Economics 325–003 **Term I 2017**

Introduction to Empirical Economics **Hiro Kasahara**

Midterm Exam[1]

1. (8 points) In one year, the average stock price of Microsoft Corp. was \$77 with the standard deviation equal to \$5. Using the empirical rule, it can be estimated that approximately 95 % of the stock price of Microsoft Corp. will be in what interval?

   Answer: $77 \pm 2 \times 5 = [67, 87]$.

2. (8 points) Multiple choice. **Write your answer (i.e., A, B, C, or D) to your booklet.** No explanation necessary. For any two events A and B, consider the following two statements:

   1. $(A \cap B) \cap (A \cap \bar{B})$ is empty.
   2. $A = (A \cap B) \cup (A \cap \bar{B})$.

   Which of the followings is true?

   A). Both 1 and 2 are true, B). Both 1 and 2 are false,

   C). 1 is true and 2 is false, D). 1 is false and 2 is true.

   Answer: A.

3. (10 points) Suppose a personnel officer has 6 candidates to fill 2 positions. Among 6 candidates, 3 candidates are men and 3 candidates are women. If every candidate is equally likely to be chosen, what is the probability that no women will be hired?

   Answer: There are $C_2^6 = \frac{6!}{(6-2)!2!} = 15$ possible basic outcomes and, among them, $C_2^3 = \frac{3!}{(3-2)!2!} = 3$ cases for which no women will be hired. Therefore, the probability that no women will be hired is $\frac{C_2^3}{C_2^6} = \frac{3}{15} = \frac{1}{5} = 0.2$.

4. Suppose you took a blood test for cancer diagnosis and your blood test was positive. This blood test is positive with probability 90 percent if you indeed have a cancer. This blood test is positive with probability 5 percent if you dont have a cancer. In population, it is known that 1 percent of people have cancer. Define $A = \{\text{positive blood test}\}$ and $B = \{\text{having a cancer}\}$.

(a) (6 points) What is the probability of your having a cancer given that your blood test is positive?

Answer: Note that $P(A|B) = 0.9$, $P(A|\bar{B}) = 0.05$, $P(B) = 0.01$, and $P(\bar{B}) = 0.99$. Therefore, using the Bayes' theorem,

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})} = \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.05 \times 0.99} = \frac{0.009}{0.009 + 0.0495} \approx 0.1538.$$

(b) (6 points) Complete the following table for the joint distribution of A and B and **write your answer to your booklet**, where $\bar{A} = \{\text{negative blood test}\}$ and $\bar{B} = \{\text{not having a cancer}\}$.

Table I: Joint distribution of A and B

	$\bar{B}$	B	Marginal Prob.
$\bar{A}$			
A			
Marginal Prob	0.99	0.01	1.00

Answer: $P(A \cap B) = P(A|B)P(B) = 0.9 \times 0.01 = 0.009$, $P(A \cap \bar{B}) = P(A|\bar{B})P(\bar{B}) = 0.05 \times 0.99 = 0.0495$, $P(\bar{A} \cap \bar{B}) = P(\bar{A}|\bar{B})P(\bar{B}) = (1 - P(A|\bar{B}))P(\bar{B}) = (1 - 0.05) \times 0.99 = 0.9405$, and $P(\bar{A}|B) = (1 - P(A|B))P(B) = (1 - 0.9) \times 0.01 = 0.001$. In sum, we have the following table for the joint distribution of A and B.

Joint distribution of A and B

	$\bar{B}$	B	Marginal Prob.
$\bar{A}$	0.9405	0.001	0.9415
A	0.0495	0.009	0.0585
Marginal Prob	0.99	0.01	1.00

5. In the household survey, we asked if the household head has completed 4 year university degree or not and asked their annual incomes. Define the random variables X and Y as follows: X = 1 if the household head has completed 4 year university; X = 0 otherwise. Y is the annual incomes. For simplicity, assume that Y takes two values: 50 and 100 in thousand \$. The joint distribution of X and Y is given in Table II.

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

	Y = 50	Y = 100	Marginal Prob. of X
X = 0	0.30	0.10	0.40
X = 1	0.20	0.40	0.60
Marginal Prob. of Y	0.50	0.50	1.00

(a) (6 points) Are X and Y stochastically independent? **Prove your claim.** [Full 6 points are given to the proof.]

