

5.8 identifiability

Chris Parrish

July 3, 2016

Contents

identifiability	1
data	1
model	1
fit	2
plot	2

5.8 identifiability

reference:

- ARM chapter 05, github

```
library(rstan)
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
library(ggplot2)
```

identifiability

data

s

```
### Data
# N <- 60
# x <- rnorm(N, mean = 1, sd = 2)
# y <- ifelse(x < 2, 0, 1)
# stan_rdump(c("N", "y", "x"), file = "separation.data.R")
source("separation.data.R", echo = TRUE)

##
## > N <- 60
##
## > y <- c(0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1,
## +      1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0,
## +      1, 0, 0, 0, 0, 1, 1, 0, 0 .... [TRUNCATED]
##
## > x <- c(-0.339249227079859, 2.16776116177459, 0.583946420212137,
## +      -1.25010955939473, -2.51903611689114, 1.81544864387954, -0.679428295955671,
## +      .... [TRUNCATED]
```

model

separation.stan

```

data {
  int<lower=0> N;
  int<lower=0,upper=1> y[N];
  vector[N] x;
}
parameters {
  vector[2] beta;
}
model {
  y ~ bernoulli_logit(beta[1] + beta[2] * x);
}

```

fit

```

## Model: y ~ x
data.list <- c("N", "y", "x")
separation.sf <- stan(file='separation.stan', data=data.list,
                    iter=1000, chains=4)

```

```

## Warning: There were 1953 divergent transitions after warmup. Increasing
## adapt_delta above 0.8 may help.

```

```

## Warning: Examine the pairs() plot to diagnose sampling problems

```

```

print(separation.sf)

```

```

## Inference for Stan model: separation.
## 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%
## beta[1] -371067.5 170958.94 401387.5 -1431396.47 -545267.96 -298391.9
## beta[2]  188032.1  86676.88 203334.8   12527.67  23056.95 148933.0
## lp__         0.0      0.00      0.0      0.00      0.00      0.0
##               75%     97.5% n_eff Rhat
## beta[1] -45454.6 -24364.36      6 1.68
## beta[2]  277287.1  724180.09      6 1.68
## lp__         0.0      0.00  2000 1.00
##
## Samples were drawn using NUTS(diag_e) at Sun Jul  3 17:35:35 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).
## The estimated Bayesian Fraction of Missing Information is a measure of
## the efficiency of the sampler with values close to 1 being ideal.
## For each chain, these estimates are
## 2 2.6 2.4 1.9

```

plot

```

## Plot
beta.post <- extract(separation.sf, "beta")$beta
b <- colMeans(beta.post)

```

```
p <- ggplot(data.frame(x, y), aes(x, y)) +  
  geom_point(shape = 20, color = "darkred") +  
  stat_function(fun = function(x) 1.0 / (1 + exp(-(b[1] + b[2] * x))))  
print(p)
```

