Statistics I, Fall 2012
Homework 06

Ling-Chieh Kung
Department of Information Management
National Taiwan University

1. (10 points) Let X follows an exponential distribution with rate λ. Prove that $\text{Var}(X) = \frac{1}{\lambda^2}$.

   Hint. Find $\mathbb{E}[X^2]$ first and then use $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. To find $\mathbb{E}[X^2]$, you need to apply integration by parts.

2. (10 points; 5 points each) Suppose that X_1 and X_2 are independent uniform random variables with lower bound 0 and upper bound 1.

   (a) Determine the value of $\mathbb{E}\big[(X_1 - 2X_2)^2\big]$.

   (b) Which part of your derivation requires independence? Note that the answer may be "it does not require independence at all."

3. (10 points) Suppose that one word is randomly selected from the sentence

   "I really love all my students because they always work hard on homework."[1]

   (a) (3 points) What is the probability that there are at least six letters in the selected word?

   (b) (3 points) What is the expectation of the number of letters in the selected word?

   (c) (4 points) What is the variance of the number of letters in the selected word?

4. (10 points) Let $X \geq 0$ be a discrete random variable. Prove that

$$\Pr(X \geq t) \leq \frac{\mathbb{E}[X]}{t} \quad \forall t > 0.$$

   Hint. Use $\mathbb{E}[X] = \sum_{x<t} x \Pr(x) + \sum_{x \geq t} x \Pr(x)$ and then find reasonable lower bounds for both terms in the right-hand-side. Checking the proof of Chebyshev's theorem may help.

5. (10 points) Consider a random sample $\{X_i\}_{i=1,2,\ldots,n}$ of sample size n from a population whose finite mean and finite variance are μ and σ^2, respectively. Assume that the population size $N >> n$ so that each X_i has mean μ and variance σ^2 and X_is are independent. Consider the *sample mean*

$$\overline{X} \equiv \frac{1}{n} \sum_{i=1}^{n} X_i.$$

   Note that the sample mean is also a random variable.

   (a) (3 points) Show that the expectation of the sample mean is $\mathbb{E}[\overline{X}] = \mu$.

   (b) (3 points) Show that the variance of the sample mean is $\text{Var}[\overline{X}] = \frac{\sigma^2}{n}$.

   (c) (4 points) Show that

$$\lim_{n \to \infty} \Pr(|\overline{X} - \mu| > \epsilon) = 0 \quad \forall \epsilon > 0.$$

   Roughly speaking, $\overline{X}$ approaches μ when the sample size becomes larger and larger.

   Note. The sample mean is one of the most important statistics used in statistical inference. The above two properties are both very important and will be used a lot later in this semester. Though we will have a formal discussion for them, at this moment you already have the ability to prove the above two results.

[1] For example, the probabilities of selecting "I" and "always" are the same.

6. (15 points) The MS Excel file "StatFa12_hw06.xlsx" contains a set of sample data. Use the data to answer the following questions.

    (a) (5 points) Draw a histogram or frequency polygon for the given set of data. Let the lower bound of the first class be -20 and the class interval be 15. Create twelve classes in total.

    (b) (10 points) What is the underlying distribution of this set of sample data? To answer this question, try to find a distribution that its histogram or frequency polygon best fits that of the sample data. Do not forget to illustrate how the distribution chosen by your fits the data.

        Note. You will get full credits as long as your approximation is not bad and makes sense.

7. (6 points; 3 points each; modified from Problem 6.11 in the textbook) Suppose you are working with a data set that is normally distributed with mean 180 and standard deviation 32. Determine the value of x from the following information. Answer the two questions independently.

    (a) 70% of the values are greater than x.

    (b) x is less than 20% of the values.

8. (9 points; modified from Problem 6.44 in the textbook) A business convention holds its registration on Wednesday morning from 9:00 AM until 12:00 noon. Past history has shown that registrant arrivals follow a Poisson distribution at an average rate of 1.6 every 15 seconds.

    (a) (2 points) What is the average number of seconds between arrivals for this conference?

    (b) (3 points) What is the probability that there are at least 10 arrivals in one minute?

    (c) (2 points) What is the probability that 25 seconds or more would pass between arrivals?

    (d) (2 points) What is the probability that fewer than five seconds will elapse between arrivals?

9. (10 points) Suppose X follows an exponential distribution with rate 3.

    (a) (2 points) What is the mean of $X + 2$?

    (b) (2 points) What is the variance of $X + 2$?

    (c) (3 points) What is the probability that $X + 2 > 1$?

    (d) (3 points) Is $X + 2$ an exponential random variable? If so, what is its rate? If not, why?

10. (10 points) Recall that in Problem 7 of Homework 5, we considered a random variable X which follows a binomial distribution with $n = 20$ and $p = 0.1$.

    (a) (3 points) If you use a normal random variable to approximate X, what should be the parameters of the normal random variable?

    (b) (3 points) WITHOUT correction of continuity, estimate $\Pr(X \leq 3)$ based on the normal distribution chosen in Part (a).

    (c) (4 points) WITH correction of continuity, estimate $\Pr(X \leq 3)$ based on the normal distribution chosen in Part (a).