Dis 11B: Variance

Riddle of the day:

poison riddle

A king has a cellar of 1000 bottles of wine. An enemy queen plots to kill the king by poisoning one of the bottles with a lethal toxin. The king learns of the poisoning, but doesn't know which bottle was poisoned. Furthermore, it takes around a month to have an effect. The king decides to have his servants each drink from some of the bottles (possibly different for each servant) to determine the poisoned bottle. What is the least amount of servants needed to carry out this task in a month's time?

Review

$f(x) = x^2, \quad x^3 + \log(x).$

(LOTUS) $\mathbb{E}[f(X)] = \sum_x f(x)\, \mathbb{P}(X = x).$

↳ $\mathbb{E}[X^2] = \sum_x x^2\, \mathbb{P}(X = x).$

Variance: $\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2]$, i.e. "expected deviation from the mean".

$= \boxed{\mathbb{E}[X^2] - \mathbb{E}[X]^2.}$

Property: $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X,Y)$,

where $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$

↳ if X, Y ind., $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$

Variance of indicators: $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

$\mathbb{E}[X^2] = \sum_{i=1}^{n} \mathbb{E}[Y_i^2] + \sum_{i \neq j} \mathbb{E}[Y_i Y_j].$

1 Variance Proofs

(a) Let X be a random variable. Prove that:

$$\mathrm{Var}(X) \geq 0$$

(b) Let $X_1, \ldots, X_n$ be random variables. Prove that:

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{cov}(X_i, X_j)$$

Hint: Without loss of generality we can assume that $\mathbb{E}[X_1] = \cdots = \mathbb{E}[X_n] = 0$. Why?

(c) Let $a_1, \ldots, a_n \in \mathbb{R}$, and $X_1, \ldots, X_n$ be random variables. Prove that:

$$\sum_{i=1}^{n} a_i^2 \cdot \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} a_i \cdot a_j \cdot \mathrm{cov}(X_i, X_j) \geq 0$$

(a) $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$

$= \sum_x \underbrace{(x - \mathbb{E}[X])^2}_{\geq 0} \cdot \mathbb{P}(X = x) \quad \geq 0$

≥ 0

$\mathbb{E}[X] = \mu$

$X' = X - \mu$

(b) $\mathrm{Var}(X + c) = \mathrm{Var}(X) \qquad \mathrm{Cov}(X + c, Y) = \mathrm{Cov}(X, Y)$

$\mathrm{Var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = \mathbb{E}[X_i^2]$

$\mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] = \mathbb{E}[X_i X_j]$

$\mathrm{Var}(X_1 + \cdots + X_n) = \mathbb{E}[(X_1 + \cdots + X_n)^2]$

$(X_1 + X_2)^2 = X_1^2 + X_2^2 + 2 \cdot X_1 X_2$ $\qquad (\underbrace{\mathbb{E}[X_1 + \cdots + X_n]}_{=0})^2$

$= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2 \sum_{i<j} \mathbb{E}[X_i X_j]$

$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j)$ ✓

(c).
$$\sum_{i=1}^{n} a_i^2 \cdot \mathrm{Var}(X_i) + 2 \sum_{1\le i<j\le n} a_i \cdot a_j \cdot \mathrm{cov}(X_i, X_j) \ge 0$$

