Answers - Probability and random variables

Year: University Year 1

Answer to Question 1

a. By the addition rule, $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.5 - 0.3 =$ **0.8**; we subtract the overlap because it was counted twice. By De Morgan's law, $A^c \cap B^c = (A \cup B)^c$, so $P(A^c \cap B^c) = 1 - 0.8 =$ **0.2**.

b. The event B splits into the disjoint parts $A \cap B$ and $A^c \cap B$. Hence $P(A^c \cap B) = P(B) - P(A \cap B) = 0.5 - 0.3 =$ **0.2**.

c. By definition, $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.3}{0.5} =$ **0.6**, which equals $P(A)$. Knowing that B happened does not change the probability of A. Equivalently $P(A)P(B) = 0.3 = P(A \cap B)$, so A and B **are independent**.

Answer to Question 2

a. The tree has first branches "urn 1" and "urn 2", each with probability $\frac{1}{2}$, and second branches red/blue with probabilities $\frac{3}{5}$, $\frac{2}{5}$ for urn 1 and $\frac{1}{5}$, $\frac{4}{5}$ for urn 2. By the law of total probability, $P(R) = \frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{1}{5} = 0.3 + 0.1 =$ **0.4**.

b. By Bayes' theorem, $P(U_1 \mid R) = \dfrac{P(U_1 \cap R)}{P(R)} = \dfrac{0.3}{0.4} =$ **0.75**. A red ball makes urn 1 more likely, because urn 1 contains more red balls.

c. $P(\text{blue}) = 1 - 0.4 = 0.6$ and $P(U_1 \cap \text{blue}) = \frac{1}{2} \times \frac{2}{5} = 0.2$. So $P(U_1 \mid \text{blue}) = \dfrac{0.2}{0.6} = \dfrac{1}{3}$; a blue ball points towards urn 2.

Answer to Question 3

a. All the probabilities are non-negative and $0.1 + 0.3 + 0.4 + 0.2 = 1$, so the distribution is valid. Then $E[X] = \sum x P(X = x) = 0(0.1) + 1(0.3) + 2(0.4) + 3(0.2) = 0 + 0.3 + 0.8 + 0.6 =$ **1.7**.

b. $E[X^2] = 0(0.1) + 1(0.3) + 4(0.4) + 9(0.2) = 0.3 + 1.6 + 1.8 = 3.7$. Using $\operatorname{Var}(X) = E[X^2] - (E[X])^2$, we get $3.7 - 1.7^2 = 3.7 - 2.89 =$ **0.81**, so the standard deviation is 0.9.

c. Expectation is linear: $E[2X - 1] = 2(1.7) - 1 =$ **2.4**. For the variance, $\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$, so $\operatorname{Var}(2X - 1) = 2^2 \times 0.81 =$ **3.24**. The constant -1 shifts the values but does not change their spread.

Answer to Question 4

a. For $X \sim B(n,p)$, $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$. Here
$P(X=3) = \binom{12}{3}(0.25)^3(0.75)^9 = 220 \times 0.015625 \times 0.07508 \approx \mathbf{0.2581}$.

b. $P(X \le 1) = P(X=0) + P(X=1)$. We have $P(X=0) = 0.75^{12} \approx 0.03168$ and $P(X=1) = 12(0.25)(0.75)^{11} \approx 0.12671$. Adding, $P(X \le 1) \approx \mathbf{0.1584}$.

c. For a binomial distribution, $E[X] = np = 12 \times 0.25 = \mathbf{3}$ and $\mathrm{Var}(X) = np(1-p) = 3 \times 0.75 = 2.25$. The standard deviation is $\sigma = \sqrt{2.25} = \mathbf{1.5}$.

Answer to Question 5

a. For $X \sim \mathrm{Po}(\lambda)$, $P(X=k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$. So
$P(X=2) = e^{-2.5}\dfrac{2.5^2}{2!} = 0.082085 \times 3.125 \approx \mathbf{0.2565}$.

b. Add the first three probabilities:
$P(X \le 2) = e^{-2.5}\left(1 + 2.5 + \dfrac{2.5^2}{2}\right) = 0.082085 \times 6.625 \approx \mathbf{0.5438}$. Using the complement, $P(X \ge 3) = 1 - P(X \le 2) = 1 - 0.5438 = \mathbf{0.4562}$.

c. For a Poisson distribution, the mean and the variance are both equal to λ, so $E[X] = \mathrm{Var}(X) = \mathbf{2.5}$.

