Economics 250 — Midterm 1 (answers) **20 October 2016**

Instructions: You may use a hand calculator.

Do not hand in the question and formula sheets. You can keep (or discard) these.

Answer all **three** questions in the answer booklet provided.

There are 80 points in total and 80 minutes to complete the exam.

Show your work: incorrect answers without any work shown cannot be given partial marks. If using your calculator, it is acceptable to only write down the first few terms of a calculation.

Formulas and tables are provided at the end of the question pages; you may detach these from the question pages for easier reference.

Good luck!

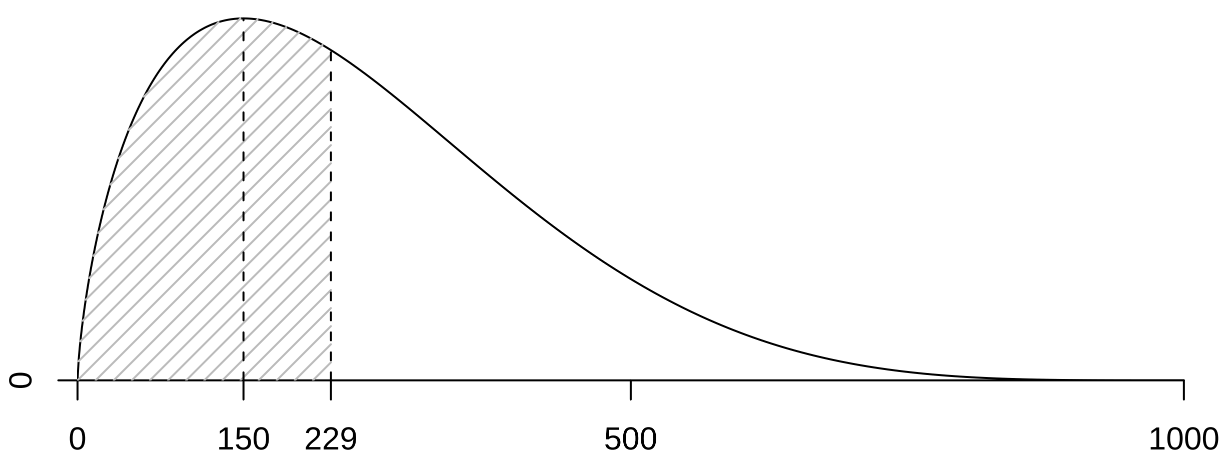
1. **[20]** The following table shows box office revenues and production costs for a sample of five recent movies (in millions of dollars).

Box office revenue	63	262	128	300	645
Production costs	90	136	9	95	282

a) Calculate the mean and standard deviation of box office revenue.

b) The mean and variance of production costs are 122.4 and 10079.3, respectively. Calculate the correlation between box office revenue and production costs.

Suppose box office revenue is known to always be between 0 and 1000 and follows the distribution with density shown below, where the shaded region is exactly half of the area under the curve.



c) What is the area under the curve between 229 and 1000?

d) Is this distribution left-skewed, symmetric, or right-skewed?

e) What are the mean, mode, and median of this distribution? If one or more of these cannot be determined directly from the graph, indicate where the value would be relative to the given values (for example "between 500 and 1000").

Answers:

a)

$$
\begin{aligned}
\overline{x} &= \frac{1}{5}(63 + 262 + 128 + 300 + 645) = \frac{1398}{5} = 279.6 \\
s_x^2 &= \frac{1}{4}\big((63 - 279.6)^2 + (262 - 279.6)^2 + (128 - 279.6)^2 + (300 - 279.6)^2 \\
&\qquad + (645 - 279.6)^2\big) \\
&= 51035.3 \\
s_x &= \sqrt{s_x^2} = \sqrt{51035.3} = 225.9
\end{aligned}
$$

b) The correlation calculation is: $r_{xy} = \frac{s_{xy}}{s_x s_y}$. The answer to part *a*) gives us $\overline{x} = 279.6$ and $s_x = 225.9$, and the question itself gives us $\overline{y} = 122.4$ and $s_y^2 = 10079.3$.

