

Some interesting problems in Probability

Márton Balázs
University of Bristol

Here are mostly probability problems which I find interesting and worth thinking about. (Some of them might appear on our problem sheets.) I'm happy to discuss about them on my office hours (unless we are too many on those). They connect to our Probability 1 material across the whole unit, come back to them later if you feel they require material not yet covered. Some of them might assume a bit more knowledge than what is given in Probability 1, we can discuss that too. *These problems are not required for any unit, they are not to be handed in.*

1. **Combinations with repetitions.** We will solve the following problem: How many ways are there to pick k objects without order out of an infinite supply of n different types of objects? Example: out of the $n = 9$ types of cakes in your favourite patisserie you wish to bring home a pack of $k = 12$ cakes. How many different choices can you make?

- First notice that the problem is equivalent with placing k indistinguishable balls (your choices) into n different urns (types of objects). The number of balls in urn i will represent the number of objects of type i selected.
- Place these n urns in a row, and think about a configuration as an ordered sequence of k identical balls, and $n - 1$ identical walls that separate the n urns. Convince yourself that every ball-urn configuration corresponds to exactly one order of these $k + n - 1$ objects.
- The number of orders of the above is now easy, a simple permutation with repetition problem: $\binom{k+n-1}{k}$. This is therefore our answer for the original problem.

To answer the cake example, we have $\binom{12+9-1}{12}$ many possible packs of cakes.

2. How many solutions does the equation

$$e_1 + e_2 + \dots + e_n = k$$

have with e_i 's non-negative integers? (Recall Problem 1.) And if we require each of them to be strictly positive?

3. A closet contains n pairs of shoes. If $2r$ shoes are randomly selected ($2r \leq n$), what is the probability that there will be (a) no complete pair, (b) exactly one complete pair, (c) exactly two complete pairs?
4. We roll a die ten times. What is the probability that each of the results $1, 2, \dots, 6$ shows up at least once? *Hint: Define the events $A_i := \{\text{number } i \text{ doesn't show up at all during the ten rolls}\}$, $i = 1 \dots 6$. Note that these events are not mutually exclusive.*
5. In a small town n TV repairmen are available, and one day k households call for TV repairmen. What is the probability that precisely i of the n repairmen will be called, $i = 1, 2, \dots, \min(n, k)$?
6.
 - **Fermi-Dirac distribution.** k balls are randomly distributed into $n \geq k$ urns in such a way that each urn contains at most one ball. (This will be the distribution after a long time if in every second a random ball is chosen, and that ball is moved to the clockwise neighboring urn, provided that that urn is empty. If the neighboring urn is occupied then nothing happens.) Find the probability that the first urn is occupied.
 - **Maxwell-Boltzmann distribution.** k distinguishable balls are randomly distributed into n urns. (This will be the distribution after a long time if in every second a random ball is chosen, and that ball is moved into the clockwise neighboring urn.) Find the probability that the first urn contains i balls, $i = 1 \dots k$.
 - **Bose-Einstein distribution.** k indistinguishable balls are randomly distributed into n urns. (This will be the distribution after a long time if in every second a random urn is chosen, and a ball, if any, from that urn is moved into the clockwise neighboring urn.) Find the probability that the first urn contains i balls, $i = 1 \dots k$.
 - Now in each of these cases consider the limit $n, k \rightarrow \infty$ such that $k/n \rightarrow \lambda$. Show that the answers for the above questions converge:
 - for Fermi-Dirac: $\mathbb{P}\{\text{the first urn contains a ball}\} \rightarrow \lambda$, this is the Bernoulli distribution;

