries and programming tools. Not only do Java’s features make it easier to write code that is error-free, but they also make it easier to locate and correct bugs that do exist. This supports construction of a reliable library and frees time otherwise spent debugging for the development of additional features. Third, Java provides a clear interface for interacting with other languages. This gives a well defined method for Hydra to be used with existing programming languages and software applications. Although early versions of Java suffered from performance problems, recent versions of the Java virtual machine (runtime) can provide speed comparable to C and C++ for numerical computations (14; 11; 21; 15).

The remainder of this text assumes a basic familiarity with the Java language on the order of 4).

3 Constructing Metropolis-Hastings Samplers Using Hydra

Hydra supports the full generality of Markov Chain Monte Carlo by providing a hierarchy of classes implementing the most common MCMC techniques. Classes exist that implement the general Metropolis-Hastings algorithm, the Metropolis sampler, and the Gibbs sampler. Hydra also implements the multi-state Adaptive Metropolis Sampling (8) algorithm that forms the basis of Adaptive Direction Sampling (5) and Normal Kernel Coupling (20). We will focus on the implementation of the Metropolis-Hastings method since it includes the others as special cases.

The Metropolis-Hastings algorithm is the most general MCMC method used in practice and includes as special cases Metropolis sampling, Gibbs sampling, and Adaptive Metropolis Sampling. The algorithm is remarkably simple. Given a target distribution of interest $\Pi$, corresponding to the statistical model, an initial starting location $X_0$, and a proposal distribution $Q(X_t)$, each iteration of the sampler consists of four steps:

1. **Propose** a candidate state $Y$ using a proposal distribution $Q(X_t)$, which may depend on the current state $X_t$:
   
   \[ Y \leftarrow Q(X_t) \]
2. **Compute** the Metropolis-Hastings acceptance probability

\[ \alpha(X_t, Y) = \min \left\{ 1, \frac{\pi(Y) q(Y \rightarrow X_t)}{\pi(X) q(X_t \rightarrow Y)} \right\} \]

\[ = \min \left\{ 1, \frac{p(Y) q(Y \rightarrow X_t)}{p(X) q(X_t \rightarrow Y)} \right\} \]

where \( \pi \) is a density corresponding to the target distribution \( \Pi \), \( p(x) \propto \pi(x) \) is an unnormalized density, and \( (Y|X_t) \) is the conditional density for \( Y \) under \( Q(X_t) \).

3. **Accept** the proposed point \( Y \) and set

\[ X_{t+1} \leftarrow Y \]

with probability \( \alpha(X_t, Y) \), otherwise **Reject** the proposed point and set

\[ X_{t+1} \leftarrow X_t. \]

4. **Increment** time: \( t \leftarrow t + 1 \).

The sequence of \( X \) values generated by this algorithm converges to a (dependent) sample from \( \Pi \) provided the proposal distribution \( Q \) meets certain conditions (19).

The *CustomMetropolisHastingsSampler* class implements the logic of Metropolis-Hastings samplers using a target distribution (model), initial state, and proposal distribution specified by the user. This is made possible by requiring the objects representing the target and proposal distributions to provide certain methods. These methods are defined by the *UnnormalizedDensity* and *GeneralProposal* interfaces, respectively. No restriction is placed on the initial state, save that it be compatible with the user-specified target and proposal distributions.

To allow flexible reporting of the progress of the MCMC sampler, the *CustomMetropolisHastingsSampler* maintains a list of user defined objects that are notified at the completion of the acceptance step of each iteration. When detailed reporting is selected, these “listeners” receive an object containing a great deal of information about each MCMC iteration.

### 3.1 The *UnnormalizedDensity* Interface for Target Distributions

Target distributions implement the *UnnormalizedDensity* interface, which defines two methods:

```java
public double unnormalizedPDF (Object state);
public double logUnnormalizedPDF(Object state);
```

These methods compute the (log) unnormalized density of the model for the state passed as a parameter.

### 3.2 The *GeneralProposal* Interface for Proposal Distributions

Proposal distributions implement the *GeneralProposal* interface, which has 4 methods:

```java
public double conditionalPDF (Object next, Object current);
public double logConditionalPDF(Object next, Object current);
public double transitionProbability (Object from, Object to);
public double logTransitionProbability (Object from, Object to);
```
The methods conditionalPDF and logConditionalPDF compute the probability of generating the object next when the current state is current. The second two methods perform the same computation, but reverse the arguments\(^1\).

### 3.3 The MCMCLListener Interface for Listener Objects

Objects that are notified at the completion of each MCMC iteration implement the MCMCLListener interface, which defines one method:

```java
public void notify(MCMCEvent event);
```

The parameter of the notify method is an object containing information about the MCMC iteration. This information can be used by the object in various ways. Possibilities include storing the current state to a file, displaying it on a plot, and computing cumulative statistics.

When detailed reporting is disabled, the object passed to notify is a GenericChainStepEvent. This object has a single field:

```java
public Object current;
```

which contains the current state \((X_t)\) of the sampler.

