PROBABILITY THEORY II

Matthew Joseph, notes by Ramdas Singh

Second Semester

List of Symbols

Ω, a sample space.

ω, an element of a sample space.

EX, the expectation of the random variable X.

$\mathrm{Var}X$, the variance of the random variable X.

$N(\mu, \sigma^2)$, a normal distribution with expectation μ and variance σ^2.

$N_n(k)$, the number of paths from $(0,0)$ to (n,k) in a simple random walk.

$N_n^+(k)$, the number of paths from $(0,0)$ to (n,k) through strictly positive values in a random walk.

p_k^X, the probability mass function for a random variable X.

Contents

Chapter 1

RANDOM WALKS AND MISC. RESULTS

January 3rd.

We first start with some initial statements. Let Ω be a countable state space, and let each $\omega \in \Omega$ have a probabiltiy $P(\omega)$ associated with it.

Lemma 1.1. *For random variables X, Y such that $X(\omega) \le Y(\omega)$ for all $\omega \in \Omega$. Then, $EX \le EY$.*

Proof. This can easily be seen by summing over all terms via the alternate definition of the expectation,

$$EX = \sum_{\omega \in \Omega} X(\omega)P(\omega) \le \sum_{\omega \in \Omega} Y(\omega)P(\omega) = EY. \tag{1.1}$$

■

We now state Markov's inequality.

Theorem 1.2 (*Markov's inequality*)**.** *If X is a non-negative randm variable, then for $a > 0$, we have*

$$P(X > a) \le \frac{EX}{a}. \tag{1.2}$$

Proof. Define an indicator function $I_a(\omega)$ as 1 if $X(\omega) \ge a$, and 0 if otherwise. We then have

$$I_a(\omega) \le \frac{X(\omega)}{a} \implies P(X \ge a) = EI_a \le \frac{1}{a}EX. \tag{1.3}$$

■

Remark 1.3. A better upper bound here may be found by starting with $I_a(\omega)X(\omega)$ instead of just $X(\omega)$.

If we have $X \sim N(0,1)$, then we can find an upper bound for its probability density function.

$$P(X > a) = \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\,dx \le \int_a^\infty \frac{1}{\sqrt{2\pi}} \frac{x}{a} e^{\frac{-x^2}{2}}\,dx = \frac{e^{\frac{-a^2}{2}}}{\sqrt{2\pi}a}. \tag{1.4}$$

Note that X here is a random variable over a continuous state space; the previous lemma and Markov's inequality also work here. We are to show them for the continuous case instead of the discrete one.

Proof. Here, we have $0 \le X(\omega) \le Y(\omega)$ for all ω in our continuous state space Ω. We see that $\{X > x\} \subseteq \{Y > x\} \implies P(X > x) \le P(Y > x)$. Integrating both sides gives us $EX \le EY$. ■

Theorem 1.4 (*Chebyshev's inequality*)**.** *Let X be a random variable with finite mean $\mu = EX$ and finite variance $\sigma^2 = Var(X)$. Then for $a > 0$,*

$$P(|X - \mu| > a) \le \frac{Var(X)}{a^2}. \tag{1.5}$$

Proof. Start with the proof of Markov's inequality, replacing the indiciator function with one that's unity when $|X - \mu| \geq a$. ■

Example 1.5. Suppose $X_1, X_2, \ldots, X_n$ are n independent and identically distributed random variables, with $EX_i = \mu$ and $\mathrm{Var}X_i = \sigma^2$. If $S_n = \sum X_i$, we then have

$$P(|S_n - n\mu| > a) \leq \frac{\mathrm{Var}S_n}{a^2} = \frac{n\sigma^2}{a^2}. \tag{1.6}$$

If we replace a with $n^{\frac{1}{2}+\varepsilon}$, we then have

$$P(|S_n - n\mu| > n^{\frac{1}{2}+\varepsilon}) \leq \frac{\sigma^2}{n^{2\varepsilon}} \to 0 \text{ as } n \to \infty. \tag{1.7}$$

Proposition 1.6. *If Var$(X) = 0$, then $P(X = EX) = 1$.*

Proof. For all $\varepsilon > 0$, we have

$$P(|X - EX| > \varepsilon) \leq \frac{\mathrm{Var}X}{\varepsilon^2} = 0. \tag{1.8}$$

Define A_n as $\{|X - EX| > \frac{1}{n}\}$. Taking $P(\bigcup A_n) = \lim_{n\to\infty} P(A_n)$, the proof follows. ■

1.1 The Law of Large Numbers

We start by stating the weak law of large numbers.

Theorem 1.7 (*Weak law of large numbers*)**.** *Let $\{X_k\}_{k\geq 1}$ be a sequence of independent and identically distributed random variables with $E|X_i| < \infty$. Let $\mu = EX_i$. Then for any $a > 0$,*

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| > a\right) = 0. \tag{1.9}$$

Proof. For now, let us assume that Ω is countable. We begin with the case where the variance of X_i, σ^2, is finite. Fix $a > 0$, and let $S_n = X_1 + X_2 + \ldots + X_n$. Then,

$$P\left(\left|\frac{S_n}{n} - \mu\right| > a\right) = P(|S_n - n\mu| > na) \leq \frac{\mathrm{Var}S_n}{n^2a^2} = \frac{n\sigma^2}{n^2a^2} \to 0 \text{ as } n \to \infty. \tag{1.10}$$