- for Maxwell-Boltzmann: $\mathbb{P}\{\text{the first urn contains } i \text{ balls}\} \rightarrow \frac{\lambda^i}{i!} \cdot e^{-\lambda}$, this is called the Poisson distribution;
 - for Bose-Einstein: $\mathbb{P}\{\text{the first urn contains } i \text{ balls}\} \rightarrow \left(\frac{\lambda}{\lambda+1}\right)^i \cdot \frac{1}{\lambda+1}$, this is called the Geometric distribution.
7. You are playing a coin tossing game with a friend. You each must choose a (different) sequence of 3 outcomes, (e.g. *HTH*). A coin will be tossed repeatedly, until one of your sequences is observed in the (possibly long) list of outcomes. You friend has chosen *HHH*. What sequence should you choose to maximise the probability your sequence comes up first?
 8. (Ross.) The king comes from a family of 2 children. What is the probability that the other child is his sister?
 9. (Ross.) A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a “false positive” result for 1 percent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability 0.01, the test result will imply he or she has the disease.) If 0.5 percent of the population actually has the disease, what is the probability a person has the disease given that the test result is positive?
 10. (Ross.) Three prisoners are informed by their jailer that one of them has been chosen at random to be executed, and the other two are to be freed. Prisoner *A* asks the jailer to tell him privately which of his fellow prisoners will be set free, claiming that there would be no harm in divulging this information because he already knows that at least one of the two will go free. The jailer refuses to answer this question, pointing out that if *A* knew which of his fellow prisoners were to be set free, then his own probability of being executed would rise from $\frac{1}{3}$ to $\frac{1}{2}$ because he would then be one of two prisoners. What do you think of the jailer’s reasoning?
 11. We repeatedly roll two dice at the same time, and only stop when at least one of them shows a six. What is the probability that the other also shows a six?
 12. **Efron’s dice.** We are given four fair dice with sides:
 - *A*: 4, 4, 4, 4, 0, 0
 - *B*: 3, 3, 3, 3, 3, 3
 - *C*: 6, 6, 2, 2, 2, 2
 - *D*: 5, 5, 5, 1, 1, 1
 Compute the probabilities $\mathbb{P}\{A > B\}$, $\mathbb{P}\{B > C\}$, $\mathbb{P}\{C > D\}$, $\mathbb{P}\{D > A\}$ for the respective numbers shown on them. Isn’t it strange?
 13. **Envelope paradox.** You will win one of 2 envelopes which contain an unknown amount of money. One of these amounts is double the other. Suppose you pick envelope *A*, see the amount, and are offered the option to swap to *B* – should you? (Why not just pick envelope *B* in the first place?!)
 14. **Pólya urn.** Initially an urn contains 1 red and 1 blue balls. In every step we choose a random ball which we replace plus we also insert another ball of the same color. (For example, if the first choice was red, then in the second step we choose from 2 red balls and one blue ball.) Show by induction that, after the n^{th} step, the urn will contain i red balls with probability $\frac{1}{n+1}$ for each $i = 1, 2, \dots, n + 1$.
 15. Urn 1 contains n blue balls, and Urn 2 contains n red balls. At each step we randomly choose a ball from Urn 1, throw it away, and move a red ball from Urn 2 into Urn 1. After the n^{th} step, Urn 2 is empty. What is the proportion of the blue balls in Urn 1 at this time? (Assume that n is large.)
 16. (Ross.) Let $S = \{1, 2, \dots, n\}$ and suppose that *A* and *B* are, independently, equally likely to be any of the 2^n subsets (including the null set and *S* itself) of *S*.
 - a) Show without computations that $\mathbb{P}\{A \subset B\} = \left(\frac{3}{4}\right)^n$.
 - b) Show, using the above, that $\mathbb{P}\{AB = \emptyset\} = \left(\frac{3}{4}\right)^n$.
 17. What is the probability that in a class of 50 students there are no coinciding birthdays?

18. Cities A, B, C, D are located (in this order) on the four corners of a square. Between them, we have the following roads: $A \leftrightarrow B, B \leftrightarrow C, C \leftrightarrow D, D \leftrightarrow A, B \leftrightarrow D$. One night each of these roads gets blocked by the snow independently with probability $1/2$. Show that the next morning city C is accessible from city A with probability $1/2$. Do this without computations, using a dual map and probabilistic coupling: instead of snow blocks, use, in a clever way, crossroads with crossings open in one direction but closed in the traverser direction. Now imagine four new cities where these crossroads meet, in a way that the map of the new cities and the crossroads be similar to the original road map. Then think about the question in terms of cities A, C and the corresponding two new cities.
19. A hunter discovers a fox in 30 yards distance, and shoots it. If the fox survives the first shot, then it tries to escape at a velocity of 10 yards/second. The hunter loads the gun and shoots again in every 3 seconds until he either kills the fox, or it disappears in the horizon. The probability that the hunter hits the fox is proportional to the inverse square of the distance: $\mathbb{P}\{\text{The hunter hits the fox from a distance } x\} = 675x^{-2}$ ($x \geq 30$). Even when the fox is hit, it is not necessary fatal: the fox survives each hit, independently of previous hits, with probability $1/4$. Is the probability that the fox survives this unpleasant adventure positive or zero?
20. Given are an urn of infinite volume and infinitely many balls numbered $1, 2, 3, \dots$. Initially the urn is empty. One minute before midnight we place balls no. $1, 2, \dots, 10$ into the urn, shake the urn well, choose a random ball which we throw away. Half a minute before midnight we place balls no. $11, 12, \dots, 20$ into the urn, shake the urn well, choose a random ball which we throw away. $\dots 2^{-n}$ before midnight we place balls no. $10n + 1, 10n + 2, \dots, 10n + 10$ into the urn, shake the urn well, choose a random ball which we throw away. We keep on this until midnight. Prove that the urn is empty at midnight with probability one.
21. A hundred passengers are in queue for boarding a hundred-seat plane. The first passenger lost his boarding pass, so he sits on a randomly chosen seat. Other passengers board one by one, and sit on their seats if they can, and sit to randomly chosen empty seats if their own is already occupied. What is the probability that the last passenger will occupy the seat that was assigned to him?
22. Fix $s \in (1, \infty)$ and define the Riemann Zeta Function

