CS 70 Discrete Mathematics and Probability Theory
Fall 2025 Course Notes **Note 18**

Random Variables: Distribution and Expectation

Recall our setup of a probabilistic experiment as a procedure of drawing a sample from a set of possible values, and assigning a probability for each possible outcome of the experiment. For example, if we toss a fair coin n times, then there are 2^n possible outcomes, each of which is equally likely and has probability $\frac{1}{2^n}$.

Now suppose we want to make a measurement in our experiment. For example, we can ask what is the number of heads in n coin tosses; call this number X. Of course, X is not a fixed number, but it depends on the actual sequence of coin flips that we obtain. For example, if $n = 4$ and we observe the outcome $\omega = HTHH$, then $X = 3$; whereas if we observe the outcome $\omega = HTHT$, then $X = 2$. In this example of n coin tosses, we only know that X is an integer between 0 and n, but we do not know what its exact value is until we observe which outcome of n coin flips is realized and count how many heads there are. Because every possible outcome is assigned a probability, the value X also carries with it a probability for each possible value it can take. The table below lists all the possible values X can take in the example of $n = 4$ coin tosses, along with their respective probabilities.

outcomes ω	value of X (# heads)	probability of occurring
$TTTT$	0	$1/16$
$HTTT$, $THTT$, $TTHT$, $TTTH$	1	$4/16$
$HHTT$, $HTHT$, $HTTH$, $THHT$, $THTH$, $TTHH$	2	$6/16$
$HHHT$, $HHTH$, $HTHH$, $THHH$	3	$4/16$
$HHHH$	4	$1/16$

Such a value X that depends on the outcome of the probabilistic experiment is called a *random variable* (abbreviated *r.v.*). As we see from the example above, a random variable X typically does not have a definitive value, but instead only has a probability *distribution* over the set of possible values X can take, which is why it is called random. So the question “What is the number of heads in n coin tosses?” does not exactly make sense because the answer X is a random variable. But the question “What is the *typical* number of heads in n coin tosses?” makes sense — it is asking what is the average value of X (the number of heads) if we repeat the experiment of tossing n coins multiple times. This average value is called the *expectation* of X, and is one of the most useful summaries (also called *statistics*) of an experiment.

1 Random Variables

Before we formalize the above notions, let us consider another example to enforce our conceptual understanding of a random variable.

Example: Fixed Points of Permutations

Question: Suppose we collect the homeworks of n students, randomly shuffle them, and return them to the students. How many students receive their own homework?

Here the probability space consists of all $n!$ permutations of the homeworks, each with equal probability $\frac{1}{n!}$. If we label the homeworks as $1, 2, \dots, n$, then each sample point is a permutation $\pi = (\pi_1, \dots, \pi_n)$ where π_i is the homework that is returned to the ith student. We call i a *fixed point* of π if $\pi_i = i$, i.e., if student i receives their own homework.

As in the coin flipping case above, our question does not have a simple numerical answer (such as 4), because the number depends on the particular permutation we choose (i.e., on the sample point). Let us call the number of fixed points X_n, which is a random variable taking values in the set $\{0, 1, 2, \dots, n\}$. (Actually the value $X_n = n-1$ is not possible: why?)

Formal Definition of a Random Variable

We now formalize the concepts discussed above.

Definition 18.1 (Random Variable)**.** *A random variable X on a sample space Ω is a function $X\colon \Omega \to \mathbb{R}$ that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.*

Until further notice, we will restrict our attention to random variables that are discrete, i.e., they take values in a range that is finite or countably infinite. This means even though we define X to map Ω to $\mathbb{R}$, the actual set of values $\{X(\omega)\colon \omega \in \Omega\}$ that X takes is a discrete subset of $\mathbb{R}$.

A random variable can be visualized in general by the picture in Figure 1.[1] Note that the term "random variable" is really something of a misnomer: it is a function so there is nothing random about it and it is definitely not a variable! What is random is which sample point of the experiment is realized and hence the value that the random variable maps the sample point to.

