

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 .

$\text{Var}X$, 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

1	RANDOM WALKS AND MISC. RESULTS	1
1.1	The Law of Large Numbers	2
1.2	Simple Random Walk	3
1.3	Erdős-Renyi Random Graph	7
2	GENERATING FUNCTIONS	11
2.1	Random Walks, with Generating Functions	13
2.2	Simple Random Walks in Higher Dimensions	15
2.3		16
2.4	Gambler's Ruin	17
2.4.1	Duration of the Game	18
3	JOINT CONTINUOUS DISTRIBUTIONS	19
3.1	Introduction	19
3.2	Some Distributions	21
3.2.1	Gamma Random Variable	21
3.3	Conditional Distribution	23
3.3.1	The t -distribution	23
3.3.2	The Bivariate Normal Distribution	24
3.4	Order Statistics	24
3.5	Joint Distribution of Functions of Random Variables	27
3.5.1	Conditional Expectation and Variance	27
4	CONVERGENCE OF RANDOM VARIABLES	31
4.1	Types of Convergence	31
4.1.1	Almost Sure Convergence	31
4.1.2	Convergence in Probability	31
4.1.3	Convergence in Distribution	34
	Appendices	37
A	Appendix	39
	Index	41

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 probability $P(\omega)$ 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,

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

■

We now state Markov's inequality.

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

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

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

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

■

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.

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

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 $\sigma^2 = \text{Var}(X)$. Then for $a > 0$,*

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

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

Example 1.5. Suppose X_1, X_2, \dots, 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

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

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

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

Proposition 1.6. If $\text{Var}(X) = 0$, then $P(X = EX) = 1$.

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

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

Define A_n as $\{|X - EX| > \frac{1}{n}\}$. Taking $P(\bigcup A_n) = \lim_{n \rightarrow \infty} P(A_n)$, 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 $a > 0$,

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

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 $S_n = X_1 + X_2 + \dots + X_n$. Then,

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

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, $U_1, V_1, \dots, U_n, V_n$, 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

$$\{|X_1 + \dots + X_n| \geq na\} \subseteq \{|U_1 + \dots + U_n| \geq \frac{na}{2}\} \cup \{|V_1 + \dots + V_n| \geq \frac{na}{2}\} \quad (1.11)$$

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

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

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

For the variance, we have

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

which transforms the previous equation as

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

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

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

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

$$\begin{aligned} P(V_1 + \dots + V_n \neq 0) &\leq P(\{V_1 \neq 0\} \cup \dots \cup \{V_n \neq 0\}) \leq nP(V_i \neq 0) = n \sum_{|x| > \delta n} P(X_i = x) \\ &\leq n \sum_{|x| > \delta n} \frac{|x|}{\delta n} P(X_i = x) = \frac{1}{\delta} E[|V_i|]. \end{aligned} \quad (1.17)$$

The rightmost term here tends to 0 as $n \rightarrow \infty$. 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, \dots, y_k , and $E[|X|] < \infty$, then $\lim_{n \rightarrow \infty} E[|X| \cdot 1_{|X| \leq n}] = E[|X|]$.*

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, \dots 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 $S_0 = k_0 = 0, S_1 = k_1, \dots, S_n = k_n$ 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 $\{X_1 = l_1, X_2 = l_2, \dots, X_n = l_n\}$, where $l_i = \pm 1$, is equally likely with probability $\frac{1}{2^n}$. Thus,

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

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

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

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 $\binom{n+k}{2}$.*

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 $\binom{n+k}{2}$. ■

Using this lemma, we find that $N_n(k) = \binom{n+k}{2}$. 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 $(b = k_0, k_1, \dots, k_{n-1}, k_n = b')$, 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 ; $(-b = -k_0, -k_1, \dots, -k_{j-1}, k_j = 0, k_{j+1}, \dots, k_n = b')$. 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 $P(S_{2\nu} = 0) = \frac{\binom{2\nu}{\nu}}{2^{2\nu}}$. 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

$$P(\text{first return to } 0 \dots) = P(S_1 \neq 0, S_2 \neq 0, \dots, S_{2\nu} \neq 0) \tag{1.21}$$

$$= P(S_1 > 0, \dots, S_{2\nu} > 0) + P(S_1 < 0, \dots, S_{2\nu} < 0)$$

$$= 2P(S_1 > 0, \dots, S_{2\nu} > 0)$$

$$= 2 \sum_{k \text{ even}, k > 0} P(S_1 > 0, \dots, S_{2\nu-1} > 0, S_{2\nu} = k)$$

$$= \frac{2}{2^{2\nu}} \sum_{k \text{ even}, k > 0} N_{2\nu}^+(k)$$

$$= \frac{2}{2^{2\nu}} \sum_{k \text{ even}, k > 0} N_{2\nu-1}(k-1) - N_{2\nu-1}(k+1)$$

$$= \frac{2}{2^{2\nu}} N_{2\nu-1}(1) = u_{2\nu}. \tag{1.22}$$

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, \dots, S_n , over n steps, achieves its maximum for the first time.

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

The key idea here is to consider the *reversed walk*; define a new walk with $X'_1 = X_m, X'_2 = X_{m-1}, \dots, X'_m = X_1$. Also define $S'_k = X'_1 + \dots + X'_k$. From here, we can deduce that $S_m > S_{m-i}$ is true if and only if $X_m + \dots + X_{m-i} > 0$ is true, which is true if and only if $S'_i > 0$ is true. So, $P(S_m > S_0, S_m > S_1, \dots, S_m > S_{m-1}) = P(S'_1 > 0, S'_2 > 0, \dots, S'_m > 0)$. If we now define $S''_k = X_{m+1} + \dots + X_{m+k}$, we have

$$\begin{aligned} P(S_m \geq S_{m+1}, S_m \geq S_{m+2}, \dots, S_m \geq S_n) &= P(X_{m+1} \leq 0, X_{m+1} + X_{m+2} \leq 0, \dots, X_{m+1} + \dots + X_n \leq 0) \\ &= P(S''_1 \leq 0, S''_2 \leq 0, \dots, S''_{n-m} \leq 0) \\ &= P(S''_1 \geq 0, S''_2 \geq 0, \dots, S''_{n-m} \geq 0) \end{aligned}$$

The first of the terms discussed, $P(S'_1 > 0, S'_2 > 0, \dots, S'_m > 0)$, 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 $\tilde{X} + X_1 + \dots + X_{i-1}$ and \tilde{S}_0 to be 0.

We then have

$$\begin{aligned}
\frac{1}{2}P(S_0 \geq 0, \dots, S_{n-m} \geq 0) &= P(\tilde{X} = 1) \cdot P(S_0 \geq 0, \dots, S_{n-m} \geq 0) \\
&= P(\tilde{X} = 1, S_0 \geq 0, S_0 \geq 0, \dots, S_{n-m} \geq 0) \\
&= P(\tilde{S}_1 = 1, \tilde{S}_2 > 0, \dots, S_{n-m+1} > 0) \\
&= P(S_1 > 0, S_2 > 0, \dots, S_{n-m+1} > 0).
\end{aligned} \tag{1.23}$$

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

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

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

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

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

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

If we look at this as a cumulative density function, 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

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

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

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

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

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

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

$$\lim_{n \rightarrow \infty} P(K_n = 0) = 0. \quad (1.31)$$

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 $2n$.

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

Lemma 1.13. *With the usual notation,*

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

Proof. We have

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

■

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 $0 < t < 1$,*

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

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

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

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

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

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

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,

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

■

January 17th.

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

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

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

Proof. Start with X_1, X_2, \dots which are independent and identically distributed Bernoulli random variables, $\text{Ber}(x)$. Let $S_n = X_1 + X_2 + \dots + X_n$. 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 $f(\frac{S_n}{n})$. We now have

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

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

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

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

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

and

$$\sum_{j \notin A_\delta} \left| f\left(\frac{j}{n}\right) - f(x) \right| P(S_n = j) \leq 2MP \left(\left| \frac{S_n}{n} - x \right| > \delta \right) \leq 2M \frac{\text{Var}(S_n)}{n^2 \delta^2} = \frac{2Mnx(1-x)}{n^2 \delta^2}. \quad (1.44)$$

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

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

■

1.3 Erdős-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

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

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

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

Also, $P(N \geq 1) \leq EN < n^3 p^3$. If $p = p_n < \frac{1}{n}$, then $P(N \geq 1) \rightarrow 0$ as $n \rightarrow \infty$. We discuss this for $p > \frac{1}{n}$. We have

$$P(N = 0) \leq P(|N - EN| \geq EN) \leq \frac{\text{Var}N}{(EN)^2} \leq \frac{(n^3 p^3 + n^4 p^5)}{\frac{n^6 p^6}{100}} \leq \frac{100}{n^3 p^3} + \frac{100}{n(np)} \rightarrow 0. \quad (1.46)$$

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*

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

Now suppose that $\frac{np_n}{\frac{1}{n}} C > 0$ as $n \rightarrow \infty$. Then we have

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

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, \dots are independent, but not necessarily identically distributed. Let $S_n = X_1 + \dots + X_n$. We say a game with accumulated entrance fees $\{\alpha_n, n \geq 1\}$ is fair if

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

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,

$$EX = \sum_{k=1}^{\infty} \frac{1}{2^k} 2^k = \infty. \quad (1.50)$$

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

$$P(|S_n - \alpha_n| > \varepsilon \alpha_n) \rightarrow 0. \quad (1.51)$$

To find this, we will define

$$\begin{aligned} U_j &= X_j 1_{\{X_j \leq a_n\}}, \\ V_j &= X_j 1_{\{X_j > a_n\}}. \end{aligned}$$

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

$$P(|S_n - \alpha_n| > \varepsilon \alpha_n) \leq P(|U_1 + \dots + U_n - \alpha_n| > \frac{1}{2} \varepsilon \alpha_n) + P(|V_1 + \dots + V_n| > \frac{1}{2} \varepsilon \alpha_n). \quad (1.52)$$

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

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

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

Thus, we will require that $a_n \gg n$. Also,

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

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

$$\alpha_n = nEU_j = n \lfloor \log_2 a_n \rfloor. \quad (1.56)$$

This gives us

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

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

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

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

$$\alpha_n = n \log_2 a_n = n \log_2 n + n \log_2 \log_2 n. \quad (1.58)$$

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

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

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 $k = 0, 1, \dots$, the function we look at is

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

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

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

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

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

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 $\mathbb{Z}_+ = \{0, 1, \dots\}$. 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,

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

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

Also note that

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

which gives us the variance of X as

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

which tells us that

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

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 $\{S_1 \leq 0, S_2 \leq 0, \dots, S_{n-1} \leq 0, S_n = 1\}$ (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, \dots, X_ν . We need another $n-\nu$ steps to reach 1; this depends on $X_{\nu+1}, X_{\nu+2}, \dots, X_n$ (Event (3)). For some $n > 1$, Event (*) occurs if and only if 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 $\nu < n$,

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

Thus,

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

We have

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

$$\implies \Phi(s) - ps = qs(\Phi(s))^2. \quad (2.21)$$

This is a standard quadratic; solving gives us

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

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,

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

which tells us that

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

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

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

January 28th.

