Simple Walk Primer

Lily Li

December 20, 2021

Let $X_1, X_2, ...$ be a sequence of independent Rademacher random variables, i.e. $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$, and let $S_n = \sum_{i=1}^n X_i$ be their partial sums. This will be our *standard setting* for a simple random walk.

1 Bounded Simple Walks

Theorem 1. (Kolmogorov's Maximal Inequality). *For independent random variables $Y_1, Y_2, ...$ such that $\mathbb{E}Y_i = 0$ and $\mathrm{Var}(Y_i) < \infty$, and $T_n = \sum_{i=1}^n Y_i$,*

$$\mathbb{P}\left(\max_{1\le i\le n} |T_i| > t\right) \le \frac{\mathrm{Var}(T_n)}{t^2}.$$

Note: that the bound on the RHS is identical to to Chebyshev's bound, but the inequality is much stronger since Chebyshev only bounds $\mathbb{P}(|T_n| > t)$ from above.

Proof. Fix some x. Let A_k be the event where $|S_k| \ge x$ but $|S_j| < x, j < k$ (we will break up the processes (S_n) up according to the time that $|S_k|$ first exceeds x). Since A_ks are disjoint and $(S_n - S_k)^2 \ge 0$,

$$\begin{aligned}\mathbb{E}S_n^2 &\ge \sum_{k=1}^n \int_{A_k} S_n^2 d\mathbb{P} = \sum_{k=1}^n \int_{A_k} S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2 d\mathbb{P} \\ &\ge \sum_{k=1}^n \int_{A_k} S_k^2 + \sum_{k=1}^n \int_{A_k} 2S_k(S_n - S_k)\end{aligned}$$

Note that $S_k \mathbb{1}_{A_k}$ and $S_n - S_k$ are independent (the former depends on the r.v. $X_1, ..., X_k$ while the latter depends on $X_{k+1}, ..., X_k$). Thus we can decompose $\int 2S_k \mathbb{1}_{A_k} \cdot (S_n - S_k) d\mathbb{P}$ as $\mathbb{E}2S_k\mathbb{1}_{A_k} \cdot \mathbb{E}(S_n - S_k) = 0$ (remember the second term is equal to zero). Since $|S_k| \ge x$ on A_k and the A_ks are disjoint,

$$\mathbb{E}S_n^2 \ge \sum_{k=1}^n \int_{A_k} S_n^2 d\mathbb{P} \ge \sum_{k=1}^n x^2 \mathbb{P}(A_k) = x^2 \mathbb{P}\left(\max_{1\le k\le n} |S_k| \ge x\right).$$ □

Lemma 2. *In our standard setting, $\mathbb{P}(|S_n| \leq \sqrt{n}/4 | |S_n| \leq \sqrt{n}/2) \geq 1/2$.*

Proof. Use the reflection principle as follows: Create a bijection between those paths where $|S_n| \in (\sqrt{n}/4, \sqrt{n}/2]$ and those where $|S'_n| \in [0, \sqrt{n}/4)$. Wlog. suppose that S_n is positive. Let $T \in [n]$ be the time step where $S_T = \sqrt{n}/4$ and for all $t > T$, $S_t > \sqrt{n}/4$. Form a unique S'_n by reflecting S_n at the point S_T across the line $\sqrt{n}/4$. The reverse process takes S'_n to a unique S_n. □

Theorem 3. *In our standard setting, $\mathbb{P}\left(\max_{1\leq i\leq n} |S_n| \geq c\sqrt{n}\right) \leq \alpha$ for constants c and $\alpha = \alpha(c)$ a function of c.*

Proof. When $c > 2$, we can use Kolmogorov's Maximal Inequality (Theorem 1) directly to get $\alpha = \frac{1}{c^2} \leq \frac{1}{4}$. When $c \leq 2$, we divide $[n]$ into intervals and upper-bound the probability of going outside the strip $\pm c\sqrt{n}$ on each interval. For concreteness, let $c = 1/2$. Divide $[n]$ into 64 pieces $T_1 = [1, n/64)$, ..., $T_{64} = [63n/64, n]$. In order to upper-bound the probability that the walk exists the strip $\pm\sqrt{n}/2$, we will upper-bound the probability that the walk exists the strip of width $\sqrt{n}/4$ from its starting position or ends outside $\pm\sqrt{n}/4$. Bound the former probability by Theorem 1, $\mathbb{P}\left(\max_{T_j<i\leq T_{j+1}} |S_i - S_{T_j}| \geq \sqrt{n}/4\right) \leq 1/4$. Suppose wlog. $S_{T_j} \geq 0$. For the latter probability, we note that it is equally likely for $S_{T_{j+1}}$ to be greater than S_{T_j} or less than S_{T_j}, so if $S_{T_j} \leq \sqrt{n}/4$, the probability that $S_{T_{j+1}} \leq \sqrt{n}/4$ — if $\mathbb{P}\left(\max_{T_j<i\leq T_{j+1}} |S_i - S_{T_j}| \geq \sqrt{n}/4\right)$ — is bounded above by $\mathbb{P}\left(S_{T_{j+1}} - S_{T_j} \leq 0\right) \leq 1/2$. Let E_i be the event $|S_{T_j}| \leq \sqrt{n}/4 \wedge \max_{1\leq i\leq T_j} |S_i| \leq \frac{\sqrt{n}}{2}$. Then, conditioned on E_i, we have that