$$\zeta(s) := \sum_{n=1}^{\infty} n^{-s}.$$

Define the probability mass function

$$p_s(n) = \frac{n^{-s}}{\zeta(s)}, \quad n = 1, 2, 3, \dots$$

and let the random variable X have this distribution. For each prime number r , define the event

$$E_r := \{X \text{ is divisible by } r\}.$$

- (a) Prove that the events $(E_r)_{r:r \text{ is prime}}$ are mutually independent.
 (b) Give a probabilistic proof of the famous Euler product formula:

$$\frac{1}{\zeta(s)} = \prod_{r:r \text{ is prime}} \left(1 - \frac{1}{r^s}\right).$$

23. As a result of a big survey it turned out that there are, on average, 2.4 children in a family, however, children have, on average, 1.55 siblings. What is the standard deviation of the number of children in a family?
24. An $\mathbb{R} \mapsto \mathbb{R}$ function is *convex*, if for every $y \in \text{dom } f$ there is a real v such that for every $x \in \text{dom } f$

$$f(x) \geq f(y) + v \cdot (x - y).$$

Suppose that f is a convex function, X is a random variable with values in $\text{dom } f$, and both X and $f(X)$ have expectations. Show *Jensen's inequality*:

$$\mathbb{E}(f(X)) \geq f(\mathbb{E} X).$$

25. Let x_1, x_2, \dots, x_n be nonnegative reals, and X a uniform choice of these numbers: $\mathbb{P}\{X = x_i\} = 1/n$, $i = 1 \dots n$. Apply Jensen's inequality for X and

- the r^{th} power function, $r > 1$;
- the reciprocal function: connection of the arithmetic and the harmonic mean;
- the logarithm: connection of the arithmetic and the geometric mean (how does Jensen's inequality work for *concave* functions?).
- The value at time t of a deposit with continuously compounded interest is

$$a(t) = a(0) \cdot e^{\int_0^t r(s) ds},$$

where $r(s)$ is the interest rate at time s . Alex always plans in advance, so he deposits an amount in the bank for exactly 10 years. Bob on the other hand only knows that he will withdraw his money from the bank at a random time T which has mean $\mathbb{E}(T) = 10$ years. Which of the two fellows can expect more money at the time of withdrawal, if

- $r(s)$ is constant 0.1, or
- $r(s)$ is of the form $0.5/(s + 1)$?

26. (Ross.) Suppose we want to generate the outcome of the flip of a fair coin, but that all we have at our disposal is a biased coin which lands on heads with some unknown probability p that need not to be equal to $\frac{1}{2}$. Consider the following procedure for accomplishing our task:

- Flip the coin.
 - Flip the coin again.
 - If both flips land on heads or both land on tails, return to step 1.
 - Let the result of the last flip be the result of the experiment.
- Show that the result is equally likely to be either heads or tails.
 - Could we use a simpler procedure that continues to flip the coin until the last two flips are different and then lets the result be the outcome of the final flip?

27. (Ross.)

- Let X be a random variable having finite expectation μ and variance σ^2 , and let $g(\cdot)$ be a twice differentiable function. Show that

$$\mathbb{E}[g(X)] \simeq g(\mu) + \frac{g''(\mu)}{2} \sigma^2.$$

Hint: Expand $g(\cdot)$ in a Taylor series about μ . Use the first three terms and ignore the remainder.

- Let X be a Poisson random variable with mean λ . Show that if λ is not too small, then $\text{Var}(\sqrt{X}) \simeq 0.25$ (!).

28. The rules of ditch-running competition are the following: the track is straight and contains infinitely many identical ditches. Two contestant of equal strength compete and run shoulder to shoulder. Each of the contestants can spring over each ditch independently with probability 2^{-L} , where L is the length of the ditches. If one of them falls in a ditch then the competition ends.

