Math 394

Solutions to Homework Assignment

Problems from Chapter 4

#17

The trick here is to notice that $P[a < X \leq b] = F_X(b) - F_X(a)$ (notice the missing/present equality sign, which is due to the definition of $F_X(x) = P[X \leq x]$).

a Clearly, $P[i - \varepsilon < X \leq i] = F_X(i) - F_X(i - \varepsilon)$, and for the different cases, we have

$$i = 1 : F_X(1) - F_X(1 - \varepsilon) = \frac{1}{2} - \frac{1 - \varepsilon}{4}$$

$$i = 2 : F_X(2) - F_X(2 - \varepsilon) = \frac{11}{12} - \left(\frac{1}{2} + \frac{1 - \varepsilon}{4}\right)$$

$$i = 3 : F_X(3) - F_X(3 - \varepsilon) = 1 - \frac{11}{12}$$

To get the required $P[X = i]$, we use the fact that

$$\{X = i\} = \bigcap_{\varepsilon \downarrow 0} \{i - \varepsilon < X \leq i\}$$

, to obtain (thanks to our assumptions about continuity of probabilities),

$$P[X = i] = \lim_{\varepsilon \downarrow 0} F_X(i) - F_X(i - \varepsilon)$$

or

$$i = 1 : \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

$$i = 2 : \frac{11}{12} - \left(\frac{1}{2} + \frac{1}{4}\right) = \frac{11}{12} - \frac{3}{4} = \frac{1}{6}$$

$$i = 3 : 1 - \frac{11}{12} = \frac{1}{12}$$

b The spirit is exactly the same: to "fix" the $<$ (instead of $\leq$) sign to the right, we write the open interval as a limit, and get

$$P\left[\frac{1}{2} < X < \frac{3}{2}\right] = \lim_{\varepsilon \downarrow 0} P\left[\frac{1}{2} < X \leq \frac{3}{2} + \varepsilon\right] = \lim\left(\frac{1}{2} + \frac{\frac{3}{2} + \varepsilon - 1}{4} - \frac{\frac{1}{2}}{4}\right) =$$

$$= \frac{1}{2} + \frac{1}{8} - \frac{1}{8} = \frac{1}{2}$$

#27

The company considers the two possible events: E, with probability p, and E^c with probability $1-p$. If the event occurs, it pays A, if it doesn't it pays nothing. Hence, the expected payment is $pA + (1-p)\cdot 0 = pA$. The expected profit will be the premium (which we'll call R) minus the expected payment, and we want this to be $.1 \cdot A$:

$$R - pA = .1 \cdot A$$

$$R = A\,(p + .1)$$

#36

When we are playing a best-out-of-3 series, with team A a winner of each game, independently, with probability p, the series will last 2 games with probability $p^2 + (1-p)^2 = 2p^2 - 2p + 1$, and 3 with probability $1 - \left\{p^2 + (1-p)^2\right\} = 2p - 2p^2$. We already checked last week that $EN = 2 + 2p - 2p^2$. We can choose to compute

$$Var\,(N) = E\left[N^2\right] - (EN)^2$$

or

$$Var(N) = E\left[(N - EN)^2\right]$$

The result won't change. For example,

$$E\left[N^2\right] - (EN)^2 = 2^2 P[N=2] + 3^2 P[N=3] - \left(2 + 2p - 2p^2\right)^2 =$$

$$4 \cdot \left(2p^2 - 2p + 1\right) + 9 \cdot \left(2p - 2p^2\right) - \left(2 + 2p - 2p^2\right)^2 =$$

$$= 8p^2 - 8p + 4 + 18p - 18p^2 - 4p^4 - 4p^2 - 4 - 8p + 8p^2 + 8p^3 =$$

$$= -4p^4 + 8p^3 - 6p^2 + 2p = 2p\left(1 - 3p + 4p^2 - 2p^3\right)$$

The expression is clearly 0 when $p = 0$, and $p = 1$ (because in both cases $N = 2$ with probability 1). We may compute the derivative, finding

$$-16p^3 + 24p^2 - 12p + 2 = 2\left(1 - 6p + 12p^2 - 8p^3\right)$$

which is indeed 0 for $p = 0.5$. Dividing this polynomial by $p - \frac{1}{2}$, we find that

$$1 - 6p + 12p^2 - 8p^3 = \left(p - \frac{1}{2}\right)\left(-8p^2 + 8p - 2\right) =$$

$$= 2\left(p - \frac{1}{2}\right)\left(-4p^2 + 4p - 1\right) = -2\left(p - \frac{1}{2}\right)(2p-1)^2 = -8\left(p - \frac{1}{2}\right)^2$$

i.e. $p = 0.5$ is a triple zero for the derivative. That this is a maximum follows obviously form the fact that the variance is zero at the boundaries and positive inside.