Recall that we used u_k denote the probability that the random walk returns to zero at step k . For unequal left-right step probabilities,

$$u_k = P(S_k = 0) = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ \binom{2k}{k} p^n q^n & \text{if } k = 2n. \end{cases}$$

Thus, the generating function for this is

$$U(s) = \sum_{n=0}^{\infty} u_{2n} s^{2n} = \sum_{n=0}^{\infty} \binom{2n}{n} (pq s^2)^n = \sum_{n=0}^{\infty} \binom{-\frac{1}{2}}{n} (-4pq s^2)^n = \frac{1}{\sqrt{1-4pq s^2}}. \quad (2.25)$$

Denote, by f_{2n} , the probability that the first return to zero occurs at step $2n$, for some $n \geq 1$. In fact, it consists of subevents; if $X_1 = 1$, denote it by f_{2n}^+ and if $X_1 = -1$, denote it by f_{2n}^- . If we also recall the definition of our ϕ_n ,

$$f_{2n}^- = P(X_1 = -1, S_2 < 0, S_3 < 0, \dots, S_{2n-1} < 0, S_{2n} = 0) = q\phi_{2n-1}. \quad (2.26)$$

The generating function of $\{f_{2n}^-\}$ will be given as

$$F^-(s) = \sum_{n=1}^{\infty} f_{2n}^- s^{2n} = q \sum_{n=1}^{\infty} \phi_{2n-1} s^{2n} = qs \sum_{n=1}^{\infty} \phi_{2n-1} s^{2n-1} = qs\Phi(s) = \frac{1}{2}(1 - \sqrt{1-4pq s^2}). \quad (2.27)$$

It can be shown that f_{2n}^+ is just f_{2n}^- with the probabilities reversed (check!). The generating function of $\{f_{2n}^+\}$ is given as

$$F^+(s) = \sum_{n=0}^{\infty} f_{2n}^+ s^{2n} = \frac{1}{2}(1 - \sqrt{1-4pq s^2}). \quad (2.28)$$

Adding both of these, we get

$$F(s) = F^+(s) + F^-(s) = 1 - \sqrt{1-4pq s^2} = 1 - \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} (-4pq s^2)^n \quad (2.29)$$

$$\implies f_{2n} = (-1)^{n+1} \binom{\frac{1}{2}}{n} (4pq)^n. \quad (2.30)$$

$F(1)$ gives us the probability that walk eventually returns to zero,

$$F(1) = \sum_{n=0}^{\infty} f_{2n} = 1 - \sqrt{1-4pq} = 1 - |p - q|. \quad (2.31)$$

$F'(1)$ gives us the expected time of return to zero,

$$F'(s) = -\frac{1}{2}(1-4pq s^2)^{-\frac{1}{2}}(-8pq s). \quad (2.32)$$

If $p = q = \frac{1}{2}$, then

$$F'(1) = \lim_{s \rightarrow 1^-} F'(s) = \infty.$$

The basic lemma can be proved using the generating functions.

2.2 Simple Random Walks in Higher Dimensions

Consider the walk in the dimension d . A walker starts at the origin in the lattice \mathbb{Z}^d . The random variables X_1, X_2, \dots are independent and identically distributed with probabilities

$$P(X_i = -e_d) + \dots + P(X_i = -e_2) + P(X_i = -e_1) + P(X_i = e_1) = P(X_i = e_2) + \dots + P(X_i = e_d) = \frac{1}{2d}.$$

for all valid i . The random walk here is defined as $S_n = X_1 + \dots + X_n$. We ask the probability that S_n returns to the origin. Denote by u_{2n} the probability that $S_{2n} = 0$, and denote by f_{2n} the probability that the first return to the origin occurs at time $2n$. By conditioning,

$$u_{2n} = \sum_{k=0}^n f_{2k} u_{2n-2k}. \quad (2.33)$$

If $U(s)$ and $F(s)$ are the appropriate generating functions, then we can show that

$$U(s) - 1 = F(s)U(s) \implies U(s) = \frac{1}{1 - F(s)}. \quad (2.34)$$

Both $U(s)$ and $F(s)$ are convergent for $|s| < 1$. For each N ,

$$\sum_{n=0}^N u_{2n} \leq \lim_{s \rightarrow 1^-} U(s) \leq \sum_{n=0}^{\infty} u_{2n}. \quad (2.35)$$

Lemma 2.6. *A random walk on \mathbb{Z}^d return to the origin with probability 1 if and only if $\sum u_{2n} = \infty$.*

Proof. Suppose $F(1) < 1$. Then, $\lim_{s \rightarrow 1^-} U(s) < \infty$ and, consequently, $\sum_{n=0}^{\infty} u_{2n} < \infty$. The converse can be proved by reversing the steps. ■

The lemma tells us that to see the probability that the random walk returns to the origin, we only need to compute $\sum_{n=0}^{\infty} u_{2n}$.

For $d = 2$, we need the number of e_i jumps to be equal to the number of $-e_i$ jumps for $i = 1, 2$. We have

$$\begin{aligned} u_{2n} &= \frac{1}{4^{2n}} \sum_{j=0}^n \binom{2n}{j} \binom{2n-j}{j} \binom{2n-2j}{n-j} \binom{n-j}{n-j} = \frac{1}{4^{2n}} \binom{2n}{n} \sum_{j=0}^n \binom{n}{j}^2 = \frac{1}{4^{2n}} \binom{2n}{n}^2 \\ &\sim \frac{2}{2\pi} \frac{n^{4n+1}}{n^{4n+2}} = \frac{1}{\pi n}. \end{aligned} \quad (2.36)$$

Since this is any asymptotic relationship, $u_{2n} \geq \frac{(1-\varepsilon)}{\pi n}$ for large n . Thus, we can show $\sum u_{2n} = \infty$.

For $d = 3$,

$$\begin{aligned} u_{2n} &= \frac{1}{6^{2n}} \sum_{j,k=0; j+k \leq n}^n \frac{(2n)!}{j!k!(n-j-k)!(n-j-k)!} = \frac{1}{6^{2n}} \sum_{j,k=0; j+k \leq n}^{\infty} \frac{(2n)!}{(j!)^2(k!)^2((n-j-k)!)^2} \\ &= \frac{1}{2^{2n}} \binom{2n}{n} \sum_{j,k; j+k \leq n} \left(\frac{n!}{j!k!(n-j-k)!} \frac{1}{3^n} \right)^2. \end{aligned} \quad (2.37)$$

$\frac{1}{2^{2n}} \binom{2n}{n}$ behaves asymptotically as $\frac{1}{\sqrt{\pi n}}$. For the rest of the term,

$$\sum_{j,k; j+k \leq n} \left(\frac{n!}{j!k!(n-j-k)!} \frac{1}{3^n} \right)^2 \leq t_n \sum_{j,k; j+k \leq n} \frac{n!}{j!k!(n-j-k)!} \frac{1}{3^n} \quad (2.38)$$

where $t_n = \max_{j,k; j+k \leq n} \frac{n!}{j!k!(n-j-k)!}$. The maximum is attained roughly when $j, k \approx \frac{n}{3}$. Also, the summation behaving as the upper bound is just unity. Thus,

$$\sum_{j,k; j+k \leq n} \left(\frac{n!}{j!k!(n-j-k)!} \frac{1}{3^n} \right)^2 \leq t_n \approx \frac{n!}{((\frac{n}{3})!)^3 3^n} \sim \frac{C}{n} \quad (2.39)$$

for some constant C . Therefore,

$$u_{2n} \leq \frac{C^*}{n^{\frac{3}{2}}} \implies \sum u_{2n} < \infty \implies F(1) < 1. \quad (2.40)$$

Theorem 2.7 (*Polya*). *A random walk in 1 or 2 dimensions will always return to the origin with probability 1. A random walk in more than 2 dimensions has a positive probability of never returning to the origin.*

2.3

January 31st.

Recall that in the first course, we studied that if $X_n \sim \text{Bin}(n, p_n)$ with $np_n \rightarrow \lambda$ as $n \rightarrow \infty$, then

$$\lim_{n \rightarrow \infty} P(X_n = k) = P(\text{Poisson}(\lambda) = k) \text{ for } k \geq 0. \quad (2.41)$$

We now extend upon this idea.

Theorem 2.8 (*Continuity theorem*). *Suppor for each n the sequence $a_{0,n}, a_{1,n}, \dots$ is a probability distribution, that is,*

$$a_{k,n} \geq 0 \text{ for all } k \text{ and } \sum_{k=0}^{\infty} a_{k,n} = 1. \quad (2.42)$$

Let $A^{(n)}(s)$ denote the generating function for $\{a_{k,n}\}_{k \geq 0}$, that is,

$$A^{(n)}(s) = \sum_{k=0}^{\infty} a_{k,n} s^k \text{ for all } n. \quad (2.43)$$

Then $a_k = \lim_{n \rightarrow \infty} a_{k,n}$ exists for all k (statement \star) if and only if $A(s) = \lim_{n \rightarrow \infty} A^{(n)}(s)$ exists for all $0 < s < 1$ (statement $\star\star$). In this case, $A(s) = \sum_{k=0}^{\infty} a_k s^k$.

Proof. Assume statement \star . Thus, $|a_{k,n} - a_k| \leq 1$ for all n large enough. If we now fix $0 < s < 1$, then for some K and a fixed $\varepsilon > 0$, we have

$$\begin{aligned} |A^{(n)}(s) - A(s)| &= \left| \sum_{k=0}^{\infty} a_{k,n} s^k - \sum_{k=0}^{\infty} a_k s^k \right| \\ &= \left| \sum_{k=0}^K a_{k,n} s^k + \sum_{k=K+1}^{\infty} a_{k,n} s^k - \sum_{k=0}^K a_k s^k - \sum_{k=K+1}^{\infty} a_k s^k \right| \\ &\leq \left| \sum_{k=0}^K a_{k,n} s^k - \sum_{k=0}^K a_k s^k \right| + \left| \sum_{k=K+1}^{\infty} (a_{k,n} - a_k) s^k \right| \\ &\leq \left| \sum_{k=0}^K a_{k,n} s^k - \sum_{k=0}^K a_k s^k \right| + \frac{s^{K+1}}{1-s}. \end{aligned} \quad (2.44)$$

We can choose K such that the second term becomes less than ε , and we can choose N such that for all $n \geq N$, the first term becomes smaller than ε . Therefore, the entire term becomes less than 2ε .

For the converse, assume statement $\star\star$. $A(s)$ is monotonic in s ; $A(0) = \lim_{s \rightarrow 0^-} A(s)$. We sandwich as follows—

$$\begin{aligned} a_{0,n} &\leq A^{(n)}(s) \leq a_{0,n} + \frac{s}{1-s} \\ \implies A^{(n)}(s) - \frac{s}{1-s} &\leq a_{0,n} \leq A^{(n)}(s). \end{aligned}$$

Letting n grow to infinity,

$$A(s) - \frac{s}{1-s} \leq \liminf_{n \rightarrow \infty} a_{0,n} \leq \limsup_{n \rightarrow \infty} a_{0,n} \leq A(s). \quad (2.46)$$

If $s \rightarrow 0$, note that $\lim_{n \rightarrow \infty} a_{0,n} = A(0)$. Now define

$$B^{(n)}(s) = \frac{A^{(n)}(s) - a_{0,n}}{s} \rightarrow \frac{A(s) - A(0)}{s} \rightarrow A'(0). \quad (2.47)$$

Working similarly,

$$a_{1,n} \leq B^{(n)}(s) \leq a_{1,n} + \frac{s}{1-s} \quad (2.48)$$

$$\implies B^{(n)}(s) - \frac{s}{1-s} \leq a_{1,n} \leq B^{(n)}(s). \quad (2.49)$$

If we again proceed as shown, we will get $B(0) = \lim_{n \rightarrow \infty} a_{1,n}$ and $a_{1,n} \rightarrow A'(0) = a_1$. Thus, induction is in play here. \blacksquare

Example 2.9. Let us work with the binomial distribution example given before. We have $X_n \sim \text{Bin}(n, p_n)$ with $np_n \rightarrow \lambda$. We have

$$\begin{aligned} A^{(n)}(s) &= \sum_{k=0}^{\infty} P(X_n = k) s^k = ((1 - p_n) + p_n s)^n = (1 + p_n(s - 1))^n \\ \implies \lim_{n \rightarrow \infty} A^{(n)}(s) &= \lim_{n \rightarrow \infty} \left(1 + \frac{np_n}{n}(s - 1)\right)^n = e^{\lambda(s-1)} = E[s^{\text{Poisson}(\lambda)}]. \end{aligned} \quad (2.50)$$

Thus, we have shown the prior statement.

