

# Math 525: Lecture 6

January 23, 2018

## 1 Moments

**Definition 1.1.** Let  $X$  be a discrete random variable and  $k$  be a positive integer. Suppose  $X^k$  is integrable. Then we call  $\mathbb{E}[|X|^k]$  the  $k$ -th *absolute moment* of  $X$ ,  $\mathbb{E}[X^k]$  the  $k$ -th *raw moment* of  $X$ , and  $\mathbb{E}[(X - \mathbb{E}[X])^k]$  the  $k$ -th *central moment* of  $X$ .

Note that the first moment is the expectation and the second central moment is the variance. The  $k$ -th raw moment is also sometimes simply called the  $k$ -th moment.

**Example 1.2.** Let  $X$  be a positive integer-valued random variable satisfying

$$\mathbb{P}(\{X = n\}) = c \frac{1}{n^3}$$

where  $c$  is a “normalizing constant” chosen such that

$$\sum_{n \geq 1} \mathbb{P}(\{X = n\}) = c \sum_{n \geq 1} \frac{1}{n^2} = 1.$$

This random variable has a finite expectation:

$$\mathbb{E}[X] = \sum_{n \geq 1} n \left( c \frac{1}{n^3} \right) = c \sum_{n \geq 1} \frac{1}{n^2} < \infty.$$

However, its variance is infinite:

$$\mathbb{E}[X^2] = \sum_{n \geq 1} n^2 \left( c \frac{1}{n^3} \right) = c \sum_{n \geq 1} \frac{1}{n} = \infty.$$

The same technique can be used to make a random variable whose first  $k$  moments are finite but all of its subsequent moments are infinite.

**Proposition 1.3.** Let  $X$  and  $Y$  be discrete random variables and  $k$  be a positive integer. If  $X^k$  and  $Y^k$  are integrable, so too is  $(X + Y)^k$ .

*Proof.* For any real numbers  $x$  and  $y$ ,

$$|x + y|^k \leq (2 \max\{|x|, |y|\})^k = 2^k \max\{|x|^k, |y|^k\} \leq 2^k |x|^k + 2^k |y|^k.$$

Therefore,

$$|X + Y|^k \leq 2^k |X|^k + 2^k |Y|^k,$$

from which the desired result follows by taking expectations of both sides.  $\square$

**Proposition 1.4.** *Let  $X$  be a discrete random variable and  $k$  be a positive integer. If  $X^k$  is integrable, so too is  $X^j$  for each  $0 \leq j \leq k$ .*

*Proof.* For any real number  $x \geq 0$ ,

$$x^j \leq \max\{x^k, 1\} \leq x^k + 1.$$

Therefore,

$$|X|^j \leq |X|^k + 1,$$

from which the desired result follows by taking expectations of both sides.  $\square$

**Corollary 1.5.** *Let  $X$  be a discrete random variable and  $k$  be a positive integer. If  $X^k$  is integrable, so too is  $(X - \mathbb{E}X)^k$  (and vice versa).*

*Proof.* Suppose  $X^k$  is integrable. Let  $Y = -\mathbb{E}X$  and apply Proposition 1.3 to see that  $(X - \mathbb{E}X)^k$  is integrable.

Suppose  $(X - \mathbb{E}X)^k$  is integrable. Then,

$$|X|^k = |(X - \mathbb{E}X) + \mathbb{E}X|^k \leq (|X - \mathbb{E}X| + |\mathbb{E}X|)^k \leq \sum_{j=0}^k \binom{k}{j} |X - \mathbb{E}X|^j |\mathbb{E}X|^{k-j}.$$

Now take expectations of both sides and apply Proposition 1.4.  $\square$

## 2 Moment generating functions

Last lecture, we looked at the probability generating function  $G$  of a discrete **nonnegative integer-valued** random variable  $X$ ,

$$G(t) = \mathbb{E}[t^X].$$

In this lecture, we start by letting  $X$  be **any** discrete random variable and examining the moment generating function  $M$  of  $X$ ,

$$M(\theta) = \mathbb{E}[e^{\theta X}].$$

As usual, we have been a bit cavalier in defining  $M$ , which is only well-defined at values of  $\theta \in \mathbb{R}$  for which the random variable  $e^{\theta X}$  is integrable. Remember the Taylor series for  $e^x$  is

