Chaining and bounds on the supremum of a random walk

Sparsho De

1 Bounding traditional random walks

Before introducing methods to bound the supremum of random walks, we demonstrate methods to bound traditional random walks. Then, we give looser bounds on the expectation of the supremum of a random walk before we introduce chaining and prove the final result: $\mathbb{E}(\sup Y_n) = \mathcal{O}(\sqrt{n})$, where Y_n is a random walk of n steps.

1.1 Martingale Concentration Inequality

We define a random walk of size n as the sum of n random signs, and then bound it via concentration inequalities. Let X_i be $+1$ with 0.5 probability and -1 with 0.5 probability. Let Y_n be the random variable defined as the position of the random walk. Then $Y_n = \sum_{i=1}^{n} X_i$. For an intuitive reasoning for why this is true, consider the following bijection. Start at position 0 on the number line. We have an input vector $\{X_1, X_2, ..., X_m\}$. For each element in the vector, if X_i is negative, it is equivalent to moving left on the number line. Else, it is equivalent to moving right, each with equal probability. This matches the definition of a random walk.

Note $\mathbb{E}(Y_{n+1}|X_1, X_2, ..., X_n) = 1/2(Y_n - 1) + 1/2(Y_n + 1) = Y_n$. Furthermore, note that Y_n is a function of i.i.d random variables $X_1, X_2, ..., X_n$ (and in particular, $Y_n = \sum_{i=1}^{n} X_i$). This means that the Y_n is a martingale, and we can apply martingale concentration inequalities (most notably, the Azuma-Hoeffding Inequality).

Before applying the inequality, it's important to note that the bound is on the *deviation* of the random walk from its mean. That is, we're bounding $\mathbb{E}(|Y_n|)$. Clearly $\mathbb{E}|Y_n)$ is just 0.

The statement of Azuma-Hoeffding is as follows:

For some martingale M_i,

$$Pr(|M_i| > t) <= 2\exp(-t^2/(2 \cdot \sum_{i=1}^{n} d_i)),$$

where $d_i <= |M_i - M_i - 1|$. This statement can be derived by applying Hoeffding's Lemma on the moment-generating function of an arbitrary martingale, and then conditioning with respect to its filtration.

Clearly for the random walk Y_n, $d_i = 1$ since at each step, Y_n changes by exactly 1. Thus, $Pr(|Y_n| > t) <= 2\exp(-t^2/(2n))$. Note that if $t = \mathcal{O}\sqrt{n}$, then $Pr(|Y_n| > t)$ is constant, implying $Pr(|Y_n| < t)$ is also constant, and deviation from the mean in a random walk is approximately a square root function of the number of steps. Therefore, $\mathbb{E}(|Y_n|) = \mathcal{O}(\sqrt{n})$.

1.2 Central Limit Theorem

A similar result can be achieved via Central Limit Theorem. For sufficiently large n, the distribution of Y_n can be modeled by a normal distribution with mean 0 and some arbitrary standard deviation σ. As mentioned in Section 1.1, a random walk of n steps can be modeled as the sum of n i.i.d random signs (referred to as X_i). Via linearity of variance, we know that $Var(Y_n) = Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i)$. By definition, $Var(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = 1 - 0 = 1$. Thus, $Var(Y_n) = n \cdot 1 = n$, meaning $\sigma = \sqrt{n}$. Thus, $\mathbb{E}(|Y_n|) = \mathcal{O}(\sqrt{n})$ because then Y_n will be a constant number of standard deviations away from the mean, giving a constant probability of the random walk staying in bounds.

2 Chaining

2.1 Introduction

Before we show that $\mathbb{E}(\sup Y_n) = \mathcal{O}(\sqrt{n})$ (which is enough to claim that $\mathbb{E}(|\sup Y_n|)$ is also $\mathcal{O}(\sqrt{n})$), we show a weaker result $\mathbb{E}(\sup Y_n) \leq \mathcal{O}(\sqrt{n \log n})$ via chernoff bounds. Then, we lead up to the tighter bound by covering the Union Bound,ϵ-net argument, and Dudley's Inequality. Actually, these approaches work for bounding the supremum of any dot product $\sigma \cdot x$ with each entry of σ being $-1, 1$ with equal probability and x being some random vector. If $x = \{1, 1, ..., 1\}$, then the dot product $\sigma \cdot x$ is a random walk of size $|X|$. We prove this statement.

