

## 1 MCMC Methodology.

Independence of  $X_1, \dots, X_n$  is not critical for an approximation of the form  $E^{\theta|x}h(X) = \frac{1}{n} \sum_{i=1}^n h(X_i)$ ,  $X_i \sim \pi(\theta|x)$ . In fact, when  $X$ 's are dependent, the ergodic theorems describe the approximation.

An easy and convenient form of dependence is Markov chain dependence. The Markov dependence is perfect for computer simulations since for producing a future realization of the chain, only the current state is needed.

### 1.1 Theoretical Background and Notation

Random variables  $X_1, X_2, \dots, X_n, \dots$  constitute a Markov Chain on continuous state space if they possess a Markov property,

$$P(X_{n+1} \in A | X_1, \dots, X_n) = P(X_{n+1} \in A | X_1, \dots, X_n) = Q(X_n, A) = Q(A | X_n),$$

for some probability distribution  $Q$ . Typically,  $Q$  is assumed a time-homogeneous, i.e., independent on  $n$  ("time"). The transition (from the state  $n$  to the state  $n+1$ ) kernel defines a probability measure on the state space and we will assume that the density  $q$  exists, i.e.,

$$Q(A | X_n = x) = \int_A q(x, y) dy = \int_A q(y | x) dy.$$

Distribution  $\Pi$  is invariant, if for all measurable sets  $A$

$$\Pi(A) = \int Q(A | x) \Pi(dx).$$

If the transition density  $\pi$  exists, it is stationary if  $q(x|y)\pi(y) = q(y|x)\pi(x)$ . Here and in the sequel we assume that the density for  $\Pi$  exists,  $\Pi(A) = \int_A \pi(x) dx$ .

A distribution  $\Pi$  is an equilibrium distribution if for  $Q^n(A|x) = P(X_n \in A | X_0 = x)$ ,

$$\lim_{n \rightarrow \infty} Q^n(A|x) = \Pi(A).$$

In plain terms, the Markov chain will forget the initial distribution and will converge to the stationary distribution.

The Markov Chain is *irreducible* if for each  $A$  for which  $\Pi(A) > 0$ , and for each  $x$ , one can find  $n$ , so that  $Q^n(A|x) > 0$ .

The Markov Chain  $X_1, \dots, X_n, \dots$  is *recurrent* if for each  $B$  such that  $\Pi(B) > 0$ ,

$$P(X_n \in B \text{ i.o.} | X_0 = x) = 1, \text{ a.s. (in distribution of } X_0)$$

It is *Harris recurrent* if  $P(X_n \in B \text{ i.o.} | X_0 = x) = 1, (\forall x)$ . The acronym *i.o.* stands for infinitely often.



Figure 1: Nicholas Constantine Metropolis, 1915-1999

## 1.2 Metropolis Algorithm

Metropolis algorithm is the fundamental to MCMC development.

Assume that the target distribution is known up to a normalizing constant. We would like to construct a chain with  $\pi$  as its stationary distribution.

As in ARM, we take a proposal distribution  $q(x, y) = q(y|x)$ , where the proposal for a new value of a chain is  $y$ , given that the chain is at value  $x$ .

Thus  $q$  defines transition kernel  $Q(A, x) = \int_A q(y|x)dy$  which is the probability of transition to some  $y \in A$ .

**Detailed Balance Equation.** A Markov Chain with transition density  $q(x, y) = q(y|x)$  satisfies detailed balance equation if there exists a distribution  $f$  such that

$$q(y|x)f(x) = q(x|y)f(y). \quad (1)$$

The distribution  $f$  is stationary (invariant) and the chain is reversible.

Indeed, if (1) holds,  $\int q(x|y)f(y)dy = \int q(y|x)f(x)dy = f(x) \int q(y|x)dy = f(x)$ , which is the definition of invariant distribution.

