S&DS 351 Homework 2 Solutions

Staff

January 2023

Problem 1

> Consider a symmetric simple random walk S with $S_0 = 0$. Let $T = \min\{n \geq 1 : S_n = 0\}$ be the time of the first return of the walk to its starting point. Show that
>
> $$\mathbb{P}(T = 2n) = \frac{1}{2n-1}\binom{2n}{n}2^{-2n},$$
>
> and deduce that $\mathbb{E}[T^\alpha] < \infty$ if and only if $\alpha < \frac{1}{2}$. you may need Stirling's formula: $n! \sim n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi}$.

Note that, by symmetry,

$$\begin{aligned}
\mathbb{P}(T = 2n) &= \mathbb{P}(T = 2n \mid X_1 = 1)\mathbb{P}(X_1 = 1) + \mathbb{P}(T = 2n \mid X_1 = -1)\mathbb{P}(X_1 = -1) \\
&= 2\mathbb{P}(T = 2n \mid X_1 = 1)\mathbb{P}(X_1 = 1)
\end{aligned}$$

What is $\mathbb{P}(T = 2n | X_1 = 1)$? By symmetry, this is the probability that a $2n-1$ step random walk from 0 to 1 does not touch 0 at any other time. Notice that the number of $2n-1$-step random walk from 0 to 1 that does not touch 0 at any other time is the same as the number of shortest paths (only horizontally and vertically) on a $\mathbb{R}^2$ plane starting from $(0, 0)$ and ending at (n, n) which stays in the region $y < x$ (except for the end points), or C_{n-1} (the $(n-1)$th Catalan number). By the Ballot Theorem, this is equal to:

$$\begin{aligned}
\mathbb{P}(T = 2n \mid X_1 = 1) &= \frac{\#\{\text{walks with property}\}}{\#\{\text{all possible walks}\}} = \frac{C_{n-1}}{2^{2n-1}} \\
&= \frac{1}{2(n-1)+1}\binom{2(n-1)+1}{n-1}2^{-2n+1} \\
&= \frac{1}{2n-1}\binom{2n-1}{n-1}2^{-2n+1} \\
&= \frac{1}{2n-1} \cdot \frac{(2n-1)!}{(n-1)!n!} \cdot 2^{-2n+1} \\
&= \frac{1}{2n-1} \cdot \frac{(2n)!}{n!n!} \cdot \frac{n}{2n} \cdot 2^{-2n+1} \\
&= \frac{1}{2n-1}\binom{2n}{n}\frac{1}{2} \cdot 2^{-2n+1} \\
&= \frac{1}{2n-1}\binom{2n}{n}2^{-2n}
\end{aligned}$$

To calculate the expected value of T^α, we look to Stirling's approximation. Indeed,

$$\mathbb{E}(T^\alpha) = \sum_{n=0}^{\infty}(2n)^\alpha \mathbb{P}(T = 2n) = \sum_{n=0}^{\infty}(2n)^\alpha \frac{1}{2n-1}\binom{2n}{n}2^{-2n}$$

Note that $\dfrac{(2n)^\alpha}{2n-1} \sim (2n)^{\alpha-1}$ in asymptotic behavior. Now, observe by Stirling's approximation that

$$\binom{2n}{n} = \frac{2n!}{(n!)^2} \sim \frac{(2n)^{2n+\frac{1}{2}}e^{-2n}\sqrt{2\pi}}{(n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi})^2} = \frac{(2n)^{2n+\frac{1}{2}}}{n^{2n+1}\sqrt{2\pi}} = \frac{(2n)^{2n+1}/\sqrt{2n}}{n^{2n+1}\sqrt{2\pi}} = \frac{2^{2n+1}}{2\sqrt{\pi n}} = \frac{2^{2n}}{\sqrt{\pi n}}.$$

