# CSCE-658 Randomized Algorithms

Lecture #10, March 8, 2016

Lecturer: Professor Jianer Chen

# 10 Variance and Chebyshev's Inequality

In the previous sections, we have seen that the expectation of a random variable sometimes may not be very informative. However, when further information about the value distribution of the random variable is provided, such as its upper bound and lower bound, then we have a better control on the random variable values based on its expectation. In this section, we further explore this line of study.

## 10.1 The variance of a random variable

The upper bound and lower bound on the values of a random variable give the maximum possible deviation of the random variable value from its expectation, which, as we have seen in the previous sections, thus provide useful information for the distribution of the random variable values in terms of its expectation. Naturally, measures on the "average deviation" of the random variable value from it e expectation will provide further information for the value distribution of the random variable. This is given by the *variance* of a random variable. A formal definition is given as follows.

**Definition 10.1** The variance of a random variable X is defined as  $\operatorname{var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$ .

We make some comments on the definition of variance. Variance is used to measure the difference between the value and the expectation of a random variable. One might think  $\mathbf{E}[X - \mathbf{E}[X]]$  is a more direct measurement for this purpose. However, by Linearity of Expectation,

$$\mathbf{E}[X - \mathbf{E}[X]] = \mathbf{E}[X] - \mathbf{E}[\mathbf{E}[X]] = \mathbf{E}[X] - \mathbf{E}[X] = 0,$$

(well, this in fact meets our intuition: the average difference between the value and the average value is 0). Thus, the expected *absolute* difference  $\mathbf{E}[|X - \mathbf{E}[X]|]$  should be used to measure the difference between the value and the expectation of the random variable X. However, the absolute value function  $|X - \mathbf{E}[X]|$  is hard to handle (e.g., it is not differentiable). On the other hand,  $(X - \mathbf{E}[X])^2$  is a much nicer function. Note that for any positive number t,  $(X - \mathbf{E}[X])^2 \leq t^2$  and  $|X - \mathbf{E}[X]| \leq t$  define the same event. So the variance  $\mathbf{var}[X]$  defined based on the expected value of  $(X - \mathbf{E}[X])^2$  serves well for measuring the absolute difference between the value and the expectation of the random variable X.

Lemma 10.1  $\operatorname{var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ , which also gives  $\mathbb{E}[X^2] \ge (\mathbb{E}[X])^2$ .

**PROOF.** By Linearity of Expectation, we have (note that  $\mathbf{E}[X]$  is a constant):

$$\mathbf{var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2 - 2X \cdot \mathbf{E}[X] + (\mathbf{E}[X])^2] = \mathbf{E}[X^2] - \mathbf{E}[2X \cdot \mathbf{E}[X]] + \mathbf{E}[(\mathbf{E}[X])^2] = \mathbf{E}[X^2] - 2\mathbf{E}[X] \cdot \mathbf{E}[X] + (\mathbf{E}[X])^2 = \mathbf{E}[X^2] - (\mathbf{E}[X])^2.$$

The inequality  $\mathbf{E}[X^2] \ge (\mathbf{E}[X])^2$  is because  $(X - \mathbf{E}[X])^2 \ge 0$  gives  $\mathbf{var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] \ge 0$ .  $\Box$ 

The variance of the sum of independent random variables satisfies the following property:

**Lemma 10.2** Let X and Y be two independent random variables, then var[X+Y] = var[X] + var[Y].

PROOF. By definition, we have:

$$\mathbf{var}[X+Y] = \mathbf{E}[((X+Y) - \mathbf{E}[X+Y])^{2}]$$
  
=  $\mathbf{E}[((X-\mathbf{E}[X]) + (Y-\mathbf{E}[Y]))^{2}]$   
=  $\mathbf{E}[(X-\mathbf{E}[X])^{2} + (Y-\mathbf{E}[Y])^{2} + 2(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$   
=  $\mathbf{E}[(X-\mathbf{E}[X])^{2}] + \mathbf{E}[(Y-\mathbf{E}[Y])^{2}] + 2\mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$   
=  $\mathbf{var}[X] + \mathbf{var}[Y] + 2\mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])].$  (23)

Since X and Y are independent, the random variables  $X' = X - \mathbf{E}[X]$  and  $Y' = Y - \mathbf{E}[Y]$  are also independent. By Lemma 10.2, we have

$$2\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] = 2 \cdot \mathbf{E}[X - \mathbf{E}[X]] \cdot \mathbf{E}[Y - \mathbf{E}[Y]].$$