Answer to Question 6

a. Standardise with $Z = \dfrac{X-50}{8} \sim N(0,1)$. Then
$P(X < 62) = \Phi\left(\dfrac{62-50}{8}\right) = \Phi(1.5) = \mathbf{0.9332}$. For the interval, the z-values are ± 1.25, so $P(40 < X < 60) = \Phi(1.25) - \Phi(-1.25) = 2(0.8944) - 1 = \mathbf{0.7887}$, using symmetry.

b. We need $\Phi(z) = 0.9$, which gives $z = 1.2816$ from tables. Then $x = \mu + z\sigma = 50 + 1.2816 \times 8 = \mathbf{60.25}$.

c. For a continuous variable, $E[Y] = \int y f(y)\,dy = \int_0^1 3y^3 dy = \dfrac{3}{4}$. Also
$E[Y^2] = \int_0^1 3y^4 dy = \dfrac{3}{5}$, so $\mathrm{Var}(Y) = \dfrac{3}{5} - \left(\dfrac{3}{4}\right)^2 = \dfrac{3}{5} - \dfrac{9}{16} = \dfrac{3}{80}$.

Answer to Question 7

a. Of the 200 ill people, 90% test positive: 180 positive and 20 negative. Of the 9 800 healthy people, 95% test negative: $9\,310$ negative and 490 positive.

b. By total probability, $P(+) = 0.02 \times 0.9 + 0.98 \times 0.05 = 0.018 + 0.049 = \mathbf{0.067}$; that is 670 positives per 10 000.

c. By Bayes' theorem, $P(\text{ill} \mid +) = \dfrac{0.018}{0.067} \approx \mathbf{0.269}$ (or $180/670$ from the table). Only about 27% of people who test positive are ill: the disease is rare, so most positives come from the large healthy group.

d. The probability 0.2687 becomes the new prior. Then $\dfrac{0.2687 \times 0.9}{0.2687 \times 0.9 + 0.7313 \times 0.05} = \dfrac{0.2418}{0.2784} \approx \mathbf{0.869}$. A second positive test makes illness very likely.

Answer to Question 8

a. By the law of total probability, $P(F) = 0.5(0.02) + 0.3(0.03) + 0.2(0.05) = 0.010 + 0.009 + 0.010 = \mathbf{0.029}$.

b. By Bayes' theorem, $P(C \mid F) = \dfrac{P(C)P(F \mid C)}{P(F)} = \dfrac{0.010}{0.029} \approx \mathbf{0.345}$.

c. In the same way $P(A \mid F) = \dfrac{0.010}{0.029} \approx 0.345$ and $P(B \mid F) = \dfrac{0.009}{0.029} \approx 0.310$. So **A and C are equally likely**, for different reasons: A supplies more chargers, while C has a higher fault rate.

d. Expected faulty chargers: $2000 \times 0.029 = \mathbf{58}$. Of these, about $2000 \times 0.5 \times 0.02 = \mathbf{20}$ come from factory A.

Answer to Question 9

a. Calls must occur independently of each other and singly (never two at exactly the same instant), at a constant average rate over the period.

b. $P(X = 0) = e^{-4}\dfrac{4^0}{0!} = e^{-4} \approx \mathbf{0.0183}$, so a quiet 10 minutes is rare.

c. Use the complement: $P(X > 6) = 1 - P(X \leq 6) = 1 - 0.8893 = \mathbf{0.1107}$, with $P(X \leq 6)$ from tables or a calculator.

d. The rate scales with time: in 5 minutes the mean is 2, so $Y \sim \text{Po}(2)$. Then $P(Y \leq 1) = e^{-2}(1 + 2) = 3e^{-2} \approx \mathbf{0.406}$.

e. Half an hour is three 10-minute periods, so the number of calls is $\text{Po}(3 \times 4) = \text{Po}(12)$, with mean and variance both equal to **12**.

Answer to Question 10

a. $Y \sim B(186,\, 0.05)$ with mean $186 \times 0.05 = \mathbf{9.3}$.
b. Overbooked when more than 180 turn up, that is $Y \leq 5$: $P(Y \leq 5)$.
c. $P(Y \leq 5) \approx \mathbf{0.0929}$, about a 9% risk.
d. $\mathrm{Po}(9.3)$ gives $P(Y \leq 5) \approx \mathbf{0.0986}$, close to the exact value because n is large and p small.
e. $186 \times 0.95 = \mathbf{176.7\ passengers}$.

