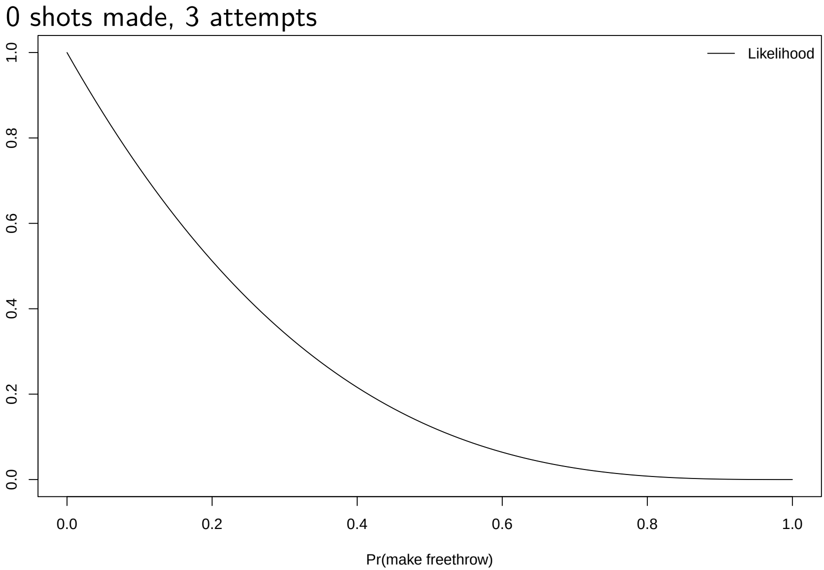
Week 3: Bayesian inference

Bayes' theorem

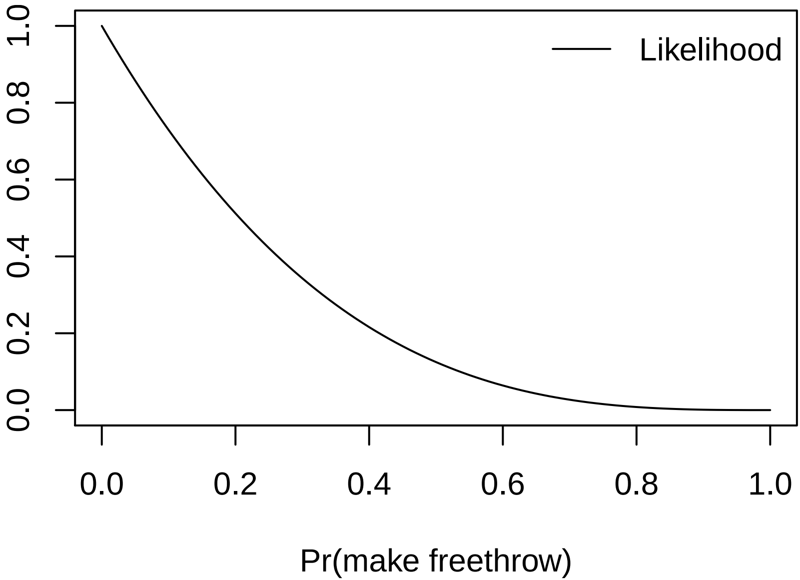
$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)}$$