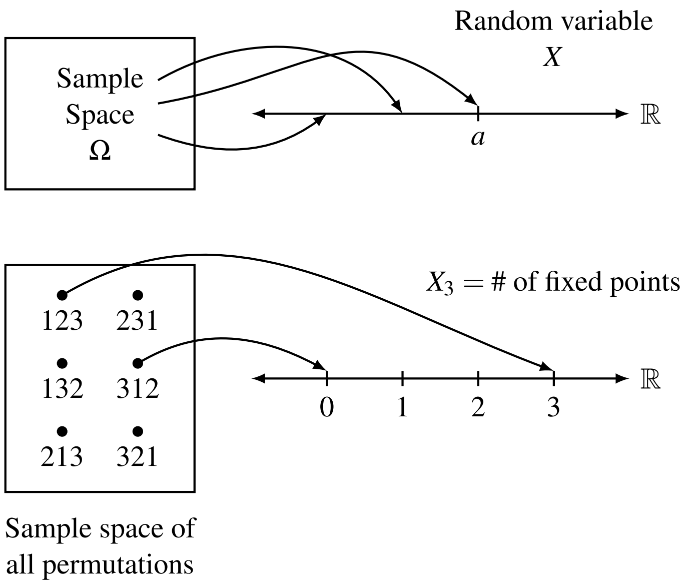

Figure 1: Visualization of how a random variable is defined on the sample space.

Exercise. By completing the lower picture in Figure 1, show that the only possible values for the r.v. X_3 are 0, 1 and 3, and that their probabilities are $\frac{1}{3}$, $\frac{1}{2}$ and $\frac{1}{6}$, respectively.

[1] The figures in this note are inspired by figures in Chapter 2 of *Introduction to Probability* by D. Bertsekas and J. Tsitsiklis.

2 Probability Distribution

When we introduced the basic probability space in an earlier note, we defined two things:

1. The sample space Ω consisting of all the possible outcomes (sample points) of the experiment.
2. The probability of each of the sample points.

Analogously, there are two important things about any random variable:

1. The set of values that it can take.
2. The probabilities with which it takes on the values.

Since a random variable is defined on a probability space, we can calculate these probabilities given the probabilities of the sample points. Let a be any number in the range of a random variable X. Then the set

$$\{\omega \in \Omega : X(\omega) = a\}$$

is an *event* in the sample space (because it is a subset of Ω). We usually abbreviate this event to simply "$X = a$". Since $X = a$ is an event, we can talk about its probability, $\mathbb{P}[X = a]$. The collection of these probabilities, for all possible values of a, is known as the *distribution* of the random variable X.

Definition 18.2 (Distribution)**.** *The distribution of a discrete random variable X is the collection of values $\{(a, \mathbb{P}[X = a]) : a \in \mathscr{A}\}$, where $\mathscr{A}$ is the set of all possible values taken by X.*

Thus, the distribution of the random variable X in our permutation example above is:

$$\mathbb{P}[X = 0] = \frac{1}{3}; \qquad \mathbb{P}[X = 1] = \frac{1}{2}; \qquad \mathbb{P}[X = 3] = \frac{1}{6}.$$

If needed, we may also write $\mathbb{P}[X = a] = 0$ for all other values of a.

The distribution of a random variable can be visualized as a bar diagram, shown in Figure 2. The x-axis represents the values that the random variable can assume. The height of the bar at a value a is the probability $\mathbb{P}[X = a]$. Each of these probabilities can be computed by looking at the probability of the corresponding event in the sample space.

Note that the collection of events $X = a$, for $a \in \mathscr{A}$, satisfy two important properties:

- Any two events $X = a_1$ and $X = a_2$ with $a_1 \neq a_2$ are disjoint.
- The union of all these events is equal to the entire sample space Ω.

The collection of events thus form a *partition* of the sample space (see Figure 2). Both properties follow directly from the fact that X is a function defined on Ω, i.e., X assigns a unique value to each and every possible sample point in Ω. So, when we sum up the probabilities of the events $X = a$, we are really summing up the probabilities of all the sample points, giving us a total of exactly 1.

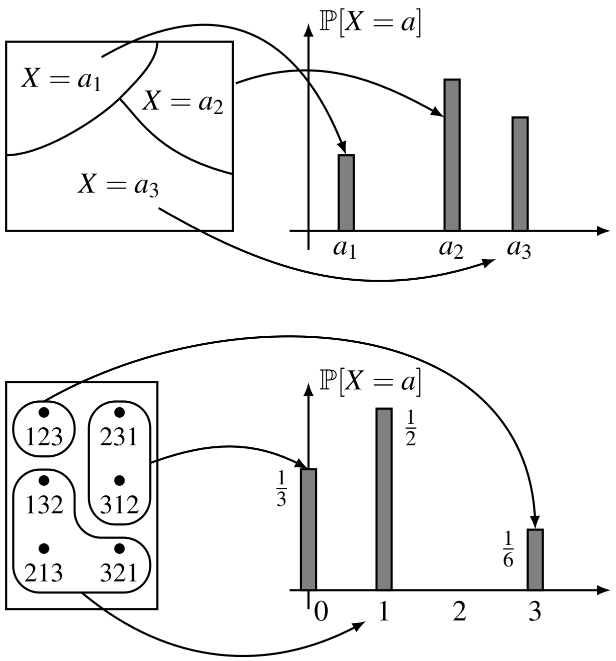

Figure 2: Visualization of how the distribution of a random variable is defined.