We now focus the case when the variance, σ^2, is infinite. Assume that the expected value, μ, is 0; if it were non-zero, we would then instead work with $X_i - \mu$. Let $\delta > 0$; we shall choose a particular δ later. For each n, define n pairs of random variables, $U_1, V_1, \ldots, U_n, V_n$, as $U_k = X_k, V_k = 0$ if $|X_k| \leq \delta n$, and $U_k = 0, V_k = X_k$ if $|X_k| > \delta n$. X_k can be rewritten as $U_k + V_k$. We then have

$$\{|X_1 + \ldots + X_n| \geq na\} \subseteq \{|U_1 + \ldots + U_n| \geq \frac{na}{2}\} \cup \{|V_1 + \ldots + V_n| \geq \frac{na}{2}\} \tag{1.11}$$

$$\implies P\left(|X_1 + \ldots + X_n| \geq na\right) \leq P\left(|U_1 + \ldots + U_n| \geq \frac{na}{2}\right) + P\left(|V_1 + \ldots + V_n| \geq \frac{na}{2}\right). \tag{1.12}$$

We focus on the first term on the right hand side. The U_i's are independently and identically distributed, so

$$P\left(|U_1 + \ldots + U_n| \geq \frac{na}{2}\right) \leq \frac{4E[|U_1 + \ldots + U_n|^2]}{a^2n^2} = \frac{4}{a^2n^2}\left(\mathrm{Var}(U_1 + \ldots + U_n) + (nEU_i)^2\right). \tag{1.13}$$

For the variance, we have

$$\mathrm{Var}(U_1 + \ldots + U_n) = n\mathrm{Var}U_i \leq nEU_i^2 \leq nE[|U_i|\,|U_i|] \leq \delta n^2 E[|U_i|] \tag{1.14}$$

which transforms the previous equation as

$$P\left(|U_1 + \ldots + U_n| \geq \frac{na}{2}\right) \leq \frac{4}{a^2n^2}\left(\delta n^2 E[|U_i|] + (nEU_i)^2\right). \tag{1.15}$$

A lemma (to be proven later) states that $E[|U_i|] = E[|X_i|]$ as $n \to \infty$, and $EU_i = EX_i = 0$ too. So,

$$P\left(|U_1 + \ldots + U_n| \geq \frac{na}{2}\right) \leq \frac{4}{a^2n^2}\left(\delta n^2 E[|U_i|] + (nEU_i)^2\right) \leq \frac{4\delta E[|U_i|]}{a^2} + \frac{4}{a^2}(EU_i)^2. \tag{1.16}$$

For the second term on the right hand side, begin with

$$P(V_1 + \ldots + V_n \neq 0) \leq P(\{V_1 \neq 0\} \cup \ldots \cup \{V_n \neq 0\}) \leq nP(V_i \neq 0) = n \sum_{|x|>\delta n} P(X_i = x)$$

$$\leq n \sum_{|x|>\delta n} \frac{|x|}{\delta n} P(X_i = x) = \frac{1}{\delta} E[|V_i|]. \tag{1.17}$$

The rightmost term here tends to 0 as $n \to \infty$. Now choose δ to be $\frac{\varepsilon a^2}{|6E|X_i||}$, and then choose N to be large enough such that for all $n > N$, both the terms are smaller than $\frac{\varepsilon}{2}$. ■

January 7th.

We now prove the lemma called upon earlier.

Lemma 1.8. *If X is a discrete random variable and takes values $y_1, y_2, \ldots, y_k$, and $E[|X|] < \infty$, then $\lim_{n\to\infty} E[|X|\, 1_{|X|\leq n}] = E[|X|]$.*

Proof. Notice that the terms on the left hand side and right hand side are $\sum_{y_k:|y_k|\leq n}$ and $\sum_{y_k} |y_k|\, P(Y = y_k)$. The condition for convergence may now be applied. ■

The above equation, begin inside absolute braces, must imply that the term $E[X \cdot 1_{|X|\leq n}]$ must also absolutely converge to EX.

1.2 Simple Random Walk

Let $X_1, X_2, \ldots$ be independent and identically distributed random variables, with $X_i = 1$ with probability $\frac{1}{2}$ and $X_i = -1$ with probability $\frac{1}{2}$. Now define $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$. The sequence $(S_n)_{n\geq 0}$ is a *simple random walk*.

Note that $S_0 = k_0 = 0, S_1 = k_1, \ldots, S_n = k_n$ can occur if and only if $|k_i - k_{i+1}| = 1$ for all $0 \leq i \leq n-1$. The sequence $(k_n)_{n\geq 0}$ is a *simple path* of the simple random walk. By the event $\{S_n = k\}$, we are concerned with the event that the random walk visits k at step n. If $(k_n)_{n\geq 0}$ is given we have $X_i = k_i - k_{i-1}$. Because the X_i's are independent and identically distributed, each event $\{X_1 = l_1, X_2 = l_2, \ldots, X_n = l_n\}$, where $l_i = \pm 1$, is equally likely with probability $\frac{1}{2^n}$. Thus,

$$P(S_n = k) = \frac{N_n(k)}{2^n} \tag{1.18}$$

where $N_n(k)$ is defined as the number of distinct of path that start at 0 and end at k at step n. We also define $N_n^+(k)$ to be the number of distinct paths that end at k at step n and stay above the x-axis up to time $n-1$. The probability of the corresponding event is

$$P(\{S_1 > 0, S_2 > 0, \ldots S_{n-1} > 0, S_n = k\}) = \frac{N_n^+(k)}{2^n}. \tag{1.19}$$

Lemma 1.9. *Suppose a, a', b, b' are integers, with $0 \leq a < a'$. Then the number of distinct path from (a, b) to (a', b') depends only on $a' - a = n$ and $b' - b = k$, and is given by $\binom{n}{\frac{n+k}{2}}$.*

Proof. Notice that we need x $+1$'s and y -1's to appear, satisfying $x + y = a' - a$ and $x - y = b' - b$. Solving, we get $x = \frac{n+k}{2}$ and $y = \frac{n-k}{2}$. Thus, the number of paths is given by $\binom{n}{\frac{n+k}{2}}$. ■

Using this lemma, we find that $N_n(k) = \binom{n}{\frac{n+k}{2}}$. The following convention is now followed; if t is not an integer, then $\binom{n}{t} = 0$.

