Examiners' commentary 2022

ST102/ST109 Elementary Statistical Theory (Part I)

General remarks

Learning outcomes

By the end of this module you should:

- be able to summarise the ideas of randomness and variability, and the way in which these link to probability theory to allow the systematic and logical collection of statistical techniques of great practical importance in many applied areas
- be competent users of standard statistical operators and be familiar with a variety of well-known distributions and their moments
- understand the fundamentals of statistical inference and be able to apply these principles to choose an appropriate model and test in a number of different settings
- recognise that statistical techniques are based on assumptions and in the analysis of real problems the plausibility of these assumptions must be thoroughly checked.

Examination structure

After an initial ten minutes of reading time, you have two hours to complete this paper, which has four compulsory questions. The questions are given equal weight, and carry 25 marks each.

What are the Examiners looking for?

The Examiners are looking for you to demonstrate command of the course material. Although the final solution is 'desirable', the Examiners are more interested in how you approach each solution, as such, most marks are awarded for the 'method steps'. They want to be sure that you:

- have covered the syllabus
- know the various definitions and concepts covered throughout the year and can apply them as appropriate to examination questions
- understand and answer the questions set.

You are not expected to write long essays where explanations or descriptions are required, and note-form answers are acceptable. However, clear and accurate language, both mathematical and written, is expected and marked.

Key steps to improvement

The most important thing you can do is answer the question set! This may sound very simple, but these are some of the things that candidates did not do. Remember:

- **Always show your working.** The bulk of the marks are awarded for your approach, rather than the final answer.
- **Write legibly!**
- **Keep solutions to the same question in one place.** Avoid scattering your solutions randomly throughout the answer booklet – the Examiners will not appreciate having to spend a lot of time searching for different elements of your solutions.
- **Where appropriate, underline your final answer.**
- **Do not waste time calculating things which are not required by the Examiners!**

Using the commentary

We hope that you find the commentary useful. For each question and subquestion, it gives:

- the answers, or keys to the answers, which the Examiners were looking for
- common mistakes, as identified by the Examiners.

Dr James Abdey, ST102/ST109 Lecturer, February 2022

Examiners' commentary 2022

ST102/ST109 Elementary Statistical Theory (Part I)

Specific comments on questions

Question 1

(a) **A box contains 12 light bulbs, of which two are defective. If a person selects 5 bulbs at random, without replacement, what is the probability that both defective bulbs will be selected?**

(7 marks)

(b) **A and B are independent events such that:**

$$P((A \cup B)^c) = \pi_1 \quad \text{and} \quad P(A) = \pi_2.$$

Determine $P(B)$ as a function of π_1 and π_2.

(7 marks)

(c) **A county is made up of three (mutually exclusive) communities A, B and C, with proportions of people living in them given by the following table:**

Community	A	B	C
Proportion	0.20	0.50	0.30

Given a person belongs to a certain community, the probability of that person being vaccinated is given by the following table:

Community given	A	B	C
Probability of being vaccinated	0.80	0.70	0.60

i. **We choose a person from the county at random. What is the probability that the person *is not* vaccinated?**

(5 marks)

ii. **We choose a person from the county at random. Find the probability that the person is in community A, given the person *is* vaccinated.**

(4 marks)

iii. **In words, *briefly* explain how the 'probability of being vaccinated' for each community would be known in practice.**

(2 marks)

Feedback on this question

Part (a) was a variant of a previously-seen exercise, requiring the application of classical probability and proved quite straightforward for candidates. Part (b) was more challenging, with the solution often not expressed as a function of π_1 and π_2, as requested. Part (c) involved applying the total probability formula and Bayes' theorem – generally this was done well, although some arithmetic mistakes occurred.

Full solutions are as follows.