When detailed reporting is enabled, the object passed to notify is a DetailChainStepEvent which has the additional fields:

```java
public Object proposed;
public Object last;
public double lastProb;
public double proposedProb;
public double forwardProb;
public double reverseProb;
public double probAccept;
public double acceptRand;
public boolean accepted;
public double acceptRate;
```

These fields provide a great deal of information about the MCMC iteration and are useful for debugging and for evaluating the performance of different proposal distributions. The interpretation of each is given in Table 1.

### 4 Example

The classes provided by Hydra can be used directly in compiled Java programs or interactively with various Java-based tools, such as JPython and the Omegahat statistical language. For ease of presentation, we will focus on the pure Java interface.

We will give an example by using Hydra to construct two different samplers for a Binomial-BetaBinomial mixture model for the loss of genetic material in esophageal cancers. We first show how to implement the unnormalized density corresponding to Binomial-BetaBinomial mixture model in Java so that it can be used with Hydra. Using this model we create a variable-at-a-time Metropolis sampler, and then a Normal Kernel Coupler.

\(^1\)The transitionProbability and logTransitionProbability are depreciated and will not be required in a future release of the software.
Table 1: Interpretation of the fields of the DetailChainStepEvent

<table>
<thead>
<tr>
<th>Field</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>public Object current;</td>
<td>$X_t$</td>
</tr>
<tr>
<td>public Object proposed;</td>
<td>$Y$</td>
</tr>
<tr>
<td>public double proposedProb;</td>
<td>$p(Y)$</td>
</tr>
<tr>
<td>public Object last;</td>
<td>$X_{t-1}$</td>
</tr>
<tr>
<td>public double lastProb;</td>
<td>$p(X_{t-1})$</td>
</tr>
<tr>
<td>public double forwardProb;</td>
<td>$q(Y</td>
</tr>
<tr>
<td>public double reverseProb;</td>
<td>$q(X_{t-1}</td>
</tr>
<tr>
<td>public double probAccept;</td>
<td>$\alpha(X_{t-1},Y)$</td>
</tr>
<tr>
<td>public boolean accepted;</td>
<td>was $Y$ accepted?</td>
</tr>
<tr>
<td>public double acceptRand;</td>
<td>uniform value used to accept/reject</td>
</tr>
<tr>
<td>public double acceptRate;</td>
<td>average value of $\alpha(X_{t-1},Y)$</td>
</tr>
</tbody>
</table>

4.1 Overview

There are four user-specified components of a Metropolis-Hastings sampler: target distribution (model), initial state, proposal distribution, and uniform random number generator. The HYDRA library provides a reliable random number generator and a selection of standard proposal distributions, leaving the user to construct the target distribution and initial state.

4.2 Creating a Target Distribution

To create an object representing the target distribution (model), the user needs to write a Java class that implements the UnnormalizedDensity interface. For our example problem, we wish to implement a class for the Bayesian hierarchical model 1

$$X_i \sim \eta \text{ Binomial}(N_i, \pi_1)$$
$$+(1 - \eta) \text{ Beta-Binomial}(N_i, \pi_2, \omega_2)$$

$$\eta \sim \text{Unif}(0, 1)$$
$$\pi_1 \sim \text{Unif}(0, 1)$$
$$\pi_2 \sim \text{Unif}(0, 1)$$
$$\omega_2 \sim \text{Unif}(0, 1/2)$$

where the density of the Beta-Binomial distribution is defined as

$$f(X_i|N_i, \pi_2, \omega_2) = \binom{n}{x} \frac{\Gamma\left(\frac{1}{\omega_2}\right)}{\Gamma\left(\frac{1}{\omega_2}\right)\Gamma\left(\frac{1}{\omega_2}\right)\Gamma\left(n-x+\frac{1}{\omega_2}\right)\Gamma(n+\frac{1}{\omega_2})}$$

A Java class implementing the density for this model is provided in table 2. We shall highlight the programming details that allow use of this class as a target distribution.

