Yale University
S&DS 551, Spring 2023
Homework 2

Chang Feng (Felix) Zhou cz397

Problem 1.

First, we note that the random walk S can only return to 0 at an even time step.

We wish to count the number of binary strings of length $2n$ that is *balanced* (equal number of 0s and 1s) subject to the condition that any prefix is not balanced. This corresponds to a random walk which returns to 0 at time $2n$ but not before that.

By symmetry, it suffices to double the number of binary strings on n 1s and $n-1$ 0s such that any prefix contains strictly more 1s than 0s (the 0-th character is prepended as 0 to ensure balance). We claim that there are

$$\frac{1}{2n-1}\binom{2n-1}{n-1}$$

such strings.

Before we prove the claim. Note that if the claim holds, then we are done. This is because the desired probability is then given by

$$\frac{\frac{1}{2n-1}\binom{2n-1}{n-1}\cdot 2}{2^{2n}} = \frac{1}{2n-1}\binom{2n}{n}2^{-2n}$$

as required.

To see the claim, we observe that there is a bijection between the number of binary strings on n 1s and $n-1$ 0s such that any prefix contains strictly more 1s than 0s and the number of binary strings on n 1s and n 0s such that any prefix contains at least as many 1s as 0s. Indeed, the bijection is obtained by prepending a 0 to the $(2n-1)$-bit string. But the cardinality of the latter set is precisely given by the well-known $(n-1)$-th Catalan number

$$C_n = \frac{1}{2n-1}\binom{2n-1}{n-1} = \frac{1}{2n-1}\binom{2n-1}{n}.$$

Finally, we have

$$
\begin{aligned}
\mathbb{E}[T^\alpha] &= \sum_{n=1}^{\infty} \mathbb{P}\{T = 2n\} \cdot (2n)^\alpha \\
&= \sum_{n=1}^{\infty} \frac{(2n)^\alpha}{2n-1} \binom{2n}{n} 2^{-2n} \\
&= \sum_{n=1}^{\infty} \frac{(2n)^\alpha}{2n-1} \frac{(2n)!}{n!n!} 2^{-2n} \\
&\approx \sum_{n=1}^{\infty} \frac{1}{2n-1} \frac{(2n)^{2n+\frac{1}{2}+\alpha} e^{-2n}\sqrt{2\pi}}{\left(n^{n+\frac{1}{2}} e^{-n}\sqrt{2\pi}\right)^2} 2^{-2n} \\
&= \frac{1}{\sqrt{2\pi}} \sum_{n=1}^{\infty} \frac{1}{2n-1} \frac{(2n)^{2n+\frac{1}{2}+\alpha}}{n^{2n+1}} 2^{-2n} \\
&= \frac{2^{\frac{1}{2}+\alpha}}{\sqrt{2\pi}} \sum_{n=1}^{\infty} \frac{1}{2n-1} n^{\alpha-\frac{1}{2}}.
\end{aligned}
$$

For $\alpha < \frac{1}{2}$, this is series is bounded above by a convergent p-series

$$\sum_{n\geq 1} \frac{1}{n^p}$$

for some $p > 1$. On the other hand, for $\alpha \geq \frac{1}{2}$, this series is bounded below by the divergent harmonic series

$$\sum_{n\geq 1} \frac{1}{n}.$$

This concludes the proof.

Problem 2.

We have that

$$\begin{aligned}\mathbb{P}\{R_n = R_{n-1}+1\} &= \mathbb{P}\{S_n \neq S_{n-1}, \dots, S_n \neq S_0\} \\ &= \mathbb{P}\{S_1 \neq 0, \dots, S_n \neq 0\} \\ &= \mathbb{P}\{S_1 S_2 \dots S_n \neq 0\}.\end{aligned}$$

Note that here we used the fact

$$\sum_{i=1}^{k} X_i \overset{d}{=} \sum_{i=1}^{k} X_{n-i+1}$$

since the X_i's are iid.

