Chapter 5

Tuesday, October 10, 2023 7:01 PM

Chapter 5

Discrete Distributions

Review of last week

Equally Likely Rule: Suppose S contains equally likely outcomes. Then

$$P(E) = \frac{|E|}{|S|}.$$

- Rules for counting: don't skip, don't double count
- **Multiplication principle**: one stage at a time
- **Binomial coefficient**: $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$ is the number of ways to choose k items from n items.

Review of last week

Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

- **Chain rule for probabilities**

$$P(A \cap B) = \boxed{P(B) \cdot P(A|B) = P(A) \cdot P(B|A).}$$

- **Bayes' rule for inverse probabilities**

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

- **Independent events** → knowing A has happened makes no diff to whether B happens ⇒ $P(A|B) = P(A)$

$$P(A \cap B) = P(A) \times P(B).$$

↳ $P(A_1 \cap A_2 \ldots \cap A_k) = P(A_1) \times P(A_2) \times \ldots \times P(A_k)$ if independent

Random Variable

A random variable is a number that is obtained as or from the result of a random experiment.

- Say we flip a coin 3 times, then our sample space has $2^3 = 8$ possible outcomes:

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

  Typically, the particular sequence of heads or tails is of little interest; what does matter is the number of heads that result.

- If we define X = number of heads in 3 tosses, we have captured the essence of the problem. We call X a random variable.

- Note: X defines a mapping (a function) from the original sample space S to a set of numbers.

$$X(HHH) = 3, X(HHT) = 2, X(HTH) = 2, X(HTT) = 1$$
$$X(THH) = 2, X(THT) = 1, X(TTH) = 1, X(TTT) = 0$$

Can have multiple random variables of interest

ex: Y = # heads − # tails

$Y(HHH) = 3$, $Y(HHT) = 1$, ... , $Y(TTT) = -3$

Random variable

Definition 5.1 Let S be a sample space associated with a random experiment, $\mathbb{R}$ is the real line and let

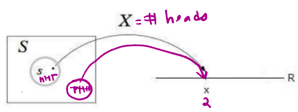
$$X : S \to \mathbb{R}.$$

Then X is called a **random variable** and

$$P(X = x) = P(s \in S : X(s) = x).$$

random var. as a f^n

number



Ex: X = # heads in 3 tosses of a fair coins (equally likely rule)

$P(X = 3) = P(HHH) = \frac{1}{8}$

$P(X = 0) = P(TTT) = \frac{1}{8}$

$P(X = 1) = P((HTT) \cup (THT) \cup (TTH))$
$= 3/8$

$P(X = 2) = P((HHT) \cup (THH) \cup (HTH))$
$= 3/8$

Note: $P(X=0) + P(X=1) + P(X=2) + P(X=3) = 1$

In defining a random variable, we have also created a new sample space (the range of the random variable).

Random variables often create a dramatically simpler sample space.

They also allow us to describe certain kinds of events very succinctly. In the coin rolling example, we can now write $P(X = 2)$ instead of P(there are two heads and one tail).

Example 5.1

Independent trials consisting of the flipping of a coin having probability $\frac{1}{3}$ of coming up heads are continually performed until a head occurs. The sample space is

$$S = \{(H); (T, H); (T, T, H); (T, T, T, H) \dots\}$$

Suppose we define the random variable X as the number of tosses for the first head to appear.

a. What is the *range*(X)?

$X(H) = 1$, $X(TH) = 2, \dots$

$\text{range}(X) = \{1, 2, 3, \dots\}$

Example 5.1

Independent trials consisting of the flipping of a coin having probability $\frac{1}{3}$ of coming up heads are continually performed until a head occurs. The sample space is

$$S = \{(H); (T, H); (T, T, H); (T, T, T, H) \dots\}$$

Suppose we define the random variable X as the number of tosses for the first head to appear.

b. Express the following event in random variable notation and calculate its probability: [at least three tosses must be made for the first head to be observed.]

E = at least 3 tosses for 1^{st} head

$= \{(TTH), (TTTH), \dots\}$

$= X \geq 3$

$P(X \geq 3) = P(E) = 1 - P(E^c) = 1 - 5/9$

$E^c = \{H, TH\}$

$P(E^c) = P(H \cup (TH))$

$= P(H) + P(TH)$ by (axiom 3)

$= \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{3}$ by (independence of tosses)

$= \frac{5}{9}$

$$\boxed{P(X \geq 3) = \frac{4}{9}}$$

We will be concerned with two main types of random variables: discrete and continuous.

