

MathStat, MORSE and Data Science: Degrees of Freedom

Department of Statistics
University of Warwick

20 / 21 June 2014

Professor David Firth, Dr Julia Brettschneider

Landscape of Mathematical Degrees at Warwick

Organised by Mathematics

- ▶ Maths (single honours)
- ▶ Maths and ... (Physics, Economics, Business, Philosophy) — joint honours

Organised by Statistics

- ▶ Mathematics and Statistics (*Mathstat*)
- ▶ Mathematics, Operational Research, Statistics and Economics (*MORSE*)
- ▶ Data Science

Organised by Computer Science

- ▶ Computer Science
- ▶ Discrete Mathematics

Mathematics and Statistics

(3 years, single-honours BSc; or 4 years, MMathStat)

A Mathematics degree for mathematically able students who have an interest in solving practical problems which arise in physical, life and social sciences, business, health and government.

<http://warwick.ac.uk/mathstat>

Two well-established Integrated Single Honours courses, leading to an extremely wide variety of careers.

MORSE

(3 years, single-honours BSc; or 4 years, MMORSE)

A Mathematics degree for those interested in a course which combines theory with modern applications.

M Mathematics

OR Operational Research

S Statistics

E Economics

<http://warwick.ac.uk/MORSE>

Data Science

(3 years, single-honours BSc)

A degree course designed for able mathematicians with an interest in pursuing sophisticated theory and methods relevant to modern applications requiring

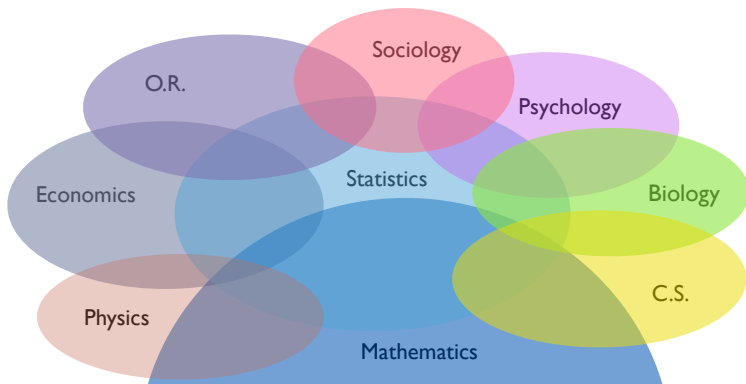
large-scale data analysis and state-of-art computer science

<http://warwick.ac.uk/datsci>

Running for the very first time in 2014–15 ...

Mathematics: applications and inspirations

David Hilbert: *“The instrument that mediates between theory and practice, between thought and observation, is mathematics; it builds the bridge and makes it stronger and stronger.”*



Mathematics of Random Sequences

Which of the sequences below are random?

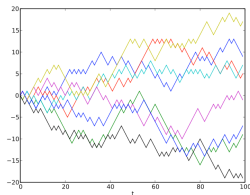


- ▶ What characterises randomness? How to measure it?
- ▶ “Real world” examples: queues, games, weather
- ▶ Modelling, analysis and prediction
- ▶ Higher dimensional random processes

Modules: e.g. **ST115 Probability**, **ST202 Stochastic Processes**

Random Sequences in Financial Markets

Random walk (e.g. random sequence white=+1, black=-1):



- ▶ Model for games, stocks
- ▶ Continuous time: Brownian motion
- ▶ Alternative models include dependency on the past

Modules: e.g. **ST339 Introduction to Mathematical Finance**,
ST906 Financial Time Series Analysis

Random Sequences in Genomics

Objective: Computational detection/discovery of DNA motifs

Rationale: Non-random structures in DNA often indicate biological meaning

Example: Key to gene regulation are promoters.



TTGTT motif for a *S. cerevisiae* (yeast) gene

Image from www.nature.com/nbt/journal/v24/n4/fig_tab/nbt0406-423_F1.html

Modules: e.g. **ST341 Statistical Genetics**, **ST323 Multivariate Statistics**

Operational Research (O.R.)

or **Management Science**: methods for solution of organisational problems of business and other enterprises.

Product pricing decisions: Competition between companies and *prisoner's dilemma*

3D printing: Major bottleneck is quality of the object

Mathematical topics in O.R.: Quality and efficiency in production and storage, queues (e.g. A & E departments, phone centres) etc.