First, we need to indicate to the Java compiler where to find the UnnormalizedDensity interface that this class will implement. This is accomplished by the line

```java
```
Table 2: Class implementing the hierarchical Binomial Beta-Binomial model for the LOH data.

```java
package org.omegahat.Simulation.MCMC.Examples;

import java.lang.Math;
import org.omegahat.GUtilties.ArrayTools;

public class Binomial_BetaBinomial_SimpleLikelihood implements UnnormalizedDensity {
    int loh[] = {7, 3, 4, 3, 5, 4, 5, 3, 6, 12, 5, 3, 1, 3, 5, 3, 11, 2, 2, 2, 3, 5, 3, 4, 6, 3, 1, 4, 5, 19, 5, 5, 6, 5, 6, 2, 0, 0, 6, 4};
    int n[] = {17, 15, 17, 18, 15, 15, 15, 19, 16, 15, 18, 19, 19, 21, 17, 16, 12, 17, 18, 18, 19, 19, 15, 12, 16, 19, 16, 19, 21, 15, 13, 20, 16, 17, 8, 7, 18, 15};

    // unnormalized binomial density
    double udb(int x, int n, double pi) { return Math.pow(pi, x) * Math.pow(1.0 - pi, n - x); }

    // unnormalized beta-binomial density
    double udbb(int x, int n, double pi, double omega) {
        int r;
        double tmp0 = 1.0;
        double tmp1 = 1.0;
        double tmp2 = 1.0;
        for (r = 0; r <= (x - 1); r++) tmp0 *= (pi + ((double) r) * omega);
        for (r = 0; r <= (n - x - 1); r++) tmp1 *= (1.0 - pi + ((double) r) * omega);
        for (r = 0; r <= (n - 1); r++) tmp2 *= (1.0 + ((double) r) * omega);
        return (tmp0 * tmp1 / tmp2);
    }

    // unnormalized binomial-beta-binomial mixture density
    double ud_b_bb(int x[], int n[], double eta, double pi0, double pi1, double omega1) {
        double retval = 1.0;
        for (int i = 0; i < x.length; i++)
            retval *= (eta) * udb(x[i], n[i], pi0) + (1.0 - eta) * udbb(x[i], n[i], pi1, omega1);
        return retval;
    }

    // Constructor
    public Binomial_BetaBinomial_SimpleLikelihood() {}

    // Log Unnormalized Density
    public double logUnnormalizedPDF(Object parms) {
        Math.log(unnormalizedPDF(parms));
    }

    // Unnormalized Density
    public double unnormalizedPDF(Object paramObj) {
        double[] parms = ArrayTools.Otod(paramObj);
        double eta = parms[0], pi0 = parms[1], pi1 = parms[2], omega = parms[3];
        // check range
        if (eta < 0.0 || (pi0 < 0.0) || (pi1 < 0.0) || (omega < 0.0))
            return 0.0;
        else
            return ud_b_bb(loh, n, eta, pi0, pi1, omega);
    }
}
```

Now we can declare the class and indicate that it implements the \texttt{UnnormalizedDensity} interface.

7 \begin{verbatim}
public class Binomial_BetaBinomial_SimpleLikelihood implements UnnormalizedDensity
\end{verbatim}

Next, the class needs a constructor that will accomplish any required initialization, such as observed data. In this case, no initialization is required since the data is hard-coded into the class, so that the line

40 \begin{verbatim}
public Binomial_BetaBinomial_SimpleLikelihood() {} 
\end{verbatim}

is sufficient.

Now, our class must provide the \texttt{unnormalizedPDF} and \texttt{logUnnormalizedPDF} methods. These methods are used by the Metropolis-Hastings sampler to compute the acceptance probability for a given proposed state.

\begin{verbatim}
// Log Unnormalized Density //
public double logUnnormalizedPDF( Object parms ) {
    return Math.log(unnormalizedPDF( parms ));
}

// Unnormalized Density //
public double unnormalizedPDF( Object paramObj ) {
    double[] parms = ArrayTools.otod( paramObj );
    double eta=parms[0], pi0=parms[1], pi1=parms[2], omega1=parms[3];

    // check range
    if ( (eta<0.0) || (pi0<0.0) || (pi1<0.0) || (omega1<0.0) ||
        (eta>1.0) || (pi0>1.0) || (pi1>1.0) || (omega1>0.5) )
        return 0.0;
    else
        return ud_bb(loh, n, eta, pi0, pi1, omega1);
}
\end{verbatim}

In this example, we have written separate functions that compute the unnormalized density, so these methods simply convert the arguments to the appropriate type (doubles), check their range, call the appropriate function.

Note that the the interface defines the argument passed to the \texttt{unnormalizedPDF} and \texttt{logUnnormalizedPDF} as an \texttt{Object}. The user must decide what type of object will represent the model parameters. For most purposes, an array of doubles (\texttt{double[]}) is an appropriate choice. For this reason, the predefined proposal methods (see section B) all operate on arrays of doubles. Any other type of object can be used, however, this requires the user to implement an appropriate proposal distribution.

### 4.3 Creating a Variable-at-a-time Metropolis Sampler

Now that we have a class that implements the unnormalized density for the Binomial-BetaBinomial model, we can construct a MCMC sampler. We first implement a variable-at-a-time Metropolis sampler. The complete class file for this sampler is shown in table 3.

Again we provide the Java compiler with the locations of the classes we will be using. This time there are five \texttt{import} statements:

\begin{verbatim}
import org.omegahat.Simulation.MCMC.*;
import org.omegahat.Simulation.MCMC.Proposals.*;
import org.omegahat.Simulation.MCMC.Listeners.*;
import org.omegahat.Simulation.RandomGenerators.*;
import org.omegahat.Probability.Distributions.*;
\end{verbatim}
Table 3: Class implementing a variable-at-a-time Metropolis sampler for the LOH model.

```java
package org.omegahat.Simulation.MCMC.Examples;

import org.omegahat.Simulation.MCMC.*;
import org.omegahat.Simulation.MCMC.Proposals.*;
import org.omegahat.Simulation.MCMC.Listeners.*;
import org.omegahat.Simulation.RandomGenerators.*;
import org.omegahat.Probability.Distributions.*;

public class Binomial_BetaBinomial_SimpleExample {
    static public void main(String[] argv) throws Throwable {
        CollingsPRNGAdministrator a = new CollingsPRNGAdministrator();
        PRNG prng = new CollingsPRNG(a.registerPRNGState());
        UnnormalizedDensity target = new Binomial_BetaBinomial_SimpleLikelihood();
        double[] diagVar = new double[]{0.083, 0.083, 0.083, 0.042};
        SymmetricProposal proposal =
            new NormalMetropolisComponentProposal(diagVar, prng);
        double[] state = new double[]{0.90, 0.23, 0.71, 0.49};
        CustomMetropolisHastingsSampler mcmc =
            new CustomMetropolisHastingsSampler(state, target, proposal, prng, true);
        MCMCListener l = new ListenerPrinter();
        MCMCListenerHandle lh = mcmc.registerListener(l);
        mcmc.iterate(10);
    }
}
```
After declaring the object, we create a main function that will get called to actually do the work of creating and running the MCMC sampler. Within main, the first object we need to create is a pseudo-random number generator. The HYDRA library provides an implementation of the Collings random number generator (3), which can be created using the 2 lines:

12 CollingsPRNGAdministrator a = new CollingsPRNGAdministrator();
13 PRNG prng = new CollingsPRNG(a.registerPRNGState());

Next, we need to instantiate (create) a copy of our class that implements the unnormalized density of the model. This is accomplished by

15 UnnormalizedDensity target = new Binomial_BetaBinomial_Likelihood();

Now we instantiate the proposal distribution. For a Metropolis-Hastings sampler, there are several choices (see Appendix B), including a variable-at-a-time random-walk proposal using a normal distribution. This is implemented by the NormalMetropolisComponentProposal class. Its constructor allows the specification of a proposal variance for each parameter. We'll use the variance of the parameters under the prior:

22 double[] diagVar = new double[]{0.083, 0.083, 0.083, 0.042};
23 SymmetricProposal proposal =
24 new NormalMetropolisComponentProposal(diagVar, prng);

Now we need to define an initial state for the sampler. We'll use the MLE, which is \eta = 0.90, \pi_1 = 0.23, \pi_2 = 0.71, \omega_2 = 0.49:

22 double[] state = new double[]{0.90, 0.23, 0.71, 0.49};

With the random number generator, initial state, target distribution, and the proposal distribution defined, we can create the actual sampler:

24 CustomMetropolisHastingsSampler mcmc =
25 new CustomMetropolisHastingsSampler(state, target, proposal,
26 prng, true);

This gives us a working MCMC sampler. The final parameter is an optional flag indicating whether the sampler should report all of the details about the MCMC iteration when it calls the listeners, or whether to just report the new state. We wish to see all of the details, so we provide the value true.

We need to attach a listener to the MCMC sampler so that we can see the results of each iteration. There is a variety of predefined listeners (see Appendix C), but we'll start with the simplest listener. Its notify method simply displays the object it receives.

28 MCMCListener l = new ListenerPrinter();
29 MCMCListenerHandle lh = mcmc.registerListener(l);

Finally, with the MCMC sampler defined and a listener attached, we are ready to run the MCMC sampler. This is accomplished by calling the MCMC sampler’s iterate method with the number of iterations to perform:

31 mcmc.iterate(10);
4.4 Running the Variable-at-a-time Metropolis Sampler

When combined with the HYDRA library, the two classes we’ve created form a complete Java program. On Unix-like systems with the standard Sun Java tools installed, the classes can be compiled using the `javac` command:

```
> javac Binomial_BetaBinomial_SimpleLikelihood.java
> javac Binomial_BetaBinomial_SimpleExample.java
```

Once the classes are compiled, the MCMC sampler can be run using the Java interpreter by

```
> java org.omegahat.Simulation.MCMC.Examples.Binomial_BetaBinomial_SimpleExample
```

This will cause the MCMC sampler to print detailed information about each of the ten iterations to the screen. The output for the first iteration is:

```
Chain Step Event (with details)
Last = ContainerState: [0.9 0.23 0.71 0.49]
Last Prob = -359.046964566765
Proposed State = ContainerState: [0.9 0.47510664155061 0.71 0.49]
Proposed Prob = -423.26454869568283
Current State = ContainerState: [0.9 0.23 0.71 0.49]
Forward Prob = -0.03694935787235
Reverse Prob = -0.03694935787235
Acceptance Prob = -64.21758412891785
Acceptance Val = 0.658405257229882
Accepted? = false
Acceptance Rate = 0.0
```

This gives the current and proposed states, the value of the unnormalized density, the forward and reverse proposal probabilities, the acceptance probability, a flag indicating whether or not the proposed state was accepted, the new state, and the cumulative acceptance rate. Note that the unnormalized density and probabilities are reported on the log scale. The output for all 10 iterations is given in Appendix D

4.5 Enhancing the Variable-at-a-time Metropolis Sampler

This example can be enhanced in a number of ways. First, the class can be modified to use a different proposal method. To use a (complete-state) random-walk Metropolis sampler, simply replace the `NormalMetropolisComponentProposal` with a `NormalMetropolisProposal`. Alternatively, the user could define a custom proposal distribution and use it instead.

