MATH CIRCLE: NOTIONS OF PROBABILITY

LAMC OLYMPIAD GROUP

Imagine a random process, such as flipping a coin or rolling a die. Such a process has a *total set of outcomes* Ω; for example, when flipping a coin we have $\Omega = \{\text{heads}, \text{tails}\}$, and when rolling a die we have $\Omega = \{1, 2, 3, 4, 5, 6\}$. When predicting weather, we could have $\Omega = \{\text{warm and rain}, \text{ warm and no rain}, \text{ cold and rain}, \text{ cold and no rain}\}$, etc.

An *event* is a set of possible outcomes, i.e. a subset of Ω. Not all subsets need to be events, but given events A and B we can form other events

$$A \text{ and } B = A \cap B, \qquad A \text{ or } B = A \cup B, \qquad \text{not } A = X \setminus A.$$

For example, when rolling a die, we have the events

$$\text{"Die lands on an even value"} = \text{"Die lands on 2, 4, or 6"} = \{2, 4, 6\},$$
$$\text{"Die lands on an odd value" and "Die lands on a value below 4"} = \{1, 3, 5\} \cap \{1, 2, 3\} = \{1, 3\}.$$

We say that an event A "happens" when the outcome of the random process is in A. Events A and B are called *disjoint* if they can never happen simultaneously, i.e. if $A \cap B = \varnothing$.

Often, one can assign a probability $0 \leq \mathbf{Pr}(A) \leq 1$ to each event $A \subset X$ in such a way that:

- $\mathbf{Pr}(\varnothing) = 0$ and $\mathbf{Pr}(\Omega) = 1$;
- If $A \subset B$ (so if A happens then B happens), then $\mathbf{Pr}(A) \leq \mathbf{Pr}(B)$.
- If A and B are disjoint, then $\mathbf{Pr}(A \cup B) = \mathbf{Pr}(A) + \mathbf{Pr}(B)$.

Then $(\Omega, \mathbf{Pr})$ is called a probability space. A common setup is when the outcomes are *uniformly random*, i.e. all outcomes are equally likely. On a finite space, this would mean that

$$\mathbf{Pr}(A) = \frac{\text{Number of outcomes in } A}{\text{Total number of outcomes}} = \frac{|A|}{|\Omega|}.$$

So for a fair die, $\mathbf{Pr}(\{1\}) = \ldots = \mathbf{Pr}(\{6\}) = 1/6$ and $\mathbf{Pr}(\text{"even value"}) = \mathbf{Pr}(\{2, 4, 6\}) = 3/6 = 1/2$. In the plane, we would say that we choose a point P *uniformly at random* from a set Ω of finite area if for any subset $A \subset \Omega$ (whose area is defined),

$$\mathbf{Pr}(\text{"}P \text{ lies in } A\text{"}) = \mathbf{Pr}(A) = \frac{\text{Area of } A}{\text{Total area of } \Omega},$$

(similarly for lengths, volumes, etc.) In practice, given an event $B \subset \Omega$ with nonzero probability, we may have some way of finding out that B certainly happens. Then we can restrict everything to a smaller probability space where B is known to happen, by defining the *conditional probabilities*

$$\mathbf{Pr}(\text{"}A \text{ given } B\text{"}) = \mathbf{Pr}(A \mid B) := \frac{\mathbf{Pr}(A \cap B)}{\mathbf{Pr}(B)} \qquad (\text{so } \mathbf{Pr}(\varnothing \mid B) = 0,\ \mathbf{Pr}(\Omega \mid B) = 1).$$

Thus for a fair die, $\mathbf{Pr}(\text{"die lands on 2 or 3, given that it lands even"}) = \mathbf{Pr}(\{2, 3\} \mid \{2, 4, 6\}) = \mathbf{Pr}(\{2\}) \ / \ \mathbf{Pr}(\{2, 4, 6\}) = (1/6) \ / \ (3/6) = 1/3$. We say that two events A and B are *independent* if whether one happens doesn't affect the likelihood of the other, i.e.

$$\mathbf{Pr}(A \mid B) = \mathbf{Pr}(A) \iff \mathbf{Pr}(A \cap B) = \mathbf{Pr}(A) \cdot \mathbf{Pr}(B) \iff \mathbf{Pr}(B \mid A) = \mathbf{Pr}(B).$$

We analogously define that n events $A_1, \ldots, A_n$ are *(mutually) independent* when

$$\mathbf{Pr}(A_1 \cap \ldots \cap A_n) = \mathbf{Pr}(A_1) \cdots \mathbf{Pr}(A_n).$$

