ECON 3040 - Chapter 2: Probability Review

Ryan T. Godwin

University of Manitoba

2.1.1 - Randomness

- Randomness
- Sample space
- Outcome
- Event

The idea of randomness helps us put mathematical structure on things that are uncertain or unpredictable.

Randomness

The inability to predict an outcome.

- If we had enough information, would it be random?
- Example: rolling two dice.
- Realization of a random event - the randomness resolves and the outcome is revealed. e.g. before and after rolling dice.
- Something could be unknown, but still non-random.

Sample space

The set of all possibilities (all outcomes) that can occur as a result of the random process.

What is the sample space when rolling two dice?

Outcome

A single point, or possibility, in the sample space.

Only one outcome can occur.

Event

A collection of outcomes. An event is a subset of the sample space.

What are the outcomes that comprise rolling higher than a 10 with two dice?

2.1.2 Probability

Probability helps us to further put mathematical structure on things that are uncertain.

- Probability assigns numbers to random events
- Between 0 and 1 (or a percentage)
- "The probability of an event is the proportion of times it occurs in the long run"

What is the probability of rolling a 7, a 12, or higher than 10, with two dice?

2.2 Random Variables (RVs)

Translates random outcomes into numerical values. (Attaches numbers to outcomes).

- Why does a die roll have numerical meaning?
- RVs are human-made constructs
- Example: temperature in Celsius, Fahrenheit, Kelvin. Different numbers representing the same random event.

2.2 Random Variables (RVs)

Translates random outcomes into numerical values. (Attaches numbers to outcomes).

- Why does a die roll have numerical meaning?
- RVs are human-made constructs
- Example: temperature in Celsius, Fahrenheit, Kelvin. Different numbers representing the same random event.

2.2.1 Discrete and continuous random variables

2.2.2 Realization of a random variable

2.2.3 Key Points

Translates random outcomes into numerical values. (Attaches numbers to outcomes).

- A random variable can take on different values (or ranges of values), with different probabilities
- There are discrete and continuous random variables
- Continuous random variables can take on an infinite number of possible values, so we can only assign probabilities to ranges of values
- We can assign probabilities to all possible values for a discrete random variable
- The realization of a random variable is just a number, it used to be random, but now we've seen the outcome

2.3 Probability function

Beware: it has many names! (probability distribution, probability distribution function (PDF), probability mass function (PMF), probability density function, density function)

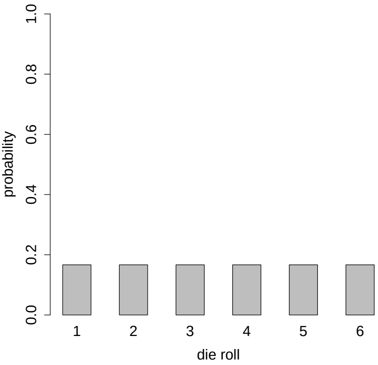
We all know what a function is, for example: $y = f(x) = 2 + 3x$. Put in an x, $f(x)$ it, get out a y. A probability function takes in an outcome, puts out a probability.

- Usually an equation
- Probability function: (i) lists all possible numerical values the RV can take; (ii) assigns a probability to each value.
- Prob. function contains all possible knowledge we can have about an RV

2.3.1 Example: die roll

$$\Pr(Y = y) = \frac{1}{6}; y = 1, \dots, 6$$

Probability function for die roll in a picture:



2.3.2 Example: a normal RV

$$f\left(y \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(y-\mu)^2}{2\sigma^2}$$

2.3.3 Probabilities of events

Probability function can be used to calculate the probability of outcomes or events occurring.

Example: Let Y be the result of a die roll. What is the probability of rolling higher than 3?

$$\Pr(Y > 3) = \Pr(Y = 4) + \Pr(Y = 5) + \Pr(Y = 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

2.4 Moments of a random variable

- "Moment" refers to a concept in physics
- 1st moment is the mean • 2nd (central) moment is the variance
- 3rd is skewness
- 4th is kurtosis
- Covariance and correlation are mixed moments

Moments summarize information about the RV. Moments are obtained from the ...

