Teaching basic statistics using exam problems

YAP Von Bing
Department of Statistics and Applied Probability, NUS

IF PRINTING NEEDED: Double-sided, four slides each side, landscape mode. See if size 200% works. All fit in 4 pieces of paper.

June 6, 2014

H2 probability syllabus

1. Addition and multiplication of probabilities
2. Mutually exclusive events and independent events
3. Use of tables of outcomes, Venn diagrams, and tree diagrams to calculate probabilities
4. Calculation of conditional probabilities in simple cases
5. Use of
   (a) $P(A') = 1 - P(A)$
   (b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
   (c) $P(A|B) = P(A \cap B)/P(B)$

Rewriting 5(c) gives valuable insight on several items:

$$P(A \cap B) = P(B)P(A|B)$$

2013 P2 Q8

For events A and B it is given that $P(A) = 0.7$, $P(B|A') = 0.8$, $P(A|B') = 0.88$. Find

(i) $P(B \cap A')$, [1]

(ii) $P(A' \cap B')$, [2]

(iii) $P(A \cap B)$. [3]

Elementary questions

In general,

1. Is $\mathrm{P}(B \cap A') = \mathrm{P}(A' \cap B)$? Why?
2. What can you say about $\mathrm{P}(B \cap A') + \mathrm{P}(A' \cap B')$? How about $\mathrm{P}(B \cap A') + \mathrm{P}(A \cap B)$? Why?

$\mathrm{P}(A) = 0.7$, $\mathrm{P}(B|A') = 0.8$, $\mathrm{P}(A|B') = 0.88$.

1. Are A and B' independent? Are they mutually exclusive?
2. Are A and B independent? Are they mutually exclusive?

A table representation

	B	B'	*Row sum*
A	$\mathrm{P}(A \cap B)$	$\mathrm{P}(A \cap B')$	
A'	$\mathrm{P}(A' \cap B)$	$\mathrm{P}(A' \cap B')$	
Column sum			1

Table : Joint probabilities in symbols, with blanks in margins.

	B	B'	*Row sum*
A			0.7
A'			
Column sum			1.0

Table : Joint probabilities in numbers.

A table representation

	B	B'	*Row sum*
A	0.26	0.44	0.70
A'	0.24	0.06	0.30
Column sum	0.50	0.50	1.00

Table : Joint probabilities in numbers.

1. $\mathrm{P}(A') = 1 - \mathrm{P}(A) = 1 - 0.7 = 0.3.$
   $\mathrm{P}(B \cap A') = \mathrm{P}(B|A') \times \mathrm{P}(A') = 0.8 \times 0.3 = 0.24.$
2. $\mathrm{P}(A' \cap B') = \mathrm{P}(A') - \mathrm{P}(A' \cap B) = 0.3 - 0.24 = 0.06.$
3. $\mathrm{P}(A'|B') = 1 - \mathrm{P}(A|B') = 1 - 0.88 = 0.12.$
   $\mathrm{P}(B') = \mathrm{P}(A' \cap B')/\mathrm{P}(A'|B') = 0.06/0.12 = 0.5$
   $\mathrm{P}(B) = 1 - \mathrm{P}(B') = 0.5.$
   Hence $\mathrm{P}(A \cap B) = \mathrm{P}(B) - \mathrm{P}(A' \cap B) = 0.5 - 0.24 = 0.26.$

2012 P2 Q5

The probability that a hospital patient has a particular disease is 0.001. A test for the disease has probability p of giving a positive result when the patient has the disease, and equal probability p of giving a negative result when the patient does not have the disease. A patient is given the test.

(i) Given that $p = 0.995$, find the probability that

(a) the result of the test is positive, [2]

(b) the patient has the disease given that the result of the test is positive. [2]

(ii) It is given that there is a probability of 0.75 that the patient has the disease given that the result of the test is positive. Find the value of p, giving your answer correct to 6 decimal places. [3]

Activity

Let D and $+$ be the events $\{$has disease$\}$ and $\{$tests positive$\}$.

- What elementary questions could be asked?
- What approaches can be used to solve the problem?

One solution

	$+$	$-$	*Row sum*
D	0.001p	0.001(1-p)	0.001
D'	0.999(1-p)	0.999p	0.999
Column sum	$0.999 - 0.998p$	$0.001 + 0.998p$	1.000

Table : Joint probabilities.

(i) (a) $\text{P}(+) = 0.999 - 0.998p$,
(b) $\text{P}(D|+) = 0.001p/(0.999 - 0.998p)$.

(ii) Set expression in (b) to 0.75, solve for p.

For $p = 0.005$, are D and $+$ independent?

Exploring further

- In general, which is preferable: (I) $\mathrm{P}(D|+) > \mathrm{P}(D)$, (II) $\mathrm{P}(D|+) < \mathrm{P}(D)$? Why? [This is not a mathematical question.]

