Answers - Probability and the binomial distribution

Year: Year 12

Answer to Question 1

a. Use the addition rule, subtracting the intersection so that it is not counted twice. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.4 - 0.2 = \mathbf{0.7}$. So $P(A \cup B) = 0.7$.

b. $A' \cap B'$ means neither A nor B happens, which is the complement of $A \cup B$. Neither event happens: $1 - P(A \cup B) = 1 - 0.7 = \mathbf{0.3}$.

c. Use the definition of conditional probability. $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.2}{0.4} = \mathbf{0.5}$. Knowing that B has happened does not change the probability of A.

d. Two events are independent when $P(A \cap B) = P(A) \times P(B)$. $P(A)P(B) = 0.5 \times 0.4 = 0.2 = P(A \cap B)$, so they **are independent** (also $P(A|B) = P(A)$).

Answer to Question 2

a. Fill in the Venn diagram from the middle outwards: start with the 9 who play both. Football only $22 - 9 = 13$; tennis only $18 - 9 = 9$; both 9. Total playing $= 13 + 9 + 9 = 31$, so neither $= 40 - 31 = \mathbf{9}$. So 9 students play neither sport.

b. Football only is the part of F outside T. $\dfrac{13}{40} = \mathbf{0.325}$. The probability is 0.325.

c. Given F, the sample space is reduced to the 22 football players, of whom 9 also play tennis. $P(T|F) = \dfrac{9}{22} = \mathbf{0.409}$.

d. Exactly one sport means football only or tennis only. Exactly one sport: $13 + 9 = 22$, probability $\dfrac{22}{40} = \mathbf{0.55}$. The probability is 0.55.

Answer to Question 3

a. Without replacement, the probabilities on the second branches change because one counter has gone. First: R with $\frac{5}{8}$, B with $\frac{3}{8}$. After a red: R $\frac{4}{7}$, B $\frac{3}{7}$. After a blue: R $\frac{5}{7}$, B $\frac{2}{7}$. The branches from each point add up to 1.

b. Multiply along the branches R then R. $P(RR) = \dfrac{5}{8} \times \dfrac{4}{7} = \dfrac{20}{56} = \dfrac{5}{14}$. The probability that both are red is $\dfrac{5}{14}$.

c. Different colours means red then blue, or blue then red; add the two paths. $P(RB) + P(BR) = \frac{5}{8} \times \frac{3}{7} + \frac{3}{8} \times \frac{5}{7} = \frac{15}{56} + \frac{15}{56} = \frac{30}{56} = \frac{15}{28}$. The probability is $\frac{15}{28}$.

d. At least one blue is the complement of both red. $1 - P(RR) = 1 - \frac{5}{14} = \frac{9}{14}$. The probability is $\frac{9}{14}$.

Answer to Question 4

a. Mutually exclusive events cannot happen together, so $P(X \cap Y) = 0$ and the probabilities simply add. $P(X \cup Y) = 0.3 + 0.5 = \mathbf{0.8}$.

b. For independent events $P(X \cap Y) = P(X)P(Y)$; then use the addition rule. $P(X \cap Y) = 0.3 \times 0.5 = 0.15$, so $P(X \cup Y) = 0.8 - 0.15 = \mathbf{0.65}$.

c. Mutually exclusive means $P(X \cap Y) = 0$. Independent means $P(X \cap Y) = P(X)P(Y) > 0$. These contradict each other. So if both probabilities are non-zero, the events cannot be both mutually exclusive and independent.

Answer to Question 5

a. Use $P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$ with $n = 8$, $p = 0.3$, $r = 2$. $\binom{8}{2}(0.3)^2(0.7)^6 = 28 \times 0.09 \times 0.1176 = \mathbf{0.2965}$. So $P(X = 2) = 0.2965$ to 4 decimal places.

b. $P(X \leq 1)$ is the sum of $P(X = 0)$ and $P(X = 1)$. $P(X = 0) + P(X = 1) = 0.7^8 + 8 \times 0.3 \times 0.7^7 = 0.0576 + 0.1977 = \mathbf{0.2553}$. So $P(X \leq 1) = 0.2553$.

c. Use the complement: $P(X \geq 3) = 1 - P(X \leq 2)$, with $P(X \leq 2) = P(X \leq 1) + P(X = 2)$. $1 - P(X \leq 2) = 1 - (0.2553 + 0.2965) =$ **0.4482**. So $P(X \geq 3) = 0.4482$.

d. For a binomial distribution $E(X) = np$ and $\mathrm{Var}(X) = np(1-p)$. $E(X) = np =$ **2.4**; $\mathrm{Var}(X) = npq = 8 \times 0.3 \times 0.7 =$ **1.68**.

