ECON 340
Economic Research Methods

Div Bhagia

Lecture 9: Distribution, Expectation, and Variance

Taking stock

We have learned how to describe variables in the data using:

- Empirical distribution
- Mean and median
- Variance and standard deviation
- Correlation and covariance

Random Sampling

- Often, data is available only for a sample of the population
- Ideally, we want a sample *representative* of the population we are interested in and not a *biased* sample
- We can achieve this by taking a *random sample* from the population
- Random sample: each unit from the population has an equal probability of being chosen

Simple Example

Say, we want to take a random sample of 2 from a population of 5.

Population:

$X_1 = \$60,000$
$X_2 = \$40,000$
$X_3 = \$40,000$
$X_4 = \$50,000$
$X_5 = \$60,000$

Population mean:

$\mu = \$50,000$

Pick a sample randomly: Spin a Wheel

Attempt 1

$\bar{X}_1 =$

Attempt 2

$\bar{X}_2 =$

Attempt 3

$\bar{X}_3 =$

Random Sampling

- We can get different values of the sample mean depending on the sample we pick
- Sample mean is a *random variable*!
- Then what can we infer about the population mean from the sample mean?
- But before that, what is a random variable?

Random Variables

- *Random variable* is a numerical summary of a random outcome.
- Examples: outcome from a coin toss or a die roll, or number of times your wireless network fails before a deadline, etc.
- Random variables can be *discrete* or *continuous*
- Discrete random variables take a discrete set of values, like $0, 1, 2, \ldots$
- Continuous random variables take on a continuum of possible values

Discrete Random Variables

- *Probability distribution* of a discrete random variable: all possible values of the variable and their probabilities.

  $$f(x) = Pr(X = x)$$

  where $0 \leqslant f(x) \leqslant 1$ for all x and $\sum_x f(x) = 1$.

- *Cumulative probability distribution* gives the probability that the random variable is less than or equal to a particular value.

  $$F(x) = Pr(X \leqslant x) = \sum_{x' \leqslant x} f(x')$$

Discrete Random Variable: Example

X: outcome from rolling a die

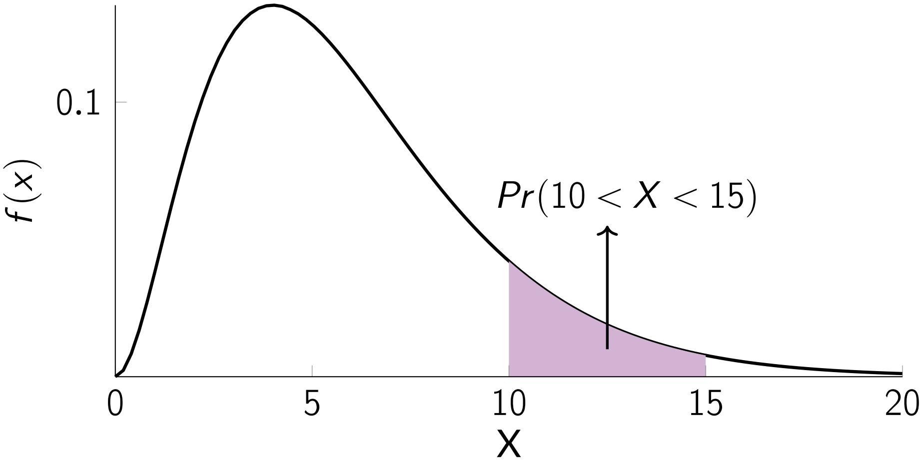
X	$f(X)$	$F(X)$
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	1

Also referred to as probability distribution function (PDF) and cumulative distribution function (CDF).

Continuous Random Variable

- For continuous random variables, due to a continuum of possible values, it is not feasible to list the probability of each possible value.
- So instead, the area under the *probability density function* $f(x)$ between any two points gives the probability that the random variable falls between those two points.
- *Cumulative probability distribution* for continuous RVs is defined as before $F(x) = Pr(X \leqslant x)$.