Answer to Question 11

a. There are 36 equally likely outcomes. Only (6,6) wins £10, so $P(W = 10) = \dfrac{1}{36}$. The other five doubles win £4, so $P(W = 4) = \dfrac{5}{36}$. The remaining 30 outcomes win nothing: $P(W = 0) = \dfrac{30}{36}$.
b. $E[W] = 10 \times \dfrac{1}{36} + 4 \times \dfrac{5}{36} = \dfrac{10 + 20}{36} = \dfrac{5}{6} \approx \mathbf{£0.83}$. Each game brings in £2, so the school expects to gain $2 - \dfrac{5}{6} = \dfrac{7}{6} \approx \mathbf{£1.17}$ per game.
c. $E[W^2] = 100 \times \dfrac{1}{36} + 16 \times \dfrac{5}{36} = \dfrac{180}{36} = 5$, so $\mathrm{Var}(W) = 5 - \left(\dfrac{5}{6}\right)^2 = 5 - \dfrac{25}{36} \approx \mathbf{4.31}$.
d. Expected profit over 300 games: $300 \times \dfrac{7}{6} = \mathbf{£350}$.

Answer to Question 12

a. Standardise: $z = \dfrac{185 - 170}{7} = \dfrac{15}{7} \approx 2.143$. Then $P(H > 185) = 1 - \Phi(2.143) \approx \mathbf{0.0161}$, about 1.6% of students.
b. The z-values are $\pm\dfrac{10}{7} \approx \pm 1.429$. By symmetry, $P(160 < H < 180) = 2\Phi(1.429) - 1 \approx \mathbf{0.847}$.
c. We need the 95th percentile: $\Phi(z) = 0.95$ gives $z = 1.6449$. Then $h = 170 + 1.6449 \times 7 \approx \mathbf{181.5\ cm}$.
d. Expected number $= np = 500 \times 0.0161 \approx \mathbf{8\ students}$ would need to duck under the old doors.

Answer to Question 13

a. $E[T] = 6$, $\mathrm{Var}(T) = \dfrac{12^2}{12} = \mathbf{12}$, $P(T > 9) = \dfrac{3}{12} = \mathbf{0.25}$.
b. $P(S > 8) = e^{-8/5} \approx \mathbf{0.202}$.
c. $\dfrac{e^{-(s+t)/5}}{e^{-s/5}} = e^{-t/5} = P(S > t)$. The exponential distribution is memoryless: having waited does not make the bus more likely to come soon.
d. Route 1: **6 minutes**; route 2: **5 minutes**. The random route has the shorter mean wait here but a much larger chance of a long wait.

Answer to Question 14

a. A sum of independent normal variables is normal. Means add: $500 + 20 = 520$. Variances add: $16 + 1 = 17$. So $M \sim N(520, 17)$, with standard deviation $\sqrt{17} \approx 4.12$ g.
b. Standardise: $P(M < 515) = \Phi\left(\dfrac{515 - 520}{\sqrt{17}}\right) = \Phi(-1.213) \approx \mathbf{0.113}$.
c. $D = M_1 - M_2$ has mean 0 and variance $17 + 17 = 34$, so $D \sim N(0, 34)$. Then $P(|D| > 10) = 2\left(1 - \Phi\left(\dfrac{10}{\sqrt{34}}\right)\right) = 2(1 - \Phi(1.715)) \approx \mathbf{0.086}$.
d. $\mathrm{Var}(M_1 - M_2) = \mathrm{Var}(M_1) + (-1)^2\,\mathrm{Var}(M_2)$. The factor -1 is squared, so it becomes $+1$: subtracting a random quantity adds uncertainty rather than removing it.

Answer to Question 15

a. The first person can have any birthday, the second must avoid one day, the third two days, and so on; the events are multiplied because birthdays are independent.
b. $1 - q_{23} \approx \mathbf{0.507}$: just over one half.
c. $q_n \approx e^{-(0+1+\cdots+(n-1))/365} = e^{-n(n-1)/730}$. Setting this equal to $\frac{1}{2}$: $n(n-1) = 730 \ln 2 \approx 506$, so $n \approx \mathbf{23}$.
d. $1 - q_{30} \approx \mathbf{0.706}$. People compare with their own birthday (only 29 chances), but there are $\binom{30}{2} = 435$ pairs that could match.

