Last time we began our discussion on probability and introduced the concept of a probability distribution function

eg Consider a baseball player who has probability $P(H)=p$ of getting a hit at every at bat. Let X = number of at bats need to get first hit. Assuming each at bat is independent what is the distribution function?

H – Hit
H^c – Hitless

$$p(1) = P(X=1) = P(H) = p$$

$$p(2) = P(X=2) = P(H^cH) = P(H^c)P(H) = (1-p)p$$

$$p(3) = P(X=3) = P(H^cH^cH) = P(H^c)P(H^c)P(H) = (1-p)^2 p$$

$$\vdots$$

general formula

$$p(x) = \begin{cases} (1-p)^{x-1} p & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

we call p a <u>parameter</u> of the distribution the formula $(1-p)^{x-1}p$ for all possible values of p is called a family of probability distributions

Cumulative Distribution Function

The cumulative distribution function (cdf) $F(x)$ of a discrete rv X with pdf $p(x)$ is defined for every number x as

$$F(x) = P(X \le x) = \sum_{y:y \le x} p(y)$$

these y are the values the X may take.

Note that $F(x) \to 1$ as $x \to \infty$
$F(x) \to 0$ as $x \to -\infty$

Also $a < b \Rightarrow F(a) < F(b)$

The CDF is the probability that the observed value of X will be less than x

eg roll two fair dice. Let X = "sum of two faces showing"

pdf

$X = x$	2	3	4	5	6	7	8	9	10	11	12
$p(x) = P(X=x)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

you can get these by counting outcomes.

$X=x$	0	2	3	4	5	6	7	8	9	10	11	12	13
$F(x)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1	1 ...

Note that $F(x)$ is defined for all possible values of x (ie beyond just the possible values of X). Compare this with $p(x)$ which equals 0 for all values of x that are not possible values of X

eg for above example

$$p(5.5) = 0 \qquad F(5.5) = P(X \leq 5.5)$$
$$= P(X=2) + P(X=3) + P(X=4) + P(X=5)$$
$$= 10/36$$

Of course we can do this with a parameterized distribution as well

eg the baseball player $p(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$

means largest integer $\leq x$

so

$$F(x) = \sum_{y \leq x} p(y) = \sum_{y=1}^{\lfloor x \rfloor} (1-p)^{y-1} p$$

$$= p \sum_{y=1}^{\lfloor x \rfloor} (1-p)^{y-1}$$

recall how to sum a geometric series

ie

$$\sum_{y=0}^{k} a^y = a^0 + a^1 + \cdots + a^k$$

(assume $0 \leq a < 1$)
k positive integer

let $S = 1 + a + a^2 + \cdots + a^k$

so $(1-a)S = 1 + a + a^2 + \cdots + a^k - a - a^2 - \cdots - a^{k+1}$

$\Rightarrow (1-a)S = 1 - a^{k+1}$

$$\Rightarrow S = \frac{(1-a^{k+1})}{(1-a)}$$

this implies

$$\sum_{y=1}^{\lfloor x \rfloor} (1-p)^{y-1} = (1-p)^0 + (1-p) + \cdots + (1-p)^{\lfloor x \rfloor - 1}$$

$$= \frac{(1-(1-p)^{\lfloor x \rfloor})}{1-(1-p)}$$

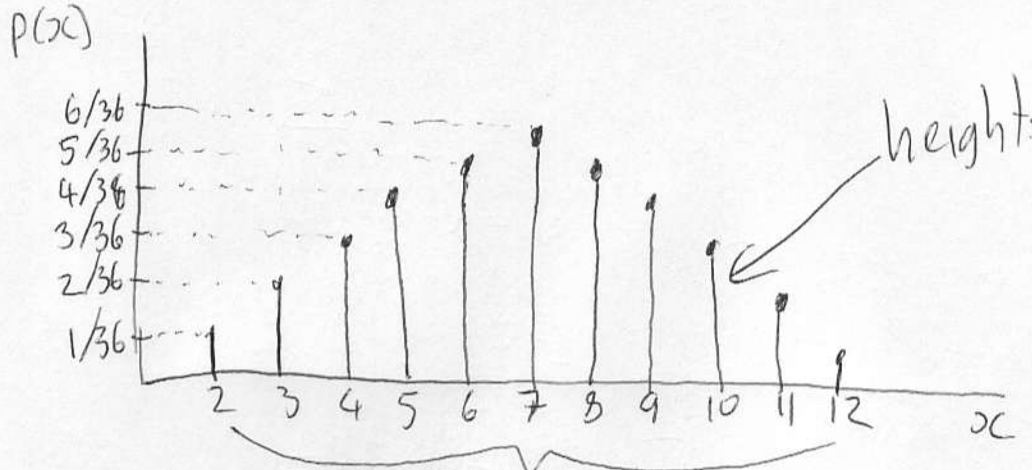
and so

$$F(x) = p\,\frac{\left(1-(1-p)^{\lfloor x \rfloor}\right)}{1-(1-p)} = \left(1-(1-p)^{\lfloor x \rfloor}\right) \quad x \geq 1$$

$$= 0 \qquad x < 1$$

Graphically representing pdf and cdf

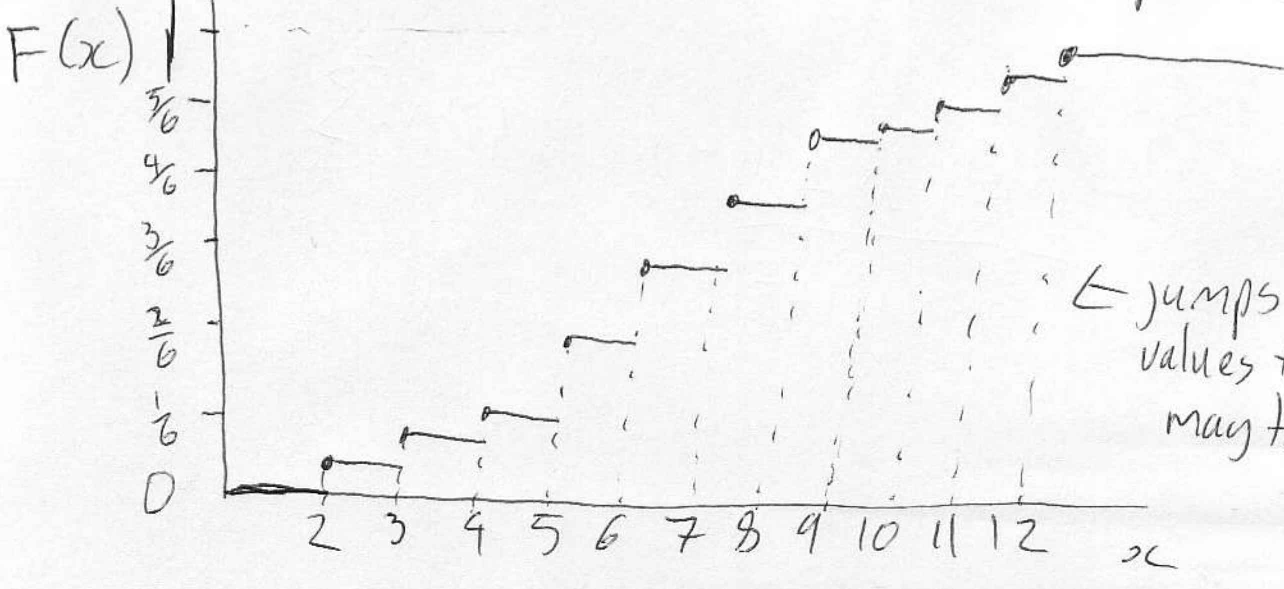
eg roll 2 dice X = "sum"



heights are the probabilities

the possible values of X

"step function"



← jumps at values that X may take

Note that you can go from the CDF back to the pdf.

eg suppose $x_1 < x_2 < x_3 < \ldots\ldots$ are the possible values of X. Then

$$p(x_i) = F(x_i) - F(x_{i-1})$$