Chapter #3: Discrete random variables and their distributions functions

- A random variable (r.v.) X is a numeric description of an experimental outcome.

  More sophisticatedly said, a r.v. X is a real-valued function defined on a sample space Ω.

$$\Omega \xrightarrow{X} \mathbb{R}\text{eal numbers}$$

- The range of X is the set of values
  $$\{X(\omega) : \omega \in \Omega\}$$
  and we denote this by R_X. $R_X \subseteq \mathbb{R}$.

  The r.v. X is a discrete r.v. if R_X is discrete; i.e., $R_X = \{x_1, x_2, \dots, x_n, \dots\}$. Hence, its graph on the real no. line looks like

  points with **well** defined gaps between them.

  The event $\{X = x\}$ has a probability mass associated with it; it is denoted by
  $$P(X = x).$$
  Of course, if $x \notin R_X$, then $\{X = x\} = \emptyset$ and

$P(X=x) = P(\phi) = 0.$

- The probability function (pf) of the r.v. X, denoted by $f_X(x)$, is defined as

$$f_X(x) = \begin{cases} P(X=x), & \text{if } x \in R_X \\ 0, & \text{if } x \notin R_X. \end{cases}$$

(Some authors use the term probability mass function; pmf.)

Of course, $\displaystyle\sum_{x \in R_X} f_X(x) = 1$

since $R_X = \{X(w) = x : w \in \Omega\}$
$= \text{all possible values } X \text{ can assume.}$

- The distribution function (df), $F_X(x)$, of a r.v. X is a real-valued function and is defined as

$$F_X(x) = P(X \le x)$$

since discrete

$$= \sum_{\text{all } x_j \le x} f_X(x_j) = \sum_{\text{all } x_j \le x} P(X = x_j)$$

This is a <u>step-function</u> and steps up at each possible value X may assume. The size of the jump is $P(X=x_j)$.

Note: In general, $f(x)$ denotes a pf. and $F(x)$ denotes a df.

<u>Example:</u>

① Experiment: Toss a "fair" coin three times.

Let X = number of heads for each outcome.

$$\Omega = \left\{ \begin{array}{cccc} \omega_1 & \omega_3 & \omega_5 & \omega_7 \\ TTT & THT & HHT & THH \\ \omega_2 & \omega_4 & \omega_6 & \omega_8 \\ TTH & HTT & HTH & HHH \end{array} \right\}$$

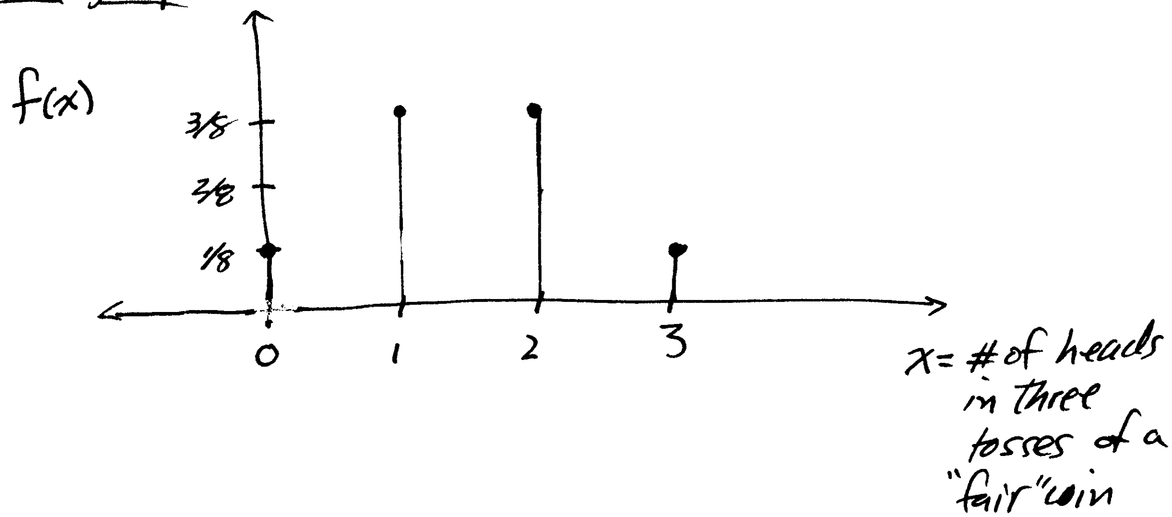
$P(\omega_j \in \Omega) = \frac{1}{8}$ since equally likely outcomes