(a) The sample space consists of all (unordered) subsets of 5 out of the 12 light bulbs in the box. There are $\binom{12}{5}$ such subsets. The number of subsets which contain the two defective bulbs is the number of subsets of size 3 out of the other 10 bulbs, $\binom{10}{3}$, so the probability we want is:

$$\frac{\binom{10}{3}}{\binom{12}{5}} = \frac{5 \times 4}{12 \times 11} = 0.1515.$$

(b) We are given that $P((A \cup B)^c) = \pi_1$, $P(A) = \pi_2$, and that A and B are independent. Hence:

$$P(A \cup B) = 1 - \pi_1 \quad \text{and} \quad P(A \cap B) = P(A)\,P(B) = \pi_2 P(B).$$

Therefore:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \pi_2 + P(B) - \pi_2 P(B) = 1 - \pi_1.$$

Solving for $P(B)$, we have:

$$P(B) = \frac{1 - \pi_1 - \pi_2}{1 - \pi_2}.$$

(c) i. Denote acceptance of vaccination by V, and not accepting by V^c. By the law of total probability, we have:

$$\begin{aligned} P(V^c) &= P(V^c \mid A)\,P(A) + P(V^c \mid B)\,P(B) + P(V^c \mid C)\,P(C) \\ &= (1 - 0.8) \times 0.2 + (1 - 0.7) \times 0.5 + (1 - 0.6) \times 0.3 \\ &= 0.31. \end{aligned}$$

ii. By Bayes' theorem, we have:

$$P(A \mid V) = \frac{P(V \mid A)\,P(A)}{P(V)} = \frac{0.8 \times 0.2}{1 - 0.31} = \frac{16}{69} = 0.2319.$$

iii. Any reasonable answer accepted, such as relative frequency estimate, or from health records.

Question 2

(a) *Use the total probability formula* to answer each of the following. A, B and C are events, with complementary events denoted by '^{c}' as per usual course notation.

i. If $P(A \mid B) = P(A \mid B^c)$, show that A and B are independent.

(7 marks)

ii. If $P(A \mid C) > P(B \mid C)$ and $P(A \mid C^c) > P(B \mid C^c)$, show that $P(A) > P(B)$.

(7 marks)

(b) The random variable X has a Poisson distribution with parameter $\lambda = 3$. Calculate $P(X \geq 2 \mid X > 0)$.

(5 marks)

(c) Surviving contestants on *Squid Game* reach the 'Glass Stepping Stones' round, where players must cross a two-panel wide bridge. One panel is made of tempered glass (supporting a player's weight) and the other panel is made of regular glass (which cannot support their weight). The bridge consists of 18 sets of two panels (so to cross the bridge a player must safely navigate the correct 'path' of 18 tempered panels). The second player follows the first player.

i. What is the probability that the *first* player crosses the bridge successfully? State any assumptions you make.

(3 marks)

ii. Would the second player be at an advantage (relative to the first player)? Briefly explain why or why not.

(3 marks)

Feedback on this question

Part (a) explicitly required the total probability formula to be used, which was sometimes not directly referenced in some solutions. Part (b) was straightforward, although sometimes the simplification of the numerator in the conditional probability expression was missed. Part (c) had been discussed in a weekly review session – regardless, I would still *not* recommend playing the game in real life!

Full solutions are as follows.

(a) i. Let $P(A \mid B) = P(A \mid B^c) = \pi$. Using the total probability formula, we have:

$$P(A) = P(A \mid B)\, P(B) + P(A \mid B^c)\, P(B^c) = \pi(P(B) + P(B^c)) = \pi.$$

Since $P(A) = P(A \mid B) = P(A \mid B^c)$, we have that A and B are independent.

ii. Using the total probability formula, we have:

$$\begin{aligned} P(A) &= P(A \mid C)\, P(C) + P(A \mid C^c)\, P(C^c) \\ &> P(B \mid C)\, P(C) + P(B \mid C^c)\, P(C^c) \\ &= P(B). \end{aligned}$$

Hence $P(A) > P(B)$.

