

# Exercise 6

## Probability Theory 2020 Autumn

Hanmo Chen  
2020214276

December 8, 2020

### 1 Problem 1

$$P(S_i \geq 0, \forall 1 \leq i \leq 2n | S_{2n} = 0) = \frac{P(S_i \geq 0, \forall 1 \leq i \leq 2n-1, S_{2n} = 0)}{P(S_{2n} = 0)} \quad (1)$$

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

$$\{S_i \geq 0, \forall 1 \leq i \leq 2n-1, S_{2n} = 0\} = \bigcup_{j=1}^n \{S_i > 0, \forall 1 \leq i \leq 2j-1, S_{2j} = 0, S_i \geq 0, \forall 2j+1 \leq i \leq 2n-1, S_{2n} = 0\} \quad (2)$$

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

$$p_{2n} = \sum_{i=1}^n p_{2n-2i} q_{2i} \quad (3)$$

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

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

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

$$P(S_i \geq 0, \forall 1 \leq i \leq 2n | S_{2n} = 0) = \frac{p_{2n}}{u_{2n}} = \frac{1}{n+1} \quad (5)$$

### 2 Problem 2

$L_{\max}^{(n)}$  is the longest length of success runs in  $X_1, X_2, \dots, X_n$ .<sup>1</sup>

---

<sup>1</sup>Reference: Erdős, 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

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

So

$$L_{\max}^{(n)} \geq k \iff \Theta(n, k) \geq 1 \quad (7)$$

$$P\left(\frac{S_{i+k} - S_i}{k} \geq 1\right) = p^k = e^{k \log p} \quad (8)$$

And

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

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

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

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

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

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

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

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

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

$$\lim_{n \rightarrow \infty} P\left(\frac{S_{i+k} - S_i}{k} < 1, 0 \leq i \leq n-k\right) \geq \lim_{n \rightarrow \infty} 1 - n^{1-c \log \frac{1}{p}} = 1 \quad (13)$$

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

### 3 Problem 3

#### 3.1 (i)

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

First we consider the probability a random walk starting 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

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

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

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

#### 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  $N_m$

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

### 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  $S_1, S_2, \dots, S_{N_T}$ , the total waiting time,

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

Due to the uniformity of previous arrival times,

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

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

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

So,

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

## 4.2 (ii)

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

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

## 5 Problem 5

In lecture 12, we have

$$\lim_{n \rightarrow \infty} P(\text{G is connected}) = \lim_{n \rightarrow \infty} P(\text{G has no iso vertices}) = \lim_{n \rightarrow \infty} P(Z_{iso} = 0) \quad (22)$$

Also,

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

$$\mathbb{E}\left(\binom{Z_{iso}}{r}\right) \sim \frac{1}{r!} e^{-cr} \quad (24)$$

So  $Z_{iso}$  converges in distribution to a  $\text{Poisson}(e^{-c})$  random variable.

$$\lim_{n \rightarrow \infty} P(\text{G is connected}) = \lim_{n \rightarrow \infty} P(Z_{iso} = 0) = \exp(-e^{-c}) \quad (25)$$

## 6 Problem 6

Denote  $Y_n$  be the number of  $K_4$  in  $G$ .

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

Obviously,

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

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

$$\lim_{n \rightarrow \infty} P(\text{G contains 4 vertices that are pairwise connected}) = \lim_{n \rightarrow \infty} P(Y_n \geq 1) \leq \lim_{n \rightarrow \infty} \mathbb{E}[Y_n] = 0 \quad (28)$$

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

$$\text{Var}(Y_n) = \sum_{S, T \in \binom{[n]}{4}} \text{Cov}(\mathbf{1}(S \in G), \mathbf{1}(T \in G)) \quad (29)$$

Consider  $|S \cap T|$ ,

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

Combining these cases,

$$\text{Var}(Y_n) \leq n^4 p^6 + n^5 p^9 + n^6 p^{11} = n^{4-6\delta} + n^{5-9\delta} + n^{6-11\delta} \quad (30)$$

And

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

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

$$\lim_{n \rightarrow \infty} P(Y_n = 0) \leq \lim_{n \rightarrow \infty} \frac{\text{Var}(Y_n)}{(\mathbb{E}[Y_n])^2} = 0 \quad (32)$$

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

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