

Lecture 3

Basics of Probability



Grace Thomson

Instructor

## Basics of Probability

In this session we will learn about:

1. Application of inferential statistics to solve business problems.
2. Rules of probability
3. Discrete and Continuous Probabilities and application to business studies.

### Definitions

First let's define Probability:

Probability (p)	Chance that a particular event will occur. Denominated as "p" It takes values from $0 \leq p \leq 1$ , therefore expressed in decimals and percentages.	$\Rightarrow$	0 --- will not occur 1 --- will occur $0 \leftrightarrow 1$ uncertainty of occurrence of an event.
--------------------	--	---------------	---

Probabilities are originated by experiments. An experiment is a process that produces an outcome that can't be predicted with certainty (the rolling of a dice, for example). In business, an experiment may be a selection or a decision. Therefore an investment decision is an experiment, as well as the production in a manufacturing line, or the location of a warehouse or the success of a retail location. In none of these cases you can be certain of how good or how bad those actions will turn out, so there is a probability attached to them, right?

In statistics, the possible results of an experiment are called **(ei) elementary events**. And a set of all the elementary events is called **Sample space**, it is the set of all possible results that a decision may yield. Within this sample space, the researcher defines exclusively what interests him, and this is called **event of interest (E)**. For example: A number two in a rolled dice, the number of damaged pieces in an industrial line, the number of delinquent accounts in a portfolio, etc.

In Probability Theory, *events* may have different nature.

Table 1

Types of events in Probability Theory

Mutually exclusive	Occurrence of one precludes occurrence of the other event.	⇒	Gender of a retail customer: E1= gender of first customer is male E2= gender of first customer is female.  E1 and E2 are mutually exclusive, this can't happen at the same time.
Independent	Occurrence of one event in no way influences the probability (p) of occurrence of the other.	⇒	E1= Gender of first customer is male E2= Gender of second customer is male  E1 and E2 are independent; E1 doesn't affect the outcome for E2.
Dependent	Occurrence of one event impacts the probability (p) of the other	⇒	E1 = Number of delinquent accounts in my portfolio  E2 = Number of new accounts without credit check.

In order to assess the probability of occurrence of an event, you may use three different methods. Each one is applied depending on the type of research you are doing, the information available, your experience, or standards of the market, if any.




### **Methods to assign probability**

There are three methods used to assess the probability of an event, a) classical b) relative frequency of occurrence and c) Subjective assessment.

Let's call  $P(E_i)$  = Probability of occurrence of event  $E_i$ ,

Table 2

## Methods to assign probability

Classical probability assessment or A priori	Measures the number of ways how an event $E_i$ may occur, based on previous experience, rules or public knowledge. <i>E.g. probabilities of a face in a rolled dice, or in a coin.</i>	$P(E_i) = \frac{\text{Number of ways } E_i \text{ can occur}}{\text{Total number of } e_i \text{ equally likely to happen.}}$
		 Classical probability is hard to apply in real life as events are not equally likely to happen. e.g. we wouldn't know the p% of success of our new product line, as there is not a standard of success or failures for this event.
Relative Frequency of Occurrence (RFO)	Measures the number of times an event of interest occurs, and is based on actual data and/or historical recollection.	$\text{RFO}(E_i) = \frac{\# \text{ of times } E_i \text{ occurs}}{(N) \text{ number of trials}}$
		 Commonly used by decision-makers, who consider their latest outcomes as probabilities.
Subjective probability Assessment	Widely used in Business fields, when experience and history is not enough to explain events or when no history is available.	 $P(E_i) \rightarrow$ decision maker's state of mind regarding the chances of occurrence, based on rapport with client, personal profile of customers, etc.

**LEARNING TEAM ACTIVITY:**

**Think about your workplace. Can you give at least 1 example of experiments for which the probability is assigned by RFO and one for which probability is assigned SUBJECTIVELY?**

