



STAT830

Prelude to Bioinformatics

S1 Day 2014

Statistics

Contents

<u>General Information</u>	2
<u>Learning Outcomes</u>	3
<u>Assessment Tasks</u>	3
<u>Delivery and Resources</u>	6
<u>Unit Schedule</u>	7
<u>Policies and Procedures</u>	9
<u>Graduate Capabilities</u>	11

Disclaimer

Macquarie University has taken all reasonable measures to ensure the information in this publication is accurate and up-to-date. However, the information may change or become out-dated as a result of change in University policies, procedures or rules. The University reserves the right to make changes to any information in this publication without notice. Users of this publication are advised to check the website version of this publication [or the relevant faculty or department] before acting on any information in this publication.

General Information

Unit convenor and teaching staff

Lecturer (STAT273)

Hilary Green

hilary.green@mq.edu.au

Contact via hilary.green@mq.edu.au

Lecturer (STAT273)

Maurizio Manuguerra

maurizio.manuguerra@mq.edu.au

Contact via maurizio.manuguerra@mq.edu.au

E4A 452

TBA

Unit Convenor

Georgy Sofronov

georgy.sofronov@mq.edu.au

Contact via georgy.sofronov@mq.edu.au

E4A536

Friday 11am-1pm

Tutor

Madawa Weerasinghe J R

madawa.weerasinghe@mq.edu.au

Contact via madawa.weerasinghe@mq.edu.au

Credit points

4

Prerequisites

Admission to MBiotech or MBiotechMCom or MLabQAMgt or PGDipLabQAMgt or PGCertLabQAMgt

Corequisites

Co-badged status

Unit description

This unit introduces the statistical and probabilistic concepts that are the basis for the study of bioinformatics. Topics include an introduction to probability and conditional probability, probability distributions, sampling distributions and an introduction to Markov processes. Particular attention is paid to how they relate to specific applications in the field of bioinformatics. A basic understanding of calculus will be an advantage.

Important Academic Dates

Information about important academic dates including deadlines for withdrawing from units are available at <https://www.mq.edu.au/study/calendar-of-dates>

Learning Outcomes

On successful completion of this unit, you will be able to:

Understand basic notions of probability theory.

Familiarity with classes of discrete and continuous random variables and their distribution functions. Evaluate probabilities of events, expected values, variances and higher moments of random variables.

For random vectors be able to evaluate joint distributions.

Understand basic properties of Markov Chains. Recognize Markov processes and understand how various situations in molecular genetics can be modelled by these processes.

Recognize how basic probability theory can be used to solve particular problems encountered in analysis of DNA sequencing.

Be familiar with basic principles of statistical data modelling.

Assessment Tasks

Name	Weighting	Due
Tests	20%	Week 3, Week 6
Final Examination	60%	To be advised
Assignments	20%	Week 8, Week 12

Tests

Due: **Week 3, Week 6**

Weighting: **20%**

There will be two mid-semester tests (10% each) of 50 minutes duration held during STAT273

lectures in week 3 and week 6.

Students are allowed to bring in one A4 page of handwritten notes, written on both sides. All necessary statistical tables and formulae will be provided. An electronic calculator is essential. Text-returnable calculators are not permitted in the tests or exam.

On successful completion you will be able to:

- Understand basic notions of probability theory.
- Familiarity with classes of discrete and continuous random variables and their distribution functions. Evaluate probabilities of events, expected values, variances and higher moments of random variables.
- For random vectors be able to evaluate joint distributions.
- Recognize how basic probability theory can be used to solve particular problems encountered in analysis of DNA sequencing.
- Be familiar with basic principles of statistical data modelling.

Final Examination

Due: **To be advised**

Weighting: **60%**

The duration of the final examination is three hours plus ten minutes' reading time. An electronic calculator and one A4 sheet of paper (written on one or both sides) may be taken in to the exam room. All material thereon must be in the student's own handwriting and not typed.

For a passing grade, satisfactory performance is required on both: (i) the average of the assignments and test; (ii) the final examination.

You are expected to present yourself for examination at the time and place designated in the University examination timetable, which will be available at <https://timetables.mq.edu.au> .