If we then substitute both of these approximations, we find,

$$\mathbb{E}(T^\alpha) \sim \sum_{n=0}^{\infty}(2n)^{\alpha-1}\left(\frac{2^{2n}}{\sqrt{\pi n}}\right)2^{-2n} = \frac{2^{\alpha-1}}{\sqrt{\pi}}\sum_{n=0}^{\infty} n^{\alpha-\frac{3}{2}}$$

Recall from your favorite calculus class that this sum will converge if and only if $\alpha - 3/2 < -1$, i.e. $\alpha < 1/2$. Notice that our approximations have, asymptotically, altered our sum by no more than some constant factor. Therefore, $\mathbb{E}(T^\alpha)$ possesses the same properties as the above sum, in regards to finiteness.

Problem 2

Let $\{X_n : n \geq 1\}$ be independent, identically distributed random variables taking integer values. Let $S_0 = 0, S_n = \sum_{i=1}^{n} X_i$. The *range* R_n of $S_0, S_1, \dots, S_n$ is the number of distinct values taken by the sequence. Show that $\mathbb{P}(R_n = R_{n-1} + 1) = \mathbb{P}(S_1 S_2 \cdots S_n \neq 0)$, and deduce that, as $n \to \infty$,

$$\frac{1}{n}\mathbb{E}[R_n] \to \mathbb{P}(S_k \neq 0 \text{ for all } k \geq 1).$$

Hence show that, for the simple random walk, $n^{-1}\mathbb{E}[R_n] \to |p - q|$ as $n \to \infty$.

Fix $S_n = a$. What is the probability that you're seeing S_n for the first time? This is the same as the probability that, when you start walking backwards from a, in n steps, you never return back to S_n on your journey to 0. The former is $\mathbb{P}(R_n = R_{n-1} + 1 | S_n = a)$. To be more concrete, let $\bar{S}_0 ... \bar{S}_n$ denote the outcomes of $0, 1, 2 ... n$ steps of a backwards random walk:

$$\begin{aligned}
&\mathbb{P}(R_n = R_{n+1} | S_n = a) \\
&= \mathbb{P}(\bar{S}_n = 0, \text{walk never revisits a} | \bar{S}_0 = a) \\
&= \mathbb{P}(\bar{S}_n = -a, \text{walk never revisits origin} | \bar{S}_0 = 0)
\end{aligned}$$

Taking expectations, we find,

$$\begin{aligned}
\mathbb{P}(R_n = R_{n+1}) &= \mathbb{E}_a[\mathbb{P}(R_n = R_{n+1} | S_n = a)] \\
&= \mathbb{E}_a[\mathbb{P}(\bar{S}_n = -a, \text{walk never revisits origin} | \bar{S}_0 = 0)] \\
&= \mathbb{P}(\bar{S}_1 ... \bar{S}_n \neq 0)
\end{aligned}$$

Finally, by symmetry, $\mathbb{P}(\bar{S}_1 ... \bar{S}_n \neq 0) = \mathbb{P}(S_1 ... S_n \neq 0)$. Why? Well, for every forwards walk $S_1, S_2 ... S_n$, there is an associated backwards random walk $-S_1 ... -S_n$; the former revisits the origin if and only if the latter does, so they occur with the same overall probability.

Part 2 First, let We now take the expectation of R_n. Observe that $R_n = 1 + \sum_{i=1}^{n} \mathbb{I}\{R_i = R_{i-1} + 1\}$, since these both count how many new values we discover. Thus,

$$\frac{1}{n}\mathbb{E}[R_n] = \frac{1}{n} + \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(R_i = R_{i-1} + 1) = \frac{1}{n} + \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(S_1 ... S_i \neq 0)$$

FACT (Cesaro Sum Converges to Limit) *If $a_n \to a$, then $\frac{1}{n}\sum_{i=1}^{n} \to a$ as well. Proof Sketch: for all $\epsilon > 0$, $|a_n - a| < \epsilon$ for all large n, upper bound $|a - \frac{1}{n}\sum_{i=1}^{n} a_n|$ by the triangle inequality, and make n large enough to account for the slack where $|a_n - a| > \epsilon$.*

