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.

Contents

Chapter 1

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 $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$. Then, $EX \leq EY$.*

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

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

■

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*

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

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

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

■

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.

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

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

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 $a > 0$,*

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

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 $\text{Var}X_i = \sigma^2$. If $S_n = \sum X_i$, we then have

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

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

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

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,

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

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

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

For the variance, we have

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

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

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

■

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 $X_m + \ldots + X_{m-i}$ 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

$$
\begin{aligned}
\frac{1}{2}P(S_0 \geq 0, \dots, S_{n-m} \geq 0) &= P(\tilde{X} = 1) \cdot P(S_0 \geq 0, \dots, S_{n-m} \geq 0) && (1.23) \\
&= P(\tilde{X} = 1, S_0 \geq 0, S_0 \geq 0, \dots, S_{n-m} \geq 0) && (1.24) \\
&= P(\tilde{S_1} = 1, \tilde{S_2} > 0, \dots, \tilde{S}_{n-m+1} > 0) && (1.25) \\
&= P(S_1 > 0, S_2 > 0, \dots, S_{n-m+1} > 0). && (1.26)
\end{aligned}
$$

Thus, we get

$$P(M_n = m) = \frac{1}{2} u_{2k} u_{2\nu - 2k} \tag{1.27}$$

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

Appendices

Chapter A

Appendix

Extra content goes here.

Appendix

Index