4.4 Random Variables

A random variable is a function from ~~the~~ a sample space S (of a random experiment) to a set T (for our purposes usually T will be the set of real numbers)

Comments

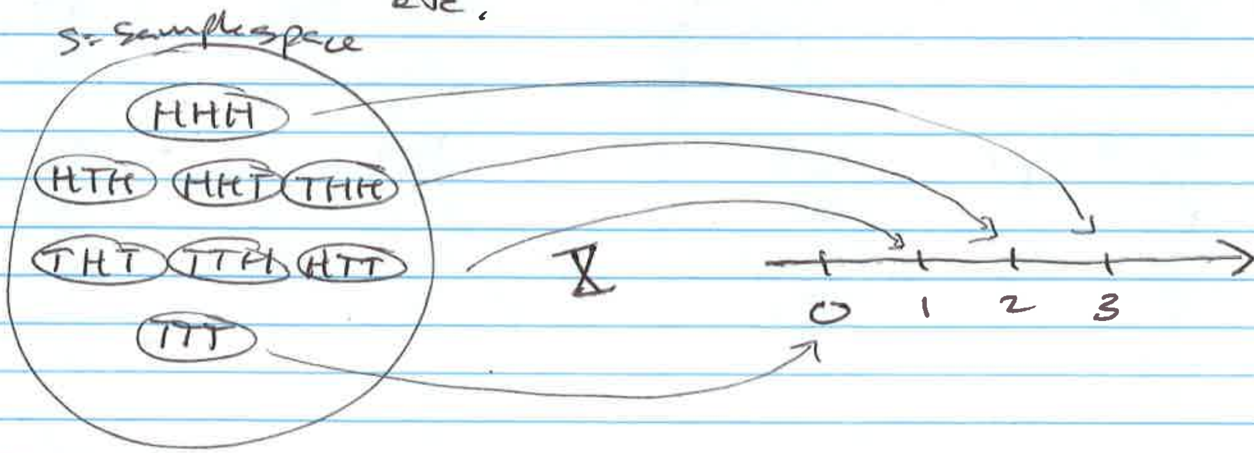
- typically a random variable is a/the variable of interest in the context of a random experiment
- the random variable is "random" in that its value depends on the outcome of the experiment

EXAMPLE (1a from book, p. 112)

Flip a coin 3 times. Let X be the number of heads that appear. So

$$X(HHH) = 3$$
$$X(HHT) = 2$$
$$X(HTH) = 2$$

etc.



In terms of probabilities

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

$$P\{X=1\} = P(\{THT, TTH, HTT\}) = \frac{3}{8}$$

$$P\{X=2\} = P(\{HTH, HHT, THH\}) = 3/8$$

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

In general

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

= Probability of outcome s in S where the value of the random variable X for s is equal to x.

EXAMPLE

A group of 100 people in Chicago (this is the sample space S). contains 42 who ~~[illegible]~~ prefer the Cubs, 35 who prefer the White Sox, and 23 who prefer other teams (or don't care). A person is selected from this group. Define the random variable

$$X = \begin{cases} 0 & \text{if the person prefers the Cubs} \\ 1 & \text{if the person prefers the White Sox} \\ 2 & \text{if the person prefers other teams or does not care.} \end{cases}$$

Then

$$P\{X=0\} = P(\{s \in S : \overbrace{s \text{ prefers Cubs}}^{X(s)=0}\}) = \frac{42}{100}$$

$$P\{X=1\} = P(\{s \in S : s \text{ prefers White Sox}\}) = \frac{35}{100}$$

$$P\{X=2\} = P(\{s \in S : s \text{ prefers other team...}\}) = \frac{23}{100}$$

<u>EX</u> Poker Hands

- Deal 5 cards sequentially without replacement from a well-shuffled deck of 52 cards (standard deck)

sample space

$$S = \{(x_1, x_2, x_3, x_4, x_5) : x_i \in D,\ x_i \neq x_j \text{ for } i \neq j\}$$

where D = set of 52 standard playing cards (i.e. the deck)

Let $\vec{X} = (X_1, X_2, X_3, X_4, X_5)$ – the sequence of cards dealt.

where $X_i \in D$ is the i^{th} card dealt.

