



STAT 830

Prelude to Bio-informatics

Semester 1 2008

Unit Outline

Unit Convenor: Stephen Brown

Students in this unit should read this unit outline carefully at the start of semester. It contains important information about the unit. You should also refer to the Stat830 web page at <http://www.stat.mq.edu.au/pg/units/stat/stat830>.

About Stat830

Stat830 is a unit run by the Statistics Department in the Division of Economic and Financial Studies specifically for the Division of Environmental and Life Sciences.

It is a core unit in the Master of Biotechnology and the Master of Biotechnology /Commerce in Business double degree courses.

Basic skills of calculus, such as elementary differentiation and integration, are essential for this course. However, the emphasis will be on the understanding of the processes, rather than the actual calculations.

Lecturers:

Stephen Brown (Fridays)
Room: E4A-543
phone: 9850 8552 fax: 9850 7669
scbrown@efs.mq.edu.au

Hilary Green (Wednesdays)
Room E4A-531
Phone 9850 8562 fax: 9850 7669
hgreen@efs.mq.edu.au

Lectures:

Wednesday: 6 - 8pm, E6A-102 For STAT273 and STAT830

Friday: 2 - 3pm, E5A-118 For STAT830 only

Students are expected to attend both of these lectures each week.

Friday: 3 - 4pm E5A-118 optional

The Wednesday lectures will consider the general theory of probability and probability distributions for both STAT273 and STAT830 students.

The Friday lectures will apply these general concepts to specific issues from bioinformatics.

In addition, an optional lecture will be held in E5A-118 on Fridays 3 - 4 pm, immediately following the earlier lecture. **No new work will be presented during this lecture.** However, it will be an opportunity to revise any assumed knowledge or work through further examples. It is especially recommended for students whose mathematical or statistical background is weak.

Texts:

Lecture notes for the Wednesday lecture will be available from the STAT273 WebCT the night before the lecture. Students should read the lecture notes before the lecture. The assignment and lecture notes pages are password-protected. The userid and password will be provided at the first lecture.

The Lecture Notes for the Friday lecture will be handed out at the beginning of each lecture and will be placed on the Website.

References that may be useful:

Basic Probability:

- Kinney, J.J. (1997) Probability - An Introduction with Statistical Applications, John Wiley and Sons QA273.K493/1997
- Scheaffer R.L. (1994) Introduction to Probability and Its Applications, (2nd Edition) Duxbury Press, QA273.S357
- Wackerly, D., Mendenhall W. Scheaffer. ()Mathematical Statistics with Applications (4th,5th or 6th Editions) QA276 .M426 2002
- Sincich, T., Levine, D.M., Stephan, D. (1999) Practical statistics by example using Microsoft *Excel* QA276.12 .S554

Copies these books are held in the Reserve section of the library.

Students will also be directed to relevant websites each week.

BioInformatics:

- Warren J. Ewens and Gregory R. Grant, Statistical methods in bioinformatics, an introduction, Springer-Verlag, New York, 2001. QH324.2 .E97 2004
- Kenneth Lange, Mathematical and statistical methods for genetic analysis, Statistics for Biology and Health, Springer-Verlag, New York, 2002. QH438.4.M33 .L36 2002
- Pak Sham, Statistics in human genetics, Arnold, 1998.
- Greg Gibson and Spencer V. Muse, A primer of genome science, Sinauer Associates, Inc (2002)

For each of these texts, only specific sections will be relevant.

Websites:

1. Stat830 homepage: <http://www.stat.mq.edu.au/pg/units/stat/stat830>
(for general information, assignments and solutions)
2. Stat273 homepage: http://www.stat.mq.edu.au/ug/units/stat_units200/stat273/
(for STAT273 general information, and STAT273 assignments and solutions)
3. Online Discussions and Review Quizzes and solutions: <https://learn.mq.edu.au/webct>

Online Quizzes:

These provide you with an opportunity to practice and gauge your basic understanding of the concepts introduced in the course. They contribute 12% towards the total assessment.

Your login ID is "gffffnnn", where "g" is the first letter of your given name, "ffff" is the first four letters of your family name (or as many as possible if less than four) and "nnn" are three digits. Your initial password "abnnnnnn" where "ab" will be two random characters and "nnnnnn" your date of birth. You will have been mailed your username and the two random characters after enrolling.

Assessment:

The unit will be assessed as follows:

Two Mid-semester Tests (6% each)	12%	
On-line quizzes:	12%	
3 Assignments	16%	(assignments not of equal value)
Final Exam	60%	

A satisfactory performance is required in all aspects of the unit. **To pass the unit you must pass the final examination.** Your final grade will reflect your performance in **all** aspects of the course. In order to attain a particular grade in the unit, the requirement for that grade must be met in **the exam and the coursework**. You should also refer to the University's rules on grades, rule 10(2), <http://handbook.mq.edu.au/PDFs/ug-bachelor-degree-rules.pdf>.

