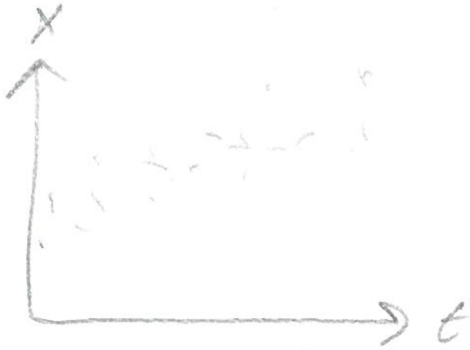
Lec 23 Math 381 5/14/18



Model $X_i \sim \text{Poisson}(a+bt_i)$

Data comes in pairs $\langle t_i, X_i \rangle$

Two parameters : a, b

Priors $P(a) \propto 1$, $P(b) \propto 1$

$P(a,b \mid X_1,\dots,X_n, t_1,\dots,t_n) \propto e^{-\sum a+bt_i} \prod_{i=1}^{n} (a+bt_i)^{X_i}$ not conj

$P(a \mid \text{——}) \propto$ " not a known distr

$P(b \mid \text{——}) \propto$ " "

What to do?

Need a means to sample a new a given —— (and the b given ——) to approx. the posterior $P(a,b \mid \text{——})$ and marginal posteriors $P(a \mid \text{——})$, $P(b \mid \text{——})$. We use the Gibbs sampling setup but for the sampling step we use a Metropolis-Hastings (M-H) Algorithm. AKA "M-H Rejection sampling".

SubSteps inside a single M-H step for sampling a new a from $p(a|—)$

Given a_{t-1}, b_{t-1} from the Gibbs sampler...

Step 1:

transition distr. — other hyper params

Draw a_t from $\theta \sim q(a_{t-1}, \phi)$ e.g. $N(a_{t-1}, 1^2)$

But a_t may not have been a good draw since q may be different from $p(a|X, b)$

Step 2. Calculate

"M-H ratio"

$$r := \frac{\dfrac{p(a = a_t, b = b_{t-1} | X)}{f_q(a_t; a_{t-1}, \phi)}}{\dfrac{p(a = a_{t-1}, b = b_{t-1} | X)}{f_q(a_{t-1}; a_t, \phi)}}$$

posterior with a_t

transition prob. at a_t

} the ratio of the prob to transition w/ proposal

} ratio of old prob. to transition prob w/o proposal

Now... so... only

Step 3: Accept a_t 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.

↳ proven to converge... proof way beyond scope of course

Note: if $f_q(q_t \mid q_{t-1}, \phi) = f_q(q_{t-1} \mid q_t, \phi)$ then r simplifies and it's called the "Metropolis alg"

