Economics 327 **Term I 2016**

Introduction to Empirical Methods **Hiro Kasahara**

Midterm Exam

1. (10 points) In one year, the average stock price of Google Inc. was \$800 with the standard deviation equal to \$100. Using the empirical rule, it can be estimated that approximately 95 % of the stock price of Google Inc. will be in what interval?

   Answer: The interval is given by $800 \pm 2 \times 100$ or $[600, 1000]$.

2. Let $X = 1$ if a person has completed 4 year university; $X = 0$ otherwise. Let Y be the annual income. For simplicity, assume that Y takes the following three values: 30, 60, and 100 in thousand \$. Suppose that the joint probability of education attainment and annual income of person is given by Table 1.

   (a) (10 points) Find the value of $E[X]$.

   Answer: $P(X = 0) = 0.24 + 0.12 + 0.04 = 0.4$ and $P(X = 1) = 0.6$. $E[X] = 0 \times 0.4 + 1 \times 0.6 = 0.6$.

   Grading: 10 points if the final answer is correct.

   (b) (10 points) Find the value of $E[Y|X = 1] - E[Y|X = 0]$.

   Answer: $E[Y|X = 0] = 30 \times 0.6 + 60 \times 0.3 + 100 \times 0.1 = 46$ and $E[Y|X = 1] = 30 \times 0.2 + 60 \times 0.6 + 100 \times 0.2 = 62$. Therefore, $E[Y|X = 1] - E[Y|X = 0] = 62 - 46 = 16$.

   Grading: 10 points if the final answer is correct.

Table I: Joint Distribution of University Degree and Annual Incomes (in thousand \$)

	Y = 30	Y = 60	Y = 100
X = 0	0.24	0.12	0.04
X = 1	0.12	0.36	0.12

3. The survey asks eligible voters in the U.S. whether he or she would vote for Clinton or Trump. An individual i's voting preference is recorded as $X_i = 1$ if she/he would vote for Clinton and as $X_i = 0$ if she/he would vote for Trump. Suppose that we randomly select two persons and let their voting preference to be represented by X_1 and X_2, where the probability distribution of X_1 is identical to that of X_2 and is given by

$$X_i = \begin{cases} 0 & \text{with probability } 1 - p \\ 1 & \text{with probability } p, \end{cases}$$

for $i = 1, 2$. Here, p represents the population fraction of voters who would vote for Clinton. Define a random variable $\bar{X}$ by $\bar{X} = (X_1 + X_2)/2$.

(a) (10 points) Derive $E[X_1]$ and $Var[X_1]$.

Answer: $E[X_1] = 0 \times (1-p) + 1 \times p = p$. $Var[X_1] = E[(X_1 - p)^2] = E[X_1^2 + p^2 - 2pX_1]$. Note $X_1^2 = X_1$ because $1 \times 1 = 1$ and $0 \times 0 = 0$. Therefore, $E[X_1^2 + p^2 - 2pX_1] = E[X_1 + p^2 - 2pX_1] = E[X_1] + p^2 - 2pE[X_1] = p + p^2 - 2p^2 = p - p^2 = p(1-p)$.

Grading: 5 points for each. Out of 5 points, 4 points are given to the final answer and 1 point is given to the derivation.

(b) (10 points) Derive $E[\bar{X}]$.

Answer: $E[\bar{X}] = (1/2)E[X_1 + X_2] = (1/2)(E[X_1] + E[X_2]) = (1/2)(p + p) = p$.

Grading: 5 points are given to the final answer and 5 points are given to the derivation.

(c) (10 points) Derive $Var[\bar{X}]$.

Answer: $Var[\bar{X}] = E[(\bar{X} - p)^2] = E[\{(1/2)(X_1 - X_2) - \}^2] = E[\{(1/2)[(X_1 - p) + (X_2 - p)]\}^2] = (1/4)E[\{(X_1 - p) + (X_2 - p)\}^2] = (1/4)E[\{[(X_1 - p) + (X_2 - p)\}^2] = (1/4)E[(X_1 - p)^2 + (X_2 - p)^2 + 2(X_1 - p)(X_2 - p)] = (1/4)\left\{E[(X_1 - p)^2] + E[(X_2 - p)^2] + 2E[(X_1 - p)(X_2 - p)]\right\} = (1/4)\{Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)\} = (1/4)(p(1-p) + p(1-p) + 0) = \frac{p(1-p)}{2}$. Note that the third equality uses $(1/2)(X_1 - X_2) - p = (1/2)(X_1 - X_2) - p = (1/2)[(X_1 - p) + (X_2 - p)]$ while the second to the last equality uses the independence between X_1 and X_2. We may also show $Var[\bar{X}] = p(1-p)$ by using $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$, i.e., $Var[\bar{X}] = (1/4)Var(X_1 + X_2) = (1/4)Var(X_1) + (1/4)Var(X_2) + (1/2)Cov(X_1, X_2) = (1/4)p(1-p) + (1/4)p(1-p) + 0 = \frac{p(1-p)}{2}$.

