Tsinghua-Berkeley Shenzhen Institute
PROBABILITY
Spring 2021

Homework 6

Feng Zhao June 10, 2021

6.1. We first consider the probability of first return to zero for $2n+2$ steps: $f_{2n+2} = P(S_1 \neq 0, S_2 \neq 0, \dots, S_{2n+1} \neq= 0, S_{2n+2} = 0) = 2P(X_1 = 1, X_1 + X_2 \geq 1, \dots, \sum_{i=1}^{2n+1} X_i = 1, X_{2n+2} = -1) = 2P(S_1 \geq 0, S_2 \geq 0, \dots, S_{2n} = 0)$. This implies that $P(S_1 \geq 0, S_2 \geq 0, \dots, S_{2n} = 0) = \frac{1}{2} f_{2n+2} = \frac{1}{2}\frac{1}{2n+1} P(S_{2n+2} = 0)$ by the formula of f_n. Therefore, $P(S_1 \geq 0, S_2 \geq 0, \dots, S_{2n} \geq 0 | S_{2n} = 0) = \frac{1}{2}\frac{f_{2n+2}}{P(S_{2n}=0)} = \frac{1}{2}\frac{1}{2n+1}\frac{P(S_{2n+2}=0)}{P(S_{2n}=0)} = \frac{1}{2}\frac{1}{2n+1}\frac{\binom{2n+2}{n+1}}{\binom{2n}{n}} = \frac{1}{n+1}$.

6.2. We claim that $f(p) = -\frac{1}{\log p}$. Let $k = c\log n$ First notice that $P(L_i \geq k) = p^k = n^{c\log p}$. Then by union bound, $P(L_{\max}^{(n)} \geq k) \leq \sum_{i=1}^n P(L_i \geq k) = np^k = n^{1+c\log p}$. When $c > f(p)$, $1 + c\log p < 0$. Therefore, $\lim_{n\to\infty} P(L_{\max}^{(n)} \geq k) = 0$. On the other hand, we will show that $\lim_{n\to\infty} P(L_{\max}^{(n)} < k) = 0$ when $1 + c\log p > 0$.

$$
\begin{aligned}
P(L_{\max}^{(n)} < k) &= P(L_i < k, \forall 1 \leq i \leq n) \\
&\leq P(L_i < k, \text{ for } i = 1, k+1, 2k+1, \dots, k\lfloor \frac{n}{k} - 1 \rfloor + 1) \\
&\overset{(a)}{=} \prod_{j=0}^{\lfloor \frac{n}{k} - 1 \rfloor} P(L_{jk+1} < k) \\
&= (1 - p^k)^{\lfloor \frac{n}{k} - 1 \rfloor} \sim \exp(-\frac{n^{1+c\log p}}{c\log n}) \to 0
\end{aligned}
$$

where (a) comes from the independent conditions of X_i. Note: the result implies that $\frac{L_{\max}^{(n)}}{\log(n)} \xrightarrow{p} -\frac{1}{\log p}$.

We can further show that

$$
\limsup_{n\to\infty} \frac{L_n}{\log(n)} = f(p) \text{ with probability } 1 \tag{1}
$$

The above equation is guaranteed if

$$
P(L_n \geq c\log(n) \text{ occurs infinitely often}) = 0 \text{ for any constant } c > f(p) \tag{2}
$$

$$
P(L_n \geq c\log(n) \text{ occurs infinitely often}) = 1 \text{ for any constant } c < f(p) \tag{3}
$$

Notice that (1) says that $\lim_{m\to\infty} \sup_{n\geq m} \frac{L_n}{\log p} = f(p)$ with probability 1. Then for any $\epsilon > 0$, $\lim_{m\to\infty} P(|\sup_{n\geq m} \frac{L_n}{\log n} - f(p)| > \epsilon) = 0$ Let

$C_n(\epsilon) = \{w \Big| |\frac{L_n(w)}{\log n} - f(p)| > \epsilon\}$. Then $P(\limsup_{n\to\infty} C_n) = 0$. Further let $A_n(\epsilon) = \{w \Big| \frac{L_n(w)}{\log n} - f(p) > \epsilon\}$ and $B_n(\epsilon) = \{w \Big| \frac{L_n(w)}{\log n} - f(p) < -\epsilon\}$. Then $\limsup_{n\to\infty} C_n(\epsilon) \subset \limsup_{n\to\infty} A_n(\epsilon) \bigcup \liminf_{n\to\infty} B_n(\epsilon)$. Therefore, (2) and (3) imply (1).

We can use Borel Cantelli lemma to prove (2). We will show that $\sum_{n=1}^{\infty} P(L_n \geq c\log(n)) = \sum_{n=1}^{\infty} n^{c\log p} < +\infty$. Since $c\log p < -1$, the converge of the series is guaranteed.

To prove (3), we cannot use Borel Cantelli lemma directly since the events are not independent. We consider $P(\liminf_{n\to\infty} B_n(\epsilon))$ with $\epsilon = f(p) - c$. That is, we will prove $\lim_{m\to\infty} P(\cap_{n=m}^{+\infty}\{w \Big| L_n(w) < c\log n\}) = 0$. Since $P(\cap_{n=m}^{+\infty}\{w \Big| L_n(w) < c\log n\})$ is an non-decreasing sequence about m, we only need to show $P(\cap_{n=m}^{+\infty}\{w \Big| L_n(w) < c\log n\}) = 0$ for any m. Let $k = c\log n$. When we consider $n = m, m+k, m+2k, \ldots$, the event series $B_n(\epsilon)$ becomes independent. Since $P(L_n < k) = 1 - p^k$, we have $P(\cap_{n=m}^{+\infty}\{w \Big| L_n(w) < c\log n\}) \leq \prod_{t=0}^{+\infty}(1 - p^{c\log(m+tk)})$. Since $c\log p > -1$, $\sum_{t=0}^{+\infty}(m+tk)^{c\log p} = +\infty \Rightarrow \sum_{t=0}^{+\infty}\log(1 - p^{c\log(m+tk)}) = -\infty \Rightarrow \prod_{t=0}^{+\infty}(1 - p^{c\log(m+tk)}) = 0$.

