Random Processes in Systems - Lecture Notes

Jean Walrand with Antonis Dimakis
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley CA 94720

August 2006

Acknowledgment

I wrote these notes as I was teaching the course in Fall 2005 to a class of bright and inquisitive students at Berkeley. I had the great luck of having Antonis Dimakis as a teaching assistant for the class. Antonis has an unusually deep understanding of probability and a knack for creating examples that illustrate the main points particularly crisply. The students and Antonis have my gratitude for their inputs. In particular, I want to single out Ming-Yang Chen for his meticulous reading and correcting of the notes.

As we use these notes in subsequent versions of the course, I hope the students will be kind enough to point out errors and needed clarifications of the material. In fact, I know they will....

Contents

Chapter 1

Introduction

This course explains models of randomness in systems. Specifically, we study detection, estimation, Markov chains, Poisson processes, and renewal processes. We choose these particular topics because of their many applications in communications, signal processing, control, and more general decision making under uncertainty.

We assume that you have taken an undergraduate course in Probability and that you have a good understanding of Linear Algebra. To help you, we include one appendix on each of these topics. We also review them briefly in the first lectures.

For convenience of the students and of the instructor, each chapter corresponds approximately to one eighty-minute lecture. Together, the notes cover a one-semester course. We deliberately left some redundancy in the notes that corresponds to repetitions in lectures. Similarly, some details are left for students to fill in. We believe that students benefit by doing some of the work instead of being presented with complete derivations.

The difficulty that we always face when teaching this course is that it covers two sets of topics usually not well represented in a single textbook. The first set of topics concerns mostly detection and estimation, including Wiener and Kalman filters. We deal with second-order properties of random variables and with linear systems. The main idea here is projection in a space of square-integrable random variables. The tools are algebraic. These topics are traditional in electrical engineering because of their many applications in communication theory and signal processing.

The second set of topics concerns Markov chains and renewal processes. Here the tools are mostly probabilistic, such as properties of individual trajectories and coupling arguments. These topics are important in computer science and in operations research. For instance, they are essential to queuing theory, stochastic scheduling, searching algorithms, learning models, and Markov decision theory.

We regularly use two different textbooks to cover the material because we have yet to find one that does a good job of presenting both. These notes are not meant to be complete, so we refer regularly to textbooks.

In the rest of this introduction we outline some of the key ideas of the course.

1.1 Detection

The *detection problem* is to guess the value of some random variable X in a finite set given some observation Y whose distribution depends on X. For instance, Y is the output of some sensor and X is either 1 or 0 to represent the occurrence of a house fire or its non-occurrence,

respectively. As another example, $X \in \{0, 1, 2, 3\}$ is the symbol that a transmitter sends and Y is the signal that a receiver gets. As yet another example, $X = 0$ indicates that some machine is working correctly and $X = 1$ that it is somehow defective; Y is a set of measurements that one performs on the output of the machine. In a medical setting, $X = 1$ could indicate that a given patient has a specific form of cancer and $X = 0$ that she does not; in that context, Y could be the output of a cat scan or some biological measurement. As a more complex example in speech recognition, X represents the sequence of syllables that someone pronounces and Y is the measurements made after processing the sound that a microphone captures.

As these examples suggest, many practical situations correspond to detection problems. The mathematical formulation of such a problem specifies the conditional distribution of Y given X. This conditional distribution is a model of the "observation channel." It specifies how the observation relates to the value X to be detected. There are two essentially different formulations of the problem that differ in what we assume known about X.

In the *Bayesian* formulation , the prior distribution of X is known. That is, we know the probability that X takes any specific value. Since we know also the conditional distribution of Y given X, we know the joint distribution of X and Y. Using this joint distribution, we can calculate the conditional distribution of X given Y. From this conditional distribution, we can determine how to calculate the value of $\hat{X}$ based on Y that minimizes some average cost $E(c(X, \hat{X}))$. For example, one may want to minimize the probability of guessing a wrong value, which corresponds to $c(X, \hat{X})$ being equal to one when $X \neq \hat{X}$ and to zero otherwise. In a communication application, this problem corresponds to minimizing the error rate of the receiver.

In the *non-Bayesian* formulation, we do not know the prior distribution of X. Such a formulation is motivated by many applications where that distribution is difficult to guess or where the system must operate satisfactorily under a wide range of possible prior distributions of X. For instance, what is the probability that your house will be on fire at noon next Tuesday? Clearly, if the design of your fire alarm relies on knowing that probability accurately, it is not likely to work satisfactorily. Similar considerations apply to medical tests and many other situations. One standard formulation, when $X \in \{0, 1\}$, is to minimize the probability of guessing $X = 0$ when in fact $X = 1$ subject to an upper bound on the probability of guessing $X = 1$ when in fact $X = 0$. In the medical setting, the first error is a "false negative" and the second is a "false positive." As you expect, there is a tradeoff between the probabilities of these two types of error. Thus, one may want to limit the rate of false positive to 5%, say, and then design the most sensitive test that minimizes the probability of false negative subject to that constraint. The acceptable probability of false positive depends on the context. For instance, when screening blood donations for HIV, one may accept a larger probability of false positive than when testing people.

1.2 Estimation

The *estimation problem* is to guess the value of X, in some subset of the real line or of some higher-dimensional space, given some observation Y whose distribution depends on X. This problem is similar to the detection problem; the key difference being that here X takes values in a continuous set instead of a finite one.

As an example, X could be the amplitude of a signal that a transmitter sends and Y the measured signal amplitude at a wireless receiver. In that situation, Y differs from X because of attenuation, noise, and various multi-path effects. As another example, X is the average

power of the noise at a receiver and Y is a set of transmitted and corresponding received signal amplitudes. As a prediction example, X is the snow depth at the base of Sugar Bowl next January 15 and Y is the set of available information from weather reports to date.

Such estimation problems occur in many fields of application. For the most part we focus on Bayesian formulations where the prior distribution of X is known. We study static and dynamic formulations. In a static formulation, the joint distribution of X and Y is specified and we estimate X based on Y. One issue is whether we can perform arbitrary calculations based on Y to estimate X or whether we are restricted to linear operations. In many applications, one considers the latter case and one derives the *linear least squares estimator* of X given Y. The restriction to linear estimators is made because it greatly simplifies the calculations. Moreover, the best linear estimator depends only on second-order statistics whereas the best nonlinear estimator generally depends on higher-order statistics that may be difficult to obtain.

In the dynamic formulation, both X and Y may change over time. For instance, say that X_n is the location at time n of some automobile. Y_n is the output of a GPS sensor at time n. One wishes to estimate X_n given all the values of Y_m for $m \leq n$. Here, the key idea is to derive a *recursive* estimator where the estimate at time n depends only on Y_n and on the estimate at time $n-1$. Thus, a recursive estimator *updates* the estimate based on the latest measurements instead of performing a new calculation every time based on all the accumulated measurements. The Kalman Filter is such a recursive estimator for linear systems.

1.3 Gaussian Random Variables

Gaussian random variables play a special role in our discussion of detection and estimation. This special role has three principal causes. The first is that Gaussian random variables occur naturally as sums of many small almost independent random variables. Thus, the thermal noise in a conductor tends to be Gaussian because it is the sum of contributions of many electrons whose motions are almost independent. Similarly, noise due to electromagnetic interference tends to be Gaussian.

A second cause for the importance of Gaussian random variables is that a linear combination of independent Gaussian random variables is again Gaussian. This fact, obviously consistent with the first cause, simplifies greatly the analysis of linear systems with Gaussian noise. We call *jointly Gaussian* random variables that are linear combinations of independent Gaussian random variables.

The third cause is that the joint distribution of jointly Gaussian random variables is completely specified by second order statistics. In particular, jointly Gaussian random variables that are uncorrelated are independent. This property reduces the analysis of Gaussian random variables to simple algebraic calculations.

1.4 Markov Chains and Renewal Processes

A *Markov chain* models the random evolution of some object in a discrete set $\mathcal{X}$. For instance, one can approximate the number of telephone calls in progress in a given office building by a Markov chain. The defining property of a Markov chain is that its evolution starts afresh from its value at any given time. Thus, in the telephone example, the evolution of the number of telephone calls after 1:00 pm depends obviously on the number, say X_0 of calls in progress at that time. However, given X_0, the values of that number of calls after 1:00 pm are independent of the values before 1:00 pm. (This statement is only approximately true, but the example

conveys the main idea.) As another example, imagine a flea that jumps on a chessboard and assume that the flea jumps randomly at each step, without remembering its previous jumps. The position of that flea at successive steps is a Markov chain.

Designate by $X_t \in \mathcal{X}$ the location of the object at time t. We call X_t the *state* of the Markov chain at time t. The questions of interest concern the fraction of time that the state has a particular value and the probability that the state has a given value at some time t. A related question is the average time it takes for the state to reach a given set of values. For instance, imagine that X_t represent the number of packets stored in a given Internet router, assuming that this can be modeled by a Markov chain. If the router transmits packets at a constant rate, there is a direct relationship between X_t and the delay of the packets through the router. Similarly, knowing the fraction of time that the router is full tells us the likelihood that the router must drop incoming packets. Accordingly, finding out the statistics of X_t is of direct relevance to estimating the performance of the router.

Think of the process of replacing a specific light bulb whenever it burns out. Assume for simplicity that the bulb is always on until it burns out and is replaced. The lifetimes of the successive bulbs are independent and identically distributed. Designate by N_t the number of bulbs that one has to use up to time t, for $t \geq 0$. Under these assumptions, N_t is a *renewal process* . We focus on two related questions. The first one is how long one can expect to wait until the bulb burns out. To be more precise, choose a large time t and let $\tau(t)$ be the random time one has to wait after time t until the bulb burns out. We explain that if the lifetime distribution is not concentrated on multiples of some constant, then the distribution of $\tau(t)$ converges as t increases. Interestingly, the mean value of $\tau(t)$ converges to a value that is larger than half the lifetime of a bulb. The second question concerns the expected number of bulbs one needs between time t and $t + L$ for $t, L > 0$. We explain that, under the same assumption as before, this expected value converges to λL as t increases. Here, λ is the reciprocal of the average lifetime of a bulb. Thus, λ is the average replacement rate of the bulbs.

Chapter 2

Mathematical Preliminaries

To make sure we all have the necessary background we start by reviewing some concepts from Set Theory, Real Numbers, Probability, and Linear Algebra.

2.1 Summary

Here are the main points reviewed in this chapter:

- Set, set operations, function
- Inverse image of a set; commutes with set operations
- Countable set; real numbers are not countable; a countable union of countable sets is countable
- Convergence of sets
- Convergence of real numbers; of a nondecreasing upper-bounded sequence; of a Cauchy sequence
- Definition of random variable and counterexample
- Almost sure convergence of random variables does not imply that of expectations
- Using the Borel-Cantelli Lemma to prove a.s. convergence
- Projection property of conditional expectation
- Using elementary row operations to solve $Ax = 0$
- Basis and dimension of a linear subspace
- A Hermitian matrix can be diagonalized by a unitary matrix

2.2 Set Theory

You recall that a set is a well-defined collection of objects. That is, a set is defined if one can determine whether any given object belongs to the set or not. Hence, the notions of element and subset of a set. One then defines basic set operations such as union, intersection, complement, and difference. The next concept is that of function f defined from a set Ω into another set S: the function assigns an element $f(\omega)$ of S to each element ω of Ω. We designate such a function by $f : \Omega \to S$.

If $f : \Omega \to S$ and $B \subset S$, then $f^{-1}(B) := \{\omega \in \Omega \mid f(\omega) \in B\}$ is the *inverse image* of B under $f(\cdot)$. You can show that the inverse image commutes with set operations. For instance, $f^{-1}(\cap_{i\in I} B_i) = \cap_{i\in I} f^{-1}(B_i)$ and $f^{-1}((A \setminus B) \cup D^c) = (f^{-1}(A) \setminus f^{-1}(B)) \cup (f^{-1}(D))^c$.

A set Ω is *countable* if one can enumerate its elements as $\Omega = \{\omega_1, \omega_2, \ldots\}$. The rational numbers are countable, the real numbers are not. A countable union of countable sets is countable.

For $n \geq 1$, let $A_n \subset \Omega$ and let also $A \subset \Omega$. We write $A_n \downarrow A$ if $A_{n+1} \subset A_n, \forall n \geq 1$ and if $\cap_{n=1}^{\infty} A_n = A$. We define $A_n \uparrow A$ similarly. More generally, we say that the sets A_n *converge to* A and we write $A_n \to A$ if $\cup_{m=n}^{\infty} A_m \downarrow A$ and $\cup_{m=1}^{n} \uparrow A$.

2.3 Real Numbers

Let $S \subset \Re$. We say that $x \in \Re$ is a *lower bound* of S if $x \leq y$ for all $y \in S$. If $x \in S$ is a lower bound of S, we write $x = \min S$ and we say that x is the *minimum* of S. The *completeness axiom* of real numbers states that if a set S of real number has a lower bound x, then it has a greatest lower bound y and we write $y = \inf S$ and we say that y is the *infimum* or the *greatest lower bound* of S. We have the similar definitions for the *maximum* $\max S$ and the *supremum* $\sup S$ or *lowest upper bound*. Thus, $0 = \min[0, 1] = \inf(0, 1]$ but $(0, 1]$ does not have a minimum element.

Let $x \in \Re$ and $x_n \in \Re$ for $n \geq 1$. We say that x_n *converges to* x and we write $x_n \to x$ if for all $\epsilon > 0$ there is some $n(\epsilon)$ such that $|x_n - x| \leq \epsilon, \forall n \geq n(\epsilon)$. We also write $x_n \to \infty$ if for all $c \in \Re$ these is some $n(c)$ such that $x_n \geq c$ for all $n \geq n(c)$.

An important fact is that a nondecreasing sequence that is bounded from above must converge.

The sequence $\{x_n, n \geq 1\}$ is *Cauchy* if $\sup_{k,m\geq n} |x_k - x_m| \to 0$ as $n \to \infty$. One has $x_n \to x$ with x finite if and only if the sequence is Cauchy. For a proof, see Lemma 21.1

2.4 Probability

We review in Appendix A the basic notions of Probability Theory. Here we highlight a few ideas.

Let $\Omega = \{1, 2, 3, 4\}$ and assume that each element has probability $1/4$. Consider the events $A = \{1, 2\}, B = \{1, 3\}, C = \{1, 4\}$. These events are pairwise independent but not mutually independent. Indeed, $P(A \cap B) = P(A)P(B)$ and similarly for the other pairs of sets, but $P(A \cap B \cap C) \neq P(A)P(B)P(C)$. Thus, $P[A \mid B \cap C] \neq P(A)$. The point of this example is that one should not confuse the statements 'A and B are independent' and 'knowing that B occurs does not tell us anything about how A occurs.'

Consider a probability space $\{\Omega, \mathcal{F}, P\}$ with $\Omega = [0, 1], \mathcal{F} = \{[0, 0.5], (0.5, 1], \emptyset, \Omega\}, P([0, 0.5]) = 0.3$. Let also $X : \Omega \to \Re$ be defined by $X(\omega) = \omega$. Then X is not a random variable

$X^{-1}([0,0.2]) = [0,0.2] \notin \mathcal{F}$. In particular, the probability that $X \leq 0.2$ is not defined. What is going on here is that X is not $\mathcal{F}$-measurable: the probability space $\{\Omega, \mathcal{F}, P\}$ is not 'rich' enough to 'measure' X.

Let $\{\Omega, \mathcal{F}, P\}$ be $[0,1]$ with the uniform probability. For $n \geq 1$, let $X_n(\omega) = n1\{\omega \leq 1/n\}$. Note that $X_n(\omega) \to 0$ as $n \to \infty$, for all $\omega \neq 0$. Thus, $P(X_n \to 0) = 1$ and we write $X_n \xrightarrow{a.s.} 0$ and we say that X_n converges to 0 *almost surely*. Note however that $E(X_n) = 1 \nrightarrow E(0) = 0$. See Lebesgue's theorem in Appendix A.

Recall Borel-Cantelli and assume that $E(X_n^2) \leq \alpha/n^2$. Fix any $\epsilon > 0$ and note that $\sum_{n=1}^{\infty} P(|X_n| > \epsilon) \leq \sum_{n=1}^{\infty} E(X_n^2)/\epsilon^2 < \infty$, so that $P(|X_n| > \epsilon, \text{ i.o.}\} = 0$. This shows that $X_n \xrightarrow{a.s} 0$.

Recall also that $E[X|Y]$ is some function $g(Y)$ with the property that $E((X - g(Y))h(Y)) = 0$ for all $h(\cdot)$. That is, $X - E[X|Y] \perp \{h(Y), h(\cdot) \text{ is a function }\}$. Thus, $E[X|Y]$ is the projection of X onto the subspace of function of Y. Also, the projection of a point onto an hyperplane is the closest element of the hyperplane to that point. Thus, $E[X|Y]$ is the function $g(Y)$ of Y that minimizes $E((X - g(Y))^2)$. These facts are the *projection property* and the *minimum mean squares estimator property* of conditional expectation.

2.5 Linear Algebra

Appendix B reviews the main ideas of Linear Algebra. Here are some points that we use frequently.

The elementary row operations (eros) consist of either adding to a row of a matrix a multiple of another row or of interchanging two rows. Consider then the homogeneous equations $Ax = 0$ where $A \in \Re^{m \times n}$ and $x \in \Re^n$. By performing eros, we can bring the matrix A to a row echelon form where row k has the form $[0, \ldots, 0, *, \ldots, *]$ and its first nonzero term, called its pivot, corresponds to column $n(k)$ where $n(k)$ is strictly increasing. The eros do not change the solutions of $Ax = 0$. Moreover, if A has the row echelon form, then we can solve the equations bottom up. Consider the last row L. Each term $m > n(L)$ corresponds to an element of x that we can set arbitrarily. The pivot then corresponds to an element of x uniquely determined from the previously selected values. Continuing in this way, we find that the free variables of x are the non-pivot elements. This argument shows that if $n > m$, then there must be non-pivot elements and there are infinitely many solutions to $Ax = 0$. See Problem 2.6 for an illustration.

A *basis* of some linear subspace $\mathcal{V}$ is a collection of linearly independent vectors of that space such that each element can be written as a linear combination of the basis vectors. Assume that there is a basis $A = [a_1|\cdots|a_m]$. Then any collection $\{v_1, \ldots, v_n\} \subset \mathcal{V}$ with $n > m$ must be linearly dependent. Indeed, $v_i = Aw_i$ for $i = 1, \ldots, n$. Hence, $0 = \sum_{i=1}^n x_i v_i = \sum_{i=1}^n x_i A w_i = A(\sum_{i=1}^n x_i w_i)$ is equivalent to $0 = \sum_{i=1}^n x_i w_i$, which is a system of m equations with $n > m$ variables, and we know from the previous discussion that it has infinitely many solutions. It follows that all bases must have the same number of elements, which is called the *dimension* of the subspace.

The following result is important in the study of Gaussian random variables.

Theorem 2.1. *Diagonalization of Hermitian Matrix*

Let $H \in \mathbb{C}^{n \times n}$ be a Hermitian matrix, *i.e., such that $H = H^*$.*

The eigenvalues $\lambda_1, \ldots, \lambda_n$ of H are real (they are not necessarily distinct);

H has n orthonormal eigenvectors $\{u_1, \ldots, u_n\}$ that form a basis for $\mathbb{C}^n$. That is, $u_i^* u_j = 1\{i = j\}$.

If $P = [u_1|\dots|u_n]$, *then*

$$P^*HP = \Lambda = diag(\lambda_1, \dots, \lambda_n)$$

and

$$H = P\Lambda P^* = \sum_{i=1}^{n} \lambda_i u_i u_i^*.$$

In particular, H *maps a unit hypercube with sides* u_i *into a box with orthogonal sides* Pu_i *and volume* $\det(H)$. *The matrix* P *is said to be* unitary *because* $P^*P = I$.

The key idea of the proof of this theorem is as follows. The eigenvalues of H are the roots of $\det(\lambda I - H) = 0$. This is an equation of degree n; it admits n roots, not necessarily distinct. Assume $Hu_1 = \lambda_1 u_1$ and that $||u_1|| = 1$. Then $u_1^* H u_1 = (u_1^* H u_1)^* = \lambda_1 ||u_1||^2$, so that λ_1 is real. Also, the subspace $\mathcal{V}_1$ orthogonal to u_1 is such that if $v \in \mathcal{V}_1$, then $Hv \in \mathcal{V}_1$. One can then continue the construction with $\mathcal{V}_1$. For details, see Theorem B.6.

2.6 Solved Problems

Problem 2.1. *Show that* $f^{-1}(\cap_{i\in I} B_i) = \cap_{i\in I} f^{-1}(B_i)$.

Solution:

By definition, $\omega \in f^{-1}(\cap_{i\in I} B_i)$ if and only if

$$f(\omega) \in \cap_{i\in I} B_i,$$

i.e., if and only if

$$f(\omega) \in B_i, \forall i \in I,$$

i.e., if and only if

$$\omega \in f^{-1}(B_i), \forall i \in I,$$

i.e., if and only if

$$\omega \in \cap_{i\in I} f^{-1}(B_i).$$

Hence,

$$f^{-1}(\cap_{i\in I} B_i) = \cap_{i\in I} f^{-1}(B_i).$$

Problem 2.2. *Exhibit a bounded set of real numbers that has a minimum but no maximum.*

Solution:
For instance $[0, 1)$.

Problem 2.3. *Prove that an upper-bounded set of real numbers has a supremum.*

Solution:
This cannot be proved; it is an axiom, part of the definition of real numbers. For instance, this property is not true for rational numbers: think of the successive decimal expansions of $\sqrt{2}$. By definition, the set or of real numbers is the smallest complete (i.e., containing the suprema) set that contains all the rational numbers.

Problem 2.4. *Show that an upper-bounded nondecreasing sequence must converge.*

Solution:
Let $\{x_n, n \geq 1\}$ be that sequence; it is bounded by b. Let x be the supremum of the sequence. The claim is that $x_n \to x$. We show this by contradiction. If x_n does not converge to x, then there is some $\epsilon > 0$ such that $x - x_n \geq \epsilon$ for all $n \geq 1$. This contradicts the fact that x is the supremum since $x - \epsilon$ is an upper bound that is smaller than x.

Problem 2.5. *Let $\{X_n, n \geq 1\}$ be i.i.d. random variables with mean zero and variance 1. Show that $Y_n = (X_1 + \cdots + X_n)/n^2$ converges almost surely to 0.*

Solution:
Note that $E(Y_n^2) = 1/n^2$, so that $\sum_n E(Y_n^2) < \infty$. We then conclude as in our discussion of Borel-Cantelli.

Problem 2.6. *Use elementary row operations to solve*

$$\begin{bmatrix} 3 & 2 & 1 \\ 1 & 2 & 2 \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

Solution:
Replace row 2 ($r2$) by $r1 - 3 \times r2$ to get

$$\begin{bmatrix} 3 & 2 & 1 \\ 0 & -4 & -5 \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

Choose x_3 arbitrarily. The second equation then implies that $-4x_2 - 5x_3 = 0$, so that $x_2 = -(5/4)x_3$. The first equation then states that $3x_1 + 2x_2 + x_3 = 0$, so that $3x_1 - (5/2)x_3 + x_3 = 0$, so that $x_1 = (3/2)x_3$. We conclude that

$$\mathbf{x}^T = [\frac{3}{2}, -\frac{5}{4}, 1]x_3.$$

Chapter 3

Gaussian Random Variables

3.1 Summary

Here are the key ideas:

- Definition 1: $W = N(0,1)$ if $f_W = \cdots$ (3.1).
- Fact 1: $W = N(0,1)$ iff $g_W(s) = \cdots$ (3.2).
- Theorem 1: If $\{X_n, n \geq 1\}$ are i.i.d. with $E(X_n) = 0$ and $E(X_n^2) = 1$, then $(X_1 + \cdots + X_n)/\sqrt{n} \to_D N(0,1)$. (Related idea: Exercise 3.)
- Definition 2: $X = N(\mu, \sigma^2)$ if $X = \mu + \sigma W$ where $W = N(0,1)$.
- Fact 2: $X = N(\mu, \sigma^2)$ iff $f_X = \cdots$ (3.3).
- Fact 3: $X = N(\mu, \sigma^2)$ iff $g_X = \cdots$ (3.4).
- Definition 3: $\mathbf{X}$ Jointly Gaussian (JG) if $\mathbf{a}^T\mathbf{X}$ is Gaussian for all vector $\mathbf{a}$.
- Fact 4: $\{X_1, \ldots, X_n\}$ i.i.d. $N(0,1) \Rightarrow \mathbf{X}$ is J.G.
- Fact 5: $\{X_1, \ldots, X_n\}$ i.i.d. $N(0,1) \Rightarrow \mu + A\mathbf{X}$ is J.G.

3.2 Why a special attention to Gaussian RVs?

- They are common (CLT, see later).
- They are a 'worst case' (max. entropy given variance).
- They are easy (conditional densities, conditional expectation).
- Preserved by linear systems.
- Elegant solutions of LQG control, Kalman and Wiener Filters.
- Very useful as models of communication links.

3.3 Standard Gaussian Random Variable

Definition 3.1. $N(0,1)$
By definition, $W = N(0,1)$ *if*

$$f_w(w) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{w^2}{2}\}, w \in \Re. \tag{3.1}$$

Exercise 1. *Check that* f_W *is a pdf.*

Fact 3.1. *MGF of* $N(0,1)$

$$g_W(s) := E(e^{sW}) = \exp\{\frac{s^2}{2}\}, s \in \mathcal{C}. \tag{3.2}$$

Exercise 2. *Check that formula.*

Observe that $g_W(-s)$ is the Laplace Transform of p_W. The MGF characterizes the pdf. It is useful to calculate moments (see below). The MGF of the sum of independent RVs is the product of their MGFs. (Convolution of densities becomes product of MGFs.)

3.4 A quick glance at CLT

Theorem 3.1. *CLT Assume* $\{X_m, m \geq 1\}$ *are i.i.d. with* $E(X_m) = 0, var(X_m) = 1$. *Then*

$$\frac{1}{\sqrt{n}} \sum_{m=1}^{n} X_m \approx N(0,1).$$

Proof: (Rough sketch...)
Note that

$$E(\exp\{s(\frac{1}{\sqrt{n}} \sum_{m=1}^{n} X_m)\}) = E(\exp\{\frac{1}{\sqrt{n}} s X_1\})^n$$
$$\approx [E(1 + \frac{s}{\sqrt{n}} X_1 + \frac{s^2}{2n} X_1^2)]^n \approx \exp\{\frac{s^2}{2}\}.$$

□

The following exercise shows the stability of the Gaussian distribution. It is another look at the CLT.

Exercise 3. *Assume that* X, Y *are i.i.d., zero-mean, unit-variance. Assume also that*

$$\frac{X+Y}{\sqrt{2}} =_D X.$$

Show that $X = N(0,1)$.

Solution:
(1) Let $g(s) = g_X(s) = g_Y(s)$. Then

$$\begin{aligned} g(s) &= g(\frac{s}{\sqrt{2}})^2 = \cdots g(\frac{s}{n})^{2n} \\ &\approx (g(0) + g'(0)\frac{s}{n} + g''(0)\frac{s^2}{2n})^{2n} \\ &= (1 + \frac{s^2}{2n})^{2n} \to \exp\{\frac{s^2}{2}\}. \end{aligned}$$

We used the facts that $g'(0) = 0$ and $g''(0) = 1$.

(2) Second solution. Let $\{X, X_n, n \geq 1\}$ be i.i.d. Then,

$$\begin{aligned} X &=_D \frac{X_1 + X_2}{\sqrt{2}} =_D \frac{\frac{X_1+X_2}{\sqrt{2}} + \frac{X_3+X_4}{\sqrt{2}}}{\sqrt{2}} \\ &=_D \frac{1}{2}\{X_1 + \cdots + X_4\} \\ &=_D \frac{1}{2^n}\{X_1 + \cdots + X_{2n}\} \to N(0,1). \end{aligned}$$

□

3.5 General Gaussian: $N(\mu, \sigma^2)$

Definition 3.2. *By definition, $Z = N(\mu, \sigma^2)$ if we can write $Z = \mu + \sigma W$ where $W = N(0, 1)$.*

Fact 3.2. *$X = N(\mu, \sigma^2)$ if and only if*

$$f_Z(z) = \frac{1}{\sigma} f_W(\frac{z-\mu}{\sigma}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{(z-\mu)^2}{2\sigma^2}\}. \tag{3.3}$$

Fact 3.3. *$X = N(\mu, \sigma^2)$ if and only if*

$$g_Z(s) = \exp\{s\mu + \frac{s^2\mu^2}{2\sigma^2}\}. \tag{3.4}$$

Remember this basic representation result contained in the definition: one can write a $N(\mu, \sigma^2)$ as a linear transformation of a $N(0, 1)$. We will see the corresponding multi-dimensional version of this result later.

3.6 Definition: Jointly Gaussian, jpdf, MGF.

The cute idea about JG RVs is that any linear combination is again G. As a consequence, the output of a linear system whose input is JG is again JG. This clean definition leads to clean consequences.

Definition 3.3. *Jointly Gaussian*
A collection $\{X_1, \ldots, X_n\}$ of RVs are jointly Gaussian if

$$\sum_{m=1}^{n} a_m X_m \text{ is a Gaussian RV}$$

for any real numbers $\{a_m, m = 1, \ldots, n\}$.

We can translate this definition in vector notation. Let $\mathbf{X}$ be the column vector with components $\{X_1, \ldots, X_n\}$. We say that $\mathbf{X}$ is JG if $\mathbf{a}^T\mathbf{X}$ is Gaussian for all $\mathbf{a} \in \Re^n$.

Fact 3.4. *Let $\mathbf{X}$ be independent $N(01)$ RVs. They are JG.*

Proof: Let $Y = \mathbf{a}^T\mathbf{X}$. We show that Y is Gaussian by computing $g_Y(s)$. We find

$$\begin{aligned} g_Y(s) &= E(e^{s(\sum_{m=1}^n a_i X_i)}) = \Pi_{m=1}^n E(s a_m X_m) \\ &= \Pi_{m=1}^n \exp\{\frac{s^2 a_m^2}{2}\} = e^{\frac{1}{2}s^2\sigma^2} \end{aligned}$$

where $\sigma^2 = \sum_{m=1}^n a_m^2$.
According to (3.4), this shows that $Y = N(0, \sigma^2)$. □

Here is a very useful fact:

Fact 3.5. *Let $\mathbf{X}$ be a vector of n i.i.d. $N(0, 1)$ RVs. Then*

$$\mu + A\mathbf{X} \tag{3.5}$$

is a vector of JG RVs.

Proof: $\mathbf{a}^T(\mu + A\mathbf{X}) = \mathbf{a}^T\mu + (A\mathbf{a})^T\mathbf{X}$ is Gaussian for all $\mathbf{a}$.

□

As we will see in Lecture 4, all JG random variables are of the form (3.5). Thus, nothing terribly mysterious. What is quite cute is that from a fairly abstract definition (Definition 3.3) one can derive a very concrete representation (3.5). That result will be a key topic in Lecture 4.

3.7 Solved Problems

Problem 3.1. *Suppose $X_1, X_2, \ldots$ are i.i.d. random variables with finite variance. If N is a r.v. taking values in a bounded subset of the nonnegative integers, is independent of $\{X_i\}$, show that*

$$E\left(\sum_{i=1}^{N} X_i\right) = E(N) \cdot E(X_1)\,, \tag{3.6}$$

$$Var\left(\sum_{i=1}^{N} X_i\right) = E(N) \cdot Var(X_1) + Var(N) \cdot (E(X_1))^2\,.$$

Give an example where (3.6) fails when N is not independent of $\{X_i\}$.

Solution:

1. Say that K is a constant for which $N < K$. Then,

$$\begin{aligned} E\left(\sum_{i=1}^{N} X_i\right) &= E\left(E\left[\sum_{i=1}^{N} X_i \middle| N\right]\right) \\ &= E\left(E\left[\sum_{i=1}^{K} 1\{i \le N\} \middle| N\right]\right) \\ &= E\left(\sum_{i=1}^{K} \{i \le N\} E[X_i|N]\right) \\ &= E\left(\sum_{i=1}^{N} E(X_i)\right) \\ &= E(N)E(X_1) \end{aligned}$$

   For a case where the above fails, let $P(X_i = 1) = 1 - P(X_i = -1) = 1/2$. Define N as follows: $N = 1$ if $X_2 = -1$, $N = 2$ if $X_2 = 1$. Then $E(N)E(X_1) = 0$, but

$$\begin{aligned} E\left(\sum_{i=1}^{N} X_i\right) &= E\left[\sum_{i=1}^{N} X_i \middle| X_2 = -1\right]\frac{1}{2} + E\left[\sum_{i=1}^{N} X_i \middle| X_2 = 1\right]\frac{1}{2} \\ &= E(X_1)\frac{1}{2} + E(X_1 + 1)\frac{1}{2} = 1/2\,. \end{aligned}$$

2. Applying part 1 for $\{X_i\}$ and $\{X_i^2\}$, yields

$$\begin{aligned} \mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) &= E\left(\sum_{i=1}^{N} X_i\right)^2 - \left(E\left(\sum_{i=1}^{N} X_i\right)\right)^2 \\ &= E\left(\sum_{i=1}^{N} X_i^2\right) + 2E\left(\sum_{i=1}^{N}\sum_{j>i}^{N} X_i X_j\right) - (E(N)E(X_1))^2 \end{aligned}$$

   Now,

$$\begin{aligned} E\left(\sum_{i=1}^{N}\sum_{j>i}^{N} X_i X_j\right) &= E\left(\sum_{i=1}^{K}\sum_{j>i}^{K} 1\{j \le N\} X_i X_j\right) \\ &= E\left(\sum_{i=1}^{K}\sum_{j>i}^{K} 1\{j \le N\} E[X_i X_j|N]\right) \\ &= (E(X_1))^2 E\left(\sum_{i=1}^{K}(N-i)1\{i \le N\}\right) \\ &= (E(X_1))^2 E(N(N-1))/2\,. \end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) &= E(N)E(X_1^2) + (E(X_1))^2 E(N(N-1)) - (E(N)E(X_1))^2 \\ &= E(N)\mathrm{Var}(X_1) + (E(X_1))^2 \mathrm{Var}(N)\,.\end{aligned}$$

Problem 3.2. *Let X be a r.v. taking values in a bounded subset of the nonnegative integers. Show that $E(X) = \sum_{i=0}^{\infty} P(X > i)$.*

Solution:
Let K be a constant with $X < K$.

$$\begin{aligned}E(X) &= E\left(\sum_{i=0}^{\infty} 1\{X > i\}\right) = E\left(\sum_{i=0}^{K} \{X > i\}\right) \\ &= \sum_{i=0}^{K} P(X > i) = \sum_{i=0}^{\infty} P(X > i)\,.\end{aligned}$$

Problem 3.3. *Find two Gaussian random variables X and Y that are uncorrelated, i.e., $E(XY) = E(X)E(Y)$, but not independent.*

Solution:
Let $X =_D N(0,1)$ and Z independent of X such that $P(Z = 1) = P(Z = -1) = 0.5$. Also, let $Y = XZ$. Note that $Y =_D N(0,1)$. Indeed,

$$\begin{aligned}P(Y \le x) &= P(Y \le x, Z = 1) + P(Y \le x, Z = -1) = P(X \le x, Z = 1) + P(X \ge -x, Z = -1) \\ &= 0.5P(X \le x) + 0.5P(X \ge -x) = P(X \le x).\end{aligned}$$

Also, $E(XY) = E(X^2Z) = E(X^2)E(Z) = 0 = E(X)E(Y)$, so that X and Y are uncorrelated. Finally, X and Y are not independent since $P(|X| < 1, |Y| > 1) = 0 \neq P(|X| < 1)P(|Y| > 1)$.

Problem 3.4. *Let X_1, X_2, X_3 be (mutually) independent $U[0,1]$ r.v., i.e., uniformly distributed in the interval $[0,1]$. Compute $E[(X_1 + X_2)^2 | X_2 + X_3]$.*

Solution:
We first note that,

$$E[(X_1 + X_2)^2 | X_2 + X_3] = E(X_1^2) + 2E(X_1)E[X_2 | X_2 + X_3] + E[X_2^2 | X_2 + X_3].$$

Next, we find the distribution of X_2 given $X_2 + X_3$. Observe that (X_2, X_3) picks a point uniformly from the unit square $[0,1] \times [0,1]$. Therefore, given $X_2 + X_3 = c$, the point (X_2, X_3) is chosen uniformly from the line segment $\{(u, c-u) : u \in \mathbb{R}\} \cap [0,1] \times [0,1]$. Thus, the marginal distribution of each coordinate, and X_2 in particular, is uniformly distributed on its range. Thus, given $X_2 + X_3 = c \le 1$ then $X_2 = U[0,c]$. In this case, $E[X_2 | X_2 + X_3 = c] = E(U[0,c]) = c/2$ and $E[X_2^2 | X_2 + X_3 = c] = E(U[0,c]^2) = c^2/3$. Thus, on $X_2 + X_3 \le 1$,

$$E[(X_1 + X_2)^2 | X_2 + X_3] = \frac{(X_2 + X_3)^2}{3} + \frac{X_2 + X_3}{2} + \frac{1}{3}\,.$$

On the other hand, on $X_2 + X_3 > 1$, we have $X_2 = U[X_2 + X_3 - 1, 1]$. But, $E[U[X_2 + X_3 - 1, 1] | X_2 + X_3] = (X_2 + X_3)/2$ and $E[U[X_2 + X_3 - 1, 1]^2 | X_2 + X_3] = (X_2 + X_3 + (X_2 + X_3 - 1)^2)/3$. Thus,

$$E[(X_1 + X_2)^2 | X_2 + X_3] = \frac{(X_2 + X_3)^2}{3} + \frac{X_2 + X_3}{6} + \frac{2}{3}\,,$$

on $X_2 + X_3 > 1$.

Problem 3.5. *Show that if X, Y, Z are mutually independent, then $f(X, Y)$ and $g(Z)$ are independent for any choice of real-valued functions f, g that make them random variables.*

Show that this may not be the case if X, Y, Z are only pairwise *independent.*

Solution:
Let A, B be Borel subsets of $\mathbb{R}$. Then,

$$\begin{aligned} P(f(X,Y) \in A, g(Z) \in B) &= P((X,Y) \in f^{-1}(A), Z \in g^{-1}(B)) \\ &= P((X,Y) \in f^{-1}(A)) P(Z \in g^{-1}(B)) \\ &= P(f(X,Y) \in A) P(g(Z) \in B)\,, \end{aligned}$$

where the last step from the first to the second line is justified because (X, Y) and Z are independent. Thus, $f(X, Y), g(Z)$ are independent.

To demonstrate that pairwise dependence is not sufficient, consider a r.v. U taking values in $\{1, 2, 3, 4\}$ with equal probability. Let X be the indicator of the event $\{U = 1 \text{ or } 2\}$, Y the indicator of $\{U = 1 \text{ or } 3\}$, and Z of $\{U = 1 \text{ or } 4\}$. It is easy to check that X, Y, Z are pairwise independent, but not mutually independent. E.g., if $X = 1, Y = 1$, then $Z = 1$. Hence, we see that on the event $XY = 1$, $Z = 1$ must hold, so XY is not independent of Z.

Chapter 4

Jointly Gaussian Random Variables

4.1 Summary

Here are the key ideas and results:

- Definition 1: Notation for covariance matrix: (4.1)
- Definition 2: $N(\mu, K)$
- Fact 3: If $\mathbf{X} = N(\mu, K)$, then $g_{\mathbf{X}} =$ (4.2)
- Theorem 1: For JG, $\perp \Rightarrow$ independence

4.2 Simple Algebra

Recall the following ideas for random vectors:

Definition 4.1. *Covariance Matrix*

$$K_{\mathbf{X},\mathbf{Y}} := cov(\mathbf{X}, \mathbf{Y}) := E((\mathbf{X} - E(\mathbf{X}))(\mathbf{Y} - E(\mathbf{Y}))^T) = E(\mathbf{X}\mathbf{Y}^T) - E(\mathbf{X})E(\mathbf{Y})^T. \tag{4.1}$$

For complex vectors, one replaces T by $*$.

You will note that the covariance matrix is the matrix of covariances. The matrix $K_{\mathbf{Y},\mathbf{Y}}$ is sometimes denoted by $K_{\mathbf{Y}}$. Some authors prefer Σ to K.

Fact 4.1. *For any random vectors $\mathbf{X}, \mathbf{Y}$ and matrices A, B, μ, ν one has*

$$cov(\mu + A\mathbf{X}, \nu + B\mathbf{Y}) = A\,cov(\mathbf{X}, \mathbf{Y})B^T.$$

4.3 Definition of Jointly Gaussian

Recall (see D3 in L3) that $\mathbf{X}$ are JG if $\mathbf{a}^T\mathbf{X}$ is Gaussian for all vector $\mathbf{a}$. We introduce a bit of notation:

Definition 4.2. *We write $\mathbf{X} = N(\mu, K)$ to indicate that $\mathbf{X}$ is JG with mean μ and covariance $K := E((\mathbf{X} - \mu)(\mathbf{X} - \mu)^T)$.*

Here is an immediate result.

Fact 4.2. *Assume that $\mathbf{X} = N(\mu, K)$. Then*

$$A\mathbf{X} + \mathbf{b} = N(A\mu + \mathbf{b}, AKA^T).$$

4.4 MGF of JG Random Variables

Fact 4.3. *Assume that $\mathbf{X} = N(\mu, K)$ is JG. Then*

$$g_{\mathbf{X}}(\sigma) := E(e^{\sigma^T\mathbf{X}}) = \exp\{\sigma^T\mu + \frac{1}{2}\sigma^T K\sigma\}. \tag{4.2}$$

Proof:
Note that $Y := \sigma^T X = N(\sigma^T\mu, \sigma^T K\sigma)$ and $g_{\mathbf{X}}(\sigma) = g_Y(1)$.

□

The main point of the Joint MGF $E(e^{\sigma^T\mathbf{X}})$ is the following.

Theorem 4.1. *The Joint MGF $E(e^{\sigma^T\mathbf{X}})$ determines uniquely the joint pdf $f_{\mathbf{X}}(\xi)$.*

The Joint MGF is the n-dimensional Laplace transform of the joint pdf. As in the one dimensional case, it has a unique inverse. That is, if two joint pdf have the same Joint MGF, they must be identical.

4.5 Uncorrelated JG are independent!

The following result is very useful.

Theorem 4.2. *JG RVs are independent iff they are uncorrelated The JG random variables $\mathbf{X}$ are independent iff they are uncorrelated.*

Proof:
$\Rightarrow$ Independent implies uncorrelated, for arbitrary RVs.

$\Leftarrow$ Assume $\mathbf{X}$ are JG and uncorrelated, so that $K = \text{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$ and

$$\sigma^T \mu + \frac{1}{2}\sigma^T K \sigma = \sum_{m=1}^{N} [s_m \mu_m + \frac{1}{2} s_m^2 \sigma_m^2].$$

Then, by (4.2),

$$g_{\mathbf{X}}(\sigma) = \Pi_{m=1}^n e^{s_m \mu_m + \frac{1}{2} s_m^2 \sigma_m^2} = \Pi_{m=1}^n g_{X_m}(s_m),$$

since $X_m = N(\mu_m, \sigma_m^2)$ for $m = 1, \ldots, n$. This implies that $g_{\mathbf{X}}(\sigma)$ is the same as that of independent $N(\mu_i, \sigma_i^2)$ random variables. Indeed, if the random variables X_m *were* independent $N(\mu_m, \sigma_m^2)$, then we would find that

$$g_{\mathbf{X}}(\sigma) = E(e^{\sigma^T \mathbf{X}}) = E(e^{\sum_m s_m X_m}) = \Pi_{m=1}^n g_{X_m}(s_m).$$

By Theorem 4.1, this implies that the random variables $\mathbf{X}$ are independent.

□

Note that the MGF of JG $\mathbf{X}$ is completely specified by μ and K. Consequently, $f_{\mathbf{X}}$ is also completely specified by μ and K. Accordingly, it is not surprising that independence is determined by the covariance matrix Σ. Taking the inverse LT, the joint pdf is also determined uniquely by μ and K; we discuss the explicit formulas in L5.

4.6 Conditional Expectation: An Example

Here is a little example that illustrates the power of Theorem 4.2.

Assume that $(X, Y)^T = N(0, K)$ with

$$K = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}.$$

We calculate $E[X|Y]$ as follows. First we find a number α such that

$$X - \alpha Y \perp Y.$$

That is, since the random variables are zero mean,

$$0 = E((X - \alpha Y)Y) = E(XY) - \alpha E(Y^2) = 1 - \alpha,$$

where we found the numerical values by looking at K. We conclude that $\alpha = 1$, so that

$$X - Y \perp Y.$$

Second, since $X - Y$ and Y are JG (being linear combinations of the JG random variables X and Y), we conclude from Theorem 4.2 that $X - \alpha Y$ and Y are independent (since they are uncorrelated). Now,

$$E[X|Y] = E[X - Y + Y|Y] = E[X - Y|Y] + E[Y|Y] = E(X - Y) + Y = Y.$$

In this derivation, the second identity comes from the following two properties of conditional expectation:

$$\text{P1: } X, Y \text{ independent } \Rightarrow E[X|Y] = E(X);$$

$$\text{P2: } E[Xg(Y)|Y] = E[X|Y]g(Y).$$

We will extend this example to the general vector case in L5.

4.7 Solved Problems

Problem 4.1. *Let $\vec{X} = A\vec{W}$, and $\vec{Y} = B\vec{W}$, where $W = N(\vec{0}, I_{n\times n})$, $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times n}, \det A \neq 0, \det B \neq 0$. What is the conditional pdf $f_{\vec{X}|\vec{Y}}(\cdot|\cdot)$, of $\vec{X}$ given $\vec{Y}$?*

Solution:
Since B is nonsingular, we see that $\vec{W} = B^{-1}\vec{Y}$, so that $\vec{X} = AB^{-1}\vec{Y}$. Thus, given $\vec{Y} = \vec{y}$, the random variable $\vec{X}$ takes the value $AB^{-1}\vec{y}$ with probability one and its density is a Dirac impulse at that value.

Problem 4.2. *Suppose (X, Y_1, Y_2) is a zero-mean jointly Gaussian (JG) random vector, with covariance matrix*

$$K = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

Find the conditional pdf $f_{X|Y_1,Y_2}(\cdot|\cdot)$ of X given (Y_1, Y_2). Calculate $E[X|Y_1, Y_2]$.

Solution:
The complication here is that $K_{\vec{Y}}$ is singular. This means that the vector $\vec{Y}$ does not have a density. In fact,

$$E(Y_1 - 2Y_2)^2 = \text{var}(Y_1) - 4\text{cov}(Y_1, Y_2) + 4\text{var}(Y_2) = 4 - 8 + 4 = 0,$$

which shows that $Y_1 = 2Y_2$. Thus we should consider $f_{X|Y_1}$. Now,

$$X - aY_1 \perp Y_1$$

if $0 = E((X - aY_1)Y_1) = \text{cov}(X, Y_1) - a\text{var}(Y_1) = 2 - a$, i.e., if $a = 2$. In that case, $X - 2Y_1$ and Y_1 are independent. But, $X = X - 2Y_1 + 2Y_1$, so that given $Y_1 = y_1$ the random variable X is $N(2y_1, \sigma^2)$ where

$$\sigma^2 = \text{var}(X - 2Y_1) = \text{var}(X) - 4\text{cov}(X, Y_1) + 4\text{var}(Y_1) = 4 - 8 + 16 = 12.$$

We conclude that

$$f_{X|Y_1,Y_2}[x|y_1, y_2] = \frac{1}{\sqrt{2\pi 12}} \exp\{-\frac{1}{24}(x - 2y_1)^2\}, \text{ if } y_1 = 2y_2.$$

Also, this expression is meaningless if $y_1 \neq 2y_2$.

Along the way, we found that $E[X|Y_1, Y_2] = E[X|Y_1] = 2Y_1$ since $X = (X - 2Y_1) + 2Y_1$ and $X - 2Y_1$ is zero-mean and independent of Y_1.

Chapter 5

Representation of Jointly Gaussian Random Variables

5.1 Summary

Here are the key ideas and results:

- Example in Section 5.2 of $E[X|Y]$ for (X,Y) JG.
- Theorem 1: If $(\mathbf{X},\mathbf{Y})$ are $N(0,K)$ with $|K|\neq 0$, then, $f_{\mathbf{X}|\mathbf{Y}}[\cdot|\mathbf{y}] = N(A\mathbf{y},\Sigma)$ where A,Σ are given by (23.2).
- Theorem 2: Under same assumptions, $E[\mathbf{X}|\mathbf{Y}] = \cdots$ (5.2).
- Theorem 3: K is a covariance matrix iff it is positive semi-definite; then $K = R^2 = Q\Lambda Q^T$ for some orthogonal matrix Q and $R = Q\Lambda^{1/2}Q^T$
- Theorem 4: If $\mathbf{X} = N(\mu,K)$ with $|K|\neq 0$, then $f_{\mathbf{X}} = \cdots$ (23.1)
- Fact 1: Sum of squares of two i.i.d. $N(0,1)$ is exponentially distributed.

5.2 Example - Continued

Recall our little example from the end of L4:

Assume that $(X,Y)^T = N(0,K)$ with

$$K = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}.$$

We found that $X-Y \perp Y$, so that $X-Y$ and Y are independent. We used that to calculate $E[X|Y] = E[X-Y+Y|Y] = Y$.

There is another useful consequence of the independence of $X-Y$ and Y: Given that $Y=y$, we see that

$$X = (X-Y)+Y\ |_{Y=y} = N(y,\sigma^2)$$

where

$$\sigma^2 = \mathrm{var}(X-Y) = var(X) + \mathrm{var}(Y) - 2cov(X,Y) = 3+1-2 = 2.$$

That is,

$$f_{X|Y}[\cdot|y] = N(y, 2).$$

Here are a few key observations:

- The mean of X given $Y = y$ depends on y (it is $E[X|Y = y] = y$).
- However, the variance of X given $Y = y$ does not depend on y! Again, this fact follows from the independence of $X - Y$ and Y. The 'noise' $X - Y$ that is added to Y to get X does not depend on Y.
- The variance of X given Y is smaller than the variance of X. (Here, it is 2 instead of 3.)

We generalize these observations in the next section.

5.3 Conditional Densities

Theorem 5.1. *Conditional Densities*
Assume that $(\mathbf{X}, \mathbf{Y})$ *are* $N(0, K)$ *with* $|K| \neq 0$. *Then, given* $\mathbf{Y} = \mathbf{y}$, $\mathbf{X} = N(A\mathbf{y}, \Sigma)$ *where*

$$A = K_{XY}K_Y^{-1} \text{ and } \Sigma = K_X - K_{XY}K_Y^{-1}K_{YX}. \tag{5.1}$$

Proof:
First note that $|K_Y| \neq 0$. Second, observe that

$$\mathbf{Z} := \mathbf{X} - A\mathbf{Y} \perp \mathbf{Y} \text{ if } K_{XY} = AK_Y, \text{ i.e., } A = K_{XY}K_Y^{-1}.$$

Then, $\mathbf{X} = A\mathbf{Y} + \mathbf{Z}$ where $\mathbf{Z}$ and $\mathbf{Y}$ are independent. Also,

$$\begin{aligned}\Sigma := K_Z &= E(\mathbf{X} - A\mathbf{Y})(\mathbf{X} - A\mathbf{Y})^T) \\ &= K_X - AK_{YX} - K_{XY}A^T + AK_YA^T \\ &= K_X - K_{XY}K_Y^{-1}K_{YX}.\end{aligned}$$

□

One interesting observation is that the variance Σ of $\mathbf{X}$ given that $\mathbf{Y} = \mathbf{y}$ does not depend on the value of $\mathbf{y}$. Another observation is the reduction of the variance K_X due to the observation.
We can also derive the following consequence.

Theorem 5.2. *Conditional Expectation*
Under the same assumptions as Theorem 5.1,

$$E[X|Y] = K_{XY}K_Y^{-1}\mathbf{Y}. \tag{5.2}$$

Exercise 4. *Extend the results of Theorems 5.1 and 5.2 to nonzero-mean random variables.*

5.4 Covariance Matrices

Assume that K is a covariance matrix. That means that $K = E(\mathbf{X}\mathbf{X}^T)$ for some zero-mean random vector $\mathbf{X}$. Here are some basic properties.

Theorem 5.3. *Properties of Covariance Matrix*

Assume that K is a covariance matrix. That matrix must have the following properties.

(1) K is positive semi-definite. That is, $\mathbf{a}^T K \mathbf{a} \geq 0$ for all $\mathbf{a} \in \Re^n$.

(2) K is positive definite if and only if $|K| \neq 0$.

(3) The eigenvalues of K are real and nonnegative. Let $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$ be the eigenvalues of K repeated according to their multiplicity. Then $|K| = \lambda_1 \lambda_2 \cdots \lambda_n$. There is an orthonormal matrix Q such that $KQ = Q\Lambda$.

(4) If K is positive definite, then $K^{-1} = Q\Lambda^{-1}Q^T$.

(5) There is a unique positive semi-definite matrix R such that $K = R^2$ and

$$R = Q\Lambda^{1/2}Q^T.$$

(6) A positive semi-definite symmetric matrix K is a covariance matrix. It is the covariance matrix of $R\mathbf{X}$ where $X = N(0, I)$.

Proof:

(1) Assume $K = E(\mathbf{X}\mathbf{X}^T)$ for some zero-mean random vector $\mathbf{X}$. For $\mathbf{a} \in \Re^n$ one has

$$\mathbf{a}^T K \mathbf{a} = E(Y^2) \text{ where } Y = \mathbf{a}^T \mathbf{X}.$$

Hence $\mathbf{a}^T \Sigma \mathbf{a} \geq 0$.

(2)-(4) Since K is positive semi-definite, (2)-(4) follow from Theorem 6 in [**?**].

(5) The issue is uniqueness. The matrix R is such that $R = V\Lambda^{1/2}V^T$ where $K = V\Lambda V^T$. Thus, V are the eigenvectors of K and $\Lambda^{1/2}$ is fixed.

(6) is immediate.

□

The above theorem tells us about the shape of $f_{\mathbf{X}}$, as stated in the next result, illustrated in Figure 5.1.

Theorem 5.4. *Assume that $\mathbf{X} = N(0, K)$. If $|K| = 0$, the RVs $\mathbf{X}$ do not have a joint density. If $|K| \neq 0$, then*

$$f_{\mathbf{X}}(\xi) = \frac{1}{(2\pi)^{n/2}|K|^{1/2}} \exp\{-\frac{1}{2}\xi^T K^{-1} \xi\}. \tag{5.3}$$

Also, the level curves of $f_{\mathbf{X}}$ are ellipses whose axes are the eigenvectors of K and dimensions scaled by the square roots of the eigenvalues of K.

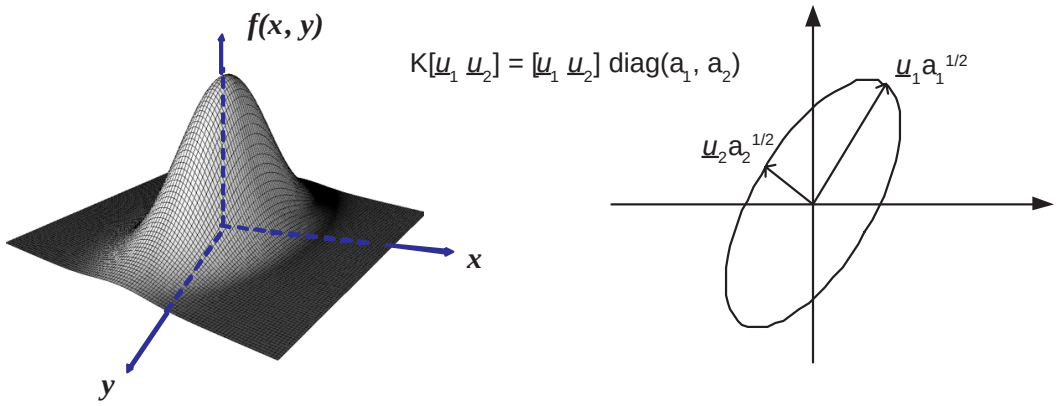


Figure 5.1: The $N(0, K)$ probability density function.

Proof:

The expression for $f_{\mathbf{X}}$ follows from the representation $X = RY$ and the observation that if $\xi = R\mathbf{y}$, then $\mathbf{y} = R^{-1}\xi$ and $\mathbf{y}^T\mathbf{y} = \xi^T R^{-2}\xi = \xi K^{-1}\xi$.

The level curves are sets of ξ such that $\xi^T K^{-1}\xi = \mathbf{y}^T\mathbf{y} = c$ where $\xi = R\mathbf{y}$. Thus, $\mathbf{y}$ belongs to a circle with radius $\sqrt{c}$ and ξ belongs to an ellipse whose axes are the eigenvectors $\mathbf{u}_i$ of R. (See Section VIII in [**?**].)