Note that, clearly, $\mathbb{P}(S_1 ... S_i \neq 0) \to \mathbb{P}(S_1, S_2 .. \neq 0)$. Therefore, by this property, $\frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(R_i = R_{i-1} + 1) \to \mathbb{P}(S_1, S_2 .. \neq 0)$ and $1/n \to 0$. Therefore, $\frac{1}{n}\mathbb{E}[R_n] \to \mathbb{P}(S_1 S_2 ... \neq 0)$.

Part 3 By Theorem 7, $\mathbb{P}(S_1 ... S_n \neq 0) = \frac{1}{n}\mathbb{E}[|S_n|]$. Assume, without loss of generality, that $p > q$. Observe that, by the strong law of large numbers, since the X_n are i.i.d., we have $\frac{1}{n}S_n \to \mathbb{E}[X_1] \to p - q$ with probability 1. Therefore, $\frac{1}{n}S_n > 0$ with probability 0. And so, as $n \to \infty$, $\frac{1}{n}S_n - \frac{1}{n}|S_n| \to 0$, meaning $\frac{1}{n}\mathbb{E}[|S_n|] \to \frac{1}{n}|\mathbb{E}[S_n]| = p - q = |p - q|$. The casework for $p < q$ and $p = q$ is similar. We conclude that $\mathbb{P}(S_1 ... S_n \neq 0) \to |p - q|$. And thus, $\frac{1}{n}\mathbb{E}[R_n] \to |p - q|$

Problem 3

Consider a simple random walk starting at 0 in which each step is to the right with probability p $(= 1 - q)$. Let T_b be the number of steps until the walk first reaches b where $b > 0$. Show that $\mathbb{E}[T_b \mid T_b < \infty] = b/|p - q|$.

Recall that $\mathbb{P}(T_b = n) = \frac{b}{n}\mathbb{P}(S_n = b)$. Also, assume that $\mathbb{P}(T < \infty) \neq 0$, so that our conditional expectation is well posed. Thus,

$$\begin{aligned}\mathbb{E}[T_b|T < \infty] &= \mathbb{P}(T < \infty)^{-1} \sum_{n=1}^{\infty} n\mathbb{P}(T_b = n) \\ &= \mathbb{P}(T < \infty)^{-1} \sum_{n=1}^{\infty} \frac{bn}{n}\mathbb{P}(S_n = b) \\ &= b\mathbb{P}(T < \infty)^{-1} \sum_{n=1}^{\infty} \mathbb{P}(S_n = b) \\ &= b\mathbb{P}(T < \infty)^{-1}\mathbb{E}\left[\sum_{n=1}^{\infty} \mathbb{I}\{S_n = b\}\right]\end{aligned}$$

If we let N_b be the number of times the walk reaches b, we find the above is all equal to $b\mathbb{P}(T < \infty)^{-1}\mathbb{E}[N_b]$. Yet, on the other hand, $\mathbb{E}[N_b] = \mathbb{E}[N_b|T < \infty]\mathbb{P}(T < \infty) + 0\mathbb{P}(T = \infty)$. Thus, the above is equal to $b\mathbb{E}(N_b|T < \infty)$. But note also, by spatial homogeneity, $\mathbb{E}(N_b|T < \infty) = 1 + \mathbb{E}[N_0]$, since $T < \infty$ tells us that our walk visits b, which contributes one visit. And from that point on, we ask, how many times does our walk revisit b? And that will be the same probability as the expected number of visits *to the origin*, if we start *from the origin*.