We still need to calculate s_y and s_{xy}:

$$
\begin{aligned}
s_y &= \sqrt{s_y^2} = \sqrt{10079.3} = 100.4 \\
s_{xy} &= \frac{1}{4}\left(\sum_{i=1}^{5}(x_i - \overline{x})(y_i - \overline{y})\right) \\
&= \frac{1}{4}\big((63 - 279.6)(90 - 122.4) + (262 - 279.6)(136 - 122.4) + \\
&\qquad (128 - 279.6)(9 - 122.4) + (300 - 279.6)(95 - 122.4) + \\
&\qquad (645 - 279.6)(282 - 122.4)\big) \\
&= \frac{81728.8}{4} \\
&= 20432.2
\end{aligned}
$$

and so putting it all together we get:

$$
r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{20432.2}{225.9 \times 100.4} = 0.9009
$$

For reference: the depicted distribution (and the sample values in part *a*)) are a "Beta" distribution with parameters $\alpha = 1.706, \beta = 5$, rescaled to cover the range 0–1000 (a basic beta distribution has range 0–1). That is obviously not a distribution we talked about, and you don't need to know that to answer the questions.

c) Since this is a density curve, we know the area under the entire curve must equal 1. Since the shaded area equals 0.5, that means the *unshaded* area also equals **0.5**.

d) **Right-skewed**: the density clearly has a longer right-tail than left-tail.

e) The mode is **150**, the median is **229**, and the mean is **between 229 and 500**.

Explanation:

The mode is the most common (or most likely) value. For a continuous distribution like this one, that is the value that has the highest probability density: 150.

The median is the value that divides the distribution in half: half the time values from this distribution will lie below the median, and half the values lie above the median. Since the shaded area equals 0.5, the median is the edge of this shaded area, i.e. 229.

The mean is between 229 and 500 (in actual fact, it is at 254.4 for this distribution).

There are two ways to get this answer: first, the mean is the "balancing point" of the distribution, and is affected by the length and size of the tails. Looking at it from the median, we have the same total area on other side, but the right tail extends further, That means the mean will be more affected by extreme values on the right of the median than on the left, and so the mean will be to the right of the median.

Another way to look at this is that, since the distribution is right-skewed, the mean lies above the median (this is a property of a right-skewed distribution).

We can also easily tell than the mean will be less than 500: 500 is the middle of the range of values of the distribution, but this is clearly not the balancing point because there is much more mass to the left of 500 than to the right.

Thus we conclude that the mean will be somewhere between 229 and 500.

2. [**30**] An investor is considering making an investment of $1000 into a stock. Suppose that X represents the dollar value of the investment after one year, and is known to be uniformly distributed between 900 and 1200, with standard deviation of 86.60.

a) What is the mean value of the investment after a year?

b) What is the probability that the investor loses value (that is, that the value after a year is less than the initial investment)?

c) What is the probability of seeing a change in value of more than $70 (either an increase *or* a decrease)?

As an alternative, the investor is considering making smaller investments of $100 into each of 10 different stocks. The value of each of these investments after a year is known to be uniformly distributed with mean 105.

d) If each of these investments has a value between 100 and 110 with probability 0.25, what are the minimum and maximum values of each investment?

Assuming the 10 stock values are independent of one another, the total value of the portfolio, W, will be very close to normally distributed with a mean of 1050 and standard deviation of 36.51. Use this approximation to answer the following questions.

e) Which of the investments (the single investment X, or the diversified investment W) is more likely to be near its mean? How do you know?

f) What is the probability that the investor loses value after a year with the diversified investment?

g) What is the probability that the portfolio has a value between 1000 and 1100 after a year?

Answers:

a) The mean of a uniform distribution is its midpoint, $\frac{900+1200}{2} = 1050$.

b) $P(X < 1000) = (1000 - 900)/(1200 - 900) = 1/3$ (or 0.333).

c)

$$\begin{aligned} P(X < 930 \cup X > 1070) &= P(X < 930) + P(X > 1070) \qquad \text{(disjoint)} \\ &= \frac{930 - 900}{1200 - 900} + \frac{1200 - 1070}{1200 - 900} \\ &= \frac{160}{300} = \frac{8}{15} \\ &= 0.533 \end{aligned}$$

d) One approach to answer this question is to reason that, since we know that a span of 10 (110-100) covers 1/4 of the range of values, it must be the case than the entire range of values is 40, and since 105 is in the middle, the range must be **[85, 125]**.

An alternative approach is to do the same reasoning as the above, but mathematically, as follows. Let Y denote the value of one of the investments, so $Y \sim \mathcal{U}[a, b]$. Since we known the mean is 105, and that it is therefore in the middle of a and b, we also know that $a = 105 - c, b = 105 + c$ for some constant c, and so we have the distribution: $\mathcal{U}[105 - c, 105 + c]$.

