REU Project: Topics in Probability

Trevor Davis

August 14, 2006

0.1 Note on document:

This document is meant for personal use only, much of the material is directly taken from or based on other sources such as S.B.S Varadhan's lecture notes on *Probability Theory*, Sheldon Ross's textbook *A First Course In Probability*, and proofs from my mentor Dani Zarnescu and my fellow students: Robert Prag and Muhammed Waliji.

0.2 Layout of Document

This document is a brief introduction to probability theory. Section 1 introduces the main ideas and theorems of probability and measure theory. Section 2 introduces various concepts of convergence dealing with probability and introduces some of the important limit theorems of probability. And finally in Section 3 we introduce the concept of conditional probability and Markov chains. Spread throughout the document there are examples illustrating many of the concepts mentioned (but in the interest of space not all concepts are illustrated in such a way).

1 Measure Theory

1.1 Measure Theory and Probability

Many undergrads learn a fair amount about probability without ever using measure theory. Why should such a student learn about Measure-Theoretic Probability? The first reason is that the extension of probability to measure-theory is fairly natural and intuitive since probabilities already measure sets. Another reason is that using measure-theory nicely generalizes probability theory. This allows discrete probabilities and continuous probabilities to be treated under the same framework (whereas without measure-theory many concepts and theorems need to be defined and proved twice - once for discrete probabilities and a second time for continous probabilities). A third and final reason is that using measure-theory allows one to use the analytic tools of measure theory.

Definition 1.1 *A class $\mathcal{B}$ of subsets $A \subset \Omega$ is a* field *if the following two properties hold: First, the whole space Ω and the empty set Φ are in $\mathcal{B}$. Second,* *for any two sets A and B in $\mathcal{B}$, the sets $A \cup B$ and $A \cap B$ are again in $\mathcal{B}$. Additionally, $\mathcal{B}$ is considered a σ-*field *if $A_n \in \mathcal{B}$ for every n, then $A = \cup_n A_n \in \mathcal{B}$.*

Definition 1.2 *A* finitely additive probability measure *is a nonnegative set function $P(\cdot)$ defined for sets $A \in \mathcal{B}$ (where $\mathcal{B}$ is a σ-*field*) that satisfies the following properties:*

$$P(A) \geq 0 \ \forall \ A \in \mathcal{B} \tag{1}$$

$$P(\Omega) = 1 \ \textit{and} \ P(\Phi) = 0 \tag{2}$$

If $A \in \mathcal{B}$ and $B \in \mathcal{B}$ are disjoint, then

$$P(A \cup B) = P(A) + P(B) \tag{3}$$

Additionally, $P(\cdot)$ is considered a countably additive probability measure *if for any sequence of pairwise disjoint sets A_n $A = \bigcup_n A_n$ then,*

$$P(A) = \sum_n P(A_n) \tag{4}$$

Definition 1.3 *A probability measure P on a field $\mathcal{F}$ is said to be* countably additive on $\mathcal{F}$ *if for any sequence $A_n \in \mathcal{F}$ with $A_n \downarrow \Phi$, we have $P(A_n) \downarrow 0$.*

Definition 1.4 *For any class $\mathcal{F}$ of subsets of Ω the σ-*field generated by $\mathcal{F}$ *(which we'll often denote $\mathcal{B}$) is the unique smallest σ-field that contains $\mathcal{F}$.*

Theorem 1.1 (Caratheodory Extension Theorem) *Any countably additive probability measure P on a field $\mathcal{F}$ extends uniquely as a countably additive probability measure to the $\sigma-$field $\mathcal{B}$ generated by $\mathcal{F}$.*

Let $\mathcal{I} = \{I_{a,b} : -\infty \leq a < b \leq \infty\}$ where $I_{a,b} = \{x : a < x \leq b\}$ if $b < \infty$, and $I_{a,\infty} = \{x : a < x < \infty\}$. The class of sets that are finite disjoint unions of members of $\mathcal{I}$ is a field $\mathcal{F}$ if the empty set Φ is added to the class. If we are given a function $F(x)$ on the real line which is nondecreasing, continuous from the right and satisfies

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to \infty} F(x) = 1$$

we can define a finitely additive probability measure P by first first defining

$$P(I_{a,b}) = F(b) - F(a)$$

for intervals and then extending it to $\mathcal{F}$ by defining it as the sum for disjoint unions from $\mathcal{I}$.

