Replication Simulation

Common Ground

We start by generating data using the MAMBA package.

```r
# simulate data with MAMBA package
mamba_data <- generate_data_mamba()

# effect sizes and variance
snpeffect <- mamba_data$betajk
snpvar <- mamba_data$sjk2
```

Then we fit the simulated data to the MAMBA model.

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

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

} else {

  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]
```

For reference, a table of the first ten SNPS and the probabilities of them being replicable is given below.

Table 1: Table continues below

snp	Oj_1	Oj_2	Oj_3	Oj_4	Oj_5	Oj_6	Oj_7
1	0.01868	0.02338	0.0202	0.01186	0.01166	0.01173	0.01261
2	0.01159	0.01176	0.01755	0.01492	0.01785	0.0116	0.0192
3	0.01702	0.01615	0.02142	0.0121	0.01167	0.01181	0.06507
4	0.01995	0.0124	0.01484	0.0221	0.03821	0.01404	0.01736
5	0.4869	0.01471	0.0245	0.02256	0.01355	0.01158	0.01163
6	0.01271	0.07303	0.0174	0.02239	0.01342	0.01584	0.0125
7	0.01674	0.01275	0.02382	0.02488	0.01229	0.03706	0.01189
8	0.01847	0.03943	0.03137	0.01272	0.07773	0.02679	0.01167
9	0.2277	0.01167	0.01473	0.0186	0.01187	0.01966	0.01253
10	0.01304	0.01809	0.01158	0.01914	0.0252	0.02125	0.2977

Oj_8	Oj_9	Oj_10	ppr
0.02198	0.0141	0.01285	0.002303
0.239	0.01222	0.01182	0.003711
0.03443	0.02324	0.01449	0.0009735
0.01177	0.01401	0.01531	0.0012
0.0119	0.01594	0.01749	0.00614
0.01162	0.02826	0.01363	0.001531
0.01267	0.01284	0.01178	0.001068
0.05198	0.0117	0.01173	0.002062
0.01307	0.01768	0.01208	0.01035
0.05586	0.01865	0.01326	0.005077

The PRP package only looks at one SNP at a time, so for comparison we'll just look at only the first SNP The posterior-probability-of-replicability (PPR) for the first SNP is 0.0023025.

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

Our posterior predictive replication p-values (posterior PRP) for the first SNP using the PRP package is 0.763.

Quantitative/Qualitative Differences

Now we look at a few examples.

Case 1: Non-Zero Effect Sizes

In the first case, we use a fake model with non-zero effect sizes containing an extreme outliers.

```r
# high value nonzero effect size
# fake effect size
fakesnp1_1 <- c(6, 7, 6, 20)
fakesnp1_2 <- c(7, 8, 8, 9)
fakesnp1_3 <- c(7, 6, 7, 8)
fakesnpeffect1 <- rbind(fakesnp1_1, fakesnp1_2, fakesnp1_3)


# fake variance
fakesnpv1_1 <- c(3, 3, 3, 3)
fakesnpv1_2 <- c(4, 5, 6, 3)
fakesnpv1_3 <- c(4, 5, 4, 4)
fakesnpvar1 <- rbind(fakesnpv1_1, fakesnpv1_2, fakesnpv1_3)
```

```r
fakemod1 <-mamba(betajk = fakesnpeffect1,
                 sjk2 = fakesnpvar1)
```

The outlier probabilities from MAMBA are given in the table below.

snp	Oj_1	Oj_2	Oj_3	Oj_4	ppr
1	1	1	1	1	4.018e-53
2	1	1	1	1	4.723e-30
3	1	1	1	1	6.262e-27

First, with the MAMBA model, we can see that the model detected an outlier in the first SNP, given by the PPR value of $4.0182081 \times 10^{-53}$.

```r
fakemod1_post_prp <- posterior_prp(beta = fakesnpeffect1[1, ],
                                   se = sqrt(fakesnpvar1[1, ]))
```

Using the PRP package, we get a PRP value of 0.002.

Case 2: Normal Distribution, Non-Zero Effect Size SNP

In the second case, we generate a normal distribution with mean = 0 and standard deviation = 1, with one SNP having a non-zero effect size

```
# generating normal data
fakesnp_norm2_1 <- rnorm(10, mean= 10, sd = 2)
fakesnp_norm2_2 <- rnorm(10, mean= 0, sd = 2)
fakesnp_norm2_3 <- rnorm(10, mean = 0, sd = 1)
fakesnpeffect2 <- rbind(fakesnp_norm2_1,
                        fakesnp_norm2_2,
                        fakesnp_norm2_3)

fakesnp_norm_var2_1 <- rnorm(10, mean = 2, sd = 0.1)
fakesnp_norm_var2_2 <- rnorm(10, mean = 1, sd = 0.001)
fakesnp_norm_var2_3 <- rnorm(10, mean = 1, sd = 0.001)
fakesnpvar2 <- rbind(fakesnp_norm_var2_1,
                     fakesnp_norm_var2_2,
                     fakesnp_norm_var2_3)
```