Modules: e.g. **IB104 Linear Programming, ST222 Games, Decisions and Behaviour, IB407 Decision Analysis, ST413 Bayesian Statistics and Decision Theory**

Optimization



Photo by futureatlas <http://futureatlas.com/blog>

Since 2001, optimization has saved Air New Zealand £8M/yr
(Source: Optima Corporation <http://www.theoptimacorporation.com/airline>)

Optimization



Photo by futureatlas <http://futureatlas.com/blog>

Since 2001, optimization has saved Air New Zealand £8M/yr
(Source: Optima Corporation <http://www.theoptimacorporation.com/airline>)

Statistics

Illustration through some recent 4th-year projects:

- ▶ *Lee-Carter Models to Study the Changing Age Profile of the UK Population*
- ▶ *Relative Importance of Variables in Rankings of Universities*
- ▶ *Quasi Variances: Methods of Calculation, and Developing Software Implementations*
- ▶ *Analysis of Paired Comparison Data using Bradley-Terry Models, with Applications to Premier League Football*
- ▶ *Election-night Forecasting: How to Learn from Previous Elections?*

Modules: e.g. **ST221 Linear Statistical Modelling**, **ST332: Medical Statistics**, **ST337 Bayesian Forecasting and Intervention**

Getting it Right on Election Night

David Firth, Warwick Statistics

Research by David Firth has developed new statistical methods that have already been highly successful — in full public view — at the last two UK General Elections.

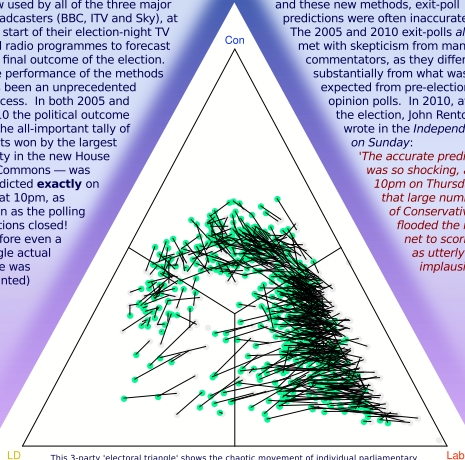


Projection of the predicted 2010 election outcome onto the tower of Big Ben, shortly after the 10pm close of polling stations. Photo courtesy of Gary White.

The new approach to exit polling is now used by all of the three major broadcasters (BBC, ITV and Sky), at the start of their election-night TV and radio programmes to forecast the final outcome of the election. The performance of the methods has been an unprecedented success. In both 2005 and 2010 the political outcome — the all-important tally of seats won by the largest party in the new House of Commons — was predicted **exactly** on air at 10pm, as soon as the polling stations closed! (before even a single actual vote was counted)

In previous elections, before 2005 and these new methods, exit-poll predictions were often inaccurate. The 2005 and 2010 exit-polls also met with skepticism from many commentators, as they differed substantially from what was expected from pre-election opinion polls. In 2010, after the election, John Rentoul wrote in the *Independent* on Sunday:

'The accurate prediction was so shocking, at 10pm on Thursday, that large numbers of Conservatives flooded the internet to scorn it as utterly implausible...'



This 3-party 'electoral triangle' shows the chaotic movement of individual parliamentary constituencies in successive general elections. The key to predictive success is a good **probabilistic model** for these changes in vote-share among the main political parties.

The new methods are fully described and analysed in **Exit Polling in a Cold Climate** by J Curtice and D Firth *Journal of the Royal Statistical Society*, 2008

For more information, please see <http://warwick.ac.uk/exitpolling>



Another example: Azhar's project. (Azhar is now Principal Statistician at the Ministry of Justice, in London.)

Modelling and Estimation of Social Networks



Azzie did a *superb* study of **exponential random graph models**, which help explain and predict network connections.

Economics

Some of our MORSE students do a *lot* of economics. Others do less (there is great flexibility).

Modules: For example, **EC310 Topics in Development Economics** . . . Population, fertility and gender; land, credit, insurance; trade and development; . . .

Another example, **EC312 International Economics** . . . Models of exchange rate determination with flexible prices, fixed prices and sticky prices . . .

How high should be the charge for using the M6 Toll Road?

Of course the toll is set to provide
a good return on private investment.
However . . .



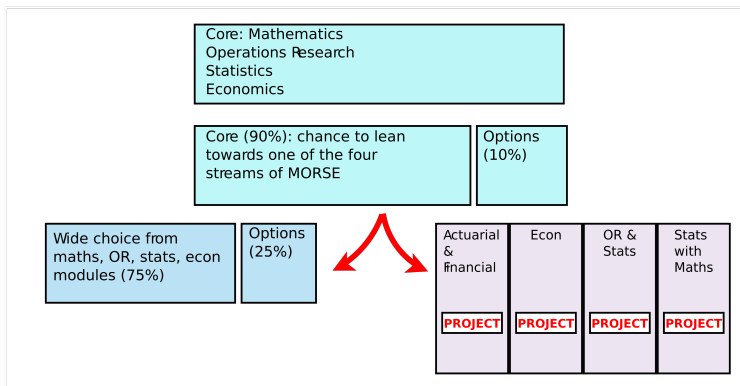
©Adrian Bailey and licensed
for reuse under CC BY-SA 2.0.

. . . if the toll is set too low, usage will rise and so will
congestion;

. . . if the toll is set too high, usage will be too low.

This kind of consideration is typical for many public goods.

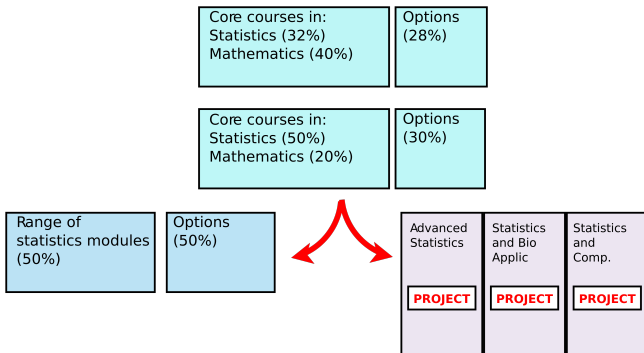
MORSE / MMORSE course structure:



Options: physics, philosophy, comp sci, maths, stats.

Other options: languages, history, chemistry . . .

MathStat / MMathStat course structure:



Options: maths, stats, OR, econ, comp sci, physics, education.
Other options: languages ...

And now a further possibility: Data Science

The future?

BSc Data Science at Warwick driven by:

- ▶ high, and still rapidly growing, **demand from employers** globally (for 'analytics', 'big data' specialists ... 'data scientists')
- ▶ strong **demand from our own students** (for more experience with data technology, programming, etc.)

Like MORSE, an innovative Warwick collaboration (this time between Computer Science and Statistics).

Data Science: What is it?

Some central themes:

- ▶ data **everywhere** (banks, supermarkets, online, transport, weather, government, genomics, ...)
- ▶ often **massive** data
- ▶ some common aims are
 - ▶ extract useful **patterns**
 - ▶ identify **what works**
 - ▶ make real-time **predictions**
- ▶ demands skill in both **computing** and **statistics** (as well as 'common sense', communication skills, domain knowledge, ...)

Data Science: the hype and the reality

Hype?

- ▶ “Data Scientist: The Sexiest Job of the 21st Century” (Harvard Business Review, October 2012)
- ▶ “Data Scientists: The New Rock Stars of the Tech World” (techopedia, 20 April 2012)
- ▶ “Data Crunchers: Now the Cool Kids on Campus” (The Wall Street Journal, 1 March 2013)

Data Science: The Hype and the Reality

Reality: Evidence-based studies show large, and increasing, demand.

McKinsey Global, 2011 report based on careful analysis of market supply and demand:

By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills.

Graduates in Data Science at Warwick will have precisely the combination of knowledge and skills needed to play leading roles in this rapidly growing field.

BSc Data Science: The Course

Year 1 Core Maths, Stats, Comp. Sci.

Year 2 Core Stats, Comp. Sci. + options (2:1 ratio)

Year 3 Data Science project (25% time commitment), and wide array of options (in Stats, Comp. Sci. and more widely)

Possibility also of *intercalated year* in industry or at an overseas partner university.

