MATH96046/MATH97073 Statistical Theory: Exam Note 2022-23

Revision class: 9:00-10:00 on **Wed 3 May 2023** in **Huxley 144**.

Format

The exam will be 2 hours (Y3) or 2.5 hours (Y4) and contain **four** (Y3) or **five** (Y4) questions, with Y4 containing **one** question on the mastery material. Each question is worth 20 marks, so you are graded on 80 (Y3) or 100 (Y4) marks . This equates to 1.5 min per mark so pace yourself accordingly (don't spend 20 min on a 2 mark question!). Calculators are **not** permitted (or needed).

Question style

The exam will be in a similar style to the exams from the last 2 years, and thus at a similar level to the "average" questions of the problem sets (the hard PS questions are *significantly* more difficult than the exam). Understanding the problem sets is therefore a good guide for the exam. Past papers are a good source of practice.

Closed book format: the exam will be **closed-book** unlike the previous two years. You are expected to know **definitions**, **theorems** and **concepts** from the module and to be able to **apply them**. You are **not** expected to memorize major proofs from the course and there will be no questions asking you to prove long results from memory.

The kind of information you are expected to know (and apply) is contained in the "Revision Handout" document.

You will be expected to know standard mathematical facts, for example

- $(1 + x/n)^n \to e^x$ as $n \to \infty$ for any $x \in \mathbb{R}$.
- The probability mass/density functions for the basic distributions: Bernoulli, Binomial, Poisson, Gaussian, exponential, uniform.
- Recall or quickly derive the mean and variance of the above distributions.
- Anything above the second moment (e.g. EX^3) will be either provided or assigned marks if the questions requires you to derive it.

Common hints

- If you are stuck on a question, move on and come back later.
- You can use previous questions even if you cannot answer them.
  e.g. if a question says "(b) Show that XX is true", then you can use XX in part (c) even if you cannot prove (b).
- Pay attention to the wording of the questions, e.g. "hence deduce" means use the previous part (even if you cannot prove it). "Prove" is different from "explain".
- If you cannot solve (part of) a question, write something down to show you understand what the problem is about.
  e.g. Suppose you have to find the limiting distribution of $\sqrt{n}(\bar{X}_n - \theta)$ for $X_1, \ldots, X_n \sim \text{Bernoulli}(\theta)$. If you cannot work out the variance of X_1, it is still better to write 'by the central limit theorem $\sqrt{n}(\bar{X}_n - \theta) \to^d N(0, \text{Var}_\theta(X_1))$ but I cannot work out $\text{Var}_\theta(X_1)$" than to put nothing. This can still get you partial marks.

Mastery material (Y4)

The Mastery Material will cover the Bootstrap and the Jackknife, which are two resampling methods widely used in computational statistics. The material is based on Chapter 3 of the book:

L. Wasserman, All of Nonparametric Statistics. Springer (2006).

More details can be found on the **module BB page**. For the exam, it is important to be familiar with these concepts and how to apply them. Any 'complicated' formulas that are needed will be provided in the question.