Only documented illness or unavoidable disruption may be used as reasons for not sitting an examination at the designated time. In these circumstances you may wish to consider applying for Special Consideration. Information about the special consideration policy and procedure is available at: http://www.mq.edu.au/policy/docs/special_consideration/policy.html

On-line submission of Special Consideration Applications is available at: http://web.science.mq.edu.au/undergraduate_programs/current/admin_central/

It is Macquarie University policy not to set early examinations for individuals or groups of students. All students are expected to ensure that they are available until the end of the teaching semester, that is, the final day of the official examination period.

On successful completion you will be able to:

- Understand basic notions of probability theory.
- Familiarity with classes of discrete and continuous random variables and their

distribution functions. Evaluate probabilities of events, expected values, variances and higher moments of random variables.

- For random vectors be able to evaluate joint distributions.
- Understand basic properties of Markov Chains. Recognize Markov processes and understand how various situations in molecular genetics can be modelled by these processes.
- Recognize how basic probability theory can be used to solve particular problems encountered in analysis of DNA sequencing.
- Be familiar with basic principles of statistical data modelling.

Assignments

Due: **Week 8, Week 12**

Weighting: **20%**

There will be two assignments (10% each), the first one due in week 8 and the second one due in week 12. On-time submission of all assignments is compulsory. Late submission of assignments will not be accepted without a good reason, and extension requests should be made directly to the Unit Convenor.

Assignment submission

Assignments are to be submitted to your tutor, in your tutorial in the week in which they are due. No extensions will be considered unless satisfactory documentation outlining illness or misadventure is submitted.

On successful completion you will be able to:

- Understand basic notions of probability theory.
- Familiarity with classes of discrete and continuous random variables and their distribution functions. Evaluate probabilities of events, expected values, variances and higher moments of random variables.
- For random vectors be able to evaluate joint distributions.
- Understand basic properties of Markov Chains. Recognize Markov processes and understand how various situations in molecular genetics can be modelled by these processes.
- Recognize how basic probability theory can be used to solve particular problems encountered in analysis of DNA sequencing.
- Be familiar with basic principles of statistical data modelling.

Delivery and Resources

Technologies used and required

All unit materials, including administrative updates, lecture notes, tutorials and assignments, will be posted on the Unit website on iLearn. The web address is <https://ilearn.mq.edu.au>.

Students will attend three one-hour STAT273 lectures, one one-hour STAT830 lecture and one one-hour STAT830 tutorial per week.

Students will be able to access STAT273 teaching materials available on its iLearn site.

The notes shown in lectures will be available on iLearn before the lecture is given.

Tutorial exercises will be set weekly and will be available on iLearn before the tutorial. Students are expected to have attempted all questions before the tutorial.

The timetable for classes can be found on the University web site at:
<http://www.timetables.mq.edu.au>

There is no required textbook for this unit.

Recommended reference books are:

Introduction to Probability

J. J. Kinney. *Probability - An Introduction with Statistical Applications*. John Wiley and Sons, 1997. QA273.K493/1997

R.L. Scheaffer. *Introduction to Probability and Its Applications*, (2nd Edition). Duxbury Press, 1994. QA273.S357 (Any edition).

D. Wackerly, W. Mendenhall and R. Scheaffer. *Mathematical Statistics with Applications*. (4th,5th or 6th Editions). Duxbury Press, 2002. QA276 .M426 (Any edition).

T. Sincich, D. M. Levine and D. Stephan. *Practical statistics by example using Microsoft Excel*. Prentice Hall, 1999. QA276.12 .S554 (Any edition).

Prelude to Bioinformatics

W. J. Ewens and G. R. Grant. *Statistical Methods in Bioinformatics, an Introduction*. Springer, 2001. QH324.2 .E97 2004

K. Lange. *Mathematical and Statistical Methods for Genetic Analysis, Statistics for Biology and Health*. Springer, 2002. QH438.4.M33 .L36 2002

D. Husmeier, R. Dybowski and S. Roberts. *Probabilistic Modeling in Bioinformatics and Medical Informatics*. Springer, 2005.