The Borel σ-field $\mathcal{B}$ on the real line is the σ-field generated by $\mathcal{F}$.

Theorem 1.2 (Lebesgue) *P is countably additive on $\mathcal{F}$ if and only if $F(x)$ is a right continuous function of x. Therefore, for each right continuous nondecreasing function $F(x)$ with $F(-\infty) = 0$ and $F(\infty) = 1$ there is a unique probability measure P on the Borel subsets of the line, such that $F(x) = P(I_{-\infty,x})$. Conversely, every countably additive probability measure P on the Borel subsets of the line comes from some F. The correspondence between P and F is one-to-one.*

The function F is called the distribution function corresponding to the probability measure P.

1.2 Exercise 2.5

Prove that if two distribution functions agree on the set of points at which they are both continuous, they agree everywhere.

Proof: Let F and G be two such distribution functions. Suppose to the contrary, then there exists a point $x_0 \in \mathbb{R}$ s.t. F and G are both discontinuous at x_0 and $F(x_0) \neq G(x_0)$. Since F and G are both distribution functions then they are both right continuous at x_0. Therefore, for *any* sequence of points $\{x_n\}$ ($x_n \neq x_0\ \forall n \in \mathbb{Z}^+$) converging to x_0 from the right then the sequence $\{F(x_n)\}$ converges to $F(x_0)$ and the sequence $\{G(x_n)\}$ converges to $G(x_0)$. Since the functions are monotone and right continous they have only countably many points where they are discontinuous. Hence by ignoring those points for both F and G we can choose a sequence $\{x_n\}$ where $F(x_n)$ and $G(x_n)$ are both continous $\forall n \in \mathbb{Z}^+$ and hence $\forall n \in \mathbb{Z}^+$ then $F(x_n) = G(x_n)$. Then $\{F(x_n)\} = \{G(x_n)\} \longrightarrow F(x_0) = G(x_0)$.
$\Rightarrow\Leftarrow$

Definition 1.5 *A* random variable *or* measurable function *is a map $f : \Omega \to \mathbb{R}$ such that for every Borel set $B \subset \mathbb{R}$, then $f^{-1}(B) = \{w : f(w) \in B\}$ is a measurable subset of Ω. The* expectation *or* mean *of a random variable is defined as $E[X] = \int X(w)\mathrm{d}P$. The* variance *is defined as $Var[X] = E[X^2] - (E[X])^2$.*

A function that is measurable and satisfies $|f(w)| \leq M\ \forall w \in \Omega$ for some finite M is called a *bounded measurable function*.

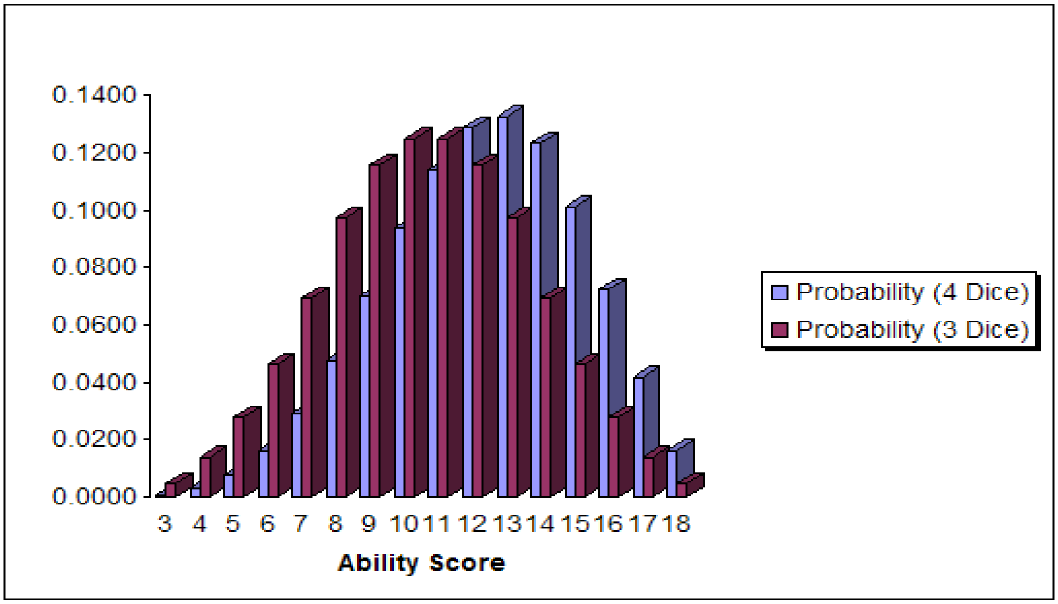
The following statements lay the foundation of an integration theory.