- Show that probability that the game ends in a draw is smaller than ε if and only if

$$L < \log_2 \left(\frac{1 + \varepsilon}{1 - \varepsilon} \right).$$

- Whenever a game ends in a draw they start a new game until a winner is found. For which value of L will the expected number of jumps of contestants be minimal?

29. Find expectation of $(1 + X)^{-1}$ when

- $X \sim \text{Binom}(n, p)$,
- $X \sim \text{Poi}(\lambda)$.

30. After a foot-race, 300 of the participants found one tick and 75 found two ticks in their bodies. Estimate the total number of participants.

31. The average density of a forest is 16 trees on every 100 square yards. The tree trunks can be considered as cylinders of a diameter of 0.2 yards. We are standing inside the forest, 120 yards from its edge. If we shoot a gun bullet out of the forest without aiming, what is the probability that it will hit a tree trunk? (Ignore the marginal fact that centers of the tree trunks cannot be closer than 0.2 yards to each other.)
32. In a country families bear children until the first boy is born. What is the distribution of sexes in this country? Are the sexes of children within a family independent?
33. **Surveys.** In a population of size N , m people smoke. A survey is made by randomly selecting and asking n different people whether they smoke or not. Let X be the smokers among those asked. Show that the distribution of X converges to $\text{Binom}(n, p)$ if $N \rightarrow \infty$, $m \rightarrow \infty$ such that $m/N \rightarrow p$ while n is kept fixed.
34. **Geometric is memoryless.** Show that, if $X \sim \text{Geom}(p)$, then

$$\mathbb{P}\{X \geq n+k \mid X > n\} = \mathbb{P}\{X \geq k\},$$

and the Geometric is the only positive integer variable with this property. The above display interprets as the *memoryless property*: given that more than n trials are needed for the first success, the probability that an additional k or more trials are needed is the same as if we just started the experiments.

35. (Ross.) Let X be a continuous random variable having cumulative distribution function F . Define the random variable Y by $Y = F(X)$. Show that Y is uniformly distributed over $(0, 1)$.
36. (Ross.) We say that X is *stochastically larger* than Y , written $X \geq_{\text{st}} Y$, if for all t ,

$$\mathbb{P}\{X > t\} \geq \mathbb{P}\{Y > t\}.$$

Show that if $X \geq_{\text{st}} Y$, then $\mathbb{E}[X] \geq \mathbb{E}[Y]$ when

- X and Y are nonnegative random variables;
- X and Y are arbitrary random variables.

Hint: Write X as $X = X^+ - X^-$ where

$$X^+ = \begin{cases} X, & \text{if } X \geq 0 \\ 0, & \text{if } X < 0 \end{cases}, \quad X^- = \begin{cases} 0, & \text{if } X \geq 0 \\ -X, & \text{if } X < 0 \end{cases}.$$

Similarly, represent Y as $Y^+ - Y^-$. Then make use of part a).

37. (Ross.) Show that $X \geq_{\text{st}} Y$ if and only if $\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)]$ for all increasing functions f .
Hint: If $X \geq_{\text{st}} Y$, show that $\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)]$ by showing that $f(X) \geq_{\text{st}} f(Y)$ and then using the previous exercise. To show the reverse, define an appropriate increasing function f .
38. (Ross.) A bus travels between the two cities A and B , which are 100 miles apart. If the bus has a breakdown, the distance from the breakdown to city A has a uniform distribution over $(0, 100)$. There is a bus service station in city A , in B , and in the center of the route between A and B . It is suggested that it would be more efficient to have the three stations located 25, 50 and 75 miles, respectively, from A . Do you agree? Why? What is the optimal configuration and in what sense?
39. Let F be a continuous distribution function with $F(0) = 0$. Show that

$$G(x) := \begin{cases} 0, & \text{if } x \leq 1, \\ F(x) - F(x^{-1}), & \text{if } x > 1 \end{cases}$$

is also a distribution function. Give a probabilistic interpretation to this formula.

40. Is there a continuous $[0, 1] \rightarrow [0, \infty)$ function for which $\int_0^1 g(x) dx = 1$, $\int_0^1 x \cdot g(x) dx = a$, $\int_0^1 x^2 \cdot g(x) dx = a^2$?
41. Let X be normally distributed with mean zero, variance σ^2 . Prove that for any $x > 0$ we have

$$\frac{1}{\sqrt{2\pi}} e^{-(x^2/2\sigma^2)} \left(\frac{\sigma}{x} - \frac{\sigma^3}{x^3} \right) < \mathbb{P}\{X > x\} < \frac{1}{\sqrt{2\pi}} e^{-(x^2/2\sigma^2)} \frac{\sigma}{x}.$$

Hint: differentiate all three terms of the inequality and compare the derivatives.

