Exercise 6
Probability Theory 2020 Autumn

Hanmo Chen
2020214276

December 8, 2020

1 Problem 1

$$P(S_i \geqslant 0, \forall 1 \leqslant i \leqslant 2n | S_{2n} = 0) = \frac{P(S_i \geqslant 0, \forall 1 \leqslant i \leqslant 2n-1, S_{2n} = 0)}{P(S_{2n} = 0)} \tag{1}$$

Consider the event $\{S_i \geqslant 0, \forall 1 \leqslant i \leqslant 2n-1, S_{2n} = 0\}$ partitioned by its first return to zero,

$$\{S_i \geqslant 0, \forall 1 \leqslant i \leqslant 2n-1, S_{2n} = 0\} = \bigcup_{j=1}^{n} \{S_i > 0, \forall 1 \leqslant i \leqslant 2j-1, S_{2j} = 0, S_i \geqslant 0, \forall 2j+1 \leqslant i \leqslant 2n-1, S_{2n} = 0\} \tag{2}$$

Denote $p_{2n} = P(S_i \geqslant 0, \forall 1 \leqslant i \leqslant 2n-1, S_{2n} = 0)$ and $q_{2n} = P(S_i > 0, \forall 1 \leqslant i \leqslant 2n-1, S_{2n} = 0) = \frac{1}{2}P(S_i \neq 0, \forall 1 \leqslant i \leqslant 2n-1, S_{2n} = 0) = \frac{1}{2}f_{2n}$,

$$p_{2n} = \sum_{i=1}^{n} p_{2n-2i} q_{2i} \tag{3}$$

Define $P(z) = \sum_{n=0}^{\infty} p_{2n} z^n, Q(z) = \sum_{n=1}^{\infty} q_{2n} z^n = \frac{1}{2}F(z)$

$$P(z) = \frac{1}{1 - Q(z)} = \frac{2}{1 + \sqrt{1-z}} = \frac{2}{z}(1 - \sqrt{1-z}) = \frac{2}{z}F(z) \tag{4}$$

So $p_{2n} = 2f_{2n+2} = \frac{2}{2n+1}\binom{2n+2}{n+1}\frac{1}{2^{2n+2}}$. Thus,

$$P(S_i \geqslant 0, \forall 1 \leqslant i \leqslant 2n | S_{2n} = 0) = \frac{p_{2n}}{u_{2n}} = \frac{1}{n+1} \tag{5}$$

2 Problem 2

$L_{\max}^{(n)}$ is the longest length of success runs in $X_1, X_2, \cdots, X_n$.[1]

[1] Reference: Erdös, P., & Rényi, A. (1970). On a new law of large numbers. Journal d'analyse mathématique, 23(1), 103-111.

Denote $S_0 = 0, S_n = \sum_{i=1}^{n} X_i$ and define

$$\Theta(n,k) = \max_{0 \leqslant i \leqslant n-k} \frac{S_{i+k} - S_i}{k} \tag{6}$$

So

$$L_{\max}^{(n)} \geqslant k \Longleftrightarrow \Theta(n,k) \geqslant 1 \tag{7}$$

$$P(\frac{S_{i+k} - S_i}{k} \geqslant 1) = p^k = e^{k \log p} \tag{8}$$

And

$$\begin{aligned} P(L_{\max}^{(n)} \geqslant k) = P(\Theta(n,k) \geqslant 1) = P\left(\max_{0 \leqslant i \leqslant n-k} \frac{S_{i+k} - S_i}{k} \geqslant 1\right) \\ = 1 - P\left(\max_{0 \leqslant i \leqslant n-k} \frac{S_{i+k} - S_i}{k} < 1\right) = 1 - P\left(\frac{S_{i+k} - S_i}{k} < 1, 0 \leqslant i \leqslant n-k\right) \end{aligned} \tag{9}$$

Now we are trying to find the upper and lower bound of $P\left(\frac{S_{i+k}-S_i}{k} < 1, 0 \leqslant i \leqslant n-k\right)$. First,

$$\begin{aligned} P\left(\frac{S_{i+k} - S_i}{k} < 1, 0 \leqslant i \leqslant n-k\right) &\leqslant P\left(\frac{S_{(i+1)k} - S_{ik}}{k} < 1, i = 0, 1 \cdots [\frac{n-k}{k}]\right) \\ &\leqslant (1 - e^{k \log p})^{n/k} = (1 - n^{c \log p})^{n/k} \\ &\leqslant \exp\left(-\frac{n^{1 - c \log \frac{1}{p}}}{c \log n}\right) \end{aligned} \tag{10}$$