By Linearity of Expectation,  $\mathbf{E}[X - \mathbf{E}[X]] = \mathbf{E}[X] - \mathbf{E}[\mathbf{E}[X]] = 0$  (similarly  $\mathbf{E}[Y - \mathbf{E}[Y]] = 0$ ). Bringing this back to (23) gives  $\mathbf{var}[X + Y] = \mathbf{var}[X] + \mathbf{var}[Y]$ .  $\Box$ 

Obviously, Lemma 10.2 can be extended to the case of more than two independent random variables. On the other hand, we should not expect that Lemma 10.2 be extended to general linear combinations of independent random variables, as we did for the linearity of expectation in Theorem 8.1: for a constant c, by the definition we will have  $\operatorname{var}[cX] = \mathbf{E}[(cX - \mathbf{E}[cX])^2] = c^2 \cdot \operatorname{var}[X]$ .

The function that directly measures the absolute difference between the values and the expectation of a random variable X is called the *standard deviation*  $\sigma[X]$  which is defined to be the positive square root of the variance of X:  $\sigma[X] = \sqrt{\operatorname{var}[X]}$ . Thus, the variance  $\operatorname{var}[X]$  is also written as  $\sigma^2[X]$ .

There are some other functions that describe the value distribution of random variables. We give the definitions below. These functions will become more important when we study probability spaces with uncountable sample spaces (for example, the sample space is the set of all points in the unit square in the 2D Euclidean space).

### **Definition 10.2** Let X and Y be random variables.

- 1. The density function for X is defined as  $p_X(u) = \Pr[X = u];$
- 2. The distribution function for X is defined as  $F_X(u) = \Pr[X \le u]$ .
- 3. The joint density function for X and Y is defined as  $p_{X,Y}(u,v) = \Pr[X = u \cap Y = v];$
- 3. The joint distribution function for X and Y is defined as  $F_{X,Y}(u,v) = \Pr[X \le u \cap Y \le v]$ .

## **10.2** Computing expectations and variances

There are certain important classes of random variables, that will be introduced in this subsection. We will show how the expectation and the variance of these random variables are computed. Throughout the discussion in this subsection, we assume that p is a real number with 0 .

We start with a very simple case. Suppose that we have a (probably biased) coin  $\kappa_p$  that shows head with probability p and tail with probability 1 - p. Thus, tossing the coin  $\kappa_p$  (only once) constitutes a probability space  $\Omega_1$  whose sample space has only two outcomes H and T such that  $\Pr[H] = p$ and  $\Pr[T] = 1 - p$ . Define a random variable  $X_1$  on this probability space such that  $X_1(H) = 1$  and  $X_1(T) = 0$ . Thus, in terms of its density function, we have  $\Pr[X_1 = 1] = p$  and  $\Pr[X_1 = 0] = 1 - p$ . In fact, we do not have to restrict to the probability space  $\Omega_1$  with only two outcomes. If a random variable  $X_1$  on an (arbitrary) probability space has its density function satisfy  $\Pr[X_1 = 1] = p$  and  $\Pr[X_1 = 0] = 1 - p$ , we will call  $X_1$  a *Bernoulli random variable* with parameter p (or say that  $X_1$  is a random variable of *Bernoulli distribution* with parameter p).

The expectation  $\mathbf{E}[X_1]$  of a Bernoulli random variable  $X_1$  with parameter p can be easily computed, using Formula (20):

$$\mathbf{E}[X_1] = \Pr[X_1 = 1] \cdot 1 + \Pr[X_1 = 0] \cdot 0 = p.$$
(24)

To compute the variance  $\mathbf{var}[X_1]$  of the Bernoulli random variable  $X_1$  with parameter p, we use Lemma 10.1, noting that  $\Pr[X_1^2 = 1] = p$  and  $\Pr[X_1^2 = 0] = 1 - p$ :

$$\mathbf{var}[X_1] = \mathbf{E}[X_1^2] - (\mathbf{E}[X_1])^2 = p - p^2 = p(1-p).$$
(25)

