

# Random Variables

Stat 241

## Discrete Random Variable

A random variable  $X$  is a **discrete random variable** if the range of  $X$  is finite or an infinite sequence of values. A discrete random variable can be characterized by a function called a **probability mass function (pmf)** which specifies the probability for each possible numerical value of the random variable.

$$p(x) = P(X = x)$$

This function can be specified with a formula or in a table of probabilities. Given the pmf we can calculate the probability any event by adding the probabilities for all of the these values in the event.

**Example 1.** Toss a fair coin 5 times. Let  $X$  = number of heads produced.

- Create a probability table for  $X$ .
- What is  $P(X \text{ is even})$ ?
- What is  $P(X \geq 2)$ ?

## Binomial Random Variables

This example is an instance of an important kind of discrete random variable called a **binomial random variable**. Binomial random variables arise in situations which have the following properties.

- The random process consists of a **predetermined number of trials** (usually denoted  $n$ ).
- Each trial has **two outcomes** (generically called success and failure).
- The **probability of success is the same for each trial** (often denoted  $p$ ).
- Each trial is **independent** of the others.
- The random variable counts the **number of successes** in the  $n$  trials.

Such a variable is denoted by  $X \sim \text{Binom}(n, p)$ .

There is a handy formula for the pmf of a  $\text{Binom}(n, p)$  random variable:

$$p(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

where  $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ .  $\binom{n}{x}$  is read “n choose x” and counts the number of ways to pick a set of  $x$  items from a collection of size  $n$ . This can be computed in R using the `choose()` function.

```
choose(5, 2)
```

```
## [1] 10
```

```
factorial(5) / (factorial(2) * factorial(3))
```

```
## [1] 10
```

**Example 2.** If we roll a fair die 10 times, what is the probability of getting 2 or more sixes?

**Example 3.** Two cards are dealt from an ordinary deck of playing cards. Let  $X$  = number of aces dealt.

- a. Create a probability table for  $X$ .
- b. Is  $X$  a binomial random variable? If so, what are  $n$  and  $p$ ? If not, why not?
- c. What is  $P(X \geq 1)$ ?

**Example 4.** A fair coin is tossed until a head is produced. Let  $X$  = number of tosses.

- a. Create a probability table for  $X$ .
- b. Can you give a formula for the pmf for  $X$ ?
- c. Can you show that the sum of all the probabilities is 1?
- d. What is  $P(X \geq 4)$ ?
- e. What is the probability it takes an even number of tosses to production first head? ( $P(X \text{ is even})$ )

## Continuous Random Variables

A **continuous random variable** is a random variable whose range is an interval of real numbers.

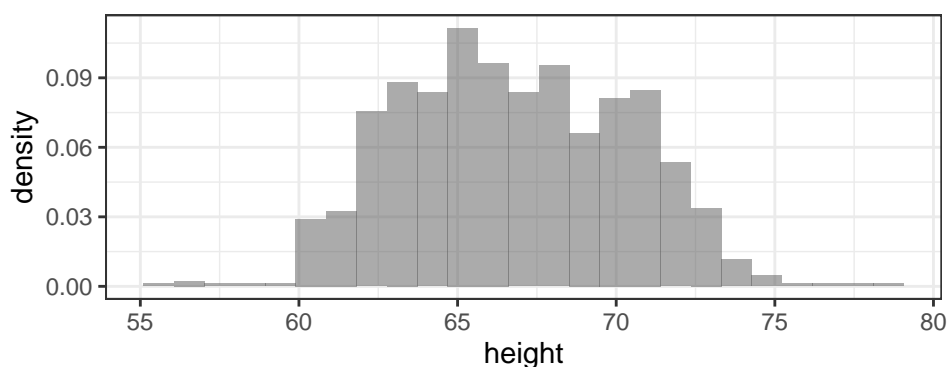
### Examples

- Randomly select an adult male: Let  $X =$  (exact) height in inches
- Randomly select a washer produced by a particular machine. Let  $X =$  (exact) diameter in mm.
- Randomly select a battery and use it to power a device. Let  $X =$  (exact) time until battery can no longer power device.

### Density histograms and density plots of data

Let's consider the heights of the adult children in the Galton data set. Here is a **density histogram**.

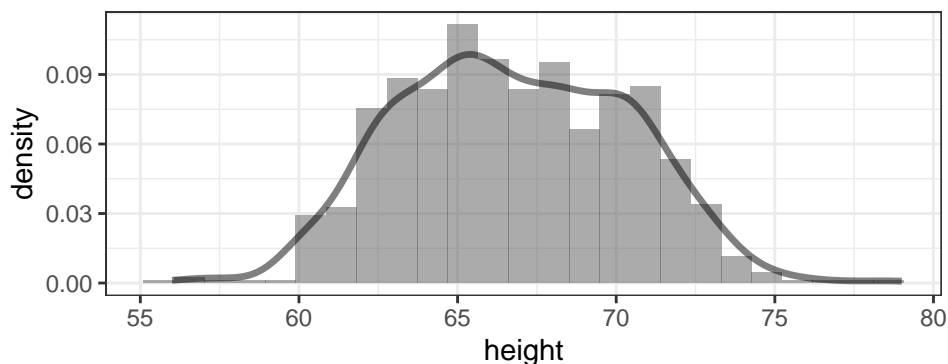
```
gf_dhistogram( ~ height, data = Galton)
```



The density scale is chosen so that \_\_\_\_\_.