```
fakemod_mamba2 <- mamba(betajk = fakesnpeffect2,
                        sjk2 = fakesnpvar2)
```

The outlier probabilities from MAMBA are given in the table below.

Table 4: Table continues below

snp	Oj_1	Oj_2	Oj_3	Oj_4	Oj_5	Oj_6	Oj_7	Oj_8
1	1	1	1	1	1	1	1	1
2	0.9998	0.9575	0.9965	0.9926	0.995	0.9977	0.9982	0.9676
3	0.9559	0.9528	0.9527	0.9571	0.9633	0.954	0.9531	0.9548

Oj_9	Oj_10	ppr
1	1	4.075e-123
0.9637	1	3.044e-06
0.955	0.9624	1

The PPR value for the first SNP is given by $4.0747326 \times 10^{-123}$

```
norm_prp_mod <- posterior_prp(beta = fakesnpeffect2[3 ],
                              se = sqrt(fakesnpvar2[3, ]))
norm_prp_mod$pvalue
```

```
## [1] 0.775
```

In contrast, the PPR value for the same SNP is 0.775.

Case 3: Normal Distribution, Zero Effect Size

In the third case, we look at SNPs with zero effect size but with differing concordant signs.

```
# generating normal data centered around zero
fakesnp_norm3_1 <- rnorm(10, mean= 0, sd = 2)
fakesnp_norm3_2 <- rnorm(10, mean= 0, sd = 1)
fakesnp_norm3_3 <- rnorm(10, mean = 0, sd = 1)
fakesnpeffect3 <- rbind(fakesnp_norm3_1,
                        fakesnp_norm3_2,
                        fakesnp_norm3_3)

fakesnp_norm_var3_1 <- rnorm(10, mean = 2, sd = 0.1)
fakesnp_norm_var3_2 <- rnorm(10, mean = 1, sd = 0.1)
fakesnp_norm_var3_3 <- rnorm(10, mean = 1, sd = 0.001)
fakesnpvar3 <- rbind(fakesnp_norm_var3_1,
                     fakesnp_norm_var3_2,
                     fakesnp_norm_var3_3)
```

```
# fitting MAMBA model
fakemod_norm3 <- mamba(betajk = fakesnpeffect3,
                       sjk2 = fakesnpvar3)
```

The outlier probabilities from MAMBA are given in the table below.

Table 6: Table continues below

snp	Oj_1	Oj_2	Oj_3	Oj_4	Oj_5	Oj_6	Oj_7
1	0.08009	0.06752	0.08215	0.06481	0.06407	0.07543	0.06273
2	0.06572	0.4541	0.1838	0.06296	0.08717	0.123	0.1237
3	0.09303	0.2129	0.06248	0.06316	0.06539	0.1129	0.0663

Oj_8	Oj_9	Oj_10	ppr
0.06657	0.0752	0.06437	0.3529
0.06614	0.0657	0.07331	0.9816
0.06155	0.06524	0.08885	0.2512

The PPR value for the first SNP is given by 0.3529286.

```
# fitting posterior prp
norm_prp_mod3 <- posterior_prp(beta = fakesnpeffect3[1, ],
                               se = sqrt(fakesnpvar3[1, ]))
```

The posterior PRP value given with these SNPs is 0.707.

Different Population Sampling

First we sample from 5000 individuals. We sample at different allele frequencies.

```r
# 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 = 0.05)
```

Using a linear equation, we get our phenotype values.

```r
beta = 0.005
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.

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

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

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

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

allele0.01	allele0.1	allele0.25
0.00103	0.008012	0.004414

Using these predictions, we can see how MAMBA and the PRP library behave.

```r
# list of our (predicted) effect size and variance
snpeffect <- c(bhat0.01, bhat0.1, bhat0.25)
fakesnp <- c(bhat0.01, bhat0.1, bhat0.25)
snpeffect_prp <- rbind(snpeffect, fakesnp)
snpvar <- c(var(y0.01), var(y0.1), var(y0.25))
snpvar_prp <- rbind(snpvar, snpvar)

# I doubled the values here because MAMBA seems to need more than one snp for it to work
```

First we fit to MAMBA.

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

```
diff_pop$outliermat
```

```
##   snp Oj_1 Oj_2 Oj_3          ppr
## 1:   1    1    1    1 1.720581e-30
## 2:   2    1    1    1 1.720581e-30
```

Next we fit to the PRP model.

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

Using MAMBA, the PPR is $1.7205812 \times 10^{-30}$. In contrast, using the PRP library our PRP value is 0.947.