Stat 400, section 3.1-3.2 Random Variables & Probability Distributions

notes by Tim Pilachowski

For a given situation, or **experiment**, observations are made and data is recorded. A **sample space** S must contain *all possible* outcomes for an experiment. An **event** (designated with a capital letter A, B, C, etc.) is a subset of the sample space, and will incorporate one or more of the outcomes.

Example A. Define X = number of heads in ten tosses of a coin. What are the values that X may assume?

Note carefully: The probabilities for the values of X are not equal!

Example B. You test electrical components until you find one that fails. Define Y = number of components tested *before* a failure comes up. What are the values that Y may assume?

Note carefully: The probabilities for the values of Y are not equal! [We'll suppose that P(individual component fails) = 0.1.]

A **random variable** is a rule/formula that assigns a number value to each outcome in a sample space S.

"Random" reminds us that we cannot predict specific outcomes, only discuss the probabilities.

Each value x_i assigned to a discrete random variable X will have an associated probability, $P(X = x_i)$.

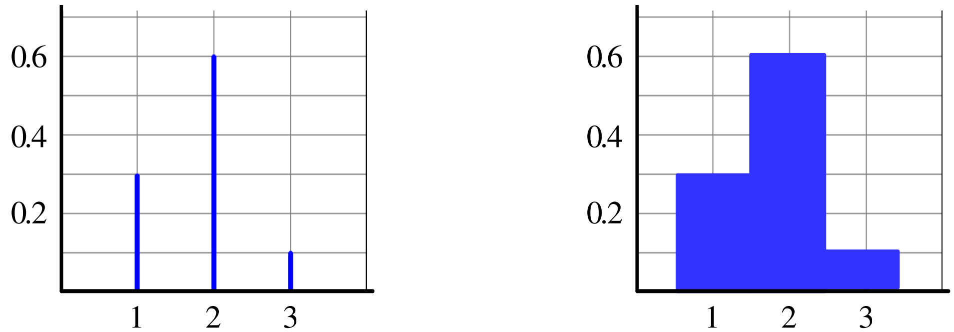
In Example A we defined our random variable as X = number of heads in ten tosses of a coin.
In Example B we defined our random variable as Y = number of components tested before a failure comes up.

Example C. Suppose that a box contains 3 blue blocks and 2 yellow blocks. You pick three blocks without replacement. Define X = number of blue blocks drawn.

What are the values that X may assume?

x			
$P(X = x)$			

The above is a probability distribution table. We could also have drawn either a line graph or a probability histogram.



Each value x_i assigned to a discrete random variable X will have an associated probability, $P(X = x_i)$.

A function $p(x_i) = P(X = x_i)$ is called (what a surprise!) a **probability distribution function** or **probability mass function**. (Some of you may recognize "probability mass function" as being a Physics-type concept.) While the text uses the abbreviations pdf and pmf, I'll try to stick with writing the phrases out.

For Example C, the formal way of writing the probability distribution function would be

$$p(x) = P(X = x) = \begin{cases} \frac{3}{10} & x = 1 \\ \frac{3}{5} & x = 2 \\ \frac{1}{10} & x = 3 \\ 0 & \text{otherwise} \end{cases}.$$

A probability distribution function must meet the basic criteria for all probabilities.

$$0 \le p(x) = P(X = x_i) \le 1 \text{ for each value } x_i \text{ , and } \sum_{\text{all } x_i} P(X = x_i) = 1$$

Why can't the table below give the probability distribution for a random variable X ?

x	0	3	5
$P(X = x)$	$-\frac{1}{5}$	$\frac{3}{5}$	$\frac{3}{5}$

Why can't the table below give the probability distribution for a random variable X ?

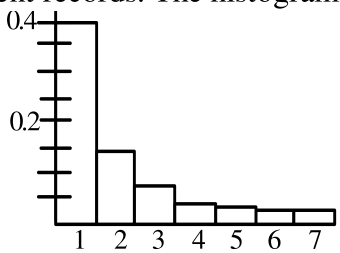
x	-3	2	4
$P(X = x)$	$\frac{3}{7}$	$\frac{2}{7}$	$\frac{4}{7}$

Most of the examples done so far illustrate **finite random variables**. The values X may assume are limited in scope (Examples A and C).

Variables that are (at least in theory) unlimited are **infinite random variables**. (In Example B, $Y = 0, 1, 2, 3, 4, 5, 6, \ldots$)

Example D. Let X = the number of days each ICU patient stays in intensive care. What are the values that X can assume?

The probabilities would be developed based on relative frequencies—observations made from hospital and patient records. The histogram might look something like this.



This histogram is extrapolated from research articles. In the literature, researchers have tried correlating length of stay with diagnosis. For your general knowledge, this probability distribution is approximately exponential, with formula $f(x) = 0.4e^{-0.4x}$. (We'll take a look at exponential distributions in chapter 4.)

Note that, for each probability, $0 \le p(x) = P(X = x) \le 1$.

All of the examples done so far illustrate **discrete random variables**. The values X may assume are distinct and countable.

Measurements such as length or distance would be **continuous random variables**.

Example E. Suppose we measure the heights of 25 people and we define X = height in inches. What are the values that X may assume?

Example E revisited. Suppose we measure the heights of 25 people and we define X = height in inches.
$X \in \{x \mid x > 0\}$

Although height is a continuous random variable, we can (and in practice, do) treat it as a discrete random variable by rounding off to a specified accuracy.

