

Assignment 3 + 4

ECE 5412 Bayesian Estimation and Learning

Wanlin Li wl598

1.

Metropolis Hastings with Hierarchical Bayes Model for posterior probability

In this paper we will discuss the approach and application of Metropolis Hastings with Hierarchical Bayes Model for estimating posterior probability in solving the real world problem. We will go through the definition of Metropolis Hastings in Bayesian Estimation, then followed by the application in our problem with the performance analyzed. Indeed, we would also go over the benefits and why we would choose this method for our problem, to be more detailed, we will investigate the constraints, precision and what could be important for future studying, potential problems and errors will be explained as well.

For estimating the problem model we have came up with Bayesian random variable with a certain distribution to estimate the real world situation, like many of the random variables with different features we choose hierarchical modeling. the model has four super-parameters as the independent variables comparing to model, and such procedure to obtain the posterior distribution of the cell frequencies for subpopulation.

$$p(\theta) = n(\theta|\alpha) / \int n(\theta|\alpha) d\theta = p(\theta|\alpha)$$

where $n(\theta|\alpha)$ is the function with parameters α that was fit to the histogram

$$p(\theta|x) \propto p(x|\theta) p(\theta|\alpha)$$

So posterior is proportional to likelihood times prior by Bayes' rule, however individual samples from a large population can exist unknown influential factor α for describing the overall situation,

so we need to assume existence of extra layer in observing spectra, which is in another word, θ is regarded as a latent variable.

$$p(\alpha, \theta | x) \propto p(x | \theta, \alpha) p(\theta | \alpha) p(\alpha)$$

To obtain the samples from our distribution, comparing to acceptance-rejection method we are going to use the Metropolis Hasting for drawing samples based on previous state or value, which means this procedure is not independent. Moreover, we will be using random-walk metropolis, so by definition q transitional probability will be symmetric for