Let's calculate $\mathbb{E}[N_0]$, the expected number of hits to the origin NOT including the trivial visit at our starting point. Now, recall from the chapter that the probability the walk returns to the origin is $1 - |p - q|$. We observe that hitting the origin is now a sort of geometric process. Think of *failure* as the event that we return to the origin at some point, and *success* as the event that we escape, never to return. In other words, we find that $\mathbb{E}[N_0] \sim \text{Geom}(|p - q|)$, under the interpretation of the geometric in which we count *failures*, not trials. This distribution has mean $\frac{1}{|p-q|} - 1$. Therefore, $\mathbb{E}(N_b|T < \infty) = 1/|p - q|$. The desired result follows immediately.

Problem 4

Consider a branching process whose family sizes have the geometric mass function $f(k) = qp^k, k \geq 0$, where $p+q=1$, and let Z_n be the size of the nth generation. Let $T = \min\{n : Z_n = 0\}$ be the extinction time, and suppose that $Z_0 = 1$. Find $\mathbb{P}(T=n)$. For what values of p is it the case that $\mathbb{E}[T] < \infty$?

Define the probability mass generating function like so:

$$\psi(\rho) = \sum_{k=0}^{\infty} f(k)\rho^k = \sum_{k=0}^{\infty} qp^k\rho^k = q\sum_{k=1}^{\infty}(p\rho)^k = \frac{q}{1-\rho p}$$

Solving for $\psi(\rho) = \rho$ (cross multiplication and basic quadratic formulas stuff), we obtain solutions to this equation $\rho = q/p$. It is a known fact that when $\mu < 1$ for the distribution that generates the branching, then $\rho < 1$. Thus, if $p > q$, $\rho = q/p$, and otherwise, $\rho = 1$. This is true by theorem 5 in the text.

Part 2 Recall from the text that Generating Functions have the following property for branching processes $G_n(s) \triangleq \mathbb{E}[s^{Z_n}] = G(G(..G(s))\ (n iterations)$, where $G(s)$ is the generating function for the children of a particular individual. Note also the moment generating function of the geometric is $G(s) = \sum_{k=0}^{\infty} qp^k s^k = \frac{q}{1-ps}$. The book claims, as we prove by induction at the end of the problem,

$$G_n(s) = \begin{cases} \frac{n-(n-1)s}{n+1-ns} & p = q \\ \frac{q(p^n - q^n - ps(p^{n-1}-q^{n-1}))}{p^{n+1}-q^{n+1}-ps(p^n-q^n)} & p \neq q \end{cases}$$

Using the GF Observe that, since $\mathbb{P}(Z_n = 0, Z_{n-1} \neq 0) = \mathbb{P}(Z_n = 0)$,

$$\mathbb{P}(T=n) = \mathbb{P}(Z_n = 0, Z_{n-1} \neq 0) = \mathbb{P}(Z_n = 0) - \mathbb{P}(Z_n = 0, Z_{n-1} = 0) = \mathbb{P}(Z_n = 0) - \mathbb{P}(Z_{n-1} = 0)$$

Furthermore, we know that $G_n(0) = \mathbb{P}(Z_n = 0)$ and likewise for $n-1$. First, in the $p = q$ symmetric case, we have $G_n(0) = \frac{n-(n-1)(0)}{n+1-n(0)} = \frac{n}{n+1}$. Similarly, we have $G_{n-1}(0) = \frac{n-1}{n}$. So then,

$$\mathbb{P}(T=n) = \frac{n}{n+1} - \frac{n-1}{n} = \frac{n^2-(n-1)(n+1)}{n(n+1)} = \frac{n^2-(n^2-1)}{n(n+1)} = \frac{1}{n(n+1)}$$

In the asymmetric case, $G_n(0) = \frac{q(p^n-q^n)}{p^{n+1}-q^{n+1}}$. Likewise, $G_{n-1}(0) = \frac{q(p^{n-1}-q^{n-1})}{p^n-q^n}$. So,

$$\mathbb{P}(T=n) = \frac{q(p^n-q^n)}{p^{n+1}-q^{n+1}} - \frac{q(p^{n-1}-q^{n-1})}{p^n-q^n} = \frac{q[(p^n-q^n)^2-(p^{n-1}-q^{n-1})(p^{n+1}-q^{n+1})]}{(p^{n+1}-q^{n+1})(p^n-q^n)} = \frac{p^{n-1}q^n(p-q)^2}{(p^{n+1}-q^{n+1})(p^n-q^n)}$$

