**Ecol 145 Assignment 2**
Dahl Winters
2/3/06

Questions

(The R code can be found at the end of the homework.)

1. *Using the full data set for the 48 different reefs evaluate the mean-variance relationship for the variable PREV_1. Parallel what we did in class except use more categories than we were able to use. Deciles might be a good place to start. Produce a plot that compares a Poisson, NB1, and NB2 model for the mean-variance relationship.*

<table>
<tr><td align="center">10 categories</td><td align="center">8 categories</td></tr>
</table>
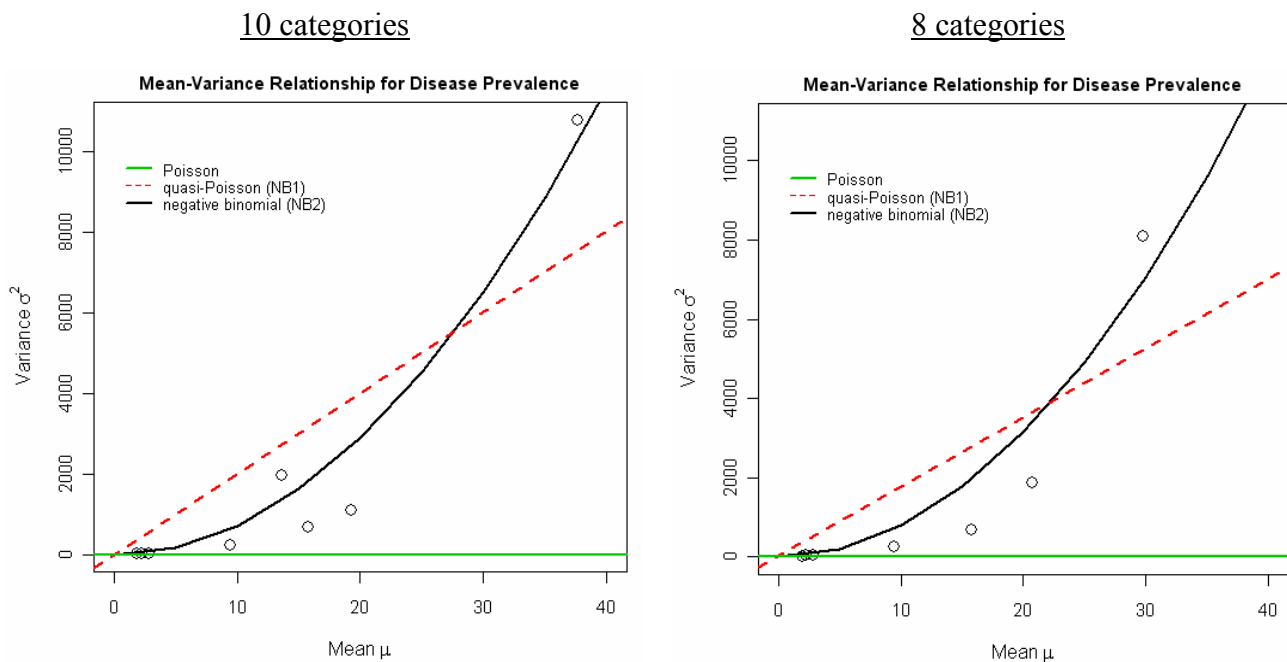
First I tried using deciles to evaluate the mean-variance relationship (left), which is necessary to determine which probability model best describes how disease occurrence (PREV_1) correlates with water temperature (WSSTA). The NB2 model visually appears to be the best model describing the mean-variance relationship. The NB1 model looks worse, and the Poisson is by far the worst (since the variance is constrained to equal the mean, but the variance increases sharply with only small increases in the mean).