Bernoulli Distribution

A simple yet very useful probability distribution is the *Bernoulli* distribution of a random variable which takes value in $\{0,1\}$:

$$\mathbb{P}[X = i] = \begin{cases} p, & \text{if } i = 1, \\ 1 - p, & \text{if } i = 0, \end{cases}$$

where $0 \le p \le 1$. We say that X is distributed as a *Bernoulli* random variable with parameter p, and write

$$X \sim \text{Bernoulli}(p) \qquad \text{or} \qquad X \sim \text{Ber}(p).$$

Binomial Distribution

Let us return to our coin tossing example above, where we defined our random variable X to be the number of heads. More formally, consider the random experiment consisting of n independent tosses of a biased coin that shows H with probability p. Each sample point ω is a sequence of tosses, and $X(\omega)$ is defined to be the number of heads in ω. For example, when $n = 3$, $X(THH) = 2$.

To compute the distribution of X, we first enumerate the possible values that X can take. They are simply $0, 1, \ldots, n$. Then we compute the probability of each event $X = i$ for $i = 0, 1, \ldots, n$. The probability of the event $X = i$ is the sum of the probabilities of all the sample points with exactly i heads (for example, if $n = 3$ and $i = 2$, there would be three such sample points $\{HHT, HTH, THH\}$). Any such sample point has probability $p^i(1-p)^{n-i}$, since the coin flips are independent. There are exactly $\binom{n}{i}$ of these sample points. Hence,

$$\mathbb{P}[X = i] = \binom{n}{i} p^i (1-p)^{n-i}, \qquad \text{for } i = 0, 1, \ldots, n. \tag{1}$$

This distribution, called the *binomial* distribution, is one of the most important distributions in probability. A random variable with this distribution is called a *binomial* random variable, and we write

$$X \sim \text{Bin}(n, p),$$

where n denotes the number of trials and p the probability of success (observing an H in the example). An example of a binomial distribution is shown in Figure 3. Notice that due to the properties of X mentioned above, it must be the case that $\sum_{i=0}^{n} \mathbb{P}[X = i] = 1$, which implies that $\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = 1$. This provides a probabilistic proof of the Binomial Theorem from an earlier note where we saw it combinatorially, for $a = p$ and $b = 1 - p$.

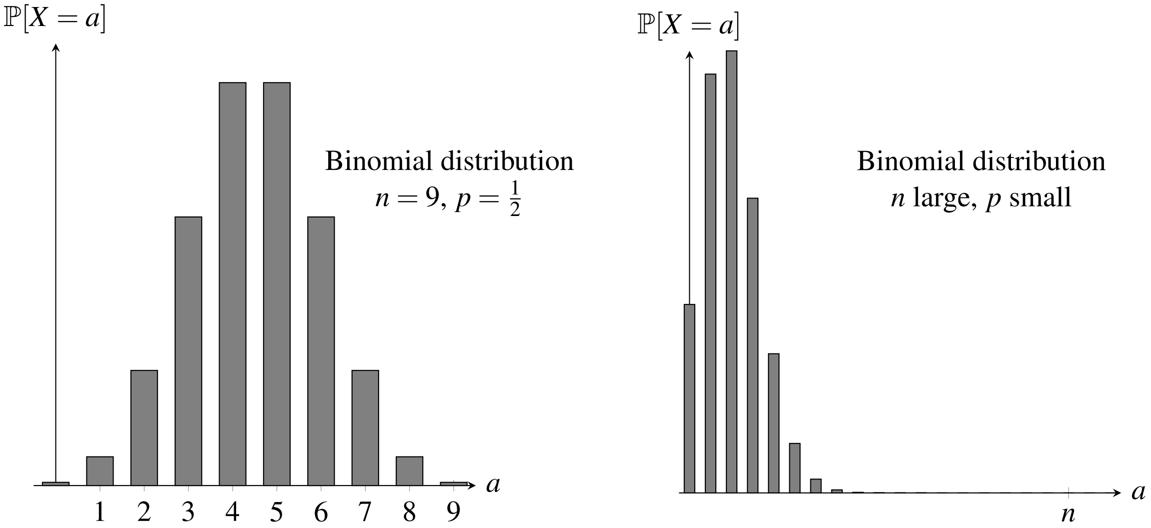

Figure 3: The binomial distributions for two choices of (n, p).