For a given target distribution  $\pi$ , the proposal  $q$  is admissible if

$$\text{supp } \pi(x) \subset \cup_x \text{supp } q(\cdot|x).$$

Metropolis-Hastings Algorithm is universal. One can select an arbitrary proposal distribution that is admissible. Of course such arbitrary distribution/kernel cannot be expected to satisfy the detailed balance equation (1) for the target distribution  $\pi$ , i.e.,

$$q(y|x)\pi(x) \neq q(x|y)\pi(y).$$

Suppose (wlog)

$$q(y|x)\pi(x) > q(x|y)\pi(y).$$

Then there is a factor  $\rho(x, y) \leq 1$  such that the above inequality is balanced,

$$q(y|x) \cdot \rho(x, y) \cdot \pi(x) = q(x|y)\pi(y) \cdot 1.$$

By solving with respect to  $\rho(x, y)$  one obtains,

$$\rho(x, y) = \frac{q(x|y)\pi(y)}{q(y|x)\pi(x)} \wedge 1,$$

where  $a \wedge b$  denotes  $\min\{a, b\}$ . What is the transition kernel corresponding to modified equation?  $q_M(y|x) = q(y|x)\rho(x, y) + \mathbf{1}(y = x)(1 - \int q(y|x)\rho(x, y)dy)$ .

### Metropolis-Hastings Algorithm.

Assume that target distribution  $\pi$  is known up to the normalizing constant. This may be the case of posteriors which are always known up to the proportionality constant as products of the likelihood and a prior.

STEP 1	Start with arbitrary $x_0$ from the support of target distribution.
STEP 2	At stage $n$ , generate proposal $y$ from $q(y x_n)$ .
STEP 3	Take $x_{n+1} = y$ with probability $\rho(x_n, y) = \frac{q(x_n y)\pi(y)}{q(y x_n)\pi(x_n)} \wedge 1$ . Otherwise, take $x_{n+1} = x_n$ . This random acceptance is done by generating a uniform on (0,1) random variable $U$ and accepting the proposal $y$ if $U \leq \rho(x_n, y)$ .
STEP 4	Increase $n$ and return to STEP 2.

### Some Common Choices for $q$ .

If  $q(x|y) = q(y|x)$ , i.e. if the kernel is symmetric, the acceptance ratio  $\rho(x, y)$  simplifies to

$$\frac{\pi(y)}{\pi(x)} \wedge 1,$$

since the proposal kernels from the numerator and denominator cancel. If in addition  $q$  depends on  $(x, y)$  via  $|y - x|$ , i.e.,  $q(x, y) = q^*(|y - x|)$ , for some distribution  $q^*$ , the algorithm is called the *Metropolis random walk*. A symmetric kernel is the original proposal from Metropolis et al. (1953).

If the proposal  $q(x, y)$  does not depend on  $x$ , i.e.,

$$q(y|x) = q(y),$$

the algorithm is called the *independence Metropolis*. It is similar to the acceptance/rejection method (ARM) but unlike the ARM, every step produces a realization from the target distribution. That realization may be repeated many times which is the case when proposal is not accepted and current state is repeatedly taken to be the new state.

### 1.2.1 Examples

[From Johnson and Albert (1999)] A small company improved a product and wants to infer about the proportion of potential customers who will buy the product if the new product is preferred to the old one. The company is certain that this proportion will exceed 0.5, i.e. and uses the uniform prior on  $[0.5, 1]$ . Out of 20 customers surveyed, 12 prefer the new product. Find the posterior for  $p$ .