Software

The statistical software R will be used. This is a free software environment for statistical computing and graphics and is downloadable from the website

<http://www.r-project.org/>

in versions for Windows, MacOS and Unix platforms. R is also available in the computer labs in E4B. It is convenient to bring a memory stick when using these computers. MS Excel will also be used in this unit.

Students may find useful the link to online answer engine Wolfram Alpha:

<http://www.wolframalpha.com/>

Changes since the last offering of this unit

In current offering of STAT830, two class tests have been introduced instead of four random short tests. The statistical software R will be used.

Unit Schedule

STAT273 Introduction to Probability Schedule

Week	Lecture Topics
<u>Module 1: Introduction to first probability concepts</u>	
1	Experiments, sample spaces, probability rules, permutations and combinations
2	Conditional Probability. Bayes' Theorem
<u>Module 2: Discrete random variables</u>	
3	Random Variables, Probability Functions, Cumulative Distribution functions, Expected value and Variance
4	Important Discrete Distributions: Bernoulli, Binomial, Geometric and Poisson distributions.
5	More Discrete Distributions, Negative Binomial and Hypergeometric distributions.
<u>Module 3: Continuous random variables</u>	
6	Introduction to Continuous random variables. Cumulative distribution function.
<i>Mid-semester break</i>	
7	Important Continuous Distributions: Uniform, Exponential and Normal distributions
8	More Continuous Distributions. Gamma and Beta Distributions. Tchebysheff's Theorem
<u>Module 4: Samples and tests</u>	
9	Functions of Random Variables. Central Limit Theorem, Normal Approximations.
10	Chi-squared Distribution; Distributions of sample mean and variance; F-Distribution; Test for equality of variances.
<u>Module 5: Joint distributions and Markov chains</u>	

11	Joint Distributions: Discrete and Continuous distributions.
12	Introduction to Markov Chains. Transition probabilities, state vector, equilibrium and absorbing states.
13	Public holiday

STAT830 Prelude to Bioinformatics Schedule

WEEK	Lecture Topics
1	Introduction to STAT830
2-4	Hardy-Weinberg Equilibrium, recombination rates and Haldane's function and marker assisted selection
5-7	Statistical problems in DNA sequencing
8-9	Basic principles of hypothesis testing
10-12	Applications of Markov Processes
13	Revision

Policies and Procedures

Macquarie University policies and procedures are accessible from [Policy Central](#). Students should be aware of the following policies in particular with regard to Learning and Teaching:

Academic Honesty Policy http://mq.edu.au/policy/docs/academic_honesty/policy.html

Assessment Policy <http://mq.edu.au/policy/docs/assessment/policy.html>

Grading Policy <http://mq.edu.au/policy/docs/grading/policy.html>

Grade Appeal Policy <http://mq.edu.au/policy/docs/gradeappeal/policy.html>

Grievance Management Policy http://mq.edu.au/policy/docs/grievance_management/policy.html

Disruption to Studies Policy http://www.mq.edu.au/policy/docs/disruption_studies/policy.html *The Disruption to Studies Policy is effective from March 3 2014 and replaces the Special Consideration Policy.*

In addition, a number of other policies can be found in the [Learning and Teaching Category](#) of Policy Central.

Student Code of Conduct

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: https://students.mq.edu.au/support/student_conduct/

Student Support

Macquarie University provides a range of support services for students. For details, visit <http://students.mq.edu.au/support/>

Learning Skills

Learning Skills (mq.edu.au/learningskills) provides academic writing resources and study strategies to improve your marks and take control of your study.

- [Workshops](#)
- [StudyWise](#)
- [Academic Integrity Module for Students](#)
- [Ask a Learning Adviser](#)

Student Services and Support

Students with a disability are encouraged to contact the [Disability Service](#) who can provide appropriate help with any issues that arise during their studies.

Student Enquiries

For all student enquiries, visit Student Connect at ask.mq.edu.au

IT Help

For help with University computer systems and technology, visit <http://informatics.mq.edu.au/help/>.

When using the University's IT, you must adhere to the [Acceptable Use Policy](#). The policy applies to all who connect to the MQ network including students.

Graduate Capabilities

PG - Discipline Knowledge and Skills

Our postgraduates will be able to demonstrate a significantly enhanced depth and breadth of knowledge, scholarly understanding, and specific subject content knowledge in their chosen fields.

