LECTURE 6
VARIANCE, MOMENTS, AND MOMENT GENERATING FUNCTION (MGF) OF DISCRETE RANDOM VARIABLES

1 Variance

Consider the following two distributions:

$$S_X = \{-1, 1\}, \quad p_X(-1) = p_X(1) = 1/2, \tag{1}$$
$$S_Y = \{-2, 2\}, \quad p_Y(-2) = p_Y(2) = 1/2. \tag{2}$$

Recall from Lecture 5 that the expected value of a random variable captures the center (location) of its distribution. If X is distributed according to the PMF p_X described above, and Y is distributed according to the PMF p_Y, then

$$EX = EY = 0.$$

Thus, the two distributions have the same center, but they are obviously different: the distribution of Y is more spread out. One of the measures of spread of a distribution is *variance*.

Definition 1. (Variance and standard deviation) (a) Let X be a random variable with the support $S_X = \{x_1, x_2, \ldots\}$ and PMF p_X. The variance of X is defined as

$$\begin{aligned} Var(X) &= E(X - EX)^2 \\ &= \sum_{x \in S_X} (x - EX)^2 p_X(x) \\ &= (x_1 - EX)^2 p_X(x_1) + (x_2 - EX)^2 p_X(x_2) + \ldots. \end{aligned}$$

(b) The *standard deviation* of X is defined as

$$\sigma_X = \sqrt{Var(X)}. \tag{3}$$

Remark.

1. Similarly to the expectation, the variance is a number capturing one of the properties of the distribution of a random variable.

2. From its definition, one can see that the variance is a weighted sum of squared distances from the center of a distribution. Since realizations x_i's are on the right and on the left of the expectation EX, some of the differences $(x_i - EX)$ are positive and some are

negative. Squaring makes all deviations positive. It does not matter for the variance if x_i is on the right or left of the mean, only squared distances from the mean matter.

3. Because it is defined in terms of squared distances, the variance is not measured in the same units as X. For example, if X is measured in dollars, then the variance is measured in squared dollars. To obtain a measure of spread of the distribution of X that is measured in the same units as X, we compute the standard deviation of X (the square-root of the variance).

4. From equation (3), $\sigma_X^2 = Var(X)$. We, therefore, often denote the variance of X as σ_X^2.

5. A unit-free measure of the spread of a distribution is given by the *coefficient of variation.* Let $\mu_X = EX$. When $\mu_X \neq 0$, the coefficient of variation is defined as

$$\sigma_X/\mu_X.$$

Example. Suppose that X and Y are distributed as in (1) and (2).

$$\begin{aligned} Var(X) &= (-1-0)^2 p_X(-1) + (1-0)^2 p_X(1) \\ &= 1 \times \frac{1}{2} + 1 \times \frac{1}{2} \\ &= 1. \end{aligned}$$

Thus, $\sigma_X = 1$.

$$\begin{aligned} Var(Y) &= (-2-0)^2 \times p_Y(-2) + (2-0)^2 p_Y(2) \\ &= 4 \times \frac{1}{2} + 4 \times \frac{1}{2} \\ &= 4, \end{aligned}$$

and $\sigma_Y = 2$. Here, $Var(X) < Var(Y)$ is a simple expression of the fact that the distribution of Y is more disperse than that of X.

In finance, the variance (or the standard deviation) is often used as a measure of *risk*, and investment strategies are evaluated in terms of their variances and expected returns. A higher expected return makes the investment strategy more attractive. Since individuals are typically risk-averse, a higher variance makes the strategy less attractive.

Some basic properties of the variance are given below.

Theorem 2. *Let c be a constant.*
(a) $Var(c) = 0$.
(b) $Var(c + X) = Var(X)$.

(c) $Var(cX) = c^2 Var(X)$.

(d) $Var(X) = EX^2 - (EX)^2$.

Proof. (a) By Definition (1),

$$\begin{aligned} Var(c) &= E(c - Ec)^2 \\ &= E(c - c)^2 \\ &= E0 \\ &= 0, \end{aligned}$$

where the second and last equalities hold because the expectation of a constant is the constant itself.

(b) Again, we start from the definition of the variance:

$$\begin{aligned} Var(c + X) &= E((c + X) - E(c + X))^2 \\ &= E(c + X - (c + EX))^2 \\ &= E(X - EX)^2 \\ &= Var(X), \end{aligned}$$

where the equality in the second line holds by the linearity of expectation (Theorem 3, Lecture 5).