$$R_x = \{ X(\omega) : \omega \in \Omega \} = \{0, 1, 2, 3\}$$

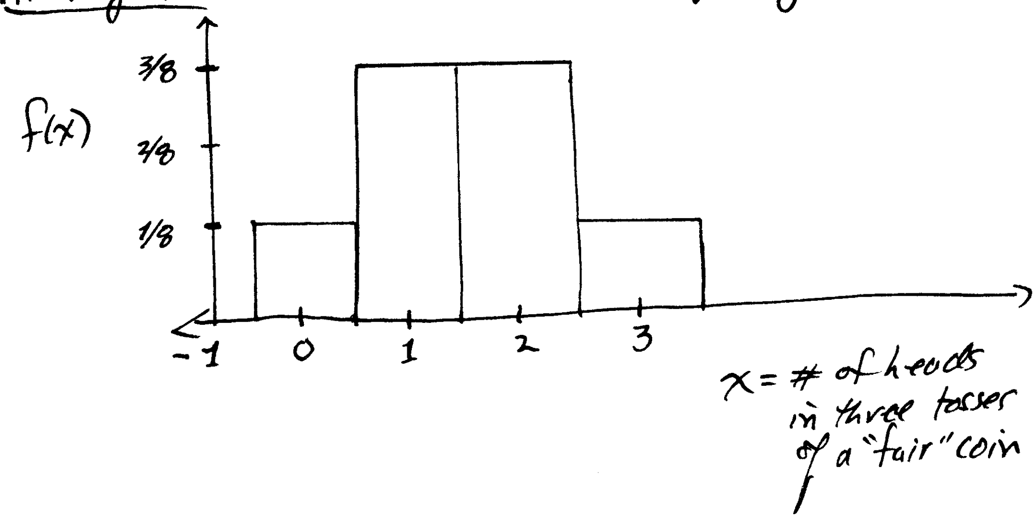
$X = x$	$f(x) = P(X=x)$
0	$\frac{1}{8}$
1	$3/8$
2	$3/8$
3	$1/8$
	$8/8 = 1$

This is a tabular display of this probability distribution of the r.v. X.

- <u>line graph</u> as a visual display



- <u>Histogram</u> as a visual display



- <u>Mathematical formula</u> to <u>describe</u> <u>a</u> <u>probability distribution</u>

$$f(x) = P(X=x) = \binom{3}{x}\left(\frac{1}{2}\right)^3, \quad x = 0, 1, 2, 3$$

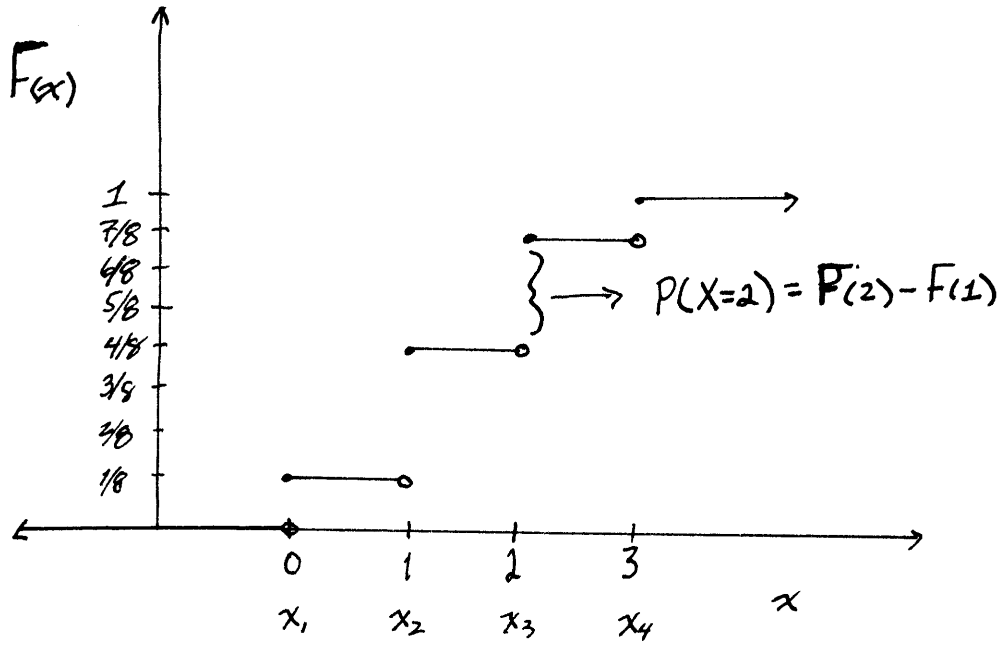
check this: $f(0) = \binom{3}{0}\left(\frac{1}{2}\right)^3 = \frac{3!}{0!3!}\cdot\frac{1}{8} = \frac{1}{8}$ ✓