Values of p for finite extinction time When $p = q$, we have already known that $\mathbb{P}(T=n) = \frac{1}{n(n+1)}$, therefore:

$$\mathbb{E}(T) = \sum_{n=1}^{\infty} \frac{n}{n(n+1)} = \sum_{n=1}^{\infty} \frac{1}{n+1} = +\infty.$$

When $p > q$, we know that $\mathbb{P}(T=\infty) = \frac{q}{p}$, which makes $\mathbb{E}(T) = \infty$. When $p < q$, denote $\gamma = p/q < 1$. Since

$$\mathbb{P}(T=n) = \frac{p^{n-1}q^n(p-q)^2}{(p^{n+1}-q^{n+1})(p^n-q^n)} = \frac{\gamma^{n-1}}{(1+\gamma+\ldots+\gamma^{n-1})(1+\gamma+\ldots+\gamma^n)} < \gamma^{n-1},$$

we have:

$$\mathbb{E}(T) \leqslant \sum_{n=1}^{\infty} n\gamma^{n-1} = \frac{1}{(1-\gamma)^2} < \infty.$$

To sum up, for all p such that $p < \frac{1}{2}$, we have $\mathbb{E}(T) < \infty$.x

Proof of Generating Function Let us show both of these by induction. First, let us show the $p = q$ case. Note, subbing in $n = 1$, $\frac{n-(n-1)s}{n+1-ns} = \frac{1}{2-s} = \frac{1/2}{1-1/2s} = G(s)$. Now, for the inductive step, we assume the claim is true of $n-1$, and we want to prove it for n. Then,

$$G_n(s) = G(G_{n-1}(s)) = \frac{q}{1 - pG_{n-1}(s)} = \frac{1}{2 - G_{n-1}(s)}$$
$$= \frac{1}{2 - \frac{(n-1)-(n-2)s}{(n-1)+1-(n-1)s}} = \frac{1}{\frac{2[(n-1)+1-(n-1)s]-(n-1)+(n-2)s}{(n-1)+1-(n-1)s}}$$
$$= \frac{(n-1)+1-(n-1)s}{2(n-1)+2-2(n-1)s-(n-1)+(n-2)s}$$
$$= \frac{n-(n-1)s}{(n+1)-ns}$$

Completing the inductive step. Now we show the proof for the asymmetric case. Observe that, at $n = 1$,

$$\frac{q(p^n - q^n - ps(p^{n-1} - q^{n-1})}{p^{n+1} - q^{n+1} - ps(p^n - q^n)} = \frac{q(p - q - ps(1-1))}{p^2 - q^2 - ps(p-q)} = \frac{q(p-q)}{(p+q)(p-q) - ps(p-q)} = \frac{q}{1-ps} = G(s)$$

As desired. Now, for the inductive step, we have,

$$G_n(s) = \frac{q}{1 - pG_{n-1}(s)} = \frac{q}{1 - p\left(\frac{q(p^{n-1}-q^{n-1}-ps(p^{n-2}-q^{n-2})}{p^n-q^n-ps(p^{n-1}-q^{n-1})}\right)}$$
$$= \frac{q(p^n - q^n - ps(p^{n-1} - q^{n-1}))}{p^n - q^n - ps(p^{n-1} - q^{n-1}) - pq(p^{n-1} - q^{n-1} - ps(p^{n-2} - q^{n-2}))}$$
$$= \frac{q(p^n - q^n - ps(p^{n-1} - q^{n-1}))}{p^n - q^n + (-ps - pq)(p^{n-1} - q^{n-1}) + (pq)(ps)(p^{n-2} - q^{n-2})}$$

