stan examples

Yanruyu Zhu (yaz4004)

12/8/2021

This file contains examples of 'Stan'.
The first example is for Gaussian model.
The second example is for Beta-binomial model.

Set up

```
library('rstan')
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)
seed = 2021
```

Ex_1 Gaussian

The mock data is 1,000 i.i.d. X ~ Normal(μ=5,σ=1).

```
X = rnorm(n=1000, mean=5, sd=1) # fake data
mydata = list(N=1000, X=X)      # stan takes data format as 'List'
fit = stan(file='my_model_norm.stan', data=mydata, seed = seed)
#print(fit)
```

Template Result:

```
print(fit)
```

```
## Inference for Stan model: my_model_norm.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##          mean se_mean   sd    2.5%     25%     50%     75%   97.5% n_eff Rhat
## mu       5.02    0.00 0.03    4.95    5.00    5.02    5.04    5.08  3007    1
## sigma    1.03    0.00 0.02    0.98    1.01    1.03    1.04    1.07  3600    1
## lp__  -525.24    0.03 1.06 -528.04 -525.66 -524.92 -524.49 -524.22  1603    1
##
## Samples were drawn using NUTS(diag_e) at Sun Dec 12 00:04:47 2021.
## 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).
```

Pair plots:
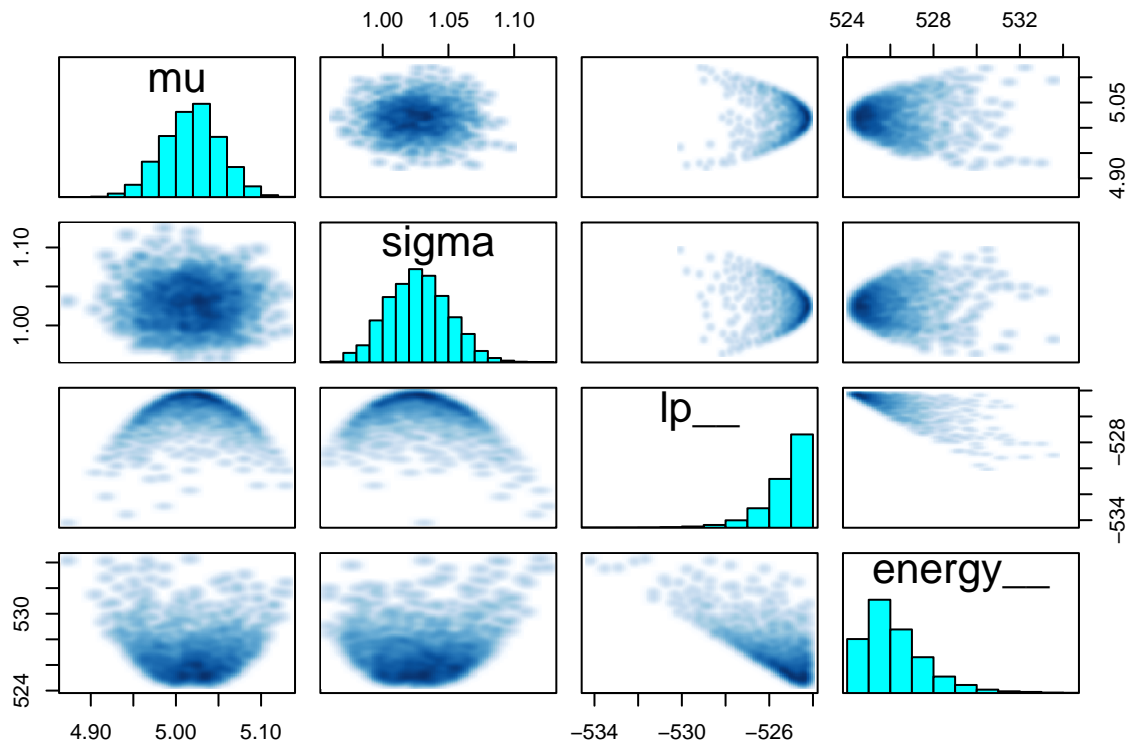
```
pairs(fit)
```


Figure 1: Pair plots for Gaussian model

Try to extract sample results

```
param = extract(fit)
param$mu[1:10]
```

```
##  [1] 5.01870 4.99595 4.97067 5.04062 5.03137 4.99058 5.03702 4.96794 4.99763
## [10] 5.00286
```

