Course:	Mathematical Statistics
Term:	Fall 2018
Instructor:	Gordan Žitković

Lecture 1
Discrete random variables

1.1 Random Variables

A large chunk of probability is about random variables. Instead of giving a precise definition, let us just mention that a **random variable** can be thought of as an uncertain (usually numerical, i.e., with values in $\mathbb{R}$, but not always) quantity.

While it is true that we do not know with certainty what value a random variable Y will take, we usually know how to assign a number - the probability - that its value will be in some some[1] subset of $\mathbb{R}$. For example, we might be interested in $\mathbb{P}[Y \geq 7]$, $\mathbb{P}[Y \in [2, 3.1]]$ or $\mathbb{P}[Y \in \{1, 2, 3\}]$.

Random variables are usually divided into **discrete** and **continuous**, even though there exist random variables which are neither discrete nor continuous. Those can be safely neglected for the purposes of this course, but they play an important role in many areas of probability and statistics.

1.2 Discrete random variables

Before we define discrete random variables, we need some vocabulary.

Definition 1.2.1. Given a set B, we say that the random variable Y is B**-valued** if $\mathbb{P}[Y \in B] = 1$.

In words, Y is B-valued if we know for a fact that Y will never take a value outside of B.

Definition 1.2.2. A random variable is said to be **discrete** if there exists a set $\mathcal{S}$ such that $\mathcal{S}$ is either finite or countable[a] and Y is $\mathcal{S}$-valued.

[a] Countable means that its elements can be enumerated by the natural numbers. The only (infinite) countable sets we will need are $\mathbb{N} = \{1, 2, \dots, \}$ or $\mathbb{N}_0 = \{0, 1, 2, \dots\}$.

[1] We will not worry about measurability and similar subtleties in this class.

Definition 1.2.3. The **support** $\mathcal{S}_Y$ of the discrete random variable Y is the smallest set $\mathcal{S}$ such that Y is $\mathcal{S}$-valued.

Example 1.2.4. *A die is thrown and the number obtained is recorded and denoted by Y.* The possible values of Y are $\{1, 2, 3, 4, 5, 6\}$ and each happens with probability $1/6$, so Y is certainly $\mathcal{S}$-valued. Since $\mathcal{S}$ is finite, Y is discrete.

One still needs to argue that $\mathcal{S}$ is the support $\mathcal{S}_Y$ of Y. The alternative would be that $\mathcal{S}_Y$ is a proper subset of $\mathcal{S}$, i.e., that there are redundand elements in $\mathcal{S}$. This is not the case since all elements in $\mathcal{S}$ are "important", i.e., happen with positive probability. If we remove anything from $\mathcal{S}$, we are omitting a possible value for Y.

On the other hand, it is certainly true that Y always takes its values in the finite set $\mathcal{S}' = \{1, 2, 3, 4, 5, 6, \mathbf{7}\}$, i.e., that Y is $\mathcal{S}'$-valued. One has to be careful with the terminology here: it is correct to say that Y is an $\mathcal{S}'$-valued (or even $\mathbb{N}$-valued) random variable, even though it only takes the values $1, 2, \dots, 6$ with positive probabilities.

Discrete random variables are very nice due to the following fact: in order to be able to compute any conceivable probability involving a discrete random variable Y, it is enough to know how to compute the probabilities $\mathbb{P}[Y = y]$, for all $y \in \mathcal{S}$. Indeed, if you are interested in figuring out what $\mathbb{P}[Y \in B]$ is, for some set $B \subseteq \mathbb{R}$ (e.g., $B = \{5, 6, 7\}$, $B = [3, 6]$, or $B = [-2, \infty)$), we simply pick all $y \in \mathcal{S}_Y$ which are also in B and sum their probabilities. In mathematical notation, we have

$$\mathbb{P}[Y \in B] = \sum_{y \in \mathcal{S}_Y \cap B} \mathbb{P}[Y = y]. \qquad (1.2.1)$$

