First-year Statistics for Psychology Students through Worked Examples

2. Probability and Bayes' Theorem

Charles McCreery, D.Phil.

Formerly Lecturer in Experimental Psychology
Magdalen College
Oxford

Oxford Forum

Psychological Paper No. 2007-2

Acknowledgements

I am grateful to the following for comments and guidance at various points in the evolution of this tutorial: Dr Fabian Wadel, Dr Paul Griffiths, and Professor David Popplewell.

I am also indebted to Andrew Legge for help with the formatting of mathematical formulae and symbols.

Most recently, I have become much indebted to Dr Ed Knorr, who very kindly read through the complete typescript and made numerous suggestions and corrections, both large and small.

Any remaining errors or omissions are my responsibility. I would be pleased to receive information from anyone who spots any error, mathematical or otherwise. I can be contacted via e-mail at:

charles.mccreery@oxford-forum.org

I should also be pleased to hear from anyone who finds this tutorial helpful, either for themselves or for their students.

Charles McCreery

General Introduction[1]

There are usually three complementary methods for mastering any new intellectual or artistic task; these are, in ascending order of importance:

- reading books about it
- observing how other people do it
- actually doing it oneself

These tutorials focus on the second of these methods. They are based on handouts that I developed when teaching first-year psychology students at Magdalen College, Oxford. The core of each tutorial is a worked example from an Oxford University Prelims Statistics examination paper. I have therefore placed this section in prime position; however, in teaching the order of events was different, and more nearly corresponded to the three-fold hierarchy of methods given above:

1. Students were invited to read one of the chapters on the Recommended Reading list, given at the end of each tutorial. They were also expected to attend a lecture on the topic in question at the Department of Experimental Psychology.

2. Students would attend a tutorial, in which we would go through the worked example shown here. They would take away the handouts printed as Appendices at the end of each chapter, which were designed to give structure to the topic and help them when doing an example on their own.

3. They would be given another previous examination question to take away and do in their own time, which would be handed in later for marking.

I am strongly in favour of detailed worked examples; following one is the next best thing to attempting a question oneself. Even better than either method is doing a statistical test on data which one has collected oneself, and

[1] This is a general introduction to a series of three tutorials available here: http://www.celiagreen.com/charlesmccreery.html

which therefore has some personal significance to one, but that is not usually practicable in a first-year course.

I list three books in the General Bibliography at the end of this tutorial which give worked examples. One of these is Spiegel (1992), in which each chapter has numerous ‘solved problems’ on the topic in question. These worked problems occupy more than half of each chapter. However, the solutions to the individual problems are not as detailed and discursive as the ones I give here.

Another book which is based on worked examples on each of the topics covered is Greene and D’Oliveira (1982), also listed in the General Bibliography. Their examples are as detailed as those I give here. However, they do not cover probability and Bayes’ theorem or Analysis of Variance.

Finally, I strongly recommend the *Introductory Statistics Guide* by Marija Norusis, designed to accompany the statistical package *SPSS-X*, and based on worked examples throughout. Even if the student does not have access to a computer with the *SPSS-X* package on it, this instruction manual contains excellent expositions of all the basic statistical concepts dealt with in my own examples.

Probability and Bayes' Theorem

Contents

1. Introduction

A word of encouragement

Probability theory is a subject which is well-known for producing what seem at first sight to be counter-intuitive results. In addition, Bayes' theorem may seem difficult to grasp at first, because it seems to involve us in 'thinking backwards' in a way we are not used to. However, like most ideas, it is actually quite simple, and indeed obvious, once grasped. The problem is that having grasped it once may not guarantee that the understanding of it sticks – it may be necessary to think it through a second and even a third time. To this end some people may find the visual method of representing probabilities, via 'probability trees', helpful.

Concerning the layout of this tutorial

Although section 2.2 is called 'Answer', it is not intended to be a model answer, such as one might give in an examination. Even without the sections of commentary which I have hived off within square brackets for the sake of clarity, section 2.2 contains much more information than one would need to give in an examination, since I am attempting to explain what I am doing as I go along.

2. A worked examination question

2.1 The question[2]

M&M's and Smarties[3] are two different brands of small milk chocolates in a crisp coloured shell. Each item of confectionery is about the same size and each brand comes in a mixture of colours.

A large bowl contains a mixture of the two brands in the ratio of five M&M's to four Smarties in just four colours - red, yellow, orange and green.

The proportions of the M&M's which are red, yellow, orange and green are 0.3, 0.4, 0.1 and 0.2 respectively, while the equivalent proportions for Smarties are 0.25, 0.2, 0.3 and 0.25.

