Mini-course notes: Introduction to probabilistic combinatorics [0010]

July 29, 2026 · *Alexandros Singh*

In this course we will be working with random combinatorial objects, mainly random graphs. Our goal will be to get a glimpse into their structure studying various "combinatorial statistics". To put it informally, we imagine drawing random objects from a sack and recording for each one some quantity of interest.

To formalise these ideas, we introduce the notions of *discrete probability spaces* and *discrete random variables*. The serve as a model of the "sacks" from which we draw objects and of the "quantities of interest" which we record.

Definition 1. Discrete probability space
A discrete probability space is a tuple $(\Omega, \mathbb{P})$ of

- A finite or countably infinite set Ω.
- A function $\mathbb{P} : 2^{\Omega} \to [0, 1]$ such that:
  - $\mathbb{P}(\Omega) = 1$,
  - $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$.

The set Ω is called the *sample space*. The function $\mathbb{P}$ is called a *probability function* (or *measure*). Elements of 2^{Ω}, i.e. subsets of Ω, are called *events.* We'll say that two events $A, B \subseteq \Omega$ are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Notice how knowing $\mathbb{P}(\{\omega\})$ for all singletons $\omega \in \Omega$ suffices to reconstruct $\mathbb{P}(A)$ for any $A \in 2^{\Omega}$. You'll often see the term *probability mass function* used to refer to $\mathbb{P}(\{\omega\})$. If you are familiar with probability theory, then you can think of this as a density function with respect to the *counting measure* over Ω.

As an aside, we note that we took our space to be 2^{Ω}. This is what people often tacitly choose, except for when they want to model situations of "limited information". Indeed, if you're familiar with probability theory you know that more generally we consider triples $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F} \subseteq 2^{\Omega}$ is a *σ-algebra*.

The following inequality is basic but quite powerful.

Theorem 2. Union bound
Let $(\Omega, \mathbb{P})$ be a discrete probability space. If $A_i, i = 1 \dots n$ are n events, then:

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$

A nice application of the union bound is given in the following exercise, based on a result by Erdős published in 1947.

Exercise 3. A bound for the Ramsey numbers $R(k, k)$.
The *diagonal Ramsey number* $R(k, k)$ is the smallest positive integer n such that every red-blue edge-coloring of the complete graph K_n contains a monochromatic clique K_k. For example, $R(3, 3) = 6$.

Show that if $\binom{n}{k} 2^{1-\binom{n}{2}} < 1$, then $R(k, k) > n$. Deduce that $R(k, k) > 2^{k/2}$ for all $k \geq 3$.

Hint: Color each edge red or blue with probability $1/2$, independently for each edge. Let E be the event that a monochromatic K_k exists. Let $S_1, \ldots, S_p$ be an arbitrary exhaustive listing of all possible subsets of k vertices in K_n. Then $\mathbb{P}(E) = \mathbb{P}(M_1 \cup \cdots \cup M_p)$, where M_p is the event that S_p is monochromatic. What is the probability that a subset S of k edges is monochromatic? How many such subsets exist (i.e. what's p)? Use the above together with the union bound to derive the result.

If all went well, you've proven that a large graph exists which avoids monochromatic cliques of size k. Yet you have not explicitly constructed it! This is a hallmark of the *probabilistic method.* Indeed, no known deterministic construction exists that gets anywhere close to this bound!

Definition 4. Discrete random variable, distribution
Given a discrete probability space $(\Omega, \mathbb{P})$, a *discrete random variable* X is a function $X : \Omega \to \mathbb{R}$. The distribution (or probability mass function) of X is $\mathbb{P}(X = x) = \mathbb{P} \circ X^{-1}, x \in \mathbb{R}$. Put otherwise:

$$\mathbb{P}(X = x) = \mathbb{P}(\{\omega | X(\omega) = x\}) = \sum_{\omega \text{ s.t } X(\omega) = x} \mathbb{P}(\{\omega\}).$$

Here's some examples of distributions which we'll encounter in our study of random graphs.

Definition 5. Bernoulli distribution
Let $p \in [0, 1]$ be given. We say that a random variable X with support $k \in \{0, 1\}$ follows the *Bernoulli distribution* with parameter p, $X \sim \text{Bernoulli}(p)$, if:

$$\mathbb{P}(X = 0) = 1 - p,$$
$$\mathbb{P}(X = 1) = p.$$

The Bernoulli distribution is often interpreted as a "biased coin toss".

Definition 6. Binomial distribution

Let $n \in \mathbb{N}, p \in [0,1]$ be given. We say that a random variable X with support $k \in \{0, \dots, n\}$ follows the *binomial distribution* with parameters n, p, $X \sim \mathrm{Bin}(n,p)$, if:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^k.$$

The binomial distribution is often interpreted as the distribution of the number of successes in a sequence of n independent Bernoulli trials with parameter p.