Although we define the binomial distribution in terms of an experiment involving tossing coins, this distribution is useful for modeling many real-world problems. Consider for example the error correction problem studied earlier. Recall that we wanted to encode n packets into $n+k$ packets such that the recipient can reconstruct the original n packets from any n packets received. But in practice, the number of packet losses is random, so how do we choose k, the amount of redundancy? If we model each packet getting lost with probability p and the losses are independent, then if we transmit $n+k$ packets, the number of packets received is a random variable X with binomial distribution: $X \sim \text{Bin}(n+k, 1-p)$ (we are tossing a coin $n+k$ times, and each coin turns out to be a head (packet received) with probability $1-p$). So the probability of successfully decoding the original data is:

$$\mathbb{P}[X \geq n] = \sum_{i=n}^{n+k} \mathbb{P}[X = i] = \sum_{i=n}^{n+k} \binom{n+k}{i} (1-p)^i p^{n+k-i}.$$

Given fixed n and p, we can choose k such that this probability is no less than, say, 0.99.

Hypergeometric Distribution

Consider an urn containing $N = B + W$ balls, where B balls are black and W are white. Suppose you randomly sample $n \leq N$ balls from the urn *with* replacement, and let X denote the number of black balls in your sample. What is the probability distribution of X? Since the probability of seeing a black ball is

B/N for each draw, independently of all other draws, X follows the binomial distribution $\mathrm{Bin}(n,p)$, where $p = B/N$.

What if you randomly sample $n \le N$ balls from the urn *without* replacement? In this case, the probability of seeing a black ball in the i-th draw depends on the colors of the $i-1$ balls already drawn; that is, unlike in the case of sampling with replacement, the samples are not independent. The probability distribution of the number Y of black balls in this setting can be found as follows.

Since all ways of drawing n balls are equally likely, we are in a uniform probability space Ω, so we can compute probabilities using counting. Specifically, for any $k = 0,1,\ldots,n$, we have

$$\mathbb{P}[Y = k] = \frac{|E_k|}{|\Omega|}, \tag{2}$$

where $E_k \subseteq \Omega$ is the set of outcomes that contain exactly k black balls. The total number of possible outcomes is $|\Omega| = \binom{N}{n}$. The total number of outcomes containing exactly k black balls is

$$|E_k| = \binom{B}{k}\binom{N-B}{n-k};$$

to see this, note that there are $\binom{B}{k}$ ways to choose the k black balls out of the B in the urn, and $\binom{N-B}{n-k}$ ways to choose the remaining $n-k$ balls out of the $N-B$ white balls in the urn. Plugging these calculations into (2), we conclude that

$$\mathbb{P}[Y = k] = \frac{\binom{B}{k}\binom{N-B}{n-k}}{\binom{N}{n}},$$

for $k \in \{0,1,\ldots,n\}$. (Note that $\binom{m}{j} = 0$ if $j > m$, so $\mathbb{P}[Y = k] \neq 0$ only if $\max(0, n+B-N) \le k \le \min(n,B)$.)

This probability distribution is called the *hypergeometric distribution* with parameters N, B, n, and we write

$$Y \sim \mathrm{Hypergeometric}(N,B,n).$$

3 Expectation

The distribution of a r.v. contains *all* the information about the r.v. In most applications, however, the complete distribution of a r.v. is very hard to calculate. For example, consider the homework permutation example with $n = 20$. In principle, we would have to enumerate $20! \approx 2.4 \times 10^{18}$ sample points, compute the value of X at each one, and count the number of points at which X takes on each of its possible values (though in practice we could streamline this calculation a bit)! Moreover, even when we can compute the complete distribution of a r.v., it is often not very informative.

For these reasons, we seek to *summarize* the distribution into a more compact, convenient form that is also easier to compute. The most widely used such form is the *expectation* (or *mean* or *average*) of the r.v.

