# ORF 524, Statistical Theory and Methods. (Fall 2017)

## Schedule

Monday-Wednesday 11 am to 12:20 pm, Friend 109.

**Instructor:** Samory Kpotufe.
*Office hours:* Monday 2:30 pm to 3:45 pm, Sherrerd 327.

**Assistant in Instruction:** Guillaume Martinet
*Office hours:* Wednesday 4 pm to 5:30 pm, Sherrerd library (107).

## Reading

There are no required books. However, you might find the following books useful. In particular, the first three are closest to much of the material we will cover (Mathematical Statistics). The last two are useful in the last 2-3 weeks of class (Nonparametrics). The material is pulled from all these books and other sources, so we'll regularly post class notes and other relevant reading material.

- *Statistical Inference (2nd ed.)*, by G. Casella and R. L. Berger.
- *Mathematical Statistics (2nd ed.)*, by J. Shao.
- *All of Statistics*, by L. Wasserman.
- *A Distribution-Free Theory of Nonparametric Regression*, by L. Györfi, M. Kohler, A. Krzyżak, H. Walk.
- *Introduction to Nonparametric Estimation*, by A. Tsybakov.

## Grading

**Homeworks** (70%)**.** Roughly biweekly, consisting of a number of exercises, each exercise worth 10 pts.

You may collaborate, but turn in your own written answers. You are strongly encouraged to first try doing exercises on your own. They will be designed to help you assess and cement your understanding of the material.

**Midterm** (15%)**.** This will be at a level similar to your homework exercises.

**Final homework** (15%)**.** Collaboration is not allowed on this last homework.

---

**Background.** Familiarity with basic probability concepts, e.g. common r.v.s, moments, ...

---

## Overview of topics

The course is concerned with how to assess the quality of solutions to statistical problems. The emphasis is not on developing *solutions*, but rather on developing intuition on *how good* of a solution is possible for various statistical problems, and on the nature of such solutions.

Thus, we will be developing the basic mathematical tools to answer questions about the nature and quality of solutions and what makes a statistical problem inherently hard or easy.

Much of the course will cover what's often called *Mathematical Statistics*, but also delve into the more modern subject of *Nonparametrics* towards the end.

## 1 Intro (0.5 week)

*Course overview, and basic probability tools.*

## 2 Point estimation (3-3.5 weeks)

We want to estimate a *point of interest* from data. This *point* might be a real value or vector $\theta = \theta(P)$ where $P$ is the unknown data distribution.

**Data reduction through statistics** (1 week)

**Estimation principles** (1) week

**Estimation quality**(1-1.5 week)

## 3 Inference: Confidence Regions and Hypothesis testing (2 weeks)

We want to estimate a *region* of interest from data. This region might be a subset of $\mathbb{R}$ or a high-dimensional region $R = R(P)$, where $P$ is the unknown data-distribution.

**Hypothesis testing** (1 week)

**Confidence regions** (1 week)

## 4 Large sample regimes (1.5-2 weeks)

We want to estimate or infer something from data, but $P$ is unknown; luckily we have a large sample so everything is sort of Gaussian ...?

**Random sequences and notions of convergence** (1 week)

**Large sample inference** (0.5-1 week)

## 5 Nonparametrics (3-4 weeks)

We want to estimate an *infinite* parameter $\theta$, e.g. $\theta \equiv$ a highly nonlinear function $f = f(P)$, where $P$ is the unknown data-distribution.

**Arbitrarily slow rates (a.k.a. no-free-lunch)** (0.5-1 weeks)

**Nonparametric regression and classification** (1.5 - 2 weeks)

**Nonparametric density estimation** (1 week)

Complexity regularization (if time permits).

□