FINAL

Lecture 21 Math 391 5/16/17

Recall Bayesian protocol:

① Pick $\mathcal{F}$, the likelihood model

② Pick $P(\theta)$, your prior

③ Collect data X

④ Obtain posterior $P(\theta | X)$ for inference via – direct closed form

– conjugation

– Grid sampling

– Gibbs sampling

What if 1 & 2 were wrong? We should have some means at our disposal to check.

First check (easy to pass). Recall $P(X) = \int_{\Theta} P(X|\theta) P(\theta) d\theta$

the elusive denominator seldom computed.

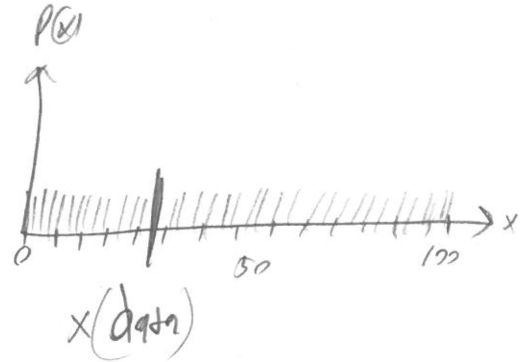
We called this the "prior predictive distr.". Why? It shows you what data looks like coming from your model $\mathcal{F}$ subject to parameters from your prior idea

e.g. $P(X|\theta) = \text{Binom}(\overset{n}{100}, \theta)$

$P(\theta) = U(0,1) = \text{Beta}(1,1)$

$\Rightarrow P(X) = \text{BetaBin}(100, 1, 1)$

x = 19



Does the data plausibly come from $p(x)$? Yes... so far so good.

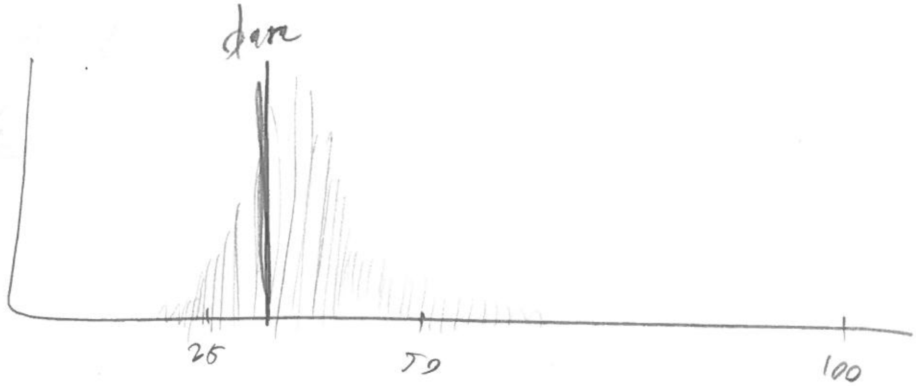
<u>Second Check</u> (Harder to pass) Recall $p(x^*|x) = \int p(x|\theta)\, p(\theta|x)\, d\theta$

the post. predictive distr.

AKA the post. replicative distr.

where x^* is "replicated data that could be observed tomorrow if the experiment that produced x today was repeated tomorrow. In which case..

$$p(x^*|x) = \text{BetaBin}(100, 30, 62)$$



Does the data look like other replicas of the data? If so... we're probably okay...

We are only assessing model plausibility, not model truth in an absolute sense...

~~Lec 21 Math 341 5/16/17~~

Systematic sweep Gibbs Sampler

to estimate $P(\theta_1, \ldots, \theta_p \mid x)$, guess $\vec{\theta}$ and

sample from conditionals $P(\theta_j \mid \theta_{-j})$ while iteratively updating

We will now build another Gibbs sampler ..

$$X_1,\dots,X_n \mid \theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho \overset{iid}{\sim} \rho N(\theta_1,\sigma_1^2) + (1-\rho) N(\theta_2,\sigma_2^2)$$

But we can use data augmentation again $I_1,\dots,I_n$.

$$P(I_1,\dots,I_n,\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho \mid X_1,\dots,X_n) \propto P(X_1,\dots,X_n \mid I_1,\dots,I_n,\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho)$$

$$\cdot \underbrace{P(I_1,\dots,I_n,\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho)}$$

$$\underbrace{P(I_1,\dots,I_n \mid \theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho)}_{\prod_{i=1}^n \rho^{I_i}(1-\rho)^{1-I_i}} \underbrace{P(\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho)}_{\propto 1 \propto 1 \propto \frac{1}{\sigma_1^2} \propto \frac{1}{\sigma_2^2} \propto 1}$$

$$= \left(\prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(X_i-\theta_1)^2}\right)^{I_i} \left(\frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(X_i-\theta_2)^2}\right)^{1-I_i} \rho^{I_i}(1-\rho)^{1-I_i}\right) \frac{1}{\sigma_1^2}\frac{1}{\sigma_2^2}$$

$$= \left(\rho\frac{1}{\sqrt{2\pi\sigma_1^2}}\right)^{\Sigma I_i} e^{-\frac{1}{2\sigma_1^2}\Sigma I_i(X_i-\theta_1)^2} \left((1-\rho)\frac{1}{\sqrt{2\pi\sigma_2^2}}\right)^{n-\Sigma I_i} e^{-\frac{1}{2\sigma_2^2}\Sigma(1-I_i)(X_i-\theta_2)^2} \frac{1}{\sigma_1^2}\frac{1}{\sigma_2^2}$$

$$P(\theta_1 \mid \cdots\cdots) \propto e^{-\frac{1}{2\sigma_1^2}\Sigma I_i(X_i^2-2X_i\theta_1+\theta_1^2)} \propto e^{-\frac{1}{2\sigma_1^2}(-2\theta_1\Sigma I_iX_i+\theta_1^2\Sigma I_i)} = e^{\frac{\Sigma I_iX_i}{\sigma_1^2}\theta_1 - \frac{\Sigma I_i}{\sigma_1^2}\theta_1^2}$$

$$\propto N\left(\frac{\Sigma I_iX_i}{\Sigma I_i}, \frac{\sigma_1^2}{\Sigma I_i}\right)$$

$$P(\theta_2 \mid \cdots) \propto N\left(\frac{\Sigma(1-I_i)X_i}{\Sigma 1-I_i}, \frac{\sigma_2^2}{\Sigma 1-I_i}\right)$$

