Outline of Probability

Christopher Enright

Legal Group

for smarter lawyers

www.legalskills.com.au

Outline of Probability

Christopher Enright

I feel like a fugitive from the law of averages.[1]

Sinch
for smarter lawyers

Sinch	
Business Name	Sinch
Firm Name	Sinch Software Pty Ltd
Seminars	www.sinch.com.au
Electronic Products	www.sinch.com.au
Books	www.legalskills.com.au
Email	info@legalskills.com.au

First Edition	2016
Copyright	© Christopher Enright 2016

Citation
Enright, Christopher (2016) ‘ Outline of Probability’ Sinch, Canterbury

Author
Christopher Enright is a barrister, solicitor and chartered accountant. Chris has a Master of Commerce (Management) degree from the University of New England. In an earlier life he lectured in law and management at various universities.

1. Bill Maudlin *Up Front* (1946) p 39

Preface

Relevance of Probability

Australian law schools tend to be light in their treatment of the associated fields of legal reasoning and legal skills. In an attempt to rectify part of this neglect this article provides an elementary account of probability. The legal system uses probability in several circumstances. To facilitate this the article provides a basic account of probability.

One common circumstance for using probability is projections by actuaries for the purpose of assessing damages for personal injury. Another major use is in findings facts in a court case. Since this is fundamental to the legal system it is worth outlining how it works.

Model for Proof of Facts

There is a model for proof of facts with a scale of proof and five steps that utilise this scale of proof.[1] The following diagram sets out the titles of these steps:

Steps	**Content**
Step 0	Scale of Proof
Step 1	Starting Point: Burden of Proof
Step 2	Versions of Truth
Step 3	Probability of Truth
Step 4	Finishing Point: Standard of Proof

Model for Proof of Facts

Step 0 Scale of Proof

Proof of facts is a matter of probability because the concept of probability underlies the reasoning process. To explain this further the diagram below sets out a scale of proof with probabilities that run from 0% to 100%:

0%	←→	25%	←→	50%	←→	**51%**	←→	75%	←→	**99%**	←→	100%

Scale of Proof

The four steps in the model enable a lawyer or a court to approach the task of proving facts in a structured way. The basis of this explanation is as follows.

1. There is a more detailed accounnt of proof of facts in Christopher Enright (2015) *Proof of Facts* 2nd ed Sinch, Canterbury.

Step 1 Starting Point: Burden of Proof

Step 1 views a case from the perspective of the party who initiates the case – the author labels this person as the initiator. Step 1 lays down the rule as to where on the scale of proof a case starts. At the commencement of a case the initiator is at the 0% mark on the scale for most issues in a case. So, the 0% mark constitutes the starting point for the initiator – they have it all before them. This is why the 0% mark represents the burden of proof or onus of proof as it is also called. It shows that at the outset no part of the initiator's case is proved – this is because the initiator bears the onus of proof.

Step 2 Versions of Truth

Step 2 involves the parties putting versions of truth before the court. These versions of the truth are the versions of the facts of the case that each party argues represent the truth of the matter. At the same time as the parties present their versions of the facts they also present evidence for the disputed facts of the case in their attempt to prove their version of the facts.

Step 3 Probability of Truth

Step 3 entails the court considering the evidence before it. The court uses this evidence to make an assessment of how probable are the sets of facts that each party presents.

Step 4 Finishing Point: Standard of Proof

To perform Step 4 for a case there are three tasks.

<u>Task 1. Identifying the Required Standard of Proof</u>

It is necessary to determine the standard of proof required for the type of case before the court. Achieving the required standard of proof is the finishing point for a party presenting a case to a court. This law relating to proof lays down two very common standards:

1. Civil Cases. The general standard of proof required for civil cases is proof on the balance of probabilities. This is taken to be 51%.
2. Criminal Cases. The general standard of proof required for civil cases is proof beyond reasonable doubt. There is no percentage that officially portrays this standard.

<u>Task 2. Determining the Actual Standard of Proof</u>

Step 3 in the model for proving facts involves determining the actual standard of proof. Step 3 is labelled 'Probability of Truth'.

Task 3. Making the Comparison
It is necessary to compare the standard of proof that the law requires with the standard of proof that the initiating party has attained.

When the court has done this there will be one of two possible outcomes:
1. Achievement of Standard of Proof. The initiating party has achieved the required standard of proof. Because of this the initiating party wins the case.
2. Non Achievement of Standard of Proof. The initiating party fails to achieve the required standard of proof. Because of this the initiating party loses the case.

Christopher Enright
Stockton
11 November 2016

Contents

Table of Cases

Abbreviation	Meaning
ANRC	Australian National Railways Commission

Table of Diagrams

Labels

Section 2 Measuring Probability

Section 5 Deriving Probability

Labels

Introduction
Describing Items
Listing Items
Diagrams
Probability

Introduction

Discussion in this publication refers to items such as a statute or a meaning of an ambiguous provision. Often these are part of a collection, list, range or set of items. Frequently the text puts them in a diagram where they represent a model or a step on the way to explaining a model. The purpose here is to explain the labelling system used to refer to these items.

Describing Items

Labelling Items

There are several aspects to labelling the items in a set, range, list or collection. These are name, number, letter and designating a set of items.

Name

The name of an item commences with a capital letter. Some examples are Element, Statute and Meaning.

Number

Items in a set, range, list or collection are generally numbered. For example, the elements of a legal rule are labelled Element 1, Element 2, Element 3 and so on. These numbers are ways of identifying elements and distinguishing one from another. They are generally not intended to create any list according to preferences or values.

Letter

Items in a set, range, list or collection can be lettered. For example a list of statutes can be Statute A, Statute B and so on.

Designating a Set of Items

It is useful to designate a set of items with a single and simple tag. Here is an outline. The basic proposition is that a simple and obvious tag has two aspects:
1. Description. Use a written label on the items as a tag or description. Put it in plural form. Thus a tag for a set of statutes would be ‘Elements’.

2. Numbers. After the tag add a space then a compound numerical tag consisting of three items:
   2.1 The number of the first item in the set.
   2.2 A hyphen.
   2.3 The number of the last item in the set.

Here are two illustrations:
1. A set of six elements would be Elements 1-6.
2. A set of elements where the number can vary from situation to situation is written as Elements 1-n.

1. Naming the Items

The item has a name, which is usually obvious. For example each statute in a set of statutes would bear the name 'Statute', and each elements in a set of elements would be 'Element'.

2. Numbering the Items

There are two possibilities for the numbering of a set, list or range of items:
1. There can be a fixed number in the set.
2. There can be a variable number in the set.

2.1 Fixed Number in the Set

In a particular instance there may be a specific number of items in a set. For example a particular legal rule might be composed of five elements. In this case the first and last numbers designate the number of items in the set or range. In this example of a set of five elements, one would designate the set as 'Elements 1-5'.

2.2 Variable Number in the Set

Sometimes the text refers to a set or a list in general terms in cases where the number of items in the set can vary from situation to situation. In this event, the way to go is to number the last item with the symbol 'n'. To refresh readers, 'n' stands for however many there are on a particular occasion. An example would be a general discussion about elements of a legal rule. In this case the possibilities vary from legal rule to legal rule. Thus the designation of this set of items is Elements 1-n.