Second, it is impractical to store and interpret all of the detailed information produced by using the `StepListenerPrinter` listener for more than a few iterations. Instead, we would like to store just the current state to a disk file. This is accomplished by replacing the `StepListenerPrinter` object with a `StrippedListenerWriter`. Change lines 28 and 29 to

```
28 ListenerWriter l = new StrippedListenerWriter("MCMC.output");
29 MCMCListenerHandle lh = mcmc.registerListener(l);
```

and replace line 32 with

```
32 lh.close();
```

The `lh.close();` command makes sure that the file that is used to store the MCMC output is properly closed once the MCMC iterations are complete.

Now that the output is being stored to a disk file, it is reasonable to increase the number of iterations. Naturally, this is done by changing the value in the `mcmc.iterate` call to the desired value, say 10,000.
Compiling and running the modified class now generates a data file containing 10,000 MCMC iterations. This output can be read into a standard statistical package for computation of diagnostics or to perform inference. For this, we have found the CODA package of MCMC diagnostics, which exists in versions for both R and S-PLUS, particularly helpful (see the appendix for information on obtaining CODA). For either version, the commands

```r
library(data)
mcmc.data <- mcmc(as.matrix(read.table("MCMC.output")));
```

will load the CODA library (provided it is installed) and properly import the MCMC data. A selection of diagnostics, plots, and summaries is then available. For instance, the default CODA plots and summary statistics for our 10,000 iterations are shown in Figure 1.

Figure 1: CODA plots for 10,000 MCMC iterations.
4.6 Implementing the Normal Kernel Coupler

To implement the Normal Kernel Coupler (NKC) introduced by \(20\), only a three changes need to be made to our example class. First, we use a different proposal distribution. Second, we initialize a set of initial values rather than a single value. Third, we use the `CustomHastingsCoupledSampler` class instead of the `CustomMetropolisHastingsSampler` class. The complete source code for the modified class is given in table 4.

The proposal distribution for the NKC is implemented by the class `NormalKernelProposal`. Its constructor requires two arguments, a random number generator, and a matrix that specifies the variance for the normal kernel. For our example, the proposal is instantiated by the lines:

```java
double[][] Var = {
    {0.003, 0.0, 0.0, 0.0},
    {0.0, 0.001, 0.0, 0.0},
    {0.0, 0.0, 0.012, 0.0},
    {0.0, 0.0, 0.0, 0.007}
};
HastingsCoupledProposal proposal = new NormalKernelProposal(Var, prng);
```

The NKC maintains a set of current states that must be initialized. We use a `MultiDoubleState`, which holds a list of `double` values, to hold the initial values.

```java
int numComponents = 200;
MultiDoubleState state0 = new MultiDoubleState(numComponents);
for (int i = 0; i < numComponents / 2; i++)
    state0.add(new double[]{0.903, 0.228, 0.708, 0.486});
for (int i = numComponents / 2; i < numComponents; i++)
    state0.add(new double[]{0.078, 0.831, 0.230, 4.5e-9});
```

In this case, we’ve initialized half of the values to each of the two local maxima.

The logic of multi-state MCMC samplers is implemented by the `CustomHastingsCoupledSampler` class. This class is instantiated using 6 parameters, the set of initial states, the number of current states to maintain, the target (model) distribution, the proposal distribution, a random number generator, and a flag indicating whether to report the details of the iteration:

```java
CustomHastingsCoupledSampler mcmc =
    new CustomHastingsCoupledSampler(state0, numComponents,
    target, proposal, prng, true);
```

Although we could have simply used a `StrippedListenerWriter` this would generate very large output file by writing out the entire set of 200 current states at each iteration. Instead, we use using `ThinningProxyListener` class, which “thins” the events it receives by a specified factor before passing them on:

```java
ThinningProxyListener pL = new ThinningProxyListener(numComponents);
MCMCListenerHandle pLh = mcmc.registerListener(pL);
MCMCListenerWriter lw = new StrippedListenerWriter("NKC.output");
MCMCListenerHandle lh = pl.registerListener(lw);
```

