

Homework 5

Zachary Foster

11/06/2014

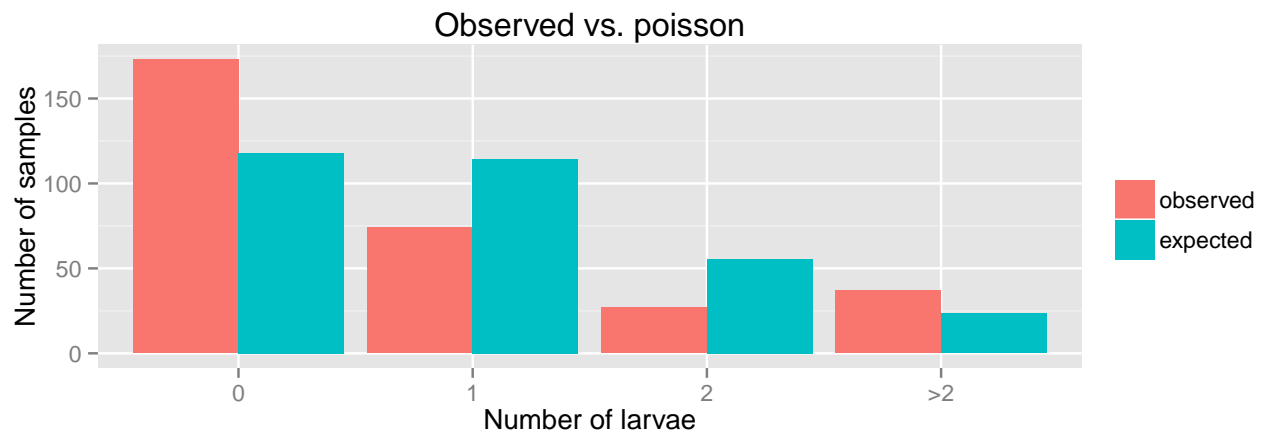
Preparation of input data

```
larvae <- 0:12
counts <- c(173, 74, 27, 14, 9, 5, 4, 2, 0, 0, 1, 1, 1)
data <- rep(larvae, counts)
```

Goodness of fit of the Poisson distribution

```
# Generate poisson distribution based on mean larvae count -----
poisson <- dpois(larvae, lambda = mean(data))
poisson <- c(poisson, 1 - sum(poisson))
poisson <- poisson * sum(counts)
# Compare distributions -----
compare_dist <- function(expected) {
  # Find which values to lump so that there are at least 2 -----
  lump_at <- sum(sapply(seq_along(expected),
    function(i) sum(expected[i:length(expected)]) > 2)) - 1
  lumped <- lump_at:length(expected)
  # Create data frame with observed and expected -----
  comp <- data.frame("observed" = c(c(counts, 0)[-lumped], sum(c(counts, 0)[lumped])),
    "expected" = c(expected[-lumped], sum(expected[lumped])),
    "count" = c(larvae[-lumped], paste0(">", larvae[lump_at - 1])))
  # Do a Chi-squared test comparing the distributions -----
  p_value <- 1 - pchisq(sum((comp$observed-comp$expected)^2/comp$expected), 2)
  print(paste("Chi-squared p-value:", signif(p_value)))
  return(comp)
}
comp <- compare_dist(poisson)
```

```
## [1] "Chi-squared p-value: 2.32037e-14"
```

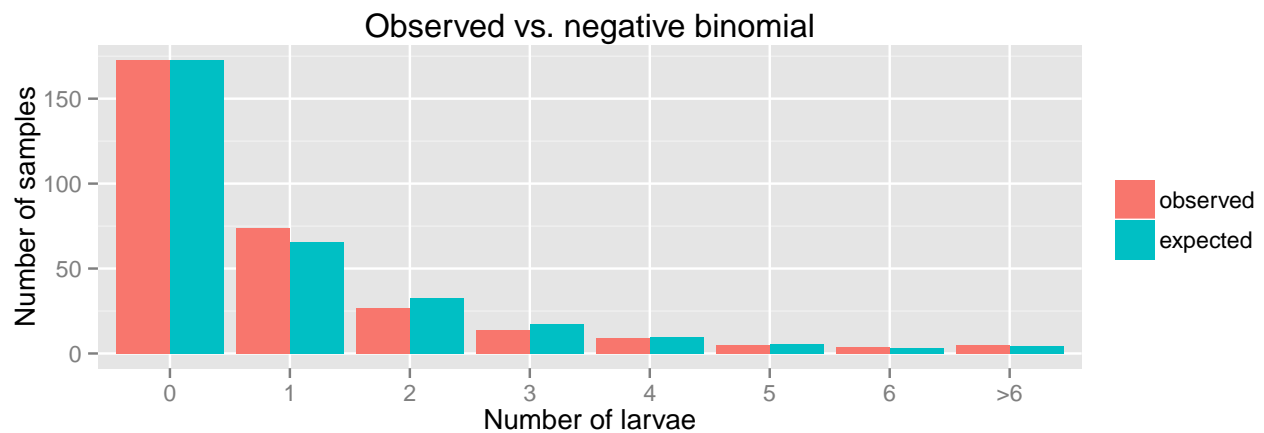


Judging from both visual inspection of the observed distribution compared to the predicted poisson distribution as well as the chi-squared test results ($p = 2.32 \cdot 10^{-14}$), I believe that the poisson is not a good fit for the observed data.

Goodness of fit of the negative binomial distribution

```
# Estimate k from number of zeros -----
nbinome_k_estimate <- function(data) {
  # Find the k that makes the following function equal to zero -----
  f <- function(k, N = length(data), NO = sum(data == 0), xbar = mean(data))
    log(N/NO) - k * log(1 + xbar/k)
  uniroot(f=f, interval=c(0.001, 100))$root
}
k_hat <- nbinome_k_estimate(data)
# Generate negative binomial distribution -----
nbinome <- dnbinom(x = larvae, size = k_hat, prob = k_hat/(k_hat + mean(data)))
nbinome <- c(nbinome, 1 - sum(nbinome))
nbinome <- nbinome * sum(counts)
# Compare distributions -----
comp <- compare_dist(nbinome)
```

```
## [1] "Chi-squared p-value: 0.214412"
```



The negative binomial seems a much better fit to the observed data than the poisson. The values look much more similar to the observed values and the chi-squared test for a difference between the two distributions was insignificant ($p = 0.21$), indicating that there is no reason to exclude the negative binomial as a potential model. Without considering other potential models, the negative binomial is acceptable.