The *product* of two finite probability spaces $(\Omega_1, \mathbf{Pr}_1)$ and $(\Omega_2, \mathbf{Pr}_2)$ is $(\Omega_1 \times \Omega_2, \mathbf{Pr})$, where $\Omega_1 \times \Omega_2$ is the Cartesian product and

$$\forall a \in \Omega_1, b \in \Omega_2, \qquad \mathbf{Pr}(\{(a,b)\}) := \mathbf{Pr}_1(\{a\}) \cdot \mathbf{Pr}_2(\{b\}).$$

From here we can deduce that for events $A \subset \Omega_1$ and $B \subset \Omega_2$, $\mathbf{Pr}(A \times B) = \mathbf{Pr}_1(A) \cdot \mathbf{Pr}_2(B)$. So we can identify an event $A \subset \Omega_1$ with the event $A \times \Omega_2 \subset \Omega_1 \times \Omega_2$, which has probability $\mathbf{Pr}_1(A) \cdot 1 = \mathbf{Pr}_1(A)$.

Problem 1. We pick a 5-digit positive integer uniformly at random. What's the probability that all 5 digits are distinct?

Problem 2. In a box there are 20 red balls, 30 yellow balls and 50 green balls. We shake the box to shuffle the balls, and pick 3 balls without looking.

(a) What's the probability that all 3 balls have the same color?

(b) Given that the second ball we pick is red, what's the probability that the third one is green?

Problem 3. We're given a fair die, that takes values from 1 to 6 uniformly at random.

(a) If we throw the die we'll get a value x. Consider the events A : "x is even", B : "x is a multiple of 3", C : "$x = 2$". Which of them are independent? What's $\mathbf{Pr}(C \mid A)$?

(b) We'll throw the die twice and look at the total score (between 2 and 12). What's the most likely value we'll obtain, and what's its probability? (Assume the two throws are independent.)

Problem 4. (*Bayes' rule*). Given events A and B (not necessarily independent) with $\mathbf{Pr}(B) \neq 0$, show that

$$\mathbf{Pr}(A \mid B) = \mathbf{Pr}(B \mid A) \cdot \frac{\mathbf{Pr}(A)}{\mathbf{Pr}(B)}.$$

For example: say that when it's cold there's a 30% probability of raining, but in general there's a 10% chance of raining and a 40% chance of it being cold. What's the probability of it being cold *given* that it's raining? (Here, % just means dividing by 100, e.g. 100% = 1.)

Problem 5. There are 30 students in a class. What's the probability that 2 of them have the same birthday? (Assume the birthdays of different students are independent, and each of the 365 days of a year is equally likely.)

Problem 6. We are given a regular n-gon (where $n \geq 3$), and we choose three *distinct* vertices of it, uniformly at random (that is, every set of 3 vertices is equally likely). What's the probability that the triangle formed by these vertices is equilateral?

Problem 7. Consider an equilateral triangle cut into 9 equal equilateral triangles, by lines parallel to the sides; overall these give us 10 vertices. We pick a point P uniformly at random inside the triangle; what's the probability that out of the 10 vertices, P is closest to the center G of the triangle?

Problem 8. We have a biased coin, that lands on heads with probability $0 < p < 1$, and tails with probability $1 - p$. What's the probability that after $n \geq 1$ throws of the coin, we got heads at least once? (Assume different throws are independent.)

Problem 9. We pick a number n uniformly at random from $\{1, 2, \ldots, 3000\}$. Show that the probability that n is prime is less than 1/3. Can you improve this?

Problem 10. We throw a fair coin 2020 times. What's the probability that we get heads exactly 1010 times?

Problem 11. We pick a nonnegative integer n strictly less than 1 000 000, uniformly at random. What's the probability that n is divisible by 3 *and* doesn't contain the digit 3?

Problem 12. For events A and B (not necessarily disjoint or independent), show that

$$\mathbf{Pr}(A \cup B) + \mathbf{Pr}(A \cap B) = \mathbf{Pr}(A) + \mathbf{Pr}(B).$$

In particular, $\mathbf{Pr}(A \cup B) \leq \mathbf{Pr}(A) + \mathbf{Pr}(B)$.

Now that you're more familiar with basic concepts, we can introduce random variables. Fix a probability space $(\Omega, \mathbf{Pr})$, where Ω is the set of outcomes.