Null Option

There is a special case with options where one of the options is to do nothing and leave things as they are. This occurs, for example, with the proposed making of a statute where one option is just not to enact a statute. In a case such as this the option is labelled with the symbol for nought, namely '0'. Thus the option not to enact a statute is designated as Statute 0. Statute 0 represents the null option – it is the option for a legislature not to enact a statute on a topic whereas Statutes 1, 2 3 and so on are options for different versions of a statute on a topic (on the basis that there is no form of a statute that can better present conditions). Given this the full set or range of possible statutes for a legislature to enact consists of Statutes 0-n.

Corresponding Items

Sometimes there are sets with corresponding items. This can occur for a number of reasons. Here are two examples:

1. For making and interpreting law, items correspond because of causation. Each version of a statute on a subject and each meaning of an ambiguous provision will cause an effect if a legislature enacts the statute or if a court declares the meaning to be legally correct.
2. In the model for litigation, elements and facts correspond because each element delineates a category of facts so that in a particular case the element is satisfied by a fact that falls within that category. Similarly, facts and evidence correspond because each fact is proved or potentially provable by some evidence.

Single Relationships

Corresponding items are labelled with the same number or letter. Here are some illustrations:

1. Statutes, Meanings and their Predicted Effects. Statute 0 is predicted to cause Effect 0, Statute 1 is predicted to cause Effect 1, Statute 2 is predicted to cause Effect 2 and so on. Meaning 1 is predicted to causes Effect 1, Meaning 2 is predicted to cause Effect 2 and so on. Similarly, Statute X (or Meaning X) is predicted to cause Effect X while Statute Y (or Meaning Y) is predicted to cause Effect Y.
2. Facts Satisfying Elements. Fact 1 is the label given to a fact that fits within or satisfies Element 1, Fact 2 is the label given to a fact that fits within or satisfies Element 2 and so on.
3. Evidence Proving Facts. Evidence 1 is the label given to evidence that might prove or has proved Fact 1, Evidence 2 is the label given to evidence that might prove or has proved Fact 2, and so on.

Collective Relationships

It is possible to use labels of correspondence to make collective statements. Here are some examples: Statutes 0-n are predicted to cause Effects 0-n, while Evidence 1-n is capable of proving Facts 1-n. To construe these collective statements properly it is necessary to apply the maxim *reddendo singula singulis*. Literally this says that each is rendered on their own. In plainer language, the items are to be taken singularly so the each item in the first list is paired with the corresponding item in the second list. The adverb ‘respectively’ captures this notion.

Two or More Version of an Item

There may be two or more versions of an item. Additional letters or numbers can distinguish the different versions. For example:

1. If Element 2 is ambiguous because it has two meanings, the versions of Element 2 can be designated Element 2A and Element 2B.
2. There can be two versions of a fact. There are two major possibilities:

2.1 In a case there may be two versions of Fact 2 because the plaintiff propounds one and the defendant propounds the other. These can be designated 'P' and 'D' to signify the plaintiff and defendant's version. Thus the two versions are Fact 2P and Fact 2D.

2.2 After investigating the facts of a case the defendant may find that there is evidence to support two versions of one of the facts in their case. These are facts that the defendant could use to rebut the plaintiff's satisfying Element 3. The defendant or the court could designate these as Fact 3D.1 and Fact 3D.2.

Subdivisions of Items

It is possible to designate subdivisions of an item with a numbering system that invokes the form but not the meaning of decimal points. Thus if Element 2 has three sub-elements, one can designate them as Element 2.1, Element 2.2, and Element 2.3. If Element 2.2 has three sub-elements we can designate these as Element 2.2.1, Element 2.2.2 and Element 2.2.3. Obviously this form of numbering adapts to any number of levels of subdivision.

Possibilities: 'X', 'Y', Etc

Sometimes the text needs to refer to any option, that is, to an option in general terms. Conveniently this is labelled with a capital letter. Commonly, this is the letter X, so that a general option for a legislature wishing to pass a statute is Statute X. Naturally, if there is a need to refer to more than one option additional letters may be used. For example, there could be reference to Statute X and Statute Y; in this case Statute X is one possible statute and Statute Y is another possible statute.

Signifying Relationships

Sometimes it is necessary to signify a relationship between two items. This can be done using standard symbols. This table sets out the major possibilities:

Symbol	Relationship	Illustration
$<$	Less than	$X<Y$. X is less than Y.
$>$	Greater than	$X>Y$. X is greater than Y.
$=$	Equals	$X=Y$. X equals Y,
$\neq$	Not Equals	$X\neq Y$. X does not equal Y.
$\approx$	Approximately Equals	$X\approx Y$. X is approximately equal to Y.

≡	Congruence Relationship	X≡Y. X is congruent with Y.
≅	Isomorphic	X≅Y. X is structurally identical to Y
Labels Diagram 1. Symbols for Relationships		

Listing Items

Where there is a list, for example a list of the meanings of an ambiguous provision, we can set these out in the text as a series – Meaning 1, Meaning 2 ... Meaning n. In the text, as we have noted, the range can be efficiently represented as Meanings 1-n. In a table they are set out as a list in the following way:

Meanings
Meaning 1
Meaning 2
Meaning n
Labels Diagram 2. List of Meanings

In this presentation it is not strictly necessary to include Meaning 2. Indeed, it is actually redundant, when n=2. However, it usefully emphasises the sense of a list that sets out the range of options or possibilities.

Diagrams

Lists in a table can be connected to become a diagram or figure. This can involve corresponding items. A useful illustration consists of a diagram that has two major columns that match corresponding items. One column sets out the meanings of an ambiguous provision in a statute in Statute X and the other sets out the effect for the whole statute that each meaning is predicted to cause.

Here is the illustration:

1	**2**	**3**	
Meanings	→	**Effects**	**1**
Meaning 1		Effect 1	**2**
Meaning 2		Effect 2	**3**
Meaning n		Effect n	**4**
Labels Diagram 3. Meanings and Effects			

This diagram functions in the following way:
* Column 1 shows the meanings of the ambiguous provision, being Meanings 1-n.
* Column 3 shows the effect of the statue that each meaning is predicted to cause if a court chooses them as the legally correct meaning of the ambiguous

provision. Let us flesh this out. Every statute that is enacted causes a number of outcomes. The author refers to the full collection of outcomes that a statute is predicted to cause as an effect. When a court interprets a statute it is faced with the basic options in terms of the range of meanings of the ambiguous provision that gives rise to the need to interpret the statute. The diagram labels these meanings as Meanings 1-n. If a court decides that Meaning 1 is the legally correct meaning of the ambiguous provision that decision is likely to have an impact on the effect that the whole statute will cause. Column 3, as stated, sets out this effect, the effect of the whole statute, for Meaning 1. In a similar way it sets out the effect for each other meaning of the ambiguous provision. This method of identifying the effects of each meaning caters for the constitutional rule in each Australian jurisdiction that requires a court to interpret a statute in the way that will 'best achieve' the purpose and object for which the legislature enacted the statute. Now the purpose or object of a statute is to cause some effect or outcome. Hence the term 'Effect' aligns directly with purpose and object (which of course is why the table includes it).