Lemma 1.10 (The *method of images*)**.** *Suppose a, a', b, b' are integers, with $0 \leq a < a'$ and $b, b' > 0$. Then the number of distinct paths from (a, b) to (a', b') that intersect the x-axis is equal to the number of paths from $(a, -b)$ to (a', b').*

Proof. Consider any path $(b = k_0, k_1, \ldots, k_{n-1}, k_n = b')$, from (a, b) to (a', b'), that intersects the x-axis. Let j be the smallest index for which $k_j = 0$. For ease, denote (a, b) by A, (a', b') by A', $(a + j, 0)$ by B, and $(a, -b)$ by A''. Reflect the segment from A to B about the x-axis to obtain a 'mirrored-path' from A'' to B; $(-b = -k_0, -k_1, \ldots, -k_{j-1}, k_j = 0, k_{j+1}, \ldots, k_n = b')$. There is now a one-to-one correspondence between the paths from A to A' that intersect the x-axis, and the paths from A'' to A'. ■

We can now easily compute $N_n^+(k)$; it simply the number of paths from $(1,1)$ to (n,k) that do not intersect the x-axis.

Theorem 1.11 (*Ballot theorem*)**.** *The number of paths that progress from $(0,0)$ to (n,k) through strictly positive values is given by $N_n^+(k) = \frac{k}{n}N_n(k)$.*

Proof. We have

$$\begin{aligned} N_n^+(k) &= \text{ number of paths from } (1,1) \text{ to } (n,k) - \text{ number of such paths that intersect the } x\text{-axis} \\ &= N_{n-1}(k-1) - N_{n-1}(k+1) \\ &= \binom{n-1}{\frac{n+k}{2}-1} - \binom{n-1}{\frac{n+k}{2}} = \frac{k}{n}\binom{n}{\frac{n+k}{2}} = \frac{k}{n}N_n(k). \end{aligned} \tag{1.20}$$

■

Suppose $n = 2\nu$. Define $u_{2\nu}$ to be $P(S_{2\nu} = 0) = \frac{\binom{2\nu}{\nu}}{2^n}$. The question we ask is to compute the probability that the first return to 0, if at all, occurs after step n. It can be found out as

$$\begin{aligned} P(\text{first return to 0...}) &= P(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2\nu} \neq 0) \qquad (1.21) \\ &= P(S_1 > 0, \ldots, S_{2\nu} > 0) + P(S_1 < 0, \ldots, S_{2\nu} < 0) \\ &= 2P(S_1 > 0, \ldots, S_{2\nu} > 0) \\ &= 2 \sum_{k \text{ even}, k>0} P(S_1 > 0, \ldots, S_{2\nu-1} > 0, S_{2\nu} = k) \\ &= \frac{2}{2^{2\nu}} \sum_{k \text{ even}, k>0} N_{2\nu}^+(k) \\ &= \frac{2}{2^{2\nu}} \sum_{k \text{ even}, k>0} N_{2\nu-1}(k-1) - N_{2\nu-1}(k+1) \\ &= \frac{2}{2^{2\nu}} N_{2\nu-1}(1) = u_{2\nu}. \qquad (1.22) \end{aligned}$$

We state this down as a lemma.

Lemma 1.12 (*Basic lemma*)**.** *For n even, the probability that the first return to 0, if at all, occurs after step n is the same as the probability that the location at step n is 0. For n odd, it is the probability that the location at step $n-1$ is 0.*

We ask another question; for a fixed n, where does the random walk achieve its first maximum upto time n? For this, denote by M_n the index m at which the walk $S_0, S_1, \ldots, S_n$, over n steps, achieves its maximum for the first time.

For $0 < m < n$, $M_n = m$ if and only if $S_m > S_0,\ S_m > S_1, \ldots, S_m > S_{m-1}$ and $S_m \geq S_{m+1},\ S_m \geq S_{m+2}, \ldots, S_m \geq S_n$. Notice that the first of these two conditions depends only on $X_1, X_2, \ldots, X_m$, and the second condition depends only on $X_{m+1}, X_{m+2}, \ldots, X_n$. So, $P(M_n = m) = P(S_m > S_0,\ S_m > S_1, \ldots, S_m > S_{m-1}) \cdot P(S_m \geq S_{m+1},\ S_m \geq S_{m+2}, \ldots, S_m \geq S_n)$.

The key idea here is to consider the *reversed walk*; define a new walk with $X_1' = X_m,\ X_2' = X_{m-1}, \ldots, X_m' = X_1$. Also define $S_k' = X_1' + \ldots + X_k'$. From here, we can deduce that $S_m > S_{m-i}$ is true if and only if $X_m + \ldots + X_{m-i} > 0$ is true, which is true if and only if $S_i' > 0$ is true. So, $P(S_m > S_0,\ S_m > S_1, \ldots, S_m > S_{m-1}) = P(S_1' > 0,\ S_2' > 0, \ldots, S_m' > 0)$. If we now define $S_k'' = X_{m+1} + \ldots + X_{m+k}$, we have

$$\begin{aligned} P(S_m \geq S_{m+1},\ S_m \geq S_{m+2}, \ldots, S_m \geq S_n) &= P(X_{m+1} \leq 0,\ X_{m+1} + X_{m+2} \leq 0, \ldots, X_{m+1} + \ldots + X_n \leq 0) \\ &= P(S_1'' \leq 0,\ S_2'' \leq 0, \ldots, S_{n-m}'' \leq 0) \\ &= P(S_1'' \geq 0,\ S_2'' \geq 0, \ldots, S_{n-m}'' \geq 0) \end{aligned}$$

The first of the terms discussed, $P(S_1' > 0,\ S_2' > 0, \ldots, S_m' > 0)$, can be computed for $m = 2\nu, 2\nu+1$; it is simply $\frac{1}{2}u_{2\nu}$. For the latter of these terms, we introduce a new random variable $\tilde{X}$ which has the same distribution as the X_i's and is independent. Also define $\tilde{S}_i$ to be $\tilde{X} + X_1 + \ldots + X_{i-1}$ and $\tilde{S}_0$ to be 0.

