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).

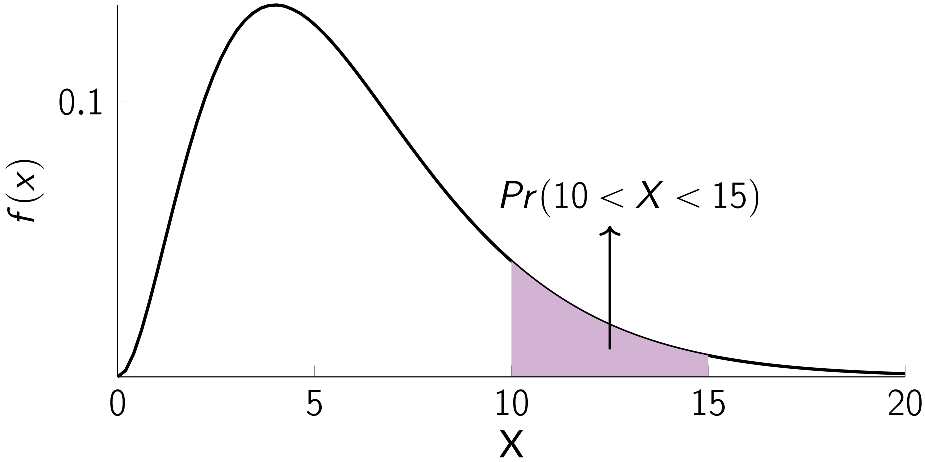
Continuous Random Variable

- 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 $E(X)$

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

Roll of a Die

The expected value from the roll of a die

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{21}{6} = 3.5$$

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 square of the deviation of X from its mean.

Roll of a Die

Expected value

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{21}{6} = 3.5$$

Variance

$$Var(X) = (-2.5)^2.\frac{1}{6} + (-1.5)^2.\frac{1}{6} + (-0.5)^2.\frac{1}{6} + (0.5)^2.\frac{1}{6}$$
$$+(1.5)^2.\frac{1}{6} + (2.5)^2.\frac{1}{6} = \frac{17.5}{6}$$

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

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

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:

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

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:

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

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

Flip a coin to gain \$10 or loose \$10.

x	$f(x)$	$xf(x)$	$(x - E(X))^2$	$f(x)(x - E(X))^2$
10	0.5	5	$(10)^2$	50
-10	0.5	-5	$(-10)^2$	50
		0		100

So we have that $E(X) = 0$ and $var(X) = 100$.

Now say, $Y = X - 5$ and $Z = 0.5X$. What is the expectation and variance of Y and Z?

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 $\sigma_Z = 1$.