1. If $A \in \Sigma$ (where Σ are the measurable sets of Ω), then the indicator function A:
$$\mathbf{1}_A(w) = \begin{cases} 1 & \text{if } w \in A \\ 0 & \text{if } w \notin A \end{cases}$$
is bounded and measurable.

2. Sums, products, limits, compositions and reasonable elementary operations like min and max performed on measurable functions lead to measurable functions.

Example of concepts: In a popular role-playing game players create characters with various 'ability scores' (determining that characters strength, intelligence and other physical or mental attributes) that take integer values from 3 to 18 (with 10 or 11 considered average) – these numbers will be our **sample space**. One method of determining these 'ability scores' is to roll 4 dice, discard the lowest dice, and use the sum of the three remaining dice. The **probability** of getting each of the possible scores from 3 to 18 using this method is conveniently graphed below and compared to the method of only rolling 3 dice and using the sum. We also find that the **expectation** of the original method is approximately 12.2446 (compared to 10.5 for only using 3 dice). We also give an example of an **distribution function** F(z) for the values of z between the various ability scores x (for completeness F(z) equals zero for all z less than 3 and one for all z greater than 18).

Ability Score (x)	4-Dice Ways to Roll	4-Dice Probability p(x)	4-Dice x*p(x)	4-Dice F(z) $x \le z < (x+1)$
3	1	0.0008	0.0023	0.0008
4	4	0.0031	0.0123	0.0039
5	10	0.0077	0.0386	0.0116
6	21	0.0162	0.0972	0.0278
7	38	0.0293	0.2052	0.0571
8	62	0.0478	0.3827	0.1049
9	91	0.0702	0.6319	0.1752
10	122	0.0941	0.9414	0.2693
11	148	0.1142	1.2562	0.3835
12	167	0.1289	1.5463	0.5123
13	172	0.1327	1.7253	0.6451
14	160	0.1235	1.7284	0.7685
15	131	0.1011	1.5162	0.8696
16	94	0.0725	1.1605	0.9421
17	54	0.0417	0.7083	0.9838
18	21	0.0162	0.2917	1.0000
Total	1296	1.0000	**12.2446**	



3. If $\{A_j : 1 \le j \le n\}$ is a finite disjoint partition of Ω into measurable sets, the function $f(w) = \sum_j c_j \mathbf{1}_{A_j}(w)$ is a measurable function and is referred to as a *simple* function.

4. Any bounded measurable function f is a uniform limit of simple functions.

5. For simple functions $f = \sum c_j \mathbf{1}_{A_j}$ the integral $\int f(w) \mathrm{d}P$ is defined to be $\sum_j c_j P(A_j)$. It enjoys the following properties:

   (a) If f and g are simple, so is any linear combination $af + bg$ for real constants a and b and $\int (af + bg) \mathrm{d}P = a \int f \mathrm{d}P + b \int g \mathrm{d}P$.

   (b) If f is simple so is $|f|$ and $|\int f \mathrm{d}P| \le \int |f| \mathrm{d}P \le \sup_w |f(w)|$

6. If $\{f_n\}$ is a sequence of simple functions converging to f uniformly, then $\{a_n\} = \int f_n \mathrm{d}P$ is a Cauchy sequence of real numbers and therefore has a limit a as $n \to \infty$. The integral $\int f \mathrm{d}P$ of f is defined to be this limit a. It can be shown that choosing a different approximating sequence $\{g_n\}$ for f leads to the same $a = \int f \mathrm{d}P$.