A **discrete** random variable is one that can only take on a finite or countably infinite set of values.

- X, the number of heads in 3 flips of a coin is clearly a discrete random variable since $range(X) = \{0, 1, 2, 3\}$.

A **continuous** random variable can take on all the values in an interval.

- Y, the life length of a randomly selected bulb is theoretically a continuous random variable since $range(Y) = [0, \infty)$.

Example 5.2

In each case, state whether the random variable is discrete or continuous.

a. The distance traveled by a football when thrown. continuous.
   range (X) = [0, ∞)

b. Toss a coin repeatedly until the first head appears and record the number of tails. discrete
   Y = # tails preceding 1st head
   range(Y) = {0, 1, 2, . . . }

Probability Mass Function

One useful way to describe the distribution of a discrete random variable, especially one with a finite range, is by way of a table.

X	Number of heads in 3 tosses			
x_i	0	1	2	3
probability $f(x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$\sum_{i=1}^{4} f(x_i) = 1$

This is an example of a **Probability Mass Function** or PMF. As the name suggests, the PMF is associated with "point probabilities". Note that the table only shows the values which have positive probability.

$f(0) = 1/8$, $f(1) = 3/8$

Probability Mass Function

Definition 5.2 For a discrete random variable X, we define the **Probability Mass Function (PMF)** $f(x)$ by:

$$f_X(x) = P(X = x), \ \forall\, x.$$

We will write f_X when we want to emphasize the random variable.

The PMF $f(x)$ of a discrete random variable can be positive for at most a countable number of values. That is, if X can take values x_1, x_2,x_3, ... then

$$\begin{aligned} f(x_i) &> 0, \ i = 1, 2, 3, \ldots \\ &= 0 \text{ otherwise.} \end{aligned}$$

Furthermore, since X must take one of the values x_i, we have

$$\sum_{i=1}^{\infty} f(x_i) = 1.$$

This is useful in the following ex:

x_i	3	4
$f(x_i)$	$\frac{c}{2}$	c

where c is unknown. Find c

X takes values $x_1 = 3$ and $x_2 = 4$

$f(x_i) > 0 \quad \forall i$

$\Rightarrow c > 0$

Also: $\sum_{i=1}^{2} f(x_i) = 1 \Rightarrow f(x_1) + f(x_2) = 1$

$\Rightarrow \frac{c}{2} + c = 1$

$\Rightarrow \boxed{c = \frac{2}{3}}$

Example 5.3

A store manager receives a shipment of 30 microwave ovens, 5 of which are (unknown to the manager) defective. The store manager takes a random sample of 4 ovens from the shipment and tests them to see if they are defective. Let X denote the number of defective ovens found. Write the PMF of X in a tabular format. (You may assume that every sample of 4 ovens is equally likely to be selected.)

Expt

D D D D D N ... N
1 2 3 4 5 6 ... 30

sample of 4 ovens

ex: D D N N

X = # defectives in sample

range$(X) = \{0, 1, 2, 3, 4\}$

S = all possible samples of size 4 from 30

$|S| = \binom{30}{4}$ all equally likely

$P(X=0) = P(E_0) = .4616$

Define E_0 = no defectives in sample (all 4 are chosen from non-defectives)

$P(E_0) = \frac{|E_0|}{|S|}$ (equally likely rule)

$P(X=0) = \frac{\binom{5}{0}\binom{25}{4}}{\binom{30}{4}} = .4616$

$P(X=1) = \frac{\binom{5}{1}\cdot\binom{25}{3}}{\binom{30}{4}} = .4196$

$P(X=2) = \frac{\binom{5}{2}\cdot\binom{25}{2}}{\binom{30}{4}} = .1095$

x	0	1	2	3	4
$f(x)$	.4616	.4196	.1095	.0091	.0002

Formula

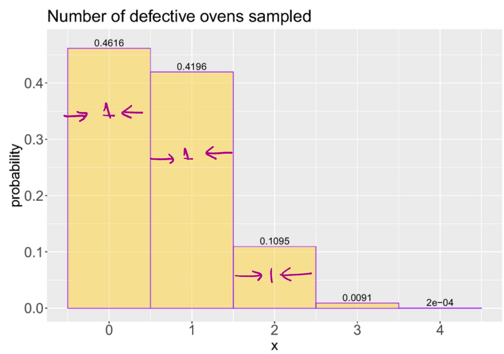
$f(x) = \frac{\binom{5}{x}\cdot\binom{25}{4-x}}{\binom{30}{4}}$ $x = 0, 1, 2, 3, 4$