42. There are two identical insurance companies, both having ten thousand customers. At the beginning of the year every customer pays \$500 to his/her insurance company. During the year each customer, independently, establishes a claim to \$1500 for damages. Both companies also have \$50 000 in reserve from the previous year. A company goes bankrupt, if it cannot settle the claims. Does it make sense for the two companies to join? Let p_1 be the probability that at least one of the two companies goes bankrupt, and let p_2 be the probability that the joint companies would go bankrupt. Compute p_1 and p_2 (approximately) and draw the conclusion.
43. A random variable is said to have a *symmetric distribution*, if there is a $\mu \in \mathbb{R}$ for which $X - \mu$ and $\mu - X$ have the same distribution. Denote by F the distribution function of X .
- Compute $F(\mu)$. (*Careful, tricky question.*)
 - Compute $\int_{\mu-a}^{\mu+a} F(x) dx$.
44. For which $c > 0$ value will $F(x) = \frac{e^x}{e^x+c}$ be the distribution function of a symmetric random variable?
45. Let X be a random variable such that $\mathbb{P}\{X = 0\} = 0$, and $Y := X^{-1}$. Under what conditions will X and Y have the same distribution?
46. Prove that if ξ is a Cauchy random variable with density $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$, then $X := 1/\xi$, $Y := 2\xi/(1-\xi^2)$, and $Z := (3\xi - \xi^3)/(1-3\xi^2)$ are also Cauchy distributed. *Hint: use the trigonometric identity: if $\xi = \tan(\alpha)$, then $1/\xi = \tan(\frac{\pi}{2} - \alpha)$, $2\xi/(1-\xi^2) = \tan(2\alpha)$, and $(3\xi - \xi^3)/(1-3\xi^2) = \tan(3\alpha)$.*
47. Let X be a Cauchy random variable with density $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$. We know that $\mathbb{E}|X| = \infty$, but $\mathbb{E}(|X|^{1-\varepsilon}) < \infty$ for any $\varepsilon > 0$. Determine the limit $\lim_{\varepsilon \searrow 0} \varepsilon \mathbb{E}(|X|^{1-\varepsilon})$. *HINT: Split the integral at a well chosen point.*
48. Let X be uniformly distributed on the interval $[-3, 4]$, and $\Psi(x) = |x - 1| + |x + 1|$. Determine the distribution function $G(y)$ of the random variable $Y = \Psi(X)$. Is Y a continuous random variable? Give the decomposition of G to the sum of a discrete, an absolutely continuous, and a continuous but singular, nondecreasing function.
49. Benford's distribution describes the fact that many real-world quantity (e.g., physical constants, salaries, stocks) behave in a scale-invariant way within some boundaries. Let $0 < a < b$ be real numbers, then the precise definition is that X has *Benford distribution*, if it is continuous, almost surely $a \leq X < b$, and for all $a < z < b$ we have

$$\mathbb{P}\{z \leq X < \alpha z\} \text{ does not depend on } z \text{ whenever } 1 < \alpha < b/z.$$

(As an example, consider the case when $a \leq 1$ and $b \geq 4$. Then X has the same probability of falling between 1 and 2 as falling between 2 and 4.)

- Formulate the above statement in terms of the distribution function of X .
- Differentiate the equation of part a) with respect to z , and draw the conclusion: the density of X is of the form C/x ($a \leq x < b$). Find the value of C .
- Find the distribution function of X .
- From now on we fix the value $a = 1$. Let $b = 10^n$ ($n > 1$ integer), and show that the distribution of the first digit of X does not depend on n .
- Therefore, we also fix the value of b now: let $a = 1$ and $b = 10$. Compute the probabilities

$$\mathbb{P}\{\text{the first digit of } X \text{ is } i\}, \quad i = 1, 2, \dots, 9.$$

50. (Ross.) Suppose that X and Y are independent geometric random variables with the same parameter p .
- Without any computations, what do you think is the value of $\mathbb{P}\{X = i | X + Y = n\}$? *Hint: Imagine that you continually flip a coin having probability p of coming up heads. If the second head occurs on the n^{th} flip, what is the probability mass function of the time of the first head?*
 - Verify your conjecture in part a).
51. Let X and Y be i.i.d. geometric random variables, both with parameter p . Define $U := \min\{X, Y\}$ and $V := X - Y$. Show that U and V are independent.