I also tried doing just 8 categories instead of 10 as suggested in Hint #3 to see how robust the results would be with a different number of categories (right). The results look virtually identical in terms of the shapes of the models fitted. However, there is just one data point above both the NB1 and NB2 curves compared to two when there were 10 categories. Qualitatively, it seems these curves are less justified by the data because there are so few categories. If we had more data, we could have more categories than just 8 or 10, and there would be more points on the above plots to justify using NB2 over NB1 as our probability model.

2. Repeat the plot we did in class for the negative binomial distribution but add a Poisson distribution with λ = 1 and another with λ = 5. Compare these to the corresponding negative binomial distributions. Remembering that the Poisson distribution corresponds to a negative binomial with an infinite value of θ, what seems to be the primary effect of decreasing the value of θ for each value of λ? Submit your answer to this question along with your plot. Modify the legend accordingly.
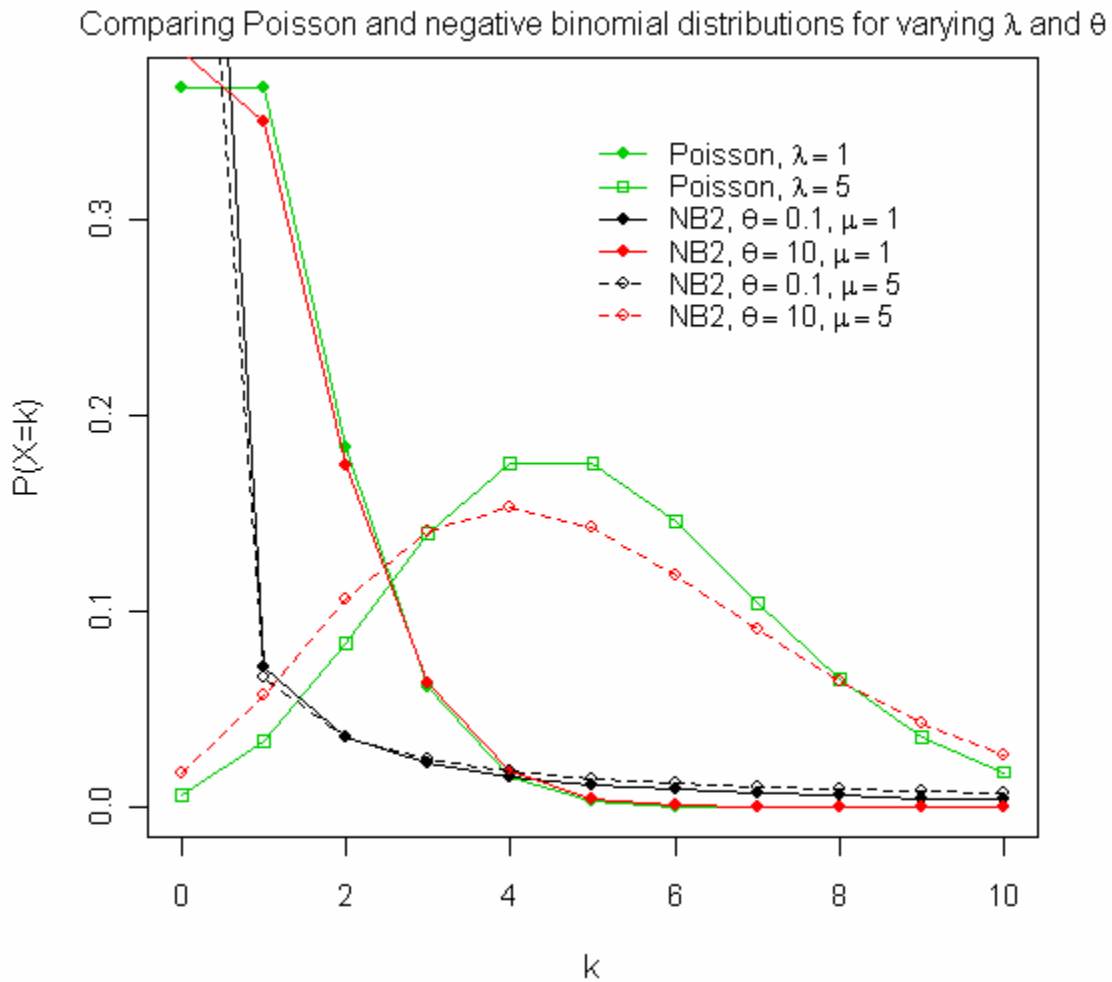
Since the Poisson distribution corresponds to a negative binomial with an infinite value of θ, we expect that the Poisson should look less like the negative binomial as θ decreases (thus, it will look more like the NB2 when θ = 10 than when θ = 0.1). This can be seen by comparing the 2 Poisson distributions (green) with the NB2 distributions that have θ = 10 (red). The Poisson with λ = 5 has some resemblance to the NB2 with θ = 10 and μ = 5, but the Poisson with λ = 1 has a strong resemblance to the NB2 with θ = 10 and μ = 1. If θ were even larger, we likely would be unable to tell the Poisson and NB2 distributions apart.

