Additive Alleles – Addendums

Silas Tittes

August 22, 2015

Since my last post, I've learned about the option to add figure captions to plots (figure 1) in Rmarkdown – fancy! I also realized one reason why the frequency of phenotypic classes didn't seem more similar to the densities taken from the normal distribution with the same parameters. Here's where I left off (typos and all):

My realization came from plotting the density values from the normal distribution with the parameters from the data against the actual frequencies (figure 2) – like so:

```
plot(den.theor, out$class.freqs,
    pch = 19)
abline(0, 1, lwd = 2, col = "red")
```

The cool thing is, the density and actual values seem to have a linear relationship (figure 2), but it's just not one-to-one. Time for a regression!

```
mod <- lm(out$class.freqs ~ den.theor)
round(mod$coefficients)</pre>
```

```
## (Intercept) den.theor
## 0 19
```

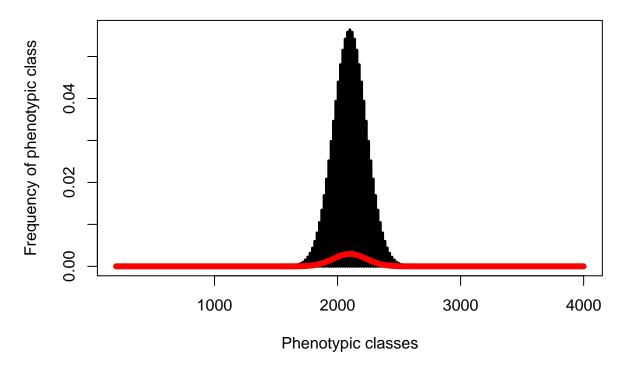


Figure 1: A trip down memory lane (my last post). The distribution of phenotypic classes that result from the additive contributions of 100 loci. The red line shows the density from normal distribution with the same mean and standard deviation as the data.

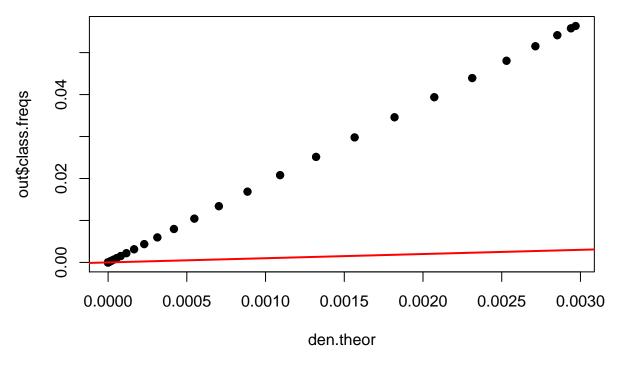


Figure 2: Density of normal distribution with mean and standard deviation taken from the phenotypic classes, against the acutal frequency of the phenotypic classes. Line shows 1:1 relationship I was hoping for.

After a bit of farting around, it became clear that the slope of the regression line was the difference between the allele contributions! Try it yourself if you don't believe me. I was rather excited to learn this, and it seems to hold true for any case I try (I call this a "pudding proof" because it's not a mathematical proof, but the proof is in there somewhere – just keep eating. I also like pudding quite a bit).

```
n.loci <- 100 #choose the number of loci
allele.types <- c(10, 5) #choose the allelic contributions
allele.diff <- abs(allele.types[2] -</pre>
    allele.types[1])
out <- pheno.dist(n.loci, allele.types)</pre>
# plot
plot(out$class.types, out$class.freqs,
    type = "h", lwd = 3, col = "grey50",
    ylab = "Frequency of phenotypic class",
    xlab = "Phenotypic classes")
# See what a normal
# distribution looks like on
# top of the data
avg <- sum(out$class.types * out$class.freqs)</pre>
stdev <- sqrt(sum(((out$class.types -</pre>
    avg)^2 * out$class.freqs)))
den.theor <- dnorm(out$class.types,</pre>
    mean = avg, sd = stdev)
mod <- lm(out$class.freqs ~ den.theor)</pre>
scaler <- allele.types[1] - allele.types[2]</pre>
lines(out$class.types, den.theor *
    scaler, col = "red", lwd = 6)
```

```
plot(den.theor * scaler, out$class.freqs,
    pch = 19)
abline(0, 1, lwd = 2, col = "red")
```