Running the sampler now will output the complete state once every 200 iterations.
Table 4: Class implementing a Normal Kernel Coupler for the LOH model.

```java
package org.omegahat.Simulation.MCMC.Examples;

import org.omegahat.Simulation.MCMC.*;
import org.omegahat.Simulation.MCMC.Proposals.*;
import org.omegahat.Simulation.MCMC.Listeners.*;
import org.omegahat.Simulation.RandomGenerators.*;
import org.omegahat.Probability.Distributions.*;

public class Binomial_BetaBinomial_SimpleExample_NKC {
    static public void main(String[] argv) throws Throwable {
        CollingsPRNGAdministrator a = new CollingsPRNGAdministrator();
        PRNG prng = new CollingsPRNG(a.registerPRNGState());
        UnnormalizedDensity target = new Binomial_BetaBinomial_SimpleLikelihood();
        double[][] Var = {{0.003, 0.0, 0.0, 0.0},
                          {0.0, 0.001, 0.0, 0.0},
                          {0.0, 0.0, 0.012, 0.0},
                          {0.0, 0.0, 0.0, 0.007}};
        HastingsCoupledProposal proposal = new NormalKernelProposal(Var, prng);
        int numComponents = 200;
        MultiDoubleState state0 = new MultiDoubleState(numComponents);
        state0.add(new double[] {0.903, 0.228, 0.708, 0.486});
        for(int i=numComponents/2; i < numComponents; i++)
            state0.add(new double[] {0.078, 0.831, 0.230, 4.5e-9});
        CustomHastingsCoupledSampler mcmc =
            mcmc = new CustomHastingsCoupledSampler(state0, numComponents,
                                                    target, proposal, prng,
                                                    false);
        ThinningProxyListener pL = new ThinningProxyListener(numComponents);
        MCMCListenerHandle plh = mcmc.registerListener(pL);
        MCMCListenerWriter l = new StrippedListenerWriter("NKC.output");
        MCMCListenerHandle lh = plh.registerListener(l);
        mcmc.iterate(10000);
        l.close();
    }
}
```
5 Conclusions and Future Directions

Our example has shown that the HYDRA MCMC library makes it easy to create different Metropolis-Hastings samplers without extensive programming. This should encourage additional statisticians to experiment with and use the Metropolis-Hastings method.

We hope that the HYDRA library will form the basis of a set of MCMC tools that are easy to use, robust, and complete. In particular we intend to integrate HYDRA with the statistical tools R, Splus, and SAS, as well as the new Omegahat statistical computing language (18; 2; 1). These interfaces promise to provide flexible and powerful interactive environments for MCMC.

Other goals for the HYDRA library include

- visual tools for specifying and monitoring MCMC simulations
- support for distributed/parallel computing
- a library of target distributions corresponding to common statistical models, such as GLM’s and mixture models.

References


## A Installing HYDRA

The HYDRA Java package is available in two forms, as a Jar file (Hydra.jar) containing only the compiled classes and as a gzipped tar file (Hydra.current.tar.gz) containing the full source code as well as the compiled classes. Both files are available from the HYDRA web page located at http://www.warnes.net/Hydra.

### A.1 Installing the jar File

Download Hydra.jar and append the full path to the jar file to the CLASSPATH. For example, if Hydra.jar has been placed in the directory /home/user/jars/, the proper command for setting the CLASSPATH using sh compatible shells is

```
> CLASSPATH=$CLASSPATH:/home/user/jars/Hydra.jar
> export CLASSPATH
```
and using `csh` compatible shells

```
> setenv CLASSPATH $CLASSPATH:/home/user/jars/Hydra.jar
```

### A.2 Installing the Full Source

Download `Hydra.current.tar.gz`. It can then be unpacked using GNU tar via

```
> tar -xvzf Hydra.current.tar.gz
```

which will unpack a directory tree with root “Hydra”. The Java files and source code are contained in directories under `Hydra/org/omegahat`.

The location of this directory then needs to be added to the Java class path. If the directory tree was unpacked in `/home/user/jsrc/` this can be accomplished using `sh` compatible shells by

```
> CLASSPATH=$CLASSPATH:/home/user/jsrc/Hydra
> export CLASSPATH
```

and using `csh` compatible shells

```
> setenv CLASSPATH $CLASSPATH:/home/user/jsrc/Hydra
```

### A.3 Other Packages

Two additional Java packages may be required to use particular features of the HYDRA library, Visual Numerics’ JNL and Omegahat. In addition, the CODA package, in conjunction with either R or SPLUS statistical packages, provides a useful suite of tools for evaluating and making inference using MCMC output.

- Visual Numerics’ JNL library is required for several of the HYDRA classes, in particular those used in the Binomial-BetaBinomial example given below. JNL is available free of charge from the Visual Numerics web site:
  

- The HYDRA MCMC classes were designed to be compatible with the Omegahat statistical programming system. Omegahat provides an interactive environment for statistical programming and analysis and is under active development by the Omegahat project, [http://www.omegahat.org](http://www.omegahat.org).

- The statistical package R is a free re-implementation of the S language and may be obtained free of charge from [http://www.r-project.org](http://www.r-project.org).

- The CODA package of MCMC diagnostics and other tools for Splus can be obtained from [http://www.mrc-bsu.cam.ac.uk/bugs/classic/coda04/readme.shtml](http://www.mrc-bsu.cam.ac.uk/bugs/classic/coda04/readme.shtml).

A version for R can be obtained from [http://www-fis.iarc.fr/coda](http://www-fis.iarc.fr/coda).
B Predefined Proposal Distributions

HYDRA provides a selection of predefined proposal methods for the Metropolis, Metropolis-Hastings, and Hastings-Coupled techniques.