A sweet is chosen at random from the bowl. What is the probability that it is (i) a green Smartie; (ii) green; (iii) a Smartie if it is green?

The shades of green used by the two brands are very different and can be readily identified, but it is not possible to differentiate the two brands from the other colours. A single sweet is drawn at random from the container and a statistician is shown its colour. If it is green, the statistician will identify its brand correctly, but if it is not green, the statistician will toss a fair coin to decide which brand she thinks it is. What is the probability that she correctly identifies the brand?

If four sweets are chosen from this very large pool of sweets, what is the probability that three will be of one brand and the other sweet of the other brand?

---

[2] The question is taken from the Prelims Statistics paper for first-year psychology students at Oxford University, Hilary Term, 1999.

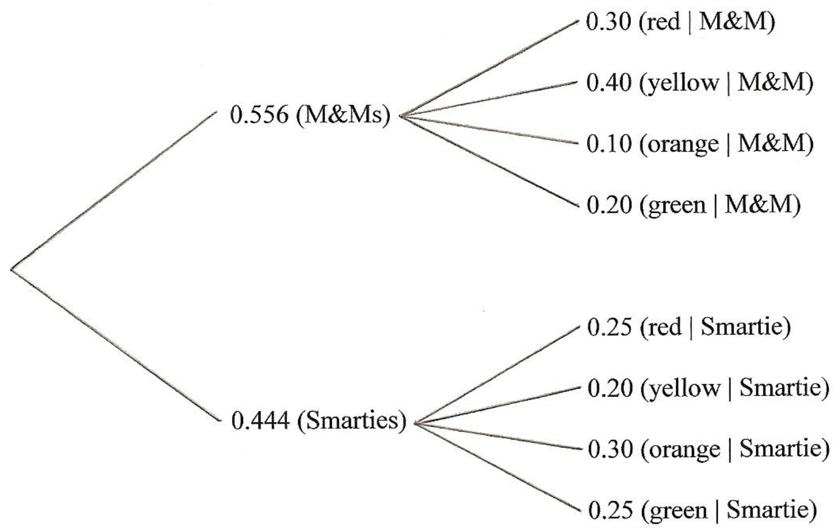
[3] 'M&M's' and 'Smarties' are registered trade names.

2.2 The answer

2.2.1 The data, and probability tree

Out of every nine items in the bowl five are M&M's and four are Smarties.
So the probability of picking an M&M's at random is 5/9 = 0.556.
The probability of picking a Smartie is 4/9 = 0.444.

Probability Tree:



2.2.2 Probability of randomly choosing a green Smartie

[p = probability
| = given that
x = multiplied by]

p (Smartie & green)
= p(Smartie) x p(green | Smartie)
= 0.444 x 0.25
= 0.111

[Illustrates the mnemonic: 'multiply for AND'. See paragraph 3.3 below.]

2.2.3 Probability of randomly choosing green

There are two ways of getting a green sweet: one can choose **EITHER** a green M&M **OR** a green Smartie, but not both at once—the possibilities are mutually exclusive, or 'disjunctive'. So this is a case of **addition** of probabilities.

[Remember the mnemonic: 'Add for OR'.]

p(Smartie and green) = 0.111 (from the first part of the question, above).
p(M&M and green) = p(M&M) x p(green | M&M)
= (0.556 x 0.20)
= 0.111

p(green) = p(Smartie and green) + p(M&M and green)
= 0.111 + 0.111
= 0.222

2.2.4 Probability of Smartie given green

[This is a Bayes-type question, since we are asked to find the probability of one of a number of possible antecedents to a given consequent.]

There are two ways of being green (as we saw in the previous part of the question): being an M&M and green, and being a Smartie and green. The probability that a given green sweet is a Smartie is therefore the probability of being a green Smartie expressed as a proportion of all the ways of being green; i.e., the probability of being a green Smartie divided by the probability of being green. More formally:

p(Smartie | green) = p(green | Smartie) / (p(green | Smartie) + p(green | M&M))

= 0.111 [from first part of question] / 0.222 [from second part]
= 0.5

[Notice that this result makes intuitive sense if you look at the probability tree above. The proportion of M&M's to Smarties in the bowl is very roughly equal; the proportion of greens within the M&M group is very roughly equal to the proportion of greens within the Smarties group; and the disproportions in the two cases work in opposite directions to each other (i.e., more M&M's than Smarties overall, but more greens among the Smarties than among the M&M's). So it makes sense that the resulting probability (Smartie given green) is exactly half.