2.4.1 Mean (expected value)

- Value that is expected
- Average through repeated realizations of the RV
- Determined from the probability function (do some math to it)
- Mean is summarized info that is already contained in the prob. function
- Let Y be the RV. Then the mean of Y = expected value of Y = $\mu_Y = E[Y]$

If Y is discrete: **the mean is the weighted average of all possible outcomes, where the weights are the probabilities of each outcome.**

The equation for the mean of Y (Y is discrete):

$$\mathrm{E}[Y] = \sum_{i=1}^{K} p_i Y_i$$

- p_i probability of the i^{th} event
- Y_i is the value of the i^{th} outcome
- K is the total number of outcomes

Study this equation. It is a good way to understand the mean.

Exercise: calculate the mean die roll

Properties of the mean

- $\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y]$
- $\mathrm{E}[cY] = c\mathrm{E}[Y]$, where c is a constant
- $\mathrm{E}[c + Y] = c + \mathrm{E}[Y]$
- $\mathrm{E}[c] = c$

The mean when y is a *continuous* random variable:

$$\mathrm{E}[y] = \int y f(y) \mathrm{d}y$$

What the mean is not

The mean is different from the median and the mode, although all are measures of central tendency.

The mean is different from the sample mean or sample average. The mean comes from the probability function. The sample mean/average comes from a sample of data.

2.4.3 Variance

- Measure of the spread or dispersion of a RV
- Denoted by σ^2. The variance of y would be σ_y^2 and the variance of x would be σ_x^2
- Variance is the expected squared difference of a variable from its mean
- Equation:

$$\mathrm{Var}(Y) = \mathrm{E}\left[(Y - \mathrm{E}[Y])^2\right]$$

When Y is a discrete random variable, this equation becomes:

$$\mathrm{Var}(Y) = \sum_{i=1}^{K} p_i \times \left(Y_i - \mathrm{E}\left[Y_i\right]\right)^2$$

- For variance (the 2nd moment), we are taking the expectation of a squared term
- For skewness (the 3rd moment), we would take the expectation of a cubed term, etc.

Exercise: calculate the variance of a die roll

Properties of the variance

- $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \times \text{Cov}[X, Y]$
- $\text{Var}[cY] = c^2\text{Var}[Y]$, where c is a constant
- $\text{Var}[c + Y] = \text{Var}[Y]$
- $\text{Var}[c] = 0$

Exercise: I change the sides of the die to equal 2,4,6,8,10,12. What is the mean and variance of the die roll?

Exercise: What is the mean and variance of the sum of two dice?

2.4.5 Covariance

- Measures the relationship between two random variables
- Random variables Y and X have a joint probability function
- Joint prob. func.: (i) lists all possible combos of Y and X; (ii) assign a probability to each combination
- A useful summary of a joint probability function is the covariance
- The covariance between Y and X is the expected difference of Y from its mean, multiplied by the expected difference of X from its mean
- Covariance tells us something about how two variables are related, or how they move together
- Tells us about the direction and strength of the relationship between two variables

$$\mathrm{Cov}(Y, X) = \mathrm{E}\left[(Y - \mu_Y)(X - \mu_X)\right]$$

The covariance between Y and X is often denoted as σ_{YX}. Note the following properties of σ_{YX}:

- σ_{YX} is a measure of the *linear* relationship between Y and X. Non-linear relationships will be discussed later.
- $\sigma_{YX} = 0$ means that Y and X are linearly independent.
- If Y and X are independent (neither variable causes the other), then $\sigma_{YX} = 0$. The converse is not necessarily true (because of non-linear relationships).
- The $\mathrm{Cov}(Y, Y)$ is the $\mathrm{Var}(Y)$.
- A positive covariance means that the two variables tend to differ from their mean in the *same* direction.
- A negative covariance means that the two variables tend to differ from their mean in the *opposite* direction.

2.4.6 Correlation

- Correlation usually denoted by ρ
- Similar to covariance, but is easier to interpret

$$\rho_{YX} = \frac{\mathrm{Cov}(Y, X)}{\sqrt{\mathrm{Var}(Y)\,\mathrm{Var}(X)}} = \frac{\sigma_{YX}}{\sigma_Y \sigma_X}$$

The difficulty in interpreting the value of covariance is because $-\infty < \sigma_{YX} < \infty$. Correlation transforms covariance so that it is bound between -1 and 1. That is, $-1 \leq \rho_{YX} \leq 1$.