(c)

$$\begin{aligned} Var(cX) &= E(cX - E(cX))^2 \\ &= E(cX - cEX)^2 \\ &= E(c(X - EX))^2 \\ &= E(c^2(X - EX)^2) \\ &= c^2 E(X - EX)^2 \\ &= c^2 Var(X). \end{aligned}$$

(d) First, write

$$E(X - EX)^2 = E(X - EX)(X - EX),$$

and

$$\begin{aligned} (X - EX)(X - EX) &= X(X - EX) - (EX)(X - EX) \\ &= X^2 - (EX)X - (EX)(X - EX). \end{aligned}$$

Next,

$$\begin{aligned} E(X^2 - (EX)X) &= EX^2 - E\left((EX)X\right) \\ &= EX^2 - (EX)EX \\ &= EX^2 - (EX)^2, \end{aligned}$$

where the equality in the second line holds by the linearity of expectation since (EX) is a constant. Lastly,

$$E\left((EX)(X - EX)\right) = (EX)E(X - EX),$$

however,

$$\begin{aligned} E(X - EX) &= EX - E(EX) \\ &= EX - EX \\ &= 0. \end{aligned} \tag{4}$$

□

Remark.

1. All the proofs follow directly from the definition of the variance and linearity of expectation.

2. In the proof of part (d), we showed that $X - EX$ is a random variable that has mean zero (equation (4)).

3. According to part (a), a constant has zero variance. As a matter of fact, the result holds if and only if: when the variance of a random variable is zero, we can conclude that the random variable is a constant. To show that, suppose that $Var(X) = 0$, and assume that X is discrete. Let's try to find the support of X.

$$\begin{aligned} 0 &= E(X - EX)^2 \\ &= (x_1 - EX)^2 p_X(x_1) + (x_2 - EX)^2 p_X(x_2) + \dots . \end{aligned} \tag{5}$$

   Since

$$\begin{aligned} (x_i - EX)^2 &\geq 0, \\ p_X(x_i) &> 0 \end{aligned}$$

   for any x_i, the sum in (5) can be equal zero if and only if each individual element is zero.

We conclude that for all x_i's,

$$(x_i - EX)^2 = 0$$

or

$$x_i = EX$$

for all x_i's. Thus, the support of X can include only one point equal to the mean of X:

$$S_X = \{EX\}.$$

We conclude that X is a constant.

4. According to part (b), shifting a distribution does not change its spread (dispersion).

5. Parts (b) and (c) together imply

$$Var(a + bX) = b^2 Var(X),$$

where a and b are constants.

6. Part (c) implies that the standard deviation of cX is equal to

$$|c|\sigma_X,$$

where σ_X is the standard deviation of X.

7. Part (d) implies that when $EX = 0$, the variance of X is given by

$$Var(X) = EX^2.$$

8. For distributions with infinite support, it is possible to have a finite mean ($-\infty < EX < \infty$) but infinite variance ($Var(X) = \infty$).

2 Moments and moment generating function (MGF)

While the mean and variance are important characteristics of a distribution, they capture only its certain properties. Since the mean and variance are determined by EX and EX^2, generalizing the approach we can try to capture the additional characteristics of a distribution by considering other expressions of the form EX^r.

Definition 3. (Moments) (a) The r-th moment of X is defined as EX^r.

(b) The r-th *central* moment of X is defined as $E(X - EX)^r$.

Note that the first moment is just the mean of the random variable, and the second central moment is the variance. Also, the first central moment is zero.

It turns out that the information about all (existing) moments can be captured by a certain function defined below.

Definition 4. (Moment Generating Function (MGF)) Let X be a random variable with the support $S_X = \{x_1, x_2, \ldots\}$ and PMF p_X. Its MGF is defined as

$$\begin{aligned} M_X(t) &= Ee^{tX} \\ &= \sum_{x \in S_X} e^{tx} p_X(x) \\ &= e^{tx_1} p_X(x_1) + e^{tx_2} p_X(x_2) + \ldots, \end{aligned}$$

where $t \in \mathbb{R}$.