2.2 Chernoff Bounds

Suppose we have taken $m \leq n$ steps in a random walk Y_m. By Azuma-Hoeffding (and Chernoff), $Pr(|Y_m| > t) <= 2\exp(-t^2/(2m))$. Note, if $t = C\sqrt{n \log n}$, $Pr(|Y_n| > t) <= 2\exp(-C \cdot n \log n/(2m)) \leq 2\exp(-C \log n) = 2/n^C$, where we can make C arbitrary large. The probability that $Y_n > C\sqrt{n \log n}$ for all $1 \leq m \leq n$ is $n \cdot 2/n^c = 2/n^{c-1}$ which converges to 0 for sufficiently large random walks.

2.3 Union Bound

We use the union bound to establish a slightly worse bound than that can be done via chaining. Intuitively, the union bound states that for a finite number of events, the probability of at least one event happening is less than or equal to the sum of the probabilities of each event happening. Equality is only reached if the events are disjoint. This can be proved by Principle of Inclusion and Exclusion (PIE). By PIE, $P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n) - X$, where X is some combination of the sum of the intersection of events. In particular $X \geq 0$, so $P(A_1 \cup A_2 \cup ... \cup A_n) \leq P(A_1) + P(A_2) + ... + P(A_n)$.

We define $r(T)$ as $\mathbb{E}(\sup_{x\in T}(\sigma \cdot x))$, where T is the collection of vectors of size T of the form $\frac{1}{\sqrt{|T|}}\{1, 1, 1, ..., 1, 0, 0, ..., 0\}$, with $n1$ terms and n ranging from 0 to $|T|$. After normalizing the vector, we expect to show that $r(T) = \mathcal{O}(1)$.

Note that

$$\mathbb{E}(r(T)) \leq \mathbb{E}(|r(T)| = \int_0^\infty P(|r(T)| > u)du.$$

This is true because $\mathbb{E}X = \int_0^\infty P(X > x)dx$ due to integration by parts. Thus, we can have a sufficient upper bound by evaluating this integral. The difficulty in evaluating the integral is dealing with $P(|r(T)| > u)$ explicitly. Note that $P(|r(T)| > u) = P(\exists x \in T||\sigma \cdot x| > 0)$. We are left to evaluate

$$\int_0^\infty P(\exists x \in T||\sigma \cdot x| > u)du.$$

Now, we can employ the union bound. This integral is less than or equal to

$$\int_0^\infty \sum_{x\in T} P(|\sigma \cdot x| > u)du = \sum_{x\in T} \int_0^\infty P(|\sigma \cdot x| > u)du \leq 2|T| \int_0^\infty e^{-u^2/2}du.$$

The last inequality was found by applying Azuma-Hoeffding for random walks, or Khintchine inequality for random vectors generally, to all $x \in T$. That is, using the result from 1.1, we have $Pr(|\sigma \cdot x| > u) \leq 2\exp\left(-\sqrt{n} \cdot u^2/2m\right) \leq 2\exp\left(-u^2/2\right)$ for a random walk of any size m.

Actually, evaluating this integral directly gives $O(n)$, which is not a very good bound. To remedy this, we take advantage of the fact that a probability is at most 1. Instead, we split the initial integral

$$\int_0^\infty P(\exists x \in T||\sigma \cdot x| > u)du$$

into

$$\int_0^{C\sqrt{lg|T|}} P(\exists x \in T||\sigma \cdot x| > u)du + \int_{C\sqrt{lg|T|}}^\infty P(\exists x \in T||\sigma \cdot x| > u)du.$$