Answer to Question 6

a. Yes: $X \sim B(10, \frac{1}{6})$ (fixed number of independent trials, two outcomes, constant probability).

b. Not exactly: the probability changes after each card is removed, so the trials are not independent. (It is a hypergeometric situation.)

c. No: the number of trials is not fixed; it is geometric.

d. A fixed number n of trials; each trial has two outcomes (success/failure); the trials are independent; the probability of success p is the same for each trial.

Answer to Question 7

a. The three days are independent, so multiply the probabilities of a dry day. $0.7^3 =$ **0.343**. The probability of three dry days is 0.343.
b. The number of rainy days in 5 days is $B(5, 0.3)$. $X \sim B(5, 0.3)$: $\binom{5}{2}(0.3)^2(0.7)^3 = 10 \times 0.09 \times 0.343 =$ **0.3087**. The probability of exactly 2 rainy days is 0.3087.
c. At least one rainy day is the complement of 7 dry days. $1 - 0.7^7 = 1 - 0.0824 =$ **0.9176**. It is very likely (about 92%) to rain at least once in a week.
d. The expected value of $B(n, p)$ is np. Dry days $\sim B(30, 0.7)$, mean $30 \times 0.7 =$ **21**. About 21 dry days are expected in a 30-day month.

Answer to Question 8

a. There are 20 independent bulbs, each faulty or not, with the same probability 0.04. $X \sim B(20, 0.04)$, where X is the number of faulty bulbs in a box.
b. No faulty bulbs means all 20 are good. $0.96^{20} =$ **0.442**. The probability is about 0.442.
c. Add $P(X = 0)$, $P(X = 1)$ and $P(X = 2)$ using the binomial formula. $P(X \leq 2) = 0.96^{20} + 20(0.04)(0.96)^{19} + 190(0.04)^2(0.96)^{18} = 0.4420 + 0.3683 + 0.1458 =$ **0.956**. The probability of at most 2 faulty bulbs is about 0.956.
d. More than 2 is the complement of at most 2. $P(X > 2) = 1 - 0.9561 =$ **0.0439**. $E(X) = 20 \times 0.04 =$ **0.8**. About 4.4% of boxes lead to a refund, and on average there are 0.8 faulty bulbs per box.

Answer to Question 9

a. The first set of branches shows whether a person has the disease; the second shows the test result. First branch: disease 0.02, no disease 0.98. Second branch: positive 0.95 (if disease), 0.03 (if no disease).
b. A positive test can come from a person with the disease or from a healthy person; add the two paths. $P(+) = 0.02 \times 0.95 + 0.98 \times 0.03 = 0.019 + 0.0294 =$ **0.0484**. So about 4.84% of people test positive.
c. Use $P(D|+) = \dfrac{P(D \cap +)}{P(+)}$. $P(D|+) = \dfrac{0.019}{0.0484} =$ **0.393**. Only about 39% of people who test positive actually have the disease.
d. The disease is rare, so most positive results come from the large healthy group even though the false-positive rate is small: 0.0294 of the population against 0.019 genuinely ill.

Answer to Question 10

a. There are 12 independent questions, each correct with probability $\frac{1}{4}$. $X \sim B(12, 0.25)$. This is the number of correct answers.
b. Use the binomial formula with $r = 3$. $\binom{12}{3}(0.25)^3(0.75)^9 = 220 \times 0.015625 \times 0.07508 =$ **0.258**. The probability of exactly 3 correct answers is about 0.258.
c. Passing means 6 or more correct; use the complement of $P(X \leq 5)$. $P(X \geq 6) = 1 - P(X \leq 5) = 1 - 0.9456 =$ **0.0544** (about 5.4%). So she has only about a 5% chance of passing by guessing.
d. The expected value of $B(n, p)$ is np. $E(X) = 12 \times 0.25 =$ **3**. She expects 3 marks, half the pass mark, so passing by guessing is unlikely.

Answer to Question 11

a. The number of late days in 5 days is $B(5, 0.2)$; never late means 0 late days. $0.8^5 =$ **0.328**. The probability is about 0.328.
b. Use the binomial formula with $r = 2$. $\binom{5}{2}(0.2)^2(0.8)^3 = 10 \times 0.04 \times 0.512 =$ **0.2048**. The probability is 0.2048.
c. At least 3 means exactly 3, 4 or 5; add the three probabilities. $P(X \geq 3) = \binom{5}{3}(0.2)^3(0.8)^2 + \binom{5}{4}(0.2)^4(0.8) + (0.2)^5 = 0.0512 + 0.0064 + 0.00032 =$ **0.05792**. The probability is about 0.0579.
d. Multiply the probability of a letter in one week by the number of weeks. $38 \times 0.05792 = 2.2$, so about **2 letters**. About 2 letters would be expected in a school year.

