5 Probability: the mathematics of randomness

The theory of probability is the mathematician's description of random variation. This chapter introduces enough probability to serve as a minimum background for making formal statistical inferences.

5.1 (Discrete) random variables

The concept of a random variable is introduced in general terms and the special case of discrete data is considered.

5.1.1 Random variables and distributions

It is helpful to think of data values as subject to chance influences. Chance is commonly introduced into the data collection process through

1.

2.

3.

Definition 5.1. A *random variable* is a quantity that (prior to observation) can be thought of as dependent on chance phenomena.

Definition 5.2. A *discrete random variable* is one that has isolated or separated possible values (rather than a continuum of available outcomes).

Definition 5.3. A *continuous random variable* is one that can be idealized as having an entire (continuous) interval of numbers as its set of values.

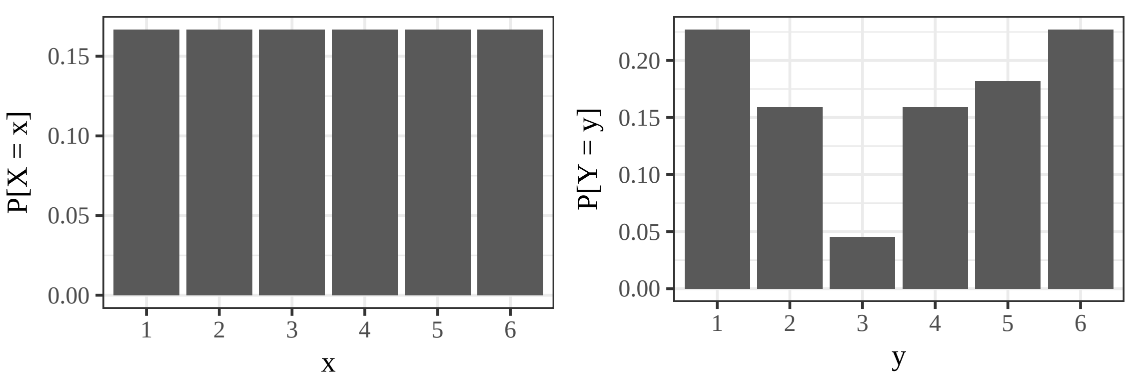
Example 5.1 (Roll of a die)**.**

Definition 5.4. To specify a *probability distribution* for a random variable is to give its set of possible values and (in one way or another) consistently assign numbers between 0 and 1 - called *probabilities* - as measures of the likelihood that the various numerical values will occur

Example 5.2 (Roll of a die, cont'd)**.**

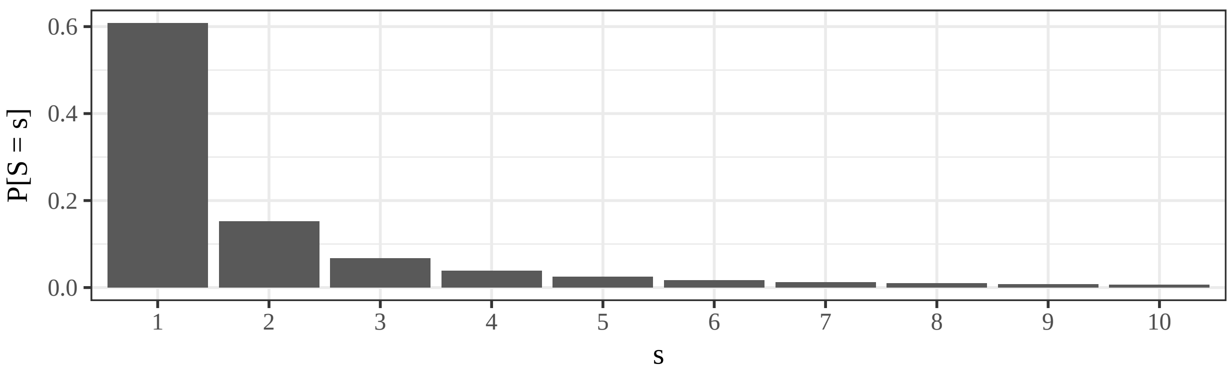
x	1	2	3	4	5	6
$P[X = x]$	1/6	1/6	1/6	1/6	1/6	1/6

y	1	2	3	4	5	6
$P[Y = y]$	5/22	7/44	1/22	7/44	2/11	5/22



Example 5.3 (Shark attacks)**.** Suppose S is the number of provoked shark attacks off FL next year. This has an infinite number of possible values. Here is one possible (made up) distribution:

s	1	2	3	$\cdots$	k	$\cdots$
$P[S=s]$	$\frac{6}{\pi^2}$	$\frac{1}{2^2}\frac{6}{\pi^2}$	$\frac{1}{3^2}\frac{6}{\pi^2}$	$\cdots$	$\frac{1}{k^2}\frac{6}{\pi^2}$	$\cdots$



5.1.2 Probability mass functions and cumulative distribution functions

The tool most often used to describe a discrete probability distribution is the *probability mass function*.