We also know that $P(100 < Y < 110) = 0.25$, and from this:

$$\begin{aligned} \frac{110 - 100}{(105 + c) - (105 - c)} &= 0.25 \\ \frac{10}{2c} &= 0.25 \\ 10 &= 0.5c \\ c &= 20 \end{aligned}$$

and therefore the minimum and maximum are $105 - c = 85$ and $105 + c = 125$.

e) To answer this, compare the standard deviations for X and W (both given): $s_x = 86.60, s_w = 36.51$. Since $s_x > s_w$ that means that, on average, values of X are farther from the mean of X, or in other words, that **values of W are more likely to be closer to the mean.**

f) The initial investment is 1000, and $W \sim \mathcal{N}(1050, 36.51)$, so:

$$\begin{aligned} P(W < 1000) &= P\left(Z < \frac{1000 - 1050}{36.51}\right) \\ &= P(Z < -1.37) \\ &= .0853 \end{aligned}$$

g) We know that $P(1000 < W < 1100) = P(W < 1100) - P(W < 1000)$. The last of these probabilities we calculated in the previous section, so:

$$\begin{aligned} P(1000 < W < 1100) &= P\left(Z < \frac{1100 - 1050}{36.51}\right) - 0.0853 \\ &= P(Z < 1.37) - 0.0853 \\ &= 0.9147 - 0.0853 \\ &= 0.8294 \end{aligned}$$

You could also take a shortcut by realizing that, since W is symmetric around 1050 and the values are the same distance on other size of 1050, the answer will be simply $1 - 2 * .0853 = 0.8294$.

3. **[30]** A deck of standard playing cards has 52 cards. Each card has one of four suits ($\spadesuit, \heartsuit, \diamondsuit$, or $\clubsuit$) and one of thirteen ranks (A, 2, 3, …, 9, 10, J, Q, or K). Thus there are 13 cards of each suit, and 4 cards of each rank.

a) Suppose you draw one card from the deck at random. Let A denote the event that the card is a $\spadesuit$, and let B denote the event that the card is a 7. Find the following:

(i) $P(A)$ *(ii)* $P(B)$ *(iii)* $P(A^{\complement})$ *(iv)* $P(A \cap B)$ *(v)* $P(B \cup A)$

b) Suppose you draw four cards at random from a deck, and look at them. The cards are: $5\heartsuit, Q\heartsuit, Q\spadesuit, 2\heartsuit$. Let C denote the event of drawing these 4 cards.

Let D denote the event that the next card you draw (i.e. the fifth card) is a Q. Write down a probability expression in terms of C and D that expresses the statement "the probability that, after drawing the four cards shown above, the next card is a Q", and find this probability.

c) Let E be the event that a card drawn at random is black ($\spadesuit$ or $\clubsuit$), let F be the event that the card is a face card (J, Q, or K), and let G denote the event that the card is the $Q\spadesuit$.

(i) Find $P(G|E \cap F)$

(*ii*) Find $P(E \cap F|G)$

d) For each of the following pairs of events, indicate whether the two events are disjoint, and whether the two events are independent.

(*i*) A: the first card drawn is a ♠.
B: the first card drawn is a black card (♠ or ♣).

(*ii*) A: the first card drawn has rank 3, 4, or 5.
B: the first card drawn is 6♢.

(*iii*) A: the first card drawn is a ♠.
B: the first card drawn is a 7.

(*iv*) A: the *first* card drawn is a ♢.
B: the *second* card drawn is a ♢.

Now suppose that the deck is missing several cards. You know, however, that the probability of drawing a J is 1/8, and that the probability of drawing a ♡ is 1/2. You also know that if the card is a J there is a probability of 1/3 that the card is a ♡.

e) If a card drawn at random is a ♡, what is the probability that the card is a J?

f) Show that, with this deck, drawing a ♡ and drawing a J are *not* independent.

Answers: (Note that leaving answers as fractions for this question is perfectly acceptable: the decimal value calculations shown are not required).

a) (*i*) There are 13 ♠s out of 52 cards in total, so: $P(A) = 13/52 = 1/4 = 0.25$

(*ii*) There are 4 "7"s out of 52 cards in total, so: $P(B) = 4/52 = 1/13 = 0.0769$

(*iii*) $P(A^{\complement}) = 1 - P(A) = 1 - 1/4 = 3/4 = 0.75$

Alternatively: there are 13 ♡s, 13 ♢s, and 13 ♣s out of 52 cards, and so the probability equals $39/52 = 3/4$.

(*iv*) $A \cap B$ is the event that the card is the 7♠, i.e. this is the event that you draw one particular card, thus the value is $1/52 = 0.0192$.

(*v*) Cards that satisfy $A \cup B$ include all the ♠s and all the "7"s, and so 16 cards in total (not $13 + 4 = 17$ because that counts the 7♠ card twice), and so $P(A \cup B) = 16/52 = 4/13 = 0.3077$.

You could also use the formula: $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52}$ to get the same answer.

b) The associated probability expression is: $P(D|C)$. Its value is: $P(D|C) = 2/48 = 1/24 = 0.04167$: given that we know the cards satisfying C are no longer in the deck, there are two Q cards left out of the 48 cards remaining in the deck.