We then have

$$\begin{aligned}\frac{1}{2}P(S_0 \geq 0, \ldots, S_{n-m} \geq 0) &= P(\tilde{X} = 1) \cdot P(S_0 \geq 0, \ldots, S_{n-m} \geq 0) \\ &= P(\tilde{X} = 1, S_0 \geq 0, S_0 \geq 0, \ldots, S_{n-m} \geq 0) \\ &= P(\tilde{S_1} = 1, \tilde{S_2} > 0, \ldots, \tilde{S}_{n-m+1} > 0) \\ &= P(S_1 > 0, S_2 > 0, \ldots, S_{n-m+1} > 0). \end{aligned} \tag{1.23}$$

Thus, we get

$$P(M_n = m) = \frac{1}{2}u_{2k}u_{2\nu-2k} \tag{1.24}$$

where m is of the form $2k$ or $2k+1$, and n is of the form 2ν, with $1 < k < \nu$.

January 10th.

Plugging in $m = 0$, we get $P(M_n = 0) = P(S_1 \leq 0, \ldots, S_{2\nu} \leq 0) = \frac{1}{2}u_{2\nu}$. For $m = n$, we have $P(M_n = n) = P(S_1 \leq 0, \ldots, S_{2\nu} \leq 0) = \frac{1}{2}u_{2\nu}$. Let us first compute u_{2k}.

$$\begin{aligned} u_{2k} = P(S_{2k} = 0) = \frac{\binom{2k}{k}}{2^{2k}} &= \frac{(2k)!}{(k!)^2 2^{2k}} \\ \sim \frac{(2k)^{2k+\frac{1}{2}}e^{-2k}\sqrt{2\pi}}{(\sqrt{2\pi}k^{k+\frac{1}{2}}e^{-k})^2 2^{2k}} &= \frac{1}{\sqrt{\pi k}}. \end{aligned} \tag{1.25}$$

For $0 < a < b < 1$, we have

$$\begin{aligned} P(an \leq M_n \leq bn) &= \sum_{m=an}^{bn} P(M_n = m) = \sum_{k=a\nu}^{b\nu} u_{2k}u_{2\nu-2k} \\ &\sim \sum_{k=a\nu}^{b\nu} \frac{1}{\sqrt{\pi k}}\frac{1}{\sqrt{\pi(\nu-k)}} = \sum_{k=a\nu}^{b\nu} \frac{1}{\nu\sqrt{\pi\frac{k}{\nu}}\sqrt{\pi(1-\frac{k}{\nu})}} \\ &\to \frac{1}{\pi}\int_a^b \frac{dx}{\sqrt{x(1-x)}} = \frac{2}{\pi}(\arcsin\sqrt{b} - \arcsin\sqrt{a}). \end{aligned} \tag{1.26}$$

In fact, this is the *arcsin law for maxima*; for $0 \leq t \leq 1$, we have

$$\lim_{n\to\infty} P\left(\frac{M_n}{n} \leq t\right) = \frac{2}{\pi}\arcsin\sqrt{t}. \tag{1.27}$$

If we look at this as a cumulative density funtion, the probability density function becomes $\frac{d}{dt}\frac{2}{\pi}\arcsin\sqrt{t} = \frac{1}{\pi\sqrt{t(1-t)}}$.

We are now interested in $\tilde{M_n}$, the last time when maximum up to time n is attained. We can just look at the walk backwards again; in this case, we get

$$P\left(\frac{\tilde{M_n}}{n}\right) = P\left(\frac{n - \tilde{M_n}}{n} \leq t\right) \to \frac{2}{\pi}\arcsin\sqrt{t}. \tag{1.28}$$

We now ask the probability that the random walk of $n = 2\nu$ steps last visit 0 at time $2k$. We denote by K_n the location of the last return to 0 in a walk of n steps. Now look at

$$\begin{aligned} \alpha_{2k,2\nu} = P(K_n = 2k) &= P(S_{2k} = 0, S_{2k+1} \neq 0, \ldots, S_{2\nu} \neq 0) \\ &= P(S_{2k} = 0) \cdot P(X_{2k+1} \neq 0, \ldots, X_{2k+1} + \ldots + X_{2\nu} \neq 0) \\ &= P(S_{2k} = 0) \cdot P(S_1 \neq 0, \ldots, S_{2\nu-2k} \neq 0) = u_{2k}u_{2\nu-2k}. \end{aligned} \tag{1.29}$$

We can also state an *arcsin law for last visit* here; for $0 < t < 1$

$$\lim_{n\to\infty} P(K_n \leq tn) = \frac{2}{\pi}\arcsin\sqrt{t}. \tag{1.30}$$

If we set the an additional limit that says t tends to 0, replacing t by an arbitrary $\varepsilon > 0$, we have

$$\lim_{n\to\infty} P(K_n = 0) = 0. \tag{1.31}$$

Given enough time, a simple random walk must return to 0.