Probability Density Function



How to calculate the area under the curve?

- Area under the curve is calculated using an *integral* (just like a sum but for continuous variables)
- However, don't sweat, in most cases a statistical program or old-school tables (in the back of textbooks) can help us find these areas for commonly used distributions
- We will now define expectation and variance for the discrete case, but it is generalizable to continuous RVs

Expectation

- *Expectation*: average value of the random variable over many repeated trials or occurrences
- Computed as a weighted average of the possible outcomes, where the weights are the probabilities
- The expectation of X is also called the *expected value* or the *mean* and is denoted by μ_X or $E(X)$

$$\mu_X = E(X) = \sum_x f(x)x$$

Example

An outdoor market vendor sells handmade crafts.

Weather	**Probability**	**Sales ($)**
No rain	0.6	300
Light rain	0.3	150
Heavy rain	0.1	50

$$E(Sales) = (0.6 \times 300) + (0.3 \times 150) + (0.1 \times 50) = \$230$$

Example

An outdoor market vendor sells handmade crafts.

Weather	**Probability**	**Sales ($)**
No rain	0.6	300
Light rain	0.3	150
Heavy rain	0.1	50

$$E(Sales) = (0.6 \times 300) + (0.3 \times 150) + (0.1 \times 50) = \$230$$

Daily sales will fluctuate due to randomness of the weather. However, if this day were to repeat itself multiple times, the vendor would, on average, expect to achieve daily sales of $230.

Variance and Standard Deviation

The variance and standard deviation measure the dispersion or the "spread".

$$\sigma_X^2 = Var(X) = E[(X - \mu_X)^2] = \sum_x (x - \mu_X)^2 f(x)$$

Variance is the expected value of the squared deviation of X from its mean.

In our example, the variance will capture the potential variability or fluctuation in sales on any given day compared to the average (expected value).

Example

An outdoor market vendor sells handmade crafts.

Weather	Probability	Sales ($)
No rain	0.6	300
Light rain	0.3	150
Heavy rain	0.1	50

$$E(Sales) = 0.6 \cdot (300) + 0.3 \cdot (150) + 0.1 \cdot (50) = \$230$$

$$\begin{aligned} Var(Sales) &= 0.6 \cdot (300 - 230)^2 + 0.3 \cdot (150 - 230)^2 + 0.1 \cdot (50 - 230)^2 \\ &= 8100 \end{aligned}$$

Variance and Standard Deviation

Alternative formula for the variance:

$$Var(X) = E[(X - \mu_X)^2] = E(X^2) - \mu_X^2$$

Since variance is in units of the square of X, therefore we use *standard deviation* which is the square-root of variance

$$\sigma_X = \sqrt{\sigma_X^2}$$

For our example, standard deviation of sales is $\sqrt{8100} = \$90$.

Transformations of Random Variables

If you shift every outcome by some constant a,

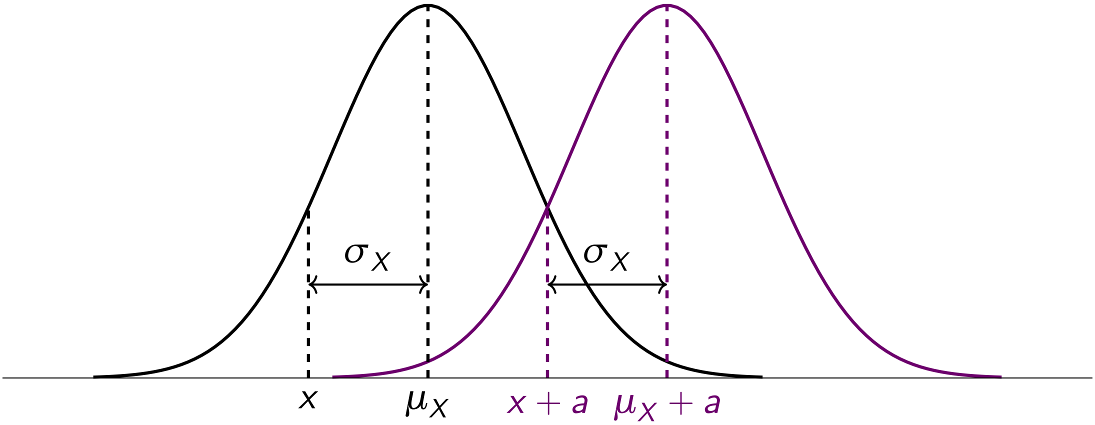
- Mean also shifts by a:

