STATS 325: Information about the Exam

The exam is three hours long. You will need a calculator. The final mark for the course is:

EITHER 15% assignments, 10% term test, 75% exam,

OR if you qualify for plussage, 100% exam, if this mark is better.

Remember that you need **70% on coursework** (assignments and test) to be considered for plussage, and you need **50% on the final exam** to pass the course.

Format and Preparation

The format of the exams is exactly the same as the format of the exams for 2006 to 2013, given out in class or available from `www.stat.auckland.ac.nz/~stats325/exams.php`. Most of the content in the exams from 2003 to 2005 is also still relevant, with a few changes. All answers to past exams are available on the website.

You should attempt **all** questions on the exam. Marks for each question part are shown on the exam paper. An Attachment with useful formulae is provided. The Attachment is copied at the back of this leaflet. You should read over the Attachment carefully to be sure of what is there.

The best possible preparation for the exam is **practice.** You should attempt as many of the past exams as you can. There are very few questions for which learning notes is helpful (or even possible), and you will find these anyway on the past papers, e.g. Exam 2008, Q. 2(a); Exam 2006, Q. 8(e). Start practising exam questions from day one, and just have your notes nearby to refer to as you need them. If you spend time reading your notes instead of practising questions, you will be at a disadvantage to other students.

Question difficulty

The past exams have 'difficulty' gradings next to the marks: *E* (easy), *M* (moderate), *H* (hard). These are my guesses, so don't worry if you disagree! The other students are more likely to agree with you than with me, and this comes out in the marking. Here are some guidelines:

1. **Easy questions.** Everybody should be able to do these. If a question marked E seems quite substantial, it probably means it is *predictable*. Watch out for predictable questions: anything that has appeared very often on the past exams is quite likely to appear on your exam too.
2. **Moderate questions.** Half of the paper is in this category. Some *M* questions border on *H*, others border on *E*, and most are in-between. You can pass the paper if you can get all of the *E* marks and a bit over half of the *M* marks.
3. **Hard questions.** These are designed to challenge students at the $A/A+$ end of the spectrum. Sometimes they require more careful thought than other questions; sometimes they just require independent thought — for example, being able to work through a question from beginning to end, without getting much guidance. Put a lot of effort into practising these if you are aiming for A or $A+$.

On the real exam, **there are no difficulty gradings.** My informal percentage gradings are: 27% *E*, 44% *M*, and 29% *H*. There are 6 questions. The harder marks come more commonly towards the end, but **there are easy and hard questions interspersed throughout the exam.**

Try to be calm, thorough and careful from the very first question. From past experience, lots of students throw away easy marks at the beginning of the exam as they settle in. If you have spare time, **check your answers**, **including** any questions you thought were easy. These are the easiest marks to pick up by checking, so they are the most useful checks to make!

Try to attempt all parts of each question. Sometimes there are easy marks available after a harder part. The marks allotted to each question give a clue about how much work is expected.

Your exam is of a similar standard to the exams from 2006 to 2013. Informally, I think the H questions on the 2012 exam were a bit too hard, and my own judgement is that your H questions are easier.

If you find the past exams hard to complete in 3 hours, you should practise several times the 'predictable' types of questions so that you are fully up to speed. Look at the model solutions on the webpage: your answers should contain about the same level of detail. If you put in much more detail, you will waste time. If you put in much less, you will be at risk of losing marks for working. People who skip too many lines of working often make mistakes, and then there is **nothing correct written down** that allows me to award part-marks — make sure you're not in this position!

If you get stuck...

If you get stuck in the middle of a question, **leave that part and keep going.** There is usually enough information given so that you can do the later parts of a question even if you don't get the answers to the earlier parts. If not, then wrong answers will be carried through, so you won't lose marks for a later part by guessing – or just making up – the answer to an earlier part and carrying on. But please be sure to **tell me** what you're doing — a quick written note, like *'Can't do (b), assume answer is 0.2'*, is fine. On the other hand, no marks can be awarded for a blank page.

If you realise something has gone wrong in a question but can't work out where, **let me know.** If you've just made a numerical or transcription error, you won't lose many (if any) marks. But you are very likely to lose marks if you fiddle your answers, or leave a ludicrous answer (e.g. a negative probability) without a note to say you realise it can't be right. Remember that Stats and Maths are about **communication** as much as thought and reasoning — you need to **communicate** with the person reading your script by using helpful words like "So..." or "But...", or "LHS = ...", which make all the difference in understanding what you are doing.

Targeted Revision for 2014

The exam revisits several of the questions/ideas covered in the Assignments and Test. Being able to reproduce the assignment questions will not be enough: you need to **understand** how the questions work so that you can apply your understanding in different (but similar) situations on the exam paper. Here is some advice for targeted revision for the 2014 exam. The material on Assignments 3 and 4 is particularly important this year.

- Test: revise everything, especially applications of the Law of Total Expectation.
- Assignments 1 and 2: check you understand everything, especially in Assignment 1.
- Assignment 3: revise the whole assignment thoroughly, including all applications of LoTE and the disease model.
- Assignment 4: revise the whole assignment thoroughly, including Total Progeny.
- Assignment 5: check you understand everything.