Neither of the Poisson distributions resemble the NB2 distributions that have θ = 0.1 (black), because θ is so small. For those two, we can observe that shape of the distribution depends little on the value of the mean (μ), though as k increases, there is a slightly sharper decline for μ = 1 vs. for μ = 5.

3. Produce a single plot that compares four different gamma distributions with different shape and scale parameters. The relevant function in R is dgamma. Examine the help window for dgamma and notice that the shape parameter is identical to the one used in class but that the scale parameter used in dgamma is the reciprocal of ours. To start with try α = 1, 2, and 5 with β = 1. For your fourth curve select another value of β that seems to generate an interesting distribution. What effect does the scale parameter seem to have on the distribution?
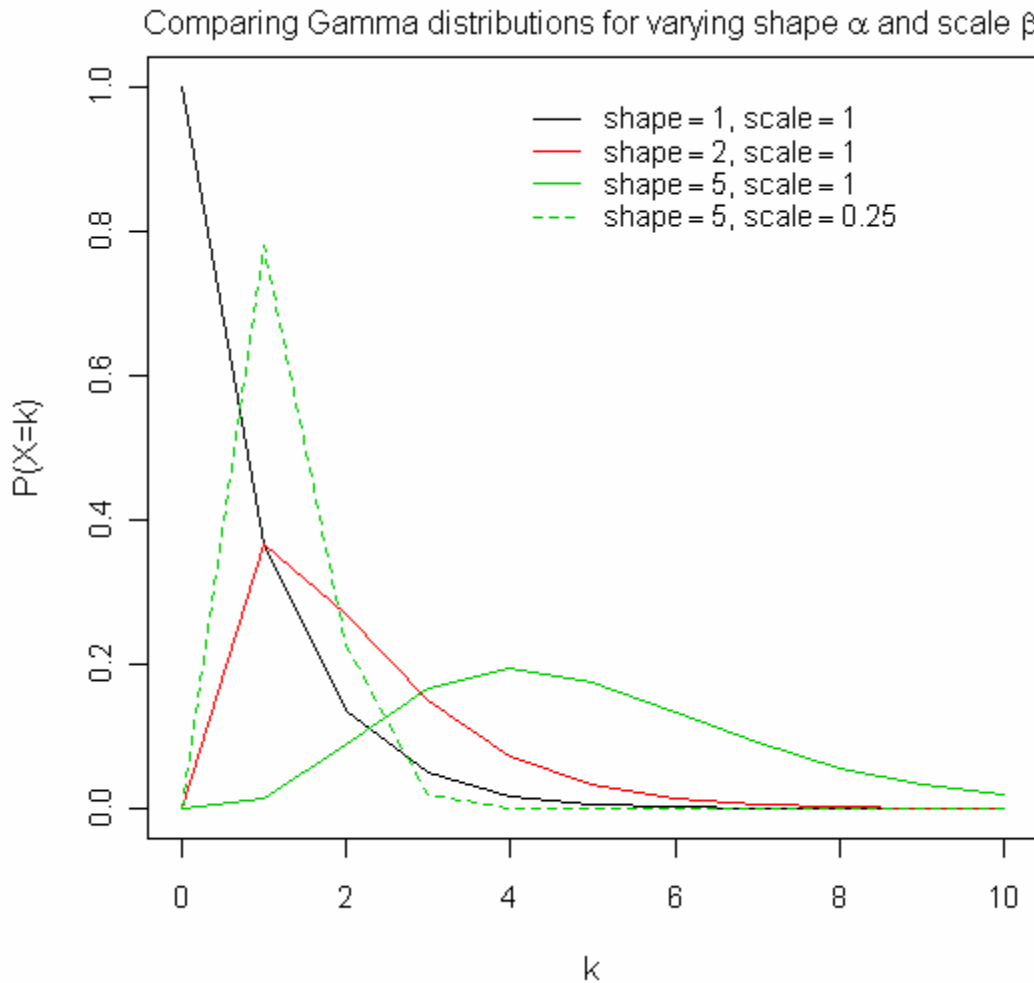
The scale parameter (β) controls the spread of the gamma distribution along the k axis, as well as the probabilities that x will equal smaller or larger values of k. As β increases, the gamma distribution is stretched farther along the k axis. This means the probabilities of x equaling smaller values of k decrease, and the probabilities of x equaling larger values of k increase. As β approaches infinity, the distribution approaches a straight line stretching to k = infinity, with infinitesimal probabilities that x equals any value of k. Decreasing β gave the most interesting shapes of distributions, I thought. When β decreases, the probabilities that x equals smaller values of k increases, and gives a peak (as shown from $0 < k < 3$).

