

Population Genetics Homework - Week 2

Christian Polania

9/5/2021

Question 1

```
dat <- read.structure("http://darwin.eeb.uconn.edu/eeb348-resources/repens-outliers.stru", n.ind=662, n
dat <- genind2hierfstat(dat)
ci <- boot.vc(dat[, 1], dat[, -1], diploid = TRUE)$ci[,c(2,5)]
colnames(ci) <- c("Fst", "Fis")
wc(dat)
```

```
## $FST
## [1] 0.288204
##
## $FIS
## [1] 0.2155412
```

```
ci
```

```
##           Fst    Fis
## 2.5%  0.2740 0.1783
## 50%   0.2883 0.2152
## 97.5% 0.3049 0.2582
```

95% of hierfstat's estimations of F_{st} were between 0.273 and 0.304, with a mean at 0.288. The corresponding estimations of F_{is} were between 0.176 and 0.255, with a mean of 0.214.

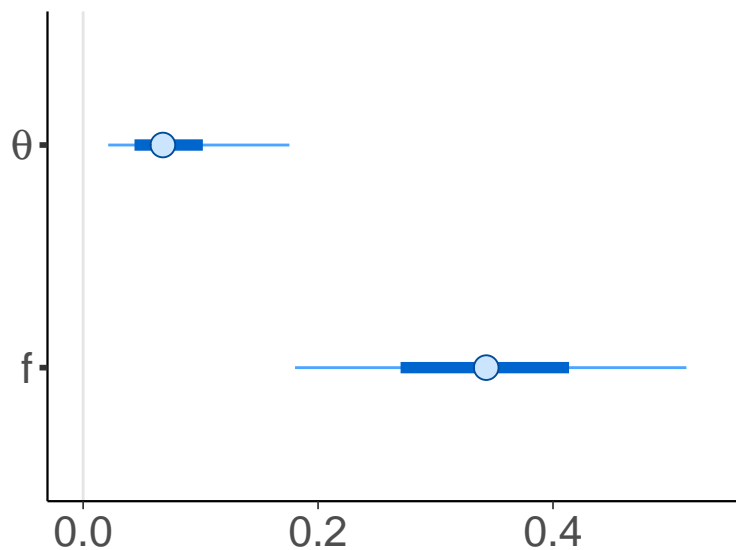
Question 2

```
options(mc.cores = parallel::detectCores())
genos <- read_marker_data("http://darwin.eeb.uconn.edu/eeb348-resources/isotoma.csv")
fit <- analyze_codominant(genos, refresh = 0)
```

```
## Warning: Tail Effective Samples Size (ESS) is too low, indicating posterior variances and tail quant
## Running the chains for more iterations may help. See
## http://mc-stan.org/misc/warnings.html#tail-ess
```

```
## Inference for Stan model: analyze_codominant.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
```

```
##
##          mean se_mean    sd    2.5%    25%    50%    75%    97.5% n_eff
## f          0.343   0.002 0.103   0.143   0.270   0.343   0.414   0.547  4030
## theta      0.079   0.002 0.051   0.017   0.044   0.068   0.102   0.215   452
## lp__ -121.915   0.260 4.592 -132.714 -124.401 -121.271 -118.742 -114.698   313
##          Rhat
## f          1.001
## theta      1.007
## lp__      1.011
##
## Samples were drawn using NUTS(diag_e) at Fri Sep 10 21:17:57 2021.
## 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).
```



95% of Hickory's estimations of F_{st} were between 0.017 and 0.186, with a mean at 0.073. The corresponding estimations of F_{is} were between 0.152 and 0.546, with a mean of 0.347.

Question 3 The high F_{is} value (>0.3) implies that, on average, inbreeding is common within subpopulations.

Question 4 The low F_{st} value (<0.1) from Hickory implies that there is little genetic differentiation (and likely little gene flow) between subpopulations.

Question 5 F_{is} didn't change much between hierfstat and Hickory, whereas F_{st} did significantly. Hickory's interpretation implies little population structuring, whereas hierfstat's implies genetic differentiation to the point of reproductive isolation.

Question 6 I played around with each of the settings for `analyze_codominant()` and even after substituting some values that seemed unrealistic, I couldn't get F_{st} to budge past 0.1, still much less than hierfstat's. I would have thought that the reason for the difference was some assumption that Hickory made that hierfstat didn't, but now I have no idea! Looking forward to hearing the right answer.