Grading: 2 points are given to the final answer and 8 points are given to the derivation.

4. (10 points) For any random variable X, define $W = \frac{X - E(X)}{\sqrt{Var(X)}}$. Derive $E[W]$ and $Var[W]$.

Answer: $E[W] = E\left(\frac{X - E(X)}{\sqrt{Var(X)}}\right) = \frac{E(X - E(X))}{\sqrt{Var(X)}} = \frac{E(X) - E(X)}{\sqrt{Var(X)}} = 0$.

$Var(W) = Var\left(\frac{X - E(X)}{\sqrt{Var(X)}}\right) = \frac{Var(X - E(X))}{Var(X)} = \frac{E\big((X - E(X) - E[X - E(X)])^2\big)}{Var(X)} = \frac{E\big((X - E(X))^2\big)}{Var(X)} = \frac{Var(X)}{Var(X)} = 1$.

Grading: 5 points for each. Out of 5 points, 4 points are given to the final answer and 1 point is given to the derivation.

5. (10 points) Let X and Y be two discrete random variables. The set of possible values for X is $\{x_1, \ldots, x_n\}$; and the set of possible values for Y is $\{y_1, \ldots, y_m\}$. The joint function of X and Y is given by $p_{ij}^{X,Y} = P(X = x_i, Y = y_j)$ for $i = 1, \ldots n; j = 1, \ldots, m$. The marginal probability function of X is $p_i^X = P(X = x_i) = \sum_{j=1}^m p_{ij}^{X,Y}$ for $i = 1, \ldots n$, and the marginal probability function of Y is $p_j^Y = P(Y = y_j) = \sum_{i=1}^n p_{ij}^{X,Y}$ for $j = 1, \ldots m$. Let $g_1(x)$ and $g_2(y)$ are some functions of x and y, respectively. Prove that, if random variable X and Y are independent, then $\text{Cov}(g_1(X), g_2(Y)) = 0$. **Please use the summation operator in the proof for this question**.

Answer:

$$
\begin{aligned}
& Cov(g_1(X), g_2(Y)) \\
&= E[(g_1(X) - E(g_1(X)))(g_2(Y) - E(g_2(Y)))] \\
&= \sum_{i=1}^n \sum_{j=1}^m [g_1(x_i) - E(g_1(X))][g_2(y_j) - E(g_2(Y))] p_i^X p_j^Y \quad \text{(because } X \text{ and } Y \text{ are independent)} \\
&= \sum_{i=1}^n [g_1(x_i) - E(g_1(X))] p_i^X \left\{ \sum_{j=1}^m [g_2(y_j) - E(g_2(Y))] p_j^Y \right\}
\end{aligned}
$$

(we can move $[g_1(x_i) - E(g_1(X))]p_i^X$ outside of $\sum_{j=1}^m$ given that it does not depend on j's)

$$
\begin{aligned}
&= \sum_{i=1}^n [g_1(x_i) - E(g_1(X))] p_i^X \left\{ \sum_{j=1}^m g_2(y_j) p_j^Y - \sum_{j=1}^m E(g_2(Y)) p_j^Y \right\} \\
&= \sum_{i=1}^n [g_1(x_i) - E(g_1(X))] p_i^X \left\{ E(g_2(y_j)) - E(g_2(Y)) \underbrace{\sum_{j=1}^m p_j^Y}_{=1} \right\} \\
&= \sum_{i=1}^n [g_1(x_i) - E(g_1(X))] p_i^X \times 0 = 0.
\end{aligned}
$$

Grading: If the definition is wrong, you get zero. Another key part is $p_{ij}^{X,Y} = p_i^X p_j^Y$. If this is not written, then you lose at least 5 points.

6. (10 points) Prove that $Cov(X, Y) = E[XY] - E[X]E[Y]$. In this proof, you don't necessarily use the summation operator.

Answer: $Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E[XY - E(X)Y - E(Y)X + E(X)E(Y)] = E[XY] - E[E(X)Y] - E[E(Y)X] + E[E(X)E(Y)] = E[XY] - E(Y)E(X) - E(X)E(Y) + E(X)E(Y) =$

$E[XY] - E(X)E(Y)$.

Grading: If the definition is wrong, you get zero.

7. (10 points) Let a, b, and c be some constant. Prove that $Var(aX + bY + c) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$. In this proof, you don't necessarily use the summation operator.

   Answer: $Var(aX+bY+c) = E[(aX+bY+c-E[aX+bY+c])^2]$. Note that $aX+bY+c-E[aX+bY+c] = a(X-E(X))+b(Y-E(Y))$. Therefore, $E[(aX+bY+c-E[aX+bY+c])^2] = E[(a(X-E(X))+b(Y-E(Y)))^2] = E[a^2(X-E(X))^2+b^2(Y-E(Y)))^2+2ab(X-E(X))(Y-E(Y))] = a^2E[(X-E(X))^2] + b^2E[(Y-E(Y)))^2] + 2abE[(X-E(X))(Y-E(Y))] = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$.

   Grading: If the definition is wrong, you get zero.