We can think of this as a random variable, but a more relevant one is

$$W = \{X_1, X_2, X_3, X_4, X_5\}$$ = unordered set of 5 cards

We can think of this random variable W (i.e. a poker hand) that takes values in the set of all possible poker hands

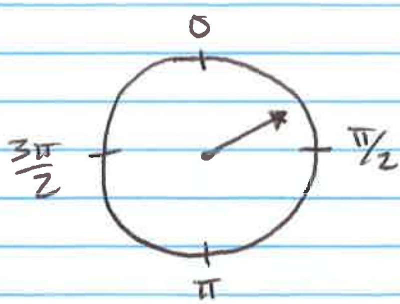
$$T = \{\{x_1, x_2, x_3, x_4, x_5\} : x_i \in D,\ x_i \neq x_j \text{ for } i \neq j\}$$

i.e. we could assign real number value to different hands

– i.e. one hand is more valuable than another hand

EXAMPLE Spinners

Z = random variable



Let Z be the place where the spinner stops. Z can take on any real number in the interval $[0, 2\pi)$

Let Z be the quadrant where the spinner stops. Z can take on one of 4 values 0, 1, 2, or 3.

4.2 Discrete Random Variables

A random variable that can take on at most a countable number of possible values is said to be <u>discrete</u>.

(see Ch. 5 for continuous random variables)

For a discrete random variable X, define the <u>probability mass function</u>, $p(a)$ of X by

$$p(a) = P\{X = a\}$$

(see text, p. 116)

EXAMPLE

Flip a fair coin until H (heads) occurs. Let X be the number of times the coin is flipped. So X is a random variable whose value is the number of times Tails appears before H, plus 1. That is

X: random variable, function of experimental outcome.

$$X(TTTH) = 4$$
$$X(TH) = 2$$
$$X(TTTTTTH) = 7$$

The range of X (remember X is a function) is the set of positive integers

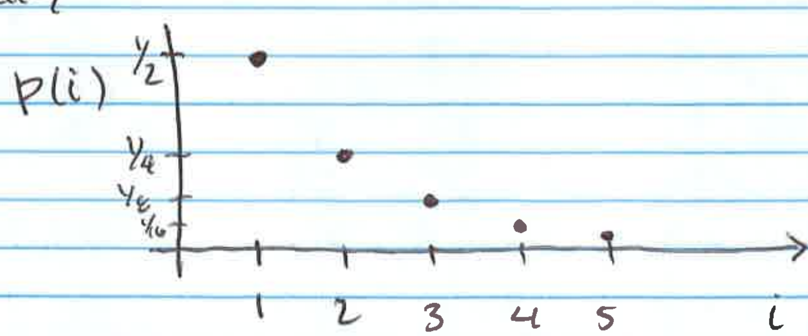
The probability mass function is, for $i = 1, 2, 3, \ldots$

$$p(i) = P\{X = i\} = \underbrace{\left(\tfrac{1}{2}\right)^{i-1}}_{i-1 \text{ tails}} \cdot \underbrace{\left(\tfrac{1}{2}\right)}_{H} = \left(\tfrac{1}{2}\right)^{i}$$

Note: $$\sum_{i=1}^{\infty} p(i) = \sum_{i=1}^{\infty} \left(\tfrac{1}{2}\right)^{i} = 1$$