(b) We have that $X \sim \text{Pois}(3)$. The required probability is:

$$\begin{aligned} P(X \geq 2 \mid X > 0) = \frac{P(\{X \geq 2\} \cap \{X > 0\})}{P(X > 0)} &= \frac{P(X \geq 2)}{P(X \geq 1)} \\ &= \frac{1 - P(X \leq 1)}{1 - P(X = 0)} \\ &= \frac{1 - \mathrm{e}^{-3} - 3\mathrm{e}^{-3}}{1 - \mathrm{e}^{-3}} \\ &= 0.8428. \end{aligned}$$

(c) i. Assuming independence across pairs of glass panels:

$$P(\text{survive}) = \left(\frac{1}{2}\right)^{18} = \frac{1}{262{,}144} = 3.8 \times 10^{-6}.$$

ii. At the very least, the second player would know the correct pane of the first pair from the first player (and possibly more, depending how far the first player got), so:

$$P(\text{second player survives}) \geq \left(\frac{1}{2}\right)^{17} = \frac{1}{131{,}072} > \frac{1}{262{,}144}.$$

Question 3

(a) A random variable, X, has the following probability density function:

$$f(x) = \begin{cases} x/5 & \text{for } 0 \leq x < 2 \\ (20 - 4x)/30 & \text{for } 2 \leq x \leq 5 \\ 0 & \text{otherwise.} \end{cases}$$

i. Sketch the graph of $f(x)$. (The sketch can be drawn on ordinary paper – no graph paper needed.)

(3 marks)

ii. Derive the cumulative distribution function of X.

(7 marks)

iii. Find the mean and the standard deviation of X.

(7 marks)

(b) i. A coffee machine can be calibrated to produce an average of μ millilitres (ml) per cup. Suppose the quantity produced is normally distributed with a standard deviation of 9 ml per cup. Determine the value of μ such that 250 ml cups will overflow only 1% of the time.

(4 marks)

ii. Suppose now that the standard deviation, σ, can be fixed at specified levels. What is the largest value of σ that will allow the amount of coffee dispensed to fall *within* 30 ml of the mean with a probability of at least 95%?

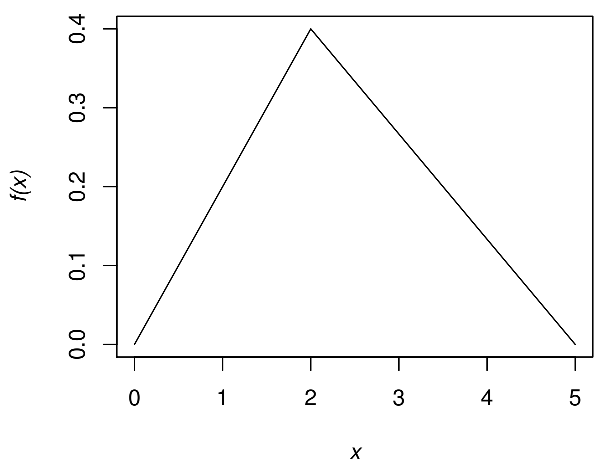
(4 marks)

Feedback on this question

In part (a), some candidates did not spot the linear nature of $f(x)$, and hence the triangular distribution which results. Specification of $F(x)$ sometimes missed the cases $x < 0$ and $x > 5$. Part iii. was fine, although there were some integration errors. Part (b) proved straightforward for most candidates.

Full solutions are as follows.

(a) i. The pdf of X has the following form:



ii. We determine the cdf by integrating the pdf over the appropriate range, hence:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ x^2/10 & \text{for } 0 \leq x < 2 \\ (10x - x^2 - 10)/15 & \text{for } 2 \leq x \leq 5 \\ 1 & \text{for } x > 5. \end{cases}$$

This results from the following calculations. Firstly, for $x < 0$, we have:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{x} 0\,\mathrm{d}t = 0.$$

For $0 \leq x < 2$, we have:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{0} 0\,\mathrm{d}t + \int_{0}^{x} \frac{t}{5}\,\mathrm{d}t = \left[\frac{t^2}{10}\right]_0^x = \frac{x^2}{10}.$$