You *could* calculate this from the formula $P(D|C) = P(D \cap C)/P(C)$, but it is more work that mostly complicates the problem unnecessarily: if you interpret C as the precise sequence of cards given, you would get $P(C) = (\frac{1}{52})(\frac{1}{51})(\frac{1}{50})(\frac{1}{49})$, and you would get $P(C \cap D) = (\frac{1}{52})(\frac{1}{51})(\frac{1}{50})(\frac{1}{49})(\frac{2}{48})$. All of these fractions except the $\frac{2}{48}$ cancel out, leaving you with the same $\frac{2}{48}$ answer.

If you interpret C as the set of 4 cards (in any order), you'd still get the same answer, but the first four probability values (which get cancelled out) would be $(\frac{4}{52})(\frac{3}{51})(\frac{2}{50})(\frac{1}{49})$.

c) (*i*) Since we know $E \cap F$ we already know that the card is both a black suit and a face card, of which there are only 6 possible cards that satisfy both conditions: $J\spadesuit$, $Q\spadesuit$, $K\spadesuit$, $J\clubsuit$, $Q\clubsuit$, $K\clubsuit$. (Every other card is either not a black card, or is not a face card).

Thus the conditional probability equals $1/6 = 0.1667$.

(*ii*) Since we already know exactly the card that was drawn ($Q\spadesuit$), there is no uncertaintly left: we already know the card is both a black suit and a face card, so the probability simply equals 1.

d) (*i*) Not disjoint (any $\spadesuit$ satisfies both).

Not independent: $P(B)$, the probability that a card is black, equals 1/2, but $P(B|A)$, the probability that a $\spadesuit$ is black card equals 1. Alternatively: $P(A)$, the probability of a $\spadesuit$ in general, equals 1/4, but $P(A|B)$, the probability of a $\spadesuit$ knowing that the card is black equals 1/2.

(*ii*) Disjoint: it is impossible for a card to be both one of 3, 4, or 5 *and* for the card to be a 6 at the same time.

Not independent. If you tell me that the card is a 3, 4, or 5, I instantly know that the card is not $6\diamondsuit$. (Alternatively, if you tell me the card is $6\diamondsuit$, I know it cannot possibly be a 3, 4, or 5).

(*iii*) Not disjoint: $7\spadesuit$ satisfies both A and B.

They *are* independent, which you can verify in several ways. For example, $P(A|B) = 1/4$ (there are 4 "7"s, and only one of them is a spade). $P(A) = 13/52 = 1/4$ (there are 13 spades out of 52 cards in total). You could similarly show $P(B|A) = P(B)$.

Alternatively: $P(A \cap B) = 1/52$, and since this equals $P(A)P(B) = (1/4)(1/13) = 1/52$, A and B are independent.

(*iv*) Not disjoint: it is possible for both cards to be diamonds.

Not independent: the probability of the second card being a diamond without knowing the first card is $13/52 = 1/4 = 0.25$, but if I know the first card is a diamond, the probability of getting a second one is $12/51 = 0.235$.

e) This question is basically asking you to apply Bayes' Rule. We are given:

$$P(\heartsuit) = 1/2$$
$$P(J) = 1/8$$
$$P(\heartsuit|J) = 1/3$$

and are asked to find:

$$\begin{aligned} P(J|\heartsuit) &= \frac{P(\heartsuit|J)P(J)}{P(\heartsuit)} \\ &= \frac{(1/3)(1/8)}{1/2} \\ &= 1/12 = 0.08333 \end{aligned}$$

If you really want to avoid Bayes' Rule, you can also answer the question like this:

$$P(J|\heartsuit) = \frac{P(J \cap \heartsuit)}{P(\heartsuit)}$$

The numerator probability will be 1 divided by the number of cards in the deck. To figure that out, we can determine that there must be three Js in the deck (since the probability of a J being a $\heartsuit$ equals 1/3), and since the probability of getting a J equals 1/8, there must be $3 \times 8 = 24$ cards in the deck in total. Therefore we know $P(J \cap \heartsuit) = 1/24$, and so $P(J|\heartsuit) = (1/24)/(1/2) = 1/12$.

Bayes' Rule, however, is definitely easier.

f) *Any* of the following inequalities can be used to show that they are **not independent**:

$$\begin{aligned} P(\heartsuit|J) &= \frac{1}{3} \neq \frac{1}{2} = P(\heartsuit) \\ P(J|\heartsuit) &= \frac{1}{12} \neq \frac{1}{8} = P(J) \\ P(J \cap \heartsuit) &= \frac{1}{24} \neq \frac{1}{16} = P(J)P(\heartsuit) \end{aligned}$$