Lecture 6

- Monte Carlo estimation
- Stan coding
- Bayesian linear regression

Recall expectation:

If X is continuous with density $f(x)$

$$E[h(X)] = \int h(x)f(x)dx$$

or

If X is discrete with probability function $P(x)$

$$E[h(X)] = \sum h(x)P(x)$$

Monte Carlo method

Law of large numbers:

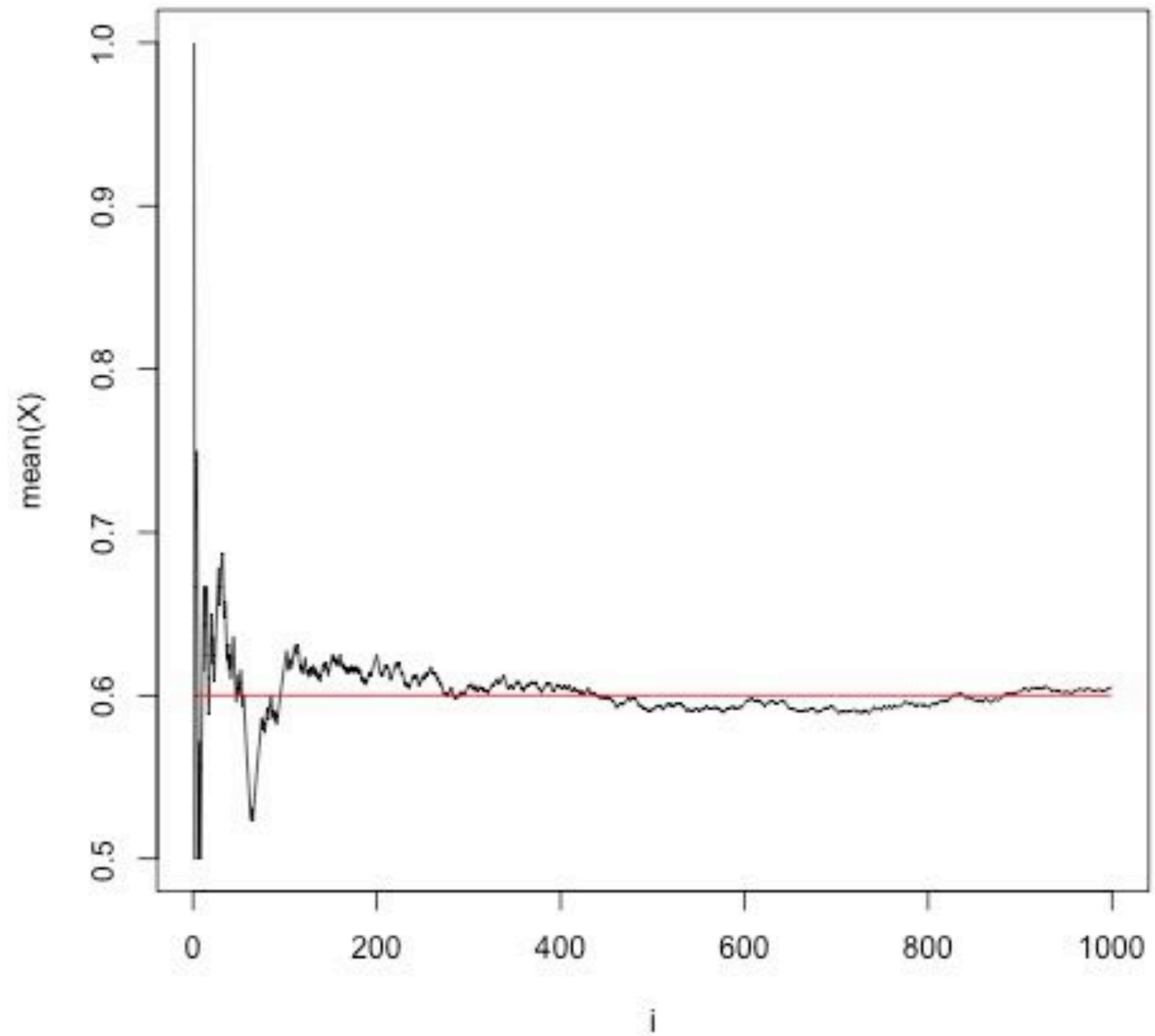
If $X_1, X_2, \ldots, X_N$ are independent random variables with density f, then as N tends to infinity

$$\frac{1}{N}\sum_{i=1}^{N} h(X_i) \rightarrow E[h(X)]$$

This means that if draw enough random samples we can compute any expectation.

Biased coin example

$$X_i \sim Bin(1, \theta = 0.6)$$



How well does the Monte Carlo sampler work?