* Column 2 contains an arrow pointing from the Column 1 to Column 3, thereby indicating that each meaning in Column 1 is predicted to cause the statute to have the corresponding effect in Column 3.

* Columns 1-3 indicate meanings and their predicted effects. Assume for the purposes of the explanation that a court is interpreting an ambiguous provision in Statute X that has Meanings 1-3:

1. If a court chooses Meaning 1 as the legally correct meaning the prediction is that Statute X will cause Effect 1.
2. If a court chooses Meaning 2 as the legally correct meaning the prediction is that Statute X will cause Effect 2.
3. If a court chooses Meaning 3 as the legally correct meaning the prediction is that Statute X will cause Effect 3.

Probability

A number of symbols are used for probability. This diagram shows the common symbols and their meanings:

Symbol	Meaning
P(A)	probability that event A occurs
P(B)	probability that event B occurs
$P(A \cup B)$	probability that event A or event B occurs (A union B)
$P(A \cap B)$	probability that event A and event B both occur (A intersection B)
P(A')	probability that event A does not occur
P(A | B)	probability that event A occurs given that event B has occurred already (conditional probability)
P(B | A)	probability that event B occurs given that event A has occurred

	already (conditional probability)
P(B | A’)	probability that event B occurs given that event A has not occurred already (conditional probability)
ϕ	the empty set = an impossible event
S	the sample space = an event that is certain to occur

Labels Diagram 4. Symbols Used for Probability

Section 1 Introduction

The account of proof of facts in this book is based on probability. This chapter provides an outline of probability for readers who may have forgotten their school mathematics and need a refresher. The last section of this chapter explains the symbols used in probability.

Probability

Probability caters for uncertainty. Probability is widely used in the human sciences such as management and psychology and also the physical sciences. As Professor Alan Hájek neatly puts it, '[p]robability is virtually ubiquitous' since it is used in so many disciplines of both academic and practical importance.[1] It is, therefore, relevant to any study of law from the human sciences perspective. It is directly relevant to proof of facts since proof is based on probability.

Commonly but loosely people treat probability as referring to the likelihood of an event's occurring. In fact, this is a fallacy, called the mind projection fallacy. It is a fallacy that probability is 'a property of objects and processes in the real world'.[2] Probability is a state of mind, not the state of the world. It is a measure of the strength of our belief that an event will occur or has occurred.[3]

There are two aspects to this:

1. Descriptive Function. There is a descriptive function. Probability describes how certain we are about the truth of something.
2. Derivative Function. The second function of probability is derivative. When one or more probabilities are known and quantified further probabilities can be derived by rules based on deduction. This chapter explains and illustrates these rules.

This chapter uses the standard symbols that are utilised in the mathematics of probability. These are set out in the preliminary pages to this book under Labels.

Uses

Introduction

There are two uses of probability that are of fundamental importance to working with law. One arises in using policy to make law, while the other is located in the task of finding facts.

---

1. Alan Hájek's 'Interpretations of Probability'
2. Robertson and Vignaux (1993) p 460
3. Robertson and Vignaux (1993) p 462

Making Law
Probability is deployed in the process of making legislative policy. Any law is essentially trying to change the future. Consequently as part of the process of making of law it is necessary to know what effects a proposed law is likely to cause. This invokes probability in two ways.

First, to identify the effects that a law will cause, it is necessary to invoke laws of behavioural science. These are derived using an experimental process that frequently relies on statistical inferences. Statistical inferences are based on probability.

Second, even if a causal law suggests that a proposed statute might cause a particular effect, there may be some uncertainty. Since probability is a means of encapsulating uncertainty surrounding future events, it can be used to factor this uncertainty into the tasks of making and interpreting a law.[4]

Finding Facts
Probability is used by courts in the process of finding facts as a means of coping with the uncertainty that is so often inherent in the task. Probability is a way of encapsulating and working with uncertain knowledge.

Function of Probability
It is a common occurrence that we humans are uncertain about something such as the happening of some event in the future. In these cases probability can be invoked as a means of catering for this uncertainty. To explain uncertainty, however, it is first necessary to consider the basis of certainty and uncertainty.

Certainty
Where a person is certain that they know something they believe that it is true with a probability of 100%. They are certain about the truth when either of two things happen. They have observed something, such as X or Y, for themselves.

Or, they have observed one thing (about which they are then certain), and they are also certain about a behavioural law. This behavioural law can takes either of two forms:
1. Another thing will happen in consequence of the first. This can be written in the form: ‘When X occurs Y will (later) occur’.
2. Another thing has happened to cause the first. This can be written in the form: ‘When Y has occurred X has occurred beforehand because X is the sole cause of Y’.

4. Enright (2015) *Legal Reasoning* Chapters 13-16

Uncertainty

In other cases a person is not certain about the truth. Why are they not certain? They are not certain for either or both of two reasons:

1. They are not certain about X or Y because they have not observed either of these for themselves, or they have observed them but have not observed them properly.
2. They are not certain about the truth of a behavioural law that would enable them to say either of two things. (a) When X occurs Y will (later) occur. (b) When Y has occurred X has occurred beforehand because X is the sole cause of Y.

In these cases when the person is not certain about X or Y they believe that X or Y is true with a probability of something less than 100%. This is something that falls between 0% and 100%.

This analysis reveals the fundamental point. We cannot predict the future or know the past with certainty when we do not have perfect information. In the absence of perfect information it is necessary to resort to probability. If we cannot know something for certain the next best thing is to know, as a plausible estimate, a reasonable belief as to the chance of its occurring. For example X is a disease and we believe, following research, that it will occur in about 37% of the population.

Using probability is reassuring because it reduces uncertainty. Reducing uncertainty is an advantage for at least two reasons. First, uncertainty makes most of us anxious. Second, individuals, institutions and society can make plans for the future. For example, if a disease will occur in 37% of the population the health system can allocate resources in advance to treat and prevent it.

A statement of probability may go in either of two directions. It may use an event to predict the probability of a second future event. This probability is based on the behavioural law in the form 'X causes Y'. Probability may also be used to determine whether a past event might have happened. This can be done, for example, when there is a causal law in the form 'X is the sole cause of Y'. When Y is observed to have happened, this causal law gives us a basis for believing that X happened, and happened at an earlier time.

Uncertainty with Probability

Probability came to prominence in the seventeenth century to explain games of chance. However, it was not until 1933 that a major attempt was made to explain it by postulating axioms on which probability properly should be built. Andre Nilolaevich Kolmogorov did this (1903-1987) in his classical text *Foundations of the Theory of Probability*. These axioms put probability on a firm footing. But notwithstanding this, the philosophical foundations of probability are still subject to major debate. This uncertainty permeates discussion and literature on probability. Yet for all this, the use of probability in the various sciences continues unabated. Given the purpose of this book it is not feasible to delve into any of the philosophical debate on probability. Instead, discussion here is confined to an attempt to explain in as clear a way as possible from a practical perspective the basic principles and rules of probability.

Section 2 Measuring Probability

Introduction

Probability is abbreviated ‘P’. It ranges between absolute certainty that an event will occur and absolute certainty that it will not occur. Probability can be measured by verbal formulas or numerical formulas. Numerical formulas can be decimals, percentages or fractions.