$$p(\theta \mid y) \propto p(y \mid \theta)p(\theta)$$

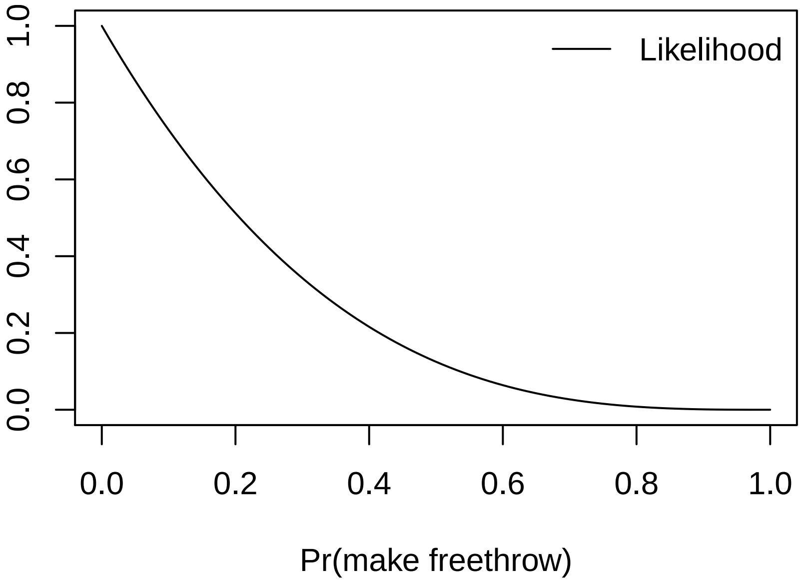
Freethrow example



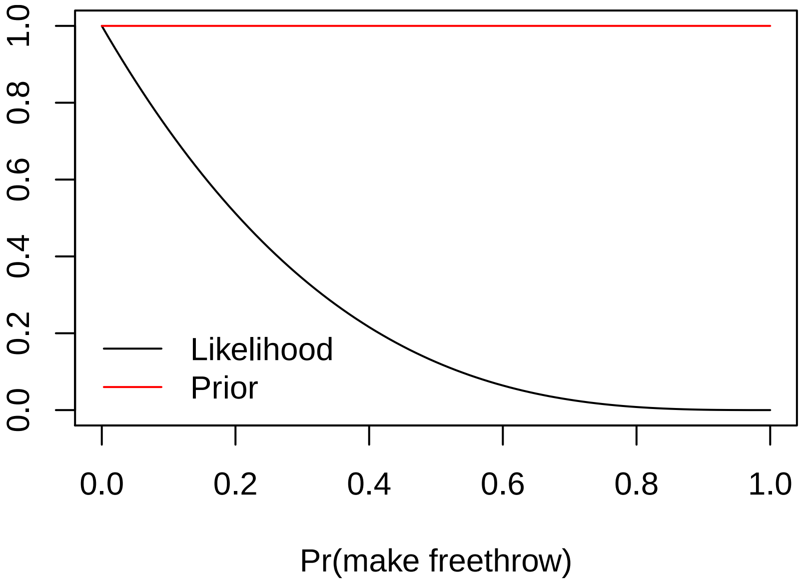
What is the MLE?



What is our prior?



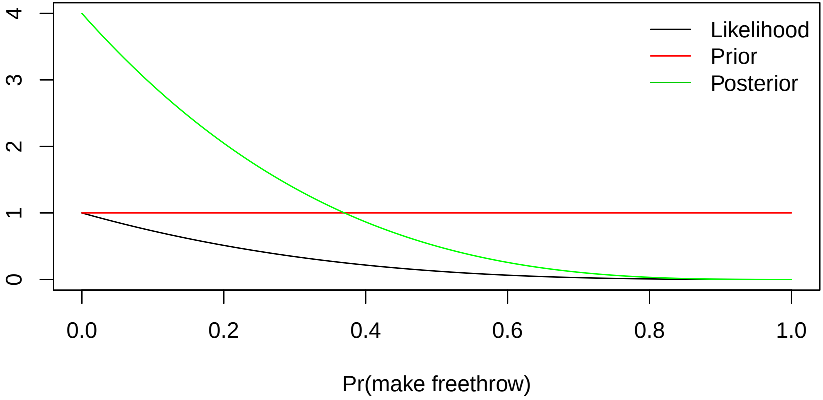
Uniform prior (never heard of a "free throw")



The posterior: uniform prior

$$y \sim Binomial(k = 3, p)$$

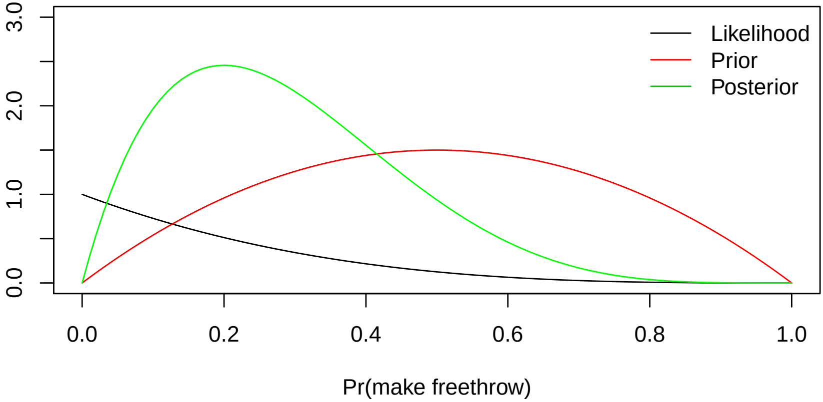
$$p \sim Beta(1, 1)$$



Non-uniform prior

$$y \sim Binomial(k = 3, p)$$

$$p \sim Beta(2, 2)$$



What if Pat takes a lot of free throws?