Example 2.10. We have $X_1^{(n)}, X_2^{(n)}, \dots, X_n^{(n)}$ independent, with $X_i^{(n)} \sim \text{Ber}(p_i^{(n)})$ for $1 \leq i \leq n$. Let $S_n = X_1^{(n)} + X_2^{(n)} + \dots + X_n^{(n)}$. We have

$$E[s^{S_n}] = \prod_{i=1}^n E[s^{X_i^{(n)}}] = \prod_{i=1}^n \left((1 - p_i^{(n)}) + p_i^{(n)} s\right) = \exp\left(\ln \prod_{i=1}^n (\dots)\right). \quad (2.51)$$

Assume that $\lim_{n \rightarrow \infty} \sum_{i=1}^n p_i^{(n)} = \lambda$ and $\lim_{n \rightarrow \infty} \max_i p_i^{(n)} = 0$. Thus,

$$\begin{aligned} \exp\left(\sum_{i=1}^n \ln(1 + p_i^{(n)}(s - 1))\right) &= \exp\left(\sum_{i=1}^n p_i^{(n)}(s - 1) - \frac{(p_i^{(n)}(s - 1))^2}{2} + \dots\right) \\ &= \exp\left((s - 1) \sum_{i=1}^n p_i^{(n)} - \sum_{i=1}^n o(p_i^{(n)}(s - 1))\right) \end{aligned} \quad (2.52)$$

$$\rightarrow e^{\lambda(s-1)}. \quad (2.53)$$

Example 2.11. Let $X^{(n)} \sim \text{NB}(r_n, p)_n$, the number of successes before the r_n^{th} success in trials with success probability p_n . Let $p_n \rightarrow 1$ and $r_n \rightarrow \infty$ such that $r_n(1 - p_n) \rightarrow \lambda$, where λ is fixed. We would then have $P(X^{(n)} = k) \rightarrow P(\text{Poisson}(\lambda) = k)$.

2.4 Gambler's Ruin

We take a look at a gambler, who has starting capital z . His probability of a success (+1) is p , and of a failure (−1) is q . We ask the probability q_z that the gambler reaches 0 before a when he starts at capital z . Note that q_z satisfies

$$q_z = pq_{z+1} + qq_{z-1} \text{ for } 1 < z < a - 1 \text{ (statement } \star), \text{ with } q_0 = 1, q_a = 0 \text{ (statement } \star\star). \quad (2.54)$$

We look at two cases, beginning with the case when $p \neq q$. Note that $q_z = 1$ for $1 \leq z \leq a - 1$ solves for statement \star , ignoring statement $\star\star$ and ignoring probability for now. $q_z = (\frac{q}{p})^z$ for $1 \leq z \leq a - 1$ also solves for statement \star . Therefore, $A + B(\frac{q}{p})^z$ solves statement \star . Now, we plug in the boundary conditions given by statement $\star\star$. Solving the equations $A + B = 1$ and $A + B(\frac{q}{p})^a = 0$ gives us

$$B = \frac{1}{1 - (\frac{q}{p})^a}, \quad A = 1 - \frac{1}{1 - (\frac{q}{p})^a}. \quad (2.55)$$

Plugging this in, gives us

$$q_z = \frac{(\frac{q}{p})^a - (\frac{q}{p})^z}{(\frac{q}{p})^a - 1}. \quad (2.56)$$

Note that we were working the case when $p \neq q$. For $p = q$, this solution does not work.

We work the case for when $p = q = \frac{1}{2}$. Again, $q_z = 1$ for $1 \leq z \leq a - 1$ satisfies statement \star . We also find that $q_z = z$ for $1 \leq z \leq a - 1$ also satisfies this statement. Hence, we look for $A + Bz$ which satisfies boundary condition given by statement $\star\star$. Solving, this gives us

$$q_z = 1 - \frac{z}{a}. \quad (2.57)$$

Note that we are yet to show $p_z + q_z = 1$. If we instead focus on a *second* gambler playing against our gambler, we would have a gambler with capital $a - z$, and probability of success q and probability of failure p . Replacing z by $a - z$ and q by p and p by q in our formed equations would give us $p_z + q_z = 1$.

Let us intuitively look at our equations with a table of examples.

p	q	z	a	q_z
0.45	0.55	9	10	0.21
0.45	0.55	90	100	0.866
0.45	0.55	99	100	0.182
0.5	0.5	9	10	0.1
0.5	0.5	90	100	0.1
0.5	0.5	99	100	0.01

Table 2.1: Probability of ruin (q_z) given initial parameters.

Note that the expected net gain is given by

$$(a - z)(1 - q_z) - zq_z = a(1 - q_z) - z. \quad (2.58)$$

If we plug in this into our first three rows of our table, we would have -1.1 , -77 , -18 . If one is gambling under such condition, we must start with big capital z and low target $a - z$.

2.4.1 Duration of the Game

We look at D_z , the expected duration of a game starting at z ; the expected time before the gambler hits a or 0. The linear recurrence satisfied here is

$$D_z = pD_{z+1} + qD_{z-1} + 1 \text{ with boundary conditions } D_0 = 0, D_a = 0. \quad (2.59)$$

For $p \neq q$,

$$D_z = \frac{z}{q - p} - \frac{a}{q - p} \left(\frac{1 - (\frac{q}{p})^z}{1 - (\frac{q}{p})^a} \right). \quad (2.60)$$

For $p = q = \frac{1}{2}$,

$$D_z = z(a - z). \quad (2.61)$$

Chapter 3

JOINT CONTINUOUS DISTRIBUTIONS

3.1 Introduction

February 4th.

Recall that $X : \Omega \rightarrow \mathbb{R}$ is *continuous random variable* if it has a probability density function $f_X : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$. In this case, if $A \subseteq \mathbb{R}$, then

$$P(X \in A) = \int_A f_X(x)dx. \quad (3.1)$$

For a minute dx ,

$$P(X \in [x, x + dx]) \approx f_X(x)dx. \quad (3.2)$$

Two random variables X and Y are termed *jointly continuous* if there exists a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that for $A \subseteq \mathbb{R}^2$,

$$P((X, Y) \in A) = \iint_A f(x, y)dx dy. \quad (3.3)$$

In this case, f is termed the *joint probability density function* of X and Y . In particular, if $B, C \subseteq \mathbb{R}$, then

$$P(X \in C, Y \in B) = P((X, Y) \in C \times B) = \int_B \int_C f(x, y)dx dy. \quad (3.4)$$

The *joint cumulative density function* is given as

$$F(a, b) = P(X \leq a, Y \leq b) = P((X, Y) \in (-\infty, a] \times (-\infty, b]) = \int_{-\infty}^b \int_{-\infty}^a f(x, y)dx dy. \quad (3.5)$$

There is, again, joint analagous versions of the single random variables cases;

$$\frac{\partial^2}{\partial a \partial b} F(a, b) = f(a, b) \quad (3.6)$$

and

$$P(X \in [a, a + da], Y \in [b, b + db]) \approx f(a, b)dadb. \quad (3.7)$$

Note that (X, Y) being jointly continuous implies that both X and Y are continuous. Indeed, if $A \subseteq \mathbb{R}$, then

$$P(X \in A) = P(X \in A, Y \in \mathbb{R}) = \int_A \int_{\mathbb{R}} f(x, y)dy dx = \int_A f_X(x)dx. \quad (3.8)$$

In this case, the inner intergral becomes the probability density funciton of X .

Example 3.1. We are given the joint probability density function of X and Y as

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y} & \text{if } 0 < x < \infty, 0 < y < \infty, \\ 0 & \text{if otherwise.} \end{cases}$$

We are to compute $P(X > 1, Y < 1)$, $P(X < Y)$ and $P(X < a)$. This is left as an exercise.

Example 3.2. Suppose (X, Y) represents a random points inside a circle of radius R . The probability density function is given by

$$f(x, y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } x^2 + y^2 \leq R^2, \\ 0 & \text{if otherwise.} \end{cases}$$

Compute f_X, f_Y and f_D where $D = \sqrt{X^2 + Y^2}$.

We call X and Y independent if, for $A, B \subseteq \mathbb{R}$,

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B). \quad (3.9)$$

If we take $A = (-\infty, a], B = (-\infty, b]$, then

$$F(a, b) = F_X(a)F_Y(b) \implies f(a, b) = f_X(a)f_Y(b).$$

In fact, all three conditions are equivalent. Note that everything discussed so far may be extended to more than two random variables. If $A \subseteq \mathbb{R}^n$ and (X_1, X_2, \dots, X_n) are jointly continuous, then

$$P((X_1, X_2, \dots, X_n) \in A) = \int \cdots \int_A f(x_1, x_2, \dots, x_n) dx_1 \cdots dx_n \quad (3.10)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$.

Example 3.3 (*Buffin's needle problem*). Suppose we have a table of width D and length sufficient. We throw a needle of length $L \leq D$ randomly (and necessarily) on this table. We are to find the probability that the needle will land over either side of the table separated by width D . Denote by X the distance of the midpoint of the needle from the nearest edge, and denote by θ the angle it subtends with respect to the vertical.

Note that $X \in [0, \frac{D}{2}]$, and $\theta \in [0, \frac{\pi}{2}]$. By the procedure, one can assume these are independent. We look at the probability that the length of half the needle is greater than the length of the hypotenuse determined by X and θ —

$$P\left(\frac{L}{2} > \frac{X}{\cos \theta}\right) = \int_0^{\frac{\pi}{2}} \int_0^{\frac{L}{2} \cos \theta} \frac{2}{D} \cdot \frac{2}{\pi} dx d\theta = \frac{2L}{D\pi} \int_0^{\frac{\pi}{2}} \cos \theta d\theta = \frac{2L}{D\pi}. \quad (3.11)$$

Proposition 3.4. *The jointly continuous random variables X and Y are independent if and only if the probability density function can be factored into functions of x and y respectively.*

Suppose that X and Y are independent and jointly continuous random variables. Then $f(x, y) = f_X(x)f_Y(y)$. Let us compute the cumulative density function of $X + Y$.

$$P(X+Y \leq a) = \int_{-\infty}^{\infty} \int_{-\infty}^{a-x} f_X(x)f_Y(y) dy dx = \int_{-\infty}^{\infty} f_X(x)F_Y(a-x) dx = \int_{-\infty}^{\infty} f_Y(y)F_X(a-y) dy. \quad (3.12)$$

Differentiating to give the probability density function of $X + Y$,

$$\frac{d}{da} F_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(x)f_Y(a-x) dx = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y) dy. \quad (3.13)$$

This is the convolution of f_X and f_Y .

February 11th.

Suppose we have U_1, U_2, \dots, U_n which are independent and have the distribution $\text{Uniform}(0, 1)$. From $\{1, 2, \dots, n\}$, if we choose a subset of size k , then each subset of size k has probability $\frac{1}{\binom{n}{k}}$. We will generate I_1, I_2, \dots, I_n , where exactly k of them are 1. I_1 is defined as

$$I_1 = \begin{cases} 1 & \text{if } U_1 < \frac{k}{n}, \\ 0 & \text{if otherwise.} \end{cases} \quad (3.14)$$

Once $I_1, I_2 < \dots, I_i$ are determined, we can define

$$I_{i+1} = \begin{cases} 1 & \text{if } U_{i+1} < \frac{k - (I_1 + I_2 + \dots + I_i)}{n-1}, \\ 0 & \text{if otherwise.} \end{cases} \quad (3.15)$$

Notice that if $I_1 + I_2 + \dots + I_i = k$, then $I_{i+1}, I_{i+2}, \dots = 0$. Note that $P(I_1 = 1) = \frac{k}{n}$, and

$$P(I_{i+1} = 1 | I_1, I_2, \dots, I_i) = \frac{k - \sum_{r=1}^i I_r}{n-i}, \text{ for } 1 < i \leq n. \quad (3.16)$$

If we do induction on $k+n$, we know the case $k+n=2$ to be true ($k=1, n=1$). We assume for all the cases $k+n \leq l$, and suppose that $k+n=l+1$. Consider any subset of size k , say, $i_1 \leq i_2 \leq \dots \leq i_k$.