Remember that questions that have been common in the past exams are also likely to occur.

In 2014, there will *not* be much calculation / computation / solving simultaneous equations, but you should bring a calculator anyway. The exam tests the later chapters most heavily (chapter 4 onwards).

Revision List

There is not much in the way of factual information that you have to learn. Most of the course is about testing understanding.

Basics

1. Formula for solving a quadratic equation: $ax^2 + bx + c = 0 \;\Rightarrow\; x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.
2. Principle of mathematical induction. Use the Maths Tutor for extra practice: link from the course website.
3. Manipulation of exponents: e.g. $(e^{-x})^2 = e^{-2x}$ while $e^x e^{2x} = e^{x+2x} = e^{3x}$
4. Differentiation using the product rule and the chain rule.
5. Integration of basic functions including x^n, e^{ax}, $\frac{1}{x}$.

Probability

1. Complement: $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.
2. Conditional probability: $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
3. Bayes Rule: $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$.
4. Partition Rule: $\mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \ldots + \mathbb{P}(B \cap A_k)$ if $A_1, \ldots, A_k$ form a partition of the sample space. Equivalently, $\mathbb{P}(B) = \mathbb{P}(B \mid A_1)\mathbb{P}(A_1) + \ldots + \mathbb{P}(B \mid A_k)\mathbb{P}(A_k)$.
5. Probability of a union: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
6. Independence: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for independent events A and B.
7. Chains of events: $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_2 \cap A_1)$.

Discrete distributions

All required information about the Geometric, Binomial, and Poisson distributions is provided in the Attachment. You might need to recognise situations where a Geometric (number of failures before the first success) or Binomial (number of successes in a fixed number of trials) is required. This information is also provided (see Attachment).

Continuous distributions

Where required, the p.d.f., distribution function and mean of the Uniform and Exponential distributions will be provided. Nothing else is required.

Random variables

1. Expectation: $\mathbb{E}(X) = \sum_x x\mathbb{P}(X = x)$ when X is discrete;
   $\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$ when X is continuous.
2. Expectation of a function of X: $\mathbb{E}(g(X)) = \sum_x g(x)\mathbb{P}(X = x)$ when X is discrete;
   $\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$ when X is continuous.
3. Variance: $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ when X is continuous or discrete.
4. Conditional expectation:
   - If Y is a ***random variable***, then $\mathbb{E}(X \mid Y = y) = \sum_x x\mathbb{P}(X = x \mid Y = y)$ when X is discrete.
   - If A is an ***event***, then $\mathbb{E}(X \mid A) = \sum_x x\mathbb{P}(X = x \mid A)$ when X is discrete.
5. Probability as a conditional expectation:
   - For any ***discrete*** random variable Y,
   $$\mathbb{P}(X \in S) = \sum_y \mathbb{P}(X \in S \mid Y = y)\mathbb{P}(Y = y).$$
   - For any ***continuous*** random variable Y,
   $$\mathbb{P}(X \in S) = \int_y \mathbb{P}(X \in S \mid Y = y) f_Y(y)\, dy.$$
   - These can be summarised for ***any*** random variables X and Y as:
   $$\mathbb{P}(X \in S) = \mathbb{E}_Y \left\{\mathbb{P}(X \in S \mid Y)\right\}.$$
6. Law of Total Expectation: $\mathbb{E}(X) = \mathbb{E}_Y(\mathbb{E}(X \mid Y))$.
7. Law of Total Variance: $\text{Var}(X) = \mathbb{E}_Y(\text{Var}(X \mid Y)) + \text{Var}_Y(\mathbb{E}(X \mid Y))$.
8. Independence: for independent X and Y, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$,
   $\mathbb{E}\Big(g(X)h(Y)\Big) = \mathbb{E}\Big(g(X)\Big)\mathbb{E}\Big(h(Y)\Big)$, and $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Probability generating functions

1. Probability generating function: $G_X(s) = \mathbb{E}\left(s^X\right)$, if X is discrete and takes values 0, 1, 2, ...
2. $P(X = 0) = G_X(0)$ for any random variable X.
3. Defective random variables: X is defective (can take the value ∞) if and only if $G_X(1) < 1$. If X is defective, then $\mathbb{P}(X = \infty) = 1 - G_X(1)$.

Other information is provided on the Attachment.

Branching processes

1. $G_n(s) = G_{n-1}\Big(G(s)\Big) = G\Big(G_{n-1}(s)\Big)$, where $G_n(s)$ is the PGF of Z_n, and $G(s)$ is the PGF of family size, Y. This is the *Branching Process Recursion Formula.*
2. Probability of extinction, γ, is the smallest non-negative solution to the equation $G(s) = s$.
3. Conditions for ultimate extinction. Let $\mu = \mathbb{E}(Y)$, then extinction is certain if $\mu \leq 1$. Extinction is less than certain (but generally possible) if $\mu > 1$.

Markov chains

1. Markov property:

$$\mathbb{P}(X_{t+1} = s \mid X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0) \;=\; \mathbb{P}(X_{t+1} = s \mid X_t = s_t),$$

for all $t = 1, 2, 3, \ldots$ and for all states $s_0, s_1, \ldots, s_t, s$.

