Warnings of ASH

Lei Sun

July 10, 2015

When investigating the null behavior of ASH, we found ASH usually generates 6-7 warnings out of 100 runs - each run consists of three ASH functions, namely ASH applied to $\hat{\beta}_1$, $\hat{\beta}_2$ jointly and separately. Most of the warnings are due to EM not converging. This mini-project is to investigate how and when it happens.

The code for an M-run simulation is as follows.

```r
library(ashr)

t<-c()

M<-1

for (i in 1:M){
  beta1<-rnorm(1000,0,1)
    beta2<-rnorm(1000,0,1)
    s<-rep(1,1000)
    betahat1<-rnorm(1000,beta1,s)
    betahat2<-rnorm(1000,beta2,s)

    r12<-ash(c(betahat1,betahat2),c(s,s),mixcompdist="normal",method="shrink")
    r1<-ash(betahat1,s,mixcompdist="normal",method="shrink")
    r2<-ash(betahat2,s,mixcompdist="normal",method="shrink")

    t<-c(t,2*(r1$loglik+r2$loglik-r12$loglik))
}
```

That is, we generate β_1, $\beta_2 \sim N(0,1)$, set $\hat{s}=1$ for both groups, $\hat{\beta}_1 \sim N(\beta_1, \hat{s}=1)$, $\hat{\beta}_2 \sim N(\beta_2, \hat{s}=1)$. Let Group 1 $= \left(\hat{\beta}_1, s=1\right)$, Group 2 $= \left(\hat{\beta}_2, s=1\right)$.
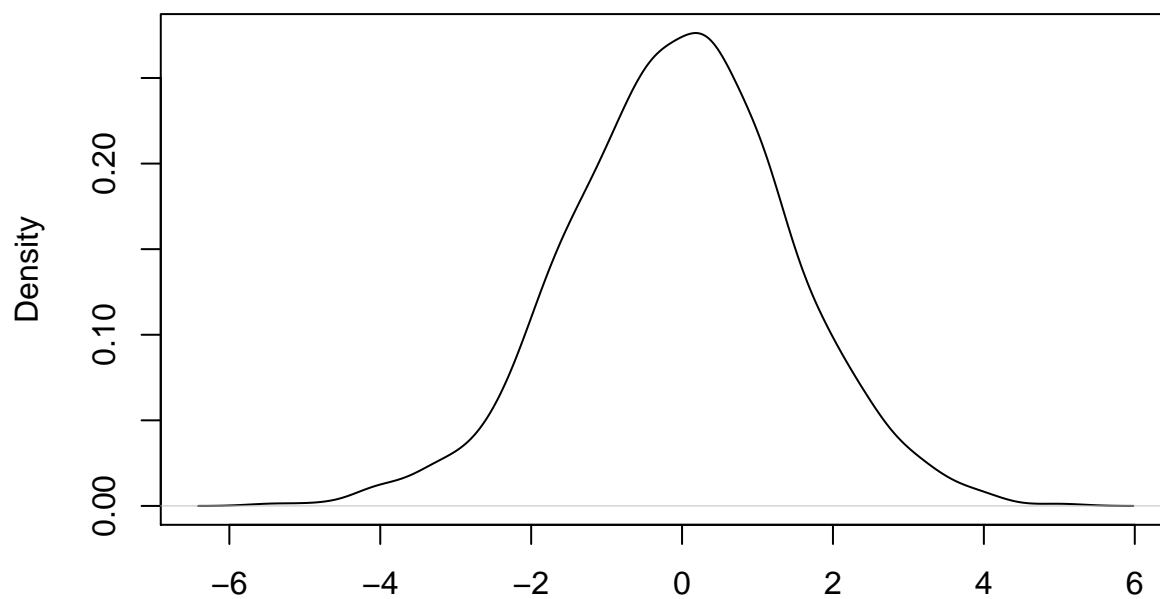
Now we do a 200-run simulation, and use `tryCatch` to record the detailed warning messages: each time ASH generates a warning, `tryCatch` record the corresponding $\hat{\beta}_1$ and/or $\hat{\beta}_2$. We've recorded 12 verified "malign" data sets which generate EM not coverging warnings and plot them as follows.
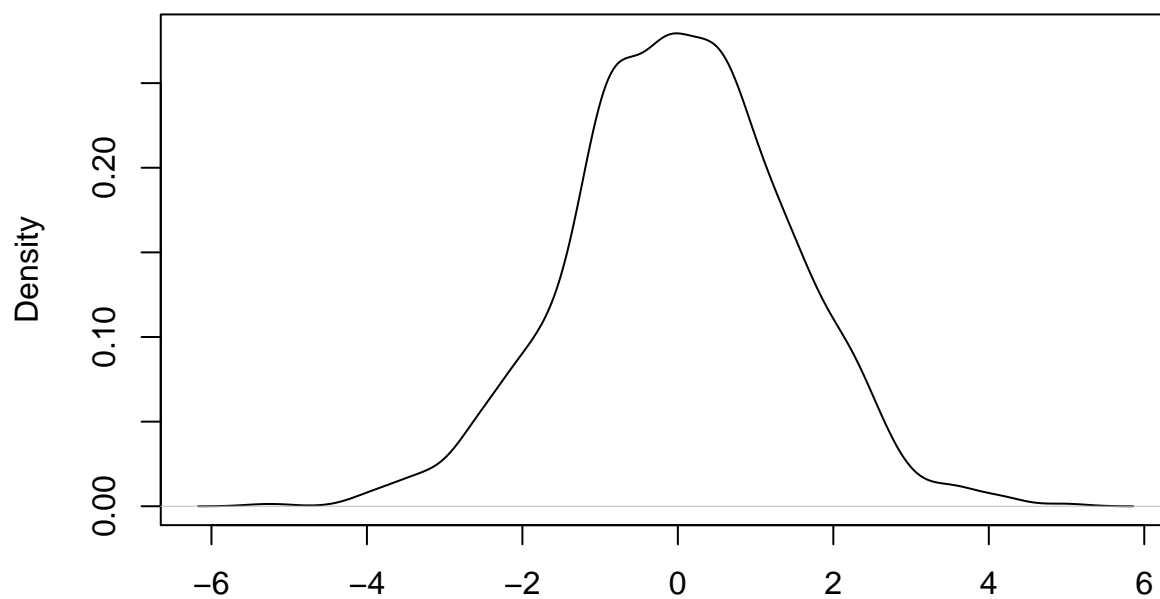
betahat_1
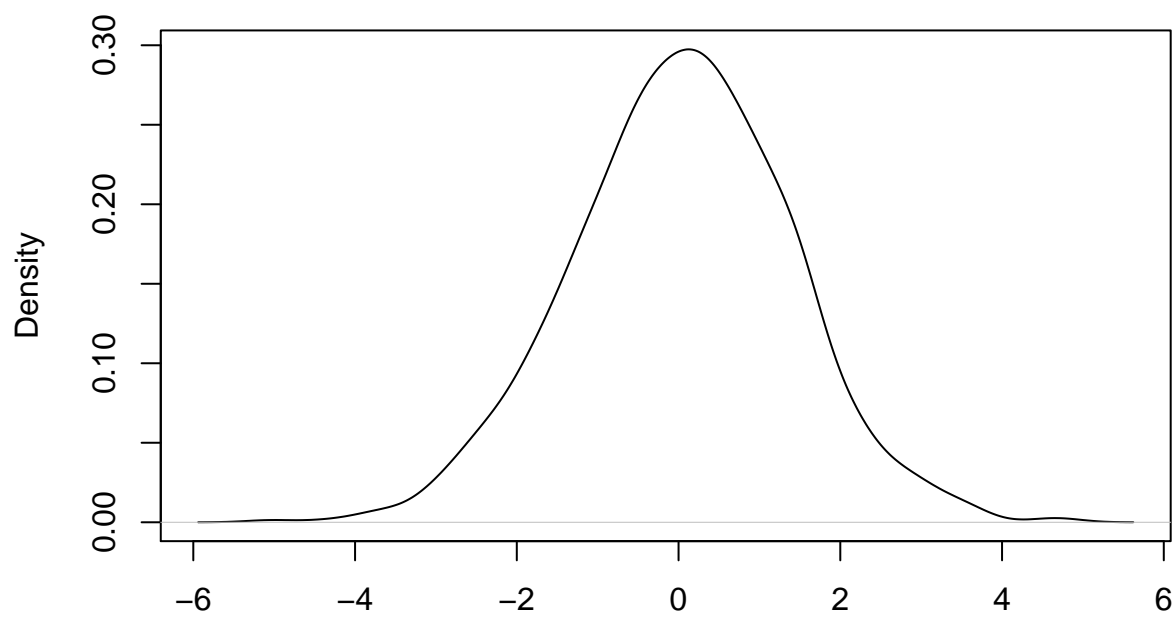
N = 1000   Bandwidth = 0.2996