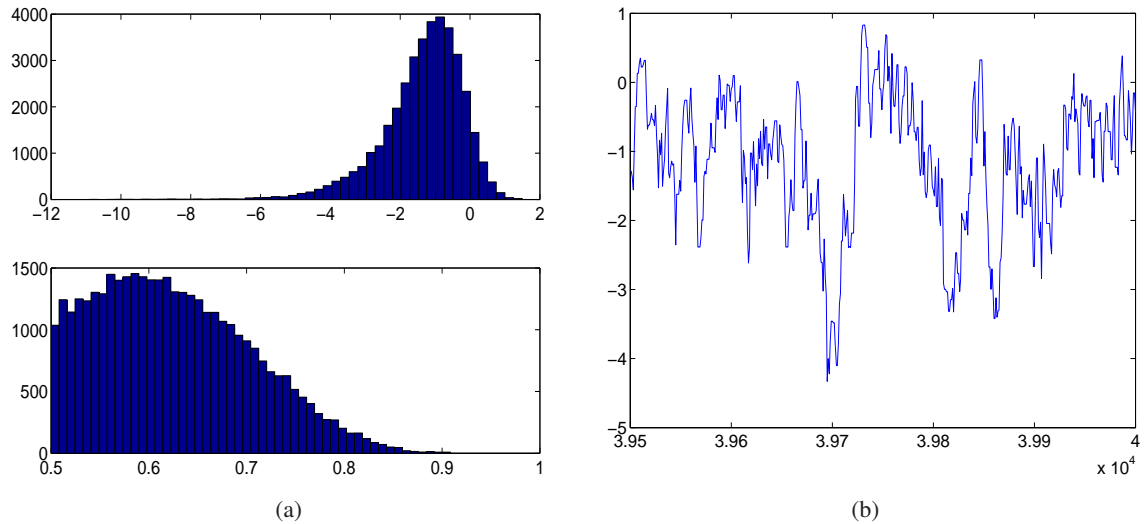


Figure 2:  $s = 0.5$

Since the support of  $p$  is  $[0.5, 1]$ , we transform the data by  $\theta = \log \frac{p-0.5}{1-p}$ , so that  $\theta \in (-\infty, \infty)$ . For such  $\theta$  it is easier to specify the proposal, although one can construct Metropolis chain for the untransformed parameter. The inverse transformation is

$$p = \frac{1/2 + \exp\{\theta\}}{1 + \exp\{\theta\}},$$

with Jacobian  $\frac{1/2 \exp\{\theta\}}{(1 + \exp\{\theta\})^2}$ , and the density for  $\theta$  is proportional to

$$\frac{(1/2 + \exp\{\theta\})^{12} \exp\{\theta\}}{(1 + \exp\{\theta\})^{22}}.$$

The proposal distribution is normal  $\mathcal{N}(\theta_n, s^2)$ , where  $\theta_n$  is current state of the chain and  $s^2$  is to be specified. Here is matlab program illustrating the sampling (`albertmcl.m` at the course web page).

```
%-----
nn = 40000; %      nn=number of metropolis iterations
s=10; %      s = std of normal proposal density
burn=2000; %      burn = burnin amount
%-----
ps=[];
thetas=[]; %transformed p's
old = 0; % start, theta_0
for i = 1:nn
    prop = old + s*randn(1,1); %proposal from N(theta_old, s^2)
    u = rand(1,1);
    ep=exp(prop); eo=exp(old);
    post_p=((1/2 + ep)^12 * ep)/((1+ep)^22);
    post_o=((1/2 + eo)^12 * eo)/((1+eo)^22);
    new = old;
    if u <= min(post_p/post_o, 1)
```

```

new = prop; %accept proposal as 'new'
old = new; % and set 'old' to be the 'new'
% for the next iteration;
end
thetas = [thetas, new]; %collect all theta's
ps=[ps, (1/2+exp(new))/(1+exp(new))]; %back-transformation to p's.
end

```

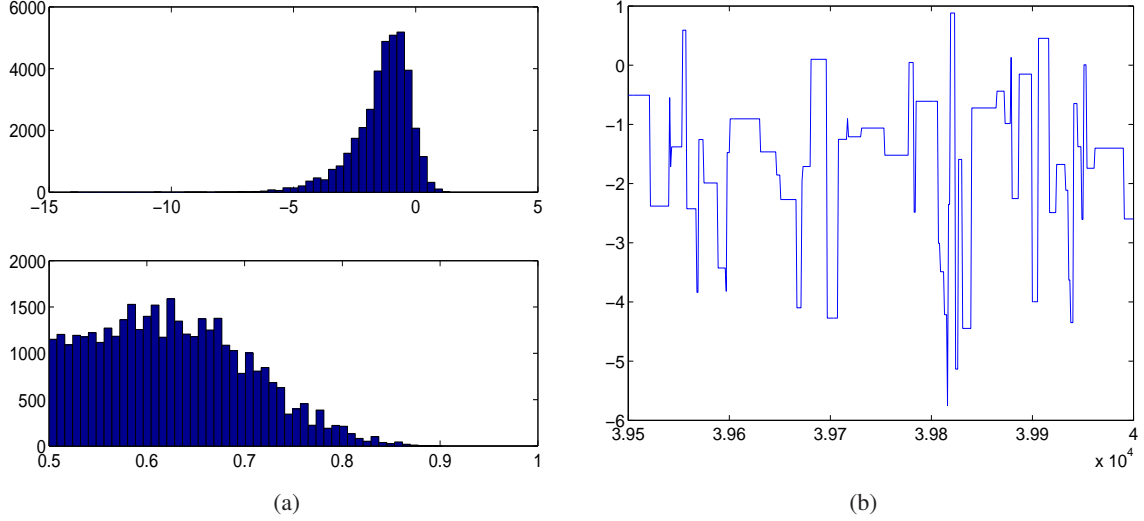


Figure 3:  $s = 10$

Figures 2 and 3 illustrate the simulation for  $s = 0.5$  and  $s = 10$  in the proposal distribution. Panels (a) depict the histograms for  $\theta$  and  $p$ , while the panel (b) depicts the last 500 simulations of the chain. Notice that the chain in Figure 2(b) mixes well and indeed reminds a random walk, while its counterpart in Figure 3(b) shows poor mixing.