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \cdots = \sum_{n \geq 0} \frac{1}{n!} x^n.$$

If we substitute this into  $M(\theta)$ , we obtain

$$M(\theta) = \mathbb{E} \left[ \sum_{n \geq 0} \frac{\theta^n}{n!} X^n \right].$$

Now, we would like to distribute the expectation over the sum to conclude

$$M(\theta) = \sum_{n \geq 0} \frac{\theta^n}{n!} \mathbb{E}[X^n]. \quad (1)$$

However, while we know from last lecture that we can distribute the expectation over a **finite** sum (from the property  $\mathbb{E}[aX + bY] = a\mathbb{E}X + b\mathbb{E}Y$ ), we cannot argue about infinite sums yet, so the conclusion (1) is just heuristic! We will defer a rigorous proof of this claim to a future lecture. For the time being, let's proceed assuming (1) is true. If we take derivatives with respect to  $\theta$ ,

$$\begin{aligned} M'(\theta) &= \sum_{n \geq 1} \frac{\theta^{n-1}}{(n-1)!} \mathbb{E}[X^n] \\ M''(\theta) &= \sum_{n \geq 2} \frac{\theta^{n-2}}{(n-2)!} \mathbb{E}[X^n] \\ &\vdots \\ M^{(k)}(\theta) &= \sum_{n \geq k} \frac{\theta^{n-k}}{(n-k)!} \mathbb{E}[X^n] \end{aligned}$$

and we can conclude

$$M^{(k)}(0) = \mathbb{E}[X^k], \quad k = 1, 2, \dots \quad (2)$$

Note that we have also ignored the fact that to evaluate the  $k$ -th derivative at  $\theta_0$ , we require  $M$  to be defined in a neighborhood of  $\theta_0$ . Regardless, if we proceed ignoring this issue, we deduce from (2) that the moment generating function generates the moments (perhaps unsurprisingly, given its name).

*Remark 2.1.* Note that  $M(0) = 1$  since  $M(0) = \mathbb{E}[X^0] = \mathbb{E}[1]$ . This is true for any random variable, since 1 is integrable.

## 3 Special discrete distributions

There are a handful of discrete distributions which come up frequently in applications. Our last topic today is to study some of these special distributions and compute their moments.

### 3.1 Bernoulli

A random variable  $X$  has a *Bernoulli distribution* if

$$\mathbb{P}(\{X = 1\}) = p \quad \text{and} \quad \mathbb{P}(\{X = 0\}) = 1 - p$$

for some  $0 \leq p \leq 1$ . We will often simply write  $X \sim \text{Bernoulli}(p)$  to indicate such a random variable.

**Example 3.1.** Toss a coin once, corresponding to the sample space  $\Omega = \{H, T\}$ . Define  $X$  by  $X(H) = 1$  and  $X(T) = 0$ . Then,  $X$  has a Bernoulli distribution.

The moment generating function of  $X \sim \text{Bernoulli}(p)$  is

$$M(\theta) = \mathbb{E}[e^{\theta X}] = e^{\theta \cdot 0} \mathbb{P}(\{X = 0\}) + e^{\theta \cdot 1} \mathbb{P}(\{X = 1\}) = (1 - p) + e^{\theta} p.$$

Note that  $M^{(k)}(\theta) = e^{\theta} p$ . Therefore,  $\mathbb{E}X^k = M^{(k)}(0) = p$  for all  $k = 1, 2, \dots$ . From this, it follows that

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

### 3.2 Binomial

A random variable  $X$  has a *binomial distribution* with parameters  $n \in \{1, 2, \dots\}$  and  $0 \leq p \leq 1$  if

$$\mathbb{P}(\{X = k\}) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

We will often simply write  $X \sim B(n, p)$  to indicate such a random variable. Note that the above implies that  $X$  only takes values in  $\{0, 1, \dots, n\}$  with positive probability:

**Proposition 3.2.** Let  $X \sim B(n, p)$ . Then,

$$\sum_{k=0}^n \mathbb{P}(\{X = k\}) = 1.$$

*Proof.* By the binomial theorem,

$$\sum_{k=0}^n \mathbb{P}(\{X = k\}) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} = (p + 1 - p)^n = 1^n = 1. \quad \square$$

**Example 3.3.** Toss the same coin  $n$  times. Let  $X$  be the number of heads witnessed in all  $n$  coin tosses. Assume that the probability of getting heads on each toss is  $0 \leq p \leq 1$ . Then,  $X \sim B(n, p)$ .