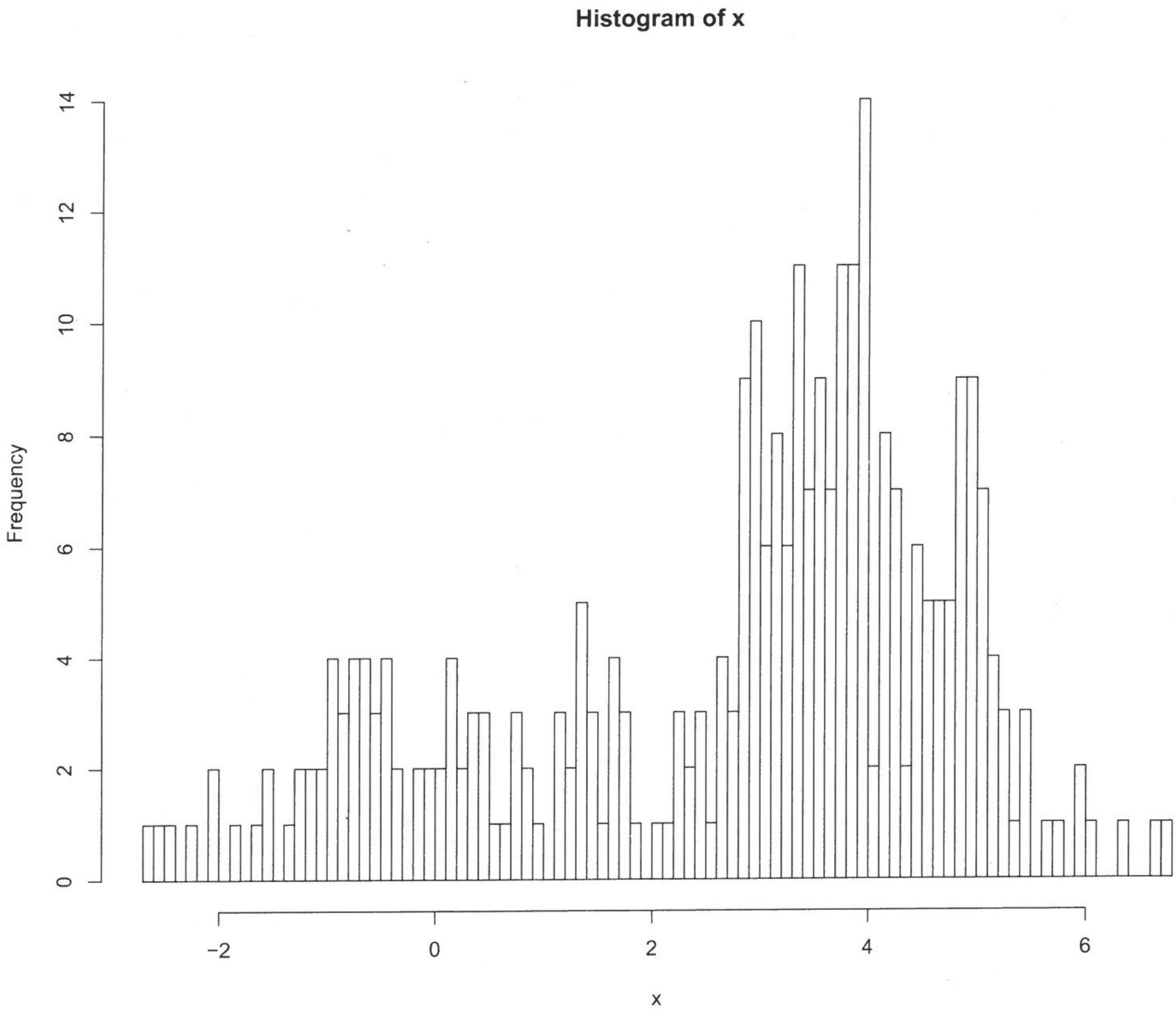
$$P(\sigma_1^2 \mid \cdots) \propto (\sigma_1^2)^{-\frac{\Sigma I_i}{2}-1} e^{-\frac{\Sigma I_i(X_i-\theta_1)^2/2}{\sigma_1^2}} \propto \text{InvGamma}\left(\frac{\Sigma I_i}{2}, \frac{\Sigma I_i(X_i-\theta_1)^2}{2}\right)$$

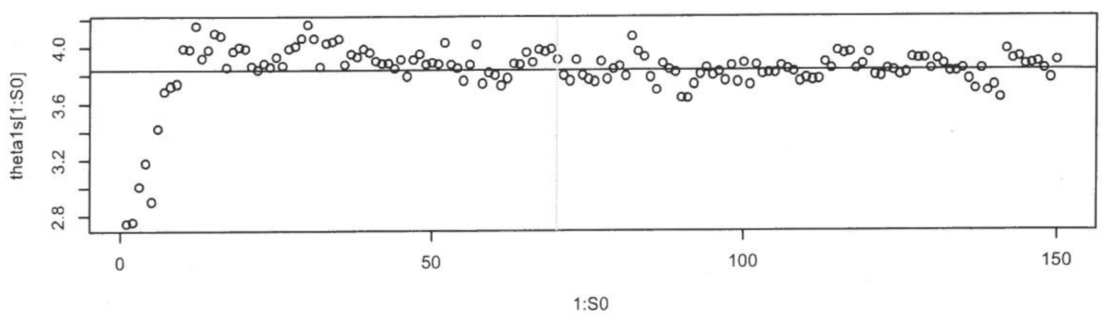
$$P(\sigma_2^2 \mid \cdots) \propto \text{InvG}\left(\frac{\Sigma 1-I_i}{2}, \frac{\Sigma(1-I_i)(X_i-\theta_2)^2}{2}\right)$$

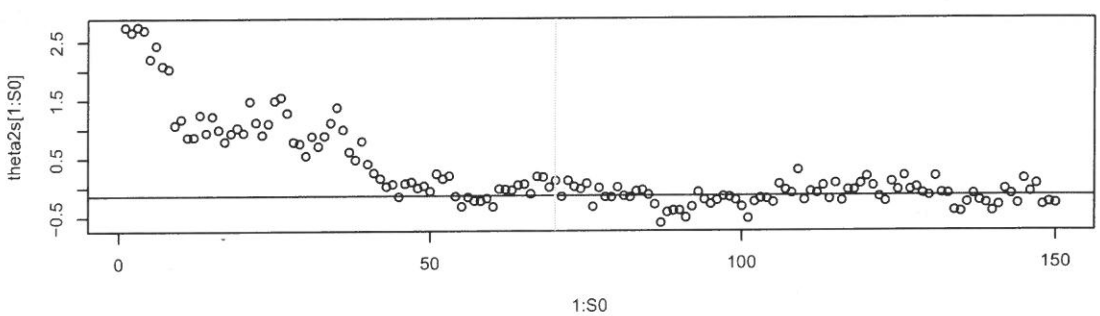
$$P(\rho \mid \cdots) \propto \rho^{\Sigma I_i}(1-\rho)^{\Sigma 1-I_i} \propto \text{Beta}(1+\Sigma I_i, 1+\Sigma 1-I_i)$$