□

5.5 Generating a $N(0, 1)$- Random Variable.

In this section, we explain a cute result that can be used to generate Gaussian random variables. We start with some general ideas.

5.5.1 Generating Random Variables

For well chosen values of α, β, N, the sequence X_n defined below looks like i.i.d. $U[0, 1]$:

$$Y_{n+1} = (\alpha Y_n + \beta) \mathrm{mod}(N), X_{n+1} = X_{n+1}/N.$$

Assume then that $U =_D U[0, 1]$. We can generate a RV X with cdf $F_X(.)$ by computing

$$X = F_X^{-1}(U).$$

Indeed, $P(X \leq x) = P(U \leq F_X(x)) = F_X(x)$, as desired. For instance,

$$X = -\frac{1}{\lambda}\ln(U)$$

generates a random variable with

$$\begin{aligned} P(X \leq x) &= P(-\frac{1}{\lambda}\ln(U) \leq x) \\ &= P(\ln(U) \geq -\lambda x) = P(U \geq e^{-\lambda x}) = 1 - e^{-\lambda x}. \end{aligned}$$

That is, X is exponentially distributed with rate λ (i.e., with mean λ^{-1}).

5.5.2 Generating Gaussian Random Variables

The general method discussed in the previous section does not work well for generating a $N(0,1)$ RV X because F_X does not have a close form and neither does its inverse. Of course, we could tabulate F_X^{-1}, but that is not very elegant. Instead, one uses the following fact.

Fact 5.1. *Let X, Y be i.i.d. $N(0,1)$. Then $Z = X^2 + Y^2$ is exponentially distributed with mean 2.*

By the way, one says that $\sqrt{Z} = ||(X,Y)||_2$ has a Raleigh distribution. Thus a Raleigh-distributed random variable is the square root of an exponentially-distributed random variable. This distribution is important in wireless system (Raleigh fading).

Proof:

We calculate the MGF of Z as follows:

$$\begin{aligned} E(e^{sZ}) &= \frac{1}{2\pi}\int_0^\infty \int_0^\infty e^{s(x^2+y^2)} e^{-x^2/2-y^2/2} dxdy \\ &= \frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} e^{sr^2-r^2/2} r dr d\theta = \int_0^\infty r e^{(s-1/2)r^2} dr \\ &= \frac{1}{2s-1}\int_0^\infty de^{(s-1/2)r^2} = \frac{1}{1-2s}. \end{aligned}$$

On the other hand, if V is exponentially distributed with mean 2, then

$$\begin{aligned} E(^{sV}) &= \int_0^\infty e^{sv}\frac{1}{2}e^{-v/2} dv \\ &= \frac{1}{2(s-1/2)}\int_0^\infty de^{(s-1/2)v} = \frac{1}{1-2s}. \end{aligned}$$

Comparing these expressions shows that $Z = X^2 + Y^2 =_D V$. □

One method to generate the random variables X and Y is then to generate Z as an exponentially distributed random variable, then to calculate

$$X = \sqrt{Z}\cos(\theta) \text{ and } Y = \sqrt{Z}\sin(\theta)$$

where θ is $U[0,2\pi]$ and is independent of Z. This procedure yields two i.i.d. $N(0,1)$ random variables X and Y.

Exercise 5. *Assume that X_1 and X_2 are two i.i.d. random variables and that their joint distribution is invariant under rotation. Show that the random variables are Gaussian.*

Once we can generate i.i.d. $N(0,1)$ random variables, we know how to generate any $N(\mu, K)$ random vector.

5.6 Solved Problems

Problem 5.1. *Suppose $\vec{Y}$ is zero-mean JG with covariance matrix*

$$K = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -3 \\ 0 & -3 & 6 \end{bmatrix}.$$

Find a matrix A such that $\vec{Y} = A\vec{Z}$ where $\vec{Z} = N(0, \mathbf{I})$.

Solution:
Here again, the complication is that K is singular so that the vector $\vec{Y}$ does not have a density. This implies that $\vec{Y}$ is a linear combination of fewer than 3 independent Gaussian random variables. To find the appropriate linear combination, we use the fact that K is Hermitian. Consequently, by Theorem 2.1, we can write $K = P\Lambda P^T$ where $P^T P = I$. Thus, $K = AA^T$ where $A = P\Lambda^{1/2}$. We can then write $\vec{Y} = A\vec{Z}$ where $\vec{Z} = N(0, \mathbf{I})$. We leave you the calculation of P and A.

Chapter 6

Detection

6.1 Summary

Here are the key ideas and results:

- Formulation of the detection problem in Section 6.2.
- Maximum Aposteriori Probability detector (MAP), in Section 6.3.
- Maximum Likelihood detector (ML), in Section 6.3.
- Examples: two basic models of binary detection.

6.2 Detection

The detection problem can be formulated as follows. (X, Y) are random variables where X takes values in a finite set $\{1, 2, \ldots, N\}$. One observes Y and one wants to guess the value of X. This estimate is a function of the observations, and we denote it by $g(Y)$. The "goodness" of each estimate is specified by a cost function $c(x, z)$ that associates a cost in guessing z when the true value is x.

The detection problem is the following: choose a function $g(\cdot)$ such that it minimizes

$$E(c(X, g(Y))) \,. \tag{6.1}$$

One interpretation of this problem is the following. Consider the case where one has to detect a sequence of i.i.d. values $X_1, X_2, \ldots$. Say that for each X_i, one observes Y_i. If the decision rule $g(\cdot)$ is used, then this will incur an average cost of

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} c(X_i, g(Y_i)) \,.$$

By the law of large numbers, this is the same as (6.1).

The key in minimizing (6.1) is to minimize $E[c(X, z)|Y = y]$ over z, for each value y of the observation Y. If z_0 is the optimum, then if we define $g_0(\cdot)$ by $g_0(y) = z_0$, then $g_0(\cdot)$ minimizes (6.1), as well, for if $h(\cdot)$ is any other function, then $E(c(X, h(Y))) = E(E[c(X, h(Y))|Y]) \geq E(E[c(X, g_0(Y))|Y]) = E(c(X, g_0(Y)))$.

Now,

$$E[c(X,z)|Y=y] = \sum_i c(i,z) P[X=i|Y=y]$$
$$= \sum_i c(i,z) \frac{f_{Y|X}(y|i)P(X=i)}{f_Y(y)} .$$

Thus, to minimize (6.1) one can equivalently solve

$$\min_z \sum_i c(i,z) P(X=i) f_{Y|X}(y|i) , \tag{6.2}$$

in the case Y is a continuous r.v., and

$$\min_z \sum_i c(i,z) P(X=i) P[Y=y|X=i] , \tag{6.3}$$

when Y is discrete.

6.3 Minimizing the probability of wrong detection

Assume $X, g(Y) \in \{1,\dots,N\}$. A common cost function for this case is $c(x,z) = 1\{x \neq z\}$. It corresponds to minimizing $P(X \neq g(Y))$, the probability of wrong detection. By (6.3), we solve,

$$\arg\min_z \sum_i (1 - 1\{i \neq z\}) P(X=i) P[Y=y|X=i] = \arg\max_z P(X=z) P[Y=y|X=z]$$
$$= \arg\max_z P[X=z|Y=y] . \tag{6.4}$$

The last expression is the posterior distribution of X, by incorporating the information included in the observation y. For this reason, we call this rule the *Maximum Aposteriori Probability* (MAP) detector. The optimum z as a function of the observation y, is denoted by $\text{MAP}[X|Y=y]$. Notice that in general, it depends on the knowledge (or a stipulation) of the prior distribution of X.

In some cases, the postulation of a prior may be quite arbitrary. Indeed, one assumes he knows the distribution of what he intends to estimate. Because of this, a distribution free rule might be desirable. One way to do this, is simply by agreeing on the use of a certain prior. When this is the uniform prior, $P(X=i) = 1/N$ for all $i = 1,\dots,N$, one then must solve

$$\max_z P[Y=y|X=z] .$$

Intuitively, this rule spits back the value z of X under which the observation is most likely. The resulting rule is called the *Maximum Likelihood* (ML) detector, and the estimate of X it provides, the ML estimate (MLE) denoted by $\text{MLE}[X|Y=y]$.

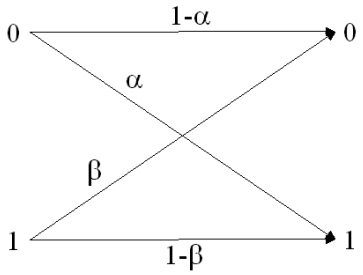


Figure 6.1: When $X = 0$, with probability α the wrong output Y will be produced. Similarly, with probability β the input $X = 1$ is wrongly produced at the output.

6.4 Binary detection models

Consider the case of the binary channel depicted in Figure 1.

Since $P[Y = 0|X = 0] = 1 - \alpha, P[Y = 0|X = 1]$, the MLE given $Y = 0$ is 1, if $\beta > 1 - \alpha$, and 0 otherwise.

To compute the MAP, one needs to further decide on a prior. Let $p_i := P(X = i), i = 0, 1$. Using the first line in (6.4), given $Y = 0$, the MAP estimate is 1, if $p_1\beta > p_0(1 - \alpha)$, and 0 otherwise.

Notice that $\text{MAP}[X|Y = 0] = \text{MLE}[X|Y = 0]$, when $p_0 = p_1$, as explained in the previous section. When $p_0 > p_1$, MAP favors 0 more than ML does.

Now, let us consider a model with continuous observations Y. In particular, when $X = 0$,

$$f_{Y|X}(y|0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} ,$$

while when $X = 1$,

$$f_{Y|X}(y|1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} ,$$

for some $\mu, \sigma > 0$.

$\text{MLE}[X|Y = y] = 1$ or 0, depending on whether the so-called *likelihood-ratio* $f_{Y|X}(y|1)/f_{Y|X}(y|0) > 1$ or < 1. This happens if $y > \mu/2$ or $< \mu/2$, respectively.

For computing $MAP[X|Y = y]$, we have,

$$P[X = 1|Y = y] = \frac{f_{Y|X}(y|1)p_1}{f_Y(y)} \text{ and } P[X = 0|Y = y] = \frac{f_{Y|X}(y|0)p_0}{f_Y(y)} .$$

Therefore, we should compare the likelihood-ratio $f_{Y|X}(y|1)/f_{Y|X}(y|0)$ with p_0/p_1 instead of 1 which we used in MLE. Plugging the expressions for $f_{Y|X}(\cdot|\cdot)$ above, we get that MAP[X—Y=y] chooses 1 if

$$y > \frac{\sigma^2}{\mu} \log \frac{p_0}{p_1} + \frac{\mu}{2} ,$$

and 0 otherwise.

6.5 Solved Problems

Solution:
Given some observation $\vec{y} = (y_1, \ldots, y_n) \in \{0,1\}^n$, the likelihood ratio $\Lambda(\vec{y})$ is

$$\Lambda(\vec{y}) = \frac{P[\vec{Y} = \vec{y}|X=1]}{P[\vec{Y} = \vec{y}|X=0]} = \frac{p_1^{\sum_{i=1}^n y_i}(1-p_1)^{n-\sum_{i=1}^n y_i}}{p_0^{\sum_{i=1}^n y_i}(1-p_0)^{n-\sum_{i=1}^n y_i}}\,. \tag{6.5}$$

The Neyman-Pearson test is to choose $Z=1$ with probability $\phi(\vec{y})$ given by

$$\phi(\vec{y}) = \begin{cases} 1, & \text{if } \Lambda(\vec{y}) > \lambda, \\ \gamma, & \text{if } \Lambda(\vec{y}) = \lambda \\ 0, & \text{otherwise,} \end{cases}$$

for λ, γ given by $P[Z=1|X=0] = \beta \Leftrightarrow P[\Lambda(\vec{Y}) > \lambda|X=0] + \gamma P[\Lambda(\vec{Y}) = \lambda|X=0] = \beta$.

From (6.5) we see that $\sum_{i=1}^n Y_i$ is a sufficient statistic.

Problem 6.1. *Repeat Problem 1 for the following situation. When $X=0$, $Y_1, \ldots, Y_n$ are i.i.d. $U[0,2]$ r.v's. When $X=1$, they are $U[0,3]$.*

What was the sufficient statistic? Explain.

Solution:
Let $3 \geq y_i \geq 0$ for all $i = 1, \ldots, n$. Then,

$$\Lambda(\vec{y}) = \frac{f_{\vec{Y}|X}(\vec{y}|1)}{f_{\vec{Y}|X}(\vec{y}|0)} = \frac{3^{-n}}{2^{-n}1\{\max_i y_i \leq 2\}} = \begin{cases} (2/3)^n & , \text{ if } \max_i y_i \leq 2 \\ \infty & , \text{ otherwise}\,. \end{cases} \tag{6.6}$$

Set $\lambda = (2/3)^n$, and $\gamma = \beta$. Then,

$$P[Z=1|X=0] = P[\Lambda(\vec{Y}) > \lambda|X=0] + \gamma P[\Lambda(\vec{Y}) = \lambda|X=0] = 0 + \gamma 1 = \beta\,.$$

From (6.6) we see that $\max_i Y_i$ is a sufficient statistic.

Problem 6.2. *Repeat Problem 2 for the case where $Y_i = N(1,1)$ when $X=0$, and $Y_i = N(0,4)$ when $X=1$.*

What was the sufficient statistic? Explain.

Solution:

$$\Lambda(\vec{y}) = \frac{f_{\vec{Y}|X}(\vec{y}|1)}{f_{\vec{Y}|X}(\vec{y}|0)} = \frac{1}{2^n}\prod_{i=1}^n \exp((3y_i^2 - 8y_i + 4)/8)\,. \tag{6.7}$$

Since $\Lambda(\vec{y})$ is a continuous r.v., γ is not needed, and λ is (uniquely) determined by $P[\Lambda(Y) > \lambda|X=0] = \beta$.

From (6.7) we see that $\sum_i (3Y_i^2 - 8Y_i)$ is a sufficient statistic.

Problem 6.3. *Let U, V be two independent $U[0,1]$ r.v's.*

1. *Compute $X = L[\cos(U)|U+V]$, the linear least squares error estimator of $\cos(U)$ given $U+V$.*
2. *Compute $Y = E[\cos(U)|U+V]$*

3. Compare $E((X-U^2)^2)$ and $E((Y-U^2)^2)$.

Solution:

1.
$$L[\cos(U)|U+V] = \frac{\mathrm{Cov}(\cos(U), U+V)}{\mathrm{Var}(U+V)}(U+V) + E(\cos(U)) - \frac{\mathrm{Cov}(\cos(U), U+V)}{\mathrm{Var}(U+V)}E(U+V)$$

Now,

$$\begin{aligned}\mathrm{Cov}(\cos(U), U+V) &= \mathrm{Cov}(\cos(U), U) + \mathrm{Cov}(\cos(U), V) \\ &= E(U\cos(U)) - E(U)E(\cos(U))\,,\end{aligned}$$

by independence of U, V. But,

$$\begin{aligned}E(U\cos(U)) &= \int_0^1 u\cos(u)du = \sin(1) - \int_0^1 \sin(u)du \\ &= \sin(1) + \cos(1) - 1\,,\end{aligned}$$

so $\mathrm{Cov}(\cos(U), U+V) = \cos(1) + 0.5\sin(1) - 1$. Also, $\mathrm{Var}(U+V) = \mathrm{Var}(U) + \mathrm{Var}(V) = 1/6$, and

$$E(\cos(U)) = \int_0^1 \cos(u)du = \sin(1)\,.$$

2. Notice that (U, V) is uniformly distributed on the unit square. On $U+V=s$, $(U,V) = (U, s-U)$ is uniformly distributed, so U is $U[0,s]$ if $s<1$, and $U[s-1,1]$ if $s \geq 1$. Hence, if $s \geq 1$,
$$E[\cos(U)|U+V] = \int_{s-1}^1 \frac{\cos(u)}{2-s}du = \frac{\sin(1)-\sin(s-1)}{2-s}\,,$$
and
$$E[\cos(U)|U+V] = \int_0^s \frac{\cos(u)}{s}du = \frac{\sin s}{s}\,,$$
if $s<1$.

3. From theorem proved in class, we expect $E((X-\cos(U))^2) \geq E((Y-\cos(U))^2)$.

Chapter 7

Binary Detection under AWGN

7.1 Summary

Here are the key ideas and results:

- Section 7.3: binary detection with vector observations under AWGN. The ideas from Chapter 6 carry over to the case of vector observations.
- Definition: Sufficient statistic, in Section 7.3.1.
- BER calculations and system design, in Section 7.4.

7.2 Example: system design under power constraints

Let us consider the basic (scalar observation) model of binary detection in Chapter 6, under Gaussian noise:

$$Y = \mu X + W \,,$$

where $X \in \{0, 1\}$ is the r.v. we want to guess, W is noise $N(0, \sigma^2)$, and Y is the observation. Assume X has the distribution $P(X = 0) = p_0$, $P(X = 1) = p_1$.

In Chapter 6 we saw that $\hat{X} := \text{MAP}[X|Y = y] = 1$ or 0 depending on whether $Y > \theta$ or $< \theta$, respectively, where

$$\theta := \frac{\mu}{2} + \frac{\sigma^2}{\mu} \log \frac{p_0}{p_1} \,.$$

Then,

$$\begin{aligned} P(X \neq \hat{X}) &= P[X \neq \hat{X}|X = 0]p_0 + P[X \neq \hat{X}|X = 1]p_1 \\ &= P(N(0, \sigma^2) \geq \theta)p_0 + P(N(\mu, \sigma^2) \leq \theta)p_1 \\ &= P\left(N(0, 1) \geq \frac{\theta}{\sigma}\right) p_0 + P\left(N(0, 1) \leq \frac{\theta - \mu}{\sigma}\right) p_1 \,. \end{aligned}$$

When $p_0 = p_1$ we have the further simplification

$$P(X \neq \hat{X}) = P\left(N(0, 1) \geq \frac{\mu}{2\sigma}\right) \,. \tag{7.1}$$

Now, if we wanted to design our communication system such that the bit-error-rate (BER) is less than, say 10^{-10}, then we could use (7.1), to calculate the average energy per bit required

to achieve this. (Note that $P(N(0,1) \geq 6) \leq 10^{-10}$.) For example, if the amplitude of the signal at the sender is A, and G the channel gain, then $\mu = AG$. Since no signal is sent when $X = 0$, which happens half of the time ($p_0 = p_1 = 1/2$), the average energy per bit is $p = \frac{1}{2}A^2$. Thus,

$$\frac{p}{\sigma^2} \geq \frac{72}{G^2}\,,$$

gives a lower bound in terms of the power of the noise.

7.3 Binary detection with vector observations

The key ideas in L7 carry over to the vector case: Let $X \in \{0,1\}$ as before, $\vec{\mu_0}, \vec{\mu_1} \in \mathbb{R}^n$ and

$$\vec{Y} = \vec{\mu_i} + \vec{Z}\,, \quad \text{when } X = i\,,$$

where $\vec{Z} = N(0,K)$. This can model a variety of situations. One can think of $\vec{Y}$ as the received signals at an antenna array, each coordinate corresponding to the output of a different antenna. Another way to look at this model, is by considering $\vec{\mu_0}, \vec{\mu_1}$ as waveforms, corrupted by noise with an autocorrelation structure specified by matrix K.

Let's compute $\text{MLE}[X|\vec{Y} = \vec{y}]$ for the case that $|K| \neq 0$.

$$\begin{aligned}\frac{f_{\vec{Y}|X}(\vec{y}|1)}{f_{\vec{Y}|X}(\vec{y}|0)} &= \exp\left(-\frac{1}{2}(\vec{y}-\vec{\mu_1})^T K^{-1}(\vec{y}-\vec{\mu_1}) + \frac{1}{2}(\vec{y}-\vec{\mu_0})^T K^{-1}(\vec{y}-\vec{\mu_0})\right) \\ &= \exp\left((\vec{\mu_1}-\vec{\mu_0})^T K^{-1}\vec{y} - \frac{1}{2}\vec{\mu_1}^T K^{-1}\vec{\mu_1} + \frac{1}{2}\vec{\mu_0}^T K^{-1}\vec{\mu_0}\right)\,.\end{aligned} \tag{7.2}$$

The last expression is the *likelihood-ratio*. Usually, we work with the exponent, the so called *log-likelihood-ratio* (LLR)

$$(\vec{\mu_1}-\vec{\mu_0})^T K^{-1}\vec{y} - \frac{1}{2}\vec{\mu_1}^T K^{-1}\vec{\mu_1} + \frac{1}{2}\vec{\mu_0}^T K^{-1}\vec{\mu_0}\,, \tag{7.3}$$

which we denote by $\text{LLR}[X|\vec{Y} = \vec{y}]$.

Thus,

$$\text{MLE}[X|\vec{Y} = \vec{y}] = \begin{cases} 1, & \text{if } \text{LLR}[X|\vec{Y} = \vec{y}] > 0 \\ 0, & \text{otherwise} \end{cases}$$

For $\text{MAP}[X|\vec{Y} = \vec{y}]$ one needs to compare $f_{\vec{Y}|X}(\vec{y}|1)p_1$ with $f_{\vec{Y}|X}(\vec{y}|0)p_0$. Hence,

$$\text{MAP}[X|\vec{Y} = \vec{y}] = \begin{cases} 1, & \text{if } \text{LLR}[X|\vec{Y} = \vec{y}] > \log(p_1/p_0) \\ 0, & \text{otherwise.} \end{cases}$$

Whereas we started with a nonlinear problem, that of minimizing the probability of incorrect detection, we see that we get simple linear rules. To see why this happens, notice that when $K = I$, from (7.2), one chooses $i = 0$ or $i = 1$ depending on which of $\vec{\mu_0}$,$\vec{\mu_1}$ is closest to the observation vector $\vec{y}$, in the sense of Euclidean distance $|\vec{y} - \vec{\mu_i}|$. Hence, the decision rule is specified by the two halfspaces divided by the line of equidistant points from $\vec{\mu_0}, \vec{\mu_1}$.

7.3.1 Sufficient statistic

Notice that even though the dimensionality n of the observations might be high, what is actually sufficient in determining the LLR -and hence MAP and MLE-, is a linear function of Y. Such a function is called a *sufficient statistic*. More precisely, we will call $\vec{g}(\vec{Y})$ a sufficient statistic[1], if

$$f_{\vec{Y}|X}(\vec{y}|x) = F(\vec{g}(\vec{y}), x)G(\vec{y}) ,$$

for some real functions F, G. There can be many sufficient statistics; a trivial one is $\vec{Y}$ itself. In the definition, X, x may be also vectors.

7.3.2 Matched filter

One common way to compute ${\vec{\mu_i}}^T\vec{y}$ needed in the sufficient statistic, is by a *matched filter*. Note that

$${\vec{\mu_i}}^T\vec{y} = \sum_{t=1}^{n} \mu_i(t)y_t = \sum_{t=1}^{n} h_i(n-t)y_t ,$$

where $\vec{\mu_i} = (\mu_i(1), \ldots, \mu_i(n))^T$, and $h_i(t) = \mu_i(n-t)$. Thus, ${\vec{\mu_i}}^T\vec{y}$ can be computed by passing $\vec{y}$ through a filter with an impulse response $h_i(\cdot)$ "matched" to the signal $\vec{\mu_i}$.

7.4 Bit error rate calculations

Assume $K = \sigma^2 I$, for simplicity. The probability of incorrect detection is

$$P(\text{error}) = P\left(\text{LLR}[X|\vec{Y}] > \log\frac{p_0}{p_1}|X=0\right)p_0 + P\left(\text{LLR}[X|\vec{Y}] < \log\frac{p_0}{p_1}|X=1\right)p_1 . \quad (7.4)$$

Hence, we need to determine the distribution of $\text{LLR}(X|\vec{Y})$. $\vec{Y}$ is jointly Gaussian, and

$$\begin{aligned} E[\text{LLR}(\vec{Y})|X=0] &= \frac{(\vec{\mu_1}-\vec{\mu_0})^T\vec{\mu_0}}{\sigma^2} + \frac{{\vec{\mu_0}}^T\vec{\mu_0}}{2\sigma^2} - \frac{{\vec{\mu_1}}^T\vec{\mu_1}}{2\sigma^2} \\ &= \frac{1}{\sigma^2}(\vec{\mu_1}-\vec{\mu_0})^T\vec{\mu_0} - \frac{1}{2\sigma^2}(\vec{\mu_1}+\vec{\mu_0})^T(\vec{\mu_0}-\vec{\mu_1}) \\ &= -\frac{1}{2\sigma^2}\|\vec{\mu_1}-\vec{\mu_0}\|_2^2 \end{aligned}$$

Similarly, we find

$$\text{Var}(\text{LLR}(\vec{Y})|X=0) = \frac{\|\vec{\mu_1}-\vec{\mu_0}\|_2^2}{\sigma^2} .$$

Both the variance and conditional mean depend only on the ratio $\|\vec{\mu_1}-\vec{\mu_0}\|_2/\sigma =: \gamma$.

Thus, (7.4) becomes

$$P(\text{error}) = P\left(N(0,1) \geq \frac{1}{\gamma}\log\frac{p_0}{p_1} + \frac{\gamma}{2}\right)p_0 + P\left(N(0,1) \leq -\frac{1}{\gamma}\log\frac{p_0}{p_1} + \frac{\gamma}{2}\right)p_1 .$$

Observe that BER depends on the energy in the difference of the two signals, $\vec{\mu_1}$ and $\vec{\mu_0}$, and how these compare to the power of the noise σ^2.

[1] This is one of many equivalent definitions, that suffices for our purposes.

When $p_0 = p_1$, the above simplifies to

$$P(\text{error}) = P\left(N(0,1) \geq \frac{\gamma}{2}\right) .$$

Usually one wants to minimize this, under some power constraint. Part of the problem is to determine the signals $\vec{\mu_i}$. One way to do this is to set $\vec{\mu_0} = 0$, i.e. allocate no power to this signal, and transmit any $\vec{\mu_1}$ at full power.

7.5 Example: relation with the scalar case

Assume that a symbol is a bit-string of fixed length n, where each of the n bits is picked independently. Independent noise $W = N(0, I_{n\times n})$ corrupts each bit in a symbol. What is the MAP detector?

In this case, we have 2^n possible symbols in the set $\{0,1\}^{\{1,\ldots,n\}} =: S$, but we don't need to have 2^n matched filters for implementing MAP. One has,

$$\begin{aligned}
\text{MAP}[\vec{X}|\vec{Y} = \vec{y}] &= \arg\max_{\vec{z}\in S} f_{\vec{Y}|\vec{X}}(\vec{y}|\vec{z}) P(\vec{X} = \vec{z}) \\
&= \arg\max_{\vec{z}\in S} \left(\prod_{i=1}^{n} f_{Y_i|X_i}(y_i|z_i)\right)\left(\prod_{j=1}^{n} P(X_j = z_j)\right) \\
&= \arg\max_{\vec{z}\in S} \prod_{i=1}^{n} f_{Y_i|X_i}(y_i|z_i) P(X_i = z_i) ,
\end{aligned}$$

where the second line follows from independence of the X_i's. Thus, $\text{MAP}[\vec{X}|\vec{Y} = \vec{y}]$ will estimate the i-th bit of $\vec{X}$ as $\hat{X}_i = \text{MAP}[X_i|Y_i = y_i]$.

Chapter 8

Hypothesis Testing; Estimation 1

8.1 Summary

Here are the key ideas and results:

- The Hypothesis Testing problem is to maximize $P[Z = 1|X = 1]$ subject to $P[Z = 1|X = 0] \leq \beta$ where Z is based on Y and $f_{Y|X}$ is known.
- The solution of the HT problem is given by the Neyman-Pearson theorem 8.1.
- Some composite HT problems have a simple solution; most do not.
- $g(Y)$ is sufficient for X if X and Y are conditionally independent given $g(Y)$.

8.2 Binary Hypothesis Testing

So far we have considered the Bayesian detection problem of minimizing $E(c(X, g(\mathbf{Y})))$ when both p_X and $f_{\mathbf{Y}|X}$ are known. Recall that, when $X \in \{0, 1\}$ the optimum decision has the form $g(\mathbf{Y}) = h(\Lambda(\mathbf{Y}))$ where $\Lambda(\mathbf{y}) = f_{\mathbf{Y}|X}[\mathbf{y}|1]/f_{\mathbf{Y}|X}[\mathbf{y}|0]$ is the likelihood ratio. In particular, if $\Lambda(\mathbf{y}) = f(k(\mathbf{y}))$, then $k(\mathbf{Y})$ is a sufficient statistic for detecting X given $\mathbf{Y}$.

In this section we explore the detection problem when p_X, the *prior*, is not known. We start with the case when $X \in \{0, 1\}$.

8.2.1 Formulation

Definition 8.1. *Binary Hypothesis Testing Problem*

One is given $f_{Y|X}$. For each observed value y of Y, one chooses $\phi(y) \in [0, 1]$ and one lets $Z = 1$ with probability $\phi(y)$ and $Z = 0$ otherwise. One is also given $\beta \in (0, 1)$. The objective is to choose ϕ to

$$\text{minimize } P[Z = 0 \mid X = 1] \text{ subject to } P[Z = 1 \mid X = 0] \leq \beta. \tag{8.1}$$

One interpretation is that $X = 1$ means that your house is on fire. In that case the problem is to design the alarm system to detect a fire with the largest probability compatible with a probability of false alarm at most equal to β.

8.2.2 Neyman-Pearson Theorem

The key result is the following.

Theorem 8.1. *Neyman-Pearson*
The solution to the binary hypothesis testing problem is as follows:

$$\phi(y) = \begin{cases} 1, & \text{if } \Lambda(y) := \frac{f_{Y|X}[y|1]}{f_{Y|X}[y|0]} > \lambda \\ 0, & \text{if } \Lambda(y) < \lambda \\ \gamma, & \text{if } \Lambda(y) = \lambda \end{cases} \tag{8.2}$$

where $\lambda > 0$ and $\gamma \in [0,1]$ are the only values such that

$$P[Z = 1 \mid X = 0] = \beta \text{ when } P[Z = 1 \mid Y = y] = \phi(y). \tag{8.3}$$

The interpretation of the result is that if $\Lambda(y) > \lambda$, then one is pretty sure that $X = 1$ because the observed value would be much less likely if $X = 0$. In that case, one safely decides $Z = 1$. If $\Lambda(y) < \lambda$, one decides $Z = 0$. If $\lambda(y) = \lambda$, one edges the bet by deciding $Z = 1$ with probability $\phi(y)$. The threshold λ is an adjustment of the 'sensitivity' of the alarm: the lower λ, the most likely the alarm is to sound. The randomization γ is designed to achieve exactly the probability of false alarm.

Before looking at the proof, let us examine two representative examples.

8.2.3 Two Examples

Example 8.1. *We consider the binary symmetric channel. That is, one is given some $\epsilon \in [0, 0.5)$ and, for $x, y \in \{0,1\}$, one has*

$$P[Y = y \mid X = x] = \begin{cases} 1-\epsilon, & \text{if } x = y \\ \epsilon, & \text{if } x \neq y. \end{cases}$$

We find

$$\Lambda(y) = \frac{P[Y = y|X = 1]}{P[Y = y|X = 0]} = \begin{cases} \frac{1-\epsilon}{\epsilon}, & \text{if } y = 1 \\ \frac{\epsilon}{1-\epsilon}, & \text{if } y = 0 \end{cases}$$

Since $\Lambda(1) > \Lambda(0)$, we see that the solution of the hypothesis testing problem must be of the following form:

$$P[Z = 1|Y = 1] = \gamma_1 \text{ and } P[Z = 1|Y = 0] = \gamma_0$$

where γ_0 and γ_1 are chosen in $[0,1]$ so that (23.4) holds. Now,

$$\begin{aligned} P[Z = 1|X = 0] &= P[Z = 1|Y = 1]P[Y = 1|X = 0] + P[Z = 1|Y = 0]P[Y = 0|X = 0] \\ &= \gamma_1\epsilon + \gamma_0(1-\epsilon). \end{aligned}$$

The values of γ_0 and γ_1 depend on β. One finds that

$$(\gamma_0, \gamma_1) = \begin{cases} (0, \frac{\beta}{\epsilon}), & \text{if } \beta \leq \epsilon \\ (\frac{\beta-\epsilon}{1-\epsilon}, 1), & \text{if } \beta > \epsilon. \end{cases}$$

Example 8.2. *In this example, $Y = X + V$ where X and V are independent and $V = N(0,1)$. Here,*

$$\Lambda(y) = \exp\{-\frac{1}{2}(y-1)^2\}/\exp\{-\frac{1}{2}y^2\} = \exp\{y - \frac{1}{2}\}.$$

Since $\Lambda(y)$ is increasing in y, we see that the solution of the hypothesis testing problem has the following form:

$$\phi(y) = \begin{cases} 1, & \text{if } y > y_0 \\ 0, & \text{if } y < y_0 \\ \gamma, & \text{if } y = y_0 \end{cases}$$

where $y_0 = \Lambda^{-1}(\lambda)$. However, since $P[Y = y_0|X = x] = 0$ for $x \in \{0,1\}$, one can ignore the last possibility. Accordingly, the solution is

$$\phi(y) = \begin{cases} 1, & \text{if } y > y_0 \\ 0, & \text{if } y < y_0. \end{cases}$$

The value of y_0 is such that $P[Z = 1|X = 0] = \beta$, i.e.,

$$\beta = P[Y > y_0|X = 0] = P(V > y_0).$$

We now turn to the proof of Theorem 8.1.

8.2.4 Proof of Neyman-Pearson Theorem

Define Z as indicated by the theorem and let V be some random variable that corresponds to another choice $\phi'(y)$ instead of $\phi(y)$. We assume that V satisfies the bound on the false alarm probability, i.e., that $P[V = 1|X = 0] \leq \beta$. We show that $P[V = 1|X = 1] \leq P[Z = 1|X = 1]$. To do this, note that $\Lambda(Y)(Z - V) \geq \lambda(Z - V)$. Hence,

$$E[\Lambda(Y)(Z - V)|X = 0] \geq \lambda E[Z - V|X = 0] \geq 0.$$

But

$$E[\Lambda(Y)(Z - V)|X = 0] = E[Z - V|X = 1].$$

Indeed,

$$\begin{aligned} E[\Lambda(Y)(Z - V)|X = 0] &= \int \Lambda(y)[\phi(y) - \phi'(y)] f_{Y|X}[y|0]dy \\ &= \int \frac{f_{Y|X}[y|1]}{f_{Y|X}[y|0]}[\phi(y) - \phi'(y)] f_{Y|X}[y|0]dy \\ &= \int [\phi(y) - \phi'(y)] f_{Y|X}[y|1]dy = E[Z - V|X = 1]. \end{aligned}$$

Hence,

$$0 \leq E[Z - V|X = 1] = P[Z = 1|X = 1] - P[V = 1|X = 1],$$

as was to be shown.

□

For an intuitive discussion of this result, see [10], page 126.

8.2.5 Important Observations

One remarkable fact is that the optimal decision is again a function of the likelihood ratio. Thus, as in the Bayesian case, if $\Lambda(\mathbf{y}) = f(k(\mathbf{y}))$, the optimal decision for the hypothesis testing problem is a function of the sufficient statistic $k(\mathbf{Y})$.

We used two simple but useful observations in the examples. The first one is that when $\Lambda(y)$ is increasing in y, the decision rule is a threshold on Y. The second is that when Y has a density, there is no need to randomize by introducing some γ.

8.2.6 Composite Hypotheses

We have considered only the case $X \in \{0, 1\}$. The general case would be that X takes values in some set $\mathcal{X}$ and one wishes to determine if $X \in \mathcal{A} \subset \mathcal{X}$ or $X \notin \mathcal{A}$. Say that one selects $Z = 1$ to mean that one thinks that $X \in \mathcal{A}$. One would then attempt to maximize $P[Z = 1|X \in \mathcal{A}]$ subject to the constraint that $P[Z = 1|X \notin \mathcal{A}] \leq \beta$. This general problem does not admit a simple answer. For instance, the likelihood ratio $P[Y = y|X \in \mathcal{A}]/P[Y = y|X \notin \mathcal{A}]$ is not defined since we do not have a prior distribution of X. However, some problems have a simple answer. We give one example.

Example 8.3. *One knows that, given X, $Y = N(X, 1)$ and we want to determine whether $X = \mu_0$ or $X > \mu_0$. First note that if the alternatives are $X = \mu_0$ or $X = \mu_1 > \mu_0$, then the optimal decision would be $Z = 1\{Y > y_0\}$ with $P[Y > y_0|X = \mu_0] = \beta$. Thus, the value of y_0 does not depend on μ_1. It follows that this decision rule is optimal for the composite problem.*

8.3 Conditional Independence and Sufficient Statistics

The notion of conditional independence is useful to clarify the concept of sufficient statistics.

Definition 8.2. *Conditional Independence*

The random variables X and Z are conditionally independent *given the random variable Y if*

$$P[X \in A, Z \in B|Y] = P[X \in A|Y]P[Z \in B|Y], \forall A, B. \tag{8.4}$$

Note the following simple fact:

Fact 8.1. *If X and Z are conditionally independent given Y, then $g(X)$ and $h(Z)$ are conditionally independent given Y, for any functions $g(\cdot)$ and $h(\cdot)$.*

The following fact is a direct consequence of the definition.

Fact 8.2. *X and Z are conditionally independent given Y if and only if*

$$E[g(X)h(Z)|Y] = E[g(X)|Y]E[h(Z)|Y], \forall g(\cdot), h(\cdot). \tag{8.5}$$

Proof:

The result is immediate from (8.4) if $g(X) = \sum_i a_i 1\{X \in A_i\}$ and $h(Z) = \sum_j b_j 1\{Z \in B_j\}$. The general case follows by approximation.

□

From this property, one finds the following result.

Fact 8.3. *Assume that X and Z are conditionally independent given Y. Then*

$$E[X|Y,Z] = E[X|Y].$$

Proof:

We must show that

$$E(f(Y,Z)E[X|Y]) = E(f(Y,Z)X), \forall f(\cdot).$$

Assume that $f(Y,Z) = g(Y)h(Z)$. Then

$$\begin{aligned} E(f(Y,Z)E[X|Y]) &= E(g(Y)h(Z)E[X|Y]) = E(E[g(Y)h(Z)E[X|Y]|Y]) \\ &= E(E[g(Y)h(Z)X|Y]), \text{ by (8.5)} \\ &= E(f(Y,Z)X). \end{aligned}$$

The general case follows by linearity and approximation. For instance, approximate $f(Y,Z)$ by polynomials.

□

Here is a different twist on the notion of sufficient statistic.

Definition 8.3. *Sufficient Statistic*

$g(Y)$ is a sufficient statistic for X if X and Y are conditionally independent given $g(Y)$.

Let us revisit the detection problem in the light of this definition. Recall that

$$MAP[X|Y=y] = \text{argmax}_x P[X=x|Y=y].$$

Now, with $z = g(y)$,

$$P[X=x|Y=y] = P[X=x|Y=y, g(Y)=z] = P[X=x|g(Y)=z],$$

so that

$$MAP[X|X=y] = h(z).$$

The connection with likelihood ratios is as follows. Assume that X and Y are conditionally independent given $g(Y)$. Then, with $z = g(y)$,

$$f_{Y|X}[y|x] = f_{g(Y)|X}[z|x] f_{Y|g(Y)}[y|z],$$

so that

$$\Lambda(x;y) := \frac{f_{Y|X}[y|x]}{f_{Y|X}[y|x_0]} = \frac{f_{g(Y)|X}[z|x]}{f_{g(Y)|X}[z|x_0]} = \phi(x;z).$$

It follows that the solution of the HT problem, the MLE, and the MAP are all functions of the sufficient statistic.

8.4 Solved Problems

Problem 8.1. *When $X=0$, $Y_1, \ldots, Y_n$ are i.i.d. Bernoulli r.v's with $P[Y_m = 1|X=0] = p_0$. When $X=1$, we have $P[Y_m = 1|X=1] = p_1$, instead. ($1 > p_1 > p_0 > 0$.)*

Find the Neyman-Pearson test, to minimize $P[Z=0|X=1]$ subject to $P[Z=1|X=0] \leq \beta$.

What was the sufficient statistic? Explain.

Chapter 9

MMSE and LLSE

9.1 Summary

Here are the key ideas and results:

- The MMSE of $\mathbf{X}$ given $\mathbf{Y}$ is $E[\mathbf{X}|\mathbf{Y}]$.
- The LLSE of $\mathbf{X}$ given $\mathbf{Y}$ is $L[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X}) + K_{\mathbf{X},\mathbf{Y}} K_{\mathbf{Y}}^{-1}(\mathbf{Y} - E(\mathbf{Y}))$ if $K_{\mathbf{Y}}$ is nonsingular.
- Theorem 9.3 states the properties of the LLSE.
- The linear regression approximates the LLSE when the samples are realizations of i.i.d. random pairs (X_m, Y_m).

9.2 Estimation: Formulation

The estimation problem is a generalized version of the Bayesian decision problem where the set of values of $\mathbf{X}$ can be $\Re^n$. In applications, one is given a *source model* $f_{\mathbf{X}}$ and a *channel model* $f_{\mathbf{Y}|\mathbf{X}}$ that together define $f_{\mathbf{X},\mathbf{Y}}$. One may have to estimate $f_{\mathbf{Y}|\mathbf{X}}$ by using a *training sequence*, i.e., by selecting the values of $\mathbf{X}$ and observing the channel outputs. Alternatively, one may be able to observe a sequence of values of $(\mathbf{X}, \mathbf{Y})$ and use them to estimate $f_{\mathbf{X},\mathbf{Y}}$.

Definition 9.1. *Estimation Problems*

One is given the joint distribution of $(\mathbf{X}, \mathbf{Y})$. *The estimation problem is to calculate* $\mathbf{Z} = g(\mathbf{Y})$ *to minimize* $E(c(\mathbf{X}, \mathbf{Z}))$ *for a given function* $c(\cdot, \cdot)$. *The random variable* $\mathbf{Z} = g(\mathbf{Y})$ *that minimizes* $E(||\mathbf{X} - \mathbf{Z}||^2)$ *is the* minimum mean squares estimator *(MMSE) of* $\mathbf{X}$ *given* $\mathbf{Y}$. *The random variable* $\mathbf{Z} = A\mathbf{Y} + \mathbf{b}$ *that minimizes* $E(||\mathbf{X} - \mathbf{Z}||^2)$ *is called the* linear least squares estimator *(LLSE) of* $\mathbf{X}$ *given* $\mathbf{Y}$*; we designate it by* $\mathbf{Z} = L[\mathbf{X}|\mathbf{Y}]$.

9.3 MMSE

Here is the central result about minimum mean squares estimation. You have seen this before, but we recall the proof of that important result.

Theorem 9.1. *MMSE*

The MMSE of $\mathbf{X}$ *given* $\mathbf{Y}$ *is* $E[\mathbf{X}|\mathbf{Y}]$.

Proof: You should recall the definition of $E[\mathbf{X}|\mathbf{Y}]$, a random variable that has the property

$$E[(\mathbf{X} - E[\mathbf{X}|\mathbf{Y}])h_1(\mathbf{Y})] = 0, \forall h_1(\cdot), \tag{9.1}$$

or equivalently

$$E[h_2(\mathbf{Y})(\mathbf{X} - E[\mathbf{X}|\mathbf{Y}])] = 0, \forall h_2(\cdot). \tag{9.2}$$

By $E(\mathbf{X}) = E[E[\mathbf{X}|\mathbf{Y}]]$, the interpretation is that $\mathbf{X} - E[\mathbf{X}|\mathbf{Y}] \perp h(\mathbf{Y})$ for all $h(\cdot)$. By Pythagoras, one then expect $E[\mathbf{X}|\mathbf{Y}]$ to be the function of $\mathbf{Y}$ that is closest to $\mathbf{X}$, as illustrated in Figure 9.1.

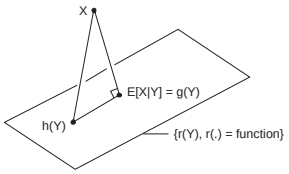


Figure 9.1: The conditional expectation as a projection.

Formally, let $h(\mathbf{Y})$ be an arbitrary function. Then

$$\begin{aligned} E(||\mathbf{X} - h(\mathbf{Y})||^2) &= E(||\mathbf{X} - E[\mathbf{X}|\mathbf{Y}] + E[\mathbf{X}|\mathbf{Y}] - h(\mathbf{Y})||^2) \\ &= E(||\mathbf{X} - E[\mathbf{X}|\mathbf{Y}]||^2) + 2E((E[\mathbf{X}|\mathbf{Y}] - h(\mathbf{Y}))^T(\mathbf{X} - E[\mathbf{X}|\mathbf{Y}])) + E(||E[\mathbf{X}|\mathbf{Y}] - h(\mathbf{Y})||^2) \\ &= E(||\mathbf{X} - E[\mathbf{X}|\mathbf{Y}]||^2) + E(||E[\mathbf{X}|\mathbf{Y}] - h(\mathbf{Y})||^2) \geq E(||\mathbf{X} - E[\mathbf{X}|\mathbf{Y}]||^2). \end{aligned}$$

In this derivation, the third identity follows from the fact that the cross-term vanishes in view of (9.2). Note also that this derivation corresponds to the fact that the triangle $\{\mathbf{X}, E[\mathbf{X}|\mathbf{Y}], h(\mathbf{Y})\}$ is a right triangle with hypothenuse $(\mathbf{X}, h(\mathbf{Y}))$. □

Example 9.1. *You recall that if* $(\mathbf{X}, \mathbf{Y})$ *are jointly Gaussian with* $K_{\mathbf{Y}}$ *nonsingular, then*

$$E[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X}) + K_{\mathbf{X},\mathbf{Y}}K_{\mathbf{Y}}^{-1}(\mathbf{Y} - E(\mathbf{Y})).$$

You also recall what happens when $K_{\mathbf{Y}}$ *is singular. The key is to find* A *so that* $K_{\mathbf{X},\mathbf{Y}} = AK_{\mathbf{Y}} = AQ\Lambda Q^T$. *By writing*

$$Q = [Q_1|Q_2], \Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix}$$

where Λ_1 *corresponds the part of nonzero eigenvalues of* $K_{\mathbf{Y}}$, *one finds*

$$K_{\mathbf{X},\mathbf{Y}} = AQ\Lambda Q^T = A[Q_1|Q_2]\begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} = AQ_1\Lambda_1 Q_1^T,$$

so that $K_{\mathbf{X},\mathbf{Y}} = AK_{\mathbf{Y}}$ *with* $A = K_{\mathbf{X},\mathbf{Y}}Q_1\Lambda_1^{-1}Q_1^T$.

9.4 LLSE

Theorem 9.2. *LLSE*

$L[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X}) + K_{\mathbf{X},\mathbf{Y}}K_{\mathbf{Y}}^{-1}(\mathbf{Y} - E(\mathbf{Y}))$ *if* $K_{\mathbf{Y}}$ *is nonsingular.* $L[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X}) + K_{\mathbf{X},\mathbf{Y}}Q_1\Lambda_1^{-1}Q_1^T(\mathbf{Y} - E(\mathbf{Y}))$ *if* $K_{\mathbf{Y}}$ *is singular.*

Proof: $\mathbf{Z} = L[\mathbf{X}|\mathbf{Y}] = A\mathbf{Y} + \mathbf{b}$ satisfies $\mathbf{X} - \mathbf{Z} \perp B\mathbf{Y} + \mathbf{d}$ for any B and $\mathbf{d}$ with $E(\mathbf{X}) = E[L[\mathbf{X}|\mathbf{Y}]]$. If we consider any $\mathbf{Z}' = C\mathbf{Y} + \mathbf{c}$, then $E((\mathbf{Z} - \mathbf{Z}')^T(\mathbf{X} - \mathbf{Z})) = 0$ since $\mathbf{Z} - \mathbf{Z}' = A\mathbf{Y} + \mathbf{b} - C\mathbf{Y} - \mathbf{c} = (A - C)\mathbf{Y} + (\mathbf{b} - \mathbf{c}) = B\mathbf{Y} + \mathbf{d}$. It follows that

$$\begin{aligned} E(||\mathbf{X} - \mathbf{Z}'||^2) &= E(||\mathbf{X} - \mathbf{Z} + \mathbf{Z} - \mathbf{Z}'||^2) \\ &= E(||\mathbf{X} - \mathbf{Z}||^2) + 2E((\mathbf{Z} - \mathbf{Z}')^T(\mathbf{X} - \mathbf{Z})) + E(||\mathbf{Z} - \mathbf{Z}'||^2) \\ &= E(||\mathbf{X} - \mathbf{Z}||^2) + E(||\mathbf{Z} - \mathbf{Z}'||^2) \geq E(||\mathbf{X} - \mathbf{Z}||^2). \end{aligned}$$

Figure 9.2 illustrates this calculation. It shows that $L[\mathbf{X}|\mathbf{Y}]$ is the projection of $\mathbf{X}$ on the set of linear functions of $\mathbf{Y}$. The projection $\mathbf{Z}$ is characterized by the property that $\mathbf{X} - \mathbf{Z}$ is orthogonal to all the linear functions $B\mathbf{Y} + \mathbf{d}$. This property holds if and only if (from (9.1))

$$E(\mathbf{X} - \mathbf{Z}) = 0 \text{ and } E((\mathbf{X} - \mathbf{Z})\mathbf{Y}^T) = 0,$$

which are equivalent to $E(\mathbf{X}-\mathbf{Z}) = 0$ and $\text{cov}(\mathbf{X}-\mathbf{Z}, \mathbf{Y}) = 0$. The comparison between $L[\mathbf{X}|\mathbf{Y}]$ and $E[\mathbf{X}|\mathbf{Y}]$ is shown in Figure 9.3. □

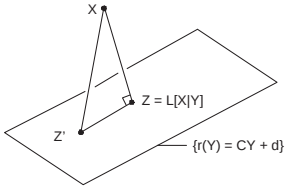


Figure 9.2: The LLSE as a projection.

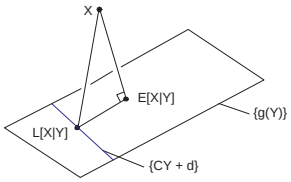


Figure 9.3: LLSE vs. MMSE.

The following result indicates some properties of the LLSE. They are easy to prove.

Theorem 9.3. *Properties of LLSE*

(a) $L[A\mathbf{X}_1 + B\mathbf{X}_2|\mathbf{Y}] = AL[\mathbf{X}_1|\mathbf{Y}] + BL[\mathbf{X}_2|\mathbf{Y}]$.

(b) $L[L[\mathbf{X}|\mathbf{Y}, \mathbf{Z}] \mid \mathbf{Y}] = L[\mathbf{X}|\mathbf{Y}]$.

(c) If $\mathbf{X} \perp \mathbf{Y}$*, then* $L[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X})$.

(d) Assume that X *and* Z *are conditionally independent given* Y*. Then, in general,* $L[X|Y, Z] \neq L[X|Y]$.

Proof: (b): It suffices to show that

$$E(L[\mathbf{X}|\mathbf{Y},\mathbf{Z}] - L[\mathbf{X}|\mathbf{Y}]) = 0 \text{ and } E((L[\mathbf{X}|\mathbf{Y}] - L[\mathbf{X}|\mathbf{Y},\mathbf{Z}])\mathbf{Y}^T) = 0.$$

We already have $E(X - L[\mathbf{X}|\mathbf{Y},\mathbf{Z}]) = 0$ and $E(X - L[\mathbf{X}|\mathbf{Y}]) = 0$, which implies $E(L[\mathbf{X}|\mathbf{Y},\mathbf{Z}] - L[\mathbf{X}|\mathbf{Y}]) = 0$. Moreover, from $E((\mathbf{X} - L[\mathbf{X}|\mathbf{Y},\mathbf{Z}])\mathbf{Y}^T) = 0$ and $E((\mathbf{X} - L[\mathbf{X}|\mathbf{Y}])\mathbf{Y}^T) = 0$, it is obtained $E((L[\mathbf{X}|\mathbf{Y}] - L[\mathbf{X}|\mathbf{Y},\mathbf{Z}])\mathbf{Y}^T) = 0$.

(d): Here is a trivial example. Assume that $X = Z = Y^{1/2}$ where X is $U[0,1]$. Then $L[X|Y,Z] = Z$. However, since $Y = X^2$, $Y^2 = X^4$, and $XY = X^3$, we find

$$\begin{aligned} L[X|Y] &= E(X) + \frac{E(XY) - E(X)E(Y)}{E(Y^2) - E(Y)^2}(Y - E(Y)) \\ &= \frac{1}{2} + \left(\frac{1}{4} - \frac{1}{2}\cdot\frac{1}{3}\right)\left(\frac{1}{5} - \frac{1}{9}\right)^{-1}\left(Y - \frac{1}{3}\right) \\ &= \frac{3}{16} + \frac{15}{16}Y. \end{aligned}$$

We see that $L[X|Y,Z] \neq L[X|Y]$ because

$$Z \neq \frac{3}{16} + \frac{15}{16}Y = \frac{3}{16} + \frac{15}{16}Z^2.$$

□

Figure 9.4 illustrates property (b) which is called the *smoothing property* of LLSE.

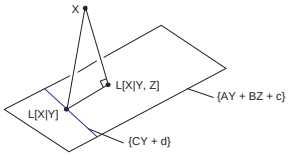


Figure 9.4: Smoothing property of LLSE.

9.5 Examples

We illustrate the previous ideas on a few examples.

Example 9.2. *Assume that U, V, W are independent and $U[0,1]$. Calculate $L[(U+V)^2 \mid V^2 + W^2]$.*

Solution: Let $X = (U+V)^2$ and $Y = V^2 + W^2$. Then

$$K_{X,Y} = \text{cov}(U^2 + 2UV + V^2, V^2 + W^2) = 2\text{cov}(UV, V^2) + \text{cov}(V^2, V^2).$$

Now,

$$\text{cov}(UV, V^2) = E(UV^3) - E(UV)E(V^2) = \frac{1}{8} - \frac{1}{12} = \frac{1}{24}$$

and

$$\text{cov}(V^2, V^2) = E(V^4) - E(V^2)E(V^2) = \frac{1}{5} - \frac{1}{9} = \frac{4}{45}.$$

Hence,

$$K_{X,Y} = \frac{1}{12} + \frac{4}{45} = \frac{31}{180}.$$

We conclude that

$$L[(U+V)^2 \mid V^2+W^2] = E((U+V)^2) + K_{X,Y}K_Y^{-1}(Y - E(Y)) = \frac{7}{6} + \frac{31}{32}\left(Y - \frac{2}{3}\right).$$

Example 9.3. *Let U, V, W be as in the previous example. Calculate $L[\cos(U+V)|V+W]$.*

Solution: Let $X = \cos(U+V)$ and $Y = V+W$. We find

$$K_{X,Y} = E(XY) - E(X)E(Y).$$

Now,

$$E(XY) = E(V\cos(U+V)) + \frac{1}{2}E(X).$$

Also,

$$\begin{aligned}
E(V\cos(U+V)) &= \int_0^1\int_0^1 v\cos(u+v)dudv = \int_0^1 [v\sin(u+v)]_0^1 dv \\
&= \int_0^1 (v\sin(v+1) - v\sin(v))dv = -\int_0^1 v\cdot d\cos(v+1) + \int_0^1 v\cdot d\cos(v) \\
&= -[v\cos(v+1)]_0^1 + \int_0^1 \cos(v+1)dv + [v\cos(v)]_0^1 - \int_0^1 \cos(v)dv \\
&= -\cos(2) + [\sin(v+1)]_0^1 + \cos(1) - [\sin(v)]_0^1 \\
&= -\cos(2) + \sin(2) - \sin(1) + \cos(1) - \sin(1).
\end{aligned}$$

Moreover,

$$\begin{aligned}
E(X) &= \int_0^1\int_0^1 \cos(u+v)dudv = \int_0^1 [\sin(u+v)]_0^1 du = \int_0^1 (\sin(u+1) - \sin(u))du \\
&= -[\cos(u+1)]_0^1 + [\cos(u)]_0^1 = -\cos(2) + \cos(1) + \cos(1) - \cos(0) = -\cos(2) + 2\cos(1) - 1.
\end{aligned}$$

In addition,

$$E(Y) = E(V) + E(U) = 1.$$

and

$$K_Y = \frac{1}{6}.$$

9.6 Linear Regression vs. LLSE

We discuss the connection between the familiar *linear regression* procedure and the LLSE.

One is given a set of n pairs of numbers $\{(x_m, y_m), m = 1, \ldots, n\}$, as shown in Figure 9.5. One draws a line through the points. That line approximates the values y_m by $z_m = \alpha x_m + \beta$. The line is chosen to minimize

$$\sum_{m=1}^n (z_m - y_m)^2 = \sum_{m=1}^n (\alpha x_m + \beta - y_m)^2. \tag{9.3}$$

That is, the linear regression is the linear approximation that minimizes the sum of the squared errors. Note that there is no probabilistic framework in this procedure.