Definition 7. Poisson distribution
Let $\lambda \in (0, +\infty)$ be given. We say that a random variable X follows the *Poisson distribution* with rate λ, $X \sim \mathrm{Pois}(\lambda)$, if:

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

The Poisson distribution $\mathrm{Pois}(\lambda)$ appears as the $n \to +\infty$ limit of the binomial distribution $\mathrm{Bin}(n, p_n)$ in the cases where the product np_n converges to some finite limit $\lim_{n \to +\infty} np_n = \lambda$. This is the so-called *Poisson limit theorem*.

We'll say that two random variables X, Y over the same space are *independent* if $\mathbb{P}(X = x, Y = y) = \mathbb{P}(\{\omega | X(\omega) = x\} \cap \{\omega | Y(\omega) = y\}) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$.

As a side note, we aren't obliged to take $\mathbb{R}$ as the codomain of X. Indeed we'll often take $\mathbb{N}$ or $\mathbb{Z}$. Whatever the codomain S is, it is important to note that we'll always have $X(S)$ be a finite or countably infinite set.

Two following two quantities associated to a random variable are its *expected value* (or *mean*, or *expectation*) and its *variance*. They serve to capture the "average outcome" around which a the distribution of a random variable is centered as well as how "spread" out it is.

Definition 8. Expectation
Given a discrete probability space $(\Omega, \mathbb{P})$, and discrete random variable X its *expected value*, if it exists, is

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x).$$

Definition 9. Variance
Given a discrete probability space $(\Omega, \mathbb{P})$, and discrete random variable X its *variance*, if it exists, is

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2 \right) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Notice how the variance measures precisely the mean of the square difference $(X - \mathbb{E}(X))^2$, which quantifies how "spread" X is around its mean.

More generally, one can define various flavours of moments for a random variable as follows.

Definition 10. Moments (raw, central, factorial)
Let $(\Omega, \mathbb{P})$ be a probability space and X a random variable on it. Given $k \in \mathbb{N}$, the k-th *raw, central* and *factorial* moment of X are, respectively:

$$\mathbb{E}(X^k), \quad \mathbb{E}((X - \mathbb{E}(X))^k), \quad \text{and} \quad \mathbb{E}(X^{(k)}),$$

where $X^{(k)} = X(X-1)(X-2)\dots(X-k+1)$.

The factorial moments will be of particular interest to us as they render a lot of the combinatorics simpler, especially for subgraph counts in $G(n,p)$. They also take a particularly convenient form for the Poisson distribution, as we'll later see.

A particularly important property of the expectation is that it is linear: $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y), \mathbb{E}(\lambda X) = \lambda \mathbb{E}(X)$. This holds *regardless* of whether X and Y are independent!

Definition 11. Indicator variable
Let $(\Omega, \mathbb{P})$ be a probability space and $A \subseteq \Omega$ and event. The *indicator variable* of A is random variable defined by:

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{otherwise.} \end{cases}$$

Calculating the mean for an indicator variable, we find $\mathbb{E}(I_A) = \mathbb{P}(A)$. We'll often exploit this fact together with the aforementioned linearity of expectation to compute the expectation of a random variable by expressing it as a sum of *indicator variables*. The following exercise presents such an example.

Exercise 12. Mean number of fixed points in random permutations
Fix $n \geq 1$ and let Ω be the set of all permutations of $\{1, \dots, n\}$. Endow Ω with the uniform probability measure:

$$\mathbb{P}(p) = \frac{1}{|\Omega|} = \frac{1}{n!},$$

to form the probability space $(\Omega, \mathbb{P})$ of *random permutations*.

Let $X(\pi)$ be the number of fixed points of a permutation $\pi \in \Omega$. What is $\mathbb{E}(X)$?

Hint: express X as a sum of indicator variables.

To better appreciate why decomposing X into indicator variables helps, you can try to re-calculate the above expectation directly using its definition, i.e. by figuring out the distribution of X. How do they two methods compare?

You should also take a moment to appreciate how the indicator variables you used in the above exercise are very much *not* independent, yet linearity of expectation still holds, which is what allows you to solve the problem!

What else can the mean $\mathbb{E}(X)$ tell us? The following basic but powerful observation is a simple manifestation of the *first moment principle* or *method*.

Lemma 13.
Let $(\Omega, \mathbb{P})$ be a probability space and X a random variable on it. Then

- $\mathbb{E}(X) \leq t \Rightarrow \mathbb{P}(X \leq t) > 0$,
- $\mathbb{E}(X) \geq t \Rightarrow \mathbb{P}(X \geq t) > 0$.

Here's an exercise to get you used to this idea.

Exercise 14. Existence of large cuts in graphs
Let $G = (V, E)$ be a given graph. A *cut* is a partition of the vertices of G into two disjoint subsets. The *size* of the cut is the number of edges that cross it (have one endpoint in each part). Show that in any given graph there exists a cut of size at least $|E|/2$.

Hint: pick a random partition of V and let X be the number of edges that cross it. Decompose X as a sum of indicator variables and then use linearity of expectation and the above lemma.

A quantititive version of the above lemma is the following famous inequality, which we often employ in the context of the first moment method.