It remains to verify our denominators are consistent. Observe that,

$$p^n - q^n + (-ps - pq)(p^{n-1} - q^{n-1}) + (pq)(ps)(p^{n-2} - q^{n-2})$$
$$= p^n - q^n - sp^n + spq^{n-1} - qp^n + pq^n + sqp^n - sp^2q^{n-1}$$
$$= p^n - q^n - qp^n + pq^n + s(-p^n + pq^{n-1} + qp^n - p^2q^{n-1})$$
$$= p^n(1-q) - q^n(1-p) + s(q^{n-1}(p-p^2) - p^n(1-q))$$
$$= p^np - q^nq + s(q^{n-1}pq - p^np) = p^{n+1} - q^{n+1} + ps(q^n - p^n)$$

As desired. This completes the induction. We will now use this fact. We will rewrite the moment generating function as a power series, which lets us trivially extract probabilities.

Problem 5

Consider a branching process G_t with $G_0 = 1$ and offspring distribution $\mathsf{Poisson}(2)$. We know that the process will either go extinct or diverge to infinity, and the probability that it is any fixed finite value should converge to 0 as $t \to \infty$. In this exercise you will investigate how fast such probabilities converge to 0. In particular, consider the probability $\mathbb{P}(G_t = 1)$, and find the limiting ratio

$$\lim_{t\to\infty} \frac{\mathbb{P}(G_{t+1} = 1)}{\mathbb{P}(G_t = 1)}.$$

Solution We will approach this also through generating functions. Note that, letting $F_t(s)$ be the generating function of G_t,

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

Furthermore, as $M_{t+1}(s) = M_1(M_t(s))$, by the product rule,

$$\mathbb{P}(G_{t+1} = 1) = M_{t+1}'(0) = M_1'(M_{t+1}(0)) \cdot M_t'(0) = M_1'(\mathbb{P}(G_{t+1} = 0))\mathbb{P}(G_t = 1)$$

Therefore,

$$\frac{\mathbb{P}(G_{t+1} = 1)}{\mathbb{P}(G_t = 1)}. = \frac{M_1'(\mathbb{P}(G_{t+1} = 0))\mathbb{P}(G_t = 1)}{\mathbb{P}(G_t = 1)} = M_1'(\mathbb{P}(G_{t+1} = 0))$$

We should pause to note that $\mathbb{P}(G_{t+1} = 0) \uparrow \mathbb{P}(\text{Extinction Occurs}) = \rho$. Since M_1 is continuously differentiable, it therefore follows from continuity that $M_1'(\mathbb{P}(G_{t+1} = 0)) \to M_1'(\rho)$

Probability of Extinction We have gotten away with it thus far, but it's time to find the generating function. Recall the PMF of a poisson distribution is $f(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ (here $\lambda = 2$). Thus, the moment generating function is $M_1(s) = \sum_{k=0}^{\infty} \frac{(s\lambda)^k e^{-\lambda}}{k!}$. You might recognize this as $e^{-\lambda}e^{\lambda s} = e^{\lambda(s-1)}$. Thus, the extinction probability ρ is given by the solution to $\rho = e^{\lambda(\rho-1)}$. Solving with $\lambda = 2$ we have a solution given numerically by solutions $\rho = 1, 0.203$. Furthermore, recall the mean of the Poisson distribution is given by λ, i.e. we have a mean of 2. It therefore follows that we take the solution $\rho = 0.203$. Thus, $\mathbb{P}(G_{t+1} = 0) \uparrow \rho = 0.203$. Now, we evaluate $M_1'(s) = \frac{d}{ds}e^{-\lambda}e^{\lambda s} = \lambda e^{-\lambda}e^{\lambda s} = \lambda M_1(s)$. Therefore, $M_1'(\rho) = \lambda M_1(\rho) = \lambda\rho$. In our case, we have that this is equal to $2\rho \approx 0.406$. This is the limiting ratio.