→ e.g. geom. series

Graphically

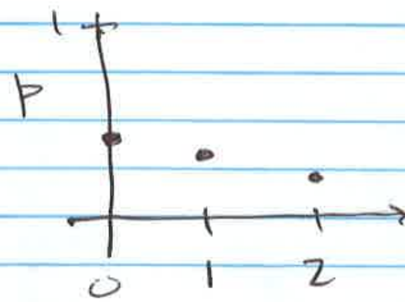


EXAMPLE (Chicago Baseball)

Here the random variable X has range $\{0,1,2\}$

The probability mass function has

$$p(0) = P\{X=0\} = \frac{42}{100}$$
$$p(1) = P\{X=1\} = \frac{35}{100}$$
$$p(2) = P\{X=2\} = \frac{23}{100}$$



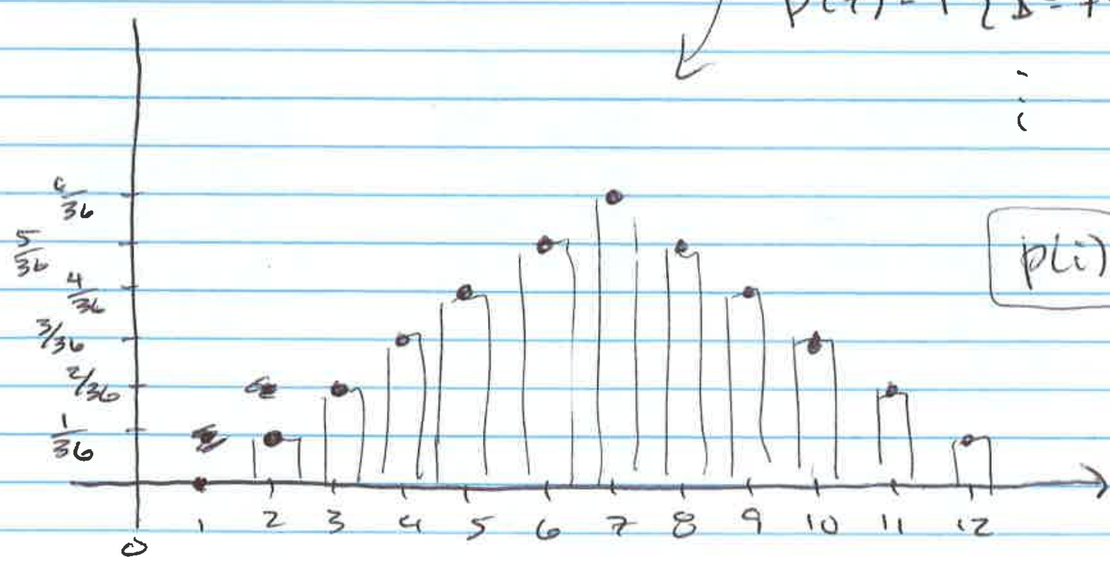
EXAMPLE (Roll 2 dice)

Let X = sum of two rolled dice. X = random variable taking on integers in $\{2,3,4,\ldots,12\}$

(e.g. $X((1,6)) = 7$
$X((5,2)) = 7$
$X((1,4)) = 5$

The probability mass function is

$$p(2) = P\{X=2\} = \frac{1}{36}$$
$$p(7) = P\{X=7\} = \frac{6}{36}$$
$$\vdots$$



$\boxed{p(i) \text{ for } i=2,\ldots,12}$

See also Fig. 4.2, p. 117

Comments:

- For a discrete random variable X, the probability mass function $p(a) = P\{X=a\}$ is positive for at most a countable number of values of a.