Definition 1.2.5. The **probability mass function (pmf)** of a discrete random variable Y is a function p_Y defined on the support $\mathcal{S}_Y$ of Y by

$$p_Y(y) = \mathbb{P}[Y = y], \; y \in \mathcal{S}_Y.$$

In practice, we usually present the pmf p_Y in the form of a table (called the **distribution table**) as

$$Y \sim \begin{array}{r|cccc} y & y_1 & y_2 & y_3 & \dots \\ \hline p_Y(y) & p_1 & p_2 & p_3 & \dots \end{array}$$

or, simply,

$$Y \sim \begin{array}{|cccc} y_1 & y_2 & y_3 & \dots \\ \hline p_1 & p_2 & p_3 & \dots \end{array},$$

where the top row lists all the elements y of the support $\mathcal{S}_Y$ of Y, and the bottom row lists their probabilities $p_Y(y) = \mathbb{P}[Y = y]$. It is easy to see that the function p_Y has the following properties:

1. $p_Y(y) \in [0, 1]$ for all y, and
2. $\sum_{y \in \mathcal{S}_Y} p_Y(y) = 1$.

Here is a first round of examples of discrete random variables and their supports.

Example 1.2.6.

1. *A fair (unbiased) coin is tossed and the value observed is denoted by Y.* Since the only possible values Y can take are H or T, and the set $\mathcal{S} = \{H, T\}$ is clearly finite, Y is a discrete random variable. Its distribution is given by the following table:

y	H	T
$p_Y(y)$	$1/2$	$1/2$

Both H and T are possible (happen with probability $1/2$), so no smaller set $\mathcal{S}$ will have the property that $\mathbb{P}[Y \in \mathcal{S}] = 1$. Consequently, the support $\mathcal{S}_Y$ of Y is $\mathcal{S} = \{H, T\}$.

2. *A die is thrown and the number obtained is recorded and denoted by Y.* The possible values of Y are $\{1, 2, 3, 4, 5, 6\}$ and each happens with probability $1/6$, so Y is discrete with support $\mathcal{S}_Y$. Its distribution is given by the table

y	1	2	3	4	5	6
$p_Y(y)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

3. *A fair coin is thrown repeatedly until the first H is observed; the number of Ts observed before that is denoted by Y.* In this case we know that Y can take any of the values $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ and that there is no finite upper bound for it. Nevertheless, we know that Y cannot take values that are not non-negative integers. Therefore, Y is $\mathbb{N}_0$-valued and, in fact, $\mathcal{S}_Y = \mathbb{N}_0$ is its support. Indeed, we have $\mathbb{P}[Y = y] = 2^{-y-1}$, for $y \in \mathbb{N}_0$, i.e.,

$$Y \sim \begin{array}{c|cccc} y & 0 & 1 & 2 & \dots \\ \hline p_Y(y) & 1/2 & 1/4 & 1/8 & \dots \end{array}$$

4. *A card is drawn randomly form a standard deck, and the result is denoted by Y.* This example is similar to the 2., above, since Y takes one of finitely many values, and all values are equally likely. The

difference is that the result is not a number anymore. The set $\mathcal{S}$ of all possible values can be represented as the set of all pairs like $(\spadesuit, 7)$, where the first entry denotes the picked card's suit (in $\{\heartsuit, \spadesuit, \clubsuit, \diamondsuit\}$), and the second is a number between 1 and 13. It is, of course, possible to use different conventions and use the set $\{2, 3, \dots, 9, 10, J, Q, K, A\}$ for the second component. The point is that the values Y takes are not numbers.

1.3 Events and Bernoulli random variables

Random variables Y which can only take one of two values $0, 1$, i.e., for which $\mathcal{S}_Y \subseteq \{0, 1\}$, are called **indicators** or **Bernoulli random variables** and are very useful in probability and statistics (and elsewhere). The name comes from the fact that you should think of such variables as signal lights; if $Y = 1$ an event of interest has happened, and if $Y = 0$ it has not happened. In other words, Y *indicates* the occurence of an event.

