1 Variance Proofs

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

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

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

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

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

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

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

Solution:

(a) By definition, we have that

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \sum_k (k - \mathbb{E}[X])^2 \cdot \mathbb{P}(X = k) \end{aligned}$$

Each term in this summation is non-negative, since it is the product of a squared number and a probability, both of which are guaranteed to be non-negative. Since the sum of a bunch of non-negative terms must itself be non-negative, we have that $\text{Var}(X) \geq 0$.

(b) We first note that for a constant c, $\text{Var}(X + c) = \text{Var}(X)$ and similarly that $\text{cov}(X + c, Y) = \text{cov}(X, Y)$. Thus, we can subtract $\mathbb{E}[X_i]$ from X_i without changing anything in our target equality; this reduces us to the case where all the means are zero. Hence, we can write

$$\begin{aligned} \text{Var}(X_1 + \cdots + X_n) &= \mathbb{E}\left[(X_1 + \cdots + X_n)^2\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 + 2 \sum_{1 \leq i < j \leq n} X_i X_j\right] \\ &= \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i X_j] \\ &= \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j) \end{aligned}$$

where we used the fact that the means were all zero in the first and last steps to simplify the definitions of variance and covariance.

(c) We start with the left hand side of our desired inequality. Factoring the constants into the variance and covariance, we get

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

By part (b), we have that

$$\sum_{i=1}^{n} \mathrm{Var}(a_i X_i) + 2 \sum_{1\le i<j\le n} \mathrm{cov}(a_i X_i, a_j X_j) = \mathrm{Var}(a_1 X_1 + \cdots + a_n X_n)$$

Since $a_1 X_1 + \ldots + a_n X_n$ is a discrete random variable, part (a) tells us that its variance is non-negative. Putting these all together, we get that

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

2 Subset Card Game

Jonathan and Yiming are playing a card game. Jonathan has $k > 2$ cards, and each card has a real number written on it. Jonathan tells Yiming (truthfully), that the sum of the card values is 0, and that the sum of squares of the values on the cards is 1. Specifically, if the card values are $c_1, c_2, \ldots, c_k$, then we have $\sum_{i=1}^{k} c_i = 0$ and $\sum_{i=1}^{k} c_i^2 = 1$.

The cards are then going to be dealt randomly in the following fashion: for each card in the deck, a fair coin is flipped. If the coin lands heads, then the card goes to Yiming, and if the coin lands tails, the card goes to Jonathan. Note that it is possible for either player to end up with no cards/all the cards.

Calculate $\mathrm{Var}(S)$, where S is the sum of value of cards in Yiming's hand. The answer should not include a summation.

Solution: Let I_i be the indicator random variable indicating whether or not card i goes to Yiming. We have $S = \sum_{i=1}^{k} c_i I_i$ as the value of Yiming's hand. Then, we see that $\mathbb{E}[S] = \sum_{i=1}^{k} c_i \cdot \frac{1}{2} = 0$ and

$$\begin{aligned} \mathrm{Var}(S) &= \sum_{i=1}^{k} \mathrm{Var}(c_i I_i) \qquad \text{(due to independence) of } I_i \\ &= \sum_{i=1}^{k} c_i^2 \,\mathrm{Var}(I_i) \end{aligned}$$

We know that I_i is a Bernoulli random variable, so its variance is $\frac{1}{4}$. Thus, we see that $\mathrm{Var}(S) = \frac{1}{4}$.

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?

Solution: Let N be the number of floors the elevator does not stop at. We can represent N as the sum of the indicator variables $I_1, \ldots, I_n$, where $I_i = 1$ if no one gets off on floor i. Thus, we have

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

and from linearity of expectation,

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

To find the variance, we cannot simply sum the variance of our indicator variables. However, since $\text{Var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2$ the only piece we don't already know is $\mathbb{E}[N^2]$. We can calculate this by again expanding N as a sum:

$$\mathbb{E}[N^2] = \mathbb{E}[(I_1 + \cdots + I_n)^2] = \mathbb{E}\left[\sum_{i,j} I_i I_j\right] = \sum_{i,j} \mathbb{E}[I_i I_j] = \sum_{i}^{n} \mathbb{E}[I_i^2] + \sum_{i \neq j} \mathbb{E}[I_i I_j].$$

The first term is simple to calculate: since I_i is an indicator, $I_i^2 = I_i$, so we have

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

meaning that

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

From the definition of the variables I_i, we see that $I_i I_j = 1$ when both I_i and I_j are 1, which means no one gets off the elevator on floor i and floor j. This happens with probability

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

Thus we now know

$$\sum_{i \neq j} \mathbb{E}[I_i I_j] = n(n-1)\left(\frac{n-2}{n}\right)^m,$$

and we can assemble everything we've done so far to see that

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