7. Now the integral is defined for all bounded measurable functions and enjoys the following properties:

   (a) If f and g are bounded measurable functions and a, b are real constants then the linear combination $af + bg$ is again a bounded measurable function, and
   $$\int (af + bg) \mathrm{d}P = a \int f \mathrm{d}P + b \int g \mathrm{d}P$$

   (b) If f is a bounded measurable functions so is $|f|$ and $|\int f \mathrm{d}P| \le \int |f| \mathrm{d}P \le \sup_w |f(w)|$.

   (c) Also, for any bounded measurable f,
   $$\int |f| \mathrm{d}P \le P(\{w : |f(w)| > 0\}) \sup_w |f(w)|$$

   (d) If f is a bounded measurable function and A is a measurable set then we define
   $$\int_A f(w) \mathrm{d}P = \mathbf{1}_A(w) f(w) \mathrm{d}P$$
   and we can write for any measurable set A,
   $$\int f \mathrm{d}P = \int_A f \mathrm{d}P + \int_{A^c} f \mathrm{d}P$$

Definition 1.6 *A sequence* $\{f_n\}$ *is said to converge to a function* f everywhere *or* pointwise *if*
$$\lim_{n \to \infty} f_n(w) = f(w)$$
for every $w \in \Omega$.

Definition 1.7 *A sequence $\{f_n\}$ of measurable functions is said to converge to a measurable function f* almost everywhere *or* almost surely *if there exists a measurable set N with $P(N) = 0$ such that*

$$\lim_{n \to \infty} f_n(w) = f(w)$$

for every $w \in N^c$

Definition 1.8 *A sequence $\{f_n\}$ of measurable functions is said to converge to a measurable function f* in measure *or* in probability *if*

$$\lim_{n \to \infty} P[w : |f_n(w) - f(w)| \geq \epsilon] = 0$$

for every $\epsilon > 0$.

Lemma 1.1 *If $\{f_n\} \to f$ almost everywhere then $\{f_n\} \to f$ in measure.*

1.3 Exercise 1.12

If $\{f_n\} \to f$ in measure it is not necessarily true that $\{f_n\} \to f$ almost everywhere:

Counterexample: Consider the interval [0,1] and divide it successively into 2,3,4 $\cdots$ parts and enumerate the intervals in succession. That is, $I_1 = [0, \frac{1}{2}], I_2 = [\frac{1}{2}, 1], I_3 = [0, \frac{1}{3}], I_4 = [\frac{1}{3}, \frac{2}{3}], I_5 = [\frac{2}{3}, 1]$, and so on. If $f_n(x) = \mathbf{1}_{I_n}(x)$ then $\{f_n\}$ tends to 0 in measure but not almost everywhere.

However the following statement is true:
If $\{f_n\} \to f$ as $n \to \infty$ in measure, then there is a subsequence $\{f_{n_j}\}$ such that $\{f_{n_j}\} \to f$ almost everywhere as $j \to \infty$.

Example: Using the previous counterexample, take the subsequence of intervals of the form $I_{n_j} = [0, \frac{1}{j+1}]$ $\forall j \in \mathbb{Z}^+$ and let $N = 0$ (which clearly has measure 0 since it is a single point). Then if $f_n(x) = \mathbf{1}_{I_n}(x)$ then $\{f_{n_j}\}$ tends to 0 $\forall x \notin N$.

Proof of claim: Since $\{f_n\} \to f$ in measure then

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} P(w : |f_n(w) - f(w)| \geq \epsilon) = 0$$

then we can choose a subsequence $\{f_{n_j}\}$ such that the following holds $\forall j \in \mathbb{Z}^+$:

$$P(w : |f_{n_j}(w) - f(w)| \geq 2^{-j}) < 2^{-j}$$

Let $E_j = \{x : |f_{n_j}(w) - f(w)| \geq 2^{-j}\}$, then it is clear that,

$$\forall w \notin \bigcup_{j\geq k} E_j \Rightarrow |f_{n_j} - f(w)| \leq 2^{-k}$$

And,

$$\forall w \notin \bigcap_{k} \bigcup_{j\geq k} \Rightarrow f_{n_j}(w) \to f(w)$$

Also,

$$\lim_{k\to\infty} P(\bigcup_{j\geq k} E_j) = 0 \text{ since } P(\bigcup_{j\geq k} E_j) \leq \sum_{j\geq k} P(E_j) \leq \sum_{j\geq k} \frac{1}{2^j}$$

Hence $\{f_{n_j}\}$ converges to an f for all x outside a set of measure zero.

