BUILDING SCIENCE
IN AFRICA

AIMS VIDEO COURSES

SUPPORTING BOOKLET

PROBABILITY & STATISTICS

WITH

PROF DAVID SPIEGELHALTER

AIMS

SOUTH AFRICA

African Institute for Mathematical Sciences

AIMS SOUTH AFRICA

6 MELROSE ROAD | MUIZENBERG | CAPE TOWN 7945 | SOUTH AFRICA
TEL: +27 (0)21 787 9320 | FAX: +27 (0)21 787 9321
EMAIL: info@aims.ac.za | WEB: www.aims.ac.za

AIMS Online Courses

The mission of the AIMS academic programme is to provide an excellent, advanced education in the mathematical sciences to talented African students in order to develop independent thinkers, researchers and problem solvers who will contribute to Africa's scientific development.

Teaching at AIMS is based on the principle of learning and understanding, rather than simply listening and writing, during classes, and on creating an atmosphere of increasing our knowledge through class discussions, through small group discussions, by formulating conjectures and assessing the evidence for them, and sometimes going down wrong paths and learning from the mistakes that led us there. The essential features of the classes at AIMS are that, in contrast to formal lecture courses, they are highly interactive, where the students engage with the lecturer throughout the class time, are encouraged to learn together in a journey of questioning and discovery, and where lecturers respond to the needs of the class rather than to a pre-determined syllabus. AIMS teaching philosophy is to promote critical and creative thinking, to experience the excitement of learning from true understanding, and to avoid rote learning directed only towards assessment.

Leading international and local experts offer the courses at AIMS, which are three weeks long (each module consisting of 30 hrs) and collectively form the coursework for a structured masters degree which also includes a research component. The advertised content is a guide, and the lecturers are encouraged, and indeed expected, to adapt daily to meet the current needs of the students.

Over the past ten years AIMS has achieved international recognition for this innovative and flexible approach. It has been the starting point for the remarkable success of our students and alumni and we all benefit from the support of many who have "witnessed the AIMS-magic and keep coming back for more."

This year we have decided to film selected courses and to make them available to a larger audience as an online facility. African universities may choose to use these courses to supplement and enhance their own postgraduate programmes. We believe this would be best achieved through engagement with AIMS. One way for this to happen, would be for AIMS to suggest or nominate a specialist tutor to spend time at the university, guiding students who follow the online programme. Where possible expert lecturers who have taught at AIMS may visit the university to give a short introduction to the course. We would welcome this interaction as well as the contribution our online courses will make to the growth of the mathematical sciences ecosystem in Africa.

Barry Green
Director & Professor of Mathematics
African Institute for Mathematical Sciences
January 2013

Probability & Statistics Video Course

http://www.youtube.com/aimsacza

Probability & Statistics Video Course List

WEEK 1

DAY 1 - 01:08:50

DAY 2 - 01:07:09

DAY 3 - 01:04:21

DAY 4 - 01:16:12

DAY 5 - 01:07:04

WEEK 2

DAY 6 - 01:11:57

DAY 7 - 01:06:09

DAY 8 - 01:12:14

DAY 9 - 01:21:08

WEEK 3

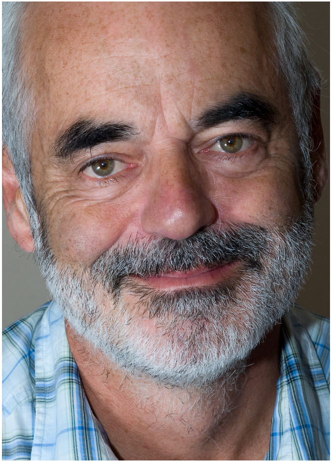
DAY 10 - 00:56:53

DAY 11 - 00:55:50

DAY 12 - 00:45:48

DAY 13 - 01:19:11

DAY 14 - 01:07:26



PROF DAVID SPIEGELHALTER

Prof David Spiegelhalter has been Winton Professor of the Public Understanding of Risk at the University of Cambridge since October 2007, which he combines with being a Senior Scientist in the MRC Biostatistics Unit.

His background is in medical statistics, with an emphasis on Bayesian methods: his MRC team developed the BUGS software which has become the primary platform for applying modern Bayesian analysis using simulation technology.