About the shape parameter (α)…as it increases, the gamma distribution goes from resembling an exponential curve for $\alpha < 1$ to a curve having a maximum, with tails that get broader as α gets larger.

R Code

Problem 1

```
#   creating deciles with roughly equal counts in each category
> quantile(corals$WSSTA,seq(0,1,0.1))
  0%  10%  20%  30%  40%  50%  60%  70%  80%  90% 100%
   0    0    1    1    2    3    5    7   10   18   30
#The 0% and 10% deciles are repeated, as well as the 20% and 30% deciles.

> quantile(corals$WSSTA,c(.1,.3,.4,.5,.6,.7,.8,.9,1))
 10%  30%  40%  50%  60%  70%  80%  90% 100%
   0    1    2    3    5    7   10   18   30

> cut(corals$WSSTA,quantile(corals$WSSTA,c(.1,.3,.4,.5,.6,.7,.8,.9,1)),
include.lowest=TRUE)
#originally gave lots of intervals and NAs, but include.lowest=TRUE removed
the NAs.
Levels: [0,1] (1,2] (2,3] (3,5] (5,7] (7,10] (10,18] (18,30]

> table(cut(corals$WSSTA,quantile(corals$WSSTA,c(.1,.3,.4,.5,.6,.7,.8,.9,1))
,include.lowest=TRUE))
  [0,1]   (1,2]   (2,3]   (3,5]   (5,7]  (7,10] (10,18] (18,30]
     89      31      25      34      25      23      30      23

> table(cut(corals$WSSTA,quantile(corals$WSSTA,c(.1,.3,.4,.5,.6,.7,.8,.9,1)),
include.lowest=TRUE,right=FALSE))
  [0,1)   [1,2)   [2,3)   [3,5)   [5,7)  [7,10) [10,18) [18,30]
     48      41      31      42      32      25      28      33
#doing right=FALSE helped equilibrate the categories more, especially the
first category.

#   getting means and vars for each category
> WSSTA.decs<-
cut(corals$WSSTA,quantile(corals$WSSTA,c(.1,.3,.4,.5,.6,.7,.8,.9,1)),include.
lowest=TRUE,right=FALSE)
> tapply(corals$PREV_1, WSSTA.decs,mean)->means
> tapply(corals$PREV_1, WSSTA.decs,var)->vars

#   plotting vars vs. means
> plot(means, vars, axes=FALSE, xlab=expression(paste("Mean ",mu)),
ylab=expression(paste("Variance ",sigma^2)),xlim=c(0,40), cex=1.5)
> axis(1,cex.axis=.9)
> axis(2,cex.axis=.9)
> box()

#   plotting the NB2 model
> quad.coef<-coef(lm(vars~offset(means)+I(means^2)-1))
> quad.func<-function(x) x+quad.coef*x^2
> lines(seq(0,150,5), quad.func(seq(0,150,5)), col=1, lty=1, lwd=2)

#   plotting the NB1 model
> abline(0, coef(lm(vars~means-1)), col=2, lty=2, lwd=2)

#   adding the Poisson
> abline(0,1,col=3,lty=1,lwd=2)
```

```
#   adding the legend and title
> legend(0,10000, c('Poisson', 'quasi-Poisson (NB1)', 'negative binomial
(NB2)'), col=c(3,2,1), lwd=c(2,1,2), lty=c(1,2,1), cex=c(.8,.8,.8), bty='n')
> mtext('Mean-Variance Relationship for Disease Prevalence', side=3, line=.5,
font=2, cex=.9)
```

