UCBadmit

Chris Parrish

June 9, 2016

Contents

UCBadmit

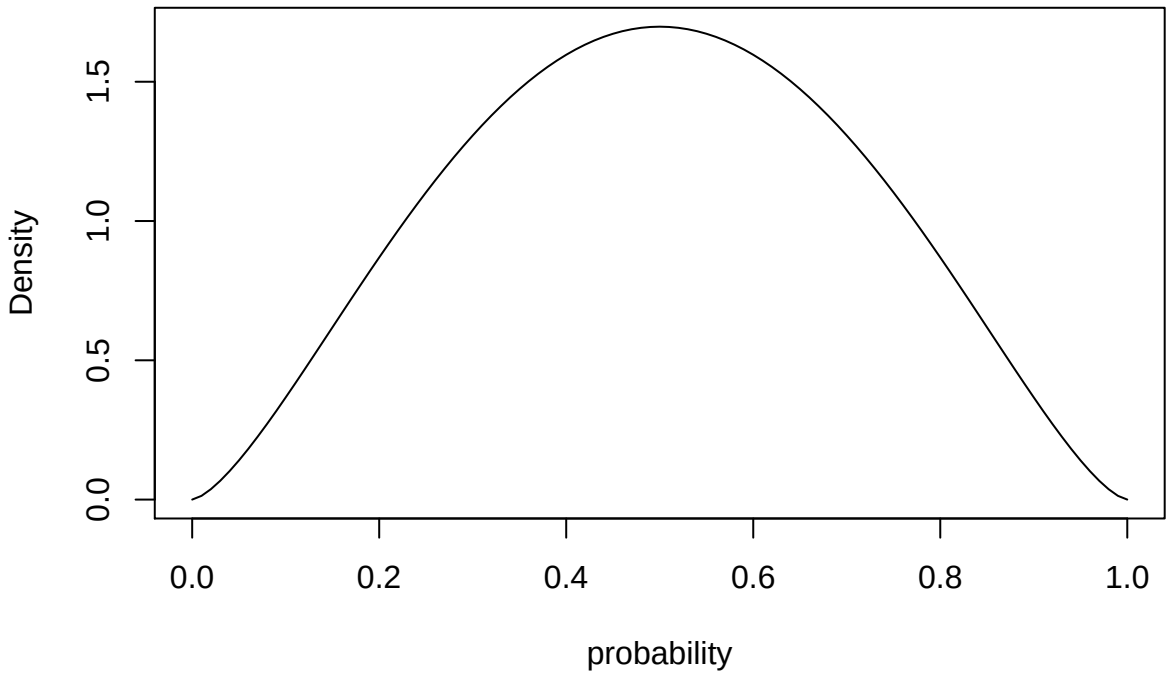
reference: McElreath, Statistical Rethinking, chap 11, p.347

```
library(rethinking)
library(ggplot2)
```

admissions

beta distribution

```
## R code 11.25
pbar <- 0.5
theta <- 5
curve( dbeta2(x,pbar,theta) , from=0 , to=1 ,
    xlab="probability" , ylab="Density" )
```



data

```
## R code 11.26
data(UCBadmit)
d <- UCBadmit
str(d)
```

```
## 'data.frame':    12 obs. of  5 variables:
##  $ dept            : Factor w/ 6 levels "A","B","C","D",..: 1 1 2 2 3 3 4 4 5 5 ...
##  $ applicant.gender: Factor w/ 2 levels "female","male": 2 1 2 1 2 1 2 1 2 1 ...
##  $ admit           : int  512 89 353 17 120 202 138 131 53 94 ...
##  $ reject          : int  313 19 207 8 205 391 279 244 138 299 ...
##  $ applications    : int  825 108 560 25 325 593 417 375 191 393 ...
```

```
d <- d[ , c("admit", "applications")]
```

model

$$
\begin{aligned}
A_i &\sim BetaBinomial(n_i, \bar{p}_i, \theta) \\
logit(\bar{p}_i) &= \alpha \\
\alpha &\sim Normal(0, 10) \\
\theta &\sim HalfCauchy(0, 1)
\end{aligned}
$$