$$y \sim Binomial(k >> 3, p)$$

$$p \sim Beta(2, 2)$$

$k \rightarrow$ inf: prior doesn't matter

Demo: freethrows in Stan

MCMC animation

http://mbjoseph.github.io/2013/09/08/metropolis.html

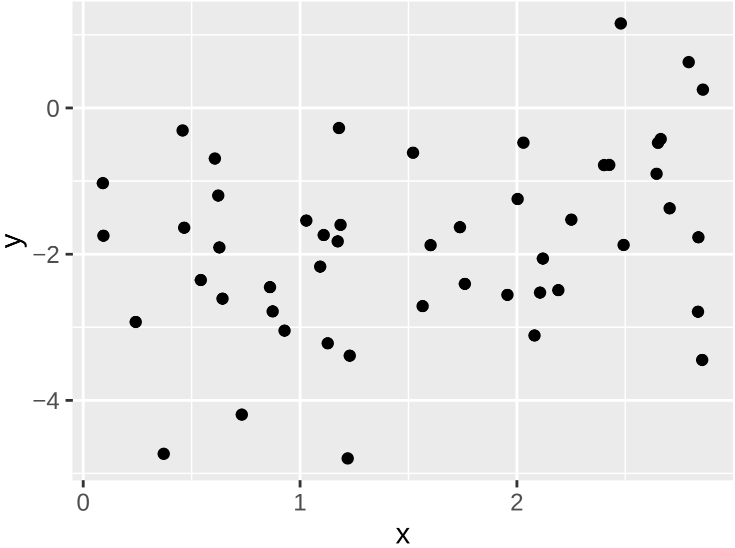
Bayes in practice

1. write model
2. translate model
3. estimate parameters

Bayes in practice

1. **write model**
2. translate model
3. estimate parameters

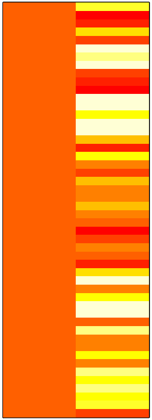
Bayesian linear regression



Writing a model

$$y \sim N(\mu, \sigma)$$

$$\mu = X\beta$$



Writing a model

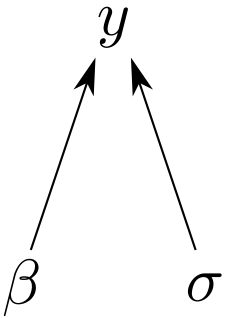
$y \sim N(\mu, \sigma)$
$\mu = X\beta$
What's missing?

Writing a model

$$
\begin{aligned}
&y \sim N(\mu, \sigma) \\
&\mu = X\beta \\
&\beta \sim N(0, 2) \\
&\sigma \sim \mathit{halfCauchy}(0, 5)
\end{aligned}
$$

Writing a model

$$y \sim N(\mu, \sigma)$$
$$\mu = X\beta$$
$$\beta \sim N(0, 2)$$
$$\sigma \sim halfCauchy(0, 5)$$



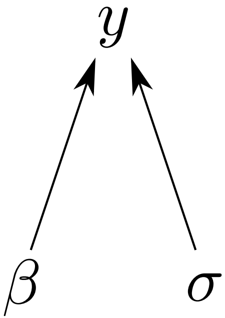
Writing a model

$$[\theta \mid y] = \frac{[\theta, y]}{[y]}$$

$$\implies [\theta \mid y] \propto [\theta, y]$$

Factoring $[\theta, y]$ with graph:

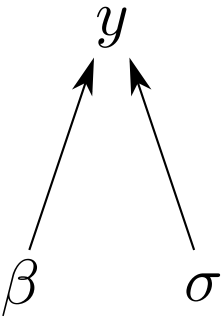
$$[\theta \mid y] \propto [y \mid \beta, \sigma]\,[\beta]\,[\sigma]$$



Components of the posterior distribution

$$[\theta \mid y] \propto [y \mid \beta, \sigma]\,[\beta]\,[\sigma]$$

$[y \mid \beta, \sigma]$: likelihood
$[\beta]$: prior for slope
$[\sigma]$: prior for standard deviation



Bayes in practice

1. write model
2. **translate model**
3. estimate parameters

Stan translation

```
data {
  int n;
  int p;
  matrix[n, p] X;
  vector[n] y;
}

parameters {
  vector[p] beta;
  real<lower=0> sigma;
}

model {
  beta ~ normal(0, 2);
  sigma ~ cauchy(0, 5);
  y ~ normal(X * beta, sigma);
}
```