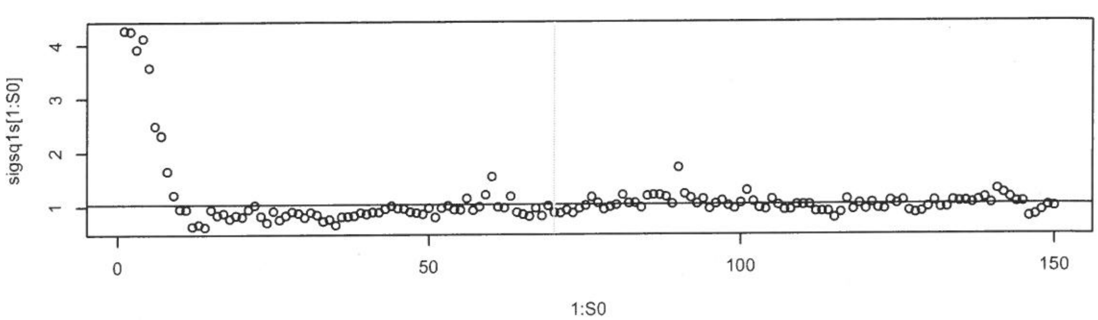
$$P(I_i \mid \cdots) \propto \Big(\underbrace{\rho\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2\sigma_1^2}(X_i-\theta_1)^2}}_{a}\Big)^{I_i} \Big(\underbrace{(1-\rho)\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{1}{2\sigma_2^2}(X_i-\theta_2)^2}}_{b}\Big)^{1-I_i} \propto \text{Bern}\left(\frac{a}{a+b}\right)$$