Then denote A_i as the event $\left\{\frac{S_{i+k}-S_i}{k} < 1\right\}$

$$\begin{aligned} P\left(\frac{S_{i+k} - S_i}{k} < 1, 0 \leqslant i \leqslant n-k\right) &= P(\bigcap_{i=0}^{n-k} A_i) = 1 - P(\bigcup_{i=0}^{n-k} \overline{A_i}) \\ &\geqslant 1 - \sum_{i=1}^{n} P(\overline{A_i}) = 1 - np^k = 1 - n^{1 - c \log \frac{1}{p}} \end{aligned} \tag{11}$$

If $c < \frac{1}{\log \frac{1}{p}}$,

$$\lim_{n \to \infty} P\left(\frac{S_{i+k} - S_i}{k} < 1, 0 \leqslant i \leqslant n-k\right) \leqslant \lim_{n \to \infty} \exp\left(-\frac{n^{1 - c \log \frac{1}{p}}}{c \log n}\right) = 0 \tag{12}$$

Thus $\forall c < \frac{1}{\log \frac{1}{p}}, \lim_{n \to \infty} P(L_{\max}^{(n)} \geqslant c \log n) = 1$

If $c > \frac{1}{\log \frac{1}{p}}$,

$$\lim_{n \to \infty} P\left(\frac{S_{i+k} - S_i}{k} < 1, 0 \leqslant i \leqslant n-k\right) \geqslant \lim_{n \to \infty} 1 - n^{1 - c \log \frac{1}{p}} = 1 \tag{13}$$

Thus $\forall c > \frac{1}{\log \frac{1}{p}}$, $\lim_{n\to\infty} P(L_{\max}^{(n)} \geqslant c\log n) = 0$. $f(p) = \frac{1}{\log \frac{1}{p}}$.

3 Problem 3

3.1 (i)

Note that $P(N_m \geqslant 1)$ is the probability X_i reaches m before returning to zero.

First we consider the probability a random walk staring from k reaches 0 before to N, which is the gambler's ruin probability, and we denote it as p_k. It is easy to see that

$$p_k = \frac{1}{2}(p_{k-1} + p_{k+1}) \tag{14}$$

And $p_0 = 1, p_N = 0$. So $p_k = \frac{N-k}{N}$

In our case, because $m > 0$, X_1 must be 1, starting from $X_1 = 1$, the probability X_i reaches m before to 0 is $\frac{1}{m}$, so

$$P(N_m \geqslant 1) = P(X_1 = 1)P(N_m \geqslant 1 \mid X_1 = 1) = \frac{1}{2m} \tag{15}$$

3.2 (ii)

From (i) we know that $P(N_m = 0) = 1 - \frac{1}{2m}$.

The event m is visit n times before returning to 0 can be splits into some phases,

1. Starting from 0, visit m before returning to 0
2. Starting from m, return to m before visiting 0 (Repeat $n-1$ times)
3. Starting from m, visit 0 before returning to m

And starting from m, return to m before visiting 0 is the same as starting from 0, return to 0 before visiting $-m$. And according to symmetry, the distribution of N_{-m} is the same with N_m

$$P(N_m = n) = P(N_m \geqslant 1)P(N_m = 0)^{n-1}P(N_m \geqslant 1) = \frac{1}{4m^2}\left(1 - \frac{1}{2m}\right)^{n-1} \tag{16}$$

4 Problem 4

4.1 (i)

If the passengers arrives after T and waits until forever, then $\mathbb{E}[W] = \infty$. In the following we assume no passenger comes after T.

Suppose there are $N(T)$ passengers waiting when the train arrives at T, and the arrival time is $S_1, S_2, \cdots, S_{N_T}$, the total waiting time,

$$W = \sum_{i=1}^{N(t)} T - S_i = \tag{17}$$

Due to the uniformity of previous arrival times,

$$\mathbb{E}[W|N(t)=n] = \mathbb{E}[\sum_{i=1}^{n} T - S_i|N(t)=n] = nT - \mathbb{E}[\sum_{i=1}^{n} U_i|N(t)=n] \tag{18}$$

where $U_i \sim \text{Uniform}[0,T], i = 1, 2, \cdots, n$. Thus,

$$\mathbb{E}[W|N(t)=n] = \frac{nT}{2} \tag{19}$$

So,

$$\mathbb{E}[W] = \mathbb{E}[\mathbb{E}[W|N(t)=n]] = \frac{T}{2}\,\mathbb{E}[N(T)] = \frac{\lambda T^2}{2} \tag{20}$$