A *random variable* X with values in a set S is just a function from Ω to S (we write $X : \Omega \to S$). In other words, X assigns a value to each outcome of our random experiment. We can then consider *events* of the form "$X = x$" (where $x \in S$), with probability

$$\mathbf{Pr}(X = x) := \mathbf{Pr}(\{\alpha \in \Omega : X(\alpha) = x\}).$$

Suppose we throw a fair die. Then one random variable could be the parity of the value that we get, and then

$$\begin{aligned}\mathbf{Pr}(\text{Parity} = \text{even}) &= \mathbf{Pr}(\{\alpha \in \{1,2,3,4,5,6\} : \text{Parity}(\alpha) = \text{even}\}) \\ &= \mathbf{Pr}(\{2,4,6\}) \\ &= 1/2.\end{aligned}$$

On the other hand, we can also take X to be the value of the die itself, which would just be the identity function $X(\alpha) = \alpha$ on $\{1,2,3,4,5,6\}$. Then $\mathbf{Pr}(X = 1) = \mathbf{Pr}(\{1\}) = 1/6$. We could also take X to be a *constant* random variable, for example $X(\alpha) = 1$ for all α. Then $\mathbf{Pr}(X = 1) = 1$ and $\mathbf{Pr}(X = 2) = \mathbf{Pr}(X = 3) = \mathbf{Pr}(X = 4) = \mathbf{Pr}(X = 5) = \mathbf{Pr}(X = 6) = 0$.

More generally, we can consider the event "$X \in A$" (where A is a subset of S), which has probability

$$\mathbf{Pr}(X \in A) := \mathbf{Pr}(\{\alpha \in \Omega : X(\alpha) \in A\}).$$

So if X is the value of a fair die, then

$$\begin{aligned}\mathbf{Pr}(X \geq 3) &= \mathbf{Pr}(\{\alpha \in \{1,2,3,4,5,6\} : \alpha \geq 3\}) \\ &= \mathbf{Pr}(\{3,4,5,6\}) \\ &= 2/3.\end{aligned}$$

Now given a random variable $X : \Omega \to S$ and a function $f : S \to T$, we can also consider the random variable $f(X)$, which is just the function $\alpha \mapsto f(X(\alpha))$. So for example if X is the value of a fair die, then $\text{Parity}(X)$, $2X$, X^2 and $6 - X$ are also random variables. We can similarly take functions of two random variables, and so on; for instance, given random variables X and Y, (X, Y) and $X + Y$ are also random variables, with

$$\mathbf{Pr}(X + Y = s) = \sum_{x+y=s} \mathbf{Pr}(X = x \text{ and } Y = y).$$

Recall that two events $A, B \subset \Omega$ are independent when $\mathbf{Pr}(A \text{ and } B) = \mathbf{Pr}(A) \cdot \mathbf{Pr}(B)$. We say that two random variables $X : \Omega \to S$, $Y : \Omega \to T$ are *independent* if for all $A \subset S$ and all $B \subset T$,

$$\mathbf{Pr}(X \in A \text{ and } Y \in B) = \mathbf{Pr}(X \in A) \cdot \mathbf{Pr}(Y \in B),$$

where the left-hand-side just means $\mathbf{Pr}((X, Y) \in A \times B)$. *Mutual independence* of $n \geq 1$ random variables is defined similarly by $\mathbf{Pr}(X_1 \in A_1, \cdots, X_n \in A_n) = \mathbf{Pr}(X_1 \in A_1) \cdots \mathbf{Pr}(X_n \in A_n)$.

A *real* random variable is one that takes real values. Given a (discrete) real random variable X, its *expected value* or *mean* is

$$\mathbf{E}[X] = \mu_X = \sum_{x \in \mathbb{R}} x \mathbf{Pr}(X = x),$$

and its *variance* (or *squared standard deviation*) is

$$\begin{aligned}\mathrm{var}(X) = \sigma_X^2 &= \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] \\ &= \sum_{x\in\mathbb{R}} (x - \mathbf{E}[X])^2 \, \mathbf{Pr}(X = x).\end{aligned}$$

So $\mathbf{E}[X]$ is in some sense our best guess for X (e.g. the expected value of a fair die is 3.5), and $\mathrm{var}(X)$ is a measure of how far X is from $\mathbf{E}[X]$. Note that $(X - \mathbf{E}[X])^2 \geq 0$, so we always have $\mathrm{var}(X) \geq 0$. We use squared distance instead of absolute value since it is more convenient.

More generally, given two random variables X and Y, their *covariance* is

$$\mathrm{cov}(X, Y) := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])], \qquad \text{thus} \qquad \mathrm{var}(X) = \mathrm{cov}(X, X).$$

This is a measure of how correlated X and Y are, as we will see later. As a final remark, expected values and covariances can be similarly defined for *continuous* random variables, but that requires some calculus (sums become integrals, etc.).

Problem 13. On a game show, you would roll a die a large number of times. Whenever you roll a 1, you lose \$10 000, but for every other value you win \$1 500. Should you play?

