Stats 210: Information about the Exam

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

EITHER 10% assignments, 8% tutorials, 7% team sessions, 75% exam,

OR 100% exam.

You need 70% on coursework (assignments, tutorials, and team sessions) to be considered for plussage. The tutorial mark is determined by the best 8 of your 10 tutorial marks (1% each). The team mark is best 5 out of 6 quizzes, plus all 5 team tasks, rescaled to 7%.

Please check your assignment and tutorial marks on Cecil against your own records. Any mistakes: please let me know.

Special note: log = ln

In all places on the exam, the term 'log' refers to the ***natural logarithm***, also written 'ln'. This means you must use the calculator key marked 'ln'.

Format

The format of the exams is exactly the same as the format of the past exams and mock exams given out in class or available from `www.stat.auckland.ac.nz/~stats210/exams.php`

There are 100 marks on the exam. You should attempt ***all*** questions. Marks for each question part are shown on the exam paper.

An Attachment is provided. The Attachment is exactly the same as the Attachments provided on the previous exams. It gives ***only*** the probability functions, means, and variances, for the Geometric, Negative Binomial, and Hypergeometric distributions. ***You must learn the relevant quantities for Binomial, Poisson, Exponential, and Uniform distributions.*** The Attachment does ***not*** give descriptions of the different distributions: you must also learn these. Note that the Hypergeometric distribution is non-examinable this semester.

It is particularly important that you learn the following:

- Descriptions of Binomial, Geometric, and Negative Binomial distributions (e.g. number of successes out of n trials, etc.)
- Roles of the Poisson and Exponential distributions in the Poisson process.
- Binomial distribution: probability function, mean, and variance.
- Poisson distribution: probability function, mean, and variance.
- Exponential distribution: probability density function, mean, and variance.
- Uniform distribution: probability density function, and mean.
- Properties of the Normal distribution, in particular the distributions of $aX + b$ and $X + Y$ if X and Y are independent Normal random variables and a and b are constants.

The items above are all featured on the Revision List attached to this document.

The best possible preparation for the exam is practice. You should attempt as much of the past exams and mock exams as you can. All the exams on the webpage are relevant. Those from the second semester 2009, 2008, and 2006, and all the Mock exams, are particularly relevant because they were written by the same examiner as yours.

Outline of the exam

The question style is very similar to the question style on the previous exams. Your exam has seven questions in total, including some fairly short questions and some more substantial questions. The questions include:

- questions on hypothesis tests, primarily discrete distributions: about 28 marks;
- questions involving likelihood and estimators, with both discrete and continuous distributions: about 38 marks, similar to past exams.
- questions on probability: about 15 marks, similar to the past exams but **not** including the permutations/combinations questions that are on some past exams from before 2008.

It would be very helpful for you to ensure you understand the principles behind the **Catch The Doublers** task (Team Task 3). You can revise the task with answers from the webpage at `www.stat.auckland.ac.nz/~stats210/quiz.php`

Question difficulty

The exam is a mixture of questions that are easy, moderate, and hard. Everyone should be able to do the easy questions. To pass the paper, you will also need to be able to make a good attempt at the moderate questions. The hard questions are designed to challenge students at the $A/A+$ end of the spectrum.

My informal question gradings on your exam are: 26% easy, 50% medium, and 24% hard.

Note that easy marks can be interspersed with hard marks, so always try to attempt all parts of each question. Questions can be attempted in any order. ***For marking purposes, please list questions on the cover-sheet of the exam paper in numerical order (1, 2, 3, . . .) instead of the order in which you have attempted them.*** If you write them in a different order, the marker has to cross them out and rewrite them in numerical order.

Rating questions according to difficulty is an imprecise science, and you will probably see discrepancies in the difficulty ratings on the mock exams. You might also have very different ratings from mine. Don't worry if this is the case: the chances are that the other students agree with you more than they agree with me, and grading of the exam marks would take this into account. Sometimes questions are marked Hard just because they require independence of thought (little guidance given), or because they cover a topic traditionally considered hard by most students.

Revision List

The first priority is to learn the revision list below. You can get many marks on the exam just by knowing the things on this list. Moreover, you won't be able to answer many of the questions unless you have learnt these items. Test yourself against this list SEVERAL TIMES before the exam. Once or twice is not enough because you will forget under pressure.