One reason the Bernoulli random variables are so useful is that they let us manipulate *events* without ever leaving the language of random variables. Here is an example:

Example 1.3.1. Suppose that two dice are thrown so that Y_1 and Y_2 are the numbers obtained (both Y_1 and Y_2 are discrete random variables with $\mathcal{S}_{Y_1} = \mathcal{S}_{Y_2} = \{1, 2, 3, 4, 5, 6\}$). If we are interested in the probability that their sum is at least 9, we proceed as follows. We define the random variable W - the sum of Y_1 and Y_2 - by $W = Y_1 + Y_2$. Another random variable, let us call it Y, is a Bernoulli random variable defined by

$$Y = \begin{cases} 1, & W \geq 9, \\ 0, & W < 9. \end{cases}$$

With such a set-up, Y signals whether the event of interest has happened, and we can state our original problem in terms of Y, namely "Compute $\mathbb{P}[Y = 1]$!".

This example is, admittedly, a little contrived. The point, however, is that anything can be phrased in terms of random variables; thus, if you know how to work with random variables, i.e., know how to compute their distributions, you can solve any problem in probability that comes your way.

Another reason Bernoulli random variables are useful is the fact that we can do arithmetic with them.

Example 1.3.2. 70 coins are tossed and their outcomes are denoted by $W_1, W_2, \ldots, W_{70}$. All W_i are random variables with values in $\{H, T\}$ (and therefore not Bernoulli random variables), but they can be easily **recoded** into Bernoulli random variables as follows:

$$Y_i = \begin{cases} 1, & \text{if } W_i = H, \\ 0, & \text{if } W_i = T. \end{cases}$$

Once you have the "dictionary" $\{1 \leftrightarrow H, 0 \leftrightarrow T\}$, random variables Y_i and W_i carry exactly the same information. The advantage of using Y_i is that the random variable

$$N = \sum_{i=1}^{70} Y_i,$$

which takes values in $\mathcal{S}_N = \{0, 1, 2, \ldots, 70\}$ counts the number of heads among $W_1, \ldots, W_{70}$. Similarly, the random variable

$$M = Y_1 \times Y_2 \times \cdots \times Y_{70}$$

is a Bernoulli random variable itself. What event does it indicate?

1.4 Some widely used discrete random variables

The distribution of a random variable is sometimes defined as "the collection of all possible probabilities associated to it". This sounds a bit abstract, and, at least in the discrete case, obscures the practical significance of this important concept. We have learned that for discrete variables the knowledge of the pmf or the distribution table (such as the one in part 1., 2. or 3. of Example 1.2.6) amounts to the knowledge of the whole distribution. It turns out that many random variables in widely different contexts come with the same (or similar) distribution tables, and that some of those appear so often that they deserve to be named (so that we don't have to write the distribution table every time). The following example lists some of those, **named, distribution**. There are many others, but we will not need them in these notes.

Example 1.4.1.

1. **Bernoulli distribution.** We have already encountered this distribution in our discussion of indicator random variables above. It is characterized by the distribution table of the form

$$\begin{array}{c|cc} & 0 & 1 \\ \hline & 1-p & p \end{array}, \qquad (1.4.1)$$

where p can be any number in $(0, 1)$. Strictly speaking, each value of p defines a different distribution, so it would be more correct to speak of a **parametric family** of distributions, with $p \in (0, 1)$ being the **parameter**.

In order not to write down the table (1.4.1) every time, we also use the notation $Y \sim B(p)$. For example, the Bernoulli random variable which takes the value 1 when a fair coin falls H and 0 when it falls T has a $B(1/2)$-distribution.