The LHS $\leq C\sqrt{lg|T|}$ and the RHS

$$\leq 2|T| \int_{C\sqrt{lg|T|}}^\infty e^{-u^2/2}du = \mathcal{O}(1).$$

This is because $\int_{C\sqrt{lg|T|}}^\infty e^{-u^2/2}$ can be interpreted as the probability that some random variable sampled from a normal distribution with mean 0 and constant variance is at least $C\sqrt{\log n}$ deviations from the mean for some constant C, which is approximately $1/n$. More rigorously, we can show that $S = x\int_{C\sqrt{\ln|x|}}^\infty e^{-u^2/2}du < A$ for some positive constants A and C by applying L'Hospital's rule and the second fundamental theorem of calculus. For example, this is true if $C = 5$ and $A = 2$. Actually, since the S strictly converges to 0, we can get a slightly tighter bound than $\sqrt{\log n}$. The sum of the bounds of both terms are known, and only the RHS is $O(1)$. The LHS is larger. That is, we don't achieve the bound of $\mathcal{O}(1)$ regardless.

2.4 ϵ-net argument

Epsilon nets are introduced to further reduce bounds. They utilize the property that $\sup(\sigma \cdot x)$ and $\sup(\sigma \cdot y)$ are close together if x and y are close together. Formally, an ϵ-net T' of T is defined as a subset of T such that any element x of T is at most ϵ away from its nearest member x' in T'. Clearly $\mathbb{E}(\sup_{x \in T}(\sigma \cdot x)) = \mathbb{E}(\sup_{x \in T}(\sigma \cdot (x' + (x - x')))) \leq r(T') + \mathbb{E}\sup(\sigma \cdot (x - x'))$. The last inequality is a result of the following intuition: The supremum of a random walk of a steps is max(supremum during first b steps, supremum during last $a - b$ steps) which is $\leq$ sup(first b steps) + the greatest movement that occurs during the last $a - b$ steps $= \sup b + \sup(a - b)$. Note that $\mathbb{E}\sup \sigma \cdot x' = r(T')$ by definition and that $\mathbb{E}\sup \sigma \cdot (x - x') \leq \mathbb{E}\sup ||\sigma|| \cdot ||(x - x')||$ by Cauchy-Schwartz. Recall that each entry of σ is either 1 or -1 with equal probability. In particular, $||\sigma|| = \sqrt{n}$. Also, $||(x - x')|| \leq \epsilon$ by definition of epsilon net. So, $\mathbb{E}\sup ||\sigma|| \cdot ||(x - x')|| = \mathbb{E}\sup \sqrt{n}\epsilon = \epsilon\sqrt{n}$.

Using the result from 2.3, $r(T') \leq C\sqrt{log|T'|}$. That is, we need to find an upper bound on the size of the epsilon net of T, referred to as K. Then, our total bound on $\sup_{x \in T} \sigma \cdot x \leq C\sqrt{\log K} + \sqrt{n}\epsilon$.

Recall T is the collection of vectors of size T of the form $1/\sqrt{|T|}\{1, 1, 1, ..., 1, 0, 0, ..., 0\}$, with $n1$ terms and n ranging from 0 to $|T|$. We claim K is at most $1/\epsilon^2$. Let T' be the set of x^t such that t is a multiple of $n \cdot \epsilon^2$ (we can set ϵ such that $n\epsilon^2$ is an integer). Now, consider x^r in T' closest to x^t in T. Since our ϵ-net contains every multiple of $n\epsilon^2$, x^t and x^r can differ in at most $n\epsilon^2$ positions. In each position the two vectors differ, one vectors contains a 0 and the other contains $1/\sqrt{n}$. Thus, the distance between them is $\sqrt{1/\sqrt{n})^2 \cdot n\epsilon^2} = \epsilon$, which guarantees T' is the ϵ-net of T. Since T contains n elements and T' contains every multiple of $n\epsilon^2$, $|T'| = n/(n\epsilon^2) = 1/\epsilon^2$. Thus, $K = 1/\epsilon^2$.

Substituting $1/\epsilon^2$ for K and taking the limit as ϵ goes to 0, we have $\sup_{x \in T} \sigma \cdot x \leq C\sqrt{\log K} + \sqrt{n}\epsilon = \sqrt{\log n}$, reaching the same bound as the union bound.