Discrete distributions

1. **Poisson distribution.**

Notation	$X \sim \text{Poisson}(\lambda)$
Description	If told that $\{N_t\}_{t \geq 0}$ follows a Poisson process, then N_t is the number of events to occur by time t, and $N_t \sim \text{Poisson}(\lambda t)$. $X \sim \text{Poisson}(\lambda)$ is also commonly used as a subjective model unrelated to the Poisson process.
Probability function	$f_X(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (x = 0, 1, 2, \ldots)$
Mean	$E(X) = \lambda$
Variance	$\text{Var}(X) = \lambda$
Sum	If X and Y are independent, and $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu)$, then $X + Y \sim \text{Poisson}(\lambda + \mu)$.

2. **Binomial distribution.**

Notation	$X \sim \text{Bin}(n, p)$ or $X \sim \text{Binomial}(n, p)$
Description	X is the number of successes out of n independent trials, each with constant probability p of success.
Probability function	$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (x = 0, 1, \ldots, n)$
Mean	$E(X) = np$
Variance	$\text{Var}(X) = npq = np(1-p)$
Sum	If X and Y are independent, and $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$, then $X + Y \sim \text{Bin}(n + m, p)$.

3. **Geometric distribution.**

Notation	$X \sim \text{Geometric}(p)$
Description	X is the number of ***failures*** before the ***first*** success in a sequence of independent trials, each with constant probability p of success.
Probability function	Given on Attachment, but you should understand why $f_X(x) = P(X = x) = p(1-p)^x \quad (x = 0, 1, 2, \ldots)$

4. **Negative Binomial distribution.**

Notation	$X \sim \text{Negative Binomial}(k, p)$ or $X \sim \text{NegBin}(k, p)$
Description	X is the number of ***failures*** before the ***kth*** success in a sequence of independent trials, each with constant probability p of success.

Continuous distributions

1. **Exponential distribution.**

Notation	$X \sim \text{Exponential}(\lambda)$
Description	If told that $\{N_t\}_{t \geq 0}$ follows a Poisson process, then X is the *waiting time* between any two events, and $X \sim \text{Exponential}(\lambda)$. $X \sim \text{Exponential}(\lambda)$ is also commonly used as a subjective model unrelated to the Poisson process.
Probability density function	$f_X(x) = \lambda e^{-\lambda x} \quad (x > 0)$
Mean	$E(X) = 1/\lambda$
Variance	$\text{Var}(X) = 1/\lambda^2$

2. **Uniform distribution.**

Notation	$X \sim \text{Uniform}(a, b)$ or $X \sim \text{U}(a, b)$
Probability density function	$f_X(x) = \dfrac{1}{b-a} \quad (a < x < b)$
Mean	$E(X) = \dfrac{a+b}{2}$

3. **Normal distribution.**

Notation	$X \sim \text{Normal}(\mu, \sigma^2)$ or $X \sim \text{N}(\mu, \sigma^2)$
Probability density function	You do not need to learn this, but for the record: $f_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \quad (-\infty < x < \infty)$
Mean	$E(X) = \mu$
Variance	$\text{Var}(X) = \sigma^2$
Sum	If X and Y are independent, and $X \sim \text{N}(\mu_1, \sigma_1^2)$, $Y \sim \text{N}(\mu_2, \sigma_2^2)$, then $X + Y \sim \text{N}(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2)$.
Scaling	If $X \sim \text{N}(\mu, \sigma^2)$ and a and b are constants, then $aX + b \sim \text{N}(a\mu + b,\ a^2\sigma^2)$.

Probability

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

   — For all other events, $P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$.
7. Chains of events: $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_2 \cap A_1)$.

Random variables

Discrete random variable:

- Probability function: $f_X(x) = P(X = x)$.
- Distribution function: $F_X(x) = P(X \leq x) = \sum_{u \leq x} P(X = u)$.

Continuous random variable:

- Distribution function: $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)\, du$.
- Probability density function: $f_X(x) = F_X'(x)$.

Probability for continuous r.v.'s: $P(a < X < b) = \int_a^b f_X(x)\, dx$,
or $P(a < X < b) = F_X(b) - F_X(a)$.

Median: the median m of the distribution of X
satisfies $F_X(m) = 0.5$.

Expectation:

- — Discrete X: $E(X) = \sum_x x P(X = x)$
- — Continuous X: $E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$.

Expectation of $g(X)$:

- — Discrete X: $E(g(X)) = \sum_x g(x) P(X = x)$
- — Continuous X: $E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$.

Variance: $\mathrm{Var}(X) = E(X^2) - (EX)^2$ when X is continuous or discrete.

Expectation of $aX + b$: $E(aX + b) = aE(X) + b$ for any X; a and b constants.

Variance of $aX + b$: $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$ for any X; a and b constants.