Now we consider the experiment of tossing the coin  $\kappa_p n$  times. Thus, the outcomes of the sample space are all strings of length n in  $\{H, T\}^n$ , where an outcome containing h H's and n - h T's has a probability  $p^h(1-p)^{n-h}$  (students should verify that this truly makes a valid probability space  $\Omega_2$  because  $\sum_{h=0}^n {n \choose h} p^h(1-p)^{n-h} = 1$ ). Define a random variable  $X_2$  that for an outcome  $\omega$  in the probability space  $\Omega_2$ ,  $X_2(\omega)$  is equal to the number of H's in  $\omega$ . Since there are  ${n \choose h}$  ways to pick h positions in a string of length n and place the symbol H in these h positions (and place the symbol T in the other n - h positions), we derive that the density function for the random variable  $X_2$  is  $\Pr[X_2 = h] = \binom{n}{h}p^h(1-p)^{n-h}$  for all integers  $0 \le h \le n$ . Again, for any random variable  $X_2$  on an arbitrary probability space (not necessary the above probability space  $\Omega_2$ ), whose range is  $\{0, 1, 2, \ldots, n\}$  and density function is  $\Pr[X_2 = h] = \binom{n}{h}p^h(1-p)^{n-h}$  for all  $0 \le h \le n$ , we will call  $X_2$  a binomial random variable with parameters n and p (or say that the random variable  $X_2$  has a binomial distribution with parameters n and p).

Note that the binomial random variable  $X_2$  can be represented by a sum of independent Bernoulli random variables. For this, define the Bernoulli random variable  $Y_i$ , for  $1 \le i \le n$ , on the probability space  $\Omega_2$  such that for any outcome  $\omega$  in  $\Omega_2$ ,  $Y_i = 1$  if the *i*-th symbol in  $\omega$  is H, and  $Y_i = 0$  otherwise. Clearly,  $\Pr[Y_i = 1] = p$  and  $\Pr[Y_i = 0] = 1 - p$ . Thus,  $Y_i$  is a Bernoulli random variable with parameter p. Moreover, since the n tosses of the coin  $\kappa_p$  are independent, for any  $i \ne j$ , the random variables  $Y_i$ and  $Y_j$  are independent. Now the binomial random variable  $X_2$  is clearly equal to  $Y_1 + Y_2 + \cdots + Y_n$ .

Therefore, by Linearity of Expectation and using (24), we have

$$\mathbf{E}[X_2] = \mathbf{E}[Y_1 + Y_2 + \dots + Y_n] = \mathbf{E}[Y_1] + \mathbf{E}[Y_2] + \dots + \mathbf{E}[Y_n] = np.$$
(26)

By Lemma 10.2 (see also the remark after the proof of Lemma 10.2) and using (25), we have

$$\mathbf{var}[X_2] = \mathbf{var}[Y_1 + Y_2 + \dots + Y_n] = \mathbf{var}[Y_1] + \mathbf{var}[Y_2] + \dots + \mathbf{var}[Y_n] = np(p-1).$$
(27)

We consider the binomial random variable with a special (and the most popular) parameter p = 1/2. Then (26) shows that if you toss the coin  $\kappa_{1/2}$  *n* times, then the expected number of times you will see the head is np = n/2. This certainly matches our intuition. Moreover, by (27), the variance is np(p-1) = n/4. Remember that the variance measures the square of the deviation of the random variable values from its expectation. Thus, the expected deviation of the binomial random variable with parameter 1/2 from its expectation is really  $\sqrt{n/4} = \sqrt{n}/2$ . Note that compared to the expectation n/2 of the random variable,  $\sqrt{n}/2$  is much smaller. This gives us a feeling that the binomial random variable is quite concentrated around its expectation.

Now we return back to the game of tossing the coin  $\kappa_p$ , but this time we change the game rule: we repeatedly toss the coin  $\kappa_p$  but stop when the first head turns up. Thus, the outcomes of this probability space  $\Omega_3$  are the strings of the form  $TT \cdots TH$  (i.e., an arbitry number of T's followed by a single H). The probability on an outcome  $T^hH$  is  $(1-p)^hp$ . (Again, students should verify that this makes a valid probability space  $\Omega_3$  because  $\sum_{h=0}^{\infty}(1-p)^hp=1$ ). Define a random variable  $X_3$  on  $\Omega_3$ that for an outcome  $\omega$  in  $\Omega_3$ ,  $X_3(\omega)$  is equal to the length of  $\omega$  (thus,  $X_3(\omega)$  is equal to the number of tosses given by the outcome  $\omega$ ). Again, for any random variable  $X_3$  with the range  $\{1, 2, 3, \ldots\}$  and density function  $\Pr[X_3 = h] = (1-p)^{h-1}p$  for all  $h \ge 1$ , we will call  $X_3$  a geometric random variable with parameter p (or say that the random variable  $X_3$  has a geometric distribution with parameter p).