An experiment (random occurrence) which can end in two possible ways (usually called *success* and *failure*, even though those names should not always be taken literally) is often called a **Bernoulli trial**. If we "encode" success as 1 and failure by 0, each Bernoulli trial gives rise to a Bernoulli random variable.

2. **Binomial distribution.** A random variable whose distribution table looks like this

0	1	...	$(n-1)$	n
q^n	$\binom{n}{1} p q^{n-1}$	...	$\binom{n}{n-1} q p^{n-1}$	p^n

for some $n \in \mathbb{N}$, $p \in (0, 1)$ and $q = 1 - p$, is called the **binomial distribution**, usually denoted by $b(n, p)$. Remember that the *binomial coefficient* $\binom{n}{k}$ is given by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \text{ where } n! = n(n-1)(n-2) \cdot \ldots \cdot 2 \cdot 1.$$

Binomial distribution(s) form a parametric family with two parameters $n \in \mathbb{N}$ and $p \in (0, 1)$, and each pair (n, p) corresponds to a different binomial distribution.

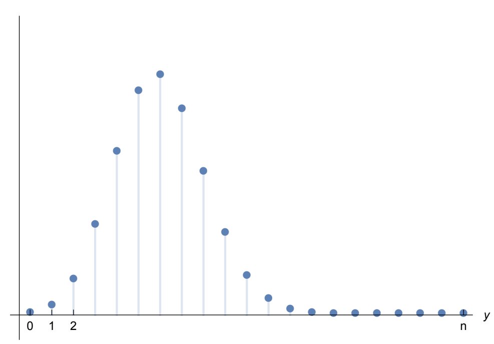

Figure 1. The probability mass function (pmf) of a typical binomial distribution.

Recall that the binomial distribution arises as the "number of successes in n independent Bernoulli trials", i.e., it counts the number of H in n independent tosses of a biased coin whose probability of H is p.

3. **Geometric distribution.** The geometric distribution is similar to the binomial in that it counts the number of "successes" in independent, repeated Bernoulli trials. The difference is that the number of trials is no longer fixed (i.e., $= n$), but we keep tossing until we get our first success. Since the trials are independent, if the probability of success in each trial is $p \in (0,1)$, the probability that it will take exactly k failures before the first success is $q^k p$, where $q = 1 - p$. Therefore, so the geometric distribution - denoted by $g(p)$ - comes with the following table

$$\begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & \dots \\ \hline & p & qp & q^2 p & q^3 p & \dots \end{array}.$$

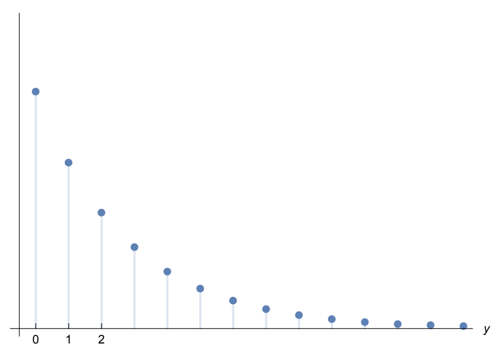

Figure 2. The probability mass function (pmf) of a typical geometric distribution.

> **Caveat:** When defining the geometric distribution, some books count the number of trials to the first success, i.e., add the final success into the count. This shifts everything by 1 and leads to a distribution with support $\mathbb{N}$ (and not $\mathbb{N}_0$). While this is no big deal, this ambiguity tends to be confusing at times and leads to bugs in software. For us, the geometric distribution will always start from 0. The distribution which counts the final success will be referred to as the **shifted geometric distribution**, but we'll try to avoid it altogether.

