R Notebook

First we simulate data according to MAMBA.

```r
# fitting simulated MAMBA data to model
if (file.exists("../data/mamba_data.rda") &&
    file.exists("../data/sim_mamba_mod.rda") ) {

  load(file = "../data/mamba_data.rda")
  load(file = "../data/sim_mamba_mod.rda")

  snpeffect <- mamba_data$betajk
  snpvar <- mamba_data$sjk2

} else {
  # simulate data with MAMBA
  mamba_data <- generate_data_mamba()
  save(mamba_data, file = "../data/mamba_data.rda")

  # save effects and variance
  snpeffect <- mamba_data$betajk
  snpvar <- mamba_data$sjk2

  # fit mamba model
  sim_mod <- mamba(betajk = snpeffect, sjk2 = snpvar)
  save(sim_mod, file = "../data/sim_mamba_mod.rda")

}

# take values from model
pprs <- sim_mod$ppr
effectsize <- sim_mod$mu.hat
outliers <- sim_mod$outliermat[ppr < 0.5]
```

```r
# fitting with prp library
post_prp <- posterior_prp(beta = snpeffect[1, ],
              se = sqrt(snpvar[1, ]))
post_prp$pvalue
```

```
## [1] 0.189
```

```r
# showing only first ten rows
# first ten rows of data
snpeffect1 <- snpeffect[1:10, ]
snpvar1 <-snpvar[1:10, ]
```

```r
# convert matrix to list of doubles, each element is a row in the dataset
test_vec_eff = lapply(seq_len(nrow(snpeffect1)),
                      function(x) {
                        snpeffect1[x,]
                      })
test_vec_se = lapply(seq_len(nrow(snpvar1)),
                      function(x) {
                        sqrt(snpvar1[x,])
                      })

list1 = t(mapply(posterior_prp,
                  beta = test_vec_eff,
                  se = test_vec_se))
```

```r
# indices for snps w/ and w/o outliers, first 500 snps
out_studies <- mamba_data$Ojk
out_rows_ind <- which(rowSums(out_studies == 0) > 0) # indices of snps with outliers
no_out_rows_ind <-which(rowSums(out_studies) == 10) # indices of rows w/o outliers
```

```r
# loading data from r scripts
load(file = "../data/post_prp_data.rda") # post_prp_data
load(file = "../data/post_prp_data_pval.rda") # post_prp_data_pval
post_prp_data_ch <- read.csv(file = "../data/post_prp_data.csv") # post_prp_data_ch
```

```r
# pprs for snps w/ and w/o outliers
mean(pprs[out_rows_ind])
```

```
## [1] 0.009836762
```

```r
mean(pprs[no_out_rows_ind])
```

```
## [1] 0.009037079
```

```r
# prps for snps w/ and w/o outliers
median(post_prp_data_pval[out_rows_ind])
```

```
## [1] 0.449
```

```r
median(post_prp_data_pval[no_out_rows_ind])
```

```
## [1] 0.449
```

Different Population Sampling

We sample from 5000 individuals. For each "study," we sample at different allele frequencies.

```
# allele sampling
n = 5000
allele0.01 <- rbinom(n, 2, 0.01)
allele0.1 <- rbinom(n, 2, 0.1)
allele0.25 <- rbinom(n, 2, 0.25)

# noise
epsilon <- rnorm(5000, mean = 0, sd = 1)
```

Using a linear equation, we get our phenotype values.

```
beta = rnorm(1, mean = 0, sd = 1)
y0.01 = (beta * allele0.01) + epsilon
y0.1 = (beta * allele0.1) + epsilon
y0.25 = (beta * allele0.25) + epsilon
```

Finally we fit a linear model to get an effect size estimate.

```
lm_model0.01 <- lm(y0.01 ~ allele0.01)
bhat0.01 <- lm_model0.01$coefficients[2]
se0.01 <- summary(lm_model0.01)$sigma

lm_model0.1 <- lm(y0.1 ~ allele0.1)
bhat0.1 <- lm_model0.1$coefficients[2]
se0.1 <- summary(lm_model0.1)$sigma

lm_model0.25 <- lm(y0.25 ~ allele0.25)
bhat0.25 <- lm_model0.25$coefficients[2]
se0.25 <- summary(lm_model0.25)$sigma
```

The effect size estimates are given in the table below. Our true effect size is `beta`.

allele0.01	allele0.1	allele0.25
-1.074	-0.9315	-1.011

Their respective standard errors are given in the table below.

0.9816, *0.9814* and *0.9816*

Using these predictions, we can see how MAMBA and the PRP library behave. We treat each allele frequency as its own study, and so we compare the three studies with different allele frequencies to each other.

```
# list of our (predicted) effect size and variance
snpeffect <- c(bhat0.01, bhat0.1, bhat0.25)
snpeffect_prp <- rbind(snpeffect, snpeffect)
snpvar <- c(se0.01**2, se0.1**2, se0.25**2)
snpvar_prp <- rbind(snpvar, snpvar)
```

First we fit these results to MAMBA.

```
diff_pop <- mamba(beta = snpeffect_prp,
                  sjk2 = snpvar_prp)
```

```
diff_pop$outliermat
```

```
##    snp      Oj_1      Oj_2      Oj_3        ppr
## 1:   1 0.03322111 0.03295981 0.03310119 0.005052808
## 2:   2 0.03322111 0.03295981 0.03310119 0.005052808
```

Next we fit to the PRP model.

```
diff_pop_prp <- posterior_prp(beta = snpeffect,
                              se = sqrt(snpvar))
diff_pop_prp$pvalue
```

```
## [1] 0.797
```