Problem 14. Assume that we work in a finite probability space $(\Omega, \mathbf{Pr})$ such that each outcome has positive probability. For a (discrete) random variable $X : \Omega \to \mathbb{R}$, show that if $\mathrm{var}(X) = 0$ then X is constant.

Problem 15. If X and Y are independent random variables, and $f(X), g(Y)$ are functions of them, show that $f(X)$ and $g(Y)$ are also independent.

Problem 16. (*Linearity of expectation*) Let X and Y be real random variables.

(a) Show that $\mathbf{E}[X+Y] = \mathbf{E}[X]+\mathbf{E}[Y]$. For $r \in \mathbb{R}$, show also that $\mathbf{E}[rX] = r\mathbf{E}[X]$. For instance, the expected value of the sum of two dice is 7.

(b) If X and Y are *independent*, show that $\mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$.

Problem 17. Given real random variables X and Y, we defined their covariance $\mathrm{cov}(X, Y)$ as $\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$. Show that

$$\mathrm{cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X] \cdot \mathbf{E}[Y].$$

So in particular, $\mathrm{var}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2$. Since variance is nonnegative, this also shows that $\mathbf{E}[X^2] \geq \mathbf{E}[X]^2$. Problem 16(b) hence shows that $\mathrm{cov}(X, Y) = 0$ when X and Y are independent, but the converse isn't necessarily true.

Problem 18. (*Bilinearity of covariance*) Let X, X' and Y be real random variables.

(a) Show that $\mathrm{cov}(X + X', Y) = \mathrm{cov}(X, Y) + \mathrm{cov}(X', Y)$. For $r \in \mathbb{R}$, show that $\mathrm{cov}(rX, Y) = r \cdot \mathrm{cov}(X, Y)$. So in particular $\mathrm{var}(rX) = \mathrm{cov}(rX, rX) = r^2\mathrm{var}(X)$.

(b) If X and Y are independent, show that $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.

Problem 19. Let $X_1, X_2, X_3, \ldots$ be mutually independent random variables, all with mean μ and variance σ^2. For $n \geq 1$, let $S_n := (X_1 + X_2 + \ldots + X_n)/n$.

(a) What are the mean and variance of S_n? The *law of large numbers* states that under some additional assumptions, S_n approaches the constant μ as $n \to \infty$.

(b) What are the mean and variance of $\sqrt{n}(S_n - \mu)$? The *central limit theorem* states that under some additional assumptions, $\sqrt{n}(S_n - \mu)$ approaches a *normal distribution* as $n \to \infty$.

Problem 20. We are given a fair die. We throw the die multiple times until we roll a 6. What's the expected value of the *number of throws* n after which we stop?

Problem *21. (*Markov chains*) We are given a fair coin.
(a) We throw the coin until we hit tails and heads consecutively, in this order (write this as TH). Prove that the expected value of the *number of throws* n after which we stop (i.e., we hit TH for the first time, with the H on the nth throw) is 4.
(b) We throw the coin until we hit tails, heads and then tails consecutively (THT). Prove that the expected value of the *number of throws* n after which we stop is 10.

Problem *22. **(a)** We choose two points A and B independently, uniformly at random from the unit circle. What's the expected value of AB^2?
(b) We are given n fixed points $A_1, \ldots, A_n$ on the circle. Show that there is a point P on the circle such that $PA_1^2 + PA_2^2 + \ldots + PA_n^2 \geq 2n$.

Problem *23. (*Markov's inequality*) Let X be a real random variable and $t > 0$. Show that

$$\mathbf{Pr}\left(|X| > t\right) \leq \frac{1}{t}\mathbf{E}\left[|X|\right].$$

If $\mu := \mathbf{E}[X]$ is the mean and $\sigma := \sqrt{\operatorname{var}(X)}$ is the standard deviation, use this to show that

$$\mathbf{Pr}\left(|X - \mu| > t\sigma\right) \leq \frac{1}{t^2}.$$

This gives an application for considering the standard deviation: we can quantify the fact that it's unlikely for the value of X to be more than $t\sigma$ away from its mean.

Problem *24. Given real random variables X and Y, show that

$$\operatorname{cov}(X, Y)^2 \leq \operatorname{var}(X) \cdot \operatorname{var}(Y).$$

Hint: consider the random variable $Z := \operatorname{var}(Y)X - \operatorname{cov}(X, Y)Y$*, use Problem 17 to compute* $\operatorname{var}(Z)$ *and use that variance is nonnegative.*
Because of this inequality, it makes sense to define the *correlation* between X and Y as

$$\operatorname{corr}(X, Y) := \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}} = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y},$$

which is a number between -1 and 1. The correlation is 0 when X and Y have "very little in common" (in particular when they're independent since then $\operatorname{cov}(X, Y) = 0$). Here's a picture: