4.3. Expected value.

X : random variable.

$E[X]$ = "expected value", "mean", "average"

Ex # of apartments in Ann Arbor:

Bedrooms	0	1	2	3	4
Apts.	1,200	2,572	3,307	3,198	1,806

Total = 12,083

Average # of bedrooms in an AA apt?

$$\text{Ans} = \frac{\text{total \# bdrms}}{\text{total \# apts}} = \frac{0\cdot 1200 + 1\cdot 2572 + 2\cdot 3307 + 3\cdot 3198 + 4\cdot 1806}{12{,}083} \approx 2.15$$

- Probabilistic view: experiment = choose an apt at random
  r.v. X = # bdrms in the apt.

pmf: $p(0) = \frac{1200}{12083} \approx 0.10$, $p(1) = \frac{2572}{12{,}083} \approx 0.21$, $p(2) = \frac{3307}{12083} \approx 0.27$, $p(3) = \frac{3198}{12083} = 0.26$, $p(4) = \frac{1806}{12083} \approx 0.15$

Average: $E[X] = 0\cdot p(0) + 1\cdot p(1) + 2\cdot p(2) + 3\cdot p(3) + 4\cdot p(4) \approx 2.15$.

Def Let X be a discrete r.v. with pmf $p(x)$. The expected value of X is defined as

$$E[X] = \sum_x x\, p(x) = \sum_k x_k\, p(x_k)$$

Remark 1. If all N values of X are equally likely, then $p(x_k) = \frac{1}{N}$ $\Rightarrow$

$$E[X] = \frac{1}{N}\sum_{k=1}^{N} x_k \quad \text{(average)}.$$

- In general, the sum $\sum_k x_k\, p(x_k)$ is the weighted average which puts more weight of more likely values.

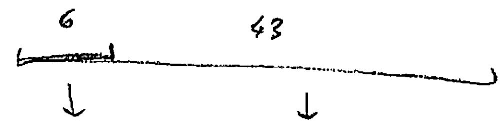
Ex X = # heads in 2 coin flips (see Ex. p. 28)

$$E[X] = 0\cdot\frac{1}{8} + 1\cdot\frac{3}{8} + 2\cdot\frac{3}{8} + 3\cdot\frac{1}{8} = 1.5$$

Ex (Lottery). 6 out of 49 (order irrelevant)

# correct	Prize
all 6	\$1,200,000
5	\$800
4	\$35

Expected $\underbrace{\text{amount a player wins}}_{\substack{\parallel \\ X}}$ = ?



pmf: $p(1{,}200{,}000) = P\{\text{all 6 correct}\} = \dfrac{1}{\binom{49}{6}}$

$p(800) = P(5\text{ correct}, 1\text{ incorrect}) = \dfrac{\binom{6}{5}\binom{43}{1}}{\binom{49}{6}}$

$p(35) = P(4\text{ correct}, 2\text{ incorrect}) = \dfrac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}}$

$p(0) = 1 - p(1200000) - p(800) - p(35)$

$\mathbb{E}[X] = 1{,}200{,}000 \cdot p(1{,}200{,}000) + 800\, p(800) + 35 \cdot p(35) + 0 \cdot p(0) \approx 0.13$

Ans: 13¢.

(If the lottery ticket costs < 13¢, it is advantageous to play. Otherwise not).

Property: If $X_1, X_2, \ldots$ are random variables, then
$$\boxed{E\left[\sum_i X_i\right] = \sum_i E(X_i).}$$

(To be proved later)

Example (Group Testing)

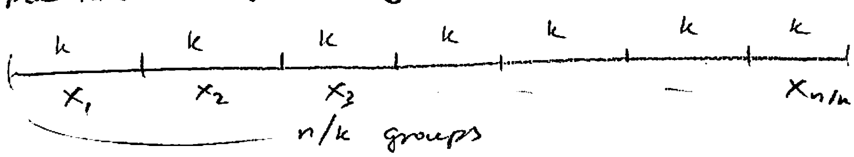
Need to test blood of n people for a rare disease (syphilis in men drafted during WWII). $p = P(\text{positive})$.

Method 1: test all people individually ($\Rightarrow$ n tests.)

Method 2: draw and mix blood of k people.
(Group testing)
If test is negative $\Rightarrow$ no more tests needed ($\Rightarrow$ 1 test)
If positive $\Rightarrow$ test all k people individually ($\Rightarrow$ $k+1$ tests).
Repeat for next k people, etc.

Expected $\underbrace{\#\text{ of tests in Method 2}}_{X}$?

Divide n people into n/k groups of k:



$\Rightarrow$ $X = X_1 + X_2 + \ldots + X_{n/k}$

X_i = # of tests needed for group i. Values 1, $k+1$

$P\{X_i = 1\} = P\{\text{all } k \text{ people are negative}\} = (1-p)^k$

$P\{X_i = k+1\} = 1-(1-p)^k.$

$E[X_i] = 1\cdot(1-p)^k + (k+1)\left[1-(1-p)^k\right].$ (same for each group i)

$$\Rightarrow E[X] = \sum_{i=1}^{n/k} E(X_i) = \boxed{\frac{n}{k}\left((1-p)^k + (k+1)\left(1-(1-p)^k\right)\right)}$$

For example, if $n = 100{,}000$, $p = 10^{-4}$ (on ave, 10 sick).

With $k = 120$ $\Rightarrow$ $E[X] = 2026$

Compare to method 1 where # = 100,000.