$\beta \sim N(0, 2)$

$\sigma \sim halfCauchy(0, 5)$

$y \sim N(X\beta, \sigma)$

Bayes in practice

1. write model
2. translate model
3. **estimate parameters**

Estimating parameters

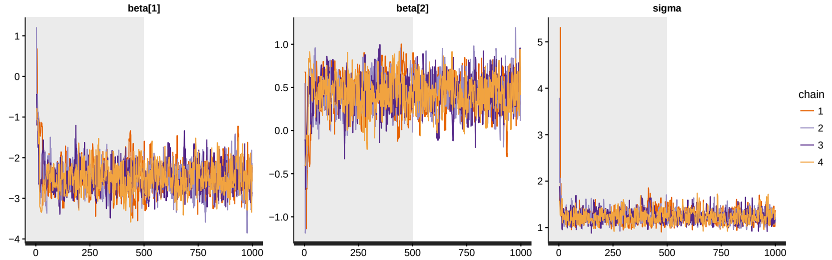
```
library(rstan)

stan_d <- list(y = y, X = X, n = nrow(X), p = ncol(X))

m <- stan('lm.stan', data = stan_d, iter=1000)
```

The last line does the following:

- generates MCMC algorithm for your model
- compiles it into fast C code
- initializes parameters
- runs MCMC algorithm
- formats output into a `stanfit` model

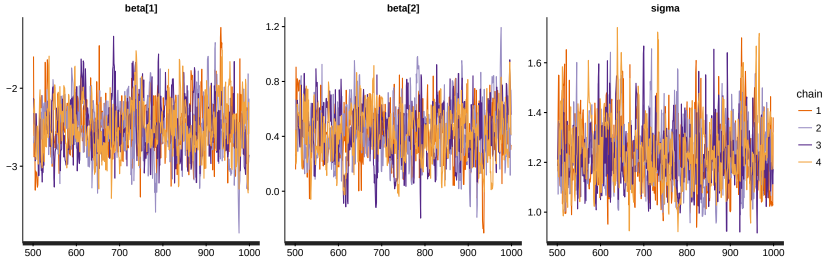
Evaluating convergence

```
traceplot(m, inc_warmup = TRUE)
```



