

# Homework 4

Biology 697

9/27/2012

## 1 Probability

Complete problems 21 & 25 on pg. 124-125 of Whitlock & Schluter. You should be able to do these with pencil and paper (or briefly in R).

## 2 Distributions

### 2.1 Continuous

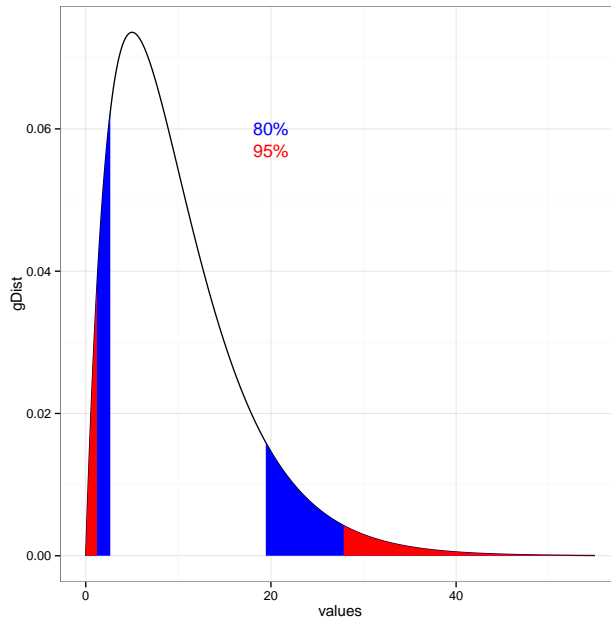
Using whatever plotting package you would like, draw a Gamma distribution (dgamma). Your choice of parackage. Visually show both its 80% confidence interval and 95% confidence intervals. Note, geom\_area takes x and y as it's aesthetics. In the base plotting package, look at polygon.

```
library(ggplot2)
values<-seq(0, 55, 0.01)
gDist<-dgamma(values, shape=2, scale=5)

gDistPlot<-ggplot(mapping=aes(x=values, y=gDist))+
  geom_line() + theme_bw()
#gDistPlot

#get the upper and nlower CIs
CIGs<-qgamma(c(0.025, 0.1, 0.9, 0.975), shape=2, scale=5)

#plot it using area
gDistPlot + geom_area(mapping=aes(x=seq(CIGs[3],55,.01),
  y=dgamma(seq(CIGs[3],55,.01), shape=2,scale=5)), fill="blue") +
  geom_area(mapping=aes(x=seq(CIGs[4],55,.01),
  y=dgamma(seq(CIGs[4],55,.01), shape=2,scale=5)), fill="red") +
  geom_area(mapping=aes(x=seq(0,CIGs[2],.01),
  y=dgamma(seq(0,CIGs[2],.01), shape=2,scale=5)), fill="blue") +
  geom_area(mapping=aes(x=seq(0,CIGs[1],.01),
  y=dgamma(seq(0,CIGs[1],.01), shape=2,scale=5)), fill="red") +
  annotate("text", x=20, y=0.06, color="blue", label="80%") +
  annotate("text", x=20, y=0.057, color="red", label="95%")
```



## 2.2 Extra Credit: Discrete

Now do the same for a Negative Binomial Distribution. You'll have to play around a bit here with `rect`, `barplot`, or `geom_rect`.

```

valuesNB<-0:30
nbDist<-dnbinom(valuesNB, mu=10, size=4)

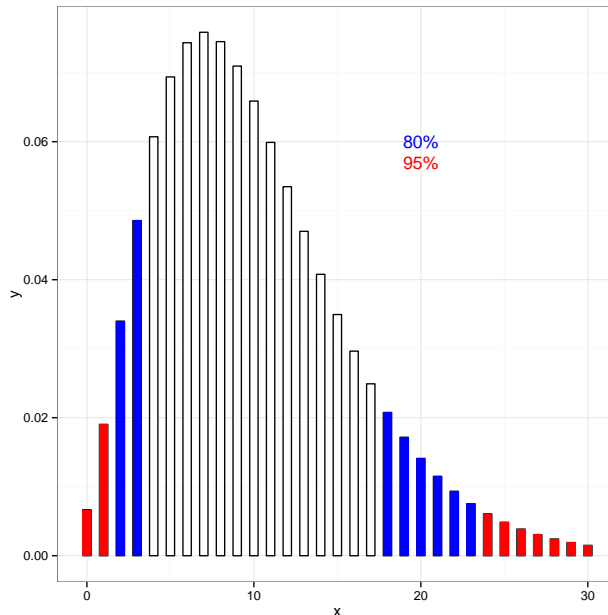
nbinDistPlot<-ggplot(mapping=aes(xmin=valuesNB-0.25, xmax=valuesNB+0.25,
                                ymin=rep(0, length(valuesNB)), ymax=nbDist))+
  geom_rect(fill="white", color="black") +
  theme_bw()

#nbinDistPlot

#get the confidence intervals
CINBs<-qnbinom(c(0.025, 0.1, 0.9, 0.975), mu=10, size=4)
u95 <- CINBs[4]:30
u80 <- CINBs[3]:30
l80 <- 0:CINBs[2]
l95 <- 0:CINBs[1]
nbinDistPlot +
  geom_rect(mapping=aes(xmin=u80-0.25, xmax=u80+0.25, ymax=dnbinom(u80, mu=10, size=4),
                        ymin=rep(0, length(u80))), fill="blue") +
  geom_rect(mapping=aes(xmin=u95-0.25, xmax=u95+0.25, ymax=dnbinom(u95, mu=10, size=4),
                        ymin=rep(0, length(u95))), fill="red") +
  geom_rect(mapping=aes(xmin=l80-0.25, xmax=l80+0.25, ymax=dnbinom(l80, mu=10, size=4),
                        ymin=rep(0, length(l80))), fill="blue") +
  geom_rect(mapping=aes(xmin=l95-0.25, xmax=l95+0.25, ymax=dnbinom(l95, mu=10, size=4),
                        ymin=rep(0, length(l95))), fill="red") +
  annotate("text", x=20, y=0.06, color="blue", label="80%") +