Denote by f_{2n} the probability that the first return to 0 occurs at time $2n$.

$$\begin{aligned} f_{2n} &= P(S_1 \neq 0, \dots, S_{2n-1} \neq 0, S_{2n} = 0) \\ &= P(S_1 \neq 0, \dots, S_{2n-1} \neq 0) - P(S_1 \neq 0, \dots, S_{2n} \neq 0) \\ &= P(S_1 \neq 0, \dots, S_{2n-2} \neq 0) - P(S_1 \neq 0, \dots, S_{2n} \neq 0) \\ &= u_{2n-2} - u_{2n} = \frac{1}{2n-1} u_{2n}. \end{aligned} \tag{1.32}$$

Lemma 1.13. *With the usual notation,*

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + \dots + f_{2n} u_0. \tag{1.33}$$

Proof. We have

$$\begin{aligned} P(S_{2n} = 0) &= \sum_{k=1}^{n} P(S_{2n} = 0, \text{ first return at } 2k) \\ &= \sum_{k=1}^{n} P(\text{first return at 2k}) \cdot P(S_{2n} = 0 \mid \text{first return at 2k}) \\ \implies P(S_n = 0) &= \sum_{k=1}^{n} f_{2k} u_{2n-2k}. \end{aligned} \tag{1.34}$$

■

Theorem 1.14. *The probability that in the time interval* 0 *to* $n = 2\nu$*, the random walk spends* $2k$ *amount of time on the positive side and* $2\nu - 2k$ *amount of time on the negative side is* $\alpha_{2k,2\nu}$*.*

Corollary 1.15. *For* $0 < t < 1$*,*

$$P(\textit{random walk spends less than tn time on positive side}) \to \frac{2}{\pi} \arcsin \sqrt{t}. \tag{1.35}$$

Proof. This is the proof of the theorem. We introduce $b_{2k,2\nu}$; it is defined as the probability that the random walk of length 2ν and $2k$ sides above the x-axis. We need to show that $b_{2k,2\nu} = \alpha_{2k,2\nu}$. We have

$$b_{2\nu,2\nu} = P(S_1 \geq 0, S_2 \geq 0, \dots, S_{2\nu} \geq 0) = u_{2\nu}, \tag{1.36}$$

$$b_{0,2\nu} = P(S_1 \leq 0, \dots, S_{2\nu} \leq 0) = u_{2\nu}. \tag{1.37}$$

We are left to prove it for $1 \leq k \leq \nu - 1$. Assume that exactly $2k$ out of 2ν time are spent above the x-axis, with $1 \leq k \leq \nu - 1$. Suppose first return to 0 occurs at time $2r < 2\nu$. We deal in cases.

- Case I: $2r$ time units upto first return are on the positive side. Then, $r \leq k \leq \nu - 1$. The time from $2r$ to 2ν has to be above the x-axis, $2k - 2\nu$ time. The number of such paths is $(\frac{1}{2} 2^{2r} f_{2r})(2^{2\nu-2r} b_{2k-2r,2\nu-2r})$.
- The $2r$ time units upto the first return are on the negative side. The nubmer of such paths is $(\frac{1}{2} 2^{2r} f_{2r})(2^{2\nu-2r} b_{2k,2\nu-2r})$. Also, $\nu - r \geq k$.

Thus, we have

$$b_{2k,2\nu} = \frac{1}{2} \sum_{r=1}^{k} f_{2r} b_{2k-2r,2\nu-2r} + \frac{1}{2} \sum_{r=1}^{\nu-k} f_{2r} b_{2k,2\nu-2r}. \tag{1.38}$$

We now proceed with induction on ν. We have already shown this for $\nu = 1$; assume that this is true for $\nu \leq V - 1$. By induction,

$$\begin{aligned} b_{2k,2V} &= \frac{1}{2} \sum_{r=1}^{k} f_{2r} \alpha_{2k-2r,2V-2r} + \frac{1}{2} \sum_{r=1}^{V-k} f_{2r} \alpha_{2k,2V-2r} \\ &= \frac{1}{2} u_{2V-2k} \sum_{r=1}^{k} f_{2r} u_{2k-2r} + \frac{1}{2} u_{2k} \sum_{r=1}^{V-k} f_{2r} u_{2V-2k-2r} \\ &= u_{2k} u_{2\nu-2k} = \alpha_{2k,2\nu}. \end{aligned} \tag{1.39}$$

■

January 17th.

Theorem 1.16 (*Weirstrass's polynomial approximation.*)**.** *Let $f : [0,1] \to \mathbb{R}$ be a continuous function. Then for every $\varepsilon > 0$, there is a polynomial P, dependent on f and ε, such that*

$$|f(x) - P(x)| < \varepsilon \textit{ for all } x \in [0,1]. \tag{1.40}$$

Remark 1.17. Any continuous function $f : [0,1] \to \mathbb{R}$ is bounded and uniformly continuous. This fact will be useful in proving the previous theorem.

Proof. Start with $X_1, X_2, \ldots$ which are independent and identically disitributed Bernoulli random variables, $\mathrm{Ber}(x)$. Let $S_n = X_1 + X_2 + \ldots + X_n$. From the weak law of large numbers, we know that $\frac{S_n}{n}$ is approximately x. We can expect that $f(x)$ will also be approximately $f\left(\frac{S_n}{n}\right)$. We now have

$$\begin{aligned} f_n(x) = Ef\left(\frac{S_n}{n}\right) &= \sum_{j=0}^{n} f\left(\frac{j}{n}\right) P(S_n = j) \\ &= \sum_{j=0}^{n} f\left(\frac{j}{n}\right) \binom{n}{j} x^j (1-x)^{n-j}. \end{aligned} \tag{1.41}$$

This is now a polynomial; we wish to see how close this is to f. Define A_δ to be $\{j : \left|\frac{j}{n} - x\right| \le \delta\}$

$$\begin{aligned} |f_n(x) - f(x)| &= \left|\sum_{j=0}^{n} \left(f\left(\frac{j}{n}\right) - f(x)\right)\right| P(S_n = j) \\ &= \left|\sum_{j\in A_\delta} \left(f\left(\frac{j}{n}\right) - f(x)\right) + \sum_{j\notin A_\delta} \left(f\left(\frac{j}{n}\right) - f(x)\right)\right| P(S_n = j) \\ &\le \sum_{j\in A_\delta} \left|f\left(\frac{j}{n}\right) - f(x)\right| P(S_n = j) + \sum_{j\notin A_\delta} \left|f\left(\frac{j}{n}\right) - f(x)\right| P(S_n = j). \end{aligned} \tag{1.42}$$