- Case I, where $i_1 = 1$. Here,

$$\begin{aligned} & P(I_{i_1} = I_{i_2} = \dots = I_{i_k} = 1, I_j = 0 \text{ otherwise}) \\ &= P(I_1 = 1) \cdot P(I_{i_2} = \dots = I_{i_k} = 1, I_j = 0 \text{ otherwise} | I_1 = 1) \\ &= \frac{k}{n} \frac{1}{\binom{n-1}{k-1}} = \frac{1}{\binom{n}{k}}. \end{aligned}$$

- Case II, where $i_1 > 1$. In this case,

$$\begin{aligned} & P(I_{i_1} = 0, I_{i_2} = \dots = I_{i_k} = 1, I_j = 0 \text{ otherwise}) \\ &= P(I_1 = 0) \cdot P(I_{i_2} = \dots = I_{i_k} = 1, I_j = 0 \text{ otherwise} | I_1 = 0) \\ &= \left(1 - \frac{k}{n}\right) \frac{1}{\binom{n-1}{k}} = \frac{1}{\binom{n}{k}}. \end{aligned}$$

3.2 Some Distributions

3.2.1 Gamma Random Variable

We have $X \sim \text{Gamma}(t, \lambda)$ where $\lambda > 0$ and $t > 0$. The probability density function is defined as

$$f_X(x) = \frac{\lambda e^{-\lambda y} (\lambda y)^{t-1}}{\Gamma(t)} \text{ for } 0 < y < \infty. \quad (3.17)$$

Here, Γ is the gamma function.

Proposition 3.5. *If $X \sim \text{Gamma}(s, \lambda)$ and $Y \sim \text{Gamma}(t, \lambda)$, then $X + Y \sim \text{Gamma}(s+t, \lambda)$.*

Proof. Via convolution, we know that

$$\begin{aligned} f_{X+Y}(a) &= \int_0^a f_X(a-y) f_Y(y) dy \\ &= \frac{1}{\Gamma(t)\Gamma(s)} \int_0^a \lambda e^{-\lambda(a-y)} (\lambda(a-y))^{t-1} \lambda e^{-\lambda y} (\lambda y)^{s-1} dy \\ &= \frac{\lambda e^{-\lambda a} (\lambda a)^{s+t-1}}{\Gamma(t+s)} \frac{\Gamma(t+s)}{\Gamma(t)\Gamma(s)} \int_0^a \frac{(a-y)^{t-1} y^{s-1}}{a^{s+t-1}} dy \\ &= \frac{\lambda e^{-\lambda a} (\lambda a)^{s+t-1}}{\Gamma(t+s)} \frac{\Gamma(t+s)}{\Gamma(t)\Gamma(s)} \int_0^1 (1-u)^{t-1} u^{s-1} du. \end{aligned}$$

If we integrate this probability distribution function, we must have

$$\int_0^\infty f_{X+Y}(a) da = 1 \quad (3.18)$$

$$\implies \frac{\Gamma(s+t)}{\Gamma(s)\Gamma(t)} \int_0^1 (1-u)^{t-1} u^{s-1} du = 1. \quad (3.19)$$

This gives rise to the Beta function, defined as $\int_0^1 (1-u)^{t-1} u^{s-1} du = B(t, s) = \frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}$. ■

One corollary that can be inferred from here is that if X_1, X_2, \dots, X_n are independent $\text{Gamma}(t_i, \lambda)$ distributions, then $\sum_{i=1}^n X_i \sim \text{Gamma}(\sum_{i=1}^n t_i, \lambda)$. We also notice that $X \sim \text{Exp}(\lambda) = \text{Gamma}(1, \lambda)$. If we take the X_i 's to be all the exponential distribution, then $S_n = X_1 + X_2 + \dots + X_n \sim \text{Gamma}(n, \lambda)$. The density of S_n is given by

$$g_n(y) = \frac{\lambda(\lambda y)^{n-1} e^{-\lambda y}}{\Gamma(n)} = \frac{\lambda(\lambda y)^{n-1} e^{-\lambda y}}{(n-1)!} \text{ for } 0 < y < \infty \quad (3.20)$$

with the cumulative distribution function

$$G_n(y) = \frac{1}{(n-1)!} \int_0^y \lambda(\lambda a)^{n-1} e^{-\lambda a} da. \quad (3.21)$$

Example 3.6. Suppose we have buses that each take time $X_i \sim \text{Exp}(\lambda)$. Suppose we fix a time t , and define $N(t)$ to be the number of buses seen up to time t . We ask the probability $P(N(t) = n)$.

$$P(N(t) = n) = P(X_1 + X_2 + \dots + X_n \leq t, X_1 + X_2 + \dots + X_{n+1} > t) \quad (3.22)$$

$$= P(X_1 + \dots + X_n \leq t) - P(X_1 + \dots + X_n \leq t, X_1 + \dots + X_{n+1} \leq t)$$

$$= P(X_1 + \dots + X_n \leq t) - P(X_1 + \dots + X_{n+1} \leq t)$$

$$= G_n(t) - G_{n+1}(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}. \quad (3.23)$$

Notice that there is a unique k for which $S_{k-1} < t \leq S_k$. Define X_k to be $S_k - S_{k-1}$.

Proposition 3.7. The X_k satisfying $S_{k-1} < t \leq S_k$ has density

$$v_t(x) = \begin{cases} \lambda^2 x e^{-\lambda x} & \text{if } 0 < x \leq t, \\ \lambda(1 + \lambda t) e^{-\lambda x} & \text{if } x > t. \end{cases} \quad (3.24)$$

Proof. Let k denote the index for which $S_{k-1} < t \leq S_k$. Define $L_t = S_k - S_{k-1}$. Let us first compute the cumulative distribution function.

- Case I, when $x < t$. Note that $L_t \leq x \iff$ there exist n, y such that $t - x \leq y < t$, $S_n = y$, $t - y \leq X_{n+1} \leq x$.

$$\begin{aligned} P(L_t \leq x) &= \sum_{n=1}^{\infty} \int_{t-x}^t \int_{t-y}^x f_{S_n, X_{n+1}}(y, z) dz dy \\ &= \sum_{n=1}^{\infty} \int_{t-x}^t \int_{t-y}^x g_n(y) f_{X_{n+1}}(z) dz dy \\ &= \sum_{n=1}^{\infty} \int_{t-x}^t g_n(y) [e^{-\lambda(t-y)} - e^{-\lambda x}] dy \\ &= \lambda \int_{t-x}^t [e^{-\lambda(t-y)} - e^{-\lambda x}] dy. \end{aligned} \quad (3.25)$$

- Case II, when $x > t$. Note that

$$\{L_t \leq x\} = \{t < S_1 \leq x\} \cup \bigcup_{n=1}^{\infty} \{\text{bus } n \text{ arrives at } y < t \text{ and } t - y < X_{n+1} \leq x\}. \quad (3.26)$$

These are all disjoint events. $P(t < S_1 \leq x) = e^{-\lambda t} - e^{-\lambda x}$, and

$$P(\text{bus } n \text{ arrives at } y < t \text{ and } t - y < X_{n+1} \leq x) = \int_0^t \int_{t-y}^x g_n(y) f_{X_{n+1}}(z) dz dy. \quad (3.27)$$

Adding up the probabilities of the disjoint events, we have

$$\begin{aligned} P(L_t \leq x) &= e^{-\lambda t} - e^{-\lambda x} + \sum_{n=1}^{\infty} \int_0^t g_n(y) [e^{-\lambda(t-y)} - e^{-\lambda x}] dy \\ &= e^{-\lambda t} - e^{-\lambda x} + \lambda \int_0^t [e^{-\lambda(t-y)} - e^{-\lambda x}] dy \\ &= e^{-\lambda t} - e^{-\lambda x} + 1 - e^{-\lambda t} - \lambda t e^{-\lambda x}. \end{aligned} \quad (3.28)$$

■

3.3 Conditional Distribution

February 25th.

Let X and Y have a joint probability density function given by $f(x, y)$. The *conditional probability density function* of X given $Y = y$ is

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} \quad (3.29)$$

and it is defined for $f_Y(y) > 0$. The motivation behind defining this is as follows—

$$f_{X|Y}(x|y)dx = \frac{f(x, y)dxdy}{f_Y(y)dy} = \frac{P(x \leq X \leq x+dx, y \leq Y \leq y+dy)}{P(y \leq Y \leq y+dy)} = P(x \leq X \leq x+dx | y \leq Y \leq y+dy). \quad (3.30)$$

If one is given the conditional probability density function, we can do the following computation as well,

$$P(X \in A | Y = y) = \int_A f_{X|Y}(x|y)dx. \quad (3.31)$$

We can also make sense of a *conditional cumulative distribution function* of X given $Y = Y$.

$$F_{X|Y}(a|y) = P(X \leq a | Y = y) = \int_{-\infty}^a f_{X|Y}(x|y)dx \quad (3.32)$$

Remark 3.8. If X and Y are independent, then the joint density factorizes into the product of the marginals which results in

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x). \quad (3.33)$$

Example 3.9. Suppose the joint density of X and Y is given as

$$f(x, y) = \begin{cases} e^{-\frac{x}{y}} e^{-y} y^{-1} & \text{if } 0 < x < \infty, 0 < y < \infty \\ 0 & \text{if otherwise.} \end{cases} \quad (3.34)$$

Let us compute $P(X > 1 | Y = y)$ for $0 < y < \infty$.

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{e^{-\frac{x}{y}} e^{-y}}{y \int_0^\infty \frac{e^{-\frac{x}{y}} e^{-y}}{y} dx} = \frac{1}{y} e^{-\frac{x}{y}} \text{ for } 0 < x < \infty. \quad (3.35)$$

Thus,

$$P(X > 1 | Y = y) = \int_1^\infty \frac{1}{y} e^{-\frac{x}{y}} dx = e^{-\frac{1}{y}}. \quad (3.36)$$

3.3.1 The t -distribution

Suppose we have $Y \sim \chi_n^2 \equiv \text{Gamma}(\frac{n}{2}, \frac{1}{2})$ and $Z \sim N(0, 1)$ with both independent. Then the t -distribution with n degrees of freedom is defined as

$$T = \frac{Z}{\sqrt{Y/n}} = \sqrt{n} \frac{Z}{\sqrt{Y}}. \quad (3.37)$$

If a $Y = y$ is fixed, then $T = \sqrt{\frac{n}{y}} Z \sim N(0, \frac{n}{y})$. Thus,

$$f_{T|Y}(t|y) = \frac{1}{\sqrt{2\pi \frac{n}{y}}} e^{-\frac{1}{2} \frac{t^2 y}{n}} \quad (3.38)$$

$$\implies f_{T,Y}(t, y) = f_{T|Y}(t|y) f_Y(y) = \frac{1}{\sqrt{2\pi \frac{n}{y}}} e^{-\frac{1}{2} \frac{t^2 y}{n}} \left(\frac{e^{-\frac{y}{2}} y^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \right) \text{ for } t \in \mathbb{R}, y > 0. \quad (3.39)$$

Thus, the probability density function for T can be found out as

$$f_T(t) = \int_0^\infty f_{T|Y}(t|y)dy = \int_0^\infty \frac{1}{\sqrt{2\pi n}} \cdot \frac{e^{-\frac{1}{2}(1+\frac{t^2}{n})y} y^{\frac{n-1}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} dy. \quad (3.40)$$

If we let $c = \frac{1}{2}(1 + \frac{t^2}{n})$ and make the change of variable from x to cy , the intergral transforms as

$$\frac{c^{-\frac{n+1}{2}}}{\sqrt{\pi n} 2^{\frac{n+1}{2}} \Gamma(\frac{n}{2})} \int_0^\infty e^{-x} x^{\frac{(n-1)}{2}} dx = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n} \Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}. \quad (3.41)$$

Note that as $n \rightarrow \infty$,

$$f_T(t) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}. \quad (3.42)$$

3.3.2 The Bivariate Normal Distribution

Five parameters are used here— $\mu_x, \mu_y \in \mathbb{R}$, $\sigma_x, \sigma_y > 0$, and $-1 < \rho < 1$. The joint probability density function is given by

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\right)\right). \quad (3.43)$$

We find the conditional density of X given $Y = y$.

