EM384: Analytical Methods for Engineering Management

Lesson 28: Random Variables and Probability Distributions

6 April 2023

Table of contents

1. Lesson Objectives

2. Stochastic Models

3. Random Variables

4. Conclusion

Lesson Objectives

Lesson Objectives

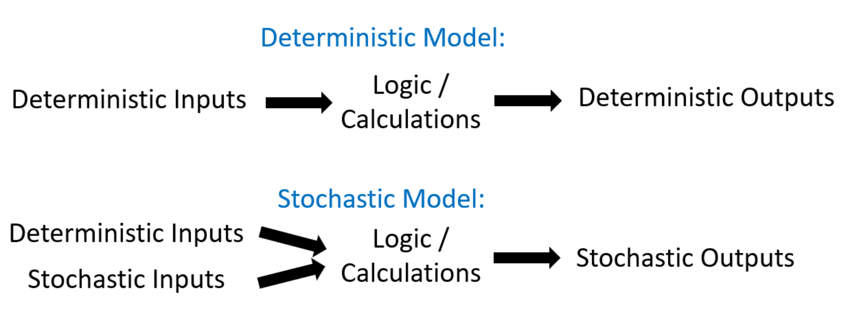
- Understand the concept of a stochastic model.
- Understand random variables and their role in simulation.
- Apply randomness to Excel models.
- Execute Excel commands RAND, RANDBETWEEN

Stochastic Models

Stochastic Models

The term stochastic refers to randomness in a model.

- A model that has *no randomness* in it is called a deterministic model (These are the models we have been doing in blocks 1-3).
- A model that has *randomness* built in is called a stochastic model. Stochastic models still have a structure, i.e. rules, logic and calculations that you put in place, but some variables are allowed to be random, within the confines of a probability distribution.



Examples

Examine whether the following model is deterministic or stochastic

- We take a brand new deck of cards and we draw the top card. Our model output is the face value of the card we drew.

Examples

Examine whether the following model is deterministic or stochastic

- We take a brand new deck of cards and we draw the top card. Our model output is the face value of the card we drew.
- Deterministic (as long as a new pack of cards is always in the same order)

Examples

Examine whether the following model is deterministic or stochastic

- We take a shuffled deck of cards and we draw the top card. Our model output is the face value of the card we drew.

Examples

Examine whether the following model is deterministic or stochastic

- We take a shuffled deck of cards and we draw the top card. Our model output is the face value of the card we drew.
- Stochastic, because there are different possible outputs.

Examples

Examine whether the following model is deterministic or stochastic

- We make a model that records the orders of customers at a burger joint and outputs the price each one pays.

Examples

Examine whether the following model is deterministic or stochastic

- We make a model that records the orders of customers at a burger joint and outputs the price each one pays.
- Deterministic, because we know the price each customer pays at the time they pay.

Examples

Examine whether the following model is deterministic or stochastic

- We make a model to predict the average price that customers will pay at a burger joint next week.

Examples

Examine whether the following model is deterministic or stochastic

- We make a model to predict the average price that customers will pay at a burger joint next week.
- Stochastic, if assume a probability distribution for the price each customer pays (perhaps based on previous data).

Examples

Examine whether the following model is deterministic or stochastic

- We make a model that calculates the net present value of a firm's holdings in 10 years, based on a forecast of revenue and expenses that we have in Excel.

Examine whether the following model is deterministic or stochastic

- We make a model that calculates the net present value of a firm's holdings in 10 years, based on a forecast of revenue and expenses that we have in Excel.
- Deterministic since all parameters for our model are in the Excel file (unless we have random parameters in the Excel file).

Examples

Examine whether the following model is deterministic or stochastic

- We make a model to minimize the cost of a spaghetti dinner for the Corps of cadets, while meeting nutritional requirements.

Examples

Examine whether the following model is deterministic or stochastic

- We make a model to minimize the cost of a spaghetti dinner for the Corps of cadets, while meeting nutritional requirements.

- Deterministic, as long as the linear program parameters don't change.

Examples

Examine whether the following model is deterministic or stochastic

- We make a model to simulate (1000 times) the balance of our Roth IRA in 30 years based on the annual return of the SP 500 from the past 30 years, and fixed contributions for the next 10 years.

Examples

Examine whether the following model is deterministic or stochastic

- We make a model to simulate (1000 times) the balance of our Roth IRA in 30 years based on the annual return of the SP 500 from the past 30 years, and fixed contributions for the next 10 years.
- Stochastic, as we have 30 different years of returns for the SP 500 and we have to create a probability distribution from which to sample to simulate the next 30 years).

Random Variables

Random Variables

A Random Variable is a variable whose possible values are numerical outcomes of an experiment or a random phenomenon. A random variable is generally represented by a capital letter, i.e. X, and the outcome of the random variable by a lower case letter, i.e. x.

- A random variable X is discrete if it can assume only a finite or countably infinite number of distinct values.
- A random variable X is continuous if it can assume an uncountably infinite number of distinct values.

Random variables are defined on a sample space, which is the set of all possible outcomes for that random variable.

- Discrete random variables have a sample space marked as a set, i.e. $S = \{0, 1, 2, ...\}$, or $S = \{1, 2, 3, 4, 5, 6\}$.
- Continuous random variables have a sample space that is an interval, i.e. $S = [0, 1]$, or $S = [0, \infty)$.

Random Variables

- We use a capital letter, like X, to denote a random variable
- The values of a random variable are denoted with a lowercase letter, in this case x
- For example, $P(X = x)$ which means *the probability that random variable X takes on the value x.*
- For example, $P(X \leq 20)$ which means *the probability that random variable X takes on a value less than or equal to* 20.
- For example, $P(10 \leq X \leq 20)$ which means *the probability that random variable X takes on a value between* 10 *and* 20, *included.*

Probability of an Event

An event refers to a set of one or more possible outcomes. The probability of an event is the extent to which an event is likely to occur.