Lemma 15. Markov's inequality
Let $(\Omega, \mathbb{P})$ be a probability space and X a *non-negative* random variable on it. Then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Proof: Since X is never negative,

$$\mathbb{E}(X) = \sum_i i\mathbb{P}(X = i) \geq \sum_{i \geq t} i\mathbb{P}(X = i) \geq t\mathbb{P}(X \geq t).$$

As a first exercise, try to re-derive the union bound by applying Markov's inequality to carefuly constructed random variable.

Let us generalise a bit and (instead of a single random variable X) look at *sequences* of X_n, which we will assume to be non-negative as is often the case in combinatorics. We frequently want to study $\mathbb{P}(X_n > 0)$, the probability that X_n is positive.

If we want to show that $\mathbb{P}(X_n > 0) \to 0$ (as $n \to +\infty$), then we can use the first moment method: $\mathbb{E}(X) \to 0$ implies $\mathbb{P}(X_n > 0) \leq \mathbb{E}(X) \to 0$.

What if we want to show that $\mathbb{P}(X_n > 0) \to 1$ (or that, at least, its bounded away from 0)? One might think that a large mean or even $\mathbb{E}(X_n) \to +\infty$ suffices. After all, for any fixed n we can apply the aforementioned argument to $X = X_n$: $\mathbb{E}(X) > 0$ does imply that $\mathbb{P}(X > 0) > 0$ as we saw above. The following (counter-)example shows that this is, unfortunately, not the case *asymptotically*!

Example 16. [ssscc_e1]
Let X_n be a non-negative random variable satisfying

$$X_n = \begin{cases} n^2, & \text{with probability } 1/n, \\ 0, & \text{otherwise.} \end{cases}$$

Then $\mathbb{E}(X_n) \to +\infty$ but $\mathbb{P}(X_n > 0) \to 0$.

If, however, we manage to show that the variance of X_n is relatively small with respect to its mean, then we can salvage the above idea. This is the basis for the *second moment method* (or principle), which the following lemmas make concrete.

Lemma 17. Chebyshev's inequality
Let $(\Omega, \mathbb{P})$ be a probability space and X a random variable on it, having finite non-zero variance. Then for every $t > 0$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\mathrm{Var}(X)}{t}.$$

Proof: Use Markov's inequality on $(X - \mathbb{E}(X))^2$.

Specifically for the events $X = 0$ and its complement $X > 0$, we obtain as a corollary the following.

Corollary 18.
Let X be a random variable satisfying the requirements for Chebyshev's inequality to be applied. Then

$$P(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2},$$

and

$$P(X > 0) \geq 1 - \frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2}.$$

Proof: the event $X = 0$ implies $|X - \mathbb{E}(X)| = \mathbb{E}(X)$, from which the above are obtained via Chebyshev's inequality and taking the complement.

More generally, for sequences of random variables we have.

Corollary 19.
Let X_n be a sequence of random variables with $\mathbb{E}(X_n) \geq 0$ and $\mathrm{Var}(X_n) = o(\mathbb{E}(X_n)^2)$. Then

$$P(X_n = 0) \to 0.$$

A stronger version of the above is given by the Paley-Zygmund inequality, which is what we often see used in the context of the second moment method. For this mini-course, we content ourselves with the weaker formulation above.

Notice how the first moment method gives *upper bounds* while the second moment method gives *lower bounds* for $P(X_n > 0)$.

Oftentimes we want to go beyond the computation of the asymptotic mean or variance for a sequence X_n: we want to prove that X_n converges to a limiting random variable X. There are various notions of convergence, among which we often find *convergence in distribution, convergence in probability*, and *almost sure convergence*. For our course we will limit ourselves to proving convergence in probability, which is in some sense the weakest form of the above (it is implied by the others).

The typical definition of convergence in distribution involves the cumulative distribution function $\mathbb{P}(X \leq x)$. In our discrete setting however, we can formulate it directly in terms of $\mathbb{P}(X = k)$ as follows.

Definition 20.
We say that a sequence X_n of non-negative integer-valued random variables *converges in distribution*, denoted as $X_n \xrightarrow{d} X$, to a random variable X if

$$\lim_{n \to +\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$$

Of course the above definition can be adapted to work with other types of discrete random variables taking values in some some finite or countably infinite set.

Working directly with $\mathbb{P}(X_n = k)$ can often get messy. A particularly handy tool for showing convergence in distribution is the *method of (factorial) moments*: sometimes proving $\mathbb{E}(X_n^{(r)}) = \mathbb{E}(X^{(r)})$ for ever r is enough! The "sometimes" refers to a crucial condition: the distribution of X must be *uniquely determined by its moments*! In general, figuring out if this is the case for a given distribution is a *moment problem*, a classical question in probability theory.

We will however not worry about this here, since all the distributions that interest us are uniquely defined by their moments. In particular, in studying subgraph counts in $G(n,p)$ we will often encounter the Poisson distribution as our limit distribution, for which we have the following.

Proposition 21.
The Poisson distribution $\mathrm{Pois}(\lambda)$ is uniquely defined by its factorial moments, which are as follows (assuming $X \sim \mathrm{Pois}(\lambda)$):

$$\mathbb{P}(X^{(r)}) = \lambda^r.$$

Table of Contents