Lecture 5

Random Variables, Densities & Distributions

Often we are not interested in the details of the random experiment or its resulting probability space S, per se. Instead, we may be interested only in a particular *quantifiable feature* of the outcomes.

For instance, an insurance company may not care much about the underlying probability space when it sells an auto insurance policy to a customer. It may only be interested in the number of accidents, or loss per accident, the policy holder will have during the lifetime of the policy. These are quantifiable features even though the underlying probability space may be extremely complex.

Random variables help us collect the probabilities of such a quantifiable feature from any probability space. For the insurance example the company may need to compute

$$\begin{aligned}&\mathbb{P}(\text{policy holder will have no accident}),\\ &\mathbb{P}(\text{policy holder will have one accident}),\\ &\mathbb{P}(\text{policy holder will have two accidents}),\quad \text{et cetera.}\end{aligned}$$

The company would like to have all these probabilities in hand in advance so that they can figure out what premiums they should charge the client. A collection of all these probabilities associated to a random variable can be displayed by a function, called its *density*.

A random variable (rv) is itself a *real valued function* defined over a sample space S. It automatically creates several related concepts, whether we use them or not. So, the study of a random variable involves studying all these related concepts that the random variable creates, five of which are listed below.

- (i) As a real valued function over S (i.e., the definition of an rv).
- (ii) As a numerically labeled partition of the sample space S.
- (iii) The density of the random variable, which presents the probabilities of the partitioning events created by the random variable.

- (iv) The cumulative distribution function (cdf) that the random variable produces.
- (v) As a random draw from a particular population of numbers (perhaps constructed artificially).

So think of an rv as not an octopus but rather a "*pentapus*".

5.1 Discrete Random Variables

A random variable is nothing but an assignment of numbers (i.e., a numerical valued function) to all the outcomes of the random experiment.

Example - 5.1.1 - **(Random variable as a real valued function)** Consider the sample space, S, consisting of all possible outcomes of four tosses of a coin. That is,

$$S=\left\{\begin{array}{cccc} HHHH, & HHHT, & HHTH, & HTHH,\\ THHH, & HHTT, & HTHT, & THHT,\\ TTHH, & THTH, & HTTH, & TTTH,\\ TTHT, & THTT, & HTTT, & TTTT \end{array}\right\}.$$

Question: which outcome of this sample space do you like the most?

Obviously, unless we bring in some preference mechanism, the question is vague and uninteresting. But if I inform you that, if you perform this random experiment once, I will give you as many dollars as the number of heads in the outcome you observe, then the obvious answer to the question is the outcome $HHHH$. You will hate to observe the outcome $TTTT$. How do you compare the outcomes

$$TTTH,\quad TTHT,\quad THTT,\quad HTTT?$$

You don't! You attach the same importance to each of these four outcomes. This numerical ranking is best captured by a random variable. The random variable labels each outcome by the amount of money it will generate, as collect below.

$$\begin{array}{cccc} HHHH\to 4 & HHHT\to 3 & HHTH\to 3 & HTHH\to 3\\ THHH\to 3 & HHTT\to 2 & HTHT\to 2 & THHT\to 2\\ TTHH\to 2 & THTH\to 2 & HTTH\to 2 & TTTH\to 1\\ TTHT\to 1 & THTT\to 1 & HTTT\to 1 & TTTT\to 0. \end{array}$$

By the way, we usually denote a random variable (function) by any one of the last few capital letters of the English alphabet, such as X or Y or Z or W etc. So the above view of a random variable, say X, is nothing but a real-valued function defined on the sample space S describing the amount generated by the random experiment. The next example provides a slightly different perspective.

Example - 5.1.2 - **(The partition created by a random variable)** For the random variable X of the last example $X(TTTT)=0$. In other words, $X=0$ if and only if the outcome $TTTT$ occurs. In inverse image notation $X^{-1}(0)=\{TTTT\}$.

Watermark

Probabilists usually abbreviate this by writing $\{X = 0\}$. Similarly, $\{X = 1\}$ stands for the event

$$\{TTTH,\ TTHT,\ THTT,\ HTTT\},$$

since $X(TTTH) = 1$ and $X(TTHT) = 1$ and so on.

