Homework 3
Probability: a Graduate Course

---

1. Self-normalized sums

Let $X_1, \dots, X_n$ be identically distributed random variables that take only positive values. Consider the partial sums

$$S_m := X_1 + X_2 + \ldots + X_m.$$

(a) Assume that the random variables X_i are independent. Show that

$$\mathbb{E}[S_m/S_n] = m/n \quad \text{for all } m \le n.$$

(b) Show by example that without independence assumption, the result of part (a) may fail.

2. Characterization of independence

Suppose the joint density $f(x_1, \dots, x_n)$ of random variables $X_1, \dots, X_n$ can be factored as

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

for some measurable functions $f_i : \mathbb{R} \to [0, \infty)$. Prove that $X_1, \dots, X_n$ are independent. (Note that the functions f_i are not assumed to be probability densities.)

3. Binomial coefficients, simplified

Prove that the partial sums of binomial coefficients

$$\binom{n}{\le m} := \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{m}.$$

satisfy

$$\left(\frac{n}{m}\right)^m \le \binom{n}{m} \le \binom{n}{\le m} \le \left(\frac{en}{m}\right)^m.$$

for integers $1 \le m \le n$.

Hint: To prove the upper bound, multiply both sides by the quantity $(m/n)^m$, replace this quantity by $(m/n)^k$ in the left side, and use the binomial theorem. To prove the lower bound, use the definition of the binomial coefficient to express it as a product of m fractions; check that each fraction is lower bounded by n/m.

4. Stochastic dominance

We say that random variable X stochastically dominates random variable Y, denoted $X \succeq Y$, if

$$\mathbb{P}\{X \geq t\} \geq \mathbb{P}\{Y \geq t\} \quad \text{for all } t \in \mathbb{R}. \tag{1}$$

Show that $X \succeq Y$ if and only if

$$\mathbb{E}\, f(X) \geq \mathbb{E}\, f(Y) \quad \text{for all nondecreasing measurable functions } f : \mathbb{R} \to \mathbb{R}. \tag{2}$$

Hint for the direction (1) $\Rightarrow$ (2): First prove this for $f(x) = x$ and for nonnegative random variables X and Y. Extend this for arbitrary random variables by decomposing into positive and negative part. Finally, note that $X \succeq Y$ implies $f(X) \succeq f(Y)$.

5. Small world

Consider an Erdös-Rényi random graph $G_n \sim G(n, p_n)$. Show that if $2\sqrt{\ln(n)/n} < p_n < 0.99$ then the diameter[1] of G equals 2 with probability that converges to 1 as $n \to \infty$.

6. Group testing

Imagine we need to test n people for a rare disease, which affects every person independently with probability p. Instead of testing everyone individually, we can mix the samples obtained from any group of k people and test the mix. Such test will be positive if and only if at least one person in that group is sick.

Describe a procedure that allows us to determine the health status of each of the n people by using only $Cn\sqrt{p}$ tests on average, and by testing each person at most twice.[2]

Hint: test a mix of k samples. If positive, test each of these k people individually. Move on to the next group of k people. Compute the expected total number of tests. Optimize in k.

7. Critical random graphs may have isolated vertices

(a) Let X be a nonnegative random variable. Prove that

$$\mathbb{P}\{X > 0\} \geq \frac{(\mathbb{E}\, X)^2}{\mathbb{E}[X^2]},$$

as long as the denominator is nonzero.

---

[1] The distance between a pair of vertices u, v is defined as the smallest number of edges of a path that connects u and v. The diameter of G is defined as the largest distance between a pair of vertices.

[2] Here C is an absolute constant of your choice. For example, the solution is acceptable if you prove the conclusion with $C = 100$. I think the optimal constant is $C = 2$.

(b) Show that an Erdös-Rényi random graph $G_n \sim G(n, p_n)$ with $p_n = \ln(n)/n$ satisfies

$$\liminf_{n\to\infty} \mathbb{P}\left\{G \text{ has an isolated vertex}\right\} > 0.$$

8. Extremal variance

(a) Prove that any random variable X that takes values in the interval $[0,1]$ satisfies

$$\mathrm{Var}(X) \le \frac{1}{4}.$$

(b) Prove that if an equality holds above, then X must have Bernoulli distribution with parameter $1/2$.

Hint for part (a): note that $X(1-X) \ge 0$ and take expected value on both sides.

9. Order statistic

Let $X_1, \ldots, X_n$ be independent random variables that are uniformly distributed in $[0,1]$. Let $X_{(k)}$ denote a k-th smallest among them. Show that

$$\mathbb{E}\, X_{(k)} = \frac{k}{n+1}.$$

10. Cauchy distribution

Let X and Y be independent $N(0,1)$ random variables. Prove that $\mathbb{E}[X/Y]$ does not exist.