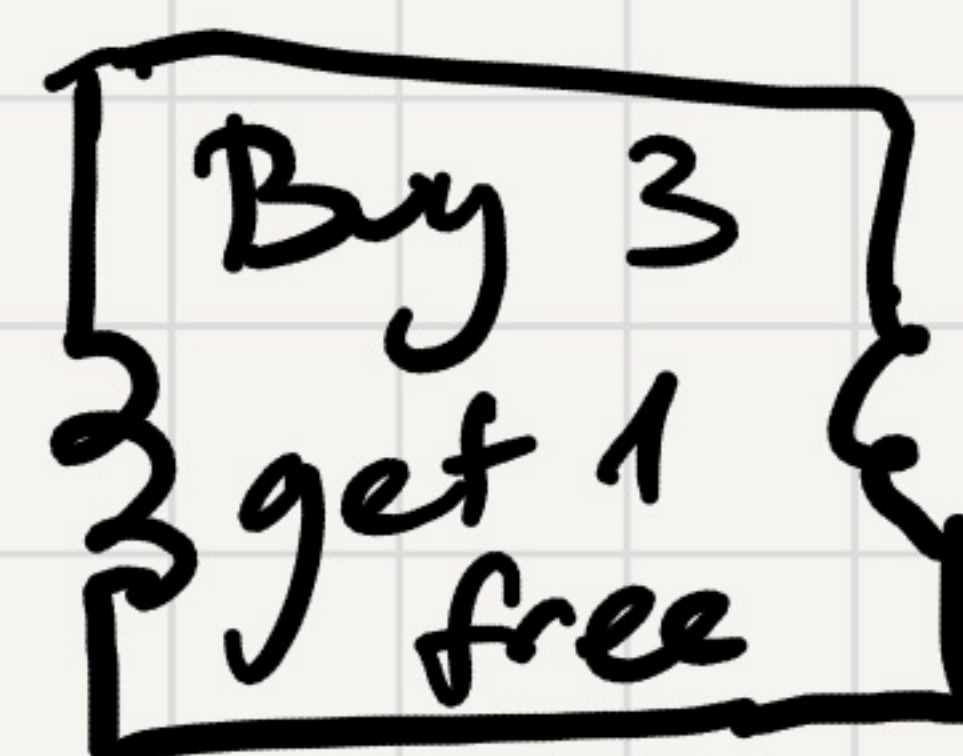
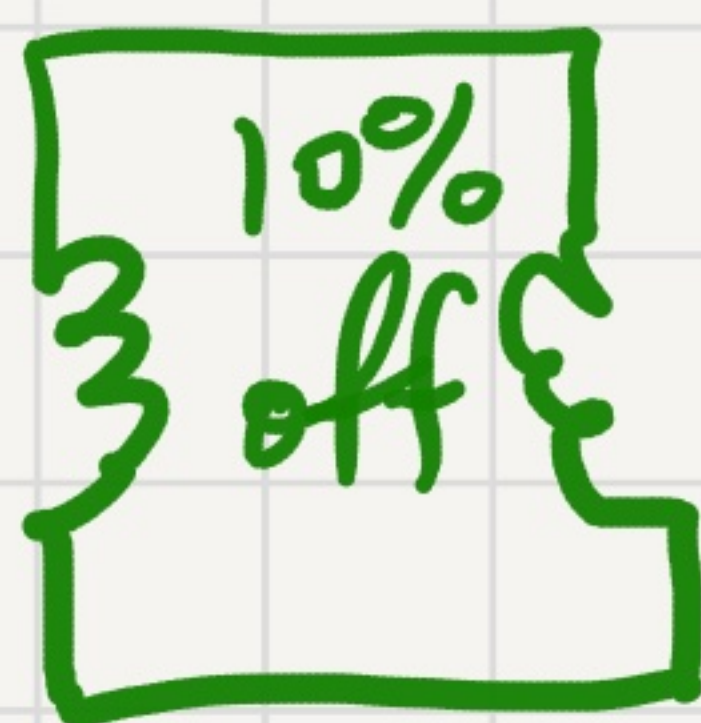


Coupon Collector

Suppose there's a coupon inside every cereal box.



Suppose there are n types of coupons
and each box has a random type.