Verbal Formulas

Probability can be measured, in a rough sense, by verbal formulas. We can and do use phrases such as ‘well founded’,[5] ‘fairly certain’, ‘most likely’, ‘with a strong chance’, ‘reasonably likely’, ‘beyond all reasonable doubt’ (the standard of proof in criminal cases), ‘glaringly improbable’ and so on, to indicate the strength of our belief.[6]

Numerical Formulas

More commonly, however, probability is measured with numbers. This has two advantages: numbers can be compared, and numbers can be used to calculate other related or derived probabilities. There is, however, an underlying disadvantage, because measuring probability with numbers is not an exact process when the numbers are just estimates, which is often the case. Hence the mathematical processes used in deriving and comparing probabilities can convey a false sense of certainty.

Three scales are commonly used when probability is measured with numbers, although they amount to the same thing. These consist of:

1. Decimals. Decimals range between nought and one.
2. Percentages. Percentages, true to label, range between nought and one hundred.
3. Fractions. Where probability is based on fractions the denominator of the fraction is based on the number of possibilities.

Decimals

Decimals utilise a range between nought (0) and one (1), which is labelled a decimal measure because it entails using the decimal point. To illustrate:

1. A probability of 0 is stating that the event will certainly not occur.
2. A probability of 1 is that the event will definitely occur.
3. If the event will occur one (1) time in four (4) it has a probability of 0.25.

5. *Minister for Immigration v Guo* (1996) 144 ALR 567, 578

6. A finding of fact being ‘glaringly improbable’ is a ground on which an appellate court can overturn a finding of fact by a lower court – see *House v The King* (1936) 55 CLR 499, 505, *Devries v Australian National Railways Commission* (1993) 177 CLR 472, 479.

Percentages

Probability can be measured by percentages. To illustrate:

1. An event that will certainly occur has a probability of 100%.
2. An event that will certainly not occur has a probability of 0%.
3. An event that occurs one time in four has a probability of 25%.
4. An event that occurs two times in five has a probability of 40%.

Fractions

Probability can be expressed as a fraction. In some cases there is a practical advantage in using fractions. One example is racing. For the ordinary punter it makes sense to say, for example, that a racehorse has a one in four chance of winning. A bookmaker indicates this probability by quoting the horse at odds of three to one, that is 3-1 (against). This says that the horse has three chances of losing compared to one chance of winning.

Another example involves the probability of drawing a card of a particular type from a pack of 52. This probability is 1 in 52, that is, 1/52. This can be expressed as a percentage, ie 0.9230769, but this is a clumsy number and is also an imperfect approximation. Moreover, most of us know that a deck has 52 cards, so fractions with 52 as the base or denominator make immediate sense.

Illustration

To illustrate these methods, we can show the three means of measuring probability in a table that portrays the probability of an event that has a one in four chance of occurring:

	Will Occur	**Will Not Occur**	**Total**
Percentage	25%	75%	100%
Decimal	0.25	0.75	1
Fraction	1/4	3/4	1

Diagram 2.1 Measures of Probability

The total, 100% or 1, represents certainty. It is the sum of the probability that an event will occur or will not occur. Since an event can only occur or not occur, there is absolute certainty that there will be one or other outcome.

Use in Practice

All three measures are used in practice, and in any event, one can be converted to the other. In this regard, percentage and decimal probabilities, as the table above shows, are fundamentally the same, except that the denominator of the fraction is one (1) for decimals and one hundred (100) for percentages.

Section 3 Applying Probability

Introduction

Probability, as has been stated, accounts for uncertainty. In fact it covers uncertainty in at least three areas:

1. How Likely. It can indicate how likely is an event to occur.
1. How Often. It can indicate how often an event will occur.
1. How True. It can indicate how true is a proposition.

It is convenient to treat these areas separately since they emphasise different applications of probability.

Likelihood

Will an event happen or will it not? Probability can answer this question by stating the likelihood that a particular event will happen or has happened. For example if a person tosses a coin there is a 50% chance that the coin will come down heads and a 50% chance that it will come down tails.

Frequency

Probability may be used to state the number of expected occurrences of an event when an experiment is repeated. For example if we toss a coin 10,000 times, there will be something close to 5,000 cases of heads and 5,000 cases of tails. In medicine, for every 1,000 people who take some harmful substance, X, it may be estimated from research that 565 will become infected with disease Y.

Truth

Probability can be used to describe the degree of confidence that a proposition such as a behavioural law is true. Sometimes a behavioural law is established by an experiment using samples. The researcher systematically manipulates a variable that is labelled the independent variable. The researcher then observes and measures the response of another variable to this manipulation; this variable is called the dependent variable. These responses may be just a 'yes' or a 'no,' but they can also consist of measurable responses. In this latter case, an issue arises as to whether any changes to the dependent variable are due to variations between samples or are genuine products of changes to the independent variable.

In these circumstances researchers utilise confidence intervals to determine if results can be accepted as demonstrating the plausibility of the truth of the causal law. Confidence intervals quantify the degree of certainty with which an experimental outcome can be believed. Common levels are 90%, 95%, 99% and 99.9%. Confidence levels determine the probability that the experiment has demonstrated the causal law to be consistent with the observed data.

Relationship of Likelihood and Frequency

The relationship between these two forms, 'how often' and 'how likely' can be neatly summarised in two propositions. 'How often' can predict 'how likely'. 'How likely' can predict 'how often'.

How Likely Predicting How Often

Probability of 'how likely' can be a predictor of 'how often'. To illustrate, let us return to the example of the deck of cards. The probability of drawing any particular card (for example the Ace of Hearts) from a deck of 52 is 1/52. This probability of 1/52 has a 'physical' interpretation: it means that we would expect, on average, to draw the Ace of Hearts once in every 52 times we pulled a card from a shuffled pack. In fact, this interpretation gives us an easy bridge from 'how likely' to 'how often.'

Assume now that we keep on drawing one card from this deck until we have drawn a card in this way 52 million times. At this point, utilising the law of large numbers, we can predict that each card will have been drawn approximately [52,000,000 x (1/52)] times, that is, 1 million times. In other words, we now have a prediction of 'how often' based on 'how likely.'

This example can be generalised to postulate that 'how likely' is a good predictor of 'how often' especially in the long run. The point to using the long run is that freak results, for example turning up the same card on ten consecutive occasions, do not greatly distort the figures because of the large number of occasions on which a card has been drawn. To illustrate, assume after 52 million draws the Ace of Hearts is then drawn ten times. This will have negligible impact on the relative frequency, that is the number of occasions when the Ace of Hearts has been drawn compared to other cards, because 10 is such a small number compared to 1 million.

How Often Predicting How Likely

Probability of 'how often' is a predictor of 'how likely'. Assume that the medical data for the occurrence of a disease ('how often') is that 56% of people who have done X, for example reached age 70, have contracted disease Y. Jack is now 70. What does the data say about the likelihood that Jack has disease Y?