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x, y)}{f_Y(y)} = c_1(y)f(x, y) \\ &= c_2(y) \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{x(y-\mu_y)}{\sigma_x\sigma_y}\right)\right) \\ &= c_2(y) \exp\left(-\frac{1}{2(1-\rho^2)\sigma_x^2}\left(x^2 - 2x\mu_x + \mu_x^2 - 2\rho\frac{\sigma_x}{\sigma_y}x(y-\mu_y)\right)\right) \\ &= c_3(y) \exp\left(-\frac{1}{2(1-\rho^2)\sigma_x^2}\left(x^2 - 2x(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y))\right)\right) \\ &= c_4(y) \exp\left(-\frac{1}{2(1-\rho^2)\sigma_x^2}\left(x - \left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right)\right)^2\right). \end{aligned} \quad (3.44)$$

The integral of the last term must be 1, so we can conclude that it is the probability density function must be that of the normal distribution.

$$X|Y = y \sim N\left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y), \sigma_x^2(1 - \rho^2)\right), \quad (3.46)$$

$$Y|X = x \sim N\left(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x), \sigma_y^2(1 - \rho^2)\right). \quad (3.47)$$

Note that X and Y are independent if and only if $\rho = 0$. Also,

$$f_X(x) = \frac{f(x, y)}{f_{Y|X}(y|x)} = \frac{1}{\sqrt{2\pi}\sigma_x^2} \exp\left(-\frac{1}{2\sigma_x^2}(x - \mu_x)^2\right). \quad (3.48)$$

3.4 Order Statistics

February 28th.

Suppose we have X_1, X_2, \dots, X_n independent and identically distributed continuous random variables with common probability density function f and cumulative distribution function F . We order the X_i 's in such a way that $X_{(1)} < X_{(2)} < \dots < X_{(n)}$. These are termed the *order statistics* corresponding to X_1, X_2, \dots, X_n , and $X_{(k)}$ is termed the k^{th} order statistic. Note that $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ takes the

values $x_1 \leq x_2 \leq \dots \leq x_n$ if and only if for some permutation (i_1, i_2, \dots, i_n) of $(1, 2, \dots, n)$ we have $X_1 = x_{i_1}, X_2 = x_{i_2}, \dots, X_n = x_{i_n}$. Also, for sufficiently small $\varepsilon > 0$,

$$\begin{aligned} P\left(x_{i_1} - \frac{\varepsilon}{2} < X_1 < x_{i_1} + \frac{\varepsilon}{2}, x_{i_2} - \frac{\varepsilon}{2} < X_2 < x_{i_2} + \frac{\varepsilon}{2}, \dots, x_{i_n} - \frac{\varepsilon}{2} < X_n < x_{i_n} + \frac{\varepsilon}{2}\right) \\ \approx \varepsilon^n f(x_{i_1})f(x_{i_2}) \cdots f(x_{i_n}) = \varepsilon^n f(x_1)f(x_2) \cdots f(x_n). \end{aligned} \quad (3.49)$$

Therefore, for $x_1 < x_2 < \dots < x_n$, we have

$$\begin{aligned} P\left(x_1 - \frac{\varepsilon}{2} < X_{(1)} < x_1 + \frac{\varepsilon}{2}, x_2 - \frac{\varepsilon}{2} < X_{(2)} < x_2 + \frac{\varepsilon}{2}, \dots, x_n - \frac{\varepsilon}{2} < X_{(n)} < x_n + \frac{\varepsilon}{2}\right) \\ \approx n! \varepsilon^n f(x_1) \cdots f(x_n) \end{aligned} \quad (3.50)$$

$$\implies f_{(X_{(1)}, X_{(2)}, \dots, X_{(n)})}(x_1, x_2, \dots, x_n) = n! f(x_1)f(x_2) \cdots f(x_n). \quad (3.51)$$

Example 3.10. Three people are distributed on a 1 mile long rong, uniformly. Fix $d \leq \frac{1}{2}$. We are to find the probability that no two people are less distance d apart. Note that the probability density function is $f_{(X_{(1)}, X_{(2)}, X_{(3)})}(x_1, x_2, x_3) = 6$ for $0 \leq x_1 < x_2 < x_3 \leq 1$. The probability is then given us

$$P(X_{(2)} - X_{(1)} \geq d, X_{(3)} - X_{(2)} \geq d) = \int_0^1 \int_{x_1+d}^1 \int_{x_2+d}^1 6 dx_3 dx_2 dx_1 = (1-2d)^3. \quad (3.52)$$

The marginal density of $X_{(k)}$ is given as $f_{X_{(k)}}(y_k) =$

$$\int \cdots \int_{x_1 < \dots < x_{k-1} < y_k < x_{k+1} < \dots < x_n} f_{(X_{(1)}, \dots, X_{(n)})}(x_1, \dots, x_{k-1}, y_k, x_{k+1}, \dots, x_n) dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_n. \quad (3.53)$$

If we note that

$$F_{(X_{(n)})}(y) = P(X_{(n)} \leq y) = P(X_1 \leq y, \dots, X_n \leq y) = [F(y)]^n \quad (3.54)$$

then

$$f_{X_{(n)}}(y) = n[F(y)]^{n-1}f(y). \quad (3.55)$$

Similarly,

$$1 - F_{X_{(1)}}(y) = P(X_{(1)} > y) = P(X_1 > y, X_2 > y, \dots, X_n > y) = [1 - F(y)]^n \quad (3.56)$$

gives us

$$f_{X_{(1)}}(y) = n[1 - F(y)]^{n-1}f(y). \quad (3.57)$$

For the other marginal densities, we work as follows—

$$\begin{aligned} F_{X_{(k)}}(y) &= P(X_{(k)} \leq y) = P(\text{at least } k \text{ of } X_1, \dots, X_n \leq y) \\ &= \sum_{j=k}^n P(\text{exactly } j \text{ of } X_1, \dots, X_n \leq y) \\ &= \sum_{j=k}^n \binom{n}{j} F(y)^j (1 - F(y))^{n-j}. \end{aligned} \quad (3.58)$$

Differentiating this to find the probability density function is a tedious task. We work around this. Note that for $X_{(k)}$ to attain the value y_k , one of the X_1, \dots, X_k should equal y_k , $k-1$ should be less than y_k , and $n-k$ should be greater than y_k . For a fixed partition to satisfy these conditions, we have

$$P(X_{i_1}, X_{i_2}, \dots, X_{i_{k-1}} < y_k, X_{i_k} = y_k, X_{j_1}, X_{j_2}, \dots, X_{j_{n-k}} > y) = F(y)^{k-1} f(y) (1 - F(y))^{n-k}. \quad (3.59)$$

Thus, we can choose any partition in the above mentioned way to get

$$f_{X_{(k)}}(y) = \binom{n}{k-1, n-k, 1} F(y)^{k-1} f(y) (1 - F(y))^{n-k}. \quad (3.60)$$

For $i < j$ the joint density function of $(X_{(i)}, X_{(j)})$, $f_{X_{(i)}X_{(j)}}(y_i, y_j)$ exists for $y_i < y_j$. From the same intuitive reasoning as before, we have

$$f_{X_{(i)}X_{(j)}}(y_i, y_j) = \binom{n}{i-1, 1, j-i-1, 1, n-j} F(y_i)^{i-1} f(y_i) (F(y_j) - F(y_i))^{j-i-1} f(y_j) (1 - F(y_j))^{n-j}. \quad (3.61)$$

Example 3.11. We are to find the cumulative distribution function of $R = X_{(n)} - X_{(1)}$, the range of the X_i 's. We simply have

$$\begin{aligned} P(R \leq a) &= P(X_{(n)} - X_{(1)} \leq a) = \iint_{x_n - x_1 \leq a} f_{X_{(1)}X_{(n)}}(x_1, x_n) dx_1 dx_n \\ &= \int_{-\infty}^{\infty} \int_{x_1}^{x_1+a} \frac{n!}{(n-2)!} [F(x_n) - F(x_1)]^{n-2} f(x_1) f(x_n) dx_n dx_1 \end{aligned} \quad (3.62)$$

Making the substitution $y = F(x_n) - F(x_1)$ with $dy = f(x_n) dx_n$ will help compute the function. The computation is left as an exercise to the reader.

March 4th.

Let X_1, X_2, \dots, X_n be independent and identically distributed as $\text{Exp}(\alpha)$. These will denote the service times of n counters in a post office, commencing at time 0. Thus, $X_{(i)}$ will denote the time of the i^{th} discharge. We have

$$P(X_{(1)} > t) = P(X_1 > t, \dots, X_n > t) = \prod_{i=1}^n P(X_i > t) = e^{-n\alpha t} \implies X_{(1)} \sim \text{Exp}(n\alpha). \quad (3.63)$$

Also,

$$P(X_{(n)} \leq t) = P(X_1 \leq t, \dots, X_n \leq t) = \prod_{i=1}^n P(X_i \leq t) = (1 - e^{-\alpha t})^n.$$

Intuitively, from the memoryless property of the exponential distribution, we can infer a proposition.

Proposition 3.12. *The n variables $X_{(1)}, X_{(2)} - X_{(1)}, \dots, X_{(n)} - X_{(n-1)}$ are independent, and the density of $X_{(k+1)} - X_{(k)}$ is $(n-k)\alpha e^{-(n-k)\alpha t}$.*

Proof. We show the case for $n = 3$. The proof can then be generalized. We have

$$f_{X_{(1)}X_{(2)}X_{(3)}}(z_1, z_2, z_3) = 6f_{X_1, X_2, X_3}(z_1, z_2, z_3) = 6\alpha^3 e^{-\alpha} e^{-\alpha(z_1+z_2+z_3)}. \quad (3.64)$$

Also,

$$\begin{aligned} P(X_{(1)} > t_1, X_{(2)} - X_{(1)} > t_2, X_{(3)} - X_{(2)} > t_3) &= 6 \int_{t_1}^{\infty} \alpha e^{-\alpha z_1} \int_{z_1+t_2}^{\infty} \alpha e^{-\alpha z_2} \int_{z_2+t_3}^{\infty} \alpha e^{-\alpha z_3} dz_3 dz_2 dz_1 \\ &= e^{-\alpha t_3} e^{-2\alpha t_2} e^{-3\alpha t_1}. \end{aligned} \quad (3.65)$$

Therefore, $X_{(1)} \sim \text{Exp}(3\alpha)$, $X_{(2)} - X_{(1)} \sim \text{Exp}(2\alpha)$, and $X_{(3)} - X_{(2)} \sim \text{Exp}(\alpha)$. ■

We ask a few questions; suppose A and B are currently being served at the office, and a third clerk C enters. The probability that C leaves last is one-half due to the memoryless property of the exponential distribution. To find the total time spent by C , we have $T = X_{(1)} + Z$, where Z is the distribution of C 's time spent being served. The distribution of T is given as

$$f(t) = \int_{\mathbb{R}} f_{X_{(1)}}(x) f_Z(t-x) dx = \int_0^t 2\alpha e^{-2\alpha x} \alpha e^{-\alpha(t-x)} dx = 2\alpha e^{-\alpha t} (1 - e^{-\alpha t}).$$

Let us also compute the distribution of the time of last departure, \tilde{T} . Note that C enters only when one of A or B is served. The first discharge has distribution $Z \sim \text{Exp}(2\alpha)$. When C enters, we have 2 independent $\text{Exp}(\alpha)$ random variables, where the last discharge is given as $X_{(2)}$. Note that X_1, X_2 are independent of Z . From the previous question, we know that $X_{(2)} = X_{(2)} - X_{(1)} + X_{(1)}$ has distribution

$\text{Exp}(2\alpha) + \text{Exp}(\alpha)$, and density $2\alpha e^{-\alpha t}(1 - e^{-\alpha t})$. Therefore, the distribution of \tilde{T} matches that of $Z + X_{(2)}$. Integrating gives us

$$f_{\tilde{T}}(\tilde{t}) = 4\alpha(e^{-\alpha\tilde{t}} - e^{-2\alpha\tilde{t}}) = \alpha\tilde{t}e^{-2\alpha\tilde{t}}. \quad (3.66)$$

Moving away from the exponential distribution, let us consider X_1, X_2, \dots, X_n to be independent and identically distributed to $\text{Unif}[0, 1]$. Note that $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ partition the interval $[0, 1]$ into subintervals of length $l_1 = X_{(1)}, l_2 = X_{(2)} - X_{(1)}, \dots, l_n = X_{(n)} - X_{(n-1)}$. The lengths l_1, l_2, \dots, l_n are not independent since they have to satisfy $\sum l_i = 1$. The distributions of l_i 's turn out to be identical.