$$E(X + a) = E(X) + a$$

- Variance is unchanged:

$$Var(X + a) = Var(X)$$

Shifting the Distribution



Transformations of Random Variables

If you scale every outcome by some constant b,

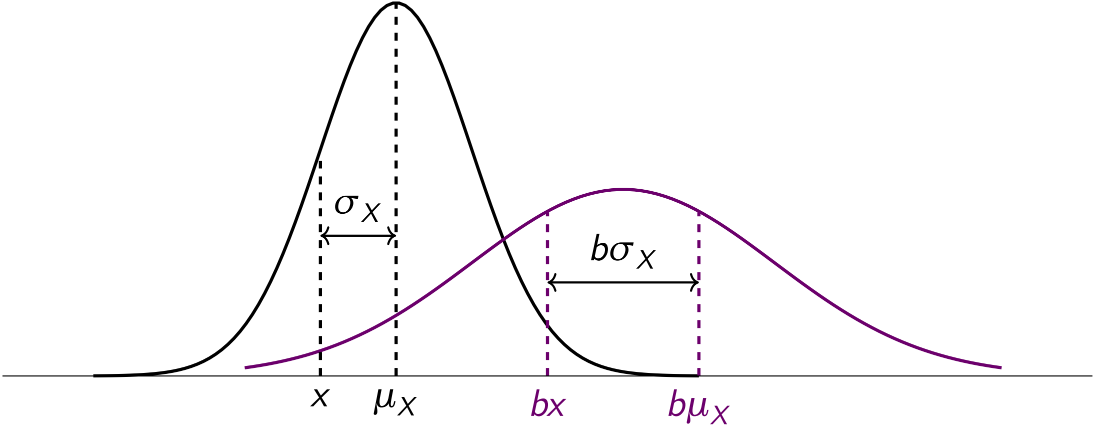
- Mean is also scaled by b:

$$E(bX) = bE(X)$$

- Variance is scaled by b^2:

$$Var(bX) = b^2 Var(X)$$

Scaling the Distribution



Transformations of Random Variables

More generally, if X is a random variable and

$$Y = a + bX$$

Then Y is also a random variable with

$$E(Y) = a + bE(X) \qquad Var(Y) = b^2 Var(X)$$

In addition, a linear transformation of a random variable does not change the shape of the distribution.

Example

The market vendor offered you a job, you get \$20 and 10% commission (Y) on sales (X) per day.

$$Y = 20 + 0.1X$$

Weather	Probability	X	Y
No rain	0.6	300	20 + 0.1(300) = 50
Light rain	0.3	150	20 + 0.1(150) = 35
Heavy rain	0.1	50	20 + 0.1(50) = 25

$$E(Y) = 20 + 0.1E(X) = 20 + 0.1(230) = 43$$

$$Var(Y) = 0.1^2 \cdot Var(X) = 0.01(8100) = 81$$

$$SD(Y) = \sqrt{81} = 9 \; (= 0.1SD(X))$$

Standardized Random Variables

A random variable can be transformed into a random variable with mean 0 and variance 1 by subtracting its mean and then dividing by its standard deviation, a process called standardization.

$$Z = \frac{X - \mu_X}{\sigma_X}$$

Here, $E(Z) = 0$ and $\sigma_Z = 1$.