If we have no better information on 'how likely' than 'how often,' our best guess is to use 'how often' all the time. In the long run, that is if we do this consistently and over a reasonable number of trials, we minimise the extent of errors in our prediction more than if we had consistently used some other measure. Thus, in the absence of additional information, our best guess is that Jack's chance of having disease Y is 56%.

Section 4 Assigning Probability

Introduction

Aside from any outstanding philosophical issues with the foundations of probability, there is an issue with how one assigns an initial probability to the happening of an event or to the truth of (that is, the validity and reliability of) a causal law. Our concern with the probability of the happening of an event arises because that is the task a court has to do when it find facts.

Our concern with the probability that causal or behavioural law is true arises because a legislature enacting a statute or a court interpreting a statute or common law rule needs to know this type of probability in order to perform its task.

Happening of an Event

Several means are proposed for determining the probability that some event will happen. These are single observation, systematic observation and symmetrical evidence.

Single Observation

Individual

In the real world we all live our lives on the basis that when we observe something we know what we have observed. As the saying goes, ‘seeing is believing’ or ‘I know what I saw’. This is why one reason that a person may believe that they know something is that they have observed it. They can observe it with any of their five senses, namely sight, sound, touch, taste and hearing. The person’s account of these observations is admissible as evidence in court, provided it is relevant to the case and not forbidden by some exclusionary rule of evidence.

To illustrate observation, a person sees X and they know X for certain. X can be many things, for example the sunset this evening, a tree in the local park or the bowl of fruit on the kitchen table. In each case the person sees something, and then knows that what they have seen is true. For example, they know that something is located in a particular place or that an event has happened. In these cases a person can easily be 100% satisfied as to the truth of what they saw.

Of course it is always possible that while the person ardently believes that something is 100% true, they are wrong. And it is possible that having seen something they are not quite sure as to what really happened. So they acknowledge the fallibility of their observation and say to themselves something like this: ‘I thought I saw a rabbit dash out of the bush but I could not be sure’.

Court
Just as humans rely on the evidence of the human senses so do courts. A court receives in evidence an account of what a witness has observed. Their rationale for this is encapsulated in the saying that 'seeing is believing'. That said, while a court can believe a witness it is not obliged to do so. When the witness gives their evidence to the court the witness may believe that they have observed something correctly, that they have remembered it accurately and that they have told it truthfully. However, the court is entitled to make up its own mind on these matters. The point is that there is no guarantee that observational evidence is correct or that a court will find it so.

When courts scrutinise observational evidence to determine whether or the extent to which it is true, they resort to two means. One is the lay person's understanding of cognitive science. This directly scrutinises the evidence. For this the court asks whether the evidence of the witness is consistent with the capacity of humans to observe events, to remember them and to recount them in court. The other means of assessing truth is induction. Here the court considers whether the evidence of the witness describing events squares with the way things usually happen.

Systematic Observation

Sometimes there is empirical data on the happening of events based on serious and systematic observation. Examples are data from surveys of weather, disease, life expectancy, physiological characteristics (such as blood groups), rates of divorce and the occurrence of motor vehicle accidents. Literally, these observations say something like this: 'Of all the marriages in the years 1990-1999, 43% ended in divorce within 10 years'. This data has been gathered in a scientific method and provides some evidence for assigning probabilities to the events. The problem is that to use past data to assign probabilities in the present or future – the process of extrapolation – is not fully scientific, but it is sometimes the best means that is available.

Symmetrical Evidence

If observation is possible we can observe something and therefore know that it is true. Sometimes, however, proper observation is not possible. This happens in the case of shuffling a deck of 52 playing cards and drawing one card from the pack when the cards are face down. If a person could observe what happened to every card as it was shuffled into the pack and the pack cut, and if they could see which card, for example the 14th, was being drawn, they could tell you for sure what card it was. This, however, is not the case so there is uncertainty about our observation. We have seen the cards shuffled, but we do not fully know what has happened to each card in the process.

How do we handle this uncertainty? The key to it is that there are many ways in which the cards can be shuffled, cut and drawn. Since an ordinary human (as distinct from a card sharp) cannot control the ways in which they shuffle, cut and draw a deck of cards and since there are no special factors involved which favour one way of doing these tasks over another, there is a reasonable assumption that each way is equally likely. Put another way, evidence for the probability of each event or possible outcome is symmetrically balanced so that each outcome is equally likely. Thus, the probability of drawing any particular card from a deck of 52 is one in 52, that is, 1/52.

This form of probability was the first to excite curiosity, which it did in the seventeenth century. Aristocratic speculation about games of chance such as cards was the initial motivating force. This led to more serious considerations and much of the early work was done by correspondence between two major mathematicians, Blaise Pascal (1623-1662) and Pierre de Fermat (1601-1665).

This version of probability is variously called *a priori* probability, classical probability or the frequentist view. It is used in 'cases where there are a finite number of possible outcomes each of which is assumed to be equally probable'.[7]

Gambling with cards and dice, which prompted initial interest in probability, also furnishes simple and excellent illustrations of probability at work. With a deck of 52 cards, the probability of drawing a specific card, for example the Queen of Hearts is 1/52. With a six-sided die, the probability of throwing a specific number, for example a '3,' is 1/6.

Truth of a Causal Law

A legislature or court may invoke a causal or behavioural law when making or interpreting law. A court may invoke such a law when trying to decide the facts of a case. In the ideal case such laws would be established to the point of certainty. Unfortunately behavioural science does not normally yield absolute certainty about the truth of a behavioural law. Instead, to put it broadly, it can indicate the probability that the law is true based on the experimental methods used.

In areas where there is no scientific research at all or insufficient research, the existence of truth of causal laws is much more speculative. Ideally this uncertainty would be captured and given a numerical value. Again, behavioural science generally cannot do this. One can merely make or find a considered assessment of why a particular behavioural law may or may not be true.

7. Robertson and Vignaux (1993) p 459

Section 5 Deriving Probability

Introduction

Previous discussion has explained methods of estimating the probability that something is true. Once we have estimated a probability, by whatever method, there are a number of rules of probability that we can apply to the figure so obtained to derive, calculate or compute further information about the probability of events. Although the logic behind these computations is sound (being based on deduction), the results are still no better than the estimates of probability that we use. So, if an estimate is unsound, the result of a computation based on this estimate will also be unsound. Indeed, the computation may magnify the extent or effect of the error.

Complementarity Rule

In measuring the probability of an event's happening we are also measuring the probability that it will not happen. The probability of an event's not happening is easily derived from the probability of its happening because an event can only happen or not happen. Therefore there is absolute certainty, that is, a probability of 100%, that one or other outcome will occur. So, if the probability of an event's happening is 0.3 the probability of its not happening is 1 minus 0.3, namely 0.7. This is the complementarity rule. In formal terms, if A is an event with probability P(A), then the probability that A will not occur is [1 – P(A)].

Multiplication Rule

Introduction

Sometimes we are interested in probability where there are two or more independent events. Events A and B are independent when the happening of one of them has no bearing on the happening of the other. To illustrate with proof of facts, facts are independent for the purposes of proof when the truth of Fact A has no bearing on the truth of Fact B. An illustration from gambling is two successive throws of a die. No matter what the result of the first throw, it has no bearing on the second.