#41

If our supposed ESP man had been guessing, he would have had probability $\frac{1}{2}$ to get the guess right. The probability that he guesses right 7 times or more is thus (we are not given the specific sequence of rights and wrongs, hence we need to use the binomial distribution!)

$$\sum_{k=7}^{10} \binom{10}{k} \left(\frac{1}{2}\right)^{10} = \frac{1}{2^{10}}\left(\frac{10 \cdot 9 \cdot 8}{6} + \frac{10 \cdot 9}{2} + 1\right) = \frac{1}{2^{10}}(120 + 45 + 1) = \frac{166}{2^{10}} = .16211$$

which is not large, but it's not terribly small either. By the way, to speed up the calculations, we used the identity (obvious, if you look at the definition)

$$\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{k!\,(n-k)!}$$

#58

We calculate the corresponding quantities for the binomial distribution (b), and the Poisson distribution (p):

a $P[X=2]=$ b: $\binom{8}{2}.1^2.9^6=.1488$

p: $\lambda=np=.8\rightarrow\frac{.8^2}{2}e^{-.8}=.14379$. The relative error caused by the Poisson approximation is $\varepsilon=\frac{.14379-.1488}{.1488}=-0.033669$, i.e. about 3%

b $P[X=9]=$ b: $\binom{10}{9}.95^9.05=.31512$

p: we compute the probability that $10-X$ (which is the sum of events of small probability $1-p$) is 1: $\lambda=n(1-p)=.5\rightarrow.5\cdot e^{-.5}=.30327$. $\varepsilon=\frac{.30327-.31512}{.31512}=-.037605$

c $P[X=0]=$ b: $\binom{10}{0}.1^0.9^{10}=.34868$

p: $\lambda=np=1\rightarrow e^{-1}=.36788$. $\varepsilon=\frac{.36788-.34868}{.34868}=.055065$

d $P[X=4]=$ b: $\binom{9}{4}.2^4.8^5=.06606$

p: $\lambda=np=1.8\rightarrow\frac{1.8^4}{4!}e^{-1.8}=.072302$. $\varepsilon=\frac{.072302-.06606}{.06606}=.09449$.

In these examples, relative errors range form around 3% to more than 9%. Note that n is relatively small in all of them.

#61

With such small probability, $1.4\cdot10^{-3}$, we can use the Poisson approximation (try to use the binomial formula, and you will agree that it's best to go Poisson).

Hence we have $\lambda = np = 10^3 \cdot 1.4 \cdot 10^{-3} = 1.4$, and the probability of exactly **2** successes is

$$\frac{1.4^2}{2} e^{-1.4} = .24167$$

The requested probability, however, is for ***at least*** two successes, and so is best computed by

$$1 - e^{-1.4} - 1.4 \cdot e^{-1.4} = .40817$$

#65

a As usual, the probability of "at least one" is best computed as 1-probability of "none". The probability that none of the 500 soldiers is positive will be $.999^{500}$, so the probability for at least one positive will be $1 - .999^{500} \simeq 1 - e^{-500 \cdot .001} = 1 - e^{-0.5} = 1 - .60653 = .39347$ (the last expression uses the Poisson approximation)

b Again using the Poisson approximation, we will have that the probability for more than one positive is

$$1 - e^{-0.5} - 0.5e^{-0.5} = .090204 \tag{4}$$

Since we already know that one is positive, we need to condition on this fact (since $X \geq 2 \Rightarrow X \geq 1$, the intersection has probability given by (4)):

$$\frac{1 - e^{-0.5} - 0.5e^{-0.5}}{1 - e^{-0.5}} = 1 - \frac{0.5}{e^{0.5} - 1} = .22925$$

(the increase is significant - it corresponds to having had to rule out the highly probable event that *nobody* is positive)

c The question is somewhat ambiguous. As far as Jones knows, the probability that someone else has the disease is $1 - .999^{499} \simeq 1 - e^{-499 \cdot .001} = 1 - e^{-.499} = 1 - .60714 = .39286$

d Now, we have $500 - i$ individuals, and, for all we know, they each have a probability of 10^{-3} to carry the disease. Hence, the probability that at least one does, is

$$1 - .999^{500-i} \simeq 1 - e^{-10^{-3}(500-i)} = 1 - e^{-0.5} e^{\frac{i}{1000}}$$

Theoretical Exercises

#13

We have that to maximize $P[X = k]$, we only need to look at $p^k(1-p)^{n-k}$. The derivative is equal to zero when

$$kp^{k-1}(1-p)^{n-k} - (n-k)p^k(1-p)^{n-k-1} = 0$$
$$k(1-p) - (n-k)p = 0$$
$$np = k \Rightarrow p = \frac{k}{n}$$

In other words, the *Maximum Likelihood Estimator* for p is the *frequency* with which we observed a success.