Now let $x_n := \mathbb{E}[R_n]$ and $p_n := \mathbb{P}\{R_n = R_{n-1}+1\}$. By the downward continuity of measure, we have that as $n \to \infty$,

$$p_n \to \ell =: \mathbb{P}\{\forall k \geq 1, S_k = 0\}.$$

By the law of total expectation,

$$\begin{aligned}x_n &= \mathbb{E}[R_n \mid R_n = R_{n-1}] \cdot (1-p_n) + \mathbb{E}[R_n \mid R_n = R_{n-1}+1] \cdot p_n \\ &= x_{n-1}(1-p) + (x_{n-1}+1)p_n \\ &= x_{n-1} + p_n \\ &= \sum_{k=1}^{n} p_k.\end{aligned}$$

Recall that if a real-valued sequence converges, then the average of the partial sums also converge to the same limit. It follows that

$$\frac{1}{n}\mathbb{E}[R_n] = \frac{1}{n}\sum_{k=1}^{n} p_k \to \ell$$

as $n \to \infty$.

Finally, we wish to show that in the case of simple random walks,

$$\ell = |p-q|.$$

Without loss of generality, let us assume that $p \geq q$ and show that $\ell = p - q$.

Equivalently, we can show that the probability of eventually hitting 0 is $2q$. Then $\ell = 1 - 2q = p - q$. Let p_k denote the probability of eventually hitting 0 starting at position $k > 0$. Then we have the recurrence

$$\begin{aligned}p_1 &= q + pp_2 \\ &= q + pp_1^2.\end{aligned}$$

To see this remark that the increments of the random walk are independent, thus to arrive at 0 starting at 2 is equivalent to two independent random walks hitting 0 when starting at 1.

Solving this quadratic equation yields solutions $1, \frac{q}{p}$. In the case of $p \geq q$, we take $\frac{q}{p}$ and in the case of $p < q$, we take 1. The probability of eventually returning to 0 assuming that $p \geq q$ is thus

$$q \cdot 1 + p \cdot \frac{q}{p} = 2q$$

as desired. To see this, we condition on the first step being to the left or the right and apply the appropriate solution to the quadratic equation above.

Problem 3.

From our work in class, we know that the expected hitting time for $p = q = \frac{1}{2}$ is infinity. But then the expected hitting time for $p < q$ can only be greater and is thus not finite either. We focus on the case where $p > q$.

Let us rewrite this as the expected hitting time of 0 starting at position b by flipping the values of p, q. Our strategy to compute this value is to compute the expected hitting time by setting an artificial boundary at n and then letting $n \to \infty$, similar to our work in class. This strategy is justified by the monontone convergence theorem.

Let $T_{b,n}$ denote the hitting time of one of two barriers $0, n$ starting at $b \in [0, n]$. We have

$$\begin{aligned}
\mathbb{E}[T_{b,n}] &= q\mathbb{E}[T_{b-1,n}] + p\mathbb{E}[T_{b+1,n}] + 1 \\
\mathbb{E}[T_{0,n}] &= 0 \\
\mathbb{E}[T_{n,n}] &= 0.
\end{aligned}$$

Solving this recurrence under the boundary conditions yield

$$\begin{aligned}
\mathbb{E}[T_{b,n}] &= \frac{b}{q-p} - \frac{n}{q-p} \cdot \frac{\left(\frac{q}{p}\right)^b - 1}{\left(\frac{q}{p}\right)^n - 1} \\
&\to \frac{b}{q-p}. && n \to \infty
\end{aligned}$$

Keeping in mind that we flipped the values of p, q, we conclude that the desired expectation is

$$\frac{b}{p-q}$$

as desired.

Problem 4.

Write X_n to denote the size of the population at time n. Then

$$\mathbb{P}\{T = n\} = \mathbb{P}\{X_n = 0\} - \mathbb{P}\{X_{n-1} = 0\}$$

with the base case $\mathbb{P}\{T = 0\} = 0$.