Using the Central limit theorem one can show that:

$$V\left[E[h(X)] - \frac{1}{N}\sum_{i=1}^{N} h(X_i)\right] = \frac{C}{N}$$

For some C >0,
so the variance of the error decreases as one over N.

Stan

Stan can be used to generate samples from complicated posterior distributions.

Stan uses Markov Chain Monte Carlo methods to simulate from the posterior distribution.

There exits a Interface that we will spend a big part of todays lecture talking about.

In R we build the specific model using a string

The stan model has many separate parts,
we will start looking at the example from the book
at page 407.

The goal is to simulate the posterior of the probability of getting tails after a series of coin flips, given a Beta prior.

The first thing one need to build is the data.

```
model.string =
“
  data {
  int<lower = 0> N;
   int y[N];
  }
“
```

int = integer $\{\ldots, 2, -1, 0, 1, 2, \ldots\}$

int<lower = 0> is any integer greater or equal to zero.

float - any number

float<lower=2> is any number greater or equal to two.

int y[N] - a vector of length N containing integers

The second thing is to define the parameters, i.e the r.v. :

```
"
parameters {
    real<lower = 0, upper = 1> theta;
  }
"
```

real<lower = 0, upper = 1> - any value between [0,1]

Finally the model:

```
"
model{
    theta ~ beta(1, 1);
    y      ~ bernoulli(N,theta);
  }
"
```

here we have beta prior and a binomial likelihood
(note n bernoulli trials is one Binomial observations)

```
modelString = "
  data {
  int<lower = 0> N;
   int y[N];
  }
  parameters {
    real<lower = 0, upper = 1> theta;
  }
  model{
    theta ~ beta(1, 1);
    y     ~ bernoulli(theta);
  }
"
```

Now we have the complete model, but we need move data into stan:

```
n <- 10
y <- rbinom(n, 1, 0.5) # n - bernoulli trials
data.list <- list(y = y,
                  N = n )
```

Finally to run stan:

```
library(rstan)
result_stan <-stan(model_code = modelString,
                   data = data.list,
                   chains = 1,
                   iter   = 1000)
```

analysis of output:

summary(result_stan,par = "theta")$summary

posterior mean

monte carlo error

posterior standard dev

quantiles

how much each
MCMC sample is worth
discuss this next mote later.

```
          mean     se_mean        sd      2.5%       25%       50%      75%     97.5%    n_eff     Rhat
theta 0.426007 0.006442914 0.1349636 0.1851459 0.3284235 0.4226521 0.518616 0.6913102 438.8021 1.017182
```

If you want other quantiles use probs:

```
summary(result_stan, par = "theta", probs = c(0.1,0.9))
$summary
```

	mean	se_mean	sd	10%	90%	n_eff	Rhat
theta	0.426007	0.006442914	0.1349636	0.2503379	0.607655	438.8021	1.017182

