

**STAT 515**  
**Homework #1 WITH SOLUTIONS**

1. A fair 6-sided die is rolled repeatedly. Let  $X$  equal the number of rolls required to obtain the first 5 and  $Y$  the number required to obtain the first 6. Calculate

(a)  $E(X)$

**Solution:** Since  $X$  is a geometric random variable with parameter  $p = 1/6$ ,  
 $E(X) = 1/p = 6$ .

(b)  $E(X | Y = 1)$

**Solution:** The fact that  $Y = 1$  is equivalent to the fact that a 6 is rolled on the first roll. Starting with the second roll, however, we're back to the same situation in part (a), so all we need to do is add in the first roll, which is obviously not a 5:

$$E(X | Y = 1) = 1 + E(X) = 7$$

(c)  $E(X | Y = 5)$

**Solution:** Here,  $Y = 5$  means that there are no 6's on rolls 1 through 4, so conditioning on  $I\{X \leq 4\}$  gives

$$P(X = k | Y = 5, 1 \leq X \leq 4) = \frac{(4/5)^{k-1}(1/5)}{1 - (4/5)^4} \quad \text{for } k = 1, 2, 3, 4.$$

We may therefore use brute force to calculate

$$E(X | Y = 5, X \leq 4) = \frac{(1/5)}{1 - (4/5)^4} [1 + 2(4/5) + 3(4/5)^2 + 4(4/5)^3] = 2.225.$$

On the other hand, if  $X > 5$ , then we're back to the same situation as part (a) starting at the sixth roll, which means that

$$E(X | Y = 5, X > 5) = 5 + E(X) = 11.$$

Thus, the final answer is a weighted average of the two:

$$\begin{aligned} E(X | Y = 5) &= E[E(X | Y = 5, I\{X \leq 4\})] \\ &= 2.225P(X \leq 4 | Y = 5) + 11P(X > 5 | Y = 5) \\ &= 2.225[1 - (4/5)^4] + 11(4/5)^4 = 5.8192. \end{aligned}$$

2. Suppose that  $X$  is exponentially distributed with parameter  $\lambda$ ; i.e.,  $E(X) = 1/\lambda$ . Calculate  $P(X - 2 \leq t | X > 2)$  for an arbitrary  $t > 0$ . Based on your answer, what is the conditional distribution of  $X - 2$  given  $X > 2$ ?

**Solution:** The cumulative distribution function of  $X$  is  $F(t) = 1 - e^{-t\lambda}$  for  $t > 0$ . Thus,

$$P(X - 2 \leq t | X > 2) = \frac{P(2 < X \leq t + 2)}{P(X > 2)} = \frac{F(t + 2) - F(2)}{1 - F(2)} = \frac{e^{-2\lambda} - e^{-(t+2)\lambda}}{e^{-2\lambda}} = 1 - e^{-t\lambda}.$$

This is the original cdf of  $X$ , which means that given  $X > 2$ , the conditional distribution of  $X - 2$  is simply exponential with mean  $1/\lambda$ .

3. A coin having probability  $p$  of resulting in heads is successively flipped until the  $r$ th head appears. Let  $X$  be the total number of flips required. Then  $X$  is called a *negative binomial* random variable with parameters  $r$  and  $p$ . (NB: A geometric random variable results in the special case  $r = 1$ .)

Argue that the probability mass function of  $X$  is given by

$$p(k) = \binom{n-1}{r-1} p^r (1-p)^{n-r} \quad \text{for } k=r, r+1, \dots$$

**Hint:** How many successes must there be in the first  $n - 1$  trials?

**Solution:** First, please note that the pmf I gave (above) is completely incorrect: It should not use both  $n$  and  $k$ , since there should be only a single variable used for the total number of flips. It is probably better to use  $k$  (in the hint too!), avoiding  $n$  altogether because of its association with the binomial distribution. Here is the correct mass function:

$$p(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \quad \text{for } k=r, r+1, \dots$$

To obtain  $X = k$ , we must get a string of exactly  $r$  heads and  $k-r$  tails, where the last flip is heads. Each such string occurs with probability  $p^r (1-p)^{k-r}$ , so to find  $P(X = k)$  it remains only to count the number of such strings and then multiply by  $p^r (1-p)^{k-r}$ . Since the last flip must be heads, the number of strings is equal to the number of ways to get  $r-1$  heads in the first  $k-1$  flips, which is simply  $\binom{k-1}{r-1}$ .

4. Suppose that  $X_1$  and  $X_2$  are independent geometric random variables with the same parameter  $p$ . Prove that the conditional distribution of  $X_1$  given  $X_1 + X_2 = n + 1$ , for some positive integer  $n$ , is discrete uniform.