General point: the answer to a Bayes-type question may always be thought of as a fraction, in which the numerator (on the top) is the probability of the outcome arriving via the particular route you are interested in, while the denominator (on the bottom) consists of that same probability plus the probabilities of that same outcome arriving by all the other possible routes.

N.B. The probability questions in this paper are often progressive in the way this one is: i.e., successive parts of the question often make use of the results of previous parts, which can save you work if you notice it.]

2.2.5 Penultimate part of question

A single sweet is taken at random.

Probability it is green = 0.222 (from section 2.2.3). This means that fraction 0.222 of the time, we can correctly identify the sweet *without* having to flip a coin.

Probability it is not green = 1 − 0.222
[because the probabilities must sum to 1]
= 0.778.

Probability of guessing whether a non-green sweet is an M&M or a Smartie correctly by chance (e.g., by tossing a coin) = 0.5.

Therefore the probability of guessing a not-green sweet correctly
= 0.778 x 0.5
= 0.389

[Because we have just established above that 77.8% of the sweets are not-green, and that every time she has to guess a not-green sweet she is right half the time.]

0.222 [probability of correctly guessing green] + 0.389 [probability of correctly guessing not-green] = 0.611

So, the overall probability of correctly guessing the brand is 0.611.

[Here we have applied the additive rule, ‘Add for OR’. The two possibilities on each trial - having to guess a green sweet, or having to guess a not-green sweet – are disjunctive, or mutually exclusive. So we have to add the probabilities of the two possible results.

One way of thinking of the logic of this part of the question is as follows: imagine the person choosing a succession of 100 sweets. 22 will be green and will therefore be correctly identified. 78 will be other colours, so only half of them will be correctly identified: 78/2 = 39. Thus 22 + 39 = 61. So 61 out of the hundred sweets will be correctly identified. 61/100 = 0.61.

In general it is often helpful in probability questions to imagine the probabilities as percentages, i.e., as actual numbers of people/things out of a population of 100. This is equivalent to thinking of probabilities as areas, such as subsections of a pie-chart, which can also be helpful.]

2.2.6 Final part: probability of choosing three of one brand and one of the other

There are 8 ways of choosing 3 sweets of one brand and 1 of the other, which may be represented as follows:

MMMS
MMSM
MSMM
SMMM

SSSM
SSMS
SMSS
MSSS

Note that you are not likely to encounter each arrangement equally by chance.

The first four possible arrangements each have probability 0.555^3 x 0.444 = 0.0759, because this reflects the proportion of M&M's (0.555) and Smarties (0.444) in the overall collection of sweets.

Similarly, the second four each have probability 0.444^3 x 0.555 = 0.0486.

4 x 0.0759 = 0.304
4 x 0.0486 = 0.194
0.304 + 0.194 = 0.498

[Note that this last part of the question refers to a 'very large' pool of sweets. This is to allow you to assume that, although this is not a case of random selection 'with replacement', the pool is sufficiently large for the picking of four sweets from it not to affect the ratio of M&M's to Smarties in what remains.]

3. Recommended reading

Hoel, Paul G. (1976). *Elementary Statistics* (4th edition). New York: Wiley. Chapter 3.

> Covers the basic ideas of probability, the addition and multiplication rules, probability trees, and Bayes' theorem.

Hays, William L. (1994). *Statistics* (5th edition). Orlando, Florida: Harcourt Brace. Chapter 1.

> Also covers the basic ideas of probability, the addition and multiplication rules, and Bayes' theorem, but does not introduce probability trees. Has a fuller discussion than Hoel of the concept of conditional probability.

Appendix

Probability and Bayes' theorem: summary of some key points

▪ Some key concepts

- Independence (of two or more events)

  Criterion: "If Al occurs, does that change the chances that A2 will occur from what they would be if Al were completely ignored?" (Hoel, 1976, p.55)

- Mutual exclusion (of two or more events)

  Criterion: "If Al occurs, does that make it impossible for A2 to occur?" (Ibid.)

- Conditional probability

  The probability of B given A, represented as $p(B \mid A)$.

