MATH 420 Homework #2 SOLUTIONS

Instructions: A printed copy of your homework should be handed in at **the start of class** the day it is due. If you have any supplementary electronic files you wish to turn in (e.g. R scripts or wxMaxima files) email them to the instructor prior to class with file name format: `Lastname-hwX.ext`. Each part of each exercise is worth 10 points unless stated otherwise.

Exercise 1: List three probability distributions (besides the Normal distribution) commonly used in your major field of study, and for each of these describe

(a) the physical/real-world processes that lead to each distribution, or otherwise justify their widespread use;

(b) any useful associations between those real-world processes, the parameters of the distribution, and the mean and/or variance (or similar quantities) of that distribution; and

(c) select plausible parameter values (one set for each distribution) and plot the histogram of a large sample and the corresponding density/mass function.

Ans: *Various responses.*

Exercise 2: (Ross, 1.14) The probability of winning on a single toss of the dice is p. Player A starts, and if they fail, they pass the dice to player B, who then attempts to win on their toss. They continue tossing the dice back and forth until one of them wins. What are their respective probabilities of winning?

Ans: This is essentially a Geometric distribution for the time until someone wins, with A winning on the even turns, B winning on the odd turns.

$$P(A\ wins) = \sum_{n=0}^{\infty} p\,(1-p)^{2n} = \frac{p}{1-(1-p)^2} = \frac{1}{2-p}$$

$$P(B\ wins) = \sum_{n=0}^{\infty} p\,(1-p)^{2n+1} = \frac{p\,(1-p)}{1-(1-p)^2} = \frac{1-p}{2-p}$$

Exercise 3: (Ross, 1.21) Suppose that 5% of men and 0.25% of women are color-blind (assume there are equal numbers of men and women). A color-blind person is chosen at random. What is the probability of this person being male?

Ans: One approach is to apply Bayes' Theorem. A second (equivalent) approach is to calculate this by directly calculating the proportion of color blind individuals who are male directly. For example, a population of size $2N$ has $2N/20$ color-blind men, $2N/400$ color-blind women, and thus $21N/400$ color-blind people total. Diving, we get

$$P(\text{male}|\text{color-blind}) = \frac{2N/20}{21N/400} = \frac{1}{1+1/20} = \frac{20}{21} \approx 95.2381\%.$$

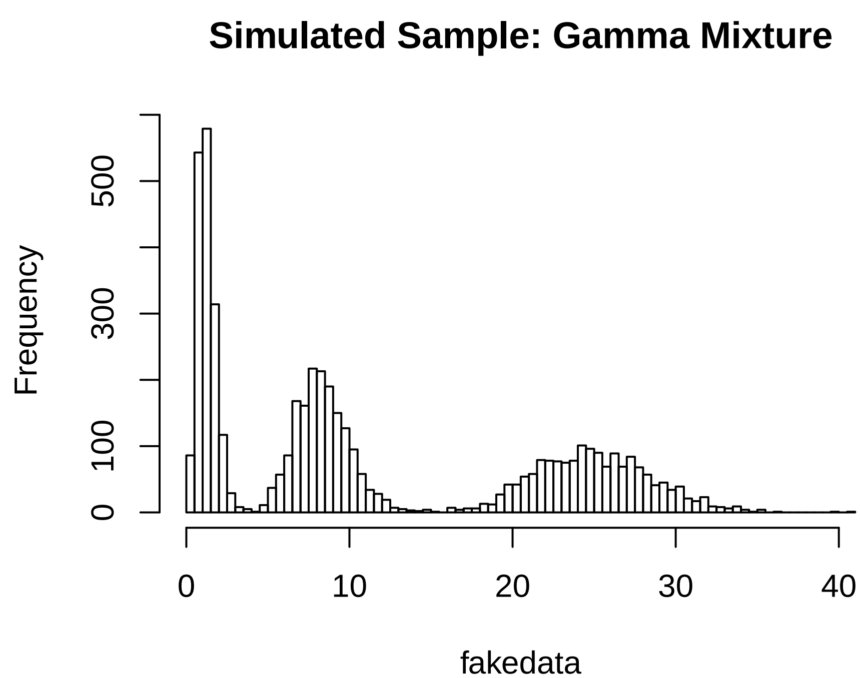
Exercise 4: Simulating mixture distributions.

(a) The R code on the last page of this assignment generates a random sample from a mixture of two Normal distributions. Modify it to generate samples from a mixture of three (or more) Gamma Distributions.

Ans:

```
rgammamix <- function(n, rates, shapes, weights) {
if(length(rates) != length(shapes) & length(rates)!=length(weights)) {
        error("rates, shapes, weights must have equal length!")
}
# First pick which distribution each observation comes from...
indx = sample(1:length(rates),n,replace=TRUE,prob=weights/sum(weights))
return(rgamma(n, rate=rates[indx], shape=shapes[indx]))
}

# Example:
set.seed(4)
fakedata = rgammamix(5000, rates=2:4, shapes=c(50,25,5), weights=c(1,1,1))
hist(fakedata,100,main="Simulated Sample: Gamma Mixture")
```