We have two terms to deal with now. For the first term, choose $\delta > 0$ such that $|x - y| < \delta \implies |f(x) - f(y)| < \varepsilon$; this δ can be chosen since f is uniformly continuous. Similarly, also choose $M = \sup_{x\in[0,1]} |f(x)|$. M is finite since f is bounded. Thus, we have

$$\sum_{j\in A_\delta} \left|f\left(\frac{j}{n}\right) - f(x)\right| P(S_n = j) \le \sum_{j\in A_\delta} \varepsilon P(S_n = j) \le \varepsilon \tag{1.43}$$

and

$$\sum_{j\notin A_\delta} \le 2MP\left(\left|\frac{S_n}{n} - x\right| > \delta\right) \le 2M\frac{\mathrm{Var}(S_n)}{n^2\delta^2} = \frac{2Mnx(1-x)}{n^2\delta^2}. \tag{1.44}$$

Combining the two, and choosing n large enough, we have

$$|f_n(x) - f(x)| \le \varepsilon + \frac{2Mx(1-x)}{n\delta^2} \le \varepsilon + \frac{M}{2n\delta^2} \le 2\varepsilon. \tag{1.45}$$

■

1.3 Erdös-Renyi Random Graph

We first discuss the setup; start with n vertices of an empty graph. For any pair of points (i, j), with $i \neq j$, join these vertices with an edge with probability p independently for all such pairs. Such a graph is denoted by $G_{n,p}$.

A collection of three points $S = \{i, j, k\}$ form a triangle if $G_{n,p}$ has the edges $\{i, j\}$, $\{j, k\}$, and $\{i, k\}$. We question the probability that such a graph has no formed triangles. Can we find $p = p_n$ such that triangles begin to appear at p_n? Let S be any set of three vertices. Define X_S to be the indicator function; 1 if S forms a triangle, and 0 otherwise. We note that $X_S \sim \text{Ber}(p^3)$. We note that

$$EX_S = p^3,\ \text{Var}X_S = p^3(1-p^3) \le p^3.$$

Denote by N the number of triangles in the graph $G_{n,p}$. Clearly,

$$N = \sum_{S:|S|=3} X_S,\ EN = \binom{n}{3}p^3 < n^3p^3,\ \text{Var}N = \sum_S \text{Var}X_S + \sum_S \sum_{T\neq S} \text{Cov}(X_S X_T) \le n^3p^3 + n^4p^5$$

ALso, $P(N \ge 1) \le EN < n^3p^3$. If $p = p_n << \frac{1}{n}$, then $P(N \ge 1) \to 0$ as $n \to \infty$. We discuss this for $p >> \frac{1}{n}$. We have

$$P(N=0) \le P(|N-EN| \ge EN) \le \frac{\text{Var}N}{(EN)^2} \le \frac{(n^3p^3+n^4p^5)}{\frac{n^6p^6}{100}} \le \frac{100}{n^3p^3} + \frac{100}{n(np)} \to 0. \tag{1.46}$$

We can state this as a theorem.

Theorem 1.18. *Consider G_{n,p_n}. Let E be the event that the graph is triangle free. We then have*

$$P(E) \to \begin{cases} 0 & \text{if } \frac{p_n}{\frac{1}{n}} \to \infty, \\ 1 & \text{if } \frac{p_n}{\frac{1}{n}} \to 0. \end{cases} \tag{1.47}$$

Now suppose that $\frac{np_n}{\to} C > 0$ as $n \to \infty$. Then we have

$$N \approx \text{Poisson}\left(\frac{C^3}{6}\right). \tag{1.48}$$

January 21st.

Remark 1.19. For this next ‘game’, we will think of X_i’s as the winnings in game i and μ to be the entrance fees for a game.

Definition 1.20. Suppose that $X_1, X_2, \ldots$ are independent, but not necessarily identically distributed. Let $S_n = X_1 + \ldots + X_n$. We say a game with accumulated entrance fees $\{\alpha_n, n \ge 1\}$ is fair if

$$P(\left|\frac{S_n}{\alpha_n} - 1\right| > \varepsilon) \to 0 \tag{1.49}$$

for all $\varepsilon > 0$.

Using this definition of ‘fair’, we look at an example.

Example 1.21. This is the St. Petersburg’s paradox. This is the game; toss a coin repeatedly until the first head is observed. If this head occurs at the k^{th} toss, the amount paid out is $X = 2^k$. Let us find a fair accumulated entrance fees. In this case,

$$EX = \sum_{k=1}^{\infty} \frac{1}{2^k} 2^k = \infty. \tag{1.50}$$

Suppose we play this game n times. We are to find a fair accumulated sum $\{\alpha_n\}$ such that

$$P(|S_n - \alpha_n| > \varepsilon\alpha_n) \to 0. \tag{1.51}$$

To find this, we will define

$$U_j = X_j 1_{\{X_j \le a_n\}},$$
$$V_j = X_j 1_{\{X_j > a_n\}}.$$

a_n shall be determined later. Note that $S_n = X_1 + \ldots + X_n = U_1 + \ldots + U_n + V_1 + \ldots V_n$. Then,

$$P(|S_n - \alpha_n| > \varepsilon\alpha_n) \le P(|U_1 + \ldots + U_n - \alpha_n| > \frac{1}{2}\varepsilon\alpha_n) + P(|V_1 + \ldots + V_n| > \frac{1}{2}\varepsilon\alpha_n). \tag{1.52}$$

