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 $0 \le \mathbf{Pr}(A) \le 1$ 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 $\mathbf{Pr}(A) \le \mathbf{Pr}(B)$.
- If A and B are disjoint, then $\mathbf{Pr}(A \cup B) = \mathbf{Pr}(A) + \mathbf{Pr}(B)$.

Then $(\Omega, \text{Events}, \mathbf{Pr})$ 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.

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

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?