- The probability mass function has the property (see also, p. 117)

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

where x_i are values taken on by X

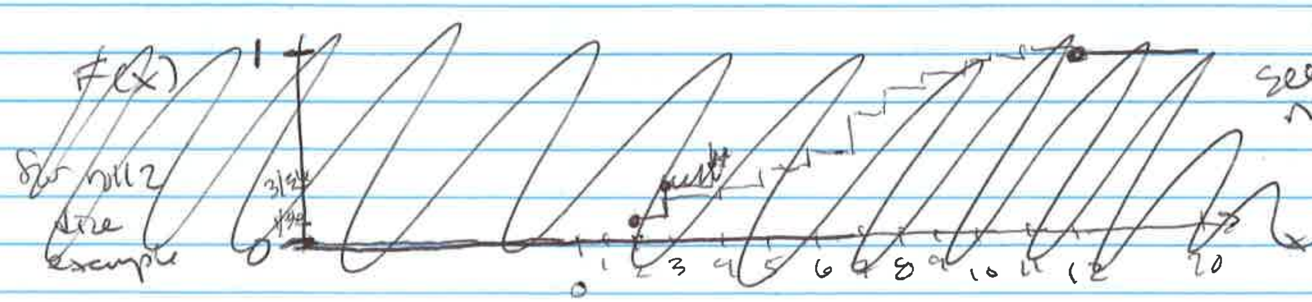
  - this may simplify to a finite sum if there are a finite # of values x_i that may be taken by X.

Def: Cumulative Distribution Function

For any random variable X, the cumulative distribution function, F, is defined by

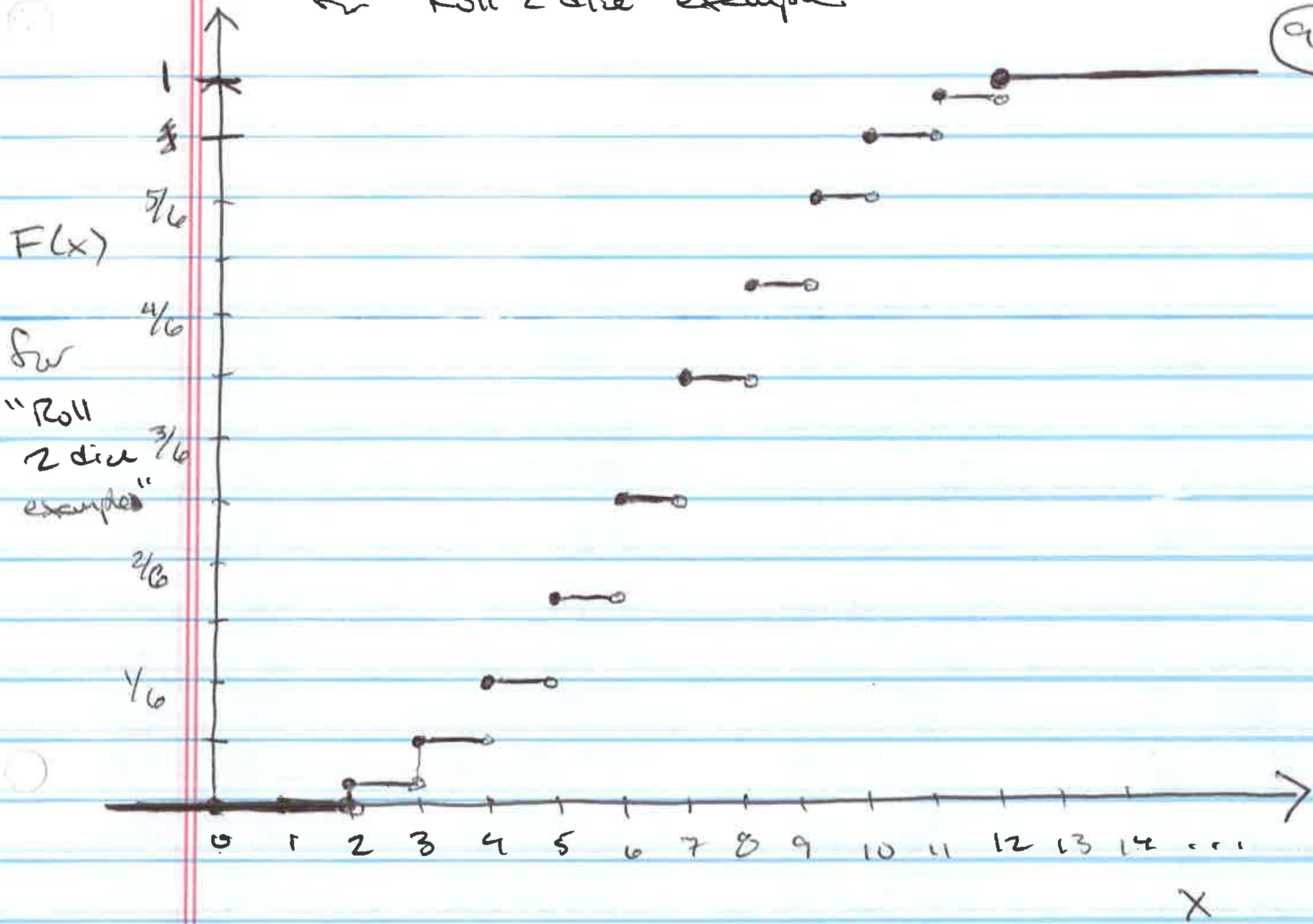
$$F(x) = P\{X \le x\} \qquad -\infty < x < \infty$$