Before we discuss expectation and rariance of geometric random variables, we prove a useful lemma.

**Lemma 10.3** For any  $p, 0 , <math>\sum_{h=1}^{\infty} h(1-p)^{h-1} = 1/p^2$ .

PROOF. Let  $S = \sum_{h=1}^{\infty} h(1-p)^{h-1}$ . We have

$$S = \sum_{h=1}^{\infty} h(1-p)^{h-1} = 1 + \sum_{h=2}^{\infty} h(1-p)^{h-1} = 1 + \sum_{h=2}^{\infty} ((h-1)+1)(1-p)^{h-2}(1-p)$$
$$= 1 + (1-p)\sum_{h=2}^{\infty} (h-1)(1-p)^{h-2} + (1-p)\sum_{h=2}^{\infty} (1-p)^{h-2}$$
$$= 1 + (1-p)\sum_{h=1}^{\infty} h(1-p)^{h-1} + (1-p) \cdot \frac{1}{p} = 1 + (1-p) \cdot S + \frac{1-p}{p}$$

Solving S = 1 + (1 - p)S + (1 - p)/p gives  $S = 1/p^2$ .

The trick in the proof of Lemma 10.3 is to set the infinite sum as a variable S, then rewrite the infinite sum so that S is given in an equation of a losed form and we can solve the equation to get the value of S. This trick will be useful in our combinatorial analysis, in particular it will be used when we compute the expectation and rariance of geometric random variables.

By definition, the expectation of the geometric random variable  $X_3$  with parameter p is given by

$$\mathbf{E}[X_3] = \sum_{h=1}^{\infty} h \cdot \Pr[X_3 = h] = \sum_{h=1}^{\infty} h(1-p)^{h-1}p = p \sum_{h=1}^{\infty} h(1-p)^{h-1}.$$

Applying Lemma 10.3, we get  $\mathbf{E}[X_3] = p \sum_{h=1}^{\infty} h(1-p)^{h-1} = p/p^2 = 1/p$ . Now we consider the variance of  $X_3$ . By Lemma 10.1,  $\mathbf{var}[X_3] = \mathbf{E}[X_3^2] - (\mathbf{E}[X_3])^2$ . By the above discussion, we know  $(\mathbf{E}[X_3])^2 = 1/p^2$ . Thus, we only need to compute  $\mathbf{E}[X_3^2]$ .

The random variable  $X_3^2$  only takes values that are perfect squares. Moreover, for any positive integer h,  $[X_3^2 = h^2]$  and  $[X_3 = h]$  define the same event. Thus,  $\Pr[X_3^2 = h^2] = (1-p)^{h-1}p$ . So we have

$$\mathbf{E}[X_3^2] = \sum_{h=1}^{\infty} h^2 \cdot \Pr[X_3^2 = h^2] = \sum_{h=1}^{\infty} h^2 (1-p)^{h-1} p = p \sum_{h=1}^{\infty} h^2 (1-p)^{h-1}.$$
 (28)

So we need to compute the sum  $\sum_{h=1}^{\infty} h^2 (1-p)^{h-1}$ . As we did in the proof of Lemma 10.3, let  $T = \sum_{h=1}^{\infty} h^2 (1-p)^{h-1}$ . Then we have

$$T = \sum_{h=1}^{\infty} h^2 (1-p)^{h-1} = 1 + \sum_{h=2}^{\infty} h^2 (1-p)^{h-1} = 1 + (1-p) \sum_{h=2}^{\infty} ((h-1)+1)^2 (1-p)^{h-2}$$
  
$$= 1 + (1-p) \sum_{h=1}^{\infty} (h+1)^2 (1-p)^{h-1} = 1 + (1-p) \sum_{h=1}^{\infty} (h^2+2h+1)(1-p)^{h-1}$$
  
$$= 1 + (1-p) \left( \sum_{h=1}^{\infty} h^2 (1-p)^{h-1} + 2 \sum_{h=1}^{\infty} h(1-p)^{h-1} + \sum_{h=1}^{\infty} (1-p)^{h-1} \right)$$
  
$$= 1 + (1-p) \left( T + \frac{2}{p^2} + \frac{1}{p} \right)$$

where in the last equality, we replaced the infinite sum  $\sum_{h=1}^{\infty} h^2 (1-p)^{h-1}$  by T, applied Lemma 10.3 to get  $\sum_{h=1}^{\infty} h(1-p)^{h-1} = 1/p^2$ , and used the formula of sum for the geometric seriese  $\sum_{h=1}^{\infty} (1-p)^{h-1} = 1/(1-(1-p)) = 1/p$ . Now Solving  $T = 1 + (1-p)(T+2/p^2+1/p)$  gives  $T = (2-p)/p^3$ . Bringing this in (28) we get  $\mathbf{E}[X_3^2] = (2-p)/p^2$ . So  $\mathbf{var}[X_3] = \mathbf{E}[X_3^2] - (\mathbf{E}[X_3])^2 = (2-p)/p^2 - 1/p^2 = (1-p)/p^2$ . We summarize the discussion in this subsection in the following theorem.