For $2 \leq x \leq 5$, we have:

$$\begin{aligned} F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t &= \int_{-\infty}^{0} 0\,\mathrm{d}t + \int_{0}^{2} \frac{t}{5}\,\mathrm{d}t + \int_{2}^{x} \frac{20-4t}{30}\,\mathrm{d}t \\ &= 0 + \frac{4}{10} + \left[\frac{2t}{3} - \frac{t^2}{15}\right]_2^x \\ &= \frac{4}{10} + \left(\frac{2x}{3} - \frac{x^2}{15}\right) - \left(\frac{4}{3} - \frac{4}{15}\right) \\ &= \frac{2x}{3} - \frac{x^2}{15} - \frac{2}{3} \\ &= \frac{10x - x^2 - 10}{15}. \end{aligned}$$

iii. To find the mean we proceed as follows:

$$\begin{aligned} \mu = \mathrm{E}(X) = \int_{-\infty}^{\infty} x\,f(x)\,\mathrm{d}x &= \int_{0}^{2} \frac{x^2}{5}\,\mathrm{d}x + \int_{2}^{5} \frac{20x - 4x^2}{30}\,\mathrm{d}x \\ &= \left[\frac{x^3}{15}\right]_0^2 + \left[\frac{x^2}{3} - \frac{2x^3}{45}\right]_2^5 \\ &= \frac{8}{15} + \left(\frac{25}{3} - \frac{250}{45}\right) - \left(\frac{4}{3} - \frac{16}{45}\right) \\ &= 7/3 \text{ or } 2.3333. \end{aligned}$$

Similarly:

$$\begin{aligned} \mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x &= \int_{0}^{2} \frac{x^3}{5}\,\mathrm{d}x + \int_{2}^{5} \frac{20x^2 - 4x^3}{30}\,\mathrm{d}x \\ &= \left[\frac{x^4}{20}\right]_0^2 + \left[\frac{2x^3}{9} - \frac{x^4}{30}\right]_2^5 \\ &= \frac{16}{20} + \left(\frac{250}{9} - \frac{625}{30}\right) - \left(\frac{16}{9} - \frac{16}{30}\right) \\ &= \frac{13}{2} \text{ or } 6.5. \end{aligned}$$

Hence the variance is:

$$\sigma^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{13}{2} - \left(\frac{7}{3}\right)^2 = \frac{117}{18} - \frac{98}{18} = \frac{19}{18} \approx 1.0555.$$

Therefore, the standard deviation is $\sigma = \sqrt{1.0555} = 1.0274$.

(b) i. Let $X =$ volume filled, so that $X \sim N(\mu, 81)$. We require that:

$$P(X > 250) = 0.01.$$

Standardising, we have:

$$P(Z > 2.33) = 0.01$$

hence:

$$2.33 = \frac{250 - \mu}{9} \quad \Rightarrow \quad \mu = 229.03.$$

ii. We have that $X \sim N(\mu, \sigma^2)$. It follows that:

$$0.95 \leq P(|X - \mu| < 30) = P\left(|Z| < \frac{30}{\sigma}\right)$$

so that $30/\sigma = 1.96$ or $\sigma = 30/1.96 = 15.31$.

Question 4

(a) Suppose that X is a discrete random variable for which the moment generating function is:

$$M_X(t) = \frac{1}{4}(e^{3t} + e^{6t} + e^{9t}) + \frac{1}{8}(e^{2t} + e^{4t})$$

for $-\infty < t < \infty$. Write down the probability distribution of X.

(6 marks)

(b) Consider two random variables, X and Y. They both take the values -1, 0 and 1. The joint probabilities for each pair of values, (x, y), are given in the following table.

	$X = -1$	$X = 0$	$X = 1$
$Y = -1$	0.09	0.16	0.15
$Y = 0$	0.09	0.08	0.03
$Y = 1$	0.12	0.16	0.12

i. Determine the marginal distributions and calculate the expected values of X and Y, respectively.

(4 marks)

ii. Calculate the covariance of the random variables X and Y.

(4 marks)

iii. Calculate $E(X \mid Y = 0)$ and $E(X \mid X + Y = 1)$.

(6 marks)

iv. Define $U = |X|$ and $V = Y$. Calculate $E(U)$ and the covariance of U and V. Are U and V correlated?

(5 marks)

Feedback on this question

Part (a) was a variant of a previously-seen exercise. Some candidates forgot the '0 otherwise' case. Part (b) was standard, with most errors being arithmetic.

Full solutions are as follows.