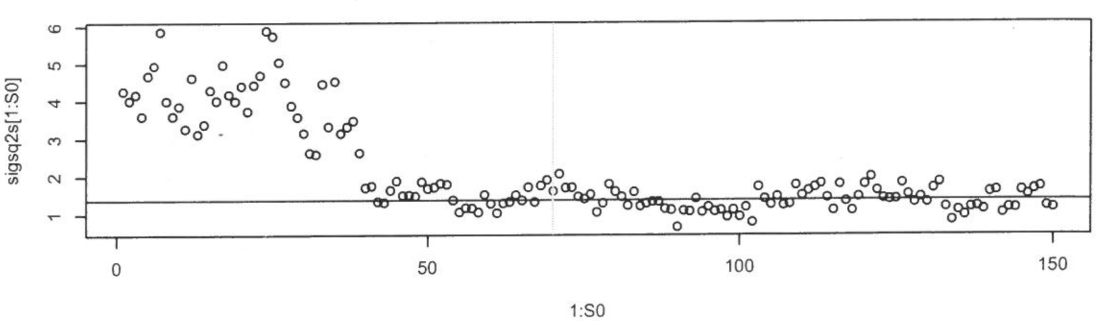
Mixture Model Gibbs Sampler

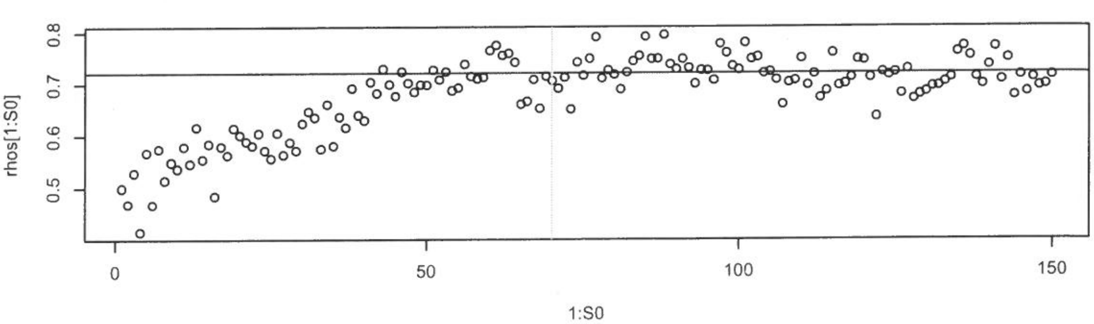


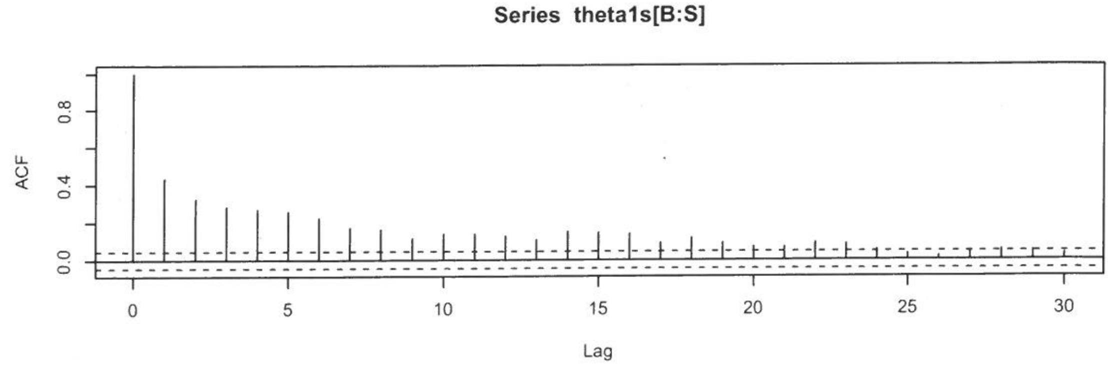

theta1s[1:S0]
2.8
3.2
3.6
4.0
0
50
100
150
1:S0

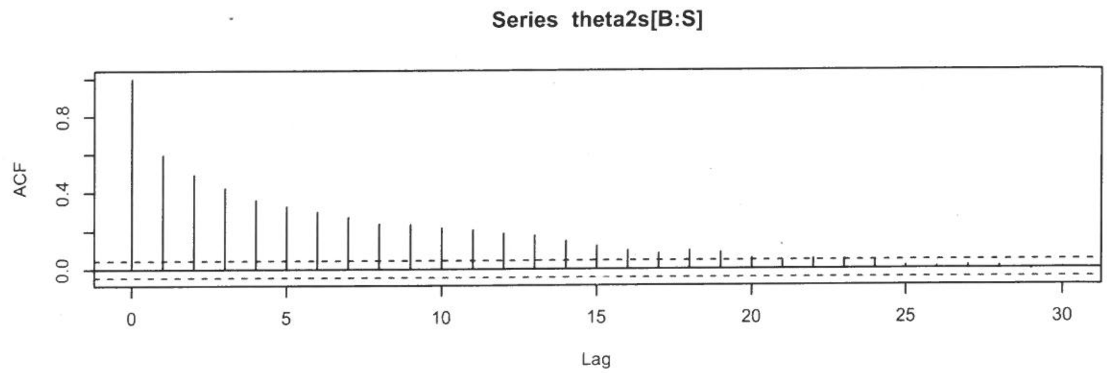

theta2s[1:S0]
-0.5
0.5
1.5
2.5
0
50
100
150
1:S0

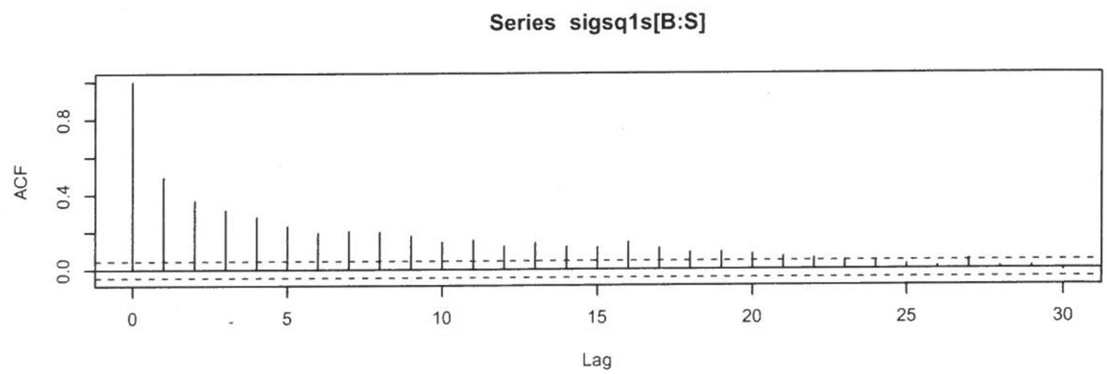

sigsq1s[1:S0]
1
2
3
4
0
50
100
150
1:S0

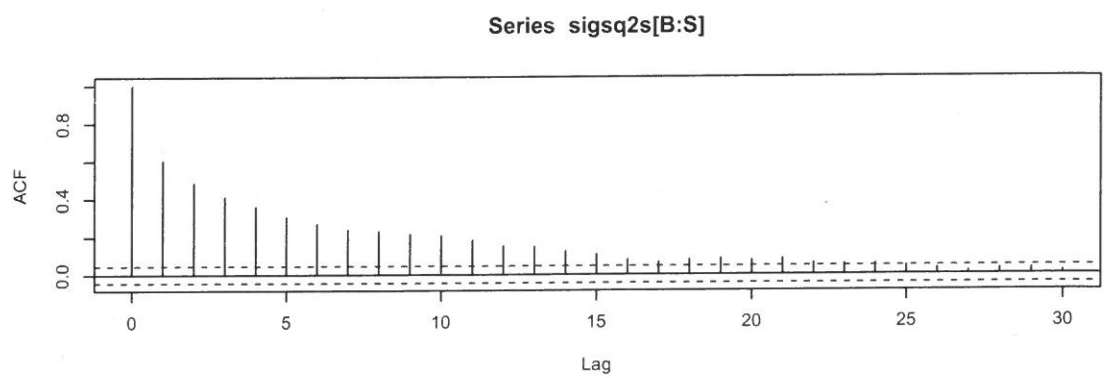

sigsq2s[1:S0]
1
2
3
4
5
6
0
50
100
150
1:S0

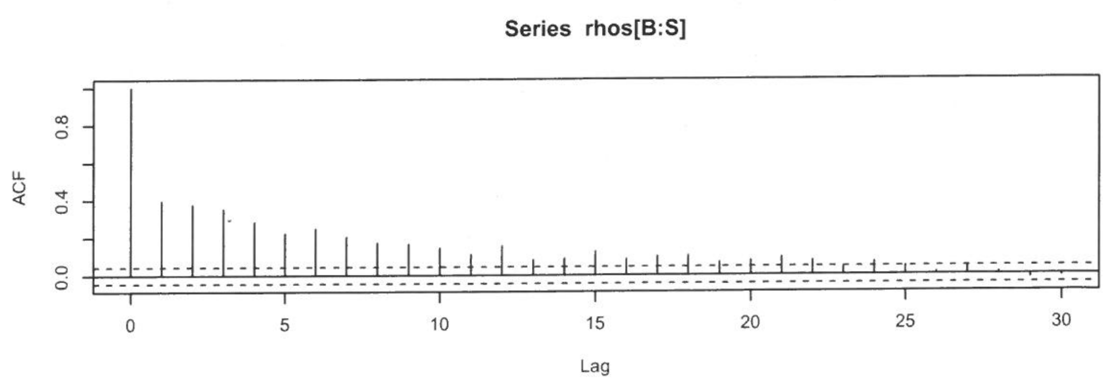

rhos[1:S0]
0.5
0.6
0.7
0.8
0
50
100
150
1:S0

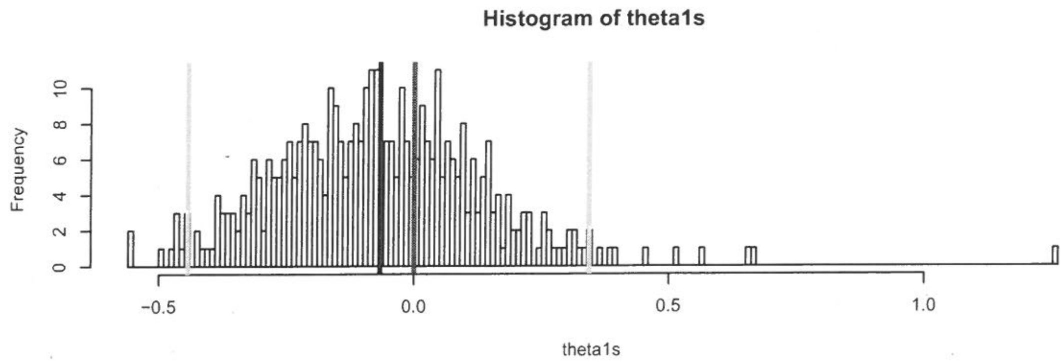

Series theta1s[B:S]
ACF
0.0
0.4
0.8
0
5
10
15
20
25
30
Lag

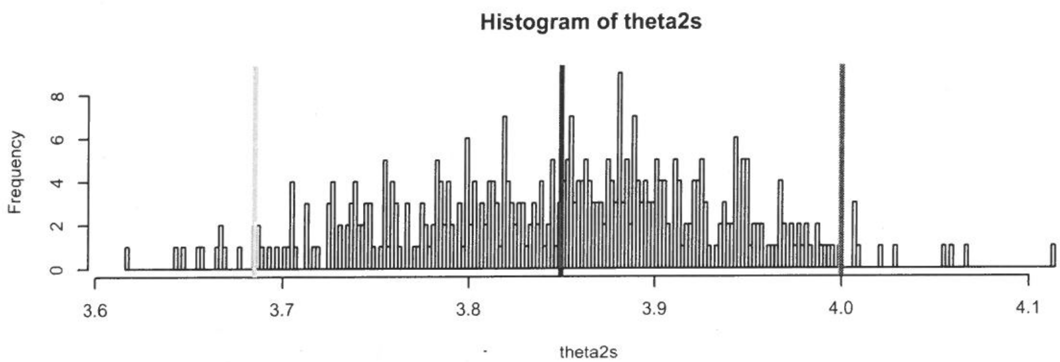

Series theta2s[B:S]
ACF
0.0
0.4
0.8
0
5
10
15
20
25
30
Lag

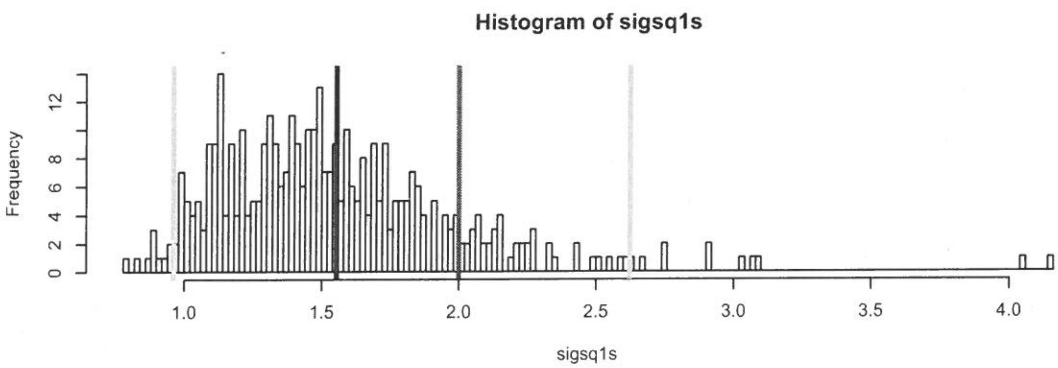

Series sigsq1s[B:S]
ACF
0.0
0.4
0.8
0
5
10
15
20
25
30
Lag

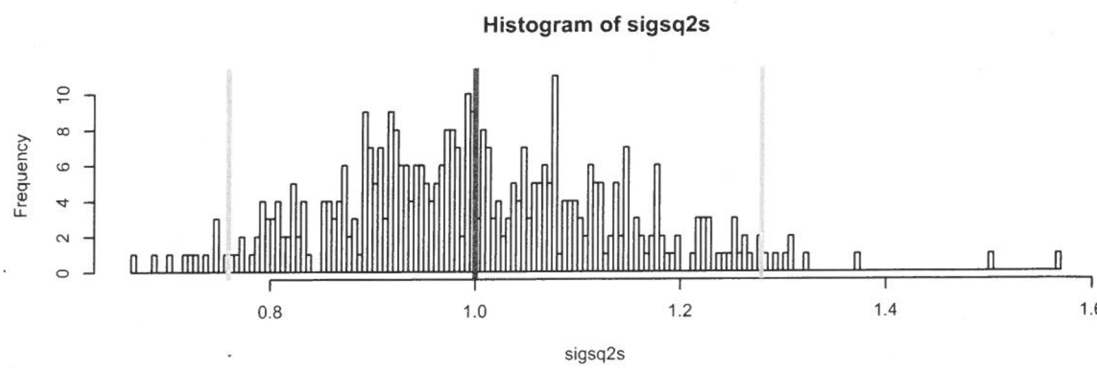

Series sigsq2s[B:S]
ACF
0.0
0.4
0.8
0
5
10
15
20
25
30
Lag

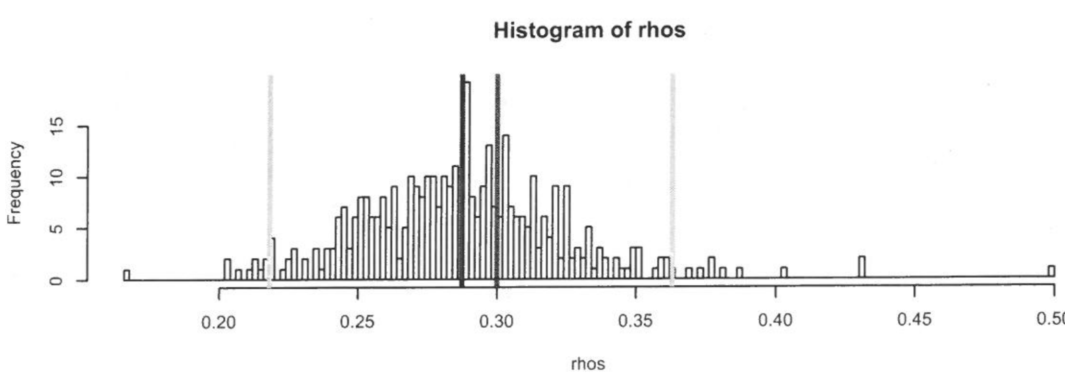

Series rhos[B:S]
ACF
0.0
0.4
0.8
0
5
10
15
20
25
30
Lag


Histogram of theta1s
Frequency
0
2
4
6
8
10
-0.5
0.0
0.5
1.0
theta1s


Histogram of theta2s
Frequency
0
2
4
6
8
3.6
3.7
3.8
3.9
4.0
4.1
theta2s


Histogram of sigsq1s
Frequency
0
2
4
6
8
12
1.0
1.5
2.0
2.5
3.0
3.5
4.0
sigsq1s


Histogram of sigsq2s