## Rules of probability

Table 3

### Summary of Rules of Probability

Rule 1 Possible values of p	$P(\bar{E}) = 1 - \sum_{i=1}^k P(E_i) = 1$	Probability can't be negative, but can't be more than 1.
Rule 2 Sum of elementary events $\rightarrow \Sigma = 1$		K=# elementary events in the sample ei= ith elementary event
Rule 3 Addition rule for elementary events	$E_i = \{e_1, e_2, e_3\}$ Then $P(E_i) = P(e_1) + P(e_2) + P(e_3)$	Add the probabilities of all the elementary events within an Event of interest
Complement rule	Probability of complement	
Rule 4 Addition rule for any 2 events $E_1$ & $E_2$	$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2) - P(E_1 \text{ and } E_2)$	Add probabilities of each event and subtract the probability of common events.
Rule 5 Addition rule for mutually exclusive events	$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2)$	Simply add the probabilities of each event.
Rule 6 Conditional probability for any 2 events $\rightarrow$ sucession	$P(E_1/E_2)$ is the probability that event $E_1$ will occur given that some other event has already happened  $P(E_1/E_2) = \frac{P(E_1 \text{ and } E_2)}{P(E_2)} \rightarrow \begin{matrix} \text{joint probab.} \\ \text{marginal prob.} \end{matrix}$	$P(E_2) > 0$
Rule 7 Conditional probability for independent events	$P(E_1/E_2) = P(E_1)$ ; $P(E_2) > 0$ $P(E_2/E_1) = P(E_2)$ ; $P(E_1) > 0$	

## Bayes' Theorem

This is a very simple concept, based on **conditional probabilities**. It relies on the fact that in real life additional information about events is always available, and once it's provided, the assessment of the probability is different. This is very common in business environments where managers rely on studies to enhance their analysis

Let's start by looking at the formula. What is the probability of an Event  $i$  given that Event  $B$  has occurred?

$P(E_i/B)$  → Probability of  $E_i$  given that  $B$  has occurred.

$P(E_i/B)$  is simply a **revised probability** → with a conditional probability in the numerator, and the probability of all the possible events in the denominator.

$$P(E_i/B) = \frac{P(E_i)P(B/E_i)}{P(E_1)P(B/E_1) + P(E_2)P(B/E_2) + \dots + P(E_k)P(B/E_k)}$$

ith event of interest                      New information  
 ↓    ↙

### Application of bayes' theorem to business

A company that has two soap manufacturing plants in Ohio and Virginia needs to know "what is the probability that given a report about defective soaps ( $D$ ), the soaps were originated in Ohio ( $O_i$ )?  **$P(\text{Ohio}/D)$** .

The key element here is that they don't have that direct report, but they do have information about how many defective soaps within each location they usually have  $P(\text{Defective}/\text{Ohio})$  and  $P(\text{Defective}/\text{Virginia})$  and they also know about the general participation of the Ohio plant in the total production  $P(\text{Ohio})$ .

#### Participation of each plant in production

$P(\text{Ohio}) =$  probability that soap is produced in Ohio → 0.60

$P(\text{Virginia}) =$  probability that soap is produced in Virginia → 0.40



Sum of these two probabilities must be 1 → 0.6 + 0.4

**Probability of defective soaps within each plant**

$P(\text{Defective}/\text{Ohio}) = \text{Probability of defective soaps in the Ohio plant} \rightarrow 0.05$

$P(\text{Defective}/\text{Virginia}) = \text{Probability of defective soaps in the Virginia plant} \rightarrow 0.10$

With this information we build the following table:

On the first column write the names of the plants. On the 2<sup>nd</sup>. column write the probability that soaps come from those plants. On the 3<sup>rd</sup>. column write the conditional probability of having defective soaps within each plant (this is the conditional probability). On the 4<sup>th</sup> column operate a **joint probability** by simply multiplying the prior probability times the conditional probability; add that column up. Use that total to compute the last column of revised probability, by dividing the joint probability of Ohio by the total. Now you have a 0.4286 probability that soaps that are defective come from Ohio.

**Table 3****Summary Table for Bayes Theorem Application**

Events	Prior probability	Conditional probability	Joint Probability (prior * conditional)	Revised probability
Ohio	0.60	0.05	$0.60 * 0.05 = 0.03$	$0.03/0.07 = 0.4286$
Virginia	0.40	0.10	$0.40 * 0.10 = 0.04$	
	1.00		0.07	

LEARNING TEAM ACTIVITY: You can use this same table to compute  $P(\text{Virginia}/D)$ : The probability that defective soaps come from Virginia. Give it a try

R: 0.5714

*Notice that this is just the complement of the previous probability*

Let's now play some Virtual Games to see how probability problems, estimates and inference arise. All of these games are available at **SOCR Experiments**

