Statistics I, Fall 2012
Homework 05

Ling-Chieh Kung
Department of Information Management
National Taiwan University

1. (15 points) Consider two random variables X and Y with means μ_X and μ_Y, respectively. Their covariance is defined as
$$\text{Cov}(XY) \equiv \mathbb{E}\Big[(X-\mu_X)(Y-\mu_Y)\Big].$$
Please remind yourself how this definition for two random variables looks like the one we introduced in Chapter 3 for two sets of data. Let σ_X and σ_Y be the standard deviations of X and Y, respectively, then the correlation coefficient of X and Y is defined as $\frac{\text{Cov}(XY)}{\sigma_X\sigma_Y}$.

   (a) (5 points) Suppose the joint probability table of X and Y is in Table 1. For example, the joint probability $\Pr(X=2, Y=1) = 0.4$. Find $\text{Cov}(XY)$ and the correlation coefficient.

x	y			Total
	1	4	5	
2	0.4	0.2	0.1	0.7
3	0.05	0.1	0.15	0.3
Total	0.45	0.3	0.25	1

Table 1: Joint probability table for 1a.

w	z			Total
	1	4	5	
2	0.2	0.1	0.3	0.6
3	0.1	0.2	0.1	0.4
Total	0.3	0.3	0.4	1

Table 2: Joint probability table for 1c.

   (b) (2 points) Are X and Y independent? Why?

   (c) (5 points) Suppose the joint probability table of another two random variables W and Z is in Table 2. For example, the joint probability $\Pr(W=2, Z=1) = 0.2$. Find $\text{Cov}(WZ)$ and the correlation coefficient.

   (d) (3 points) Are W and Z independent? Why?

2. (10 points) The variance of a hypergeometric random variable with parameters N, A, and n is
$$np(1-p)\left(\frac{N-n}{N-1}\right).$$

   (a) (5 points) When $n = N$, the variance becomes 0. Briefly explain why there is no variability.

   (b) (5 points) When $n = 1$, the variance becomes $np(1-p)$. Briefly explain why the variance is the same as that in a sampling with replacement.

3. (15 points) Let X be a hypergeometric random variable with population size N, number of "1"s A, and the sample size n. Let $p = \frac{A}{N}$. In this problem, we will derive the mean and variance of X.

   (a) (3 points) Let X_i be the outcome of the ith trial, $i = 1, 2, ..., n$. We may show that $\Pr(X_i = 1) = p$ for all $i = 1, ..., n$. To see this, first note that this is trivial for $i = 1$. Now let's focus on $i = 2$. Use the law of total probability
$$\Pr(X_2 = 1) = \Pr(X_2 = 1 | X_1 = 1)\Pr(X_1 = 1) + \Pr(X_2 = 1 | X_1 = 0)\Pr(X_1 = 0)$$
to show that $\Pr(X_2 = 1) = p$. Then convince yourself (WITHOUT writing anything on your answer sheet) that we may repeat this process (with tedious calculations) to show that $\Pr(X_i = 1) = p$ for all $i = 3, ..., n$.

   (b) (2 points) Show that $\mathbb{E}[X_i] = p$ and $\text{Var}(X_i) = p(1-p)$ for all $i = 1, ..., n$.

(c) (2 points) Show that $\mathbb{E}[X] = np$.
Hint. Use $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

(d) (3 points) Now we want to calculate $\text{Var}(X)$. Unfortunately, even though all X_is have the same mean and variance, they are still dependent. Therefore, we need to use the following general formula

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \text{Cov}(X_i, X_j).$$

Prove a special case of the above formula: $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$.

(e) (2 points) Consider our sampling without replacement experiment. It can be shown (with tedious calculations) that $\text{Cov}(X_i, X_j) = \text{Cov}(X_1, X_2)$ for all $i \neq j$. Take this result as given (so you do not need to prove it), show that $\text{Var}(X) = np(1-p) + n(n-1)\text{Cov}(X_1, X_2)$.

(f) (3 points) Find $\text{Cov}(X_1, X_2)$ by testing the result you obtained in Part (e) with a very special case: $n = N$. Then show that, for any $n < N$, we have $\text{Var}(X) = np(1-p)\left(\frac{N-n}{N-1}\right)$.

4. Let X follows a uniform distribution with the lower bound a and the upper bound b.

(a) Prove that $\mathbb{E}[X] = \frac{a+b}{2}$.

(b) Prove that $\text{Var}(X) = \frac{(b-a)^2}{12}$.

5. (10 points) Let X be a continuous random variable whose pdf is

$$f(x) = \begin{cases} ke^{2x} & \text{if } x \in [1,2] \\ 0 & \text{otherwise} \end{cases}.$$

(a) (5 points) Find the value of k.

(b) (5 points) Find $\mathbb{E}[X]$.
Hint. You may need to use integration by parts.

6. (10 points) Let X be a continuous random variable whose pdf is

$$f(x) = \begin{cases} \frac{kx}{4} & \text{if } x \in [0,1] \\ \frac{k(4-x)}{12} & \text{if } x \in [1,4] \\ 0 & \text{otherwise} \end{cases}.$$

(a) (3 points) Find the value of k.
Hint. You may need to express $\int_0^4 f(x)dx$ as the sum of two integrals.

(b) (3 points) Find $\mathbb{E}[X]$.

(c) (4 points) Find the cdf $F(x)$.

7. (10 points) Suppose that X follows a binomial distribution with $n = 20$ and $p = 0.1$.

(a) (5 points) Find the exact value of $\Pr(X \leq 3)$.

(b) (5 points) Use a Poisson random variable to approximate X and then estimate $\Pr(X \leq 3)$ based on a Poisson distribution. Is the approximation good? Why or why not?

8. (10 points) The number of typing errors made by a typist follows a Poisson distribution with an average of two errors per page.

(a) (5 points) If more than four errors appear on a page, the typist will not get her salary for that page. What is the probability that this happens for a given page?

(b) (5 points) Suppose that she is working on a booklet with 20 pages and she may get an extra bonus if there are no more than 20 errors in the whole booklet. What is the probability that she earns the bonus?

9. (10 points) At a gasoline station, the daily demand follows a uniform distribution with mean 13 kiloliters and standard deviation $\sqrt{3}$ kiloliters.

   (a) (3 points) What are the two parameters of the uniform distribution?

   (b) (3 points) Suppose the station has prepared 15 kiloliters of gasoline at the beginning of one day. What is the probability that it runs out of gasoline by the end of that day?

   (c) (4 points) Suppose the station wants to achieve a service level of 99%, i.e., the probability of running out of gasoline is only 1%. Find the amount of gasoline that they should prepare.