Continuing this way, we obtain a partition of S that the random variable, X, has created. The components of this partition are as follows:

$$\begin{aligned}
\{X = 0\} &= \{TTTT\},\\
\{X = 1\} &= \{TTTH,\ TTHT,\ THTT,\ HTTT\},\\
\{X = 2\} &= \{HHTT,\ HTHT,\ THHT,\ TTHH,\ THTH,\ HTTH\},\\
\{X = 3\} &= \{HHHT,\ HHTH,\ HTHH,\ THHH\},\\
\{X = 4\} &= \{HHHH\}.
\end{aligned}$$

Every random variable creates a partition of the associated sample space S. You should have notice that so far the concept of probability has not come under discussion. The next example gives yet another perspective of a random variable by bringing probabilities of the partitioning events into focus.

Example - 5.1.3 - **(The density created by a random variable)** Consider the last example one more time. The event $\{X = 2\}$, i.e., "the experiment will yield 2 dollars" is the same event as

$$\{X = 2\} = \{HHTT,\ HTHT,\ THHT,\ TTHH,\ THTH,\ HTTH\}.$$

Hence, we immediately see that, when the coin is fair, (by using the counting technique of computing probabilities)

$$\mathbb{P}(X = 2) = \frac{6}{16} = \frac{3}{8}.$$

The same reasoning gives us the probabilities of all the numerical labels (rankings):

$$\begin{aligned}
&\mathbb{P}(X = 0) = \frac{1}{16}, \quad \mathbb{P}(X = 1) = \frac{4}{16}, \quad \mathbb{P}(X = 2) = \frac{6}{16},\\
&\mathbb{P}(X = 3) = \frac{4}{16}, \quad \mathbb{P}(X = 4) = \frac{1}{16}.
\end{aligned}$$

The collection of all the values (labels) that X used (which form the range of the function X), along with probabilities of their corresponding events, constitutes the density[1] of the random variable. Here is this density of X in a tabular form.

Values of X	0	1	2	3	4
Probabilities	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

[1] This kind of density is also called a probability mass function in some text books.

Example - 5.1.4 - Let us consider another example. Suppose we roll two fair dice and note the two faces that come up. In this case, the sample space consists of 36 elements. Suppose we were only interested in knowing probabilities of certain events which deal with the sum of the two face values that come up on the two dice. In this case our random variable X measures the sum of the two face values. As a function, the random variable X can be stated as

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$$X(i, j) = i + j,$$

where i = outcome of first die, and j = outcome of second die. Note that the event $\{X = 2\}$ is the same as the event $\{(1, 1)\}$ in the original sample space and the event $\{X = 3\}$ is the same event as $\{(1, 2), (2, 1)\}$ and so on.

The partition of S that X created is as follows:

$$\begin{aligned}
\{X = 2\} &= \{(1, 1)\}\\
\{X = 3\} &= \{(1, 2), (2, 1)\}\\
\{X = 4\} &= \{(1, 3), (2, 2), (3, 1)\}\\
\{X = 5\} &= \{(1, 4), (2, 3), (3, 2), (4, 1)\}\\
\{X = 6\} &= \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}\\
\{X = 7\} &= \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}\\
\{X = 8\} &= \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}\\
\{X = 9\} &= \{(3, 6), (4, 5), (5, 4), (6, 3)\}\\
\{X = 10\} &= \{(4, 6), (5, 5), (6, 4)\}\\
\{X = 11\} &= \{(5, 6), (6, 5)\}\\
\{X = 12\} &= \{(6, 6)\}.
\end{aligned}$$

The density that X created, of course, is

Values of X	2	3	4	5	6	7	8	9	10	11	12
Probabilities	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Once the density of X is in our hand, we no longer have to go back to the sample space to find the probabilities of these events. This is the main benefit of having the density of a random variable.

Remark - 5.1.1 - **(Caution: Density vs relative frequency distribution)** In the last two examples we obtained the densities without rolling a die or tossing a coin. Conceptual understanding of the rules of probability were enough! Had we tossed the coin or rolled the die (i.e. performed the experiment) a large number of

Watermark

times, we would have obtained the data distribution written as a relative frequency-distribution. The density and the relative frequency distribution differ the same way as the relative frequency of an event and its probability differed (as explained in earlier lectures). The density being a limiting form of the relative frequency distribution.

A computer rolled a pair of fair dice 10,000 times and obtained the following relative frequency distribution of the sum of the two face values. The last column gives the exact values of the density in fractions for comparison purposes.