If we wish to calculate the probability that each of two or more independent events will occur, we use the product or multiplication rule. This involves taking the probability of each individual event and multiplying them together. We will deal with this in stages, stating the main rule, then stating some subsidiary rules.

Main Rule

There are three aspects to the main version of the product rule – derivation, statement and illustration.

Derivation of Rule

Before we formally state this rule we will show how to derive it by using the game of two-up. This involves throwing two coins into the air together, with a spinning action, and betting on the outcome, that is, the sides of the coins that face upwards. (In many countries, the sides of a coin are labelled 'heads' and 'tails' and we use this terminology here.) If we ignore the very remote chance that a coin can land on its edge there are two possibilities, which are taken to be equally likely, heads with a probability of 50% and tails with a probability of 50%.

Now consider throwing the two coins. There are four outcomes, which are all equally likely. We can set this out in the following table:

First Coin	**Second Coin**	**Probability**
Heads	Heads	25%
Tails	Tails	25%
Heads	Tails	25%
Tails	Heads	25%
		100%

Diagram 5.1 Derivation of the Multiplication Rule

How did we calculate the probability for each combination? We reasoned in the following way, using the probability of two tails as an example:

1. When the first coin is thrown the chances are equal that it will come down heads or tails, for example the probability of tails is 50%.
2. Assume now that the first coin is tails. On 50% of occasions the second coin is also tails. Hence the probability of two tails is 50% of 50%, namely 25%.

Statement of Rule

Having shown how the product or multiplication rule is derived, we can now formally state the rule. When two or more events are independent the probability that all of the events will occur is equal to the multiple of the probability of each event. Thus if the events (designating an event by E) and their probabilities are E_1 (A%), E_2 (B%) … E_n (N%), the probability of all of them happening (E_1, E_2 and E_n) is A% x B% x … N%.

In formal terms, the rule can be stated in the following way. If A and B are independent events, the probability that both A and B occur simultaneously is P(A) x P(B).

Illustration of Rule

One area in which the product rule is used is the process of finding facts in a court case. Therefore we will illustrate the rule by showing how it is used to do this.

For this illustration assume that there is a case with the following characteristics:
1. The case involves a cause of action that has four elements, Elements 1-4.
2. The facts that satisfy these elements are Facts 1-4. Thus Fact 1 satisfies Element 1, Fact 2 satisfies Element 2, Fact 3 satisfies Element 3 and Fact 4 satisfies Element 4.
3. We assume that the facts are independent.
4. We assume that this is a civil action where the standard of proof is proof on the balance of probabilities, that is, proof to a probability of 51%. Although we assume that this is a civil action the reasoning applies equally to a criminal action, although we have to adjust the standard of proof to the criminal standard, namely proof beyond all reasonable doubt.[8]

When the court hears the case it finds the probability of truth for each fact. These findings are set out in the following table:

Fact	**True**	**False**	**Total**
Fact 1	60%	40%	100%
Fact 2	80%	20%	100%
Fact 3	75%	25%	100%
Fact 4	60%	40%	100%

Diagram 5.2 Independent Facts

Having found the probability that each specific fact is true or (false), the court now has to proceed to its main task in finding facts, to determine the probability that all of the facts are true (because by doing this it can determine if the plaintiff has sufficiently proved their case). Since the facts are independent we use the product rule. To calculate this overall probability we multiply together the probability for each specific fact, Facts 1-4 in the illustration. This gives us the probability that all of the facts are true:

Probability	=	P (Fact 1) x P (Fact 2) x P (Fact 3) x P (Fact 4)
	=	(0.60) x (0.80) x (0.75) x (0.60)
	=	0.216
	=	21.6%

Diagram 5.3 Illustration of the Multiplication Rule

This shows us that the probability that each of the four facts in the plaintiff's case is absolutely true is 21.6%. To round off the illustration let us explain how the legal system should use this information to work out who wins the case and who loses. In a civil case, as we noted above, a plaintiff must prove their case on

8. To avoid confusion, it is worth noting that proof 'beyond reasonable' doubt can in principle be stated as a percentage, but there is no formal designation of what that percentage is. The author suggests that it should be around 99%.

the balance of probabilities, ie to a probability of 51%. Here the probability of the plaintiff's case is only 21.6%. This is less than the 51%, which the plaintiff requires, so the plaintiff fails to prove her case and loses.

Subsidiary Rules

In the illustration above, the probabilities for Facts 1-4 were respectively 60%, 80%, 75% and 60%. The probability that all four facts, Facts 1-4, were absolutely true is 21.6%. We can use this result to illustrate two subsidiary rules, which are of general application but which also help us to understand more about how to use probability in fact-finding.

Rule 1: Result Lower Than Lowest Individual Score

In using the product rule we multiply a number of specific probabilities together. These are essentially fractions. Because of this, by multiplying them we obtain a figure that can never be higher than the lowest individual measure. Where X is a fraction, half of X is always less than X. (This point, of course, applies to any fraction, not just a half.)

To illustrate from the example above, the overall probability that all facts are true is 21.6%, which is still lower, indeed considerably lower, than the lowest individual figure of 60% (which is the probability for both Fact 1 and Fact 4).

Rule 2: Each Additional Fact Lowers Probability

Each additional fact introduces an additional probability. This additional probability lowers the value of the final result unless its value is 100%. To illustrate, start with Fact 1 then add the remaining facts one at a time and see what happens. We can set out the results as follows:

Facts	*Probability:*	*Probability:*	*Probability:*
	Additional Fact	*Overall*	*Reduction in Overall Probability*
Fact 1	60%	60%	Not applicable
Fact 1-2	80%	48%	20%
Fact 1-3	75%	36%	25%
Fact 1-4	60%	21.6%	40%

Diagram 5.4 Independent Facts

We will illustrate the necessary calculations using Facts 1-3:

1. Overall probability is given by multiplying together the probability for the individual facts. This is (60 x 80 x 75)%, namely 36%.
2. Reduction. There is a reduction from 48% for Facts 1-2, to 36% for Facts 1-3. This is a [(48-36)/48]% reduction, namely 25%.

Explanation of a Paradox

Looking at the individual probabilities on the surface (60%, 80%, 75%, 60%) may suggest that the plaintiff has more than satisfied the standard of proof of 51%. This, however, is not the case when the facts are independent (as is the assumption here) because the probability that all four facts are true is only 21.6%. As Rule 1 says, the final result when these individual probabilities are multiplied together is lower than, often considerably lower than, the lowest individual measure.

To illustrate this further, we will construct a hypothetical case using the lowest individual probability in the example above, namely 60%. Assume that there is just one other fact and the probability that this fact is true is 80%. In this case the overall probability is (60 x 80)% namely 48%, which means that the plaintiff fails to make the necessary standard of 51%. This is an extreme case because there are only two facts and the second had a high probability. This shows how easy it is to get a figure such as 60% 'down' when the probabilities are multiplied together.

Sum Rule

Assume that there are two events, A and B. The sum rule determines either the probability that event A or event B occurs, or the probability that both occur. There are two possibilities – that the two events are mutually exclusive (disjoint) or not mutually exclusive (conjoint).