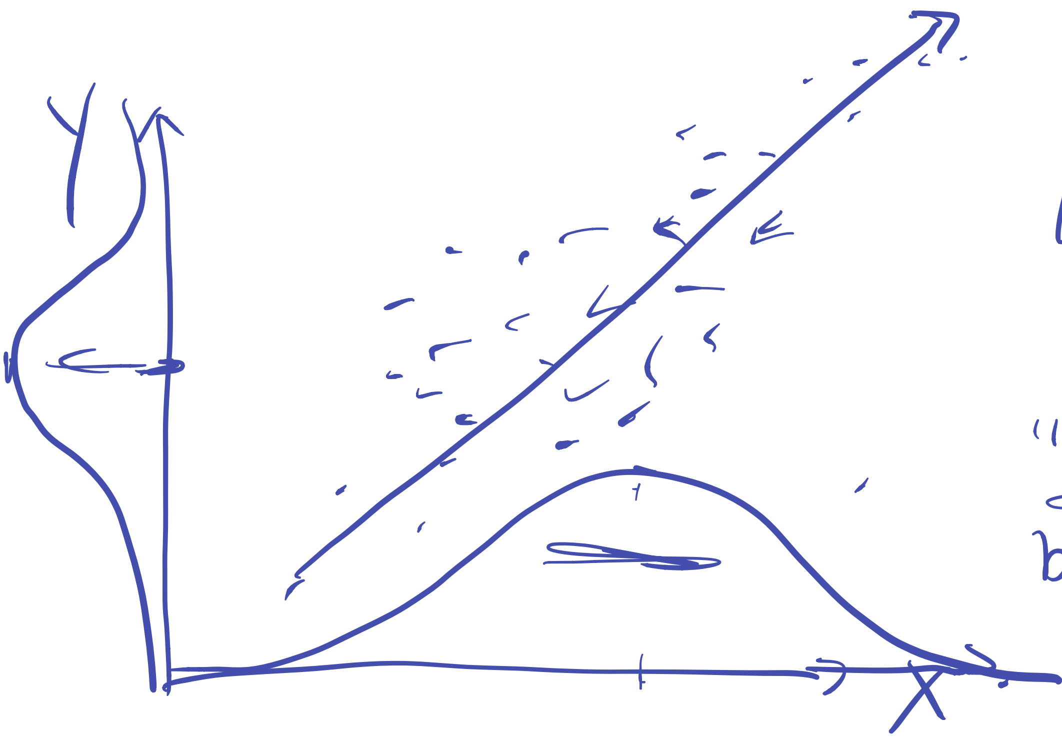
$$Var(aX) = a^2 Var(X).$$

$$ab\, Cov(X,Y) = Cov(aX, bY).$$

$$\sum_{i=1}^{n} Var(a_i X_i) + 2\sum_{i<j} cov(a_i X_i, a_j X_j) \ge 0.$$

part b). $\hookrightarrow = Var(a_1X_1 + a_2X_2 + \dots + a_nX_n) \geq 0$

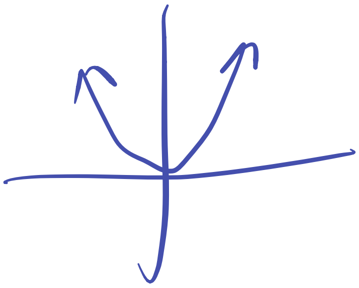
part a). ✓.



Cov(X,Y) measures the "linear association" b/w X and Y.

$$\frac{Cov(X,Y)}{\sigma_X \sigma_Y} = corr(X,Y).$$

$X = \{1, -1\}$



$Y = X^2.$

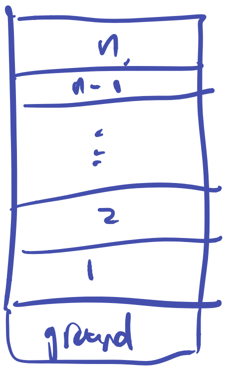
$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

$$= 0 - 0$$

$$= 0.$$

3 Variance

A building has n upper floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each person gets off at one of the n upper floors uniformly at random and independently of everyone else. What is the *variance* of the number of floors the elevator *does not* stop at?



m people

indicators?

$X :=$ # of floors the elevator does not stop at.

$X = X_1 + \cdots + X_n$

$$X_i = \begin{cases} 1 & \text{if elevator does not stop at floor } i \\ 0 & \text{o/w.} \end{cases}$$

$$\mathbb{P}(X_i = 1) = \left(1 - \frac{1}{n}\right)^m = \mathbb{E}[X_i].$$

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n \cdot \left(1 - \frac{1}{n}\right)^m.$$

$$\mathrm{Var}(X) = \underbrace{\mathbb{E}[X^2]} - \underline{\mathbb{E}[X]^2}$$

$X_i^2 = X_i$

$$\mathbb{E}[X^2] = \mathbb{E}[(X_1 + X_2 + \cdots + X_n)^2]$$

$$= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2 \sum_{i<j} \mathbb{E}[X_i X_j].$$

$$\mathbb{E}[X_i^2] = \mathbb{E}[X_i] = \left(1 - \frac{1}{n}\right)^m.$$

$$\mathbb{E}[X_i X_j] = \mathbb{P}(X_i = 1 \cap X_j = 1) = \left(\frac{n-2}{n}\right)^m.$$

$$\mathbb{E}[X^2] = \sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^m + 2\sum_{i<j} \left(\frac{n-2}{n}\right)^m$$

$$= n\left(1 - \frac{1}{n}\right)^m + 2 \cdot \binom{n}{2} \cdot \left(\frac{n-2}{n}\right)^m$$

$$\boxed{\mathrm{Var}(X) = n\left(1 - \frac{1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m - n^2\left(1 - \frac{1}{n}\right)^{2m}}$$