4. **Poisson distribution.** This is also a family of distributions, parameterized by a single parameter $\lambda > 0$, and denoted by $P(\lambda)$. Its support is $\mathbb{N}_0$ and the distribution table is given by

0	1	2	3	4	$\ldots$
$e^{-\lambda}$	$e^{-\lambda}\lambda$	$e^{-\lambda}\frac{\lambda^2}{2}$	$e^{-\lambda}\frac{\lambda^3}{3!}$	$e^{-\lambda}\frac{\lambda^4}{4!}$	$\ldots$

The closed form for the pmf is

$$p_Y(y) = e^{-\lambda}\frac{\lambda^y}{y!},\ y \in \mathbb{N}.$$

The Poisson distribution arises as a limit when $n \to \infty$ and $p \to 0$ while $np \sim \lambda$ in the Binomial distribution.

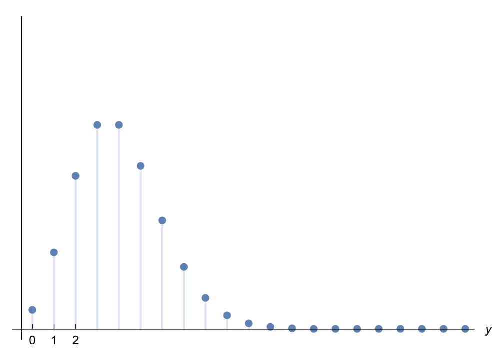

Figure 3. The probability mass function (pmf) of a typical Poisson distribution with $\lambda > 1$.

1.5 Expectations and standard deviations

Expectations and standard deviations provide summaries of numerical random variables - they give us some information about them without overwhelming us with the entire distribution table. The expectation can be thought of as a *center* of the distribution, while the standard deviation gives you an idea about its *spread*[2]

Definition 1.5.1. For a discrete random variable Y with support $\mathcal{S}_Y \subseteq \mathbb{R}$, we define the **expectation** $\mathbb{E}[Y]$ of Y by

$$\mathbb{E}[Y] = \sum_{y \in \mathcal{S}_y} y\, p_Y(y), \tag{1.5.1}$$

if the (possibly) infinite sum $\sum_{y\in\mathcal{S}} y\, p_Y(y)$ *absolutely converges*, i.e., as long as

$$\sum_{y\in\mathcal{S}} |y|\, p_Y(y) < \infty. \tag{1.5.2}$$

When the sum in (1.5.2) above diverges (i.e., takes the value $+\infty$), we say that the expectation of Y **is not defined**.

Perhaps the most important property of the expectation is its linearity:

Theorem 1.5.2. *If $\mathbb{E}[Y_1]$ and $\mathbb{E}[Y_2]$ are both defined then so is $\mathbb{E}[\alpha Y_1 + \beta Y_2]$, for any two constants α, β. Moreover,*

$$\mathbb{E}[\alpha Y_1 + \beta Y_2] = \alpha\mathbb{E}[Y_1] + \beta\mathbb{E}[Y_2].$$

In order to define the standard deviation, we first need to define the variance. Like the expectation, the variance may or may not be defined (depending on whether the sums used to compute it converge absolutely or not). Since we will be working only with distributions for which the existence of expectation(s) is never a problem, we do not mention this issue in the sequel.

Definition 1.5.3. The **variance** of the random variable Y is

$$\mathrm{Var}[Y] = \mathbb{E}[(Y-\mu_Y)^2] = \sum_{y\in\mathcal{S}_Y} (y-\mu_Y)^2 p_Y(y) \text{ where } \mu_Y = \mathbb{E}[Y].$$

The **standard deviation** of Y is

$$\mathrm{sd}[Y] = \sqrt{\mathrm{Var}[Y]}.$$

[2]this should be taken with a grain of salt. After all, what exactly do we mean by a *center* or a *spread* of a distribution?