Values of X	Relative Frequencies	Probabilities
2	0.0256	$0.0278 = 1/36$
3	0.0565	$0.0556 = 2/36$
4	0.0858	$0.0833 = 3/36$
5	0.1120	$0.1111 = 4/36$
6	0.1367	$0.1389 = 5/36$
7	0.1650	$0.1667 = 6/36$
8	0.1376	$0.1389 = 5/36$
9	0.1147	$0.1111 = 4/36$
10	0.0820	$0.0833 = 3/36$
11	0.0554	$0.0556 = 2/36$
12	0.0287	$0.0278 = 1/36$

Our relative frequencies are close to the actual values of the density. The moral of the story is that once we have the density of a random variable in our hand, we do not have to perform the random experiment at all to find the probabilities of various events since we know how the relative frequencies will behave, more or less.

Definition - 5.1.1 - **(A discrete rv & its density)** *Let S be a sample space. A discrete random variable, X, is a numerical valued function over S which partitions the sample space S into countably many disjoint events. The discrete probability density*[2] *of the random variable consists of two arrays of numbers, $a_1, a_2, \cdots$, representing all the distinct values of the range of X, and their corresponding probabilities, $p_1, p_2, \cdots$, such that*

- *all $p_1, p_2, \cdots$, are nonnegative numbers, and*
- $p_1 + p_2 + \cdots = 1$.

We call p_1 the probability assigned to a_1, and represent this association by writing $\mathbb{P}(X = a_1) = p_1$. Similarly, p_2 is called the probability assigned to a_2, written as $\mathbb{P}(X = a_2) = p_2$, etc. This information is sometimes presented in a tabular form:

Values of X	a_1	a_2	a_3	$\cdots$	a_n	$\cdots$
Probabilities	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$

[2] Also known as a probability mass function, or a probability distribution or just a density.

If A is a subset of the real line, the probability, $\mathbb{P}(X \in A)$, is obtained by adding the p's that are assigned to the a's lying in A. That is,

$$\mathbb{P}(X \in A) = \sum_{a \in A} \mathbb{P}(X = a).$$

Example - 5.1.5 - For the random variable X of the last example of rolling two fair dice and observing the sum of the face values, $\mathbb{P}(X \geq 10) = \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{1}{6}$.

5.2 The Cumulative Distribution Function (cdf)

The cumulative distribution function (cdf) of a discrete random variable X is a function defined over the whole real line as follows:

$$F(t) := \mathbb{P}(X \leq t) = \sum_{a:\ a \leq t} \mathbb{P}(X = a), \qquad -\infty < t < \infty.$$

Note $F(t)$ is always a nondecreasing function of t, and $0 \leq F(t) \leq 1$.

Example - 5.2.1 - Consider the random variable X of Example 5.1.3 counting the total number of heads when a fair coin is tossed four times. From its density we see that its cdf is

$$F(t) = \begin{cases} 0 & \text{if } t < 0, \\ \frac{1}{16} & \text{if } 0 \leq t < 1, \\ \frac{1}{16} + \frac{4}{16} = \frac{5}{16} & \text{if } 1 \leq t < 2, \\ \frac{1}{16} + \frac{4}{16} + \frac{6}{16} = \frac{11}{16} & \text{if } 2 \leq t < 3, \\ \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} = \frac{15}{16} & \text{if } 3 \leq t < 4, \\ 1 & \text{if } t \geq 4. \end{cases}$$

Both the density and the cdf can be plotted as shown below. The density is just

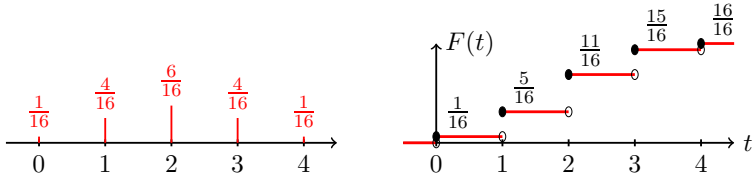


Figure 5.1: The Density and the Cumulative Distribution Function.

a "stick graph", where the height of each stick is the probability amount. The location of the stick is the corresponding value of the random variable.

Watermark

Example - 5.2.2 - **(The secretary problem revisited)** Consider the secretary's matching problem, Example 2.1.4, once again. We have n letters and their corresponding n envelopes. Letters are placed into the envelopes randomly. We would like to know the likelihoods of the number of letters which will go into their own envelopes. Define the random variable

$$X := \text{the number of correctly placed letters.}$$

In this case, to obtain the density one needs sophisticated counting methods. However, when n is small, by listing all the outcomes and using equally likely sample spaces, one can find the density of this random variable.

