Information about the Exam

The exam is two hours long.

You will need a calculator.

You must score at least 50% on the exam to pass the course.

Please check your marks on CANVAS against your own records. Any mistakes:
please let me know at c.giurcaneanu@auckland.ac.nz.

Special note

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. Note also that to find the value of the constant e = 2.718…, you need to use the calculator button e^x and enter value 1 (i.e. $e = e^1$). Note that e^x and exp(x) have the same significance. Hence, e^x=exp(x).

Format

The format of the exams is similar to the format of the recent exams available on CANVAS.

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 recent past 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.

It is particularly important that you learn the following:

Descriptions of Binomial, Geometric, Negative Binomial, and Hypergeometric distributions (e.g. number of successes out of n trials, etc.)

Role of the Poisson distribution and the Exponential distribution 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.

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 recent Past Exams as you can.

Outline of the exam

You should be prepared to answer:

Questions on probability, involving manipulations

Questions involving likelihood, estimation and estimators

Questions on hypothesis tests

Revision List

The first priority is to learn the revision list overleaf. You can get 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	$f_X(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (x = 0, 1, 2, \dots)$
Mean	$E(X) = \lambda$
Variance	$\text{Var}(X) = \lambda$

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	$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (x = 0, 1, \dots, n)$
Mean	$E(X) = np$
Variance	$\text{Var}(X) = npq = np(1-p)$

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 $f_X(x) = P(X = x) = p(1-p)^x \quad (x = 0, 1, 2, \dots)$

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.

5. **Hypergeometric distribution.**

Notation	$X \sim \text{Hypergeometric}(N, M, n)$
Description	There are N objects, of which M are special. A sample of size n is drawn. X is the number of the n sampled objects that are special.

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	$\text{Var}(X) = \sigma^2$

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.

Random variables

1. Discrete random variable. Probability function: $f_X(x) = P(X = x)$.
   Distribution function: $F_X(x) = P(X \le x) = \sum_{u \le x} P(X = u)$.

2. Continuous random variable. Probability density function: $f_X(x)$.
   Distribution function: $F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(u)\, du$.
   Probability density function: $f_X(x) = F'_X(x)$.

3. Probability for continuous random variables: $P(a < X < b) = \int_a^b f_X(x)\, dx$, or $P(a < X < b) = F_X(b) - F_X(a)$.

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

5. Expectation: $E(X) = \sum_x xP(X = x)$ when X is discrete;
   $E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$ when X is continuous.

6. Expectation of a function of X: $E(g(X)) = \sum_x g(x)P(X = x)$ when X is discrete;
   $E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$ when X is continuous.

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

8. Independence: $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ for all x, y, for *independent* X and Y, discrete or continuous.

9. Expectation of a sum: $E(X + Y) = E(X) + E(Y)$ for *all* X and Y (discrete, continuous, independent, or non-independent).

10. Variance of a sum: $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ *only* when X and Y are independent.

11. Expectation of a product: $E(XY) = E(X)E(Y)$ *only* when X and Y are independent.

12. Change of variable formula: for X and Y continuous, $f_Y(y) = f_X\big(x(y)\big) \left|\frac{dx}{dy}\right|$ if $Y = g(X)$ with g monotone.

13. Central Limit Theorem: If $X_1, \ldots, X_n$ are independent and identically distributed, with mean $E(X) = \mu$, and variance $\mathrm{Var}(X) = \sigma^2$, and if n is large ($n \ge 30$ is usually safe but smaller n is often sufficient), then:

    $$T = X_1 + \ldots + X_n \sim \text{approx Normal}(n\mu,\ n\sigma^2)\,;$$

    and

    $$\overline{X} = \left(\frac{X_1 + \ldots + X_n}{n}\right) \sim \text{approx Normal}\left(\mu,\ \frac{\sigma^2}{n}\right)\,.$$

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. Most people did a good job of writing out clear working in the Term Test, so don't be the only person to hand in a sloppy exam script.

2. The = sign has a specific meaning. You will be amazed how 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, or even just make one up, and carry on. As long as you tell me what you're doing, you won't lose marks for the later part. (For example, write "*I can't do part (b) so I'll assume the answer to (b) is 2*" ...)

5. Check that your answers make sense: e.g. probabilities always lie between 0 and 1; an answer for $E(X)$ is unlikely to 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. If you are told "Using the result R, show that ...", then you may start from result R without proving that it is true. 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.

7. 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.

8. 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.

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!