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

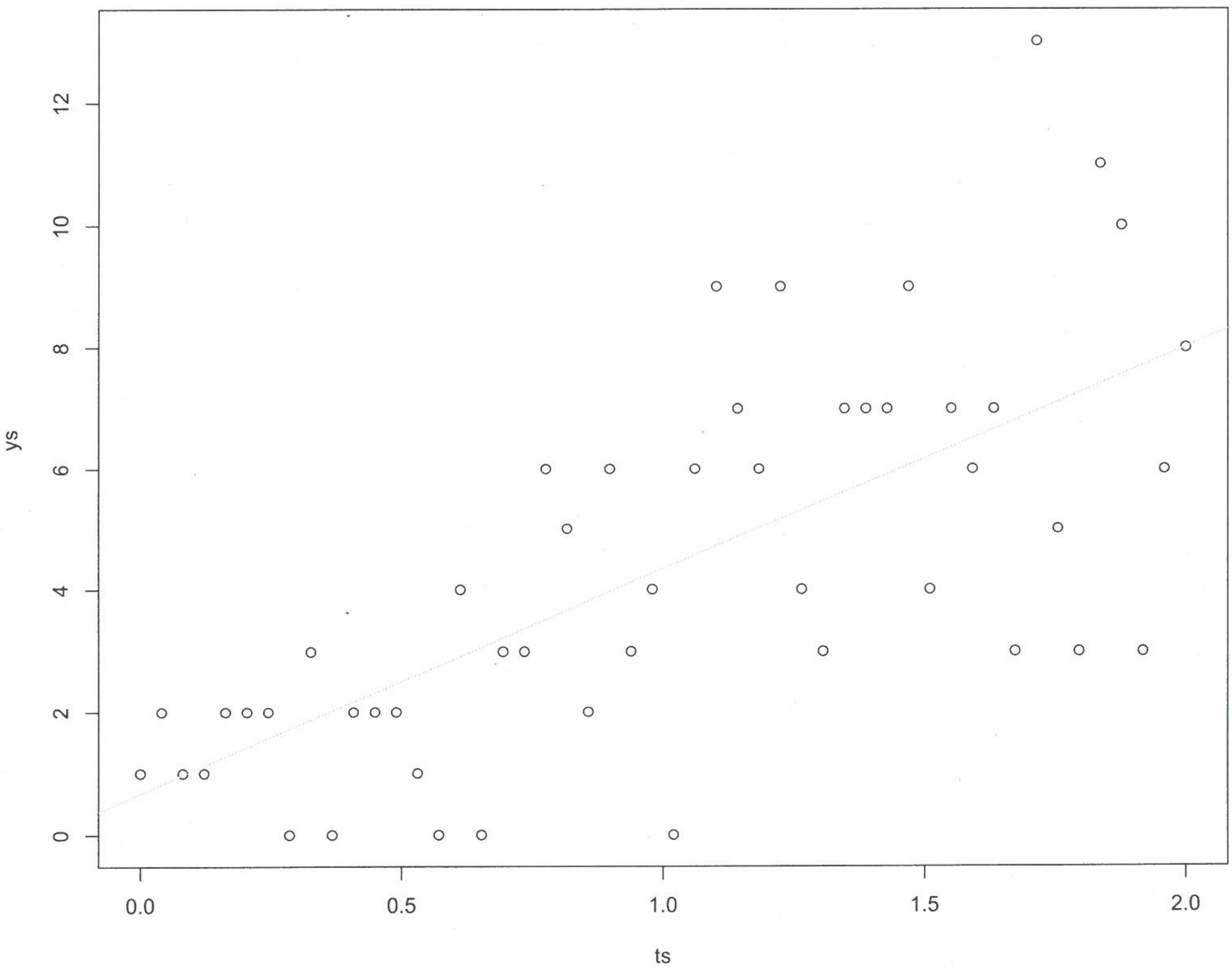
Using $q = N(\theta_{t-1}, 1^2)$, we can now have inference for the linear in time Poisson model!

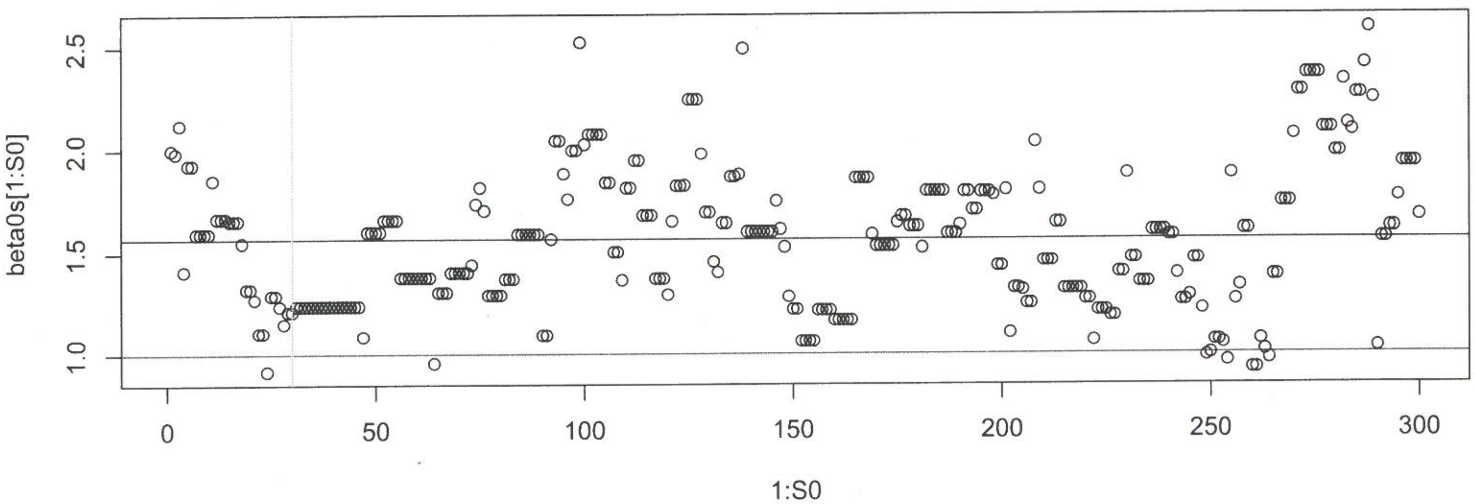
Note: $f_q(q_t \mid q_{t-1}, \phi) = f_q(q_{t-1} \mid q_t, \phi)$ for the normal distr.