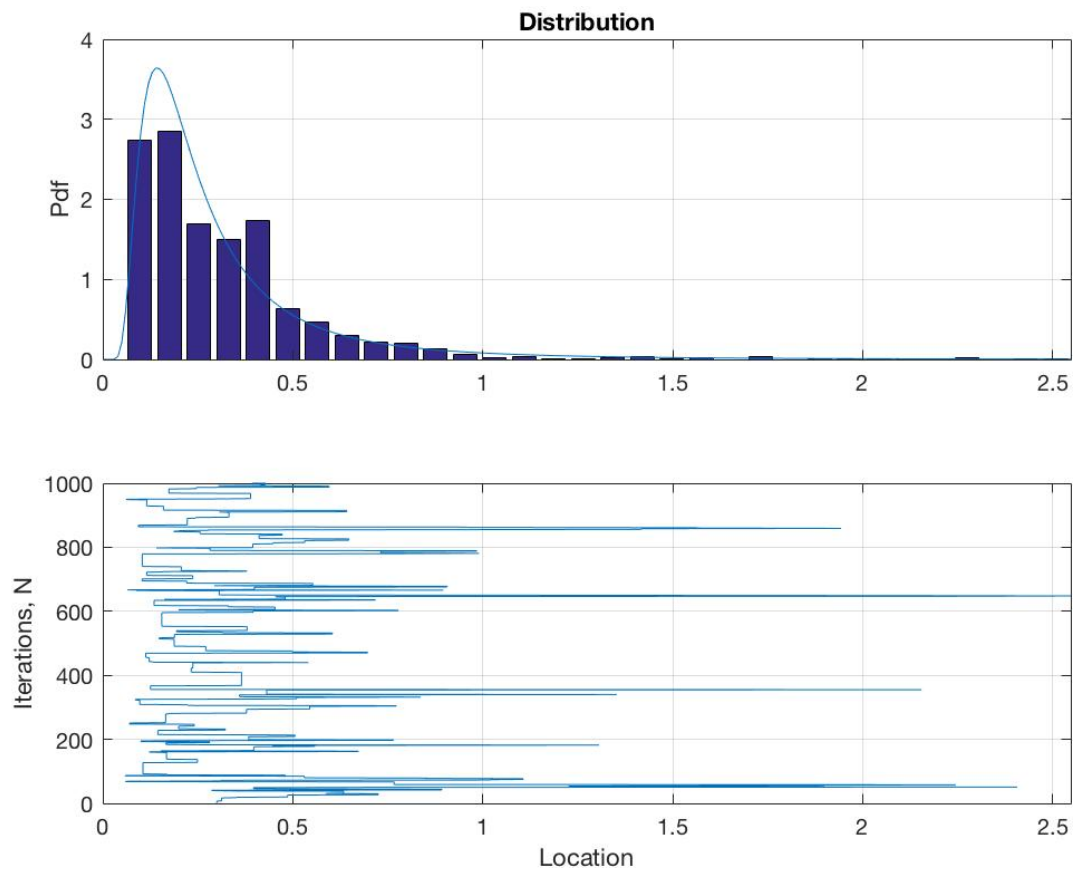
$$p(x | x') \text{ and } p(x' | x)$$

So the density function explains that when the density is larger than current density, we will accept such value into our model, if not it will be accepted under a probability. The prior information are about both allele frequencies and inter-population divergence.

After we have drawn enough samples from the sequence and we can apply Monte Carlo to estimate the result. The iterations do display some characters on the result, for example, the certain range of values are stuck at a value different from the rest of the dataset, but with the increase number of iterations the density function approach closely to the theoretical pdf function.

The structure of our model can be different based on dependence and correlations for a dataset when multiple variables are not independent. To be noticed, Monte Carlo method is essentially a stochastic Markov Chain, which balanced off the randomness and simplicity within the range of

our current studying. One of the advantages: the components have different dependence structures across sub-population.



In conclusion, we have analyzed Bayesian Hierarchical Model for Allele Frequencies and studied direct probabilistic relationship between theory and observation, Monte Carlo gave the studying enough room yielding the corresponding testing result. The model will become relatively simple when the stages reduce.

Reference

<https://eml.berkeley.edu/reprints/misc/understanding.pdf>

<http://www.stat.cmu.edu/~roeder/publications/lrd2001.pdf>

Appendix

```
clear all;
```

```

close all;
md= 5;
p= @(t) (t.^(-md/2-1)).*(exp(-1./(2*t)));
proposal_PDF = @(mu) unifrnd(0,3); %pdf
N= 1000;
nn= 0.1*(N);
theta= zeros(1,N);
theta(1)=0.3;
for i = 1:N
    theta_ast = proposal_PDF([]);
    alpha      = p(theta_ast)/p(theta(i));
    if rand <= min(alpha,1)
        theta(i+1) = theta_ast;
    else
        theta(i+1) = theta(i);
    end
end
xx = eps:0.01:10;
figure;
subplot(2,1,1);
[n1 x1] = hist(theta, ceil(sqrt(N)));
bar(x1,n1/(N*(x1(2)-x1(1))));
hold on;
plot(xx, p(xx)/trapz(xx,p(xx)));
grid on;
xlim([0 max(theta)]);
title('Distribution');
ylabel('Pdf');
% Samples
subplot(2,1,2);
plot(theta, 1:N+1);
xlim([0 max(theta)]);
ylim([0 N]);
grid on;
xlabel('Location');
ylabel('Iterations, N');

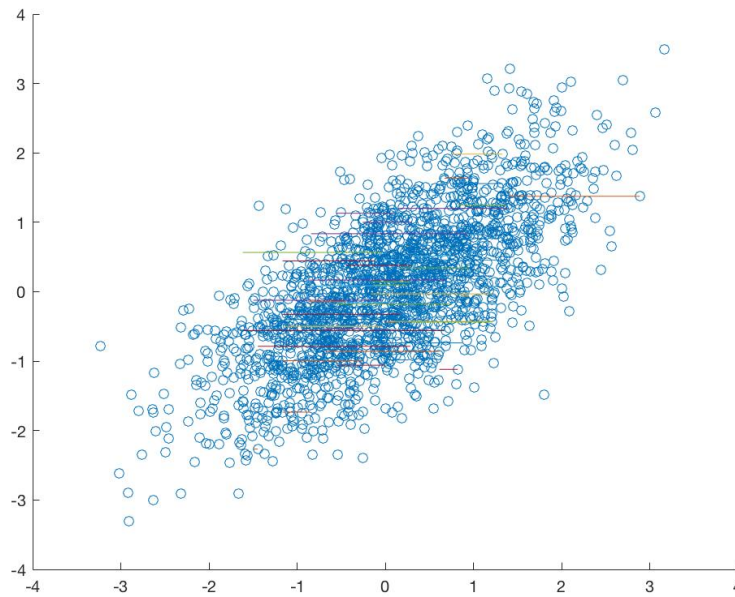
```

