

Lecture 26

Lecturer: Debmalya Panigrahi

Scribe: Kevin Sun

1 Overview

In the last lecture, we began studying random variables and various aspects associated with them, such as independence, expectation, and variance. In this lecture, we will continue studying properties about variance, and we will also see some useful inequalities concerning random variables.

2 Variance of Random Variables

Recall that a *random variable* is a total function whose domain is a sample space S . The codomain of a random variable is often \mathbb{R} or $\{0, 1\}$, but in general, it can be any set.

Let $X : S \rightarrow \mathbb{R}$ be a random variable. Recall that the *expectation* of X represents the “average value” of X , and can be written in the following two ways:

$$E[X] = \sum_{\omega \in S} X(\omega) \cdot \Pr[\omega] = \sum_{x \in \mathbb{R}} x \cdot \Pr[X = x]. \quad (1)$$

Another important quantity associated with X is its *variance*, given as follows:

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

(Note that $E^2[X]$ denotes the quantity $(E[X])^2$.) Intuitively, the variance of X measures how “spread out” the values of X are: if $\text{Var}[X] = 0$, then X is a constant, and if $\text{Var}[X]$ is high, then X often takes value far from its expectation. Finally, the *standard deviation* of X is the positive square root of its variance.

Geometric random variable: Recall the following random process: we have a coin that results in H with probability p , and we repeatedly flip this coin until we obtain a head H . Thus, each outcome in the sample space is a sequence of (possibly 0) tails, followed by a single head. We can define a random variable X on this sample space as follows:

$$\forall s \in S. X(s) \text{ is equal to the length of } s.$$

Notice that $\Pr[X = x] = (1 - p)^{x-1}p$ for every $x \in \mathbb{Z}^+$. Such a random variable X is often known as a *geometric* random variable with parameter p . (Intuitively, the value of p can be interpreted as the probability of “success,” i.e., the probability that the experiment ends at each flip.)

Recall that the expectation of this random variable is $1/p$. Now we will calculate its variance:

$$\text{Var}[X] = E[X^2] - E^2[X] = E[X^2] - \frac{1}{p^2}. \quad (3)$$

Let $S = E[X^2]$, and notice from the right-hand side of (2), we can write S as follows:

$$\begin{aligned} S &= 1^2p + 2^2(1-p)p + 3^2(1-p)^2p + 4^2(1-p)^3p + \cdots \\ (1-p) \cdot S &= 1^2(1-p)p + 2^2(1-p)^2p + 3^2(1-p)^3p + \cdots \end{aligned}$$

Subtracting the second equation from the first equation yields

$$pS = p + 3(1-p)p + 5(1-p)^2p + 7(1-p)^3p + \cdots,$$

which implies

$$S = 1 + 3(1-p) + 5(1-p)^2 + 7(1-p)^3 + \cdots.$$

and

$$(1-p)S = (1-p) + 3(1-p)^2 + 5(1-p)^3 + \cdots.$$

Again, we subtract $(1-p)S$ from S to obtain the following:

$$\begin{aligned} pS &= 1 + 2(1-p) + 2(1-p)^2 + 2(1-p)^3 + \cdots \\ &= 1 + 2(1-p) \left[1 + (1-p) + (1-p)^2 + \cdots \right] \\ &= 1 + 2(1-p) \cdot \frac{1}{1 - (1-p)} \\ &= \frac{2-p}{p}. \end{aligned}$$

Since we initially let $S = E[X^2]$, substituting this into (3) yields

$$\text{Var}[X] = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

We will now prove a theorem that can often be applied to calculate the variance of random variables more generally. Recall that if two random variables X and Y are *independent*, then $E[XY] = E[X] \cdot E[Y]$.

Theorem 1. *If X and Y are independent random variables, then*

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y].$$

Remark: This statement of this theorem is similar to that of linearity of expectation $E[X + Y] = E[X] + E[Y]$. However, for variance, independence of X and Y is required, whereas linearity of expectation holds regardless of whether the random variables in consideration are independent.

Proof. By the equation given in (2), we have

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

Consider the first term of the expression above:

$$E[(X + Y)^2] = E[X^2 + 2XY + Y^2] = E[X^2] + 2 \cdot E[XY] + E[Y^2], \quad (4)$$

and now the second:

$$E^2[X + Y] = (E[X + Y])^2 = E^2[X] + 2 \cdot E[X] \cdot E[Y] + E^2[Y]. \quad (5)$$

Since X and Y are independent, we know that $2 \cdot E[XY] = 2 \cdot E[X] \cdot E[Y]$. Thus, subtracting (5) from (4) yields

$$\begin{aligned} \text{Var}[X + Y] &= E[X^2] - E^2[X] + E[Y^2] - E^2[Y] \\ &= \text{Var}[X] + \text{Var}[Y], \end{aligned}$$

where the second equality again holds from (2). \square

Binomial random variable: Now consider the following process: we flip a coin n times, and each result is H with probability p and independent of the other results. We define a random variable X that maps each outcome to the number of H 's it contains. Such a random variable is known as a *binomial* random variable with parameters n and p . We've already seen that this random variable has expectation np ; we now consider its variance.