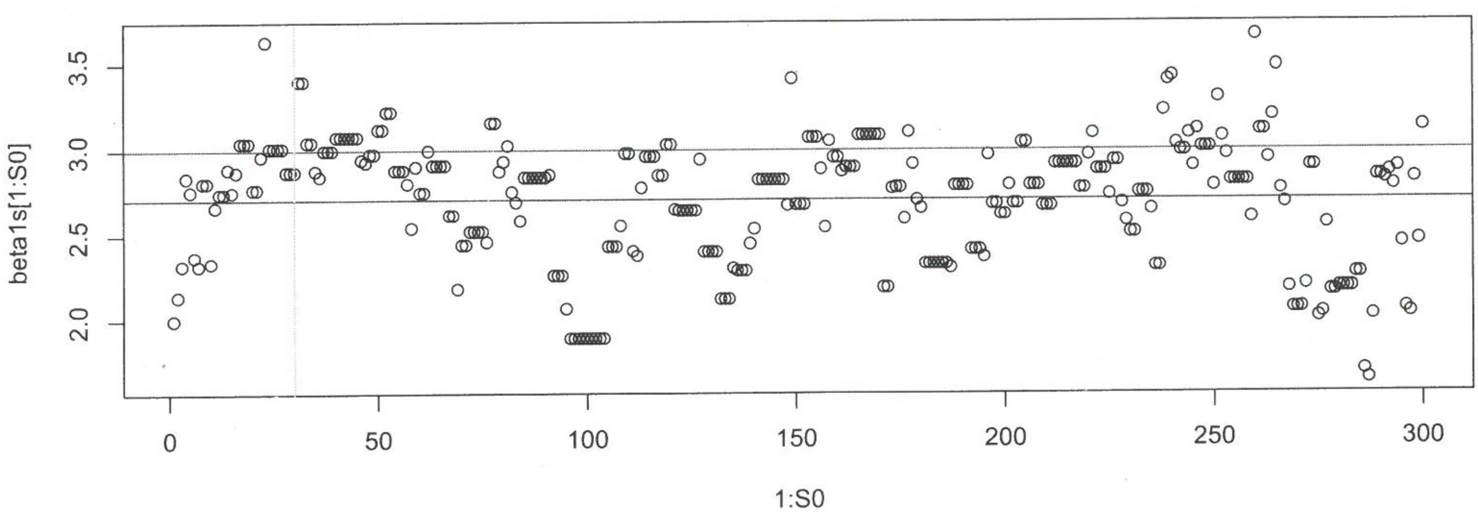
So it's Metropolis

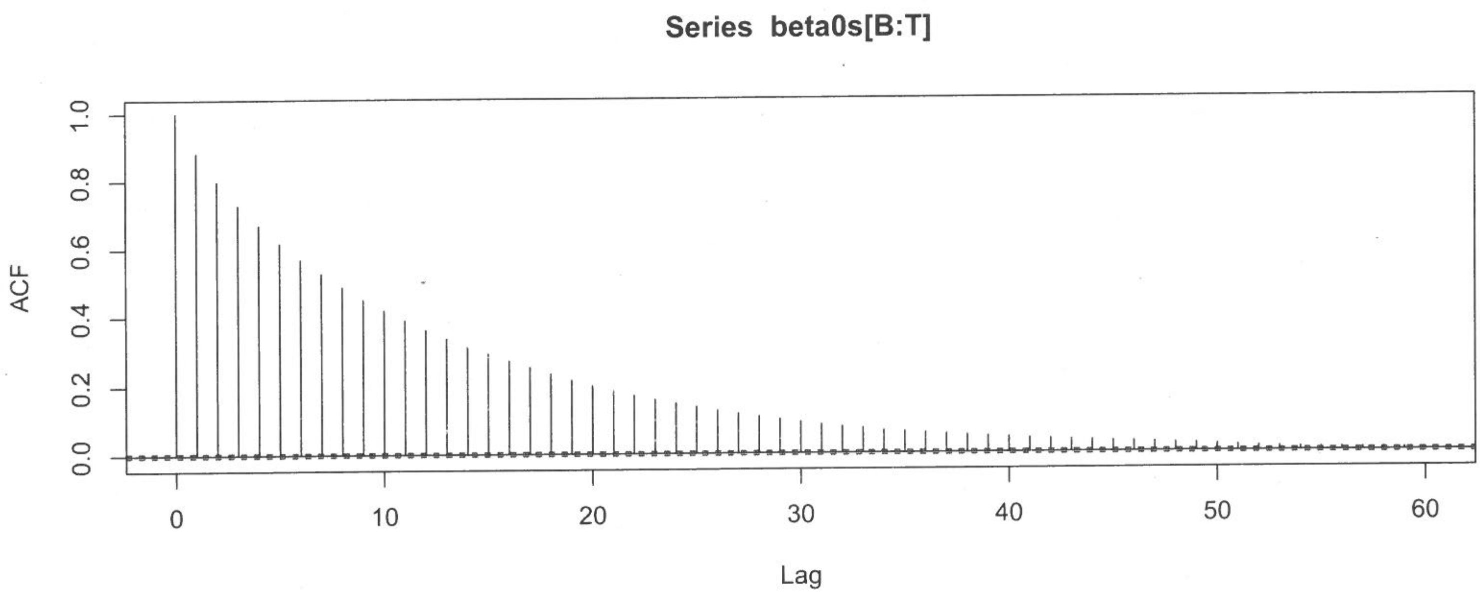
Poisson Regression

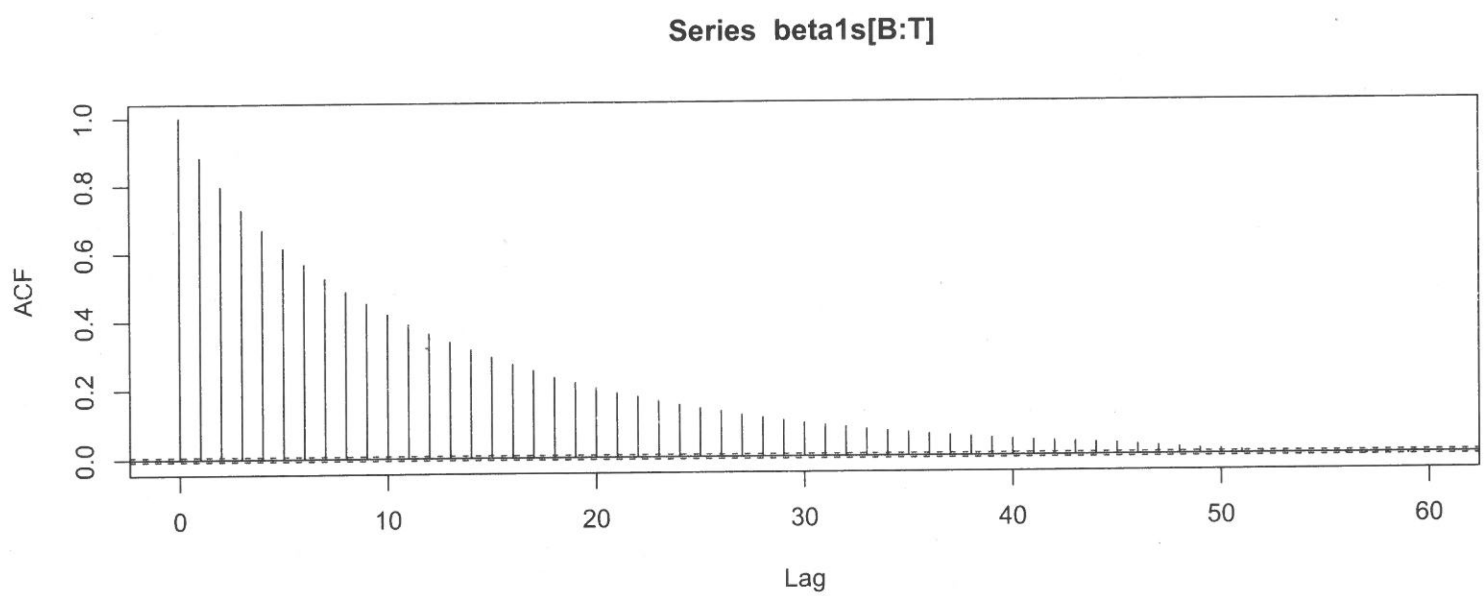
Metropolis - Hastings Sampler

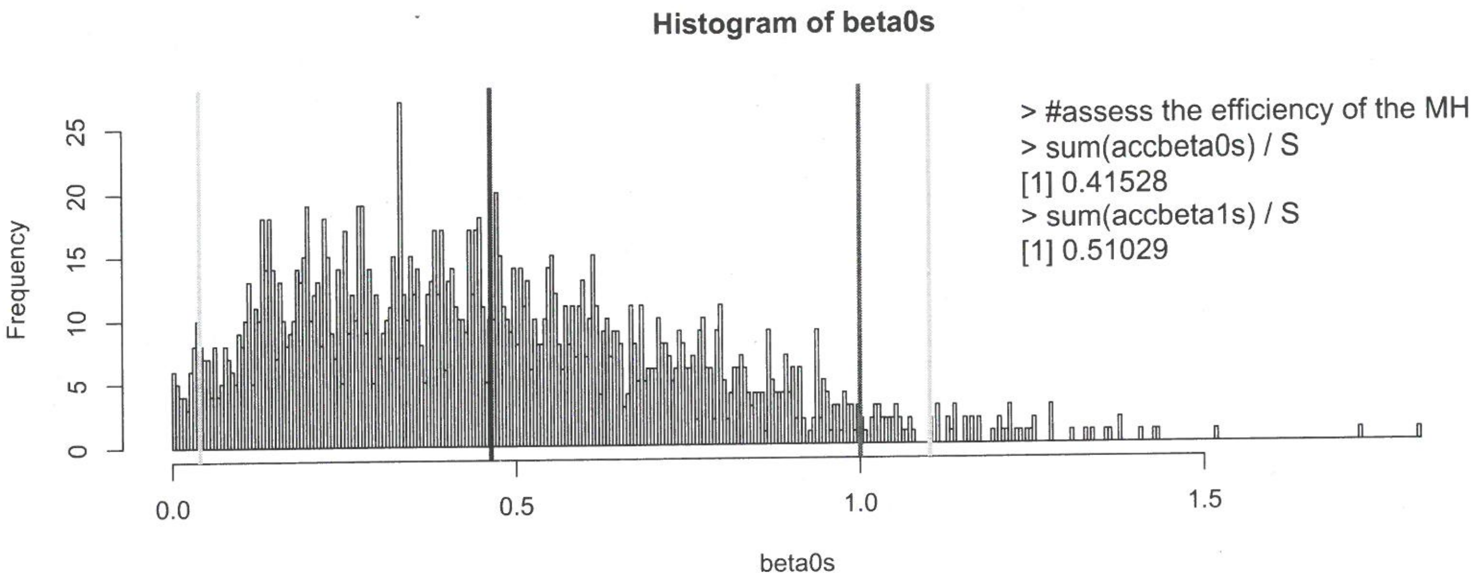


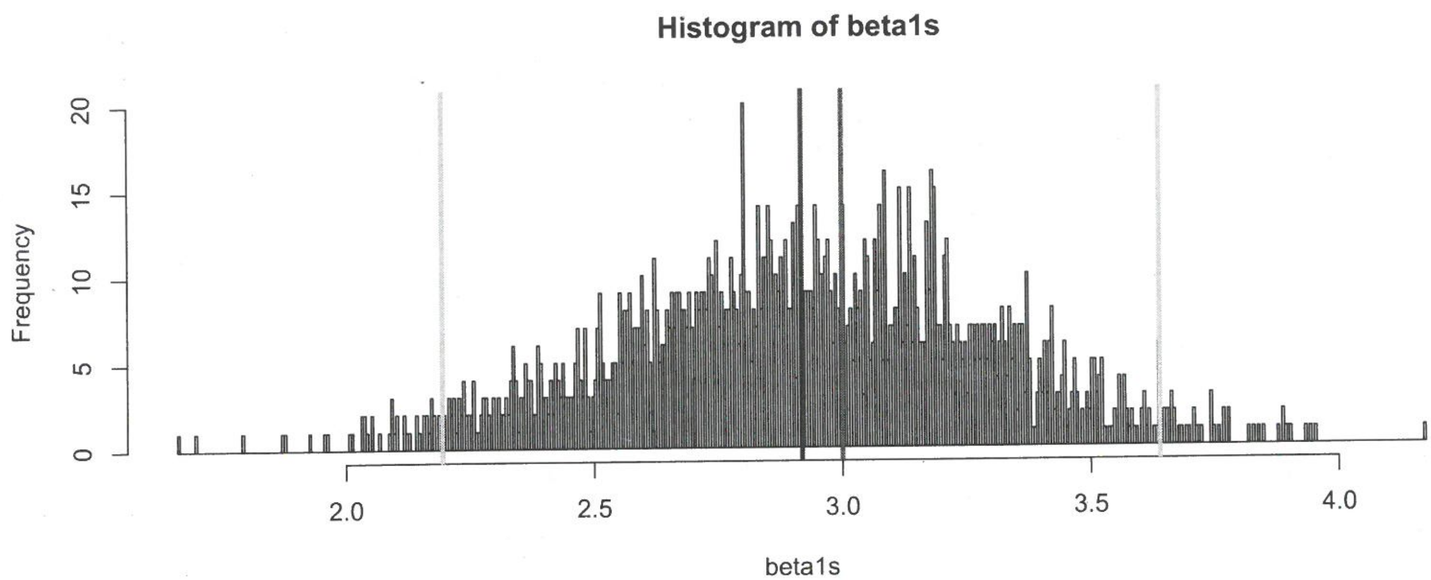

beta0s[1:S0]
1:S0
1.0
1.5
2.0
2.5
0
50
100
150
200
250
300


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


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