$$\mathbb{P}\left(\max_{T_j<i\leq T_{j+1}} |S_i| > \frac{\sqrt{n}}{2} \mid E_i\right) \leq \mathbb{P}\left(\max_{T_j<i\leq T_{j+1}} |S_i - S_{T_j}| \geq \sqrt{n}/4\right) + \mathbb{P}\left(|S_{T_{j+1}}| > \sqrt{n}/4\right)$$
$$\leq \frac{1}{4} + \frac{1}{2} = \frac{3}{4}.$$

Let $p_j = \mathbb{P}\left(\max_{T_{j-1}<i\leq T_j} |S_i - S_{T_{j-1}}| > \frac{\sqrt{n}}{4}\right) + \mathbb{P}\left(|S_{T_j}| > \sqrt{n}/4\right) \leq \frac{3}{4}$. Thus, over all the intervals,

$$\mathbb{P}\left(\max_{1\leq i\leq n} |S_n| \geq \frac{\sqrt{n}}{2}\right) \leq p_1 + (1-p_1)\left(p_2 + (1-p_2)\left(\cdots\left(p_{63} + (1-p_{63})p_{64}\right)\right)\right).$$

Since $p_i = p_j$ for $i, j \in [64]$, let $p := p_1$ and write the above bound as

$$p\left(1 + (1-p) + \cdots + (1-p)^{63}\right) = 1 - (1-p)^{64}.$$

Thus $\mathbb{P}\left(\max_{1\leq i\leq n} |S_n| \leq \frac{\sqrt{n}}{2}\right) \geq (1/4)^{64}$. More generally, for any constant c, we would divide $[n]$ into $t = 16/c^2$ intervals and get a lower bound of $\mathbb{P}\left(\max_{1\leq i\leq n} |S_n| \geq c\sqrt{n}\right) \geq (1/4)^t$. □

2 Ballot Theorem and its Implications

In an election A gets α votes and B gets β votes for $\alpha > \beta$. Let a *dominating ballot sequence* be a sequence of the As and Bs such that in any prefix of the sequence there are more As than Bs.

Theorem 4. (Ballot Theorem). *The probability that a random sequence is dominating is* $(\alpha - \beta)/(\alpha + \beta)$.

Proof. Here we will use the reflection principle. Let $N_{i,j}$ be the number of paths from $(0,0)$ to (i,j). Let $y = \alpha - \beta$ and $n = \alpha + \beta$. The number of dominant ballot sequences is equivalent to the number of path from $(1,1)$ to (n,y) which do not touch the x-axis. We can write this as $N_{n-1,y-1} - N_{n-1,y+1}$ since the latter term is exactly the number of paths which intersect the x-axis. It follows that

$$\begin{aligned} N_{n-1,y-1} - N_{n-1,y+1} &= \binom{n-1}{a-1} - \binom{n-1}{a} \\ &= \frac{(n-1)!}{(\alpha-1)!(n-\alpha)!} - \frac{(n-1)!}{\alpha!(n-\alpha-1)!} \\ &= \binom{n}{a} \cdot \left(\frac{a-(n-a)}{n}\right) = \binom{n}{a} \cdot \left(\frac{y}{n}\right) \end{aligned}$$ □

There is another beautiful proof of Dvoretzky and Motzkin which goes as follows: Arrange the α As and β B in a circle. Repeatedly remove pairs of AB until only $\alpha - \beta$ As remain. Note that each A is the start of a dominating ballot sequence.

From this, we can compute the distribution of the time to hit 0 for a simple random walk.

Lemma 5. $\mathbb{P}(S_1 \neq 0, ..., S_{2n} \neq 0) = \mathbb{P}(S_{2n} = 0)$.

Proof. The proof structure is very similar to that of the Ballot Theorem (Theorem 4). Note that $\mathbb{P}(S_1 \neq 0, ..., S_n \neq 0) = \mathbb{P}(S_1 > 0, ..., S_n > 0) + \mathbb{P}(S_1 < 0, ..., S_n < 0)$. Since the two terms are symmetric, it suffices to compute $\mathbb{P}(S_1 > 0, ..., S_n > 0)$. Let $p_{n,x} = \mathbb{P}(S_n = x)$. Condition on the value of S_n to obtain $\mathbb{P}(S_1 > 0, ..., S_{2n-1} > 0, S_{2n} = 2r)$ we note that

$$\mathbb{P}(S_1 > 0, ..., S_{2n-1} > 0, S_{2n} = 2r) = \frac{N_{2n-1,2r-1} - N_{2n-1,2r+1}}{2^{2n}} = \frac{(p_{2n-1,2r-1} - p_{2n-1,2r+1})}{2}.$$

