

Lab 3. Probabilistic Programming

Roi Naveiro

Stan in Linear Regression

Libraries

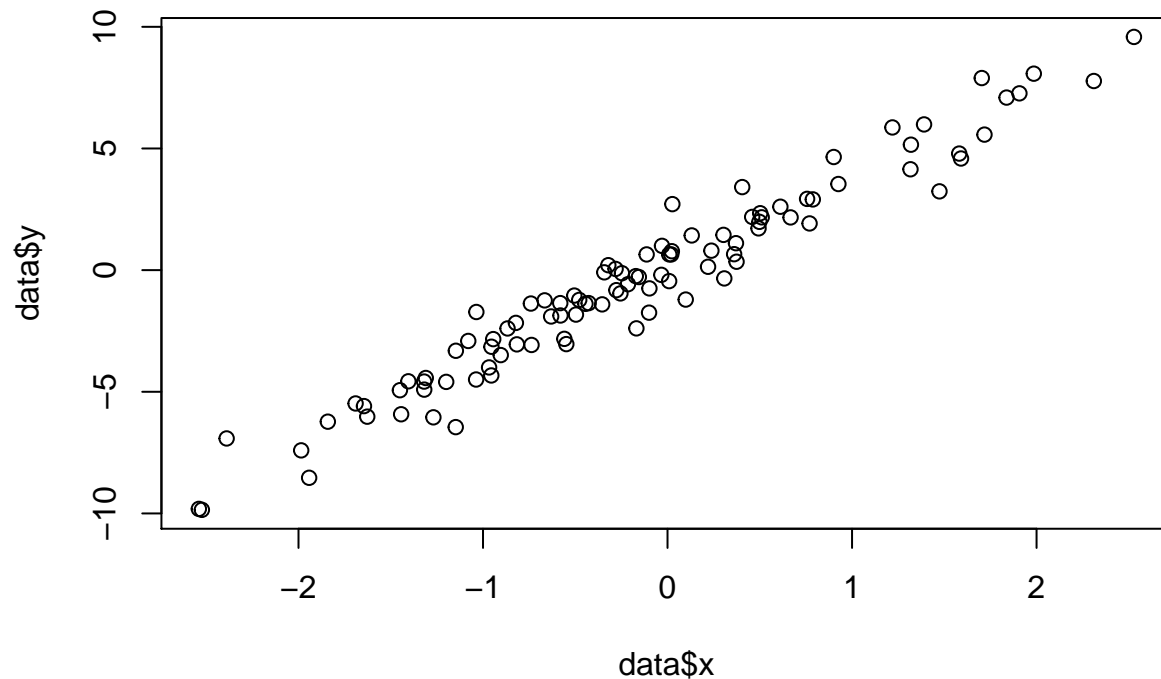
```
library(cmdstanr)
library(posterior)
library(bayesplot)
color_scheme_set("brightblue")
```

We first generate data from easy linear regression model

```
gen_dat <- function(n, beta, sigma) {
  x <- rnorm(n)
  y <- 0 + beta*x + rnorm(n, 0, sigma)
  data.frame(x = x, y = y)
}
```

We visualize them

```
data = gen_dat(100, 3.5, 0.85)
plot(data$x, data$y)
```



Sample from posterior

Compile model

```
mod <- cmdstan_model("lr.stan")
mod$print()

## // Linear Model with Normal Errors
## data {
##   // number of observations
##   int N;
##   // response
##   vector[N] y;
##   // covariate
##   vector[N] x;
## }
## parameters {
##   // regression coefficient vector
##   real beta;
##   real<lower=0> sigma;
## }
## transformed parameters {
##   vector[N] mu;
##
##   mu = x * beta;
## }
## model {
##   // priors
##   beta ~ normal(0., 2.0);
##   sigma ~ exponential(0.01);
##   // likelihood
##   y ~ normal(mu, sigma);
## }
```

Run model

```
data_1 <- list(N=100, y=data$y, x=data$x )
fit <- mod$sample(
  data = data_1,
  seed = 123,
  chains = 4,
  parallel_chains = 4,
  refresh = 500
)
```

```
## Running MCMC with 4 parallel chains...
##
## Chain 1 Iteration:    1 / 2000 [  0%] (Warmup)
## Chain 1 Iteration:   500 / 2000 [ 25%] (Warmup)
## Chain 1 Iteration:  1000 / 2000 [ 50%] (Warmup)
## Chain 1 Iteration:  1001 / 2000 [ 50%] (Sampling)
## Chain 1 Iteration:  1500 / 2000 [ 75%] (Sampling)
## Chain 1 Iteration:  2000 / 2000 [100%] (Sampling)
## Chain 2 Iteration:    1 / 2000 [  0%] (Warmup)
## Chain 2 Iteration:   500 / 2000 [ 25%] (Warmup)
## Chain 2 Iteration:  1000 / 2000 [ 50%] (Warmup)
```

```
## Chain 2 Iteration: 1001 / 2000 [ 50%] (Sampling)
## Chain 2 Iteration: 1500 / 2000 [ 75%] (Sampling)
## Chain 2 Iteration: 2000 / 2000 [100%] (Sampling)
## Chain 3 Iteration:   1 / 2000 [  0%] (Warmup)
## Chain 3 Iteration:  500 / 2000 [ 25%] (Warmup)
## Chain 3 Iteration: 1000 / 2000 [ 50%] (Warmup)
## Chain 3 Iteration: 1001 / 2000 [ 50%] (Sampling)
## Chain 3 Iteration: 1500 / 2000 [ 75%] (Sampling)
## Chain 3 Iteration: 2000 / 2000 [100%] (Sampling)
## Chain 4 Iteration:   1 / 2000 [  0%] (Warmup)
## Chain 4 Iteration:  500 / 2000 [ 25%] (Warmup)
## Chain 4 Iteration: 1000 / 2000 [ 50%] (Warmup)
## Chain 4 Iteration: 1001 / 2000 [ 50%] (Sampling)
## Chain 4 Iteration: 1500 / 2000 [ 75%] (Sampling)
## Chain 4 Iteration: 2000 / 2000 [100%] (Sampling)
## Chain 1 finished in 0.1 seconds.
## Chain 2 finished in 0.1 seconds.
## Chain 3 finished in 0.1 seconds.
## Chain 4 finished in 0.1 seconds.
##
## All 4 chains finished successfully.
## Mean chain execution time: 0.1 seconds.
## Total execution time: 0.3 seconds.
```