```

```
annotate("text", x=20, y=0.057, color="red", label="95%")
```



### 3 P Values and Hypothesis Testing

Complete problems 15 & 18 on pg. 147 of Whitlock & Schluter. When possible, compare values obtained from tables (the old school way) to values generated by R.

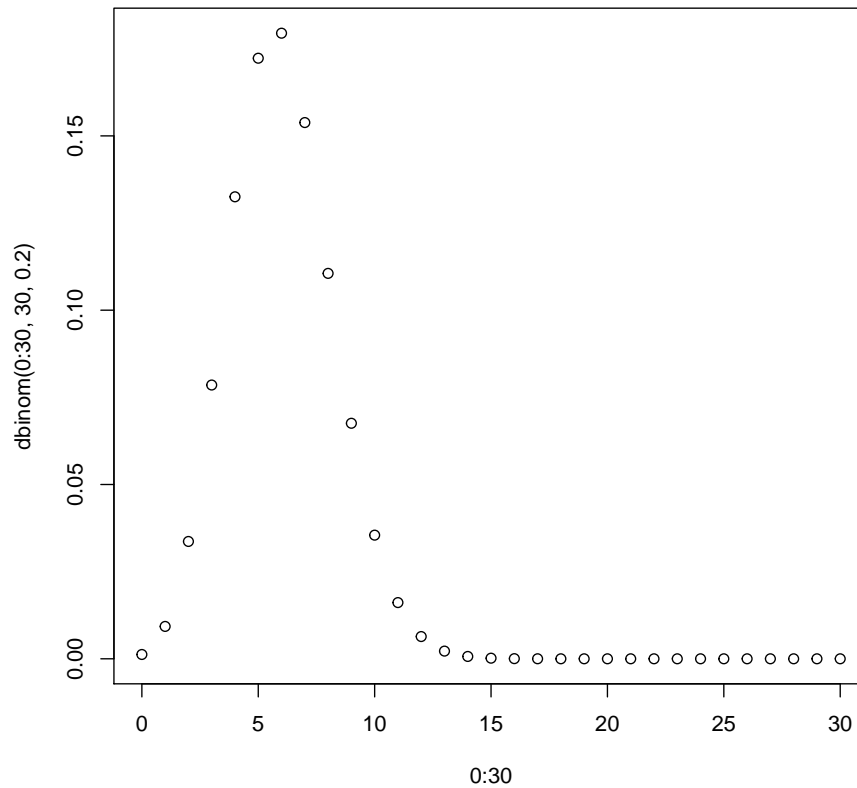
### 4 Hypothesis Testing & Power

A paleoecologist is sampling fossilized snails across the boundary of an ocean acidification event. After sampling 3 sites, she has found a total of 30 unique taxa (10 per site). Of those, 10 appear to go extinct. Normally, she would expect 20% of the species to go extinct over the time period sampled.

#### 4.1

Assuming the extinction of each species extinction or survival was independent of the others - like a coin flip, does it appear that acidification may have affected snail extinction? Use an  $\alpha$  of 0.05? Why or why not?

```
# upper or lower tail?  
plot(0:30, dbinom(0:30, 30, 0.2))
```



```
(1 - pbinom(10, 30, 0.2)) * 2
## [1] 0.05123
```

## 4.2

Assuming a 39% extinction rate due to acidification, what was the  $\beta$  and power of her analysis given an  $\alpha$  of 0.05? Use 5000 simulated draws. Was she just unlucky?