How many boxes do you need to purchase
until you get all coupons?

Coupon Collector

Experiment: Get random coupons from $\{1, \dots, n\}$ until you have all types of coupons.

X - # of boxes until you get all n types.

Example: $n=3$

2 | 2 2 1 | 1 2 1 3 |
 X_1 X_2 X_3

$$\begin{aligned} E[X] &= E[X_1] + E[X_2] + E[X_3] \\ &= 1 + \frac{3}{2} + 3 = 5.5 \end{aligned}$$

Idea: Write X as the sum $X_1 + X_2 + X_3$

1 = X_1 - # boxes until we get first type

$\text{Geo}(\frac{2}{3}) \leftarrow X_2$ - # boxes until we get second type after we got first type.

$\text{Geo}(\frac{1}{3}) \leftarrow X_3$ - # boxes until we get third type after we got second type

General Case:

Write $X = X_1 + X_2 + \dots + X_n$

X_i - #boxes until we get i unique types after we got $(i-1)$ unique types.

- $X_1 = 1$
- $X_2 = \begin{cases} 1 & \text{w.p. } \frac{n-1}{n} \\ 2 & \text{w.p. } \frac{1}{n} \cdot \frac{n-1}{n} \\ \vdots & \end{cases} \quad X_2 \sim \text{Geo}\left(\frac{n-1}{n}\right)$
- $X_3 \sim \text{Geo}\left(\frac{n-2}{n}\right)$
- $X_i \sim \text{Geo}\left(\frac{n-(i-1)}{n}\right)$

$$E[X] = E[X_1] + \dots + E[X_n] =$$

Balls in Bins

Suppose you throw balls in n bins

How many balls do you need to throw until
all bins are non-empty?

Functions of Random Variables

Suppose X is a r.v. and $g: \mathbb{R} \rightarrow \mathbb{R}$

Then $Y = g(X)$ is also a r.v.

$$Y(\omega) = g(X(\omega)). \quad \text{Q: Calculate } E[Y]$$

Method 1: Compute the dist of Y from the dist of X

Method 2: $E[Y] = \sum_{\omega \in \Omega} g(X(\omega)) \cdot \text{Pr}[\omega]$

"LOTUS"

$$= \sum_x \sum_{\omega: X(\omega)=x} g(x) \cdot \text{Pr}[\omega]$$

$$= \sum_x g(x) \cdot \sum_{\omega: X(\omega)=x} \text{Pr}[\omega] = \sum_x g(x) \cdot \text{Pr}[X=x]$$