  "Most of the information we deal with in everyday life as the basis for the choices we must make has a conditional character [...] The very best information we have to go on is usually no more than a probability. These probabilities are conditional, because virtually all our information is of an 'if-then' character." (Hays, 1994, p. 45)

▪ The Additive Rule

When Al and A2 are mutually exclusive events,

$$p\{A1 \text{ or } A2\} = p\{A1\} + p\{A2\}.$$

"Given a set of mutually exclusive events, the occurrence of one event or another is equal to the sum of their separate probabilities." (Howell, 1997, p.112)

- **The Multiplicative Rule**

"The probability of the joint occurrence of two or more independent events is the product of their individual probabilities." (Hoel, op.cit., p.113)

Mnemonic: Add for 'OR',
Multiply for 'AND'

- **Bayes' Theorem**

The theorem applies when we are 'working backwards' from a known outcome to the probability of one of several possible antecedent events.

General formulation:

p(A|B) = p(B|A) x p(A) / p(B)

Where p = probability
| = given that
x = multiplied by
and B is a consequent of some antecedent A

E.g., A is one of the two possibilities on the left of the probability tree in section 2.2.1 above, and B is one of the four possibilities branching out from A, on the right of the tree.

General Bibliography

Textbooks of the kind listed below are usually updated every few years. If the reader finds there is an edition later than the one listed here, he or she is recommended to buy the latest version.

Greene, Judith and D'Oliveira, Manuela (1982). *Learning to Use Statistical Tests in Psychology*. Milton Keynes: Open University Press.

Hays, William L. (1994). *Statistics* (5th edition). Orlando, Florida: Harcourt Brace.

Hoel, Paul G. (1976). *Elementary Statistics* (4th edition). New York: Wiley.

Howell, David C. (1997). *Statistical Methods for Psychology* (4th edition). London: Duxbury Press.

Norusis, Marija J. (1988). *SPSS-X Introductory Statistics Guide, for SPSS-X Release 3*. Chicago, Illinois: SPSS Inc.

Spiegel, Murray R. (1992). *Schaum's Outline of Theory and Problems of Statistics* (2nd edition). New York: McGraw-Hill.

Tabachnick, Barbara G. and Fidell, Linda S. (1983). *Using Multivariate Statistics*. London: Pearson Education Ltd.

Charles McCreery is a Research Director of Oxford Forum, an independent association of academics, set up to research and publish in currently neglected areas of psychology, theoretical physics, philosophy and economics.

If you feel have derived some benefit from the preceding tutorial, please consider making a donation to Oxford Forum's work, via the PayPal button on the following webpage:

http://celiagreen.blogspot.co.uk/

Some other publications by members of Oxford Forum are described in the following pages. Unless otherwise indicated, these publications are available from either Book Systems Plus or Amazon.

Particulars of how to contact Book Systems Plus are given below.

Book Systems Plus Ltd
Station Road
Linton
Cambridgeshire
CB21 4UX
United Kingdom

Telephone: +44(0)1440 706716
Website: http://www.booksystemsplus.com/

Celia Green
Lucid Dreams
Foreword by Professor H.H. Price, FBA

Hamish Hamilton, reissued by Institute of Psychophysical Research

ISBN 978 09000760 08 (hardback)

The original scientific study of lucid dreams: dreams in which the subject is aware that he or she is dreaming, and sometimes able to control the course of the dream.

"[Lucid dream research] rests almost entirely on the meticulous descriptions and original classifications of types and subtypes put forward by Green in her initial publications ...All of us 'second generation' researchers have found ourselves continuously in her debt."
Professor Harry Hunt, Brock University

"A fascinating subject together with a wealth of equally fascinating examples."
J.B. Priestley

* * *

Charles McCreery

Dreams and Psychosis
A new look at an old hypothesis

Oxford Forum

Price £4.95; 34 pages

ISBN: 978 09536772 83

This paper proposes a theory of psychosis based on a link between sleep and hyperarousal. It is argued that the phenomenological similarities between psychosis and dreams arise from the fact that sleep can occur, not only in states of deafferentation and low arousal, but also in states of hyperarousal resulting from extreme stress (Oswald, 1962).

It is proposed that a tendency to hyperarousal leaves certain individuals vulnerable to 'micro-sleeps' (Oswald, 1962) in everyday life, with the attendant phenomena of hallucination and other sorts of reality-distortion. Delusional thinking may follow as an attempt to rationalise these intrusions of dream-phenomena into daylight hours.

Celia Green

Out-of-the-Body Experiences
Foreword by Professor H.H. Price, FBA

Hamish Hamilton, reissued by Institute of Psychophysical Research

ISBN 978 09000760 15 (hardback)

An analysis of four hundred first-hand case histories in which people seemed to leave their body and see it from outside.

