TRINITY COLLGE DUBLIN

School of Computer Science and Statistics

Extra Questions ST3009: Statistical Methods for Computer Science

NOTE: There are more example questions in Chapter 8 of the course textbook "A First Course in Probability" by Sheldon Ross, but ignore those questions involving continuous-valued random variables.

Question 1. Consider a six sided die and let X be the number that we observe when it is thrown. We know that $E[X] = 3.5$.

(a) What is $P(X \geq 5)$?

(b) Using Markov's inequality obtain a bound on $P(X \geq 5)$. How does it compare with the exact value in (a) ?

Question 2. Sometimes I forget a few items when I leave the house in the morning. For example, here are the probabilities that I forget various pieces of footwear: left sock 0.2, right sock 0.1, left shoe 0.1, right shoe 0.3. Let X be the number of these that I forget.

(a) What is $E[X]$? Hint. Let X_1 be 1 when I forget my left sock and 0 otherwise, similarly $X_2 = 1$ when I forget my right sock, $X_3 = 1$ when I forget my left shoe and $X_4 = 1$ when I forget my right shoe. Then $X = X_1 + X_2 + X_3 + X_4$.

(b) Use the Markov Inequality to upper bound the probability that I forget 3 or more items.

Now suppose that I forget each item independently.

(c) What is Var(X) ?

(d) Use Chebyshev's inequality to upper bound the probability that I forget 2 or more items.

Question 3. A post office handles, on average, 10000 letters a day.

(a) Using Markov's inequality, what can be said about the probability that it will handle at least 15000 letters tomorrow ?

(b) Suppose now that the variance σ^2 in the number of letters per day is 2000. Using Chebyshev's inequality what can be said about the probability that this post office handles between 8000 and 12000 letters tomorrow?

(c) Using Chebyshev's inequality how can we bound the probability that it will handle at least 15000 letters tomorrow ? How does it compare with the bound in (a).

Question 4. A biased coin, which lands heads with probability 1/10 independently each time it is flipped, is flipped 200 times consecutively. Using Markov's inequality give a bound on the probability that it lands heads 120 times or more.

Question 5. Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

(a) What can be said about the probability that this weeks production will exceed 75?

(b) If the variance of a weeks production is known to equal 25, then what can be said about the probability that this weeks production will be between 40 and 60?

Hint: use Markov and Chebyshev inequalities.

Question 6. You would like to estimate the average number of hours p per day that a TCD student spends on youtube. To do this you plan to carry out a survey of the students by sampling n students independently and uniformly at random from the population. Letting

X_i be the number of hours spent by student i in the sample, suppose the mean can be estimated as $X = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(a) X is a Binomial random variable, with the X_i's all having mean p. Express $Var(X)$ in terms of p.

(b) Using the fact that $x(1-x) \leq 0.25$ for all $0 \leq x \leq 1$ (this can be verified using matlab), what is the maximum value of $Var(X)$?

(c) Using Chebyshev's inequality discuss how the value of n can be selected so as to ensure $P(|X-p| \geq 0.05) \leq 0.05$. Recall Chebyshev's inequality is $P(|X - E[X]| \geq k) \leq Var(X)/k^2$.

Question 7. In a study on cholestrol levels a sample of 12 men and women was chosen. The plasma cholestrol levels (mmol/L) of the subjects were as follows:

6.0,6.4,7.0,5.8,6.0,5.8,5.9,6.7,6.1,6.5,6.3,5.8

(a) Estimate the mean of the plasma cholestrol levels with a 95% confidence interval.

(b) What assumptions did you make about the sample in order to make your estimate ?