The fundamental properties of the variance/standard deviation are given in the following theorem:

Theorem 1.5.4. *Suppose that Y_1 and Y_2 are random variables and that α is a constant. Then*

1. $\mathrm{Var}[\alpha Y_1] = \alpha^2 \mathrm{Var}[Y_1]$, *and*
2. *if, additionally, Y_1 and Y_2 are independent, then*

$$\mathrm{Var}[Y_1 + Y_2] = \mathrm{Var}[Y_1] + \mathrm{Var}[Y_2].$$

> ***Caveat:*** *These properties are not the same as the properties of the expectation. First of all the constant comes out of the variance* with a square, *and second, the variance of the sum is the sum of the individual variances only if additional assumptions, such as the* independence *between the two variables, are imposed.*

Finally, here is a very useful alternative formula for the variance of a random variable:

Proposition 1.5.5. $\mathrm{Var}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$.

Let us compute expectations and variances/standard deviations for our most important examples.

Example 1.5.6.

1. **Bernoulli distribution.** Let $Y \sim B(p)$ be a Bernoulli random variable with parameter p. Then (remember q is a shortcut for $1 - p$)

$$\mathbb{E}[Y] = 0 \times q + 1 \times p = p.$$

Using (1.5.5), we get

$$\begin{aligned}\mathrm{Var}[Y] &= \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = 0^2 \times q + 1^2 \times p - p^2 \\ &= p - p^2 = p(1-p) = pq,\end{aligned}$$

and, so, $\mathrm{sd}[Y] = \sqrt{pq}$.

2. **Binomial distribution.** Moving on to the binomial, $Y \sim b(n, p)$, we could either use the formula (1.5.1) and try to evaluate the sum

$$\mathbb{E}[Y] = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k},$$

or use some of the properties of the expectation of Theorem 1.5.2. To do the latter, we remember that the distribution of a binomial is the same as the distribution of a sum of n (independent) Bernoullies. So if we write $Y = Y_1 + \cdots + Y_n$, and each $Y_1 \ldots Y_n$ has the $B(p)$ distribution, Theorem 1.5.2 yields

$$\mathbb{E}[Y] = \mathbb{E}[Y_1] + \mathbb{E}[Y_2] + \cdots + \mathbb{E}[Y_n] = np. \tag{1.5.3}$$

A similar simplification can be achieved in the computation of the variance, too. While it was unimportant in that $Y_1, \ldots, Y_n$ are independent in (1.5.3), it is crucial for Theorem 1.5.4:

$$\mathrm{Var}[Y] = \mathrm{Var}[Y_1] + \cdots + \mathrm{Var}[Y_n] = npq,$$

and, so, $\mathrm{sd}[Y] = \sqrt{npq}$.

3. **Geometric distribution.** The trick from 2. above cannot be applied to the geometric random variables. If nothing else, this is because Theorem 1.5.2 can only be applied to a given (fixed, nonrandom) number n of random variables. We can still use the definition (1.5.1) and evaluate an infinite sum:

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} kpq^k.$$

Instead of doing that, let us proceed somewhat informally and note that we can think of a geometric random variable as follows:

> With probability p our first throw is a success and $Y = 0$. With probability q our first throw is a failure and we restart the experiment on the second throw, making sure to add the first failure to the count.

Therefore,

$$\mathbb{E}[Y] = p \times 0 + q \times (1 + \mathbb{E}[Y]),$$

and, so, $\mathbb{E}[Y] = q/p$.

Similar reasoning can be applied to obtain

$$\begin{aligned}\mathbb{E}[Y^2] &= p \times 0 + q\mathbb{E}[(1+Y)^2] = q + 2q\,\mathbb{E}[Y] + q\mathbb{E}[Y^2] \\ &= q + 2q^2/p + q\mathbb{E}[Y^2],\end{aligned}$$

which yields $\mathrm{Var}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = q/p^2$ and $\mathrm{sd}[Y] = \sqrt{q}/p$.