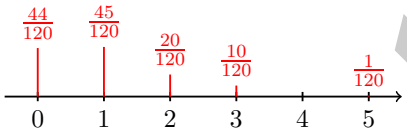
For example, when we have $n = 3$ letters along with three corresponding envelopes addressed to Tom, Dick and Harry, then there are six possible ways of putting the letters into the envelopes. By direct enumeration the reader can verify that $\mathbb{P}(X = 0) = \frac{2}{6}$, $\mathbb{P}(X = 1) = \frac{3}{6}$, etc., giving the density

Values of X	0	1	2	3
Probabilities	$\frac{2}{6}$	$\frac{3}{6}$	0	$\frac{1}{6}$

When we have $n = 5$ letters, there are $5! = 120$ possible ways we can put the 5 letters into the 5 envelopes, so S has 120 outcomes. Let X_5 be the number of correctly placed letters. The density of X_5 can be enumerated similarly:

Values of X_5	0	1	2	3	4	5
Probabilities	$\frac{44}{120}$	$\frac{45}{120}$	$\frac{20}{120}$	$\frac{10}{120}$	0	$\frac{1}{120}$

A stick graph plot of the density is shown below. The cumulative distribution



function (cdf) is useful, for instance, to compute the probability that the majority of the recipients got their own letters is

$$\mathbb{P}(X_5 \geq 3) = \frac{10}{120} + 0 + \frac{1}{120} = \frac{11}{120} = 0.09.$$

Note that the event $\{X_5 \geq 3\}$ stands for "3 or more people, among the 5 people, received their own letters". So,

$$F(2) \;=\; \mathbb{P}(X_5 \leq 2) \;=\; 1 - \mathbb{P}(X5 > 2) \;=\; 1 - \mathbb{P}(X_5 \geq 3) \;=\; 0.91.$$

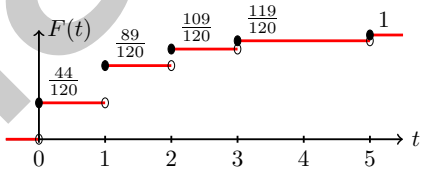
If we were asked to predict the likelihood of at most one person receiving his letter then the event is $\{X \leq 1\}$ with probability

$$F(1) \;=\; \mathbb{P}(X_5 = 0) + \mathbb{P}(X_5 = 1) = \frac{44}{120} + \frac{45}{120} = \frac{89}{120} \approx 0.7416.$$

So, there is about 75% chance that at most one person will get his letter. The cdf of this random variable is as follows.

$$F(t) \;=\; \begin{cases} 0 & \text{if } t < 0, \\ \frac{44}{120} & \text{if } 0 \leq t < 1, \\ \frac{44}{120} + \frac{45}{120} = \frac{89}{120} & \text{if } 1 \leq t < 2, \\ \frac{44}{120} + \frac{45}{120} + \frac{20}{120} = \frac{109}{120} & \text{if } 2 \leq t < 3, \\ \frac{44}{120} + \frac{45}{120} + \frac{20}{120} + \frac{10}{120} = \frac{119}{120} & \text{if } 3 \leq t < 5, \\ 1 & \text{if } 5 \leq t. \end{cases}$$

Its plot is shown below. The reader should see how the density can be pulled out of the cdf, making the two concepts equivalent.



Definition - 5.2.1 - **(The distribution function)** *The distribution function (or cumulative distribution function, cdf) of any random variable, X, is*

$$F(t) \;=\; \mathbb{P}(X \leq t), \qquad t \in \mathbb{R}.$$

For some random variables one can write $F(t)$ in a closed form, but for many we cannot do so. A function $F(t)$ is the cdf of a discrete random variable if and only if it has the following properties:

- *$F(t)$ is a nondecreasing function of t,*
- *$F(-\infty) = \lim_{x \to -\infty} F(x) = 0$, $\qquad F(\infty) = \lim_{x \to +\infty} F(x) = 1$,*
- *$F(t)$ is a right continuous function,*
- *There exists a countable set Δ of real numbers so that $\mathbb{P}(X = a) = F(a) - F(a^-) > 0$, for any $a \in \Delta$ and $\sum_{a \in \Delta} \mathbb{P}(X = a) = 1$.*

Remark - 5.2.1 - **(Random selection from a population)** Now we give another view of a random variable and its density. Consider tossing four fair coins and observing the number of heads that appear. The resulting random variable, say X, has the following density

Values of X	0	1	2	3	4
Probabilities	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Now consider another experiment in which sixteen identically shaped marbles are placed in a box, one marble is labeled 0, four marbles are labeled 1, six marbles

Watermark

are labeled 2, four marbles are labeled 3, and one marble is labeled 4. We reach in and randomly draw one marble and note its value, denote it by Y. The collection of all the numbered marbles is our population

$$\text{Population} = \{0,\ 1,\ 1,\ 1,\ 1,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 3,\ 3,\ 3,\ 3,\ 4\}.$$

It is clear that the density of Y is the same as the above density of X.