Let us look at the Poisson example

1875	1876	1877	1878	1879	1880	1881	1882	1883	1884
3	5	7	9	10	18	6	14	11	9

1885	1886	1887	1888	1889	1890	1891	1892	1893	1894
5	11	15	6	11	17	12	15	8	4

$$P(Y = k|\theta) = \frac{\theta^k e^{-\theta}}{k!}$$

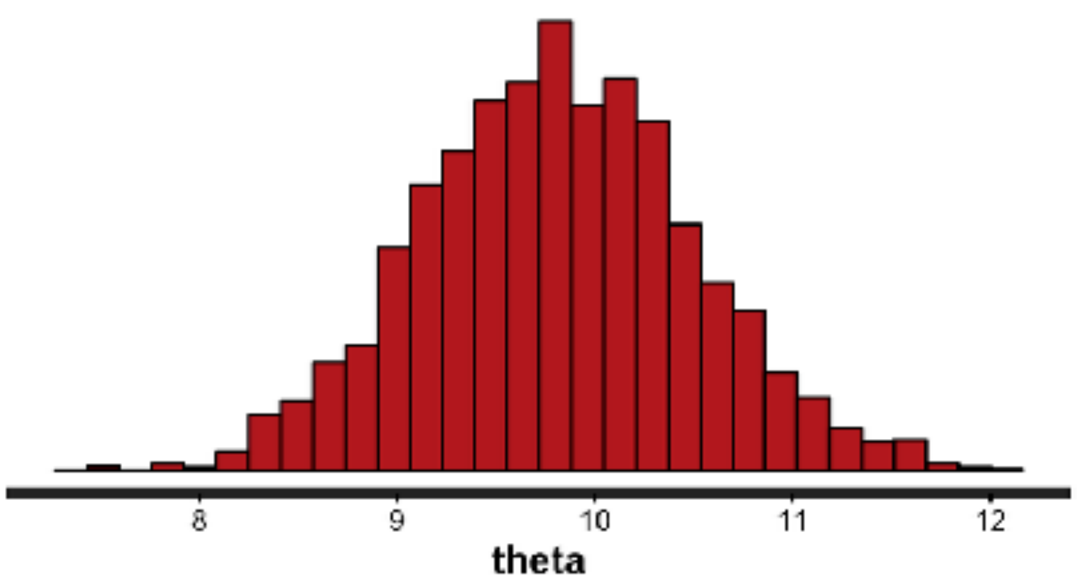
$$\theta \sim \Gamma(\alpha, \beta)$$

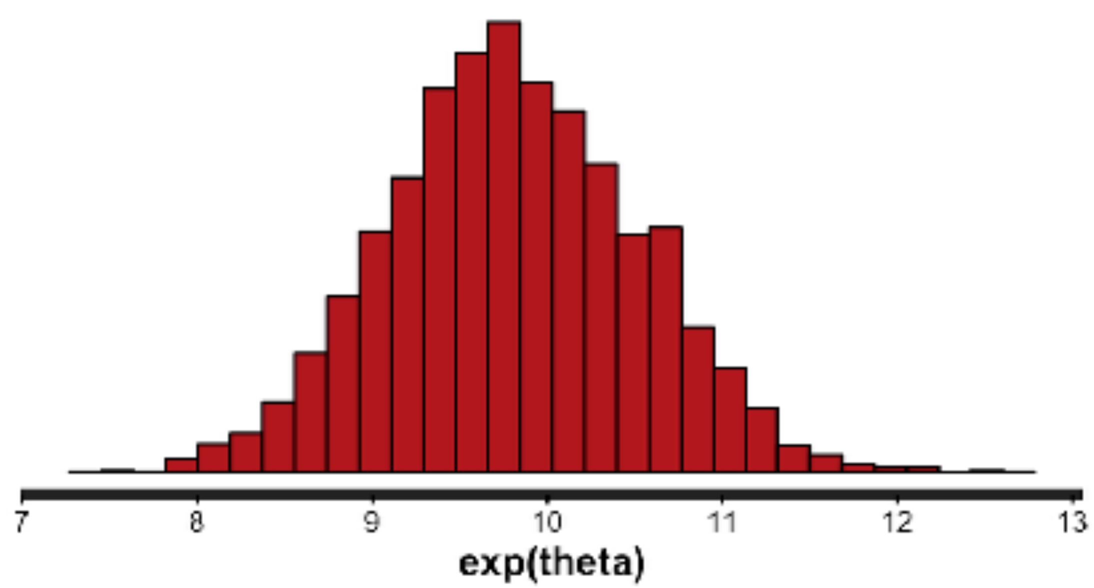
$$P(Y = k|\theta) = \frac{\theta^k e^{-\theta}}{k!}$$

$$\theta \sim \Gamma(\alpha = 1, \beta = 0.1)$$

$$P(Y = k|\theta) = \frac{e^{k\theta} e^{-e^{\theta}}}{k!}$$

$$\theta \sim N(0, 10)$$





Normal fixed mean unknown variance

Suppose one observe n iid normal zero mean random variables

$$\{x_1, x_2, \ldots, x_n\}$$

We are interested in estimate the variance of x.