Mutually Exclusive

If events A and B are mutually exclusive then:

P(A or B)	=	P(A) + P(B)

As an illustration, assume that there is a group of 150 students where 30 are freshmen and 60 are sophomores. Assume that we need to determine the probability that a student picked from this group at random is either a freshman or sophomore. In this case the individual probabilities are:

P (freshman)	=	30/150
P (sophomore)	=	60/150
P(freshman or sophomore)	=	30/150 + 60/150
	=	90/150
	=	60%

In other words 90 of the 150, students that is 60%, are freshmen or sophomores.

Not Mutually Exclusive

If events A and B are not mutually exclusive then:

P(A or B)	=	P(A) + P(B) – P(A and B)

As an illustration, assume that there is a group of 150 students where 40 are juniors, 50 are female, and 30 are both female and juniors. The task is to find the probability that a student picked from this group at random is either a junior or female.

Here the individual probabilities are as follows:

P(junior)	=	40/150
P(female)	=	50/150
P(junior and female)	=	30/150
P(junior or female)	=	40/150 + 50/150 – 30/150
	=	60/150
	=	40%

This makes sense since 90 of the 150 students are juniors or female. The point to subtracting the percentage of students who are junior and female is to avoid double counting. When we add 40 juniors to 50 females and get a total of 90, we have over-counted. The 30 female juniors were counted twice; 90 minus 30 is 60, so there are 60 students who are juniors or female.

Bayes Theorem

Bayes Theorem was formulated by the Rev Thomas Bayes (c 1702-1761). Thomas Bayes was born in London in about 1702. He became a Presbyterian minister. As far as is known, during his lifetime he published two works: *Divine Benevolence, or an Attempt to Prove That the Principal End of the Divine Providence and Government is the Happiness of His Creatures* (1731), and *An Introduction to the Doctrine of Fluxions, and a Defence of the Mathematicians Against the Objections of the Author of the Analyst* (published anonymously in 1736), in which he defended the logical foundation of Isaac Newton's calculus against the criticism of George Berkeley, author of *The Analyst*. Bayes was elected as a Fellow of the Royal Society in 1742 possibly on the strength of the *Introduction to the Doctrine of Fluxions*. Bayes died in Tunbridge Wells, Kent in 1761. He is buried in Bunhill Fields Cemetery in London where many Nonconformists are buried. So in death, as in life, he was separated from the Church of England.

The fame of Thomas Bayes rests on a posthumously published masterwork, *An Essay Toward Solving a Problem in the Doctrine of Chances.*[9] In this he enunciated what is now known as Bayes Theorem. Bayes Theorem can be deployed in abductive reasoning in trying to ascertain the best explanation for something. According to one view it can be deployed in the process of fact-finding where there is a lack of direct evidence for vital facts so that they fall to be ascertained by inference.

Operation of Theorem

To explain Bayes Theorem let us work with a simple example.[10] Assume that in a country the probability that any person has 'lurgi' (a fictitious disease) is 5 percent. Thus if:

A_1 refers to the event of having the disease

A_2 refers to the event of not having the disease

then

$P(A_1) = 0.5$

$P(A_2) = 0.95$

In the context of Bayes Theorem these probabilities are called prior probabilities because they are the existing data before additional information is discovered about these probabilities. Thus if we select a person at random the best estimate we now have is that the person has a 0.05 probability of having lurgi.

Assume now that someone develops a diagnostic test for lurgi, but it is not completely accurate. Let B denote the event that the test shows the disease is present. Assume that if the person has the disease that the evidence clearly shows that the probability of the test indicating the presence of the disease is 0.90. Assume that if the person actually does not have the disease but the test indicates that the disease is present is 0.15. Using the standard notation these probabilities can be written as follows:

$P(B|A_1) = 0.90$

$P(B|A_2) = 0.15$

The point to Bayes Theorem is that it enables us to upgrade the relevant probabilities. To illustrate, assume that a person X is selected at random and the diagnostic test is performed. The test indicates that the disease is present. What is the probability that the person actually has the disease? In symbolic form, we want to know $P(A_1|B)$. This is called a posterior or revised probability, because it is revised following the discovery of additional information since the original

9. Bayes (1764)

10. This example in the text of Bayes Theorem is taken from Robert D Mason, Douglas A Lind *Statistical Techniques in Business and Economics* (1993) 8th ed Sydney: Irwin, pp 185-186.

prior probability was determined. In this case where there are only two possible events, having the disease (A_1) or not having the disease (A_2), the probability is given by the formula:

$P(A_1 \mid B)$	=	$\dfrac{P(A_1) \text{ x } P(B \mid A_1)}{P(A_1) \text{ x } P(B \mid A_1) + P(A_2) \text{ x } P(B \mid A_2) + \ldots + P(A_n) \text{ x } P(B \mid A_n)}$

Diagram 5.5 Formula for Bayes Theorem

In this illustration there were only two events (having the disease (A_1) or not having the disease (A_2)) Obviously, it is possible that there can be more than two possible events. In this case the denominator needs to be adjusted. Calculations for the answer to the question are set out in the following table:

Events	**Prior Probability**	**Conditional Probability**	**Joint Probability**	**Posterior Probability**
A_1	$P(A_1)$	$P(B \mid A_1)$	$P(A_1 \text{ and } B)$	$P(A_1 \mid B)$
Disease	0.05	0.90	0.0450	0.0450/.1875 = 0.24
No Disease	0.95	0.15	0.3425	0.1425/.1875 = 0.76
			P(B) = 0.1875	Total 1.00

Diagram 5.6 Calculation of Bayesian Probabilities

The final column Posterior Probability gives the answer. The probability that X has the disease is 0.24 or 24%, while the probability that X does not have the disease is 0.76 or 76%.

Operation of Theorem

This example above gives something of the flavour of Bayes' Theorem. It permits mathematical calculation of the likelihood of one event given another event. This is how Bayes' Theorem enters fact-finding. It is relevant when facts are based only on indirect or circumstantial evidence. For example, given that John is a bank robber, has been in the area of a bank robbery and owns a pistol looking like the one that the masked robber used to threaten the bank teller, what is the probability that John was the robber on this occasion?

Bayes' Theorem enables us to calculate this probability. However, while the mathematical calculations involved are precise, they are based on subjective probabilities fed into the relevant equations. Results from the equations are flawed in that the data that they use is flawed, being mere estimates. Nevertheless, there is a benefit in that the equations show how the specific probabilities can be combined to give the overall probability of guilt or innocence.

Expected Value

A bird in the hand is worth two in the bush.[11]

To explain how expected value works, assume that a firm is considering an expansion of its market. It has two options and manages to calculate that the potential profit for Option 1 is $250,000 and for Option 2 is $400,000. However, there is no certainty that this profit will eventuate. In fact, on the information that is available to the firm, there is only a 75% chance of the $250,000 and a 40% chance of the $400,000.

What this firm now needs is a mechanism for taking this uncertainty into account as it faces a choice to expand or not expand its market. Indeed such a mechanism is needed in a range of legal and non-legal activities which include whether to pass one statute or another, to litigate or not litigate, to invest in shares or bonds, or to stay where we are rather than take a new job. Fortunately there is such a mechanism, which is known as expected value. It enables us to adjust the return for each possibility by factoring in the uncertainty. We do this by measuring the return as a probable or expected return rather than by reference simply to the dollar value of the return. Expected value of an outcome is the probability of the outcome multiplied by the net value of the outcome.

