§3.2: Probability Distributions for Discrete R.V.

Def: The <u>probability distribution</u> of a random var. X is $f(x) = P(X = x)$

[for <u>discrete</u> random variables this is often called the "<u>probability mass function</u>" (<u>pmf</u>)]

The <u>cumulative distribution function</u> (<u>cdf</u>) of X is
$$F(x) = P(X \le x) = \sum_{k \le x} f(k)$$

(Over time you will see that cdf $F(x)$ is more useful and important, but pmf $f(x)$ is easier to define & work with.)

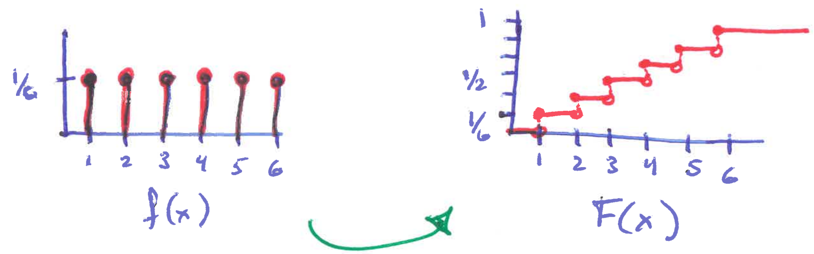
Example Roll a die $X = \#$ rolled

$f(1) = 1/6$	$F(1) = 1/6$
$f(2) = 1/6$	$F(2) = 2/6$
$f(3) = 1/6$	$F(3) = 3/6$
$f(4) = 1/6$	$F(4) = 4/6$
$f(5) = 1/6$	$F(5) = 5/6$
$f(6) = 1/6$	$F(6) = 6/6$

(<u>Note:</u> $F(x)$ is <u>increasing</u>
$0 \le F(x) \le 1$
<u>Final value</u> of $F(x)$ <u>is 1</u>)

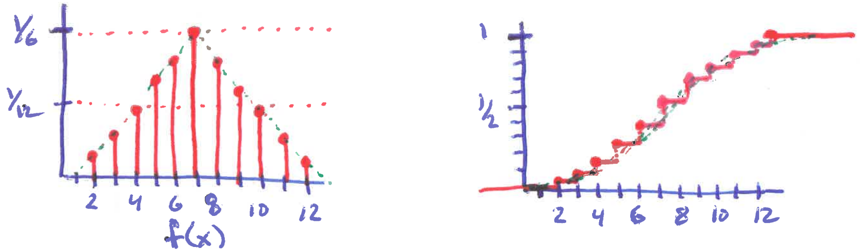
Often we will graph the pmf $f(x)$ & cdf $F(x)$

Example Roll a die $X = \#$ rolled



[$F(x)$ steps up from 0 to 1
height of each step is value $f(x) = P(X = x)$]

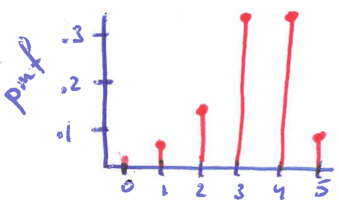
Example Roll two dice $X = $ sum of #s

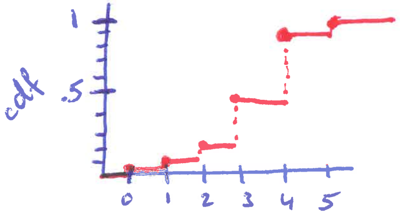


R code:
```
library(discreteRV)
(X <- RV(1:6))
(f <- SofI(X, X))
plot(f)
(F <- c(0, cumsum(probs(f))))
plot(stepfun(2:12, F, right=TRUE), main="cdf",
     ylab = "Probabilities")
```

<u>Example</u> Flip an unfair coin 5 times, X = #heads

↳ suppose $P(H) = 2/3$





$P(X=0) = (1/3)^5$

$\approx .004$

$P(X=1) = (1/3)^4 (2/3) \binom{5}{1}$

$= \frac{2}{3^5} \cdot 5 \approx .041$

$P(X=2) = (1/3)^3 (2/3)^2 \binom{5}{2}$

$= \frac{2^2}{3^5} \cdot \frac{5 \cdot 4}{2} \approx .16$

$P(X=3) = (1/3)^2 (2/3)^3 \binom{5}{3}$

$= \frac{2^3}{3^5} \cdot \frac{5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3} \approx .33$

⋮

<u>R code:</u>

```
library(discreteRV)
(X <- RV(0:1, probs=c(1/3, 2/3)))
(f <- SofIID(X, 5))
plot(f)
F <- c(0, cumsum(probs(f)))
plot(stepfun(min(f):max(f),
             F, right=TRUE),
     main="cdf",
     ylab="Probabilities")
```

<u>Note:</u> The example above is identical to "Roll 5 dice X = #rolls ≥ 3"

↳ Sum of 5 Bernoulli random variables with $P(1) = 2/3$

When doing statistics the <u>actual experiment</u> isn't important — only the <u>distribution of probabilities</u> of the different results matter.

Distributions of probability are organized into <u>families</u>
– specific members of families are distinguished by values of <u>parameters</u>.

<u>Example</u> The Bernoulli family of distributions are random variables

$$X = \begin{cases} 0 & \text{with probability } 1-\alpha \\ 1 & \text{with probability } \alpha \end{cases}$$

For this family there is only <u>one</u> parameter

$\alpha = P(X=1)$

Often we will write $X \sim \text{Bernoulli}(\alpha)$

"<u>X is distributed as</u> Bernoulli with parameter α"

<u>Subexample</u> A fair coin is tossed. $X = \begin{cases} 1 & \text{Heads} \\ 0 & \text{Tails} \end{cases}$

Then $X \sim \text{Bernoulli}(\alpha = 1/2)$

<u>Subexample</u> A die is rolled. $X = \begin{cases} 1 & \geq 3 \text{ rolled} \\ 0 & \leq 2 \text{ rolled} \end{cases}$

Then $X \sim \text{Bernoulli}(\alpha = 2/3)$

<u>Subexample</u> Four coins tossed. $X = \begin{cases} 1 & \geq 2 \text{ Heads} \\ 0 & \leq 1 \text{ Heads} \end{cases}$

Then $X \sim \text{Bernoulli}(\alpha = 11/16)$