2.

Bayesian Inference on evidence in Gibbs samplings

In this paper we have studied forensic evidence that has been sampled from Gibbs sampling method. Without any sample size correction, the MCSE tends to underestimate the standard error for the Bayes' Factor. We will introduce the definition and implementation approach, then followed by the analysis of benefits and potential shortcomings with this method.

Gibbs sampling can give effective result without too much complexity as non-stochastic MCMC, and we obtain the converged solution as our result. In the paper, the bayes' factor are discussed as common factor in methods to rule out misleading evidence. During criminal trials, the qualitative study scandalized priors and nuisance parameters.



As one of the most common Monte Carlo Markov Chain simulation method, Gibbs sampling is one of the special condition of Metropolis Hasting, by observing existing variables we sample from a dimension that is lower comparing to the current dimension, we make Bayesian inference based on posterior distribution.

Since many of the terms are similar to the previous paper we have discussed, the basics about MCMC, Metropolis Hasting and Bayesian random variables will not be repeated in this analysis.

And we will get on to the example from the paper:

From the studying on XXX, we have observed that the sampling issue with the issue by

Set initial value $\theta(0) = (\theta(0), \dots, \theta(0))$

Draw $\theta(t)|\theta(t), \dots, \theta(t), Y$

Posterior draws $\theta(1), \dots, \theta(S)$

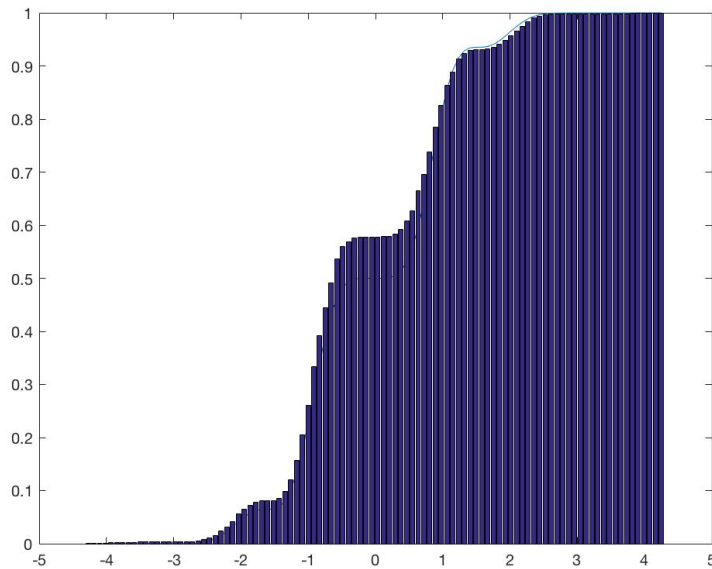
<http://www.mit.edu/~ilkery/papers/GibbsSampling.pdf>