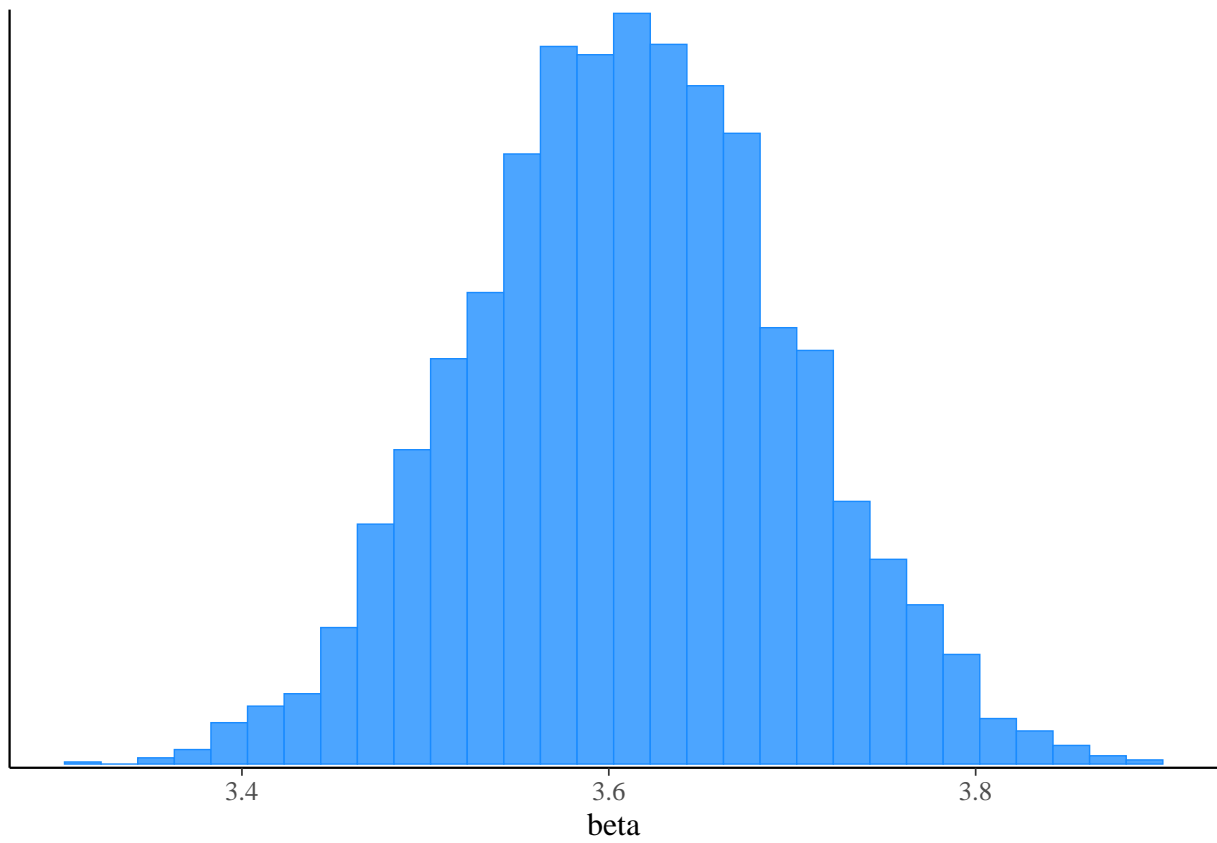
Summary of data

```
fit$summary()
```

```
## # A tibble: 103 x 10
##   variable      mean  median      sd      mad      q5      q95  rhat ess_bulk
##   <chr>      <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <dbl>   <dbl>
## 1 lp__      -43.1    -42.8    0.994    0.724   -45.2   -42.2    1.00   2038.
## 2 beta        3.61     3.61    0.0867   0.0867    3.47    3.76    1.00   3196.
## 3 sigma       0.924    0.920    0.0652   0.0640    0.823    1.04    1.00   2859.
## 4 mu[1]      -2.10    -2.10    0.0504   0.0504   -2.19   -2.02    1.00   3196.
## 5 mu[2]      -2.95    -2.95    0.0708   0.0708   -3.07   -2.83    1.00   3196.
## 6 mu[3]      -4.76    -4.76    0.114    0.114   -4.96   -4.58    1.00   3196.
## 7 mu[4]      -2.41    -2.41    0.0577   0.0577   -2.50   -2.31    1.00   3196.
## 8 mu[5]       0.0722   0.0723  0.00173  0.00173   0.0694   0.0752    1.00   3196.
## 9 mu[6]       4.76     4.77    0.114    0.114    4.58    4.96    1.00   3196.
## 10 mu[7]     -3.90    -3.90    0.0936   0.0936   -4.06   -3.75    1.00   3196.
## # ... with 93 more rows, and 1 more variable: ess_tail <dbl>
```

Posterior samples

```
mcmc_hist(fit$draws("beta"))
```



```
mcmc_hist(fit$draws("sigma"))
```