$$f(x;\theta) \propto \frac{1}{\sqrt{\theta}} e^{-\frac{x^2}{2\theta}}$$

First we assume a Inverse Gamma prior

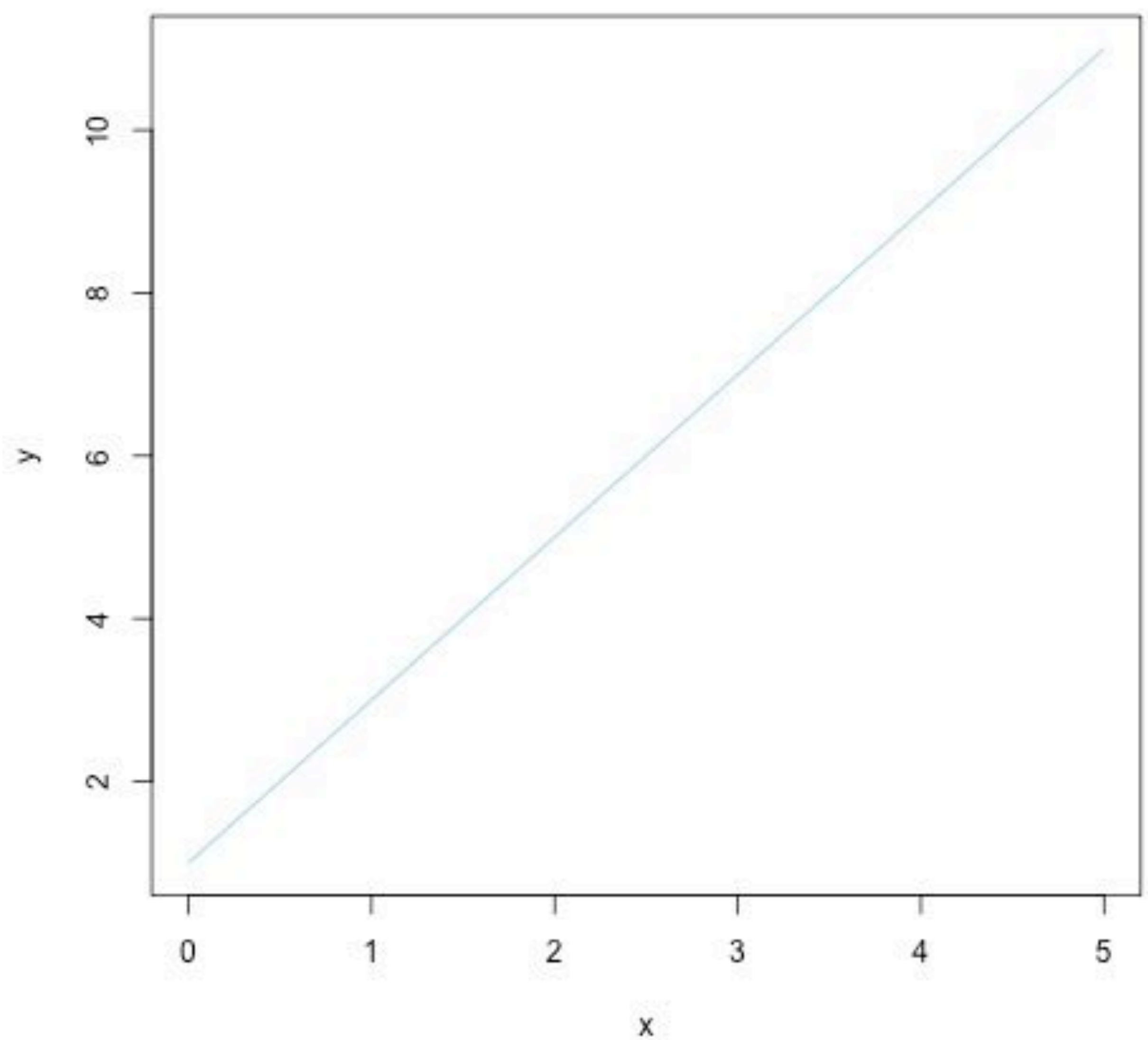
$$f(\theta;\alpha,\beta) = \Gamma^{-1}(\theta;\alpha,\beta) \propto \theta^{-\alpha-1} e^{-\frac{\beta}{\theta}}$$

Then one can show that posterior distribution is again gamma:

$$f(\theta|Y,\alpha,\beta) = \Gamma^{-1}(\theta;\alpha+\frac{n}{2},\beta+\sum_{i=1}^{n}\frac{x_i^2}{2})$$