([http://socr.ucla.edu/htmls/SOCR\\_Experiments.html](http://socr.ucla.edu/htmls/SOCR_Experiments.html)):

### 1. Let's-Make-A-Deal or Monty Hall Paradox (Monty Hall Experiment):

The screenshot shows the SOCR Experiments web interface in a Netscape browser. The browser window title is "SOCR: Statistics Online Computational Resource - Netscape". The address bar shows "http://socr.ucla.edu/htmls/SOCR\_Experiments.html". The page has a navigation menu with "SOCR Home", "Tools", "Distributions", "Experiments", "Analyses", "Games", "Modeler", "Charts", and "More". The main content area is titled "SOCR Experiments" and features a banner image with a car and a goat. Below the banner is the "Virtual Experiments" section, which is currently displaying the "Monty Hall Experiment".

The "Monty Hall Experiment" interface includes a control panel on the left with buttons for "Reset", "Help", "About", "Snapshot", and "COPY". It also has a "Show Model Distribution" option with radio buttons for "ON" and "OFF". The main simulation area is divided into three panels: "1" (a yellow door labeled "Door 1"), "2 Player" (a blue car), and "3 Host" (a grey goat). Above the simulation area are controls for "Update 1" and "Stop 10", and a "Standard" dropdown menu. A progress bar shows "p = 1.00".

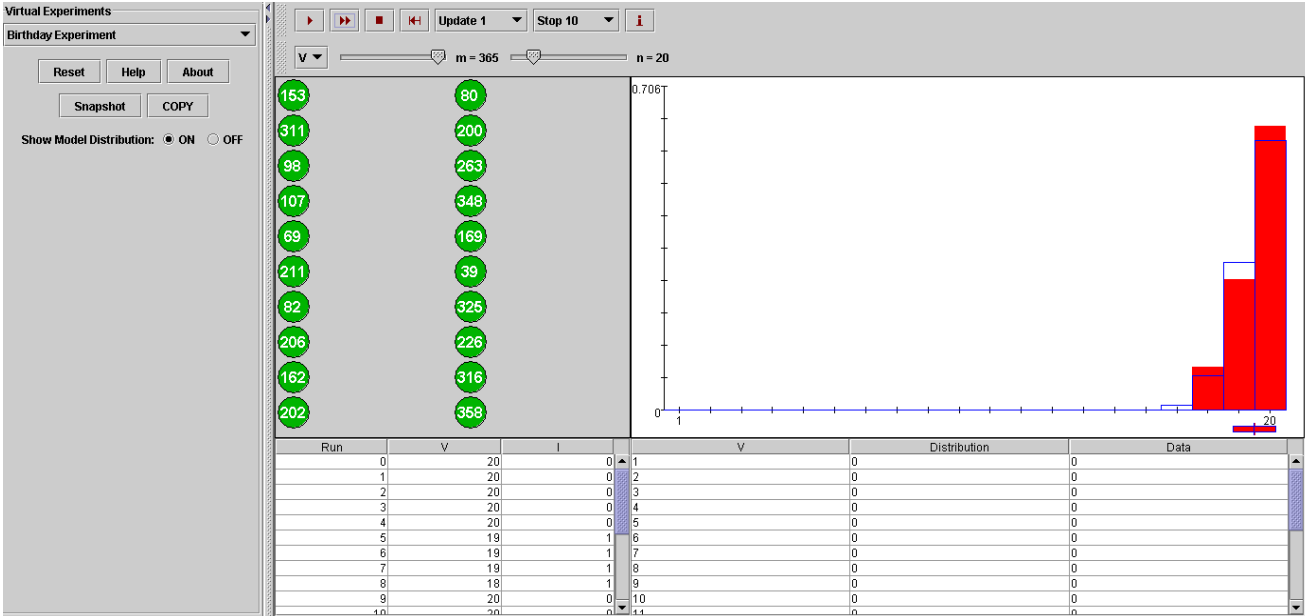
Below the simulation area is a table with the following data:

