DATA 252 / DATA 551: Homework 3 Solution

1. In *Introduction to Probability*, skim through pages 183 to 187; don't worry about the technical details, but make sure that you understand figures 5.1 and 5.2. Read the section on **Possin Distribution**, pages 187 to 192 (some of the ensuing homework problems are based on examples from this section).

Solution NA.

2. A typesetter makes, on average, one mistake per 1000 word. Let X be the number of mistakes that he makes on a page with n words.

   a. Suppose we model X with a binomial distribution. What are the model parameters? Discuss the binomial model assumptions in the context of the problem.

   b. With a binomial model, calculate the probabilities that: (i) he makes no mistake on a page with 100 words; (ii) he makes exactly 1 mistake on a page with 100 words; (iii) he makes 1 or more mistake on a page with 100 words; (iv) he makes no mistake on a page with 200 words.

   c. Suppose we model X with a poisson distribution. What is the model parameter? Discuss the poisson model assumptions in the context of the problem.

   d. Re-calculate the probabilities in b using a poisson model.

   e. Simulate 1000 random observations from the binomial model in a and plot a histogram of these observations; simulate another 1000 random observations from the poisson model in c and also plot a histogram. Compare the binomial and poisson models by comparing the two histograms.

Solution: We can model X with binomial $(n, p = 0.001)$. We assume that (1) whether he makes a mistake on one word is *independent* from whether he makes mistakes on other words, and (2) each word has *equal probability* that he makes a mistake. Probabilities in b are calculated below.

```
dbinom(0,size=100,prob=0.001)
```

```
## [1] 0.9047921
```

```
dbinom(1,size=100,prob=0.001)
```

```
## [1] 0.09056978
```

```
1-dbinom(0,size=100,prob=0.001)
```

```
## [1] 0.09520785
```

```
dbinom(0,size=200,prob=0.001)
```

```
## [1] 0.8186488
```

We can also model X with poisson ($\lambda = 0.001n$). This is an approximation because the sample space of number of mistakes he can make is $0, 1, \ldots, n$, but the poisson distribution has sample space $0, 1, \ldots, \infty$. We assume that (1) whether he makes a mistake on one word is *independent* from whether he makes mistakes on other words, and (2) he makes mistakes at a *constant rate*. Probabilities in d are calculated below. Note that the probabilities are very close to the binomial model.

```r
dpois(0,lambda=0.1)
```

```
## [1] 0.9048374
```

```r
dpois(1,lambda=0.1)
```

```
## [1] 0.09048374
```

```r
1-dpois(0,lambda=0.1)
```