betahat_2
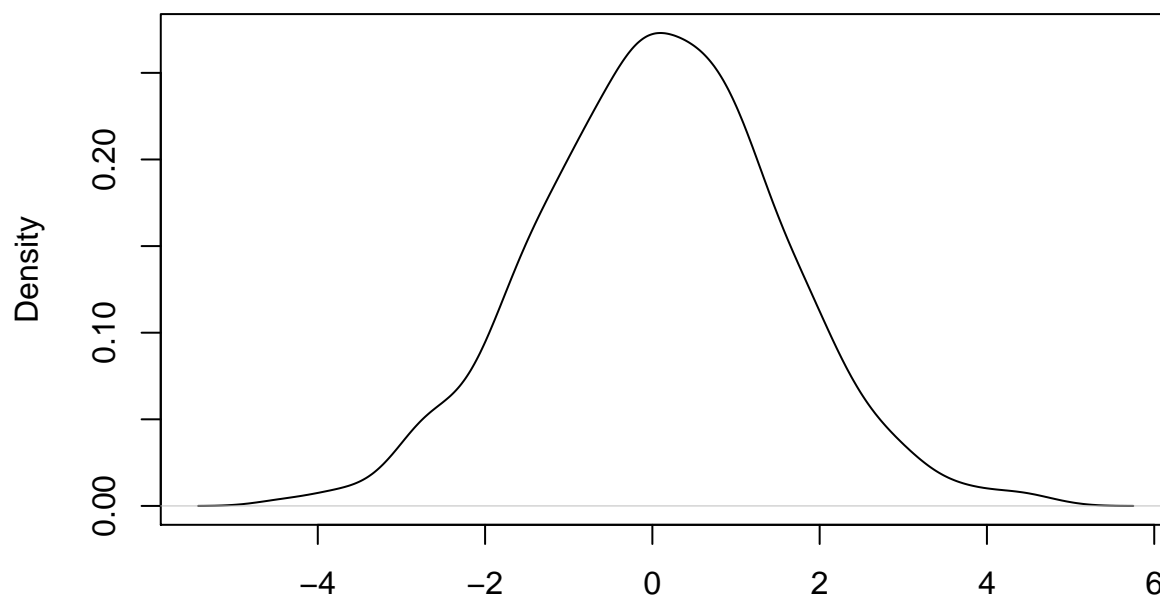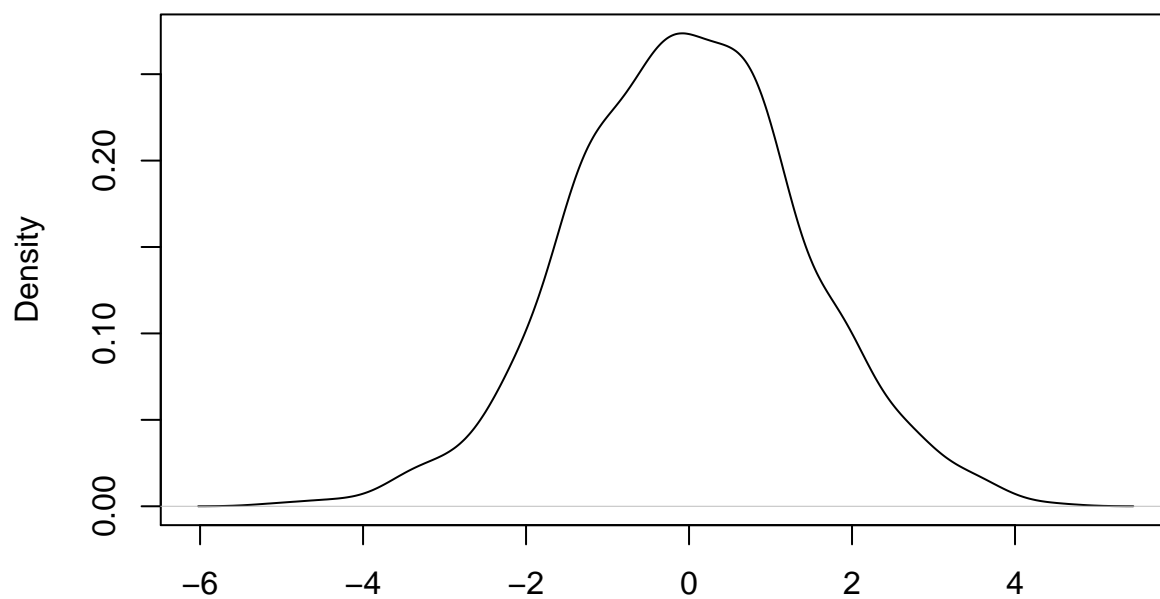