Notice that the solution depends on k, that is the observed value of X. In a statistical experiment, we would call this quantity, $\frac{X}{n}$ (in statistical lingo, this is called, confusingly, a *statistic*), the MLE for the unknown parameter p, and would use *its* distribution to give quantitative statements about the likely values of p.

#18

We have another example of a *Maximum Likelihood Estimator*. As before, we only need to concentrate on the factors in $P[X = k]$ that actually involve λ, that is, $\lambda^k e^{-\lambda}$. Setting the derivative equal to zero, we have

$$k\lambda^{k-1}e^{-\lambda} - \lambda^k e^{-\lambda} = 0$$
$$k - \lambda = 0$$
$$\lambda = k$$

The MLE for the Poisson parameter is the number of occurrences of the "rare" event. This is consistent with our result in Exercise #13 and the connection between Binomial and Poisson random variables. Recall that, in the Poisson approximation $\lambda = np$. Since the MLE for p is $\frac{k}{n}$, it is very natural that the MLE for λ is $n \cdot \frac{k}{n}$

#25

This is a remarkable and important feature of the Poisson distribution: if we *censor* its observations with (independent) probability p, the result is still Poisson, albeit with a smaller parameter.

First of all, observe how this is again consistent with the connection between Binomial and Poisson distributions. The binomial case would work as follows. Suppose $X = \sum_{k=1}^{n} X_k$ is the sum of n independent Bernoulli variables with parameter q, and hence $Bin(n,q)$. Suppose now that, independently, each Bernoulli variable can be turned to zero with probability p. This can be represented, for example, by introducing another sequence of independent Bernoulli variables with parameter p, and considering $Z = \sum_{k=1}^{n} X_k Y_k$. For $X_k Y_k = 1$ we need both Bernoulli variables to be equal to 1. Since they are independent,

$$P[X_k = 1, Y_k = 1] = P[X_k = 1][Y_k = 1] = qp$$

and consequently Z is binomial with parameters n, and pq. A Poisson limit will have parameter $\lambda = npq = p\mu$ if $\mu = nq$

Now for a direct proof. Let N be a Poisson variable with parameter μ. Consider the censored variable X. To compute $P[X = k]$, we will condition on $\{N = n\}$ (obviously, we have to require $n \geq k$), and apply the total probability theorem:

$$P[X = k] = \sum_{n=k}^{\infty} P[X = k \mid N = n]P[N = n]$$

Now observe that, conditioned on a fixed value of n, X is the result of n independent observations that result in a success with probability p: in other words, its *conditional* distribution will be binomial with parameters n and p. Substituting, we have

$$P[X = k] = \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{\mu^n}{n!} e^{-\mu} = e^{-\mu} \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \mu^k \mu^{n-k} \frac{1}{n!} =$$

$$= e^{-\mu}\frac{(p\mu)^k}{k!}\sum_{n=k}^{\infty}\frac{(\mu(1-p))^{n-k}}{(n-k)!} = e^{-\mu}\frac{(p\mu)^k}{k!}\sum_{m=0}^{\infty}\frac{(\mu(1-p))^m}{m!} = e^{-\mu}\frac{(p\mu)^k}{k!}e^{\mu(1-p)} = e^{-\lambda}\frac{\lambda^k}{k!}$$

where (after a simple change of summation index to $m = n - k$), we defined $\lambda = \mu p$

#27

We actually discussed the core of this fact in class ("The Geometric distribution has no memory"). It follows directly from the definition of conditional probability, and the obvious fact that $\{X = n + k\} \subseteq \{X > n\}$, so that $\{X = n + k\} \bigcap \{X > n\} = \{X = n + k\}$. Consequently the required conditional probability is

$$\frac{P[X = n + k]}{P[X > n]} = \frac{(1-p)^{n+k-1} - (1-p)^{n+k}}{(1-p)^n} = (1-p)^{k-1} - (1-p)^k = P[X = k]$$