(a) If X can take only a finite number of values $x_1, x_2, \ldots, x_k$ with probabilities $p_1, p_2, \ldots, p_k$, respectively, then the mgf of X will be:

$$M_X(t) = p_1 e^{tx_1} + p_2 e^{tx_2} + \cdots + p_k e^{tx_k}.$$

By matching this expression for $M_X(t)$ with:

$$M_X(t) = \frac{1}{4}(e^{3t} + e^{6t} + e^{9t}) + \frac{1}{8}(e^{2t} + e^{4t})$$

it can be seen that X can take only the five values 2, 3, 4, 6 and 9 and hence:

$$p(x) = \begin{cases} 1/8 & \text{for } x = 2 \text{ and } 4 \\ 1/4 & \text{for } x = 3, 6 \text{ and } 9 \\ 0 & \text{otherwise.} \end{cases}$$

(b) i. The marginal distribution of X is:

X	-1	0	1
$p_X(x)$	0.30	0.40	0.30

The marginal distribution of Y is:

Y	-1	0	1
$p_Y(y)$	0.40	0.20	0.40

Hence:

$$\mathrm{E}(X) = \sum_x x\, p_X(x) = (-1 \times 0.30) + (0 \times 0.40) + (1 \times 0.30) = 0$$

and:

$$\mathrm{E}(Y) = \sum_y y\, p_Y(y) = (-1 \times 0.40) + (0 \times 0.20) + (1 \times 0.40) = 0.$$

ii. We have:

$$\begin{aligned}\mathrm{E}(XY) &= (-1 \times -1 \times 0.09) + (-1 \times 1 \times 0.12) + (1 \times -1 \times 0.15) + (1 \times 1 \times 0.12)\\ &= 0.09 - 0.12 - 0.15 + 0.12\\ &= -0.06.\end{aligned}$$

Therefore:

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) = -0.06 - 0 \times 0 = -0.06.$$

iii. We have $P(Y = 0) = 0.09 + 0.08 + 0.03 = 0.20$, hence:

$$P(X = -1 \mid Y = 0) = \frac{0.09}{0.20} = 0.45$$

$$P(X = 0 \mid Y = 0) = \frac{0.08}{0.20} = 0.40$$

$$P(X = 1 \mid Y = 0) = \frac{0.03}{0.20} = 0.15$$

and therefore:

$$\mathrm{E}(X \mid Y = 0) = -1 \times 0.45 + 0 \times 0.4 + 1 \times 0.15 = -0.30.$$

We also have $P(X + Y = 1) = 0.16 + 0.03 = 0.19$, hence:

$$P(X = 0 \mid X + Y = 1) = \frac{0.16}{0.19} = \frac{16}{19}$$

$$P(X = 1 \mid X + Y = 1) = \frac{0.03}{0.19} = \frac{3}{19}$$

and therefore:

$$\mathrm{E}(X \mid X + Y = 1) = 0 \times \frac{16}{19} + 1 \times \frac{3}{19} = \frac{3}{19} = 0.1579.$$

iv. Here is the table of joint probabilities:

	$U = 0$	$U = 1$
$V = -1$	0.16	0.24
$V = 0$	0.08	0.12
$V = 1$	0.16	0.24

We then have that $P(U = 0) = 0.16 + 0.08 + 0.16 = 0.40$ and also that $P(U = 1) = 1 - P(U = 0) = 0.60$. Also, we have that $P(V = -1) = 0.40$, $P(V = 0) = 0.20$ and $P(V = 1) = 0.40$. So:

$$\mathrm{E}(U) = 0 \times 0.40 + 1 \times 0.60 = 0.60$$

$$\mathrm{E}(V) = -1 \times 0.40 + 0 \times 0.20 + 1 \times 0.40 = 0$$

and:

$$\mathrm{E}(UV) = -1 \times 1 \times 0.24 + 1 \times 1 \times 0.24 = 0.$$

Hence $\mathrm{Cov}(U, V) = \mathrm{E}(UV) - \mathrm{E}(U)\mathrm{E}(V) = 0 - 0.60 \times 0 = 0$. Since the covariance is zero, so is the correlation coefficient, therefore U and V are uncorrelated.