This can be illustrated from the example above where the expected values for Option 1 and Option 2 are respectively ($250,000 x 75%) and ($400,000 x 40%). To choose between these options it is necessary to determine the expected values of Options 1 and 2. The diagram below sets these out:

Option	*Profit*	*Probability*	*Expected Value*
Option 1	250,000	75%	187,500
Option2	400,000	40%	160,000

Diagram 5.7 Expected Values

At the diagram reveals, the expected value of Option 1 is $187,500 and of Option 2 $160,000. This indicates that Option 1 is a better investment than Option 2. This example shows how expected value produces a measure of the return for each outcome, which enables us to compare the returns. In turn this enables us to make a decision because we take the outcome with the best return. Thus, expected value is a method, involving a calculation, which is used to take into account uncertainty. By their nature, decisions take their effect in the future and the future is inherently uncertain.

This is why expected value is used in business to make investment decisions that are needed when a new project, be it large or small, is contemplated. In law

11. This is a proverb.

expected value, or at least the reasoning process that underlies it, can be used for making and interpreting law. Those making and interpreting law can use expected value to factor in the possibility that the predicted costs and benefits of a law or an interpretation of a law may not come about.

Expected value can also be deployed in making the decision whether to litigate. Further, on one analysis the tort of negligence incorporates expected value.[12]

While expected value is an extremely useful tool in principle, it has two major limitations. First, the probability that proposed action will incur a cost or return a benefit can rarely be known precisely. Second, it can be difficult to compute net benefit, because costs and benefits are not always measurable.[13]

Section 6 Fallacies in Probability

Introduction

Careless thinking about probability can lead to errors. These errors involve arguing from one established probability to a second probability. These fallacies are of extreme concern when the second probability involves guilt or innocence in a criminal trial. Two common forms of this fallacious reasoning have been identified and labelled the prosecutor's fallacy and the defendant's fallacy.[14]

Prosecutor's Fallacy

An illustrative version of the prosecutor's fallacy is as follows. Assume that there is a one in two million chance of a match of evidence at the crime scene if a defendant is innocent. Based on this, the prosecutor argues that there is a one in two million chance of innocence and consequently a 1,999,999 in two million (99.99995%) chance of guilt, which is well past the point of proof beyond all reasonable doubt.

But assume further that the crime took place in a city of ten million people, any of whom might be the perpetrator. If the authorities tested each person one would expect five matches with the evidence. On this basis the possibility of a match taken on its own there is only a one in five (20%) chance of guilt, which is way below reasonable doubt. In other words, when one views this evidence properly it is apparent that the probability of guilt is small.

---

12. Computer programs, called Bayesian Belief Networks (BBN), have been designed to handle the complex interaction of the probabilities of multiple pieces of evidence. For an illustration see Huygen (2002).
13. Christopher Enright *Legal Reasoning* (2015) Chapter 12 Measurement of Net Benefit
14. These were identified and labelled by Thompson and Schumann (1987).

Defendant's Fallacy

The defendant's fallacy can be illustrated from the example used for the prosecutor's fallacy. There it was concluded that taking the evidence of a match on its own, there is only a one in five (20%) chance of guilt. It is, however, a fallacy to use this figure of 20% when there is other circumstantial evidence pointing to the guilt of the defendant. To state the obvious, each piece of circumstantial evidence for the defendant's guilt increases the probability of guilt.

Sally Clark Case

In the United Kingdom there was a famous conviction of Sally Clark, based on the prosecutor's fallacy. In 1998 Sally Clark was accused of killing her first child, Christopher, at 11 weeks of age and then conceiving a second child, Harry, and killing him at 8 weeks. For the defence it was argued that both deaths were cases of sudden infant death syndrome (SIDS). An expert witness testified that the chance of two deaths in the same family from SIDS was about 1 in 73 million. Sally Clark was convicted of murder of these two children in 1999.

Following her conviction the Royal Statistical Society issued a press release pointing out two errors of reasoning.[15] First, they criticised the figure of 1 in 73 million for the frequency of two cases of SIDS in such a family. The Society said: 'This approach is, in general, statistically invalid. It would only be valid if SIDS cases arose independently within families, an assumption that would need to be justified empirically. Not only was no such empirical justification provided in the case, but there are very strong *a priori* reasons for supposing that the assumption will be false. There may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes much more likely'.

Second, they pointed out that 'figures such as the 1 in 73 million are very easily misinterpreted. Some press reports at the time stated that this was the chance that the deaths of Sally Clark's two children were accidental. This (mis-) interpretation is a serious error of logic known as the Prosecutor's Fallacy. The jury needs to weigh up two competing explanations for the babies' deaths: SIDS or murder. Two deaths by SIDS or two murders are each quite unlikely, but one has apparently happened in this case. What matters is the relative likelihood of the deaths under each explanation, not just how unlikely they are under one explanation (in this case SIDS, according to the evidence as presented)'.

Eventually the law caught up with statistics, and Sally Clark was freed when the Court of Appeal quashed her conviction in January 2003. Further medical

15. Royal Statistical Society (2001)

analysis of the case by experts highlighted both the difficulty of deriving firm conclusions from post mortem examinations on infants and the highly equivocal or fragile nature of the pathological evidence that was used against Sally Clark. Moreover, later consideration of the medical evidence indicated that the body of the second child to die, Harry, had the infection staphylococcal aureus, raising a strong possibility that he died from staphylococcal sepsis.[16] After winning the case and being released a journalist said to Sally Clark: ‘So, you finally won’ to which Sally Clark said: ‘There are no winners here’. Sally Clark died aged 42 years on 16 March 2007. Some of her friends believe that Sally died of a broken heart.

References

Bayes, Thomas (1731) *Divine Benevolence, or an Attempt to Prove That the Principal End of the Divine Providence and Government is the Happiness of His Creatures*

Bayes, Thomas (1736) *Introduction to the Doctrine of Fluxions, and a Defence of the Mathematicians Against the Objections of the Author of the Analyst*

Bayes, Thomas (1764) *An Essay Toward Solving a Problem in the Doctrine of Chances*

Byard, Roger W (2004) ‘Unexpected infant death: lessons from the Sally Clark case’ Medical Journal of Australia 181 (1): 52-54

Enright, Christopher (2015) *Legal Reasoning* Sinch: Newcastle

Hájek, Alan (2003) ‘Interpretations of Probability’ See /archives- http://-platostanford-edu /sum2003/entries/probability-interpret/

Mason, Robert D; Lind, Douglas A (1993) *Statistical Techniques in Business and Economics* 8th ed Sydney: Irwin

Robertson, B and Vignaux, GA (1991) ‘Inferring Beyond Reasonable Doubt’ 11 *Oxford Journal of Legal Studies* 431

Royal Statistical Society (2001) ‘News Release’ 23 October 2001, see http://wwwrssorg-uk/docs/RoyalStatisticalSocietydoc

Thompson, EL; Shumann, EL (1987) ‘Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy’ *Law and Human Behavior* (Springer) II (3): 167

16. Byard (2004)