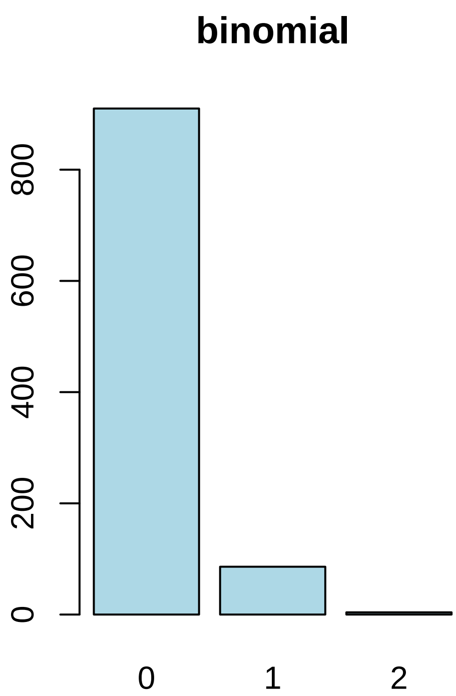
```
## [1] 0.09516258
```

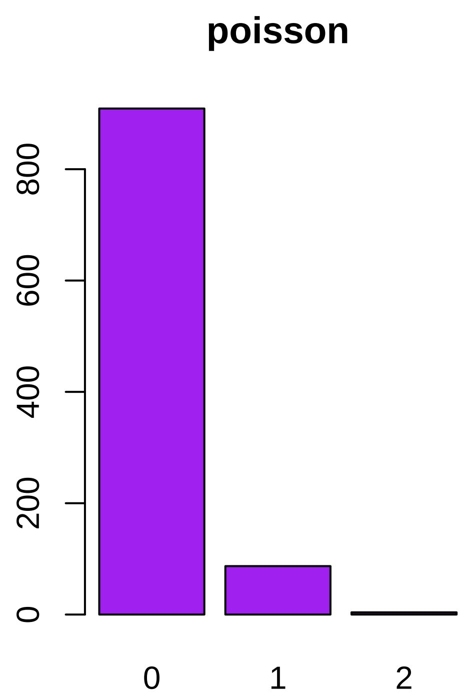
```r
dpois(0,lambda=0.2)
```

```
## [1] 0.8187308
```

Histograms to compare the binomial and poisson models are given below. Comment: it's hard to estimate tail probabilities like $P(X = 3, 4, \ldots, 100)$ using simulation because the probabilities are very small.

```r
par(mfrow=c(1,2))
plot(as.factor(rbinom(1000,size=100,prob=0.001)),
     col='lightblue',main='binomial')
plot(as.factor(rpois(1000,lambda=0.1)),
     col='purple',main='poisson')
```





3. Do question 18 (page 199) in *Introduction to Probability*.

Solution:

(a) The number of raisins in a randomly picked cookie, X, can be modeled using a poisson distribution with $\lambda = 600/500$, so $P(X = 0)$ is `dpois(0,lambda=6/5)`=0.3011942.

(b) Similar to (a), we use poisson with $\lambda = 400/500$.
$P(X = 2) =$ `dpois(2,lambda=4/5)`=0.1437853.

(c) $P(X \geq 2) = 1 - P(X = 0, 1)$.
`1-dpois(0,lambda=1000/500)-dpois(1,lambda=1000/500)`=0.5939942.

4. Under the setup of question 26 (page 201) in *Introduction to Probability*, the number of hits in each of the 576 small areas can be modeled by a poisson distribution. What is the parameter? Calculate the probabilities that a small area would have 0, 1, 2, 3, 4, and 5 or more hits under the poisson model, and compare the probabilities to the actual proportions.

Solution: The number of hit can be modeled by poisson with $\lambda = 537/576$. Calculation below shows that the poisson model works very well.

```r
#estimated probabilities
estimated=dpois(c(0,1,2,3,4),lambda=537/576)
estimated[6]=1-sum(dpois(c(0,1,2,3,4),lambda=537/576))

#actual proportions
actual=c(229,211,93,35,7,1)/537

result=cbind(estimated,actual)
```

```r
rownames(result)=c('0 hit','1 hit','2 hits','3 hits','4 hits', '5 or more hits')
round(result,3)
```

```
##                  estimated actual
## 0 hit                0.394  0.426
## 1 hit                0.367  0.393
## 2 hits               0.171  0.173
## 3 hits               0.053  0.065
## 4 hits               0.012  0.013
## 5 or more hits       0.003  0.002
```

5. Do question 33 (page 202) in *Introduction to Probability* by simulating the process and making random draws from the distribution of X. The random draws can be used to empirically study the distribution of X. For instance, plot a histogram of the random draws to estimate the shape of the distribution; use observed proportions to estimate the probability mass function.

