Problem set for Probability and Statistics 1 — 30 March 2026

Warm-up

1. Three six-sided dice are rolled. Is it more likely for the sum of the three rolls to be 11 or 12?

2. Which is more likely: getting at least one 6 when rolling four dice, or at least three 6s when rolling twelve dice?

3. Prove that if the events A, B are independent, then so are the events A, B^c; and so are the events A^c, B^c.

4. There are 100 coins in the chest. 99 of them are normal, but one has an eagle (which counts as tails) on both sides. We pull out a random coin and flip it six times, each time it lands tails. What is the probability that we have pulled a two-eagle coin? (Try to guess first, then calculate.)

5. For disease C we have two tests, A and B. The A test has both sensitivity and specificity $p = 0.95$. The B test always says the patient is healthy. Assume that $\mathbb{P}[C] = 0.01$.

(a) Calculate for both tests the probability of success (i.e., the correct answer) if we use them on a random patient. Think about what this says about the usefulness of both tests.

(b) For what value of p is the probability of success of both tests the same?

6. Let X, Y be the results of two independent rolls of a four-sided die (with numbers 1, ..., 4).

(a) What is the pdf of $Z_1 = \max(X, Y)$?

(b) What is the pdf of $Z_2 = XY$?

7. Historical data show that our server receives an average of 30 requests per minute. Use the Poisson distribution to determine the probability that the server will receive exactly 40 requests in the next minute.

Bonus problems

8. The logical formula $A \Rightarrow B$ is equivalent to $\neg B \Rightarrow \neg A$. We will look at analogies involving probability.

(a) Show that if $\mathbb{P}[B|A] = 1$, then $\mathbb{P}[A^c|B^c] = 1$.

(b) Show, however, that it is possible to have $\mathbb{P}[B|A] \geq 0.9$ but $\mathbb{P}[A^c|B^c] = 0$.

9. In a TV quiz show, a participant will be presented with two questions to answer, the show ends when he answers wrong. For question A, worth 1,000 Kč, he estimates that he will answer correctly 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 total winnings if he chooses to answer question A first?

(b) What if he picks question B first?

(c) Bonus: if the success probabilities are p_A, p_B, and the rewards m_A, m_B, how should the participant decide?

10. * Consider a permutation of $\{1, 2, \dots, n\}$ chosen uniformly at random from all possible permutations. Let the random variable X be the number of fixed points in this random permutation. Find $\mathbb{E}[X]$ and $\mathrm{var}(X)$.

11. * Let M be the number of emails received per day, S be the number of spam emails among them, H be the number of "ham" emails – those that are not spam. Assume that $M \sim \mathsf{Poi}(\lambda)$ and that each email, independently of others, has a probability p of being spam.

(a) Express $\mathbb{P}[S = k]$ (as an infinite sum) using the joint distribution of M and S.

(b) Derive that $S \sim \mathsf{Poi}(p\lambda)$.

(c) Derive that $H \sim \mathsf{Poi}((1 - p)\lambda)$ and also that H, S are independent random variables.

12. ***(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 as n increases? (See the hints at the end for help.)

Further practice

13. The probability of a data breach at our server for each given day is 0.01, independently for each day. Let T be the number of days until the first data breach.

(a) What is the distribution of T?

(b) Calculate $\mathbb{E}[T]$ and $Var(T)$.

(c) What is the probability that the server remains secure for an entire year?

14. Each software test can either find a bug (which we count as success) or not (this we count as failure). Assume the probability of finding a bug in one test is 0.05, and a developer performs 20 independent tests. Let X be the number of bugs found.

(a) What is the distribution of X?

(b) Calculate $\mathbb{E}[X]$ and $Var(X)$.

(c) What is the probability of finding exactly three bugs?

15. Independent random variables $X_1, \ldots, X_n$ have geometric distribution with parameters $p_1, \ldots, p_n$. What is the distribution of $X = \min(X_1, \ldots, X_n)$?

16. Consider a group of m married couples (i.e., a total of $2m$ individuals). Suppose that after ten years, each of these $2m$ people will still be alive with probability p, independently of the others. We assume every couple stays together until one of them dies.

Let L be the set of people who will be alive after ten years, and A their number (i.e., $A = |L|$). Furthermore, let B be the number of couples where both partners will be alive; thus, A, B are random variables satisfying $0 \leq A \leq 2m$ and $0 \leq B \leq m$. For each $a = 0, \ldots, 2m$, we want to compute $\mathbb{E}[B|A = a]$.

(a) Let us consider a specific individual. What is the probability that they will be alive after ten years, given that $A = a$? In other words, labelling that person as x, what is $\mathbb{P}[x \in L|A = a]$?

(b) Let us consider a specific married couple. What is the probability that both partners will be alive, given that $A = a$?

(c) Express B as the sum of m suitable indicator random variables.

(d) The linearity of expected value also holds for conditional expected value, i.e.,

$$\mathbb{E}[\sum_{i=1}^{m} X_i|J] = \sum_{i=1}^{m} \mathbb{E}[X_i|J]$$

for any event J and random variables $X_1, \ldots, X_m$. Utilize this to compute $\mathbb{E}[B|A = a]$.

(e) What is the distribution of the random variable A? (Either name it or write the probability function, i.e., determine $\mathbb{P}[A = a]$.)

(f) For a chosen a-element set of people M, what is the probability that it exactly corresponds to the set of survivors? In other words, what is $\mathbb{P}[L = M]$? And what about $\mathbb{P}[L = M|A = a]$?

For problem 12: 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.
