Lecture 6: One more example for Bayes' formula, discrete random variables, probability mass function

1 One more example for Bayes' formula

Example 1.1. A student is working on a particular multiple choice question in an exam. Suppose that there are m choices in total for the question. The probability that the student knows the correct answer is p. If the student does not know the answer, he/she will make a random guess. Given that the student marks the correct answer, what is the conditional probability that he/she is guessing?

Solution. We first write down the basic events:

$$\begin{aligned} H &: \text{the correct answer is marked,} \\ A_1 &: \text{the student knows the answer,} \\ A_2 &: \text{the student is guessing.} \end{aligned}$$

By the assumptions, we have

$$\mathbb{P}(A_1) = p,\ \mathbb{P}(H|A_1) = 1,\ \mathbb{P}(H|A_2) = \frac{1}{m}.$$

According to Bayes' formula, the desired conditional probability is given by

$$\begin{aligned} \mathbb{P}(A_2|H) &= \frac{\mathbb{P}(A_2) \cdot \mathbb{P}(H|A_2)}{\mathbb{P}(A_1) \cdot \mathbb{P}(H|A_1) + \mathbb{P}(A_2) \cdot \mathbb{P}(H|A_2)} \\ &= \frac{(1-p) \times \frac{1}{m}}{p \times 1 + (1-p) \times \frac{1}{m}} = \frac{1-p}{mp + 1 - p}. \end{aligned}$$

From the above expression, it is clear that as m gets larger, the conditional probability $\mathbb{P}(A_2|H)$ gets smaller. This is not surprising, since if the number of choices gets larger, it is less likely that the student is guessing if he/she marks the correct answer. On the other hand, we can also write

$$\mathbb{P}(A_2|H) = \frac{1-p}{mp+1-p} = 1 - \frac{m}{m-1+\frac{1}{p}}.$$

From this, we see that if p increases, $\mathbb{P}(A_2|H)$ decreases. This is also not surprising: the more likely that the student knows the correct answer, the less likely that he/she is guessing if the correct answer is marked.

2 Random variables: general definitions

Given a random experiment and a sample space, sometimes we are interested in certain functions of the outcome rather than the outcome itself. For instance, when we play the board game Monopoly, we throw a pair of dice to decide the number of moves to take. In this case, we are only interested in the sum $i+j$ of the two dice but not the actual outcome (i,j). The consideration of functions of outcomes leads us to the important notion of random variables.

We begin with the definition of a random variable. Let Ω be a give sample space.

Definition 2.1. A *random variable* is a function (or a mapping)

$$X : \Omega \to \mathbb{R},\ \omega \mapsto X(\omega)$$

defined on the sample space Ω and taking values in $\mathbb{R}$.

Given a random variable $X : \Omega \to \mathbb{R}$, we use S_X to denote the set of its possible values, namely $S_X = \{X(\omega) : \omega \in \Omega\}$. Note that S_X is a subset of $\mathbb{R}$.

Example 2.1. Toss a pair of coins. The sample space is

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\}.$$

Consider the random variable X defined by the total number of heads in the outcome. More precisely, X is defined by

$$X((H,H)) = 2,\ X((H,T)) = X((T,H)) = 1,\ X((T,T)) = 0.$$

Apparently, $S_X = \{0,1,2\}$.

Remark 2.1. As seen from the above example, a random variable needs not be one-to-one (different oucomes can give the same value of X).

Notation. We often use X, Y, Z etc. to denote a random variable (i.e. the function) and x, y, z etc. to denote its generic values. Let X be a given random variable. It is often important to understand the following types of events:

$$\{\omega : X(\omega) = x\},\ \{\omega : X(\omega) \leqslant x\},\ \{\omega : x < X(\omega) \leqslant y\}\ \{\omega : X(\omega) \in A\} \text{ etc.}$$

where x, y are given real numbers. The probabilities of these events encode essential information about the "distribution" of the random variable X. This will be clear when we introduce the notion of cumulative distribution functions later on. We often use the short-handed notation

$$\{X = x\},\ \{X \leqslant x\},\ \{x < X \leqslant y\}\ , \{X \in A\} \text{ etc.}$$

to denote these events and

$$\mathbb{P}(X = x),\ \mathbb{P}(X \leqslant x),\ \mathbb{P}(x < X \leqslant y),\ \mathbb{P}(X \in A) \text{ etc.}$$

to denote their probabilities.

Example 2.2. In the game of Monopoly, we throw a pair of fair dice. The sample space is

$$\Omega = \{(1,1), (1,2), \cdots (6,6)\}.$$

Consider the random variable X defined by the sum of the two upturned faces. More precisely, X is defined by

$$X(\omega) = i + j, \text{ for each } \omega = (i, j) \in \Omega.$$

We have $S_X = \{2, 3, 4, \cdots, 12\}$. It is clear that

$$\{X = 2\} = \{(1,1)\},\ \{X = 3\} = \{(1,2), (2,1)\},$$
$$\{X = 4\} = \{(1,3), (2,2), (3,1)\}.$$

Therefore,

$$\mathbb{P}(X = 2) = \frac{1}{36},\ \mathbb{P}(X = 3) = \frac{2}{36},\ \mathbb{P}(X = 4) = \frac{3}{36}$$

and

$$\mathbb{P}(X \leqslant 4) = \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36}.$$

3 Discrete random variables and the probability mass function

There are two fundamental types of random variables. The study of these two types of random variables and their distributions consistutes a substantial part of the current subject. The basic idea of classifying a random variable is based on whether it takes values discretely or continuously. In the former case, we use simple counting and summation to study its properties. In the latter case, we rely on differentiation and integration from calculus.