Also called the "distribution function". This function specifies the probability that the random variable X (function of outcomes in sample space) is less than or equal to the value x.



see next page

EXAMPLE Cumulative Distribution Function, $F(x)$, for "Roll 2 dice" example



$F(2) = P\{X \le 2\} = 1/36$

$F(3) = P\{X \le 3\} = 3/36$

$F(4) = P\{X \le 4\} = 6/36$

$F(5) = P\{X \le 5\} = \frac{10}{36}$

$F(6) = P\{X \le 6\} = \frac{15}{36}$

$F(7) = P\{X \le 7\} = \frac{21}{36}$

$F(8) = P\{X \le 8\} = \frac{26}{36}$

$F(9) = P\{X \le 9\} = \frac{30}{36}$

$F(10) = P\{X \le 10\} = \frac{33}{36}$

$F(11) = P\{X \le 11\} = \frac{35}{36}$

$F(12) = P\{X \le 12\} = \frac{36}{36} = 1$

$F(x \ge 12) = 1$

EXAMPLE

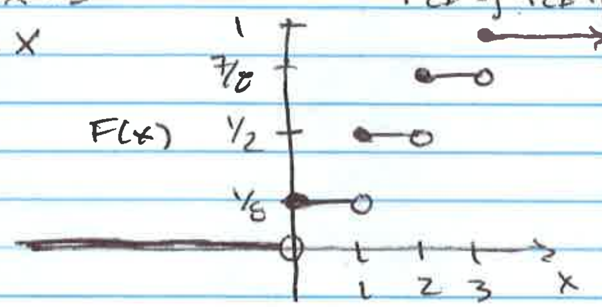
Flip coin 3 times. X = # of heads that appear.

Recall
$$P\{X=0\} = \tfrac{1}{8}$$
$$P\{X=1\} = 3/8$$
$$P\{X=2\} = 3/8$$
$$P\{X=3\} = 1/8$$

$$F(x) = P\{X < x\}$$

$$F(x) = \begin{cases} 0 & x<0 \\ 1/8 & 0 \le x < 1 \\ 1/8 + 3/8 & 1 \le x < 2 \\ 1/8 + 3/8 + 3/8 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

e.g. $P\{X<2\} = P\{X=0 \text{ or } X=1\}$
$= P\{X=0\} + P\{X=1\}$



Note:

- In general for $a<b$

$$P\{X \le b\} = P\{X \le a\} + P\{a < X \le b\}$$

and since $P\{a < X \le b\} \ge 0$ it follows that

$$P\{X \le b\} \ge P\{X \le a\}$$

That is, if $a<b$ then $F(b) \ge F(a)$ so

the cumulative distribution function is a nondecreasing function