2 Weak Convergence

Definition 2.1 *If α is a probability distribution on the line, its characteristic function is defined by*

$$\phi(t) = \int \exp[itx] \mathrm{d}\alpha$$

Definition 2.2 *A sequence $\{\alpha_n[I]\}$ of probability distributions on $\mathbb{R}$ is said to converge weakly to a probability distribution α ($\{\alpha_n\} \Rightarrow \alpha$) if,*

$$\lim_{n\to\infty} \alpha_n[I] = \alpha[I]$$

for any interval $I = [a, b]$ such that the single point sets a and b have probability 0 under α.

Definition 2.3 *A sequence $\{\alpha_n\}$ of probability measures on the real line $\mathbb{R}$ with distribution functions $F_n(x)$ is said to converge weakly to a limiting probability measure α with distribution function $F(x)(\{F_n\} \Rightarrow F)$ if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for every x that is a continuity point of F.

Theorem 2.1 (Lévy-Cramér continuity theorem) *The following are equivalent:*

1. *$\{\alpha_n\} \Rightarrow \alpha$ or $\{F_n\} \Rightarrow F$*
2. *For every bounded continuous function f(x) on $\mathbb{R}$*

$$\lim_{n\to\infty} \int_{\mathbf{R}} f(x) \mathrm{d}\alpha_n = \int_{\mathbb{R}} f(x) \mathrm{d}\alpha$$

3. *If $\phi_n(t)$ and $\phi(t)$ are respectively the characteristic functions of α_n and α, for every real t,*

$$\lim_{n\to\infty} \phi_n(t) = \phi(t)$$

Definition 2.4 *Two events A and B are said to be independent if $P[A \cap B] = P[A]P[B]$.*

Example: Let us roll two dice. Let A be the event that the first roll is a 1 and B be the event that the sum of the two dice is a 2. Then $P[A \cap B] = P[B] = \frac{1}{36} \neq \frac{1}{216} = \frac{1}{6}\frac{1}{36} = P[A]P[B]$ and hence A and B are not independent. However, suppose instead that B is the event that the sum of the two dice is 7. Then $P[A \cap B] = \frac{1}{36} = \frac{1}{6}\frac{1}{6} = P[A]P[B]$ and now A and B are independent.

Definition 2.5 *Two random variables X and Y are independent if the events $X \in A$ and $Y \in B$ are independent for any two Borel sets A and B on the line i.e.*

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B]$$

for all Borel sets A and B.

Definition 2.6 *A finite collection $\{X_j : 1 \leq j \leq n\}$ of random variables are said to be independent if for any n Borel sets $A_1, \ldots, A_n$ on the line*

$$P[\cap_{1 \leq j \leq n}[X_j \in A_j]] = \prod_{1 \leq j \leq n} P[X_j \in A_j]$$

Theorem 2.2 (Weak Law of Large Numbers) *If $X_1, X_2, \ldots, X_n$ are independent and identically distributed with a finite first moment and $E(X_i) = m < \infty$, then $\frac{X_1+X_2+\cdots+X_n}{n}$ converges to m in probability as $n \to \infty$.*

Theorem 2.3 (Strong Law of Large Numbers) *If $X_1, X_2, \ldots, X_n$ are independent and identically distributed with $E|X_i|^4 = C < \infty$, then $\frac{X_1+X_2+\cdots+X_n}{n}$ converges to $E[X_1]$ almost everywhere as $n \to \infty$.*

Theorem 2.4 (Lévy's Theorem) . *If $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of independent random variables, then the following are equivalent:*

1. *The distribution α_n of $S_n = X_1 + \cdots + X_n$ converges weakly to a probability distribution α on $\mathbb{R}$.*
2. *The random variable S_n converges in probability to a limit $S(w)$.*
3. *The random variable S_n converges with probability 1 to a limit $S(w)$.*

Let us look at the space Ω of real sequences $\{x_n : n \geq 1\}$. There is $\mathcal{B}$, the product σ-field on Ω. In addition there are the sub σ-fields $\mathcal{B}^n$ generated by $x_j : j \geq n$. $\mathcal{B}^n$ are $\downarrow$ with n and $\mathcal{B}^\infty = \cap_n \mathcal{B}^n$ (which is also a σ-field) is called the tail σ-field. The typical set in $\mathcal{B}^\infty$ is a set depending only on the tail behavior of the sequence.

