1. Let $0 < p < 1$ be a constant. Let $\omega_n$ denote the clique number (i.e. size of the maximum clique) in the Erdős-Rényi graph $G(n,p)$. Show that as $n \to \infty$, $\frac{\omega_n}{\log n}$ converges in probability to a constant as a function of $p$. Find the limit.

2. (Binomial vs Hypergeometric: stochastic dominance). Binomial and Hypergeometric distributions arise from sampling a finite population with and without replacements, respectively. The next two problems deal with their comparison.

Consider an urn consisting of $N$ balls in total among which $k$ are red, and $N-k$ are blue. Let $X$ denote the number of red balls obtained by sampling $n$ balls from the urn without replacements. Let $Y$ denote the number of red balls obtained by sampling $n$ balls from the urn with replacements. Then $X \sim \text{Hypergeometric}(N,k,n)$ and $Y \sim \text{Binom}(n, \frac{k}{N})$. Here $N,k,n$ are integers such that $0 \leq k \leq N$ and $0 \leq n \leq N$.

(a) For any real-valued random variable $X$ and $Y$, we say that $X$ is \textit{stochastically dominated} by $Y$, denoted by $X \preceq Y$, if $F_Y(t) \leq F_X(t)$ for every $t$, where $F_X(t) \triangleq \mathbb{P}[X \leq t]$ is the CDF of $X$. Note that this is a statement about comparing distributions, rather than random variables. Nevertheless, show that $X \preceq Y$ if and only if there exists a coupling (joint distribution) between $X$ and $Y$, that is, a probability space on which $X$ and $Y$ are defined, such that $X \leq Y$ almost surely. (Hint: how to generate random variables from uniform distribution?)

(b) Show that $\text{Bern}(p) \preceq \text{Bern}(q)$ if $p \leq q$. Describe the coupling explicitly.

(c) Show that both binomial and hypergeometric can be written as a sum of Bernoulli random variables:

$$X = X_1 + \ldots + X_n, \quad Y = Y_1 + \ldots + Y_n$$

where $X_i$’s are $Y_i$’s are distributed as $\text{Bern}(\frac{k}{N})$, and $Y_i$’s are independent.

(d) Show that

$$\text{Hypergeometric}(N,k,n) \preceq \text{Binom} \left( n, \frac{k}{N-n} \right).$$

(Hint: use part (b) and consider the conditional law of $X_t$ given $X_1, \ldots, X_{t-1}$.)

3. (Binomial vs Hypergeometric: convex ordering).
(a) For any real-valued random variable $X$ and $Y$, we say that $X$ is dominated by $Y$ in the convex ordering, denoted by $X \leq_{cvx} Y$, if
\[
\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]
\]
for every convex function $f$. Again, this is a statement about comparing distributions, rather than random variables. Nevertheless, show that $X \leq_{cvx} Y$ if there exists a coupling between $X$ and $Y$, such that
\[
\mathbb{E}[Y|X] = X, \text{ a.s.}
\]
(b) Next we construct such a coupling for binomial and hypergeometric distributions. If you can construct another coupling that works, you can skip these two parts.
Show that one can simulate sampling with replacements from sampling without replacements as follows: In the context of (1), show that one can generate $(Y_1, \ldots, Y_n)$ from $(X_1, \ldots, X_n)$ by resampling
\[
Y_i = \begin{cases} 
X_i, & \text{with probability } 1 - \frac{i-1}{k}, \\
X_m, & \text{with probability } \frac{1}{k}, \quad m = 1, \ldots, i-1.
\end{cases}
\]
In other words, show that $(Y_1, \ldots, Y_n)$ defined in (4) are indeed iid $\text{Bern}(\frac{k}{N})$.
(c) Use (b) to construct an explicit coupling between $X \sim \text{Hypergeometric}(N, k, n)$ and $Y \sim \text{Binom}(n, \frac{k}{N})$, such that (3) holds, thereby proving Hoeffding’s inequality:
\[
\text{Hypergeometric}(N, k, n) \leq_{cvx} \text{Binom}\left(n, \frac{k}{N}\right)
\]
(Hint: To make the coupling symmetric in $Y_1, \ldots, Y_n$, randomize their ordering.)
(d) Invoke Hoeffding’s inequality to compare the variance: $\text{Var}(X) \leq \text{Var}(Y)$.
(e) Invoke Hoeffding’s inequality to show that hypergeometric distribution satisfies the same binomial tail bound:
\[
\mathbb{P}\left[|X - \frac{nk}{N}| \geq \epsilon\right] \leq 2 \exp\left(-\frac{2\epsilon^2}{n}\right), \quad \epsilon > 0.
\]
4. (Weyl’s inequality)

(a) Prove the following Courant-Fischer’s variational representation of eigenvalues: Let $A$ be an $n \times n$ real symmetric matrix, with eigenvalues $\lambda_1(A) \geq \ldots \geq \lambda_n(A)$. Then for each $i \in [n],$
\[
\lambda_i(A) = \sup_{\dim(V) = i} \inf_{v \in V : \|v\|_2 = 1} v^\top Av.
\]
(Hint: use the EVD of $A$).
(b) Prove that: if $A, B$ are both real symmetric, then
\[
|\lambda_i(A) - \lambda_i(B)| \leq \|A - B\|_{op}
\]
\[\text{Is the “only if” part also correct?} \]