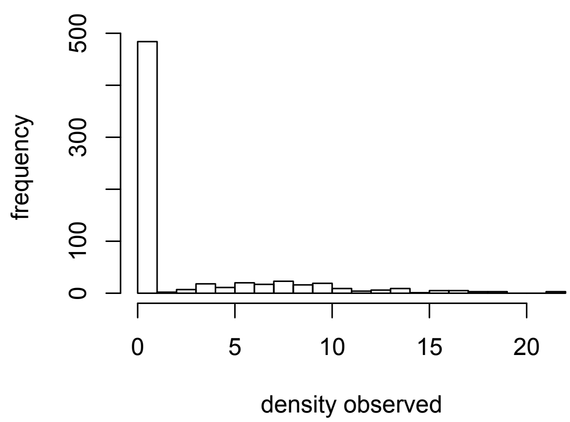
Week 9: Resampling and empirical estimation

For the hierarchical tadpole problem shown in lecture, there is a solved beta-binomial density formula, so you could have used that and avoided the ugly simulation. But now for a problem you don't have a closed probability expression for.

The data in `willowtitn.csv` are counts of observed Willow Tits (*Parus montanus*, aka *Poecile montanus*, it is a very much like the North American Chickadee) seen during transect walks at a number of different sites. The column `n` is the number seen on a given walk, `elev` is the elevation (in meters) of the site, and `percforest` is the forest coverage of the site.

Your mission is to model the abundance counts, plotted below. The working hypothesis is that the over-abundance of zeros in the data are due to a two-stage model. First, the observer either sees a group of birds or does not. This is a binomial process. Second, there is a Poisson distribution of bird densities, once they are found. So there are two parameters in this process: the probability of finding a flock (call this ϕ) and the mean of the Poisson distribution of flock sizes (call this λ). The resulting distribution is known as a *zero-inflated Poisson distribution*.



(a) Simulate data from your hypothesized two-stage bird observation process. Plot these simulated distributions of counts, for various values of ϕ and λ. Can you find reasonable starting values for these parameters, by comparing your simulations to the plotted data? Better yet, can you guess good starting values, just with simple calculations on the raw data?

(b) Write a density function—call it `dsimzeropois`—to simulate data from your hypothesized sampling process and return empirically estimated likelihoods of the observed data. Use the beta-binomial example from lecture as a guide to the coding strategy. You will, however, need to think through the probability model correctly. Once you have your density function, estimate the model on the data, using `mle2`. Remember to use `method="SANN"`.

NB: You will need to constrain λ of the Poisson to positive reals, or the estimation may not go well. Try doing this with a transformation you have already seen in the class, by using `lambda=exp(lambda)`. Similarly constrain ϕ by using a logit function. Also, it may help to add the parameter `skip.hessian=TRUE` to your `mle2` call. All this does is avoid a complex calculation that is unnecessary in our case. (Like George Washington did, we still have problems with Hessians.)

(c) Compare the fit of the model from (b) to a pure Poisson model, estimated using the `dpois` function. Interpret the coefficients of both models, and explain how the zero-inflated distribution changes inferences about the population size and structure.

(d) SUPER BONUS PROBLEM: Can you write an analytical density function that computes the likelihoods for the zero-inflated Poisson, but does not use any random number functions? You can verify that your analytical function—call it `dzeropois`—and your simulation function yield the (nearly) same estimated likelihoods with the following code:

```
kphi <- 0.5
kl <- 3
l <- sapply( 0:30 , function(z) dzeropois(z,phi=kphi,lambda=kl) )
plot( 0:30 , l , type="l" )
ls <- sapply( 0:30 , function(z) dsimzeropois(z,phi=kphi,lambda=kl,R=9999) )
lines( 0:30 , ls , lty=2 , col="red" )
```