4. **Poisson distribution.** We know that the Poisson distribution arises as a limit of binomial distributions when $n \to \infty$, $p \to 0$ and

$np \sim \lambda$. We can expect, therefore, that its expectation and variance should behave accordingly, i.e., for $Y \sim P(\lambda)$, we have

$$\mathbb{E}[Y] = \lambda \text{ and } \operatorname{Var}[Y] = \lambda. \tag{1.5.4}$$

The reasoning behind $\operatorname{Var}[Y] = \lambda$ uses the formula $\operatorname{Var}[Y] = npq$ when $Y \sim b(n, p)$ and plugs in $q \equiv 1$, since $q = 1 - p$ and $p \to 0$. A more rigorous way of showing that (1.5.4) is correct is to evaluate the sums

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} k p_Y(k) = \sum_{k=0}^{\infty} k e^{-\lambda} \lambda^k / k! \text{ and}$$
$$\mathbb{E}[Y^2] = \sum_{k=0}^{\infty} k^2 p_Y(k) = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \lambda^k / k!.$$

and use Proposition 1.5.5. The sums can be evaluated explicitly, but since the focus of these notes is not on evaluation of infinite sums, so we skip the details.

1.6 Problems

Problem 1.6.1. A die is rolled 5 times; let the obtained numbers be given by $Y_1, \dots, Y_5$. Use counting to compute the probability that

1. all $Y_1, \dots, Y_5$ are even?
2. at most 4 of $Y_1, \dots, Y_5$ are odd?
3. the values of $Y_1, \dots, Y_5$ are all different from each other?

Problem 1.6.2. Identify the supports of the following random variables:

1. $Y + 1$, where $Y \sim B(p)$ (Bernoulli),
2. Y^2, where $Y \sim b(n, p)$ (binomial),
3. $Y - 5$, where $Y \sim g(p)$ (geometric),
4. $2Y$, where $Y \sim P(\lambda)$ (Poisson).

Problem 1.6.3. Let Y denote the number of tosses of a fair die until the first 6 is obtained (if we get a 6 on the first try, $Y = 0$). The support $\mathcal{S}_Y$ of Y is

(a) $\{0, 1, 2, 3, 4, \dots\}$

(b) $\{1, 2, 3, 4, 5, 6\}$

(c) $\{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}$

(d) $\{\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \left(\frac{5}{6}\right)^3 \times \frac{1}{6}, \dots\}$

(e) none of the above

Problem 1.6.4. The probability that Janet makes a free throw is 0.6. What is the probability that she will make at least 16 out of 23 (independent) throws? Write down the answer as a sum - no need to evaluate it.

Problem 1.6.5. Let Y_1 and Y_2 be random variables with distributions

$$Y_1 \sim \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline & 1/4 & 1/4 & 1/4 & 1/4 \end{array} \text{ and } Y_2 \sim \begin{array}{c|cc} & 1 & 2 \\ \hline & 1/2 & 1/2 \end{array}.$$

Then

(a) $Y_1 + Y_2 \sim \begin{array}{c|ccccc} & 2 & 3 & 4 & 5 & 6 \\ \hline & 1/8 & 1/4 & 1/4 & 1/4 & 1/8 \end{array}$

(b) $\mathcal{S}_{Y_1+Y_2} = \mathcal{S}_{Y_1} \cup \mathcal{S}_{Y_2}$

(c) Y_1 is binomially distributed

(d) the events $\{Y_1 = 1\}$ and $\{Y_2 = 2\}$ are mutually exclusive

(e) none of the above

Problem 1.6.6. (*) Bob and Alice alternate taking customer calls at a call center, with Alice always taking the first call. The number of calls during a day has a Poisson distribution with parameter $\lambda > 0$.

1. What is the probability that Bob will take the last call of the day (that includes the case when there are 0 calls). (*Hint:* What is the Taylor series for the function $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$ around $x = 0$?)

2. Who is more likely to take the last call? Alice or Bob? As above, if there are no calls, we give the "last call" to Bob.

Problem 1.6.7. Three unbiased and independent coins are tossed. Let Y_1 be the total number of heads on the first two coins, and let Y be the random variable which is equal to Y_1 if the third coin comes up *heads* and $-Y_1$ if it comes up *tails*. Compute $\text{Var}[Y]$.