Answer: No. To prove this, it suffices to show that $P(X,Y) \neq P(X)P(Y)$ for some (X,Y). $P(X = 0, Y = 50) = 0.30$ and $P(X = 0)P(Y = 50) = 0.4 \times 0.5 = 0.20$. Therefore, because $P(X = 0, Y = 50) \neq P(X = 0)P(Y = 50)$, X and Y are not stochastically independent.

(b) (6 points) What is the conditional probability mass function of Y given $X = 1$? Answer:

$$f_{Y|X=1}(y) = \begin{cases} 0.2/0.6 = 1/3 & \text{if y=50} \\ 0.4/0.6 = 2/3 & \text{if y=100.} \end{cases}$$

(c) (6 points) Find the value of $E[Y|X = 1] - E[Y|X = 0]$. Answer: From the previous question, $E[Y|X = 1] = 50 \times (1/3) + 100 \times (2/3) = 250/3 \approx$. For $E[Y|X = 0]$, note that the conditional probability mass function of Y given $X = 0$ is

$$f_{Y|X=1}(y) = \begin{cases} 0.3/0.4 = 3/4 & \text{if y=50} \\ 0.1/0.4 = 1/4 & \text{if y=100.} \end{cases}$$

Therefore, $E[Y|X = 0] = 50 \times (3/4) + 100 \times (1/4) = 250/4$. Finally, $E[Y|X = 1] - E[Y|X = 0] = 250(1/3 - 1/4) = 250/12 \approx 20.83$.

(d) (6 points) Show that the law of iterated expectations $E_X[E_Y[Y|X]] = E_Y[Y]$ holds for this example.

Answer: Using the marginal probability mass function of Y, we have $E_Y[Y] = 50\times0.5+100\times0.5 = 75$. On the other hand, $E_X[E_Y[Y|X]] = E_Y[Y|X=0]P(X=0) + E_Y[Y|X=1]P(X=1) = (250/4)\times0.4+(250/3)\times0.6 = 75$. [Full points are allocated to the understanding of the meaning of $E_X[E_Y[Y|X]]$, i.e., namely whether the student correctly expresses $E_X[E_Y[Y|X]] = E_Y[Y|X = 0]P(X=0) + E_Y[Y|X=1]P(X=1)$.]

6. (8 points) The daily stock prices of a company named "Data Science" are known to be normally distributed with a mean of \$100 and a standard deviation of \$10. What proportion of daily stock prices is larger than \$120? Answer: The z-value corresponding to $X = 120$ is $Z = \frac{120-100}{10} = 2$. Therefore, $P(X > 120) = P(Z > 2) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$.

7. (10 points) Suppose that $\{X_1, X_2, ..., X_n\}$ is a random sample (so that X_1, X_2, ..., X_n are stochastically independent), where X_i takes a value of zero or one with probability $1-p$ and p, respectively. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Show that $E(\bar{X}) = p$ and $\text{Var}(\bar{X}) = \frac{p(1-p)}{n}$.

Answer:

$$\begin{aligned}
E(\bar{X}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}E\left(X_1 + X_2 + ... + X_n\right) \\
&= \frac{1}{n}\left\{E(X_1) + E(X_2) + ... + E(X_n)\right\} \\
&= \frac{1}{n}\left\{p + p + ... + p\right\} \\
&= \frac{np}{n} = p.
\end{aligned}$$

Let $\tilde{X}_i = X_i - E[X_i]$.

$$
\begin{aligned}
Var(\bar{X}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \left(\frac{1}{n}\right)^2 Var\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2} E\left(\left\{\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right]\right\}^2\right) \\
&= \frac{1}{n^2} E\left(\{(X_1 - E[X_1]) + (X_2 - E[X_2]) + ... + (X_n - E[X_n])\}^2\right) \\
&= \frac{1}{n^2} E\left(\left\{\tilde{X}_1 + \tilde{X}_2 + ... + \tilde{X}_n\right\}^2\right) \\
&= \frac{1}{n^2} E\left(\tilde{X}_1^2 + \tilde{X}_2^2 + ... + \tilde{X}_n^2 + 2\tilde{X}_1\tilde{X}_2 + ... + 2\tilde{X}_{n-1}\tilde{X}_n\right) \\
&= \frac{1}{n^2} \left(Var(X_1) + Var(X_2) + ... + Var(X_n) + 2Cov(X_1, X_2) + ... + 2Cov(X_{n-1}, X_n)\right) \\
&= \frac{1}{n^2} \left(p(1-p) + p(1-p) + ... + p(1-p) + 0 + ... + 0\right) \\
&= \frac{np(1-p)}{n^2} \\
&= \frac{p(1-p)}{n}.
\end{aligned}
$$

8. 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 $j = 1, \ldots m$.