Example. If $S_X = \{-1, 1\}$ and $p_X(-1) = p_X(1) = 1/2$,

$$M_X(t) = (e^{-t} + e^{t})/2.$$

If $S_Y = \{-2, 2\}$ and $p_Y(-2) = p_Y(2) = 1/2$,

$$M_Y(t) = (e^{-2t} + e^{2t})/2.$$

Remark. When the support of a distribution is infinite, Ee^{tX} can be infinite for large values of t. In this case, we would restrict the range of t to sufficiently small values where the MGF is finite, when this is possible.

The relationship between moments and the MGF is established in the following result.

Theorem 5. *The r-th moment of X (when exists) is equal to the r-th derivative of the MGF of X evaluated at zero:*

$$\left. \frac{d^r M_X(t)}{dt^r} \right|_{t=0} = EX^r,$$

where $r = 1, 2, \ldots$.

Proof. We consider here the case of a discrete X with the support and PMF $S_X = \{x_1, x_2, \ldots\}$

and $p_X(x)$ respectively. Consider the case of $r = 1$.

$$\begin{aligned}
\frac{dM_X(t)}{dt} &= \frac{d}{dt}\left(\sum_{x \in S_X} e^{tx} p_X(x)\right) \\
&= \frac{d}{dt}\left(e^{tx_1} p_X(x_1) + e^{tx_2} p_X(x_2) + \ldots\right) \\
&= \frac{de^{tx_1}}{dt} p_X(x_1) + \frac{de^{tx_2}}{dt} p_X(x_2) + \ldots \\
&= x_1 e^{tx_1} p_X(x_1) + x_2 e^{tx_2} p_X(x_2) + \ldots \\
&= \sum_{x \in S_X} x e^{tx} p_X(x) \\
&= E\left(X e^{tX}\right).
\end{aligned}$$

Next, substituting $t = 0$, we obtain:

$$E\left(X e^{0 \times X}\right) = EX.$$

Hence,

$$\frac{dM_X(0)}{dt} = EX.$$

For $r = 2$,

$$\begin{aligned}
\frac{d^2 M_X(t)}{dt^2} &= \frac{d}{dt}\left(\frac{dM_X(t)}{dt}\right) \\
&= \frac{d}{dt}\left(\sum_{x \in S_X} x e^{tx} p_X(x)\right) \\
&= \sum_{x \in S_X} x \left(\frac{de^{tx}}{dt}\right) p_X(x) \\
&= \sum_{x \in S_X} x \left(x e^{tx}\right) p_X(x) \\
&= \sum_{x \in S_X} x^2 e^{tx} p_X(x) \\
&= E\left(X^2 e^{tX}\right).
\end{aligned}$$

Again, substituting $t = 0$, we obtain:

$$\begin{aligned}
\frac{d^2 M_X(0)}{dt^2} &= E\left(X^2 e^{0 \times X}\right) \\
&= EX^2.
\end{aligned}$$

More generally,

$$\frac{d^r M_X(t)}{dt^r} = E\left(X^r e^{tX}\right),$$

from which the result follows. □

Example. Suppose that $S_X = \{-1, 1\}$ and $p_X(-1) = p_X(1) = 1/2$, and therefore

$$M_X(t) = (e^{-t} + e^t)/2.$$

We have:

$$\begin{aligned} \frac{dM_X(t)}{dt} &= \frac{d}{dt}\left(e^{-t} + e^t\right)/2 \\ &= (-e^{-t} + e^t)/2, \end{aligned}$$

and

$$\begin{aligned} EX &= \frac{dM_X(0)}{dt} \\ &= (-e^0 + e^0)/2 \\ &= (-1 + 1)/2 \\ &= 0. \end{aligned}$$

Next,

$$\begin{aligned} \frac{d^2 M_X(t)}{dt^2} &= \frac{d}{dt}(-e^{-t} + e^t)/2 \\ &= (e^{-t} + e^t)/2. \end{aligned}$$

Hence,

$$\begin{aligned} EX^2 &= \frac{d^2 M_X(0)}{dt^2} \\ &= (e^0 + e^0)/2 \\ &= 1. \end{aligned}$$

More generally,

$$\frac{d^r M_X(t)}{dt^r} = \frac{1}{2}\left((-1)^r e^{-t} + e^t\right),$$

and

$$\begin{aligned} EX^r &= \frac{d^r M_X(0)}{dt^r} \\ &= \frac{1}{2}((-1)^r + 1). \end{aligned}$$

Thus, in this example, $EX^r = 0$ if r is odd, and $EX^r = 1$ if r is even.