**Hint:** The sum of independent geometric random variables with the same mean is negative binomial.

**Solution:** Since  $X_1$  can take values  $1, \dots, n$ , we calculate for  $k = 1, \dots, n$ :

$$\begin{aligned} P(X_1 = k \mid X_1 + X_2 = n + 1) &= \frac{P(X_1 = k)P(X_2 = n + 1 - k)}{P(X_1 + X_2 = n + 1)} \\ &= \frac{p(1-p)^{k-1}p(1-p)^{n-k}}{\binom{n}{1}p^2(1-p)^{n-1}} = \frac{1}{n}. \end{aligned}$$

5. Suppose that you arrive at a party, along with a random number,  $X$ , of additional people, where  $X \sim \text{Poisson}(10)$ . The times at which people (including you) arrive at the party are independent uniform(0,1) random variables.

(a) Find the expected number of people who arrive before you.

(b) Find the variance of the number of people who arrive before you.

**Solution:** Let  $N$  denote the number of people who arrive before you and  $T$  the time when you arrive. I will give three different solutions:

- i. Conditional on  $X$ , by symmetry it is equally likely that you are the 1 through  $(X + 1)$ th person to arrive. Therefore,  $N$  is discrete uniform on  $0, \dots, X$  conditional on  $X$ . This means that  $E(N \mid X) = X/2$  and  $\text{Var}(N \mid X) = X(X + 2)/12$ . Thus, we obtain

$$\begin{aligned} E(N) &= E[E(N \mid X)] = E[X/2] = 5, \\ \text{Var}(N) &= E[\text{Var}(N \mid X)] + \text{Var}[E(N \mid X)] = \frac{1}{12}E[X^2 + 2X] + \text{Var}[X/2] \\ &= \frac{110 + 20}{12} + \frac{10}{4} = \frac{40}{3}. \end{aligned}$$

- ii. Conditional on  $X$  and  $T$ ,  $N$  is binomial with parameters  $X$  and  $T$ . Furthermore,  $X$  and  $T$  are independent. Therefore, using  $E(X^2) = 110$  and  $E(T^2) = 1/3$ ,

$$\begin{aligned} E(N) &= E[E(N \mid X, T)] = E[XT] = E[X]E[T] = 5, \\ \text{Var}(N) &= E[\text{Var}(N \mid X, T)] + \text{Var}[E(N \mid X, T)] = E[XT(1 - T)] + \text{Var}[XT] \\ &= E(XT) - E(XT^2) + E(X^2T^2) - [E(XT)]^2 \\ &= 5 - \frac{10}{3} + \frac{110}{3} - 25 = \frac{40}{3} \end{aligned}$$

- iii. This idea is from one of the students who came to office hours: Conditional on  $T$ ,  $N$  is  $\text{Poisson}(10T)$ . One way to see this is to note that the arrival times of the other people define a Poisson process, which is something we'll study later in the course. We conclude that

$$\begin{aligned} E(N) &= E[E(N | T)] = E[10T] = 5, \\ \text{Var}(N) &= E[\text{Var}(N | T)] + \text{Var}[E(N | T)] = E[10T] + \text{Var}[10T] = 5 + \frac{100}{12} = \frac{40}{3}. \end{aligned}$$

6. A short simulation exercise: Estimate the answers you obtained for Problem 5 above via simulation. You already have the answer so you can compare your estimates with the answer. Use 1000 replications of the process (one replicate of the process = one randomly sampled party.)

- First download and install R; see the course webpage, <http://sites.stat.psu.edu/~dhunter/515/>, for a link to “R statistical software links” provided by Dr. Haran.
- You can find a simple example for random variate simulation here: <http://www.stat.psu.edu/~dhunter/515/hw/hw1ex.R>. You can adapt this example to estimate the expectation and variance for this problem.

Ideally, you should also be reporting simulation (Monte Carlo) standard errors for your estimates; we will discuss this later in the course.

Because this is your first assignment and your R code here will be quite short, please include a printout of your R code with the assignment.

**Solution:** Using R code that can be found in <http://sites.stat.psu.edu/~dhunter/515/hw/sol01Code.R>, I obtained the sample mean and variance shown below. (NB: You can type the following `source` command directly into R to obtain similar results.)

```
> source("http://sites.stat.psu.edu/~dhunter/515/hw/sol01Code.R")
[1] 4.909
[1] 13.47019
```

Since these are only sample estimates, it makes sense to report measures of the uncertainty of these estimates. For instance, we might consider a simple 95% confidence interval for the true mean, which is given by  $4.91 \pm 2\sqrt{13.47/1000}$ , or (4.68, 5.14).

For the sake of completeness, here is the code from <http://sites.stat.psu.edu/~dhunter/515/hw/sol01Code.R>:

```
# Exercise 6:
N <- 1:1000
for(i in 1:1000) {
  X <- rpois(1,10)
  mytime <- runif(1)
  N[i] <- sum(runif(X)<mytime)
}
print(mean(N))
print(var(N))
```