Variance

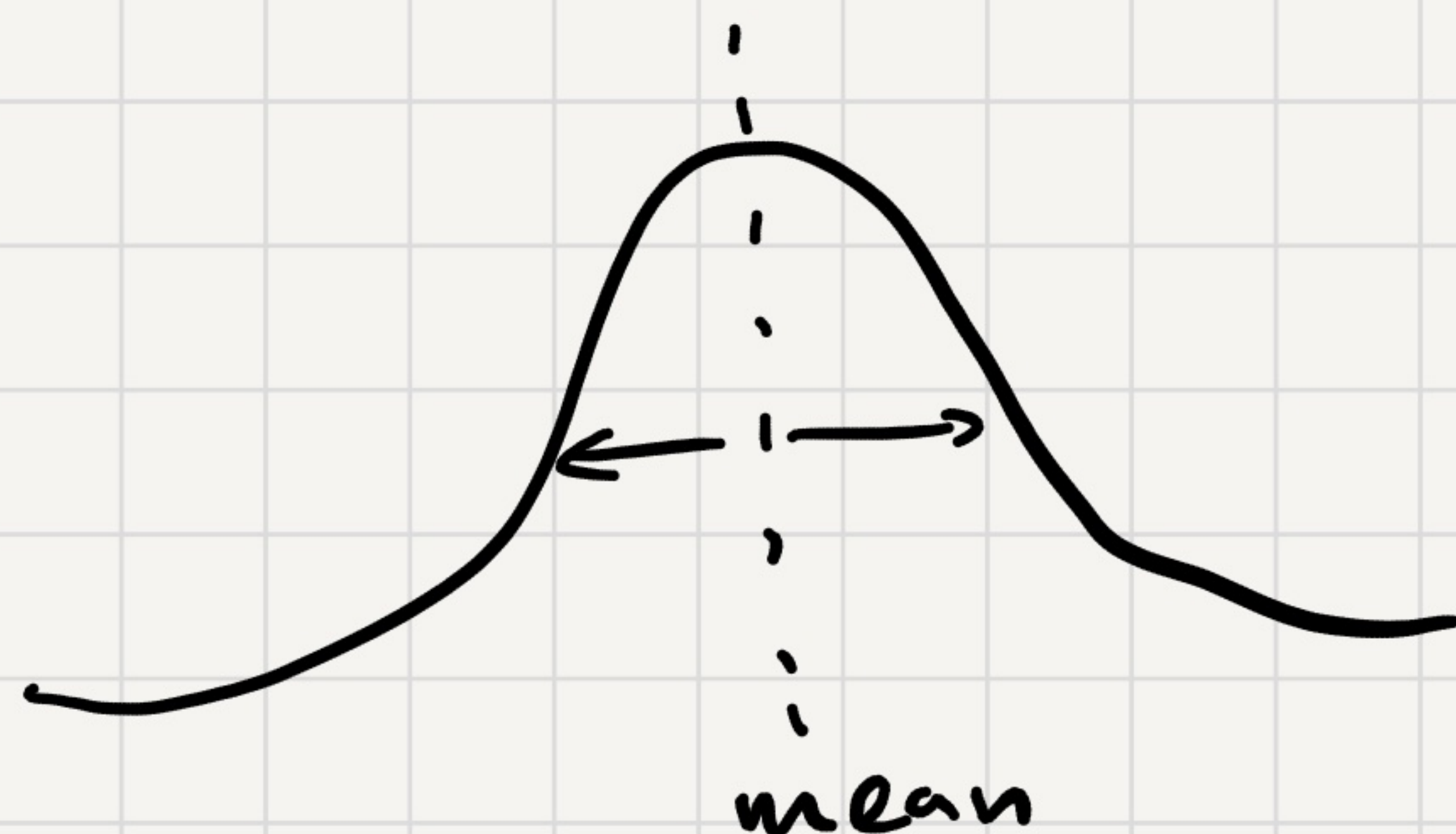
Suppose we have some r.v. X

$E[X]$ - the average value X would get if we run the experiment many times.

But if we run the experiment only once, we still would like a guarantee on how close X is to its mean.

Variance

The variance measures the ^{expected} deviation ^{squared} of X from its mean.



Def'n: The variance of a r.v. X is defined as

$$\text{Var}(X) = E[(X - EX)^2]$$

Measures the expected deviation squared.

$$\sigma(X) = \text{stdev}(X) = \sqrt{\text{Var}(X)}$$

Variance

Def'n: The variance of a r.v. X is defined as

$$\text{Var}(X) = E[(X - E[X])^2]$$

Fact:

$$\text{Var}(X) = E[X^2] - E[X]^2.$$

Proof:

Let $\mu = E[X]$.