N = 1000   Bandwidth = 0.3281

betahat_3

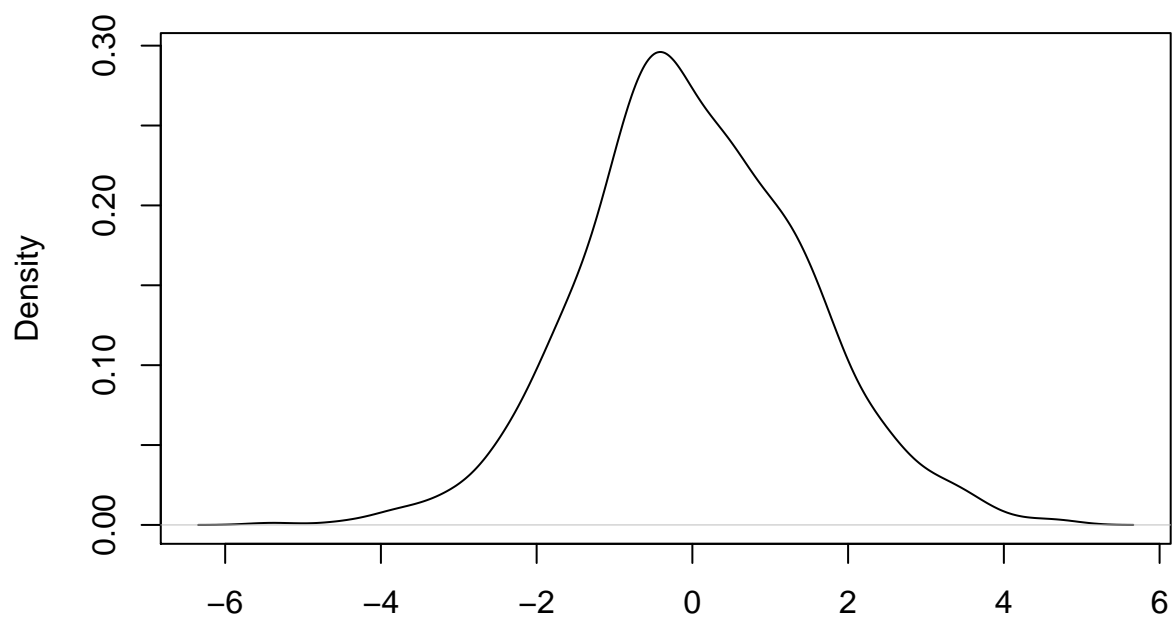N = 2000   Bandwidth = 0.2669

betahat_4

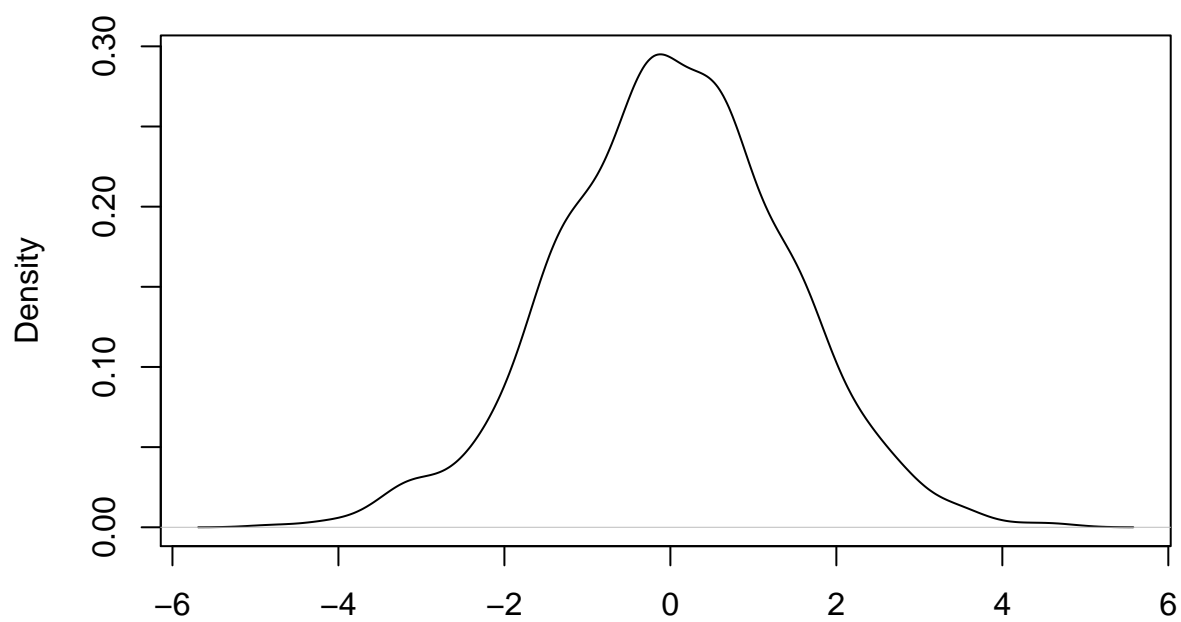N = 1000   Bandwidth = 0.3219

3

betahat_5

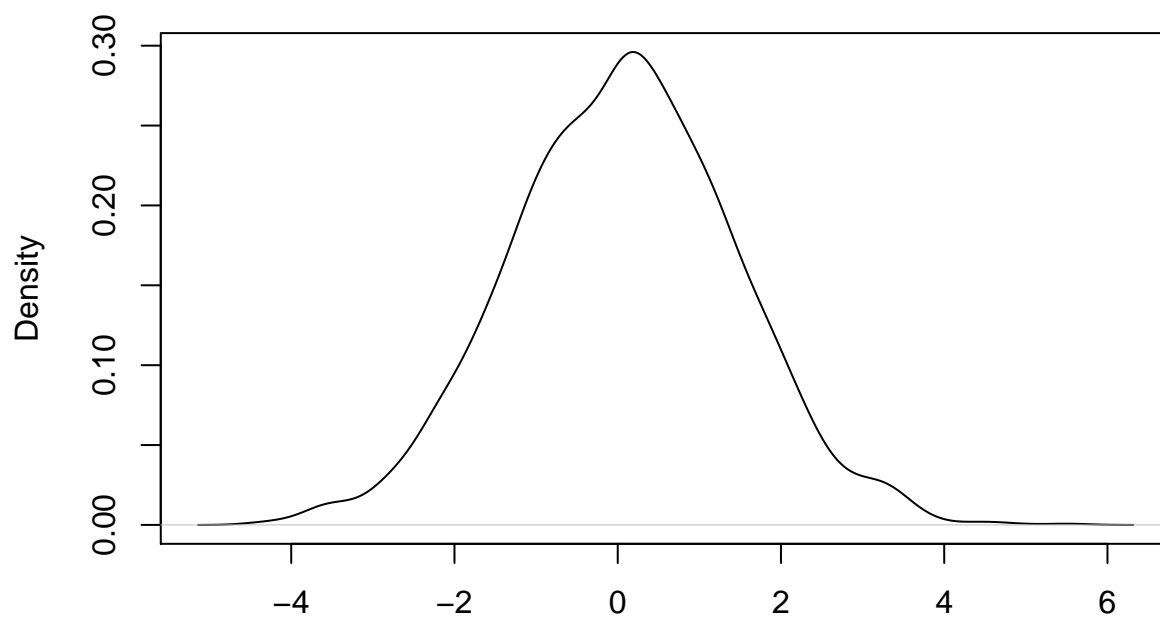N = 1000   Bandwidth = 0.3214

betahat_6