$f(1) = \binom{3}{1}\left(\frac{1}{2}\right)^3 = \frac{3!}{1!2!}\cdot\frac{1}{8} = \frac{3}{8}$ ✓

$f(2) = \binom{3}{2}\left(\frac{1}{2}\right)^3 = 3/8$ ✓

$f(3) = \binom{3}{3}\left(\frac{1}{2}\right)^3 = 1/8$ ✓

- The d.f. $F(x)$ for this example:



$$R_x = \{x_1, x_2, x_3, x_4\} = \{0, 1, 2, 3\}$$

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1/8, & \text{if } 0 \le x < 1 \\ 4/8, & \text{if } 1 \le x < 2 \\ 7/8, & \text{if } 2 \le x < 3 \\ 1, & \text{if } 3 \le x < \infty \end{cases}$$

$F(x)$ is increasing, from $0 \to 1$, step function. It jumps up at each value in R_x and is right continuous

Note that $F(x)$ is defined now over the entire real number line.

Note: $f(x_j) = P(X = x_j) = F(x_j) - F(x_{j-1})$;

of course, we assume the range is written in increasing order so that

$$x_1 < x_2 < \cdots < x_{j-1} < x_j < \cdots < x_n < \cdots$$

$F(x)$ is the r.v. analogue to the ~~ogiv~~ relative cumulative ogive and to the empirical distribution function $F_n(x)$.

Note then:

(1) If X is a discrete r.v. with pf $f_X(x)$, then

$$f_X(x) = P(X=x) \quad \text{for } x \in R_X.$$

The pf $f_X(x) \geq 0$ and $\sum_{\text{all } x \in R_X} f_X(x) = 1.$

(2) Let $x_1 < x_2 < \cdots < x_{j-1} < x_j < \cdots < x_n < \cdots$ denote values x in R_X. Then the df $F_X(x)$ determines the values $f_X(x)$ since

$$f_X(x_j) = P(X = x_j) = F_X(x_j) - F_X(x_{j-1}).$$

Let $\varphi(X)$ denote a function of a discrete r.v. X.

Let X have the following probability distribution:

x	-2	-1	0	1	2
$f(x)$	0.1	0.2	0.3	0.3	0.1

Compute the pf of $Y = X^2$. $\boxed{\text{Here } y = \varphi(x) = x^2.}$

y	0	1	4
$f_Y(y)$	0.3	0.5	0.2

Since, e.g.,
$$\begin{aligned} P(Y=4) &= P(X=-2 \text{ or } 2) \\ &= P(X=-2) + P(X=2) \\ &= 0.1 + 0.1 = .2 \end{aligned}$$

I.e., the event $Y=4$ occurs when and only when $X=-2$ or $X=2$ occurs.

Recall

- the sample mean $\bar{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

- the sample mean for data grouped into a frequency table where
  x is the district value
  $f(x)$ = number ~~of~~ times x occurs in the data set
  $\hat{f}(x) = \dfrac{f(x)}{n}$ = the relative frequency.