Frequency
0
2
4
6
8
10
0.8
1.0
1.2
1.4
1.6
sigsq2s


Histogram of rhos
Frequency
0
5
10
15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
rhos

$$P(Q) \propto \overbrace{0.1^{Q}\, 0.2^{1-Q}}^{=k(Q)} \qquad \text{Supp}(Q) = \{0,1\}$$

$$= ?$$

$$P(Q=1) = \frac{k(1)}{k(0)+k(1)} = \frac{0.1}{0.1+0.2}$$

$$P(I_2 \mid \cdots) = \text{same}$$

$$\vdots$$

$$P(I_n \mid \cdots) = \text{same}$$

COURSE OVER

Recall change pt model...

$\overset{iid}{\sim} \text{Poisson}(\lambda_1)$ $\qquad$ $\overset{iid}{\sim} \text{Poisson}(\lambda_2)$

What if instead...

e.g.
the drowses...

$$P(X_t) = \text{Poisson}(\overset{\lambda}{\overset{\|}{a + bt}})$$

and dependent

i.e. λ is a function of time,

a linear function of time.

this is called a "Poisson Regression"

t	X
1	3
1.3	0
2	5
2.7	7
3.1	6
4.9	5
5.0	5
7.5	9

the likelihood

$$P(X \mid a,b) = \prod_{i=1}^{n} \frac{(a+bt_i)^{x_i} e^{-(a+bt_i)}}{x_i!} = \frac{\left(\prod (a+bt_i)^{x_i}\right) e^{-\sum a+bt_i}}{\prod x_i!}$$

You care about λ i.e. you care about finding a, b; $\propto e^{-\sum a+bt_i} \prod (a+bt_i)^{x_i}$

$$P(a,b \mid X) \propto P(X \mid a,b)\, P(a,b) = P(X \mid a,b)$$

prior? $P(a) \propto 1$, $P(b) \propto 1$

$a, b \in \mathbb{R}$

$$\Rightarrow P(a,b \mid x) \propto e^{-\sum a+bt_i} \prod (a+bt_i)^{x_i}$$ non-conj.