- In (i), which of (I) or (II) is true?

- Is there a p such that D and $+$ are independent? If yes, what is the implication?

- If $\mathrm{P}(D|+) > \mathrm{P}(D)$, what can be said about $\mathrm{P}(D|-)$ and $\mathrm{P}(D)$? Why?

Even further

- Verify: For any A and B,

$$P(A|B)P(B) + P(A|B')P(B') = P(A)$$

- Let $0 < a < 1$, and x, y, z be such that

$$ax + (1 - a)y = z$$

  Suppose that $x > z$. Show that $y < z$.
- Deduce that for any A and B with $0 < P(B) < 1$, if $P(A|B) > P(A)$, then $P(A|B') < P(A)$. This answers the last question in previous slide.

2011 P2 Q9

Camera lenses are made by two companies, A and B. 60% of all lenses are made by A and the remaining 40% by B. 5% of the lenses made by A are faulty. 7% of the lenses made by B are faulty.

(i) One lens is selected at random. Find the probability that
  (a) it is faulty, [2]
  (b) it was made by A, given that it is faulty. [2]

(ii) Two lenses are selected at random. Find the probability that
  (a) exactly one of them is faulty, [2]
  (b) both were made by A, given that exactly one is faulty. [2]

Table

	Faulty	*Not faulty*	*Row sum*
A	3.0%	57.0%	**60%**
B	2.8%	37.2%	**40%**
Column sum	5.8%	94.2%	**100%**

Table : Joint percentages, given information in bold.

$$\begin{aligned} 3\% &= 60\% \times 5/100 \\ 57\% &= 60\% - 3\% \\ 2.8\% &= 40\% \times 7/100 \\ 37.2\% &= 40\% - 2.8\% \end{aligned}$$

Backward questions

(i) (a) 0.058. (b) $3/5.8 \approx 0.517$.

(ii) (a)

$$\binom{2}{1} \times 0.058 \times 0.942 \approx 0.109.$$

What does the binomial coefficient mean in this problem?

(b)

$$\frac{2 \times 0.03 \times 0.57}{0.109} \approx 0.313.$$

Why use 0.03 and 0.57?

2010 P2 Q7

For events A and B it is given that $\mathrm{P}(A) = 0.7$, $\mathrm{P}(B) = 0.6$ and $\mathrm{P}(A|B') = 0.8$. Find

(i) $\mathrm{P}(A \cap B')$, [2]

(ii) $\mathrm{P}(A \cup B)$, [2]

(iii) $\mathrm{P}(B'|A)$. [2]

For a third event C, it is given that $\mathrm{P}(C) = 0.5$ and that A and C are independent.

(iv) Find $\mathrm{P}(A' \cap C)$. [2]

(v) Hence state an inequality satisfied by $\mathrm{P}(A' \cap B \cap C)$. [1]

Solution, more backward questions

	B	B'	*Row sum*
A	0.38	0.32	**0.70**
A'	0.22	0.08	0.30
Column sum	**0.60**	0.40	**1.00**

Table : Joint probabilities, given information in bold.

(i) $\mathrm{P}(A \cap B') = 0.8 \times 0.4 = 0.32$.
(ii) $\mathrm{P}(A \cap B) = 0.7 - 0.32 = 0.38$.
$\mathrm{P}(A' \cap B) = 0.6 - 0.38 = 0.22$.
$\mathrm{P}(A \cup B) = 0.32 + 0.38 + 0.22 = 0.92$, or
$\mathrm{P}(A \cup B) = 0.70 + 0.22 = 0.92$.
What formulae are used here?

General questions

- Is this possible? $\mathrm{P}(A) = 0.3$ and $\mathrm{P}(A \cap B) = 0.4$.
- If $\mathrm{P}(A) = 0.4$ and $\mathrm{P}(B) = 0.8$, can they be mutually exclusive? Can they be independent?
- Repeat the previous, if $\mathrm{P}(A) + \mathrm{P}(B) > 1$.

More on two events

Suppose A and B are independent events.

- Are A and B' independent? Why?

- Are A' and B' independent? Why?

An addition formula

For any events A and B,

$$P(A \cap B) + P(A \cap B') = P(A)$$

- Is this argument correct? Or is there something wrong?

$$\begin{aligned} P(A \cap B) + P(A \cap B') &= P(A)P(B) + P(A)P(B') \\ &= P(A)[P(B) + P(B')] = P(A) \end{aligned}$$

- Can it be deduced from

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

2012 P2 Q9

In an opinion poll before an election, a sample of 30 voters is obtained. The number of voters in the sample who support the Alliance Party is denoted by A. State, in context, what must be assumed for A to be well modelled by a binomial distribution. Assume now that A has the distribution $\mathrm{B}(30, p)$.

What kind of questions can we ask?