Run	G	S	W	W	Distribution	Data
1	1	1	0	0	0.33333	0.5
2	1	1	1	1	0.66667	0.5

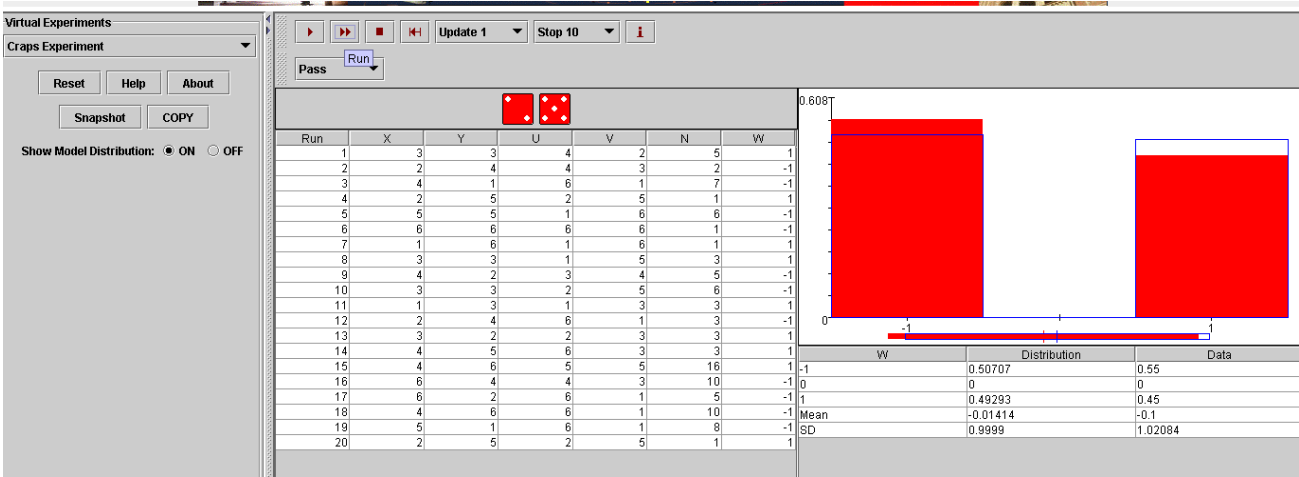
To the right of the table is a histogram showing a step function with a red area under the curve. The x-axis is labeled "0" and "1", and the y-axis is labeled "1".