**Weibull Example.** The Weibull distribution is used extensively in reliability, queueing theory, and many other engineering applications, partly for its ability to describe different hazard rate behavior and partly for historic reasons. The Weibull distribution parameterized by  $\alpha$  - the shape or slope, and  $\eta^{-1/\alpha}$  - the scale,

$$f(x|\alpha, \eta) = \alpha \eta x^{\alpha-1} e^{-x^\alpha \eta},$$

is not a member of the exponential family of distributions and explicit posteriors for  $\alpha$  and  $\eta$  are impossible.

Consider the prior  $\pi(\alpha, \eta) \propto e^{-\alpha} \cdot \eta^{\beta-1} e^{-\beta \eta}$ , and observations `data = [0.200 0.100 0.250];`. Imagine these data are extremely expensive – obtained by performing a destructive inspection of the pricey products.

Construct MCMC based on the Metropolis-Hastings algorithm and approximate posteriors for  $\alpha$  and  $\eta$ . Assume the hyperparameter `beta = 2;` and proposal distribution

$$q(\alpha', \eta' | \alpha, \eta) = \frac{1}{\alpha \eta} \exp \left\{ -\frac{\alpha'}{\alpha} - \frac{\eta'}{\eta} \right\}$$

(product of two exponentials with means  $\alpha$  and  $\eta$ ). Note that  $q(\alpha', \eta' | \alpha, \eta) \neq q(\alpha, \eta | \alpha', \eta')$  and  $q$  does not cancel in the expression for  $\rho$ .

Some hints that are checked and should work well:

(i) Start with arbitrary initial values, say:

```
alpha = 2; eta = 2;
```

(ii) Set `alphas = []`; `etas = []`; and do:

```
for i = 1:10000 alpha_prop = - alpha * log(rand); eta_prop = -
eta * log(rand);
%-----
    prod1 = prod(data);
    prod2 = prod( exp( eta * data.^alpha - eta_prop * data.^alpha_prop));
%-----
rr = (eta_prop/eta)^(beta-1) * exp(alpha - alpha_prop - beta * ...
    (eta_prop - eta)) * exp(- alpha/alpha_prop - eta/eta_prop + ...
    alpha_prop/alpha + eta_prop/eta)*prod1.^(alpha_prop - alpha) * ...
    prod2 * ((alpha_prop * eta_prop)/(alpha * eta))^(n-1);
%-----
r = min( rr ,1);
    if (rand < r)
        alpha = alpha_prop; eta = eta_prop;
    end
alphas = [alphas alpha]; etas = [etas eta]; end
```

Values `alpha_prop` and `eta_prop` are proposals from independent exponential distributions with means `alpha` and `eta`. “Burn in” 5000 out of 10000 simulations (usually 100-500 is enough) to make sure that there is no influence of the initial values for  $\alpha$  and  $\eta$ , and plot the histograms of their posterior distributions.

```
figure(1) subplot(1,2,1) hist(alphas(5000:end),100) subplot(1,2,2)
hist(etas(5000:end),100)
```

Finally, report the mean and variance of `alphas` and `etas`. These are desired Bayes estimators with their posterior precisions.

### 1.3 Gibbs Sampler

The Gibbs sampler, introduced in [11], is a special case of A Single Component Metropolis algorithm. Define  $\mathbf{X} = (X_1, X_2, \dots, X_p)$ . Each step of the algorithm will consist of  $p$  coordinatewise updates.

Define  $\mathbf{X}_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p)$ , and  $\mathbf{X}_{-i}^n = (X_1^{n+1}, \dots, X_{i-1}^{n+1}, X_{i+1}^n, \dots, X_p^n)$ ,  $i = 1, \dots, p$ .  $\mathbf{X}_{-i}^n$  is the update at step  $n$ , where first  $i - 1$  coordinates are updated to their values at step  $n + 1$  and the coordinates at positions  $i + 1, i + 2, \dots, p$  are not updated and still are at step  $n$ .

