Problem set for Probability and Statistics 1 — 16/18 March 2026

Basic problems about expected value

Recall the formula for expected value:

$$\mathbb{E}[X] = \sum_{x \in \mathsf{Im}(X)} x \cdot \mathbb{P}[X = x],$$

1. (a) Let $\mathbb{P}[X = 100] = p$, $\mathbb{P}[X = 0] = 1 - p$. Determine $\mathbb{E}[X]$.
(b) Let $\mathbb{P}[Y = 100] = p$, $\mathbb{P}[Y = 99] = 1 - p$. Determine $\mathbb{E}[Y]$.

2. Suppose solving one problem takes X minutes, where $X \in \{8, 9, 10, 11, 12\}$. The time required for solving the problem is random (for instance, it may depend on current weather), and the probability function is $p_X(8) = p_X(9) = 0.1$, $p_X(10) = p_X(11) = 0.2$, $p_X(12) = 0.4$. Compute $\mathbb{E}[X]$.

3. (a) We have an unlimited number of black and red socks in a drawer, unpaired. We take out socks without looking, with both colours being equally likely to be drawn. How many socks do we have to take out before we have two of the same colour in the worst case, and how many on average?
(b) Solve the same problem for three different colours.

Linearity of $\mathbb{E}$

$$\mathbb{E}[a_1 X_1 + \cdots + a_n X_n] = a_1 \mathbb{E}[X_1] + \cdots + a_n \mathbb{E}[X_n]$$

4. We have a biased coin, which lands heads with probability p. We toss it n times.
(a) Let X be the number of positions where the coin landed heads first and then immediately tails. (For example, if $n = 6$ and the sequence is HTHHTH, then $X = 2$.) Determine $\mathbb{E}[X]$.
(b) Now let Y be the number of occurrences of the sequence HTH (here "occurrence" means that the three symbols occupy three consecutive positions, in this order). What is $\mathbb{E}[Y]$?

5. We again have a coin which lands head with probability p, and toss it $\binom{n}{2}$ times. Based on the outcomes, we construct a graph with vertices $V = \{1, 2, \ldots, n\}$. For all pairs $\{i, j\} \in \binom{V}{2}$, we determine if they are connected by an edge – this happens when the corresponding coin toss results in heads. The resulting graph is called a random graph $G(n, p)$.
(a) Show that the expected value of the number of edges in the graph is $p\binom{n}{2}$.
(b) Show that the expected value of the number of triangles in the random graph is $p^3\binom{n}{3}$.
(c) Let $3 \le k \le n$ be such that $\binom{n}{k}(p^{\binom{k}{2}} + (1-p)^{\binom{k}{2}}) < 1$. Argue that there is a graph on n vertices such that among any k vertices, some two of them are connected by an edge but not all of them are connected by an edge.
(d) For any $3 \le k \le n$ such that $n \ge \binom{n}{k}(p^{\binom{k}{2}} + (1-p)^{\binom{k}{2}})$, argue that there is a graph on $n - \lfloor \binom{n}{k}(p^{\binom{k}{2}} + (1-p)^{\binom{k}{2}}) \rfloor$ vertices such that among any k vertices, some two of them are connected by an edge but not all of them are connected by an edge.

Conditional expectation

$$\mathbb{E}[X|B] = \sum_{x \in \mathsf{Im}(X)} x \cdot \mathbb{P}[X = x|B]$$

$$\mathbb{E}[X] = \sum_i \mathbb{P}[B_i] \cdot \mathbb{E}[X|B_i] \qquad B_1, B_2, \ldots \text{ is a partition of } \Omega$$

6. In a quiz, there are 20 multiple-choice questions with four choices each. A correct answer gains 1 point, an incorrect answer deducts 1/4 point, and an unanswered question earns zero points. Each question has a probability q of being one of those known by Quizo and hence has a known correct answer. If Quizo does not know the correct answer, he is aware of it and can decide whether to guess.

(a) What is the expected value of the number of points Quizo will score if he answers only the questions for which he knows the answer?

(b) What if he decides to guess when he doesn't know the answer?

