Discrete Random Variables

Random Variables

Many problems in probability theory revolve around pure numbers rather than arbitrary elements of a sample space (which can be arbitrary objects, such as tuples, or other objects). It is therefore useful to introduce functions that take elements of a sample space and map them into a subset of the real numbers, i.e.,

$$X \colon S \to \mathbb{R}.$$

where such a function X is said to be a ***random variable***.

The term "random variable" originates from the idea that X has numerical values ("variable") that are derived from the outcome of a random experiment ("random").

Random Variables

2.1. Example. Suppose we flip a coin three times. Then the sample space may be given by

$$S = \big\{(t,t,t),\ (t,t,h),\ (t,h,t),\ (t,h,h),\\ (h,t,t),\ (h,t,h),\ (h,h,t),\ (h,h,h)\big\}$$

with t denoting “tails” and h denoting “heads”.

We might now define X as follows:

$$\begin{aligned} X(t,t,t) &= 0, & X(t,t,h) &= 1, & X(t,h,t) &= 1, & X(t,h,h) &= 2, \\ X(h,t,t) &= 1, & X(h,t,h) &= 2, & X(h,h,t) &= 2, & X(h,h,h) &= 3. \end{aligned}$$

Clearly, X denotes the number of heads in three coin flips.

Random Variables

We can now ask what the probability is that X takes on the value 1, which can be found from the probability of each event in the sample space:

$$P[X = 1] = P\Big[\{(t, t, h), (t, h, t), (h, t, t)\}\Big].$$

The notation used on the left is the standard notation for denoting probabilities of random variables. For example, we write

$$P[X = x] = P[A]$$

where $x \in \mathbb{R}$ and $A \subset S$ is the event containing all sample points p such that $X(p) = x$.

More generally, we may write

$$P[a \le X \le b]$$

to denote the probability that the values of X lie between a and b.

Random Variables and Probability Density Functions

Hence, the probability that a random variable takes on values in a certain range is in principle determined from the probability space $(S, \mathscr{F}, P)$. However, to ensure that this works consistently, a lot of mathematical theory is required if the range of X is (for example) an arbitrary subset of $\mathbb{R}$.

Therefore, we will make two assumptions:

1) We distinguish between
   - ***discrete random variables***, defined as having a countable range in $\mathbb{R}$
   - ***continuous random variables***, defined as having range equal to $\mathbb{R}$

   (In principle, a random variable can be of neither of these types, but we will not discuss such cases here.)
2) We assume that a random variable comes with a ***probability density function*** that allows the calculation of probabilities directly, without recourse to the probability space.

Discrete Random Variables

2.2. Definition. Let S be a sample space and Ω a countable subset of $\mathbb{R}$. A ***discrete random variable*** is a map

$$X\colon S \to \Omega$$

together with a function

$$f_X\colon \Omega \to \mathbb{R}$$

having the properties that

(i) $f_X(x) \geq 0$ for all $x \in \Omega$ and

(ii) $\sum_{x\in\Omega} f_X(x) = 1$.

The function f_X is called the ***probability density function*** or ***probability distribution*** of X.

We often say that a random variable is given by the pair (X, f_X).

Density for Discrete Random Variables

For discrete random variables, we define the density function f_X in such a way that

$$f_X(x) = P[X = x].$$

In the following slides, we will introduce various concepts based on examples of discrete random variables. We will derive the density function based on the probabilities of the sample space on which the random variables are defined.

Bernoulli Random Variable

Consider an experiment that can result in two possible outcomes, e.g., success or failure, heads or tails, even or odd. Suppose that the probability of success is p, where $0 < p < 1$. Such an experiment is said to be a ***Bernoulli trial***.

2.3. Definition. Let S be a sample space and

$$X\colon S \to \{0,1\} \subset \mathbb{R}.$$

Let $0 < p < 1$ and define the density function

$$f_X\colon \{0,1\} \to \mathbb{R}, \qquad f_X(x) = \begin{cases} 1-p & \text{for } x = 0 \\ p & \text{for } x = 1. \end{cases}$$

Then X is said to be a ***Bernoulli random variable*** or follow a ***Bernoulli distribution*** with parameter p. We indicate this by writing

$$X \sim \text{Bernoulli}(p)$$

Independent and Identical Trials

More generally, we frequently discuss a sequence of n independent and identical Bernoulli trials. Here,

- ***independent*** means that the outcome of one trial does not influence the outcome of the following trials.
- ***identical*** means that each trial has the same probability of success.

2.4. Example.

- If we flip two fair coins, the two trials are independent and identical.
- If we flip a coin that is fair and another coin that is not fair, the trials are independent but not identical.
- Suppose a box is filled with 10 red balls and 10 black balls. Twice, we draw a ball out of the box but do not replace it. This is a Bernoulli trial where drawing a red ball counts as a "success". The probability of success on the first draw is the same as on the second draw (prove this!). Hence the two trials are identical, but they are clearly not independent. (Since the result of the first draw influences the probability of success in the second draw.)

Counting Successes in a Sequence of Trials

Suppose that we perform a sequence of n independent and identical Bernoulli trials. After recording the results, we define X to be the random variable giving the number of successes in n trials.