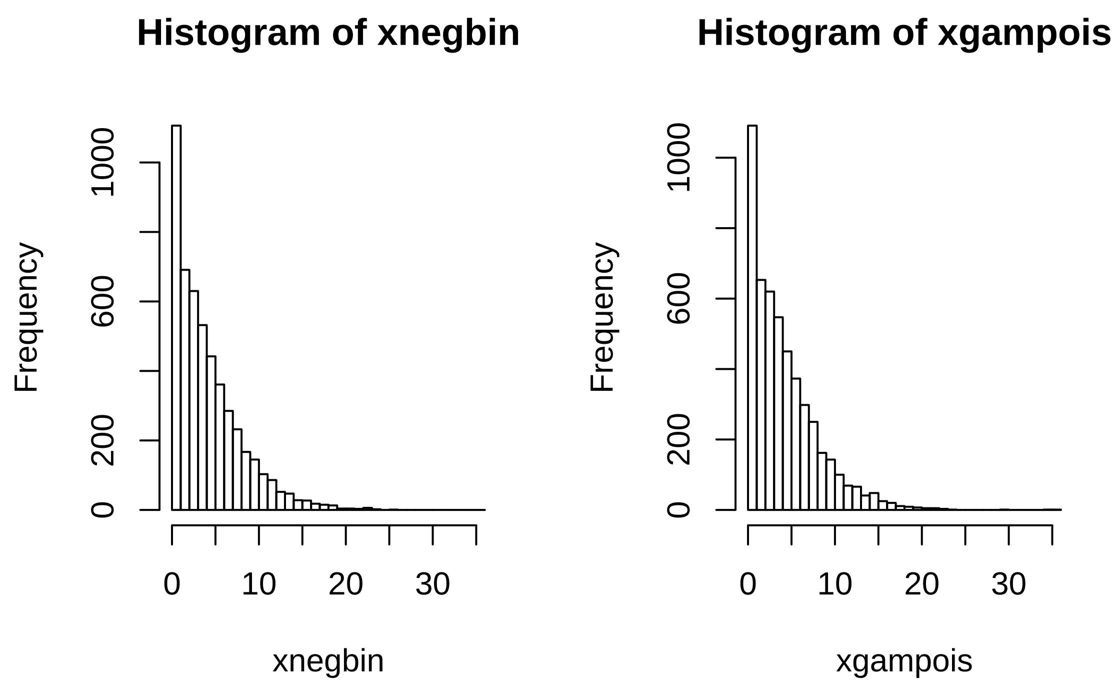
(b) The Negative Binomial distribution with rate r and probability p can be viewed as a *compound* distribution (aka a *continuous mixture* distribution) where each observation is drawn from a Poisson whose rate parameter λ is not constant, but instead is sampled from a Gamma distribution, which -- to quote the Wikipedia page https://en.wikipedia.org/wiki/Negative_binomial_distribution as of 5pm on 9/8/2017 -- is parameterized by "shape $= r$ and scale $\theta = p/(1-p)$ or correspondingly rate $\beta = (1-p)/p$". The R code below compares samples from these two distributions using the above parameterization, but there's a problem! **Correct the code, and explain the source of the error.**

Ans: The wikipedia version defined p as the probability of failures, not successes as in the R implementation, therefore "p" and "$1-p$" are being confused! The solution is to recognize that "p" in one case is "$1-p$" in the other, thus

```
rgampois <- function (n, r, prob) {
# use a different lambda for each observation
# parameterized using scale=(1-p)/(1-(1-p)), to align with R's
# definition of 'p', not wikipedia's scale=p/(1-p)...
rpois(n, lambda = rgamma(n, shape=r, scale=(1-prob)/prob))
}

r=2
p=0.3
xnegbin  <- rnbinom(5000, size = r, prob = p)
xgampois <- rgampois(5000, r, p)

par(mfrow=c(1,2)) # Two subplots, arranged in 1 row, 2 columns
hist(xnegbin,breaks = 0:max(xnegbin,xgampois))
hist(xgampois,breaks = 0:max(xnegbin,xgampois))
```



As shown above, now the two ways of sampling from a Negative Binomial agree.

The original code given with the assignment...

```
# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
# Exercise 4a

rnormix <- function(n, mean1, mean2, sd1, sd2, p1, p2) {
# n = sample size, ps=c(p1, p2) are the mixing probabilities
# means = c(mean1,mean2) and sds = c(sd1,sd2) the distributions.
ps=c(p1,p2)/(p1+p2) # This ensures these sum to 1!
means=c(mean1,mean2)
sds=c(sd1,sd2)
# First pick which distribution each observation comes from...
indx = sample(c(1,2),n,replace=TRUE,prob=ps)
# next give the corresponding vector of means and sds to rnorm()...
return(rnorm(n, mean=means[indx], sd=sds[indx]))
# see ?rnorm for details.
}

# Example:
fakedata = rnormix(5000, mean1=10, mean2=20, sd1=2, sd2=1, p1=0.2, p2=0.8)
hist(fakedata,50)


# # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # # #
# Exercise 4b

rgampois <- function (n, r, prob) {
# use a different lambda for each observation
# parameterized using scale=p/(1-p) as described on wikipedia
rpois(n, lambda = rgamma(n, shape=r, scale=prob/(1-prob)))
}

r=2
p=0.3
xnegbin  <- rnbinom(5000, size = r, prob = p)
xgampois <- rgampois(5000, r, p)

x11() # try quartz() if this doesn't work on your mac!
par(mfrow=c(1,2)) # Two subplots, arranged in 1 row, 2 columns
hist(xnegbin,breaks = 0:max(xnegbin,xgampois))
hist(xgampois,breaks = 0:max(xnegbin,xgampois))
```