Metropolis Samplers

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NormalMetropolisProposal</td>
<td>normal random-walk proposal</td>
</tr>
<tr>
<td>NormalMetropolisComponentProposal</td>
<td>variable-at-a-time random-walk proposal</td>
</tr>
</tbody>
</table>

Metropolis-Hastings Samplers

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NormalProposal</td>
<td>(fixed) normal proposal</td>
</tr>
<tr>
<td>NormalMetropolisProposal</td>
<td>normal random-walk proposal</td>
</tr>
<tr>
<td>NormalMetropolisComponentProposal</td>
<td>variable-at-a-time random-walk proposal</td>
</tr>
<tr>
<td>MixtureProposal</td>
<td>finite mixture proposal using specified components</td>
</tr>
</tbody>
</table>
**Hastings-Coupled (Multi-Chain) Samplers**

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IndependentHastingsCoupledProposal</td>
<td>Wrapper for independent Metropolis-Hastings Samplers</td>
</tr>
<tr>
<td>AdaptiveNormalMetropolisProposal</td>
<td>(Variance) Adaptive Normal Metropolis Proposal</td>
</tr>
<tr>
<td>AdaptiveNormalProposal</td>
<td>(Mean, Variance) Adaptive Normal Proposal</td>
</tr>
<tr>
<td>NormalKernelProposal</td>
<td>Normal Kernel Coupler</td>
</tr>
<tr>
<td>AdaptiveNormalKernelProposal</td>
<td>(Variance) Adaptive Normal Kernel Coupler</td>
</tr>
<tr>
<td>LocallyAdaptiveNormalKernelProposal</td>
<td>Locally-Adaptive Normal Kernel Coupler</td>
</tr>
<tr>
<td>KernelDirectionSampler</td>
<td>Kernel Direction Sampler</td>
</tr>
</tbody>
</table>
C Predefined Listeners

A variety of predefined listeners are available. These allow monitoring various features of the MCMC simulation and give several storage methods.

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcceptanceWriter</td>
<td>Stores the cumulative acceptance rate to a file</td>
</tr>
<tr>
<td>CovarianceWriter</td>
<td>Stores the cumulative covariance matrix to a file</td>
</tr>
<tr>
<td>DistanceListener</td>
<td>Computes the observed and expected acceptance rate, step distance, step distance conditional on acceptance</td>
</tr>
<tr>
<td>DistanceWriter</td>
<td>Stores the observed and expected acceptance rate, step distance, step distance conditional on acceptance to a file</td>
</tr>
<tr>
<td>HistogramWriter</td>
<td>Stores a cumulative histogram of the current states a file</td>
</tr>
<tr>
<td>ListenerGzipWriter</td>
<td>Stores the current state to GZIP compressed file</td>
</tr>
<tr>
<td>ListenerPrinter</td>
<td>Prints the event passed to notify()</td>
</tr>
<tr>
<td>ListenerWriter</td>
<td>Stores the event passed to notify() to a file</td>
</tr>
<tr>
<td>MeanWriter</td>
<td>Stores the cumulative mean vector to a file</td>
</tr>
<tr>
<td>PosteriorProbWriter</td>
<td>Stores the (unnormalized) posterior probability of the current state to a file</td>
</tr>
<tr>
<td>QuantileWriter</td>
<td>Stores the cumulative quantiles to a file</td>
</tr>
<tr>
<td>StepListenerPrinter</td>
<td>Prints MCMCStepEvents</td>
</tr>
<tr>
<td>StrippedListenerGzipWriter</td>
<td>Stores the current state to a GZIP compressed file</td>
</tr>
<tr>
<td>StrippedListenerWriter</td>
<td>Stores the current state to a file</td>
</tr>
<tr>
<td>ThinningProxyListener</td>
<td>A proxy for other listeners that thins the reported events by a specified factor, eg 1 out of every 100</td>
</tr>
</tbody>
</table>
D Output from Binomial_BetaBinomial_Example.java

$>$ java org.omegahat.Simulation.MCMC.Examples.Binomial_BetaBinomial_Example

Chain Step Event (with details)
Last = ContainerState: | 0.9 0.23 0.71 0.49 |
Last Prob = -359.046964566765
Proposed State = ContainerState: | 0.9 0.4751068415506061 0.71 0.49 |
Proposed Prob = -423.26454869568283
Current State = ContainerState: | 0.9 0.23 0.71 0.49 |
Forward Prob = -0.0369493853787235
Reverse Prob = -0.0369493853787235
Acceptance Prob = -64.21758412891785
Acceptance Val = 0.658405257229882
Accepted? = false
Acceptance Rate = 0.0

Chain Step Event (with details)
Last = ContainerState: | 0.9 0.23 0.71 0.49 |
Last Prob = -359.046964566765
Proposed State = ContainerState: | 0.9 0.23 0.71 0.49 |
Proposed Prob = -360.0165066537818
Current State = ContainerState: | 0.9 0.23 0.4561096830883196 |
Forward Prob = -0.06327351354448152
Reverse Prob = -0.06327351354448152
Acceptance Prob = -0.969542087016805
Acceptance Val = 0.31056162077494043
Acceptance Rate = 0.0