$$\text{Var}(X) = E[(X - \mu)^2]$$

$$= E[X^2 - 2\mu X + \mu^2]$$

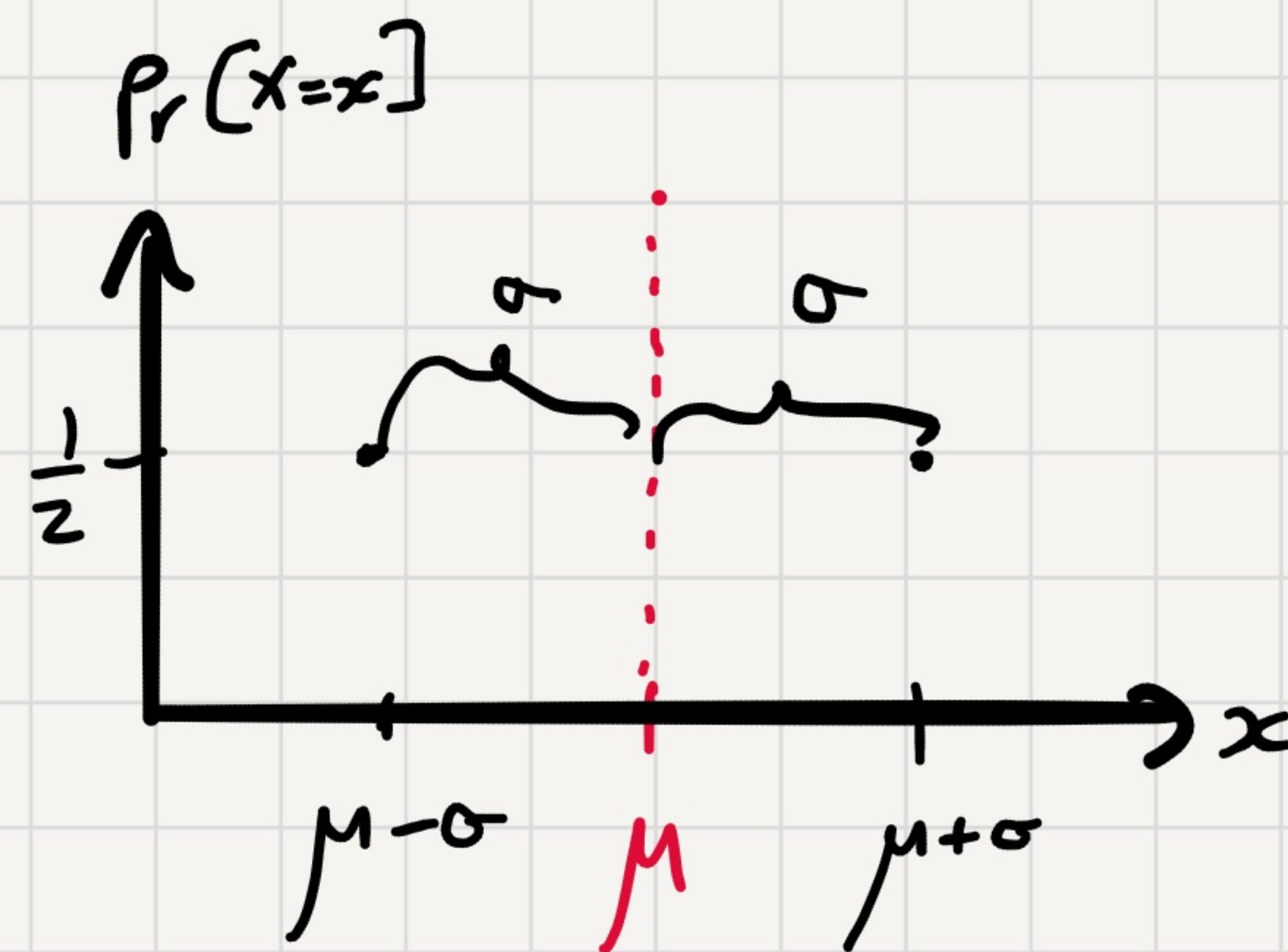
$$= E[X^2] - 2\mu \underbrace{E[X]}_{\mu} + \mu^2$$

$$= E[X^2] - \mu^2 = E[X^2] - E[X]^2.$$

Simple Example

$$X = \begin{cases} \mu + \sigma & \text{w.p. } \frac{1}{2} \\ \mu - \sigma & \text{w.p. } \frac{1}{2} \end{cases}$$

$$E X = \frac{(\mu + \sigma) + (\mu - \sigma)}{2} = \mu$$



$$\begin{aligned} \text{Var}(X) &= E[(X - E X)^2] = \frac{1}{2} \cdot (\mu + \sigma - \mu)^2 + \frac{1}{2} \cdot (\mu - \sigma - \mu)^2 \\ &= \frac{1}{2} \sigma^2 + \frac{1}{2} \sigma^2 = \sigma^2 \end{aligned}$$

Facts: For any $c \in \mathbb{R}$, r.v. X

1. $\text{Var}(c \cdot X) = c^2 \cdot \text{Var}(X)$. Scales by c^2

2. $\text{Var}(c + X) = \text{Var}(X)$. shifts center

Proof:

1.
$$\begin{aligned} \text{Var}(c \cdot X) &= \mathbb{E}[(cX)^2] - \mathbb{E}[cX]^2 \\ &= c^2 \cdot \mathbb{E}[X^2] - c^2 \cdot \mathbb{E}[X]^2 = c^2 \cdot \text{Var}(X). \end{aligned}$$