Independence:	$f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all x, y, for *independent* X and Y, discrete or continuous.
Expectation of a sum:	$E(X + Y) = E(X) + E(Y)$ for *all* X and Y (discrete, continuous, independent, or non-independent).
Variance of a sum:	$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ *only* when X and Y are independent. (One mark will be deducted if this formula is used without justifying by independence.)
Expectation of a product:	$E(XY) = E(X)E(Y)$ *only* when X and Y are independent. (One mark deducted for not justifying as above.)
Change of variable formula: For X and Y continuous,	$f_Y(y) = f_X(x(y)) \left\| \frac{dx}{dy} \right\|$ if $Y = g(X)$ with g monotone. Quote range of values of y in the *final* answer.
Sample mean: For any $X_1, X_2, \ldots, X_n$:	sample mean $= \overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$.
Central Limit Theorem: For $X_1, \ldots, X_n$ independent, $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$, and for large n ($n \geq 30$):	$T = X_1 + \ldots + X_n \sim$ approx Normal$(n\mu,\ n\sigma^2)$; and $\overline{X} = \frac{1}{n}(X_1 + \ldots + X_n) \sim$ approx N$\left(\mu,\ \frac{\sigma^2}{n}\right)$.

Differentiation, integration, mathematical manipulations

You should be able to:

- differentiate $y = x^n$, $y = \frac{1}{x}$, $y = e^{ax}$, $y = \log(x)$;
- integrate $y = x^n$, $y = e^{ax}$, $y = \frac{1}{x}$;
- differentiate using the product rule and the chain rule;
- integrate by parts;
- manipulate exponents: e.g. $(e^{-x})^2 = e^{-2x}$ while $e^x e^{2x} = e^{x+2x} = e^{3x}$;
- manipulate logs: e.g. $-\log(x) = \log\left(\frac{1}{x}\right)$, $a\log(x) = \log(x^a)$, and $e^{\log x} = x$.

General notes

Here are a few general rules to follow.

1. Layout is ***important***. You ***must*** set out your answers clearly to gain marks for working. Don't stop in the middle of one calculation to start evaluating something different without telling me what you're doing.

2. The = sign has a specific meaning. Many people misuse this symbol without realising it. For example, writing

$$\begin{aligned} L(p\,;\,x) &= p(1-p)^x \\ &= (1-p)^x - xp(1-p)^{x-1} \end{aligned}$$

   is plain wrong: you have differentiated the expression on the second line. Instead, you should write:

$$\begin{aligned} L(p\,;\,x) &= p(1-p)^x \\ \frac{dL}{dp} &= (1-p)^x - xp(1-p)^{x-1}. \end{aligned}$$

   If you misuse the = sign you are sure to lose marks for working.

3. Please be honest. It will be obvious if you've gone wrong somewhere and are fiddling your solutions to get the answer required, and it will irritate me that you think I won't notice. Instead, if you realise something has gone wrong but can't work out where, just say so. If you've just made a small numerical or transcription error, you won't lose many (if any) marks. Fiddling answers will always lose marks!

4. 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. Just guess the answer to the earlier part, and carry on. As long as you tell me what you're doing, you won't lose marks for the later part.*** A short written note, like "*Can't do (b), assume answer to (b) is 2*", is fine.

5. Check that your answers make sense: e.g. probabilities always lie between 0 and 1; an answer for $E(X)$ should not involve an x; in a change of variable question, the answer for $f_Y(y)$ should involve y and not x; and so on.

6. The marks allotted for a question are a clue about how much work you need to do. If you are doing long calculations for one or two marks, you are probably doing something wrong. However, there are sometimes 2 or 3 marks available for almost no calculation, if the answer requires some careful thought.

7. Follow the wording of questions carefully. If you are asked to "State the value of ...", you should be able to state it without a calculation. If you are asked to "Find the value of ...", a calculation (or at least an explanation) is usually required. If you are told "Using the result R, show that ...", then you may start from result R without proving that it is true.

8. If you are asked to *state* or *name* a distribution, this means giving a name like Binomial, Uniform, Exponential, with appropriate parameters. It does not mean writing a formula for the pdf or distribution function.

9. Remember always to give the range of values when you quote a final answer. This is particularly important in change of variable questions, or when you have to find a distribution function or a p.d.f. Note that for distribution functions, you might need to specify

$$F_X(x) = \begin{cases} 0 & \text{for } x \leq \ldots, \\ \ldots & \text{for } \ldots < x < \ldots, \\ 1 & \text{for } x \geq \ldots. \end{cases}$$

Good luck! ☺