Let $\eta_n := \mathbb{P}\{X_n = 0\}$. From our work in class,

$$\begin{aligned}
\eta_0 &= 0 \\
\eta_{n+1} &= \sum_{k=0}^{\infty} \eta_n^k f(k) \\
&= \sum_{k=0}^{\infty} \eta_n^k \cdot qp^k \\
&= \frac{q}{1 - \eta_n p}.
\end{aligned}$$

Define $r := p/q$. We argue by induction that

$$\eta_n = \begin{cases} \frac{n}{n+1}, & p = q = \frac{1}{2} \\ \frac{r^n - 1}{r^{n+1} - 1}, & p \neq q \end{cases}$$

Case I: $p = q = \frac{1}{2}$ We see that the formula is correct for the base case of $n = 0$. Suppose inductively that the formula holds up to some $n \in \mathbb{N}$. Then

$$\begin{aligned}
\eta_{n+1} &= \frac{q}{1 - \eta_n p} \\
&= \frac{1}{1 - \eta_n} \\
&= \frac{n+1}{2(n+1) - n} \\
&= \frac{n+1}{n+2}
\end{aligned}$$

as required.

In this case for $n \geq 1$,

$$\begin{aligned}
\mathbb{P}\{T = n\} &= \frac{n}{n+1} - \frac{n-1}{n} \\
&= \frac{n^2 - (n^2 - 1)}{n(n+1)} \\
&= \boxed{\frac{1}{n(n+1)}}.
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}[T] &= \sum_{n \geq 1} n \cdot \frac{1}{n(n+1)} \\
&= \sum_{n \geq 1} \frac{1}{n+1}
\end{aligned}$$

is the diverging harmonic series so the expectation is infinite.

<u>Case II: $p \neq q$</u> We see that the formula is correct for the base case of $n = 0$. Suppose inductively that the formula holds up to some $n \in \mathbb{N}$. Then

$$\begin{aligned}
\eta_{n+1} &= \frac{q}{1 - \eta_n p} \\
&= \frac{q}{1 - \frac{r^n - 1}{r^{n+1} - 1} p} \\
&= \frac{r^{n+1} - 1}{\frac{1}{q}(r^{n+1} - 1) - (r^n - 1)r} \\
&= \frac{r^{n+1} - 1}{r^{n+1}(1/q - 1) + (r - 1/q)} \\
&= \frac{r^{n+1} - 1}{r^{n+2} - 1}
\end{aligned}$$

as desired.

In this case for $n \geq 1$,

$$\begin{aligned}
\mathbb{P}\{T = n\} &= \frac{r^n - 1}{r^{n+1} - 1} - \frac{r^{n-1} - 1}{r^n - 1} \\
&= \boxed{\frac{r^{n-1}(r-1)^2}{(r^{n+1} - 1)(r^n - 1)}}.
\end{aligned}$$

Recall for non-negative discrete random variables we have the identity

$$\begin{aligned}
\mathbb{E}[T] &= \sum_{n \geq 1} n \cdot \mathbb{P}\{T = n\} \\
&= \sum_{n \geq 1} \mathbb{P}\{T \geq n\} \\
&= \sum_{n \geq 1} 1 - \mathbb{P}\{T < n\} \\
&= \sum_{n \geq 1} 1 - \mathbb{P}\{T \leq n - 1\} \\
&= \sum_{n \geq 0} 1 - \mathbb{P}\{T \leq n\} \\
&= \sum_{n \geq 0} 1 - \mathbb{P}\{X_n = 0\} \\
&= \sum_{n \geq 0} 1 - \frac{r^n - 1}{r^{n+1} - 1} \\
&= \sum_{n \geq 0} \frac{r^{n+1} - 1 - r^n + 1}{r^{n+1} - 1} \\
&= \sum_{n \geq 0} \frac{r^n(r - 1)}{r^{n+1} - 1}.
\end{aligned}$$

For $r < 1$, this series converges by the ratio test

$$\begin{aligned} \frac{r^{n+1}(r-1)}{r^{n+2}-1} \cdot \frac{r^{n+1}-1}{r^n(r-1)} &= \frac{r^{n+2}-r}{r^{n+1}-1} \\ &\to r \\ &< 1. \end{aligned}$$

For $r > 1$, the series diverges as it is element-wise bounded below by a divergent series

$$\frac{r^n(r-1)}{r^{n+1}-1} \geq \frac{r^n(r-1)}{r^{n+1}} = \frac{r-1}{r} \not\to 0.$$

Problem 5.