Definition 18.3 (Expectation)**.** *The expectation of a discrete random variable X is defined as*

$$\mathbb{E}[X] = \sum_{a \in \mathscr{A}} a \times \mathbb{P}[X = a], \tag{3}$$

where the sum is over all possible values taken by the r.v.

Technical Note. Expectation is well defined provided that the sum on the right hand side of (3) is absolutely convergent, i.e., $\sum_{a\in\mathscr{A}} |a| \times \mathbb{P}[X = a] < \infty$. There are discrete random variables for which expectations do not exist, such as the r.v. X with distribution $\mathbb{P}[X = i] = \frac{6}{\pi^2 i^2}$ for all positive integers i. (The reason for the factor $\frac{6}{\pi^2}$ here is to make the probabilities sum to 1, since $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$.)

For our simpler permutation example with only 3 students, the expectation is

$$\mathbb{E}[X] = \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{2}\right) + \left(3 \times \frac{1}{6}\right) = 0 + \frac{1}{2} + \frac{1}{2} = 1.$$

That is, the expected number of fixed points in a permutation of three items is exactly 1.

The expectation can be seen in some sense as a "typical" value of the r.v. (though note that $\mathbb{E}[X]$ may not actually be a value that X can take). The question of how typical the expectation is for a given r.v. is a very important one that we shall return to in a later lecture.

Here is a physical interpretation of the expectation of a random variable: imagine carving out a wooden cutout figure of the probability distribution as in Figure 4. Then the expected value of the distribution is the balance point (directly below the center of gravity) of this object.

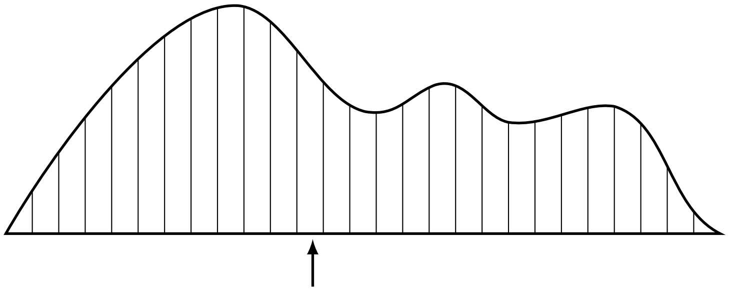

Figure 4: Physical interpretation of expected value as the balance point.

3.1 Examples

1. **Single die.** Throw a fair die once and let X be the number that comes up. Then X takes on values $1,2,\ldots,6$ each with probability $\frac{1}{6}$, so

$$\mathbb{E}[X] = \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = \frac{7}{2}.$$

   Note that X never actually takes on its expected value $\frac{7}{2}$.

2. **Two dice.** Throw two fair dice and let X be the sum of their scores. Then the distribution of X is

a	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}[X = a]$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

   The expectation is therefore

$$\mathbb{E}[X] = \left(2 \times \frac{1}{36}\right) + \left(3 \times \frac{1}{18}\right) + \left(4 \times \frac{1}{12}\right) + \cdots + \left(12 \times \frac{1}{36}\right) = 7.$$

3. **Roulette.** A roulette wheel is spun (recall that a roulette wheel has 38 slots: the numbers $1, 2, \dots, 36$, half of which are red and half black, plus 0 and 00, which are green). You bet \$1 on Black. If a black number comes up, you receive your stake plus \$1; otherwise you lose your stake. Let X be your net winnings in one game. Then X can take on the values $+1$ and -1, and $\mathbb{P}[X=1] = \frac{18}{38}$, $\mathbb{P}[X=-1] = \frac{20}{38}$. Thus,
$$\mathbb{E}[X] = \left(1 \times \frac{18}{38}\right) + \left(-1 \times \frac{20}{38}\right) = -\frac{1}{19};$$
i.e., you expect to lose about a nickel per game. Notice how the zeros tip the balance in favor of the casino!

Geometric and Poisson Distributions

Recall our basic probabilistic experiment of tossing a biased coin n times. This is a very simple model, yet surprisingly powerful. Many important probability distributions that are widely used to model real-world phenomena can be derived from looking at this basic coin tossing model.

The first example is the Binomial(n, p) distribution, introduced earlier. This is the distribution of the number of Heads, S_n, in n tosses of a biased coin with probability p to be Heads. Recall that the distribution of S_n is $\mathbb{P}[S_n = k] = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \dots, n\}$. Below, we will see two additional distributions also derived from coin flips.