stan

```
m11.5 <- map2stan(
    alist(
        admit ~ dbetabinom(applications,pbar,theta),
        logit(pbar) <- a,
        a ~ dnorm(0,2),
        theta ~ dexp(1)
    ),
    data=d,
    constraints=list(theta="lower=0"),
    start=list(theta=3),
    iter=4000 , warmup=1000 , chains=2 , cores=2 )
```

```
##
## SAMPLING FOR MODEL 'admit ~ dbetabinom(applications, pbar, theta)' NOW (CHAIN 1).
## WARNING: No variance estimation is
##          performed for num_warmup < 20
##
##
## Chain 1, Iteration: 1 / 1 [100%]  (Sampling)
##  Elapsed Time: 6e-06 seconds (Warm-up)
##                7e-05 seconds (Sampling)
##                7.6e-05 seconds (Total)
```

```
## Computing WAIC
```

```
## Constructing posterior predictions
```

```
## [ 600 / 6000 ]
[ 1200 / 6000 ]
[ 1800 / 6000 ]
[ 2400 / 6000 ]
[ 3000 / 6000 ]
[ 3600 / 6000 ]
[ 4200 / 6000 ]
[ 4800 / 6000 ]
[ 5400 / 6000 ]
[ 6000 / 6000 ]
```

```
## R code 11.27
precis(m11.5)
```

```
##        Mean StdDev lower 0.89 upper 0.89 n_eff Rhat
## theta  2.76   0.99       1.20       4.19  4610    1
## a     -0.37   0.31      -0.89       0.10  4437    1
```

analysis

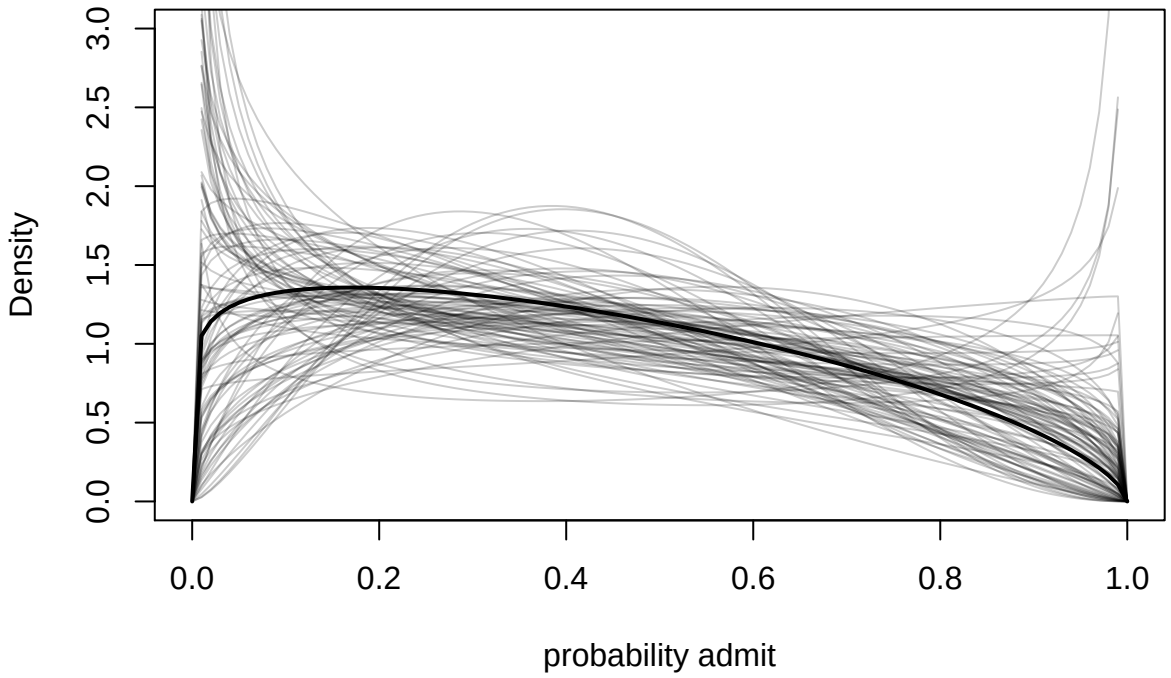
```
## R code 11.28
post <- extract.samples(m11.5)
quantile( logistic(post$a) , c(0.025,0.5,0.975) )
```

```
##      2.5%       50%     97.5%
## 0.2731602 0.4079590 0.5583337
```

```
## R code 11.29
post <- extract.samples(m11.5)

# draw posterior mean beta distribution
curve( dbeta2(x,mean(logistic(post$a)),mean(post$theta)) , from=0 , to=1 ,
    ylab="Density" , xlab="probability admit", ylim=c(0,3) , lwd=2 )

# draw 100 beta distributions sampled from posterior
for ( i in 1:100 ) {
    p <- logistic( post$a[i] )
    theta <- post$theta[i]
    curve( dbeta2(x,p,theta) , add=TRUE , col=col.alpha("black",0.2) )
}
```



postcheck

```
## R code 11.30
postcheck(m11.5)
```

```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]
## [ 100 / 1000 ]
[ 200 / 1000 ]
```

```
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]
```