To find the values of α and β that minimize the sum of the squared errors, one differentiates (9.3) with respect to α and β and sets the derivatives to zero. One finds

$$\sum_{m=1}^{n}(\alpha x_m + \beta - y_m) = 0 \text{ and } \sum_{m=1}^{n} x_m(\alpha x_m + \beta - y_m) = 0.$$

Solving these equations, we find

$$\alpha = \frac{A(xy) - A(x)A(y)}{A(x^2) - A(x)^2} \text{ and } \beta = A(y) - \alpha A(x).$$

In these expressions, we used the following notation:

$$A(y) := \frac{1}{n}\sum_{m=1}^{n} y_m, A(x) := \frac{1}{n}\sum_{m=1}^{n} x_m, A(xy) := \frac{1}{n}\sum_{m=1}^{n} x_m y_m, \text{ and } A(x^2) := \frac{1}{n}\sum_{m=1}^{n} x_m^2.$$

Thus, the point y_m is approximated by

$$z_m = A(y) + \frac{A(xy) - A(x)A(y)}{A(x^2) - A(x)^2}(x_m - A(x)).$$

Note that if the pairs (x_m, y_m) are realizations of i.i.d. random variables $(X_m, Y_m) =_D (X, Y)$ with finite variances, then, as $n \to \infty$, $A(x) \to E(X)$, $A(x^2) \to E(X^2)$, $A(y) \to E(Y)$, and $A(xy) \to E(XY)$. Consequently, if n is large, we see that

$$z_m \approx A(Y) + \frac{\text{cov}(X, Y)}{\text{var}(X)}(x_m - E(X)) = L[Y|X = x_m].$$

See [4] for examples.

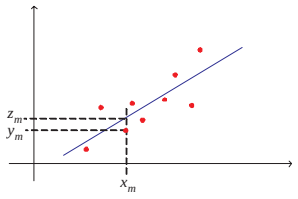


Figure 9.5: Linear Regression.

Chapter 10

Kalman Filter - 1

10.1 Summary

Here are the key ideas and results:

- We are looking for recursive estimators.
- One key idea is that if $E(\mathbf{X}) = E(\mathbf{Y}) = E(\mathbf{Z}) = 0$ and if $\mathbf{Y} \perp \mathbf{Z}$, then $L[\mathbf{X}|\mathbf{Y},\mathbf{Z}] = L[\mathbf{X}|\mathbf{Y}] + L[\mathbf{X}|\mathbf{Z}]$.
- An application is Kalman Filter.

10.2 Updating a LLSE

In this section we derive how to update a LLSE. Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be three random vectors. Can we express $L[\mathbf{X}|\mathbf{Y},\mathbf{Z}]$ in terms of $L[\mathbf{X}|\mathbf{Y}]$ and $\mathbf{Z}$? Here is the key result.

Theorem 10.1. *Assume that the random vectors $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ are zero-mean and such that $\mathbf{Y} \perp \mathbf{Z}$. Then*

$$L[\mathbf{X}|\mathbf{Y},\mathbf{Z}] = K_{\mathbf{X},\mathbf{Y}} K_{\mathbf{Y}}^{-1} \mathbf{Y} + K_{\mathbf{X},\mathbf{Z}} K_{\mathbf{Z}}^{-1} \mathbf{Z} = L[\mathbf{X}|\mathbf{Y}] + L[\mathbf{X}|\mathbf{Z}]. \tag{10.1}$$

Moreover,

$$cov(\mathbf{X} - L[\mathbf{X}|\mathbf{Y},\mathbf{Z}]) = cov(\mathbf{X} - L[\mathbf{X}|\mathbf{Y}]) - K_{\mathbf{X},\mathbf{Z}} K_{\mathbf{Z}}^{-1} K_{\mathbf{Z},\mathbf{X}}. \tag{10.2}$$

Proof:

If you look at Figure 10.1, you should see (10.1).

To verify (10.2), we do the algebra:

$$\begin{aligned}
\text{cov}(\mathbf{X} - L[\mathbf{X}|\mathbf{Y},\mathbf{Z}]) &= \text{cov}(\mathbf{X} - K_{\mathbf{X},\mathbf{Y}} K_{\mathbf{Y}}^{-1} \mathbf{Y} - K_{\mathbf{X},\mathbf{Z}} K_{\mathbf{Z}}^{-1} \mathbf{Z}) \\
&= \text{cov}(\mathbf{X} - K_{\mathbf{X},\mathbf{Y}} K_{\mathbf{Y}}^{-1} \mathbf{Y}) + \text{cov}(K_{\mathbf{X},\mathbf{Z}} K_{\mathbf{Z}}^{-1} \mathbf{Z}) - 2\text{cov}(\mathbf{X} - K_{\mathbf{X},\mathbf{Y}} K_{\mathbf{Y}}^{-1} \mathbf{Y}, K_{\mathbf{X},\mathbf{Z}} K_{\mathbf{Z}}^{-1} \mathbf{Z}) \\
&= \text{cov}(\mathbf{X} - L[\mathbf{X}|\mathbf{Y}]) + K_{\mathbf{X},\mathbf{Z}} K_{\mathbf{Z}}^{-1} K_{\mathbf{Z},\mathbf{X}} - 2K_{\mathbf{X},\mathbf{Z}} K_{\mathbf{Z}}^{-1} K_{\mathbf{Z},\mathbf{X}} \\
&= \text{cov}(\mathbf{X} - L[\mathbf{X}|\mathbf{Y}]) - K_{\mathbf{X},\mathbf{Z}} K_{\mathbf{Z}}^{-1} K_{\mathbf{Z},\mathbf{X}},
\end{aligned}$$

as was to be shown. □

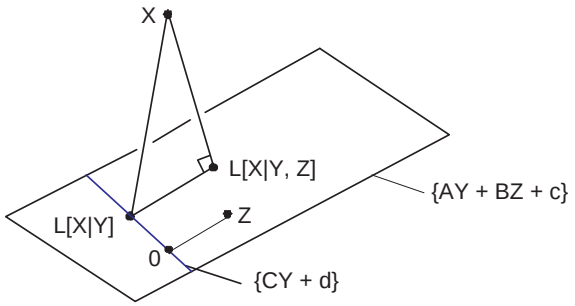


Figure 10.1: Updating an LLSE with additional uncorrelated observations.

10.3 Kalman Filter

The setup is that X_n describes the state of a system whose dynamics are linear. One observes Y_n, a noisy version of a linear observation of X_n. The problem is to calculate recursively $\hat{X}_n := L[X_n|Y_1, \ldots, Y_n]$. By *recursively*, we mean that $\hat{X}_{n+1}$ should be a function of $\hat{X}_n$ and Y_{n+1}. The key idea is then to update the LLSE $\hat{X}_n$ with the new observation Y_{n+1}.

For ease of notation, we consider a linear system with matrices that do not depend on time. That is,

$$\mathbf{X}_{n+1} = A\mathbf{X}_n + \mathbf{V}_n \text{ and } \mathbf{Y}_n = C\mathbf{X}_n + \mathbf{W}_n, n \geq 1,$$

where $\{\mathbf{X}_1, \mathbf{V}_n, \mathbf{W}_n, n \geq 1\}$ are all orthogonal and zero-mean with $\text{cov}(\mathbf{V}_n) = K_{\mathbf{V}}$ and $\text{cov}(\mathbf{W}_n) = K_{\mathbf{W}}$.

Theorem 10.2. *Kalman Filter*

$$\hat{X}_n = L[X_n|Y_1, \ldots, Y_n] = L[X_n|Y^n]$$

is obtained by the Kalman Filter:

$$\hat{X}_n = A\hat{X}_{n-1} + R_n(Y_n - CA\hat{X}_{n-1})$$

where

$$\begin{aligned} R_n &= S_n C^T [CS_n C^T + K_W]^{-1}, \\ S_n &= A\Sigma_{n-1}A^T + K_V, \\ \Sigma_n &= (I - R_n C)S_n. \end{aligned}$$

Thus,

$$S_{n+1} = K_V + AS_nA^T - AS_nC^T[CS_nC^T + K_W]^{-1}CS_nA^T.$$

Moreover, the matrices S_n and Σ_n have the following significance:

$$\begin{aligned} S_n &= cov(X_n - A\hat{X}_{n-1}), \\ \Sigma_n &= cov(X_n - \hat{X}_n). \end{aligned}$$

Proof:

Let

$$U_n = Y_n - L[Y_n|Y^{n-1}] = Y_n - L[CX_n + W_n|Y^{n-1}] = Y_n - CL[X_n|Y^{n-1}] = Y_n - CA\hat{X}_{n-1} = CX_n + W_n - CL[X_n|Y^{n-1}]$$

Then, by (10.1),

$$\hat{X}_n = L[X_n|Y^n] = L[X_n|Y^{n-1}] + L[X_n|U_n] = A\hat{X}_{n-1} + R_n U_n = A\hat{X}_{n-1} + R_n(Y_n - CA\hat{X}_{n-1})$$

where

$$R_n = \text{cov}(X_n, U_n)\text{cov}(U_n)^{-1}.$$

Now,

$$\begin{aligned}\text{cov}(X_n, U_n) &= \text{cov}(X_n, CX_n + W_n - CL[X_n|Y^{n-1}]) \\ &= \text{cov}(X_n - L[X_n|Y^{n-1}], CX_n + W_n - CL[X_n|Y^{n-1}]) = S_n C^T.\end{aligned}$$

Also,

$$\text{cov}(U_n) = \text{cov}(CX_n + W_n - CL[X_n|Y^{n-1}]) = CS_nC^T + K_W.$$

Thus,

$$R_n = S_nC^T[CS_nC^T + K_W]^{-1}.$$

In addition,

$$S_n = \text{cov}(X_n - L[X_n|Y^{n-1}]) = \text{cov}(AX_{n-1} + V_{n-1} - A\hat{X}_{n-1}) = A\Sigma_{n-1}A^T + K_V.$$

Finally, by (10.2),

$$\Sigma_n = \text{cov}(X_n - L[X_n|Y^n]) = \text{cov}(X_n - L[X_n|Y^{n-1}]) - \text{cov}(X_n, U_n)\text{cov}(U_n)^{-1}\text{cov}(U_n, X_n) = S_n - R_nCS_n.$$

□

See [8] for a more detailed discussion.

10.4 Solved Problems

Problem 10.1. *Give an example of linear system $X_{n+1} = AX_n + V_n, Y_n = CX_n + W_n$, where*

$$Cov(X_n - L[X_n|Y_0, \ldots, Y_{n-1}]) \to \infty, \quad \text{as } n \to \infty\,.$$

Solution:
Let $X_{n+1} = X_n + V_n, Y_n = W_n$, for i.i.d. $N(0,1)$ noise (V_n) and (W_n). Also, let $X_0 = N(0,1)$.
Then,

$$\text{Cov}(X_n - L[X_n|Y^{n-1}]) = \text{Var}(X_n - E(X_n)) = n\,.$$

Problem 10.2. *In the setting of the previous problem, give an example where*

$$Cov(X_n - L[X_n|Y_0, \ldots, Y_{n-1}]) \to 0, \quad \text{as } n \to \infty\,.$$

Solution:
Consider $X_{n+1} = X_n, Y_n = X_n + W_n$, where noise is as in the previous problem.
Then,

$$\begin{aligned}\text{Cov}(X_n - L[X_n|Y^{n-1}]) &\le \text{Cov}\left(X_n - \frac{1}{n}\sum_{i=1}^n Y_i\right) \\ &= \text{Cov}\left(\frac{1}{n}\sum_{i=1}^n W_i\right) = \frac{1}{n}\end{aligned}$$

Problem 10.3. *Give an example of (A, B) not reachable, with state space $\mathbb{R}^2$.*

Solution:
$A = I_2, B = (1, 0)^T$. Then,

$$\text{rank}[B \quad AB] = \text{rank} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = 1 < 2 \, .$$

Problem 10.4. *Give an example of (A, C) not observable, with state space $\mathbb{R}^2$.*

Solution:
$A = I_2, C = (1, 0)$. Then,

$$\text{rank}[C^T \quad A^T C^T] = \text{rank} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = 1 < 2 \, .$$

Problem 10.5. *Let X, V, W be r.v's with $V \perp W$. Show that $Var(X + V - L[X + V|X + W])$ is increasing in $Var(X)$.*

Solution:
Without loss of generality, we can assume that (X, V, W) is JG. (E.g., take (X_0, V_0, W_0) JG with the same mean and covariance matrix as (X, V, W).) Let $\xi = N(0, \sigma^2)$, independent of V, W, and define $X' = X + \xi$. Then, $\text{Var}(X') > \text{Var}(X)$, and

$$\begin{aligned} \text{Var}(X' + V - L[X' + V|X' + W]) &\geq \text{Var}(X' + V - L[X' + V|X' + W, \xi]) \\ &= \text{Var}(X + \xi + V - L[X + V|X + W, \xi] - \xi) \\ &= \text{Var}(X + V - L[X + V|X + W]) \, . \end{aligned}$$

Problem 10.6. *Consider the (scalar) system,*

$$X_{n+1} = a_n X_n + V_n \, ,$$

where $(a_0, a_1, \ldots)$ are i.i.d. $N(\mu, \sigma^2)$, independent of $(X_0, V_0, V_1, \ldots)$. As usual, assume $(V_0, V_1, \ldots)$ are i.i.d. independent of X_0.

Find conditions on μ, σ^2 so that $Cov(X_n) \to K$ as $n \to \infty$, for some constant $K < \infty$.

Solution:
Assume $V_n = N(0, K_V)$ for all n, $E(X_0) = 0$.

$$\begin{aligned} \text{Var}(X_{n+1}) &= \text{Var}(a_n X_n + V_n) \\ &= \text{Var}(a_n X_n) + K_V \\ &= E(a_n^2)\text{Var}(X_n) + K_V \\ &= (\sigma^2 + \mu^2)\text{Var}(X_n) + K_V \, . \end{aligned}$$

Thus, $\text{Var}(X_n)$ converges if $\sigma^2 + \mu^2 < 1$, diverges for $\sigma^2 + \mu^2 > 1$. If $\sigma^2 + \mu^2 = 1$ it converges if and only if $K_V = 0$.

Problem 10.7. *Consider the (scalar) system,*

$$\begin{aligned} X_{n+1} &= aX_n + V_n \\ Y_n &= c_n X_n + W_n \, , \end{aligned}$$

under the usual independence assumptions. Further, assume $c_0, c_1, \ldots$ are i.i.d. with mean λ and variance u^2, independent of the noise and X_0.

Derive recursive equations for $\hat{X}_n = L[X_n|Y_0, \ldots, Y_n]$.

Solution:
$\hat{X}_{n+1} = L[X_{n+1}|Y^n, Y_{n+1}] = a\hat{X}_n + L[X_{n+1}|U]$, where $U = Y_{n+1} - E(Y_{n+1}|Y^n)$. Now,

$$\begin{aligned} U &= Y_{n+1} - E(Y_{n+1}|Y^n) \\ &= c_{n+1}aX_n + c_{n+1}V_n + W_{n+1} - a\lambda\hat{X}_n \,, \end{aligned}$$

so $\hat{X}_{n+1} = a\hat{X}_n + R_n(Y_{n+1} - a\lambda\hat{X}_n)$, where $R_n = E(X_{n+1}U)/E(U^2)$. Now,

$$\begin{aligned} E(U^2) &= a^2\mathrm{Var}(c_{n+1}X_n - \lambda X_n) + \lambda^2 a^2 \mathrm{Var}(X_n - \hat{X}_n) + E(c_{n+1}^2 V_n^2) \\ &= a^2u^2E(X_n^2) + \lambda^2a^2\Sigma_n + (\lambda^2 + u^2)K_V + K_W \\ &= \lambda^2 S_n + u^2(K_V + a^2E(X_n^2)) + K_W \,, \end{aligned}$$

where we defined $\Sigma_n = E((X_n - \hat{X}_n)^2)$, $S_n = a^2\Sigma_n + K_V$. In addition,

$$\begin{aligned} E(X_{n+1}U) &= E((aX_n + V_n)(c_{n+1}aX_n + c_{n+1}V_n + W_{n+1} - a\lambda\hat{X}_n)) \\ &= a^2E(X_n((c_{n+1}X_n - \lambda X_n)) + a^2\lambda E(X_n(X_n - \hat{X}_n)) + \lambda K_V \\ &= a^2\lambda\Sigma_n + \lambda K_V = S_n\lambda \,. \end{aligned}$$

Finally,

$$\begin{aligned} \Sigma_{n+1} &= E((aX_n + V_n - a\hat{X}_n - E_nU)^2) \\ &= a^2\Sigma_n + K_V + R_n^2E(U^2) - 2aR_nE(U(X_n - \hat{X}_n)) - 2R_nE(V_nU) \\ &= (1 - \lambda R_n)S_n \,, \end{aligned}$$

where we have used

$$\begin{aligned} E(U(X_n - \hat{X}_n)) &= E((c_{n+1}aX_n - \lambda aX_n + a\lambda(X_n - \hat{X}_n))(X_n - \hat{X}_n)) \\ &= a\lambda E((X_n - \hat{X}_n)^2) = a\lambda\Sigma_n \,, \end{aligned}$$

and $E(V_nU) = \lambda K_V$.

Problem 10.8. *Consider the (scalar) system*

$$\begin{aligned} X_{n+1} &= aX_n + V_n \\ Y_n &= \begin{cases} cX_n + W_n & w.p.\ 1-p \\ \text{“error”} & w.p.\ p \end{cases} \end{aligned}$$

The interpretation of “error” is that an observation is lost or discarded independently, with probability $p > 0$.

How would you calculate $E[X_n|Y_0, \ldots, Y_n]$?

Solution:
Assume $X_0, (W_n), (V_n)$ are independent Gaussian. Introduce the measurement $Y_n' = C_n[X_n\ 0]^T + [1\ 0]^T W_n$, where

$$C_n = \begin{cases} [c \quad 0] & \text{,on } Y_n = \text{“error”} \\ [0 \quad 1] & \text{,on } Y_n \neq \text{“error”} \,. \end{cases}$$

Notice that the (optimal) KF estimate for $E[X_n|Y^n]$ does not depend on $C_{n+1}, C_{n+2}, \ldots$, so we don't loose anything by not knowing whether $Y_n =$ “error” or not, ahead of time. Thus, the

estimate $E[X_n|(Y')^n] = E[X_n|Y^n]$ is as good as the estimate provided by an "oracle" KF where all C_n's are known ahead of time. Thus, $\hat{X}_n = E[X_n|Y^n]$ is given by the KF corresponding to the system where one knows C_n ahead of time, i.e., assumes they are constants.

Observe that $\hat{X}_n$ is not necessarily equal to $L[X_n|Y^n]$, as the latter must be a nonrandom linear combination of $Y_0, \ldots, Y_n$.

Problem 10.9. *Suppose $\eta_0, \eta_1, \ldots$ are i.i.d. $N(\mu, \sigma^2)$ for some unknown parameter μ -σ is known- which we wish to estimate by observing the sequence $\xi_0, \xi_1, \ldots$ given as $\xi_{n+1} = a\xi_n + \eta_n$, $\xi_0 = 0$.*

How to estimate μ using Kalman filtering?

Solution:
The idea is to consider the system with state $X_n = (\mu, \xi_n)$, so

$$X_{n+1} = \begin{pmatrix} 1 & 0 \\ 1 & a \end{pmatrix} X_n + \begin{pmatrix} 0 \\ 1 \end{pmatrix} V_n \,,$$
$$Y_n = [0 \quad 1] X_n \,.$$

Thus, the state estimator given by the KF is

$$\hat{X}_{n+1} = \begin{pmatrix} 1 & 0 \\ 1 & a \end{pmatrix} \hat{X}_n + R_n(\xi_{n+1} - [1 \quad a]\hat{X}_n) \,,$$

where R_n is calculated by the KF recursive equations for this system.

Chapter 11

Kalman Filter: Convergence

11.1 Summary

Here are the key ideas and results of this important topic.

- A system is observable if its state can be determined from its outputs (after some delay).
- A system is reachable if there are inputs to drive it to any state.
- We explore the evolution of the covariance in a linear system in Section 11.3.
- The error covariance of a Kalman Filter is bounded if the system is observable.
- The covariance increases if it starts from zero.
- If a system is reachable and observable, then everything converges and the stationary filter is asymptotically optimal.

In the following sections, we explore what happens as $n \to \infty$. Specifically, we want to understand if the error covariance Σ_n blows up or if it converges to a finite value. First, we recall some key notions about linear systems.

11.2 Observability and Reachability

Lemma 11.1. *Cayley-Hamilton*
Let $A \in \Re^{m \times m}$ *and* $\det(sI - A) = \alpha_0 + \alpha_1 s + \cdots + \alpha_{m-1} s^{m-1} + s^m$. *Then*

$$\alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1} A^{m-1} + A^m = 0.$$

In particular, $span\{I, A, A^2, \ldots\} = span\{I, A, \ldots, A^{m-1}\}$.

For a proof, see B.2 in Appendix B.

Definition 11.1. *Observability and Reachability*
The linear system

$$\mathbf{X}_{n+1} = A\mathbf{X}_n, \mathbf{Y}_n = C\mathbf{X}_n, n \geq 1 \tag{11.1}$$

is observable *if* $\mathbf{X}_1$ *can be determined exactly from* $\{\mathbf{Y}_n, n \geq 1\}$. *We then say* (A, C) *is observable.*

The linear system

$$\mathbf{X}_{n+1} = A\mathbf{X}_n + C\mathbf{U}_n, n \geq 1 \tag{11.2}$$

is reachable *if, for every state* $\mathbf{X}$*, there is a sequence of inputs* $\{\mathbf{U}_n, n \geq 1\}$ *that drives the system from* $\mathbf{X}_1 = 0$ *to state* $\mathbf{X}$*. We say that* (A, C) *is reachable.*

Fact 11.1. *(a)* (A, C) *is observable if and only if* $[C^T|A^TC^T|\cdots|(A^{m-1})^TC^T]$ *is of full rank.*

(b) (A, C) *is reachable if and only if* (A^T, C^T) *is observable, i.e., if and only if* $[C|AC|\cdots|A^{m-1}C]$ *is of full rank. In that case,*

$$\sum_{p=0}^{m-1} A^p CC^T (A^p)^T \text{ is positive definite.}$$

Proof: To see (a) note that (11.1) implies that $\mathbf{Y}_n = C\mathbf{X}_n = CA^{n-1}\mathbf{X}_1$. Consequently, observability is equivalent to the null space of $[C|CA|CA^2|\cdots]$ being $\{0\}$. Accordingly, this is equivalent to the matrix $[C^T|A^TC^T|(A^2)^TC^T|\cdots]$ being of full rank. The conclusion then follows from Lemma 11.1.

For (b), observe that (11.2) implies that

$$\mathbf{X}_n = A^{n-1}\mathbf{X}_1 + \sum_{k=1}^{n-1} A^{n-k-1}C\mathbf{U}_k = \sum_{k=1}^{n-1} A^{n-k-1}C\mathbf{U}_k.$$

Therefore, the system is reachable if and only if $[C|AC|A^2C|\cdots]$ is of full rank. The conclusion then follows again from Lemma 11.1. □

11.3 System Asymptotics

Our discussion is borrowed from [8]. First, let us examine the evolution of the unobserved system

$$\mathbf{X}_{n+1} = A\mathbf{X}_n + \mathbf{V}_n, n \geq 1,$$

where $\{\mathbf{X}_1, \mathbf{V}_n, n \geq 1\}$ are all orthogonal and zero-mean with $\text{cov}(\mathbf{V}_n) = K_{\mathbf{V}}$ and $K_n = \text{cov}(\mathbf{X}_n)$. Note that

$$K_{n+1} = AK_nA^T + K_{\mathbf{V}}. \tag{11.3}$$

The following theorem describes the evolution of K_n. Recall that a matrix $A \in \Re^{m\times m}$ is said to be *stable* if its eigenvalues all have magnitude strictly less than 1.

Theorem 11.1. *(a) If* A *is* stable*, then there is a positive semidefinite matrix* K *such that* $K_n \to K$ *as* $n \to \infty$*. Moreover,* K *is the unique solution of the equation*

$$K = AKA^T + K_{\mathbf{V}}. \tag{11.4}$$

(b) Let $K_{\mathbf{V}} = QQ^T$*. Assume that* (A, Q) *is reachable. Then* A *is stable if and only if the equation*

$$K = AKA^T + QQ^T$$

has a positive definite solution K*.*

Proof: (a) By induction we see that the solution of (11.3) is

$$K_n = A^{n-1}K_1(A^{n-1})^T + \sum_{p=0}^{n-2} A^p K_{\mathbf{V}}(A^p)^T. \tag{11.5}$$

If A is stable, then $|(A^p)_{i,j}| \leq C\alpha^p$ for some finite C and some $\alpha \in (0,1)$, as it can be seen by considering the SVD of A. This implies that the first term in (11.5) vanishes and the sum converges to some $K = \sum_{p=0}^{\infty} A^p K_{\mathbf{V}}(A^p)^T$ that is clearly positive semidefinite. The convergence also implies (11.4). It remains to show that (11.4) has a unique solution. Let K' be another solution. Then $\Delta = K' - K$ satisfies $\Delta = A\Delta A^T$. Recursive substitutions show that $\Delta = A^n\Delta(A^n)^T$. Letting $n \to \infty$ shows that $\Delta = 0$ and the solution of (11.4) is unique.

(b) If A is stable, we know from (a) that (11.4) has a unique positive semidefinite solution $K = \sum_{p=0}^{\infty} A^p QQ^T(A^p)^T$. Since (A, Q) is reachable, the null space of $[Q^T|Q^TA^T|Q^T(A^2)^T|\cdots]$ is $\{0\}$. This implies K is positive definite.

To show the converse, assume that K is a positive definite solution of (11.4). Then we know that $K = A^nK(A^n)^T + \sum_{p=0}^{n-1} A^pQQ^T(A^p)^T, n \geq 1$. Let λ be an eigenvalue of A^T with eigenvector v. Then $A^Tv = \lambda v$ and

$$v^*Kv = |\lambda|^{2n}v^*Kv + v^*\left(\sum_{p=0}^{n-1} A^pQQ^T(A^p)^T\right)v.$$

But $\sum_{p=0}^{n-1} A^pQQ^T(A^p)^T$ is positive definite from the reachability of (A, Q), which implies that the last term in the above identity is positive. Consequently, it must be that $|\lambda| < 1$. □

The statements of the theorem should be intuitive. If the system is stable, then the state tries to go to 0 but is constantly pushed by the noise. That noise cannot push the state very far and one can expect the variance of the state to remain bounded. The convergence is a little bit more subtle. If the system is reachable, then the noise pushes the state in all directions and it is not surprising that the variance of the state is positive definite if the system is stable. If the system is not stable, the variance would explode.

11.4 Filter Asymptotics

We now explore the evolution of Kalman Filter.

Theorem 11.2. *Let $K_{\mathbf{V}} = QQ^T$. Suppose that (A, Q) is reachable and (A, C) is observable. If $S_1 = 0$, then*

$$\Sigma_n \to \Sigma, R_n \to R, S_n \to S, \text{ as } n \to \infty.$$

The limiting matrices are the only solutions of the equations

$$\Sigma = (I - RC)S, R = SC^T(CSC^T + K_{\mathbf{W}})^{-1}, \text{ and } S = A\Sigma A^T + K_{\mathbf{V}}.$$

Equivalently, S is the unique positive semidefinite solution of

$$S = A(S - SC^T\left(CSC^T + K_{\mathbf{W}}\right)^{-1}CS)A^T + K_{\mathbf{V}}. \tag{11.6}$$

Moreover, the time-invariant filter $\mathbf{Z}_n = A\mathbf{Z}_{n-1} + R(\mathbf{Y}_n - CA\mathbf{Z}_{n-1})$ satisfies $cov(\mathbf{Z}_n - \hat{\mathbf{Z}}_n) \to \Sigma$, as $n \to \infty$.

Comments: The reachability implies that the noise excites all the components of the state. The observability condition guarantees that the observations track all the components of the state and imply that the estimation error remains bounded. Note that the state could grow unbounded, if A is unstable, but the estimator tracks it even in that case. The time-invariant filter has the same asymptotic error as the time-varying one.

Proof: The proof has the following steps. For two positive semidefinite matrices S and S', we say that $S \leq S'$ if $S' - S$ is positive semidefinite. Similarly, we say that the positive semidefinite matrices $\{S_n, n \geq 1\}$ are bounded if $S_n \leq S$ for some positive semidefinite matrix S.

- (a) The matrices S_n are bounded.
- (b) If $S_1 = 0$, then S_n is nondecreasing in S_1.
- (c) If $S_1 = 0$, then $S_n \uparrow S$, where S is a positive semidefinite matrix.
- (d) The matrix $A - ARC$ is stable.
- (e) For any S_1, $S_n \to S$.
- (f) Equation (23.8) has a unique positive semidefinite solution S.
- (g) The time-invariant filter has the same asymptotic error covariance as the time-varying filter.

We outline these steps.

(a) The idea is that, because of observability, $\mathbf{X}_{n+m}$ is a linear function of $\{\mathbf{Y}_p, \mathbf{V}_p, \mathbf{W}_p, p = n, \ldots, n+m-1\}$. The covariance S_{n+m} must be bounded by that of $\mathbf{X}_{n+m}$ given $\{\mathbf{Y}_p, p = n, \ldots, n+m-1\}$, which is a linear combination of the covariances of $2m$ random variables and is therefore uniformly bounded for all n.

(b) That is, if S_1 is replaced by $S_1' \geq S_1$, then S_n is replaced by $S_n' \geq S_n$. The proof of this fact is pretty neat. One could try to do it by induction, based on the algebra. That turns out to be tricky. Instead, consider the following argument. Since we worry only about covariances, we may assume that all the random variables are jointly Gaussian. In that case, increasing S_1 to $S_1' > S_1$ can be done by replacing $\mathbf{X}_1$ by $\mathbf{X}_1' = \mathbf{X}_1 + \xi_1$, where $\xi_1 = N(0, S_1' - S_1)$ and is independent of $\{\mathbf{X}_1, \mathbf{V}_n, \mathbf{W}_n, n \geq 0\}$. Now, let $\mathbf{X}_n'$ be the system state corresponding to $\mathbf{X}_1'$ and $\mathbf{Y}_n'$ be the corresponding observation. Note that $\mathbf{X}_{n+1}' = \mathbf{X}_{n+1} + A^n \xi_1$. Consequently,

$$L[\mathbf{X}_{n+1}'|\mathbf{Y}^n, \xi_1] = L[\mathbf{X}_{n+1}|\mathbf{Y}^n] + L[\mathbf{X}_{n+1}|\xi_1] + L[A^n\xi_1|\mathbf{Y}^n] + L[A^n\xi_1|\xi_1] = L[\mathbf{X}_{n+1}|\mathbf{Y}^n] + A^n\xi_1$$

and

$$S_{n+1} = \text{cov}(\mathbf{X}_{n+1}' - L[\mathbf{X}_{n+1}'|\mathbf{Y}^n, \xi_1])$$

while

$$S_{n+1}' = \text{cov}(\mathbf{X}_{n+1}' - L[\mathbf{X}_{n+1}'|(\mathbf{Y}')^n]).$$

Since, for each $n \geq 1$, one can express $\mathbf{Y}_n'$ as a linear function of $(\mathbf{Y}_n, \xi_1)$, it follows that $S_{n+1} \leq S_{n+1}'$.

(c) Note that $S_1 = 0 \leq S_2$. From part (b), $S_n \leq S_{n+1}$. However, $S_n \leq B$, for some positive semidefinite B. Thus, $S_n(i,i) := e_i^T S_n e_i \leq S_{n+1}(i,i) \leq B(i,i)$. This implies that $S_n(i,i) \uparrow S(i,i)$ for some nonnegative $S(i,i)$. Similarly, $\alpha(n) := S_n(i,i) + 2S_n(i,j) + S_n(j,j) = (e_i+e_j)^T S_n(e_i+e_j) \leq \alpha(n+1) \leq (e_i+e_j)^T B(e_i+e_j)$. This implies that $\alpha(n) \uparrow \alpha$. We conclude that $S_n(i,j)$ must converge to some $S(i,j)$. Hence $S_n(i,j) \to S(i,j)$.

(d) Simple algebraic manipulations show that

$$S = (A - ARC)S(A - ARC)^T + ARK_{\mathbf{W}}R^TA^T + K_{\mathbf{V}}.$$

Assume that $(A - ARC)^T v = \lambda v$ with $|\lambda| > 1$ and nonzero v. Then

$$v^*Sv = v^*(A-ARC)S(A-ARC)^Tv+v^*ARK_{\mathbf{W}}R^TA^Tv+v^*K_{\mathbf{V}}v = |\lambda|^2v^*Sv+v^*ARK_{\mathbf{W}}R^TA^Tv+v^*K_{\mathbf{V}}v.$$

Since $|\lambda| > 1$, it is obtained that $v^*Sv = v^*ARK_{\mathbf{W}}R^TA^Tv = v^*K_{\mathbf{V}}v = v^*QQ^Tv = 0$, which implies $Qv = 0$ and (A, Q) is not reachable.

(e) We borrow this proof from [7]. The notation is the same as before. Consider the system $\xi_{n+1} = A^T\xi_n + C^T\zeta_n + \mathbf{W}_n$. The problem is to find the $\zeta_k(\xi_k)$ for $k = 1, \ldots, n$ to achieve

$$\Phi_n(\xi) = \min\, E[\sum_{k=1}^{n-1}\{\xi_k^TK_{\mathbf{V}}\xi_k + \zeta_k^TK_{\mathbf{W}}\zeta_k\} + \xi_n^TS_1\xi_n \mid \xi_1 = \xi].$$

By considering all possible choices for ζ_1 and assuming that the cost is minimized over $\{2, \ldots, n+1\}$, one sees that

$$\Phi_{n+1}(\xi) = \min_\zeta\{\xi^TK_{\mathbf{V}}\xi + \zeta^TK_{\mathbf{W}}\zeta + E(\Phi_n(A^T\xi + C^T\zeta + \mathbf{W}_1))\}.$$

A direct calculation (see below) allows to verify that

$$\Phi_n(\xi) = \xi^TS_n\xi + \sum_{k=1}^{n-1}\text{trace }(S_kK_{\mathbf{V}})$$

and the minimizing values of ζ_n are given by

$$\zeta_{n-k} = -[CS_kC^T + K_{\mathbf{W}}]^{-1}CS_kA^T\xi_{n-k}.$$

Instead of using this optimal control, one uses

$$\varepsilon_{n-k} = -[CSC^T + K_{\mathbf{W}}]^{-1}CSA^T\xi_{n-k}.$$

In that case, the cost is $\xi^TG_n\xi$ where

$$G_{n+1} = \Gamma G_n\Gamma^T + ARK_{\mathbf{W}}R^TA^T + K_{\mathbf{V}},\ \text{with } G_1 = S_1 \text{ and } \Gamma = A - ARC.$$

Since Γ is stable from part (d), we see that G_n converges to S (Theorem 11.1). Now,

$$\Phi_n(\xi) = \xi^TS_n\xi + \sum_{k=0}^{n-1}\text{trace }(S_kK_{\mathbf{V}}) \le \xi^TG_n\xi,$$

for all ξ. This implies that $S_n \le G_n$. Recall that if $S_1 = 0$, then the resulting covariance matrices, S_n', are such that $S_n' \uparrow S$. Moreover, we see that $S_n' \le S_n \le G_n$ with $S_n' \uparrow S$ and $G_n \to S$. It follows that $S_n \to S$.

In the calculations above, we used the fact that, if A_1 is symmetric and invertible, then

$$\min_\zeta\{\xi^TA_0\xi + \zeta^TA_1\zeta + 2\xi^TA_2\zeta\} = \xi^T[A_0 - A_2^TA_1^{-1}A_2]\xi$$

and is achieved by

$$\zeta = -A_1^{-1}A_2\xi,$$

as a direct calculation shows.

(f) Assume that S' is another positive semidefinite solution of (23.8). If $S_1 = S'$, then $S_n = S'$. However, we know from (e) that $S_n \to S$. Hence, $S' = S$.

(g) This follows immediately because the time-invariant filter is a special case of the time-varying filter. □

Chapter 12

Wiener Filter

12.1 Summary

Here are the key ideas and results.

- Definition of LTI system
- Definition of wide sense stationary processes
- Definition of transfer function and power spectral density
- How an LTI system modifies the spectral density of a process
- Derivation of the Wiener Filter

12.2 Overview

One is given two sequences of random vectors $\{\mathbf{X}_n, \mathbf{Y}_n, n \in \mathbb{Z}\}$. The random variables are specified through their first and second moments. The problem is to calculate $\hat{\mathbf{X}}_n := L[\mathbf{X}_n|\mathbf{Y}^m]$ where $\mathbf{Y}^n := \{\mathbf{Y}_m, m \leq n\}$.

Thus, whereas Kalman assumes a dynamic model of the sequences, Wiener starts with the specification of the first and second moments. For the problem to be tractable, one assumes that the first and second moments time-invariant. One can then expect a result of the form

$$\hat{\mathbf{X}}_n = \sum_{m=-\infty}^{n} \phi(n-m)\mathbf{Y}_m.$$

This expression is called the Wiener filter. Thus, $\hat{X}_n$ is the output at time n of a causal linear time invariant system whose input is the sequence of observations $\{Y_n\}$. Figure 12.1 illustrates the idea.

Figure 12.1: The Wiener Filter.

In this lecture, we explain how to derive the Wiener filter.

We start with a brief review of linear time invariant systems. We then discuss what it means for the first and second order moments to be time-invariant.

12.3 Linear Time Invariant Systems

Definition 12.1. *LTI System*

A system *is a mapping that transforms an arbitrary* input *sequence* $\{X_n, n \in \mathbb{Z}\}$ *into an* output *sequence* $\{Y_n, n \in \mathbb{Z}\}$. *The system is* linear *if to a linear combination of inputs corresponds the same linear combination of outputs. It is* time invariant *if delaying the input by* m *time units results in delaying the output by* m *time unit, for any* $m \in \mathbb{Z}$.

Example 12.1. *(a) A system with output* $Y_n = a_0 X_n + a_1 X_{n-1} + a_2 X_{n-2}$ *is LTI.*

(b) A system with output $Y_n = (X_n + X_{n-1})^2$ *is time invariant but not linear.*

(c) A system with output $Y_n = X_n \cos(2\pi f_0 n)$ *is linear but not time invariant.*

(d) A system such that $Y_n = (1-\alpha)Y_{n-1} + \alpha X_n$ *for* $n \geq 0$ *where* $Y_{-1} := 0$ *is LTI.*

Definition 12.2. *Impulse Response*

Consider an LTI system. Let $\{h_i(n), n \in \mathbb{Z}\}$ *to be its output when its input is* $\{e_i \delta_n, n \in \mathbb{Z}\}$ *where* $\delta_n := 1\{n = 0\}$ *for* $n \in \mathbb{Z}$. *The* impulse response *of the system is defined as* $\{h(n), n \in \mathbb{Z}\}$ *where* $h(n)$ *is a matrix whose* i*-th column is* $h_i(n)$.

Example 12.2. *A system such that* $Y_n = X_{n-n_0}$ *for* $n \in \mathbb{Z}$ *delays its input by* n_0 *time units. Its impulse response is* $H(n) = I1\{n = n_0\}$ *since* $e_i 1\{n = n_0\}$ *is the output at time* n *when the input is* $e_i \delta_n$.

Fact 12.1. *Consider an LTI system with impulse response* $\{h(n), n \in \mathbb{Z}\}$ *and input* $\{X_n, n \in \mathbb{Z}\}$. *Its output is* $\{Y_n, n \in \mathbb{Z}\}$ *where*

$$Y_n = \sum_{m=-\infty}^{\infty} h(n-m) X_m, n \in \mathbb{Z}.$$

Proof:

The input is the linear combination $\sum_n X_n \theta_n \delta$ where θ_n is the operator that delays a function by m time units. Consequently, the output must be $\sum_n X_n \theta_n h$ whose value at time m is $Y_m = \sum_n h(m-n) X_n$.

□

12.4 Wide Sense Stationary

Definition 12.3. *Wide Sense Stationary*

The sequences of complex valued random vectors $\{\mathbf{X}_n, \mathbf{Y}_n, n \in \mathbb{Z}\}$ *are* wide sense stationary *(wss) if*

$$E(\mathbf{X}_n) = \mu_{\mathbf{X}}, E(\mathbf{Y}_n) = \mu_Y, E(X_{n+m} X_n^*) = R_{\mathbf{X}}(m), E(\mathbf{Y}_{n+m} \mathbf{Y}_n^*) = R_Y(m), E(\mathbf{X}_{n+m}, \mathbf{Y}_n^*) = R_{XY}(m), \forall n, m \in \mathbb{Z}.$$

In particular, each sequence is also said to be wide sense stationary. (Some authors say that $(\mathbf{X}, \mathbf{Y})$ *are jointly wide sense stationary, but we use the simpler terminology.)*

For simplicity of notation, we assume throughout that $\mu_X = \mu_Y = 0$.

Example 12.3. *Let* $\{X_n, n \in \mathbb{Z}\}$ *be a sequence of zero mean uncorrelated random variables with unit variance. Then* X *is wss and* $R_X(n) = 1\{n = 0\}$. *This process is called a* white noise.

Example 12.4. *Let $X_n = \xi e^{2\pi f_0 n+\theta}$ where $f_0 \in (-1/2, 1/2)$, $\theta = U[0, 2\pi]$, $E(\xi) = 0$, $var(\xi) = \sigma^2$, and ξ, θ are independent. Then $\{X_n\}$ is wss and $R_X(m) = \sigma^2 e^{j2\pi f_0 m}$. The process $\{X_n\}$ is a sinusoid with frequency f_0 and a random phase (to make it wss).*

We need to clarify the notions of bounded and causal systems.

Definition 12.4. *Bounded; Causal*

A linear system is bounded *if its output is bounded whenever the input is bounded.*

A system is causal *if its output at time n depends only on its inputs at times $m \leq n$, for all $n \in \mathbb{Z}$.*

We have the following lemma.

Lemma 12.1. *(a) A linear time invariant system is bounded if and only if its impulse response $\{h(n)\}$ is summable, i.e., such that $||h||_1 := \sum_n ||h(n)||_1 < \infty$ where $||v||_1 := \sum_i |v_i|$ for any vector v.*

(b) A linear time invariant system is causal if and only if $h(n) = 0$ for $n < 0$.

Proof:

(a) If h is summable, then $||Y_n||_1 \leq ||h||_1 \max_n\{||X_n||_1\}$. Conversely, let $(X_n)_i = \text{sign}(h_i(-n))$, so that $h(-n)X_n = ||h(-n)||_1$. If the system is bounded, since $\{X_n\}$ is bounded, we must have $||Y_0||_1 = ||h||_1 < \infty$.

(b) This statement is obvious.

□

12.5 Frequency Domain

Definition 12.5. *Transfer Function*

The transfer function is the Fourier transform of the impulse response. That is

$$H(f) = \sum_{m=-\infty}^{\infty} h(m)e^{-j2\pi mf}, f \in \Re.$$

Note that

$$h(n) = \int_{-1/2}^{1/2} H(f)e^{j2\pi f} df,$$

since

$$\int_{-1/2}^{1/2} e^{j2\pi nf} df = 1\{n = 0\}.$$

We say that $H(f)$ is rational *if it is the ratio of two polynomials in $e^{-2\pi f}$.*

Note that $e^{j2\pi f} = e^{j2\pi(f+k)}$ for any integer k. Consequently, $H(f)$ is periodic with period 1.

Example 12.5. *Consider the system with $Y_n = (1-\alpha)Y_{n-1} + \alpha X_n$. Assume that $\alpha \in (0, 1)$. Find the transfer function.*

If the input is δ, then the output is such that $Y_0 = \alpha, Y_1 = (1-\alpha)Y_0 + \alpha X_1 = (1-\alpha)\alpha, Y_2 = (1-\alpha)Y_1 = (1-\alpha)^2\alpha$. Continuing in this way, we see that $Y_n = (1-\alpha)^n\alpha 1\{n \geq 0\}$. Hence,

$$h(n) = (1-\alpha)^n \alpha 1\{n \geq 0\},$$

so that

$$H(f) = \sum_{n=0}^{\infty} (1-\alpha)^n \alpha e^{-j2\pi fn} = \frac{1}{1-(1-\alpha)e^{-j2\pi f}} = \frac{\alpha}{1-(1-\alpha)z^{-1}}.$$

Example 12.6. *The system with*

$$b_0 Y_n + b_1 Y_{n-1} + \cdots + b_q Y_{n-q} = a_0 X_n + a_1 X_{n-1} + \cdots + a_p X_{n-p}, n \in \mathbb{Z}$$

has transfer function

$$H(f) = \frac{a_0 + a_1 z^{-1} + \cdots + a_p z^{-p}}{b_o + b_1 z^{-1} + \cdots + b_q z^{-q}}. \tag{12.1}$$

Example 12.7. *In the examples below, A is a stable matrix.*
(a) $h(n) = A^n 1\{n \geq 0\} \leftrightarrow H(f) = [I - Ae^{-j2\pi f}]^{-1}$;
(b) $h(n) = A^{-n} 1\{n \leq 0\} \leftrightarrow H(f) = [I - Ae^{j2\pi f}]^{-1}$.

Definition 12.6. *Poles*
The poles *of a transfer function $H(f)$ are the values of $z := e^{j2\pi f}$ for which $H(f) = \infty$.*

In Example 12.7(a), the poles of $H(f)$ are the eigenvalues of A and are inside the unit circle. Indeed, $H(f) = [I - Az^{-1}]^{-1}$ which is infinite when $I - Az^{-1}$ is singular, i.e., when $|zI - A| = 0$. On the other hand, in Example 12.7(b), the poles are the reciprocals of the eigenvalues of A and are outside the unit circle.

Note that Example 12.7 (a) is a causal transfer function whereas Example 12.7 (b) is not. We suspect that the transfer function is causal iff its poles are inside the unit circle. We discuss that result below. We have the following lemma.

Lemma 12.2. *A bounded system with a rational transfer function is causal if and only if its poles are strictly inside the unit circle.*

Proof:

We consider the one-dimensional case. The multidimensional case is derived by considering each component of h. By performing a partial fraction expansion of $H(z^{-1})$, one finds that it is a linear combination of terms of the form $(1 - az^{-1})^{-k}$. Now,

$$\frac{1}{1-az^{-1}} = \begin{cases} \sum_n a^n 1\{n \geq 0\} z^{-n}, & \text{if } |a| < 1; \\ \sum_n a^n 1\{n \leq 0\} z^{-n}, & \text{if } |a| > 1. \end{cases}$$

Taking the derivative with respect to z, we find

$$\frac{z^{-1}}{(1-az^{-1})^2} = \begin{cases} \sum_n n a^{n-1} 1\{n \geq 0\} z^{-n}, & \text{if } |a| < 1; \\ \sum_n n a^{n-1} 1\{n \leq 0\} z^{-n}, & \text{if } |a| > 1. \end{cases}$$

Taking multiple derivatives shows that the inverse transform of $(1 - az^{-1})^{-k}$ is causal whenever $|a| < 1$ and anti-causal whenever $|a| > 1$.

□

Next we define the spectral density.

Definition 12.7. *Spectral Density*
Let $\{\mathbf{X}_n, \mathbf{Y}_n, n \in \mathbb{Z}\}$ *be wss. Then*

$$S_X(f) := \sum_{m=-\infty}^{\infty} R_X(m)e^{-j2\pi mf} \text{ and } S_{XY}(f) := \sum_{m=-\infty}^{\infty} R_X(m)e^{-j2\pi mf}, f \in \Re. \tag{12.2}$$

S_X *is called the* spectral density *of* X *and* S_{XY} *the* cross spectral density *of* $\mathbf{X}$ *and* $\mathbf{Y}$. *(Some authors say power spectral density.)*

We say that $S_X(f)$ *is* rational *if it is the ratio of two polynomials in* $e^{-j2\pi f}$.

Note that $e^{j2\pi f} = e^{j2\pi(f+k)}$ *for any integer* k. *Consequently,* S_X *and* S_{XY} *are periodic with period 1.*

We have a simple inversion lemma.

Lemma 12.3. *Basic Properties*
One has

$$R_X(n) = \int_{-1/2}^{1/2} S_X(f)e^{j2\pi nf}df \text{ and } R_{XY}(n) = \int_{-1/2}^{1/2} S_{XY}(f)e^{j2\pi nf}df. \tag{12.3}$$

Moreover,

$$E(||X_n||^2) = \int_{-1/2}^{1/2} trace(S_X(f))df. \tag{12.4}$$

Proof:
To verify (12.4), note that

$$E(||X_n||^2) = E(X_n^*X_n) = \text{trace}(E(X_nX_n^*)) = \text{trace}(R_X(0)) = \text{trace}(\int_{-1/2}^{1/2} S_X(f)df).$$

□

The spectral density is a convenient way to study the effect of linear time invariant systems on random sequences, as the following result shows.

Theorem 12.1. *Assume that*

$$\mathbf{V}_n = \sum_{m=-\infty}^{\infty} h_1(n-m)\mathbf{X}_m \text{ and } \mathbf{W}_n = \sum_{m=-\infty}^{\infty} h_2(n-m)\mathbf{Y}_m$$

where $\sum_{m=-\infty}^{\infty} ||h_k(m)||^2 < \infty$ *for* $k = 1, 2$. *Then*

$$S_{VW}(f) = H_1(f)S_{XY}(f)H_2^*(f)$$

where $H_k(f)$ *is the* transfer function *that corresponds to* h_k.

Proof:
The result says that time-domain convolutions become products in the frequency domain.

We have

$$\begin{aligned} S_{VW}(f) &= \sum_m R_{VW}(m)z^{-m} = \sum_m E(V_{n+m}W_n^*)z^{-m} \\ &= \sum_m E(\sum_k h_1(n+m-k)X_k \sum_p Y_p^* h_2^*(n-p))z^{-m} \\ &= \sum_m \sum_k \sum_p h_1(n+m-k)z^{-(n+m-k)} R_{XY}(k-p)z^{-(k-p)f} h_2^*(n-p)z^{-(n-p)} \\ &= H_1(f)S_{XY}(f)H_2^*(f), \end{aligned}$$

as claimed.

□

The following examples illustrate the ideas.

Example 12.8. *Let $\{X_n, n \in \mathbb{Z}\}$ be a* white noise *, i.e., a sequence of zero mean uncorrelated random variables with unit variance. Then $S_X(f) = 1$. Indeed, $R_X(n) = 1\{n = 0\}$, so that $S_X(f) := \sum_{m=-\infty}^{\infty} R_X(m)z^{-m} = 1$.*

Example 12.9. *Let $X_n = \xi e^{2\pi f_0 n + \theta}$ where $f_0 \in (-1/2, 1/2)$, $\theta = U[0, 2\pi]$, $E(\xi) = 0$, $var(\xi) = \sigma^2$, and ξ, θ are independent. Then $S_X(f) = \sigma^2\delta(f - f_0)$.*

Indeed,

$$R_X(m) = \sigma^2 e^{j2\pi f_0 m} = \int_{-1/2}^{1/2} \sigma^2 \delta(f - f_0)e^{j2\pi mf} df.$$

The process $\{X_n\}$ is a sinusoid with frequency f_0 and a random phase (to make it wss). The expression for S_X indicates that all the power is concentrated on the frequency f_0.

Example 12.10. *Let $\{X_n\}$ be wss and one-dimensional. Then $Y_n = \sum_k h(n-k)X_k$ is such that $S_Y(f) = |H(f)|^2 S_X(f)$. Assume now that h is a filter that lets the frequencies in $[f_0, f_0 + \epsilon]$ go through without modifying their amplitude and blocks all other frequencies, with $0 < \epsilon \ll 1$. Then $H(f) = 1\{f_0 \le f \le f_0 + \epsilon\}$. Consequently,*

$$E(|Y_n|^2) = \int_{-1/2}^{1/2} S_Y(f)df = \int_{f_0}^{f_0+\epsilon} S_X(f)df = S_X(f_0)\epsilon.$$

Accordingly, we see that $S_X(f)\epsilon$ is the power contained in $\{X_n\}$ in the frequencies $[f_0, f_0 + \epsilon]$. Hence the terminology (power) spectral density. A similar justification can be made for the meaning of cross spectral density.

Example 12.11. *Assume that the input of system (12.1) is white. Find $S_Y(f)$ where Y is the output of the filter. We know that*

$$S_Y(f) = |H(f)|^2 = |\frac{a_0 + a_1 z^{-1} + \cdots + a_p z^{-p}}{b_o + b_1 z^{-1} + \cdots + b_q z^{-q}}|^2 \text{ where } z = e^{j2\pi f}.$$

Note that S_Y is rational.

12.6 Wiener Filter

We start with a representation result.

Theorem 12.2. *Whitening Filter*

Consider a wss process X with rational spectral density S_X. One assumes that the process satisfies some technical conditions known as the Paley-Wiener conditions.

In the scalar case, one can write

$$S_X(f) = |H(f)|^2$$

where $H(f) = N(f)/D(f)$ with N and D being polynomials in z^{-1} with zeros inside the unit circle. (The zeros are the values of z for which the polynomials are zero.)

In the multidimensional case, one can write

$$S_X(f) = H(f)KH^*(f)$$

where $H(f)$ has its poles inside the unit circle, $H^{-1}(f)$ has its poles on or inside the unit circle, and K is positive definite.

This result shows that one can represent X as the output of a linear time invariant filter whose input is white; the filter is causal and its inverse is causal. This filter is called the *whitening filter*.

This result is easy if the process is the output of a filter with rational transfer function and white input noise. The result is non trivial even in that case because the filter may not be causal nor causally invertible. In the the scalar case, we know that $S_X(f) = |\Phi(f)|^2$ where Φ is the transfer function of the filter. Since Φ is rational, one can separate the zeros and poles so that H has the indicated properties. Because $E(|X_n|^2) < \infty$, S_X cannot have poles on the unit circle.

The multidimensional case is similar.

□

Here is an example.

Example 12.12. *Assume that*

$$S_Y(f) = \frac{1}{|(2-z)(3-z)|^2}.$$

In that case, we can write $S_Y(z^{-1}) = |H(z^{-1})|^2$ with

$$H(z^{-1}) = \frac{1}{(2-z^{-1})(3-z^{-1})} = \frac{1}{2-z^{-1}} - \frac{1}{3-z^{-1}}$$

where H is causal and causally invertible.

12.6.1 Projection

We explain how to calculate the LLSE of X_n given all values of Y (not only the past values).

Theorem 12.3. *Let $\{X, Y\}$ be wss. Then*

$$\tilde{X}_n = L[X_n | Y_m, m \in \mathbb{Z}] = \sum_m g(n-m) Y_m$$

where

$$G(f) = S_{XY}(f) S_Y^{-1}(f).$$

Proof:

We know that G should be such that $X_n - \tilde{X}_n \perp Y_m$ for all $m \in \mathbb{Z}$. That is,

$$0 = S_{(X-\tilde{X}),Y}(f) = S_{XY}(f) - G(f) S_Y(f),$$

which proves the result.

□

12.6.2 Whitening and Projection

Now we consider the situation illustrated in Figure 12.2.

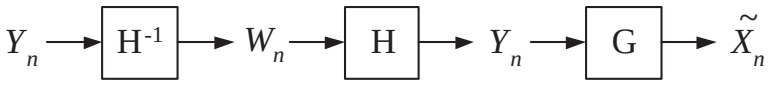


Figure 12.2: A non-causal filter.

This filter calculates $L[X_n | Y_m, m \in \mathbb{Z}] = L[X_n | W_m, m \in \mathbb{Z}]$ where W is a white noise. The following result explains how to extract the optimal causal filter.

Theorem 12.4. *Wiener Filter*

Let

$$K(f) = H(f) G(f).$$

Then

$$\hat{X}_n = L[X_n | Y^n] = \sum_{m \le n} k(n-m) W_m.$$

Let also $K_+(e^{-j2\pi f})$ be the transfer function of the filter with impulse response $k(n)1\{n \ge 0\}$. Then the Wiener filter *has transfer function*

$$K_+(f) H^{-1}(f) \text{ where } K(f) = S_{XY}(f) S_Y^{-1}(f) H(f).$$

The filter is illustrated in Figure 12.3.

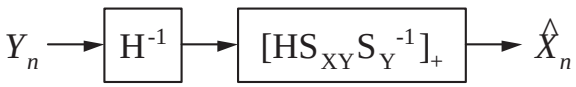


Figure 12.3: The construction of the Wiener filter.

Proof:

Because H is causal and causally invertible, the set of linear combinations of W^n is exactly the set of linear combinations of Y^n. Moreover, W is white. Consequently,

$$L[\sum_m k(n-m)W_m|Y^n] = L[\sum_m k(n-m)W_m|W^n] = \sum_{m\le n} k(n-m)W_m.$$

□

Example 12.13. *Assume that*

$$S_Y(f) = \frac{1}{|(2+z)(3+z)|^2}.$$

and

$$X_n = Y_{n+n_0}, n \in \mathbb{Z}.$$

We calculate the Wiener filter to find $L[Y_{n+n_0}|Y^n]$.

