Math 734 – Spring 2024

Homework 1

Due: Wed, Feb. 28, 10PM CT. Homework should be submitted as a single PDF file via Canvas. Please read the additional instructions in the syllabus. Late homework will **not** be accepted.

[Durrett] refers to the course textbook Richard Durrett: Probability: Theory and Examples, 5th edition, 2019

Exercise 1 (kth hitting time)**.** Let $(X_n)_{n\geq 0}$ be a stochastic process defined on a measurable space $(\mathscr{S}, \mathscr{G})$ adapted to a fltration $(\mathscr{F}_n)_{n\geq 0}$. Let $A \subseteq \S$ be a measurable subset of the state space. For each $k \geq 1$, let $T_A^{(k)}$ denote the kth time that the process X_n visits some state in A. That is,

$$T_A^{(m)} = \begin{cases} \inf\{n > T_A^{(m-1)} : X_n \in A\} & \text{if } T_A^{(m-1)} < \infty \\ \infty & \text{otherwise.} \end{cases}$$

Show that $T_A^{(k)}$ is a stopping time for all $k \geq 1$.

Exercise 2 (Casino always win)**.** Let $X = (X_n)_{n\geq 0}$ be a supermartingale w.r.t. a filtration $\mathscr{F}_n$ and let $H = (H_n)_{n\geq 1}$ be any predictable sequence w.r.t. $(\mathscr{F}_n)_{n\geq 1}$. Suppose that H_n is bounded and nonnegative for $n \geq 1$. Show that $\int_0^n H\, dX$ is a supermartingale w.r.t. $\mathscr{F}_n$. (*Hint:* Mimic the proof of Theorem 5.2.18.) Also show the similar results for submartingales and martingales. (For the martingale case, it holds without assuming $H_n \geq 0$.)

Exercise 3. Find an instance of martingale converging in probability but not almost surely.

Exercise 4. Use your favorate programming language (e.g., python, R, matlab, C++) and reproduce plots similar to the ones in Figure 5.3.1.

Exercise 5 (A variational Jensen's inequality)**.** Let X be a mean zero RV taking values from an interval $[-A, B]$. Fix a convex function $\varphi : \mathbb{R} \to \mathbb{R}$. We will show that

$$\mathbb{E}[\varphi(X)] \leq \varphi(-A)\frac{B}{A+B} + \varphi(B)\frac{A}{A+B}. \tag{1}$$

In words, over all possible distributions of X over $[-A, B]$, the most extreme distribution that maximizes $\mathbb{E}[\varphi(X)]$ is the one that puts point mass on $-A$ and B as in the right-hand side.

(i) Let Y be a RV taking values from $[0,1]$ and mean $p \in [0,1]$. Suppose that for any convex function $\psi : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}[\psi(Y)] \leq (1-p)\psi(0) + p\psi(1). \tag{2}$$

Then deduce (1) from this. (*Hint:* Rescale X and make appropriate change to φ.)

(ii) Here we will deduce (2). Let Y be as before. Let $U \sim \text{Uniform}(0,1)$ independent from Y. Argue that

$$\mathbf{1}(U \leq Y) \mid Y \sim \text{Bernoulli}(Y) \quad \text{and} \quad \mathbf{1}(U \leq Y) \sim \text{Bernoulli}(p).$$

(You may use Ex. 5.1.14 for the first part.) Then use Jensen's inequality to deduce

$$(1-p)\varphi(0) + p\varphi(1) = \mathbb{E}[\varphi(\mathbf{1}(U \leq Y))] \geq \mathbb{E}[\varphi(Y)].$$

(iii) (Hoeffding's lemma) Let $\varphi(x) = e^{\theta x}$ for a fixed $\theta > 0$ and assume $A = B > 0$. Deduce that

$$\mathbb{E}[\exp(\theta X)] \leq \frac{\mathbb{E}[\exp(-\theta A)] + \mathbb{E}[\exp(\theta A)]}{2} \leq \exp(\theta^2 A^2/2).$$

Exercise 6 (Number of triangles in $G(n,p)$)**.** Let $T = T(n,p)$ denote the total number of triangles in $G(n,p)$.

(i) For each three distinct nodes i, j, k in G, let $Y_{ijk} := \mathbf{1}(ij, jk, ki \in E)$, which is the indivator variable for the even that there is a triangle with node set $\{i, j, k\}$. Show that

$$Y_{ijk} \sim \text{Bernoulli}(p^3).$$

(ii) Show that we can write

$$T = \sum_{1 \le i < j < k \le n} \mathbf{1}(ij, jk, ki \in E). \tag{3}$$

Deduce that the expected number of triangles is

$$\mathbb{E}[T] = \binom{n}{3} p^3.$$

(iii) Show that

$$\text{Var}(T(n,p)) = \binom{n}{3}(p^3 - p^6) + 12\binom{n}{4}(p^5 - p^6) \sim \frac{n^4}{2}(p^5 - p^6).$$

(*Hint:* First compute $\mathbb{E}[T^2]$ and use the fact that $\text{Var}(T) = \mathbb{E}[T^2] - \mathbb{E}[T]^2$. For computing $\mathbb{E}[T^2]$, use (3) and consider possible cases according to the number of overlapping edges.) Thus $\text{Std}(T(n,p)) = \Theta(n^2)$. If CLT holds for $T(n,p)$, then $T(n,p)$ should fluctuate around its mean by $\Theta(n^2)$. Can we conclude this by CLT?

(iv) Show that for each $t \ge 0$,

$$\mathbb{P}\left(\left|T(n,p) - \binom{n}{3}p^3\right| \ge t\right) \le 2\exp\left(-\frac{t^2}{n(n-1)(n-2)^2}\right).$$

Deduce that the above probability is $o(1)$ if $t \gg n^2$. Specifically, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left|T(n,p) - \binom{n}{3}p^3\right| \ge n^{2+\varepsilon}\right) \le 2\exp\left(-n^{2\varepsilon}\right).$$

Thus, McDirmid's inequality almost confirms the upper tail of fluctuation of $T(n,p)$ predicted by CLT. (*Hint:* Let $X_1, \dots, X_{\binom{n}{2}}$ denote the indicator of there being an edge for the kth pair of distinct nodes. Let $f(X_1, \dots, X_{\binom{n}{2}})$ denote the number of triangles using the edges indicated by X_ks. Consider the "edge exposure filtration" $(\mathscr{F}_n)_{0 \le n \le \binom{n}{2}}$, where we reveal the connectedness of every pair of distinct nodes (i, j) sequentially. Argue that there at most $n-2$ triangles that contains a given edge. Then use Theorem 5.4.3.)

Exercise 7 (Durrett)**.** 4.3.1, 4.3.2, 4.3.3, 4.3.4