Some questions

(i) What is the sampling method?

(ii) How is the sampling method related to the binomial distribution?

(iii) How can p be interpreted?

Random draws

Make 5 random draws with replacement from a box of 7 white and 3 black balls.

- Find the probability of the event WWWBB.
- Are the events WWWBB and WBWBW mutually exclusive?
- The probability of getting exactly three white balls is

$$\binom{5}{3}\left(\frac{7}{10}\right)^3\left(\frac{3}{10}\right)^2$$

Why is this formula correct?

Poisson probabilities

Let X and Y be independent Poisson random variables with rates 1 and 2. Let $Z = X + Y$.

- Find $\text{P}(Z = 0)$, $\text{P}(Z = 1)$ and $\text{P}(Z = 2)$.
- Do the answers above confirm a general fact?

2012 P2 Q6

On a remote island a zoologist measures the tail lengths of a random sample of 20 squirrels. In a species of squirrel known to her, the tail lengths have mean 14.0cm. She carries out a test, at the 5% significance level, of whether squirrels on the island have the same mean tail length as the species known to her. She assumes that the tail lengths of squirrels on the island are normally distributed with standard deviation 3.8cm.

Let L be the mean tail length. Under the null hypothesis,

(i) what is the expectation of L?

(ii) what is the variance of L?

(iii) find $\mathrm{P}(L > 15.8)$.

2011 P2 Q10

In a factory, the time in minutes for an employee to install an electronic component is a normally distributed continuous variable T. The standard deviation of T is 5.0 and under ordinary conditions the expected value of T is 38.0.

Under ordinary conditions,

(i) what is the distribution of the mean of 50 independent random times?

(ii) what is the chance that it exceeds 37.1 minutes?

2011 P2 Q11

A committee of 10 people is chosen at random from a group consisting of 18 women and 12 men. The number of women on the committee is denoted by R. Find $P(R = 4)$.

Questions:

(i) The process is the same as drawing without replacement. What is the probability of getting 4 women followed by 6 men? What about 6 men followed by 4 women?

(ii) Assuming that the probability of any sequence of 4 women and 6 men is the same, find $P(R = 4)$.

Random draws again

Make 3 random draws without replacement from a box of 7 white and 3 black balls. Let X_i be the result of draw number i.

- The probability that $X_1 = \text{W}$, $X_2 = \text{W}$ and $X_3 = \text{B}$ is

$$\frac{7}{10} \times \frac{6}{9} \times \frac{3}{8}$$

  True or false: $3/8 = \text{P}(X_3 = \text{B})$.

- Find the distribution of X_1, and of X_2.

- Are X_1 and X_2 independent? Or some simpler questions.

Simulation

Trial	X_1	X_2	X_3
1	W	W	W
2	B	W	W
3	W	W	B
4	W	W	W
⋮	⋮	⋮	⋮

Table : Results of many trials, showing first 4 trials.

(i) In the first column, roughly what fraction of rows have W? What about the second or the third column?

(ii) In the first two columns, roughly what fraction of rows have WW? What about WB?

(iii) Among rows where the first column is W, roughly what fraction of rows have W in the second column?

Coin tossing

Toss a fair coin 9 times, then count the number of heads obtained. Repeat the experiment 500 times. What if the number of tosses is 100 times? In the latter case, roughly how many of them are larger than 60?

- What can you say about these 500 numbers?
- What if the number of tosses is 100 times?
- In the latter case, roughly how many of them are larger than 60?

Playing cards

From a well-shuffled deck of cards, deal two cards. Let

$$\begin{aligned} A &= \{\text{first card is the ace of heart}\} \\ B &= \{\text{second card is the ace of diamond}\} \end{aligned}$$

- Find $\mathrm{P}(B)$ and $\mathrm{P}(B|A)$.
- Find $\mathrm{P}(A|B)$.
- Are A and B independent? Are they mutually exclusive?

Playing cards

From a well-shuffled deck of cards, deal one. Let

$$\begin{aligned} A &= \{\text{first card is a heart}\} \\ B &= \{\text{first card is an ace}\} \end{aligned}$$

Are A and B independent? Someone works it out this way:

Since $P(A) = 1/4$ *and* $P(B) = 4/52 = 1/13$,

$$P(A \cap B) = \frac{1}{4} \times \frac{1}{13} = \frac{1}{52}$$

A and B are independent.

If the argument correct? Is the conclusion correct?

Conclusion

- To solve an examination problem effectively, students need to solve a series of elementary problems.
- Variable question styles help consolidate basic skills. If questions are too hard, possible to simplfy.
  1. Backward questions are useful, and allow students to evaluate reasoning too complicated for them to construct.
  2. Similar point for simulation questions, though the teacher has to do some computer coding.
- Construction of questions from exam or tutorial problems helps detect possible errors in problems.
- Apply to topics outside probability and statistics?