Assignments:

There will be three compulsory assignments. They are to be handed in at ERIC (E4B-106) assignment by 10 a.m. on the due date (a Friday). The assignment will not be marked unless it is accompanied by the appropriate assignment cover sheet. Failure to submit assignments may result in automatic exclusion from the unit. The due dates are marked on the Schedule below.

Mid-semester Tests:

Two mid-semester tests of 50 minutes will be held in the first lecture hour. The tests will be with the STAT273 students. They will be on:

Week 6, 2nd April

Week 10, 14th May

Any student who cannot attend the test due to unavoidable disruption must report the circumstances (supported by medical certificate or other proper evidence) in writing to the Registrar as soon as possible.

Final Examination:

This will be an exam specifically for STAT830. It will be of 3 hours duration with 10 minutes reading time.

For the Final examination you are allowed to bring in one A4 page of handwritten notes, written on both sides. All necessary statistical tables will be provided.

An electronic calculator is essential and will be required for the final examination and class tests. Text-returnable calculators are not permitted in the tests or exam.

Plagiarism:

The University defines plagiarism in its rules: "Plagiarism involves using the work of another person and presenting it as one's own." Plagiarism is a serious breach of the University's rules and carries significant penalties. You must read the University's practices and procedures on plagiarism. These can be found on the web at: <http://www.student.mq.edu.au/plagiarism/>. Penalties may include a deduction of marks, failure in the unit, and/or referral to the University Discipline Committee.

Additional Syllabus

The topics for the Friday lectures will include:

- The Hardy Weinberg Principle.
- Recombination rates and Haldane's function.
- Marker assisted selection.
- Revision of the basic principles of hypothesis testing.
- Type I and Type II errors and problems with multiple testing of large data sets.
- Statistical problems in sequencing large strands of DNA
- Hidden Markov Processes

These topics will be covered after the basic principles have been covered in the Wednesday lectures.

Learning Outcomes:

By the end of this unit, students should be able to:

- Have a solid understanding of introductory probability theory.
- Be able to recognize the different discrete and continuous distributions.
- By identifying the distribution, use the appropriate function to obtain probabilities, expected values and variances.
- Recognize Markov processes and hidden Markov processes and understand how various situations in molecular genetics follow these processes.
- Recognize how basic theory can be used to solve particular problems encountered in sequencing DNA.
- Understand the main problems particular to analyzing large data sets and interpret the analysis accordingly.

Students with Disabilities:

Students with disabilities are encouraged to contact the Equity Support Unit to determine whether they are eligible for support service. The Equity Support Staff can be contacted on 9850 7497. The web page is <http://www.mq.edu.au/uchs/equity.html>. The Advice of Disability form may be downloaded from <http://www.sss.mq.edu.au/equity/policies/>

Misadventure:

Should you suffer illness or other misadventure which affects your performance during term or for the exam, then you should fill in an [Advice of Absence or other Circumstances](#) and /or [Request for Special Consideration](#).

Students should read the DEFS' information regarding supplementary exams, http://www.efs.mq.edu.au/services/policies_consideration.htm

Schedule for the Wednesday Probability Section

WEEK	LECTURE TOPIC	To Do
25 Feb <i>W1</i>	Experiments, sample spaces, Probability Rules, Permutations and Combinations Theoretical vs. Empirical probability	Quiz 1
3 March <i>W2</i>	Conditional Probability Independence, Bayes' Theorem	Quiz 2
10 March <i>W3</i>	Random Variables Probability Functions, Discrete Probability Distributions, Cumulative Distribution functions, Expected value and Variance	Quiz 3
17 March <i>W4</i>	Discrete Distributions; Bernoulli, Binomial, Geometric, Poisson.	Quiz 4 Thurs 20th: Assign1 due
24 March <i>W5</i>	More Discrete Distributions; Negative Binomial and Hypergeometric.	Quiz 5
31 March <i>W6</i>	Introduction to Continuous random variables	Quiz 6 Test 1
7 April <i>W7</i>	Cumulative distribution function Functions of Random Variables, Sampling distributions, Uniform and Exponential Distributions.	Quiz 7
Midsemester Break: 12th – 27th April		
28 April <i>W8</i>	Normal Distribution Model checking, Central Limit Theorem, Normal Approximations	Quiz 8 2nd: Assign2 due
5 May <i>W9</i>	Gamma Distributions, Beta Distributions Tchebysheff's Theorem	Quiz 9
12 May <i>W10</i>	Chi-squared Distribution, Distribution of sample variance, F-Distribution, Test for Equality of Variance, t- Distribution, Distribution of sample mean (σ unknown)	Quiz 10 Test 2
19 May <i>W11</i>	Joint Distributions: Discrete and Continuous cases	Quiz 11
26 May <i>W12</i>	Introduction to Markov Chains States, Transition probabilities, State vectors, Equilibrium, Absorbing States	Quiz 12 30th: Assign3 due
2 June <i>W13</i>	Review	Review Quiz,

The schedule for the Friday series will be determined as the course proceeds.