Answer to Question 12

a. The number of baskets is $B(10, 0.7)$; use the binomial formula with $r = 8$. $\binom{10}{8}(0.7)^8(0.3)^2 = 45 \times 0.0576 \times 0.09 =$ **0.2335**. The probability of exactly 8 baskets is about 0.2335.
b. Use the complement of $P(X \leq 7)$, from a calculator or tables. $P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.6172 =$ **0.3828**. The probability is about 0.383.
c. Use the cumulative binomial function on a calculator. $P(X \leq 6) =$ **0.3504**. The probability is about 0.350.
d. For $B(n, p)$ the mean is np and the variance is $np(1-p)$. Mean $= 10 \times 0.7 =$ **7**. Variance $= 10 \times 0.7 \times 0.3 = 2.1$, standard deviation $= \sqrt{2.1} =$ **1.45**.

Answer to Question 13

a. There are $6 \times 6 = 36$ equally likely outcomes; the totals of 7 are (1,6), (2,5), (3,4), (4,3), (5,2), (6,1). 6 outcomes out of 36: $\frac{1}{6}$. The probability is $\frac{1}{6}$.

b. At least one six is the complement of no sixes, and $P(\text{no six}) = \frac{5}{6} \times \frac{5}{6}$. $1 - \left(\frac{5}{6}\right)^2 = \frac{11}{36}$. The probability is $\frac{11}{36}$.

c. Given at least one six, the sample space is reduced to 11 outcomes; count those with a total of at least 10. With at least one six there are 11 outcomes. Total ≥ 10 among them: (6,4), (4,6), (6,5), (5,6), (6,6): 5. So $\frac{5}{11}$. The conditional probability is $\frac{5}{11}$.

d. Let A be at least one six and B a total of at least 10; compare $P(A \cap B)$ with $P(A)P(B)$. $P(\text{total} \geq 10) = \frac{6}{36} = \frac{1}{6}$ (the outcomes (4,6), (5,5), (5,6), (6,4), (6,5), (6,6)). $P(A)P(B) = \frac{11}{36} \times \frac{1}{6} = \frac{11}{216}$, but $P(A \cap B) = \frac{5}{36} = \frac{30}{216}$. Not equal, so **not independent**.

Answer to Question 14

a. Without replacement, the probabilities change after each sock is taken; multiply along the path. $\frac{6}{10} \times \frac{5}{9} \times \frac{4}{8} = \frac{120}{720} = \frac{1}{6}$. The probability that all three are red is $\frac{1}{6}$.

b. Exactly two green socks can happen in three orders, and each order has the same probability. Three orders (GGR, GRG, RGG), each with probability $\frac{4}{10} \times \frac{3}{9} \times \frac{6}{8} = \frac{72}{720} = 0.1$. Total $= 3 \times 0.1 =$ **0.3**. The probability is 0.3.

c. At least one green is the complement of all three red. $1 - \frac{1}{6} = \frac{5}{6}$. The probability is $\frac{5}{6}$.

d. The probability of green changes as socks are removed (no replacement), so the trials are not independent. So the conditions for a binomial model are not satisfied.

Answer to Question 15

a. Read the totals from the table and divide by 150. $P(\text{bus}) = \frac{60}{150} = 0.4$. $P(\text{Y13} \cap \text{car}) = \frac{25}{150} = \frac{1}{6}$. So $P(\text{bus}) = 0.4$ and $P(\text{Y13 and car}) = \frac{1}{6}$.

b. Given Year 13, only the 80 Year 13 students are considered. $\frac{25}{80} =$ **0.3125**. So 31.25% of Year 13 students come by car.

c. Given walk, only the 50 walkers are considered. $\frac{30}{50} = \mathbf{0.6}$. So 60% of walkers are in Year 12.
d. Independence requires $P(\text{Y13} \cap \text{car}) = P(\text{Y13}) \times P(\text{car})$.
$P(\text{Y13}) \times P(\text{car}) = \frac{80}{150} \times \frac{40}{150} = 0.142$, but $P(\text{Y13} \cap \text{car}) = 0.167$. They differ, so the events are **not independent**: Year 13 students are more likely to come by car.

Answer to Question 16

a. At least one advert is the complement of no advert in n independent opens. $P(\text{none}) = 0.9^n$, so $P(\geq 1) = 1 - 0.9^n$. For $n = 20$: $1 - 0.1216 = 0.8784$ ✓. The entry in the table is correct.
b. Solve the inequality with logarithms; dividing by $\ln 0.9$ reverses the sign.
$1 - 0.9^n > 0.99 \Rightarrow 0.9^n \frac{\ln 0.01}{\ln 0.9} = 43.7$. So $n = \mathbf{44}$.
c. Substitute $n = 43$ and $n = 44$ into $1 - 0.9^n$. $n = 43$: $1 - 0.9^{43} = 0.9892$ (below 0.99). $n = 44$: 0.9903 (above) ✓. So 44 is the smallest number of opens.
d. The number of adverts is $B(44, 0.1)$, with mean np. $44 \times 0.1 = \mathbf{4.4}$. On average 4.4 adverts would be shown in 44 opens.