2. Distribution of X_t: if $X_0 \sim \boldsymbol{\pi}^T$, then $X_t \sim \boldsymbol{\pi}^T P^t$.

3. Elements of P^t: $\mathbb{P}(X_t = j \mid X_0 = i) = \left(P^t\right)_{ij}$.

4. Hitting probabilities: if h_{iA} is the probability of hitting set A, starting from state i, then the vector of hitting probabilities $\boldsymbol{h_A} = (h_{iA} \,:\, i \in S)$ is the minimal non-negative solution to the following first-step analysis equations:

$$h_{iA} = \begin{cases} 1 & \text{for} \quad i \in A, \\ \sum_{j \in S} p_{ij} h_{jA} & \text{for} \quad i \notin A. \end{cases}$$

In practice, it is often easy to see that $h_{kA} = 0$ for some states k by inspection.

5. Mean hitting times: if m_{iA} is the expected number of steps required to hit set A, starting from state i, then the vector of mean hitting times $\boldsymbol{m_A} = (m_{iA} \,:\, i \in S)$ is the minimal non-negative solution to the following first-step analysis equations:

$$m_{iA} = \begin{cases} 0 & \text{for} \quad i \in A, \\ 1 + \displaystyle\sum_{j \notin A} p_{ij} m_{jA} & \text{for} \quad i \notin A. \end{cases}$$

6. Communicating classes: two states are in the same class if they are each accessible from the other. A Markov chain is ***irreducible*** if there is only one class. A ***state*** is ***aperiodic*** if the greatest common divisor of return times to the state is 1. Any two states in the same class have the same period.

7. Equilibrium distribution: $\boldsymbol{\pi}$ is an equilibrium distribution for transition matrix P if $\boldsymbol{\pi}^T P = \boldsymbol{\pi}^T$ and $\sum_i \pi_i = 1$.

   Alternatively, if you can show that $X_{t+1} \sim X_t$ for some distribution of X_t, then this distribution must be an equilibrium distribution.

8. Convergence to equilibrium: if P is irreducible and aperiodic, and an equilibrium distribution exists, then the distribution of X_t converges to the equilibrium distribution as $t \to \infty$, regardless of its start state.

   If an equilibrium distribution exists but the irreducible and aperiodic criteria are **not** satisfied, X_t might or might not converge to equilibrium regardless of start state. In such a case, you will have to work by common sense and inspection.

   If X_t does converge, then the long-run probability that X_t is in a particular state is given by the equilibrium distribution. If you are asked for a long-run probability, **you must justify your answer by checking that convergence does occur.** The equilibrium distribution represents long-run probabilities only when the chain **converges** to equilibrium.

Good luck! ☺

ATTACHMENT

1. **Discrete Probability Distributions**

Distribution	$\mathbb{P}(X=x)$	$\mathbb{E}(X)$	Var(X)	PGF, $\mathbb{E}(s^X)$
Geometric(p)	pq^x (where $q=1-p$), for $x=0,1,2,\dots$	$\frac{q}{p}$	$\frac{q}{p^2}$	$\frac{p}{1-qs}$
	Number of failures before the first success in a sequence of independent trials, each with $\mathbb{P}(\text{success})=p$.			
Binomial(n,p)	$\binom{n}{x}p^x q^{n-x}$ (where $q=1-p$), for $x=0,1,2,\dots,n$.	np	npq	$(ps+q)^n$
	Number of successes in n independent trials, each with $\mathbb{P}(\text{success})=p$.			
Poisson(λ)	$\frac{\lambda^x}{x!}e^{-\lambda}$ for $x=0,1,2,\dots$	λ	λ	$e^{\lambda(s-1)}$

2. **Uniform Distribution:** $X \sim$ Uniform(a,b).
   Probability density function, $f_X(x)=\frac{1}{b-a}$ for $a<x<b$. Mean, $\mathbb{E}(X)=\frac{a+b}{2}$.

3. **Properties of Probability Generating Functions**

Definition:	$G_X(s)=\mathbb{E}(s^X)$	
Moments:	$\mathbb{E}(X)=G_X'(1)$	$\mathbb{E}\left\{X(X-1)\dots(X-k+1)\right\}=G_X^{(k)}(1)$
Probabilities:	$\mathbb{P}(X=n)=\frac{1}{n!}G_X^{(n)}(0)$	

4. **Geometric Series:** $1+r+r^2+r^3+\dots \;=\; \sum_{x=0}^{\infty} r^x \;=\; \frac{1}{1-r}$ for $|r|<1$.

   Finite sum: $\sum_{x=0}^{n} r^x = \frac{1-r^{n+1}}{1-r}$ for $r \neq 1$.

5. **Binomial Theorem:** For any $p,q \in \mathbb{R}$, and integer $n>0$, $(p+q)^n \;=\; \sum_{x=0}^{n}\binom{n}{x}p^x q^{n-x}$.

6. **Exponential Power Series:** For any $\lambda \in \mathbb{R}$, $\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}=e^{\lambda}$.