(c) How should the penalty for incorrect answers change to make the answers in parts a) and b) the same?

7. In a TV quiz show, a participant can choose one of two questions to answer. For question A, worth 1,000 Kč, he estimates that he will correctly answer with a probability of 0.8. For question B, his success probability is only 0.5 but for a correct answer he receives 2,000 Kč.

(a) What is the expected value of the winnings if he picks question A?

(b) What if he picks question B?

(c) Bonus: if the success probabilities are p_A, p_B, and the rewards m_A, m_B, how should the participant decide? * And what if there are more than two questions?

8. We roll a standard six-sided die; we terminate on the first roll whose result is not 6. Let X be the sum of all the rolls. Compute $\mathbb{E}[X]$.

Bonus problems

9. An indicator random variable I_A for an event A is defined as follows:

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

(a) What is $\mathbb{E}[I_A]$?

(b) Let $A = A_1 \cup A_2 \cup \cdots \cup A_n$. Verify the equality

$$1 - I_A = \prod_{i=1}^{n} (1 - I_{A_i}).$$

(c) Expand the above product and use the linearity of expectation. You will obtain the inclusion-exclusion principle which you know from discrete mathematics.

10. (Two envelopes problem) You are presented with two envelopes, both of which contain a positive amount of money, with one of them containing exactly twice as much as the other. You are then asked to pick one of them. However, after you have made your decision and before you have opened the envelope, you are offered the opportunity to switch to the other envelope. Supposing the envelope you picked contains k Kč, what is the expected value of the contents of the other envelope? Does this make sense? If not, where might the issue lie?

11. ***(Coupon collector problem)** There are n different types of Pokémon in a game. Each time you play, you get one random Pokémon. Every Pokémon is equally likely to appear. How many times do you need to play, on average, to collect all the Pokémon? How fast does the expectation grow? (See the hints at the end for help.)

12. * A fair coin is tossed repeatedly. Label the number of heads among the first i tosses as H_i and the number of tails among the first i tosses as T_i. Let N be the random variable for the smallest number for which $H_N > T_N$.

(a) For which values of n do we have $\mathbb{P}[N = n] = 0$? (See the hints at the end for help.)

(b) Let $R = \frac{H_N}{N}$. What is the distribution of R, given $N = n$?

(c) Let $f(n)$ be the number of sequences of n coin tosses which lead to the event $N = n$. Compute the values of $f(1)$, $f(2)$ and $f(3)$.

(d) Show that f satisfies the recurrence relation $f(n) = \sum_{k=2}^{n-1} f(k-1)f(n-k)$. (See the hints at the end for help.)

(e) Given that the *Catalan numbers* $C_m = \frac{1}{m+1}\binom{2m}{m}$ satisfy $C_0 = 1, C_m = \sum_{i=1}^{m} C_{i-1}C_{m-i}$, determine $f(n)$ for all n.

(f) Given the Maclaurin series $\arcsin(x) = \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(2n+1)(n!)^2} x^{2n+1}$, determine $\mathbb{E}[R]$.

For practice

13. King Louis wants to have a male heir so that he can name him Louis again. Each year, his wife gives birth to exactly one child, which is equally likely to be a boy or a girl, independent of previous attempts. All children survive. If a boy is born, Louis will not have any more children. Let S be the number of sons born and D be the number of daughters born.

(a) Determine $\mathbb{E}[S]$.

(b) Determine $\mathbb{E}[D]$.

14. We toss a coin with probability p of landing heads. What is the expected number of tosses we make if we stop when

(a) two consecutive tosses give different results?

(b) two consecutive tosses give the same result?

For problem 11: for $1 \leq i \leq n$, define variables T_i that track the number of times played between getting the $(i-1)^{\text{th}}$ Pokémon and getting the i^{th} Pokémon.

For part a) of problem 12: find expressions for $H_N + T_N$ and $H_N - T_N$ in terms of N.

For part d) of problem 12: For $N \geq 3$, let K be the smallest positive integer such that $H_K = T_K$, and consider cases based on the value of K.