4 Geometric Distribution

Consider repeatedly tossing a biased coin with Heads probability p. Let X denote the number of tosses *until the first Head appears*. Then X is a random variable that takes values in $\mathbb{Z}^+$, the set of positive integers. The event that $X = i$ is equal to the event of observing Tails for the first $i-1$ tosses and getting Heads in the ith toss, which occurs with probability $(1-p)^{i-1}p$. Such a random variable is called a *geometric* random variable. Note that X is unbounded: i.e., for any positive integer i (no matter how large), there is some positive probability that $X = i$.

The geometric distribution frequently occurs in applications because we are often interested in how long we have to wait before a certain event happens: how many runs before the system fails, how many shots before one is on target, how many poll samples before we find a Democrat, how many retransmissions of a packet before successfully reaching the destination, etc.

Definition 18.4 (Geometric Distribution)**.** *A random variable X for which*
$$\mathbb{P}[X = i] = (1-p)^{i-1}p, \qquad \textit{for } i = 1, 2, 3, \dots,$$
is said to have the geometric distribution with parameter p. This is abbreviated as $X \sim$ Geometric(p) *or $X \sim$* Geom(p).

As a quick check, we can verify that the total probability of X is equal to 1:
$$\sum_{i=1}^{\infty} \mathbb{P}[X = i] = \sum_{i=1}^{\infty} (1-p)^{i-1}p = p \sum_{i=1}^{\infty} (1-p)^{i-1} = p \times \frac{1}{1-(1-p)} = 1,$$
where in the second-to-last step we have used the formula for the sum of a geometric series.

If we plot the distribution of X (i.e., the values $\mathbb{P}[X = i]$ against i) we get a curve that decreases monotonically by a factor of $1-p$ at each step, as illustrated in Figure 5.

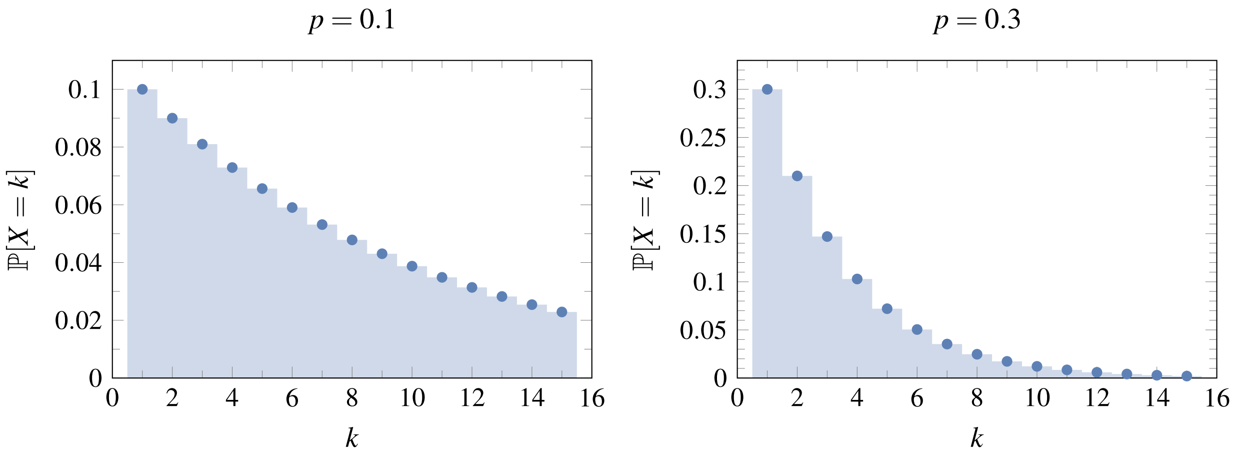

Figure 5: Illustration of the Geometric(p) distribution for $p = 0.1$ and $p = 0.3$.

We will consider the expectation of a geometric random variable in the next set of notes.

4.1 Memoryless Property

The geometric distribution has a special property called *memorylessness*[2]. To get some intuition for this property, suppose we revisit our earlier motivational example of repeatedly tossing a biased coin with heads probability p, and let X denote the number of tosses until the first head appears. Thus $X \sim \text{Geometric}(p)$.

As we've seen, the probability that we need *more than* n tosses before getting the first head is $\mathbb{P}[X > n] = (1-p)^n$, since the first n tosses must all have been tails.