If we know the theoretical distribution of possible outcomes, then we can find the theoretical probability of event A happening:

$$P(A) = \frac{\text{number of ways event A can happen}}{\text{total number of possible outcomes}}$$

If we use simulation to obtain n outcomes of the same experiment, we can find the empirical probability of event A happening:

$$P(A) = \frac{\text{number of times A happened}}{n}$$

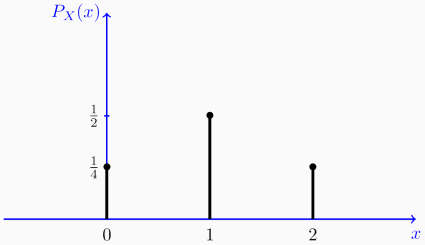
Random Variables

There are two types of random variables:

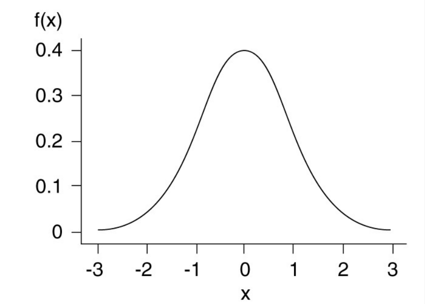
- Discrete random variables often take only integer values
  - Example: Number of credit hours, Difference in number of credit hours this term vs last
- Continuous random variables take real (decimal) values
  - Example: Cost of books this term, Difference in cost of books this term vs last

Characterizing Random variables

- A random variable (RV) has a probability mass function (PMF) (for discrete RVs) and a probability density function (PDF) (for continuous RVs) which describes the possible outcomes and their probability.
- A probability is always a value in the interval $[0, 1]$.
- The total probability of all outcomes of a RV is 1.



Example Probability Mass Function (PMF)



Example Probability Density Function (PDF)

PE1

Open a blank Excel sheet and generate the following random variables, using the =RAND(), =RANDBETWEEN(), and =INDEX() Excel functions:

1. A continuous uniform random variable between 0 and 1
2. A continuous random variable between 2 and 5
3. A discrete uniform random variable between 2 and 10
4. A discrete random variable taking on the values 3, 4, and 8 with equal probability.

Expected Value of a Random Variable by Hand

- We are often interested in the average outcome of a random variable.
- We call this the expected value, and it is a probability weighted sum of the possible outcomes.

For a discrete random variable:

$$E(X) = \sum_{x} x\, p(x)$$

Example, what is the expected value of a dice roll?

$$1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = 3.5$$

The expected value does NOT have to be a possible value, rather, it represents the average you would get after many trials.

Expected Value and Standard Deviation in Excel

To find the expected value of a RV in Excel, we first have to generate many samples of the RV, then use the **=AVERAGE( )** command on the samples.

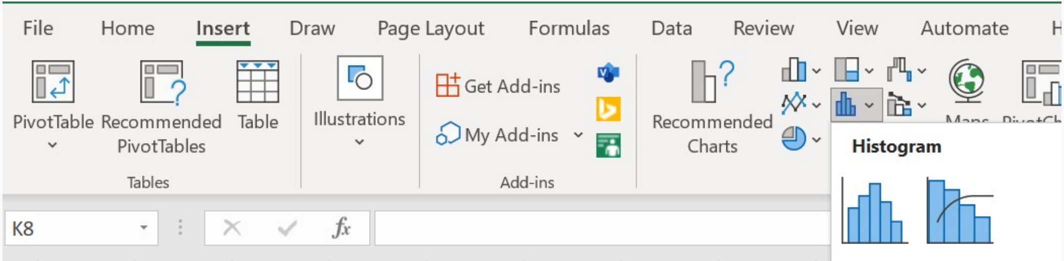
We are often interested in the standard deviation (measure of spread) of a random variable.

To find the standard deviation of a RV in Excel, we first have to generate many samples of the RV, then use the **=STDEV.S( )** command.

Histogram of Outcomes

- The expected value and standard deviation only tell part of the story. A histogram of outcomes (which will have the same shape as the PMF or PDF) describes all of the outcomes from many trials of the same random variable.

If we generate a certain number of outcomes for a random variable (i.e., we simulate outcomes, then we can plot a histogram of the results.



PE2

Create a new Excel sheet and simulate 4000 outcomes for a standard normal random variable, using =norm.inv(rand(),0,1).

1. calculate the expected value of this random variable based on your simulation. Compare this to the theoretical mean of 0.
2. Calculate the standard deviation of the outcomes. Compare this to the theoretical standard deviation of 1.
3. Create a histogram of your simulation outcomes.

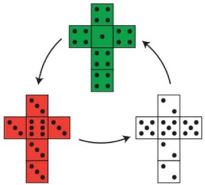
PE3

Consider the problem below,

1. Without doing any calculations, which die would you expect to be better in the game?
2. In Excel, find the expected value of each die. Does this change your mind?
3. In Excel, simulate 20 outcomes for each of the game types: red vs. green, red vs. white, and white vs. green.

Jon's dice

Jon has three six-sided dice with unusual numbering.



A game consists of two players each choosing a die. They roll once and the highest number wins.

Conclusion

Next Class

Homework:

- Read PDF on probability distributions (will be posted on Teams)

Next Lesson:

- Describe the characteristics of four discrete probability distributions that can be incorporated into a simulation (General, Discrete Uniform, Bernoulli, Binomial, Poisson) using the PMF and CDF.
- Describe the characteristics of three continuous probability distributions that can be incorporated into a simulation (Uniform, Normal, Exponential) using the PDF and CDF
- Use Excel to simulate random variables from these distributions