Chain Step Event (with details)
Last = ContainerState: | 0.9 0.23 0.4561096830883196 |
Last Prob = -360.0165066537818
Proposed State = ContainerState: | 0.9 0.23 0.4561096830883196 |
Proposed Prob = -360.1951841070053
Current State = ContainerState: | 0.9 0.23 0.4561096830883196 |
Forward Prob = 0.6154440530408976
Reverse Prob = 0.6154440530408976
Acceptance Prob = -0.17867745322348583
Acceptance Val = 0.13272539253007873
Accepted? = true
Acceptance Rate = 0.5

Chain Step Event (with details)
Last = ContainerState: | 0.9 0.23 0.4561096830883196 |
Last Prob = -360.1951841070053
Proposed State = ContainerState: | 0.9 0.23 0.4561096830883196 |
Proposed Prob = -363.1065248979661
Current State = ContainerState: | 0.9 0.23 0.4561096830883196 |
Forward Prob = 0.23049155040119496
Reverse Prob = 0.23049155040119496
Acceptance Prob = -Infinity
Acceptance Val = 0.335025050367706
Accepted? = false
Acceptance Rate = 0.5

Chain Step Event (with details)
Last = ContainerState: | 0.9 0.23 0.4561096830883196 |
Last Prob = -360.1951841070053
Proposed State = ContainerState: | 0.9 0.23 0.4561096830883196 |
Proposed Prob = -363.1065248979661
Current State = ContainerState: | 0.9 0.23 0.4561096830883196 |
Forward Prob = 0.23049155040119496
Reverse Prob = 0.23049155040119496
Acceptance Prob = -Infinity
Acceptance Val = 0.335025050367706
Accepted? = false
Acceptance Rate = 0.5
Acceptance Prob = -2.9113407909608213
Acceptance Val = 0.14726731467399157
Accepted? = false
Acceptance Rate = 0.4

Chain Step Event (with details)
Last = ContainerState: [0.9 0.23 0.45610096830883196 0.4225189574152196]
Last Prob = -360.1951841070053
Proposed State = ContainerState: [0.9 0.23 0.1825266980628881 0.4225189574152196]
Proposed Prob = -364.9924350935826
Current State = ContainerState: [0.9 0.23 0.45610096830883196 0.4225189574152196]
Forward Prob = -0.1255457772139429
Reverse Prob = -0.1255457772139429
Acceptance Prob = -4.797250988657353
Acceptance Val = 0.4310649477090058
Accepted? = false
Acceptance Rate = 0.3333333333333333

Chain Step Event (with details)
Last = ContainerState: [0.9 0.23 0.45610096830883196 0.4225189574152196]
Last Prob = -360.1951841070053
Proposed State = ContainerState: [0.9 0.23 0.45610096830883196 0.5879703533000055]
Proposed Prob = -359.85442835072524
Current State = ContainerState: [0.9 0.23 0.45610096830883196 0.5879703533000055]
Forward Prob = 0.3415983954458079
Reverse Prob = 0.3415983954458079
Acceptance Prob = 0.0
Acceptance Val = 0.36058319097411967
Accepted? = true
Acceptance Rate = 0.42857142857142855

Chain Step Event (with details)
Last = ContainerState: [0.9 0.23 0.45610096830883196 0.5879703533000055]
Last Prob = -360.1951841070053
Proposed State = ContainerState: [0.9 0.23 0.45610096830883196 0.5879703533000055]
Proposed Prob = -361.2666664185844
Current State = ContainerState: [0.9 0.23 0.45610096830883196 0.5879703533000055]
Forward Prob = 0.9268299028867995
Reverse Prob = 0.9268299028867995
Acceptance Prob = 0.0
Acceptance Val = 0.42857142857142855
Accepted? = false
Acceptance Rate = 0.375

Chain Step Event (with details)
Last = ContainerState: [0.9 0.23 0.45610096830883196 0.5879703533000055]
Last Prob = -359.85442835072524
Proposed State = ContainerState: [0.9 0.23 0.45610096830883196 0.5879703533000055]
Proposed Prob = -360.42346255905113
Current State = ContainerState: [0.9 0.23 0.45610096830883196 0.5879703533000055]
Forward Prob = 0.3211093780366665
Reverse Prob = 0.3211093780366665
Acceptance Prob = -0.5690342083258884
Acceptance Val = 0.3211093780366665
Accepted? = true
Acceptance Rate = 0.4444444444444444

Chain Step Event (with details)
Last = ContainerState: [0.9 0.23 0.45610096830883196 0.5879703533000055]
Last Prob = -360.42346255905113
Proposed State = ContainerState: [0.9 0.23 0.45610096830883196 0.5879703533000055]
Proposed Prob = -365.44521171685466
Current State = ContainerState: [0.9 0.23 0.45610096830883196 0.5879703533000055]
| 0.58797035333000055 | Forward Prob  = \(-0.2084655958574132\)
| Reverse Prob   = \(-0.2084655958574132\)
| Acceptance Prob = \(-5.0217491578035265\)
| Acceptance Val = 0.7418864833851747
| Accepted?      = false
| Acceptance Rate = 0.4 |