$$P(X=2) = \frac{\binom{5}{2}\cdot\binom{25}{2}}{\binom{30}{4}} = .1095 \qquad f(x) = \frac{\binom{5}{x}\cdot\binom{25}{4-x}}{\binom{30}{4}} \qquad x = 0,1,2,3,4$$

$$P(X=3) = \frac{\binom{5}{3}\cdot\binom{25}{1}}{\binom{30}{4}} = .0091$$

$$P(X=4) = \binom{5}{4}\cdot\binom{25}{0} \Big/ \binom{30}{4} = .0002$$

Probability Histogram



Probability histogram: code

```
# create data frame consisting of xi and f(xi)
library(tidyverse) → umbrella packages
ovens <- data.frame(
  x = 0:4,
  f = c(0.4616, 0.4196, 0.1095, 0.0091, 0.0002)
)

# make probability histogram
ggplot(data = ovens, mapping = aes(x = x, y = f)) +
  geom_col(
    width = 1, alpha = 0.5,
    fill = "gold", color = "purple"
  ) +
  geom_text(mapping = aes(label = round(f, 4), y = f + 0.01)) +   optional
  labs(     → necessary
    x = "x",
    y = "probability",
    title = "Number of defective ovens sampled"
  )
```

x	f
0	.4616
1	
2	
3	
4	

For random variables with infinitely many possible values, a formula provides the most concise representation of the PMF.

Example 5.4

Let X denote the number of tosses until the first head when tossing a fair coin. Find the P.M.F. of X. You may assume the outcome on one toss is independent of the outcome on a different toss.

$S = \{H, TH, TTH, \dots\}$

$\text{range}(X) = \{1, 2, 3, \dots\}$

$P(X=1) = \frac{1}{2}$

$P(X=2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2^2}$

$P(X=3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2^3}$

$f(x) = P(X=x) = \left(\frac{1}{2}\right)^x, \quad x = 1, 2, 3, \dots$

variation: $P(H) = 1/3$

$f(x) = P(X=x) = \left(\frac{2}{3}\right)^{x-1} \cdot \frac{1}{3} \quad x = 1, 2, 3, \dots$

$P(X=1) = P(H) = \frac{1}{3}$

$P(X=2) = P(TH) = P(T) \cdot P(H) = \frac{2}{3} \cdot \frac{1}{3}$

$P(X=3) = P(TTH) = \left(\frac{2}{3}\right)^2 \cdot \frac{1}{3}$

Cumulative Distribution Function

There is yet one more important way to describe the distribution of a discrete random variable, with a **cumulative distribution function (CDF)**

Definition 5.3 The **Cumulative Distribution Function** F of a random variable X is defined by

$$F(x) \quad = \quad P(X \leq x), \quad \forall\, x.$$

Probability Mass Function: $P(X=x) = f(x)$

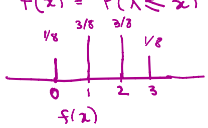
PMF for number of heads in 3 tosses

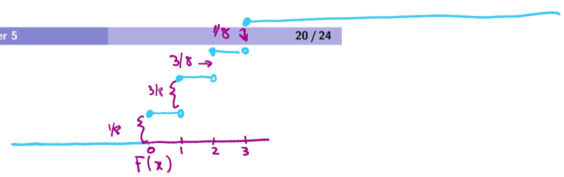
X	Number of heads in 3 tosses			
x	0	1	2	3
probability	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

CDF for number of heads in 3 tosses

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \begin{cases} 0 & x < 0 \\ \frac{1}{8} & 0 \leq x < 1, \\ \frac{4}{8} & 1 \leq x < 2 \\ \frac{7}{8} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases} \end{aligned}$$

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





$x<0$: ex: $F(-1) = P(X \leq -1) = 0$

$F(x) = 0 \qquad x < 0$

$x=0$: $F(0) = P(X \leq 0) = P(X=0) = 1/8$ $\Big\}$ $F(x) = \frac{1}{8} \qquad 0 \leq x < 1$

$F(1/2) = P(X \leq 1/2) = P(X=0) = 1/8$

$F(x) = 0 \qquad x < 0$

$x = 0: \quad F(0) = P(X \le 0) = P(X=0) = 1/8$

$0 < x < 1: \quad \text{ex: } F(1/2) = P(X \le 1/2) = P(X=0) = 1/8$

$\therefore \quad F(x) = 1/8$

$\Big\} \; F(x) = \frac{1}{8} \qquad 0 \le x < 1$

$x = 1: \quad F(1) = P(X \le 1) = P(X=0) + P(X=1) = 1/8 + 3/8 = 1/2$

$1 < x < 2: \quad \text{ex: } F(3/2) = P(X \le 3/2) = P(X=0) + P(X=1) = 1/8 + 3/8 = 1/2$