We first bound the second term on the right hand side. We have

$$P(|V_1 + \ldots + V_n| > \frac{1}{2}\varepsilon\alpha_n) \leq P(\bigcup_{i=1}^{n}\{V_i \neq 0\}) \leq nP(V_1 \neq 0) = nP(X_1 > a_n) \tag{1.53}$$

$$= n \sum_{2^k > a_n} P(X = 2^k) \leq \frac{2n}{a_n}. \tag{1.54}$$

Thus, we will require that $a_n >> n$. Also,

$$EU_1 = \sum_{k \leq \log_2 a_n} 2^k \cdot 2^{-k} = \lfloor \log_2 a_n \rfloor, \ \mathrm{Var}U_1 \leq E[U_1^2] = \sum_{k \leq \log_2 a_n} (2^k)^2 \cdot 2^{-k} = 2^{\lfloor \log_2 a_n \rfloor + 1} - 1 < 2a_n. \tag{1.55}$$

$\frac{1}{n}(U_1 + \ldots + U_n) \approx EU_j = \lfloor \log_2 a_n \rfloor$, so we should choose

$$\alpha_n = nEU_j = n \lfloor \log_2 a_n \rfloor. \tag{1.56}$$

This gives us

$$P(|U_1 + \ldots + U_n - \alpha_n| > \frac{1}{2}\varepsilon\alpha_n) \leq \frac{n(2a_n)}{\frac{1}{4}\varepsilon^2\alpha_n^2}. \tag{1.57}$$

Thus, we have another condition where we require that $\frac{na_n}{\alpha_n^2} \to 0$. The conditions we require are

$$\frac{n}{a_n} \to 0 \text{ and } \frac{na_n}{n^2(\log_2 a_n)^2} \to 0.$$

The sequence $\{a_n\}$ defined as $a_n = n\log_2 n$ satisfies these properties. The sequnce α_n is thus

$$\alpha_n = n\log_2 a_n = n\log_2 n + n\log_2\log_2 n. \tag{1.58}$$

Chapter 2

GENERATING FUNCTIONS

January 24th.

Definition 2.1. For a sequence $\{a_n\}_{n\geq 0}$, the *generating function* of $\{a_n\}$ is given as

$$A(s) = \sum_{n=0}^{\infty} a_n s^n \tag{2.1}$$

for some $-s_0 < s < s_0$.

For this probability course, we will be interested in a particular form; for a random variable X that takes values $k = 0, 1, \ldots$, the function we look at is

$$\sum_{k=0}^{\infty} P(X = k)s^k \text{ for } -1 \leq s \leq 1. \tag{2.2}$$

Suppose we have two sequences $\{a_n\}$ and $\{b_n\}$ with generating functions $A(s)$ and $B(s)$, respectively. If we define a new sequence $\{c_n\}$ as

$$c_n = a_0 b_n + a_1 b_{n-1} + \ldots + a_{n-1} b_1 + a_n b_0 \text{ for all } n \geq 0, \tag{2.3}$$

then the sequence $\{c_n\}$ is termed the *convolution* of the sequences $\{a_n\}$ and $\{b_n\}$, and we shall denote it as

$$\{c_n\} = \{a_n\} * \{b_n\}.$$

Note that this convolution operation is both associative and commutative. We are now interested in finding the generating function of $\{c_n\}$. We have

$$\begin{aligned} C(s) &= \sum_{n=0}^{\infty} c_n s^n = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_k b_{n-k} \right) s^n \\ &= \sum_{n=0}^{\infty} \sum_{k=0}^{n} a_k s^k b_{n-k} s^{n-k} = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} a_k s^k b_m s^m \\ \implies C(s) &= \left(\sum_{k=0}^{\infty} a_k s^k \right) \cdot \left(\sum_{m=0}^{\infty} b_m s^m \right) = A(s) \cdot B(s). \end{aligned} \tag{2.4}$$

We state this down as a theorem.

Theorem 2.2. $C(s) = A(s) \cdot B(s)$ *when* $\{c_n\} = \{a_n\} * \{b_n\}$.

Suppose X takes values in $\mathbb{Z}_+ = \{0, 1, \ldots\}$. Denote $P(X = k)$ as p_k. The generating function is, thus,

$$\mathcal{P}(s) = \sum_{k=0}^{\infty} p_k s^k = E[s^X].$$

Also,

$$\mathcal{P}(1) = 1, \tag{2.5}$$

$$\mathcal{P}'(1) = \sum_{k=1}^{\infty} k p_k s^{k-1}|_{s=1} = EX. \tag{2.6}$$

Also note that

$$E[X^2] = \sum_{k=0}^{\infty} k^2 p_k = \sum k(k-1)p_k + \sum k p_k = \mathcal{P}''(1) + \mathcal{P}'(1) \tag{2.7}$$

which gives us the variance of X a

$$\mathrm{Var}X = E[X^2] - (EX)^2 = \mathcal{P}''(1) + \mathcal{P}'(1) - (\mathcal{P}'(1))^2. \tag{2.8}$$

The individual probabilities of $X = k$ may also be found as

$$p_k = P(X = k) = \frac{1}{k!} \cdot \frac{d^k}{ds^k} \mathcal{P}(s)|_{s=0}. \tag{2.9}$$

Now suppose that X and Y are two independent variables, taking values in $\mathbb{Z}_+$. Let $Z = X + Y$. We ask the probability that Z equals k. We can find this as

$$P(Z = k) = \sum_{m=0}^{k} P(X = m, Y = k - m) = \sum_{m=0}^{k} P(X = m) \cdot P(Y = k - m). \tag{2.10}$$

Therefore, denoting $p_k^{(X)}$ to be the probability mass function of X, we have

$$\{p_k^{(Z)}\} = \{p_k^{(X)}\} * \{p_k^{(Y)}\} \implies \mathcal{P}^{(Z)}(s) = \mathcal{P}^{(X)}(s) \cdot \mathcal{P}^{(Y)}(s). \tag{2.11}$$

There is an easier way to see the last equation; we could have started with $Es^Z = E[s^X \cdot s^Y] = E[s^X]E[s^Y]$.