As another exercise, let X_1, X_2 be independent and randomly chosen uniformly on the unit circle. Let $r(x, y)$ denote the clockwise circular distance from points x to y on the unit circle. Note that $r(X_1, X_2)$ has the same distribution as $r(X_2, X_1)$, computed as $\text{Unif}[0, 1]$.

3.5 Joint Distribution of Functions of Random Variables

March 7th.

Suppose (X_1, X_2) has joint probability density function $f_{X_1 X_2}$, and suppose $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ for some functions g_1, g_2 . We assume that g_1 and g_2 satisfy the following conditions:

- The equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 , say, $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$.
- g_1 and g_2 have continuous partial derivatives at all (x_1, x_2) and the determinant of the *Jacobian matrix* is

$$J(x_1, x_2) = \det \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{pmatrix} = \frac{\partial g_1}{\partial x_1} \frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2} \frac{\partial g_2}{\partial x_1} \neq 0. \quad (3.67)$$

Then (Y_1, Y_2) has the joint density

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1 X_2}(x_1, x_2) |J(x_1, x_2)|^{-1} \quad (3.68)$$

where $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$. The proof of this comes from multivariate analysis.

Remark 3.13. The above can be extended to more than 2 random variables. Suppose that we have the probability density function $f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$ and we have $Y_1 = g_1(X_1, X_2, \dots, X_n), Y_2 = g_2(X_1, X_2, \dots, X_n), \dots, Y_n = g_n(X_1, X_2, \dots, X_n)$. We have to assume that all the g_i 's have continuous partial derivatives, and we have to assume that

$$J(x_1, x_2, \dots, x_n) = \det \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \dots & \frac{\partial g_n}{\partial x_n} \end{pmatrix} \neq 0 \text{ for all } (x_1, x_2, \dots, x_n). \quad (3.69)$$

We also suppose that that $y_1 = g_1(x_1, \dots, x_n), \dots, y_n = g_n(x_1, \dots, x_n)$ has a unique solution given by $x_1 = h_1(y_1, \dots, y_n), \dots, x_n = h_n(y_1, \dots, y_n)$. The joint density function of (Y_1, \dots, Y_n) is then given by

$$f_{Y_1 Y_2 \dots Y_n}(y_1, y_2, \dots, y_n) = f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) |J(x_1, \dots, x_n)|^{-1} \quad (3.70)$$

where $x_i = h_i(y_1, \dots, y_n)$.

3.5.1 Conditional Expectation and Variance

Let (X, Y) have a joint probability density function $f(x, y)$, and suppose $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a function.

Proposition 3.14. *The expectation of $g(X, Y)$ is given as*

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dy dx.$$

Remark 3.15. By this proposition, if we pick $g(x, y) = x$ for all $(x, y) \in \mathbb{R}^2$, we get

$$EX = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dy dx = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Proof. Assume that $g(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$. We have

$$\begin{aligned} E[g(X, Y)] &= \int_0^{\infty} P(g(X, Y) > t) dt \\ &= \int_0^{\infty} \left(\iint_{\{(x, y): g(x, y) > t\}} f(x, y) dy dx \right) dt \\ &= \iint \int_0^{g(x, y)} f(x, y) dt dy dx \\ &= \iint g(x, y) f(x, y) dy dx. \end{aligned} \quad (3.71)$$

For a general g , we simply work with $g = g^+ - g^-$ where $g^+(x, y) = \max\{g(x, y), 0\}$ and $g^-(x, y) = \max\{-g(x, y), 0\}$. ■

If (X, Y) have a joint probability density function $f(x, y)$, then it can be shown that $E[X + Y] = EX + EY$.

Example 3.16. Suppose $X, Y \sim \text{Unif}[0, L]$. We compute $E|X - Y|$. Noting that the joint probability density function is $f(x, y) = \frac{1}{L^2}$ for all $(x, y) \in [0, L] \times [0, L]$, we have

$$\begin{aligned} E|X - Y| &= \int_0^L \int_0^L \frac{|x - y|}{L^2} dy dx = \frac{1}{L^2} \int_0^L \left(\int_0^x (x - y) dy + \int_x^L (y - x) dy \right) dx \\ &= \frac{1}{L^2} \int_0^L \left(\frac{x^2}{2} + \frac{(L - x)^2}{2} \right) dx = \frac{L}{3}. \end{aligned} \quad (3.72)$$

Recall that the conditional density of X given $Y = y$ is given as

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

provided that $f_Y(y) > 0$.

Definition 3.17. The conditional expectation is given as

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$$

provided that $f_Y(y) > 0$.

Example 3.18. We compute $E[X|Y = y]$ for $f(x, y) = \frac{e^{-\frac{x}{y}} e^{-y}}{y}$ for $x, y > 0$. It can be shown that the marginal density $f_Y(y)$ is $f_Y(y) = e^{-y}$. Thus, the conditional density is

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{e^{-\frac{x}{y}} e^{-y}}{y e^{-y}} = \frac{1}{y} e^{-\frac{x}{y}}. \quad (3.73)$$

The conditional expectation is then given as

$$E[X|Y = y] = \int_0^{\infty} \frac{x e^{-\frac{x}{y}}}{y} dx = y.$$

It can be shown that

$$E[g(X)|Y = y] = \int g(x) f_{X|Y}(x|y) dx. \quad (3.74)$$

Proposition 3.19. *We have*

$$EX = E[E[X|Y]].$$

Proof. Since $E[X|Y = y] = \int x f_{X|Y}(x|y) dx$ is a function of y , we have

$$\begin{aligned} E[E[X|Y]] &= \int_{-\infty}^{\infty} f_Y(y) \left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx \right) dy = \int_{-\infty}^{\infty} f_Y(y) \int_{-\infty}^{\infty} x \frac{f(x, y)}{f_Y(y)} dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dy dx = \int_{-\infty}^{\infty} x f_X(x) dx = EX. \end{aligned}$$

■

The conditional variance is also given as

$$\text{Var}(X|Y = y) = E[X^2|Y = y] - (E[X|Y = y])^2 = \int x^2 f_{X|Y}(x|y) dx - \left(\int x f_{X|Y}(x|y) dx \right)^2. \quad (3.75)$$

Also recall the conditional variance formula—

$$\begin{aligned} \text{Var}(X) &= E[(X - EX)^2] = E[(X - E[X|Y] + E[X|Y] - EX)^2] \\ &= E[(X - E[X|Y])^2 + (E[X|Y] - EX)^2 + 2(X - E[X|Y])(E[X|Y] - EX)] \\ &= E[\text{Var}(X|Y)] + \text{Var}(E[X|Y]) + 0. \end{aligned} \quad (3.76)$$

Example 3.20. We look at the bivariate normal distribution. The probability density function is given by

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left(-\frac{1}{2(1-\rho^2)} \left(\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right) \right). \quad (3.77)$$

It can be shown that the correlation between X and Y is indeed ρ , that is, $\frac{E(XY) - \mu_x\mu_y}{\sigma_x\sigma_y} = \rho$. $E(XY)$ is computed by considering $E(XY) = E[E(XY|Y)]$.

Example 3.21. Suppose (S, R) has the following distributions; $R|S = s \sim N(s, 1)$ and $S \sim N(\mu, \sigma^2)$. We wish to compute $E[S|R = r]$ and $\text{Var}(S|R = r)$. The former can be found out by first computing

$$f_{S|R}(s|r) = \frac{f_{S,R}(s, r)}{f_R(r)} = \frac{f_{R|S}(r|s)f_S(s)}{f_R(r)}. \quad (3.78)$$

March 18th.

For random variables X and Y which are jointly distributed, the following hold true:

- $E[(X - a)^2] \geq E[(X - EX)^2]$,
- $E[(Y - g(X))^2] \geq E[(Y - E[Y|X])^2]$.

The best predictor of Y using X is $E[Y|X]$.

Example 3.22. Suppose X is a continuous random variable that we wish to discretize. We first fix an increasing set of numbers $\dots < a_{-2} < a_{-1} < a_0 < a_1 < a_2 < \dots$ such that $\lim_{i \rightarrow \infty} a_i = \infty$ and $\lim_{i \rightarrow -\infty} a_i = -\infty$. Let Y be a random variable taking the value y_i when $a_i < X \leq a_{i+1}$. We wish to choose y_i that minimizes $E[(X - Y)^2]$. For these optimal y_i 's one can also show that $EY = EX$ and $\text{Var}Y = \text{Var}X - E[(X - Y)^2]$.

Start with

$$\begin{aligned} E[(X - Y)^2] &= \sum_i E[(X - Y)^2 | a_i < X \leq a_{i+1}] P(a_i < X \leq a_{i+1}) \\ &= \sum_i E[(X - y_i)^2 | a_i < X \leq a_{i+1}] P(a_i < X \leq a_{i+1}). \end{aligned} \quad (3.79)$$

It is enough to minimize $E[(X - y_i)^2 | a_i < X \leq a_{i+1}]$. y_i is then computed to be

$$y_i = E[X | a_i < X \leq a_{i+1}] = \frac{E[X \cdot 1_{\{a_i < X \leq a_{i+1}\}}]}{P(a_i < X \leq a_{i+1})} = \frac{\int_{a_i}^{a_{i+1}} x f_X(x) dx}{\int_{a_i}^{a_{i+1}} f_X(x) dx}. \quad (3.80)$$

To show that the expectations of Y and X are equal, we have

$$EY = \sum_i y_i P(Y = y_i) = \sum_i E[X | a_i < X \leq a_{i+1}] P(a_i < X \leq a_{i+1}) = EX. \quad (3.81)$$

Define a random variable I to be $I = i$ if $a_i < X \leq a_{i+1}$. Then

$$\text{Var} X = \text{Var}(E[X|I]) + E[\text{Var}(X|I)] \quad (3.82)$$

can be used. We can compute $E[X|I]$ as $y_I = Y$ and $\text{Var}(X|I)$ as $E[(X - Y)^2 | I]$. Thus, the variance equality follows.

Chapter 4

CONVERGENCE OF RANDOM VARIABLES

4.1 Types of Convergence

We begin with a probability space (Ω, P) , and define a sequence of random variables $X_n : \Omega \rightarrow \mathbb{R}$ for $n \geq 1$ and also have another $X : \Omega \rightarrow \mathbb{R}$.

4.1.1 Almost Sure Convergence

Definition 4.1. We first discuss the notion of *almost sure convergence*. A sequence of random variables X_n converges almost surely to X , that is, $X_n \xrightarrow{\text{a.s.}} X$ if

$$P(\{\omega : X_n(\omega) \rightarrow X(\omega)\}) = 1.$$

Example 4.2. Suppose that $(\Omega, P) = ([0, 1], \text{Unif})$. Suppose that q_n is an enumeration of the rationals in $[0, 1]$. We define $X(\omega) = 1$ and $X_n(\omega) = 1$ if $\omega \in [0, 1] \setminus \{q_n\}$ and 0 otherwise. We find that

$$A = \{\omega : X_n(\omega) \rightarrow X(\omega)\} = [0, 1] \quad (4.1)$$

since for any rational ω , further enough terms lead to $X_n(\omega) = 0$ and for any irrational ω , we always have $X_n(\omega) = 1$. Thus, $P(A) = 1$.

Example 4.3. As a continuation to the previous example, if we define the sequence of random variables as

$$X_n(\omega) = \begin{cases} 0 & \text{if } \omega \in \{q_1, \dots, q_n\}, \\ 1 & \text{if otherwise,} \end{cases} \quad (4.2)$$

we find that $\{\omega : X_n(\omega) \rightarrow X(\omega)\} = [0, 1] \setminus \mathbb{Q}$. The sequence converges to 0 for rational ω . But, in our probability space, $P([0, 1] \setminus \mathbb{Q}) = 1$.

Example 4.4. If we define the sequence of random variables as $X_n = n1_{[0, \frac{1}{n}]}$, we find that $X_n \xrightarrow{\text{a.s.}} 0$. However, $EX_n = 1$ and $E[0] = 0$. The expectation is not preserved in almost sure convergence

4.1.2 Convergence in Probability

Definition 4.5. We discuss the notion of *convergence in probability*. We say that a sequence of random variables X_n converges in probability to X , that is, $X_n \xrightarrow{P} X$ if for every $\varepsilon > 0$,

$$P(|X_n - X| > \varepsilon) \rightarrow 0.$$

To show that a sequence of random variables X_n converges in probability to X , it is enough to show that $P(|X_n - X| > \frac{1}{k}) \rightarrow 0$ for all $k \geq 1$.

