Ex In a group of n people, each person is sick with prob. p, independently. What is the expected $\underbrace{\text{\# of sick people}}_{=X}$?

$X = \sum_{i=1}^{n} \mathbb{E}[X_i]$ where $X_i = \begin{cases} 1 & \text{if person } i \text{ is sick} \\ 0 & \text{if not} \end{cases}$

$\mathbb{E}(X_i) = 1 \cdot p + 0 \cdot (1-p) = p$

$\mathbb{E}(X) = np$

Ex (Group Testing) Need to test n people for a disease (ex: syphilis in WWII soldiers).

$p = P\{\text{test is positive}\}$.

Method 1 : test everyone ($\Rightarrow$ n tests) Expensive.

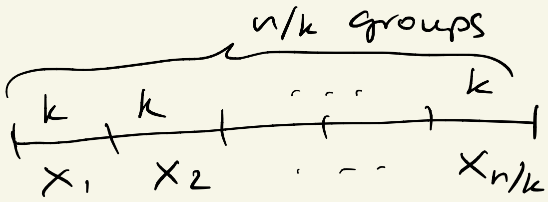
Method 2 : • mix the samples of k people, test the mix.

If test negative $\Rightarrow$ no more tests needed ($\Rightarrow$ 1 test)

If positive $\Rightarrow$ test all k people individually ($\Rightarrow$ $k+1$ test)

• Repeat for next k people, etc.

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



$$X = X_1 + \dots + X_{n/k}$$

where $X_i = \#(\text{tests needed for group } i) \in \{1, k+1\}$.

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

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

$$\Rightarrow \mathbb{E}[X_i] = 1 \cdot (1-p)^k + (k+1)\left[1-(1-p)^k\right] = 1 + k\left[1-(1-p)^k\right] \quad \forall i$$

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

- Optimize in k? Simplify first:

Bernoulli inequality: $(1-p)^k \geq 1 - pk \quad \Rightarrow \mathbb{E}[X] \leq n\left(\frac{1}{k} + pk\right)$

Minimize in k. Optimal choice: $k = 1/\sqrt{p} \Rightarrow$

$$\mathbb{E}[X] \leq \boxed{2n\sqrt{p}} \ll n \text{ for small } p.$$

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

$\Rightarrow k = 100$, $\mathbb{E}[X] \leq 2{,}000$ tests $\lll 100{,}000$ in Method 1.

Note: Multistage $\Rightarrow O(np \log \frac{1}{p})$: almost achieves lower bd of np.
(wikipedia)

$\mathbb{E}\,\#(\text{sick people})$