Now let's look at a density plot.

```
gf_dhistogram( ~ height, data = Galton) %>%  
gf_dens( ~ height, data = Galton, size = 1.2)
```



This provides a “smooth” version of the histogram and also has the property that

## Density Curves and Density Functions

A continuous random variable is described by a **probability density function (pdf)**. The plot of a pdf will look just like curve in a density plot.

Probability density functions always have two important properties:

- 1.
- 2.

We determine probabilities from a pdf by taking the area under the curve over the region corresponding to our event (ie, by integration).

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

**Example 5.** Let  $X$  be a number randomly chosen from the interval  $[0, 2]$  in such a way that all numbers are equally likely. We call  $[0, 2]$  the **support** of  $X$  is  $[0, 2]$ .

Since no value of  $X$  is more likely to be selected than any other value, the density function  $f(x)$  must be a constant on  $[0, 2]$ .

- a. What is the constant value?
- b. What is  $P(X = 1)$ ?
- c. What is  $P(0 \leq X \leq 1)$ ?
- d. What is  $P(1 \leq X \leq 3/2)$ ?
- e. What is  $P(1 \leq X \leq 3)$ ?

A random variable  $X$  whose pdf is constant (where it is non-zero) is said to have a **uniform distribution**. We will denote this as  $X \sim \text{Unif}(a, b)$ , where  $a$  and  $b$  are the upper and lower limits of the support.

**Example 6.** Let  $f(x) = x/2$  for  $x \in (0, 2)$  (and 0 elsewhere). We can write this as  $f(x) = x/2 \cdot \mathbb{I}[x \in (0, 2)]$  or  $f(x) = x/2$  on  $(0, 2)$ .

- a. Verify that  $f(x)$  is a probability density function.
- b. Compute  $P(1 \leq X \leq 3/2)$ .
- c. Compute  $P(1 \leq X \leq 4)$ .

### The Cumulative Distribution Function (cdf) for a continuous random variable

The **cdf** for  $X$  is defined by

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

**Example 7.** Let  $X$  have a uniform distribution on  $[0, 4]$ .

- a. pdf:  $f(x) =$

- b. cdf:  $F(x) =$

**Example 8.** Let  $X$  be the random variable whose pdf is  $f(x) = x/2 \cdot \mathbb{I}[x \in (0, 2)]$ . Find the cdf for  $X$ .

### Using the cdf to compute probabilities

If  $F(x)$  is the cdf for the random variable  $X$ , then  $P(a \leq X \leq b) =$  \_\_\_\_\_

**Example 9.** Let  $F(x) = x^2$  on  $[0, 1]$

- a. What is  $F(x)$  when  $x < 0$ ?
- b. What is  $F(x)$  when  $x > 1$ ?
- c. What is  $P(X \leq 1/2)$ ?
- d. What is  $P(1/2 \leq X \leq 3/4)$ ?
- e. What is  $P(-2 \leq X \leq 1/2)$ ?
- f. What is the pdf for  $X$ ?

**Example 10.** Let  $f(x) = e^{-x}$  on  $(0, \infty)$  (and 0 elsewhere).

- a. Show that  $f$  is a pdf.
- b. Let  $X$  be the corresponding random variable. What is  $P(X > 2)$ ?
- c. What is the cdf for  $X$ ?
- d. Use the cdf to find  $P(X < 1)$ .

## More Practice

1. Parts coming off an assembly line have a 1% chance of being defective. If 3 parts are randomly chosen from this line and  $X$  is the number of defective parts.
  - a. Compute the probability function  $p(x)$  for  $X$ .
  - b. What is the probability that at least one of the three is defective?
2. Parts coming off an assembly line have a 1% chance of being defective. All of the parts coming off the line are inspected. Let  $X$  be the number parts inspected up to and including the first defective part.
  - a. Is  $X$  continuous or discrete?
  - b. What is the support of  $X$ ?
  - c. Find the probability mass function  $p(x)$  for  $X$ .
  - d. What is the probability that the first defective part is the 100th part?
3. A biased coin has a 40% chance of producing a head. If it is tossed 10 times,
  - a. What is the probability of getting exactly 3 heads?
  - b. What is the probability of getting 3 or more heads. (This can be calculated in two different ways. The easier way uses the complement rule.)
4.
  - a. Find the value of  $C$  for which the function  $f(x)Cx^2 \cdot \mathbb{I}_{x \in (0, 2)}$  is a pdf.
  - b. Use the pdf to find  $P(0 \leq X \leq 1)$  and  $P(1 \leq X \leq 5)$ .
  - c. Use the pdf to compute the cumulative distribution function  $F(x)$ .
5.
  - a. Find the value of  $C$  for which the function  $f(x) = \frac{C}{x^2} \cdot \mathbb{I}_{x \geq 1}$  is a pdf.
  - b. What is  $P(X \leq 2)$ ?
  - c. What is  $P(X > 3)$ ?
  - d. Find the cumulative distribution function  $F(x)$ .
  - e. Use the cumulative distribution function to find  $P(2 \leq X \leq 5)$ .