### 2. The Birthday Paradox

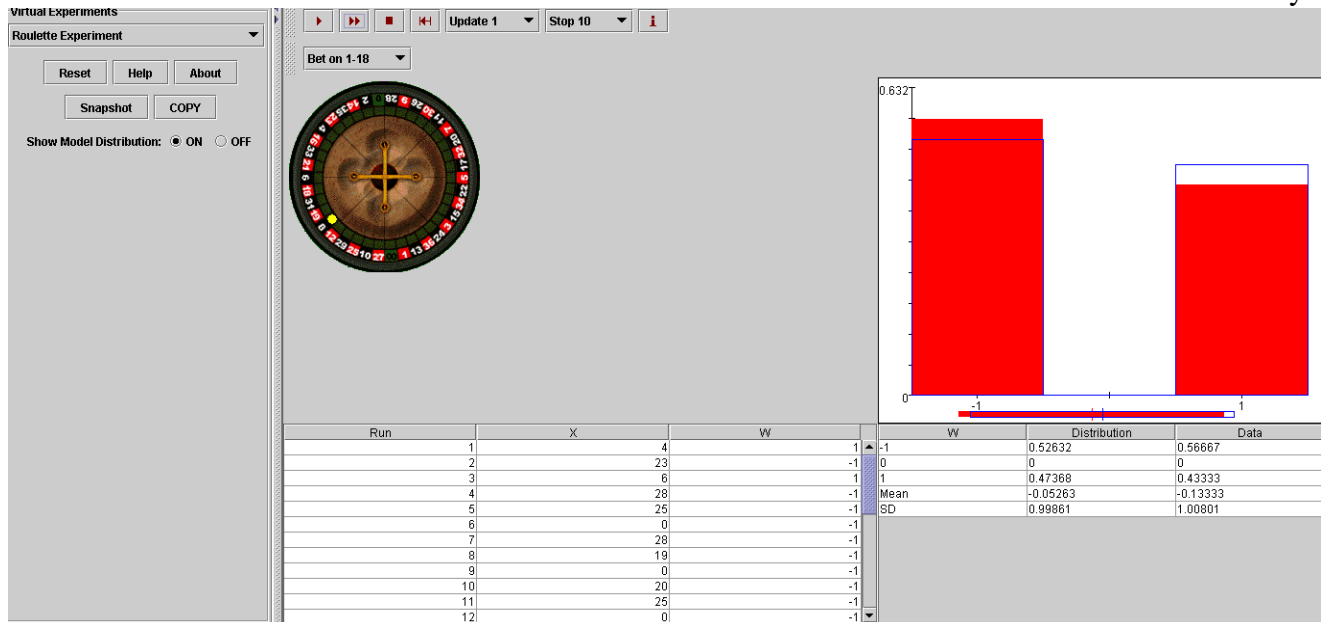




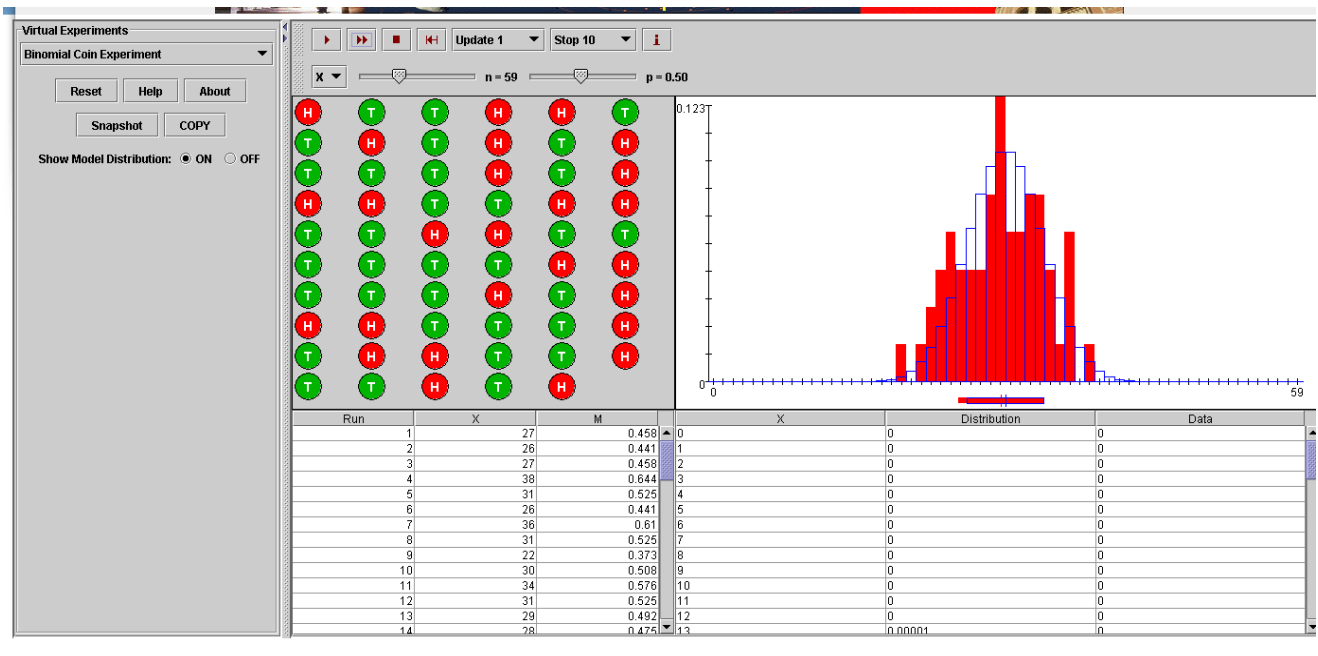
3. Game of Crabs



4. Roulette Game – what are the odds of winning if we bet on an outcome between 1-18?



5. Coin Toss



Here is more information about these SOCR Virtual experiments:

[http://wiki.stat.ucla.edu/socr/index.php/SOCR\\_EduMaterials\\_Activities\\_Birthday](http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_Birthday)

[http://wiki.stat.ucla.edu/socr/index.php/SOCR\\_EduMaterials\\_Activities\\_BinomialCoinExperiment](http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_BinomialCoinExperiment)

[http://wiki.stat.ucla.edu/socr/index.php/SOCR\\_EduMaterials\\_Activities\\_RouletteExperiment](http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_RouletteExperiment)

[http://wiki.stat.ucla.edu/socr/index.php/SOCR\\_EduMaterials\\_Activities\\_MontyHall](http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_MontyHall)

[http://wiki.stat.ucla.edu/socr/index.php/SOCR\\_EduMaterials\\_Activities\\_Discrete\\_Probability\\_examples](http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_Discrete_Probability_examples)