N = 1000   Bandwidth = 0.3167
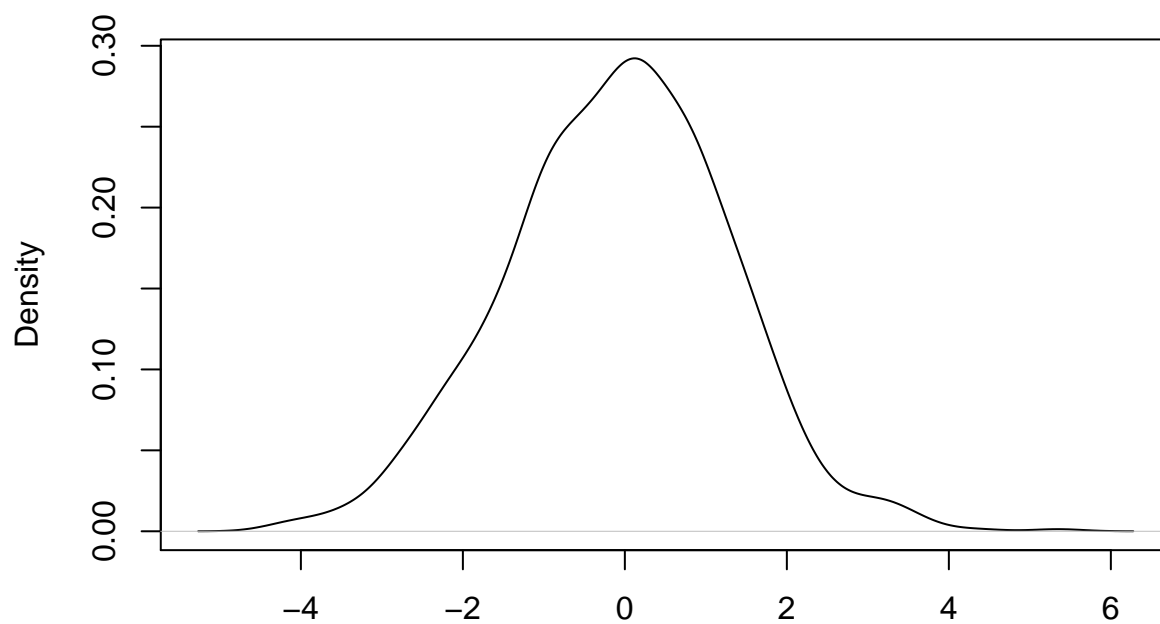
betahat_7

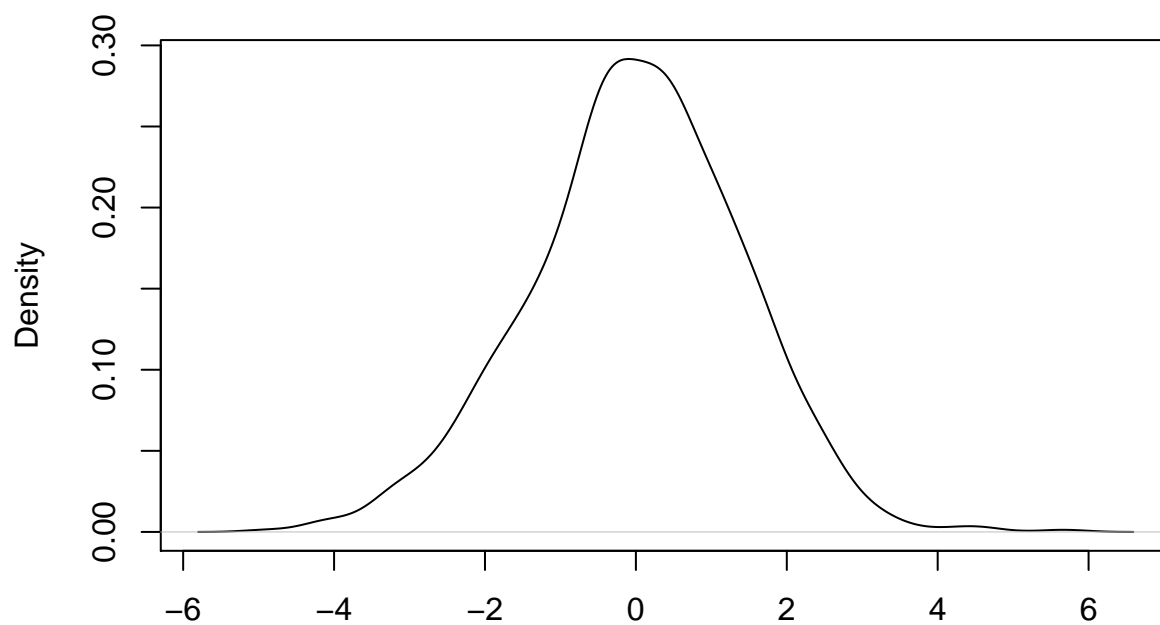N = 1000   Bandwidth = 0.301

betahat_8

N = 2000   Bandwidth = 0.27

betahat_9
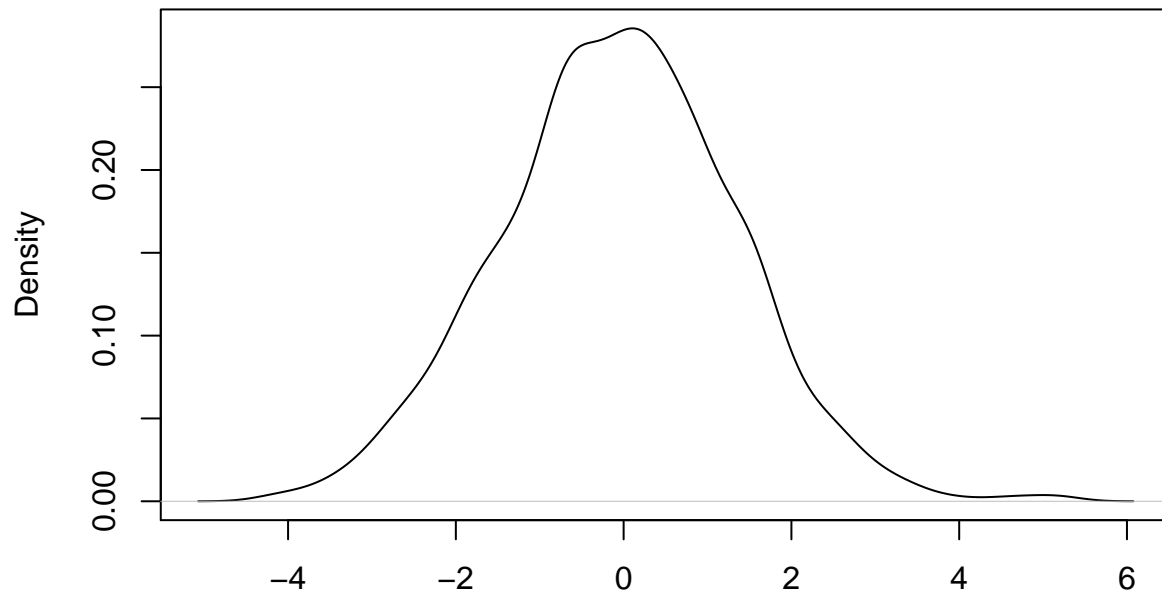
Density

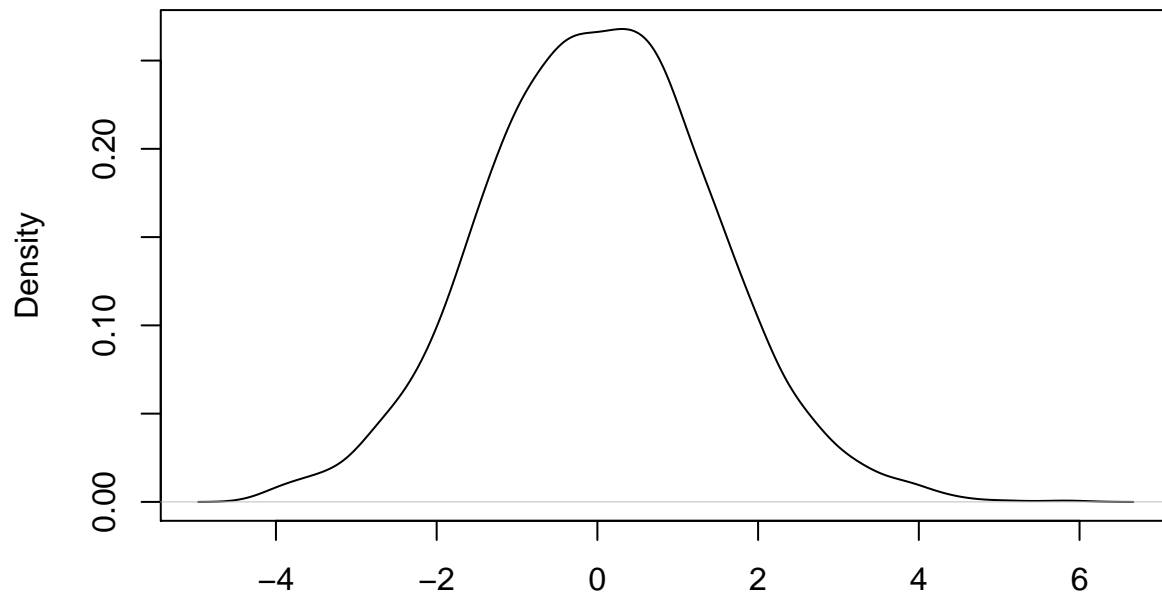N = 1000   Bandwidth = 0.3092