Corollary 2. *The variance of a binomial random variable X with parameters n and p is $np(1 - p)$.*

Proof. We shall write X as the sum of n Bernoulli random variables: for every $i \in \{1, 2, \dots, n\}$, we define a random variable X_i as 1 if the i -th flip returned H , and 0 otherwise. Now notice that

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

In other words, the total number of heads can be calculated by summing independent and identical Bernoulli random variables over the n flips. We've seen that the variance of each X_i is $p(1 - p)$.

Furthermore, since the outcome of the coin flips are independent of one another, we can apply Theorem 1 (repeatedly):

$$\text{Var}[X] = \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_n] = np(1 - p). \quad \square$$

3 Deviation from Expectation

We will conclude our study of random variables by looking at some useful and well-known bounds. These bounds apply to a broad class of random variables and are generally used to understand the probability that a random variable deviates from its expectation.

Theorem 3 (Markov's Bound). *Let X be a non-negative random variable, and let k be a positive real number. Then*

$$\Pr[X > k] < \frac{E[X]}{k}.$$

Proof. Let p denote the probability that X is at least k , i.e., $p = \Pr(X > k)$. Recall from (1) that we can write the expectation of X as follows:

$$\begin{aligned} E[X] &= \sum_{x \in \mathbb{R}} x \cdot \Pr[X = x] \\ &= \sum_{x > k} x \cdot \Pr[X = x] + \sum_{x \leq k} x \cdot \Pr[X = x] \\ &> k \sum_{x > k} \Pr[X = x] + \sum_{x < k} 0 \cdot \Pr[X = x] \\ &= k \cdot \Pr[X > k], \end{aligned}$$

Dividing the expressions above by k yields the desired result. \square

We now apply Theorem 3 on a particular non-negative random variable to obtain another well-known bound that is often more useful.

Theorem 4 (Chebyshev's Bound). *Let X be a random variable, and let σ_X denote the standard deviation of X . Then for any $c \in \mathbb{R}^+$, the following inequality holds:*

$$\Pr \left[|X - E[X]| > c \cdot \sigma_X \right] < \frac{1}{c^2}.$$

Proof. Consider the random variable $(X - E[X])^2$. Notice that the expectation of this random variable is precisely σ_X^2 , i.e., the variance of X . Furthermore, notice that this real variable is always non-negative. Thus, for any $k \in \mathbb{R}^+$, the following holds:

$$\Pr \left[(X - E[X])^2 > k \right] < \frac{E[(X - E[X])^2]}{k} = \frac{\text{Var}[X]}{k}.$$

In particular, we can set $k = c^2 \cdot \text{Var}[X]$ to obtain

$$\Pr \left[(X - E[X])^2 > c^2 \cdot \text{Var}[X] \right] < \frac{1}{c^2}.$$

Since we know that c is positive, the event considered above is precisely the event considered in the statement of the theorem. Thus, the inequality holds. \square

Example 1: Now let us apply these bounds in the following scenario: we toss a fair coin n times, and let X denote the number of resulting H 's (i.e., X is a binomial random variable with parameters n and $p = 1/2$). Then $E[X] = np = n/2$, that is, we expect to see $n/2$ heads. But in reality, the number of heads almost certainly will not be exactly $n/2$, so we ask ourselves: what is the probability that X exceeds its expectation by over \sqrt{n} ? In other words, we seek the value of

$$\Pr \left[X > \frac{n}{2} + \sqrt{n} \right].$$

Observe that Theorem 3 tells us

$$\Pr \left[X > \frac{n}{2} + \sqrt{n} \right] < \frac{n/2}{n/2 + \sqrt{n}},$$

but this bound is not very illustrative. If n is large, then this number is only slightly smaller than 1, which is a trivial upper bound on the value we seek. (For example, if $n = 1024$, then the above bound tells us the probability is at most 0.94, which is close to 1.)

In contrast, Theorem 4 gives a much tighter bound. The variance of X is $np(1-p) = n/4$ (see Corollary 2) so its standard deviation is $\sigma_X = \sqrt{n}/2$. This implies the following:

$$\Pr \left[X > \frac{n}{2} + \sqrt{n} \right] < \Pr \left[\left| X - \frac{n}{2} \right| > \sqrt{n} \right] < \frac{1}{4}.$$

The first inequality holds because the event on the left-hand side is a strict subset of the event on the right-hand side, and the second inequality follows from Theorem 4 by setting $c = 2$ so that $c \cdot \sigma_X = \sqrt{n}$.

4 Summary

In this lecture, we concluded our study of random variables. We computed the variance of well-known random variables, proved a general property about the variance of the sum of independent random variables, and derived some well-known inequalities that bound the probability a random variable deviates from its mean.