2.
$$\begin{aligned} \text{Var}(c+X) &= \mathbb{E}[(c+X - \mathbb{E}(c+X))^2] \\ &= \mathbb{E}[(c+X - c - \mathbb{E}X)^2] \\ &= \mathbb{E}[(X - \mathbb{E}X)^2] = \text{Var}(X). \end{aligned}$$

Facts: For any $c \in \mathbb{R}$, r.v. X

1. $\text{Var}(c \cdot X) = c^2 \cdot \text{Var}(X)$.

Scales by c^2

2. $\text{Var}(c + X) = \text{Var}(X)$.

shifts center

If X and Y are r.v.s.

what's $\text{Var}(X + Y) = ?$

Recap

Independent Random Variables

Def'n: We say that two r.v.s X and Y are independent if

$$\Pr[X=a, Y=b] = \Pr[X=a] \cdot \Pr[Y=b]$$

for all $a \in \text{range}(X)$
 $b \in \text{range}(Y)$.

Fact: If X & Y are indep.

$$\Pr[X=a | Y=b] = \Pr[X=a]$$

Theorem: If X and Y are independent r.v.s

then
$$E[X \cdot Y] = E[X] \cdot E[Y]$$

Proof:

$$\begin{aligned} E[X \cdot Y] &= \sum_{a,b} a \cdot b \cdot \Pr[X=a, Y=b] \\ &= \sum_{a,b} a \cdot b \cdot \Pr[X=a] \cdot \Pr[Y=b] \\ &= \left(\sum_a a \cdot \Pr[X=a] \right) \cdot \left(\sum_b b \cdot \Pr[Y=b] \right) \\ &= E[X] \cdot E[Y]. \end{aligned}$$

Theorem: If X and Y are independent r.v.s, then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y).$$

Proof:

$$\begin{cases} \mathbb{E}[(X+Y)^2] = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] \\ (\mathbb{E}[X+Y])^2 = (\mathbb{E}X + \mathbb{E}Y)^2 \\ = (\mathbb{E}X)^2 + 2(\mathbb{E}X)(\mathbb{E}Y) + (\mathbb{E}Y)^2 \end{cases}$$

$$\begin{aligned} \text{Var}(X+Y) &= \mathbb{E}[X^2] - (\mathbb{E}X)^2 + 2\mathbb{E}[XY] - 2(\mathbb{E}X)(\mathbb{E}Y) \\ &\quad + \mathbb{E}[Y^2] - (\mathbb{E}Y)^2 \\ &= \text{Var}(X) + 0 + \text{Var}(Y) \end{aligned}$$

↑
since X, Y are indep.

Theorem: If X_1, X_2, \dots, X_n are pairwise independent r.v.s then

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$$

Proof: (Similar)

$$E[(X_1 + \dots + X_n)^2] = \sum_{i=1}^n E[X_i^2] + \sum_{i \neq j} E[X_i X_j]$$

$$- E[X_1 + \dots + X_n]^2 = \sum_{i=1}^n E[X_i]^2 + \sum_{i \neq j} E[X_i] E[X_j]$$

$$\begin{aligned} \text{Var}(X_1 + \dots + X_n) &= \sum_{i=1}^n E[X_i^2] - E[X_i]^2 + \sum_{i \neq j} E[X_i X_j] - (E[X_i] E[X_j]) \\ &= \sum_{i=1}^n \text{Var}(X_i) + 0. \end{aligned}$$

Facts: For any $c \in \mathbb{R}$, r.v. X

1. $\text{Var}(c \cdot X) = c^2 \cdot \text{Var}(X)$.

Scales by c^2

2. $\text{Var}(c + X) = \text{Var}(X)$.

shifts center

Q: If X and Y are r.v.s.
what's $\text{Var}(X + Y) = ?$

A: It depends.

Facts: For any $c \in \mathbb{R}$, r.v. X

1. $\text{Var}(c \cdot X) = c^2 \cdot \text{Var}(X)$.

