Probability and Statistics

Solve every problem.

Part I: Probability

Problem 1. Suppose that a sequence $\{X_n\}$ of real-valued random variables converges to X in distribution and there are positive constants r and C such that $\mathbb{E}|X_n|^r \leq C$ for all n. Show that

$$\lim_{n \to \infty} \mathbb{E}|X_n|^s = \mathbb{E}|X|^s$$

for all $0 < s < r$.

Problem 2. Let $p(x, y)$ be the (one-step) transition function of a Markov chain on a discrete state space S and $p_n(x, y)$ be the n-step transition function. Show that for any positive integers L and N and any two states x and y we have

$$\sum_{n=L}^{N+L} p_n(x, y) \leq \sum_{n=0}^{N} p_n(y, y).$$

Problem 3. Let $\{X_n\}$ be an independent, identically distributed sequence of random variables with the symmetric Bernoulli distribution

$$\mathbb{P}\{X = 1\} = \mathbb{P}\{X = -1\} = \tfrac{1}{2}.$$

Let $S_n = \sum_{i=1}^{N} X_i$ be the partial sum. Show that for all $\alpha > \tfrac{1}{2}$,

$$\mathbb{P}\left\{\lim_{n \to \infty} \frac{S_n}{n^\alpha} = 0\right\} = 1.$$

Problem 4. Let $X^n = \{X_{ij}\}$ be an $n \times n$ random matrix whose entries are independent and identically distributed random variables with the symmetric Bernoulli distribution

$$\mathbb{P}\{X = 0\} = \mathbb{P}\{X = 1\} = \tfrac{1}{2}.$$

Let $p_n = \mathbb{P}\{\det X_N \text{ is odd}\}$. Show that $\lim_{n \to \infty} p_n > 0$.

Part II: Statistics

Problem 5. You have been asked to help design a randomized trial of a new drug, call it drug A, to be used in place of the current drug, call it drug B, for a particular medical condition. The budget is fixed to have 1000 patients treated with A and 1000 treated with drug B. The issue is how to do the allocation of patients, because we have many pre-randomization measurements on each patient, roughly 200, such as blood pressure

recordings, age, sex, and a large collection of genetics measurements. Obviously it is desirable to have the A group similar to the B group with respect to all pre-treatment covariates and non-linear functions of them that are expected to influence the effectiveness of the drugs with respect to the outcome variables.

Complete (or simple) randomization does this in expectation, but with many covariates, some covariates will not be balanced between the A and B groups in any one single randomized allocation. Standard blocking used in traditional experimental design can force balance on a few covariates, but the designer of drug A wants to have an experimental design that creates balance on many covariates, and feels that you, as a modern applied mathematician/statistician, should be able to do this.

Describe a class of methods that achieves this goal where each patient has a positive probability of receiving drug A and a positive probability of receiving drug B. Provide enough detail that you are describing an explicit algorithm.

Problem 6. You are given the results of a randomized experiment of two drugs, A and B. The experiment was not conducted in the usual way, however, but rather by allocating patients by a machine-learning algorithm under which each patient has a positive probability of receiving A and of receiving B; moreover the algorithm is completely specified and is built to create better than random balance on the covariates.

(a) Can unbiased estimates of the causal effect of drug A versus B be found, and if so, show why.

(b) Can exact small sample, non-parametric inferences for the causal effect in part (a) be derived, based solely on the randomization distribution of some statistic? For example, can we find exact significance levels under a sharp null hypothesis? If so, outline how to accomplish this goal.