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


Histogram of beta0s
> #assess the efficiency of the MH
> sum(accbeta0s) / S
[1] 0.41528
> sum(accbeta1s) / S
[1] 0.51029
Frequency
0
5
10
15
20
25
0.0
0.5
1.0
1.5
beta0s


Histogram of beta1s
Frequency
0
5
10
15
20
2.0
2.5
3.0
3.5
4.0
beta1s

Switching to a new unit of model "validation"...

Recall Bayesian protocol:

1. Pick $\mathcal{F}$, the likelihood model
2. Pick $P(\theta)$, your prior
3. Collect data X
4. Obtain posterior $P(\theta \mid X)$ for inference via — direct closed form
   - conjugacy
   - 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 $P(X) = \int_{\Theta} P(X \mid \theta) P(\theta)\, d\theta$

the elusive denominator seldom computed.

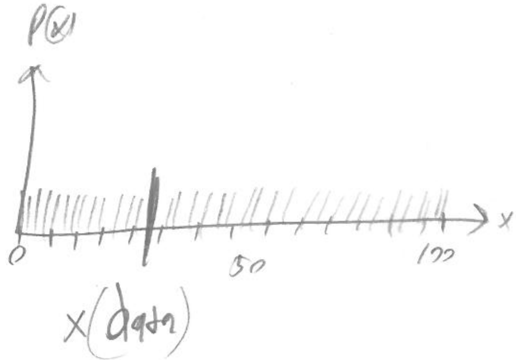
We called this the "prior predictive distr.". Why? It shows you what data looks like coming from your model $\mathcal{F}$ subject to parameters from your prior idea

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

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

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

$x = 19$



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

Second Check (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.

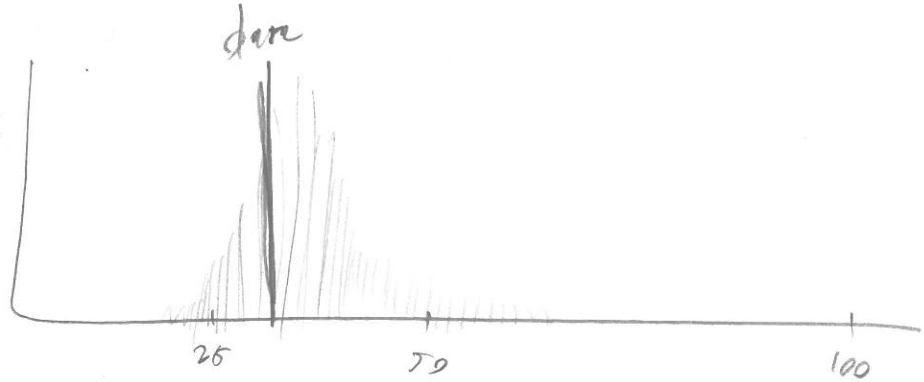
where x^* is "replicated data

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...