**Theorem 10.4** Let p be a real number, 0 , and let <math>n be an integer,

- (1) for Bernoulli random variable  $X_1$  with parameter p,  $\mathbf{E}[X_1] = p$  and  $\mathbf{var}[X_1] = p(1-p)$ ;
- (2) for binomial random variable  $X_2$  with parameters n and p,  $\mathbf{E}[X_2] = np$  and  $\mathbf{var}[X_2] = np(1-p)$ ;
- (3) for geometric random variable  $X_3$  with parameter p,  $\mathbf{E}[X_2] = 1/p$  and  $\mathbf{var}[X_3] = (1-p)/p^2$ .

#### 10.3Chebyshe's Inequality

When the variance of a random variable X is known, we have the following Chebyshe's Inequality that provides useful information about the value distribution of the random variable X in terms of its expectation  $\mathbf{E}[X]$  and variance  $\mathbf{var}[X]$ .

**Theorem 10.5** (Chebyshe's Inequality) Let X be a random variable and let t > 0 be a constant. Then  $\Pr[|X - \mathbf{E}[X]| \ge t] \le \operatorname{var}[X]/t^2.$ 

Since X is a random variable,  $X' = (X - \mathbf{E}[X])^2$  is also a random variable that takes PROOF.

only non-negative values. Moreover,  $t^2$  is a positive number. Thus, we can apply Markov Inequality (Theorem 9.1) on X', which gives

$$\Pr[X' \ge t^2] \le \frac{\mathbf{E}[X']}{t^2} = \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{t^2} = \frac{\mathbf{var}[X]}{t^2}$$

Since t > 0,  $[X' \ge t^2]$  (which is  $[(X - \mathbf{E}[X])^2 \ge t^2]$ ) and  $[|X - \mathbf{E}[X]| \ge t]$  define the same event so  $\Pr[X' \ge t^2] = \Pr[|X - \mathbf{E}[X]| \ge t]$ , the theorem gets proved.  $\Box$ 

Similar to the way we handled expectation, if we let  $t = r \cdot \sigma[X]$  in Theorem ??, we will get the following result (note that  $\sigma^2[X] = \operatorname{var}[X]$ ), which is also called Chebyshev' Inequality.

**Corollary 10.6** (Chebyshev's Inequality) Let X be a random variable and let r > 0 be a constant. Then  $\Pr[|X - \mathbf{E}[X]| \ge r \cdot \sigma[X]] \le 1/r^2$ .

We show how Chebyshev's Inequality is able to provide more precise information for a random variable when it is compared with Markov Inequality. Consider the binomial random variable  $X_2$  with parameter p = 1/2. By Theorem 10.4, the expectation of  $X_2$  is  $\mathbf{E}[X_2] = n/2$ . Suppose that we ask "what is the probability that the value of  $X_2$  is larger than or equal to 3n/4?"

Using Markov Inequality, we get

$$\Pr[X_2 \ge 3n/4] \le \frac{n/2}{3n/4} = \frac{2}{3}.$$
(29)

On the other hand, if we use Chebyshev's Inequality, noting that  $[X_2 \ge 3n/4]$  and  $[(X_2 - n/2) \ge n/4]$  define the same event, and that  $\operatorname{var}[X_2] = n/4$ , then we get

$$\Pr\left[X_2 \ge \frac{3n}{4}\right] = \Pr\left[\left(X_2 - \frac{n}{2}\right) \ge \frac{n}{4}\right] \le \Pr\left[\left|X_2 - \frac{n}{2}\right| \ge \frac{n}{4}\right] \le \frac{\operatorname{var}[X_2]}{(n/4)^2} = \frac{4}{n}.$$
(30)

Comparing (29) and (30), you can see that Chebyshev's Inequality gives a much tighter bound for the probability for the event  $[X_2 \ge 3n/4]$ .