 #   trying 8 categories instead of 10
```
> quantile(corals$WSSTA,seq(0,1,0.1))
   0%  10%  20%  30%  40%  50%  60%  70%  80%  90% 100%
    0    0    1    1    2    3    5    7   10   18   30
#The 0% and 10% deciles are repeated, as well as the 20% and 30% deciles.

> quantile(corals$WSSTA,seq(0,1,0.125))
   0% 12.5%   25% 37.5%   50% 62.5%   75% 87.5%  100%
    0     0     1     2     3     5     8    17    30
#Only the 0% and 12.5% octiles are repeated, so I will use the c function to
merge those two.

> cut(corals$WSSTA,quantile(corals$WSSTA,c(.125,.25,.375,.5,.625,.75,.875,1))
,include.lowest=TRUE)
#originally gave lots of intervals and NAs, but include.lowest=TRUE removed
the NAs.
Levels: (0,1] (1,2] (2,3] (3,5] (5,8] (8,17] (17,30]

> table(cut(corals$WSSTA,quantile(corals$WSSTA,c(.125,.25,.375,.5,.625,.75,
.875,1)),include.lowest=TRUE))
  [0,1]   (1,2]   (2,3]   (3,5]   (5,8]  (8,17] (17,30]
     89      31      25      34      32      36      33
#the counts in each category are roughly the same except for the first
category.

> table(cut(corals$WSSTA,quantile(corals$WSSTA,c(.125,.25,.375,.5,.625,.75,
.875,1)),include.lowest=TRUE,right=FALSE))
  [0,1)   [1,2)   [2,3)   [3,5)   [5,8)  [8,17) [17,30]
     48      41      31      42      42      38      38
#doing right=FALSE helped equilibrate the categories more, especially the
first category.

#   getting means and vars for each category
> WSSTA.octs<-
cut(corals$WSSTA,quantile(corals$WSSTA,c(.125,.25,.375,.5,.625,.75,.875,1)),
include.lowest=TRUE,right=FALSE)
> tapply(corals$PREV_1, WSSTA.octs,mean)->means
> tapply(corals$PREV_1, WSSTA.octs,var)->vars

#   plotting vars vs. means
> plot(means, vars, axes=FALSE, xlab=expression(paste("Mean ",mu)),
ylab=expression(paste("Variance ",sigma^2)),xlim=c(0,40), ylim=c(0,11000),
cex=1.5)
> axis(1,cex.axis=.9)
> axis(2,cex.axis=.9)
> box()

#   plotting the NB2 model
> quad.coef<-coef(lm(vars~offset(means)+I(means^2)-1))
> quad.func<-function(x) x+quad.coef*x^2
> lines(seq(0,150,5), quad.func(seq(0,150,5)), col=1, lty=1, lwd=2)
```

```
#   plotting the NB1 model
> abline(0, coef(lm(vars~means-1)), col=2, lty=2, lwd=2)

#   adding the Poisson
> abline(0,1,col=3,lty=1,lwd=2)
```