   (a) (10 points) Prove that, if X and Y are stochastically independent, then $Cov(g(X), a + bY) = 0$ for any function g and any constant a and b. **Please use the summation operator in the proof for this question**.

Answer:

$$
\begin{aligned}
Cov(g(X), a+bY) &= E[(g(X)-E(g(X)))(a+bY-E(a+bY))] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}[g(x_i)-E(g(X))][a+by_j-E(a+bY)]p_i^X p_j^Y \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\{[g(x_i)-E(g(X))]p_i^X\}b\{[y_j-E(Y)]p_j^Y\} \\
&= b\sum_{i=1}^{n}[g(x_i)-E(g(X))]p_i^X\left\{\sum_{j=1}^{m}[y_j-E(Y)]p_j^Y\right\} \\
&= b\left\{\sum_{j=1}^{m}[y_j-E(Y)]p_j^Y\right\}\left\{\sum_{i=1}^{n}[g(x_i)-E(g(X))]p_i^X\right\} \\
&= b\left\{\sum_{i=1}^{n}g(x_i)p_i^X-\sum_{i=1}^{n}E(g(X))p_i^X\right\}\cdot\left\{\sum_{j=1}^{m}y_jp_j^Y-\sum_{j=1}^{m}E(Y)p_j^Y\right\} \\
&= b\left\{E(g(X))-\sum_{i=1}^{n}E(g(X))p_i^X\right\}\cdot\left\{E(Y)-\sum_{j=1}^{m}E(Y)p_j^Y\right\} \\
&= b\left\{E(g(X))-E(g(X))\sum_{i=1}^{n}p_i^X\right\}\cdot\left\{E(Y)-E(Y)\sum_{j=1}^{m}p_j^Y\right\} \\
&= b\left\{E(g(X))-E(g(X))\cdot 1\right\}\cdot\left\{E(Y)-E(Y)\cdot 1\right\} \\
&= 0\cdot 0 = 0.
\end{aligned}
$$

(b) (10 points) Define $Z=\frac{X-E(X)}{\sqrt{Var(X)}}$ and $W=\frac{X-a}{\sqrt{Var(X)}}$ for some constant a. Prove that $\mathrm{Corr}(Z,W)=1$ for any a. (You don't necessarily need to use the summation operator).

Answer: We may compute $Cov(Z,W)$, $Var(Z)$ and $Var(W)$.

$$
\begin{aligned}
Cov(Z,W) &= E\left((Z-E(Z))(W-E(W))\right) \\
&= E\left(\left\{(X-E(X))/\sqrt{Var(X)}-0\right\}\left\{(X-a)/\sqrt{Var(X)}-E((X-a)/\sqrt{Var(X)})\right\}\right) \\
&= \frac{E\left(\{X-E(X)\}\{(X-a)-E(X-a)\}\right)}{\sqrt{Var(X)}\times\sqrt{Var(X)}} \\
&= \frac{Var(X)}{Var(X)} = 1.
\end{aligned}
$$

$$Var(Z) = Var\left((X - E(X))/\sqrt{Var(X)}\right) = \left(\frac{1}{\sqrt{Var(X)}}\right)^2 Var\left((X - E(X))\right) = \frac{Var(X)}{Var(X)} = 1,$$

$$Var(W) = Var\left((X - a)/\sqrt{Var(X)}\right) = \left(\frac{1}{\sqrt{Var(X)}}\right)^2 Var\left((X - a)\right) = \frac{Var(X)}{Var(X)} = 1,$$

where we used the fact that $Var\left((X - a)\right) = Var\left(X\right)$ holds for any a because $Var\left((X - a)\right) = E[\{X - a - E[X - a]\}^2] = E[\{X - E[X]\}^2] = Var(X)$. Therefore,

$$Corr(Z, W) = \frac{Cov(Z, W)}{\sqrt{Var(Z)}\sqrt{Var(W)}} = \frac{1}{1 \times 1} = 1.$$