Scales by c^2

2. $\text{Var}(c + X) = \text{Var}(X)$.

shifts center

Q: If X and Y are r.v.s.
what's $\text{Var}(X + Y) = ?$

A: It depends.

Example 1: X, Y ind.

Example 2: X, Y r.v.s $Y = X$

Example 3: X, Y r.v.s $Y = -X$

Variance of a Binomial Distribution

$X \sim \text{Bin}(n, p)$ models # of heads
in n coin tosses w. heads prob. p .

$$P[X=i] = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$$

$$X = X_1 + \dots + X_n \quad X_i = \text{indicator that } i\text{th flip is heads.}$$

$$E[X] = E[X_1 + \dots + X_n] = n \cdot p.$$

$$\text{Var}(X) = \text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$$

$$\text{Var}(X_i) = E[X_i^2] - E[X_i]^2 = p - p^2 = p \cdot (1-p)$$

Variance of a Binomial Distribution

$X \sim \text{Bin}(n, p)$ models # of heads
in n coin tosses w. heads prob. p .

$$\Pr[X=i] = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$$

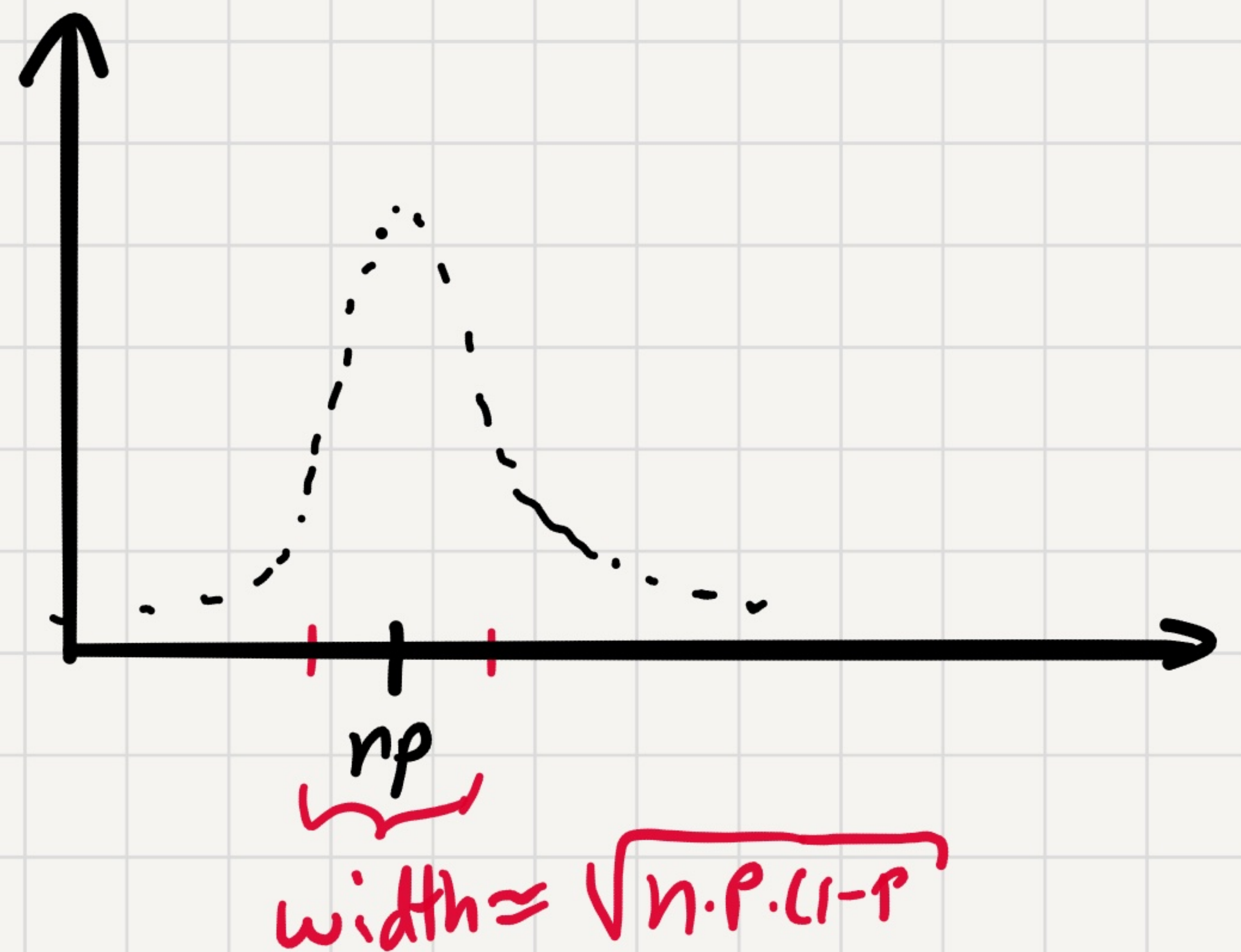
$$\mathbb{E}X = np$$

Scales
linearly
with n

$$\text{Var}(X) = n \cdot p \cdot (1-p) \leq \frac{n}{4}$$

$$\text{stdev}(X) = \sqrt{n \cdot p \cdot (1-p)} \leq \frac{\sqrt{n}}{2}$$

Scales
like \sqrt{n}



Number of Fixed Points - Variance

Handout assignment at random to n students

X = number of students who got their own assignment.

$\Omega = \{ \pi : \pi \text{ is a permutation of } \{1, \dots, n\} \}$

$X = X_1 + X_2 + \dots + X_n$ $X_i = \text{indicator that } \pi(i) = i$

$$\text{Var}(X) = E[X^2] - E[X]^2$$

$$\begin{aligned} E[X^2] &= E[(X_1 + \dots + X_n)^2] = E\left[\sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j\right] \\ &= \sum_{i=1}^n E[X_i^2] + \sum_{i \neq j} E[X_i X_j] \end{aligned}$$

$$E[X_i X_j] = \text{Pr}[\pi(i) = i, \pi(j) = j] = \frac{1}{n} \cdot \frac{1}{n-1}$$