52. (Ross.) Let X be a normal random variable with parameters $\mu = 0$ and $\sigma^2 = 1$ and let I , independent of X , be such that $\mathbb{P}\{I = 1\} = \frac{1}{2} = \mathbb{P}\{I = 0\}$. Now define Y by

$$Y = \begin{cases} X, & \text{if } I = 1 \\ -X, & \text{if } I = 0. \end{cases}$$

In words, Y is equally likely to equal either X or $-X$.

- Are X and Y independent?
 - Are I and Y independent?
 - Show that Y is normal with mean 0 and variance 1.
 - Show that $\text{Cov}(X, Y) = 0$.
53. We break a stick at two random, uniformly and independently chosen points.
- What is the probability that, considering the three parts obtained as edges, a triangle can be assembled?
 - What is the probability that each of the three parts is shorter than a (a is a fixed number, more than the third of the length of the stick)?
54. Ants are dropped on a metre stick uniformly at random, (and in a uniformly random direction: left or right). If they meet another ant they will turn around and start travelling in the opposite direction. What is the average distance travelled by an ant before it reaches an end of the stick?
55. Let X_1, X_2, X_3 be i.i.d. $\mathcal{N}(0, 1)$ random variables. Define

$$\varrho := \sqrt{X_1^2 + X_2^2 + X_3^2}, \quad \xi_i := X_i/\varrho, \quad i = 1, 2, 3.$$

- Compute the density of ϱ .
 - Show that ϱ and the vector $\underline{\xi} = (\xi_1, \xi_2, \xi_3)$ are independent, and $\underline{\xi}$ is uniformly distributed on the surface of the unit sphere.
56. Let X and Y be i.i.d. $\text{Exp}(\lambda)$ random variables. Prove that $U := X + Y$ and $V := X/(X + Y)$ are independent.
57. The *Monte Carlo integration* is basically the following. Suppose that f is a $[0, 1] \rightarrow [0, 1]$ continuous function, and we want to approximate the integral

$$I = \int_0^1 f(x) dx.$$

We generate a uniform random point on the unit square $[0, 1]^2$ n times independently (for example by taking independent $\text{Uniform}(0, 1)$ variables U_i and V_i , $i = 1, \dots, n$), and we denote by X_n the number of such points that fall below the graph of the function f (that is, the number of indices i with $f(U_i) < V_i$). Our estimate for I is then X_n/n .

How large should we choose n for the case of the function $f(x) = x^2$, if we want 0.1 accuracy with at least 98% probability?

- We take three uniform independent points in $[0, 1]$. Find the distribution function of the middle one.
 - We take n uniform independent points in $[0, 1]$. Find the distribution function of the k^{th} one.
59. (Ross.) Let X_1, X_2, \dots be a sequence of independent and identically distributed continuous random variables. Let $N \geq 2$ be such that

$$X_1 \geq X_2 \geq \dots \geq X_{N-1} < X_N.$$

That is, N is the point at which the sequence stops decreasing. Show that $\mathbb{E}[N] = e$.

Hint: First find $\mathbb{P}\{N \geq n\}$.

60. With every step while hiking, independently, Johnny might fall on his face and hurt his knee, or he also might fall back and hurt his elbow. Every ten miles, on average, he hurts his knee three times and his elbow two times. What is the longest hike his mother can let him go if she wants no injuries with probability at least $2/3$?

61. The number of accidents in a city is Poisson(10) distributed. After each accident, with probability 0.6 the parties can come to an agreement themselves, but with probability 0.4 the police is called. Yesterday the police was called 5 times. What is the probability that 10 accidents occurred?

62. Let X_1, X_2, \dots, X_n be

- independent geometrically distributed with respective parameters p_1, p_2, \dots, p_n , or
- independent exponentially distributed with respective parameters $\lambda_1, \lambda_2, \dots, \lambda_n$.

In both cases find the distribution of $\min\{X_1, X_2, \dots, X_n\}$. For the exponential case also find the probability that the minimum is achieved by X_i ($i = 1 \dots n$).

63. What is the probability that the sum of n independent Geometric(p) random variables is odd?

64. A fair coin is flipped repeatedly. Find the expected number of flips needed to first see the sequence HHH . How about HTH ?

65. A type of sand consists of sphere-shaped grains with lognormally distributed diameters of parameter $\mu = -0.5$ and $\sigma = 0.3$. What *percentage of the weight* of the sand is in grains with diameter less than 0.5?

66. Let X and Y be independent and identically distributed, non negative random variables. Find $\mathbb{E}\left(\frac{X}{X+Y}\right)$.