6.3. (a) $P(N_m \geq 1 | X_1 = -1) = 0$ while $P(N_m \geq 1 | X_1 = 1)$ is the probability that the particle first hits m before hitting 0 starting from position 1. This probability is $\frac{1}{m}$ from the gambler's win problem with two players. Then $P(N_m \geq 1) = P(N_m \geq 1 | X_1 = 1)P(X_1 = 1) = \frac{1}{2m}$

(b) $P(N_m = n) = P(N_m = n-1) \cdot \frac{1}{2}(\frac{m-1}{m} + 1) = \frac{2m-1}{2m} P(N_m = n-1) = \frac{(2m-1)^{n-1}}{(2m)^{n+1}}$.

6.4. (a) $W = \sum_{i=1}^{N(T)}(T - T_i)$

$$\begin{aligned}\mathbb{E}[W] &= \sum_{s=1}^{+\infty} P(N(T) = s)\,\mathbb{E}[W | N(T) = s] \\ &= \sum_{s=1}^{+\infty} \frac{(\lambda T)^s e^{-\lambda T}}{s!}\left(\sum_{i=1}^{s}[T - \mathbb{E}[T_i | N(T) = s]]\right)\end{aligned}$$

Given $N(T) = s$, T_i is uniformly distributed in the interval $[0, T]$. Then $\mathbb{E}[T_i | N(T) = s] = \frac{T}{2}$, and

$$\begin{aligned}\mathbb{E}[W] &= \frac{T}{2}\sum_{s=1}^{+\infty}\frac{(\lambda T)^s e^{-\lambda T}}{(s-1)!} \\ &= \frac{\lambda T^2}{2}\end{aligned}$$

(b) Using the conclusion from (a) and the memoryless property of Poisson process, $\mathbb{E}[W] = \frac{\lambda S^2}{2} + \frac{\lambda(T-S)^2}{2}$

6.5. Since $p = (1+o(1))\frac{\log n}{n}$, the probability that G has an isolated subset with size more than 2 tends to zero as $n \to \infty$. Therefore $\lim_{n\to\infty} P(G \text{ is connected}) = \lim_{n\to\infty} P(G \text{ has no isolated vertex})$. Let $Z_{\text{iso}} = \sum_{v\in[n]} \mathbb{1}(v \text{ is isolated})$. Then $P(G \text{ has no isolated vertex}) = P(Z_{\text{iso}} = 0)$. We will show that $Z_{\text{iso}} \xrightarrow{d} \text{Pois}(e^{-c})$, and obtain $P(Z_{\text{iso}} = 0) = \exp(-e^{-c})$.

To achieve such purpose, we only need to show that $\mathbb{E}\binom{Z_{\text{iso}}}{r} \to \frac{e^{-rc}}{r!}$.

$$\begin{aligned}\mathbb{E}\binom{Z_{\text{iso}}}{r} &= \sum_{1\le v_1<v_2<\cdots<v_r\le n} P(v_i \text{ is isolated for all } i \in [r]) \\ &= \binom{n}{r}(1-p)^{r(n-1)-\binom{r}{2}} \to \frac{e^{-rc}}{r!}\end{aligned}$$

In conclusion, $\lim_{n\to\infty} P(G \text{ is connected}) = \exp(-e^{-c})$

6.6. We claim that $\delta_0 = \frac{2}{3}$. Define a set

$$\binom{\lfloor n \rfloor}{4} := \{\{i,j,k,l\} : i,j,k,l \in [n], i<j<k<l\}$$

Let Z_n be the number of 4-vertex cliques. Then

$$Z_n := \sum_{T\in\binom{\lfloor n \rfloor}{4}} \mathbb{1}(T \in G)$$

By linearity of expectation, $\mathbb{E}[Z_n] = \binom{n}{4}p^6 = O(n^{4-6\delta})$. If $\delta > \delta_0$, $\mathbb{E}[Z_n] \to 0$. By Markov's inequality, $P(Z_n \neq 0) \le \mathbb{E}[Z_n] \to 0$. Therefore, we have shown that $\lim_{n\to\infty} P(G \text{ contains 4 vertices that are pairwise connected }) = 0$ if $\delta > \delta_0$. For the other part, we use the Chebyshev's inequality for Z_n: $P(Z_n = 0) \le \frac{\text{Var}[Z_n]}{\mathbb{E}[Z_n]^2}$. Notice that

$$\text{Var}[Z_n] = \sum_{S,T\in\binom{\lfloor n \rfloor}{4}} \text{Cov}(\mathbb{1}(S\in G), \mathbb{1}(T \in G))$$

We split the summation according to $|S\cap T| = 0,1,\ldots,4$. Then we have $\text{Var}[Z_n] \le c_1n^4p^6 + c_2n^5p^9 + c_3n^6p^{11}$ where c_1, c_2, c_3 are permutation constant. Then $P(Z_n = 0) \le \frac{c_1}{n^4p^6} + \frac{c_2}{n^3p^3} + \frac{c_3}{n^2p} = c_1n^{6\delta-4} + c_2n^{3\delta-3} + c_3n^{\delta-2} \to 0$ when $\delta < \delta_0 = \frac{2}{3}$. Therefore, we have $\lim_{n\to\infty} P(G \text{ contains 4 vertices that are pairwise connected }) = 1$ if $\delta < \delta_0$.