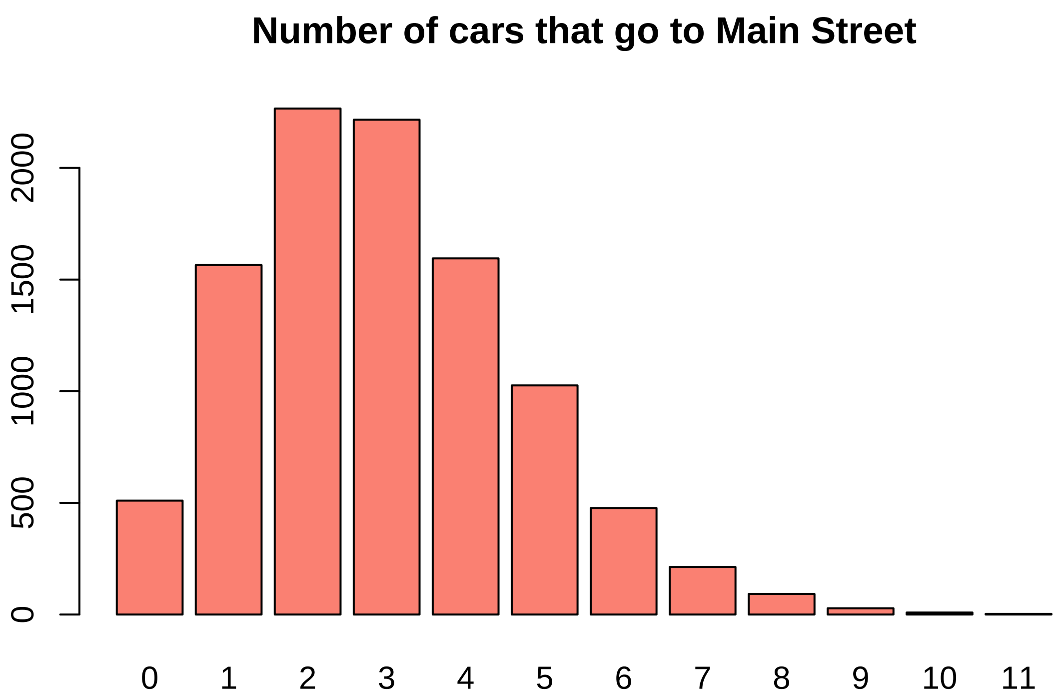
Solution: Simulation given below. We can estimate, for instance, $P(X = 0) = 0.05, P(X = 1) = 0.15$,etc.

```r
nsim=10000
result=rep(NA,nsim)

for(i in 1:nsim){
  numcar=rpois(1,lambda = 4) #total number of cars
  result[i]=rbinom(1,size=numcar,prob=3/4) #number of cars that choose Main Street
}
table(result)/nsim #estimated probabilities
```

```
## result
##      0      1      2      3      4      5      6      7      8      9
## 0.0510 0.1565 0.2266 0.2216 0.1595 0.1026 0.0477 0.0213 0.0092 0.0028
##     10     11
## 0.0009 0.0003
```

```r
plot(as.factor(result),main='Number of cars that go to Main Street',col='salmon')
```



6. Suppose that the probabilities of having a male or a female child are both 0.5, and you do not have to consider cases like twins. Calculate the following probabilities using appropriate models.

a. If you plan to have three children, what is the probability that you have exactly one daughter? What is the probability that you have at least one daughter?

b. What is the probability that your first daughter is your third child? (Hint: imagine the scenario of keeping having children until you have a daughter, and model the number of children when the first daughter "occurs").

c. What is the probability that your second daughter is your third child?

d. How many children do you need to have to guarantee with more than 95% chance that you will have at least a daughter?

Solution: For a, b, c:

```
dbinom(1,size=3,prob=0.5)   #a, exactly 1
```

```
## [1] 0.375
```

```
1-dbinom(0,size=3,prob=0.5)   #a, at least 1
```

```
## [1] 0.875
```

```
dgeom(2,prob=0.5) #b
```

```
## [1] 0.125
```

```r
dnbinom(1,size=2,prob=0.5) #c
```

```
## [1] 0.25
```

For d, let N be the number of children you have, then the number of daughters, X, can be modeled as binomial with $n = N$ and $p = 0.5$. We want $P(X \geq 1) = 1 - P(X = 0)$ to be greater than 95%. We calculate this probability for different values of N, and see that $N = 5$ is the smallest number of children to satisfy the condition.

```r
1-dbinom(0,size=c(1:10),prob=0.5)
```

```
##  [1] 0.5000000 0.7500000 0.8750000 0.9375000 0.9687500 0.9843750 0.9921875
##  [8] 0.9960938 0.9980469 0.9990234
```