  Then $\bar{x} = \sum x \hat{f}(x) = \dfrac{\sum x f(x)}{n}$

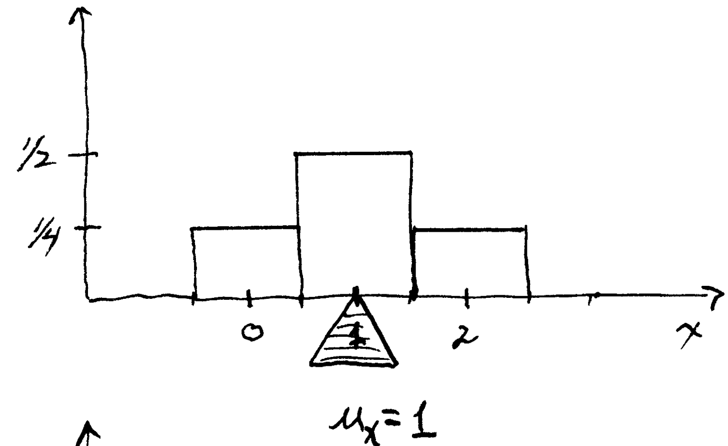
- Definition:
  The expected value (or mean) of a discrete r.v. X with pf $f_X(x)$ is ~~det~~ denoted $E(X)$ and is defined to be

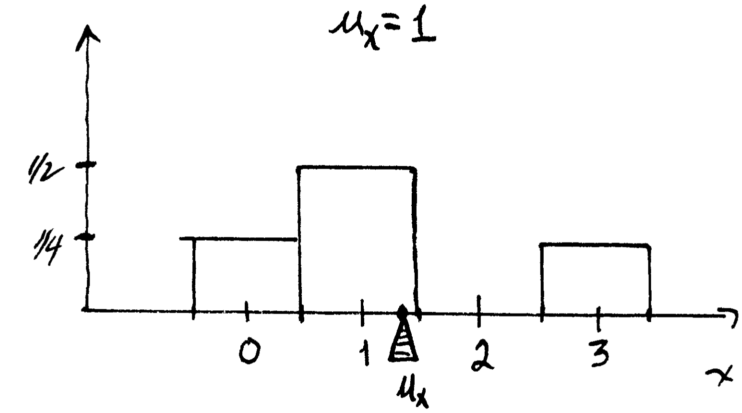
$$E(X) = \sum_{\text{all } x} x P(X=x) = \sum_{x} x f_X(x).$$

Often we use $\boxed{\mu_X = E(X)}$.

Note: $E(X) < \infty$ provided $\sum_x |x| f_X(x) < \infty$.
If $E(X) < \infty$, we say the mean exists. If $E(X) = \infty$, we say it does not exist.

$E(X)$ is the value where the histogram balances.
I.e., It is the center of mass of a distribution of weights $f(x_i)$ located at the points x_i on the line.





Here, $E(X) = \mu_X = 0(\frac{1}{4}) + 1(\frac{1}{2}) + 3(\frac{1}{4}) = \frac{5}{4} = 1.25$

- $E(\varphi(X)) = \sum_{\text{all } x} \varphi(x) f_X(x)$. Says we do <u>not</u> need to know the pf of $Y = \varphi(X)$.

Back to page 6 → $Y = X^2$

$$E(Y) = E(X^2) = \sum_x x^2 f_X(x)$$

$$= (-2)^2(.1) + (-1)^2(.2) + 0(.3) + 1^2(.3) + 2^2(.1)$$

$$= .4 + .2 + .3 + .4 = 1.3$$

or $E(Y) = \sum_y y f_Y(y) = 0(.3) + 1(.5) + 4(.2)$

$$= 0 + .5 + .8 = 1.3$$

- The K^{th} moment μ_K, $K=1,2,3,\ldots$ of a r.v. X is defined to be

$$\mu_K = E(X^K) \quad \text{where } K=1,2,3,\ldots$$

$$= \sum_x x^K f_X(x) \quad \text{when } X \text{ is discrete}$$

Note: Here $\varphi(x) = x^k$.

- Some facts from intro. calculus or even algebra II:

① $$\sum_{i=1}^{N} i = \frac{N(N+1)}{2}$$

② $$\sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}$$

③ $$\sum_{x=0}^{\infty} p^x = 1 + p + p^2 + p^3 + \cdots$$

$$= \frac{1}{1-p}$$

for $0<p<1$,

This is just the geometric series.

- The variance of a r.v. X is defined to be

$$E[(X-\mu_X)^2] = \sigma_X^2.$$

For a discrete r.v. with pf $f_X(x)$ we have

$$\sigma_X^2 = \sum_x (x-\mu_X)^2 f_X(x).$$

Fact:

$$\sigma^2 = E[(X-\mu_x)^2]$$
$$= E(X^2) - [E(X)]^2$$
$$= 2^{nd} \text{ moment} - 1^{st} \text{ moment squared}$$
$$= \sum_x x^2 f(x) - \left(\sum x f(x)\right)^2$$ when X is discrete with pf $f(x)$.