are X_i, X_j indep?

$$E[X^2] = \sum_{i=1}^n E[X_i^2] + \sum_{i \neq j} \frac{1}{n} \cdot \frac{1}{n-1} = n \cdot \frac{1}{n} + n \cdot (n-1) \frac{1}{n} \frac{1}{n-1} = 2.$$

Covariance

If X and Y are r.v.s their co-variance is

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)]$$

Fact: $\text{Cov}(X, Y) = E[XY] - (EX)(EY)$

Proof: let $\mu_x = EX$ $\mu_y = EY$

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_x)(Y - \mu_y)] \\ &= E[XY] - \mu_x E[Y] - \mu_y E[X] + \mu_x \mu_y \\ &= E[XY] - \mu_x \mu_y = E[XY] - (EX)(EY).\end{aligned}$$

Covariance

If X and Y are r.v.s their co-variance is

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)]$$

Fact: $\text{Cov}(X, Y) = E[XY] - (EX) \cdot (EY)$

Corollary: If X and Y are independent then

$$\text{Cov}(X, Y) = 0.$$

Recap

Theorem: If X and Y are independent r.v.s, then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y).$$

Proof:

$$\begin{cases} E[(X+Y)^2] = E[X^2] + 2E[XY] + E[Y^2] \\ (E[X+Y])^2 = (EX + EY)^2 \end{cases}$$

$$= (EX)^2 + 2(EX)(EY) + (EY)^2$$

$$\begin{aligned} \text{Var}(X+Y) &= E[X^2] - (EX)^2 + 2E[XY] - 2(EX)(EY) \\ &\quad + E[Y^2] - (EY)^2 \\ &= \text{Var}(X) + 0 + \text{Var}(Y) \end{aligned}$$

The only point where we used that X & Y are indep.

Theorem: If X and Y are r.v.s

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

Proof:
$$\left\{ \begin{aligned} E[(X+Y)^2] &= E[X^2] + 2E[XY] + E[Y^2] \\ (E[X+Y])^2 &= (EX + EY)^2 \\ &= (EX)^2 + 2(EX) \cdot (EY) + (EY)^2 \end{aligned} \right.$$

$$\begin{aligned} \text{Var}(X+Y) &= E[X^2] - (EX)^2 + 2E[XY] - 2(EX)(EY) \\ &\quad + E[Y^2] - (EY)^2 \\ &= \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y) \end{aligned}$$