

Problem set for Probability and Statistics 1 — 23 March 2026

Warm-up

Recall:

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

$$\mathbb{E}[X] = \sum_i \mathbb{P}[B_i] \cdot \mathbb{E}[X|B_i]$$

B_1, B_2, \dots is a partition of Ω

1. My computer occasionally misbehaves: every day, there is a probability p that it freezes. If it freezes two days in a row, I start addressing the issue. What is the expected number of days before I have to address the issue?
2. 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 $\mathbb{E}[X]$.
3. 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 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 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? * And what if there are more than two questions?

Variance

Recall:

- (definition) $\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$
- (theorem) $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- (standard deviation) $\sigma(X) = \sqrt{\text{var}(X)}$

4. Assume that solving one problem takes X minutes, where $X \in \{8, 9, 10, 11, 12\}$. The duration is random and the probability density function is $p_X(8) = p_X(9) = 0.1$, $p_X(10) = p_X(11) = 0.2$, $p_X(12) = 0.4$. (From last week, we know that $\mathbb{E}[X] = 10.7$.) Find $\text{var}(X)$ and $\sigma(X)$.
5. Do the following *alternative definitions of variance* make sense? That is, do you learn anything interesting about random variable X when you find
 - (a) $\mathbb{E}[X - \mathbb{E}[X]]$?
 - (b) $\mathbb{E}[|X - \mathbb{E}[X]|]$?
 - (c) $\mathbb{E}[(X - \mathbb{E}[X])^k]$ (for $k = 3, 4, \dots$)?
6. Let $X \sim \text{Bin}(100, 0.5)$ and $Y \sim 10\text{Bin}(100, 0.05)$ (thus, $Y/10$ has a binomial distribution $\text{Bin}(100, 0.05)$). Compute $\mathbb{E}[X]$, $\text{var}(X)$, $\sigma(X)$ and the same for Y .
7. Let $X \sim \text{Pois}(\lambda)$. Show that $\mathbb{E}[X] = \lambda$ and $\text{var}(X) = \lambda$.

Joint distribution

For n discrete random variables $X_1, \dots, X_n : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$, their *joint probability mass function* is $p_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow [0, 1]$,

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \mathbb{P}[X_1 = x_1 \wedge \dots \wedge X_n = x_n]$$

Marginal distribution: For discrete r.v.'s X , and Y with joint pmf $p_{X,Y}$, the marginal of X is

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

8. We draw two cards from the standard 52-card deck. Let X denote the number of aces drawn and Y denote the number of kings drawn. Determine the associated probability function $p_{X,Y}$ and the marginals p_X and p_Y .

9. We toss a fair coin three times. Let X be the number of heads in the first two tosses, and Y be the number of tails in the last two tosses.

(a) Determine the joint probability function $p_{X,Y}$ and also marginals p_X, p_Y .

(b) Are X and Y independent?

(c) Determine $\mathbb{P}[X < Y]$.

(d) Determine the conditional probability function $p_{X|Y}$, i.e., the numbers $\mathbb{P}[X = x|Y = y]$ for all values of x, y .

10. We say that the random variables X and Y are *uncorrelated* if $Cov(X, Y) = 0$ where the *covariance* $Cov(X, Y)$ is defined as $Cov(X, Y) := \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

Define the random variable X that takes values in $\{1, 2, 3\}$ with $\mathbb{P}[X = i] = 1/3$ for each $i \in \{1, 2, 3\}$. Consider the random variable $Y = 1$ if $X = 2$, and $Y = 0$ otherwise. Show that X and Y are uncorrelated. Are X and Y independent random variables?

11. Let X, Y be two independent random variables having Poisson distribution with parameters λ and μ respectively.

(a) Find the distribution of the convolution $Z = X + Y$.

(b) Calculate $\mathbb{E}[Z(Z - 1)]$.

(c) Compute $\text{var}(Z)$.

(d) If $\mu = 2\lambda$, compute $Cov(X, Z)$.

12. * 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 $M \sim \text{Poi}(\lambda)$ 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 $S \sim \text{Poi}(p\lambda)$.

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

More practice problems

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 $\text{var}(T)$.

(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 $\text{var}(X)$.

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

15. (Jensen's inequality) Let g be a convex function and X be a random variable. Prove that $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$. (See the hints at the end for help.)

16. * 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 $\text{var}(X)$.

17. 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 $a = 0, \dots, 2m$, 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.,

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

for any event J and random variables X_1, \dots, 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 15: consider the tangent to the graph of g at the point $x = \mathbb{E}[X]$.