We use the probability generating function

$$\Psi_t(x) := \mathbb{E}\left[x^{G_t}\right]$$

as the main tool to tackle this problem.

First, let us show that

$$\Psi_{t+1}(x) = \Psi_1(\Psi_t(x)).$$

Indeed, consider the more general scenario that $Z = \sum_{k=1}^{N} Y_k$ where Y_k, N are non-negative discrete variables such that the Y_k's are iid. We claim that the generating function Ψ_Z of Z satisfies

$$\Psi_Z(x) = \Psi_N(\Psi_Y(x)).$$

First recall that the generating function of independent random variables is the product

$$\Psi_{\sum_{k=1}^{m} Y_k}(x) = \mathbb{E}\left[x^{\sum_{k=1}^{m} Y_k}\right] = \prod_{k=1}^{n} \mathbb{E}\left[x^{Y_k}\right].$$

Observe here that independence is crucial so that we can factorize the expected values. We can now compute

$$\begin{aligned}
\Psi_Y(x) &= \sum_{k\geq 0} x^k \mathbb{P}\{Y = k\} \\
&= \sum_{k\geq 0} x^k \sum_{m\geq 0} \mathbb{P}\{Y = k \mid N = m\}\mathbb{P}\{N = m\} \\
&= \sum_{m\geq 0} \mathbb{P}\{N = m\} \cdot \sum_{k\geq 0} x^k \mathbb{P}\{Y = k \mid N = m\} \\
&= \sum_{m\geq 0} \mathbb{P}\{N = m\} \cdot \Psi_{\sum_{k=1}^{m} Y_k}(x) \\
&= \sum_{m\geq 0} \mathbb{P}\{N = m\} \cdot \Psi_Y(x)^k \\
&= \Psi_N(\Psi_Y(x)).
\end{aligned}$$

Recall that we can write

$$X_{t+1} = \sum_{k=1}^{X_t} X_{t+1,k}$$

where $X_{t+1,k} \sim \mathrm{Po}(2)$ iid. It follows that

$$\begin{aligned}
\Psi_{t+1}(x) &= \Psi_t(\Psi_1(x)) \\
&\cdots \\
&= (\Psi_1 \circ \cdots \circ \Psi_1)(x) \\
&= \Psi_1(\Psi_t(x)).
\end{aligned}$$

Next, we remark that

$$\mathbb{P}\{G_t = 1\} = \Psi_t'(0).$$

To see this observe that

$$\begin{aligned}\Psi_t'(0) &= \sum_{k\ge 1} k(0)^{k-1}\mathbb{P}\{G_t = k\} \\ &= \mathbb{P}\{G_t = 1\}.\end{aligned}$$

By the chain rule,

$$\begin{aligned}\Psi_{t+1}'(x) &= \frac{d}{dx}\Psi_1(\Psi_t(x)) \\ &= \Psi_1'(\Psi_t(x)) \cdot \Psi_t'(x) \\ \Psi_{t+1}'(0) &= \Psi_1'(\Psi_t(0)) \cdot \Psi_t'(0).\end{aligned}$$

It follows that

$$\begin{aligned}\rho_t &:= \frac{\mathbb{P}\{G_{t+1} = 1\}}{\mathbb{P}\{G_t = 1\}} \\ &= \Psi_1'(\Psi_t(0)) \\ &= \Psi_1'(\eta_t) && \eta_t := \mathbb{P}\{G_t = 0\} \\ &= \sum_{k\ge 1} k\eta_t^{k-1}\mathbb{P}\{G_1 = k\} \\ &= 2\sum_{k\ge 1} \frac{(2\eta_t)^{(k-1)}e^{-2}}{(k-1)!} \\ &= 2\exp(2\eta_t - 2) \\ &\to 2\exp(2\eta - 2).\end{aligned}$$

Finally, from our work in class, the ultimate extinction probability satisfies

$$\eta = \Psi_1(\eta) = \exp(2\eta - 2).$$

Hence

$$\boxed{\lim_{t\to\infty} \rho_t = 2\eta}.$$