This graduate capability is supported by:

Learning outcomes

- Understand basic notions of probability theory.
- Familiarity with classes of discrete and continuous random variables and their distribution functions. Evaluate probabilities of events, expected values, variances and higher moments of random variables.
- For random vectors be able to evaluate joint distributions.
- Understand basic properties of Markov Chains. Recognize Markov processes and understand how various situations in molecular genetics can be modelled by these processes.
- Recognize how basic probability theory can be used to solve particular problems encountered in analysis of DNA sequencing.
- Be familiar with basic principles of statistical data modelling.

Assessment tasks

- Tests
- Final Examination
- Assignments

PG - Critical, Analytical and Integrative Thinking

Our postgraduates will be capable of utilising and reflecting on prior knowledge and experience, of applying higher level critical thinking skills, and of integrating and synthesising learning and knowledge from a range of sources and environments. A characteristic of this form of thinking is the generation of new, professionally oriented knowledge through personal or group-based critique of practice and theory.

This graduate capability is supported by:

Learning outcomes

- Understand basic notions of probability theory.
- Familiarity with classes of discrete and continuous random variables and their distribution functions. Evaluate probabilities of events, expected values, variances and

higher moments of random variables.

- For random vectors be able to evaluate joint distributions.
- Understand basic properties of Markov Chains. Recognize Markov processes and understand how various situations in molecular genetics can be modelled by these processes.
- Recognize how basic probability theory can be used to solve particular problems encountered in analysis of DNA sequencing.

Assessment tasks

- Tests
- Final Examination
- Assignments

PG - Research and Problem Solving Capability

Our postgraduates will be capable of systematic enquiry; able to use research skills to create new knowledge that can be applied to real world issues, or contribute to a field of study or practice to enhance society. They will be capable of creative questioning, problem finding and problem solving.

This graduate capability is supported by:

Learning outcomes

- Understand basic notions of probability theory.
- Familiarity with classes of discrete and continuous random variables and their distribution functions. Evaluate probabilities of events, expected values, variances and higher moments of random variables.
- For random vectors be able to evaluate joint distributions.
- Understand basic properties of Markov Chains. Recognize Markov processes and understand how various situations in molecular genetics can be modelled by these processes.
- Recognize how basic probability theory can be used to solve particular problems encountered in analysis of DNA sequencing.

Assessment tasks

- Tests
- Final Examination
- Assignments

PG - Effective Communication

Our postgraduates will be able to communicate effectively and convey their views to different social, cultural, and professional audiences. They will be able to use a variety of technologically supported media to communicate with empathy using a range of written, spoken or visual formats.

This graduate capability is supported by:

Learning outcomes

- Understand basic notions of probability theory.
- Familiarity with classes of discrete and continuous random variables and their distribution functions. Evaluate probabilities of events, expected values, variances and higher moments of random variables.
- For random vectors be able to evaluate joint distributions.
- Understand basic properties of Markov Chains. Recognize Markov processes and understand how various situations in molecular genetics can be modelled by these processes.

Assessment tasks

- Tests
- Final Examination
- Assignments

PG - Engaged and Responsible, Active and Ethical Citizens

Our postgraduates will be ethically aware and capable of confident transformative action in relation to their professional responsibilities and the wider community. They will have a sense of connectedness with others and country and have a sense of mutual obligation. They will be able to appreciate the impact of their professional roles for social justice and inclusion related to national and global issues

This graduate capability is supported by:

Learning outcome

- Be familiar with basic principles of statistical data modelling.

Assessment tasks

- Tests
- Final Examination
- Assignments

PG - Capable of Professional and Personal Judgment and

Initiative

Our postgraduates will demonstrate a high standard of discernment and common sense in their professional and personal judgment. They will have the ability to make informed choices and decisions that reflect both the nature of their professional work and their personal perspectives.

This graduate capability is supported by:

Learning outcomes

- Recognize how basic probability theory can be used to solve particular problems encountered in analysis of DNA sequencing.
- Be familiar with basic principles of statistical data modelling.

Assessment tasks

- Tests
- Final Examination
- Assignments