Problem 2

```
> #   set up plot area
> plot(0:10, dpois(0:10,1), axes=FALSE, type='n', xlab='k',
ylab='P(X=k)',cex=1.5)
> axis(1,cex.axis=.9)
> axis(2,cex.axis=.9)
> box()

> #   plot both Poisson distributions
> points(0:10,dpois(0:10,1),pch=16,col=3)
> lines(0:10,dpois(0:10,1),col=3)
> points(0:10,dpois(0:10,5),pch=22,col=3)
> lines(0:10,dpois(0:10,5),col=3)

> #   plot all 4 of the negative binomial distributions we did in class
> points(0:10,dnbinom(0:10,size=.1,mu=1),pch=16)
> lines(0:10,dnbinom(0:10,size=.1,mu=1))
> points(0:10,dnbinom(0:10,size=10,mu=1),pch=16,col=2)
> lines(0:10,dnbinom(0:10,size=10,mu=1),col=2)
> points(0:10,dnbinom(0:10,size=.1,mu=5),pch=21)
> lines(0:10,dnbinom(0:10,size=.1,mu=5),lty=2)
> points(0:10,dnbinom(0:10,size=10,mu=5),pch=21,col=2)
> lines(0:10,dnbinom(0:10,size=10,mu=5),col=2,lty=2)

> #   add legend for all 4 distributions
> legend(5,.35, c(
+ expression(paste('Poisson, ', lambda==1)),
+ expression(paste('Poisson, ', lambda==5)),
+ expression(paste('NB2, ', theta==.1,', ', mu==1)),
+ expression(paste('NB2, ', theta==10,', ', mu==1)),
+ expression(paste('NB2, ', theta==.1,', ', mu==5)),
+ expression(paste('NB2, ', theta==10,', ', mu==5))), col=c(3,3,1,2,1,2),
pch=c(16,22,16,16,21,21), lty=c(1,1,1,1,2,2), cex=rep(.9,6), bty='n')

> #   add title
> mtext(expression(paste('Comparing Poisson and negative binomial
distributions for varying ',lambda,' and ',theta)), side=3, line=.5)
```

Problem 3

```
> #   set up plot area
> plot(0:10, dgamma(0:10,1), axes=FALSE, type='n', xlab='k',
ylab='P(X=k)',cex=1.5)
> axis(1,cex.axis=.9)
> axis(2,cex.axis=.9)
> box()

> #   plot all 4 gamma distributions
> lines(0:10,dgamma(0:10,shape=1,scale=1))
```

```
> lines(0:10,dgamma(0:10,shape=2,scale=1),col=2)
> lines(0:10,dgamma(0:10,shape=5,scale=1),col=3)
> lines(0:10,dgamma(0:10,shape=5,scale=.25),col=3,lty=2)

> #   add legend for all 4 distributions
> legend(4,1, c(
+ expression(paste(shape==1,', ', scale==1)),
+ expression(paste(shape==2,', ', scale==1)),
+ expression(paste(shape==5,', ', scale==1)),
+ expression(paste(shape==5,', ', scale==.25))),
+ col=c(1,2,3,3), lty=c(1,1,1,2), cex=rep(.9,4), bty='n')

> #   add title
> mtext(expression(paste('Comparing Gamma distributions for varying shape
',alpha,' and scale ',beta)), side=3, line=.5)
```