More information: warwick.ac.uk/datsci

Sample first-year timetable

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
9:00		...still working on it...		Lin. Algebra Lecture	Running	Catching up sleep	Dream about solution to Q4
10:00	Library	Finishing it off for deadline	Maths supervision	Finishing off assignment...	Analysis Study Group	Still sleeping	Getting this on paper...
11:00	Probability Lecture	Analysis Lecture	Probability Study Group	...for noon deadline!	Revise Analysis	Still sleeping	...doesn't actually work
12:00	Working on assignment...	Probability Lecture	Probability Lecture	Practice clarinet	Analysis Lecture		Running club
13:00	...for Stat Lab...			Probability Tutorial	Stat Lab Lecture	Out and about	
14:00		Stat Lab Computer Lab	Back to assignment...		Lin. Algebra Lecture	Out and about	Fix bicycle
15:00	Probability Problem Class	Stat Lab Study Group	...still working on it	Analysis Lecture		Out and about	Install R-packages,...
16:00	Spanish Class	Stat Lab Lecture	Athletics	Stat Lab Lecture	Library	Start Analysis assignment...	...try sample code
17:00	Spanish Class	Lin. Algebra Lecture	Athletics	Write Blog entry	Econ Lecture	...still working on it until...	Practice clarinet
18:00		Econ Lecture	Nap (unplanned)		Grocery shopping group	...stuck in Question 4	Laundry
19:00	...still working on it...		still napping	Ensemble practice...	Cooking and cleaning team		Reading for Econ
20:00	...still working on it...	Probability assignment...	Society Social	Ensemble practice...	Dinner...	Off to Leamington...	Film club...
21:00	...still working on it...	...still working on it...	Society Social	Watch tele	together...	...clubbing...	Film club...

Careers

A few examples among the destinations of our recent graduates:

- ▶ management consultancy
- ▶ investment banking
- ▶ market research
- ▶ 'big data' in commerce, science, government, ...
- ▶ medical research
- ▶ insurance and actuarial work
- ▶ social or economic research
- ▶ engineering
- ▶ sport, entertainment

etc., etc.

Our established single-honours BSc and Integrated Masters courses **Mathematics and Statistics** and **MORSE** attract excellent students, and produce graduates who are in great demand. And our new BSc Honours course, **Data Science**, is designed specifically to allow graduates to quickly become highly-skilled leaders in the "big data" era.

The emphasis in all of our courses is on mathematically based learning that is of very direct relevance to the demands of the real world. This helps explain why our graduates are so highly sought-after, even relative to graduates from other maths degrees.

The **range of destinations** of MathStat and MORSE graduates is very wide indeed. Some of the most popular:

Actuary • Investment Analyst • Statistician or Statistical Researcher • Software Engineer • Management Consultant • Marketing Analyst • Credit Risk Analyst • Accountant

Postgraduate Study, often at top graduate schools – last year including Cambridge, Oxford, ETH (Zurich), Cornell (New York), HKUST (Hong Kong) etc.

According to a report published in The Telegraph in January 2013, **Warwick is the number one target** for graduate recruitment by top employers. *The Graduate Market in 2013* reports that large, high-profile UK employers are focusing on finding graduate recruits at a small handful of elite universities, headed by Warwick. For more details, follow the links at

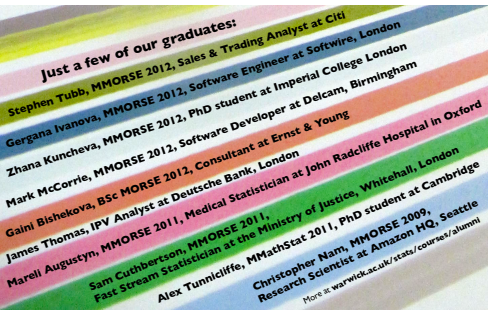
warwick.ac.uk/stats/courses/warwick-graduates-in-demand

Some data from Unistats (as at 2014-03-10): graduates surveyed 6 months after graduation

	In work or further study	(of those employed) In a professional or managerial job
MMORSE	95%	88%
BSc MORSE	85%	89%
MMathStat	96%	88%
BSc MathStat	96%	88%

Compares favourably with some of Warwick's main competitors, e.g.,

LSE		
Bus Math Stat	88%	88%
Stat with Finance	84%	89%
Actuarial Science	84%	88%
Math & Econ	81%	85%
Imperial College		
Math with Stat	84%	89%
Math with Stat For Finance	84%	89%



Entry Requirements

MathStat, MORSE and Data Science

Typical entry requirements:

**A* (Maths) + A (Further Maths) + A OR
A (Maths) + AA + 2 in STEP or Merit in AEA**

Other qualifications (e.g. IB) see:

<http://warwick.ac.uk/stats/courses/offer>

Economics A-level is *not* a prerequisite for MORSE.

Computing A-level is *not* a prerequisite for Data Science.

We encourage taking **STEP**, note September workshops.

Numbers in our current first year are 85 Home, 5 EU, and 75 other international students. Our current 4th year has 34 students.

Applications

Application Guidance can be found online:
<http://warwick.ac.uk/stats/courses/faq>

Personal statement My other choices will be straight maths
... *Don't worry!*
We welcome statements which are mainly or
exclusively about mathematical interests and
achievements.

Interview? No. Offer-Holders' Day in Spring.

Thank you for coming. Any questions?



Staff available for discussion here and in the atrium under the Statistics Department sign