<https://openprairie.sdstate.edu/cgi/viewcontent.cgi?article=2780&context=etd>

Appendix

```
rand('seed' ,12345);
n = 2000;
mu = [0 0];
w(1) = 0.5;
w(2) = 0.7;
propSigma = 1;
minn = [-4 -4];
maxx = [4 4];
%scatter
x = zeros(n,2);
x(1,1) = unifrnd(minn(1), maxx(1));
x(1,2) = unifrnd(minn(2), maxx(2));
dims = 1:2;
t = 1;
while t < n
    t = t + 1;
    T = [t-1,t];
    for iD = 1:2
        nIx = dims~iD;
        muCond = mu(iD) + w(iD)*(x(T(iD),nIx)-mu(nIx));
        varCond = sqrt(1-w(iD)^2);
        x(t,iD) = normrnd(muCond,varCond);
    end
end
figure;
h1 = scatter(x(:,1),x(:,2));
hold on;
for t = 1:50
    plot([x(t,1),x(t+1,1)],[x(t,2),x(t,2)]);
    h2 = plot(x(t+1,1),x(t+1,2));
end
legend([h1,h2,h3],{'Samples','1st 50 Samples','x(t=0)'})
hold off;
xlabel('x_1');
ylabel('x_2');
```

4.



4

I have applied Metropolis Hastings to simulate $\pi(x) \propto \sin^2(x) \times \sin^2(2x) \times \psi(x)$ with Metropolis Hasting algorithm, so we the plot as below, the curve shows theoretical cdf and the histogram shows the empirical plot.

```
clear all;
close all;
%simulation
%step1
a=zeros(1,10001);
xe=zeros(1,10001);
xe(1)=3.14;
y = zeros(1,10001);
% step 2
u = rand(1,10001);
u2 = rand(1,10001);
%MH
% step 3
for i=2:1:10001
    y(i)=(u2(i)-0.5)*2+xe(i-1);
    piy = ((sin(y(i))^2)*(sin(2*y(i))^2)*(1/sqrt(2*pi)*exp(-y(i)^2/2)));
    pix = ((sin(xe(i-1))^2)*(sin(2*xe(i-1))^2)*(1/sqrt(2*pi)*exp(-
xe(i-1)^2/2)));
    a(i)=min(1,piy/pix);
    if u(i)<a(i)
        xe(i)=y(i);
    else
        xe(i)=xe(i-1);
    end
end
[e,center]=hist(xe,100);
e=e./sum(e);
ecdf=zeros(1,length(e));
for index=1:length(e)
    ecdf(index)=sum(e(1:index));
end
```

```

%theoretical
f = 1/sqrt((2*pi))*exp(-center.^2/2);
a = sin(center).^2; b = sin(2*center).^2;
%pdf plot of theoritical curve
p2 = a.*b.*f;
cdf = zeros(1,length(p2));
for i = 1:100
    cdf(i) = sum(p2(1:i));
end
figure;
cdf = cdf/cdf(100);
plot(center,cdf);
hold on
bar(center,ecdf);
hold off

```

3. $P(x=y) = \pi_{ij}$

transitional probability $Q_{ij} = \frac{b_j}{b_i + b_j}$

$$\begin{aligned}
 P_{1,i_1} P_{i_1,i_2} \dots P_{i_k,i_{k+1}} &= Q_{i_1,i_2} \frac{b_{i_2}}{b_{i_1} + b_{i_2}} Q_{i_2,i_3} \frac{b_{i_3}}{b_{i_2} + b_{i_3}} \dots Q_{i_k,i_{k+1}} \frac{b_{i_{k+1}}}{b_{i_k} + b_{i_{k+1}}} \\
 &= Q_{i_1,i_2} \frac{b_{i_2}}{b_{i_1} + b_{i_2}} Q_{i_2,i_3} \frac{b_{i_3}}{b_{i_2} + b_{i_3}} \dots Q_{i_k,i_{k+1}} \frac{b_{i_{k+1}}}{b_{i_k} + b_{i_{k+1}}} \\
 &= Q_{i_1,i_2} \frac{b_{i_2}}{b_{i_1} + b_{i_2}} Q_{i_2,i_3} \frac{b_{i_3}}{b_{i_2} + b_{i_3}} \dots Q_{i_k,i_{k+1}} \frac{b_{i_{k+1}}}{b_{i_k} + b_{i_{k+1}}} \\
 &= P_{i_1,i_2} P_{i_2,i_3} \dots P_{i_k,i_{k+1}}
 \end{aligned}$$