Theorem 2.5 (Kolmogorov's Zero-One Law) *If $A \in \mathcal{B}^\infty$ and P is any product measure then $P(A) = 0$ or 1.*

Corollary 2.1 *Any random variable measurable with respect to the tail σ-field $\mathcal{B}^\infty$ is equal with probability 1 to a constant relative to any given product measure.*

2.1 Exercise 3.16

How can different product measures cause these constants to be different?
It is clear that if X is a random variable measurable with respect to the tail σ-field $\mathcal{B}^\infty$ is equal with probability 1 to a constant m relative to a given product measure P then $E[X] = m$. But $E[X] = \int X(w)\mathrm{d}P$. Clearly this expectation depends on the product measure P and hence the constants m can be different depending on the product measure.

Theorem 2.6 (Central Limit Theorem) *Let $X_1, \ldots, X_n, \ldots$ be a sequence of independent identically distributed random variables with $E[X_i] = 0$ and $0 < Var[X] = \sigma^2 < \infty$. Then the distribution of $\frac{X_1 + \cdots + X_n}{n}$ converges as $n \to \infty$ to the normal distribution with density*

$$p(x) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp[-\frac{x^2}{2\sigma^2}]$$

3 Dependent Random Variables

Definition 3.1 Conditional probability *is defined to be*

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

if $(\Omega, \mathcal{F}, P)$ is a probability space, $A \in \mathcal{F}$ is a set of positive measure, and if B is an arbitrary set in $\mathcal{F}$

If Ω is partitioned into a finite or countable number of disjoint measurable sets $A_1, \cdots, A_j, \cdots$ then

$$P(B) = \sum_j P(A_j)P(B|A_j)$$

.

Example: Suppose zombies attack the University of Chicago and start to bite students who eventually (if not eaten entirely) transform into new zombies after a short period of time. It is well known that such an infected student will start to drool from his mouth before completely turning into a zombie with a probability of 0.90 whereas the normal University of Chicago student (excluding Computer Science majors) only drools from his mouth with a probability of 0.05. If we notice that our fellow student John Gauss (who is not a Computer Science major) is drooling and we previously had an a priori belief that he has been bitten by a zombie (and hence infected) of 0.4 (he has bandages on his arm) then what is the conditional probability that he is infected given that he is drooling?

Answer: Given that John Gauss is drooling there is about a 92.3% chance that he has been infected, you should consider killing him with an Ax(iom) before he bites you and turns you into a zombie!

Let B = the event that John has been bitten and D = the event that John drools.

$$P(B|D) = \frac{P(B \cap D)}{P(D)} = \frac{P(B)P(D|B)}{P(D)} = \frac{P(B)P(D|B)}{P(B)P(D|B) + P(B^C)P(D|B^C)}$$

$$= \frac{(0.4)(0.90)}{(0.4)(0.90) + (0.6)(0.05)} \approx 0.923$$

Consider a sequence of random variables $X_0, X_1, \dots$ and suppose the set of possible values of these random variables is $\{0, 1, \dots, M\}$ (this can be relaxed further - i.e. to $\mathbf{Z}$. We say that the system is in state i at time n if $X_n = i$. The sequence of random variables is said to form a *Markov chain* if each time the system is in state i there is some fixed probability (which I will denote P_{ij}) that it will next be in state j. That is, for all $i_0, \dots, i_{n-1}, i, j$,

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} = P_{ij}$$

The values $P_{ij}, 0 \leq i \leq M, 0 \leq j \leq N$, are called the *transition probabilities* of the Markov chain. For each $i = 0, 1, \dots, M$ then $P_{ij} \geq 0$ and $\sum_{j=0}^{M} P_{ij} = 1$.