Gibbs sampling to the rescue?

$P(a \mid X, b) \propto$ ditto

$P(b \mid X, a) \propto$ ditto

In order to use Gibbs sampling, you would have to grid sample $k(a \mid X, b)$ and $k(b \mid X, a)$ ... TOO SLOW!!!

Need something else ... Metropolis - Hastings Algorithm

E. Teller!

Metropolis et al 1953,
Hastings, 1970

What does Gibbs sampling do? It moves around the space...
Why not move around another way?

Steps in M-H

① Initialize a_0, b_0, the two params

② Step 2:

Draw a_1 from $q(a_0, \overset{\text{other hyper params}}{\phi})$ e.g. $N(a_0, 1^2)$

but a_1 may not have been a good draw since q is divorced from $p(a \mid X, b)$

③ Step 3. Calculate

$$r := \frac{\dfrac{p(a = a_1, b = b_0 \mid X)}{q(a_1 ; a_0, \phi)}}{\dfrac{p(a = a_0, b = b_0 \mid X)}{q(a_0 ; a_0, \phi)}}$$

posterior with a_1

transition prob at a_1

} the ratio of the prob to transition w/ proposal

} ratio of ~~this~~ the prob. to transition prob without proposal

④ Accept a_1 proposal w.p. r (if $r \geq 1$, accept automatically).

⑤ Repeat Steps 2-~~4~~ for b,

⑥ Repeat steps 2-5 many times

⑦ Burn, thin

Metropolis-within-Gibbs: do Gibbs sampler but if one conditional is not available to be sampled from, do a Metropolis step.

Note: Gibbs Sampler is a special case of M-H.

(2) Draw from

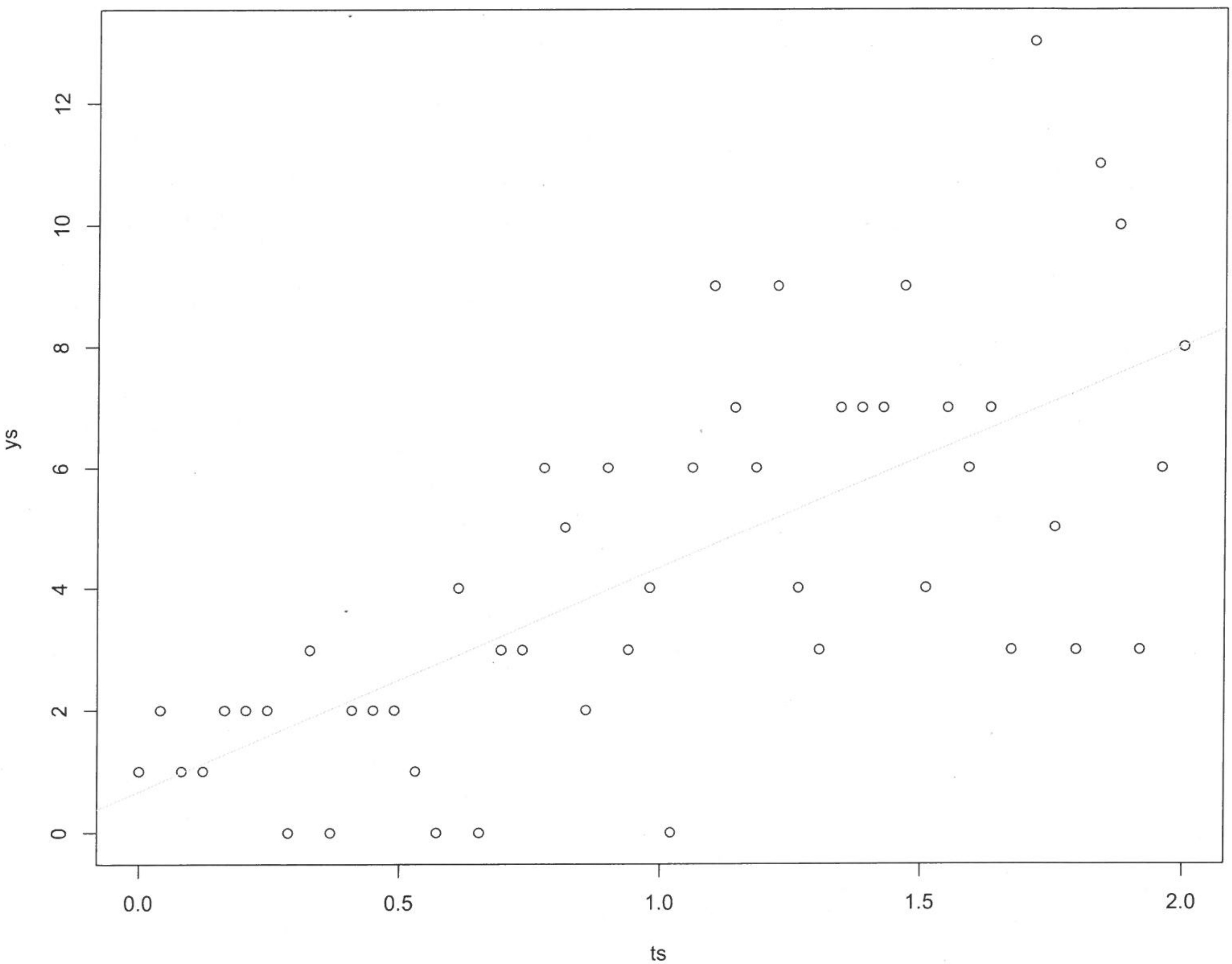
$$q(\theta_j, \delta) = P(\theta_j \mid \theta_{-j})$$