67. Let X_1, \dots, X_n be i.i.d. integer valued random variables, $m \leq n$, and $S_m = \sum_{i=1}^m X_i$. Show that for any k with $\mathbb{P}\{S_n = k\} > 0$,

$$\mathbb{E}(S_m | S_n = k) = \frac{m \cdot k}{n}.$$

68. Let X and Y be random variables with finite mean on a probability space. Prove that if $\mathbb{E}(X | Y) = Y$ and $\mathbb{E}(Y | X) = X$, then $X = Y$ a.s.

69. Let X and Y be both random variables that can only take on two values: $X = x_1, x_2$ and $Y = y_1, y_2$. Show that, as opposed to the general case, they are independent if and only if they are uncorrelated.

70. **The matching problem.** n gentlemen go out for dinner, and they leave their hats in the cloakroom. After the dinner (and several glasses of wine) they pick their hats completely randomly. Denote by X the number of gen's who take their own hats.

- Find the probability mass function of X .
- What happens as $n \rightarrow \infty$?
- Find $\mathbb{E} X$ and $\text{Var} X$.

71. Let X and Y be the two coordinates of a point that is uniformly distributed within the circle of radius one, centered at $(1, 1)$. Compute $\text{Cov}(X, Y)$.

72. Solve Buffon's needle problem without integrating, by the following argument.

- Let X be the number of times the needle intersects a line. By linearity of expectations, it is clear that $\mathbb{E} X$ is proportional to ℓ/d , where ℓ is the length of the needle and d is the distance between the parallel lines on the sheet.
- Let us now consider a smooth curve. By "dividing it into infinitesimal pieces" conclude that the same proportionality holds.
- The circle of diameter d is a nice curve where the constant of proportionality is easy to find out. Conclude that for any smooth curve we have

$$\mathbb{E} X = \frac{2}{\pi} \cdot \frac{\ell}{d}.$$

In particular, for a needle of length $\ell < d$ we have $X = 0$ or 1 , hence this also agrees with the probability that the curve intersects a line.

73. A closed loop of thread, shorter than $2r\pi$, is dropped on the surface of a sphere of radius r . Show that the loop is entirely contained in the surface of a half-sphere. *Of course, this is a probability problem.*

74. We are given a set $\underline{v}^1, \underline{v}^2, \dots, \underline{v}^n$ of unit vectors in \mathbb{R}^d . Show that there exist $\varepsilon_i = \pm 1$ numbers ($i = 1 \dots n$) such that

$$\left| \sum_{i=1}^n \varepsilon_i \cdot \underline{v}^i \right| \leq \sqrt{n}.$$

Of course, this is also a probability problem.

75. Let $A_1 \dots A_n$ be events. Their indicator variables will be denoted by $\mathbf{1}\{A_i\}$ ($i = 1 \dots n$) for this problem. Which event does

$$\prod_{i=1}^n (1 - \mathbf{1}\{A_i\})$$

indicate? On one hand, taking expectation will give the probability of this event. On the other hand, expand the above product according to a 1, singletons of $\mathbf{1}\{A_i\}$'s, products of pairs $\mathbf{1}\{A_i\}\mathbf{1}\{A_j\}$, products of triplets $\mathbf{1}\{A_i\}\mathbf{1}\{A_j\}\mathbf{1}\{A_k\}$, etc. Apply an expectation on each of these terms and conclude the inclusion-exclusion formula.

76. Twelve people enter the elevator on the ground floor of a ten-storey building. Compute the expectation and variance of the number of times the elevator stops.
77. (Ross.) Let X_1, X_2, \dots, X_n be independent and identically distributed positive random variables. Find, for $k \leq n$,

$$\mathbb{E} \left[\frac{\sum_{i=1}^k X_i}{\sum_{i=1}^n X_i} \right].$$

78. The *generating function* of a non-negative integer random variable X is given by

$$P(s) = \mathbb{E} s^X.$$

It shares many nice features with the moment generating function. In particular, the generating function of convolutions equals the product of the generating functions. Use them to find a way to paint positive integers on two 6 sided fair dice whose sum has the same distribution as the sum of two standard 6 sided dice. (The lowest number on each die should be 1). Hint: $s + s^2 + s^3 + s^4 + s^5 + s^6 = (1 + s^3)(s + s^2 + s^3) = (1 + s)(s + s^3 + s^5)$.

