MA/ACM/IDS 140C: PROBABILITY (SPRING 2025)
PROBLEM SET 1

The difficulty of the problems may vary, so try to solve as many as you can. Most of the techniques have been covered in lectures, but not all. You are encouraged to consult the list of reference materials on the course website, as well as any other relevant textbooks. However, you should not use AI chatbots such as ChatGPT. Discussion with others is allowed, but you must not share intermediate work or final solutions. Feel free to come see me if you'd like to discuss any of the problems or get some hints.

(Due: by the end of May 18)

Problem 1: Degree sequence. For the Erdős–Rényi graph $G\left(n, \frac{d}{n}\right)$ with fixed $d > 0$, show that as $n \to \infty$, the empirical distribution of the degrees converges to Poisson(d) in probability.
In other words, show that for any $k \in \mathbb{N}$,

$$\frac{1}{n}\#\{v : \deg(v) = k\} \to \frac{d^k}{k!}e^{-d},$$

in probability as $n \to \infty$.

Problem 2: Large deviation of coupon collector. Let $X_1, X_2, ...$ be i.i.d. uniformly random from $\{1, \dots, n\}$; and let T_n be the smallest time t such that each $i \in \{1, \dots, n\}$ has appeared in $X_1, \dots, X_t$. We have seen that $\frac{T_n}{n\log(n)} \to 1$ in probability, and our question is to figure out how unlikely that T_n is atypically small.
Namely, find

$$\lim_{n\to\infty} \frac{1}{n} \log(\mathbb{P}[T_n < 2n]).$$

Problem 3: Clique number. Fix $0 < a \leq 1$. For the Erdős–Rényi graph $G\left(n, n^{-a}\right)$, let C_n be its clique number, i.e., the largest integer k such that $G\left(n, n^{-a}\right)$ contains a complete graph of order k. Find the limit $\lim_{n\to\infty} C_n$.
Hint: The limit may be deterministic or random, depending on a. You may find FKG or Janson's inequality (Roch, Theorem 4.2.36) useful.

Problem 4: Summation of heavy tail. Let μ be a probability distribution on $\mathbb{R}$ with $\mathbb{E}_\mu X = 0$, and that its density satisfies $x^a \frac{1}{dx}\mu([x, x+dx]) \to 1$ as $x \to \infty$, for some $a > 3$. Let $X_1, \dots, X_n$ be i.i.d. $\sim \mu$. Denote by A_n and B_n be the largest and second largest numbers in $X_1, \dots, X_n$. Show that for any $\epsilon > 0$, as $n \to \infty$,

$$\mathbb{P}\left[B_n > \epsilon n \,\middle|\, \sum_{i=1}^n X_i > n\right] \to 0,$$

and

$$\mathbb{P}\left[A_n < (1-\epsilon)n \,\middle|\, \sum_{i=1}^n X_i > n\right] \to 0.$$

(In words, to make the summation large, typically one of the numbers make the most contribution.)

Problem 5: Random surfaces. Take a finite connected graph (V, E), and $B \subset V$. Let Ω be a finite subset of $\mathbb{R}$, and $f = \{f_v\}_{v \in B}, g = \{g_v\}_{v \in B} \in \Omega^B$, such that $f_v \le g_v$ for each $v \in B$. For each edge $(u, v) \in E$, take a convex function $\Phi_{(u,v)} : \mathbb{R} \to \mathbb{R}$. Consider two probability measures μ and ν on $\Omega^{V \setminus B}$, given by

$$\mu(h) = \frac{1}{Z_\mu} \exp\Bigg(- \sum_{(u,v) \in E: u,v \in V \setminus B} \Phi_{(u,v)}(h_u - h_v) - \sum_{(u,v) \in E: u \in B, v \in V \setminus B} \Phi_{(u,v)}(f_u - h_v) \Bigg),$$

$$\nu(h) = \frac{1}{Z_\nu} \exp\Bigg(- \sum_{(u,v) \in E: u,v \in V \setminus B} \Phi_{(u,v)}(h_u - h_v) - \sum_{(u,v) \in E: u \in B, v \in V \setminus B} \Phi_{(u,v)}(g_u - h_v) \Bigg).$$

for any $h \in \Omega^{V \setminus B}$. (Here Z_μ and Z_ν are renormalization constants.) Show that ν stochastically dominates μ.

Problem 6: Sample covariance matrix with 'not-so-high' dimension. Let X be a $p \times n$ matrix with i.i.d. Rademacher entries (i.e., each $\mathbb{P}[X_{ij} = 1] = \mathbb{P}[X_{ij} = -1] = \frac{1}{2}$, independently). Show that as $n \to \infty$ while $p/n \to 0$, there is

$$\left\| \frac{1}{n} X X^T - I_p \right\|_2 \to 0$$

in probability. Here $\| \cdot \|_2$ denotes the operator norm in $L^2(\mathbb{R}^p)$.
Hint: Use ε-net arguments. (You may want to compare this with the Marchenko-Pastur law.)

Problem 7: Cut corners in GUE. Let $Z = (Z_{ij})_{i,j=1}^n$ be the $n \times n$ GUE matrix; i.e., Z_{ii} are i.i.d. $\sim \mathcal{N}(0, 1)$ and Z_{ij} (for $i < j$) are i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$, and $Z_{ij} = \overline{Z_{ji}}$. Let its eigenvalues be $\lambda_1 \ge \cdots \ge \lambda_n$. Also let $\mu_1 \ge \cdots \mu_{n-1}$ be the eigenvalues of the upper-left $(n-1) \times (n-1)$ corner $(Z_{ij})_{i,j=1}^{n-1}$.
Find the joint probability density of $(\lambda_i)_{i=1}^n$ and $(\mu_i)_{i=1}^{n-1}$ (up to multiplying a constant factor).
Hint: Diagonalize $(Z_{ij})_{i,j=1}^{n-1} = O^* D O$ such that $D = \mathrm{diag}(\mu_1, \ldots, \mu_{n-1})$. Then show that for each $1 \le i \le n-1$,

$$|u_i|^2 = -\frac{\prod_{j=1}^n (\mu_i - \lambda_j)}{\prod_{1 \le j \le n-1; j \ne i} (\mu_i - \mu_j)},$$

where $u_1, \ldots, u_{n-1}$ are the first $n-1$ entries of the last column of $\tilde{O} Z \tilde{O}^*$, for $\tilde{O}$ being the $n \times n$ orthonormal matrix with O being its upper-left $(n-1) \times (n-1)$ corner, and $\tilde{O}_{nn} = 1$; and

$$Z_{nn} = \sum_{j=1}^n \lambda_j - \sum_{j=1}^{n-1} \mu_j.$$

These identities can be proved by computing the characteristic polynomial of Z in terms of $(\lambda_i)_{i=1}^n$ and $(\mu_i)_{i=1}^{n-1}$, $(u_i)_{i=1}^{n-1}$, Z_{nn} respectively. Then compute the Jacobian of the map $\big((u_i)_{i=1}^{n-1}, Z_{nn}\big) \mapsto (\lambda_i)_{i=1}^n$, for given $(\mu_i)_{i=1}^{n-1}$.