Theorem 4.6. $X_n \xrightarrow{a.s.} X$ implies $X_n \xrightarrow{P} X$.

Proof. We have

$$\begin{aligned} \{\omega : X_n(\omega) \rightarrow X(\omega)\} &= \{\omega : \text{for all } k \geq 1, \exists N \text{ such that for all } n \geq N, |X_n(\omega) - X(\omega)| \leq \frac{1}{k}\} \\ &= \bigcap_{k \geq 1} \bigcup_N \bigcap_{n \geq N} \{|X_n - X| \leq \frac{1}{k}\}. \end{aligned} \quad (4.3)$$

The probability of this set is 1 due to almost sure convergence. Since the intersection of these events has a probability of 1, each inside event must also have a probability of 1; thus,

$$P\left(\bigcup_N \bigcap_{n \geq N} \{|X_n - X| \leq \frac{1}{k}\}\right) = 1. \quad (4.4)$$

We note that $\bigcap_{n \geq N} \{|X_n - X| \leq \frac{1}{k}\}$ are events increasing in N . Therefore,

$$P\left(\bigcap_{n \geq N} \{|X_n - X| \leq \frac{1}{k}\}\right) \rightarrow 1 \implies P(|X_n - X| \leq \frac{1}{k}) \rightarrow 1. \quad (4.5)$$

■

However, the converse is not true.

March 21st.

Theorem 4.7. Let X_1, X_2, \dots be a sequence of random variables. If $EX_n \rightarrow c$ and $\text{Var}(X_n) \rightarrow 0$, then $X_n \xrightarrow{P} c$.

Proof. We have

$$P(|X_n - c| > \varepsilon) = P(|X_n - EX_n + EX_n - c| > \varepsilon) \leq P(|X_n - EX_n| + |EX_n - c| > \varepsilon). \quad (4.6)$$

Choose a natural N such that $|EX_n - c| < \frac{\varepsilon}{2}$, which gives us

$$P(|X_n - c| > \varepsilon) \leq P(|X_n - EX_n| > \frac{\varepsilon}{2}) \leq \frac{E[|X_n - EX_n|^2]}{\varepsilon^2/4} = \text{Var}X_n \frac{4}{\varepsilon^2} \rightarrow 0. \quad (4.7)$$

■

A consequence of this is the weak law of large numbers.

Corollary 4.8. Let X_1, X_2, \dots be independent and identically distributed random variables, with $EX_1 = \mu$ and $\text{Var}X_1 = \sigma^2 < \infty$. Let $\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$. Then $\bar{X}_n \xrightarrow{P} \mu$.

Proof. We note that $E\bar{X}_n = \mu$ and $\text{Var}\bar{X}_n = \frac{1}{n}\sigma^2$. The above theorem proves the law. ■

Example 4.9. We put n balls randomly into n boxes. Here, let X_n denote the number of empty boxes. We are interested in $Y_n = \frac{X_n}{n}$, the proportion of empty boxes. We claim that Y_n converges to e^{-1} . To compute the mean, we write $EX_n = E(\sum_{i=1}^n 1_{A_i})$, where A_i is the event that the i^{th} box is empty. Thus, this is $EX_n = \sum_{i=1}^n \frac{(n-1)^n}{n^n} = n(1 - \frac{1}{n})^n \implies EY_n = (1 - \frac{1}{n})^n \rightarrow \frac{1}{e}$.

For the variance, $\text{Var}X_n = \sum_{i,j=1}^n \text{Cov}(1_{A_i}, 1_{A_j})$. For $i = j$, this is

$$\text{Cov}(1_{A_i}, 1_{A_j}) = E[1_{A_i}] - (E[1_{A_i}])^2 = P(A_i)(1 - P(A_i)) = \frac{(n-1)^n}{n^n} \left(1 - \frac{(n-1)^n}{n^n}\right). \quad (4.8)$$

For $i \neq j$, we have

$$\text{Cov}(1_{A_i}, 1_{A_j}) = E[1_{A_i} 1_{A_j}] - P(A_i)P(A_j) = \left(\frac{n-2}{n}\right)^n - \left(\frac{n-1}{n}\right)^{2n}. \quad (4.9)$$

Thus, the final variance is calculated as

$$\text{Var}X_n = n \left(\frac{n-1}{n}\right)^n \left(1 - \left(\frac{n-1}{n}\right)^n\right) + n(n-1) \left(\left(\frac{n-2}{n}\right)^n - \left(\frac{n-1}{n}\right)^{2n}\right) \quad (4.10)$$

$$\implies \text{Var}Y_n = \frac{1}{n^2} \text{Var}X_n \rightarrow 0. \quad (4.11)$$

Thus, $Y_n \xrightarrow{P} e^{-1}$.

Example 4.10. We toss a p -coin, and look at X_n , the number of head runs in n tosses. We claim that $Y_n = \frac{1}{n}X_n \xrightarrow{p} p(1-p)$. We rewrite X_n as $X_n = \sum_{i=1}^n 1_{A_i}$ where A_i is the event that a head run starts at position i . We have

$$EX_n = \sum_{i=1}^n P(A_i) = P(A_1) + \sum_{i=2}^n P(A_i) = p + (n-1)(1-p)p \implies EY_n \rightarrow p(1-p). \quad (4.12)$$

For the variance,

$$\text{Var}X_n = \sum_{i,j=1}^n \text{Cov}(1_{A_i}, 1_{A_j}) = \sum_{i=1}^n \text{Var}1_{A_i} + 2 \sum_{i<j} \text{Cov}(1_{A_i}, 1_{A_j}). \quad (4.13)$$

The variance of a single indicator, $\text{Var}1_{A_i}$ is $p - p^2$ if $i = 1$, and $(1-p)p - ((1-p)p)^2$ if $i > 1$. For the covariance, if $|i-j| \geq 2$, the events A_i and A_j are clearly independent, so $\text{Cov}(1_{A_i}, 1_{A_j}) = 0$.

$$\text{Var}Y_n = \frac{1}{n^2} \text{Var}X_n = \frac{1}{n^2} (p - p^2 + (n-1)((1-p)p - (1-p)^2 p^2)) + \frac{1}{n^2} (\dots) \rightarrow 0. \quad (4.14)$$

Example 4.11. We revisit the coupon collector problem; there are n types of coupons, and each purchase results in one of the n types with equal probability. T_n denotes the number of purchases to get n types. The claim is that $\frac{1}{n \log n} T_n \xrightarrow{p} 1$.

Clearly, $T_1 = 1$, $T_2 - T_1 \sim \text{Geom}(\frac{n-1}{n})$, $T_3 - T_2 \sim \text{Geom}(\frac{n-2}{n})$, etc. These are all independent quantities, and the expectation of a geometric distribution is $\frac{1}{p}$. Thus,

$$\begin{aligned} T_n &= T_1 + (T_2 - T_1) + \dots + (T_n - T_{n-1}) \\ \implies ET_n &= n \left(1 + \frac{1}{2} + \dots + \frac{1}{n} \right) \implies \frac{1}{n \log n} ET_n = \frac{1 + \frac{1}{2} + \dots + \frac{1}{n}}{\log n} \rightarrow 1. \end{aligned} \quad (4.15)$$

and the variance is

$$\frac{1}{n^2 (\log n)^2} \text{Var}T_n = \frac{1}{(n \log n)^2} n \sum_{i=1}^n \frac{(i-1)}{(n-(i-1))^2} \leq \frac{1}{n^2 (\log n)^2} n^2 \sum_{i=1}^n \frac{1}{i^2} \rightarrow 0. \quad (4.16)$$

Theorem 4.12 (The *strong law of large numbers*). Let X_1, X_2, \dots be independent and identically distributed random variables with $EX_1 = \mu \in (-\infty, \infty)$. Let $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$. Then $\bar{X}_n \xrightarrow{\text{a.s.}} \mu$.

Proof. We prove this assuming that $K = EX_1^4 < \infty$. Without the loss of generality, let $\mu = 0$. Denote $S_n = \sum_{i=1}^n X_i$. The terms inside ES_n^4 are any one of the form $X_i^4, X_i^3 X_j, X_i^2 X_j^2, X_i^2 X_j X_k, X_i X_j X_k X_l$, where i, j, k, l are distinct. We have assumed that $\mu = 0$, so

$$E[X_i^3 X_j] = E[X_i^3] EX_j = 0, \quad E[X_i^2 X_j X_k] = E[X_i^2] EX_j EX_k = 0, \quad E[X_i X_j X_k X_l] = 0. \quad (4.17)$$

Thus, the only remaining terms are

$$\begin{aligned} ES_n^4 &= n EX_1^4 + \binom{n}{2} \binom{4}{2} (EX_1^2)^2 \\ \implies E\bar{X}_n^4 &\leq \frac{1}{n^3} EX_1^4 + \frac{3}{n^2} (EX_1^2)^2 \implies \sum_{n=1}^{\infty} E\bar{X}_n^4 = \sum_{n=1}^{\infty} \frac{1}{n^3} EX_1^4 + \frac{3}{n^2} (EX_1^2)^2 < \infty. \end{aligned} \quad (4.18)$$

The expectation of the infinite sum is finite, which means

$$E \left(\sum_{n=1}^{\infty} \bar{X}_n^4 \right) < \infty \implies P \left(\sum_{n=1}^{\infty} \bar{X}_n^4 < \infty \right) = 1 \implies P(\bar{X}_n \rightarrow 0) = 1. \quad (4.19)$$

■

4.1.3 Convergence in Distribution

March 28th.

Definition 4.13. We discuss the notion of *convergence in distribution*. Suppose X_1, X_2, \dots are a sequence of random variables, and let X be another random variables. Denote the cumulative distribution function of X_n by F_n and that of X by F . We say that the sequence of random variables X_n converges to X if $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ for all x at which F is continuous.

We write $X_n \Rightarrow X$ or $F_n \Rightarrow F$, or even $X_n \xrightarrow{d} X$ if X_n converges to X in distribution.

Remark 4.14. When X has a density function f , then F is continuous everywhere and we need the sequence to converge everywhere.

When F is continuous, we have $P(a < X_n \leq b) = P(X_n \leq b) - P(X_n \leq a) = F_n(b) - F_n(a) \rightarrow F(b) - F(a)$. Similarly, we can show that $P(a \leq X_n \leq b) \rightarrow P(a < X \leq b)$ since $P(X = a) = 0$; we wish to show that $P(X_n = a) = 0$ is X_n converges to a X with continuous F . This can be shown as

$$P(X_n = a) \leq P(a - \varepsilon < X_n \leq a) = F_n(a) - F_n(a - \varepsilon) \implies \limsup_{n \rightarrow \infty} P(X_n = a) \leq F(a) - F(a - \varepsilon).$$

Letting $\varepsilon \downarrow 0$, we find $\limsup_{n \rightarrow \infty} P(X_n = a) = 0 \implies \lim_{n \rightarrow \infty} P(X_n = a) = 0$.

Example 4.15. Suppose $X_n \sim \text{Bin}(n, \frac{\lambda}{n})$ and $X \sim \text{Poisson}(\lambda)$. We have already seen that $P(X_n = k) \rightarrow P(X = k)$. For $x \geq 0$, we have

$$\lim_{n \rightarrow \infty} P(X_n \leq x) = \lim_{n \rightarrow \infty} \sum_{0 \leq k \leq x} P(X_n = k) = \sum_{0 \leq k \leq x} \lim_{n \rightarrow \infty} P(X_n = k) = \sum_{0 \leq k \leq x} P(X = k) = P(X \leq x). \quad (4.20)$$

Suppose X_1, X_2, \dots are independent and identically distributed random variables with finite mean μ and finite variance σ^2 . From the law of large numbers, we know that $\frac{S_n}{n} \rightarrow \mu$ where S_n are the partial sums. The central limit theorem roughly states that $\frac{S_n}{n} = \mu + O(\frac{1}{\sqrt{n}})$. We state this more formally.