What if we've already tossed the coin m times without getting a head? What is the probability that we need more than n *additional* tosses before getting our first head? We can calculate this probability as:

$$\begin{aligned}\mathbb{P}[X > n+m \mid X > m] &= \frac{\mathbb{P}[X > n+m]}{\mathbb{P}[X > m]}\\ &= \frac{(1-p)^{n+m}}{(1-p)^m}\\ &= (1-p)^n\\ &= \mathbb{P}[X > n]\end{aligned}$$

This gives us the formal statement of the memoryless property of geometric distributions:

$$\mathbb{P}[X > n+m \mid X > m] = \mathbb{P}[X > n]..$$

The geometric distribution is memoryless in the sense that the length of time we have already been waiting does not affect the additional time we have to wait until we get our first head. (This, of course, defies the "common wisdom" that if we've tossed a coin ten times and it's come up tails every time, then surely it's more likely to come up heads on the 11th toss!)

[2] This property is shared by the *exponential distribution*, the continuous analog of the geometric distribution, which we'll cover in a future note. Indeed, these are the *only* two distributions that are memoryless.

5 Poisson Distribution

Another distribution that emerges from consideration of coin flips is the Poisson distribution. In these notes, the connection with coins will not be clear, but we will discuss this in the future.

Consider the number of clicks of a Geiger counter, which measures radioactive emissions. The average number of such clicks per unit time, λ, is a measure of radioactivity, but the actual number of clicks fluctuates according to a certain distribution called the Poisson distribution. What is remarkable is that the average value, λ, completely determines the probability distribution on the number of clicks X.

Definition 18.5 (Poisson distribution)**.** *A random variable X for which*

$$\mathbb{P}[X = i] = \frac{\lambda^i}{i!}e^{-\lambda}, \qquad \textit{for } i = 0, 1, 2, \ldots \tag{4}$$

is said to have the Poisson distribution with parameter λ. This is abbreviated as $X \sim$ Poisson(λ).

To make sure this is a valid definition, let us check that (4) is in fact a distribution, i.e., that the probabilities sum to 1. We have

$$\sum_{i=0}^{\infty} \frac{\lambda^i}{i!}e^{-\lambda} = e^{-\lambda}\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} \times e^{\lambda} = 1.$$

In the second-to-last step, we used the Taylor series expansion $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$.

The Poisson distribution is a very widely accepted model for so-called “rare events,” such as disconnected phone calls, radioactive emissions, crossovers in chromosomes, the number of cases of disease, the number of births per hour, etc. This model is appropriate whenever the occurrences can be assumed to happen randomly with some constant density λ in a continuous region (of time or space), such that occurrences in disjoint subregions are independent. One can then show that the number of occurrences in a region of unit size should obey the Poisson distribution with parameter λ.

Example

Suppose when we write an article, we make an average of 1 typo per page. We can model this with a Poisson random variable X with $\lambda = 1$. So the probability that a page has 5 typos is

$$\mathbb{P}[X = 5] = \frac{1^5}{5!}e^{-1} = \frac{1}{120e} \approx \frac{1}{326}.$$

Now suppose the article has 200 pages. If we assume the numbers of typos on each page are independent, then the probability that there is at least one page with exactly 5 typos is

$$\begin{aligned}
\mathbb{P}[\exists \text{ a page with exactly 5 typos}] &= 1 - \mathbb{P}[\text{every page has} \neq 5 \text{ typos}] \\
&= 1 - \prod_{k=1}^{200} \mathbb{P}[\text{page } k \text{ has} \neq 5 \text{ typos}] \\
&= 1 - \prod_{k=1}^{200} (1 - \mathbb{P}[\text{page } k \text{ has exactly 5 typos}]) \\
&= 1 - \left(1 - \frac{1}{120e}\right)^{200} \approx 0.46,
\end{aligned}$$

where in the last step we have used our earlier calculation for $\mathbb{P}[X = 5]$. □

5.1 Expectation of a Poisson Random Variable

Let us now calculate the expectation of a Poisson random variable.