Let  $q_i(Y_i | X_i^n, \mathbf{X}_{-i}^n)$  be the proposal distribution that generates proposals for  $i$ th coordinate only. Then the single component Metropolis proceeds as follows:

Generate a candidate  $Y_i \sim q_i(Y_i | X_i^n, \mathbf{X}_{-i}^n)$ .

Accept the candidate  $Y_i$  with probability

$$\rho(X_i^n, \mathbf{X}_{-i}^n, Y_i) = \frac{\pi(Y_i | \mathbf{X}_{-i}^n) q_i(X_i^n | Y_i, \mathbf{X}_{-i}^n)}{\pi(X_i^n | \mathbf{X}_{-i}^n) q_i(Y_i | X_i^n, \mathbf{X}_{-i}^n)}.$$

as  $X_i^{n+1} = Y_i$ . Otherwise set  $X_i^{n+1} = X_i^n$ .

Go to the next coordinate at step  $n$ .

Let  $\pi(X_i | \mathbf{X}_{-i}) = \frac{\pi(\mathbf{X})}{\int \pi(\mathbf{X}) dX_i}$  be the full conditional. The Gibbs sampler is a single step Metropolis algorithm with  $q_i(Y_i | X_i, \mathbf{X}_{-i}) = \pi(Y_i | \mathbf{X}_{-i})$ .

Obviously, at each step  $n$ ,  $\rho(X_i^n, \mathbf{X}_{-i}^n, Y_i) = 1$ , and each update in Gibbs algorithm is accepted.

In more familiar notation, suppose that  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$  is multidimensional parameter of interest. Each component can be univariate or multivariate. Suppose that we can simulate from the conditional densities  $\pi(\theta_i | \boldsymbol{\theta}_{-i})$ , where as the above,  $\boldsymbol{\theta}_{-i}$  denotes the parameter vector  $\boldsymbol{\theta}$  without the component  $i$ . If the current state of  $\boldsymbol{\theta}$  is  $\boldsymbol{\theta}^n = (\theta_1^n, \theta_2^n, \dots, \theta_p^n)$ , the Gibbs sampler produces  $\boldsymbol{\theta}^{n+1}$  in the following way:

---


$$\begin{aligned} & \text{Draw } \theta_1^{n+1} \text{ from } \pi(\theta_1 | \theta_2^n, \theta_3^n, \dots, \theta_p^n) \\ & \text{Draw } \theta_2^{n+1} \text{ from } \pi(\theta_2 | \theta_1^{n+1}, \theta_3^n, \dots, \theta_p^n) \\ & \text{Draw } \theta_3^{n+1} \text{ from } \pi(\theta_3 | \theta_1^{n+1}, \theta_2^{n+1}, \theta_4^n, \dots, \theta_p^n) \\ & \quad \vdots \\ & \text{Draw } \theta_{p-1}^{n+1} \text{ from } \pi(\theta_{p-1} | \theta_1^{n+1}, \theta_2^{n+1}, \dots, \theta_{p-2}^{n+1}, \theta_p^n) \\ & \text{Draw } \theta_p^{n+1} \text{ from } \pi(\theta_p | \theta_1^{n+1}, \theta_2^{n+1}, \dots, \theta_{p-1}^{n+1}) \end{aligned}$$


---

Above, we have assumed a fixed updating order. This may not always be the case, since it is possible to generalize the Gibbs Sampler in a number of ways. It is possible to assume the random updating order, i.e. pick the block to update randomly. It is also possible to update only one block per iteration and to choose the block to update with some preassigned probability, see Gilks et al. [16] for discussion.

### 1.3.1 Finding the Full Conditionals

The full conditionals, needed for implementation of the Gibbs sampler are *conceptually* easy to find. From the joint distribution of all variables, only expressions that contain the particular variable are entering to the conditional distribution. The difficulty is (as always) in finding normalizing constants.