Suppose we measure the heights of 25 people to the nearest inch and get the following results:

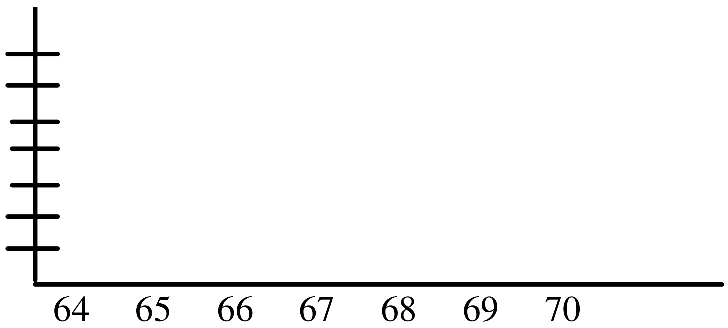
height (in.)	64	65	66	67	68	69	70
frequency	5	7	6	4	1	0	2
probability							

For Example E, the formal way of writing the probability distribution function would be

$$p(x) = \begin{cases} 0.2 & x = 64 \\ 0.28 & x = 65 \\ 0.24 & x = 66 \\ 0.16 & x = 67 \\ 0.04 & x = 68 \\ 0.08 & x = 70 \\ 0 & \text{otherwise} \end{cases} .$$

You can, perhaps, see why another method (table, line graph, histogram) is usually preferred in a case like this.

The probability histogram would look like this:



We can calculate probabilities for various values of X.

a) The probability that X is not more than 66 is

b) The probability that X is at least 68 is

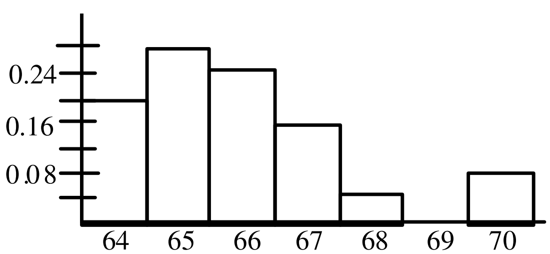
Now is as good a time as any to note that the areas of the rectangles in the probability histogram represent probabilities, and that the sum of the areas of the rectangles = 1.

{area of each bar} = {percentage of people having that height}
= {relative frequency of that height}
= {probability of a person picked at random having that height}.

We can use the idea of "area under the curve" to introduce a **cumulative probability function (text abbreviation cdf)**.

Example E continued. Suppose we measure the heights of 25 people and we define X = height rounded to the nearest inch.

Define a cumulative distribution function $F(x) = P(X \leq x)$.



$F(64) = P(X \leq 64) =$

$F(65) = P(X \leq 65) =$

$F(67) = P(X \leq 67) =$

$F(70) = P(X \leq 70) =$

The formal expression of the cumulative distribution function is as follows.

$$F(x) = \begin{cases} 0, & x < 64 \\ 0.2, & 64 \leq x < 65 \\ 0.48, & 65 \leq x < 66 \\ 0.72, & 66 \leq x < 67 \\ 0.88, & 67 \leq x < 68 \\ 0.92, & 68 \leq x < 70 \\ 1, & 70 \leq x \end{cases}$$

Now for some more challenging questions:

first method: $P(65 \le X \le 68) =$

second method: $P(65 \le X \le 68) =$

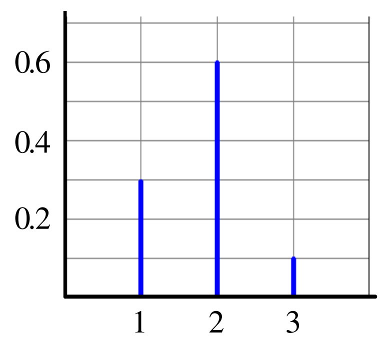
Note that when using the cumulative distribution function is used to find the probability for an interval, we needed to be careful in choosing the values we used. In general,

$$P(a \le X \le b) = F(b) - F(\text{largest possible value for } X \text{ that is strictly less than } a).$$

Example C revisited. Suppose that a box contains 3 blue blocks and 2 yellow blocks. You pick three blocks without replacement. Define X = number of blue blocks drawn.

We have the following probability distribution function and line graph.

$$p(x) = P(X = x) = \begin{cases} \frac{3}{10} & x = 1 \\ \frac{3}{5} & x = 2 \\ \frac{1}{10} & x = 3 \\ 0 & \text{otherwise} \end{cases}.$$



Find the cumulative distribution function and draw its graph.

Example B revisited. You test electrical components until you find one that fails. Define

$$X = \begin{cases} 1 & \text{if the component works (success)} \\ 0 & \text{if the component fails} \end{cases}.$$

This is an example of a **Bernoulli random variable** whose only possible values are 0 and 1.

If we know that P(individual component fails) = 0.1, then the probability density function is

An equivalent way of writing the probability density function is

In general, for a Bernoulli random variable where we specify a parameter α = probability of success,

$$p(x) = \begin{cases} 1-\alpha & \text{if } x = 0 \\ \alpha & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

Now define Y = number of components tested before a failure comes up.

$p(0) = P(Y = 0) =$

$p(1) = P(Y = 1) =$

$p(2) = P(Y = 2) =$

The probability distribution function for this Example would be

$p(y) =$

(Your text has a general version of the probability distribution function for a Bernoulli experiment as part of example 3.12.)

The cumulative distribution function for this Example would be

$F(y) = P(Y \leq y) =$

Perhaps you recognize that this is a geometric series. We could sum the infinite series to show that the sum of the probabilities equals 1.

(Your text has a general version of the cumulative distribution function for a Bernoulli experiment as part of example 3.14.)