In Example 12.12, we found

$$H(z^{-1}) = \frac{1}{(2-z^{-1})(3-z^{-1})} = \frac{1}{2-z^{-1}} - \frac{1}{3-z^{-1}},$$

so that

$$h(n) = (\frac{1}{2})^{n+1}1\{n \ge 0\} - (\frac{1}{3})^{n+1}1\{n \ge 0\}.$$

To find S_{XY} *we note that* X *is obtained from* Y *by a linear time invariant filter with impulse response* $\phi(n) = 1\{n = n_0\}$, *so that*

$$\Phi(f) = \sum_{m=-\infty}^{\infty} \phi(m)e^{-j2\pi fm} = e^{-j2\pi fn_0}.$$

Accordingly,

$$S_{XY}(f) = \Phi(f)S_Y(f) = e^{-j2\pi n_0 f}S_Y(f).$$

Next, we calculate

$$K(f) = S_{XY}(f)S_Y^{-1}(f)H(f) = e^{-j2\pi n_0 f}H(f),$$

so that

$$k(n) = h(n+n_0), n \in \mathbb{Z}.$$

It follows that

$$K_+(f) = \sum_{n=0}^{\infty} h(n+n_0)e^{-j2\pi nf} = \frac{2^{-n_0}}{2-z^{-1}} - \frac{3^{-n_0}}{3-z^{-1}}.$$

Finally, we get the Wiener filter:

$$K_+(f)H^{-1}(f) = [\frac{2^{-n_0}}{2-z^{-1}} - \frac{3^{-n_0}}{3-z^{-1}}][(2-z^{-1})(3-z^{-1})] = 2^{-n_0}(3-z^{-1}) - 3^{-n_0}(2-z^{-1}).$$

Hence,

$$\hat{X}_n = (3.2^{-n_0} - 2.3^{-n_0})Y_n - (2^{-n_0} - 3^{-n_0})Y_{n-1}.$$

12.7 Solved Problems

Problem 12.1. *Let (X_n) be wide-sense stationary (WSS), and $X_n = Y_{n+n_0}$ for some fixed $n_0 \geq 0$.*

Let

$$S_Y(f) = \frac{|2-z|^2}{|4-z|^2|3-z|^2}\,,$$

where $z = e^{j2\pi f}$.

Find the Wiener filter for $\hat{X}_n = L[X_n|Y^n]$.

Solution:
Notice that $S_Y(f) = |H(f)|^2$, for

$$H(f) = \frac{2-z^{-1}}{(4-z^{-1})(3-z^{-1})}\,,$$

where H, H^{-1} are causal, since all poles and zeros are inside the unit circle.

Also, can write $H(f)$ as

$$\begin{aligned} H(f) &= \frac{2}{4-z^{-1}} - \frac{1}{3-z^{-1}} \\ &\longleftrightarrow \frac{1}{2}\left(\frac{1}{4}\right)^n 1\{n\geq 0\} - \frac{1}{3}\left(\frac{1}{3}\right)^n 1\{n\geq 0\} =: h(n)\,. \end{aligned}$$

The impulse response of the system giving Y from X is $h_0(n) = 1\{n = -n_0\}$. But, $S_{XY}(f) = S_Y(f)H_0(f)$, so the Wiener filter is $H^{-1}(f)[H(f)H_0(f)]_+$, where $[\cdot]_+$ denotes the causal part. Let $K(f) = H(f)H_0(f)$.

Now $k(n) = h(n+n_0)$, so

$$\begin{aligned} k_+(n) &= h(n+n_0)1\{n\geq 0\} \\ &\longleftrightarrow \sum_{n\geq 0}\left(\frac{1}{2}\left(\frac{1}{4}\right)^{n+n_0} z^{-n} - \frac{1}{3}\left(\frac{1}{3}\right)^{n+n_0} z^{-n}\right) \\ &\quad \frac{1}{2}\left(\frac{1}{4}\right)^{n_0}\frac{1}{1-\frac{1}{4}z^{-1}} - \frac{1}{3}\left(\frac{1}{3}\right)^{n_0}\frac{1}{1-\frac{1}{3}z^{-1}} =: K_+(f)\,. \end{aligned}$$

The Wiener filter is $H^{-1}K_+$.

Problem 12.2. *Cosider two processes X, Y such that*

$$\begin{aligned} X_n &= \frac{1}{3}X_{n-1} + W_{n-1} \\ Y_n &= \frac{1}{2}Y_{n-1} + \frac{1}{2}X_n + V_n\,, \end{aligned}$$

where (V, W) are uncorrelated white noise sequences with known variances.

1. *Assume (X, Y) is WSS.*

   Calculate S_X, S_{XY}, S_Y.

2. *Find the Wiener filter for estimating X.*

3. *Augment state to bring the model into the form of a linear system such as the one in the formulation of Kalman filter.*

4. *Calculate the Kalman filter for the system in (c).*

5. *Show that the* stationary *Kalman filter is the same as the Wiener filter found in (b).*

Solution:
With no loss of generality, assume $E(W_n) = 0, E(W_n^2) = 1, E(V_n) = 0, E(V_n^2) = \sigma_V^2$.

1. X_n is given by passing W_n through the filter with transfer function $3z^{-1}/(3 - z^{-1})$, where $z = e^{j2\pi f}$, and f is the frequency. Hence,

$$S_X(f) = \left| \frac{3z^{-1}}{3 - z^{-1}} \right|^2 \cdot S_W(f) = \frac{9}{|3 - z^{-1}|^2} .$$

Similarly, Y_n is given by passing $\frac{1}{2}X_n + V_n$ through a filter with transfer function $2/(2 - z^{-1})$, so

$$S_{XY}(f) = S_{X,\frac{1}{2}X+V}(f)\frac{2}{2 - z} ,$$

by using the formulas from Lecture 15. But, since $V \perp X$, $S_{X,\frac{1}{2}X+V}(f) = \frac{1}{2}S_X(f)$. Thus,

$$S_{XY}(f) = \frac{9}{(2 - z)(3 - z^{-1})(3 - z)} .$$

Moreover,

$$\begin{aligned} S_Y(f) &= \left| \frac{2}{2 - z^{-1}} \right|^2 S_{X,\frac{1}{2}X+V}(f) \\ &= \frac{4\sigma_V^2}{|2 - z^{-1}|^2} + \frac{9}{|2 - z^{-1}|^2|3 - z^{-1}|^2} \\ &= 9\frac{4\sigma_V^2|3 - z^{-1}|^2/9 + 1}{|2 - z^{-1}|^2|3 - z^{-1}|^2} . \end{aligned}$$

2. We first derive the transfer function for the whitening filter H.

$$S_Y(f) = 9\frac{(a - bz^{-1})(a - bz)}{|2 - z^{-1}|^2|3 - z^{-1}|^2} ,$$

where a, b are determined by $a^2 + b^2 = 40\sigma_V^2/9 + 1, ab = 12\sigma_V^2/9$, and the requirement that $|b| < |a|$. In particular,

$$b = \sqrt{\frac{40\sigma_V^2 + 9 - \sqrt{(9 + 16\sigma_V^2)(9 + 64\sigma_V^2)}}{18}} ,$$

$$a = \sqrt{\frac{40\sigma_V^2 + 9 + \sqrt{(9 + 16\sigma_V^2)(9 + 64\sigma_V^2)}}{18}} .$$

Since $0 < b < a$, we see that by defining

$$H(f) = 3\frac{a - bz^{-1}}{(2 - z^{-1})(3 - z^{-1})}\,,$$

H, H^{-1} are causal. This will be our whitening filter.

Now, the non-causal filter is

$$G(f) = \frac{S_{XY}(f)}{S_Y(f)} = \frac{2 - z^{-1}}{|a - bz^{-1}|^2}\,,$$

so

$$G(f)H(f) = \frac{3}{(a - bz)(3 - z^{-1})} = 3\left(\frac{c}{a - bz} + \frac{dz^{-1}}{3 - z^{-1}}\right)\,,$$

for $c = a/(3a - b)$, $d = 1/(3a - b)$.

To take the causal part of $G(f)H(f)$, we turn into the time domain.

$$G(f)H(f) \longleftrightarrow \frac{3c}{a}\left(\frac{a}{b}\right)^t 1\{t \le 0\} + 3d\left(\frac{1}{3}\right)^t 1\{t \ge 1\}\,,$$

so the causal part $[GH]_+$ is

$$3d\left(\frac{1}{3}\right)^t 1\{t \ge 0\} \longleftrightarrow \frac{3d}{1 - \frac{1}{3}z^{-1}}\,,$$

where we used $c = ad$.

Finally,

$$\begin{aligned} H^{-1}[GH]_+ &= \frac{(2 - z^{-1})(3 - z^{-1})}{3(a - bz^{-1})}\frac{9d}{3 - z^{-1}} \\ &= \frac{3d(2 - z^{-1})}{a - bz^{-1}} = \frac{6d}{a - bz^{-1}} - \frac{3dz^{-1}}{a - bz^{-1}}\,, \end{aligned}$$

so the Wiener filter is

$$\check{X}_n = \frac{b}{a}\check{X}_{n-1} + \frac{6d}{a}Y_n - \frac{3d}{a}Y_{n-1}\,,$$

where a, b, d are determined above.

3. Let (X_n, Y_n) be the state. Then,

$$\begin{aligned} \begin{pmatrix} X_n \\ Y_n \end{pmatrix} &= \begin{pmatrix} 1/3 & 0 \\ 1/6 & 1/2 \end{pmatrix}\begin{pmatrix} X_{n-1} \\ Y_{n-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1/2 & 1 \end{pmatrix}\begin{pmatrix} W_{n-1} \\ V_n \end{pmatrix} \\ Z_n &= \begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \end{aligned}\,,$$

where Z_n is the observation.

4. Let $(\hat{X}_n, \hat{Y}_n)$ be the estimator for (X_n, Y_n). Then,

$$\begin{aligned} \hat{X}_n &= \frac{1}{3}\hat{X}_{n-1} + R_n(Y_n - \frac{1}{6}\hat{X}_{n-1} - \frac{1}{2}Y_{n-1}) \\ &= \frac{2 - R_n}{6}\hat{X}_{n-1} + R_nY_n - \frac{R_n}{2}Y_{n-1}, \end{aligned}$$

where the first equality is because $\hat{Y}_n = Y_n$. R_n is determined for every n, by the KF equations for this system.

5. First, note that the system in (c) is observable and reachable, so KF converges. To find the stationary KF, we solve the Ricatti equation

$$S = A(S - SC^T[CSC^T + K_W]^{-1}CS)A^T + K_V\,,$$

where for the system in part (d), we have

$$A = \begin{pmatrix} 1/3 & 0 \\ 1/6 & 1/2 \end{pmatrix}\,, \quad C = \begin{pmatrix} 0 & 1 \end{pmatrix}\,, \quad K_V = \begin{pmatrix} 1 & 1/2 \\ 1/2 & \sigma_V^2 + 1/4 \end{pmatrix}\,, \quad K_W = 0\,.$$

Let

$$S = \begin{pmatrix} s_1 & s_2 \\ s_3 & s_4 \end{pmatrix}\,,$$

and substitute for the right-hand-side of the Ricatti equation to get

$$\begin{pmatrix} s_1 & s_2 \\ s_3 & s_4 \end{pmatrix} = \frac{1}{36}\begin{pmatrix} s_1 - \dfrac{s_2 s_3}{s_4} \end{pmatrix}\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1/2 \\ 1/2 & \sigma_V^2 + 1/4 \end{pmatrix}\,.$$

So

$$S = \begin{pmatrix} 4x + 1 & 2x + 1/2 \\ 2x + 1/2 & x + \sigma_V^2 + 1/4 \end{pmatrix}\,, \tag{12.5}$$

where

$$\begin{aligned} x &= \frac{1}{36}\left(s_1 - \frac{s_2 s_3}{s_4}\right) \\ &= \frac{1}{36}\left(4x + 1 - \frac{(2x + 1/2)^2}{x + \sigma_V^2 + 1/4}\right)\,. \end{aligned}$$

The last line comes from using (12.5) again. This gives an equation in x, with solution

$$x = \frac{\sqrt{(9 + 16\sigma_V^2)(9 + 64\sigma_V^2)} - 32\sigma_V^2 - 9}{72}\,.$$

Now R_n converges to R, given by

$$R = \frac{s_2}{s_4} = \frac{2\sqrt{(9 + 16\sigma_V^2)(9 + 64\sigma_V^2)} - 64\sigma_V^2 + 18}{\sqrt{(9 + 16\sigma_V^2)(9 + 64\sigma_V^2)} + 40\sigma_V^2 + 9}\,.$$

To see that the stationary KF,

$$\hat{X}_n = \frac{2 - R}{6}\hat{X}_{n-1} + RY_n - \frac{R}{2}Y_{n-1},$$

coincides with the Wiener filter in part (b), note that

$$\frac{2a - 6b}{a} = \frac{2a^2 - 6ba}{a^2} = \frac{-64\sigma_V^2 + 18 + 2\sqrt{(9 + 16\sigma_V^2)(9 + 64\sigma_V^2)}}{40\sigma_V^2 + 9 + \sqrt{(9 + 16\sigma_V^2)(9 + 64\sigma_V^2)}} = R\,.$$

Thus, $(2 - R)/6 = b/a$. In the same way -or by looking only at the definition of d, a, b- we see that $R = 6d/a$.

Hence stationary KF=Wiener.

Problem 12.3. *Find a WSS process X which is not stationary. (We call X* stationary - *or* stationary in the strict sense-*, when for any $k \geq 1$, any Borel sets $A_1, \ldots, A_k$, we have $P(X_1 \in A_1, \ldots, X_k \in A_k) = P(X_{1+m} \in A_1, \ldots, X_{k+m} \in A_k)$ for all $m \in \mathbb{Z}$.)*

Solution:
For odd $n \in \mathbb{Z}$ consider i.i.d., Bernoulli X_n s.t. $X_n = 1$ w.p. $1/2$, and $X_n = -1$ otherwise. For even $n \in \mathbb{Z}$ consider i.i.d., Gaussian X_n s.t. $X_n = N(0,1)$. Then, $E(X_n X_m) = 1\{n - m = 0\}$, $E(X_n) = 0$, but $P(X_1 > 1) = 0 < P(X_2 > 1)$.

Problem 12.4. *Let $(Z_n : n \in \mathbb{Z})$ be a sequence of i.i.d. geometric r.v's with parameter p, i.e., $P(Z_n = k) = (1-p)^{k-1}p, k \geq 1$, p a constant in $(0,1)$.*

For $n \geq 0$, define $X_n = 1$ if $\inf\{k \geq 0 : \sum_{i \geq 0}^{k} Z_i > n\}$ is an even number, and $X_n = 0$ otherwise. Similarly for $n < 0$, define $X_n = 0$ if $\inf\{k \geq 1 : \sum_{i=-k}^{-1} Z_i \geq -n\}$ is an odd number, and $X_n = 1$ otherwise.

1. *Show that $(X_n : n \in \mathbb{Z})$ is not stationary. (For the definition of stationarity, cf. Problem 3.)*
2. *Find r.v. W such that $(X_{n+W} : n \in \mathbb{Z})$ is stationary.*

Solution:

1. Notice that $X_0 = 1, X_{-1} = 0$, so X cannot be stationary.
2. It will prove convenient to define the following first. Let $(Y'_n : n \in \mathbb{Z})$ be Bernoulli(p), i.e., Y'_n are i.i.d. with $Y'_n = 1$ w.p. p and $Y'_n = 0$ w.p. $1-p$. Also, let $X'_0 = 0$ w.p. $1/2$ and $X'_0 = 1$ w.p. $1/2$, independent of Y'. Now for $n > 0$, define $X'_n = X'_0$ if $\sum_{i=1}^{n} Y'_i$ is even, and $X'_n = 1 - X'_0$ otherwise. Similarly, define $X'_n = X'_0$ if $\sum_{i=-n}^{0} Y'_i$ is even and $X'_n = 1 - X'_0$ otherwise.

   Then it is easy to show that X' is stationary. E.g., note that $P(X'_n = 0) = P(X'_n = 1) = 1/2$, by symmetry. For $n < 0$, $P(X'_n = 0, X'_{n+1} = 1) = P(X'_n = 0 | X'_{n+1} = 1) P(X'_{n+1} = 1) = p/2$ which does not depend on n, and so on...

   Now, we will construct a r.v. W such that $(X_{n+W} : n \in \mathbb{Z})$ has the same distribution as X', thereby showing that the former is stationary too.

   Let $W_0 = \text{Geometric}(p)$ independent of all alse, and

$$W = \begin{cases} W_0 & \text{w.p. } 1/2 \\ W_0 + Z_0 & \text{w.p. } 1/2\,, \end{cases}$$

   independent of all else. Define $Y_n = |X_n - X_{n-1}|$, for all $n \in \mathbb{Z}$.

   First we show that $(Y_{n+W_0} : n \in \mathbb{Z})$ is stationary.

$$\begin{aligned} P(Y_n = 0) &= \sum_{k=0}^{\infty} P(\sum_{i=0}^{k} Z_i < n < n+1 \leq \sum_{i=0}^{k+1} Z_i) \\ &= \sum_{k=0}^{\infty} P[Z_{k+1} \geq n - \sum^{k} Z_i + 1 | \sum^{k} Z_i < n \leq \sum^{k+1} Z_i] P(\sum^{k} Z_i < n \leq \sum^{k+1} Z_i) \\ &= \sum_{k=0} P(Z_0 \geq 2) P(\sum^{k} Z_i < n \leq \sum^{k+1} Z_i) \\ &= 1 - p\,. \end{aligned}$$

Also, from the above calculations, it is obvious that the $Y_n's$ are independent. Hence, $(Y_n : n \in \mathbb{Z}, n \neq 0)$ are i.i.d. 0-1 Bernoulli with parameter p. Now, let $k_0 = \inf\{k : W_0 < \sum_{i=0}^{k} Z_i\}$, and define $\tilde{Z}_0 = \sum_{i=0}^{k_0} Z_i - W_0, \tilde{Z}_1 = Z_{k_0+1}, \tilde{Z}_2 = Z_{k_0+2}, \ldots$ and $\tilde{Z}_{-1} = W_0 - \sum_{i=0}^{k_0-1} Z_i, \tilde{Z}_{-2} = Z_{k_0-2}, \ldots$. Then, all $\tilde{Z}$ are i.i.d. geometric with parameter p. We need to show it only for $\tilde{Z}_0, \tilde{Z}_{-1}$: for $m > 0$,

$$\begin{aligned} P(\tilde{Z}_{-1} \geq m) &= P[W \geq \sum_{i=0}^{k_0} Z_i + m | \sum_{}^{k_0} Z_i < W, W \leq \sum^{k_0+1} Z_i] \\ &= \sum_{k=0}^{\infty} P[W \geq \sum_{i=0}^{k} Z_i + m | \sum^{k} Z_i < W, W \leq \sum^{k+1} Z_i, k_0 = k] P(k_0 = k) \\ &= P(W \geq m)\,. \end{aligned}$$

The proof for $\tilde{Z}_0$ is easier -the distribution of W does not matter-, so we ommit it.

But then (Y_{n+W}) has the same distribution as (Y'_n) because the number of Y_n between 1's is (i.i.d.) geometric with the same parameter, p. Also, $P(X_{W_0} = i) = P(X_{W_0+Z_0} = 1-i)$ for $i = 0, 1$, so $X_W = 0$ or 1 w.p. $1/2$ independently of the Y_{n+W} sequence. Hence $(X_W, (Y_{n+W} : n \in \mathbb{Z}))$ has the same distribution as $(X'_0, (Y'_n : n \in \mathbb{Z}))$. Recall that $Y_{n+W} = |X_{n+W} - X_{n+W-1}|$ for all $n \in \mathbb{Z}$, so $(X_{n+W} : n \in \mathbb{Z})$ is uniquely determined by $(X_W, (Y_{n+W} : n \in \mathbb{Z}))$. Similarly, $(X'_n : n \in \mathbb{Z})$ is determined *in the same way* by $(X'_0, (Y'_n : n \in \mathbb{Z}))$. Thus, the distribution of $(X_{n+W} : n \in \mathbb{Z})$ is the same as that of $(X'_n : n \in \mathbb{Z})$.

Chapter 13

Markov Chains - Discrete Time

13.1 Summary

The book [3] is very clear, so my notes will be very succinct. You can also check [10] for examples. Here are the key ideas of this lecture.

- Definition of Markov chain;
- Passage times
- Strong Markov Property
- irreducible; transient or recurrent.
- Examples; Random Walk.

13.2 Definitions

A Markov chain models the random memoryless evolution of an object in a countable set.

Definition 13.1. *Distribution, Transition Matrix*

Let $\mathcal{X}$ be a finite or countable set. A distribution *on $\mathcal{X}$ is a collection $\pi = \{\pi_i, i \in \mathcal{X}\}$ of nonnegative numbers that sum to one.*

A transition probability matrix*, or* transition matrix*, or* stochastic matrix *on $\mathcal{X}$ is a collection $P = \{P(i,j), i,j \in \mathcal{X}\}$ of nonnegative numbers such that $\sum_{j\in\mathcal{X}} P(i,j) = 1$ for all $i \in \mathcal{X}$. We think of P as a matrix, even when $\mathcal{X}$ is infinite.*

Definition 13.2. *Markov Chain*

Let $\mathcal{X}$ be a finite or countable set, π_0 a distribution of $\mathcal{X}$, and P a transition matrix on $\mathcal{X}$. A sequence of random variables $X = \{X_n, n \geq 0\}$ taking values in $\mathcal{X}$ is a Markov chain *with initial distribution π_0 and probability transition matrix P if*

$$P(X_0 = i) = \pi_0(i), \forall i \in \mathcal{X} \tag{13.1}$$

and

$$P[X_{n+1} = j | X_n = i, X_m, m \leq n-1] = P(i,j), \forall i,j \in \mathcal{X} \text{ and } n \geq 0. \tag{13.2}$$

The possible values of X_n are called *states*. In this definition, (13.1) states that π_0 is the distribution of X_0. The identities (13.2) express that the evolution of X_n starts afresh from X_n at time n, independently of the values prior to time n. This is the *Markov property* that says that the past and the future are independent given the present state.

Definition 13.3. *Irreducibility*

A Markov chain with transition matrix P is said to be irreducible *if it can go from any i to any other j in $\mathcal{X}$.*

Note that

$$\begin{aligned} P[X_{n+1}=j|X_0=i] &= \sum_{k\in\mathcal{X}} P[X_{n+1}=j, X_n=k|X_0=i] \\ &= \sum_{k\in\mathcal{X}} P[X_{n+1}=j|X_n=k, X_0=i]P[X_n=k|X_0=i] \\ &= \sum_{k\in\mathcal{X}} P[X_n=k|X_0=i]P(k,j). \end{aligned}$$

This identity shows, by induction on n, that

$$P[X_n=j|X_0=i] = P^n(i,j), \forall i,j \in \mathcal{X}. \tag{13.3}$$

This expression shows the following fact.

Fact 13.1. *The Markov chain is irreducible if and only if*

$$\sum_{n\geq 0} P^n(i,j) > 0, \forall i,j \in \mathcal{X}.$$

We first define some random times.

Definition 13.4. *First Return Time and First Passage Time*

Let X_n be a Markov chain on $\mathcal{X}$. The first return time *to state i is τ_i defined as follows:*

$$\tau_i := \min\{n>0|X_n=i\}.$$

The first passage time *of state i, T_i, is defined as follows:*

$$T_i := \min\{n\geq 0|X_n=i\}.$$

We extend these definitions to T_S and τ_S for $S \subset \mathcal{X}$.

Note that $\tau_i = T_i$ if $X_0 \neq i$. However, if $X_0 = i$, $T_i = 0 < \tau_i$.

Theorem 13.1. *Strong Markov Property*

Consider a MC $(X_n, n \geq 0)$ with transition matrix P. Assume $P[T_i < \infty | X_0 = k] = 1$, for some i, k. Then

$$P[X_{T_i+1}=j|X_0=i] = P(i,j), \forall j \in \mathcal{X}.$$

That is, the evolution of X starts afresh from i at the random time T_i. In particular, the times between successive returns to i are i.i.d.

Proof: Observe that

$$\begin{aligned}
P[X_{T_i+1} = j | X_0 = k] &= P[X_{T_i+1} = j, T_i < \infty | X_0 = k] \\
&= \sum_{t=0}^{\infty} P[X_{t+1} = j, T_i = t | X_0 = k] \\
&= \sum_{t=0}^{\infty} P[X_{t+1} = j | X_t = i, X_s \neq i \forall s = 0, \ldots, t-1, X_0 = k] P[T_i = t | X_0 = k] \\
&= \sum_{t=0}^{\infty} P[X_{t+1} = j | X_t = i] P[T_i = t | X_0 = k] \\
&= P(i,j) \sum_{t=0}^{\infty} P[T_i = t | X_0 = k] = P(i,j) P[T_i < \infty | X_0 = k] \\
&= P(i,j) \, .
\end{aligned}$$

□

Definition 13.5. *We now define some key properties of a Markov chain.*

- *Transient: State i is transient if $P[\tau_i < \infty | X_0 = i] < 1$.*
- *Recurrent: State i is recurrent if $P[\tau_i < \infty | X_0 = i] = 1$.*
- *Positive Recurrent: State i is PR if $E[\tau_i | X_0 = i] < \infty$.*
- *Null Recurrent: State i is NR if it is recurrent and $E[\tau_i | X_0 = i] = \infty$.*
- *Infinitely Often: $\{X_n = i, \ i.o. \ \} = \{\omega \mid X_n(\omega) = i \text{ for infintely many } n's\} = \{\omega \mid \sum_{n=0}^{\infty} 1\{X_n(\omega) = i\} = \infty\}$.*

We state the following results that we leave you as an exercise to prove:

Fact 13.2. *State i is transient if*

$$P[X_n = i, \ i.o. \ | X_0 = i] = 0.$$

Theorem 13.2. *Solidarity*

(a) If (X_n) is irreducible M.C. and has some recurrent state, all states must be recurrent.

(b) Moreover, if it has some positive recurrent state, then all states must be positive recurrent.

Proof:

(a) Assume state j is recurrent, and let i be any other state. We use the notation in Problem 2, with $T_i^0 = 0$. For any $k \geq 0$, define

$$V_k = \sum_{n=T_j^k+1}^{T_j^{k+1}} 1\{X_n = i\} \, .$$

Now since the past and the present are conditionally independent on knowing the present state, by the Strong Markov Property, if we condition on $X_{T_j^k}$, we see that $V_0, V_1, \ldots, V_{k-1}$ are

independent of V_k. But they are also, *unconditionally* independent since $X_{T_j^k} = j$. Thus, (V_k) are mutually independent. They are also identically distributed since $P[V_k = v] = P[V_k = v|X_{T_j^k} = j] = P[V_0 = v|X_0 = j]$.

Because of irreducibility, $p := p(j, s_1)p(s_1, s_2)\cdots p(s_L, i) > 0$ for some $L > 0$, and states $s_1, \ldots, s_L$, with $s_m \neq j$ for all $m = 1, \ldots L$. But $P[V_0 > 0|X_0 = j] \geq p$, since $V = 0 > 0$ can happen by taking the path $j, s_1, \ldots, s_L, i$.

Also observe,

$$\sum_{k=0}^{\infty} V_k \geq \sum_{k=0}^{\infty} 1\{V_k > 0\} = \infty\,,$$

where the last equality holds because $1\{V_k > 0\}$ are i.i.d. with $P(V_k > 0) = p > 0$. [We cannot have $\sum_{k=0}^{\infty} 1\{V_k > 0\} < \infty$, since $\sum_{k=0}^{K} 1\{V_k > 0\}/K \to p > 0$.]

Therefore, $N_i = \infty$. This shows that all states are recurrent.

(b) Now assume moreover that j, positive recurrent.

$$\frac{1}{K}\sum_{k=1}^{K} V_k \geq \frac{1}{K}\sum_{k=1}^{K} Z_k \to \infty\,,$$

as $K \to \infty$, from the above. But LHS is also written as

$$\frac{T_j^K}{K}\frac{\sum_{k=1}^{K} V_k}{T_j^K}\,,$$

and by positive recurrence of j, $T_j^K/K \to E[T_j^1|X_0 = j]$, as $K \to \infty$. Hence, $\sum_{k=1}^{K} V_k/K$ must converge to a positive number. So we must have also,

$$\lim_K \frac{1}{K}\sum_{n=T_i^1+1}^{T_i^K} 1\{X_n = i\} > 0\,,$$

by recurrence of i. But,

$$\lim_K \frac{1}{K}\sum_{n=T_i^1+1}^{T_i^K} 1\{X_n = i\} = \frac{T_i^K}{K}\frac{\sum_{k=1}^{K}\sum_{n=T_i^k+1}^{T_i^{k+1}}}{T_i^K}\,, \tag{13.4}$$

Arguing in a similar way as in the recurrent case, we get

$$\lim_K \frac{1}{K}\sum_{k=1}^{K}\sum_{n=T_i^k+1}^{T_i^{k+1}} 1\{X_k = i\} > 0\,.$$

By comparing the above with (17.1), we get that T_i^K/K must converge to a positive limit as well. This limit must be $E[T_i^1|X_0 = i]$ by the strong law of large numbers, so state i is positive recurrent.

□

13.3 Example

We look at one example:

Fact 13.3. *Assume $P[X_{n+1} = i+1|X_n = i] = p$ and $P[X_{n+1} = i-1|X_n = i] = q = 1-p$ for $i \in \mathcal{X} = \mathbb{Z}$. Then all the states are transient if $p \neq 1/2$.*

Proof:

Note that $X_{n+1} = X_n + V_n$ where $\{V_0, V_n, n \geq 0\}$ are independent and $P(V_n = 1) = p = 1 - P(V_n = -1)$. Consequently, of $X_0 = 0$, $X_n = V_0 + V_1 + \cdots + V_{n-1}$. Hence,

$$\frac{X_n}{n} = \frac{V_0 + V_1 + \cdots + V_{n-1}}{n} \to E(V_1) = p - q \neq 0.$$

This shows that $X_n \to +\infty$ if $p > 1/2$ and $X_n \to -\infty$ if $p < 1/2$. Now, if $X_n(\omega) \to \infty$, this implies that $X_n > 10$ for all $n \geq n(\omega)$ where $n(\omega) < \infty$. Consequently,

$$\sum_{n \geq 0} 1\{X_n = 0\} < n(\omega) < \infty.$$

Since this happens with probability one, we see that

$$P[\sum_{n \geq 0} 1\{X_n = 0\} < n(\omega) < \infty | X_0 = 0] = 1,$$

which shows, by definition, that 0 is transient. By symmetry, all states are transient. The same argument applies for $p < 1/2$.

□

13.4 Solved Problems

Problem 13.1. *Show that an irreducible Markov chain (M.C.) (X_n) with finite state space cannot be transient.*

Solution:
Assume MC is transient. Then $\sum_{n=0}^{\infty} 1\{X_n = i\} < \infty$ for any state i. But this is not possible, since $\sum_i \sum_{n=0}^{N} 1\{X_n = i\} \geq N \to \infty$ as $N \to \infty$.

Problem 13.2. *Let X be a MC and define, for a fixed i, $p = P[X_n = i \text{ for some finite } n > 0 | X_0 = i]$. Show that*

$$p < 1 \Rightarrow P[X_n, \ i.o. \ |X_0 = i] = 0$$

and

$$p = 1 \Rightarrow P[X_n, \ i.o. \ |X_0 = i] = 1.$$

Solution:
For any state i, define $N_i = \sum_{n=0}^{\infty} 1\{X_n = i\}$, the number of visits to state i. Also, let T_i^k be the time of the $k-th$ visit to state i, where if $N_i < k$ we let $T_i^k = \infty$. Then for any j,

$$\begin{aligned}
P[N_j \geq k+1 | X_0 = i] &= P[N_j \geq k+1, T_j^k | X_0 = i] \\
&= P[N_j \geq k+1 | X_{T_j} = j, T_j^k < \infty, X_0 = i] P[T_j^k < \infty | X_0 = i] \\
&= P[T_j^{k+1} < \infty | X_{T_j} = j] P[N_j \geq k | X_0 = i] \\
&= P[T_j^1 < \infty | X_0 = j] P[N_j \geq k | X_0 = i] \\
&= (P[T_j^1 < \infty | X_0 = j])^{k-1} P[T_j^1 < \infty | X_0 = i]
\end{aligned}$$

For $j = i$ we have

$$P[N_i = \infty | X_0 = i] = \lim_k P[N_i \geq k+1 | X_0 = i] = \lim_k (P[T_i^1 < \infty | X_0 = i])^k = \lim_k p^k = 0 \text{ or } 1\,,$$

depending on whether $p < 0$ or $p = 1$.

Now observe that $\{X_n = i \text{ i.o.}\} = \{N_i = \infty\}$.

Problem 13.3. *Show that if (X_n) is an irreducible M.C. and has one positive recurrent state, all states must be positive recurrent.*

Solution:
This follows from the result of the previous problem.

Problem 13.4. *Show that if the M.C. (X_n) is irreducible and positive reccurent, then*

$$\lim_n \frac{1}{n} \sum_{m=0}^{n} 1\{X_m = i\} = \frac{1}{E[\tau_i | X_0 = i]}\,,$$

where $\tau_i = \min\{n > 0 : X_n = i\}$, the time of first return *to state i.*

Solution:
We use the same notation as above. Notice that

$$\begin{aligned} P[T_i^{n+1} - T_i^n > t | T_i^n, T_i^{n-1}, \ldots] &= P[T_i^{n+1} - T_i^n > t | X_{T_i^n = i}] \\ &= P[\tau_i > t | X_0 = i]\,, \end{aligned}$$

so the times taken for excursions out of state i and back again are i.i.d. with finite mean, by positive recurrence.

Now, $T_i^n < \infty$ almost surely, since otherwise $E[\tau_i | X_0 = i] = \infty$, a fact that contradicts positve recurrence. Also, $T_i^{n+1} \geq T_i^n + 1$, so $T_i^n \uparrow \infty$. Now,

$$\frac{1}{T_i^n} \sum_{m=0}^{T_i^n} 1\{X_m = i\} = \frac{1+n}{T_i^n} \to \frac{1}{E[\tau_i | X_0 = i]}\,,$$

by the strong law of large numbers. Thus,

$$\lim_n \frac{1}{n} \sum_{m=0}^{n} 1\{X_m = i\} = \lim_n \frac{1}{T_i^n} \sum_{m=0}^{T_i^n} 1\{X_m = i\} = \frac{1}{E[\tau_i | X_0 = i]}\,.$$

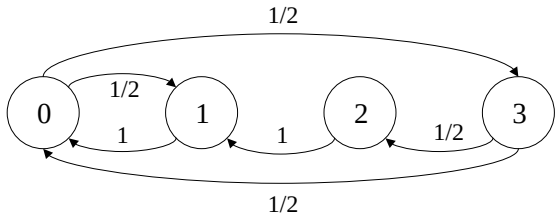
Problem 13.5. *Consider the M.C. (X_n) with state-transition probabilities given by the diagram.*

Find $E[T_2 | X_0 = 0]$, where $T_i = \min\{n \geq 0 : X_0 = i\}$ is the first passage *time to state i.*

Solution:
For any state i, we define $\beta(i) = E[T_2 | X_0 = i]$. By Kolmogorov's equations:

$$\begin{aligned} \beta(2) &= 0 \\ \beta(3) &= \frac{1}{2}(1 + \beta(2)) + \frac{1}{2}(1 + \beta(0)) \\ \beta(1) &= 1 + \beta(0) \\ \beta(0) &= \frac{1}{2}(1 + \beta(3)) + \frac{1}{2}(1 + \beta(1))\,, \end{aligned}$$

giving $\beta(0) = 8$.



Problem 13.6. *For the M.C. of Problem 5, find $P[X_n$ hits 2 before hitting 1 twice$|X_0 = 0]$.*

Solution:
According to the notation of problem 2, $P[X_n$ hits 2 before hitting 1 twice$|X_0 = 0] = P[T_1^2 < T_2^1|X_0 = 0]$. Now,

$$P[T_2^1 < T_1^2|X_0 = 0] = P[T_2^1 < T_1^1|T_1^1 < T_2^1, X_0 = 0]P[T_1^1 < T_2^1|X_0 = 0] + P[T_2^1 < T_1^2, T_1^1 > T_2^1|X_0 = 0] .$$

But,

$$\begin{aligned} P[T_2^1 < T_1^2|T_1^1 < T_2^1, X_0 = 0] &= P[T_2^1 < T_1^2|X_{T_1^1} = 1, X_n \neq 2, n = 1, \ldots, T_1^1, X_0 = 0] \\ &= P[T_2^1 < T_1^2|X_{T_1^1} = 1] \\ &= P[T_2^1 - T_1^1 < T_1^2 - T_1^1|X_{T_1^1} = 1] \\ &= P[T_2^1 < T_1^1|X_0 = 0] , \end{aligned}$$

and

$$P[T_2^1 < T_1^2, T_1^1 > T_2^1, X_0 = 0] = P[T_2^1 < T_1^1|X_0 = 0] =: a .$$

Thus, $P[T_2^1 < T_1^2|X_0 = 0] = a(1 - a) + a$. To compute a, we use Kolmogorov's equations: let $\alpha(0) = P[T_2^1 < T_1^1|X_0 = i]$, and notice that $a = \alpha(0) = \alpha(3)/2, \alpha(3) = 1/2 + \alpha(0)/2$, giving $a = 1/3$. Thus, $P[T_2^1 < T_1^2|X_0 = 0] = 5/9$.

Problem 13.7. *Assume you observe a M.C. $X_0, X_1, \ldots$. How would you estimate $P(i, j)$ for a given pair (i, j)?*

Solution:
Assume X_n is a positive recurrent chain. We have

$$\frac{1}{N}\sum_{n=0}^{N} 1\{X_n = i, X_{n+1} = j\} \to P(i, j) , \text{ as } N \to \infty .$$

The see why this is true, notice that $((X_n, X_{n+1}, n \geq 0)$ defines a positive reccurent chain.

If we empirically estimate the fraction of times the transition $i \to j$ is taken, it will converge to the unknown $P(i, j)$.

Chapter 14

Markov Chains - Part 2

14.1 Summary

Here are the key ideas of this lecture.

- Function of Markov chain.
- Kolmogorov Equations
- Reflected Random Walk.

14.2 Function of Markov Chain

Fact 14.1. *Generally, a function of a Markov chain is not a Markov chain.*

We look at a few examples.

Example 14.1. *Consider the Markov chain X with $X_{n+1} = (X_n + 1) mod(3)$ where X_0 is uniformly distributed in $\{0, 1, 2\}$. Let $f(x) = 0$ for $x = 0, 1$ and $f(2) = 1$. Then $Y_n = f(X_n)$ is not a Markov chain. Indeed,*

$$P[Y_2 = 0|Y_1 = 0, Y_0 = 0] = 0 \neq P[Y_2 = 0|Y_1 = 0] = \frac{1}{2}.$$

Intuitively, $f(X_n)$ contains less information than X_n and it may happen that $\{f(X_m), m \leq n\}$ has information about X_n that $f(X_n)$ does not contain, as the example shows.

Example 14.2. *As a trivial example where $f(X_n)$ is a Markov chain even though $f(\cdot)$ is not one-to-one, take f to be constant.*

14.3 Kolmogorov Equations

First passage times satisfy simple first step equations that we can use for their analysis.

Fact 14.2. *Kolmogorov Equations*
(a) One has, for any $A, B \subset \mathcal{X}$ and $i \in \mathcal{X}$,

$$\alpha(i) := P[T_A < T_B|X_0 = i] = \begin{cases} \sum_j P(i,j)\alpha(j), & \text{if } i \notin A \cup B; \\ 1, & \text{if } i \in A \setminus B; \\ 0, & \text{if } i \in B. \end{cases}$$

(b) One has, for any $A \subset \mathcal{X}$ and $i \in \mathcal{X}$,

$$\beta(i) := E[T_A | X_0 = i] = \begin{cases} 1 + \sum_j P(i,j)\beta(j), & \text{if } i \notin A; \\ 0, & \text{if } i \in A. \end{cases}$$

(c) Assume that $\mathcal{X} = \mathbb{Z}$ or $\mathcal{X} = \mathbb{Z}_+$ and that $T_i \uparrow \infty$ as $i \to \infty$. Then, if $i \notin A$,

$$P[T_A < T_i | X_0 = j] \uparrow P[T_A < \infty | X_0 = j] \text{ as } i \to \infty$$

and

$$E[\min\{T_A, T_i\} | X_0 = j] \uparrow E[T_A | X_0 = j] \text{ as } i \to \infty.$$

Proof:

Everything is obvious, except for (c) which relies on part (a) of the following result.

□

Theorem 14.1. *Lebesgue Convergence Theorem*

(a) Assume that $0 \le X_n \uparrow X$ as $n \to \infty$. Then $E(X_n) \uparrow E(X)$ as $n \to \infty$.

(b) Assume that $X_n \to X$ as $n \to \infty$ and $|X_n| \le Y$ with $E(Y) < \infty$. Then $E(X_n) \to E(X)$ as $n \to \infty$.

Remark. You might be tempted to think that $X_n \to X$ as $n \to \infty$ implies $E(X_n) \to E(X)$ as $n \to \infty$. However, this is not true as the following example illustrates.

Example 14.3. *Let $\Omega = (0, 1]$ and assume that ω is picked uniformly in Ω. For $n \ge 1$, let $X_n(\omega) = n1\{\omega \le 1/n\}$. Then $P(X_n = n) = 1/n = 1 - P(X_n = 0)$, so that $E(X_n) = n \times (1/n) = 1$ for $n \ge 1$. Moreover, $X_n \to X = 0$ as $n \to \infty$. However, $E(X_n) = 1$ does not converge to $E(X) = 0$ as $n \to \infty$. You see that this example violates both assumptions of Lebesgue's theorem.*

14.4 Random Walk

Fact 14.3. *Assume $P[X_{n+1} = i + 1 | X_n = i] = p$ and $P[X_{n+1} = i - 1 | X_n = i] = q = 1 - p$ for $i \in \mathcal{X} = \mathbb{Z}$. Here, $p \in (0, 1)$.*

(a) If $p \ne 0.5$, the Markov chain is transient.

(b) If $p = 0.5$, the Markov chain is null recurrent.

Proof:

(a) Assume $p < 0.5$. Fix $a, b \in \{1, 2, 3, \ldots\}$. Define $\alpha(i) = P[T_a < T_{-b} | X_0 = i], -b \le i \le a$. Then, according to Kolmogorov's equations,

$$\alpha(i) = \begin{cases} \sum_j P(i,j)\alpha(j), & \text{if } i \notin A \cup B; \\ 1, & \text{if } i = a; \\ 0, & \text{if } i = -b. \end{cases} \tag{14.1}$$

Solving these equations, assuming $p < 0.5$, we find

$$P[T_a < T_{-b} | X_0 = i] = \frac{\rho^i - \rho^{-b}}{\rho^a - \rho^{-b}}, -b \le i \le a \text{ where } \rho := \frac{q}{p} > 1. \tag{14.2}$$

Now we let $b \to \infty$. We use part (c) of Fact 14.2 to conclude that

$$P[T_a < \infty | X_0 = i] = \rho^{i-a}, i \leq a.$$

Since this value is less than 1, we see that

$$P[\tau_a < \infty | X_0 = a] = pP[T_a < \infty | X_0 = a+1] + qP[T_a < \infty | X_0 = a-1] < 1,$$

so that a is transient.

(b) Assume $p = 0.5$. In that case, the solution of (14.1) is

$$P[T_a < T_{-b} | X_0 = i] = \frac{i+b}{a+b}.$$

Consequently, as $b \to \infty$, we find

$$P[T_a < \infty | X_0 = i] = 1$$

and we conclude that every state is recurrent. To show positive recurrence, we solve the Kolmogorov equations for $\beta(i) = E[\min\{T_a, T_{-b}\} | X_0 = i]$. We find

$$E[\min\{T_a, T_{-b}\} | X_0 = i] = (a-i)(i+b), -b \leq i \leq a.$$

Letting $b \to \infty$, we find

$$E[T_a | X_0 = i] = \infty, i \neq a,$$

which proves null recurrence.

□

Chapter 15

Markov Chains - Part 3

15.1 Summary

Here are the key ideas of this lecture.

- Stationary; Invariant Distribution
- Classification theorem.

15.2 Stationary; Invariant Distribution

Definition 15.1. *We define*

- *Invariant Distribution π: A nonnegative solution of $\pi P = \pi$ and $\pi \mathbf{1} = 1$.*
- *Stationary: $X = \{X_n, n \geq 0\}$ is stationary if it has the same finite dimensional distributions as $\{X_{n+m}, n \geq 1\}$ for all m, i.e., if*

  $$P(X_0 \in A_0, \ldots, X_n \in A_n) = P(X_m \in A_0, \ldots, X_{m+n} \in A_n), \forall n, m \in \mathbb{Z}_+ \text{ and } A_0, \ldots, A_n \subset \mathcal{X}.$$

We start with a simple fact.

Fact 15.1. *(a) Let $\pi_n(i) := P(X_n = i), i \in \mathcal{X}$ and define π_n as the row vector with components $\pi_n(i)$. Let also P be the matrix with entries $P(i,j)$ for $i, j \in \mathcal{X}$. Then*

$$\pi_{n+1} = \pi_n P, n \geq 0.$$

(b) $X = \{X_n, n \geq 0\}$ is stationary if and only if π_0 is invariant.

Example 15.1. *We look at a few examples of MC with many, zero, and one invariant distribution. We also examine the empirical distribution and the behavior of π_n as $n \to \infty$ for these examples.*

The examples above suggest the classification theorem.

15.3 Classification Theorem

Theorem 15.1. *(a) If a Markov chain is irreducible, then all the states have the same recurrence properties (T, NR, or PR). The MC is then said to be T, NR, or PR, correspondingly.*

(b) If a MC is irreducible, then it is T if and only if $P[X_n = i, \text{ i.o. } |X_0 = j] = 0, \forall i, j$.

(c) If a MC is irreducible, then it is R if and only if $P[X_n = i, \text{ i.o. } |X_0 = j] = 1, \forall i, j$.

(d) If an irreducible MC is T or NR, then

$$\frac{1}{N}\sum_{n=1}^{N} 1\{X_n = i\} \to 0, \forall i \in \mathcal{X}.$$

Moreover, there is no invariant distribution. A finite irreducible MC cannot be T or NR.

(e) If an irreducible MC is PR, then

$$\frac{1}{N}\sum_{n=1}^{N} 1\{X_n = i\} \to \pi(i) > 0, \forall i \in \mathcal{X}.$$

Moreover, π *is the unique invariant distribution. A finite irreducible MC is necessarily PR.*

(f) If the MC is irreducible, PR, and aperiodic, then

$$P(X_n = i) \to \pi(i), \forall i \in \mathcal{X}.$$

Proof:

The basic steps for (a)-(c) are sketched in HW7.

(d)-(e). By considering the successive visits to i we find cycles with iid durations τ_n. Thus, with $\sigma_n = \sum_{m=1}^{n} \tau_m$ and $S_n = \sum_{k=0}^{\sigma_n} 1\{X_k = i\}$, we have

$$\frac{S_n}{\sigma_n} = \frac{(S_n/n)}{(\sigma_n/n)} \to \frac{1}{E[\tau|X_0 = i]}.$$

Thus,

$$\frac{1}{n}\sum_{m=0}^{n} 1\{X_k = i\} \to \frac{1}{E[\tau|X_0 = i]}.$$

Moreover, this limit does not depend on X_0 since X reaches i in finite time.

Now, if π is invariant and if we choose $\pi_0 = \pi$, then we find

$$\pi(i) = E(\frac{1}{n}\sum_{m=0}^{n} 1\{X_k = i\}) \to \frac{1}{E[\tau|X_0 = i]}$$

where the limit is justified by the Lebesgue convergence theorem (b).

(f) Imagine X and X' as two independent Markov chains with ptm P. We claim that if P is irreducible, PR, and aperiodic, then $Y_n = (X_n, X'_n)$ is irreducible. To see this, take two pairs of states (i, i') and (j, j'). There is some finite n such that $P^n(i, i') > 0$ and $P^n(j, j') > 0$. This follows from the fact that P is aperiodic (see Lemma (d) below). Moreover, this MC admits an invariant distribution $\pi(i)\pi(i')$, so that it is positive recurrent. Modify slightly the construction so that X starts with the invariant distribution π and X' starts from any given distribution π'. Moreover, assume that X and X' stick together after they first meet and designate their first

meeting time by τ, which does not change the fact that both X and X' are Markov chain with ptm P. We know that τ is finite, by recurrence. Then

$$|\pi(i) - P(X'_n = i)| = |P(X_n = i) - P(X'_n = i)| \leq P(\tau \geq n) \to 0. \tag{15.1}$$

The limit holds because τ is finite, so that $1\{\tau \geq n\} \to 0$ as $n \to \infty$; the expectation goes to zero by the Lebesgue convergence theorem (b). To see the inequality, note that

$$|1\{X_n = i\} - 1\{X'_n = i\}| \leq 1\{\tau \geq n\}.$$

Indeed, the left-hand side is zero if $\tau < n$ since then $X_n = Y_n$. Also, left-hand side is always less then or equal to 1. Hence,

$$1\{X_n = i\} - 1\{X'_n = i\} \leq 1\{\tau \geq n\},$$

which implies, by taking expectation, that

$$P(X_n = i) - P(X'_n = i) \leq P(\tau \geq n).$$

Similarly, we get $P(X'_n = i) - P(X_n = i) \leq P(\tau \geq n)$, which, together with the previous inequality, proves (15.1).

□

It remains to show the irreducibility of Y we used in (f). That result follows from the following lemma.

Lemma 15.1. *(a) Assume p and q are coprime. Then $1 = mp + nq$ for some $m, n \in \mathbb{Z}$.*

(b) Assume $\{p_1, p_2, \ldots, p_k\}$ are coprime, then there are some integers m_i such that $1 = m_1 p_1 + \ldots + m_k p_k$.

(c) Under the assumption of (b), there must be two consecutive integers N and $N+1$ that are positive integer linear combinations of the p_i.

(d) Again under the same assumption, all integers larger than some n_0 are positive integer linear combinations of the p_i.

Proof:

(a) For two numbers $a \neq b$, define $f(a, b) = (a - b, b)$ if $a > b$ and $(a, b - a)$ if $a < b$. We stop if $a = b$. Starting with $(a, b) = (p, q)$ and applying f repeatedly, one must end up with $a = 1$ or $b = 1$. To see that, assume that we end up with $a = b > 1$. At the previous step, we had $(2a, a)$ or $(a, 2a)$. At the step before, we had $(2a, 3a)$, and so on. By induction, we see that p and q must be multiples of a, a contradiction. Thus, we end up with $a = 1$ or $b = 1$. But, at each step, a and b are integer combinations of p and q.

(b) We extend this idea to k coprime numbers. Starting with $p_1, \ldots, p_k$, at each step, we replace the largest number with the difference between it and the smallest number. We stop when the numbers are equal or when we reach a 1. At each step, we get integer multiples of the original numbers. As in (a), we claim that we must end up with some number equal to 1. Otherwise, all the original numbers must be multiples of the same $a > 1$, a contradiction.

(c) We know $1 = \sum_j m_j p_j = \sum_j (m_j)^+ p_j - \sum_j (m_j)^- p_j$. Thus, we can choose $N = \sum_j (m_j)^- p_j$ and $N + 1 = \sum_j (m_j)^+ p_j$.

(d) In the set of integer linear combinations of the p_i, we have N and $N + 1$. Therefore, for $m > N - 1$, we also have $mN, (m-1)N + (N+1) = mN + 1, (m-2)N + 2(N+1) =$

$mN+2, \ldots, (m-N+1)N + (N-1)N = mN+N$ and these sets cover all the integers larger than N^2.

□

15.4 Solved Problems

Problem 15.1. *Show that any MC* $(X_n, n \geq 0)$ *with transition matrix* P*, can be written as*

$$X_{n+1} = f(X_n, V_n)\,, n \geq 0\,,$$

for some function $f(\cdot,\cdot)$*, where* $(V_n, n \geq 0)$ *are i.i.d. r.v's, independent of* X_0*.*

Solution:
The idea is that V_n is a collection of die rolls, one for each state. For state i, the roll yields j with probability $P(i,j)$. Then $f(i, V_n)$ is the outcome of the roll that corresponds to state i.

Chapter 16

Markov Chains - Part 4

16.1 Summary

Here are the key ideas of this lecture.

- Reflected Random Walk
- Hidden Markov Chain: Estimation and MAP
- Time-Reversibility
- Time Reversal
- Guessing (π, P') - Kelly's Trick

16.2 Reflected Random Walk

We have the following definition.

Definition 16.1. *The* reflected random walk *is the MC X on $\mathcal{X} = \{0, 1, 2, \ldots\}$ with the following transition matrix where $p \in (0, 1)$:*

$$P(i,j) = \begin{cases} p, & \text{if } j = i+1 \\ q = 1-p, & \text{if } j = (i-1)^+ \end{cases}$$

Fact 16.1. *The reflected random walk is transient if $p > 1/2$, null recurrent if $p = 1/2$, and positive recurrent if $p < 1/2$.*

Proof:

(a) $p > 1/2$. Note that $X_{n+1} \geq X_n + V_n$ where $\{X_0, V_n, n \geq 0\}$ are independent and $P(V_n = 1) = p = 1 - P(V_n = -1)$. Consequently, $X_n \geq X_0 + V_0 + \cdots + V_n$. Dividing by n, letting $n \to \infty$, and using the SLLN, we find that $X_n \to \infty$. Thus, the reflected random walk is larger than the non-reflected walk and the latter goes to infinity.

(b) $p = 1/2$. We note that $P[T_0 < \infty | X_0 = i]$ and $E[T_0 | X_0 = i]$ for $i \geq 0$ are the same as for the non-reflected random walk that we studied in Fact 3 of Lecture 17. Since the non-reflected random walk is null recurrent, we see that so is X.

(c) $p < 1/2$. The equations $\pi P = \pi$ are

$$\begin{aligned} \pi(0) &= q\pi(0) + q\pi(1); \\ \pi(k) &= p\pi(k-1) + q\pi(k+1), k \geq 1. \end{aligned}$$

The solution is seen to be

$$\pi(k) = (1-\rho)\rho^k, k \geq 0 \text{ where } \rho := \frac{p}{q} < 1.$$

Since X is irreducible and has an invariant distribution, it must be positive recurrent.

□

16.3 Hidden Markov Chain

This class of models is useful in speech recognition and in digital communication.

16.3.1 Definition

Definition 16.2. *A* hidden Markov chain *is a pair or processes (X, Y) such that X is a MC on $\mathcal{X}$ with transition probabilities P and*

$$P[Y_n = j | X_n = i; X_m, Y_m, m \neq n] = Q(i,j), i \in \mathcal{X}, j \in \mathcal{Y}.$$

In this definition, Q is a nonnegative matrix whose rows add up to one.

16.3.2 Estimation

Assume we observe Y. What does that tell us about X? Here is one formulation.

Fact 16.2. *Let $Y^n = (Y_0, \ldots, Y_n)$ for $n \geq 0$ and*

$$\hat{\pi}_n(i) := P[X_n = i | Y^n], i \in \mathcal{X}, n \geq 0.$$

Then

$$\hat{\pi}_{n+1}(i) = \frac{\sum_{i'} \hat{\pi}_n(i') P(i',i) Q(i, Y_{n+1})}{\sum_{i''} \sum_{i'} \hat{\pi}_n(i') P(i',i'') Q(i'', Y_{n+1})}.$$

Proof:

The derivation is a straightforward application of Bayes' rule. Note that

$$P[X_{n+1} = i | Y^{n+1}] = \frac{\sum_{i'} P[X_{n+1} = i, Y_{n+1}, X_n = i' | Y^n]}{\sum_{i''} \sum_{i'} P[X_{n+1} = i'', Y_{n+1}, X_n = i' | Y^n]}.$$

Now,

$$\begin{aligned} P[X_{n+1} = i, Y_{n+1}, X_n = i' | Y^n] &= P[Y_{n+1} | X_{n+1} = i] P[X_{n+1} = i, Y_{n+1} | X_n = i'] P[X_n = i' | Y^n] \\ &= P(i',i) Q(i, Y_{n+1}) \hat{\pi}_n(i'). \end{aligned}$$

Putting these equalities together proves the fact.

□

You will note that these equations that update the conditional distribution of X_n given Y^n are recursive but are nonlinear. We can view them as a "nonlinear filter." In the case of the Kalman filter, if the random variables are all Gaussian, then the distribution of X_n given Y^n is $N(\hat{X}_n, \Sigma_n)$ where $\hat{X}_n$ is computed recursively from the observations. Here, the situation is similar, but the filter is nonlinear. Also, its dimension is the number of states of the Markov chain (minus one, since the probabilities add up to one).