$\Big\} \; F(x) = \frac{1}{2} \qquad 1 \le x < 2$

$x = 2: \quad F(2) = P(X \le 2) = P(X=0) + P(X=1) + P(X=2) = 1/8 + 3/8 + 3/8 = 7/8$

$2 < x < 3: \quad \text{ex: } F(2\tfrac{1}{2}) = P(X \le 2\tfrac{1}{2}) = P(X=0) + P(X=1) + P(X=2) = 1/8 + 3/8 + 3/8 = 7/8$

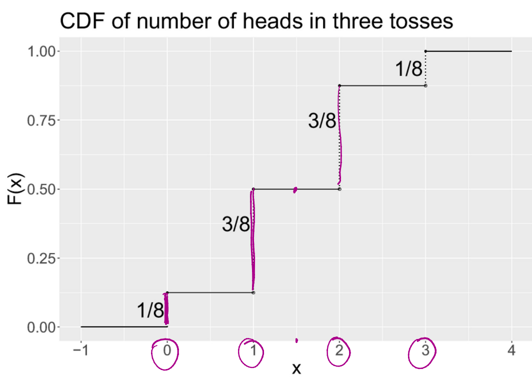
$\Big\} \; F(x) = \frac{7}{8} \qquad 2 \le x < 3$

$x = 3: \quad F(3) = P(X \le 3) = 1$

$x > 3: \quad \text{ex } F(5) = P(X \le 5) = 1$

$\Big\} \; F(x) = 1 \qquad x \ge 3$

The CDF for the number of heads in 3 tosses is graphed below:



What is the PMF for this CDF?

$f(x) =$	$1/8$	$x = 0$
	$3/8$	$x = 1$
	$3/8$	$x = 2$
	$1/8$	$x = 3$

- The graph of the CDF of a discrete random variable is a step function.
- The step function is non decreasing, has jumps at each of the possible values x and the size of the jump is equal to $P(X = x)$.
- Note that, at the jump point F takes the value at the top of the jump. This is known as *right continuity*.

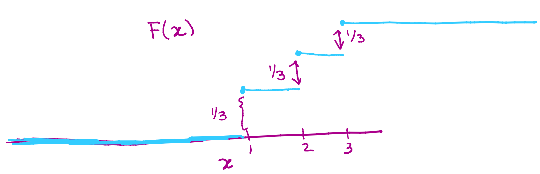
- as $x \to \infty$, $F(x) = P(X \leq x) \longrightarrow 1$
- as $x \to -\infty$ $F(x) = P(X \leq x) \longrightarrow 0$

Example 5.5

A discrete uniform random variable X has a PMF of the form

$$f(x) = \frac{1}{n}, \quad x = 1, 2, \ldots, n. \rightarrow \boxed{n=3}$$

Find the CDF of X.



$x<1$: $F(x) = P(X \leq x) = 0$

$x=1$: $F(1) = P(X \leq 1) = P(X=1) = 1/3$

$F(x) = \frac{1}{3}$ $\quad 1 \leq x < 2$

$1<x<2$: $F(x) = P(X \leq x) = P(X=1) = 1/3$

$x=2$: $F(2) = P(X \leq 2) = P(X=1) + P(X=2) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$

$F(x) = \frac{2}{3}$ $\quad 2 \leq x < 3$

$2<x<3$: $F(x) = P(X \leq x) = P(X=1) + P(X=2) = 2/3$

$F(x) = P(X \leq 3) = 1$ $\quad F(x) = 1$ $\quad 3 \leq x$

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{3} & 1 \leq x < 2 \\ \frac{2}{3} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

$2 < x < 3$: $F(x) = P(X \le x) = P(X=1) + P(X=2)$
$= 2/3$

$x = 3$: $F(3) = P(X \le 3) = 1$
$x > 3$: $F(x) = 1$
$\Big\}$ $F(x) = 1$ $3 \le x$

CDF and PMF

There is of course a connection between the PMF and CDF of a given random variable:

- To get the CDF from the ~~PDF~~ PMF, we simply add up the probabilities for all possible values up to and including x.
- To get the PMF from the CDF, we look at how much the CDF has changed from the last jump.