Websites
Statistics Laboratory: http://www.statslab.cam.ac.uk/Dept/People/Spiegelhalter/davids.html
Twitter: http://twitter.com/#!/undunc
Blog: http://understandinguncertainty.org/blog

PROBABILITY & STATISTICS
2012

PROF DAVID SPIEGELHALTER

DAY 1

AIMS
SOUTH AFRICA

Probability

David Spiegelhalter

AIMS - University of Cambridge

September 2012

Summary of course (possibly)

Weeks:

1. *Probability*: basic laws, random variables (discrete and continuous), moments, change-of-variable, Bayes theorem, *distributions*
2. *Introductory statistics*: sampling theory, maximum likelihood, estimation, hypothesis testing
3. *Statistical modelling*: regression, generalised linear models, Bayesian analysis

Associated practicals using SciPy, and questions of varying difficulty

Will try and relate to real-world problems

Please help each other!

Probability: summary of 5 lectures (possibly)

1. Intuitive probability
2. Equally-likely events
3. Discrete random variables (more formal)
4. Continuous random variables
5. Bayes theorem

Focus on skills necessary for statistics: so not covering important topics such as random walks, branching processes, Markov chains etc

Useful books in the library (in order of mathematical subtlety):

- QA276.C1X JA Crawshaw *A Concise Course in A-level Statistics with worked examples*[Basic, A-level, pages 134-259 for great worked probability examples]
- QA273.H655X SE Hodge and ML Seed *Statistics and Probability* [Basic, A-level, recommended, uses probability trees]
- QA273.G735 G Grimmett and D Welsh *Probability: an Introduction*. Recommended
- QA273.F3712 W Feller: *An Introduction to Probability Theory and its Applications* Wonderful classic book: more advanced
- QA273.W626 D Williams *Weighing the Odds: a Course in Probability and Statistics*. A very individual, even eccentric book, more mathematical but friendly

1: Intuitive probability

1. Early ideas of randomness
2. Cardano in 1560s
3. Chevalier de Méré: Pascal and Fermat in 1650s
4. Probability as proportion
5. Frequency trees
6. Multiplication and addition rules
7. Joint probability and cross-tabulations
8. Conditional probability
9. Reversing the conditioning
10. What is probability?

Early ideas of randomness

Egyptians and Romans gambled and played games using randomising devices

Astragali (knucklebones)  and dice were both used

Astragali could land on one of four faces: non-symmetric

People were obsessed with gaming, but nobody did the maths!

Gerolamo Cardano (1501-1576)

Throw two dice and add total. How likely is 7 compared to 2?

Listed all possibilities 1-1, 1-2, 1-3 etc

Idea of *sample space of equally likely events*

Probability as proportion: frequency tree

Cardano's problem: think of what you expect to happen in 36 throws.

Frequency tree displays a set of sequences of events: sequences are grouped into branches that represent different joint outcomes

- Express the number of each joint outcome as a fraction of total sequences
- This fraction can be considered the probability of the outcome
- Have implicitly assumed *equally-likely outcomes*

Probability as proportion: probability tree

Probability tree express as fraction at each 'branch'

Then we obtain three rules of probability *intuitively*

- **Complement rule**: the probability of a joint outcome is (1 - the probability of NOT getting the outcome)
- **Addition rule**: for the probability of a set of outcomes, add up the probabilities over the ends of the relevant branches
- **Multiplication rule**: for the probability of a joint outcome, multiply together the probabilities on the splits

Some notation

Suppose we have

- an experiment $\mathcal{E}$ with
- a set of possible outcomes Ω ('something happens')
- The impossible event $\emptyset$ ('nothing happens')

Intuitively, we want to assign a probability $P(A)$ for a set A.

- $P(\Omega) = 1, P(\emptyset) = 0$
- **Complement rule**: $P(A) = 1 - P(A^c)$
- **Addition rule**: If A and B are disjoint (ends of branches) $P(A \cup B) = P(A) + P(B)$
- **Multiplication rule**: if $P(B|A)$ denotes the probability of B given A has occurred, then $P(A \cap B) = P(B|A)P(A)$. Say A and B are *independent* if $P(B|A) = P(B)$, and so $P(A \cap B) = P(A)P(B)$.
- Can rearrange to give definition of *conditional probability* : $P(B|A) = P(A \cap B)/P(A)$

Long-term relative frequency (LTRF) (Williams)

We have used the intuitive idea that if we repeat an experiment lots of times, the *proportion* tends to the *probability*

- Let A be an event associate with an experiment $\mathcal{E}$
- Consider a 'super-experiment' $\mathcal{E}^{\infty}$ of infinite independent replications of $\mathcal{E}$
- $N(A, n)$ is the number of occasions A occurs in first n replicates
- The LTRF idea is that $\frac{N(A,n)}{n}$ converges to $P(A)$

So if we repeat something repeatable, enough times, the *proportion* tends to the *probability*

This is only *motivation*: will be made more rigorous later as 'laws of large numbers'

Pascal and Fermat in 1650s

Chevalier de Méré  asked Pascal to help him with his gambling

- Bet 1: throw a 'six' in 4 throws of a single dice
- Bet 2: throw '2 sixes' in 24 throws of two dice

The Chevalier found he tended to win Bet 1, and lose Bet 2. Why?