It is possible to arrange the transition probabilities P_{ij} as follows:

$$\begin{vmatrix} P_{00} & P_{01} & \cdots & P_{0M} \\ P_{10} & P_{11} & \cdots & P_{1M} \\ \vdots & \vdots & \ddots & \vdots \\ P_{M0} & P_{M1} & \dots & P_{MM} \end{vmatrix}$$

Knowledge of the transition probability matrix and the distribution of X_0 allows us to compute many probabilities of interest. For instance, the joint probability mass function of $X_0, \dots, X_n$ is given by

$$\begin{aligned} & P\{X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} \\ &= P\{X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0\} P\{X_{n-1} = i_{n-1}, \dots, X_0 = i_0\} \\ &= P_{i_{n-1}, i_n} P\{X_{n-1} = i_{n-1}, \dots, X_0 = i_0\} \\ & \vdots \\ &= P_{i_{n-1}, i_n} P_{i_{n-2}, i_{n-1}} \cdots P_{i_1, i_2} P_{i_0, i_1} P\{X_0 = i_0\} \end{aligned}$$

Example:

Suppose that whether the stockmarket rises or falls tomorrow depends on previous stockmarket movements only through whether or not it fell or rose today. Suppose further that if it rose today, then it will rise tomorrow with probability α, and if it fell today then it will rise tomorrow with probability β.

If we say that the system is in state 0 when the market rises and state 1 when it does not, then the preceding system is a two-state Markov chain having transition probability matrix:

$$\begin{vmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{vmatrix}$$

That is, $P_{00} = \alpha = 1 - P_{01}, P_{10} = \beta 1 - P_{11}$.

We can also define the n-stage transition probability, $P_{ij}^{(n)}$, that a system presently in state i will be in state j after n additional transitions.

Theorem 3.1 (The Chapman-Kolmogorov equations)

$$P_{ij}^{(n)} = \sum_{k=0}^{M} P_{ik}^{(r)} P_{kj}^{(n-r)} \quad \forall\ 0 < r < n$$

Although the $P_{ij}^{(n)}$ denote conditional probabilities we can use them to derive expressions for unconditional probabilities:

$$P\{X_n = j\} = \sum_i P\{X_n = j | X_0 = i\} P\{X_0 = i\}$$

$$= \sum_i P_{ij}^{(n)} P\{X_0 = i\}$$

It turns out that if $P_{ij}^{(n)} > 0 \quad \forall\ i, j = 0, 1, \ldots, M$ for some $n > 0$ (such a Markov chain is said to be *ergodic*) then $P_{ij}^{(n)}$ converges as $n \to \infty$ to a value π_j that depends only on j. That is, for large values of n, the probability of being in state j after n transitions is approximately equal to π_j no matter what the initial state was.

Theorem 3.2 *For an ergodic Markov chain*

$$\pi_j = \lim_{n \to \infty} P_{ij}^{(n)}$$

exists, and the π_j, $0 \le j \le M$, are the unique nonnegative solutions of

$$\pi_j = \sum_{k=0}^{M} \pi_k P_{kj}$$

$$\sum_{j=0}^{M} \pi_j = 1$$

Consider our example on stockmarkets, from the previous theorem it follows that the limiting probabilities of the stockmarket rising or falling (π_0 and π_1), are given by

$$\begin{aligned}
\pi_0 &= \alpha\pi_0 + \beta\pi_1 && (5)\\
\pi_1 &= (1-\alpha)\pi_0 + (1-\beta)\pi_1 && (6)\\
\pi_0 + \pi_1 &= 1 && (7)
\end{aligned}$$

which yields:

$$\pi_0 = \frac{\beta}{1+\beta-\alpha} \quad \pi_1 = \frac{1-\alpha}{1+\beta-\alpha}$$

For instance, if $\alpha = 0.7$, $\beta = 0.3$, then the limiting probability of the stock market rising on the nth day is $\pi_0 = \frac{1}{2}$.

3.1 Note on document:

This document is meant for personal use only, much of the material is directly taken from or based on other sources such as S.B.S Varadhan's lecture notes on *Probability Theory*, Sheldon Ross's textbook *A First Course In Probability*, and proofs from my mentor Dani Zarnescu and my fellow students: Robert Prag and Muhammed Waliji.