16.3.3 MAP

One interesting question is the MAP of $X^n := \{X_0, \ldots, X_n\}$ of the stationary MC given $Y^n := \{Y_0, \ldots, Y_n\}$. We have

Fact 16.3. *The MAP of the hidden MC X given the observation Y is given as follows:*

$$MAP[X^n | Y^n = (j_0, \ldots, j_n)] = argmax_{i_0,\ldots,i_n} \pi(i_0)Q(i_0, Y_0)P(i_0, i_1)Q(i_1, Y_1) \cdots P(i_{n-1}, i_n)Q(i_n, Y_n). \tag{16.1}$$

Proof:

We have

$$P[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n | Y^n = (j_0, \ldots, j_n)] = \frac{P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n, Y_0 = j_0, \ldots, Y_n = j_n)}{P(Y_0 = j_0, \ldots, Y_n = j_n)}.$$

Also,

$$\begin{aligned} &P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n, Y_0 = j_0, \ldots, Y_n = j_n) \\ &\quad = P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) P[Y_0 = j_0, \ldots, Y_n = j_n | X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] \\ &\quad = \pi(i_0)Q(i_0, j_0)P(i_0, i_1)Q(i_1, j_1) \cdots P(i_{n-1}, i_n)Q(i_n, j_n). \end{aligned}$$

□

This fact leads to a simple shortest path algorithm called *Viterbi's Algorithm* for finding the MAP. One constructs a graph with one copy of $\mathcal{X}$ at each time $0, 1, \ldots, n$ and one additional state θ at time -1. There is a link from θ to every state i at time 0 and a link from every state at time $m-1$ to every state at time m, for $m = 1, \ldots, n$. The link lengths are defined as follows. From θ to i at time 0, the length is $d_0(i)$ and from i at time $m-1$ to i' at time m it is $d_m(i, i')$ where

$$d_0(i) = \ln(\pi(i_0)Q(i_0, j_0)) \text{ and } d_m(i, i') = \ln(P(i, i')Q(i', j_m)), m = 1, \ldots, n.$$

To find the longest path, one defines $D(i, m)$ as the length of the longest path from θ to i at time m in the graph. We see that

$$D(i, 0) = d_0(i) \text{ and } D(i, m) = \max_{i_{m-1}} \{D(i_{m-1}, m-1) + d_m(i_{m-1}, i)\}, m = 1, \ldots, n.$$

Solving these equations, we find the longest path to stage n. It is such that i_{m-1} achieves the maximum in the above identities and $D(i_m, m) = \max_i D(i, m)$. That longest path maximizes the logarithm of the expression in (16.1) and is the MAP.

Chapter 17

Poisson Process

17.1 Summary

Here are the key ideas.

- Exponential distribution.
- Poisson process.
- Poisson measure.
- Application to a queuing model.

17.2 Exponential Distribution

Recall the following definition.

Definition 17.1. *The random variable τ is exponentially distributed with rate $\lambda \geq 0$ if $P(\tau \geq t) = e^{-\lambda t}$ with $t \geq 0$. We write $\tau =_D Exp(\lambda)$. Equivalently, $\tau = \infty$ if $\lambda = 0$ and $f_\tau(t) = \lambda e^{-\lambda t} 1\{t \geq 0\}$ if $\lambda > 0$.*

Fact 17.1. *Assume $\tau =_D Exp(\lambda)$ for some $\lambda > 0$. Then*

(a) Characteristic Function and Moments:

$$E(e^{s\tau}) = \frac{\lambda}{\lambda - s}, \forall s \text{ with } Re(s) < \lambda$$

and

$$E(\tau^n) = n!\lambda^{-n}, \text{ for } n \geq 0.$$

(b) Scaling Property:

$$\tau =_D \frac{1}{\lambda}\xi \text{ where } \xi =_D Exp(1).$$

(c) Generation from Uniform:

$$\tau =_D -\frac{1}{\lambda}\ln(U) \text{ where } U =_D U[0,1].$$

(d) Memoryless Property:

$$P[\tau > t + s | \tau > t] = P(\tau > s), \forall s, t \geq 0.$$

Proof: (a) We find

$$E(e^{s\tau}) = \int_0^\infty e^{st}\lambda e^{-\lambda t}dt = \frac{\lambda}{\lambda - s}, \forall s \text{ with } \mathrm{Re}(s) < \lambda.$$

This identity implies that

$$E(\tau^n) = \frac{d^n}{ds^n}E(e^{s\tau})|_{s=0} = \frac{n!\lambda}{(\lambda - s)^{n+1}}|_{s=0} = n!\lambda^{-n} \text{ for } n \geq 0.$$

(b) Indeed,

$$P\left(\frac{1}{\lambda}\xi > t\right) = P(\xi > \lambda t) = e^{-1(\lambda t)} = e^{-\lambda t}, \forall t \geq 0.$$

(c) One has

$$P\left(-\frac{1}{\lambda}\ln(U) > t\right) = P(\ln(U) < -\lambda t) = P(U < e^{-\lambda t}) = e^{-\lambda t}, \forall t \geq 0.$$

(d) We find, for $s, t \geq 0$,

$$P[\tau > t + s | \tau > t] = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = P(\tau > s).$$

□

The following results are useful.

Fact 17.2. *Assume τ_1 and τ_2 are independent with $\tau_i =_D Exp(\lambda_i)$ for $i = 1$ and 2.*
(a) $\min\{\tau_1, \tau_2\} =_D Exp(\lambda_1 + \lambda_2)$.
(b) Also,

$$P[\tau_1 < \tau_2 | \min\{\tau_1, \tau_2\} = t] = P[\tau_1 \leq \tau_2 | \min\{\tau_1, \tau_2\} = t] = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \forall t \geq 0.$$

(c) Moreover,

$$E(\max\{\tau_1, \tau_2\}) = \frac{1}{\lambda_1 + \lambda_2}\left(1 + \frac{\lambda_1}{\lambda_2} + \frac{\lambda_2}{\lambda_1}\right).$$

Proof: (a) We have $P(\min\{\tau_1, \tau_2\} > t) = P(\tau_1 > t, \tau_2 > t) = P(\tau_1 > t)P(\tau_2 > t) = e^{-\lambda_1 t}e^{-\lambda_2 t} = e^{-(\lambda_1+\lambda_2)t}$.

(b) First we note that $P(\tau_1 \in (t, t+\epsilon), \tau_2 > t) \approx \epsilon\lambda_1 e^{-\lambda_1 t} \times e^{-\lambda_2 t} = \epsilon\lambda_1 e^{-(\lambda_1+\lambda_2)t}$. Then

$$P[\tau_1 < \tau_2 | \min\{\tau_1, \tau_2\} \in (t, t+\epsilon)] \approx \frac{P(\tau_1 \in (t, t+\epsilon), \tau_2 > t)}{P(\min\{\tau_1, \tau_2\} \in (t, t+\epsilon))} \approx \frac{\epsilon\lambda_1 e^{-(\lambda_1+\lambda_2)t}}{(\lambda_1 + \lambda_2)\epsilon e^{-(\lambda_1+\lambda_2)t}} = \frac{\lambda_1}{\lambda_1 + \lambda_2},$$

which implies the result.

(c) $E(\max\{\tau_1, \tau_2\}) = E(\min\{\tau_1, \tau_2\}) + E(\max\{\tau_1, \tau_2\} - \min\{\tau_1, \tau_2\})$. On the other hand,

$$\begin{aligned}
& E(\max\{\tau_1, \tau_2\} - \min\{\tau_1, \tau_2\} | \min\{\tau_1, \tau_2\}) \\
= \; & E(\tau_2 - \tau_1, \tau_1 < \tau_2 | \min\{\tau_1, \tau_2\}) + E(\tau_1 - \tau_2, \tau_2 \leq \tau_1 | \min\{\tau_1, \tau_2\}) \\
= \; & E(\tau_2 - \tau_1 | \tau_1 < \tau_2, \min\{\tau_1, \tau_2\})P[\tau_1 < \tau_2 | \min\{\tau_1, \tau_2\}] + E(\tau_1 - \tau_2 | \tau_2 \leq \tau_1, \min\{\tau_1, \tau_2\})P[\tau_2 \leq \tau_1 | \min\{\tau_1, \tau_2\}] \\
= \; & \frac{\lambda_1}{\lambda_1 + \lambda_2} \cdot \frac{1}{\lambda_2} + \frac{\lambda_2}{\lambda_1 + \lambda_2} \cdot \frac{1}{\lambda_1}.
\end{aligned}$$

The result follows from the fact that $E[E[X|Y]] = E[X]$. □

17.3 Poisson Process

Definition 17.2. *Let $\{\tau_n, n \in \mathbb{N}\}$ be i.i.d. Exp(λ). Define, for $t \geq 0$,*

$$N_t = \max\{n \geq 1 \mid T_n := \tau_1 + \cdots + \tau_n \leq t\} \text{ if } t \geq \tau_1 \text{ and } N_t = 0 \text{ if } t < \tau_1.$$

The process $\{N_t, t \geq 0\}$ is called a Poisson process with rate λ. *The random times $\{T_n, n \in \mathbb{N}\}$ are the* jump times *of the process. Thus, N_t is the number of jumps in $[0, t]$. We also say that N_t is a* counting process *since it counts jumps.*

Theorem 17.1. *Let $\{N_t, t \geq 0\}$ be a* Poisson process with rate λ.

(a) For any $s > 0$, $\{N_{t+s} - N_s, t \geq 0\}$ is a Poisson process with rate λ and is independent of $\{N_t, t \leq s\}$.

(b) For any $n \geq 2$ and $0 < t_1 < t_2 < \cdots < t_n$, the random variables $N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$ are mutually independent and Poisson distributed with respective means $\lambda t_1, \lambda(t_2 - t_1), \ldots, \lambda(t_n - t_{n-1})$.

(c) Given that $N_t = n$, the jump times $\{T_1, \ldots, T_n\}$ are the ordered values of n i.i.d. $U[0, t]$ random variables.

(d) Color each jump time T_n independently red with probability α and blue with probability $1 - \alpha$ where $\alpha \in (0, 1)$. For $t \geq 0$, let A_t be the number of red jumps and B_t the number of blue jumps in $[0, t]$. Then $\{A_t, t \geq 0\}$ and $\{B_t, t \geq 0\}$ are independent Poisson processes with respective rates $\alpha\lambda$ and $(1 - \alpha)\lambda$.

Proof: (a) Suppose for concreteness that by time s there have been n arrivals occurring at times $T_1, \ldots, T_n$. We know the waiting time for the $(n+1)$-th arrival must have $\tau_{n+1} > s - T_n$, but by the lack of memory property of the exponential distribution, $P[\tau_{n+1} > s - T_n + t | \tau_{n+1} > s - T_n] = P(\tau_{n+1} > t) = e^{-\lambda t}$. This shows that the distribution of the first arrival after s is Exp(λ) and independent of $T_1, \ldots, T_n$. Therefore, it is clear that $\tau_{n+1}, \tau_{n+2}, \ldots$ are independent of $T_1, \ldots, T_n$ and τ_n, and the interarrival times after s are independent Exp(λ).

(b) The independence follows from (a). To get the distribution, it suffices to focus on N_t. We find $g(n, t + \epsilon) := P(N_{t+\epsilon} = n) \approx P(N_t = n)(1 - \lambda\epsilon) + P(N_t = n - 1)(\lambda\epsilon) + o(\epsilon)$. Thus, $g(n, t + \epsilon) = g(n, t)(1 - \lambda\epsilon) + g(n - 1, t)\lambda\epsilon + o(\epsilon)$. Hence,

$$\frac{d}{dt} g(n, t) = -\lambda g(n, t) + \lambda g(n - 1, t).$$

For $n = 0$, this gives $g(0, t) = e^{-\lambda t}$. Assume as an induction hypothesis that the result hods for $n - 1$. For n we get

$$\frac{d}{dt} g(n, t) = -\lambda g(n, t) + \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}.$$

One can check that $g(n, t) = (\lambda t)^n e^{-\lambda t}/n!$ solves this equation.

As an alternate proof, note that

$$\begin{aligned} & P(T_1 \in (t_1, t_1 + \epsilon), \ldots, T_n \in (t_n, t_n + \epsilon), T_{n+1} > t) \\ \approx\ & P(\tau_1 \in (t_1, t_1 + \epsilon), \tau_2 \in (t_2 - t_1, t_2 - t_1 + \epsilon), \ldots, \tau_n \in (t_n - t_{n-1}, t_n - t_{n-1} + \epsilon), \tau_{n+1} > t - t_n) \\ \approx\ & \lambda\epsilon e^{-\lambda t_1} \lambda\epsilon e^{-\lambda(t_2 - t_1)} \cdots \lambda\epsilon e^{-\lambda(t_n - t_{n-1})} e^{-\lambda(t - t_n)} = (\lambda\epsilon)^n e^{-\lambda t}. \end{aligned}$$

Integrating this constant value over all the possible values of $t_1 < \cdots < t_n$ in $[0, t]^n$ and observing that these values occupy a fraction $1/n!$ of the cube $[0, t]^n$, we get the result.

(c) We use the result of the calculation above and find $P[T_1 \in (t_1, t_1+\epsilon), \ldots, T_n \in (t_n, t_n+\epsilon) | N_t = n] = \frac{(\lambda\epsilon)^n e^{-\lambda t}}{P(N_t=n)} = \frac{n!\epsilon^n}{t^n}$, which proves the statement of the theorem.

(d) For small ϵ, the probability of a new jump occurring in the time interval $(t, t+\epsilon)$ given T_{N_t} is

$$\int_{t-T_{N_t}}^{t+\epsilon-T_{N_t}} \lambda e^{-\lambda s} ds = e^{-\lambda(t-T_{N_t})} - e^{-\lambda(t+\epsilon-T_{N_t})} \approx \lambda\epsilon$$

from the exponential distribution. Thus, the probability of a red (blue, resp.) jump occurring in the time interval $(t, t+\epsilon)$ is approximately $\alpha\lambda\epsilon$ ($(1-\alpha)\lambda\epsilon$, resp.). It is clear that $\{A_t, t \geq 0\}$ and $\{B_t, t \geq 0\}$ are Poisson processes with respective rates $\alpha\lambda$ and $(1-\alpha)\lambda$. Let $0 < t_1 < \cdots < t_n$ and $0 < s_1 < \cdots < s_m$, with no common values. The probability that A has the jumps t_i in $[0, t]$ and B the jumps s_j in $[0, t]$, within small ϵ, is

$$\alpha^n (1-\alpha)^m (\lambda\epsilon)^{n+m} e^{-\lambda t} = (\alpha\lambda\epsilon)^n e^{-\alpha\lambda t} ((1-\alpha)\lambda\epsilon)^m e^{-(1-\alpha)\lambda t},$$

which proves the independence. □

17.4 Poisson Measure

We generalize the previous definition to non-homogeneous and multi-dimensional situations. We limit the discussion to a concrete situation. Let λ be a σ-finite measure on $(\Re^d, \mathcal{B}(\Re^d))$. For instance,

$$\lambda(A) = \int_A \phi(x) dx,$$

where ϕ is nonnegative and bounded. Being σ-finite means that we can cover $\Re^d$ with a countable union of pairwise disjoint sets $C_n \in \mathcal{B}(\Re^d)$ such $\lambda(C_n) < \infty$ for each n. Let $\eta(n)$ be a Poisson random variable with mean $\lambda(C_n)$. Let the random variables $X_1, \ldots, X_{\eta(n)}$ be i.i.d. and distributed in C_n so that $P(X_k \in A) = \lambda(A)/\lambda(C_n)$ for every $A \subseteq C_n$. Finally, let

$$N(A) = \sum_{k=1}^{\eta(n)} 1\{X_k \in A\}, A \subseteq C_n.$$

The counting process $N(\cdot)$ is called a Poisson measure on C_n with intensity λ. By doing this construction independently in each C_n, we define a Poisson measure in $\Re^d$ with intensity $\lambda(\cdot)$.

Theorem 17.2. *Let $N(\cdot)$ be a Poisson measure on $\Re^d$. For any $n \geq 2$ and any pairwise disjoint sets $A_1, \ldots, A_n$ in $\mathcal{B}(\Re^d)$, the random variables $N(A_1), \ldots, N(A_n)$ are mutually independent and $N(A_m)$ is Poisson with mean $\lambda(A_m)$.*

Proof: Let $A_m \subseteq C_i$. Note that

$$\begin{aligned} P(N(A_m) = k_1, N(C_i) = k_1 + k_2) &= P(N(C_i) = k_1 + k_2) P[N(A_m) = k_1 | N(C_i) = k_1 + k_2] \\ &= \frac{(\lambda(C_i))^{k_1+k_2} e^{-\lambda(C_i)}}{(k_1+k_2)!} \binom{k_1+k_2}{k_1} \left(\frac{\lambda(A_m)}{\lambda(C_i)}\right)^{k_1} \left(1 - \frac{\lambda(A_m)}{\lambda(C_i)}\right)^{k_2} \\ &= \left(\frac{(\lambda(A_m))^{k_1} e^{-\lambda(A_m)}}{k_1!}\right) \left(\frac{(\lambda(C_i) - \lambda(A_m))^{k_2} e^{-\lambda(C_i)+\lambda(A_m)}}{k_2!}\right). \end{aligned}$$

Summing the above probability over all non-negative integers k_2, it is clear that $N(A_m)$ is Poisson with mean $\lambda(A_m)$. To show the independence, it suffices to look at $n = 2$. Let A_1 and A_2 partition some C_i. Then $\lambda(C_i) = \lambda(A_1) + \lambda(A_2)$ and

$$\begin{aligned} P(N(A_1) = k_1, N(A_2) = k_2) &= P(N(A_1) = k_1, N(C_i) = k_1 + k_2) \\ &= P(N(C_i) = k_1 + k_2) P[N(A_1) = k_1 | N(C_i) = k_1 + k_2] \\ &= \frac{(\lambda(C_i))^{k_1+k_2} e^{-\lambda(C_i)}}{(k_1+k_2)!} \binom{k_1+k_2}{k_1} \left(\frac{\lambda(A_1)}{\lambda(C_i)}\right)^{k_1} \left(1 - \frac{\lambda(A_1)}{\lambda(C_i)}\right)^{k_2} \\ &= \left(\frac{(\lambda(A_1))^{k_1} e^{-\lambda(A_1)}}{k_1!}\right) \left(\frac{(\lambda(A_2))^{k_2} e^{-\lambda(A_2)}}{k_2!}\right). \end{aligned}$$

□

17.5 Application

We consider a service system where customers arrive according to a Poisson process with rate λ and arrival times $\{T_n, n \in \mathbb{Z}\}$. Thus, $\{T_n - T_{n-1}, n \in \mathbb{Z}\}$ are i.i.d. Exp(λ). Each customer requires a service time S_n that is i.i.d. with some distribution $G(\cdot)$. For $n \in \mathbb{Z}$, customer n leaves at time $T'_n = T_n + S_n$.

Fact 17.3. *(a) The departure process with jumps $\{T'_n, n \in \mathbb{Z}\}$ is Poisson with rate λ.*

(b) The number of customers in the queue at time t is independent of the departure times up to time t and is Poisson with mean $\lambda E(S_1)$.

Proof: The first observation is that $\{(T_n, S_n), n \in \mathbb{Z}\}$ is a Poisson measure on $\Re^2$. To see this, let $C_n = [t_n, t_n + 1) \times \Re$, where $\{t_n, n \in \mathbb{Z}\}$ is an increasing sequence divergent to infinite. The number of points in C_n is Poisson with mean $\lambda(t_n + 1 - t_n) = \lambda$ from the fact that the arrival process is Poisson with rate λ. Given the number of points in $\{C_n, n \in \mathbb{Z}\}$, the number of points in C_n are distributed independently according to the intensity $(t, t+dt) \times (u, u+du) \mapsto \lambda dt dG(u)$ (in the sense of probability). This corresponds precisely to the definition of a Poisson measure.

Now we look at Figure 17.1. For a fixed time t, the points in the triangle A correspond to the customers in the queue at time t. The points in the set B_1 (B_2 and B_3, resp.) correspond to the customers who will leave during (t_1, t) ((t_2, t_1) and (t_3, t_2), resp.). Since the measure is Poisson, the number of points in these disjoint sets are all independent and Poisson distributed. Hence the departure process is Poisson. To find the corresponding rate we compute the mean of B_3 with $t_2 = t_3 + 1$,

$$\int_{B_3} \lambda dt dG(u) = \int_0^1 \lambda(t_2 - t_3) dG(u) = \lambda.$$

However, this particular B_3 indicates the departure process is incremented by a unit time, the rate is λ.

It remains to compute the mean of A, which is

$$\int_A \lambda dt dG(u) = \int_0^\infty \int_{t-u}^t \lambda g(u) dt du = \lambda \int_0^\infty u g(u) du = \lambda E(S_1).$$

□

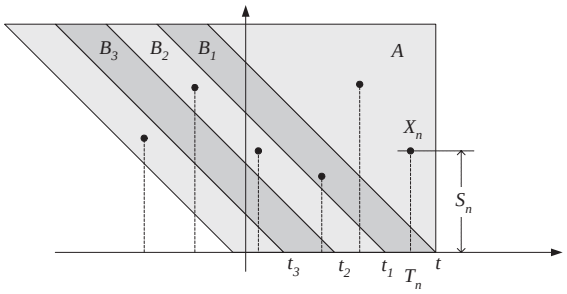


Figure 17.1: The arrivals with the service times.

17.6 Solved Problems

Problem 17.1. *Let N be a Poisson r.v. with mean λ, and M a binomial r.v. $B(m,p)$, with $mp = \lambda, m > 0$ i.e.,*

$$P(M = n) = \binom{m}{n} p^n (1-p)^{m-n} \,, \text{ for all } n = 0, \ldots, m \,.$$

Show that $P(N \le n) \le P(M \le n)$, for any $n \ge 0$.

(Hint: write N as $N = N(1) = \sum_{i=1}^{m} (N(i/m) - N((i-1)/m))$, where $N(t)$ is a Poisson process with rate λ. Define a r.v. M for which $N \ge M$, and $M =_D B(m,p)$.)

Solution:
Let $N(t)$ be a Poisson process with rate λ, and let $N = N(1)$. Observe that for any $m > 0$,

$$N = \sum_{i=1} \big(N\big(\frac{i}{m}\big) - N\big(\frac{i-1}{m}\big)\big) \ge \sum_{i=1} 1\big\{N\big(\frac{i}{m}\big) - N\big(\frac{i-1}{m}\big) > 0\big\} =: M \,.$$

Since the r.v's $(N(i/m) - N((i-1)/m), i = 1, \ldots, m)$ are independent, $M =_D B(m,p)$, with

$$p = P\big(N\big(\frac{i}{m}\big) - N\big(\frac{i-1}{m}\big) = 0\big) = 1 - e^{-\lambda/m} \,.$$

Thus, for all $n \ge 0$, $P(N \le n) \le P(M \le n)$.

Problem 17.2. *In this problem we want to estimate the rate λ of some Poisson process N_t, given the observation $(N_s : 0 \le s \le t)$, for some fixed $t > 0$.*

1. *Show that N_t/t is the maximum-likelihood estimate of λ.*
2. *Compute $Var(N_t/t)$, for $t > 0$.*

Solution:

1. Let $T_1, T_2, \ldots$ be the arrival times of N_s, and suppose we observe these to be $t_1, t_2, \ldots$, and $N_t = n$. From class,

$$P[T_1 \in (t_1, t_1 + dt), \ldots, T_n \in (t_n, t_n + dt)] = \lambda^n e^{-\lambda t} (dt)^n \,.$$

Thus, the value of λ at maximizes the likelihood must satisfy

$$n\lambda^{n-1} e^{-\lambda t} - \lambda^n t e^{-\lambda t} = 0 \implies \lambda = \frac{n}{t} \,.$$

2.
$$\mathrm{Var}\big(\frac{N_t}{t}\big) = \frac{\lambda t}{t^2}\,.$$

Problem 17.3. *Under hypothesis $X = 0$, N_t is a Poisson process with rate λ_0; under $X = 1$, N_t is a Poisson process with rate λ_1. Suppose we observe $(N_s : 0 \le s \le t)$, for some fixed $t > 0$.*

1. *Solve the hypothesis-testing problem, i.e., find a test that minimizes $P[\hat{X} = 1|X = 0]$, such that $P[\hat{X} = 0|X = 1] \le \beta$.*
2. *Find t such that $P[\hat{X} = 1|X = 0] \le \alpha$.*

Solution:

1. Assume $\lambda_1 > \lambda_0$. By Problem 5, the likelihood-ratio is given by
$$\Lambda((N_s, 0 \le s \le t)) = \frac{\lambda_1^{N_t} e^{-\lambda_1 t}}{\lambda_0^{N_t} e^{-\lambda_0 t}}\,,$$
so the Neyman-Pearson test is
$$\hat{X} = \begin{cases} 1 & ,\Lambda((N_s, 0 \le s \le t)) > \kappa \\ 1\{U > \gamma\} & ,\Lambda((N_s, 0 \le s \le t)) = \kappa \\ 0 & ,\Lambda((N_s, 0 \le s \le t)) < \kappa \end{cases},$$
where U is an independent r.v. uniform in $[0, 1]$, and κ, γ are determined by $P[\hat{X} = 0|X = 1] \le \beta$.

2. We will use a suboptimal test to bound the errors. The test will be:
$$\tilde{X} = \begin{cases} 1 & ,\frac{N_t}{t} > \frac{\lambda_1+\lambda_0}{2} \\ 0 & ,\text{ otherwise} \end{cases}$$
Notice that for any $a > 0$,
$$1\{|N_t/t - \lambda_0|\} \ge a\} \le \frac{(N_t/t - \lambda_0)^2}{a^2} 1\{|N_t/t - \lambda_0|\} \ge a\}\,,$$
so taking expectations, conditional on $X = 0$, on both sides yields
$$\begin{aligned} P[|N_t/t - \lambda_0| \ge a|X = 0] &\le \frac{E[(N_t/t - \lambda_0)^2|X = 0]}{a^2} \\ &= \frac{\lambda_0}{ta^2}\,, \end{aligned}$$
by Problem 5(b). Using $a = (\lambda_1 - \lambda_0)/2$, we can compute the error as
$$P[\tilde{X} = 1|X = 0] \le \frac{4\lambda_0}{t(\lambda_1 - \lambda_0)^2}\,.$$
In the same way, we can compute
$$P[\tilde{X} = 0|X = 1] \le \frac{4\lambda_1}{t(\lambda_1 - \lambda_0)^2}\,.$$

Thus, for $t \geq 4\max\{\lambda_1/\beta, \lambda_0/\alpha\}/(\lambda_1 - \lambda_0)^2$, we have $\max\{P[\tilde{X} = 0|X = 1], P[\tilde{X} = 1|X = 0]\} \leq \min\{\beta, \alpha\}$.

Since $P[\hat{X} = 0|X = 1] = \beta \geq P[\tilde{X} = 0|X = 1]$, the Neyman-Pearson test guarantees that $P[\hat{X} = 1|X = 0] \leq P[\tilde{X} = 1|X = 0]$.

Problem 17.4. [Two-sided stationary Poisson process.] *Let N_t^+, N_t^- be two independent Poisson processes with rate λ, and corresponding arrival times $(T_n^+, n \geq 1)$, $(T_n^-, n \geq 1)$ respectively. For $n > 0$, define $T_n = T_n^+$, $T_{1-n} = -T_n^-$.*

One can think of $(T_n, n \leq 0)$, $(T_n, n > 0)$ as describing arrivals occuring before and after time 0, respectively. You will show that the arrival statistics do not depend on the choice of $t = 0$, i.e., $(T_n, n \in \mathbb{Z})$ is stationary.

For any fixed $s \in \mathbb{R}$, let $n_0 = \max\{n \in \mathbb{Z} : T_n \leq s\}$, and consider the sequence $(T_{n+n_0} - s, n \in \mathbb{Z})$.

Show that $(T_{n+n_0} - s, n \in \mathbb{Z}) =_D (T_n, n \in \mathbb{Z})$.

Observe $E(T_1 - T_0) = 2\lambda^{-1}$, while $E(T_{n+1} - T_n) = \lambda^{-1}$ for all $n \neq 0$! How do you explain this?

Solution:
We must show $s - T_{n_0} =_D T_{n_0+1} - s =_D T_{n+1+n_0} - T_{n+n_0} =_D \text{Exp}(\lambda)$ for $n \in \mathbb{Z} \setminus \{0\}$, and that these are mutually independent.

We may assume $s \geq 0$, since the setup is symmetric around 0.

Now, $P(s - T_{n_0} > t) = E(P[s - T_{n_0} > t|n_0])$. On $n_0 = 0$, $P[s - T_{n_0} > t|n_0] = P(-T_0 > t - s) = 1\{t > s\}$. On $n_0 = m > 0$, the arrival times $(T_1, \cdot, T_m)$ are distributed as the order statistics of i.i.d. uniform on $[0, s]$, since $N_s^+ = n_0$. Thus,

$$P[s - T_{n_0} > t|n_0] = 1\{s > t\}\big(\frac{s-t}{s}\big)^m .$$

Combining all the above yields,

$$\begin{aligned} P(s - T_{n_0} > t) &= P(N_s^+ = 0)(e^{-\lambda(t-s)}1\{t > s\} + 1\{t \leq s\}) \\ &\quad + 1\{t < s\} \sum_{m=1}^{\infty} P(N_s^+ = m)\big(\frac{s-t}{s}\big)^m \\ &= e^{-\lambda s}(e^{-\lambda(t-s)}1\{t > s\} + 1\{t \leq s\}) \\ &\quad + 1\{t < s\} \sum_{m=1}^{\infty} \frac{e^{-\lambda s}(\lambda s)^m}{m!}\big(\frac{s-t}{s}\big)^m \\ &= e^{-\lambda t} . \end{aligned} \tag{17.1}$$

We still need to show that $T_{n+n_0} - T_{n+n_0-1}$ are i.i.d. $\text{Exp}(\lambda)$, for all $n \leq 0$.

On $T_{n+n_0} < 0$, $T_{n+n_0} - T_{n+n_0-1}$ is independent of $(T_m : m \geq n + n_0)$ and exponentially distributed with rate λ. This is because n_0 depends only on $(T_m : m \geq 0)$, and $(-T_m : m \leq 0)$ give the arrivals of a Poisson process of rate λ, which are independent of $(T_m : m \geq 0)$.

Let's assume $T_{n+n_0} \geq 0$ (so $n + n_0 \geq 1$) now. Then,

$$P[T_{n+n_0} - T_{n+n_0-1} > t|T_{n+n_0}] = E[P[T_{n+n_0} - T_{n+n_0-1} > t|T_{n+n_0}, n_0]|T_{n+n_0}] ,$$

but $\{T_{n+n_0} = u, N_u^+ = m, n_0 = k\} = \{T_{n+n_0} = u, n_0 = -n + m\}$, so

$$\begin{aligned} P[T_{n+n_0} - T_{n+n_0-1} > t | T_{n+n_0} = u, n_0 = k, (N_r^+, u \le r \le s)] = \\ P[u - T_{k+n-1} > t | T_{k+n} = u, N_u^+ = k+n, N_s - N_u^+ = -n, (N_r^+, u \le r \le s)] \\ = P[u - T_{k+n-1} > t | N_{u-}^+ = k+n-1] \\ = P[u - T_{n_0} > t | N_u^+ = k+n-1] \,, \end{aligned} \tag{17.2}$$

where $N_{u-}^+ = \lim_{\epsilon \downarrow 0} N_{u-\epsilon}^+$. Thus, by letting $s = u$ in (17.1), $P[T_{n+n_0} - T_{n+n_0-1} > t | T_{n+n_0}] = e^{-\lambda t}$, where we have always conditioned on $\{T_n + n_0 \ge 0\}$.

Now, by combining the above,

$$P(T_{n+n_0} - T_{n+n_0-1} > t) = e^{-\lambda t} \,.$$

Also, (17.2) implies that $T_{n+n_0} - T_{n+n_0-1}$ is independent of $s - T_{n_0}, (T_m - T_{m-1}, m = n + n_0 + 1, \ldots, n_0$.

Finally, note that $T_{n_0+1} - s, (T_{n+n_0+1} - T_{n+n_0}, n > 0)$ are i.i.d. Exp(λ), independent of everything that came before time s. This follows from the fact that $(N_{t+s}^+ - N_s^+, t \ge 0)$ is a Poisson process, independent of $(N_r^+ : 0 \le r \le s)$.

$E(T_1 - T_0) > E(T_{n+1} - T_n), \forall n \ne 0$ is called the "inspection" paradox: When an observer suddenly looks at a system of arrivals, he sees that the spread between the next and previous arrival, is "larger" than the average spread of the other inter-arrival times! The reason that this occurs is not particular to Poisson; it is because the observer is more likely to fell into a longer-than-usual interval. E.g., consider two types of alternating interarrival intervals; one that lasts 1 second, and another lasting 100 seconds. Although half of the intervals are short, there is a higher likelihood of encountering a long one because they occupy more time. In this example, the observed average interarrival interval is *not* $\frac{1}{2}1\text{s} + \frac{1}{2}100\text{s} = 50.5\text{s}$, but $\frac{1}{101}1\text{s} + \frac{100}{101}100\text{s} \approx 99\text{s}$.

Problem 17.5. *Let $\tau_1, \tau_2, \ldots$ be i.i.d. Exp(λ) r.v's, independent of a geometric r.v. Z with parameter p, $0 < p < 1$, i.e., $P(Z = k) = (1-p)^{k-1}p$, for $k \ge 1$.*

Show that

$$\sum_{i=1}^{Z} \tau_1 =_D Exp(\lambda p) \,.$$

Solution:

Let $N_t, t \ge 0$ be a Poisson process with rate λ, and interarrival times $\tau_1, \tau_2, \ldots$. Then, for all $s, t \ge 0$,

$$\begin{aligned} P\big[\sum_{i=1}^{Z} \tau_i \ge t + s \big| \sum_{i=1}^{Z} \tau i \ge s\big] &= P[Z \ge (N_{t+s} - N_s) + N_s | Z \ge N_s] \\ &= P(Z \ge N_{t+s} - N_s) = P(Z \ge N_t) \\ &= P\big(\sum_{i=1}^{Z} \tau_t \ge t\big) \,, \end{aligned}$$

so $\sum_{t=1}^{Z} \tau_i$ has the memoryless property, so it is exponentially distributed with mean

$$E\big(\sum_{i=1}^{Z} \tau_i\big) = E(Z)E(\tau_i) = p^{-1}\lambda^{-1} \,.$$

Problem 17.6. *Let $\tau_1, \ldots, \tau_n$ be n i.i.d. Exp(λ) r.v's.*
Calculate $E(\max\{\tau_1, \ldots, \tau_n\})$.

Solution:
Let $n > 1$. Then,

$$E(\max\{\tau_1, \ldots, \tau_n\}) = E(\max\{\tau_1, \ldots, \tau_{n-1}\}) + E(\max\{\tau_1, \ldots, \tau_n\} - \max\{\tau_1, \ldots, \tau_{n-1}\}) .$$

But,

$$E(\max\{\tau_1, \ldots, \tau_n\} - \max\{\tau_1, \ldots, \tau_{n-1}\}) = E((\tau_n - \max\{\tau_1, \ldots, \tau_{n-1}\}) 1\{\tau_n > \max\{\tau_1, \ldots, \tau_n\}\}),$$

and

$$P(\tau_n \geq \max\{\tau_1, \ldots, \tau_{n-1}\} + t | \tau_n \geq \max\{\tau_1, \ldots, \tau_{n-1}\}) = P(\tau_n \geq t) ,$$

so

$$E[\max\{\tau_1, \ldots, \tau_n\} - \max\{\tau_1, \ldots, \tau_{n-1}\} | \tau_n > \max\{\tau_1, \ldots, \tau_n\}]) = \frac{1}{\lambda} .$$

Hence,

$$E(\max\{\tau_1, \ldots, \tau_n\}) = E(\max\{\tau_1, \ldots, \tau_{n-1}\}) + \frac{1}{\lambda n} = \frac{1}{\lambda} \sum_{i=1}^{n} \frac{1}{i} .$$

Chapter 18

Continuous Time Markov Chains: Examples and Definition

18.1 Summary

We look at examples to illustrate the following key ideas.

- Exponential Holding Times
- Rate Matrix
- Invariant Distribution, Convergence
- Explosions
- Definition

18.2 Examples

We discuss a few simple examples.

18.2.1 Two-State MC

We define a process $\{X_t, t \geq 0\}$ as follows. X_0 is chosen in $\{0, 1\}$ with $P(X_0 = 0) = \pi_0(0)$ and $P(X_0 = 1) = \pi_0(1)$ where $\pi_0(0) \geq 0, \pi_0(1) \geq 0$, and $\pi_0(0) + \pi_0(1) = 1$. If $X_0 = i$, then $X_t = i$ for $t \in [0, T_1)$ where $T_1 =_D Exp(q_i)$ with $q_i > 0$. Also, $X_{T_1} = 1 - i$. The construction of X_t for $t \geq T_1$ then continues as before, independently of T_1.

Figure 18.1 illustrates the construction.

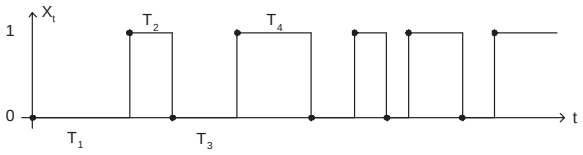


Figure 18.1: A two-state Markov chain.

The interesting questions concern the long-term fraction of time that $X_t = i$ and the evolution of $P(X_t = i)$, for $i = 0, 1$. To study these questions, we establish a few preliminary facts.

Fact 18.1. *Fix $t > 0$. Given $\{X_s, s \leq t; X_t = i\}$, the process $\{X_s, s \geq t\}$ has the same distribution as $\{X_s, s \geq 0\}$ given that $X_0 = i$, for $i \in \{0, 1\}$.*

Proof:
The exponential distribution is memoryless.

□

That property is called the *Markov property.* It states that X *restarts afresh* from X_t at time t, independently of the past.

Fact 18.2. *Let $\pi_t = (\pi_t(0), \pi_t(1))$ for $t \geq 0$ where $\pi_t(i) = P(X_t = i)$. Then*

$$\frac{d}{dt}\pi_t = \pi_t Q$$

where

$$Q = \begin{bmatrix} -q_0 & q_0 \\ q_1 & -q_1 \end{bmatrix}.$$

The matrix Q is called the rate matrix *or the* generator *of X.*

Proof:
We find that

$$\begin{aligned} P(X_{t+\epsilon} = 0) &= P(X_{t+\epsilon} = 0, X_t = 0) + P(X_{t+\epsilon} = 0, X_t = 1) \\ &= P[X_{t+\epsilon} = 0 | X_t = 0]P(X_t = 0) + P[X_{t+\epsilon} = 0 | X_t = 1]P(X_t = 1) \\ &= P[T_1 > \epsilon | X_0 = 0]\pi_t(0) + P[T_1 < \epsilon | X_0 = 1]\pi_t(1) + o(\epsilon) \\ &= (1 - q_0\epsilon)\pi_t(0) + q_1\epsilon\pi_t(1) + o(\epsilon) = \pi_t(0) - q_0\epsilon\pi_t(0) + q_1\epsilon\pi_t(1). \end{aligned}$$

Hence,

$$\frac{d}{dt}\pi_t(0) = -q_0\pi_t(0) + q_1\pi_t(1).$$

□

Using the previous fact, one easily derives the following result.

Fact 18.3. *(a) The distribution π_t of X_t is given as follows:*

$$\pi_t = \pi_0 e^{Qt}, t \geq 0.$$

(b) $\pi_t = \pi_0, \forall t \geq 0$ if and only if

$$\pi_0 Q = 0.$$

(c) The unique solution of the equations above is

$$\pi = [\frac{q_1}{q_0 + q_1}, \frac{q_0}{q_0 + q_1}].$$

We note also the following fact.

Fact 18.4. *(a) One has*

$$\lim_{T\to\infty} \int_0^T 1\{X_s = i\}ds = \pi(i), i \in \{0, 1\}, \text{ almost surely}$$

where π is as given in the previous fact.
(b) Moreover, for any π_0, one has

$$\pi_t \to \pi, \text{ as } t \to \infty.$$

Proof:
(a) Assume that $X_0 = 0$. Then

$$\frac{1}{T_1 + \cdots + T_{2n}} \int_0^{T_1+\cdots+T_n} 1\{X_t = 0\}dt = \frac{T_1 + T_3 + \cdot + T_{2n-1}}{T_1 + T_2 + \cdots + T_{2n}} \to \frac{q_0^{-1}}{q_0^{-1} + q_1^{-1}} = \frac{q_1}{q_0 + q_1}.$$

(b) A simple proof is by *coupling*. Let X_t be stationary version of this MC and Y_t an independent version of the MC with $Y_0 = 0$. Define Z_t as follows. Let $\tau = \min\{t \geq 0 | X_t = Y_t\}$. Let then $Z_t = Y_t$ for $t \leq \tau$ and $Z_t = X_t$ for $t \geq \tau$. The claim is that Z_t is again a MC with rate matrix Q and with $Z_0 = 0$, as you can easily verify. Now, since $\tau < \infty$, it is clear that $P(X_t = Z_t) \to 1$ as $t \to \infty$. Since the distribution of X_t is π for all $t \geq 0$, it follows that the distribution of Z_t converges to π. However, the distribution of Z_t is the same as that of Y_t. This concludes the proof.

□

18.2.2 Uniformization

Let $Y = \{Y_n, n \geq 0\}$ be a Markov chain with transition matrix P and $N = \{N_t, t \geq 0\}$ a Poisson process with rate λ that is independent of Y.

Fact 18.5. *Define $X = \{X_t = Y_{N_t}, t \geq 0\}$. Then*

$$P[X_{t+\epsilon} = j | X_t = i; X_s, s \leq t] = \begin{cases} q(i,j)\epsilon + o(\epsilon) := \lambda P(i,j)\epsilon + o(\epsilon), & \text{if } i \neq j \\ 1 + q(i,i)\epsilon + o(\epsilon) := 1 - \lambda \sum_{k\neq i} P(i,k)\epsilon + o(\epsilon), & \text{if } i = j. \end{cases}$$

Proof:
The key idea is the memoryless property of the Poisson process.

□

18.2.3 Explosion

Here is a cute example. X_0 is chosen in $\mathcal{X} = \{0, 1, 2, \ldots\}$ according to some distribution π_0. If $X_0 = i$, then X keeps that value for an $Exp(q_i)$ random time. It then jumps to $i + 1$, and the construction continues as before. We assume that

$$\sum_{i=0}^{\infty} q_i^{-1} < \infty.$$

For instance, $q_i = (i+1)^2$. In that case, one can see that X makes infinitely many jumps in finite time. For instance, if $X_0 = 0$. we have, for $s > 0$,

$$E(T_1 + \cdot + T_n) = q_0^{-1} + q_1^{-1} + \cdots + q_{n-1}^{-1} \to \sum_{i=0}^{\infty} q_i^{-1}, \text{ as } n \to \infty.$$

Hence, by the Lebesgue convergence theorem,

$$E(\sum_{n=1}^{\infty} T_n) = \sum_{i=0}^{\infty} q_i^{-1} < \infty$$

This implies that $P(\sum_{n=1}^{\infty} T_n < \infty) = 1$.

Thus our construction defines the process X only on $[0, \tau)$ where $\tau = \sum_{n=1}^{\infty} T_n < \infty$. To continue the construction, we restart it at time τ with the distribution π_0.

What is quite interesting is that we have defined a process that makes infinitely many jumps in finite time. Also, the fraction of time that this process spends in state 0 certainly depends on π_0. Consequently, this fraction of time is not uniquely determined by Q, even though the process is irreducible and positive recurrent if $\pi_0(0) = 0$, with the definitions that we adapt in the obvious way from the discrete time case....

18.2.4 Definition and Construction

We define and construct a Markov chain.

Definition 18.1. *Let $\mathcal{X}$ be a countable set. A* rate matrix *Q on $\mathcal{X}$ is a matrix $Q = \{q(i,j), i, j \in \mathcal{X}\}$ such that $q(i,j) \geq 0$ for all $i \neq j$ and $\sum_j q(i,j) = 0$ for all i.*

A process $\{X_t, t \geq 0\}$ that takes values in $\mathcal{X}$ is a Markov chain *with rate matrix Q if*

$$P[X_{t+\epsilon} = j | X_t = i; X_s, s \leq t] = \begin{cases} q(i,j)\epsilon + o(\epsilon), & \text{if } i \neq j \\ 1 + q(i,i)\epsilon + o(\epsilon), & \text{if } i = j. \end{cases}$$

We already saw a few examples. We now construct the general case.

Definition 18.2. *Construction*

Let Q be a rate matrix and π_0 a distribution on $\mathcal{X}$. We define a process $X = \{X_t, t \geq 0\}$ as follows.

First we choose X_0 in $\mathcal{X}$ according to π_0. Second, if $X_0 = i$, then we define an exponentially distributed random variable τ with rate $-q(i,i)$. Third, we define $X_t = i$ for $t \in [0, \tau)$. Fourth, given $X_0 = i$ and τ, we define $X_\tau = j$ with probability $P(i,j) := -q(i,j)/q(i,i)$ for $j \neq i$. Fifth, we continue the construction of X_t for $t \geq \tau$ as before, independently of X_0 and of τ.

This construction defines the process X as long as there are no explosions.

We have the following fact.

Fact 18.6. *The process X that we constructed is a Markov chain with rate Q if there are no explosions.*

Proof:

Exercise.

□

What about explosions? If they occur, we can restart after the explosion as we explained in the example. Can we say when explosions occur? The following fact provides a strong clue.

Fact 18.7. *For $n \geq 1$, let $\tau_n =_D Exp(\lambda_n)$ and assume that these random variables are independent. Then*

$$P(\tau := \sum_{n=1}^{\infty} \tau_n < \infty) = \begin{cases} 1, & \text{if } \sum_{n=0}^{\infty} \lambda_n^{-1} < \infty \\ 0, & \text{if } \sum_{n=0}^{\infty} \lambda_n^{-1} = \infty. \end{cases}$$

Proof:

We saw the first part earlier.

Assume $\sum_{n=0}^{\infty} \lambda_n^{-1} = \infty$. Then, for all $s > 0$,

$$E(e^{-s(\tau_1 + \cdots + \tau_n)}) = \Pi_{m=1}^{m}(1 + \lambda_m^{-1}s)^{-1} \leq (1 + \sum_{m=1}^{n} \lambda_m^{-1}s)^{-1} \to 0 \text{ as } n \to \infty.$$

Consequently, by Lebesgue's Theorem, $E(e^{-s\tau}) = 0$ for all $s > 0$. This implies that $P(e^{-s\tau} = 0) = 1$, so that $P(\tau = \infty) = 1$.

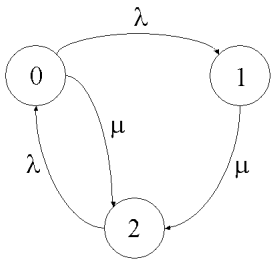
□

18.3 Solved Problems

Problem 18.1. *How would you simulate a continuous-time Markov chain (CTMC) with rate matrix Q and state-space $\{1, 2, \ldots, N\}$?*

Solution:

Problem 18.2. *Consider the CTMC with rate-matrix given by the figure.*



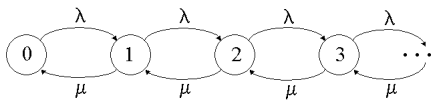
Find the invariant distribution.

Solution:

Problem 18.3. *Let X_t be a CTMC with state-space $\mathcal{X}$. Assume that $-q(i,i) \leq \lambda, \forall i \in \mathcal{X}$.*

Show that one can write $X_t = Y_{N_t}$, where $(N_t, t \geq 0)$ is a Poisson process of rate λ, and $(Y_n, n \geq 0)$ is a discrete-time MC.

Solution:



Problem 18.4. *Consider the CTMC X_t with rate-matrix given by the figure on the next page.*

1. *Show that one can write*
$$X_t = f(X_0; N_s^\lambda, N_s^\mu, s \le t)\,,$$
where N_s^λ, N_s^μ are independent Poisson processes of rate λ, μ, respectively.
2. *How would you generalize (a) to any CTMC X_t?*

Solution:

Problem 18.5. *Let X_t be a CTMC on $\mathcal{X}$ with rate-matrix $Q = [q(i,j)]_{i,j\in\mathcal{X}}$, and define $\tau_A = \min\{t > 0 | X_t \in A\}$ for $A \subset \mathcal{X}$.*
Show that

$$P[X_{\tau_A+\epsilon} = j | X_{\tau_A} = i; X_s, s \le \tau_A] = q(i,j)\epsilon + o(\epsilon)\,, i \ne j\,,\ \text{as } \epsilon \downarrow 0\,.$$

Solution:

Problem 18.6. *Provide the details for the coupling argument we used in class to show that*

$$P(X_t = 0) \to \frac{\mu}{\lambda+\mu}\,,\ \text{as } t \to \infty\,,$$

for the CTMC with rate matrix $\left[\begin{smallmatrix} -\lambda & \lambda \\ \mu & -\mu \end{smallmatrix}\right]$.

Chapter 19

Continuous Time Markov Chains: Key Results

19.1 Summary

We explain the following ideas and results:

- Invariant Distribution
- Classification Theorem
- M/M/1 queue
- Time-Reversal

19.2 Invariant Distribution

We have the following definition.

Definition 19.1. *Invariant Distribution*
The distribution $\pi = \{\pi(i), i \in \mathcal{X}\}$ *is said to be* invariant *for* Q *if*

$$\pi Q = 0,$$

i.e., if

$$\pi(i)(-q(i,i)) = \sum_{j \in \mathcal{X}} \pi(j) q(j,i), i \in \mathcal{X}. \tag{19.1}$$

The relations (19.1) are called the balance equations.

The relevance is because of the following results.

Theorem 19.1. *Kolmogorov Equations*
Let X *be a regular Markov chain (meaning, no explosions are possible) with rate matrix* Q. *For* $t \geq 0$, *let* $\pi_t = \{\pi_t(i), i \in \mathcal{X}\}$ *where* $\pi_t(i) = P(X_t = i)$. *We consider* π_t *as a row vector. Then*

$$\frac{d}{dt}\pi_t = \pi_t Q.$$

Consequently, $\pi_t = \pi$ *for all* $t \geq 0$ *if and only if* $\pi_0 = \pi$ *and* π *is invariant.*

We discuss an interpretation of the balance equations (19.1). The number of transitions from $\{i\}$ to $\{i\}^c := \{j \in \mathcal{X} \mid j \neq i\}$ and the number of transitions from $\{i\}^c$ to $\{i\}$ over any interval $[0, T]$ differ by at most one. If the Markov chain is stationary with invariant distribution π, then the rates of those transitions are given by both sides of (19.1).

Theorem 19.2. *Let X be a regular Markov chain with rate matrix Q and initial distribution π. It is stationary if and only if π is invariant.*

19.3 Classification Theorem

We define *irreducible, transient, null recurrent,* and *positive recurrent* as in discrete time. We have the following result.

Theorem 19.3. *Classification*

Let $X = \{X_t, t \geq 0\}$ be an irreducible Markov chain on $\mathcal{X}$.

(a) The states are either all transient, all null recurrent, or all positive recurrent. We then say that the Markov chain is ....

(b) If X is transient or null recurrent, then

$$\frac{1}{T}\int_0^T 1\{X_t = i\}dt \to 0, \text{ as } T \to \infty, \forall i \in \mathcal{X}, a.s.$$

Moreover, there is no invariant distribution and

$$P(X_t = i) \to 0, \forall i \in \mathcal{X}.$$

(c) If X is positive recurrent, then

$$\frac{1}{T}\int_0^T 1\{X_t = i\}dt \to \pi(i) > 0, \text{ as } T \to \infty, \forall i \in \mathcal{X}, a.s.$$

Moreover, π is the unique invariant distribution and

$$P(X_t = i) \to \pi(i), \forall i \in \mathcal{X}.$$

Proof:

Consider the *jump chain* $Y = \{Y_n, n \geq 0\}$ that specifies the sequence of successive values of X. We can see that X and Y must be of the same type.

We leave the details as an exercise.

□

19.4 M/M/1 Queue

Customers arrive according to a Poisson process with rate λ and get served one by one, with independent service times that are exponentially distributed with rate μ, by a single server.

Theorem 19.4. *The number of customers in the queue at time t, X_t, is a Markov chain with rate matrix $q(n, n+1) = \lambda$, $q(n+1, n) = \mu$ for $n \geq 0$. All the other non-diagonal terms are zero.*

This Markov chain is positive recurrent if and only if $\lambda < \mu$. In that case, the unique invariant distribution is

$$\pi(n) = (1-\rho)\rho^n, n \geq 0.$$

Proof:

Write the balance equations. You find $\pi(n) = \rho^n \pi(0), n \geq 0$. There is a solution only if $\rho < 1$.

□

19.5 Time Reversal

We want to examine the Markov chain X_t observed in reverse time. Imagine a movie of the Markov chain that you play in reverse.

The first observation may be a bit surprising.

Fact 19.1. *Assume $\{X_t, t \geq 0\}$ is a Markov chain. In general, $\{Y_t := X_{T-t}, 0 \leq t \leq T\}$ is not a Markov chain.*

Proof:

Exercise.

□

However, we have the nice result.

Theorem 19.5. *Assume X is a regular stationary Markov chain with rate matrix Q and invariant distribution π. Then Y is a regular stationary Markov chain with invariant distribution π and rate matrix Q' where*

$$q'(i,j) = \frac{\pi(j)q(j,i)}{\pi(i)}.$$

Definition 19.2. *Time-Reversible A random process X is time-reversible if $\{X_t, t \in \Re\}$ and $\{X_{T-t}, t \in \Re\}$ have the same distribution for all $T \in \Re$.*

Note that a time-reversible process must be stationary. (Prove that fact.)

Theorem 19.6. *A Markov chain with rate matrix Q is time-reversible iff it is stationary and its invariant distribution π is such that*

$$\pi(i)q(i,j) = \pi(j)q(j,i), \forall i, j \in \mathcal{X}. \tag{19.2}$$

These relations are called the detailed balance equations.

Here is a cute application to the M/M/1 queue.

Theorem 19.7. *The stationary M/M/1 chain is time-reversible. In particular, the departures from the stationary queue form a Poisson process with rate λ whose past up to time t is independent of X_t. It follows that two M/M/1 queue in tandem, when stationary, have independent queue lengths at any given time.*

Proof:

The stationary distribution satisfies the detailed balance equations (19.2), so that the stationary queue length process is time-reversible.

The departures up to time t become the arrivals after time t for the time-reversed queue. Since the time-reversed queue is again an M/M/1 queue, the arrivals after time t are a Poisson process independent of X_t. Therefore, the same is true of the departures before time t.

□

19.6 Solved Problems

Problem 19.1. *Let X_t be the queue length of an M/M/1 queue with arrival and service rates λ and μ, respectively. Explain how to simulate X_t with one Poisson process and a sequence of i.i.d. Bernoulli random variables.*

Solution:

Problem 19.2. *Assume $(N_t, t \geq 0)$ counts arrivals in $[0, t]$ and is such that*

$$P[N_{t+\epsilon} = N_t + 1 | N_s, s \leq t] = \lambda\epsilon + o(\epsilon)\,,$$
$$P[N_{t+\epsilon} = N_t | N_s, s \leq t] = 1 - \lambda\epsilon + o(\epsilon)\,.$$

Show that $(N_t, t \geq 0)$ is a Poisson process with rate λ.

Solution:

Problem 19.3. *Show that a stationary CTMC X_t is time-reversible if and only only if*

$$q(i_0, i_1)q(i_1, i_2)\cdots q(i_{n-1}, i_n)q(i_n, i_0) = q(i_0, i_n)q(i_n, i_{n-1})\cdots q(i_2, i_1)q(i_1, i_0)\,,$$

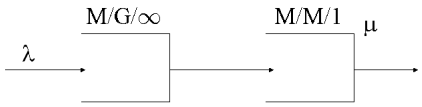
for any finite sequence of states $i_0, i_1, \ldots, i_n$. [Hint: to show the 'if' part, fix a reference state r, and for each j, define

$$\pi(j) = \alpha\frac{q(r, j_0)q(j_1, j_2)\cdots q(j_n, j)}{q(j, j_n)q(j_n, j_{n-1})\cdots q(j_0, r)}\,,$$

where $j_0, j_1, \ldots, j_n$ is a sequence of states leading from r to j.]

Solution:

Problem 19.4. *Consider an M/G/∞ and an M/M/1 queue, in tandem as in the figure. Let λ be the rate of (Poisson) arrivals in the M/G/∞ system, and μ the service rate of the M/M/1 queue.*



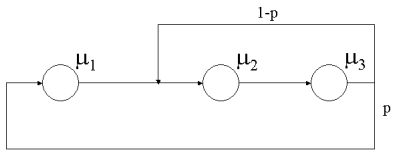
1. *How would you define the "state" X_t^1 of the M/G/∞ system? [Knowledge of the state should determine the future evolution, statistically.]*

2. *Let X_t^2 be the state of the M/M/1 system.*

   Show that if the system is started at $-\infty$, so stationarity is reached at any finite time, X_t^1 and X_t^2 are independent.