Definition 5.5. A *probability mass function (pmf)* for a discrete random variable X, having possible values $x_1, x_2, \ldots$, is a nonnegative function $f(x)$ with $f(x_1) = P[X = x_1]$, the probability that X takes the value x_1.

Properties of a mathematically valid probability mass function:

1.

2.

A probability mass function $f(x)$ gives probabilities of occurence for individual values. Adding the appropriate values gives probabilities associated with the occurence of multiple values.

Example 5.4 (Torque)**.** Let Z = the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

z	11	12	13	14	15	16	17	18	19	20
$f(z)$	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

Calculate the following probabilities:

$P(Z \leq 14)$

$P(Z > 16)$

$P(Z \text{ is even})$

$P(Z \text{ in } \{15, 16, 18\})$

Another way of specifying a discrete probability distribution is sometimes used.

Definition 5.6. The *cumulative probability distribution (cdf)* for a random variable X is a function $F(x)$ that for each number x gives the probability that X takes that value or a smaller one, $F(x) = P[X \leq x]$.

Since (for discrete distributions) probabilities are calculated by summing values of $f(x)$,

$$F(x) = P[X \leq x] = \sum_{y \leq x} f(y)$$

Properties of a mathematically valid cumulative distribution function:

1.

2.

3.

4.

Example 5.5 (Torque, cont'd)**.** Let Z = the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

z	11	12	13	14	15	16	17	18	19	20
$F(z)$	0.03	0.06	0.09	0.15	0.41	0.50	0.62	0.82	0.97	1

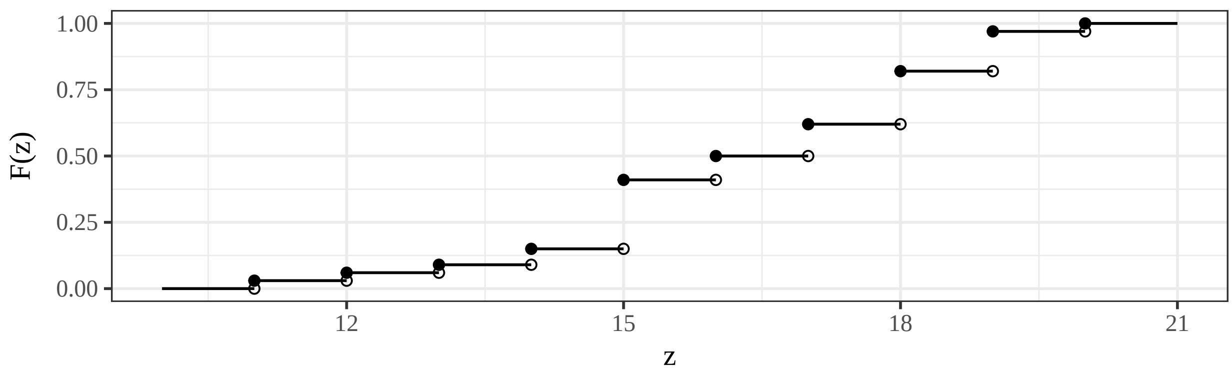


Figure 1: Cdf function for torques.

Calculate the following probabilities using the **cdf only**:

$F(10.7)$

$P(Z \leq 15.5)$

$P(12.1 < Z \leq 14)$

$P(15 \leq Z < 18)$

Example 5.6. Say we have a random variable Q with pmf:

q	$f(q)$
1	0.34
2	0.1
3	0.22
7	0.34

Draw the cdf.

5.1.3 Summaries

Almost all of the devices for describing relative frequency (empirical) distributions in Ch. 3 have versions that can describe (theoretical) probability distributions.

1.

2.

3.

Definition 5.7. The *mean* or *expected value* of a discrete random variable X is

$$\mathrm{E}X = \sum_x x f(x)$$

Example 5.7 (Roll of a die, cont'd). Calculate the expected value of a toss of a fair and unfair die.

x	1	2	3	4	5	6
$P[X = x]$	1/6	1/6	1/6	1/6	1/6	1/6

y	1	2	3	4	5	6
$P[Y = y]$	5/22	7/44	1/22	7/44	2/11	5/22

Example 5.8 (Torque, cont'd)**.** Let Z = the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

z	11	12	13	14	15	16	17	18	19	20
$f(z)$	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

Calculate the expected torque required to loosen the next bolt.

Definition 5.8. The *variance* of a discrete random variable X is

$$\mathrm{Var}X = \sum_x (x - \mathrm{E}X)^2 f(x) = \sum_x x^2 f(x) - (\mathrm{E}X)^2.$$

The *standard deviation* of X is $\sqrt{\mathrm{Var}X}$.

Example 5.9. Say we have a random variable Q with pmf:

q	$f(q)$
1	0.34
2	0.1
3	0.22
7	0.34

Calculate the variance and the standard deviation.

Example 5.10 (Roll of a die, cont'd). Calculate the variance and standard deviation of a roll of a fair die.

5.1.4 Special discrete distributions