Regression

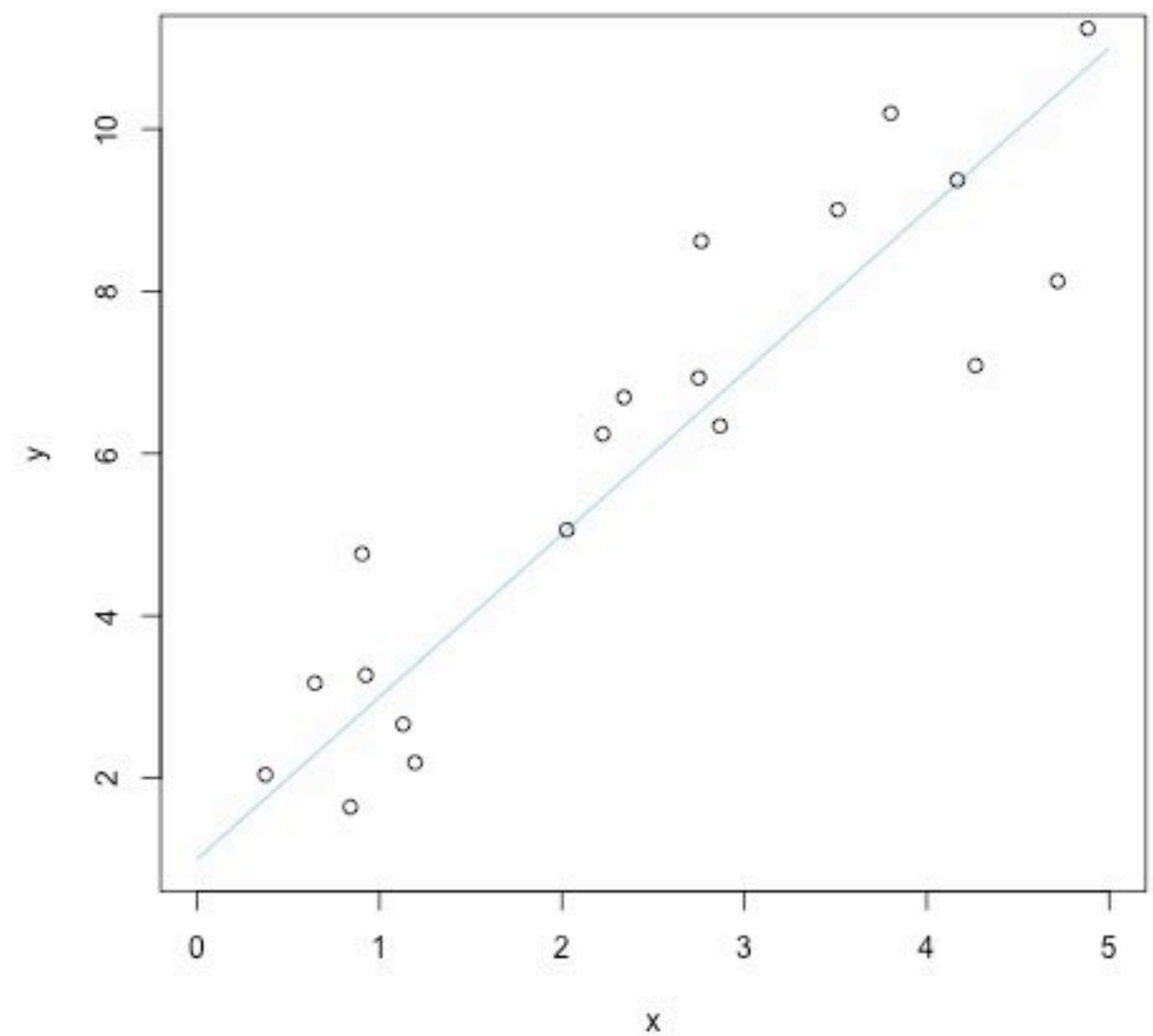
$$y = \beta_0 + x\beta_1$$



n=20

Statistical regression

$$y_j = \beta_0 + x_j\beta_1 + \epsilon_j$$
$$\epsilon_j \sim N(0, \sigma^2)$$



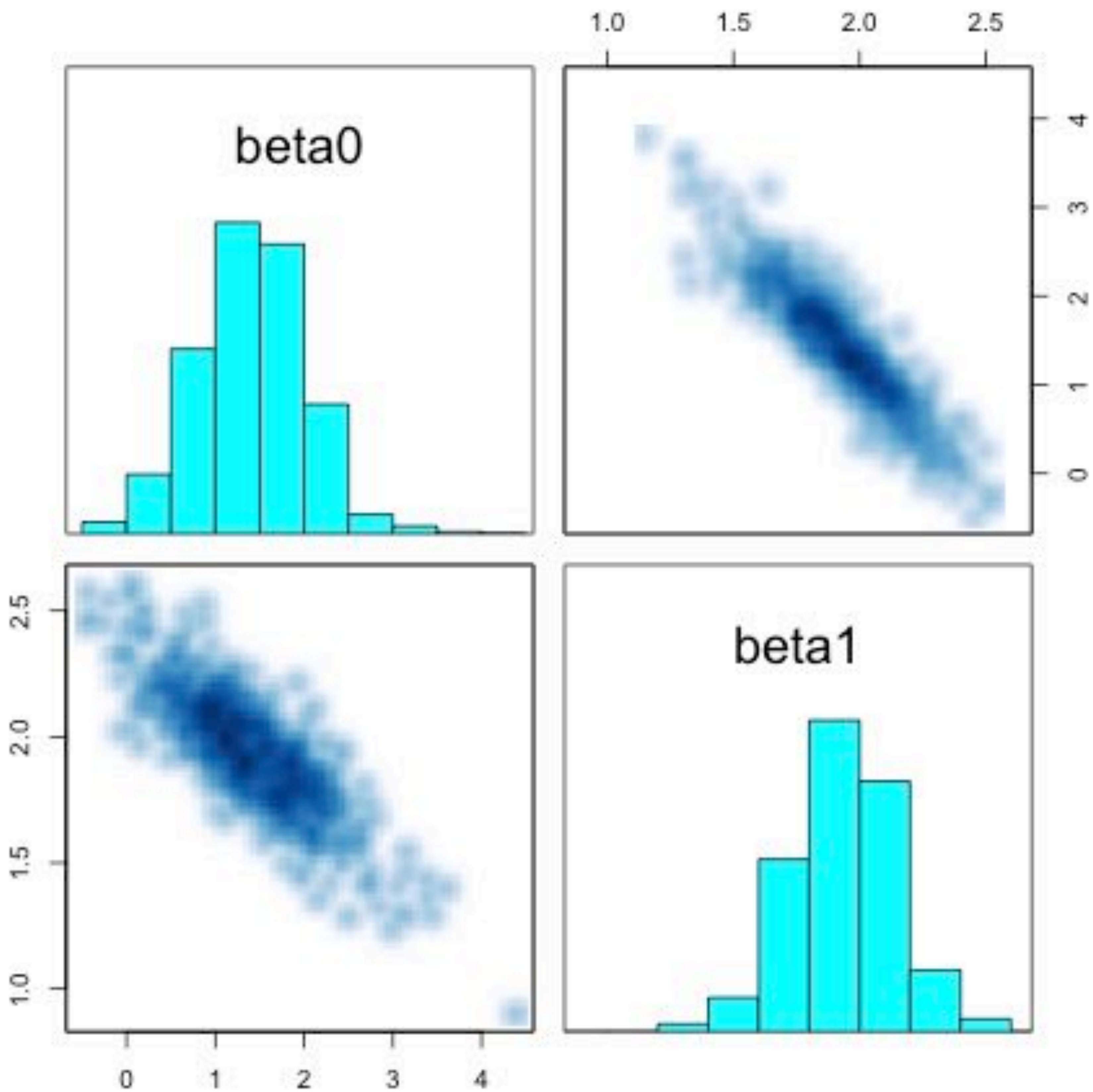
In Bayesian regression often:

$$\beta_i \sim N(0, \sigma_0^2)$$
$$\sigma^2 \sim \Gamma(\alpha, \beta)$$

The posterior we are interested in now is:

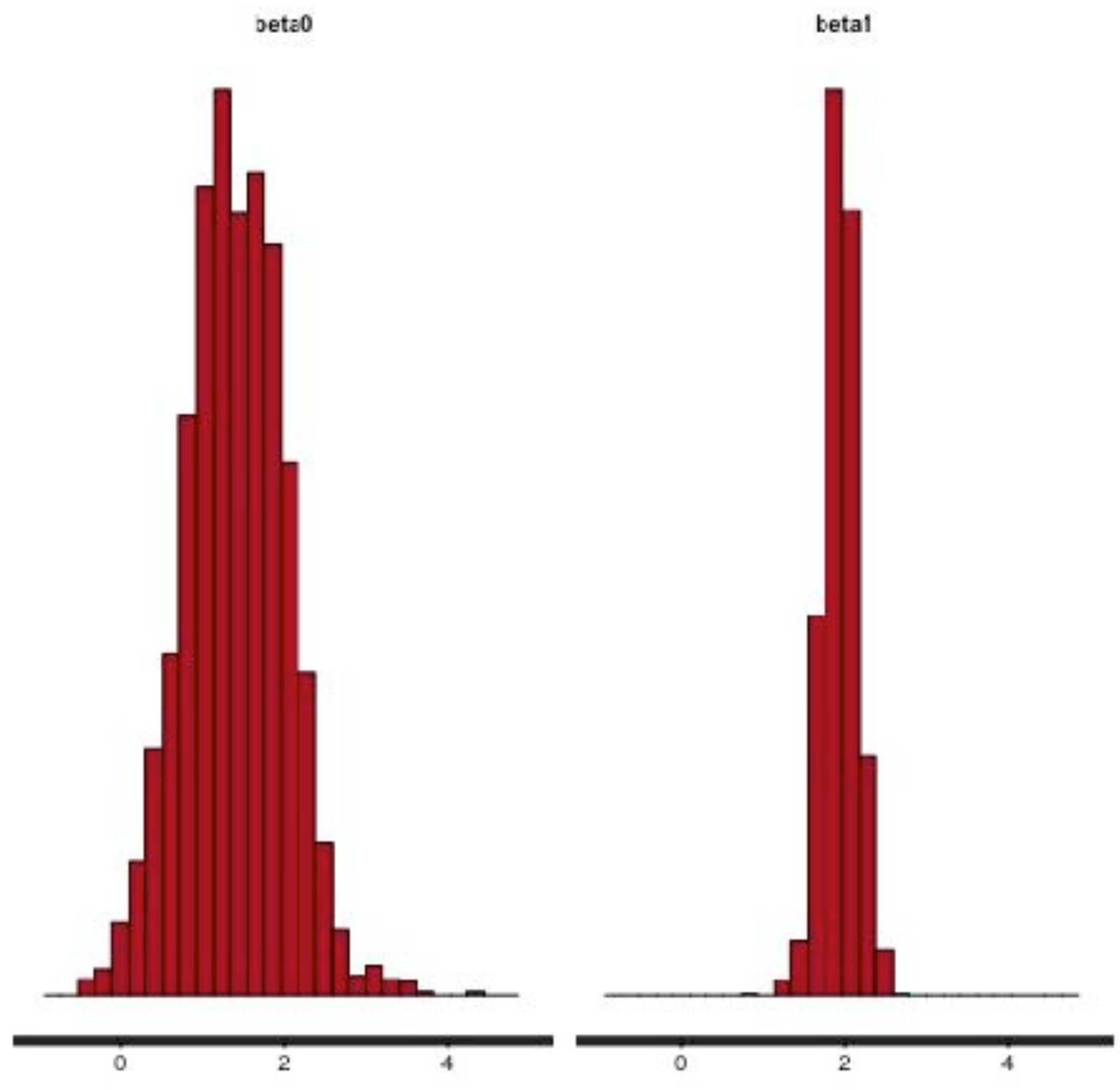
$$f(\beta_0, \beta_1, \sigma^2 | Y)$$

β is significant if the posterior interval does not contain 0

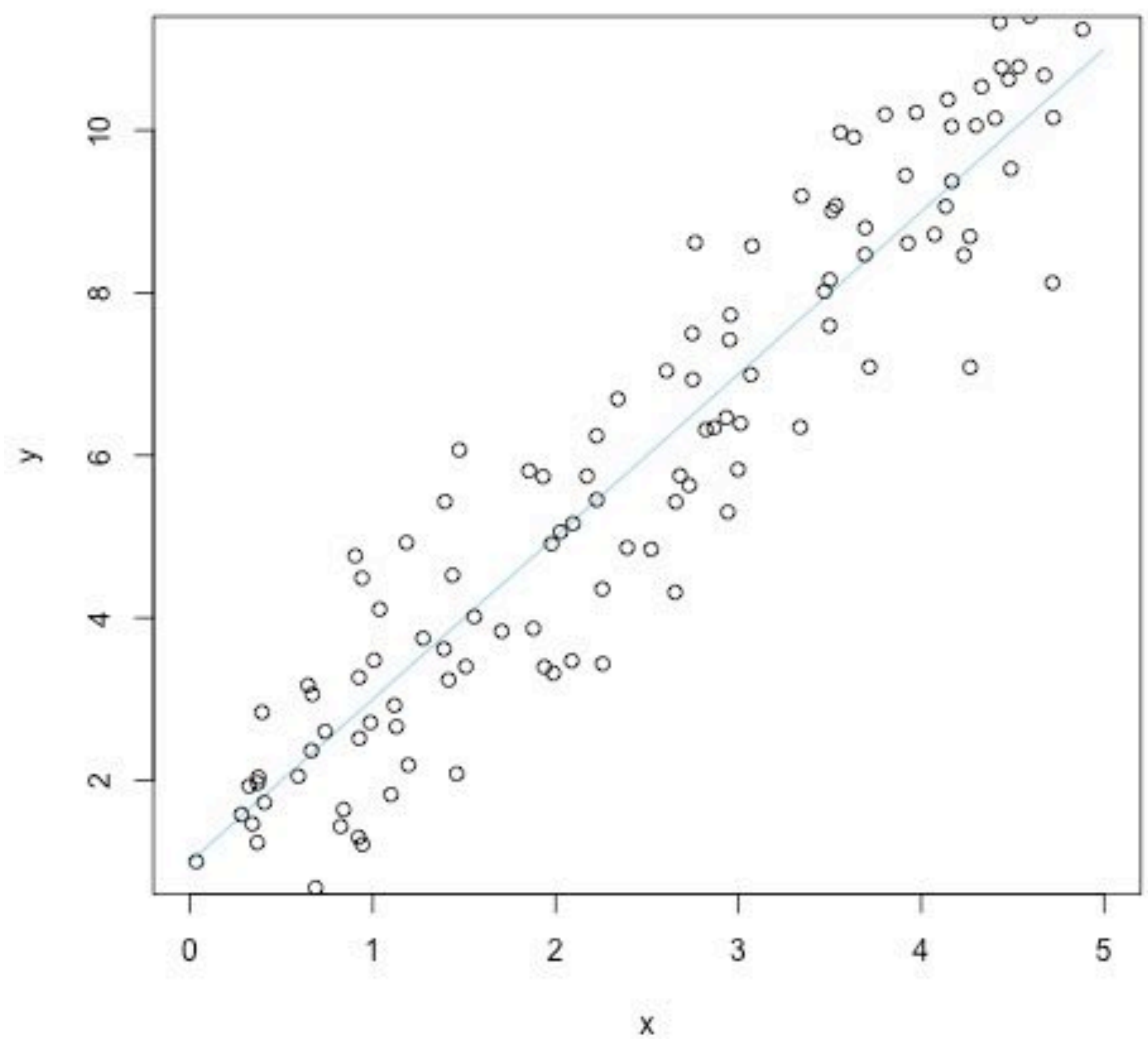


beta0
beta1