2.5 Dudley's Inequality

Dudley's inequality utilizes ϵ-nets, but to a greater extent. Instead of placing one net, we place several, further leveraging the idea that the expected supremums of random walks of similar length are also similar. Define S_k as a $1/2^k$ net of T and let x^k be the closest point to x in S_k. Then,

$$\sigma \cdot x = \sigma \cdot (x^0 + \sum_{k=1}^{\infty} x^k - x^{k-1}).$$

Now, using the results from 2.4, we have

$$\mathbb{E}\sup(\sigma \cdot \epsilon) \leq \mathbb{E}(\sigma \cdot x^0) + \mathbb{E}(\sup \sum_{k=1}^{\infty} \sigma \cdot (x^k - x^{k-1})) = \mathbb{E}(\sup \sum_{k=1}^{\infty} \sigma \cdot (x^k - x^{k-1})).$$

The last equality is because $\mathbb{E}(\sigma \cdot x^0) = \mathbb{E}(\sigma \cdot 0) = 0$. Note that

$$\mathbb{E}(\sup \sum_{k=1}^{\infty} \sigma \cdot (x^k - x^{k-1})) \leq \mathbb{E} \sum_{k=1}^{\infty} \sup(\sigma \cdot (x^k - x^{k-1}))$$

by the same intuition presented in 2.3. That is, if we arbitrarily partition a random walk into finitely many groups and add the supremum of each group, the sum will be at least as large as the supremum of the total walk. Also, note

$$\mathbb{E}\sum_{k=1}^{\infty}\sup(\sigma\cdot(x^k-x^{k-1}))\leq\sum_{k=1}^{\infty}\frac{1}{2^k}\sqrt{\log(A(1/2^k)\cdot A(1/2^{k-1}))},$$

where $A(u)$ is the size of the $u-$net. This is a direct result of the union bound presented in section 2.4. That is, we know by union bound that $\mathbb{E}(\sup_{x\in T}(\sigma\cdot x))\leq\sqrt{\log|T|}$. In the case of bounding $\mathbb{E}(\sup_{x\in T}(\sigma\cdot(x^k-x^{k-1})))$, we need to find $|T'|$, where $|T'|$ is the size of the set of (x^k,x^{k-1}). Then, $|T'|=A(2^k)\cdot A(2^{k-1})$ since we can independently pick a choice randomly from each net. The factor of $1/2^k$ arises from the fact that x^k and x^{k-1} differ from x in $1/2^k$ and $1/2^{k-1}$ positions respectively, so $x^k-x^{k-1}\leq 1/2^k+1/2^{k-1}=3/2^k$. We now reach the final result by algebraic manipulations. In section 2.4, we showed that $A(\epsilon)\leq 1/\epsilon^2$. Clearly then,

$$\sum_{k=1}^{\infty}\frac{1}{2^k}\sqrt{\log(A(1/2^k)\cdot A(1/2^{k-1}))}\leq\sum_{k=1}^{\infty}\frac{1}{2^k}\sqrt{\log A(1/2^k)}.$$

Since, $A(\epsilon)\leq 1/\epsilon^2$, $\sum_{k=1}^{\infty}\frac{1}{2^k}\sqrt{\log A(1/2^k)}\approx\sum_{k=1}^{\infty}\sqrt{k}/2^k$, which converges by ratio test. Thus, the supremum of $\sigma\cdot x$ converges to $\mathcal{O}(1)$ and we can say that the supremum of a random walk of n steps is approximately $O(\sqrt{n})$ which is our final conclusion.

Also, note that converting the sum to an integral and substituting $u=1/2^k$, we have the final result $r(T)$(as in, the notation used in 2.3) $\leq\int_0^{\infty}\sqrt{\log A(u)}du$, which is the usual statement of Dudley's Inequality.

3 Appendix

(*) https://www.sketchingbigdata.org/fall17/lec/lec8.pdf
(*) https://pdfs.semanticscholar.org/10c8/24c0920df7732c3258164669b6b2a8560267.pdf