If we have $S_n = X_1 + X_2 + \ldots + X_n$, where the X_i's are independently distributed taking values in $\mathbb{Z}_+$, it can be shown that

$$\{p_k^{(S_n)}\} = \{p_k^{(X)}\}^{n*} \tag{2.12}$$

Example 2.3. Let us compute the generating function of $X \sim \mathrm{Bin}(n, p)$. We have

$$\mathcal{P}(s) = \sum_{k=0}^{\infty} P(X = k)s^k = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} s^k = \left((1-p) + ps\right)^n. \tag{2.13}$$

This is the generating function of the binomial distribution. Clearly,

$$\begin{aligned} EX &= \mathcal{P}'(1) = np, \\ \mathrm{Var}X &= \mathcal{P}''(1) + \mathcal{P}'(1) - (\mathcal{P}'(1))^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p). \end{aligned}$$

Note that using this generating function, we can also show that $\mathrm{Bin}(n, p) + \mathrm{Bin}(m, p) = \mathrm{Bin}(m + n, p)$ when the former terms are independent.

Example 2.4. We look at $X \sim \mathrm{Poisson}(\lambda)$. We have

$$\mathcal{P}(s) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} = e^{-\lambda + \lambda s}. \tag{2.14}$$

For this, we can als verify $EX = \mathrm{Var}X = \lambda$. We can also show that $\mathrm{Poisson}(\lambda) + \mathrm{Poisson}(\mu) = \mathrm{Poisson}(\lambda + \mu)$ when the former terms are independent.

Example 2.5. We look at $X \sim \mathrm{Geo}(p)$. Denote $1 - p$ as q. The generating function is given as

$$\mathcal{P}(s) = \sum_{k=0}^{\infty} p q^k s^k = \frac{p}{1 - qs}. \tag{2.15}$$

As an extension, let X_k denote the number of failures between the $(k-1)^{\text{th}}$ and k^{th} successes. If we denote $S_r = X_1 + X_2 + \ldots + X_r$, we find that $S_r \sim \text{NB}(p, r)$. From direct computation, we know that

$$P(S_r = k) = \binom{r+k-1}{k} q^k p^r \text{ for } k = 0, 1, \ldots$$

Let us compute this in another way; S_r is the sum of independent geomtric random variables with parameter p. We have

$$\mathcal{P}^{(S_r)}(s) = \left(\frac{p}{1-qs}\right)^r = p^r(1-qs)^{-r} = p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-qs)^k \tag{2.16}$$

which tells us that

$$P(S_r = k) = p^r \binom{-r}{k} (-q)^k. \tag{2.17}$$

2.1 Random Walks, with Generating Functions

Here, we consider the paths that have a right step with probability p and a left step with probability $q = 1-p$. We first look at the waiting time for the first gain, that is, the event $\{S_1 \leq 0, S_2 \leq 0, \ldots, S_{n-1} \leq 0, S_n = 1\}$ (Event $(*)$). Denote the probability of this event by ϕ_n, and its generating function by $\Phi(s)$. Note that $\phi_0 = 0$ and $\phi_1 = p$ lead to trivial cases. We focus on $n > 1$.

We must have $S_1 = -1$ (Event (1)). Denote, by $\nu < n$, the first return to 0 (Event (2)). ν only depends on $X_0, X_1, \ldots, X_\nu$. We need another $n - \nu$ steps to reach 1; this depends on $X_{\nu+1}, X_{\nu+2}, \ldots, X_n$ (Event (3)). For some $n > 1$, Event $(*)$ occurs if and only Event (1) $\cap$ Event (2) $\cap$ Event (3) occurs for some $\nu < n$. The point here is that the three events are independent. For some fixed $\nu < n$,

$$P(\text{Event (1)}) = q, \; P(\text{Event (2)}) = \phi_{\nu-1}, \; P(\text{Event (3)}) = \phi_{n-\nu}. \tag{2.18}$$

Thus,

$$\phi_n = \sum_{\nu=2}^{n-1} q\phi_{\nu-1}\phi_{n-\nu}. \tag{2.19}$$

We have

$$\Phi(s) - ps = \sum_{n=2}^{\infty} \phi_n s^n = q \sum_{n=2}^{\infty} (\phi_1\phi_{n-2} + \ldots + \phi_{n-2}\phi_1) s^n = qs \sum_{n=1}^{\infty} \phi_n^{2*} s^n = qs(\Phi(s))^2 \tag{2.20}$$

$$\implies \Phi(s) - ps = qs(\Phi(s))^2. \tag{2.21}$$

This is a standard quadratic; solving gives us

$$\Phi(s) = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2qs}. \tag{2.22}$$

The solution with the '+' is rejected; if it was valid, then plugging in $s < 1$ would give us $\Phi(s) > 1$, which is impossible. We expand this using the binomial theorem,

$$\Phi(s) = \frac{1}{2qs}\left(1 - \sum_{k=0}^{\infty} \binom{\frac{1}{2}}{k} (-4pqs^2)^k\right) = \sum_{k=1}^{\infty} \binom{\frac{1}{2}}{k} \frac{(-1)^{k-1}(4pq)^k}{2q} s^{2k-1} \tag{2.23}$$

which tells us that

$$\phi_{2k-1} = \frac{(-1)^{k-1}}{2q} \binom{\frac{1}{2}}{k} (4pq)^k, \; \phi_{2k} = 0. \tag{2.24}$$

Thus,

$$\Phi(1) = \sum \phi_n = \frac{1 - \sqrt{1-4pq}}{2q} = \frac{1 - |p-q|}{2q} = \begin{cases} \frac{p}{q} & \text{if } p < q, \\ 1 & \text{if } p \geq q. \end{cases}$$

This gives the probability that, at some point of the random walk, the displacement 1 is reached.

Similarly, for displacement S_n, we have

$$P(S_n \leq 0 \; \forall n) = \begin{cases} \frac{q-p}{p} & \text{if } p < q, \\ 0 & \text{if } p \geq q. \end{cases}$$

Appendices

Chapter A

Appendix

Extra content goes here.

Appendix

Index