Proof:

$$\sigma^2 = \sum_x (x-\mu_x)^2 f(x)$$
$$= \sum_x \left(x^2 - 2\mu_x x + (\mu_x)^2\right) f(x)$$
$$= \sum_x x^2 f(x) - 2\mu_x \underbrace{\sum_x x f(x)}_{=\mu_x} + \mu_x^2 \overbrace{\sum_x f(x)}^{=1}$$
$$= E(X^2) - 2\mu_x^2 + \mu_x^2$$
$$= E(X^2) - \mu_x^2 = E(X^2) - (E(X))^2$$

Variance measures the spread (or dispersion) in the histogram for a random variable. The measure of spread then, back in original units of X, is

<u>standard deviation</u>, denoted by σ_x,

and is
$$\sigma_x = +\sqrt{\sigma_x^2} = +\sqrt{\text{variance}(X)}.$$

Back to example on p. 3; we toss a "fair" coin three times. $X = \#$ of heads.

$$\mu_X = E(X) = \sum_x x f(x) = 0\left(\tfrac{1}{8}\right) + 1\left(\tfrac{3}{8}\right) + 2\left(\tfrac{3}{8}\right) + 3\left(\tfrac{1}{8}\right)$$

$$= \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{12}{8} = \frac{3}{2} = 1.5$$

$$\mu_2 = E(X^2) = \sum x^2 f(x) = 1^2\left(\tfrac{3}{8}\right) + 2^2\left(\tfrac{3}{8}\right) + 3^2\left(\tfrac{1}{8}\right)$$

2nd moment

$$= \frac{3}{8} + \frac{12}{8} + \frac{9}{8}$$

$$= \frac{24}{8} = 3$$

$$\sigma_X^2 = \operatorname{var}(X) = E(X^2) - (\mu_X)^2$$

$$= 3 - \left(\tfrac{3}{2}\right)^2 = 3 - \frac{9}{4} = 3 - 2.25$$

$$= .75$$

$$\text{Stand. dev.}(X) = \sigma_X = \sqrt{.75} \approx .866$$

$\boxed{\text{Experiment}}$

Keep rolling a "fair" die until a 6 occurs.
Let $X = \#$ of rolls (trials) needed to observe a 6.

Let $S =$ success $=$ a 6 results
$F =$ failure $= 1, 2, 3, 4,$ or 5 results.

We think of this as a sequence of identical Bernoulli Trials. A Bernoulli trial is one in which results in exactly one of two possible outcomes, S or failure.
We can ~~assign~~ define a r.v. X as

$$X = \begin{cases} 1 & \text{if } S \text{ occurs} \\ 0 & \text{if } F \text{ occurs.} \end{cases}$$

Let $p =$ prob. of success. Let $q = 1-p =$ prob. of failure
Then X is called a Bernoulli r.v. with pf.

$$f_X(x) = p^x q^{1-x}, \quad x = 0, 1.$$

$$\left(\begin{array}{l} f_X(0) = P(X=0) = p^0 q^1 = q \\ f_X(1) = P(X=1) = p \end{array} \right).$$

$\Omega = \{ S, FS, FFS, FFFS, \dots \}$

So, $X = 1, 2, 3, \dots$

X is called a <u>geometric</u> r.v. as it counts the # of Bernoulli trials needed to obtain the first success. It's pf is

$$f_X(x) = P(X=x) = q^{x-1} \cdot p \,, \qquad x = 1, 2, 3, \dots$$

For the die problem, $p = P(\text{a 6 occurs})$
$= P(S) = \frac{1}{6}$
$q = P(F) = \frac{5}{6}$

$$f_X(x) = \left(\frac{5}{6}\right)^{x-1} \cdot \frac{1}{6} \,, \quad x = 1, 2, 3, \dots$$

<u>Note</u>: For a geometric pf.

$$f_X(x+1) = q^{(x+1)-1} p = q^x p = q\, q^{x-1} p = q\, f_X(x)$$

i.e. $\boxed{f(x+1) = q\, f(x)}$

$$\boxed{E(X) = \frac{1}{p} \quad \text{and} \quad Var(X) = \frac{q}{p^2}}$$