Theorem 18.1. *For a Poisson random variable $X \sim \text{Poisson}(\lambda)$, we have $\mathbb{E}[X] = \lambda$ and $\text{Var}(X) = \lambda$.*

Proof. We can calculate $\mathbb{E}[X]$ directly from the definition of expectation:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{\infty} i \times \mathbb{P}[X = i] \\
&= \sum_{i=1}^{\infty} i \times \frac{\lambda^i}{i!} e^{-\lambda} && \text{(the } i = 0 \text{ term is equal to 0 so we omit it)} \\
&= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} e^{\lambda} && \left(\text{since } e^{\lambda} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \text{ with } j = i - 1\right) \\
&= \lambda.
\end{aligned}$$

□

A plot of the Poisson distribution reveals a curve that rises monotonically to a single peak and then decreases monotonically. The peak is as close as possible to the expected value, i.e., at $i = \lfloor \lambda \rfloor$. Figure 6 illustrates the Poisson(λ) distribution for $\lambda = 1, 2, 5, 20$.

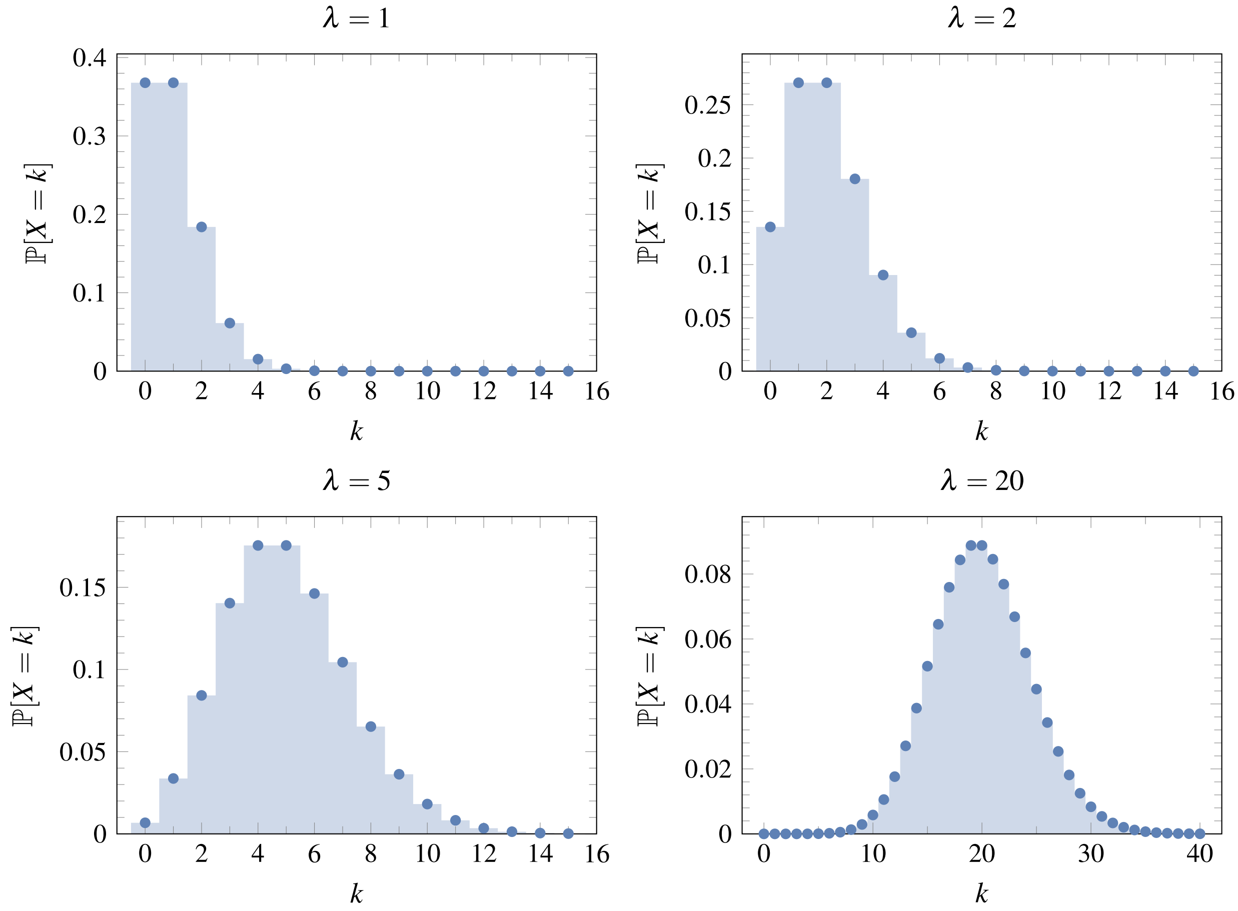

Figure 6: Illustration of the Poisson(λ) distribution for $\lambda = 1, 2, 5, 20$.