To determine the density function of X, we need to find the probability of x successes, where $x = 0, 1, \dots, n$. Note that a given sequence of results with x successes occurs with probability

$$p^x(1-p)^{n-x}$$

since the probability of success is p and the trials are independent and identical. There are $\binom{n}{x}$ ways to place x successes in n trials, hence there are that many sequences with x successes. Since the sequences are mutually exclusive, their probabilities can be added and we find

$$P[x \text{ successes in } n \text{ trials}] = \binom{n}{x} p^x(1-p)^{n-x}.$$

Binomial Random Variable

2.5. Definition. Let S be a sample space, $n \in \mathbb{N} \setminus \{0\}$, and

$$X: S \to \Omega = \{0, \ldots, n\} \subset \mathbb{R}.$$

Let $0 < p < 1$ and define the density function

$$f_X: \Omega \to \mathbb{R}, \qquad f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}. \tag{2.1}$$

Then X is said to be a ***binomial random variable*** with parameters n and p. We indicate this by writing

$$X \sim \mathrm{B}(n, p)$$

Of course, $\mathrm{B}(1, p) = \mathrm{Bernoulli}(p)$.

Binomial Random Variable

It is easy to verify that (2.1) is actually a density function: we check that $f_X(x) \geq 0$ for all $x \in \Omega$ and, furthermore,

$$\sum_{x \in \Omega} f_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p + 1 - p)^n = 1.$$

We have used the binomial theorem here and this is where the name of the distribution comes from.

Cumulative Distribution Function

In practice, we are also often interested in the ***cumulative distribution function*** of a random variable, defined as follows,

$$F_X\colon \mathbb{R} \to \mathbb{R}, \qquad\qquad F_X(x) := P[X \le x]$$

For a discrete random variable

$$F_X(x) = \sum_{y \le x} f_X(y)$$

and, in particular, in the case of the binomial distribution,

$$F_X(x) = \sum_{y=0}^{\lfloor x \rfloor} \binom{n}{y} p^y (1-p)^{n-y} \tag{2.2}$$

where $\lfloor x \rfloor$ denotes the largest integer not greater than x.

Cumulative Distribution Function

2.6. Example. Suppose a fair coin is tossed 10 times. Then the probability of obtaining not more than three heads is

$$F_X(3) = \sum_{y=0}^{3} \binom{10}{y} \frac{1}{2^{10}} = \frac{1+10+45+120}{1024} = \frac{11}{64}$$

There is no simple way of evaluating the sum (2.2), so the values have been tabulated (Table I of Appendix A in the textbook). The Mathematica command for a cumulative distribution function is is **CDF**, which for the binomial distribution, however, gives only a representation in terms of a generalized function:

```
CDF[BinomialDistribution[n, p], x]
```

$$\begin{cases} \text{BetaRegularized}[1-p, n-\text{Floor}[x], 1+\text{Floor}[x]] & 0 \le x \le n \\ 1 & x > n \\ 0 & \text{True} \end{cases}$$

The Geometric Distribution

Let us now look at another example: suppose we perform a sequence of i.i.d. Bernoulli trials which continues until a success is obtained. We then define the ***geometric random variable*** X to denote the number of trials needed to obtain the first success.

2.7. Example. A fair coin has probability $p = 1/2$ of turning heads up when flipped. The coin is flipped until the first appearance of heads, with the following result: (t, t, t, h). In this experiment, the geometric random variable X attains the value $X = 4$.

The Geometric Distribution

2.8. Definition. Let S be a sample space and

$$X \colon S \to \Omega = \mathbb{N} \setminus \{0\}.$$

Let $0 < p < 1$ and define the density function $f_X \colon \mathbb{N} \setminus \{0\} \to \mathbb{R}$ given by

$$f_X(x) = (1-p)^{x-1} p. \tag{2.3}$$

We say that X is a ***geometric random variable*** with parameter p and write $X \sim \mathrm{Geom}(p)$.

The cumulative distribution function for a geometrically distributed random variable (X, f_X) with parameter p is given by

$$F(x) = P[X \le x] = 1 - q^{\lfloor x \rfloor},$$

where $q = 1 - p$ is the probability of failure and $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

Probabilities and the Geometric Distribution

In Mathematica, the probability density function f_X is accessed through the command **PDF**, followed by the name of the distribution and the variable of f_X:

```
PDF[GeometricDistribution[p], x]
```

$$\begin{cases} (1-p)^x p & x \geq 0 \\ 0 & \text{True} \end{cases}$$

Note that this differs from (2.3): $x-1$ is replaced by x. We note: In Mathematica, the geometric distribution gives the ***number of failures*** before the first success, while our convention is to give the ***number of trials*** needed for the first success. This is a minor difference and can easily be compensated for, but it illustrates an important point:

When using a computer program, always check that the definitions in the program are the same as the ones you are using!

Probabilities and the Geometric Distribution

The concrete value $f_X(x)$ can be calculated if a given value of x is inserted:

```
PDF[GeometricDistribution[p], 4]
```

$(1 - p)^4 p$

Probability can be used to find probabilities such as $P[a \le x \le b]$:

```
Probability[1 < x ≤ 4, x ≈ GeometricDistribution[p]]
```

$(-1 + p)^2 p \left(3 - 3 p + p^2\right)$

Note that to express equalities as a condition (and not as an assignment of values), Mathematica requires the use of two equals signs:

```
Probability[x == 4, x ≈ GeometricDistribution[p]]
```

$(-1 + p)^4 p$

Probabilities and the Geometric Distribution

Question. On August 18, 1913, at the casino in Monte Carlo, a roulette wheel returned black more than 20 times in a row. Find the probability of such an event!