"While there had been stories of out-of-body experiences for centuries, Green was the first to systematically examine a large number of first-hand accounts, from more than four hundred people ..."
Professor Oliver Sacks, *Hallucinations*

"The present volume is the first in which contemporary instances are collected, collated and studied […] the results are extraordinarily interesting, stimulating and well worth examining by the reader."
Times Literary Supplement

* * *

Celia Green and Charles McCreery

Apparitions

Hamish Hamilton, reissued by Institute of Psychophysical Research

ISBN 978 09000760 91 (hardback)

An analysis of eighteen hundred first-hand accounts of experiences in which people saw, heard or sometimes even seemed to touch people or things that were not really there.

"An excellent piece of documentation, soberly treated, and well worth reading."
Anthony Powell, *Daily Telegraph*

"Enthralling"
Manchester Evening News

Celia Green

The Human Evasion, Foreword by R.H. Ward

Hamish Hamilton, reissued by Oxford Forum

ISBN 978 09536772 45 (paperback)

An attack on the way of thought of contemporary man, revealing the patterns of prejudice which underlie his most cherished and sacrosanct opinions.

"A brilliant exposition of the human predicament."
Lord Rothermere

"A subtle and sustained attack on contemporary ways of thought."
Times Literary Supplement

"Few books, long or short, are great ones; this book is short and among those few."
R.H Ward

* * *

Fabian Tassano

The Power of Life or Death: Medical Coercion and the Euthanasia Debate

Duckworths, reissued by Oxford Forum

ISBN 978 09536772 07 (hardback)

A book which argues against medical paternalism and suggests that the increasing power given to doctors to give or withhold treatment represents a dangerous infringement of individual liberty.

"A terse, clear, incisive, intellectually first-class study of the growing power of doctors and of the lack of effective checks upon the too easily concealed but surely numerous abuses of that power."
Professor Antony Flew

"His view goes straight to the medical jugular."
Nature

"Tassano presents hair-raising case studies ... his book is a timely polemic."
Literary Review

Celia Green

The Lost Cause, Causation and the Mind-Body Problem
Foreword by Professor Howard Robinson

Oxford Forum

ISBN 978 09536772 14 (hardback)

"Celia Green has succeeded in bringing together considerations from a wide range of disciplines: philosophy, obviously, but also psychology, neuroscience and fundamental physics, making skilful use of her own empirical investigations ... most impressive."
Dr Michael Lockwood, University of Oxford

"A worthwhile reminder of the various problems which surround the physicalism which currently dominates the philosophy of mind. ...Green does a good job of exposing the dogmatic underpinnings of current materialism, adherence to which makes mental causation seem deeply problematic."
The Human Nature Review

* * *

Charles McCreery

Perception and Hallucination
The case for continuity

Oxford Forum

Price £4.95; 32 pages

ISBN: 978 09536772 76

An analysis of empirical arguments for representationalism.

"I think the present paper is a very lucid and useful article. […] This is the best case I know of, of an attempt to make an empirical - as opposed to a 'philosophical' - argument against direct realism."
Howard Robinson, PhD
Professor of Philosophy, Central European University, Budapest

Celia Green

Letters from Exile
Observations on a Culture in Decline

Oxford Forum

ISBN 978 09536772 38 (hardback)

A collection of letters and essays written by Celia Green during the period 1990 to 1999, containing trenchant analyses of education, collectivised medicine, and modern ethics. The final section of the book introduces a provocative and original distinction between tribal and territorial morality.

"I have no doubt she is a genius."
Professor Antony Flew

* * *

Fabian Tassano

Mediocracy
Inversions and Deceptions in an Egalitarian Culture

Oxford Forum

ISBN 978 09536772 69 (hardback)

'A witty exposure of left wing foibles.'
Sir Samuel Brittan

'A marvellous counterblast against the psychobabble emanating from the politically correct pseudo-intellectuals who now infest British academia.'
Frederick Forsyth

'Delightfully dissects the language of modern egalitarianism and political correctness. Witty, biting and definitely not to be read by New Labour.'
Professor Patrick Minford

'Read this book and gain important insights into the way that the cultural elite's language works to disorient public debate.'
Professor Frank Furedi

Charles McCreery

The Abolition of Genius
Foreword by Professor H.J. Eysenck, PhD, DSc

Oxford Forum

Free online PDF at:
http://www.celiagreen.com/abolition6.pdf

ISBN: 978-0-9536772 69

An analysis of the relationship between genius and money. Dr McCreery puts forward the controversial thesis that the possession of a private income, either by the genius or by his or her patron, has been a necessary condition of the productivity of the great majority of geniuses throughout history.

'This is a courageous, well-argued and timely book ...'
H.J. Eysenck

124 pages