Suppose  $\boldsymbol{\theta} = (\boldsymbol{\theta}_s, \boldsymbol{\theta}_{-s})$  and we are interested in the full conditional for  $\boldsymbol{\theta}_s$ . The full conditional is

$$\pi(\boldsymbol{\theta}_s | \boldsymbol{\theta}_{-s}) = \frac{\pi(\boldsymbol{\theta}_s, \boldsymbol{\theta}_{-s})}{\int \pi(\boldsymbol{\theta}_s, \boldsymbol{\theta}_{-s}) d\boldsymbol{\theta}_s} \propto \pi(\boldsymbol{\theta}_s, \boldsymbol{\theta}_{-s}).$$

**Example.** A popular simple model to illustrate Gibbs sampler and finding the full conditionals the following [Gilks (from [12]), Chapter 5, page 76.]

$$\begin{aligned} Y_1, Y_2, \dots, Y_n & \sim \mathcal{N}(\mu, 1/\tau) \\ \mu & \sim \mathcal{N}(0, 1) \\ \tau & \sim \text{Gamma}(2, 1). \end{aligned}$$

The joint distribution is

$$\begin{aligned} f(y, \mu, \tau) &= \left\{ \prod_{i=1}^n f(y_i | \mu, \tau) \right\} \pi(\mu) \pi(\tau) \\ &= (2\pi)^{-(n+1)/2} \tau^{n/2} \exp \left\{ -\tau/2 \sum_{i=1}^n (y_i - \mu)^2 \right\} \exp\{-1/2\mu^2\} \tau \exp\{-\tau\}. \end{aligned}$$

To find the full conditional for  $\mu$  we select the terms from  $f(y, \mu, \tau)$  that contain  $\mu$  and normalize. Indeed,

$$\begin{aligned} \pi(\mu | \tau, y) &= \frac{\pi(\mu, \tau | y)}{\pi(\tau | y)} \\ &= \frac{\pi(\mu, \tau, y)}{\pi(\tau, y)} \propto \pi(\mu, \tau, y). \end{aligned}$$

Thus,

$$\begin{aligned} \pi(\mu | \tau, y) &\propto \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n (y_i - \mu)^2 \right\} \exp\{-1/2\mu^2\} \\ &\propto \exp \left\{ -\frac{1}{2}(1 + n\tau) \left( \mu - \frac{\tau \sum y_i}{1 + n\tau} \right)^2 \right\}, \end{aligned}$$

which is normal  $\mathcal{N}(\frac{\tau \sum y_i}{1 + n\tau}, \frac{1}{1 + n\tau})$  distribution. Similarly,

$$\begin{aligned} \pi(\tau | \mu, y) &\propto \tau^{n/2} \exp \left\{ -\tau/2 \sum_{i=1}^n (y_i - \mu)^2 \right\} \tau \exp\{-\tau\} \\ &= \tau^{n/2+1} \exp \left\{ -\tau \left[ 1 + \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2 \right] \right\}, \end{aligned}$$

which is unnormalized gamma  $\mathcal{Gamma}(2 + n/2, 1 + \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2)$ .

The matlab code `mcmc1.m` implements the sampler. Function `rand_gamma.m` generates random gamma variates and is a part of BayesLab. We simulated  $n = 20$  observations from  $\mathcal{N}(1, 4^2)$  distribution and started a gibs with  $\mu_0 = 0$  and  $\tau_0 = 2$ .

```
n=30; % sample size
randn('state', 10);
y = 4 * randn(1,n) + 1;
%-----
NN = 10000;
mus = []; taus = [];
suma = sum(y);
mu = 0; % set the parameters as prior means
tau = 2; %
for i = 1 : NN
    new_mu = sqrt(1/(1+n*tau)) * randn + (tau * suma)/(1+n*tau);
```



```

par      = 1+1/2 * sum ( (y - mu).^2 );
new_tau = rand_gamma(2 + n/2, par, 1,1);
mus = [mus new_mu];
taus = [taus new_tau];
mu=new_mu;
tau=new_tau;
end