So, $P_{ij} \pi_{i,j} = P_{ji} \pi_{j,i} \Rightarrow$

Because Markov chain is Reversible for a symmetric transition probability matrix.

$$\begin{aligned}
 Q_{ij} \frac{b_j}{b_i + b_j} \pi_{i,j} &= Q_{ji} \frac{b_i}{b_j + b_i} \pi_{j,i} \\
 \frac{\pi_{i,j}}{\pi_{j,i}} &= \frac{b_i}{b_j} \\
 \pi_{ij} &\propto b_j
 \end{aligned}$$

6. Prove $p(x_1, x_2)$ can be obtained from $p(x_1 | x_2)$ and $p(x_2 | x_1)$,

$$p(x_1 | x_2) = \frac{p(x_1 \cap x_2)}{p(x_2)} \quad \text{if } x_1 \text{ and } x_2 \text{ are independent.}$$

$$p(x_2 | x_1) = \frac{p(x_1 \cap x_2)}{p(x_1)} \quad p(x_1 | x_2) p(x_2 | x_1) = p(x_1 \cap x_2)$$

For gibbs sampling, x_1, x_2 are, so $p(x_1, x_2) = p(x_1 | x_2) + p(x_2 | x_1) - p(x_1 \cap x_2)$,
or conditional distribution $p(x_1, x_2 \dots x_n) \propto p(x_1, x_2)$.

5.

We are simulating Covariation for Hierarchical Bayes' Model following the Example 3 from

$$x_i, y_i | \rho \sim N(\mu, \Sigma), \quad \mu \in \mathbb{R}^2, \Sigma = \begin{bmatrix} \sigma_{xx} = 1 & \rho \\ \rho & \sigma_{yy} = 1 \end{bmatrix}$$

$$p(x_i, y_i | \rho) = \prod_{i=1}^N \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}[x_i^2 - 2\rho x_i y_i + y_i^2]\right)$$

$$\text{prior} = p(\rho) \propto |\Sigma|^{-3/2} = \begin{vmatrix} 1 & \rho \\ \rho & 1 \end{vmatrix}^{-3/2} = (1-\rho^2)^{-3/2}$$