This looks mighty similar, but the plot thickens when we look at the residuals (figure 5)!

```
plot(out$class.freqs - den.theor *
    scaler, type = "b", cex = 0.3,
    ylab = "class freqs. - (density of normal * (Allele1 - Allele2))")
```

I haven't spent very long thinking why scaling the density by the difference in allele values makes them more closely match the phenotypic classes, or why the phenotypic classes aren't

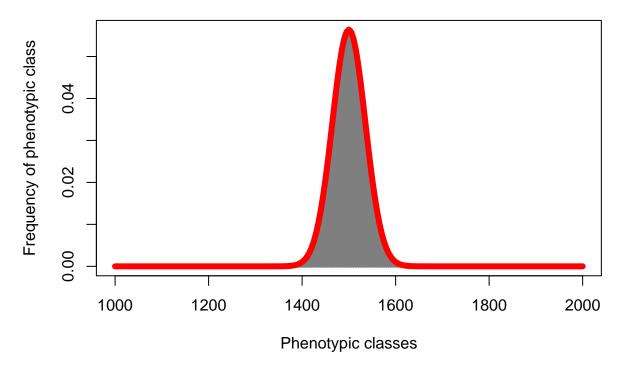


Figure 3: A new lane for new memories. The distribution of phenotypic classes that result from the additive contributions of 100 loci. The red line shows the density from normal distribution with the same mean and standard deviation as the data.

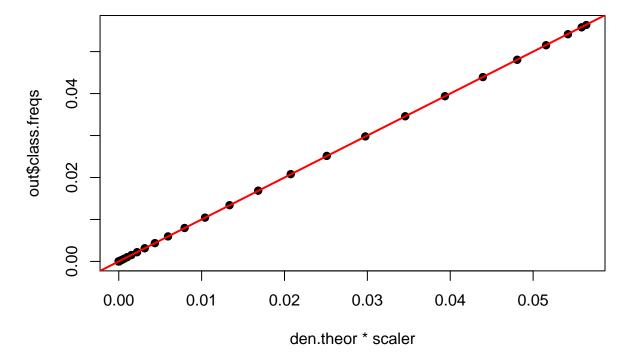


Figure 4: Density of normal distribution with mean and standard deviation taken from the phenotypic classes, against the actual frequency of the phenotypic classes. Line shows 1:1 relationship I was hoping for.

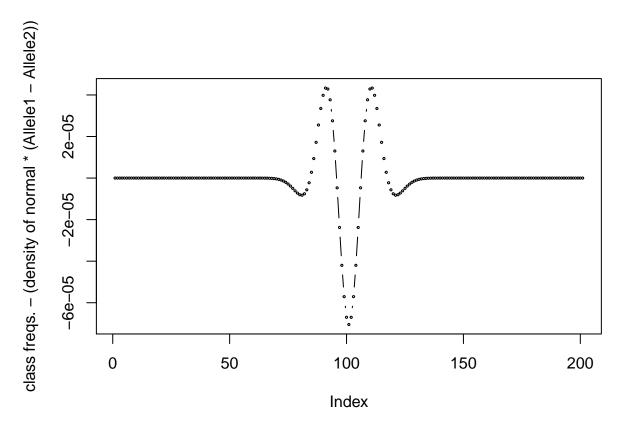


Figure 5: Difference between scaled density and the frequency phenotypic classes. Something funny is going on near the mean

quite normal. If you have any thoughts, email me! My attempt at an addendum has solved one question and made several new ones. I suppose I have heard that the distribution of phenotypic classes that results from the contribution of multiple additive alleles is *approximated* by the normal distribution. As I'll explore in another post, the term approximate becomes more important as the contributions from each allele begin to differ.

As with most things, if I spent a little time reading, I could probably find the exact form of the mean and variance, but it wouldn't be nearly as fun as struggling through it on my own. Perhaps a second addendum at a future data is in order. Toddler steps.