Solution:

Problem 19.5. *Consider a closed Jackson network where every customer can visit all the queues, but that no customer enters or exits the system. An example is given in the figure.*



Show that

$$\pi(x) = \pi_1(x_1) \cdots \pi_J(x_J) \,,$$

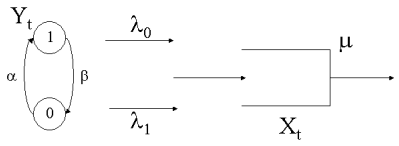
with $\pi_i(n) = \rho_i^n(1-\rho_i)$, $\rho_i = \lambda_i/\mu_i$, is invariant, for any λ_i. (λ_i must satisfy $\lambda_i = \sum_j \lambda_j R(j,i)$, $\lambda_i < \mu_i$ where $R(\cdot,\cdot)$ is the routing matrix.)

[Hint: Use Kelly's lemma.]

Solution:

Problem 19.6. *Consider a $G/M/1$ queue with service rate μ. The arrival process is modulated by the two-state CTMC Y_t as follows: when $Y_t = 0$, arrivals are Poisson with rate λ_0, while when $Y_t = 1$ arrivals are Poisson with rate λ_1.*

Let X_t be the queue length (including the customer in service, if any) at time t.



1. *Show that X_t is not a MC.*
2. *Show that $Z_t = (X_t, Y_t)$ is a MC.*
3. *What conditions on the rates are needed for positive recurrence of Z_t? [Hint: Pake's lemma.]*
4. [This part is optional.] *Write $\pi(x, y) = \pi_x(y), y \in \{0, 1\}$ and let $\pi_x = [\pi_x(0) \quad \pi_x(1)]$. Show that $\pi_x = \pi_0 R^x$ for some matrix R, under the conditions of (c). [Hint: Assume $\pi_x = \pi_0 R^x$ and write the balance equations, $(\pi Q)(x, 0) = 0, (\pi Q)(x, 1) = 0$.]*

Solution:

Chapter 20

Jackson Networks

20.1 Summary

We explain the following ideas and results:

- Kelly's Lemma
- Jackson Networks
- PASTA

20.2 Kelly's Lemma

Here is a useful result. See [6] for these ideas and more.

Lemma 20.1. *Kelly's Lemma*
Assume that Q and Q' are two rate matrices and π a distribution on $\mathcal{X}$ such that

$$\pi(i)q(i,j) = \pi(j)q'(j,i), \forall i,j \in \mathcal{X}. \tag{20.1}$$

Then $\pi Q = 0$ and Q' is the rate matrix of the Markov chain reversed in time.

Proof:
Summing (20.1) over $i \in \mathcal{X}$ and using the fact that $\sum_i q'(j,i) = 0$, we find that $\pi Q = 0$. The other statement then follows from Theorem 5 in Lecture 23.

□

This result enables us to prove that a distribution is invariant by guessing the rate matrix of the process reversed in time. We examine an example next.

Example 20.1. *Consider two M/M/1 queues in series, with arrival rate λ at the first queue and service rates μ_1 and μ_2, respectively. We assume $\lambda < \mu_1, \mu_2$. Let Q be the rate matrix of the pair (X_t^1, X_t^2) of queue lengths. Let also Q' be the rate matrix of the two queues with arrivals into queue 2 whose departures go to queue 1. For instance,*

$$q((m,n),(m-1,n+1)) = \mu_1, q'((m-1,n+1),(m,n)) = \mu_2, q((m,n),(m+1,n)) = \lambda,$$
$$q'((m+1,n),(m,n)) = \mu_1, q((m,n),(m,n-1)) = \mu_2, \text{ and } q'((m,n-1),(m,n)) = \lambda.$$

Define

$$\pi(m,n) = (1-\rho_1)\rho_1^m(1-\rho_2)\rho_2^n, m,n \geq 0 \text{ where } \rho_i = \frac{\lambda}{\mu_i}, i = 1,2.$$

We then see that

$$\pi(m,n)q((m,n),(m-1,n+1)) = \pi(m-1,n+1)q'((m-1,n+1),(m,n)).$$

Similarly,

$$\pi(m,n)q((m,n),(m+1,n)) = \pi(m+1,n)q'((m+1,n),(m,n))$$

and

$$\pi(m,n)q((m,n),(m,n-1)) = \pi(m,n-1)q'((m,n-1),(m,n)).$$

Thus, equations (20.1) hold and we conclude that π is the invariant distribution of the two queues in series. Also, the process reversed in time behaves line arrivals into queue 2 attached to queue 1.

20.3 Jackson Networks

We generalize the example of the previous section. Consider a set of J queues. Customers arrive as independent Poisson processes, with rate γ_i into queue i. Customers face independent service times in all the queues and these are exponentially distributed with rate μ_i in queue i. When a customer leaves queue i, he goes to queue j with probability $R(i,j)$ and leaves the network with probability $1 - \sum_{j=1}^{J} R(i,j)$. This system defines a *Jackson network*.

Theorem 20.1. *Assume that each customer can eventually leave.*

(a) In that case, the equations

$$\lambda_i = \gamma_i + \sum_{j=1}^{J} \lambda_j R(j,i), i = 1,\ldots,J$$

have a unique solution $(\lambda_1,\ldots,\lambda_J)$.

(b) Moreover, if $\rho_i := \lambda_i/\mu_i < 1$ for $i = 1,\ldots,J$, then the vector of queue lengths $X_t = (X_t^1,\ldots,X_t^J)$ is a positive recurrent Markov chain with invariant distribution

$$\pi(n_1,\ldots,n_J) = \pi_1(n_1)\cdots\pi_J(n_J) \text{ where } \pi_i(n) = (1-\rho_i)\rho_i^n, n \geq 0.$$

Proof:

Part (a) is easy. We focus on (b). We show that the process X_t reversed in time corresponds to the network with the flows of customers going in the reversed directions.

The routing matrix R' of the flows reversed in time is such that

$$\lambda_i R(i,j) = \lambda_j R'(j,i).$$

Indeed, the left-hand side if the rate of flow going from i to j whereas the right-hand side is the rate of flow going from j to i in the reversed network. Also, the exogenous arrival rate into queue i for the reversed network should by

$$\lambda_i' = \lambda_i(1 - \sum_{j=1}^{J} R(i,j)).$$

Designate by Q the rate matrix of the original network and by Q' the rate matrix of the reversed network. For instance, if e_i designates the unit vector in direction i, then

$$q(x+e_i, x+e_j) = \mu_i R(i,j) \text{ and } q'(x+e_j, x+e_i) = \mu_j R'(j,i).$$

We can then verify that

$$\pi(x+e_i)q(x+e_i, x+e_j) = \pi(x+e_j)q'(x+e_j, x+e_i),$$

i.e., that

$$\pi(x+e_i)\mu_i R(i,j) = \pi(x+e_j)\mu_j R'(j,i).$$

To see this, note that

$$\pi(x+e_i) = \pi(x)\frac{\lambda_i}{\mu_i} \text{ and } \pi(x+e_j) = \pi(x)\frac{\lambda_j}{\mu_j},$$

so that the previous identity reads, after simplifying by $\pi(x)$,

$$\frac{\lambda_i}{\mu_i}\mu_i R(i,j) = \frac{\lambda_j}{\mu_j}\mu_j R'(j,i),$$

which we see is satisfied. Proceeding in a similar way, we can verify that π, Q, Q' satisfy the conditions of Kelly's Lemma.

□

20.4 PASTA

How large is the backlog that a customer finds when he arrives into a queue? Let us look first at an example.

Assume customers arrive every two seconds into a queue where each service time takes exactly one second. If the queue is initially empty, every customer finds an empty queue when it arrives. However, the average queue length is 1/2 since the queue has one customer half of the time and is empty the other half. Thus, an arriving customer does not see the average queue length.

However, when the arrivals are Poisson, the situation is very simple.

Theorem 20.2. *PASTA*

Assume customers arrive as a Poisson process into a queuing system described by a stationary Markov chain X_t. Then, an arriving customer sees the state with its invariant distribution π.

Proof:

Designate the arrival process by A_t. Then

$$P[X_t = x | A_{t+\epsilon} = A_t + 1] = P(X_t = x) = \pi(x)$$

because the arrival process has independent increments, so that $\{A_{t+s}, s \geq 0\}$ and X_t are independent.

This identity shows that the state X_t just before an arrival time has its invariant distribution.

□

As an application, we study the delay of a customer through a stationary M/M/1 queue.

Theorem 20.3. *The delay of a customer through a stationary M/M/1 queue with arrival rate λ and service rate μ is exponentially distributed with rate $\mu - \lambda$.*

Proof:

Because of PASTA, an arriving customer finds n customers in the queue with probability $(1-\rho)\rho^n$ for $n \geq 0$ where $\rho = \lambda/\mu$. When he joins n other customers, because of the memoryless property of the service times, the customers finds a delay τ equal to the sum of $n+1$ independent service times that are exponentially distributed with rate μ. Hence,

$$E(e^{-s\tau}) = \sum_{n=0}^{\infty}(1-\rho)\rho^n E(e^{-sT})^{n+1}$$

where $T =_D Exp(\mu)$. Thus,

$$E(e^{-sT}) = \int_0^{\infty} e^{-st}\mu e^{-\mu t}dt = \frac{\mu}{s+\mu},$$

so that

$$E(e^{-s\tau}) = \sum_{n=0}^{\infty}(1-\rho)\rho^n(\frac{\mu}{s+\mu})^{n+1} = (1-\rho)\frac{\mu}{s+\mu}[1-\rho\frac{\mu}{s+\mu}]^{-1} = \frac{\mu-\lambda}{s+\mu-\lambda},$$

which shows that $\tau =_D Exp(\mu-\lambda)$.

A more elegant proof is to view τ as the interarrival time of a Poisson process with rate μ sampled with probability $1-\rho$, i.e., of a Poisson process with rate $\mu - \lambda$. Indeed, with probability $\rho^n(1-\rho)$, this interarrival time is the sum of $n+1$ independent exponential random variables with rate μ.

□

Chapter 21

Convergence

21.1 Summary

We explain the following ideas and results:

- Types of Convergence
- Examples
- Key Results
- Large Numbers

21.2 Overview

Our objective is to give you a sense for the concepts of convergence of random variables. The ideas are subtle but it important to understand them. The key results in probability are convergence results.

It turns out that we need a few preliminaries before we can go anywhere with this topics. We review them in Section 21.3. With these ideas out of the way, we define the different notions of convergence in Section 21.4. We illustrate these notions on some representative examples in Section 21.5. In Section 21.6 we clarify the relationships between the different modes of convergence. Finally, in Section 21.7, we discuss some results that lead to the all-important strong law of large numbers. We believe that the intermediate steps are useful to clarify the notions of convergence.

This is a difficult lecture; we hope we can help you understand these very neat ideas.

21.3 Preliminaries

We need a few tools to study limits.

We first recall the following key result about real numbers.

Lemma 21.1. *Cauchy*

The sequence $\{x_n, n \geq 1\}$ *converges to some finite* x *if and only if*

$$\epsilon_n := sup_{m,k \geq n} |x_m - x_k| \to 0 \text{ as } n \to \infty. \tag{21.1}$$

A sequence that satisfies (21.1) is said to be Cauchy. *Thus, Cauchy's Lemma states that Cauchy sequences converge in* $\Re$.

Proof:

The only if part is obvious. We show the if part. Thus assume that the sequence is Cauchy. For any n we find that

$$|x_n| < |x_1| + |x_n - x_1| \leq |x_1| + \epsilon_1.$$

Thus, the sequence $\{x_n\}$ is contained in $I_1 := [-|x_1| - \epsilon_1, |x_1| + \epsilon_1]$. Half of that interval, either the closed left-half or the closed right-half, must contain an infinite subsequence. Designate that half interval by I_2. Continuing in this way, we form a set of closed intervals I_k whose lengths go to zero and with the property that $I_{k+1} \subset I_k$ for all k. The intersection of all these sets contains some point x which is the least upper bound of the $a_n = \min\{y | y \in I_n\}$. Indeed, the a_n are nondecreasing and bounded. It is then easy to check that (21.1) implies $x_n \to x$.

□

Next, we review

Lemma 21.2. *Borel-Cantelli*

Assume that events A_n *are such that* $\sum_n P(A_n) < \infty$, *then* $P(A_n,\ i.o.) = 0$.

Proof:

Note that

$$\{A_n,\ \text{i.o.}\} = \cap_{n\geq 1} \cup_{m \geq n} A_m,\ \text{so that}\ \cup_{m\geq n} A_m \downarrow \{A_n,\ \text{i.o.}\}.$$

This implies that $P(\cup_{m\geq n} A_m) \downarrow P(A_n,\ \text{i.o.})$. But $P(\cup_{m\geq n} A_m) \leq \sum_{m\geq n} P(A_m)$, so that $P(\cup_{m\geq n} A_m) \downarrow 0$.

□

Finally, we will need the following result.

Lemma 21.3. *Kronecker*

Let $\{x_n, n \geq 1\}$ *be real number such that* $\sum_{n=1}^{\infty} x_n = s$ *is finite. Then*

$$\frac{1}{n} \sum_{k=1}^{n} k x_k \to 0.$$

Proof:

To prove the lemma, we do a simple calculation. Let $y_n = \sum_{k=n+1}^{\infty} x_k$ so that $y_0 = s$ and $x_n = y_{n-1} - y_n$. Then,

$$\sum_{k=1}^{n} k x_k = \sum_{k=1}^{n} k(y_{k-1} - y_k) = \sum_{k=0}^{n-1} (k+1) y_k - \sum_{k=1}^{n} k y_k = \sum_{k=1}^{n-1} y_k + y_0 - n y_n.$$

We note that, since $y_k \to 0$, the right-hand side divided by n goes to 0 as $n \to \infty$.

□

21.4 Types of Convergence

We explore the convergence of random variables. Since a random variable is a function, one must define the meaning of convergence. We have the following definitions.

Definition 21.1. *Types of Convergence*

Let X and $\{X_n, n \geq 1\}$ be random variables on some common probability space $\{\Omega, \mathcal{F}, P\}$.

(a) One says that the random variables X_n converge almost surely *to X as $n \to \infty$, and one writes*

$$X_n \xrightarrow{a.s.} X \text{ as } n \to \infty$$

or

$$\lim_{n\to\infty} X_n = X, \text{ a.s.}$$

if

$$P(\{\omega \mid X_n(\omega) \to \mathbf{X}(\omega)\}) = 1.$$

(b) One says that the random variables X_n converge in probability *to X as $n \to \infty$, and one writes*

$$X_n \xrightarrow{P} X \text{ as } n \to \infty$$

if

$$\lim_{n\to\infty} P(\{\omega \mid |X_n(\omega) - \mathbf{X}(\omega)| > \epsilon\}) = 0, \forall \epsilon > 0.$$

(c) One says that the random variables X_n converge in distribution *to X as $n \to \infty$, and one writes*

$$X_n \xrightarrow{D} X \text{ as } n \to \infty$$

if

$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x), \forall x \text{ where } P(X \leq x) \text{ is continuous.}$$

(d) One says that the random variables X_n converge in L_p *to X as $n \to \infty$, and one writes*

$$X_n \xrightarrow{L_p} X \text{ as } n \to \infty$$

if

$$\lim_{n\to\infty} E(|X_n - X|^p) = 0.$$

21.5 Examples

Let us quickly look at some examples to clarify these notions. In this example, $\{\Omega, \mathcal{F}, P\}$ is $[0, 1]$ with the uniform distribution. See Figure 21.1 for an illustration.

Example 21.1. *Let $X = 0$ and $X_n = n \times 1\{\omega \leq 1/n\}$. Then, as $n \to \infty$, one has*

$$X_n \xrightarrow{a.s.} X, \quad X_n \xrightarrow{P} X, \quad X_n \xrightarrow{D} X, \text{ but } X_n \not\xrightarrow{L_1} X.$$

Proof:

Note that $X_n(\omega) = 0$ for all $n > 1/\omega$, so that $X_n(\omega) \to 0$ for all $\omega \neq 0$. Hence $X_n \xrightarrow{a.s.} X$.

We find that, for any $\epsilon > 0$, $P(|X_n - X| > \epsilon) \leq 1/n \to 0$ as $n \to \infty$, so that $X_n \xrightarrow{P} X$.

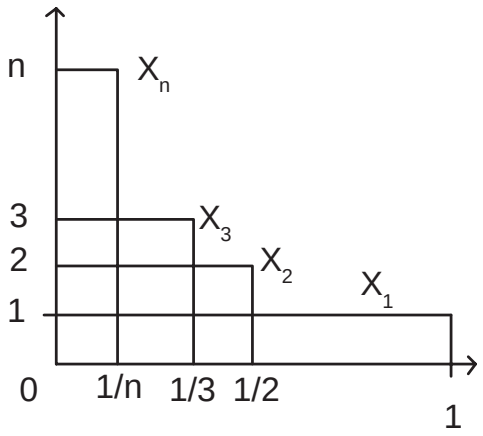


Figure 21.1: The random variables in Example 21.1.

Observe that $P(X \leq x) = 1\{x \geq 0\}$ is continuous at every $x \neq 0$. Also, for $x < 0$ we have $P(X_n \leq x) = 0 \to P(X \leq x) = 0$. Moreover, for $x > 0$, we have $P(X_n \leq x) \geq 1 - 1/n \to 1 = P(X \leq x)$, so that $X_n \xrightarrow{P} X$.

Finally, $E(|X_n - X|) = 1 \not\to 0$, so that $X_n \xrightarrow{L_1} X$.

□

Example 21.2. *Let* $\{X_n, n \geq 1\}$ *be i.i.d.* $N(0,1)$ *random variables. Then* $X_n \xrightarrow{D} X$ *but* $X_n \xrightarrow{a.s.} X$, $X_n \xrightarrow{P} X$, *and* $X_n \xrightarrow{L_1} X$.

Example 21.3. *Let* $X_1 = 1, X_2(\omega) = 1\{\omega \leq 0.5\}, X_3(\omega) = 1\{\omega > 1/2\}, X_4 = 1\{\omega \leq 1/4\}, X_5(\omega) = 1\{\omega \in (1/4, 1/2]\}, X_6(\omega) = 1\{\omega \in (1/2, 3/4]\}$, *etc, as shown in Figure 21.2. Then* $X_n \xrightarrow{P} 0$ *but* $X_n \xrightarrow{a.s.} X$.

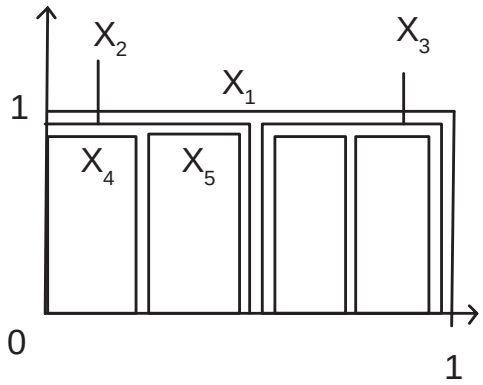


Figure 21.2: The random variables in Example 21.3.

Example 21.4. *For* $n \geq 1$, *let* $X_n =_D U[-1/(2n), 1/(2n)]$, *as shown in Figure 21.3. Let also* $X = 0$. *Then* $X_n \xrightarrow{D} X$. *Note that* $P(X_n \leq 0) = 1/2 \not\to P(X \leq 0) = 1$.

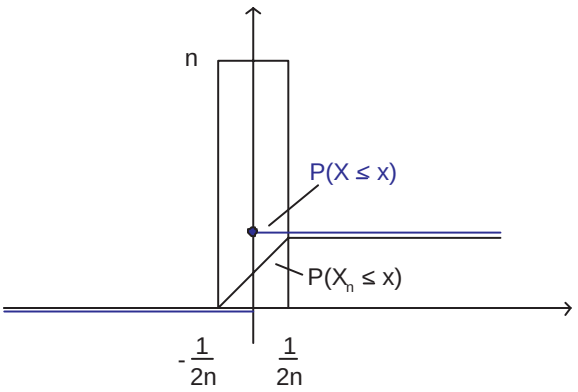


Figure 21.3: The random variables in Example 21.4.

21.6 Relationships

How do these types of convergence relate? How do we prove convergence and what can we conclude from it?

We first review the following result we saw in Lecture 17 and that we do not prove here:

Theorem 21.1. *Lebesgue Convergence Theorem*

(a) Monotone Convergence Theorem: MCT. *Assume that* $0 \leq X_n \uparrow_{a.s.} X$ *as* $n \to \infty$. *Then* $E(X_n) \uparrow E(X)$ *as* $n \to \infty$.

(b) Dominated Convergence Theorem: DCT. *Assume that* $X_n \xrightarrow{a.s.} X$ *as* $n \to \infty$ *and* $|X_n| \leq Y$ *for* $n \geq 1$ *with* $E(Y) < \infty$. *Then* $E(X_n) \to E(X)$ *as* $n \to \infty$.

We start with a counterpart of Cauchy's Lemma for random variables.

Lemma 21.4. *We say that* X_n *is Cauchy a.s., in* P, *in* L_1 *(respectively) if, as* $n \to \infty$ *and for all* $\epsilon > 0$,

$$\sup_{m,k \geq n} |X_m - X_k| \xrightarrow{a.s} 0, \ \sup_{m,k \geq n} P(|X_m - X_k| > \epsilon) \to 0, \ \sup_{m,k \geq n} E(|X_m - X_k|) \to 0, \ \textit{(respectively).}$$

The result is that a sequence that is Cauchy a.s., in P, *in* L_1 *(respectively) converges in the corresponding sense to some finite random variable.*

Proof:

(a.s.): The result is obvious.

(P): Take $\epsilon_k \downarrow 0$ with $\sum_k \epsilon_k < \infty$ and n_k such that $n, m \geq n_k$ implies that

$$P(|X_m - X_n| > \epsilon_k) \leq 2^{-k}.$$

We now use Borel-Cantelli to conclude that

$$P(|X_{n_{k+1}} - X_{n_k}| > \epsilon_k, \text{ i.o.}) = 0.$$

Thus, for all ω, one has $|X_{n_{k+1}} - X_{n_k}| \leq \epsilon_k$ for all $k > k(\omega)$. This implies that, for $m > p > k(\omega)$,

$$|X_{n_m} - X_{n_p}| \leq |X_{n_m} - X_{n_{m-1}}| + \cdots + |X_{n_{p+1}} - X_{n_p}| \leq \sum_{k=p}^{\infty} \epsilon_k \to 0 \text{ as } p \to \infty.$$

Thus, X_{n_k} is Cauchy a.s. and it follows that $X_{n_k} \xrightarrow{a.s.} X$ for some finite X. We now show that $X_n \xrightarrow{P} X$. To do this, we observe that

$$P(|X_n - X| > \epsilon) \leq P(|X_n - X_{n_k}| > \epsilon/2) + P(|X_{n_k} - X| > \epsilon/2)$$

and each term can be made as small as we wish by picking n and n_k large enough.

(L_1): Assume X_n is Cauchy in L_1. That is,

$$\sum_{m>n} E|X_m - X_n| \to 0, \text{ as } n \to \infty.$$

Using Chebyshev's inequality, we see that X_n is Cauchy in P, so that it converges in probability to some random variable X. To see that it converges in L_1 to X, we use DCT.

□

Example 21.5. *Assume that the random variables $\{X_n, n \geq 1\}$ are i.i.d. with $E(X_n) = \mu$ and $var(X_n) = \sigma^2 < \infty$. Then*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{P} \mu.$$

This result is called the weak law of large numbers.

Proof:

Recall that $\epsilon 1\{|Y| \geq \epsilon^2\} \leq Y^2$, so that $P(|Y| \geq \epsilon) \leq E(Y^2)/\epsilon^2$, which is Chebychev's inequality. It follows that

$$P(|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \to 0,$$

which proves the result.

□

We use that result in the following example.

Example 21.6. *Assume that the random variables X_n are independent and such that $E(X_n) = 0$ for all n and $\sum_{n=1}^{\infty} E(X_n^2) < \infty$. Then $\sum_{k=1}^{n} X_k \xrightarrow{P} Z$ for some finite random variable Z.*

Proof:

Let $Z_n = \sum_{k=1}^{n} X_k$ for $n \geq 1$. Fix $k < m$ and $\epsilon > 0$ and note that

$$P(|Z_m - Z_k| > \epsilon) \leq \sum_{n=k+1}^{m} E(X_n^2)/\epsilon^2.$$

Hence,

$$\sup_{m>k\geq n} P(|Z_m - Z_k| > \epsilon) \to 0 \text{ as } n \to \infty.$$

It follows that the sequence is Cauchy in P and converges in probability to some finite random variable.

□

The next result shows the relationship between the convergence types.

Theorem 21.2. *Figure 21.4 indicates the relationships between the convergence types. The positive results are the following:*

- *(1) A.s. convergence implies convergence in probability;*
- *(2) Convergence in probability implies convergence in distribution;*
- *(3) Convergence in L_1 implies convergence in probability.*

The other implications do not hold, except the following:

- *(2') Convergence in distribution to a constant implies convergence in probability to that constant;*
- *(4) A.s. convergence with differences that satisfy one of the conditions or the Lebesgue Convergence Theorem implies convergence in L_1.*

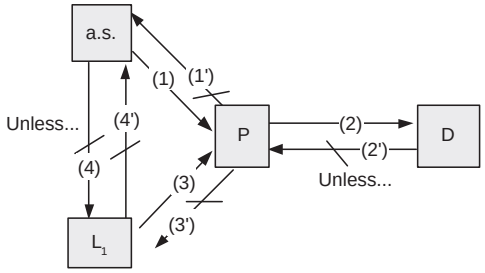


Figure 21.4: The relationships in Theorem 21.2.

Proof:

We sketch the proofs.

(1) Assume $X_n \xrightarrow{a.s.} X$ and fix $\epsilon > 0$. Then

$$Z_n := 1\{|X_n - X| > \epsilon\} \xrightarrow{a.s.} 0 \text{ as } n \to \infty.$$

Moreover, $|Z_n| \leq 1$ for $n \geq 1$. We conclude from DCT that $E(Z_n) \to 0$, i.e., that

$$P(|X_n - X| > \epsilon) \to 0, \text{ as } n \to \infty.$$

(1') Example 21.3 shows that convergence in probability does not imply almost sure convergence.

(2) We start with the following observation. Assume $P(B_n) \to 1$. Then $P(A \cap B_n) \to P(A)$. To see this, note that

$$P(A) \geq P(A \cap B_n) = 1 - P(A^c \cup B_n^c) \geq 1 - P(A^c) - P(B_n^c) = P(A) - P(B_n^c) \to P(A).$$

Assume $X_n \xrightarrow{P} X$. Then, for any $\epsilon > 0$,

$$\{X \leq x - \epsilon\} \cap \{|X_n - X| \leq \epsilon\} \subset \{X_n \leq x\}.$$

Using our previous observation, we see

$$P(X \leq x - \epsilon) \leq \liminf P(X_n \leq x).$$

Similarly, we see that

$$\{X > x + \epsilon\} \cap \{|X_n - X| \leq \epsilon\} \subset \{X_n > x\}.$$

Hence,

$$P(X > x + \epsilon) \leq \liminf P(X_n > x),$$

which is equivalent to

$$P(X \leq x + \epsilon) \geq \limsup P(X_n \leq x).$$

Letting $\epsilon \to 0$ and assuming that $P(X \leq x)$ is continuous at x, we find

$$P(X \leq x) \leq \liminf P(X_n \leq x) \leq \limsup P(X_n \leq x) \leq P(X \leq x),$$

which proves that $P(X_n \leq x) \to P(X \leq x)$.

[Recall that $\liminf x_n := \lim_{n\to\infty} \inf_{m\geq n} x_m$ is well-defined because $\inf_{m\geq n} x_m$ is nondecreasing. Similarly, $\limsup x_n = \lim_{n\to\infty} \sup_{m\geq n} x_m$. Also, $x_n \to \xi$ if and only if $\liminf x_n = x = \limsup x_n$. Finally, $\liminf(-x_n) = -\limsup x_n$.]

(2') See Example 21.2. However, if X is constant, then this implication holds. To see this, assume that $X_n \xrightarrow{D} 0$. Fix $\epsilon > 0$. We claim that $P(|X_n| > \epsilon) \to 0$. Indeed, $P(X_n \leq -\epsilon) \to 0$ and $P(X_n \leq \epsilon) \to 1$ because of the convergence in distribution.

(3) One has $\epsilon 1\{|X_n - X| \geq \epsilon\} \leq |X_n - X|$, so that $P(|X_n - X| \geq \epsilon) \leq E(|X_n - X|)/\epsilon$.

(3') See Example 21.1.

(4) In general, a.s. convergence does not imply convergence in L_1, as Example 21.1 shows. However, if $X_n - X$ satisfies the conditions of DCT or of MCT, then we can conclude that $E(|X_n - X|) \to 0$, which implies convergence in L_1.

(4') See Example 21.3.

□

21.7 Large Numbers

Our goal is to prove Theorem 21.3. We use the proof in [1], Section 3.6. Leo Breiman died on July 5, 2005 at the age of 77. He had been a Professor of Statistics at U.C. Berkeley since 1980 and made numerous contributions to statistics, notably on classification methods. You might enjoy his elegant book.

We first show that the convergence in Example 21.6 is almost sure.

Lemma 21.5. *Levy Assume that the random variables X_n are independent and such that $E(X_n) = 0$ for all n and $\sum_{n=1}^{\infty} E(X_n^2) < \infty$. Then $\sum_{k=1}^{n} X_k \xrightarrow{a.s.} Z$ for some finite random variable Z.*

Proof:

Let $Z_n = \sum_{k=1}^{n} X_k$. Assume that Z_n does not converge with some positive probability. Let $W_n := \sup_{m>n} |Z_m - Z_n|$. When Z_n does not converge, $\inf_{n\geq 1} W_n > 0$; for otherwise the sequence would be Cauchy and converges. Thus, for a fixed n, $P(W_n > 0) > 0$ and, consequently, there is some fixed $\epsilon > 0$ and some δ such that $P(W_n > \epsilon) \geq \delta > 0$.

We show that this is not possible. Define $C(n,N) = \sup_{n\le m<N} P(|Z_N - Z_m| > \epsilon)$. Let $\tau = \min\{k > n \mid |Z_k - Z_n| > 2\epsilon\}$. Then

$$P(\sup_{n<m\le N} |Z_m - Z_n| > 2\epsilon, |Z_N - Z_n| > \epsilon) = \sum_{k=n+1}^{N} P(|Z_N - Z_n| > \epsilon, \tau = k) \ge \sum_{k=n+1}^{N} P(|Z_N - Z_k| \le \epsilon, \tau = k)$$

$$= \sum_{k=n+1}^{N} P(|Z_N - Z_k| \le \epsilon) P(\tau = k) \ge (1 - C(n,N)) \sum_{k=n+1}^{N} P(\tau = k) = (1 - C(n,N)) P(\sup_{k\le N} |Z_k - Z_n| > 2\epsilon$$

Hence,

$$P(\sup_{k\le N} |Z_k - Z_n| > 2\epsilon) \le \frac{1}{1 - C(m,N)} P(|Z_N - Z_n| > \epsilon).$$

Now, observe that since the series converges in probability, $C(n,N) \to 0$ as $n \to \infty$. Similarly that convergence in probability also implies that $P(|Z_N - Z_n| > \epsilon) \to 0$ as $n \to \infty$.

Consequently, $P(\sup_{n<k\le N} |Z_k - Z_n| > 2\epsilon) \to 0$, which proves the contradiction.

□

The following theorem is very important. It "confirms" the frequentist interpretation of probability.

Theorem 21.3. *Strong Law of large Numbers*

Let X_n be i.i.d. random variables with $E(|X_n|) < \infty$. Then

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{a.s.} E(X_1) \text{ as } n \to \infty.$$

Proof:

Step 1: The result holds if $E(X_n) = 0$ and $\mathrm{var}(X_n) = \sigma^2$.

By Kronecker's Lemma, to show that $(\sum_{k=1}^{n} X_k)/n \to 0$, it suffices to show that $\sum_{k=1}^{n} (X_k)/k$ converges to a finite random variable. (Indeed, let $x_k = kX_k$ in Kronecker's Lemma.) To show that, in view of Levy's Lemma 21.5, it suffices to prove that

$$\sum_{k=1}^{\infty} E(X_k^2)/k^2 < \infty,$$

which is seen to hold.

Step 2: The result holds under the conditions of the theorem.

Define $\tilde{X}_n = X_n 1\{|X_n| \le n\}$. We also define $Y_n = \tilde{X}_n - E(\tilde{X}_n)$. We show below that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} E(\tilde{X}_n^2) < \infty. \tag{21.2}$$

Step 1 then implies that

$$\frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow{a.s.} 0, \text{ so that } \frac{1}{n}\sum_{k=1}^{n} (\tilde{X}_k - E(\tilde{X}_k)) \xrightarrow{a.s.} 0.$$

Moreover, DCT implies that $E(\tilde{X}_n) \to E(X_n)$, so that

$$\frac{1}{n}\sum_{k=1}^{n} \tilde{X}_k \xrightarrow{a.s.} E(X_1).$$

In addition, we claim that $P(|X_n| > n, \text{ i.o.}) = 0$, so that the result above proves the theorem. To see the claim, note that

$$\sum_{n=1}^{\infty} P(|X_n| > n) < \infty,$$

because $E(|X_n|) < \infty$.

It remains to prove (21.2). We have

$$\sum_{n=1}^{\infty} \frac{1}{n^2} E(\tilde{X}_n^2) = \sum_{n=1}^{\infty} \frac{1}{n^2} \int_{|x| \le n} x^2 dF(x) = \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{1}{n^2} \int_{k-1<|x|\le k} x^2 dF(x)$$
$$= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \frac{1}{n^2} \int_{k-1<|x|\le k} x^2 dF(x) \le \sum_{k=1}^{\infty} \frac{2}{k} \int_{k-1<|x|\le k} x^2 dF(x) \le 2 \sum_{k=1}^{\infty} \int_{k-1<|x|\le k} |x| dF(x) \le 2E(|X_1|) < \infty.$$

(We used the fact that $\sum_{n=k}^{\infty} 1/n^2 \le 2/k$.)

□

21.8 Solved Problems

Problem 21.1. *Suppose $(X_n, n \ge 1)$ is a sequence of (not necessarily independent) r.v's. Assume $\epsilon_n \downarrow 0$ and*

$$P(|X_k| \ge \epsilon_n) \le \frac{1}{n^2}, \forall n \ge 1.$$

Show $X_n \xrightarrow{a.s.} 0$, as $n \to \infty$.
[Hint: Use Borel-Cantelli lemma.]

Solution:

Problem 21.2. *Assume $\epsilon_n \downarrow 0$ and*

$$P(|X_k - X_m| > \epsilon_n) \le 2^{-n}, \text{ for all } k, m \ge n \ge 1.$$

Show that $(X_n, n \ge 1)$ is Cauchy, a.s., and conclude that $X_n \xrightarrow{a.s.} X$, for some finite r.v. X.
[Hint: Use Borel-Cantelli lemma.]

Solution:

Problem 21.3. *Assume*

$$\sup_{k,m \ge n} P(|X_k - X_m| > \epsilon) \to 0, \text{ as } n \to \infty, \forall \epsilon > 0.$$

Show that $X_n \xrightarrow{p} X$, for some finite r.v. X.
[Hint: Choose $\epsilon_i \downarrow 0$, and define n_i such that

$$P(|X_k - X_m| > \epsilon_i) \le 2^{-i}, \text{ for all } k, m \ge n_i.$$

Use the previous problem to conclude $X_{n_i} \xrightarrow{a.s.} X$, for some finite r.v. X. Conclude that $X_{n_i} \xrightarrow{p} X$.]

Solution:

Problem 21.4. *Construct an example of random variables* $(X_n, n \geq 1)$*, such that*

$$P(|X_n - X_m| > \epsilon) \leq 0.01\,, \text{ for all } n, m\,,$$

but

$$P(\sup_{n \neq m} |X_n - X_m| > \epsilon) = 1\,.$$

Solution:

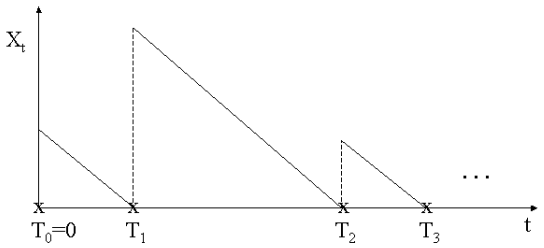
Problem 21.5. *Let* X_k *be a positive recurrent discrete-time MC, and* $f(\cdot)$ *any bounded function. Assume* Y *is distributed according to the invariant distribution of the MC.*

Use SLLN to show

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{a.s.} E(f(Y))\,.$$

Solution:

Problem 21.6. *Consider a renewal process with events at times* $T_0, T_1, \ldots$*, i.e.,* $(T_n - T_{n-1}, n \geq 1)$ *are i.i.d. Assume* $T_0 = 0$*, and define* $X_t = \min\{T_k - t | k \geq 0, T_k - t \geq 0\}$ *the residual time for the first event after time* $t > 0$*.*



Solution:

Problem 21.7. *Show*

$$\frac{1}{T}\int_0^T X_t dt \xrightarrow{a.s.} \frac{E((T_1 - T_0)^2)}{2E(T_1 - T_0)}\,, \text{ as } T \to \infty\,.$$

Solution:

Problem 21.8. *Let N_t be the number of events up to including time $t \geq 0$, i.e., $N_t = \min\{n \geq 1 : T_{n-1} \geq t\}$.*

Show

$$\frac{N_t}{t} \xrightarrow{a.s.} \frac{1}{E(T_1 - T_0)} .$$

Solution:

Problem 21.9. *Let X_t be a positive recurrent CTMC with invariant distribution π. The chain X_t modulates the arrivals of some counting process N_t as follows: while $X_t = i$, N_t increases according to an independent Poisson process of rate $\lambda(i)$. Assume $\max_i \lambda(i) < \infty$.*

Show that

$$\frac{N_t}{t} \xrightarrow{a.s.} \lambda ,$$

where $\lambda = \sum_i \lambda(i)\pi(i)$.

Solution:

Chapter 22

Renewal Processes

22.1 Summary

We introduce a few basic ideas of renewal processes:

- Renewal Process
- Variations

22.2 Overview

As the name indicates, a renewal process is one that "renews" itself regularly. That is, there is a sequence of times $\{T_n, n \in \mathbb{Z}\}$ such that the process after time T_n is independent of what happened before that time and has a distribution that does not depend on n. We have seen examples of such processes before. As a simple example, one could consider a Poisson process with jump times T_n. As another example, we could consider a positive recurrent Markov chain and T_n to be successive visits to a given state.

The basic tool to study such processes, as you probably expect, is the strong law of large numbers, with a twist. The theory we look at is very cute and certainly useful.

We start with a the simple case: the renewal process. We then examine variations.

22.3 Renewal Process

A renewal process is an arrival process with i.i.d. interarrival times.

Definition 22.1. *Renewal Process*

Let $\{T_n, n \geq 1\}$ be such that $T_n < T_{n+1}$ and $\{T_{n+1} - T_n, n \geq 1\}$ are i.i.d.. If T_1 has the same distribution as $T_2 - T_1$, this process is called a renewal process*; otherwise, it is a —em delayed renewal process. We also define $N_t = \max\{n \geq 1 \mid T_n \leq t\}$ with $N_t = 0$ for $t < T_1$ for $t \geq 0$. See Figure 22.1.*

For simplicity, we assume throughout this section that $\lambda^{-1} := E(T_2 - T_1) < \infty$.

Here are a few representative questions we can ask about a delayed renewal process.

1. In what sense is the long term rate of jumps equal to λ?

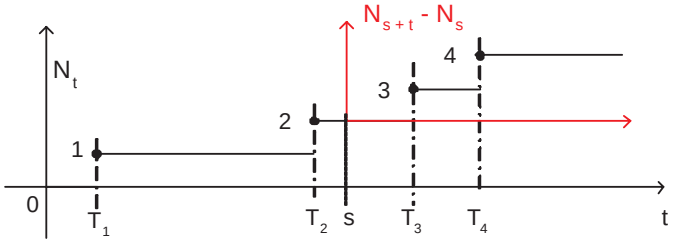


Figure 22.1: Renewal process in Definition 22.1.

2. How do we choose T_1 to make N_t stationary in the sense that the jump times after time t look like the jump times after time 0?

3. How many jumps do we expect to see in an interval of T seconds?

4. How long after t do we have to wait for the next jump?

We examine these questions below.

The rate of jumps is obtained from the strong law of large numbers that implies the following result, as you proved in a homework assignment.

Fact 22.1. *One has*

$$\frac{N_t}{t} \xrightarrow{a.s.} \lambda \text{ as } t \to \infty.$$

We now examine stationarity. We start with a definition.

Definition 22.2. *(a) A* point process *is a collection* $\mathcal{T} := \{T_n, n \geq 1\}$ *of random variables such that* $0 < T_1 < T_2 < \cdots$. *We also define the corresponding* counting process $\mathbf{N} = \{N_t, t \geq 0\}$ *where* $N_t = 0$ *if* $t < T_1$ *and* $N_t = \max\{n \geq 1 \mid T_n \leq t\}$ *otherwise.*

(b) A counting process $\mathbf{N}$ *is* stationary *if* $N \circ \theta_s := \{N_{t+s} - N_s, t \geq 0\}$ *has the same distribution for all* $s \geq 0$. *Thus, the process* $\mathbf{N} \circ \theta_s$ *is simply the jump process watched after time* s. *See Figure 22.1.*

The following result explains how to make $\mathbf{N}$ stationary.

Theorem 22.1. *Let* $\{N_t, t \geq 0\}$ *be a delayed renewal process and* $F(t) = P(T_2 - T_1 \leq t)$ *and* $G(t) = P(T_1 \leq t)$ *for* $t \geq 0$. *The process* $\{N_t, t \geq 0\}$ *is stationary if and only if*

$$G(t) = \lambda \int_0^t (1 - F(s))ds, t \geq 0 \text{ where } \lambda^{-1} = \int_0^\infty sdF(s) = E(T_2 - T_1). \tag{22.1}$$

In particular, if the process is stationary, the average time until the next jump, $E(T_1)$ *is given by*

$$E(T_1) = \frac{\lambda}{2} E((T_2 - T_1)^2). \tag{22.2}$$

Before we attempt a proof, let us look at a few examples.

Example 22.1. *(a) If* $T_2 - T_1 = 1$, *then* $T_1 =_D U[0, 1]$ *and* $E(T_1) = 1/2 = E(T_2 - T_1)/2$, *as we would expect.*

(b) If $T_2 - T_1 =_D Exp(\lambda)$, *then* $T_1 =_D Exp(\lambda)$ *and* $E(T_1) = E(T_2 - T_2)$, *which follows from the memoryless property of the exponential distribution but is nevertheless somewhat counter-intuitive when we compare with (a).*

(c) If $T_2 - T_1 =_D U[0, 1]$, *then* $F(t) = t(2 - t)$ *for* $t \in [0, 1]$ *and* $F(t) = 1$ *for* $t \geq 1$. *In this case,* $E(T_1) = 2/3$ *which is a bit larger than* $E(T_2 - T_1) = 1/2$.

(d) If $P(T_2 - T_1 = a) = p = 1 - P(T_2 - T_1 = b)$ *with* $0 < a < b$, *then* $G(t) = at + (1 - p)(\min\{t, b\} - a)^+$. *Here, we find*

$$E(T_1) = \frac{1}{2} \frac{pa^2 + (1 - p)b^2}{pa + (1 - p)b} > \frac{1}{2} E(T_2 - T_1) \text{ unless } a = b.$$

These examples show that the average time until the next jump, for a stationary process, is larger than half the average time between two successive jumps. This paradox can be explained by observing that a time t picked "randomly" is more likely to fall in a large gap between two jumps than in a small one. Thus, stationarity yields a bias in favor of the large gaps.

We can pursue this intuitive discussion by noting that the probability that t falls in a gap of length u should be proportional to u times the probability that a gap has this length, and should thus be $\gamma u dF(u)$. For this density to sum to one, we need $\gamma = \lambda$. If t falls in a gap of length u, then the time until the next jump should be uniform in $[0, u]$. This suggests that

$$P(T_1 \in (s, s + \epsilon)) = \int_s^\infty [\lambda u dF(u)] \times \frac{\epsilon}{2u} = \frac{\lambda}{2}[1 - F(s)]\epsilon.$$

This argument justifies (e.renewal1). We now turn to a proof.

Proof of Theorem 22.1

Here is a sketch. Designate by $G_s(t)$ the probability that $\mathbf{N}$ has a jump in $[s, s+t]$. This is the cpdf of the first jump time of $\mathbf{N} \circ \theta_s$. The process $\mathbf{N}$ is stationary iff G_s does not depend on s. Let $0 < \epsilon \ll 1$ and observe that

$$T_1 \circ \theta_{s+\epsilon} \leq t \Leftrightarrow T_1 \circ \theta_s \in [\epsilon, t+\epsilon) \text{ or } T_1 \circ \theta_s \in [0, \epsilon) \text{ and } (T_2 - T_1) \circ \theta_s \leq t.$$

Hence,

$$\begin{aligned} G_{s+\epsilon}(t) &= G_s(t+\epsilon) - G_s(\epsilon) + G_\sigma(\epsilon)F(t) \\ &= G_s(t) + G_s'(t)\epsilon - G_s(\epsilon)\{1 - F(t)\}. \end{aligned}$$

Subtracting $G_s(t)$ and dividing by ϵ, we find

$$\frac{G_{s+\epsilon}(t) - G_s(t)}{\epsilon} = G_s'(t) - G_s'(0)\{1 - F(t)\}.$$

The left-hand side is zero iff $G_s'(t) = G_s'(0)\{1 - F(t)\}$, which proves the result.

□

Here is a simple consequence of stationarity.

Fact 22.2. *Assume that* $\mathbf{N}$ *is a stationary point process. Then*

$$E(N_{t+L} - N_t) = \lambda L, \forall L, t \geq 0$$

where $\lambda := E(N_1)$.

Proof:

By stationarity, $E(N_{t+\epsilon} - N_t)$ does not depend on t. By decomposing the interval in segments of length ϵ, we conclude that $E(N_{t+L} - N_t)$ is proportional to L.

□

We now turn to the expected number of jumps in an interval of length L. To appreciate the issue, consider the following examples.

Example 22.2. *(a) Assume* $T_1 = T_2 - T_1 = \alpha$, *for some positive constant* α. *Then we see that* $E(N_{t+L} - N_t)$ *does not converge unless* L *is an integer multiple of* α.

(b) Assume that $T_1 = 0$ *and* $T_2 - T_1$ *takes values in a set of integer multiples of* $\alpha > 0$. *Then, the same conclusion holds as in (a).*

In case (b), we say that the distribution of $T_2 - T_1$ is "lattice." We rule out that case below to avoid having to consider special values of L. The following remarkable result is due to David Blackwell (a Professor in Statistics at Berkeley).

Theorem 22.2. *Blackwell*

Assume that F *is non-lattice. Then, for all* $L > 0$,

$$E(N_{t+L} - N_t) \to \lambda L \text{ as } t \to \infty.$$

Proof:

There are two types of proof. One is analytical and studies distributions through characteristic functions. The other is probabilistic and is essentially a coupling argument. We briefly comment on the second proof which is due to Linvall.

Consider two independent versions of the process $\mathbf{N}$. We call one version the process $\mathbf{N}$ and the other the process $\mathbf{N}'$. Assume that $\mathbf{N}'$ is stationary. Couple the two processes the first time they jump within ϵ seconds of each other. If t is after the coupling time, they have essentially the same number of jumps in $[t, t+L]$, so that the expected number is λL as we saw for a stationary point process. The trick is to show that they couple in finite time. The details are a bit too long to cover here.

□

The last question is the time until the next jump.

Theorem 22.3. *If F is non-lattice, then the distribution of the time until the next jump after time t converges to G, where G is defined in (22.1).*

Proof:

This result follows from the same coupling argument as in the proof of the Blackwell's theorem.

□

22.4 Variations

The renewal process only involves times. We can make more complex constructions.

Example 22.3. *Markov Chain*

Let $X = \{X_t, t \in \Re\}$ be a positive recurrent Markov chain on $\mathcal{X}$. Let $\{T_n, n \in \mathbb{Z}\}$ be the successive return times to a give state $i \in \mathcal{X}$ with the convention that $T_0 < 0 \leq T_1$. Then, for any $n \in \mathbb{Z}$,

$$\{X_{T_n+t}, t \geq 0\}$$

is independent of $\{X_{T_n+t}, t \leq 0\}$ and has a distribution that does not depend on n. In particular, the cycles

$$C_n := \{(t - T_n, X_t), t \in [T_n, T_{n+1}\}$$

are i.i.d. for $n \in \mathbb{Z}$. Thus, $\{T_n, n \in \mathbb{Z}\}$ is a delayed renewal process.

We can view C_n as a graph as shown in Figure 22.2.

Using the fact that the cycles are i.i.d., we can apply the law of large numbers and get results of the following form, as you did in a homework assignment.

Fact 22.3. *Let $f(.) : \mathcal{X} \to \Re$ be bounded function. Then*

$$\frac{1}{T}\int_0^T f(X_t)dt \xrightarrow{a.s.} \sum_i \pi(j)f(i)$$

where π is the invariant distribution of the Markov chain.

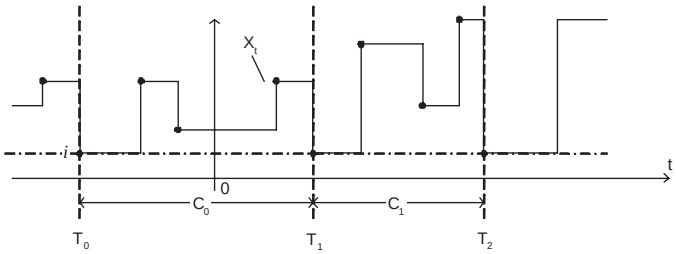


Figure 22.2: Cycle n in Example 22.3.

Another interesting variation concerns *semi-Markov processes*. Here is a definition.

Definition 22.3. *Semi-Markov Process*

Define the process $\{X_t, t \geq 0\}$ *on a countable set* $\mathcal{X}$ *as follows. One is given a transition matrix* P *on* $\mathcal{X}$. *For* $i \in \mathcal{X}$, *let* F_i *be a cpdf on* $[0, \infty)$. *Let also* π_0 *be a pmf on* $\mathcal{X}$. *Choose* X_0 *in* $\mathcal{X}$ *according to the pmf* π_0. *If* $X_0 = i$, *then pick some random variable* τ_1 *with cpdf* F_i. *Define* $X_t = i$ *for* $t \in [0, \tau_1)$. *Let* $P[X_{\tau_1} = j \mid \mathbf{X}_0 = i, \tau_1] = P(i, j)$. *Resume the construction from time* τ_1 *as if it started at time* 0.

This process is called a semi-Markov process *with transition probabilities* P, *holding time distributions* F_i, *and initial distribution* π_0.

You see that this definition is a simple extension of the continuous-time Markov chain where we replaced the exponential holding times by more general distributions.

Of course, this process is no longer Markov. Nevertheless, we expect many of the key properties to hold. For instance, if the holding times are non-lattice, finite mean, and if the jump chain is positive recurrent, then we expect asymptotic stationarity and almost sure convergence of the fraction of time in the states. All this requires a little bit of work.

Chapter 23

Review: Part 1 - Preliminaries, Detection & Estimation

23.1 Overview

These notes summarize the key results of the first part of the course. Make sure you understand these results and that you can apply them.

23.2 Linear Algebra

Linear Algebra is the study of linear transformations.

We used the following result on how to diagonalize an Hermitian matrix:

II.1. Theorem: Diagonalization of Hermitian Matrix (Notes on Linear Algebra - Theorem 6)

Let $H \in \mathbb{C}^{n\times n}$ be a Hermitian matrix, i.e., such that $H = H^*$.

The eigenvalues $\lambda_1, \ldots, \lambda_n$ of H are real (they are not necessarily distinct);

H has n orthonormal eigenvectors $\{u_1, \ldots, u_n\}$ that form a basis for $\mathbb{C}^n$. That is, $u_i^* u_j = 1\{i = j\}$.

If $P = [u_1|\ldots|u_n]$, then

$$P^*HP = \Lambda = diag(\lambda_1, \ldots, \lambda_n)$$

and

$$H = P\Lambda P^* = \sum_{i=1}^{n} \lambda_i u_i u_i^*.$$

In particular, H maps a unit hypercube with sides u_i into a box with orthogonal sides Pu_i and volume $\det(H)$.

We use the first part of II.1 to study covariance matrices. Here is the main result:

II.2. Theorem: Properties of Covariance Matrix (Lecture 6 -Theorem 1)

A covariance matrix K is a symmetric positive semi-definite matrix and conversely.

There is an orthonormal matrix Q such that $KQ = Q\Lambda$ where $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$. Here, the λ_i are the eigenvalues of K; they are real and nonnegative.

There is a unique positive semi-definite matrix R such that $K = R^2$ and

$$R = Q\Lambda^{1/2}Q^T.$$

We use the second part of II.1 in IV.1 below (change of variables).

When we studied linear systems, we used

II.3. Lemma: Cayley-Hamilton (Lemma 1 in Lectures 13-14)
If $A \in \Re^{n\times n}$, then A^k is a linear combination of $\{I, A, \ldots, A^{n-1}\}$.

A consequence of that result is

II.4. Fact: Observability and Reachability (Fact 1 of Lectures 13-14.)

$rank[C^T|A^TC^T|(A^2)^TC^T|\cdots] = rank[C^T|A^TC^T|\cdots|(A^{m-1})^TC^T]$ and $rank[C|AC|A^2C|\cdots] = rank[C|AC|\cdots|A^m$

The rank of the first (resp., second) matrix is full if and only if the corresponding system is observable (resp., reachable).

23.3 Probability

Let us review only a few of the important ideas. Typical probability spaces $\{\Omega, \mathcal{F}, P\}$ are $(\Omega, 2^\Omega, P(A) = \sum_{i\in A} p_i)$ when Ω is finite, $(\Re, \mathcal{B}(\Re), P)$ where P is σ-additive and $P(\Re) = 1$. The σ-additivity means that if $A_n \downarrow A$, then $P(A_n) \downarrow P(A)$. We use that fact many times, in particular to prove Borel-Cantelli that states that if $\sum_n P(B_n) < \infty$, then $P(B_n \text{ i.o.}) = 0$. We also use that continuity below in III.1. In the previous example, $F(x) = P((-\infty, x])$ is then a cpdf (the σ-additivity of P implies the right-continuity of F and its convergence to 0 and 1).

Key ideas are then conditional probability and independence. We explained that pairwise independence does not imply mutual independence.

A random variable on $\{\Omega, \mathcal{F}, P\}$ is a function $X : \Omega \to \Re$ such that $X^{-1}(B) \in \mathcal{F}, \forall B \in \mathcal{B}(\Re)$, equivalently, if $\{\omega : X(\omega) \le x\} \in \mathcal{F}, \forall x \in \Re$. This condition is precisely what we need to be able to define F_X and its derivative f_X. You are familiar with some standard distributions.

Expectation is the next important topic. In particular, we used a number of times, but did not prove the following result. It shows that expectation inherits the continuity of probability.

III.1. Lebesgue Convergence Theorem
(a) MCT: Assume that $0 \le X_n \uparrow X$ as $n \to \infty$. Then $E(X_n) \uparrow E(X)$ as $n \to \infty$.
(b) DCT: ssume that $X_n \to X$ as $n \to \infty$ and $|X_n| \le Y$ with $E(Y) < \infty$. Then $E(X_n) \to E(X)$ as $n \to \infty$.

23.4 Jointly Gaussian Random Variables

What makes jointly Gaussian random variables tick is that any linear combination is Gaussian.

First, one defines $X = N(\mu, \sigma^2)$ if $E(e^{sX}) = e^{s\mu + (s^2\sigma^2)/2}$.

Second, one observes that i.i.d. $N(0,1)$ random variables X_k are jointly Gaussian since

$$E(\exp\{s(a_1X_1 + \cdots + a_nX_n)\}) = \Pi_k[e^{(s^2a_k^2\sigma_k^2)/2}] == N(0, \sigma^2) \text{ with } \sigma^2 = \sum_k a_k^2\sigma_k^2.$$

Third, one has the following result that shows that this is essentially it. This result derives directly from II.2.

IV.1. Theorem: Representation of JG - (Lecture 6 - Theorem 1)
Assume $\mathbf{Y} = N(\mathbf{0}, K)$. *Then* $\mathbf{Y} =_D R\mathbf{X}$ *where* $\mathbf{X} =_D N(\mathbf{0}, \mathbf{I})$.
In particular, if $|K| \neq 0$, *then*

$$f_{\mathbf{X}}(\xi) = \frac{1}{(2\pi)^{n/2}|K|^{1/2}} \exp\{-\frac{1}{2}\xi^T K^{-1}\xi\}. \tag{23.1}$$

Fourth, we have the very useful observation, easily obtained from the joint characteristic function, that

IV.2. Theorem: Uncorrelated JG are Independent (Lecture 4 - Theorem 1)
Jointly Gaussian random variables are independent if and only if they are uncorrelated.