Problem 1.6.8. A die is thrown and a coin is tossed independently of it. Let Y be the random variable which is equal to the number on the die in case the coin comes up *heads* and twice the number on the die if it comes up *tails*.

1. What the support of $\mathcal{S}_Y$ of Y? What is its distribution (pmf)?

2. Compute $\mathbb{E}[Y]$ and $\text{Var}[Y]$.

Problem 1.6.9. n people vote in a general election, with only two candidates running. The vote of person i is denoted by Y_i and it can take values 0 and 1, depending which candidate they voted for (we encode one of them as 0 and the other as 1). We assume that votes are independent of each other and that each person votes for candidate 1 with probability p. If the total number of votes for candidate 1 is denoted by Y, then

(a) Y is a geometric random variable

(b) Y^2 is a binomial random variable

(c) Y is uniform on $\{0, 1, \dots, n\}$

(d) $\mathrm{Var}[Y] \leq \mathbb{E}[Y]$

(e) none of the above

Problem 1.6.10. A discrete random variable Y is said to have a **discrete uniform distribution** on $\{0, 1, 2, \dots, n\}$, denoted by $Y \sim u(n)$ if its distribution table looks like this:

0	1	2	...	n
$\frac{1}{n+1}$	$\frac{1}{n+1}$	$\frac{1}{n+1}$	...	$\frac{1}{n+1}$

Compute the expectation and the variance of $u(n)$. You may use the following identities: $1 + 2 + \dots + n = \frac{1}{2}n(n+1)$ and $1^2 + 2^2 + \dots + n^2 = \frac{1}{6}n(n+1)(2n+1)$.

Problem 1.6.11. (*) Let Y be a discrete random variable such that $\mathcal{S}_Y \subseteq \mathbb{N}_0$. By counting the same thing in two different ways, explain why

$$\mathbb{E}[Y] = \sum_{n \in \mathbb{N}} \mathbb{P}[Y \geq n].$$

This is called the **tail formula** for the expectation.

Problem 1.6.12. Let X be a geometric random variable with parameter $p \in (0, 1)$, i.e. $X \sim g(p)$, and let $Y = 2^{-X}$. Write down the (first few entries in) the distribution table of Y. Compute $\mathbb{E}[Y] = \mathbb{E}[2^{-X}]$.

Problem 1.6.13. Let Y_1 and Y_2 be uncorrelated discrete random variables such that $\mathrm{Var}[2Y_1 - Y_2] = 17$ and $\mathrm{Var}[Y_1 + 2Y_2] = 5$. Compute $\mathrm{Var}[Y_1 - Y_2]$. (*Note:* Y_1 and Y_2 are *uncorrelated* if $\mathbb{E}[(Y_1 - \mathbb{E}[Y_1])(Y_2 - \mathbb{E}[Y_2])] = 0$.)

(*Hint:* What is $\mathrm{Var}[\alpha Y_2 + \beta Y_2]$ in terms of $\mathrm{Var}[Y_1]$ and $\mathrm{Var}[Y_2]$ when Y_1 and Y_2 are uncorrelated?)

Problem 1.6.14. Let Y_1 and Y_2 be uncorrelated random variables such that $\mathrm{sd}[Y_1 + Y_2] = 5$. Then $\mathrm{sd}[Y_1 - Y_2] =$

(a) 1 (b) $\sqrt{2}$ (c) $\sqrt{3}$ (d) 5 (e) not enough information is given

Problem 1.6.15 (*)**.** A mail lady has $l \in \mathbb{N}$ letters in her bag when she starts her shift and is scheduled to visit $n \in \mathbb{N}$ different households during her round. If each letter is equally likely to be addressed to any one of the n households, and the letters are delivered independently of each other, what is the expected number of households that will receive at least one letter? (*Note:* It is quite possible that some households will receive more than 1 letter.)