EE 178/278A Probabilistic Systems Analysis

Spring 2014 Tse/Hussami Lecture 8

We ended the last lecture by defining the variance and standard deviation of a random variable. These notions capture the randomness in the variable or its deviation around its mean.

Definition 8.1 (Variance): The *variance* of a random variable X is defined as

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

Definition 8.2 (Standard deviation): The *standard deviation* of a random variable X is defined as

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

In the above definition of the variance, $(X-\mu)^2$ is itself a random variable because it is a function of the random variable X. This new variable measures the deviation of X around its mean. The expectation of this deviation gives a measure of the spread of the distribution.

A random variable with 0 variance assigns every outcome to the same number (and therefore not random). In general, the variance will be a positive number which models the spread of the variable. The larger the variance, the larger the spread of the random variable. The mean of a random variable is interpreted as being the center of mass, and the variance represents the moment of inertia.

The following easy observation gives us a slightly different way to compute the variance that is easier in many cases.

Theorem 8.1: *For a r.v. X with mean* $\mathbb{E}[X] = \mu$, *we have* $\text{Var}(X) = \mathbb{E}[X^2] - \mu^2$.

Proof: From the definition of variance, we have

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

In the second step here, we used linearity of expectation and that $\mathbb{E}[\mu]^2 = \mu^2$ because μ is constant.□

$\mathbb{E}[X^2]$ is called the second moment of the random variable. So the variance can be written in terms of the first and second moments of the random variable.

Let us now look at some examples:

1. **Uniform distribution.** X is a random variable that takes on values $1,\dots,n$ with equal probability $1/n$ (i.e. X has a uniform distribution). The mean of X is equal to :

$$\mathbb{E}[X] = \frac{n+1}{2}$$

To compute the variance, we just need to compute $\mathbb{E}[X^2]$, which is a routine calculation:

$$\mathbb{E}[X^2] = \sum_{\omega} \mathbf{P}(\omega) \times X^2(\omega) = \sum_{i=1}^{n} \frac{1}{n} \times i^2 = \frac{(n+1)(2n+1)}{6}.$$

Finally we get that

$$\mathrm{Var}(X) = \frac{n^2-1}{12}, \qquad \sigma_X = \sqrt{\frac{n^2-1}{12}}.$$

(You should verify these.) Note that the variance is proportional to n^2, and the standard deviation is proportional to n, the range of the r.v..

2. **Binomial distribution.** Consider $X \sim Bin(n,p)$. Then $\mathbb{E}[X] = np$, and $\mathrm{Var}(X) = \mathbb{E}[X^2] - (np)^2$.

To compute $\mathbb{E}[X^2]$, let us start by directly trying to compute it from the distribution. In that case, we should first compute the distribution of the random variable $Y = X^2$ in the following way:

$$\mathbb{E}[Y] = \sum_b b \times \mathbf{P}(Y=b) = \sum_a a^2 \mathbf{P}(X=a) = \sum_{i=0}^{n} i^2 \binom{n}{i} p^i (1-p)^{n-i}.$$

where we have used the fact that X has a binomial distribution. (You will verify the second equality in the HW.)

This sum doesn't seem easy to evaluate. The computation can be done more easily if we write $X = X_1 + X_2 + \cdots + X_n$ where

$$X_i = \begin{cases} 1 & \text{if } i\text{th toss is Heads,} \\ 0 & \text{otherwise.} \end{cases}$$

X_i is called the *indicator* r.v. for the event "the ith toss is Heads"; it indicates when the event occurs. This gives:

$$\begin{aligned} \mathbb{E}[X^2] &= \mathbb{E}[(X_1 + \cdots + X_n)^2] && (1) \\ &= \mathbb{E}[X_1^2 + \cdots + X_n^2 + X_1X_2 + X_2X_3 + \cdots] && (2) \\ &= \mathbb{E}[X_1^2] + \cdots + \mathbb{E}[X_n^2] + \sum_{i \neq j} \mathbb{E}[X_iX_j] && (3) \\ &= np + n(n-1)p^2 && (4) \end{aligned}$$

In the third line here, we have used: $\mathbb{E}[X_i^2] = p$ and $\mathbb{E}[X_iX_j] = p^2$. To show that, let $Y = X_iX_j$,

$$Y = \begin{cases} 1 & \text{if both } i\text{th and } j\text{th flips are Heads,} \\ 0 & \text{otherwise.} \end{cases}$$

Y is another indicator random variable, and we can compute:

$$\mathbb{E}[X_iX_j] = \mathbf{P}(\text{both ith and jth flips are Heads}) = \mathbf{P}(X_i = X_j = 1) = \mathbf{P}(X_i = 1) \cdot \mathbf{P}(X_j = 1) = p^2,$$

(since $X_i = 1$ and $X_j = 1$ are independent events). Note that there are $n(n-1)$ pairs i, j with $i \neq j$.