```

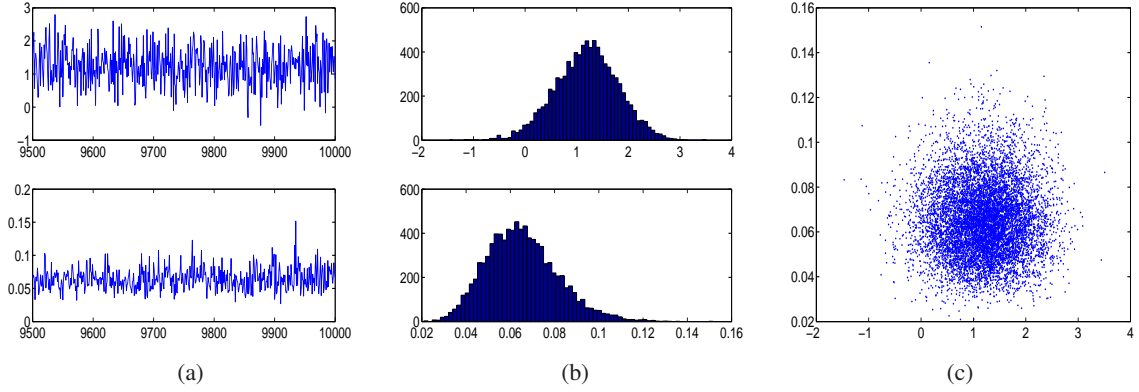


Figure 4: (a) Last 500 simulations (out of 10000) for  $\mu$  (top) and  $\tau$  (below); (b) histograms of  $\mu$  and  $\tau$ ; (c) joint simulation for  $\mu$  and  $\tau$ .

Figure 4(a) depicts last 500 simulations (out of 10000) for  $\mu$  (top) and  $\tau$  (below). Panel (b) gives histograms of  $\mu$  and  $\tau$ , and panel (c) presents a joint simulation for  $\mu$  and  $\tau$ . The burn-in period was 1000, so only 9000 variates have been used to approximate the posteriors.

The MCMC estimators are  $\hat{\mu} = 1.1938$  and  $\hat{\tau} = 0.0652$ . The performance is quite good since the theoretical parameters are 1 and  $1/16=0.0625$ , respectively.

## Exercises

**1. Beetle Mortality.** The data comes from Bliss [1] (cited in Dobson [2]) and is shown in Table 1. The data involves counting the number of beetles killed after five hours of exposure to various concentrations of gaseous carbon disulphide ( $CS_2$ ). The analysis concerns estimating the proportion  $r_i/n_i$  of beetles that are killed by the gas.

Consider the model

$$P(\text{death}|w_i) = h(w_i) = \left( \frac{\exp\{x_i\}}{1 + \exp\{x_i\}} \right)^{m_1},$$

where  $m_1 > 0$ , and  $w_i$  is a covariate (dose), and  $x_i = \frac{w_i - \mu}{\sigma}$ ,  $\mu \in \mathbb{R}$ ,  $\sigma^2 > 0$ .

Dosage ( $\log_{10} CS_2$ mg/litre)	Beetles	Killed
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

Table 1: Data on Beetle Mortality from Bliss (1935). Batches of adult beetles were exposed to gaseous carbon disulphide for five hours.

The priors are

$$\begin{aligned}
m_1 &\sim \mathcal{Gamma}(a_0, b_0), \\
\mu &\sim \mathcal{N}(c_0, d_0), \\
\sigma^2 &\sim \mathcal{IG}(e_0, f_0) \quad [\tau = \frac{1}{\sigma^2} \sim \mathcal{Gamma}(e_0, f_0)].
\end{aligned}$$

The joint posterior  $\pi(\mu, \sigma^2, m_1|y)$  is proportional to

$$\begin{aligned}
&f(y|\mu, \sigma^2, m_1)\pi(\mu, \sigma^2, m_1) \\
&\propto \left( \prod_{i=1}^k [h(w_i)]^{y_i} [1 - h(w_i)]^{n_i - y_i} \right) \times \frac{m_1^{a_0-1}}{(\sigma^2)^{e_0+1}} \exp \left\{ -\frac{1}{2} \left( \frac{\mu - c_0}{d_0} \right)^2 - \frac{m_1}{b_0} - \frac{1}{f_0 \sigma^2} \right\}.
\end{aligned}$$

The transformation  $\theta = (\theta_1, \theta_2, \theta_3) = (\mu, \frac{1}{2} \log(\sigma^2), \log m_1)$  is supported by  $R^2$  and multivariate normal proposal for  $\theta$  is possible. In the new variables,

$$\begin{aligned}
\pi(\theta|y) &\propto \left( \prod_{i=1}^k [h(w_i)]^{y_i} [1 - h(w_i)]^{n_i - y_i} \right) \times \exp\{a_0 \theta_3 - 2e_0 \theta_2\} \times \\
&\exp \left\{ -\frac{1}{2} \left( \frac{\theta_1 - c_0}{d_0} \right)^2 - \frac{\exp\{\theta_3\}}{b_0} - \frac{\exp\{-2\theta_2\}}{f_0} \right\}.
\end{aligned}$$

The acceptance probability is more computationally stable if it is represented as

$$\rho = \exp\{\log \pi^*(\theta^*|y) - \log \pi^*(\theta_n|y)\},$$

for unnormalized posterior  $\pi^*$ ,  $\theta^*$  proposal, and  $\theta_n$  current state. The choice of the hyperparameters is  $a_0 = 0.25$ ,  $b_0 = 4$  ( $m_1$  has prior mean 1, as in the standard logit model),  $c_0 = 2$ ,  $d_0 = 10$ ,  $e_0 = 2$ ,  $f_0 = 1000$ .