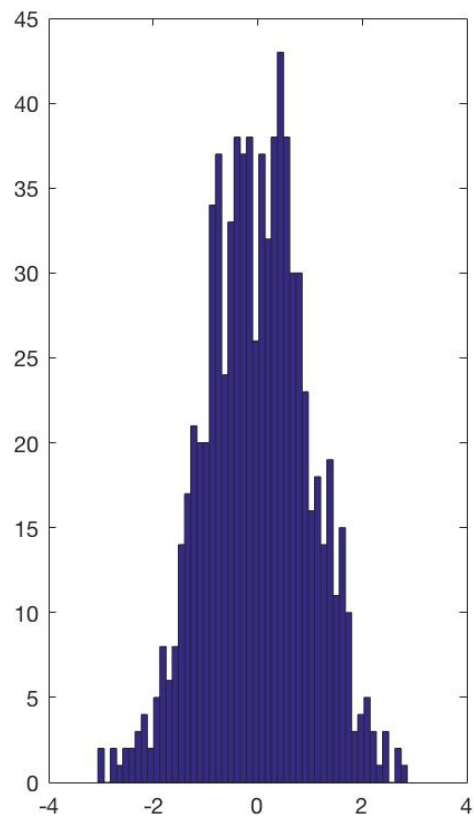
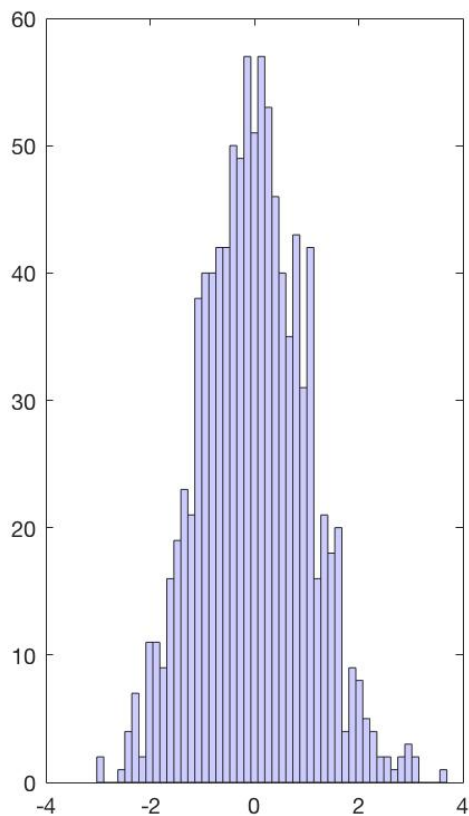
```

clear all
close all
mu = [0 0];
N=2000;
ro = 0.42;
sig = [1 ro; ro 1];
rv = mvnrnd(mu,sig,N);
i=0.45;
for t = 1:10000
    j=unifrnd(i-0.07,i+0.07);
    p1=0;
    p2=0;
    for k= 1:N
        p11=p1-1/(2*(1-i*i))*(rv(k,1)^.2;
        pp1 = rv(k,2)^2-2*i*rv(k,1)*rv(k,2)-log(2*pi*sqrt(1-i^2));
        p1 = p11+pp1;
        p22=p2-1/(2*(1-j*j))*(rv(k,1)^.2;
        pp2 = rv(k,2)^2-2*j*rv(k,1)*rv(k,2)-log(2*pi*sqrt(1-j^2));
        p2 = p22+pp2;
    end
    xx=exp(p2-p1);
    pj = ((1-j^2)^(-1.5));
    pi = (1-i^2)^(-1.5));
    yy=pj/pi;
    p=xx*yy;
    a=min(1,p);
    u=unifrnd(0,1);
    if u<a
        xe(1,t)=j;
        i=j;
    else
        xe(1,t)=i;
    end
end
end
histogram(xe)

```

7.

In comparison, it looks like the Slice sampler can sample uniformly from the region under the graph of its density function, similar to acceptance rejection. It is easier to do acceptance rejection but less precise comparing to slice sampler.



```
%slice sampler
rng default % For reproducibility
f = @(x) exp(-x.^2/2);
area = integral(f,-5,5);
N = 1000;
x = slicesample(1,N,'pdf',f,'thin',5,'burnin',1000);
figure
subplot(1,2,1)
[binheight,bincenter] = hist(x,50);
h = bar(bincenter,binheight,'hist');
h.FaceColor = [.8 .8 1];
% acceptance rejection
subplot(1,2,2)
f = @(x)x.*exp(-(x.^2)/2);
g = @(x)exp(-x);
grnd = @()rand(1);
X = accrejrnd(f,g,grnd,1.2,1000,1);
histogram(X)
```

```
function X = accrejrnd(f,g,grnd,c,m,n)
```

```
X = zeros(m,n); % Preallocate memory
for i = 1:m*n
    accept = false;
    while accept == false
```

```
        u = rand();
        v = grnd();
        if c*u <= f(v)/g(v)
            X(i) = v;
            accept = true;
        end
    end
end
```