Statistics I, Fall 2012
Homework 07

Instructor: Ling-Chieh Kung
Department of Information Management
National Taiwan University

Note. When you answer problems in this homework, you may want to use some results proved in the lecture. Feel free to do that as long as you clearly specify which results, if any, allow you to attain your conclusions.

1. (15 points; 5 points each) Let X be a random variable with the pdf

$$f(x) = 3x^2 \quad \forall x \in [0, 1].$$

Answer the following questions without using its moment generating function.

   (a) Find its first moment $\mathbb{E}[X]$.

   (b) Find its third moment $\mathbb{E}\big[X^3\big]$.

   (c) Prove that

$$\lim_{k \to \infty} \mathbb{E}\big[X^k\big] = 0.$$

2. (20 points) In this problem, we will show you that sometimes using a moment generating function to derive moments can be much harder than working with the definitions directly. Consider the random variable X in Problem 1 again.

   (a) (10 points) Prove that its moment generating function is

$$m(t) = \frac{3}{t^3}\Big[\big(t^2 - 2t + 2\big)e^t - 2\Big].$$

   (b) (5 points) Prove that

$$\frac{d}{dt}m(t) = m'(t) = \frac{3}{t^4}\Big[\big(t^3 - 3t^2 + 6t - 6\big)e^t + 6\Big].$$

   (c) (5 points) Note that we cannot plug in $t = 0$ into the above equation because the denominator will become 0. Equivalently, this is saying that $m'(t)$ is undefined at $t = 0$. Therefore, $\frac{d}{dt}m(t)|_{t=0}$ must be evaluated by calculating the limit of $m'(t)$ as t approaches 0. Prove that

$$\lim_{t \to 0} m'(t) = \frac{3}{4},$$

   which is the first moment as we calculated in Problem 1.

   Hint. You may want to use L'Hôpital's rule.

3. (15 points) Let $X \sim \text{Uni}(a, b)$. Prove that the moment generating function of X is

$$m(t) = \frac{e^{tb} - e^{ta}}{t(b - a)}.$$

4. (15 points) Let $\{X_i\}_{i=1,\dots,n}$ be a sample from a normal population with mean μ and standard deviation σ. Let $\overline{X} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean.

   (a) (10 points) Prove that

$$\overline{X} \sim \text{ND}\Big(\mu, \frac{\sigma}{\sqrt{n}}\Big).$$

(b) (5 point) Typically we denote $\frac{\sigma}{\sqrt{n}}$ as $\sigma_{\overline{X}}$. This quantity, which is called the *standard error*, is the standard deviation of the sample mean. Prove that

$$\frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \sim \mathrm{ND}(0, 1).$$

5. (10 points) Let $X_1 \sim \mathrm{Bi}(n_1, p)$ and $X_2 \sim \mathrm{Bi}(n_2, p)$. Suppose X_1 and X_2 are independent, prove that

$$X_1 + X_2 \sim \mathrm{Bi}(n_1 + n_2, p).$$

6. (25 points) Suppose we sampled n values from a normal population with mean 120 and standard deviation 40.

   (a) (5 points) What is the distribution of the sample mean?

   (b) (5 points) Suppose $n = 16$, what is the probability for the sample mean to deviate from the population mean by 6 (i.e., above 126 or below 114)?

   (c) (5 points) Suppose $n = 100$, what is the probability for the sample mean to deviate from the population mean by 6?

   (d) (10 points) What is the smallest sample size that will make the probability for the sample mean to deviate from the population mean by 6 be no greater than 1%?

Note. In practice, people use the sample mean to estimate the population mean. The above exercises show that, once we know the *sampling distribution* of a statistic (the sample mean in this example), the quality of the estimation can be ensured: As long as we can increase the sample size, we can reduce the probability that "our guess is too far from the truth." This is the central theme of statistical estimation and will be discussed further in Chapter 8. In fact, this shows why we need to spend so many weeks in learning Probability: We want to make our estimation *scientific* by controlling the degree of errors.