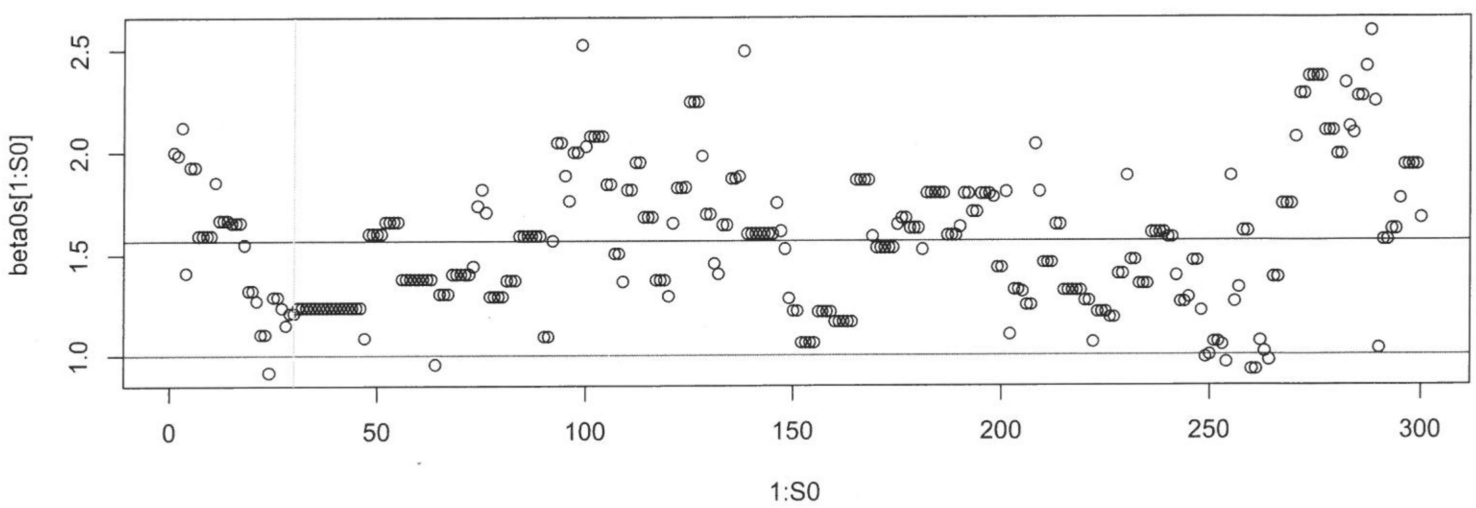
i.e. the conditional

(3)

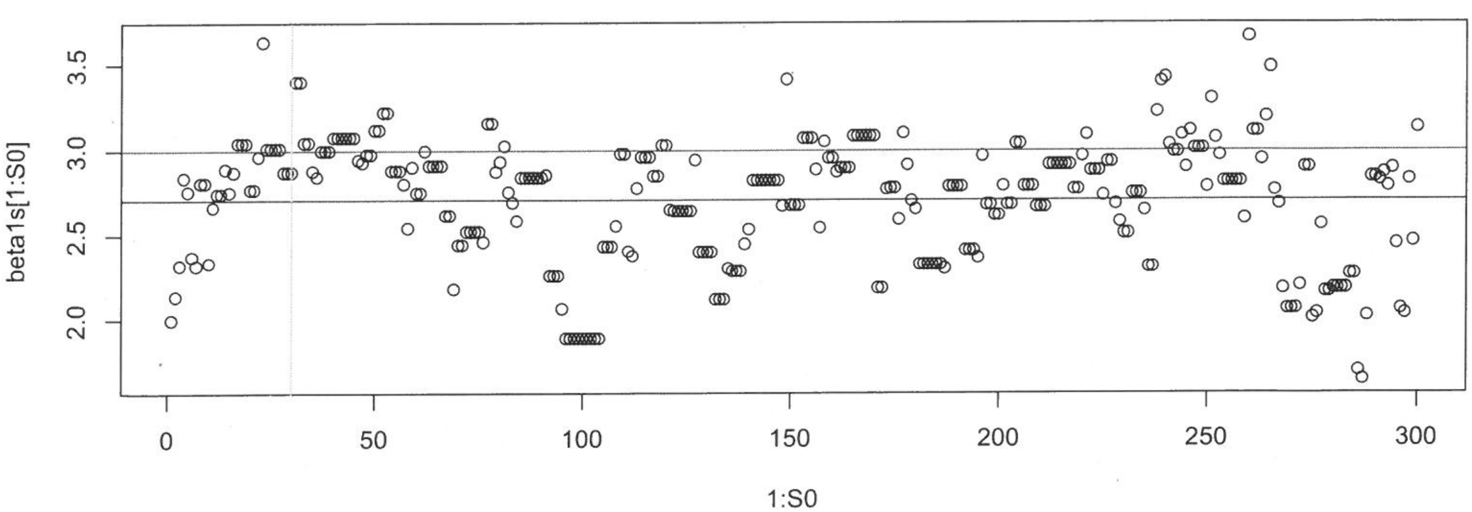
Compute $r := \dfrac{\dfrac{P(\theta_j = \theta_{j,t+1} \mid \theta_{-j})}{P(\theta_j = \theta_{j,t+1} \mid \theta_{-j})}}{\dfrac{P(\theta_j = \theta_{j,t} \mid \theta_{-j})}{P(\theta_j = \theta_{j,t} \mid \theta_{-j})}}$ $\Rightarrow$ accept w.p. 1.