To see this, consider the case in which the first  $k$  tosses result in heads ( $H$ ) and the remainder result in tails ( $T$ ). This is captured by the sample

$$\underbrace{HH \cdots H}_{k \text{ times}} \underbrace{TT \cdots T}_{n-k \text{ times}}.$$

This sample occurs with probability  $p^k (1 - p)^{n-k}$ . However, there are  $\binom{n}{k}$  permutations of the letters above, from which we obtain the expression

$$\mathbb{P}(\{X = k\}) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

The moment generating function of  $X \sim B(n, p)$  is

$$M(\theta) = \mathbb{E} [e^{\theta X}] = \sum_{k=0}^n e^{\theta k} \mathbb{P}(\{X = k\}) = \sum_{k=0}^n \binom{n}{k} (e^{\theta} p)^k (1-p)^{n-k} = ((e^{\theta} - 1)p + 1)^n.$$

Taking derivatives,

$$\begin{aligned} M'(\theta) &= e^{\theta} np M(\theta)^{(n-1)/n} \\ M''(\theta) &= M'(\theta) + e^{2\theta} (n-1) np^2 M(\theta)^{(n-2)/n}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}X &= M'(0) = M(0)^{(n-1)/n} np = np \\ \mathbb{E}[X^2] &= M''(0) = M'(0) + M(0)^{(n-2)/n} (n-1) np^2 = np(1 + (n-1)p) \end{aligned}$$

and hence

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = np(1 + (n-1)p) - (np)^2 = np(1-p).$$

### 3.3 Poisson

A random variable  $X$  has a *Poisson distribution* with parameter  $\lambda > 0$  if

$$\mathbb{P}(\{X = k\}) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

We will often simply write  $X \sim \text{Poisson}(\lambda)$  to indicate such a random variable. Note that the above implies that  $X$  only takes values in  $\{0, 1, 2, \dots\}$  with positive probability:

**Proposition 3.4.** *Let  $X \sim \text{Poisson}(\lambda)$ . Then,*

$$\sum_{k \geq 0} \mathbb{P}(\{X = k\}) = 1.$$

*Proof.* By the Taylor expansion of  $e^x$ ,

$$\sum_{k \geq 0} \mathbb{P}(\{X = k\}) = \sum_{k \geq 0} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k \geq 0} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1. \quad \square$$

Before we motivate the Poisson distribution, let's blindly compute its moment generating function:

$$M(\theta) = \mathbb{E} [e^{\theta X}] = \sum_{k \geq 0} e^{\theta k} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k \geq 0} \frac{(\lambda e^{\theta})^k}{k!} = e^{-\lambda} e^{\lambda e^{\theta}} = e^{\lambda(e^{\theta} - 1)}.$$

Taking derivatives,

$$\begin{aligned} M'(\theta) &= \lambda e^{\theta} M(\theta) \\ M''(\theta) &= M'(\theta) (\lambda e^{\theta} + 1) \end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}X &= M'(0) = \lambda e^0 M(0) = \lambda \\ \mathbb{E}[X^2] &= M''(0) = M'(0) (\lambda e^0 + 1) = \lambda(\lambda + 1)\end{aligned}$$

and hence

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda.$$

One way to motivate the Poisson distribution is through the following observation:

**Proposition 3.5.** *Let  $\lambda > 0$  and suppose that  $np \rightarrow \lambda$  as  $n \rightarrow \infty$ . Then,*

$$\lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

We recognize the left hand side in the above from  $B(n, p)$ . The above suggests that  $\text{Poisson}(np)$  captures the number of successes in  $n$  trials, each having probability  $p$ , as the number of trials becomes large.

**Example 3.6.** The number of market crashes per annum could be modelled as a  $\text{Poisson}(\lambda)$  random variable with, for example,  $\lambda = 0.1$  (one crash every ten years).