Before defining these two types of random variables, we first recall the following concept of countability.

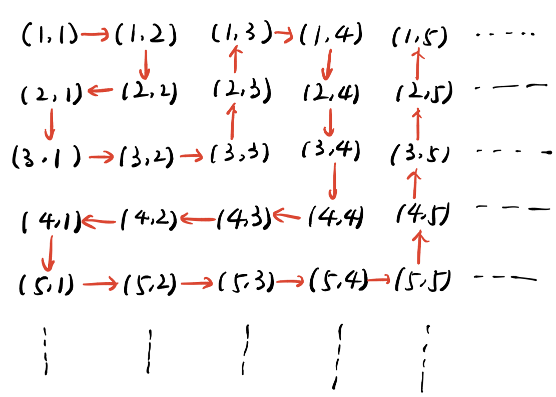
Definition 3.1. A set S is said to be *countable*, if it is either a finite set or its elements can be put in one-to-one correspondence with the set of natural numbers. In the latter case, this means that the elements of S can be labelled and enumerated by $x_1, x_2, x_3, x_4, \cdots$. We say that S is *uncountable* if it is not countable.

Remark 3.1. Intuitively, countability means that the elements of S can be enumerated one after another in a sequential way.

Example 3.1. It is clear that $\mathbb{N} = \{1, 2, \cdots\}$ is countable. The set $\mathbb{Z}$ of integers is also countable, as seen from the following listing:

$$\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \cdots\}.$$

This clearly enumerates all integers in a sequential way. The fact that $\mathbb{N} \times \mathbb{N}$ is countable is less obvious. The figure below gives a way of enumerating its elements by following the indicated arrows.



Example 3.2. It is known that $\mathbb{R}$ and $[0,1]$ are both uncountable. This is an enlightening exercise which is left to the reader. The idea is to assume on the contrary that all elements $[0,1]$ are listed in a sequential way, and then to use decimal expansions to invent an element that is not contained in the presumed sequential list. This gives a contradiction and thus $[0,1]$ cannot be countable.

Having the notion of countability, we can now introduce the first fundamental type of random variables: discrete random variables.

Definition 3.2. A random variable $X : \Omega \to \mathbb{R}$ is said to be a *discrete random variable*, if it takes values in a countable set, namely if S_X is countable.

The random variables in Example 2.1 and Example 2.2 are both discrete random variables (indeed S_X is finite in each of these examples). A simple example of a discrete random variable that takes infinitely many values is the number of tosses needed until we see a Head in the experiment of tossing a coin repeatedly. In this case, S_X is the set of positive integers.

The most important concept associated with discrete random variables is the probability mass function. This function captures all information about the "distribution" of the underlying random variable (only in the discrete case!).

Definition 3.3. Let X be a discrete random variable and S_X be the set of its possible values. The *probability mass function* (*pmf*) of X is the function $p_X : S_X \to [0,1]$ defined by

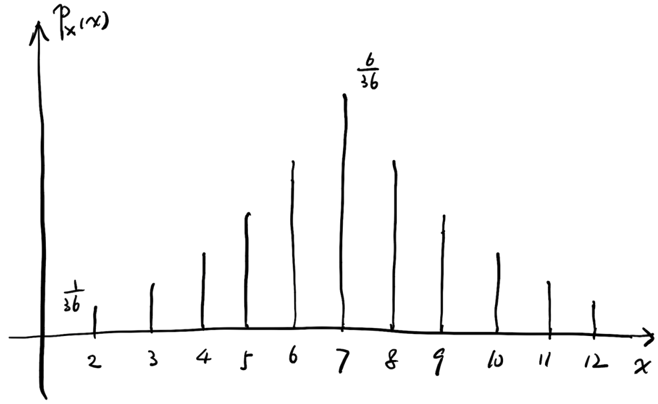
$$p_X(x) = \mathbb{P}(X = x), \quad x \in S_X.$$

Example 3.3. Recall that the Monopoly example, we have $S_X = \{2, 3, 4, \cdots, 12\}$. It is straight forward to see that the pmf of X is given by

$$p_X(2) = \mathbb{P}(X = 2) = \frac{1}{36},\ p_X(3) = \frac{2}{36},\ p_X(4) = \frac{3}{36},\ p_X(5) = \frac{4}{36},\ p_X(6) = \frac{5}{36},$$

$$p_X(7) = \frac{6}{36},\ p_X(8) = \frac{5}{36},\ p_X(9) = \frac{4}{36},\ p_X(10) = \frac{3}{36},\ p_X(11) = \frac{2}{36},\ p_X(12) = \frac{1}{36}.$$

The following figure plots the pmf of X.



Intuitively, the pmf assigns a "probability mass" of $\mathbb{P}(X = x)$ to every possible value x of X. The pmf encodes all information about the distribution of X, in the sense that we can compute the probability

$$\mathbb{P}(X \in A) \text{ for any subset } A \subseteq \mathbb{R}$$

from the knowledge of the pmf. To see this, we first decompose the event $\{X \in A\}$ according to the value of X:

$$\{X \in A\} = \bigcup_{x \in S_X \cap A} \{X = x\}.$$

It is apparent that the events $\{X = x\}$ are disjoint (for different x's) and the above union is a countable union since S_X is countable. By the countable additivity axiom, we have

$$\mathbb{P}(X \in A) = \sum_{x \in S_X \cap A} \mathbb{P}(X = x) = \sum_{x \in S_X \cap A} p_X(x). \tag{3.1}$$

In other words, the probability $\mathbb{P}(X \in A)$ is given by summing up the probability masses for those possible values in A. This is an important and useful formula for discrete random variables.