1.0
1.5
2.0
2.5
0
1
2
3
4

$f(\beta_0|Y_{1:10})$ $\qquad$ $f(\beta_1|Y_{1:10})$

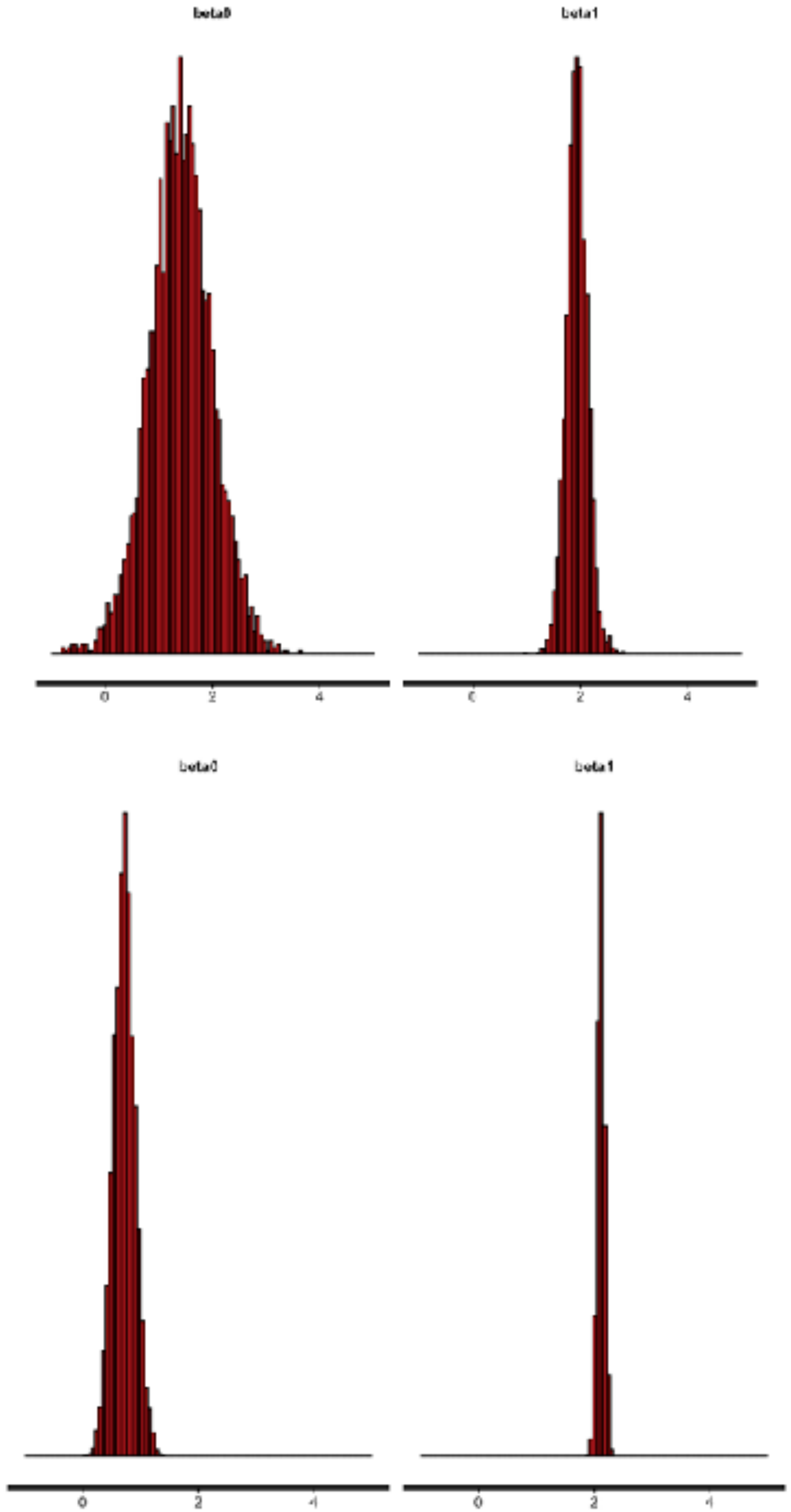


n=100



n=10

n=100



truth

$$\beta_0 = 1$$

$$\beta_1 = 2$$

one can of course have more covariates:

$$y = \sum_{i=0}^{p} x_i \beta_i + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$