Ex_2 Beta-binomial

Parameters and set up:

The number of sampling times is 1000;
Prior distribution $\theta \sim Beta(1,1)$, $Y \sim Bin(100, \theta)$;
Observed data is 82 responses in the first cohort of 100 patients;

```
iter = 1000
data_bin <- list(N = 100, y = 82)

fit_bin <- stan(file = 'beta_binom.stan', data = data_bin, iter = iter, seed = seed)
```

Posterior distribution for θ:

```
monitor(fit_bin)
```

```
## Inference for the input samples (4 chains: each with iter = 1000; warmup = 0):
##
##          Q5   Q50   Q95  Mean  SD  Rhat Bulk_ESS Tail_ESS
## theta   0.7   0.8   0.9   0.8 0.0  1.01      581      705
## lp__  -51.1 -49.3 -49.0 -49.6 0.8  1.01      805      725
##
## For each parameter, Bulk_ESS and Tail_ESS are crude measures of
## effective sample size for bulk and tail quantities respectively (an ESS > 100
## per chain is considered good), and Rhat is the potential scale reduction
## factor on rank normalized split chains (at convergence, Rhat <= 1.05).
```

```
draws <- as.data.frame(fit_bin)
hist(draws[,'theta'])
```


Histogram of draws[, "theta"]
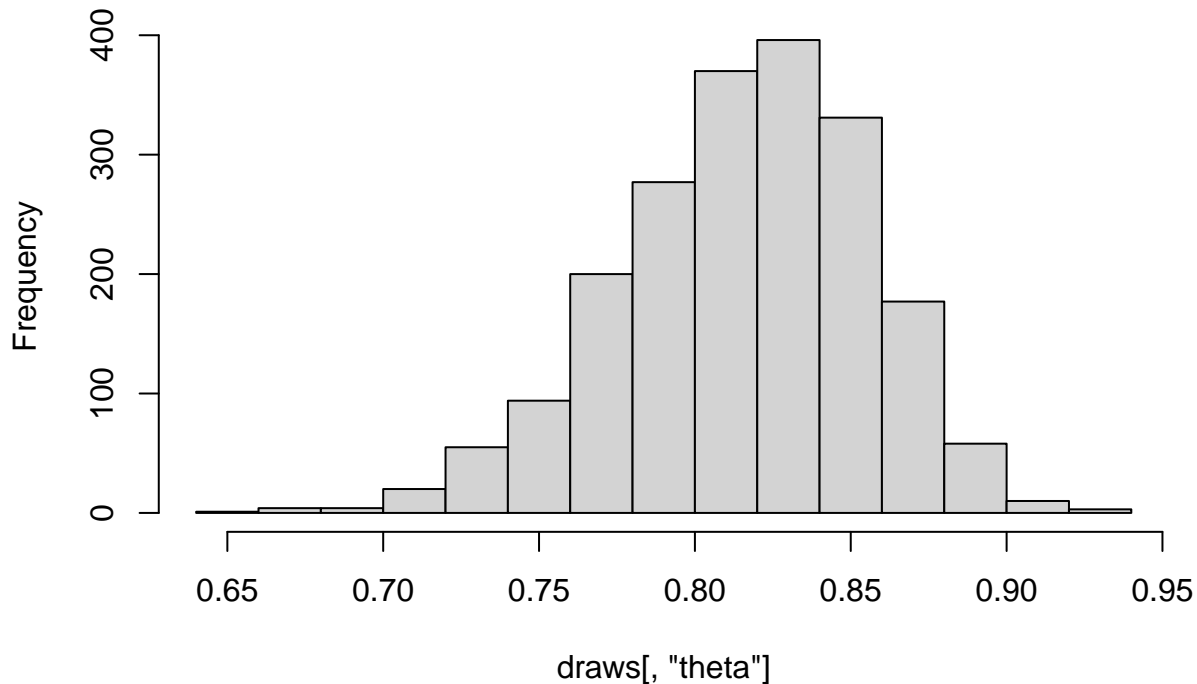
For $P(\theta \leq 0.92|Y = 82)$

```
sum(draws$theta<=0.92)/nrow(draws)
```

```
## [1] 0.9985
```

As $P(\theta \leq 0.92|Y = 82) = 0.9985 > 0.9$, we stop the trial at the first interim analysis for observing 82 responses in the first 100 patients.