Evaluating convergence

```r
traceplot(m)
```

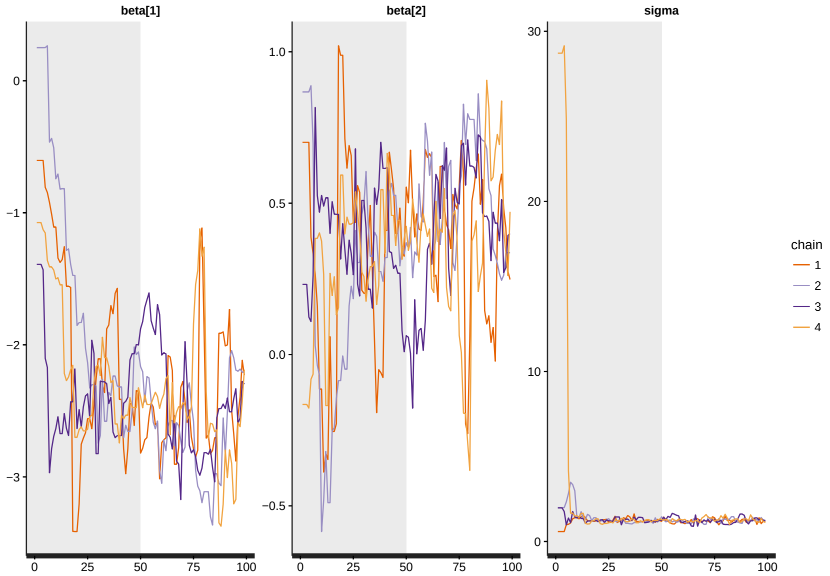


Evaluating convergence

```
m
```

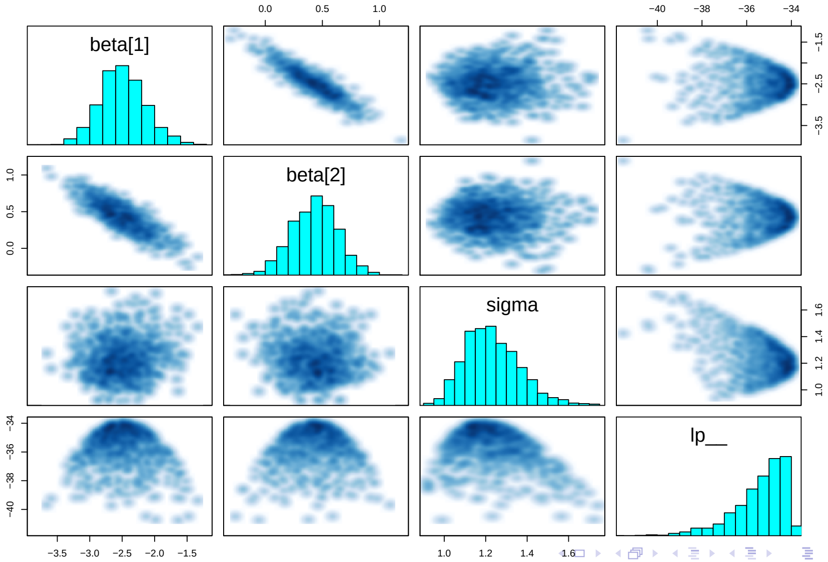
```
## Inference for Stan model: lm.
## 4 chains, each with iter=1000; warmup=500; thin=1;
## post-warmup draws per chain=500, total post-warmup draws=2000.
##
##            mean se_mean   sd   2.5%    25%    50%    75%  97.5% n_eff Rhat
## beta[1]   -2.49    0.02 0.35  -3.14  -2.73  -2.50  -2.27  -1.76   476    1
## beta[2]    0.43    0.01 0.20   0.04   0.29   0.44   0.57   0.82   457    1
## sigma      1.23    0.00 0.13   1.01   1.14   1.22   1.31   1.54   745    1
## lp__     -35.40    0.05 1.18 -38.41 -35.97 -35.09 -34.51 -33.99   615    1
##
## Samples were drawn using NUTS(diag_e) at Tue Feb  9 21:41:26 2016.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

Not enough iterations



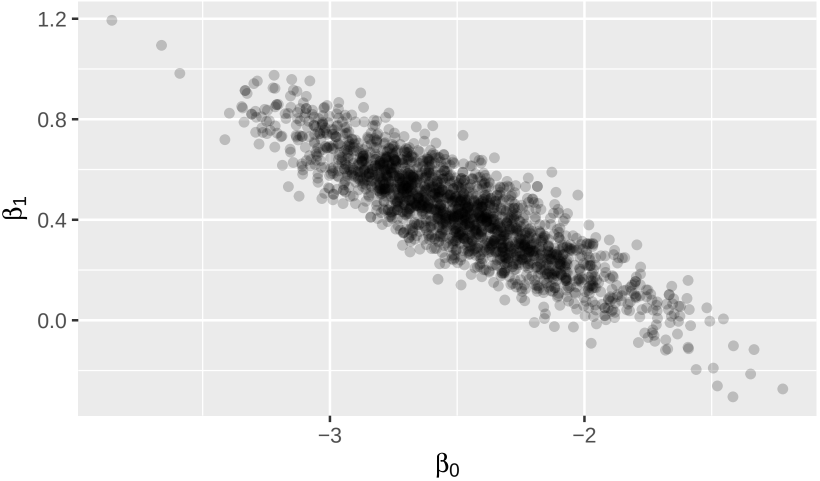
Posterior geometry

```
pairs(m)
```

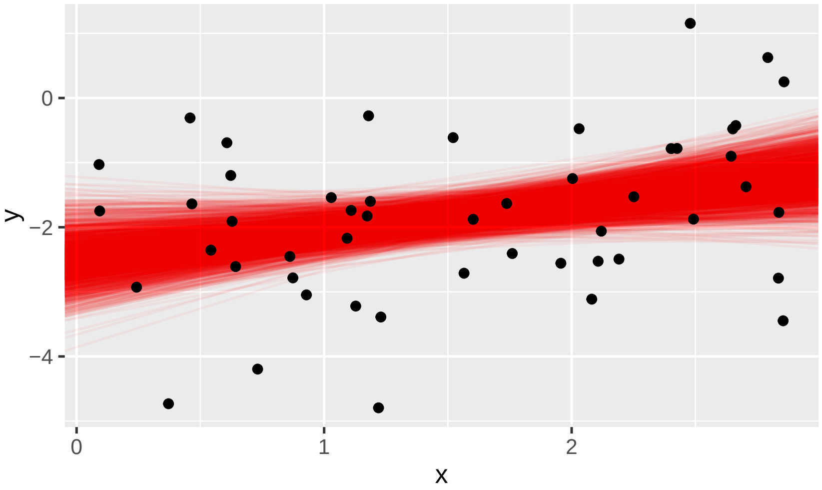


Correlation between slope and intercept

Is $\bar{x}$ positive or negative?



Visualizing posterior draws



Today's class: ladybird beetles and parasitoids