Finally,

$$\mathrm{Var}(X) = np + n(n-1)p^2 - (np)^2 = np(1-p).$$

and

$$\sigma_X = \sqrt{np(1-p)}.$$

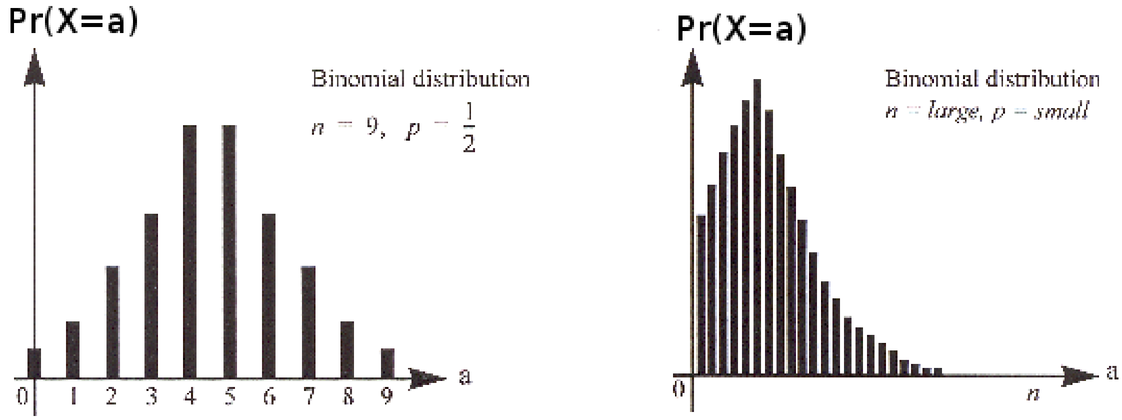

Figure 1: The binomial distributions for two choices of (n, p).

Note: Uniform and binomial random variables take values in the range from 0 to n. For the binomial distribution, the mean and variance are linear in n, and the standard deviation is linear in $\sqrt{n}$. For the uniform distribution, the mean is linear in n, the variance linear in n^2 and the standard deviation linear in n. The standard deviation of the binomial is shrunk relative to its range. This is because there is a mass concentration in a band around its mean (See figure 1). However, in the case of the uniform distribution, the standard deviation is more spread so of order n.

As another example to work out on your own: compute the variance of the number of students that get their homework back.

Some Important Distributions

We will now give some more examples of important distributions: $X \sim Bin(n, p)$,$X \sim Geom(p)$ and $X \sim Poiss(\lambda)$.

Geometric Distribution: Consider the experiment where we flip a coin until we see a head.

$$\Omega = \{H, TH, TTH, TTTH, \cdots\}$$

Let X denote the variable that counts the number of flips until and including the first Heads we obtain. There is a one to one correspondence between the outcomes and the random variable.

Assuming the results of the flips are independent, $\mathbf{P}(H) = p$, $\mathbf{P}(TH) = (1-p)p$. More generally, $\mathbf{P}(i-1 \text{ Tails before a Heads}) = (1-p)^{i-1}p$. So, the random variable X has the following distribution:

$$\mathbf{P}(X = i) = (1-p)^{i-1}p \text{ for } i = 1, 2, \cdots$$

A chance of getting a large value with a Geometric random variable is very small (See figure 2).

Now if we go back to the bit torrent servers example. A video is broken down into m chunks. Each server has a random chunk, i.e. one out of m possible choices. We are interested in the first time in which we have the whole movie (meaning all the m chunks). Let X be the number of servers we need to query before we get all m chunks. How do we analyze this problem? How do we compute $\mathbb{E}[X]$? There is a natural way of breaking this random variable into easier steps. You are making progress whenever you get a new chunk.

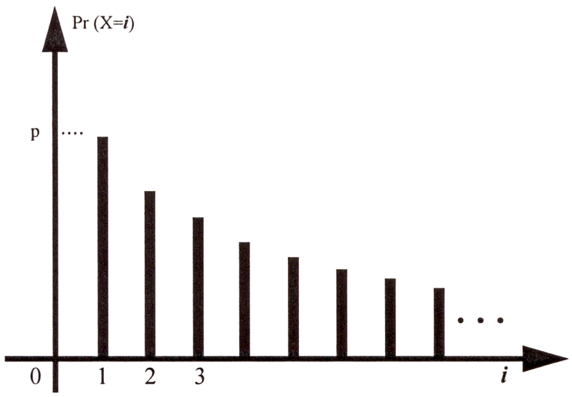

Figure 2: The Geometric distribution.

Let X_1 be the time to get a new chunk. Then $\mathbb{E}[X_1] = X_1 = 1$. Let X_2 be the time until we get the second new chunk after we already get one chunk. Can you proceed from here?