Proposal density is  $\mathcal{MVN}_3(\theta_n, \Sigma)$  with  $\Sigma = \text{diag}(0.00012, 0.033, 0.1)$ .

**2. Contingency Table.** See the matlab file `albertmc2.m` on the Bayes page. Figure 5 depicts the output.

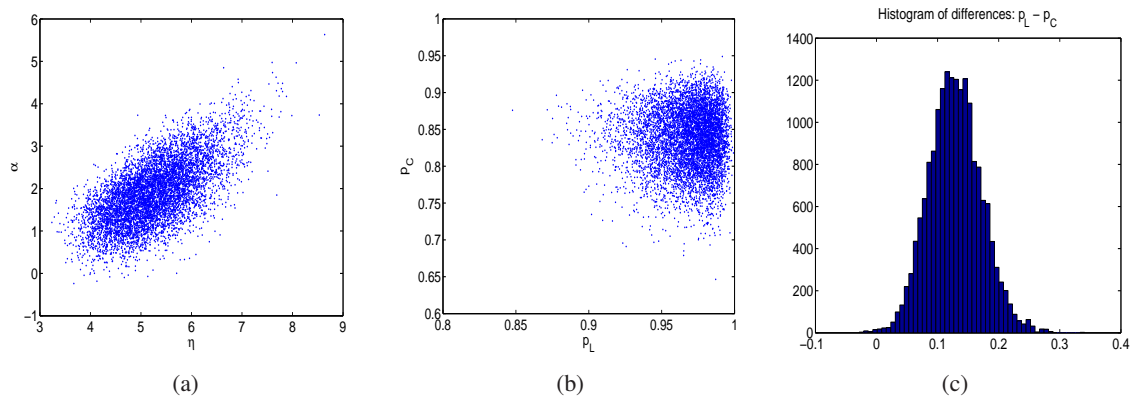


Figure 5: Graphical output from `albertmc2.m`.

## References

- [1] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B*, **36**, 192–236.
- [2] Bliss, C. I. (1935). The calculation of the dosage-mortality curve. *The Annals of Applied Biology*, **22**, 134–167.
- [3] Brooks, S.P. (1998). Monte Carlo Methods and its application. *The Statistician*, **47**, 69–100.
- [4] Brooks, S.P. (1998). Quantitative convergence assessment for Markov chain Monte Carlo via cusums. *Statistics and Computing*, **8**, 267–274.
- [5] Casella, G., George, E.I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, **46**, 167–174. (Celebrated “Gibbs for Kids”).
- [6] Chib, S., Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, **49**, 327–335.
- [7] Gelfand, A.E., Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.
- [8] Gelfand, A.E. (2000). Gibbs Sampling. *Journal of the American Statistical Association*, **95**, 1300–1304.
- [9] Robert, C. (2001). *Bayesian Choice*, Second Edition, Springer Verlag.
- [10] Dobson, A. J. (1983). *An Introduction to Statistical Modelling*. Chapman and Hall.
- [11] Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. Pami-6, No. 6, 721–741.
- [12] Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall.

- [13] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.
- [14] Johnson, V. and Albert, J. (1999). *Ordinal Data Modeling*. Springer Verlag, NY.
- [15] Metropolis, N., Rosenbluth, M.N., Teller, A.H., Teller, E. (1953). Equations of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, **21**, 1087–1092.
- [16] Robert, C.P., Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag, New York.
- [17] Roberts, G.O., Gelman, A., Gilks, W.R. (1997). Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *Annals of Applied Probability*, **7**, 110–120.
- [18] Tanner, M.A. (1996). *Tools for Statistical Inference*. Springer-Verlag, New York.
- [19] Tierney, L. (1994). Markov Chains for exploring posterior distributions( with discussion). *Annals of Statistics*, **22**, 1701–1762.