4.2 (ii)

Consider the two independent Poisson process in the time interval $[0,S]$ and $[S,T]$ with the same rate λ, applying the conclusion in (i),

$$\mathbb{E}[W] = \frac{\lambda S^2}{2} + \frac{\lambda (T-S)^2}{2} \tag{21}$$

5 Problem 5

In lecture 12, we have

$$\lim_{n\to\infty} P(\text{G is connnected}) = \lim_{n\to\infty} P(\text{G has no iso vertices}) = \lim_{n\to\infty} P(Z_{iso} = 0) \tag{22}$$

Also,

$$\mathbb{E}\binom{Z_{iso}}{r} = \binom{n}{r}(1-p)^{r(n-1)-\binom{r}{2}} \sim \frac{1}{r!}\frac{n!}{(n-r)!}\exp\left(-p(rn - \frac{r^2}{2})\right) \sim \frac{1}{r!}\frac{n!}{(n-r)!n^r}e^{-cr} \tag{23}$$

$$\mathbb{E}\binom{Z_{iso}}{r} \sim \frac{1}{r!}e^{-cr} \tag{24}$$

So Z_iso converges in distribution to a Poisson(e^{-c}) random variable.

$$\lim_{n\to\infty} P(\text{G is connnected}) = \lim_{n\to\infty} P(Z_{iso} = 0) = \exp\left(-e^{-c}\right) \tag{25}$$

6 Problem 6

Denote Y_n be the number of K_4 in G.

$$Y_n = \sum_{T \in \binom{[n]}{4}} \mathbb{1}(T \in G) \tag{26}$$

Obviously,

$$\mathbb{E}[Y_n] \leqslant n^4 p^6 = (n^{\frac{2}{3}-\delta})^6 \tag{27}$$

So if $\delta > \frac{2}{3}$,

$$\lim_{n\to\infty} P(\text{G contains 4 vertices that are pairwise connected}) = \lim_{n\to\infty} P(Y_n \geqslant 1) \leqslant \lim_{n\to\infty} \mathbb{E}[Y_n] = 0 \tag{28}$$

Then we consider $\mathrm{Var}(Y_n)$

$$\mathrm{Var}\,(Y_n) = \sum_{S,T \in \binom{[n]}{4}} \mathrm{Cov}(\mathbb{1}(S \in G), \mathbb{1}(T \in G)) \tag{29}$$

Consider $|S \cap T|$,

1. $|S \cap T| \leqslant 1$: $\mathrm{Cov}(S \cap T) = 0$
2. $|S \cap T| = 2$: There are $\leqslant \binom{n}{6}\binom{6}{2}\binom{4}{2}$ possible combinations of (S, T), and
$$\mathrm{Cov}(\mathbb{1}(S \in G), \mathbb{1}(T \in G)) \leqslant p^{11}$$
3. $|S \cap T| = 3$: There are $\leqslant \binom{n}{5}\binom{5}{3}\binom{2}{1}$ possible combinations of (S, T), and
$$\mathrm{Cov}(\mathbb{1}(S \in G), \mathbb{1}(T \in G)) \leqslant p^{9}$$
4. $|S \cap T| = 4, S = T$, There are $\binom{n}{4}$ possible S, and $\mathrm{Var}(\mathbb{1}(S \in G)) \leqslant p^6$

Combining these cases,

$$\mathrm{Var}\,(Y_n) \leqslant n^4 p^6 + n^5 p^9 + n^6 p^{11} = n^{4-6\delta} + n^{5-9\delta} + n^{6-11\delta} \tag{30}$$

And

$$\frac{\mathrm{Var}(Y_n)}{(\mathbb{E}[Y_n])^2} \leqslant \frac{n^{4-6\delta} + n^{5-9\delta} + n^{6-11\delta}}{\binom{n}{4} p^6} \sim n^{6\delta-4} o(1) \tag{31}$$

So if $\delta < \frac{2}{3}$,

$$\lim_{n\to\infty} P(Y_n = 0) \leqslant \lim_{n\to\infty} \frac{\mathrm{Var}(Y_n)}{(\mathbb{E}[Y_n])^2} = 0 \tag{32}$$

$$\lim_{n\to\infty} P(\text{G contains 4 vertices that are pairwise connected}) = 1 \tag{33}$$

So $\delta_0 = \frac{2}{3}$