Note that when I perform the four-coin experiment and obtain my X and you perform the random-draw experiment and you get your Y, most likely we will not get the same number. That is,

$$X \neq Y.$$

But what is true is that X has the same density (distribution) as Y. This type of distributional equivalence is denoted by the $\sim$ symbol, i.e.,

$$X \sim Y,$$

and we say that X and Y are identically distributed. As far as probability of events is concerned, it is irrelevant which of the two types of experiments we are imagining. This redundancy/diversity of random experiments which lead to the same probability distribution points towards why probability theory has extremely rich fields of applications.

In the field of Statistics, this random draw version is quite popular and statisticians call the randomly drawn random variable, Y, as a sample from the population. In this case a sample of size 1 since we drew only once.

If we draw and get, say Y_1, and put the marble back into the population and draw again, we get another random variable, say Y_2. Note that

$$Y_1 \ \sim \ Y_2.$$

In this case we have a sample of size two, namely Y_1, Y_2. We can draw larger samples if we continue this process. Of course if we do not replace the marble we drew on the first draw and then obtain the second marble and read its number, say Z, then Z will not have the same distribution as Y. That is,

$$Y \ \not\sim \ Z.$$

Sampling with replacement and sampling without replacement are two of the most popular sampling techniques of statistics. However, there are literally hundreds of other types of sampling techniques (which we will not go into).

Remark - 5.2.2 - **(Urn models)** When a random variable, X, takes only finitely many values, with probabilities that are rational numbers, we can always construct an artificial random experiment involving an urn containing certain number of identical looking slips of paper (or ping-pong balls). Each slip has a number written on it. We reach in and draw a slip at random. The resulting random variable Y that records the number obtained from the drawn slip, can be made to have the same density as X.

You might wonder why is this interesting? Well, modern day computers can be programmed to act like an urn and do the random drawing — over and over

again —. The moral of the story is that extremely complex random phenomena, such as weather forecasting, nuclear reactor safety, atomic weapons testing, trends in social attitudes, etc., can be studied quickly, safely and economically with this simple urn model artifact. There are books devoted to this topic, see for instance [36] and [61].

Yet another and much deeper moral of the above story is that, as far as probability statements are concerned, knowing the actual sample space is irrelevant. Having a random variable with an appropriate density/cdf is all that is needed to deal with probabilistic statements. This is how modeling comes into the picture. A correct model of any random phenomenon is judged by the accuracy of the resulting density/cdf. This is how Maxwell came up with the ideal gas law. We do not claim that the mathematical framework that Maxwell concocted is exactly how gas molecules behave. Why the model is accepted is that the density/cdf it provides matches with the experimentally observed data extremely well. No one really cares about the actual sample space that Maxwell's model sits on.

Remark - 5.2.3 - **(Which way to view a random variable?)** For many random variables all five ways of expressing a random variable can be studied interchangeably. However, for the other random variables this is not possible due to some mathematical issues and we then resort to paying more attention to the aspect which we can deal with easily.

5.3 Exercises

Exercise - 5.3.1 - *With the help of a computer, perform an experiment of tossing four fair coins* $20,000$ *times. Then write the relative frequency table of the number of heads observed.*

Exercise - 5.3.2 - *For Exercise 5.3.1, provide the exact density of number of heads observed. How does it compare with your relative frequency table?*

Exercise - 5.3.3 - *Let S be the sample space representing the outcomes of three tosses of a weighted coin for which "H" is three times as likely as a "T". Take the power set, $\mathcal{P}(S)$, as the class of events. Let X be a random variable that equals the largest number of successive heads in the outcome. For instance, $X(HHT) = 2$ and $X(HTH) = 1$. Find the density of X.*

Exercise - 5.3.4 - *Which of the following tables represent densities? Explain.*

Values of X	-1	0	1
Probabilities	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$

Values of X	0	1	2
Probabilities	$\frac{-1}{2}$	$\frac{3}{4}$	$\frac{3}{4}$

Watermark