This result directly implies the following:

IV.3. Conditional Distribution of JG (Lecture 5 - Theorem 1) *Assume that* $(\mathbf{X}, \mathbf{Y})$ *are* $N(0, K)$ *with* $|K| \neq 0$. *Then, given* $\mathbf{Y} = \mathbf{y}$, $\mathbf{X} = N(A\mathbf{y}, \Sigma)$ *where*

$$A = K_{XY}K_Y^{-1} \text{ and } \Sigma = K_X - K_{XY}K_Y^{-1}K_{YX}. \tag{23.2}$$

23.5 Detection

Recall the basic setup. The RV X takes values in a finite set and is not observed. One observes Y and the conditional distribution of Y given X is known. The detection problem is to guess X from the observed value of Y. We have two formulations: Bayesian when we know the prior distribution of X and non-Bayesian when we do not.

An example of formulation in the Bayesian case is the *Maximum A Posteriori* estimate of X given Y defined by $MAP[X|Y = y] = \arg\max_x P[X = x|Y = y]$.

An example of the non-Baesian formulation is the Maximum Likelihood Estimator of X given Y defined by $MLE[X|Y = y] = \arg\max_x P[Y = y|X = x]$. When appropriate, use conditional densities.

Another non-Bayesian formulation is the hypothesis testing problem when $X \in \{0, 1\}$: Find $\hat{X}$ calculated from Y that maximizes $P[\hat{X} = 1 \mid X = 1]$ subject to $P[\hat{X} = 1 \mid X = 1] \leq \beta$. The solution of this problem is given by

V.1. Theorem - Neyman-Pearson (Lecture 10 - Theorem 1)
The solution to the binary hypothesis testing problem is as follows:

$$\phi(y) = \begin{cases} 1, & \text{if } \Lambda(y) := \frac{f_{Y|X}[y|1]}{f_{Y|X}[y|0]} > \lambda; \\ \gamma, & \text{if } \Lambda(y) = \lambda; \\ 0, & \text{if } \Lambda(y) < \lambda, \end{cases} \tag{23.3}$$

where $\lambda > 0$ *and* $\gamma \in [0, 1]$ *are the only values such that*

$$P[Z = 1 \mid X = 0] = \beta \text{ when } P[Z = 1 \mid Y = y] = \phi(y). \tag{23.4}$$

Quite remarkably the solutions of the MLE, MAP, or HT problems are all based on the *likelihood ratio*

$$\Lambda(y;x) := \frac{f_{Y|X}[y|x]}{f_{Y|X}[y|x_0]}.$$

Thus if $\Lambda(y;x) = g(h(y);x)$, the solution is a function of $h(y)$. In that case, we say that $h(Y)$ is a *sufficient statistic* for detecting X from Y. The interpretation is that $h(Y)$ contains all the information in Y that is useful for detecting X.

23.6 Estimation

The formulation of estimation is similar to that of detection, except that X takes values in a general set. We can define MLE and MAP as before.

We focus on MMSE and LLSE.

One useful way to understand these ideas is through the concept of projections in the space of random variables with finite second moment. In that space, the scalar product of $\mathbf{X}$ and $\mathbf{Y}$ is defined as $E(\mathbf{X}^T\mathbf{Y})$ and it corresponds to a norm $||\mathbf{X}|| = \sqrt{E(\mathbf{X}^T\mathbf{X})}$. This space is Hilbert and, consequently, the projection onto a subspace is well-defined. Moreover, a projection is characterized by the fact that the difference between the random variable and its projection is orthogonal to the subspace.

Recall from Lecture 11 that the MMSE of $\mathbf{X}$ given $\mathbf{Y}$ is the function $g(\mathbf{Y})$ that minimizes the mean squared error $E(||\mathbf{X} - g(\mathbf{Y})||^2)$. We have

VI.1. Theorem: The MMSE is the Conditional Expectation (Lecture 11 - Theorem 1)

The MMSE of $\mathbf{X}$ *given* $\mathbf{Y}$ *is* $E[\mathbf{X}|\mathbf{Y}]$.

We explained that, if $h(\mathbf{Y})$ is a sufficient statistic for $\mathbf{X}$, then $E[\mathbf{X}|\mathbf{Y}] = g(h(\mathbf{Y}))$.

We then define the LLSE of $\mathbf{X}$ given $\mathbf{Y}$ as the linear function $A\mathbf{Y} + \mathbf{b}$ that minimizes the mean squared error $E(||\mathbf{X} - (A\mathbf{Y} + \mathbf{b})||^2)$. The key result is

VI.2. Theorem: Calculating the LLSE (Lecture 11 - Theorem 2) $L[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X}) + K_{\mathbf{X},\mathbf{Y}}K_{\mathbf{Y}}^{-1}(\mathbf{Y} - E(\mathbf{Y}))$ *if* $K_{\mathbf{Y}}$ *is nonsingular.*

$L[\mathbf{X}|\mathbf{Y}] = E(\mathbf{X}) + K_{\mathbf{X},\mathbf{Y}}Q_1\Lambda_1^{-1}Q_1^T(\mathbf{Y} - E(\mathbf{Y}))$ *if* $K_{\mathbf{Y}}$ *is singular.*

It follows from the projection interpretation that $E[\cdot \mid \mathbf{Y}]$ and $L[\cdot \mid \mathbf{Y}]$ are linear operations. Of course, if $\mathbf{X}$ and $\mathbf{Y}$ are JG, then $E[\mathbf{X} \mid \mathbf{Y}] = L[\mathbf{X} \mid \mathbf{Y}]$. The converse is certainly not true.

One nice result is

VI.3. Fact: Innovation (Lecture 12 - Example A)

Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ *be three random vectors such that* $\mathbf{Y} \perp \mathbf{Z}$ *and that all the random variables are zero-mean. Then*

$$L[\mathbf{X}|\mathbf{Y},\mathbf{Z}] = K_{\mathbf{X},\mathbf{Y}}K_{\mathbf{Y}}^{-1}\mathbf{Y} + K_{\mathbf{X},\mathbf{Z}}K_{\mathbf{Z}}^{-1}\mathbf{Z} = L[\mathbf{X}|\mathbf{Y}] + L[\mathbf{X}|\mathbf{Z}]. \tag{23.5}$$

Moreover,

$$cov(\mathbf{X} - L[\mathbf{X}|\mathbf{Y},\mathbf{Z}]) = cov(\mathbf{X} - L[\mathbf{X}|\mathbf{Y}]) - K_{\mathbf{X},\mathbf{Z}}K_{\mathbf{Z}}^{-1}K_{\mathbf{Z},\mathbf{X}}. \tag{23.6}$$

We also explained the relationship between linear regression and LLSE.

23.7 Kalman Filter

Consider a system whose state evolves according to linear dynamics with uncorrelated noise and assume one makes linear observations in the state, also with uncorrelated noise. The Kalman filter computes recursively the LLSE of the state given the observations to date. (Of course, in the JG case, the filter computes the conditional expectation and we know the conditional distribution.)

The key result is

VII.1. Theorem: Kalman Filter (Lectures 13-14 - Theorem 1)

Assume

$$X_{n+1} = AX_n + V_n \text{ and } Y_n = CX_n + W_n, n \geq 1$$

where X_1, V_n, W_n are all orthogonal and zero-mean, $cov(V_n) = K_V$, and $cov(W_n) = K_W$.

Then

$$\hat{X}_n = L[X_n|Y_1, \ldots, Y_n] = L[X_n|Y^n]$$

is obtained by the Kalman Filter:

$$\hat{X}_n = A\hat{X}_{n-1} + R_n(Y_n - CA\hat{X}_{n-1})$$

where

$$\begin{aligned} R_n &= S_n C^T [CS_n C^T + K_W]^{-1}, \\ S_n &= A\Sigma_n A^T + K_V, \\ \Sigma_{n+1} &= (I - R_n C)S_n. \end{aligned}$$

Thus,

$$S_{n+1} = K_V + AS_n A^T - AS_n C^T [CS_n C^T + K_W]^{-1} CS_n A^T.$$

Moreover, the matrices S_n and Σ_n have the following significance:

$$\begin{aligned} S_n &= cov(X_n - A\hat{X}_{n-1}), \\ \Sigma_n &= cov(X_n - \hat{X}_n). \end{aligned}$$

One interesting question is what happens to the estimate as time goes to infinity. One important result is

VII.2. Theorem: Filter Asymptotics (Lecture 13-14 - Theorem 3)

Let Q be a square root of K_V. Suppose that the following conditions hold:

$$(A, Q) \text{ is reachable, } (A, C) \text{ is observable, and } K_W > 0. \tag{23.7}$$

Then, if $S_0 = 0$,

$$\Sigma_n \to \Sigma, S_n \to S, R_n \to R, \text{ as } n \to \infty.$$

The limiting matrices are the only solutions of the equations

$$\begin{aligned} R &= SC^T [CSC^T + K_W]^{-1}, \\ S &= A\Sigma A^T + K_V, \\ \Sigma &= (I - RC)S. \end{aligned}$$

Equivalently, S is the unique positive semidefinite solution of

$$S = A(S - SC^T[CSC^T + K_W]^{-1}CS)A^T + K_V. \tag{23.8}$$

Moreover, the time-invariant filter

$$Z_n = AZ_{n-1} + R(Y_n - CAZ_{n-1})$$

is such that

$$cov(X_n - Z_n) \to \Sigma, \text{ as } n \to \infty.$$

We don't expect you to know the proof of this result. However, you should understand why the error covariance is bounded when the system is observable, even if the system itself is unstable. You should also realize that this result is possible because the assumptions guarantee that $A - RCA$ is stable. The convergence itself is proved through a connection with an optimal control problem that we analyze using dynamic programming and we did not have really enough time to explore these ideas.

23.8 Wiener Filter

Whereas Kalman consider processes defined by dynamic equations, Wiener explore wide sense stationary processes.

We have

VIII.1. Definition - WSS (Lecture 15 - Definitions 3 and 7)

$\{\mathbf{X}_n, \mathbf{Y}_n, n \in \mathbb{Z}\}$ *are* wide sense stationary *(wss) if*

$$E(\mathbf{X}_n) = \mu_{\mathbf{X}}, E(\mathbf{Y}_n) = \mu_{\mathbf{Y}}, E(\mathbf{X}_{n+m}\mathbf{X}_n^*) = R_{\mathbf{X}}(m), E(\mathbf{Y}_{n+m}\mathbf{Y}_n^*) = R_{\mathbf{Y}}(m), E(\mathbf{X}_{n+m}\mathbf{Y}_n^*) = R_{\mathbf{XY}}(m), \forall n, m \in \mathbb{Z}.$$

Then the spectral and cross-spectral densities *are defined as follows:*

$$S_{\mathbf{X}}(f) := \sum_{m=-\infty}^{\infty} R_{\mathbf{X}}(m)e^{-j2\pi mf} \text{ and } S_{\mathbf{XY}}(f) := \sum_{m=-\infty}^{\infty} R_{\mathbf{XY}}(m)e^{-j2\pi mf}, f \in \mathbb{R}.$$

The following result explains how linear time-invariant systems modify the spectral density.

VIII.2. Theorem - Effect of LTI System on WSS Processes (Lecture 15 - Theorem 1)

$$\mathbf{V}_n = \sum_{m=-\infty}^{\infty} \mathbf{h}_1(n-m)\mathbf{X}_m \text{ and } \mathbf{W}_n = \sum_{m=-\infty}^{\infty} \mathbf{h}_2(n-m)\mathbf{Y}_m,$$

where $\mathbf{h}_1(n)$ *and* $\mathbf{h}_2(n)$ *are summable. Then*

$$S_{\mathbf{VW}}(f) = \mathbf{H}_1(f)S_{\mathbf{XY}}(f)\mathbf{H}_2^*(f)$$

where $\mathbf{H}_i(f)$ *is the* transfer function *that corresponds to* $\mathbf{h}_i(n)$*, for* $i = 1, 2$*. That is,*

$$\mathbf{H}_i(f) = \sum_{n=-\infty}^{\infty} h_i(n)e^{-j2\pi nf}, f \in \mathbb{R}.$$

With these definitions, we can construct the Wiener filter as follows. We have processes $\mathbf{X}$ and $\mathbf{Y}$ that are WSS. The Wiener filter is a causal LTI system with input $\mathbf{Y}$ and whose output at time n is $\hat{X}_n = L[X_n | Y_m, m \leq n]$.

VIII.3. Theorem: Wiener Filter (Lecture 15, Section VI.)

Let $\mathbf{X}, \mathbf{Y}$ be WSS. Assume that $S_Y(f) = |H(f)|^2$ where $H(f) = N(f)/D(f)$ with N and D are polynomials in $z^{-1} = e^{-i2\pi f}$ with roots z inside the unit circle. The Wiener filter has transfer function

$$[S_{XY}(f) S_Y^{-1} H(f)]_+ H^{-1}(f).$$

In this expression, $[K(f)]_+$ designates the causal part of an LTI system with transfer function $K(f)$, i.e., an LTI system with impulse response $k_+(n) = k(n) 1\{n \geq 0\}$.

The key idea is that the filter H^{-1} is the *whitening filter*. When its input is $\mathbf{Y}$, its output is a white process W_n. It is easy to see that the filter $S_{XY} S_Y^{-1}$ with input Y has output $\tilde{X}_n = L[X_n \mid Y_m, m \in \mathbb{Z}]$. It follows that $\tilde{X}_n$ is the output of a filter $K(f) = S_{XY}(f) S_Y^{-1} H(f)$ when its input is W_n, so that the output of K_+ is indeed $\hat{X}_n$.

Chapter 24

Review: Part 2 - Markov Chains, Poisson Process, and Renewal Process

24.9 Overview

We review the second part of the course. As with the review of the first part, use these notes as a guide to your review. However, do not assume that what is not in the review is not part of the final.

24.10 Markov Chains: Discrete Time

A DTMC is a sequence of random variables that take value in a discrete set and are such that the evolution starts afresh from its current value, at any given time. This is the Markov property.

The basic definition is

X.1. Definition: Narkov Chain (Lecture 16)
Let $\mathcal{X}$ be a countable set and $P = \{P(i,j), i.j \in \mathcal{X}\}$ a matrix of nonnegative numbers whose rows add up to one. Let also $\pi_0 = \{\pi_0(i), i \in \mathcal{X}\}$ be a collection of nonnegative numbers that add up to one. The sequence of random variables $\mathbf{X} = \{X_n, n \geq 0\}$ is a Markov chain *on $\mathcal{X}$ with* transition probability matrix *P and* initial distribution *π_0 if*

$$P[X_{n+1} = j | X_n = i, X_m, m = 0, 1, \ldots, n-1] = P(i,j), i, j \in \mathcal{X}, n \geq 0.$$

It may help to realize the following:

X.2. Fact: Dynamic Representation (Homework)
We can represent a Markov chain $\mathbf{X}$ as

$$X_{n+1} = f(X_n, V_n), n \geq 0$$

where $\{V_n, n \geq 0\}$ are i.i.d. and independent of X_0.

In fact, one can choose $V_n =_D U[0,1]$, but this is not saying much.

A deeper observation is that a Markov chain $\mathbf{X}$ starts afresh from its value at some random times called *stopping times*. Generally, a stopping time is a random time τ that is non-anticipative. That is, we can tell that $\tau \leq n$ from $\{X_0, X_1, \ldots, X_n\}$, for any $n \geq 0$. A simple example is the first hitting time T_A of a set $A \subset \mathcal{X}$. Another simple example is T_A+5. A simple counterexample is $T_A - 1$. This property is the *strong Markov property*. It is that property that enables us to say that the successive hitting times of a given state form a renewal process. We have

X.3. Theorem: Strong Markov Property (Homework)

Let τ be a stopping time of $\mathbf{X}$. Then

$$P[X_{\tau+1} = j \mid X_\tau = i; X_m, m \leq \tau] = P(i,j), \forall i,j \in \mathcal{X}.$$

Moreover, the distribution of $\{X_{\tau+n}, n \geq 0\}$ given $\{X_\tau = i; X_m, m \leq \tau\}$ is the same as the distribution of $\mathbf{X}$ given that $X_0 = i$.

Proof:

Although you worked this out in a homework, let us give a more general and probably more compact derivation.

Let $V = 1\{X_{\tau+1} = j\}, Y = 1\{X_\tau = i\}, Z = \{X_m, m \leq \tau\}$. Then we know that

$$E[V \mid Y = 1, Z] = E[E[V \mid Y = 1, Z, \tau] \mid Y = 1, Z]].$$

However, $E[V \mid Y = 1, Z, \tau = n] = P[X_{n+1} = j \mid X_n = i, X_m, m \leq n; \tau = n] = P[X_{n+1} = j \mid X_n = i, X_m, m \leq n] = P(i,j)$. The next to last equality comes from the fact that $\{\tau = n\}$ is determined by $\{X_m, m \leq n\}$ and is consequently independent of X_{n+1} given X_n. This is precisely where we need τ to be a stopping time. Hence, it follows that $E[V \mid Y = 1, Z] = E(P(i,j)) = P(i,j)$.

□

We then defined a few concepts.

X.4. Definition: Irreducible, etc ((Definition 1 - Lecture 17)

The Markov chain $\mathbf{X}$ *is irreducible if ..., aperiodic, transient, null recurrent, positive recurrent, ...*

A distribution π *is* invariant *if it satisfies the* balance equations $\pi P = \pi$. *The Markov chain is stationary if and only if* π_0 *is invariant.*

We then get to the key result:

X.4. Theorem: Classification (Theorem 1 - Lecture 18)

(a) If a Markov chain is irreducible, then all the states have the same recurrence properties (T, NR, or PR). The MC is then said to be T, NR, or PR, correspondingly. All the states also have the same period...

(b) If a MC is irreducible, then it is T if and only if $P[X_n = i, \ i.o. \ |X_0 = j] = 0, \forall i, j$.

(c) If a MC is irreducible, then it is R if and only if $P[X_n = i, \ i.o. \ |X_0 = j] = 1, \forall i, j$.

(d) If an irreducible MC is T or NR, then

$$\frac{1}{N}\sum_{n=1}^{N} 1\{X_n = i\} \xrightarrow{a.s.} 0, \forall i \in \mathcal{X}.$$

Moreover, there is no invariant distribution. A finite irreducible MC cannot be T or NR.

(e) If an irreducible MC is PR, then

$$\frac{1}{N}\sum_{n=1}^{N} 1\{X_n = i\} \xrightarrow{a.s.} \pi(i) > 0, \forall i \in \mathcal{X}.$$

Moreover, π *is the unique invariant distribution. A finite irreducible MC is necessarily PR.*

(f) If the MC is irreducible, PR, and aperiodic, then

$$P(X_n = i) \to \pi(i), \forall i \in \mathcal{X}.$$

You understand these results and you know how to derive them. For (d)-(e) we only needed the strong law of large numbers applied to the i.i.d. inter-visit times to state i.

One cute argument to prove (f) uses coupling. We comment on this idea because it seems to confuse a few of you. By coupling we mean the construction of two random variables X, Y or two random processes $\mathbf{X}, \mathbf{Y}$ that have desired marginal distributions and have a particular joint distribution that enables to prove a relationship between the marginal distributions.

As an example, assume that $X =_D Exp(\lambda)$ and $Y =_D Exp(\mu)$ with $\lambda > \mu$. We want to show that if $f(.)$ is a nondecreasing function, then $E(f(X)) \leq E(f(Y))$. The coupling proof uses that fact that we can assume that $Y = (\lambda/\mu)X$ since this does not affect the marginal distributions of X and Y and since the inequality we want to prove depends only on the marginal distributions, not on the joint distribution. One then notices that $Y \geq X$, so that $f(Y) \geq f(X)$ and $E(f(Y)) \geq E(f(X))$. A proof based on the distributions would go as follows. We have $E(f(X)) = \int_0^\infty f(x)\lambda e^{-\lambda x} dx = f(0) + \int_0^\infty e^{-\lambda x} df(x) \leq f(0) + \int_0^\infty e^{-\mu x} df(x) = E(f(Y))$. Note that in the coupling proof, one finds an almost sure relationship between the random variables. That is, all the realizations have that relationship. In the distribution proof, one calculates.

The result (f) can be proved by coupling; so can the corresponding result for continuous-time Markov chains. We also indicated that a proof of Blackwell's renewal theorem uses coupling.

We examined a number of examples of Markov chains, such as a random walk and a reflected random walk. We also discussed the Viterbi algorithm for estimating a hidden Markov chain.

In discussions, we explained a criterion for positive recurrence.

X.5. Theorem: Criterion for Positive Recurrence (Discussion)

Let $\mathbf{X}$ *be an irreducible Markov chain on* $\mathcal{X}$*. Assume there is a function* $V : \mathcal{X} \to [0, \infty)$ *with the following properties:*

(a) $E[V(X_{n+1}) - V(X_n) \mid X_n = i] < \infty, \forall i \in \mathcal{X}$*;*

(b) $E[V(X_{n+1}) - V(X_n) \mid X_n = i] \leq -\delta < 0, \forall i \notin A$ *where* A *is a finite subset of* $\mathcal{X}$*.*

Then $\mathbf{X}$ *is positive recurrent.*

This result can be understood as follows. Since V decreases outside of A and since V is nonnegative, $\mathbf{X}$ must spend a positive fraction of time inside A. By Theorem X.4, this implies that $\mathbf{X}$ is positive recurrent.

24.11 Poisson Process

A Poisson process with rate $\lambda > 0$ is a point process with the property that a jump occurs in $[t, t+\epsilon]$ with probability $\lambda\epsilon + o(\epsilon)$, independently of the jump times less than t. Consequently, the Poisson point process is memoryless. In particular, this implies that the times between jumps must be independent and exponentially distributed with rate λ. If we define the Poisson process to have a jump at time 0, then it is not stationary. Indeed, in that case the probability that the process has a jump in $(-\epsilon, +\epsilon)$ is one, which is not the same as the probability that it has a jump in $(t-\epsilon, t+\epsilon)$ for $t \neq 0$. One the other hand, if the process has not jump at time 0, then it is stationary. This is related to the discussion of stationary renewal processes that we review later.

The key result is

XI.1. Theorem: Properties of Poisson Process (Theorem 1 - Lecture 20)

Let $\{N_t, t \geq 0\}$ *be a* Poisson process with rate λ*.*

(a) For any $s > 0$*,* $\{N_{t+s} - N_s, t \geq 0\}$ *is a Poisson process with rate* λ *and is independent of* $\{N_t, t \leq s\}$*.*

(b) For any $n \geq 2$ *and* $0 < t_1 < t_2 < \cdots < t_n$*, the random variables* $N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$ *are mutually independent and Poisson distributed with respective means* $\lambda t_1, \lambda(t_2 - t_1), \ldots, \lambda(t_n - t_{n-1})$*.*

(c) Given that $N_t = n$*, the jump times* $\{T_1, \ldots, T_n\}$ *are the ordered values of* n *i.i.d.* $U[0, t]$ *random variables.*

(d) Color each jump time T_n *independently red with probability* α *and blue with probability* $1 - \alpha$ *where* $\alpha \in (0, 1)$*. For* $t \geq 0$*, let* A_t *be the number of red jumps and* B_t *the number of blue jumps in* $[0, t]$*. Then* $\{A_t, t \geq 0\}$ *and* $\{B_t, t \geq 0\}$ *are independent Poisson processes with respective rates* $\alpha\lambda$ *and* $(1 - \alpha)\lambda$*.*

We then generalized the Poisson process to a Poisson measure in $\Re^d$. This is a distribution of isolated points with the property that the number of points in disjoint sets are mutually independent and have Poisson distribution in each set A with mean $\lambda(A)$. You see that a standard Poisson process is a Poisson measure on $\Re$ with $\lambda([t, t+s]) = \lambda s$.

We applied the notion of Poisson measure to the analysis of the M/G/∞ queue with Poisson arrivals with rate λ and i.i.d. service time distributed like S_1. There are infinitely many servers,

so that the departures are equal to the arrival times delayed by i.i.d. service times. The result is

XI.2. Fact: Property of M/G/∞ Queue (Fact 3 - Lectures 20)

(a) The departure process is Poisson with rate λ.

(b) The number of customers in the queue at time t is independent of the departure times up to time t and is Poisson with mean $\lambda E(S_1)$.

24.12 Continuous Time Markov Chains

A continuous time Markov chain is essentially a discrete time Markov chain that specifies the successive values, modified so that the process stays in state i for a random holding time that is exponentially distributed with a rate that depends on the state. Given the sequence of values of the Markov chain, all the holding times are independent.

From the memoryless property of the exponential distribution, we find that this process has the Markov property. We observed that the construction runs into a problem if jumps can accumulate (explosion). For simplicity, we assume that they cannot and we call such Markov chains *regular*. One key result is the following.

XII.1. Theorem: Invariant Distribution and Stationarity (Theorems 1-2 - Lecture 23)

Let X be a regular Markov chain (meaning, no explosions are possible) with rate matrix Q. For $t \geq 0$, let $\pi_t = \{\pi_t(i), i \in \mathcal{X}\}$ where $\pi_t(i) = P(X_t = i)$. We consider π_t as a row vector. Then

$$\frac{d}{dt}\pi_t = \pi_t Q.$$

Consequently, $\pi_t = \pi$ for all $t \geq 0$ if and only if $\pi_0 = \pi$ and π is invariant.

Let X be a regular Markov chain with rate matrix Q and initial distribution π. It is stationary if and only if π is invariant.

The classification result corresponds to Theorem X.4.

XII.2. Theorem: Classification of Markov Chains (Theorem 3 - Lecture 23)

Let $X = \{X_t, t \geq 0\}$ be an irreducible Markov chain on $\mathcal{X}$.

(a) The states are either all transient, all null recurrent, or all positive recurrent. We then say that the Markov chain is ....

(b) If X is transient or null recurrent, then

$$\frac{1}{T}\int_0^T 1\{X_t = i\}dt \to 0, \text{ as } T \to \infty, \forall i \in \mathcal{X}, a.s.$$

Moreover, there is no invariant distribution and

$$P(X_t = i) \to 0, \forall i \in \mathcal{X}.$$

(c) If X is positive recurrent, then

$$\frac{1}{T}\int_0^T 1\{X_t = i\}dt \to \pi(i) > 0, \text{ as } T \to \infty, \forall i \in \mathcal{X}, a.s.$$

Moreover, π is the unique invariant distribution and

$$P(X_t = i) \to \pi(i), \forall i \in \mathcal{X}.$$

24.13 Jackson Networks

A Jackson network is a collection of J queues. Customers arrive as independent Poisson processes, with rate γ_i into queue i. Customers face independent service times in all the queues and these are exponentially distributed with rate μ_i in queue i. When a customer leaves queue i, he goes to queue j with probability $R(i,j)$ and leaves the network with probability $1 - \sum_{j=1}^{J} R(i,j)$. The basic result is

XIII.1. Theorem: Product-Form (Theorem 1 - Lecture 24)

Assume that each customer can eventually leave.

(a) In that case, the equations

$$\lambda_i = \gamma_i + \sum_{j=1}^{J} \lambda_j R(j,i), i = 1, \ldots, J$$

have a unique solution $(\lambda_1, \ldots, \lambda_J)$.

(b) Moreover, if $\rho_i := \lambda_i/\mu_i < 1$ *for* $i = 1, \ldots, J$, *then the vector of queue lengths* $X_t = (X_t^1, \ldots, X_t^J)$ *is a positive recurrent Markov chain with invariant distribution*

$$\pi(n_1, \ldots, n_J) = \pi_1(n_1) \cdots \pi_J(n_J) \text{ where } \pi_i(n) = (1-\rho_i)\rho_i^n, n \geq 0.$$

The key idea of the proof is to guess that the process reversed in time corresponds to the Jackson network with the same queues but where the flows of customers are reversed. One uses the following lemma.

XIII.2. Lemma: Kelly (Lemma 1 - Lecture 24)

Assume that Q *and* Q' *are two rate matrices and* π *a distribution on* $\mathcal{X}$ *such that*

$$\pi(i)q(i,j) = \pi(j)q'(j,i), \forall i, j \in \mathcal{X}.$$

Then $\pi Q = 0$ *and* Q' *is the rate matrix of the Markov chain reversed in time.*

PASTA is a useful result that derives directly from the independence of the increments of a Poisson process.

XIII.3. Theorem: PASTA (Theorem 2 - Lecture 24)

Assume customers arrive as a Poisson process into a queuing system described by a stationary Markov chain X_t. *Then, an arriving customer sees the state with its invariant distribution* π.

We applied this result to show

XIII.4. Theorem: Delay through stationary M/M/1 queue (Theorem 3 - Lecture 24)

The delay of a customer through a stationary M/M/1 queue with arrival rate λ *and service rate* μ *is exponentially distributed with rate* $\mu - \lambda$.

24.14 Convergence

We observed that there are different forms of convergence and we explored the relationships between these forms. You should know the definitions of these forms of convergence. You should remember the following results: $a.s. \to P \to D$ and $L_1 \to P$. You should be able to show that the reverse implications are not true.

The following result is very useful. We don't expect you to know the proof.

XIV.1. Theorem: Strong Law of Large Numbers (Theorem 3 - Lecture 25)
Let X_n be i.i.d. random variables with $E(|X_n|) < \infty$. Then

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{a.s.} E(X_1) \text{ as } n \to \infty.$$

24.15 Renewal Processes

We start with

XV.1. Definition: Renewal Process (Definition 1 - Lecture 26)
Let $\{T_n, n \geq 1\}$ be such that $T_n < T_{n+1}$ and $\{T_{n+1} - T_n, n \geq 1\}$ are i.i.d.. If T_1 has the same distribution as $T_2 - T_1$, this process is called a renewal process*; otherwise, it is a —em delayed renewal process. We also define $N_t = \max\{n \geq 1 \mid T_n \leq t\}$ with $N_t = 0$ for $t < T_1$ for $t \geq 0$. See Figure 22.1.*

For simplicity, we assume throughout this section that $\lambda^{-1} := E(T_2 - T_1) < \infty$.

We saw two nice results. The first theorem tells us how to choose the first jump time to make the process stationary. We explained that this corresponds to picking the origin randomly along the real line among jumps separate by i.i.d. times.

XV.2. Theorem: Stationary Renewal Process (Theorem 1 - Lecture 26)
Let $\{N_t, t \geq 0\}$ be a delayed renewal process and $F(t) = P(T_2 - T_1 \leq t)$ and $G(t) = P(T_1 \leq t)$ for $t \geq 0$. The process $\{N_t, t \geq 0\}$ is stationary if and only if

$$G(t) = \lambda \int_0^t (1 - F(s))ds, t \geq 0 \text{ where } \lambda^{-1} = \int_0^\infty s dF(s) = E(T_2 - T_1).$$

In particular, if the process is stationary, the average time until the next jump, $E(T_1)$ is given by

$$E(T_1) = \frac{\lambda}{2} E((T_2 - T_1)^2).$$

The second theorem says that the process becomes stationary if the distribution is not concentrated on multiples of some fixed quantity.

XV.3. Theorem: Blackwell's Renewal Theorem (Theorem 2 - Lecture 26)
Assume that F is non-lattice. Then, for all $L > 0$,

$$E(N_{t+L} - N_t) \to \lambda L \text{ as } t \to \infty.$$

Appendix A

Notes on Probability

A.1 Introduction

Probability Theory models uncertainty and develops methods to make decision in the face of such uncertainty. We briefly review a few concepts that you learned in elementary probability and we add some important results that you may not be familiar with yet.

- Probability Space
- Random Variables
- Expectation

A.2 Probability Space

A *probability space* is a triplet $\{\Omega, \mathcal{F}, P\}$ where Ω is a nonempty set, $\mathcal{F}$ a σ-field of Ω, and P probability measure on $\mathcal{F}$. That is, $\mathcal{F}$ is a collection of subsets of Ω that is closed under countable set operations and $P : \mathcal{F} \to [0, 1]$ is a σ-additive function such that $P(\Omega) = 1$. The elements of $\mathcal{F}$ are called *events*.

The σ-additivity of P means that if $A = \cup_{n=1}^{\infty} A_n$ where $A_n \in \mathcal{F}$ for $n \geq 1$ are pairwise disjoint, then $P(A) = \sum_{n=1}^{\infty} P(A_n)$. Equivalently, assume that $A_n \downarrow A$ in $\mathcal{F}$. That is, $A_n \in \mathcal{F}$ with $A_{n+1} \subset A_n$ for $n \geq 1$ and $A = \cap_{n=1}^{\infty} A_n$. Then $P(A_n) \downarrow P(A)$, i.e., $P(A_{n+1}) \leq P(A_n)$ for $n \geq 1$ and $P(A) = \lim_{n\to\infty} P(A_n)$.

In particular, the following result holds.

Lemma A.1. *Borel-Cantelli*
Assume that events A_n are such that $\sum_n P(A_n) < \infty$, then $P(A_n,\ i.o.) = 0$.

Proof:

The notation i.o. means "infinitely often." That is, $\{A_n,\ \text{i.o. }\}$ is the collection of ω that belong to infinitely many events A_n.

One observes that *omega* belongs to infinitely many A_n is and only if for each $n \geq 1$ there is some $m \geq n$ such that ω belongs to A_m. That is,

$$\{A_n,\ \text{i.o.}\} = \cap_{n\geq 1} \cup_{m\geq n} A_m,\ \text{so that }\ \cup_{m\geq n} A_m \downarrow \{A_n,\ \text{i.o.}\}.$$

This implies that $P(\cup_{m\geq n} A_m) \downarrow P(A_n,\ \text{i.o.})$. But $P(\cup_{m\geq n} A_m) \leq \sum_{m\geq n} P(A_m)$, so that $P(\cup_{m\geq n} A_m) \downarrow 0$.

□

Here are a few commonly used probability spaces.

Example A.1. *Finite Space*

Let Ω be a countable set, $\mathcal{F} = 2^\Omega$ be the collection of all the subsets of Ω, and $\{p(\omega), \omega \in \Omega\}$ be a collection of nonnegative numbers that add up to one. One then defines $P(A) = \sum_{\omega \in A} p(\omega), A \subset \Omega$. The $\{\Omega, \mathcal{F}, P\}$ is a probability space.

The only nontrivial part of this result is that P is σ additive. To show it, assume that $A_n \subset \Omega$ with $A_{n+1} \subset A_n$ for $n \geq 1$ and let $A = \cap_{n=1}^\infty A_n$. To show that $P(A_n) \to P(A)$, we argue by contradiction. Since $P(A_n) \downarrow$, let $\alpha = \lim P(A_n)$ and assume that $\alpha > P(A)$. That implies that there must be at least some $\omega \in A_n \setminus A$ for all $n \geq 1$, a contradiction.

Example A.2. *Real Line*

Let $\mathcal{B}(\Re)$ be the smallest σ-field that contains the intervals (a, b) for all $a < b$ in $\Re$. The σ-field $\mathcal{B}(\Re)$ is called the Borel σ-field of the real line. This σ-filed exists because an arbitrary intersection of σ-fields is a σ-field. Also, $2^\Re$ is a σ-field that contains all the intervals. Thus, $\mathcal{B}(\Re)$ is the intersection of all the σ-fields that contain all the intervals.

Similarly, for $a < b \in \Re$, one defines $\mathcal{B}([a, b])$ to be the smallest σ-field of $[a, b]$ that contains all the intervals of $[a, b]$.

Recall that a set G of real numbers is open if every point $x \in G$ is surrounded by points of G. That is, for every $x \in G$, there is some $\epsilon > 0$ such that $(x - \epsilon, x + \epsilon) \subset G$. It follows that every $x \in G$ is such that $x \in (r(x) - epsilon(x), r(x) + \epsilon(x)) \subset G$ where both $r(x)$ and $\epsilon(x)$ are rational numbers. Consequently, we can write $G = \cup_{x \in G}(r(x) - epsilon(x), r(x) + \epsilon(x))$. Moreover, this union is in fact countable since all the $r(x)$ and $\epsilon(x)$ are rational and the set of rational numbers is countable. Thus, G is a countable union of intervals and this implies $G \in \mathcal{B}(\Re)$. We have shown that $\mathcal{B}(\Re)$ contains all the open sets of $\Re$.

By definition, a set is closed if its complement is open. Consequently, $\mathcal{B}(\Re)$ contains all the closed sets of $\Re$.

Example A.3. *Uniform Distribution on $[0, 1]$*

Let $\Omega - [0, 1], \mathcal{F} = \mathcal{B}([0, 1])$ and define P such that $P((a, b)) = b - a$ for $0 \leq a \leq b \leq 1$. One can show that $P(.)$ extends uniquely to a probability measure on $\mathcal{B}([0, 1])$. The resulting probability space $\{\Omega, \mathcal{F}, P\}$ describes the random experiment "choosing a point uniformly in $[0, 1]$."

Example A.4. *Uniform Distribution on $[0, 1]^2$*

The previous definitions extend to multiple dimensions. One defines $\mathcal{B}([0, 1]^2)$ as the smallest σ-filed that contains all the rectangles of the form $[a, b] \times [c, d] \subset [0, 1]^2$. The uniform distribution on $[0, 1]^2$ is the unique probability measure on $\mathcal{B}([0, 1]^2)$ such that $P([a, b] \times [c, d]) = (b - a) \times (d - c)$ for all $[a, b] \times [c, d] \subset [0, 1]^2$.

A.3 Independence

Let $\{\Omega, \mathcal{F}, P\}$ be a probability space. You recall the following definitions.

Definition A.1. *Conditional Probability*

The conditional probability of A given B is designated by $P[A|B]$ and is defined by

$$P[A|B] = \frac{P(A \cap B)}{P(B)}, \text{ if } P(B) \neq 0.$$

It follows that if $\{B_1, \ldots, B_n\} \subset \mathcal{F}$ are pairwise disjoint and such that $\cup_{m=1}^n B_m = \Omega$, then

$$P[B_m|A] = \frac{P(B_m \cap A)}{P(A)} = \frac{P[A|B_m]P(B_m)}{\sum_{k=1}^n P[A|B_k]P(B_k)}. \tag{A.1}$$

This expression is known as Bayes' Rule.

Definition A.2. *Independence*

(a) $A, B \in \mathcal{B}$ are independent *if $P(A \cap B) = P(A)P(B)$.*

(b) The events $\{A_i, i \in I\} \subset \mathcal{F}$ are pairwise independent *if A_i and A_j are independent for all $i \neq j \in I$.*

(c) The events $\{A_i, i \in I\} \subset \mathcal{F}$ are mutually independent *if, for any finite $J \subset I$, one has $P(\cap_{j \in J} A_j) = \Pi_{j \in J} P(A_j)$.*

It is important not to confuse pairwise independence and mutual independence.

A.4 Random Variables

You recall that a random variable X on a probability space $\{\Omega, \mathcal{F}, P\}$ is a function $X : \Omega \to \Re$ with the property that $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\Re)$. Here, $X^{-1}(B) := \{\omega \in \Omega \mid X(\omega) \in B\}$ is the inverse image of B under X. The point of this definition is that one can then define $P(X \in B)$ for $B \in \mathcal{B}$ as $P(X^{-1}(B))$. In particular, one defines

$$F_X(x) := P(X \leq x), x \in \Re$$

and one calls F_X the cumulative probability distribution function (c..p.d.f.) of X. Observe that $F(b) - F(a) = P(a < X \leq b)$ for $a < b \in \Re$. Note that F_X is right-continuous, that it tends to 1 at $+\infty$ and to 0 at $-\infty$. These properties follow directly from the σadditivity of P. For instance, one sees that if $x_n \downarrow x$, then $X^{-1}((-\infty, x_n]) \downarrow X^{-1}((-\infty, x])$, so that $F_X(x_n) \downarrow F_X(x)$. Also, $P(X = a) = \lim_{x \uparrow a} P(x < X \leq a) = \lim_{x \uparrow a}(F_X(a) - F_X(x)) = F_X(a) - F_X(a-)$. That is, the size of the jump (discontinuity) of F_X at a is the probability that X takes the value a.

It may happen that F_X has a derivative f_X, so that

$$F_X(x) = \int_{-\infty}^x f_X(y)dy.$$

In such a case, f_X is called the *probability density function* (p.d.f.) of X. The interpretation of f_X is that $f_X(x)dx = P(X \in (x, x + dx))$.

Consider the following simple example.

Example A.5. *Let $\Omega = [0, 1]$ and $\mathcal{F} = \{[0, 1], [0, 0.5].(0.5, 1], \emptyset\}$ where $\emptyset$ designates the empty set. Note that $\mathcal{F}$ is closed under countable set operations, so that it is a σ-field. Define $P([0, 0.5]) = 0.3, P((0.5, 1]) = 0.7, P(\emptyset) = 0$, and $P([0, 1[) = 1$. Then $\{\Omega, \mathcal{F}, P\}$ is a probability space. Consider the function $X : \Omega \to \Re$ defined by $X(\omega) = \omega$. Then we claim that X is not a random variable on $\{\Omega, \mathcal{F}, P\}$. Indeed, $X^{-1}([0, 0.2]) = [0, 0.2] \notin \mathcal{F}$. In particular, $P(X \leq 0.2)$ is not defined.*

Fact A.1. *Function of Random Variable*

Let X be a random variable on $\{\Omega, \mathcal{F}, P\}$ and $g : \Re \to \Re$ be some function such that $g^{-1}(B) \in \mathcal{B}(\Re)$ for all $B \in \mathcal{B}(\Re)$. A function g with that property is called a Borel function. *Then $Y := g(X)$ is a random variable on $\{\Omega, \mathcal{F}, P\}$.*

Proof:

One must show that $Y^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\Re)$. But $Y^{-1}(B) = X^{-1}(A)$ where $A = g^{-1}(B)$. The assumption on g implies that $A \in \mathcal{B}(\Re)$. The fact that X is a random variable implies that $X^{-1}(A) \in \mathcal{B}(\Re)$.

□

The same ideas extend to multiple random variables. Thus, if $\mathbf{X} = (X_1, \ldots, X_n)$ where each X_m is a random variable on the same probability space $\{\Omega, \mathcal{F}, P\}$, then one defines $F_{\mathbf{X}}(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$. Note that the set $\{X_1 \leq x_1, \ldots, X_n \leq x_n\} = \cap_{m=1}^{n} X_m^{-1}((-\infty, x_m])$ is in $\mathcal{F}$ since each X_m is a random variable. The function $F_{\mathbf{X}}$ is the joint cumulative probability distribution function (j.p.d.f.) of the random variables. If it has a density $f_{\mathbf{X}}$, we call it the joint probability density function (j.p.d.f.) of the random variables. You know that the j.c.p.d.f. contains much more information than the marginal c.p.d.f. F_{X_m}.

You should review the standard distributions: Bernoulli, binomial, Poisson, geometric, uniform, and exponential.

We recall the following

Definition A.3. *Independent Random Variables*

(a) Two random variables X and Y on $\{\Omega, \mathcal{F}, P\}$ are independent *if*

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B). \forall A, B \in \mathcal{B}(\Re).$$

(b) The random variables $\{X_i, i \in I\}$ on $\{\Omega, \mathcal{F}, P\}$ are pairwise independent *if X_i and X_j are independent for all $i \neq j \in I$.*

(c) The random variables $\{X_i, i \in I\}$ on $\{\Omega, \mathcal{F}, P\}$ are mutually independent *if, for any finite subset $J \subset I$, one has*

$$P(X_j \in A_j, j \in J) = \Pi_{j \in J} P(X_j \in A_j), \forall A_j \in \mathcal{B}(\Re).$$

We use the following observation repeatedly.

Theorem A.1. *Let $\{X_i, i \in I\}$ be mutually independent random variables on $\{\Omega, \mathcal{F}, P\}$. Let J_1 and J_2 be two finite subsets of I and $g : \Re^{d_1} \to \Re$ and $h : \Re^{d_2} \to \Re$ be Borel functions where $d_1 = |J_1|$ and $d_2 = |J_2|$. Then*

$$g(X_j, j \in J_1) \text{ and } h(X_j, j \in J_2)$$

are independent random variables.

The conclusion is generally not valid if the random variables are only pairwise independent.

A.5 Expectation

One defines the *expected value* $E(X)$ of a random variable X as

$$E(X) = \int_{-\infty}^{\infty} x dF_X(x),$$

provided that the integral does not yield $\infty - \infty$. If f_X exists, then

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx.$$

One can be a bit more precise here. Assume that X takes only finitely many values, so that we can write $X = \sum_{m=1}^{n} a_m 1\{\omega \in A_m\}$ for some events A_m. Then one defines $E(X) = \sum_{m=1}^{n} a_m P(A_m)$. The general definition is obtained by approximating X by simple random variables. Thus, if $X \geq 0$, one defines $X_n = \max\{k2^{-n} | k \geq 0, k2^{-2n} \leq \min\{X, n\}\}$. One sees that each X_n is a simple random variable. Moreover, $X_{n+1}(\omega) \geq X_n(\omega)$ for all ω. It is easy to show that $E(X_n)$ is nondecreasing and must therefore converge to some limit that defines $E(X)$. In the general case, one writes $X = X^+ - X^-$ where $X^+ := \max\{X, 0\}$ and $X^- := X - X^+$. One then defines $E(X^+)$ and $E(X^-)$ as before, and $E(X) := E(X^+) - E(X^-)$, provided that the result is not $\infty - \infty$.

We now state a very convenient result.

Theorem A.2. *Let X be a random variable on $\{\Omega, \mathcal{F}, P\}$ with c,p.d.f. F_X and let g be a Borel function. Then*

$$E(F(X)) = \int_{-\infty}^{\infty} g(x) dF_X(x) = \int_{-\infty}^{\infty} g(x) f_X(x) dx,$$

where the second identity holds if f_X exists.

Recall also the following definition:

Definition A.4. *The* variance *of the random variable X is designated by $var(X)$ and is defined as $var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$.*

We will use the following simple inequalities.

Fact A.2. *Let X be a random variable.*

(a) Markov: If $h : [0, \infty) \to \Re$ is nondecreasing, then

$$P(|X| \geq \epsilon) \leq \frac{E(f(|X|))}{f(\epsilon)}.$$

In particular, for $h(x) = x^2$, one gets Chebychev's inequality:

$$P(|X| \geq \epsilon) \leq \frac{E(X^2)}{\epsilon^2}.$$

(b) Jensen: Assume that $g : \Re \to \Re$ is a convex function. Then

$$E(g(X)) \geq g(E(X)).$$

Proof:

(a): Note that $f(\epsilon) 1\{|X| \geq \epsilon) \leq f(|X|)$ as you can verify by considering separately the cases $|X| < \epsilon$ and $|X| \geq \epsilon$. The result then follows by taking expectations.

(b) Since g is convex, it lies above the tangent line to its graph at the point $(E(X), g(E(X)))$. If we designate the slope of that tangent by α, we conclude that

$$g(X) \geq g(E(X)) + \alpha(X - E(X)).$$

Taking expectation on both sides yields the result.

□

The σ-additivity of P implies continuity properties of the expected value. For instance, let $X_n(\omega) = 1\{\omega \in A_n\}$ where $A_n \uparrow A$ as $n \to \infty$. Then $X_n \uparrow X$ where $X(\omega) = 1\{\omega \in A\}$ and we see that $E(X_n) = P(A_n) \uparrow E(X) = P(A)$.

However, one must be careful, as the following example demonstrates. Let $\{\Omega, \mathcal{F}, P\} = ([0,1], \mathcal{B}([0,1]), P)$ where $P([a,b]) = b - a$ for $0 \leq a \leq b \leq 1$. Define $X_n(\omega) = n \times 1\{\omega \leq 1/n\}$ and $X(\omega) = 0$ for all $\omega \in [0,1]$. We see that $X_n(\omega) \to X(\omega)$ for all $\omega \neq 0$. However, $E(X_n) = 1 \nrightarrow E(X) = 0$. Thus, in general $X_n \to X$ does not imply $E(X_n) \to E(X)$. The important positive result is as follows.

We need the following definition.

Definition A.5. *Almost Sure Convergence*

Let X_n and X be random variables on $\{\Omega, \mathcal{F}, P\}$. One says that X_n converges almost surely *to X, and one writes $X_n \xrightarrow{a.s.} X$ if the set of ω such that the numbers $X_n(\omega)$ converge to the number $X(\omega)$ has probability one. That is,*

$$X_n \xrightarrow{a.s.} X \Leftrightarrow P(\lim_{n\to\infty} X_n = X) = 1.$$

Theorem A.3. *Lebesgue Convergence Theorem*

(a) Monotone Convergence Theorem: MCT. *Assume that $0 \leq X_n \uparrow_{a.s.} X$ as $n \to \infty$. Then $E(X_n) \uparrow E(X)$ as $n \to \infty$.*

(b) Dominated Convergence Theorem: DCT. *Assume that $X_n \xrightarrow{a.s.} X$ as $n \to \infty$ and $|X_n| \leq Y$ for $n \geq 1$ with $E(Y) < \infty$. Then $E(X_n) \to E(X)$ as $n \to \infty$.*

Appendix B

Notes on Linear Algebra

B.1 Introduction

Linear Algebra is the theory of linear transformations. Applications abound in estimation, control, and Markov chains. You should be familiar with the following concepts that we review in these notes:

- Linear transformation, vector, matrix, determinant
- Basis
- Eigenvector, eigenvalue
- Range, rank, null space
- Diagonalization
- Rotation
- Projection
- Singular Value Decomposition

B.2 Preliminaries

Matrix Notation

You remember the matrix notation. For instance,

$$\begin{aligned}
&\begin{bmatrix} 3 & 4 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 7 & -2 \\ 6 & 9 \end{bmatrix} \\
&\quad = \begin{bmatrix} 3 \times 7 + 4 \times 6 & 3 \times (-2) + 4 \times 9 \\ 1 \times 7 + 5 \times 6 & 1 \times (-2) + 5 \times 9 \end{bmatrix} \\
&\quad = \begin{bmatrix} 45 & 30 \\ 37 & 43 \end{bmatrix}
\end{aligned}$$

Thus, element $(AB)_{ij}$ (meaning row i and column j) of AB is the product of row i of A times column j of B, where the product of two vectors is the sum of the products of their

corresponding elements. This makes sense if $A \in \mathbb{C}^{m\times k}$, meaning hat A has m rows and k columns, and $B \in \mathbb{C}^{k\times n}$ and then $AB \in \mathbb{C}^{m\times n}$.

Observe that, with $z \in \mathbb{C}^m$ and $v_1, \ldots, v_m \in \mathbb{C}^n$,

$$[v_1|v_2|\cdots|v_m]z = \sum_j z_j v_j.$$

Also,

$$A(B+C) = AB + BC.$$

Recall that

$$(AB)^* = B^*A^*$$

where A^* is the *transposed* of A defined by $A^*_{ij} = A_{ji}$.

Also, in general, $AB \neq BA$.

If $A \in \mathbb{C}^{n\times n}$, then A^{-1} is a matrix in $\mathbb{C}^{n\times n}$ such that AA^{-1} is the identity matrix I. It is easy to see that A^{-1} is unique, if it exists. Also, $(AB)^{-1} = B^{-1}A^{-1}$.

Inverse of 2×2

Note that, with $\Delta := ad - bc$,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} \Delta & 0 \\ 0 & \Delta \end{bmatrix},$$

so that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\Delta}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix},$$

provided that $\Delta \neq 0$.

For instance,

$$\begin{bmatrix} 3 & 5 \\ 2 & 7 \end{bmatrix}^{-1} = \frac{1}{11}\begin{bmatrix} 7 & -5 \\ -2 & 3 \end{bmatrix},$$

which enables us to state that the unique solution of the equations

$$\begin{bmatrix} 3 & 5 \\ 2 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 13 \\ 37 \end{bmatrix}$$

is

$$\begin{aligned}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 3 & 5 \\ 2 & 7 \end{bmatrix}^{-1}\begin{bmatrix} 13 \\ 37 \end{bmatrix} \\ &= \frac{1}{11}\begin{bmatrix} 7 & -5 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 13 \\ 37 \end{bmatrix} = \begin{bmatrix} -94/11 \\ 85/11 \end{bmatrix}.\end{aligned}$$

Note also that the matrix

$$\begin{bmatrix} 3 & 5 \\ 6 & 10 \end{bmatrix}$$

does not have an inverse, because $ad - bc = 3\times 10 - 6\times 5 = 0$. Observe that

$$\begin{bmatrix} 3 & 5 \\ 6 & 10 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}x_1 + \begin{bmatrix} 5 \\ 10 \end{bmatrix}x_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}(3x_1 + 5x_2).$$

Consequently, the equation

$$\begin{bmatrix} 3 & 5 \\ 6 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

has no solution unless

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

for some α. In that case, there are infinitely many solutions that correspond to $3x_1 + 5x_2 = \alpha$.

Solving Homogeneous Linear Equations

Consider the equations

$$\begin{bmatrix} 2 & 6 & 4 \\ 3 & 11 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We can subtract from the second equation the first one multiplied by $3/2$ without changing the solutions. Accordingly, these equations are equivalent to the following:

$$\begin{bmatrix} 2 & 6 & 4 \\ 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Let us fix $x_3 = u \in \mathbb{C}$. The second equation is $2x_2 + 2x_3 = 0$ and it yields

$$x_2 = -u.$$

The first equation is $2x_1 + 6x_2 + 4x_3 = 0$ and it yields

$$x_1 = -3x_2 - 2x_3 = -3(-u) - 2u = u.$$

We conclude that the general solution of the equations is

$$x_1 = u, x_2 = -u, x_3 = u.$$

The procedure that we used consists in performing elementary row manipulations of the equations to bring them into a *row echelon form.* In that form, the column index of left-most nonzero term of a row is called the *pivot.* The index of the pivot is increasing for successive rows. We then solve recursively "bottom up."

The elementary row operations consist in interchanging rows or subtracting a multiple of one row from another row. Let A be an arbitrary matrix. Assume column i is the left-most nonzero column. By interchanging rows, one makes sure that element $(1, i)$ is nonzero. By subtracting row 1 multiplied by a_{ji}/a_{1i} from row i, the new matrix has all entries ji equal to zero for $j \geq 2$. We then interchange rows below row 1 to make sure that the leftmost nonzero term in those rows is in row 2. We continue in this way to obtain the row echelon form.

Theorem B.1. *A system of m homogeneous linear equations with n variables has infinitely many solutions if $n > m$.*

Proof:

First reduce the equations to the row echelon form. Since there are more columns than rows, there must be columns without pivots. One can fix the variables x_i arbitrarily for the values of i such that column i has no pivot; these are free variables. Proceeding bottom up, one solves for the last pivot variable in terms of the free variables. One then continues recursively, solving for the other pivot variables. □

B.3 Range, Rank, Null Space, etc.

B.3.1 Linear Independence

The vectors $\{v_1, \ldots, v_k\} \in \mathbb{C}^n$ are *linearly independent* if

$$a_1 v_1 + \cdots a_k v_k = 0 \text{ only if } a_1 = \cdots = a_k = 0.$$

A *linear subspace* $\mathcal{V}$ of $\mathbb{C}^n$ is a subset of $\mathbb{C}^n$ such that a linear combination of two elements of $\mathcal{V}$ is again in $\mathcal{V}$. That is, $v_1, v_2 \in \mathcal{V}$ and $a_1, a_2 \in \mathbb{C}$ implies $a_1 v_1 + a_2 v_2 \in \mathcal{V}$. For instance, the collection of vectors $x \in \mathbb{C}^2$ such that $x_2 = 3x_1$ is a linear subspace of $\mathbb{C}^2$.

Let $\mathcal{V}$ be a subspace of $\mathbb{C}^n$. A *basis* of $\mathcal{V}$ is a collection $\{v_1, \ldots, v_m\}$ of vectors of $\mathcal{V}$ that are linearly independent and such that every $v \in \mathcal{V}$ can be written as $v = a_1 v_1 + \cdots + a_m v_m$ for some $a_i \in \mathbb{C}$.

A linear map $Q : V \to V$ is *invertible* if it is a bijection.

Fact B.1. *Let $\{v_1, \ldots, v_m\}$ be a basis of the linear subspace $\mathcal{V}$. For any $v \in \mathcal{V}$, the representation $v = a_1 v_1 + \cdots + a_m v_m$ is unique. That is, if*

$$v = a_1 v_1 + \cdots + a_m v_m = b_1 v_1 + \cdots + b_m v_m,$$

then $a_i = b_i$ for $i = 1, \ldots, m$.