Poisson Regression
Metropolis - Hastings Sampler

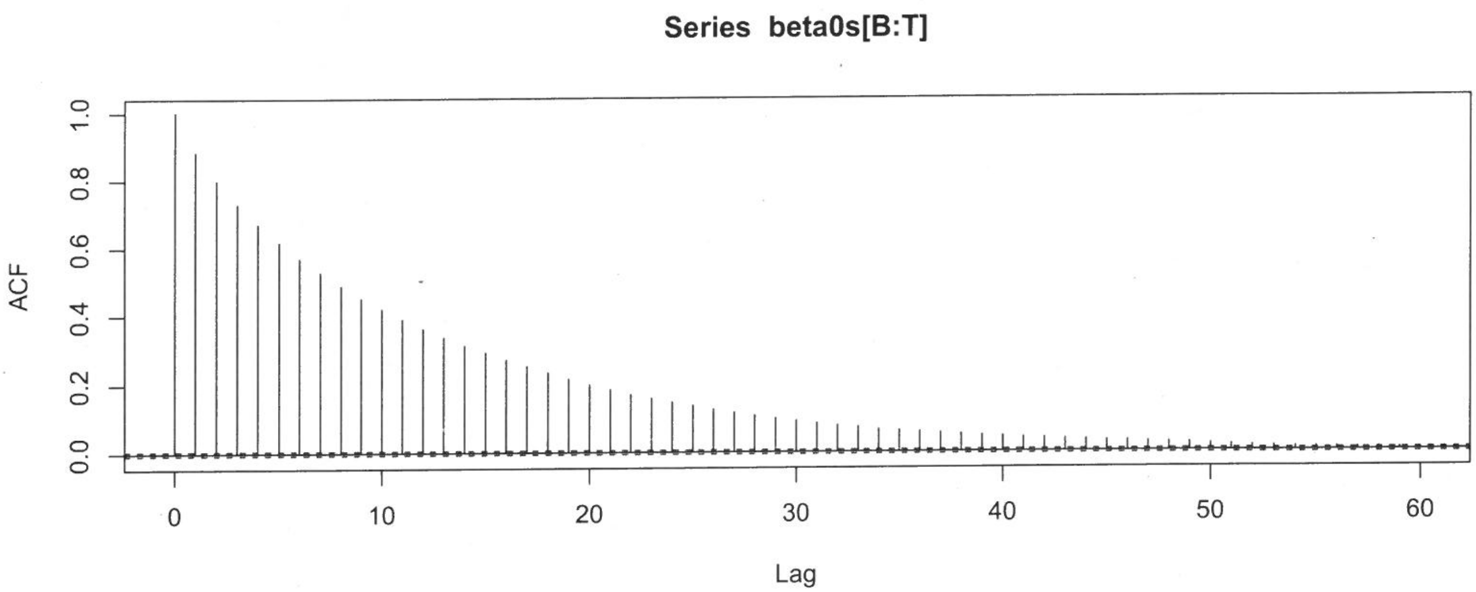


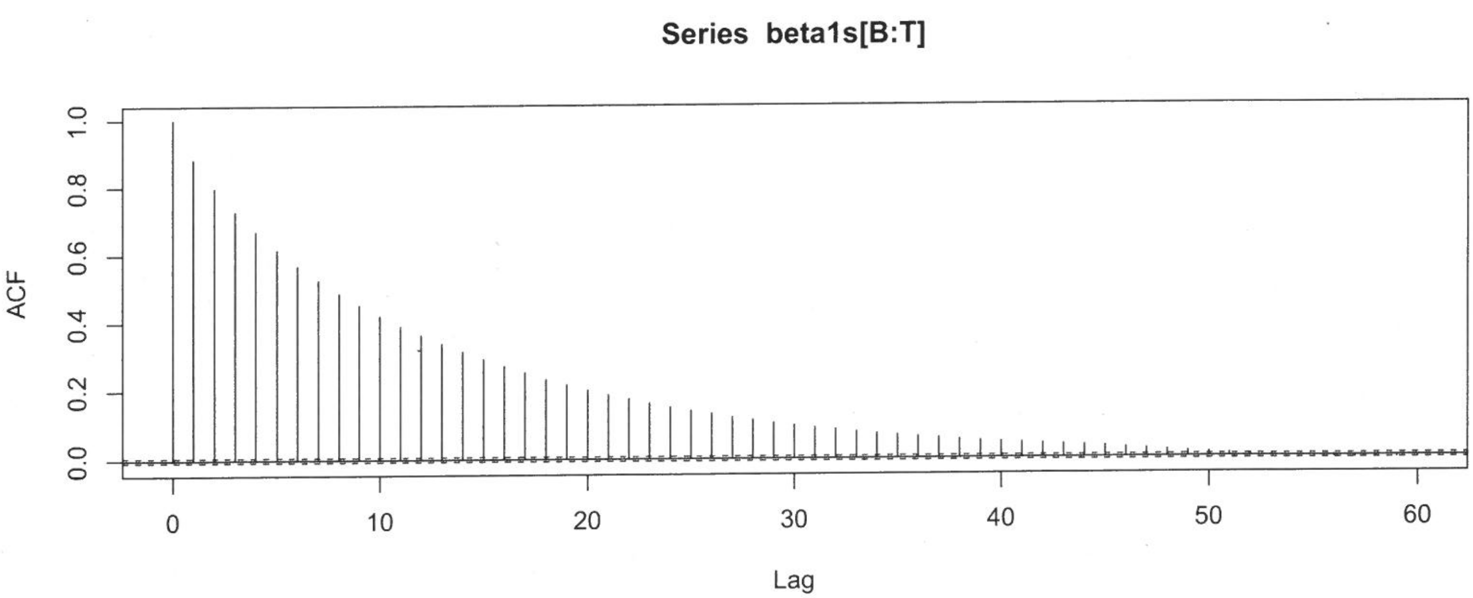


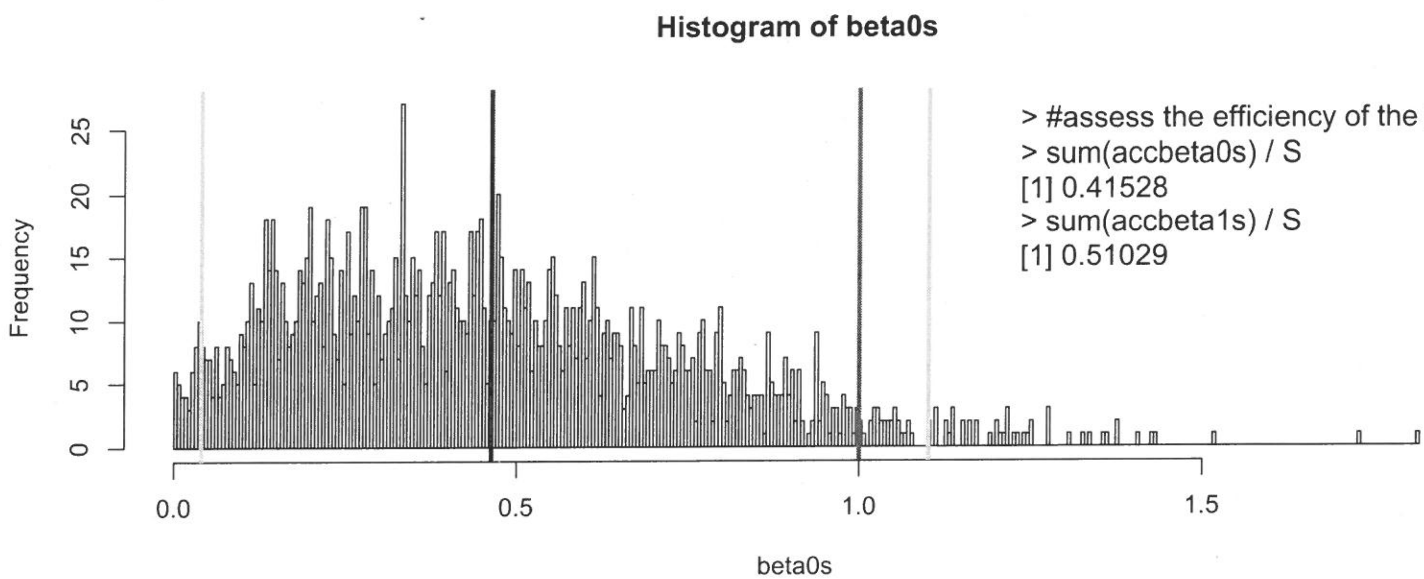
beta0s[1:S0]
2.5
2.0
1.5
1.0
0
50
100
150
200
250
300
1:S0



beta1s[1:S0]
3.5
3.0
2.5
2.0
0
50
100
150
200
250
300
1:S0


Series beta0s[B:T]
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Lag


Series beta1s[B:T]
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Lag




```
> #assess the efficiency of the MH
> sum(accbeta0s) / S
[1] 0.41528
> sum(accbeta1s) / S
[1] 0.51029
```