The Central Limit Theorem

Theorem 4.16 (The *central limit theorem*). Let X_1, X_2, \dots be independent and identically distributed random variables with with quantities defined as above. The central limit theorem states that

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow N(0, 1).$$

For $-\infty < a \leq b \leq \infty$, we have

$$P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \xrightarrow{n \rightarrow \infty} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

The left hand side may be rewritten as $P(n\mu + a\sigma\sqrt{n} \leq S_n \leq n\mu + b\sigma\sqrt{n})$ and the right hand side may be rewritten as $P(m\mu + a\sigma\sqrt{m} \leq Z_1 + \dots + Z_m \leq m\mu + b\sigma\sqrt{m})$ where $Z_1, \dots, Z_m \sim N(\mu, \sigma^2)$ are independent.

Example 4.17. Suppose a fair coin is tossed a million times, and denote S to be the number of heads. This is the sum of a million independent and identically distributed $\text{Ber}(\frac{1}{2})$ variables. Here, $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{4}$. In this case, $n\mu = 500,000$ and $\sigma\sqrt{n} = 500$. The central limit theorem states that $\frac{S - 500,000}{500} \approx N(0, 1)$. Using $a = -2.576$ and $b = 2.576$, we have

$$P(498712 \leq S \leq 501288) \approx \int_{-2.576}^{2.576} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \approx 0.99. \quad (4.21)$$

Suppose $g : \mathbb{R} \rightarrow \mathbb{R}$ is defined by $g(x) = e^{-\frac{1}{x}}$ for $x > 0$, and $g(x) = 0$ for $x \leq 0$. For $x \neq 0$, clearly $g(x)$ is infinitely differentiable. We shall show that $g(x)$ is also infinitely differentiable at $x = 0$ with $g^{(k)}(0) = 0$ always. We have

$$\lim_{x \uparrow 0} \frac{g(x) - g(0)}{x - 0} = \lim_{x \uparrow 0} \frac{0}{x} = 0 \quad (4.22)$$

and

$$\lim_{x \downarrow 0} \frac{g(x) - g(0)}{x - 0} = \lim_{x \downarrow 0} \frac{e^{-\frac{1}{x}}}{x} = 0. \quad (4.23)$$

Thus, $g(x)$ is differentiable at 0, and there exists a C such that $|g'(x)| \leq C$ for all $x \in \mathbb{R}$. For any k^{th} derivative of $g(x)$ for $x > 0$, we obtain $g^{(k)}(x) = e^{-\frac{1}{x}} p_{2k}(\frac{1}{x})$ where p_{2k} is a polynomial of degree $2k$. Again, the only problem is at 0. By the induction hypothesis, $g^{(k-1)}(0) = 0$. We have

$$\lim_{x \downarrow 0} \frac{g^{(k-1)}(x) - g^{(k-1)}(0)}{x - 0} = \lim_{x \downarrow 0} \frac{g^{(k-1)}(x)}{x} = \lim_{x \downarrow 0} \frac{e^{-\frac{1}{x}} p_{2(k-1)}(\frac{1}{x})}{x} = 0 \quad (4.24)$$

and left hand limit is easily seen to be zero. Hence, $g^{(k)}(0) = 0$ and there always exists a C_k such that $|g^{(k)}(x)| \leq C_k$ for all $x \in \mathbb{R}$. Thus, $g \in C_b^\infty$, the set of functions on \mathbb{R} that are infinitely differentiable with bounded derivatives, including the 0^{th} derivative.

For any $-\infty < a < b < \infty$, let $g_{a,b}(x) = g(x-a)g(b-x)$. We note that $g_{a,b} \equiv 0$ on $\mathbb{R} \setminus [a, b]$ and $g_{a,b} \in C_b^\infty$. Also, $g > 0$ on (a, b) . Define $C_{a,b}$ to be the integral $\int_{\mathbb{R}} g_{a,b}(x) dx > 0$. Lastly, defining $h_{a,b}(x) = \frac{1}{C_{a,b}} g_{a,b}(x)$ gives a probability density function. Denote the cumulative distribution function by $H_{a,b}$.

$$H_{a,b}(x) = \begin{cases} 0 & \text{if } x \in (-\infty, a], \\ \text{strictly increasing} & \text{if } x \in (a, b), \\ 1 & \text{if } x \in [b, \infty). \end{cases} \quad (4.25)$$

We may also note that $H_{a,b} \in C_b^\infty$. We have defined this distribution for the following lemma.

Lemma 4.18. *Let X be a random variable and suppose X_1, X_2, \dots be a sequence of random variables such that for each $f \in C_b^\infty$ we have $Ef(X_n) \rightarrow Ef(X)$. Then $X_n \rightrightarrows X$.*

Proof. Let F_n be the cumulative distribution function of X_n and F be that of X . Fix $t \in \mathbb{R}$ at which F is continuous, and also fix $s < t$. We need to show $F_n(t) \xrightarrow{n \rightarrow \infty} F(t)$. Consider $H_{s,t}$ defined as above; $H_{s,t} \in C_b^\infty$. Look at $1 - H_{s,t} \leq 1_{(-\infty, t]}$. We then have

$$\begin{aligned} E[1 - H_{s,t}(X_n)] &\leq E[1_{(-\infty, t]}(X_n)] = F_n(t) \\ \implies E[1 - H_{s,t}(X)] &= \lim_{n \rightarrow \infty} E[1 - H_{s,t}(X_n)] \leq \liminf_{n \rightarrow \infty} F_n(t) \end{aligned} \quad (4.26)$$

This gives us $F(s) \leq \liminf_{n \rightarrow \infty} F_n(t)$. Letting $s \uparrow t$, we get $F(t) \leq \liminf_{n \rightarrow \infty} F_n(t)$.

Similarly, consider $H_{t,s}$ for $s > t$. We get

$$\limsup_{n \rightarrow \infty} F_n(t) = \limsup_{n \rightarrow \infty} E[1_{(\infty, t]}(X_n)] \leq \limsup_{n \rightarrow \infty} E[1 - H_{t,s}(X_n)] = E[1 - H_{t,s}(X)] \leq E[1_{(-\infty, s]}(X)] = F(s). \quad (4.27)$$

Letting $s \downarrow t$, we have $\limsup_{n \rightarrow \infty} F_n(t) \leq F(t)$. Also considering $F(t) \leq \liminf_{n \rightarrow \infty} F_n(t)$, we obtain $F_n(t) \rightarrow F(t)$. ■

April 1st.

We now state a proof of the central limit theorem, under the assumption that $E|X_i - \mu|^3 < \infty$.

Proof of the central limit theorem. Fix n . Define $Y_i = \frac{X_i - \mu}{\sigma\sqrt{n}}$. Thus, the Y_1, \dots, Y_n are also independent and identically distributed with $EY_i = 0$ and $\text{Var}Y_i = \frac{1}{n}$. Thus, $T_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \sum_{i=1}^n Y_i$. From the previous lemma, it is enough to show that $Ef(T_n) \xrightarrow{n \rightarrow \infty} Ef(Z)$ for each $f \in C_b^\infty$ where $Z \sim N(0, 1)$. For now, fix $f \in C_b^\infty$. Let $Z_1, Z_2, \dots, Z_n \sim N(0, \frac{1}{n})$ be independent. Note that $Z = Z_1 + \dots + Z_n \sim N(0, 1)$. For $i = 0, 1, \dots, n$, define $A_i = Y_1 + \dots + Y_{i-1} + Y_i + Z_{i+1} + \dots + Z_n$. Note that $A_0 = Z$ and $A_n = T_n$. We then have

$$f(T_n) - f(Z) = f(A_n) - f(A_0) = \sum_{i=1}^n f(A_i) - f(A_{i-1}). \quad (4.28)$$

Define $B_i = Y_1 + \cdots + Y_{i-1} + Z_{i+1} + \cdots + Z_n$ where we have omitted Y_i from the summation. Then, $A_i = B_i + Y_i$ and $A_{i-1} = B_i + Z_i$. Since $f \in C_b^\infty$, there exists a real C such that $|f'''(x)| \leq C$ for all $x \in \mathbb{R}$. Via the Taylor series,

$$\left| f(A_i) - f(B_i) - Y_i f'(B_i) - \frac{Y_i^2}{2} f''(B_i) \right| \leq C \frac{|Y_i|^3}{6}. \quad (4.29)$$

Using $|EX| \leq E|X|$, and noting that B_i and Y_i are independent, we get

$$\begin{aligned} \left| Ef(A_i) - Ef(B_i) - E[Y_i f'(B_i)] - E\left[\frac{Y_i^2}{2} f''(B_i)\right] \right| &\leq C \frac{E|Y_i|^3}{6} \\ \implies \left| E \left[f(A_i) - f(B_i) - \frac{f''(B_i)}{2n} \right] \right| &\leq C \frac{E|Y_i|^3}{6}. \end{aligned} \quad (4.30)$$

Similarly, expanding A_{i-1} around B_i gives

$$\left| E \left[f(A_{i-1}) - f(B_i) - \frac{f''(B_i)}{2n} \right] \right| \leq C \frac{E|Z_i|^3}{6}. \quad (4.31)$$

Combining the last two equations, the triangle inequality gives

$$|Ef(A_i) - Ef(A_{i-1})| \leq \frac{C}{6} [E|Y_i|^3 + E|Z_i|^3]. \quad (4.32)$$

We use this result to get

$$|Ef(T_n) - Ef(Z)| = \left| \sum_{i=1}^n Ef(A_i) - Ef(A_{i-1}) \right| \leq \sum_{i=1}^n |Ef(A_i) - Ef(A_{i-1})| \leq \frac{C}{6} \sum_{i=1}^n [E|Y_i|^3 + E|Z_i|^3]. \quad (4.33)$$

Explicitly computing the required expectations gives us

$$E|Y_i|^3 = E \left| \frac{X_1 - \mu}{\sigma\sqrt{n}} \right|^3 = \frac{1}{\sigma^3 n^{3/2}} E|X_1 - \mu|^3, \quad (4.34)$$

$$E|Z_1|^3 = E \left| \frac{Z_1}{1/\sqrt{n}} \cdot \frac{1}{\sqrt{n}} \right|^3 = \frac{1}{n^{3/2}} E|N(0, 1)|^3. \quad (4.35)$$

Thus, the quantity $\frac{C}{6} \sum_{i=1}^n [E|Y_i|^3 + E|Z_i|^3]$ goes to 0, and the quantity $|Ef(T_n) - Ef(Z)|$ is minimized. ■

Definition 4.19. The *moment generating function* of a random variable X is $M_X : \mathbb{R} \rightarrow \mathbb{R}$ defined as $M_X(t) = E[e^{tX}]$ for all $t \in \mathbb{R}$.

If $\{X_n\}$ is a sequence of random variables and X is another random variables with $M_{X_n}(t) \xrightarrow{n \rightarrow \infty} M_X(t)$ for all $t \in \mathbb{R}$, then $X_n \Rightarrow X$. For an alternative proof of the central limit theorem, we can show that

$$E \left[\exp \left(t \cdot \frac{S_n - n\mu}{\sigma\sqrt{n}} \right) \right] \xrightarrow{n \rightarrow \infty} E[e^{tZ}]. \quad (4.36)$$

Appendices

Chapter A

Appendix

Extra content goes here.

Index

t -distribution, 23

almost sure convergence, 31

arcsin law for last visit, 5

arcsin law for maxima, 5

Ballot theorem, 4

Basic lemma, 4

Buffin's needle problem, 20

central limit theorem, 34

Chebyshev's inequality, 1

conditional cumulative distribution function, 23

conditional probability density function, 23

Continuity theorem, 16

continuous random variable, 19

convergence in distribution, 34

convergence in probability, 31

convolution, 11

generating function, 11

Jacobian matrix, 27

joint cumulative density function, 19

joint probability density function, 19

jointly continuous, 19

Markov's inequality, 1

method of images, 3

moment generating function, 36

order statistics, 24

Polya, 16

reversed walk, 4

simple path, 3

simple random walk, 3

strong law of large numbers, 33

Weak law of large numbers, 2

Weirstrass's polynomial approximation, 7