**Yau College Math Competition 2021**

**Final Probability and Statistics**

**Individual Overall Exam Problems (May 30, 2021)**

**Problem 1**. Let $X_1, X_2, \cdots, X_n$ be independent exponential random variables with parameter 1, and $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ be their order statistics. Let $X_{(0)} = 0$.

(1) Find the joint density function of

$$Y_k = (n+1-k)(X_{(k)} - X_{(k-1)}), \qquad k = 1, 2, \cdots, n.$$

(2) Find the limit

$$\lim_{n\to\infty} \mathbb{P}\left(X_{(n)} - \ln n \le x\right).$$

(3) Find the limit

$$\lim_{n\to\infty} \int_0^\infty \mathbb{P}\left(X_{(n)} - \ln n > x\right) \mathrm{d}x.$$

**Solution**

(1) Notice that the joint density function of $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ is

$$h(x_1, \cdots, x_n) = \begin{cases} n! e^{-\sum_{i=1}^n x_i}, & \text{if} \quad x_1 \le x_2 \le \cdots \le x_n, \\ 0, & \qquad \text{otherwise.} \end{cases}$$

Let $x_0 = 0$ and define

$$y_k = (n+1-k)(x_k - x_{k-1}), \qquad k = 1, 2, \cdots, n,$$

then

$$x_k = \sum_{i=1}^k \frac{y_i}{n-i+1}, \qquad k = 1, 2, \cdots, n,$$

and the Jacobian is $1/n!$. So the density function of $Y_1, \cdots, Y_n$ is $e^{-\sum_{i=1}^n y_i}$.

(2) Since

$$\mathbb{P}(X_{(n)} \le x) = (1 - e^{-x})^n,$$

we have

$$\mathbb{P}(X_{(n)} \le x + \ln n) = \left(1 - \frac{e^{-x}}{n}\right)^n \xrightarrow{n\to\infty} e^{-e^{-x}}.$$

(3) According to the above two steps and the lack-of-memory property, we have

$$\mathbb{E}(X_{(n)}) = 1 + \frac{1}{2} + \cdots + \frac{1}{n}.$$

Consequently,

$$\lim_{n\to\infty}\int_0^\infty \mathbb{P}\left(X_{(n)}-\ln n>x\right)\mathrm{d}x=\lim_{n\to\infty}\mathbb{E}\left(X_{(n)}-\ln n\right)=\lim_{n\to\infty}\left(1+\frac{1}{2}+\cdots+\frac{1}{n}-\ln n\right)=\gamma.$$

**Problem 2**. Let $\{X_n\}_{n\geq 1}$ be i.i.d. random variables such that $\mathbb{P}(X_1=1)=1-\mathbb{P}(X_1=-1)=p>\frac{1}{2}$. Let $S_0=0$, $S_n=\sum_{i=1}^n X_i$. Define the range of $\{S_n\}_{n\geq 0}$ by $R_n=\#\{S_0,S_1,S_2,\cdots,S_n\}$, which is the number of distinct points visited by the random walk $\{S_n\}_{n\geq 0}$ up to time $n$.

(1) Prove $\mathbb{E}(R_n)=\mathbb{E}(R_{n-1})+P(S_1S_2\cdots S_n\neq 0), \quad n=1,2,\cdots.$

(2) Find $\lim_{n\to\infty}\frac{1}{n}\mathbb{E}(R_n)$.

**Solutions**

(1)

$$\begin{aligned}
P(R_n=R_{n-1}+1)&=P(S_n\notin\{S_0,S_1,\cdots S_{n-1}\})\\
&=P(S_n\neq S_0,S_n\neq S_1,\cdots,S_n\neq S_{n-1})\\
&=P(X_1+X_2+\cdots+X_n\neq 0,X_2+X_3+\cdots+X_n\neq 0,\cdots,X_n\neq 0)\\
&=P(X_1\neq 0,X_1+X_2\neq 0,\cdots,X_1+X_2+\cdots+X_n\neq 0)\quad\text{(by i.i.d)}\\
&=P(S_1S_2\cdots S_n\neq 0).
\end{aligned}$$

Thus

$$\mathbb{E}(R_n)=\mathbb{E}(R_{n-1})+P(S_1S_2\cdots S_n\neq 0).$$

(2) Using the above relation recursively, one has

$$\frac{1}{n}\mathbb{E}(R_n)=\frac{1}{n}+\frac{1}{n}\sum_{k=1}^n P(S_1S_2\cdots S_k\neq 0)\overset{n\to\infty}{\longrightarrow}P(S_k\neq 0,\forall\ k\geq 1).$$

On the other hand, according to law of large numbers,

$$\lim_{n\to\infty}\frac{S_n}{n}=2p-1>0,\qquad\text{a.s.}$$

Thus

$$\begin{aligned}
P(S_k\neq 0,\forall\ k\geq 1)&=P(S_k>0,\forall\ k\geq 1)\\
&=\lim_{n\to\infty}P(S_k>0,\ k=1,2,\cdots,n)
\end{aligned}$$

By the reflection principle,

$$P(S_k > 0,\ k = 1, 2, \cdots, n) = \frac{1}{n}\mathbb{E}(S_n \vee 0) \overset{n\to\infty}{\longrightarrow} 2p - 1.$$

Thus $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}(R_n) = 2p - 1$.