Pascal and Fermat solved his problem

The Chevalier's Bet 1

One dice thrown 4 times. Probability of getting a six?

Can draw whole tree with 6^4 branches, but easier to consider 'six / not-six' split at each throw.

$=$ 1 - (probability of NOT getting a six) [*complement rule*]

$=$ 1 - (probability of not getting a six on 1st throw AND not on the 2nd throw AND not on 3rd throw AND not on 4th throw) [*multiplication rule for independent events*]

$= 1 - \left(\frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6}\right)$

$=$ 1 - 625/1296 $=$ 671/1296 $=$ 0.52

Slightly in the Chevalier's favour

Joint probability and cross-tabulations

Suppose we sample 100 people, and cross-classify them by whether they are male/female and do/don't like football (soccer)

Imagine we get this data

	B: like football	B^c: don't like	
A: female	40	10	50
A^c: male	20	30	50
	60	40	100

Can calculate overall fractions - under what circumstances would these represent probabilities?

Conditional probability

We can also construct a frequency / proportion / probability tree as a sequence of questions: gender? and then football?

Reversing the conditioning

Or as football? and then gender?

We shall see that this is known as 'Bayes theorem'

Decomposition result

Overall proportion liking football?

Just multiply down branches and adding across outcomes

More formally

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c).$$

Example: $P(\text{Like football}) =$
$P(\text{Like football}|\text{female})P(\text{female}) + P(\text{Like football}|\text{male})P(\text{male}).$
$= (0.8 \times 0.5) + (0.4 \times 0.5) = 0.6$

But what IS probability?

- Classical: fraction of equally likely events. But needs definition of 'equally likely'. So circular.
- Frequency: long-term relative frequency (LTRF) - what things settle down to if we wait long enough or collect enough data. But we don't have an infinite time!
- Bayesian: reasonable betting odds. What about the probability that the next man I ask likes football? Either he does or he doesn't - our uncertainty is 'epistemic' - but we can still say how likely we think the event is.

Some probability questions and practical exercises - 1

1. I flip a coin 4 times. What is the probability of getting an odd number of 'Heads'?

2. A drawer contains 3 white socks and 3 black socks. If I pick 2 socks at random (without replacement), what is the chance I get a *non-matching* pair?

3. If I put the first sock back in the drawer before taking the second, how does this change the probability of a non-matching pair?

4. If I throw 2 dice 24 times, what is the chance of getting at least one 'double-6'?

5. Write a program in SciPy to simulate throwing a dice 4 times, repeat this experiment 1,000 times and see in what proportion a '6' appeared

6. Write a program in SciPy to simulate throwing two dice 24 times, repeat this experiment 1,000 times and see in what proportion of sequences a 'double-6' appeared. Try 1,000,000 times. Compare with the exact answer (see above)

7. A college in a very foreign country is composed of 70% men and 30% women. It is known that 40% of the men, and 60% of the women, smoke cigarettes. If I see someone smoking a cigarette, what is the probability that it is a woman? Under what circumstances would this be a realistic assessment of a probability?!

8. * A drawer contains white socks and black socks. if I pick 2 socks at random, the probability of picking 2 white socks is 0.5. What is the smallest number of socks for which this is true? Find another possible solution. [You could try trial-and-error, write a program, or use some number theory!]

Useful books in the library (in order of mathematical subtlety)

- QA276.C1X JA Crawshaw *A Concise Course in A-level Statistics with worked examples*[Basic, A-level, pages 134-259 for great worked probability examples]
- QA273.H655X SE Hodge and ML Seed *Statistics and Probability* [Basic, A-level, recommended, uses probability trees]
- QA273.G735 G Grimmett and D Welsh *Probability: an Introduction.* Recommended
- QA273.F3712 W Feller: *An Introduction to Probability Theory and its Applications* Wonderful classic book: more advanced
- QA273.W626 D Williams *Weighing the Odds: a Course in Probability and Statistics.* A very individual, even eccentric book, more mathematical but friendly