Proof:

Note that

$$0 = (a_1 - b_1) v_1 + \cdots + (a_m - b_m) v_m.$$

Since the v_i are linearly independent, it follows that $a_i - b_i = 0$ for $i = 1, \ldots, m$. □

Fact B.2. *Let $\{v_1, \ldots, v_m\}$ be a basis of the linear subspace $\mathcal{V}$. Let $\{w_1, \ldots, w_k\} \subset \mathcal{V}$ with $k > m$. Then the vectors $\{w_1, \ldots, w_k\}$ are linearly dependent.*

Proof:

We want to find $\{b_1, \ldots, b_k\}$ that are not all zero and such that $b_1 w_1 + \cdots + b_k w_k = 0$. We know that

$$w_i = \sum_{j=1}^{m} a_{ij} v_j, \text{ for } i = 1, \ldots, k.$$

Hence,

$$0 = \sum_{i=1}^{k} b_i w_i = \sum_{i=1}^{k} \sum_{j=1}^{m} b_i a_{ij} v_j = \sum_{j=1}^{m} [\sum_{i=1}^{k} b_i a_{ij}] v_j.$$

It follows that

$$\sum_{i=1}^{k} b_i a_{ij} = 0, \text{ for } j = 1, \ldots, m.$$

This is a system of m homogeneous equations with $k > m$ variables. We know from Fact B.1 that there are infinitely many solutions. In particular, there are nonzero solutions $\{b_1, \ldots, b_k\}$. □

B.3.2 Dimension

Theorem B.2. *Let $\mathcal{V}$ be a subspace of $\mathbb{C}^n$. If $\{v_1, \ldots, v_m\}$ and $\{w_1, \ldots, w_k\}$ are two bases of $\mathcal{V}$, then $m = k \leq n$. The value of m is called the* dimension *of $\mathcal{V}$ that we designate by $dim(\mathcal{V})$. If $Q : \mathcal{V} \to \mathcal{V}$ is a linear invertible map, then $dim(Q\mathcal{V}) = dim(\mathcal{V})$.*

Proof:

Since $\{v_1, \ldots, v_m\}$ is a basis, Fact B.2 implies that $k \leq m$. A symmetric argument shows that $m \leq k$. It follows that $k = m$.

To show that $m \leq n$, note that $\{e_1 = (1, 0, \ldots, 0)^*, \ldots, e_n = (0, 0, \ldots, 0, 1)\}$ is a basis, so that no more than n vectors can be linearly independent, again by Fact B.2.

With $V = [v_1|\cdots|v_m]$, we claim that QV is a basis for $Q\mathcal{V}$. Indeed, $QVx = 0$ implies $Vx = 0$ because Q is invertible, which implies $x = 0$.

□

B.3.3 Range, Rank, Null Space

Let $V = \{v_1, \ldots, v_m\}$ be a collection of vectors in $\mathbb{C}^n$ and $V = [v_1|\cdots|v_m] \in \mathbb{C}^{n\times m}$ the matrix with columns v_j. By V^* we designate the *conjugate* of the matrix V. That is, $(V^*)_{ij}$ is the complex conjugate of V_{ji}. Thus, $V^* \in \mathbb{C}^{m\times n}$.

Definition B.1. *We define $R(V)$, the* range *of V, to be the collection of linear combinations of the vectors v_i. That is*

$$R(V) = \{Va \mid a \in \mathbb{C}^m\}.$$

The range of V is a linear subspace of $\mathbb{C}^n$. The dimension of that subspace is called the rank *of V.*

The null space *of V, $N(V)$, is defined as the following set:*

$$N(V) := \{z \in \mathbb{C}^m \mid Vz = 0\}.$$

Theorem B.3. *Let $V \in \mathbb{C}^{n\times m}$. Then*

$$dim(R(V)) = dim(R(V^*)).$$

Moreover,

$$dim(R(V)) + dim(N(V)) = m.$$

Proof:

Convert the matrix V to the echelon form by performing row operations. Since these operations are invertible, they do not change the dimension of the range of V. Also, they do not change $N(V)$. Now, say that the reduced matrix has k nonzero rows, and therefore k pivot variables. The k nonzero rows are linearly independent, and there are only k linearly independent row, so that $dim(R(V^*)) = k$. Also, the k columns that correspond to the pivot variables are linearly independent and there cannot be more than k linearly independent columns. Hence, $dim(R(V)) = k$.

In the previous argument, there are $m - k$ columns that do not have pivots and they correspond to independent free variables. Consequently, $dim(N(V)) = m - k$. □

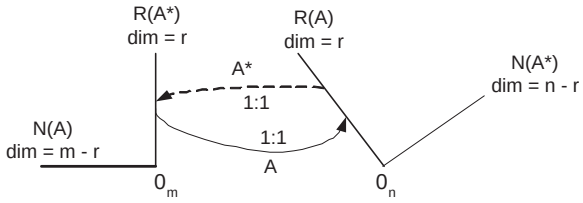


Figure B.1: The actions of A and A^*.

Lemma B.1. *Let $A \in \mathbb{C}^{m\times n}$ be a matrix of rank r.*

(a) One has the decompositions

$$\mathbb{C}^n = R(A^*) \oplus N(A) \text{ and } \mathbb{C}^m = R(A) \oplus N(A^*)$$

where $R(A^*) \perp N(A)$ and $R(A) \perp N(A^*)$. That is, any $x \in \mathbb{C}^n$ can be written uniquely as $x = u + v$ where $u \in R(A^*)$ and $v \in N(A)$ and similarly for $\mathbb{C}^m$.

(b) The restriction of A on $R(A^*)$ is a bijection of $R(A^*)$ onto $R(A)$ and

$$N(AA^*) = N(A^*), R(AA^*) = R(A).$$

(c) The restriction of A^* on $R(A)$ is a bijection of $R(A)$ onto $R(A^*)$ and

$$N(A^*A) = N(A), R(A^*A) = R(A^*).$$

Proof:

Figure B.1 illustrates the statements of the Lemma.

The key observation is that

$$R(A)^\perp := \{x \in \mathbb{C}^m \mid x^*y = 0, \forall y \in R(A)\} = N(A^*).$$

To see this, note that

$$\begin{aligned} x \in R(A)^\perp &\Leftrightarrow x^*Az = 0, \forall z \in \mathbb{C}^n \\ &\Leftrightarrow z^*A^*x = 0, \forall z \in \mathbb{C}^n \\ &\Leftrightarrow A^*x = 0 \Leftrightarrow x \in N(A^*). \end{aligned}$$

Consequently, replacing A by A^*, one finds that $R(A^*)^\perp = N(A)$.

Since $R(A^*)^\perp = N(A)$, any vector x can be written uniquely as $x = u + v$ where $u \in R(A^*)$ and $v \in N(A)$. Consequently, $Ax = Au$ and we see that A restricted to $R(A^*)$ is onto $R(A)$. Moreover, we claim that A restricted to $R(A^*)$ is one-to-one on $R(A)$. That is, if $Au_1 = Au_2$ for $u_1, u_2 \in R(A^*)$, then $u_1 = u_2$. Equivalently, let $Au = 0$ for some $u \in R(A^*)$; we show that $u = 0$. To see this, observe that $u \in N(A) \cap R(A^*) = \{0\}$.

We now verify that $N(AA^*) = N(A^*)$. Assume that $x \in N(AA^*)$. Then $AA^*x = 0$, so that $||A^*x||^2 = x^*AA^*x = 0$, which shows that $A^*x = 0$ and $x \in N(A^*)$. Hence, $N(AA^*) \subset N(A^*)$. Obviously, $N(A^*) \subset N(AA^*)$ since $A^*x = 0$ implies $AA^*x = 0$. Hence, $N(AA^*) = N(A^*)$.

To show that $R(AA^*) = R(A)$, assume that $x \in R(A)$, so that $x = Ay$. We decompose y as $y = u + v$ where $u = A^*z \in R(A^*)$ and $v \in N(A)$. Then, $x = A(u + v) = Au = AA^*z$, so that $x \in R(AA^*)$. Hence, $R(A) \subset R(AA^*)$. Also, $R(AA^*) \subset R(A)$ since any $x \in R(AA^*)$ is of the form $x = AA^*y = A(A^*y) \in R(A)$.

The other statements of the lemma are obtained by replacing A by A^*.

□

B.4 Determinant

We define the *determinant* $\det(A)$ of a square matrix A. This concept is essential to understand the inverse of matrices and also the change of variables in multivariate calculus.

Before introducing the definition, we need to review the notion of index of a permutation.

B.4.1 Permutations

Call a *transposition* of a list of numbers the interchange of two successive numbers in the list.

Consider the numbers $(1, 2, 3)$. They are in increasing order. Thus, the minimum number $s(1, 2, 3)$ of transpositions required to put these numbers in increasing order is equal to zero; that is, $s(1, 2, 3) = 0$.

Now consider $(3, 1, 2)$. To arrange these elements in increasing order we can interchange the first two elements to obtain $(1, 3, 2)$, then interchange the last two elements to obtain $(1, 2, 3)$. That is, we can reorder the three elements with two transpositions of successive elements. You can also see that this is the minimum number of such transpositions required since no single transposition does the job. Hence, $s(3, 1, 2) = 2$.

Consider a permutation $p = (p_1, p_2, \ldots, p_n)$ of $(1, 2, \ldots, n)$. There is certainly a sequence of transpositions that reorders these numbers. For instance, let $p_k = 1$. We can bring 1 back in the first position by interchanging the elements k and $k-1$, then the elements $k-1$ and $k-2$, ..., then the elements 2 and 1. After these $k-1$ transpositions, the permutation p has been modified to $(1 = p_k, p_1, \ldots, p_{k-1}, p_{k+1}, \ldots, p_n)$. One can then repeat the procedure to bring 2 in second position with a finite number of transpositions, and so on. We define $s(p)$ to be the minimum number of transpositions required to arrange the elements in p in increasing order.

Assume that one can reorder the elements of a permutation p with k transpositions. We claim that $k - s(p)$ must be an even number. Equivalently, if we perform a number k of transpositions of the elements $(1, 2, \ldots, n)$ that brings them back to the same order, then k must be even. To see this, note that the result is obvious for two elements. Assume that the result holds for n elements and we color these elements blue. Let us add one element that we color red. Consider then the set of k transpositions that bring back the elements to their original order. Among these transpositions, a number k_1 do not move the red element. By induction, k_1 must be even. The other $k_2 = k - k_1$ transpositions move the red element to the right or to the left in the set of blue elements. It is obvious that k_2 must be even to bring the red element back to its original position.

Consider $A \in \mathbb{C}^{n\times n}$. Let α be a selection of n entries of A taken from different rows and different columns. We can list the elements of α in increasing order of their rows. Thus, $\alpha = \alpha(p) := (a_{1p_1}, a_{2p_2}, \ldots, a_{np_n})$ where $p := (p_1, p_2, \ldots, p_n)$ is some permutation of $(1, 2, \ldots, n)$. Thus, there are $n!$ such selections α, one for each permutation p.

B.4.2 Definition

Definition B.2. *Let $A \in \mathbb{C}^{n\times n}$ for some $n \geq 2$. The* determinant $\det(A)$ *of matrix A is defined as*

$$\det(A) = \sum_p a_{1p_1} a_{2p_2} \cdots a_{np_n} (-1)^{s(p)} = \sum_p \alpha(p)(-1)^{s(p)}$$

where the sum is over all the possible permutations p of $(1, 2, \ldots, n)$.

To clarify this definition, let us look at some examples. For $a, b, c, d \in \mathbb{C}$,

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \alpha(1,2) - \alpha(2,1) = a_{11}a_{22} - a_{12}a_{21} = ad - bc.$$

As another example,

$$\begin{aligned} \det & \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \\ &= \alpha(1,2,3) - \alpha(1,3,2) - \alpha(2,1,3) \\ &\quad + \alpha(2,3,1) + \alpha(3,1,2) - \alpha(3,2,1) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \\ &\quad + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \\ &= aei - afh - bdi + bfg + cdh - ceg. \end{aligned}$$

Fact B.3. *(a) Assume the matrix B is obtained from $A \in \mathbb{C}^{n\times n}$ by interchanging two rows or two columns, then*

$$\det(B) = -\det(A).$$

(b) Assume the matrix B is obtained from $A \in \mathbb{C}^{n\times n}$ by multiplying one row or one column by a constant γ, then

$$\det(B) = \gamma \det(A).$$

(c) Assume the matrix A has two identical rows or two identical columns, then

$$det(A) = 0.$$

(d) Let A^* designate the transposed *of matrix A, defined so that the entry ij of A^* is the entry ji of matrix A. Then*

$$\det(A^*) = \det(A).$$

Proof:

For (a), note that the interchange modifies the index of every permutation by one. Fact (b) is obvious. Fact (c) follows from (a).

For fact (d), consider the term $\alpha(p) = a_{1p_1} \cdots a_{np_n}$ in $\det(A)$. That term appears in $\det(A^*) := \det(B)$ as $b_{p_1 1} \cdots b_{p_n n}$. Let us rearrange the terms $(b_{p_1 1}, \ldots, b_{p_n n})$ so that they are in increasing order of the rows. To do this, we need to perform $s(p)$ transpositions. We end up with the terms $b_{1q_1} \cdots b_{nq_n}$. But then we know that $s(p)$ transpositions of the terms $q = (q_1, \ldots, q_n)$ bring them back in the order $(1, 2, \ldots, n)$. That is, $s(q) = s(p)$. Accordingly, the term $a_{1p_1} \cdots a_{np_n}$ that appears with the sign $s(p)$ in $\det(A)$ appears with the sign $s(q) = s(p)$ in $\det(A^*)$ and we see that $\det(A) = \det(A^*)$.

□

Definition B.3. *Let $A \in \mathbb{C}^{n\times n}$. For $i, j \in \{1, \ldots, n\}$, the matrix $A^{ij} \in \mathbb{C}^{(n-1)\times(n-1)}$ is obtained from A by deleting row i and column j. We then define the* cofactor *of the entry a_{ij} of matrix A as*

$$c_{ij} = (-1)^{i+j} \det(A^{ij}).$$

Fact B.4.

$$\det(A) = \sum_k a_{ik}c_{ik} = \sum_k a_{kj}c_{kj}, \text{ for all } i, j.$$

Proof:

Consider a generic term in $a_{ik}c_{ik}$. Such a term is

$$\beta := (-1)^{i+k} a_{1p_1} \cdots a_{i-1,p_{i-1}} a_{i+1,p_{i+1}} a_{np_n} (-1)^{s(p)}$$

where $p = (p_1, \ldots, p_n)$ is a permutation of the numbers $(1, 2, \ldots, n)$ from which k has been removed. Consequently,

$$\beta = a_{1p_1} \cdots a_{i-1,p_{i-1}} a_{ik} a_{i+1,p_{i+1}} a_{np_n} (-1)^{s(p)+(i+k)}.$$

We claim that

$$(-1)^{s(p)+(i+k)} = (-1)^{s(p_1,\ldots,p_{i-1},k,p_{i+1},\ldots,p_n)}.$$

Let us first perform $i-1$ transpositions of $(p_1, \ldots, p_{i-1}, k, p_{i+1}, \ldots, p_n)$ to get

$$(k, p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n).$$

We then perform $s(p)$ transpositions to get

$$(k, 1, \ldots, k-1, k+1, \ldots, n).$$

Finally, with $k-1$ other transpositions, we get

$$(1, 2, \ldots, k-1, k, k+1, \ldots, n).$$

That is, we can reorder the terms $(p_1, \ldots, p_{i-1}, k, p_{i+1}, \ldots, p_n)$ with $s(p) + i + k - 2$ transpositions. We know that this may not be the minimum number of transpositions, but the difference with the minimum number must be even, as we explained when we introduced the index of permutations. This proves the claim.

It follows from this observation that the terms in $a_{ik}c_{ik}$ as k ranges over all possible values are all the terms in $\det(A)$.

□

Fact B.5. *Let B be a matrix obtained from A by subtracting from one row a multiple of a different row. Then*

$$\det(B) = \det(A).$$

Proof:

Assume that all the rows of B are the same as the rows of A, except that row i is equal to row i of A minus α times row j of A, for some $i \neq j$. Let c_{ik} be the cofactor of a_{ik} and c'_{ik} be the cofactor of b_{ik}. Then $c_{ik} = c'_{ik}$ since these cofactors involve only elements of the matrices that are not on row i. The expansion of the determinants gives

$$\det(A) = \sum_k a_{ik}c_{ik} \text{ and } \det(B) = \sum_k (a_{ik} - \alpha a_{jk})c_{ik}.$$

Now, $\sum_k a_{jk}c_{ik}$ is the determinant of the matrix obtained from A by replacing row i by a copy of row j. Since this matrix has two identical rows, its determinant is zero. We conclude that $\det(A) = \det(B)$.

□

Fact B.6. *Let $A, B \in \mathbb{C}^{n\times n}$. Then*

$$\det(AB) = \det(A)\det(B).$$

Proof:

The trick in the proof is to reduce the matrices to a row echelon form by performing simple row operations where we subtract from one row a multiple of a row above. For instance, say that we replace row 2 in matrix A by row 2 minus α times row 1. We then obtain a matrix A' with

$$a'_{ij} = a_{ij} - \alpha a_{1j} 1\{i = 2\}, j = 1, \ldots, n.$$

Note that

$$\begin{aligned}(A'B)_{ik} &= \sum_j a'_{ij} b_{jk} = \sum_j (a_{ij} - \alpha a_{1j} 1\{i = 2\}) b_{jk} \\ &= (AB)_{ik} - \alpha (AB)_{1k} 1\{i = 2\}.\end{aligned}$$

That is, the rows of $A'B$ are those of AB, except that α times row 1 is subtracted from row 2. That is, the elementary row operation performed on A results in the same row operation on AB. We know from Fact B.5 that such row operations do not modify the determinant of a matrix.

Similarly, if we interchange two rows in matrix A, then the corresponding rows of AB are interchanged. Indeed, row i of AB is equal to row i multiplied by B. Such an operation multiplies the determinant of A and that of AB by -1.

Let us then perform elementary row operations on A to reduce it to a row echelon form. The determinant of the reduced matrix A is nonzero only if the reduced matrix has a triangular form. That is, if the diagonal elements are nonzero and the terms below the diagonal are zero. The determinant of a matrix in triangular form is the product of the diagonal elements, as is easy to see by induction on the size of the matrix from Fact B.4. These operations do not modify the identity between the determinants of A and of AB,

We now perform elementary *column* operations on B to reduce it to row echelon form. These operations correspond to the same operations on AB. For instance, if $b'_{jk} = b_{jk} - \alpha b_{j1} 1\{k = 2\}$, then

$$\begin{aligned}(AB')_{ik} &= \sum_j a_i j (b_{jk} - \alpha b_{j1} 1\{k = 2\}) \\ &= (AB)_{ik} - \alpha (AB)_{i1} 1\{k = 2\}.\end{aligned}$$

Once again the determinant of B is nonzero only if after reduction B is triangular. so that B is also triangular. But now, both A and B are triangular and it is immediate that AB is triangular. Moreover, the diagonal terms of AB are the product of the corresponding diagonal terms of A and B. That is, the determinant of AB is the product of the determinants of A and B.

Since the identity between the determinant of AB and those of A and B has not been modified by the elementary row and column operations, we conclude that $\det(AB) = \det(A)\det(B)$ for the original matrices.

□

B.5 Inverse

Consider the matrix $A \in \mathbb{C}^{n\times n}$. We know that the span of A, $\text{span}(A) := \{Ax \mid x \in \mathbb{C}^n\}$, is a linear space with dimension $\text{rank}(A)$. If the columns of A are linearly independent, i.e., if $\text{rank}(A) = n$, $\text{span}(A) = \mathbb{C}^n$ and for every $y \in \mathbb{C}^n$, there is some $x \in \mathbb{C}^n$ such that $Ax = y$. Moreover, that x must be unique since we know that the representation of a vector as a linear combination of basis vectors is unique. We can then define that unique x to be the *inverse* of y under A and we designate it by $x = A^{-1}y$.

If $\text{rank}(A) < n$, then $\text{span}(A)$ is a proper subset of $\mathbb{C}^n$, so that there are some $y \in \mathbb{C}^n$ for which there is no x such that $Ax = y$. Also, if y is in $\text{span}(A)$, then there are infinitely many x such that $Ax = y$ and any two such values of x differ by an element in the null space of A.

We now discuss the calculation of A^{-1} in terms of determinants.

Theorem B.4. *Let $A \in \mathbb{C}^{n\times n}$. Then $\text{rank}(A) = n$ if and only if $\det(A) \neq 0$. Moreover, in that case, there is a unique matrix $A^{-1} \in \mathbb{C}^{n\times n}$ such that $AA^{-1} = I$ where I designates the identity matrix, i.e., the matrix whose elements are $I_{ij} = 1\{i = j\}$. That matrix has entries*

$$A^{-1}_{ij} = \frac{1}{\det(A)} c_{ji}$$

where c_{ji} is the cofactor of a_{ji} in A.

Proof:

Define B to be the matrix with entries $b_{ij} = c_{ji}$. We show that $AB = \det(A)I$. First we check the diagonal elements of AB. We find

$$(AB)_{ii} = \sum_k a_{ik}b_{ki} = \sum_k a_{ik}c_{ik} = \det(A),$$

by Fact B.4.

Second, we check the off-diagonal terms. Thus, let $i \neq j$. Then

$$(AB)_{ij} = \sum_k a_{ik}b_{kj} = \sum_k a_{ik}c_{jk}.$$

Now, consider the matrix D obtained from A by replacing row j of A by a copy of row i of A. Note that the cofactor of d_{jk} is not affected by this replacement since that cofactor involves terms that are not on row j. Accordingly, that cofactor is c_{jk}. Hence, we can write, using Fact B.4,

$$\det(D) = \sum_k d_{jk}c_{jk} = \sum_k a_{ik}c_{jk}.$$

But we know that $\det(D) = 0$, because that matrix has two identical rows (see Fact B.3(b)).

Putting these conclusions together, we see that $AB = \det(A)I$. This shows that if $\det(A) \neq 0$, then the matrix $A^{-1} := B/\det(A)$ is such that $AA^{-1} = I$. It follows that $Ax = y$ has a unique solution $x = A^{-1}y$, so that the column of A must be linearly independent and $\text{rank}(A) = n$.

Now assume that $\text{rank}(A) = n$. We want to show that $\det(A) \neq 0$. Since $\text{rank}(A) = n$, for every $i \in \{1, 2, \ldots, n\}$, there must be some $v_i \in \mathbb{C}^n$ such that $Av_i = e_i$, where $e_i(j) = 1\{i = j\}$. The matrix D with columns v_i must then be such that $AD = I$. Since $\det(AD) = \det(A)\det(D)$, this implies that $\det(A) \neq 0$. □

Theorem B.5. *Let $A \in \mathbb{C}^{n\times n}$. Then $Ax = 0$ has a nonzero solution if and only if* $\det(A) = 0$.

Proof:

If $\det(A) \neq 0$, then we know that A has an inverse A^{-1}, so that $Ax = 0$ implies $x = 0$.

Assume that $\det(A) = 0$. By performing elementary row operations, one converts A to a row echelon form B with $\det(B) = 0$. Now, $\det(B)$ is the product of the diagonal elements of B. Consequently, B has a zero diagonal element which corresponds to a nonzero x_i solving $Bx = 0$, and therefore $Ax = 0$.

□

B.6 Eigenvalues and Eigenvectors

These concepts are very useful. We start with an example.

B.6.1 Example

Consider the matrix

$$A = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}.$$

An *eigenvector* of A is a nonzero vector v such that Av is proportional to v; that is, $Av = \lambda v$. The constant λ is called the *eigenvalue* associated with the eigenvector v. You can check that

$$v_1 := \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ and } v_2 := \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

are such that

$$Av_1 = v_1 \text{ and } Av_2 = 5v_2.$$

Accordingly, v_1 and v_2 are two eigenvectors of A with respective eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 5$.

This concept is useful because we can write any other vector x as a linear combination of v_1 and v_2. For instance,

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = y_1 v_1 + y_2 v_2$$

with $y_1 = (-x_1 + 3x_2)/4$ and $y_2 = (x_1 + x_2)/4$. We then find that

$$Ax = A(y_1 v_1 + y_2 v_2) = y_1 Av_1 + y_2 Av_2 = y_1 \lambda_1 v_1 + y_2 \lambda_2 v_2.$$

We can write the above calculations in matrix form. We have

$$x = Vy = [v_1|v_2] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \text{ where } y = \begin{bmatrix} -1/4 & 3/4 \\ 1/4 & 1/4 \end{bmatrix} = V^{-1}x,$$

so that

$$Ax = AVy = AVV^{-1}x = [\lambda_1 v_1|\lambda_2 v_2]V^{-1}x = \Lambda V V^{-1}x$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2)$ is the diagonal matrix with diagonal elements λ_1, λ_2.

Repeating these operations n times, we find

$$A^n x = \Lambda^n V V^{-1}x$$

where $\Lambda^n = \text{diag}(\lambda_1^n, \lambda_2^n)$.

B.6.2 General Case

Definition B.4. *Let $A \in \mathbb{C}^n$. An* eigenvector *of A is a nonzero $v \in \mathbb{C}^n$ such that $Av = \lambda v$. The constant $\lambda \in \mathbb{C}$ is called the* eigenvalue *associated with v.*

Fact B.7. *$Av = \lambda v$ for some $v \neq 0$ if and only if $\det(A - \lambda I) = 0$.*

Proof:
$(A-\lambda I)v = 0$ admits a nonzero solution v if and only if $\det(A-\lambda I) = 0$, by Theorem B.5. □

The fact above tells us how to find the eigenvalues. One can then find the eigenvectors corresponding to an eigenvalue λ by solving the homogeneous equations $(A - \lambda I)v = 0$. One procedure to do this is to reduce the matrix to row echelon form.

Fact B.8. *Assume that A admits the linearly independent eigenvectors $\{v_1, \ldots, v_n\}$ with corresponding eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$. Then*

$$V^{-1}AV = \Lambda := diag(\lambda_1, \ldots, \lambda_n) \text{ where } V = [v_1|\cdots|v_n].$$

Proof:
This is identical to the example we explained earlier. □

Fact B.9. *Assume that the eigenvalues of A are distinct. Then the corresponding eigenvectors are linearly independent.*

Proof:
Assume otherwise, so that $\sum_{j=1}^n c_j v_j = 0$ for some nonzero c_j. Then

$$0 = A(\sum_{j=1}^n c_j v_j) = \sum_{j=1}^n c_j \lambda_j v_j.$$

It follows that

$$0 = \sum_{j=1}^n c_j \lambda_j v_j - \lambda_n \sum_{j=1}^n c_j v_j = \sum_{j=1}^{n-1} c_j (\lambda_j - \lambda_n) v_j.$$

Hence, the eigenvectors $\{v_1, v_2, \ldots, v_{n-1}\}$ are linearly dependent. We can repeat the argument and show that the eigenvectors $\{v_1, v_2, \ldots, v_{n-2}\}$ are linearly dependent. Continuing in this way shows that the eigenvector v_1 must be linearly dependent, clearly a contradiction. □

B.7 Rotation and Projection

The length of a vector $x \in \mathbb{C}^n$ is defined as

$$||x|| := (\sum_{i=1}^n x_i^2)^{1/2} = (x^* x)^{1/2}.$$

Picture a vector x and imagine that you rotate it around the origin, by some angle. Intuitively, this operation is linear. It should therefore correspond to some matrix R. The rotation must

preserve the length and also the angles between vectors. In particular, the rotated axes must be orthogonal and have a unit length. Since the rotated axes are the columns of the matrix R, this implies that

$$R^*R = I.$$

In other words, $R^* = R^{-1}$.

By definition, a *rotation matrix* is a real matrix with the property above. It is immediate that the product of two rotation matrices ia again a rotation matrix. Not surprisingly, one can represent an arbitrary rotation matrix as a product of rotations around the axes.

The concept of *projection* is also important.

Definition B.5. *Assume that $\mathcal{V}$ is a linear subspace of $\mathbb{C}^n$. Let $x \in \mathbb{C}^n$. The* projection of x onto V *is the vector $v \in V$ that is the closest to x, i.e., that achieves*

$$\min\{||x - v|| \ s.t. \ v \in \mathcal{V}\}.$$

We write $v = x|_{\mathcal{V}}$ if v is the projection of x onto $\mathcal{V}$.

Fact B.10. *(a) $v = x|_{\mathcal{V}}$ if and only if $v \in \mathcal{V}$ and $x - v \perp \mathcal{V}$, i.e.,*

$$(x - v)^* w = 0, \forall w \in \mathcal{V}.$$

(b) The projection is unique.

(c) The projection is a linear operation.

(d) Assume that $\mathcal{V} = span(V) = span(v_1, \ldots, v_k)$ where the column vectors of V are linearly independent. Then

$$x|_{\mathcal{V}} = VMx \text{ where } M = (V^*V)^{-1}V^*.$$

Proof:

(a) Take any $w \in \mathcal{V}$, so that $v + \lambda w \in \mathcal{V}$. Then

$$\begin{aligned} ||x - (v + \lambda w)||^2 &= (x - v - \lambda w)^*(x - v - \lambda w) \\ &= ||x - v||^2 - 2\lambda(x - v)^* w + \lambda^2 ||w||^2. \end{aligned}$$

Taking the derivative with respect to λ at $\lambda = 0$, we find that the value of $\lambda = 0$ achieves the minimum if and only if $(x - v)^* w = 0$.

(b) Assume that both v and w achieve the minimum distance to x in $\mathcal{V}$. We claim that $v = w$. Note that $x - v \perp \mathcal{V}$ and $x - w \perp \mathcal{V}$. It follows that $w - v = (x - v) - (x - w) \perp \mathcal{V}$. But $w - v \in \mathcal{V}$. Hence,

$$0 = (w - v)^*(w - v) = ||w - v||^2,$$

so that $v = w$.

(c) Assume $v_i = x_i|_{\mathcal{V}}$ for $i = 1, 2$. We show that

$$a_1 v_1 + a_2 v_2 = (a_1 x_1 + a_2 x_2)|_{\mathcal{V}}.$$

To show this, we verify that

$$a_1 x_1 + a_2 x_2 - (a_1 v_1 + a_2 v_2) \perp \mathcal{V}.$$

This follows from

$$[a_1 x_1 + a_2 x_2 - (a_1 v_1 + a_2 v_2)]^* v = a_1(x_1 - v_1)^* v + a_2(x_2 - v_2)^* v = 0.$$

(d) Note that $VMx \in \mathcal{V}$. Moreover, we observe that, for any $v = Vz \in \mathcal{V}$,

$$\begin{aligned}(x - VMx)^* Vz &= x^* Vz - x^* M^* V^* Vz \\ &= x^* Vz - x^* V(V^*V)^{-1} V^* Vz = 0.\end{aligned}$$

□

A little example might help. Let $v_1 = (1, 2, 3)^*$ and $v_2 = (1, 1, 2)^*$. The $\mathcal{V} = \text{span}(v_1, v_2)$ is a linear subspace of $\mathbb{C}^3$ that consists of all the vectors of the form $z_1 v_1 + z_2 v_2 = Vz$ where V is the matrix with columns v_1 and v_2 and $z = (z_1, z_2)^*$. Note that

$$V^*V = \begin{bmatrix} 14 & 9 \\ 9 & 6 \end{bmatrix},$$

so that

$$\begin{aligned} M^* &= V(V^*V)^{-1} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} \frac{1}{3} \begin{bmatrix} 6 & -9 \\ -9 & 14 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} -3 & 5 \\ 3 & -4 \\ 0 & -1 \end{bmatrix}, \end{aligned}$$

or

$$M = \frac{1}{3} \begin{bmatrix} -3 & 3 & 0 \\ 5 & -4 & 1 \end{bmatrix}.$$

We conclude that

$$x|_{\mathcal{V}} = VMx = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} \frac{1}{3} \begin{bmatrix} -3 & 3 & 0 \\ 5 & -4 & 1 \end{bmatrix} x.$$

For instance, if $x = (2, 1, 1)^*$, then

$$x|_{\mathcal{V}} = -v_1 + \frac{7}{3} v_2 \in \mathcal{V}.$$

Also, if we choose some $v \in \mathcal{V}$, say $v = Vz$, then

$$(x - x|_{\mathcal{V}})^* v = \left(\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \right)^* \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} z = 0.$$

B.8 Singular Value Decomposition

B.8.1 Some Terminology

A matrix $U \in \mathbb{C}^{n \times n}$ is said to be *unitary* if $U^*U = I$, where U^* is the complex conjugate of the transposed of U. The matrix U is said to be orthogonal if it is unitary and real.

A matrix $H \in \mathbb{C}^{n \times n}$ is called *Hermitian* if $H^* = H$; moreover, it is *positive semidefinite* if $x^* Hx \geq 0$ for all $x \in \mathbb{C}^n$; it is *positive definite* if $x^* Hx > 0$ for all nonzero $x \in \mathbb{C}^n$.

B.8.2 Decomposition of Hermitian Matrices

Theorem B.6. *Let $H \in \mathbb{C}^{n\times n}$ be a Hermitian matrix.*

(a) The eigenvalues $\lambda_1, \ldots, \lambda_n$ of H are real (they are not necessarily distinct);

(b) Eigenvectors u_i and u_j that correspond to distinct eigenvalues are orthogonal, i.e., $u_i^* u_j = 0$.

(c) H has n orthonormal eigenvectors $\{u_1, \ldots, u_n\}$ that form a basis for $\mathbb{C}^n$. That is, $u_i^* u_j = 1\{i = j\}$.

(d) If $P = [u_1|\ldots|u_n]$, then

$$P^* H P = \Lambda = diag(\lambda_1, \ldots, \lambda_n)$$

and

$$H = P\Lambda P^* = \sum_{i=1}^{n} \lambda_i u_i u_i^*.$$

(e) H is positive semidefinite (resp., definite) if and only if $\lambda_i \geq 0$ (resp., > 0) for all i.

(f) One has

$$\max_{\{x|||x||=1\}} x^* H x = \max\{\lambda_i\}$$

and

$$\min_{\{x|||x||=1\}} x^* H x = \min\{\lambda_i\}.$$

Proof:

(a) The eigenvalues of H are the zeros of $\det(\lambda I - H)$. This polynomial of degree n has n zeros (not necessarily distinct). We show that every eigenvalue λ_i must be real. Since $H = H^*$, we see that $(u_i^* H u_i)^* = u_i^* H^* u_i = u_i^* H u_i$, so that $u_i^* H u_i$ is real. However, $u_i^* H u_i = \lambda_i u_i^* u_i$ and $u_i^* u_i = ||u_i||^2$ is real. It follows that λ_i is also real.

(b) Assume that $\lambda_i \neq \lambda_j$ and if $Hu_i = \lambda_i u_i$ and $Hu_j = \lambda_j u_j$. Observe that

$$u_i^* H u_j = \lambda_j u_i^* u_j,$$

so that, by taking the complex conjugate,

$$\lambda_j u_j^* u_i = (u_i^* H u_j)^* = u_j^* H u_i = \lambda_i u_j^* u_i,$$

which shows that $(\lambda_i - \lambda_j) u_j^* u_i = 0$. Since $\lambda_i \neq \lambda_j$, it follows that $u_i^* u_j = 0$.

(c) We show that H has n linearly independent eigenvectors. Let u_1 be such that $Hu_1 = \lambda_1 u_1$ and $||u_1|| = 1$. Let also $\mathcal{V}_1$ be the space of vectors orthogonal to u_1. Note that if $x \in \mathcal{V}_1$, then $Hx \in \mathcal{V}_1$. Indeed,

$$(u_1^* H x)^* = x^* H^* u_1 = x^* H u_1 = \lambda_1 x^* u_1 = 0.$$

Pick an orthornormal basis $\{b_2, \ldots, b_n\}$ for $\mathcal{V}_1$. Let $P = [u_1|b_2|\cdots|b_n]$. Any $x \in \mathbb{C}^n$ can be written uniquely as Py with $y = P^* x$ since $P^* P = I$. Also,

$$HPy = \lambda_1 u_1 y_1 + H(\sum_{j=2}^{n} b_j y_j) = \lambda_1 u_1 y_1 + \sum_{j=2}^{n} b_j z_j$$

because $H(\sum_{j=2}^{n} b_j y_j) \in \mathcal{V}_1$. By linearity,

$$\begin{bmatrix} z_2 \\ z_3 \\ . \\ z_n \end{bmatrix} = M \begin{bmatrix} y_2 \\ y_3 \\ . \\ y_n \end{bmatrix},$$

so that

$$HP = P\tilde{H} \text{ where } \tilde{H} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & M \end{bmatrix}.$$

Note that $M^* = M$. We claim that the eigenvalues of M are the eigenvalues of H other than λ_1. Indeed,

$$\begin{aligned} \det(\lambda I - H) &= \det(\lambda I - P\tilde{H}P^*) = \det(P(\lambda I - \tilde{H})P^*) \\ &= \det(\lambda I - \tilde{H}) \end{aligned}$$

because $\det(P) = \det(P^*) = 1$ since $PP^* = I$. Now,

$$\det(\lambda I - \tilde{H}) = (\lambda - \lambda_1)\det(\lambda I_{n-1} - M).$$

Let v_2 be an eigenvector of M with eigenvalue λ_2. That is, $Mv_2 = \lambda_2 v_2$. Let also $w_2 = [0 v_2^*]^*$. Note that

$$\tilde{H}w_2 = \lambda_2 w_2,$$

so that

$$HPw_2 = P\tilde{H}w_2 = \lambda_2 Pw_2,$$

which shows that $u_2 = Pw_2$ is an eigenvector of H with eigenvalue λ_2. Moreover, $u_2 \in \mathcal{V}_1$, so that u_2 is orthogonal to u_1. We can normalize u_2 so that $||u_2|| = 1$. Continuing in this way, we define $\mathcal{V}_2$ to be the subspace of $\mathbb{C}^n$ orthogonal to u_1 and u_2. Repeating the previous steps, we find a collection of vectors $\{u_1, u_2, \ldots, u_n\}$ that are orthonormal and such that $Hu_i = \lambda_i u_i$.

(d) This is immediate.

(e) and (f) follow from (d).

□

B.8.3 Illustration of SVD

The goal of this section is to illustrate that an $m \times n$ matrix A can be decomposed as $U\Sigma V^*$ where Σ is diagonal and U, V are unitary. This decomposition, called the *singular value decomposition* of A, enables us to view the action of A as a combination of scaling and rotations. We use that result to explain the transformation of a sphere into an ellipsoid and to show how volumes gets mapped, which is essential in calculus.

Consider the matrix

$$A = \begin{bmatrix} 1 & \sqrt{3} \\ 2 & 0 \end{bmatrix}.$$

Then

$$A^* = \begin{bmatrix} 1 & 2 \\ \sqrt{3} & 0 \end{bmatrix}, AA^* = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}, A^*A = \begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix}.$$

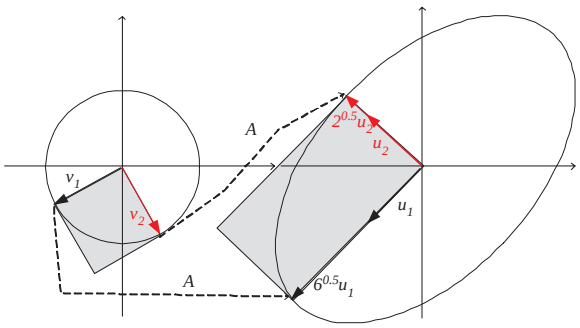


Figure B.2: The action of A combines rotations and scaling.

By solving $\det(\lambda I - AA^*) = 0$, one finds the eigenvalues $\sigma_1^2 = 6$ and $\sigma_2^2 = 2$ of AA^*. These are also the eigenvalues of A^*A. Solving $AA^*u_j = \sigma_j^2 u_j$ and $A^*Av_j = \sigma_j^2 v_j$ for $j = 1, 2$, one finds the following orthonormal eigenvectors of AA^* and A^*A:

$$U = [u_1|u_2] = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix},$$

$$V = [v_1|v_2] = \frac{1}{2} \begin{bmatrix} -\sqrt{3} & 1 \\ -1 & -\sqrt{3} \end{bmatrix}.$$

You can verify that $A = U\Sigma V^*$ with $\Sigma = \text{diag}\{\sigma_1, \sigma_2\} = \text{diag}\{\sqrt{6}, \sqrt{2}\}$.

Since $A = U\Sigma V^*$, it follows that $AV = U\Sigma$. That is, $Av_j = u_j\sigma_j$, which shows that A rotates v_j into u_j and scales it by σ_j.

For instance, consider the sphere $S = \{x \in \mathbb{C}^2 \mid ||x||^2 = 1\}$. Thus, $S = \{\alpha_1 v_1 + \alpha_2 v_2 \mid \alpha_1^2 + \alpha_2^2 = 1\}$. The mapping under A of S is $AS = \{Ax \mid x \in S\}$. Thus,

$$\begin{aligned} AS &= \{AV\alpha \mid ||\alpha||2 = 1\} = \{U\Sigma\alpha \mid ||\alpha||^2 = 1\} \\ &= \{\sigma_1\alpha_1 u_1 + \sigma_2\alpha_2 u_2 \mid \alpha_1^2 + \alpha_2^2 = 1\}. \end{aligned}$$

We conclude that AS is an ellipsoid with principal axes along u_1 and u_2 and of lengths σ_1 and σ_2.

The mapping $S \to AS$ is illustrated in Figure B.2.

As another application, consider the transformation of the square $Q = \{\alpha_1 v_1 + \alpha_2 v_2 \mid \alpha_1, \alpha_2 \in [0, 1]\}$ under the mapping A. As the figure shows,

$$AQ = \{\beta_1\sigma_1 u_1 + \beta_2\sigma_2 u_2 \mid \beta_1, \beta_2 \in [0, 1]\}.$$

We conclude that A maps the square Q with area 1 into a rectangle AQ with area $\sigma_1\sigma_2 = |\det(A)|$. We can write this result 'differentially' as

$$|Adv| = |\det(A)||dv|$$

where $dv = \{x \mid x_j \in (v_j, v_j + dv_j), j = 1, 2\}$ and $|dv| = dv_1 dv_2$. This identity states that A maps a small area dv into an area that is multiplied by $|\det(A)|$.

We now turn to the derivation of that decomposition in the general case.

B.8.4 SVD of a matrix

Theorem B.7. *Let $A \in \mathbb{C}^{m\times n}$ be a matrix of rank r. The matrix A has a* singular value decomposition

$$A = U\Sigma V^* \text{ with } \Sigma = diag\{\sigma_1, \ldots, \sigma_r, 0, \ldots, 0\}$$

where

- *$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the square roots of the common positive eigenvalues of A^*A and AA^*;*
- *$V = [V_1|V_2]$ is unitary; the columns of V_1 are a basis of $R(A^*)$ and those of V_2 are a basis of $N(A)$; the columns of V are a basis of eigenvectors of A^*A;*
- *$U_1 := AV_1\Sigma_1^{-1}$;*
- *$U = [U_1|U_2]$ is unitary; the columns of U_1 are a basis of $R(A)$ and those of U_2 are a basis of $N(A^*)$; the columns of U are a basis of eigenvectors of AA^*.*

Proof:

The matrix $H = A^*A$ is Hermitian and positive semidefinite since $x^*Hx = (Ax)^*(Ax) \geq 0$. Moreover, H has rank r because $R(A^*A) = R(A)$, by Lemma B.1. By Theorem B.6, H has the eigenvalues σ_i^2 with

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0 = \sigma_{r+1}^2 = \cdots = \sigma_n^2$$

to which corresponds an orthonormal basis of eigenvectors $\{v_1, \ldots, v_n\}$ of A^*A. Let $V = [v_1|\cdots|v_n] = [V_1|V_2]$ where $V_1 = [v_1|\cdots|v_r]$, so that $V^*V = I$. Also, by Lemma B.1, $R(A^*A) = R(A^*)$, $N(A^*A) = N(A)$, and $R(A^*)^\perp = N(A)$. It follows that V_1 is a basis of $R(A^*)$ and V_2 a basis of $N(A)$.

With $\Sigma_1 = \text{diag}\{\sigma_1, \ldots, \sigma_r\}$, one sees that $A^*AV_1 = V_1\Sigma_1^2$. Let

$$U_1 := AV_1\Sigma_1^{-1}.$$

It follows that $U_1^*U_1 = I_r$. Moreover, $AA^*U_1 = U_1\Sigma_1^2$. Let $\{u_1, \ldots, u_r\}$ be the columns of U_1. We conclude that

$$u_i^*u_j = 1\{i = j\} \text{ and } AA^*u_i = \sigma_i^2 u_i.$$

But AA^* and A^*A have exactly the same r nonzero eigenvalues. Indeed, if $A^*Ay = \lambda y$ and $x = Ay$, then $AA^*x = AA^*Ay = \lambda Ay = \lambda x$, so that the eigenvalues of A^*A are eigenvalue of AA^*. By symmetry, A^*A and AA^* have the same eigenvalue. It follows that the columns of U_1 form an orthonormal basis for $R(AA^*) = R(A)$.

Define now an $m \times (m-r)$ matrix U_2 with orthonormal columns such that $U_2^*U_1 = 0$. Then $U = [U_1|U_2]$ is a unitary matrix with $\mathbb{C}^m = R(U_1) \oplus R(U_2)$. Since $\mathbb{C}^m = R(A) \oplus N(A^*)$ and $R(U_1) = R(A)$, it follows that $R(U_2) = N(A^*)$ with the columns of U_2 forming an orthonormal basis for $N(A^*)$. We conclude that the columns of U form a basis of eigenvectors of AA^*.

Finally, since $U_1 = AV_1\Sigma^{-1}$, we find that

$$U_1\Sigma_1V_1^* = AV_1\Sigma^{-1}\Sigma_1V_1^* = A,$$

and

$$U\Sigma V^* = [U_1|U_2]\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}V_1^*\\ V_2^*\end{bmatrix} = U_1\Sigma_1V_1^* = A,$$

which completes the proof.

□

B.9 Cayley-Hamilton

The Cayley-Hamilton Theorem states that a square matrix is a root of its characteristic polynomial.

Lemma B.2. *Cayley-Hamilton*
Let $A \in \Re^{m\times m}$ *and* $f(s) := \det(sI - A) = \alpha_0 + \alpha_1 s + \cdots + \alpha_{m-1}s^{m-1} + s^m$. *Then*

$$f(A) := \alpha_0 I + \alpha_1 A + \cdots + \alpha_{n-1}A^{m-1} + A^m = 0.$$

In particular, $span\{I, A, A^2, \ldots\} = span\{I, A, \ldots, A^{m-1}\}$.

Proof:
There are many proofs of this result. We give two proofs.
Proof 1
First assume that A can be diagonalized, i.e., such that $A = P\Lambda P^{-1}$ with $\Lambda = \text{diag}\{\lambda_i\}$ where the λ_i are the eigenvalues of A . Then $A^k = P\Lambda^k P^{-1}$ and $f(A) = P\text{diag}\{f(\lambda_i)\}P^{-1} = 0$.
Second, if A is not diagonalizable, one can find a sequence of matrices A_n that are and such that $A_n \to A$ componentwise. The idea to find these matrices A_n is to perturb the entries of A slightly so that A_n has distinct eigenvalues. (See Facts B.8-B.9.) We then have $0 = f(A_n) \to f(A)$, so that $f(A) = 0$.
Proof 2
Recall that the inverse of a matrix is the adjoint divided by the determinant. Thus, if λ is not an eigenvalue of A,

$$(A - \lambda I)^{-1} = \frac{B(\lambda)}{f(\lambda)}$$

where $B(\lambda)$ is a matrix of polynomials of degrees at most $n-1$ in λ. That is,

$$B(\lambda) = B_0 + B_1\lambda + \cdots + B_{n-1}\lambda^{n-1}.$$

Now, $(A - \lambda I)B(\lambda) = f(\lambda)I$, so that

$$(A - \lambda I)(B_0 + B_1\lambda + \cdots + B_{n-1}\lambda^{n-1}) = f(\lambda)I.$$

Thus, matching coefficients,

$$\begin{aligned} AB_0 &= \alpha_0 I \\ AB_1 - B_0 &= \alpha_1 I \\ AB_2 - B_1 &= \alpha_2 I \\ &\cdots \\ AB_{n-2} - B_{n-3} &= \alpha_{n-1} I \\ AB_{n-1} - B_{n-2} &= 0. \end{aligned}$$

Taking the first identity plus A times the second, plus A^2 times the third, and so on, we find

$$(AB_0)+A(AB_1-B_0)+A^2(AB_2-B_1)+\cdots+A^{n-2}(AB_{n-2}-B_{n-3})+A^{n-1}(AB_{n-1}-B_{n-2}) = f(A),$$

but the left-hand side collapses to 0.

□

B.10 Notes on MATLAB

MATLAB is a software package that provides many mathematical functions and visualization tools. It is particularly suited for matrix operations. We illustrate some useful functions. It is not a bad idea to become familiar with MATLAB, even though we will not require it for this version of the course.

B.10.1 Matrix Notation

The matrix

$$B = \begin{bmatrix} 3 & 4 & 2 \\ 5 & 1 & 3 \\ 1 & 4 & 2 \end{bmatrix}$$

is written in MATLAB as

```
B = [3 4 2; 5 1 3; 1 4 2]
```

The product of two matrices B and C with compatible dimensions is written as

```
B*C
```

B.10.2 Eigenvectors

We find the eigenvalues of matrix A as follows:

```
eig(A)
```

which returns

```
ans =

    2.4272
   -1.4272
```

To find the eigenvectors, we write

```
[V, D] = eig(A)
```

which returns

```
V =

    0.7718   -0.5809
    0.6359    0.8140


D =

    2.4272         0
         0   -1.4272
```

B.10.3 Inverse

The inverse of A is obtained as follows:

```
C = inv(A)
```

which returns

```
C =

         0    0.5000
    0.5774   -0.2887
```

and you can verify that

```
A*C

ans =

     1     0
     0     1
```

B.10.4 Singular Value Decomposition

Here is the code you need to calculate the singular value decomposition of matrix A in Section B.8.3.

```
[U, S, V] = svd([1 3^0.5; 2 0])
```

That function returns

```
U =

   -0.7071   -0.7071
   -0.7071    0.7071


S =

    2.4495         0
         0    1.4142


V =

   -0.8660    0.5000
   -0.5000   -0.8660
```

B.11 Exercises

Calculations can be done in MATLAB. However, it is useful to know the steps to understand what is going on. We solve some of the problems and we let you work on the others.

Exercise 6. *Calculate the inverse of the following matrices:*

$$\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 5 & 4 \\ 2 & 3 \end{bmatrix}, \begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & 3 \end{bmatrix}.$$

Solutions

$$\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix}, \begin{bmatrix} 5 & 4 \\ 2 & 3 \end{bmatrix}^{-1} = \frac{1}{7}\begin{bmatrix} 3 & -4 \\ -2 & 5 \end{bmatrix},$$

Let

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & 3 \end{bmatrix}.$$

The matrix C of cofactors of A is

$$C = \begin{bmatrix} 11 & 0 & -11 \\ 0 & -3 & 1 \\ -11 & 1 & 7 \end{bmatrix}.$$

The determinant of A can be found using Fact B.4. We get, by expanding over the first column,

$$det(A) = 2 \times 11 + 1 \times 0 + 3 \times (-11) = -11.$$

Finally, we get

$$A^{-1} = \frac{1}{det(A)} C' = -\frac{1}{11} \begin{bmatrix} 11 & 0 & -11 \\ 0 & -3 & 1 \\ -11 & 1 & 7 \end{bmatrix}.$$

Exercise 7. *Perform the elementary row operations to reduce the following matrices to row echelon form:*

$$A = \begin{bmatrix} 5 & 3 & 1 \\ 2 & 4 & 2 \\ 2 & 3 & 6 \end{bmatrix}, \begin{bmatrix} 2 & 3 & 1 & 2 \\ 4 & 4 & 2 & 3 \\ 2 & 3 & 6 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 3 & 1 & 2 \\ 4 & 6 & 2 & 3 \\ 1 & 2 & 1 & 2 \end{bmatrix}.$$

Solution: For matrix A, the first step is

$$r_2' := r_2 - \frac{2}{5} r_1 = [02.80.2].$$

The second step is to cancel the first two elements of r_3. The first element is canceled as follows:

$$r_3' = r_3 - \frac{2}{5} r_1 = [01.85.2].$$

We cancel the second element by subtracting a multiple of r_2':

$$r_3'' = r_3' - \frac{1.8}{2.8} r_2' = r_3 - 0.14 r_1 - 0.64 r_2 = [004.6].$$

Thus, we perform the operation

$$\begin{bmatrix} 1 & 0 & 0 \\ -0.4 & 1 & 0 \\ -0.14 & -0.64 & 1 \end{bmatrix} A \approx \begin{bmatrix} 5 & 3 & 1 \\ 0 & 2.8 & 1.6 \\ 0 & 0 & 4.6 \end{bmatrix}.$$

We let you perform the same steps for the other matrices.

Exercise 8. *Perform the elementary row operations and the elementary column operations step by step to explain that the determinant of the product of the following two matrices is the product of their determinants:*

$$A = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 5 & 4 \\ 3 & 3 \end{bmatrix}.$$

Solution: Note that

$$\det(AB) = \det(LABU')$$

where

$$L = \begin{bmatrix} 1 & 0 \\ -0.75 & 1 \end{bmatrix} \text{ and } U' = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

Indeed, elementary row operations LAB do not modify the determinant of AB. Also, elementary column operations $LABU'$ do not modify the determinant of LAB. Now,

$$LABU = (LA)(BU') = \begin{bmatrix} 4 & 3 \\ 0 & -0.25 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 25 \\ 0 & -0.75 \end{bmatrix}.$$

But,

$$\det(A) = \det(LA) = (4)(-0.25), \det(B) = \det(BU') = (1)(3).$$

Moreover,

$$\det((LA)(BU')) = (4)(-0.25 \times 3),$$

so that

$$\det(AB) = \det(A)\det(B).$$

The same argument goes for the other examples.

Exercise 9. *Repeat the previous exercise for the following two matrices:*

$$\begin{bmatrix} 4 & 2 & 3 \\ 2 & 3 & 5 \\ 2 & 4 & 7 \end{bmatrix} \text{ and } \begin{bmatrix} 5 & 2 & 1 \\ 1 & 3 & 1 \\ 3 & 2 & 6 \end{bmatrix}.$$

Exercise 10. *Find a basis for the span of the following vectors:*

$$(1,1,3,2)^*, (2,3,1,2)^*, (2,1,2,3)^*.$$

Exercise 11. *Find the rank and the dimension of the null space of the following matrices:*

$$\begin{bmatrix} 3 & 1 & 2 \\ 4 & 2 & 3 \\ 5 & 7 & 2 \\ 2 & 6 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 2 & 4 \\ 4 & 2 & 3 \\ 5 & 1 & 2 \\ 1 & 6 & 3 \end{bmatrix}, \text{ and } \begin{bmatrix} 1 & 1 & 2 \\ 4 & 2 & 3 \\ 5 & 3 & 2 \\ 3 & 5 & 0 \end{bmatrix}.$$

Exercise 12. *Find $x|_{\mathcal{V}}$ where $\mathcal{V} = span(V)$ for the following values of x and V:*

$$V: \begin{bmatrix} 3 & 1 & 2 \\ 4 & 2 & 3 \\ 5 & 7 & 2 \\ 2 & 6 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 2 & 4 \\ 4 & 2 & 3 \\ 5 & 1 & 2 \\ 1 & 6 & 3 \end{bmatrix}, \text{ and } \begin{bmatrix} 1 & 1 & 2 \\ 4 & 2 & 3 \\ 5 & 3 & 2 \\ 3 & 5 & 0 \end{bmatrix},$$

and

$$x = (1,3,1,3)^*, (2,1,4,2)^*, \text{ and } (2,2,1,1)^*.$$

Exercise 13 (Optional). *Use MATLAB to find the decomposition $H = P\Lambda P^*$ of the following Hermitian matrices:*

$$\begin{bmatrix} 4 & 2 & 3 \\ 2 & 1 & 5 \\ 3 & 5 & 7 \end{bmatrix} \text{ and } \begin{bmatrix} 5 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 6 \end{bmatrix}.$$

Solution
Here are the steps.

```
>> [V, D] = eig([4 2 3;2 1 5;3 5 7])

V =

   -0.0553    0.9009    0.4304
    0.8771   -0.1622    0.4520
   -0.4771   -0.4025    0.7813


D =

   -1.8455         0         0
         0    2.2997         0
         0         0   11.5458
```

Hence, $H = VDV'$.

Bibliography

[1] L. Breiman: *Probability.* Addison-Wesley, 1968.

[2] F. M. Callier and C. A. Desoer: *Linear System Theory.* Springer Verlag, 1991; Appendix Ab.

[3] R. Durrett: *Essentials of Stochatic Processes.* Springer, 1999.

[4] D. Freedman: *Statistical Models: Theory and Practice.* Cambridge University Press, 2005.

[5] R. G. Gallager: *Stochastic Processes: A Conceptual Approach.* August 20, 2001.

[6] F. Kelly: *Reversibility and Stochastic networks.* Springer 1979. http://www.statslab.cam.ac.uk/ frank/BOOKS/book/whole.pdf

[7] H. Kushner: *Introduction to Stochastic Control.* Holt, Rinehart and Winston, Inc., 1971.

[8] P. R. Kumar and P. Varaiya: *Stochastic Systems: Estimation, Identification, and Adaptive Control.* Prentice-Hall, 1986.

[9] B. Noble and J. W. Daniel: *Applied Linear Algebra.* Prentice Hall, 1997.

[10] J Walrand: *Lecture Notes on Probability anb Random Processes*, On-Line, August 2004.

[11] N. Wiener: Extrapolation, Interpolation, and Smoothing of Stationary Time Series. New York: Wiley, 1949.

Index