CMPSCI 240
Reasoning Under Uncertainty
Homework 2

Prof. Hanna Wallach

Assigned: February 10, 2012
Due: February 17, 2012

Question 1: (5 points) Let N be an integer-valued random variable that takes on only positive values (i.e., $N = 1, 2, 3, 4, \ldots$). Show that

$$\mathbb{E}[N] = \sum_{i=1}^{\infty} \mathbb{P}\left[N \geq i\right].$$

Question 2: (5 points) Let $X_1, \ldots, X_n$ be independent, identically distributed random variables with common mean and variance. Find the values of scalars c and d that will make the following formula true:

$$\mathbb{E}[(X_1 + \ldots + X_n)^2] = c\,\mathbb{E}[(X_1)^2] + d\,(\mathbb{E}[X_1])^2$$

Question 3: (10 points) Your computer has been acting very strangely lately, and you suspect that it might have a virus. Unfortunately, all 12 of the different virus detection programs you own are now outdated. You know that if your computer does have a virus, each program, independently of the others, has a 0.8 probability of believing that your computer as infected and a 0.2 probability of believing that your computer is fine. If your computer does not have a virus, each program has a 0.9 probability of believing that your computer is fine and a 0.1 probability of believing that your computer is infected. Given that your computer has a 0.65 probability of being infected with some virus and that you will only believe your virus detection programs if 9 or more of them agree, find the probability that your detection programs will lead you to the right answer.

Question 4: (5 points) Let X be a discrete random variable and $Y = |X|$.

(a) Assume that the PMF of X is

$$p_X(x) = \begin{cases} Kx^2 & \text{if } x = -3, -2, 0, 1, 2, 3 \\ 0 & \text{otherwise,} \end{cases}$$

where K is a suitable constant. Determine the value of K.

(b) For the PMF of X given in part (a) calculate the PMF of Y.

(c) Give a general formula for the PMF of Y in terms of the PMF of X.

Question 5: (10 points) Phil and Sandeep play a "sudden-death" chess match. The match consists of at least one game. The first player to win a game wins the match. Each game is won by Phil with probability p, is won by Sandeep with probability q, and is a draw with probability $1 - p - q$.

(a) What is the probability that Phil wins the match?

(b) What is the PMF, mean, and variance of the match length?

Question 6: (5 points) A particular binary data transmission and reception device is prone to some error when receiving data. Suppose that each bit is read correctly with probability p. Find the value of p such that when 100,000 bits are received, the expected number of errors is at most 850.

Question 7: (10 points) Suppose that you have designed a website that describes how totally awesome you are. You are interested in the types of visitors you have. To facilitate this, you have written a program to track the top-level domain (TLD) of each visitor to the site. For example, some visitors are students at other universities, so you get visitors with a TLD of "edu"; meanwhile, some visitors are people in large companies who are interested in your resume, so you also get visitors with TLD of "com".

Suppose there are n different possible TLDs (e.g., "edu", "com", "org"). You may assume that each visitor has exactly one TLD and that the TLDs for different visitors are mutually independent. You may also assume that is equally likely for the next visitor to your site to be from any of n different TLDs. Find the expected number of visitors you need so that you have at least one visitor from each TLD. (Hint: Let X be the number of visitors needed. Represent X as a sum of geometric random variables.)