Answer to Question 17

a. An item is defective if it comes from one of the machines and is defective; add the three products from the table. $P(D) = 0.005 + 0.006 + 0.008 = \mathbf{0.019}$. So 1.9% of the factory's output is defective.
b. Use conditional probability: $P(C|D) = \frac{P(C \cap D)}{P(D)}$. $P(C|D) = \frac{0.008}{0.019} = \mathbf{0.421}$. So about 42% of defective items come from machine C, even though it makes only 20% of the output.
c. Assuming the two items are independent, multiply the probabilities. (Independent items) $0.019^2 = \mathbf{0.000361}$. The probability is 0.000361.
d. Multiply the batch size by the probability of a defect. $5000 \times 0.019 = \mathbf{95}$. About 95 defective items are expected.

Answer to Question 18

a. Read $P(X \leq 4)$ directly; for a single value subtract consecutive cumulative probabilities. $P(X \leq 4) = \mathbf{0.2173}$.
$P(X = 6) = P(X \leq 6) - P(X \leq 5) = 0.6098 - 0.4032 = \mathbf{0.2066}$.
b. $X \geq 8$ is the complement of $X \leq 7$. $P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.7869 = \mathbf{0.2131}$. The probability is 0.2131.

c. Look for the first cumulative probability above 0.5. $P(X \leq 5) = 0.4032$ and $P(X \leq 6) = 0.6098 > 0.5$, so $c =$ **6**.

d. For $B(n, p)$ the mean is np and the variance is $np(1-p)$. Mean $= 15 \times 0.4 = 6$; standard deviation $= \sqrt{15 \times 0.4 \times 0.6} = \sqrt{3.6} = 1.90$. The value $c = 6$ is the median and it equals the mean, as expected for a roughly symmetrical distribution.

Answer to Question 19

a. Use the inclusion-exclusion principle for three sets. Union $= 34 + 28 + 20 - 12 - 9 - 8 + 5 = 58$. None $= 60 - 58 =$ **2**. So 2 members do none of the sports.

b. Fill in the Venn diagram from the centre outwards: all three, then each pair only, then each sport only. All three: 5. Swim and cycle only: $12 - 5 = 7$; swim and run only: $9 - 5 = 4$; cycle and run only: $8 - 5 = 3$. Swim only: $34 - 7 - 4 - 5 = 18$; cycle only: $28 - 7 - 3 - 5 = 13$; run only: $20 - 4 - 3 - 5 = 8$. Exactly one: $18 + 13 + 8 =$ **39**. So 39 members do exactly one sport.

c. Given R, only the 20 runners are considered; 8 of them also cycle. $P(C|R) = \dfrac{8}{20} =$ **0.4**. So 40% of runners also cycle.

d. Exactly two sports means one of the three \"pair only\" regions. Exactly two: $7 + 4 + 3 = 14$; probability $\dfrac{14}{60} =$ **0.233**. The probability is $\dfrac{7}{30} \approx 0.233$.

Answer to Question 20

a. Use $E(X) = np$ and $\mathrm{Var}(X) = npq$; dividing the variance by the mean gives q. $np = 6$ and $npq = 4.2$, so $q = \dfrac{4.2}{6} = 0.7$, $p = 0.3$ and $n = \dfrac{6}{0.3} =$ **20**. So $X \sim B(20, 0.3)$.

b. Use the binomial formula with $n = 20$, $p = 0.3$, $r = 6$. $\binom{20}{6}(0.3)^6(0.7)^{14} = 38760 \times 0.000729 \times 0.0067822 =$ **0.1916**. So $P(X = 6) \approx 0.192$.

c. $\dfrac{\binom{n}{r+1}p^{r+1}q^{n-r-1}}{\binom{n}{r}p^r q^{n-r}} = \dfrac{n-r}{r+1} \times \dfrac{p}{q}$. With $n = 20$, $p/q = \frac{3}{7}$: ratio $= \dfrac{3(20-r)}{7(r+1)}$. It exceeds 1 when $60 - 3r > 7r + 7$, i.e. r, so the probabilities increase up to $r = 6$ and then fall. For $r = 5$: ratio $= \dfrac{45}{42} > 1$; for $r = 6$: $\dfrac{42}{49}$. So $P(X = 6)$ is the largest: the mode is 6.