Fun fact, TRUE in R is equivalent to 1. FALSE is equivalent to 0. So, TRUE + TRUE + FALSE = 2. Neat, huh? Also, if you ask if a vector is greater than or less than a value, the comparison will be applied across the entire vector. Try `c(1,2,3)<2` to see what I mean

```

set.seed(697)
n.sims <- 5000
p <- pbinom(rbinom(n.sims, 30, 0.39), 30, 0.2)
# fix ones in the upper tail and make it two tailed
p[which(p > 0.5)] <- 1 - p[which(p > 0.5)]
p <- 2 * p
beta <- sum(p > 0.05)/n.sims
beta

## [1] 0.3418

# power
1 - beta

## [1] 0.6582

```

### 4.3

How might her results have changed with a different alpha? Plot it. Remember, to get vectors that are not just integers, `seq` is the way to go.

`seq(from = 0, to = 1, .1)` produces `[1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0`.

```

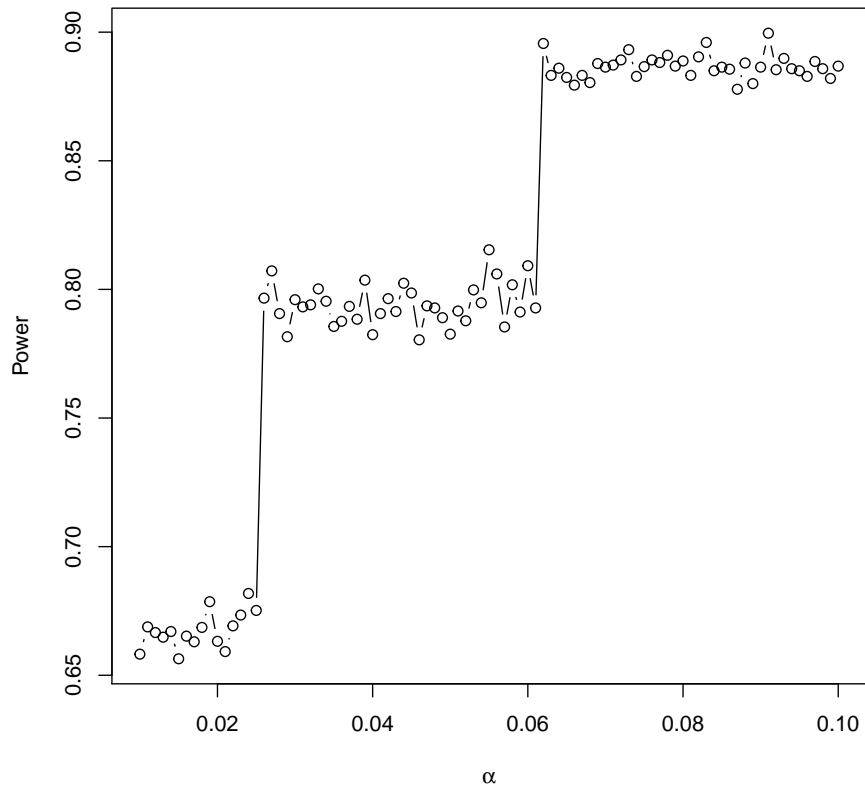
set.seed(697)
n.sims<-5000

#vector of alphas
alphaVec<-seq(0.01,0.1,0.001)

#empty vector of power
powVec<-rep(NA, length(alphaVec))

#loop over the whole thing
for(i in 1:length(alphaVec)){
  pvec<-pbinom ( rbinom(n.sims, 30,0.39), 30, 0.2)
  pvec[which(pvec>0.5)] <- 1-pvec[which(pvec>0.5)]
  powVec[i] <- 1-sum(pvec>alphaVec[i])/n.sims
}
plot(powVec ~ alphaVec, type="b", xlab=expression(alpha), ylab="Power")

```



#### 4.4 Extra Credit

Or, maybe having more sites - and hence more species - might have changed the outcome of her work. Show how number of sites and alpha interact to influence the power of her analysis, again assuming that this kind of acidification event should elevate the extinction rate to 39%.

```
#expand.grid to get all possible combinations of alpha and # of sites
conditions<-expand.grid(sites=1:15, alpha=alphaVec)
conditions$power<-rep(NA, nrow(conditions))
n.sims=5000
set.seed(697)

for(i in 1:nrow(conditions)){
  pvec<-dbinom ( rbinom(n.sims, conditions$sites[i]*10,0.39),
```

```

      conditions$sites[i]*10, 0.2)
pvec[which(pvec>0.5)] <- 1-pvec[which(pvec>0.5)]
conditions$power[i] <- 1-sum(pvec>conditions$alpha[i])/n.sims
}

qplot(sites, power, data=conditions, group=factor(alpha), color=alpha) +
  theme_bw() +
  scale_color_continuous(low="red", high="blue") +
  geom_line()

```