It follows that

$$\mathbb{P}\left(S_1 > 0, ..., S_{2n} > 0\right) = \frac{1}{2}\sum_{r=1}^{\infty} p_{2n-1,2r-1} - p_{2n-1,2r+1} = \frac{p_{2n-1,1}}{2} = \frac{\mathbb{P}(S_{2n} = 0)}{2}.$$ □

Note: S_{2n} *is distributed like a binomal with mean zero and variance* $2n$*, thus* $\mathbb{P}(S_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}$ *by the central limit theorem approximation of the central term. Generally* $\mathbb{P}(S_{2n} = 2k) \sim \frac{1}{\sqrt{\pi n}} e^{-x^2/2}$ *where* $2k/\sqrt{2n} \to x$*. You obtain this by applying Stirlings to* $\binom{2n}{n+k}$ *and arguing carefully.*[1]

[1] In particular, we have

$$\left(1 + \frac{k}{n}\right)^{-n-k} \cdot \left(1 - \frac{k}{n}\right)^{-n+k} \cdot \left(\frac{1}{\sqrt{\pi n}}\right) \cdot \left(1 + \frac{k}{n}\right)^{-1/2} \cdot \left(1 - \frac{k}{n}\right)^{-1/2}.$$

Then you use the lemma that if $c_j \to 0$, $a_j \to \infty$, and $a_j c_j \to \lambda$, then $(1 + c_j)^{a_j} \to e^{\lambda}$.

Let $L_{2n} = \sup\{m \leq 2n : S_m = 0\}$ be the last time the random walk visits zero. Let $u_{2m} = \mathbb{P}(S_{2m} = 0)$. Then $\mathbb{P}(L_{2n} = 2k) = u_{2k}u_{2n-2k}$. This is because $\mathbb{P}(L_{2n} = 2k) = \mathbb{P}(S_{2k} = 0) \cdot \mathbb{P}(S_{2k+1} \neq 0, ..., S_{2n} \neq 0) = \mathbb{P}(S_{2k} = 0) \cdot \mathbb{P}(S_{2n-2k} = 0)$.

Theorem 6. (Arcsine Law for the Last Visit to Zero). *For* $0 < a < b < 1$,

$$\mathbb{P}\left(a \leq \frac{L_{2n}}{2n} \leq b\right) \to \frac{1}{\pi}\int_a^b \frac{1}{\sqrt{x(1-x)}}dx.$$

Note: by substituting $y = \sqrt{x}$ *and* $dy = \frac{1}{2\sqrt{x}}dx$*, and changing the limits of integration to* $\sqrt{a}$ *and* $\sqrt{b}$ *respectively, we have that*

$$\frac{1}{\pi}\int_a^b \frac{1}{\sqrt{x(1-x)}}dx = \frac{2}{\pi}\int_{\sqrt{a}}^{\sqrt{b}} \frac{1}{\sqrt{(1-y)^2}}dy = \frac{2}{\pi}\left(\arcsin(\sqrt{b}) - \arcsin(\sqrt{a})\right)$$

which is a standard trig integral by further substitution $y = sin(\theta)$ *and* $dy = \cos(\theta)d\theta$.

Proof. Using Lemma 5 and the subsequent note, we obtain

$$n\mathbb{P}(L_{2n} = 2k) = nu_{2k}u_{2n-2k} \to f(x) := \frac{1}{\pi} \cdot \left(\frac{1}{\sqrt{x(1-x)}}\right)$$

when $k/n \to x$. Letting $f_n(x) = n\mathbb{P}(L_{2n} = 2k)$ for $2k/2n \leq x \leq 2(k+1)/2n$, we can approximate

$$\mathbb{P}\left(a \leq \frac{L_{2n}}{2n} \leq b\right) = \sum_{k=na'}^{nb'} \mathbb{P}(L_{2n} = 2k) \approx \int_{a'}^{b'} f_n(x)dx$$

where $2na'$ and $2nb'$ are the smallest and largest even integers greater than or equal to and $\geq 2na$ and $\leq 2nb$ respectively. We obtain the desired inequality, since $f_n(x) \to f(x)$. □

Theorem 7. (Arcsine Law for Time Above Zero). *Let* A_{2n} *be the number of segments* $(k-1, S_{k-1}) \to (k, S_k)$ *that lie above the* x*-axis, and* $u_{2m} = \mathbb{P}(S_{2m} = 0)$ *as before. Then* $\mathbb{P}(A_{2n} = 2k) = u_{2k}u_{2n-2k}$ *and so* $A_{2n} \overset{d}{=} L_{2n}$.

The proof is by induction on n. Let $\alpha_{2k,2n} = \mathbb{P}(A_{2n} = 2k)$. When $n = 1$, either both edges are above or both below the x-axis with equal probability so $\alpha_{0,2} = \alpha_{2,2} = 1/2$. Next consider $k = n$ and then all k such that $1 \leq k < n$ (this also requires the probability of first return to zero).