Consider now the example with $S_Y = \{-2, 2\}$ and $p_Y(-2) = p_Y(2) = 1/2$, and

$$M_Y(t) = (e^{-2t} + e^{2t})/2.$$

In this case,

$$\begin{aligned} \frac{dM_Y(t)}{dt} &= \frac{1}{2}\left(-2e^{-2t} + 2e^{2t}\right), \\ \frac{d^2M_Y(t)}{dt^2} &= \frac{1}{2}\left(2^2e^{-2t} + 2^2e^{2t}\right), \\ &\ldots \\ \frac{d^rM_Y(t)}{dt^r} &= \frac{1}{2}\left((-1)^r 2^r e^{-2t} + 2^r e^{2t}\right). \end{aligned}$$

It follows

$$EY^r = \frac{1}{2}\left((-1)^r 2^r + 2^r\right),$$

or

$$\begin{aligned} EY &= 0, \\ EY^2 &= \frac{1}{2}\left(2^2 + 2^2\right) = 4, \\ EY^3 &= 0, \\ EY^4 &= \frac{1}{2}\left(2^4 + 2^4\right) = 16, \\ &\ldots \end{aligned}$$

Note that all the odd moments of this distribution are exactly zero. This is explained by the fact that the distribution is *symmetric* around zero: the probability of drawing a positive value is equal to the probability of drawing a negative value of the same magnitude. Those values cancel out when computing EX^r with an odd r.

MGFs are not only useful for computing moments. As a matter of fact, the MGF provides a complete description of a distribution. In other words each distribution has a unique MGF. Thus, instead of working the PMF, we can work with the MGF without losing any information about the distribution.

Theorem 6. *Let X and Y be two random variables, and suppose that the MGFs of X and Y are equal: for all $t \in \mathbb{R}$ where the MGFs are finite,*

$$M_X(t) = M_Y(t).$$

Then X and Y have the same distribution. In particular, their PMFs satisfy $p_X(u) = p_Y(u)$ for all u's.

We are not going to provide a formal proof for the theorem, and will only illustrate why it must be true. Let $\{u_1, u_2, \ldots, u_n\}$ be the support points. Write the MGFs as

$$\begin{aligned} M_X(t) &= e^{tu_1} p_X(u_1) + e^{tu_2} p_X(u_2) + \ldots + e^{tu_n} p_X(u_n), \\ M_Y(t) &= e^{tu_1} p_Y(u_1) + e^{tu_2} p_Y(u_2) + \ldots + e^{tu_n} p_Y(u_n). \end{aligned}$$

Since $M_X(t) = M_Y(t)$ for any t, subtracting the second equation from the first, we obtain:

$$0 = e^{tu_1}(p_X(u_1) - p_Y(u_1)) + e^{tu_2}(p_X(u_2) - p_Y(u_2)) + \ldots + e^{tu_n}(p_X(u_n) - p_Y(u_n). \qquad (6)$$

The equality must hold for any value t, i.e. it holds for a continuum of values of t. This can be true if and only if

$$p_X(u) - p_Y(u) = 0$$

for all u. For example, suppose that

$$\begin{aligned} u_1 &= 1, \\ u_2 &= 2, \\ &\ldots \\ u_n &= n. \end{aligned}$$

Lets re-define

$$e^t = s.$$

Then,

$$\begin{aligned} e^{tu_1} &= e^t = s, \\ e^{tu_2} &= e^{2t} = (e^t)^2 = s^2, \\ &\ldots \\ e^{tu_n} &= e^{nt} = (e^t)^n = s^n. \end{aligned}$$

Lastly, let's define

$$\begin{aligned} c_1 &= p_X(u_1) - p_Y(u_1), \\ &\dots \\ c_n &= p_X(u_n) - p_Y(u_n). \end{aligned}$$

In this case, equation 6 becomes

$$0 = sc_1 + s^2c_2 + \ldots + s^nc_n.$$

This is a polynomial equation in s with coefficients $c_1, \ldots, c_n$. However, the equation must hold for *any* value of s. This can be true if and only if

$$c_1 = c_2 = \ldots = c_n = 0,$$

which implies that

$$p_X(u) = p_Y(u)$$

for all u.