betahat_11

Density

N = 1000   Bandwidth = 0.3092

betahat_12

N = 2000   Bandwidth = 0.2668

betahat_13

N = 2000   Bandwidth = 0.28

They don't seem very suspicious, but they tend to be non-unimodal, non-symmetric, and/or non-zero-mode. Of them, the third data set is of particular interest. This data set consists of Group 1 and Group 2. That is, the warning happens when we feed both groups to ASH jointly, but not when we feed them separately. Now we plot the density of the joint data set and the separate data sets separately.

| betahat_3 | betahat_3_1 | betahat_3_2 |

N = 2000   Bandwidth = 0.2669   N = 1000   Bandwidth = 0.3151   N = 1000   Bandwidth = 0.3038
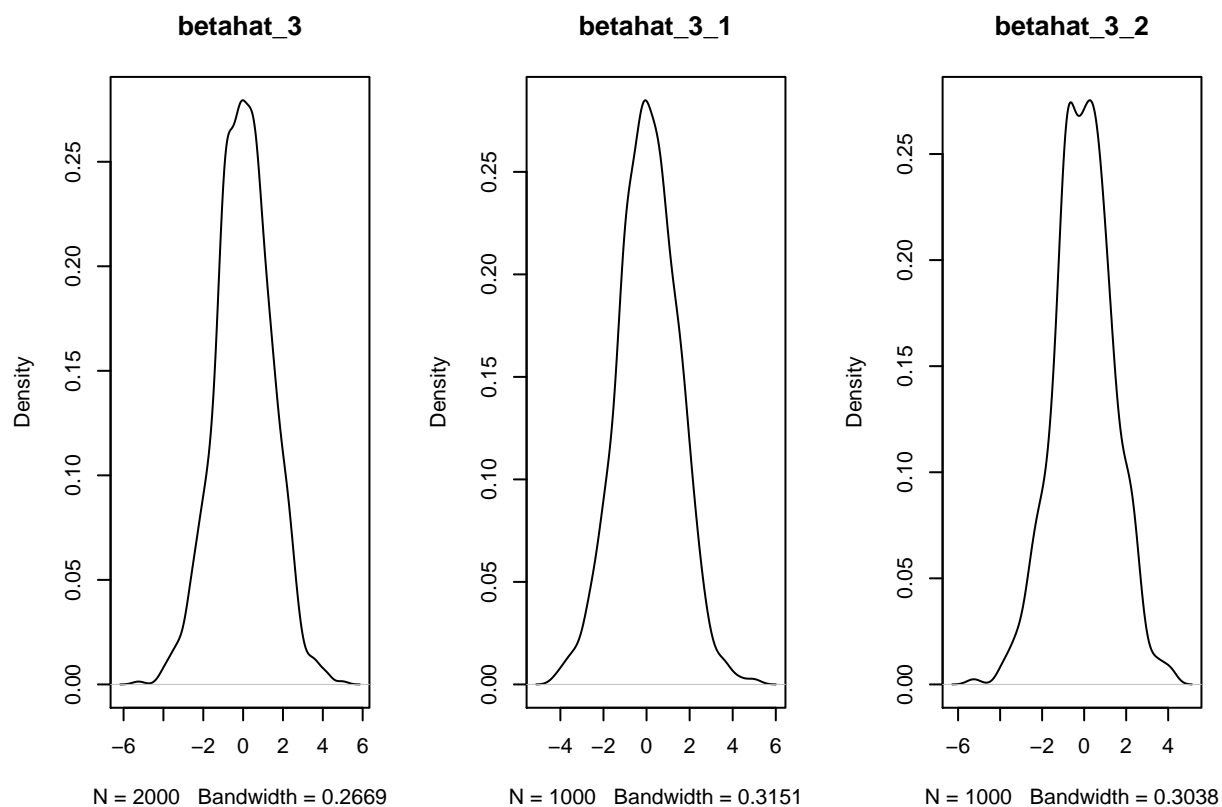
We can see that Group 2 seems more malign yet EM works with it fine. It seems that compared with non-unimodal, non-symmetry is more likely to make EM not converge.