- $\rho_{YX} = 1$ means perfect positive linear association between Y and X.
- $\rho_{YX} = -1$ means perfect negative linear association between Y and X.
- $\rho_{YX} = 0$ means no linear association between Y and X (linear independence).

2.5 Some special probability functions

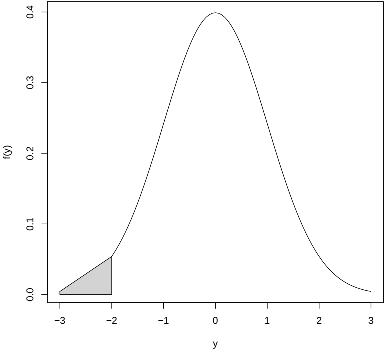
2.5.1 The normal distribution

- Common because of the "central limit theorem" (in a few slides)
- $f\left(y \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(y-\mu)^2}{2\sigma^2}$
- Mean of y is μ
- Variance of y is σ^2

2.5.2 The standard normal distribution

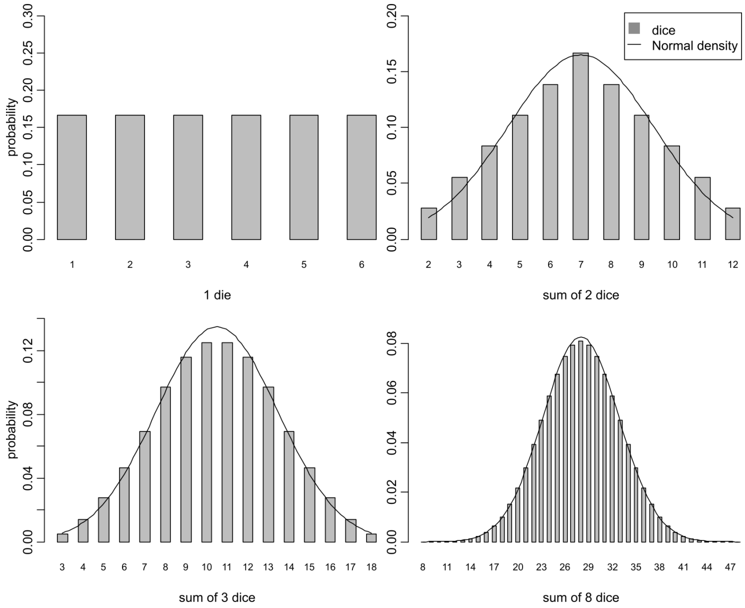
- Special case of a normal distribution, where $\mu = 0$ and $\sigma^2 = 1$
- The probability function for the Normal distribution becomes:
- $f(y) = \frac{1}{\sqrt{2\pi}} \exp \frac{-y^2}{2}$
- Any normal random variable can be "standardized"
- How to standardize?
- Standardizing has long been used in hypothesis testing (as we shall see)

Probability function for a standard Normal variable, with probability of of $y < -2$ in gray.



2.5.3 The central limit theorem

- There are hundreds of different probability functions
- Examples: Poisson, Binomial, Generalized Pareto, Nakagami, Uniform
- So why is the normal distribution so important? Why are so many RVs normal?
- Answer: CLT
- CLT (loosely speaking) – if we add up enough RVs, the resulting sum tends to be Normal


probability
0.00
0.05
0.10
0.15
0.20
0.25
0.30
1
2
3
4
5
6
1 die
0.00
0.05
0.10
0.15
0.20
dice
Normal density
2
3
4
5
6
7
8
9
10
11
12
sum of 2 dice
probability
0.00
0.04
0.08
0.12
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
sum of 3 dice
0.00
0.02
0.04
0.06
0.08
8
11
14
17
20
23
26
29
32
35
38
41
44
47
sum of 8 dice

2.5.4 The chi-square distribution

- Add to a normal RV – still normal
- Multiply a normal RV – still normal
- Square a normal RV – now it is chi-square distributed
- We will use the chi-square distribution for the F-test in a later chapter