79. A prison warden wants to make room in his prison and is considering liberating one hundred prisoners, thereby freeing one hundred cells. He therefore assembles one hundred prisoners and asks them to play the following game: he lines up one hundred urns in a row, each containing the name of one prisoner, where every prisoner's name occurs exactly once. The game is played as follows: every prisoner is allowed to look inside fifty urns. If he or she does not find his or her name in one of the fifty urns, all prisoners will immediately be executed, otherwise the game continues. The prisoners have a few moments to decide on a strategy, knowing that once the game has begun, they will not be able to communicate with each other, mark the urns in any way or move the urns or the names inside them. Choosing urns at random, their chances of survival are almost zero, but there is a strategy giving them a 30% chance of survival, assuming that the names are assigned to urns randomly – what is it?

Remark: First of all, the survival probability using random choices is $1/2^{100}$ so this is definitely not a practical strategy. 30% seems ridiculously out of reach – but yes, you read the problem correctly.

80. **Law of Large Numbers for renewal processes.** Let τ_1, τ_2, \dots be nonnegative i.i.d. random variables (*waiting times between renewals*), and suppose that $\mathbb{E} \tau_i = \mu < \infty$. Let $T(n) := \sum_{i=1}^n \tau_i$ (*the time of the n^{th} renewal*), and $N(t) = \max\{n : T(n) \leq t\}$ (*the number of renewals up to time t*). The Weak Law of Large Numbers states that $T(n)/n \rightarrow \mu$ in probability. Prove the following dual statement: for any $\delta > 0$,

$$\lim_{t \rightarrow \infty} \mathbb{P} \left\{ \left| \frac{N(t)}{t} - \frac{1}{\mu} \right| > \delta \right\} = 0.$$

81. a) Let X_1, X_2, \dots, X_n be i.i.d. random variables, $\mathbb{E} X_i = 0$. Let also $S_n = \sum_{i=1}^n X_i$. Express the third and fourth moments of S_n in terms of those of the X_i 's.
- b) Let $f : [0, 1] \rightarrow \mathbb{R}$ be a four times continuously differentiable function. Compute

$$\lim_{n \rightarrow \infty} n \int_0^1 \int_0^1 \dots \int_0^1 \left\{ f \left(\frac{x_1 + x_2 + \dots + x_n}{n} \right) - f \left(\frac{1}{2} \right) \right\} dx_1 dx_2 \dots dx_n.$$

82. Prove that

$$\lim_{n \rightarrow \infty} \int_0^1 \int_0^1 \cdots \int_0^1 \frac{x_1^2 + x_2^2 + \cdots + x_n^2}{x_1 + x_2 + \cdots + x_n} dx_1 dx_2 \cdots dx_n = \frac{2}{3}.$$

83. Let f be a $[0, 1] \rightarrow \mathbb{R}$ continuous function. Prove that

$$\lim_{n \rightarrow \infty} \int_0^1 \int_0^1 \cdots \int_0^1 f\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) dx_1 dx_2 \cdots dx_n = f\left(\frac{1}{2}\right),$$

$$\lim_{n \rightarrow \infty} \int_0^1 \int_0^1 \cdots \int_0^1 f((x_1 x_2 \cdots x_n)^{1/n}) dx_1 dx_2 \cdots dx_n = f\left(\frac{1}{e}\right).$$

84. **Conjecturing the true order of magnitude of normal fluctuations.** Let X_1, X_2, \dots be i.i.d. random variables with mean $\mathbb{E} X_i = 0$ and variance $\text{Var} X_i = \sigma^2 < \infty$. Let also $S_n = \sum_{i=1}^n X_i$. The WLLN states that for any fixed $\delta > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{\frac{|S_n|}{n} > \delta\right\} = 0.$$

Prove the following, much stronger, statement: for any sequence b_n that converges to ∞ we have, for any $\delta > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{\frac{|S_n|}{b_n \cdot \sqrt{n}} > \delta\right\} = 0.$$

85. Construct a coupling of the pair (X_n, Y_n) such that

- each X_n and Y_n are simple symmetric random walks on \mathbb{Z}^2 ,
- their distance goes a.s. to infinity.

Not probability, but still nice:

86. Show explicitly an $\mathbb{R} \rightarrow \mathbb{R}$ function that is

- (a) 0, for $x \leq 0$,
- (b) 1, for $x \geq 1$,
- (c) smooth on the whole of \mathbb{R} (each derivative exists).

87. A disk of radius 1 can be covered by $2n$ long slabs of width $\frac{1}{n}$ each. Can it also be covered by less than $2n$ of these?

88. The “ordinary” way of hanging a picture on the wall with a loop of cord and two nails is such that removing any one of the two nails the picture will not fall as the other nail will still hold it. Find a way to hang the picture such that removing *any* of the two nails makes the picture fall down from the wall.