Answer to Question 16

a. $e^{-0.3} = 0.7408$, $0.3e^{-0.3} = 0.2222$, $0.045e^{-0.3} = 0.0333$, and $1 - 0.9964 = 0.0036$.
b. Expected cost $1500 \times 0.3 = \mathbf{£450}$; standard deviation $1500\sqrt{0.3} \approx \mathbf{£822}$.
c. $Y \sim \mathrm{Po}(300)$.
d. $Y \approx N(300, 300)$: $P(Y \geq 331) \approx P\left(Z > \dfrac{330.5 - 300}{\sqrt{300}}\right) = P(Z > 1.761) \approx \mathbf{0.039}$.

e. Premiums bring in £480 000 and claims cost £450 000 on average. The premium covers claims unless $Y > 320$; $P(Y \geq 321) \approx P(Z > 1.184) \approx 0.12$, so in roughly one year in eight claims exceed premiums: the company needs reserves.

Answer to Question 17

a. Stick: $\dfrac{49}{150} \approx 0.33$. Switch: $\dfrac{101}{150} \approx 0.67$.

b. $P(H_3 \mid C_1) = \frac{1}{2}$ (the host chooses at random between 2 and 3), $P(H_3 \mid C_2) = 1$, $P(H_3 \mid C_3) = 0$.

c. $P(H_3) = \frac{1}{3}\left(\frac{1}{2} + 1 + 0\right) = \frac{1}{2}$, so $P(C_2 \mid H_3) = \dfrac{\frac{1}{3} \times 1}{\frac{1}{2}} = \dfrac{2}{3}$.

d. Your first choice is right with probability $\frac{1}{3}$, and the host's action does not change that. Switching wins whenever your first choice was wrong, with probability $\frac{2}{3}$. The simulation (0.67) agrees well.

Answer to Question 18

a. $X \sim B(400, 0.45)$: mean 180, variance $400 \times 0.45 \times 0.55 = \mathbf{99}$.

b. $np = 180$ and $n(1-p) = 220$ are both large, so the distribution is nearly symmetric and bell-shaped.

c. $P(X \geq 200) \approx P\left(Z > \dfrac{199.5 - 180}{\sqrt{99}}\right) = P(Z > 1.960) \approx \mathbf{0.025}$.

d. $\sigma = \sqrt{99} \approx 9.95$: from about 160 to 200, i.e. a sample proportion between **40% and 50%** in about 95% of polls.

e. $\sqrt{\dfrac{0.45 \times 0.55}{n}} \leq 0.01$ gives $n \geq \dfrac{0.2475}{0.0001} = \mathbf{2\ 475}$.

Answer to Question 19

a. Add the columns for X: $P(X=0) = 0.25$, $P(X=1) = 0.45$, $P(X=2) = 0.30$. Add the rows for Y: $P(Y=0) = 0.4$, $P(Y=1) = 0.6$.

b. $E[X] = 0(0.25) + 1(0.45) + 2(0.30) = \mathbf{1.05}$ and $E[Y] = \mathbf{0.6}$. Only cells with $Y = 1$ and $X \neq 0$ contribute to $E[XY]$: $1 \times 0.25 + 2 \times 0.20 = \mathbf{0.65}$.

c. $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0.65 - 1.05 \times 0.6 = \mathbf{0.02}$, a very weak positive association.

d. **No.** Independence needs every cell to equal the product of its marginals. $P(X=1, Y=0) = 0.20$ but $P(X=1)P(Y=0) = 0.18$. (Note that $P(X=0, Y=0) = 0.10 = 0.25 \times 0.4$: one matching cell is not enough.)

e. By linearity, $E[2.5X + 3Y] = 2.5(1.05) + 3(0.6) = \mathbf{£4.425}$, about £4.43 per customer.

Answer to Question 20

a. Each new box gives a figure not yet owned with probability $\frac{6-k}{6}$, independently of the past: the waiting time to the first success is geometric.

b. $E[G] = \sum_{g \geq 1} g(1-p)^{g-1}p = p \cdot \dfrac{1}{p^2} = \dfrac{1}{p}$.

c. Summing the means
$\frac{6}{6} + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1} = 6\left(1 + \frac{1}{2} + \cdots + \frac{1}{6}\right) = 6 \times 2.45 = \mathbf{14.7\ boxes}$.

d. $6 \ln 6 \approx 10.75$; adding the correction $0.577n$ (Euler's constant) gives 14.2, close to 14.7. The last figures take the longest: the final one alone needs 6 boxes on average.