Chapter 1

Introduction

This is an introductory course in probability and random processes for imaging scientists. The purpose is to develop an understanding and ability in modeling noise and random processes within the context of imaging systems. The focus will be on stationary random processes in both one dimension (time) and two dimensions (spatial). Power spectrum estimation will be developed and applied to signal characterization in the frequency domain. The effect of linear filtering will be modeled and applied to signal detection and maximization of SNR. The matched filter and the Wiener filter will be developed. Signal detection and amplification will be modeled, using noise figure and SNR as measures of system quality. At completion of the course, the student should have the ability to model signals and noise within imaging systems.

The particular question of detection and measurement of quantum flux will be addressed through the mechanism of Poisson models. The concepts behind the performance of quantum-counting detectors will be developed and modeled. The nature of thresholds will be used to investigate saturation effects. Detective Quantum Efficiency (DQE) will be used to characterize the performance of detectors.

Repeated Bernouilli trials will be used to introduce the concept of random processes. This naturally extends to the Gaussian noise model and also links to the Poisson model. This combination of related concepts is very important in modeling the performance of imaging systems.

Random processes come in many flavors. We will concentrate on wide-sense stationary random processes because they are the most widely useful. Restricting our focus to this class enables us to cover a broad territory in the small amount of time that is available in this introductory course. There are many ways to model wss processes, with two major categories being parametric source models and spectral models. The parametric source models take the viewpoint that if one can describe a machine that could have generated the process, then one will know a lot about its output. The spectral model describes the observed power vs frequency distribution without reference to the source description. These approaches are, of course, closely related. We shall show the relationship with a variety of filtered white noise systems.

The detection of phenomena, namely signals, in noise is of great interest. An extension is the estimation of the parameters of signals in noise. We shall introduce the classical systems for the detection of signals in noise and the estimation of parameter values.

Most of the discussion is done within the framework of one-dimensional processes. This might seem limiting in an imaging systems context. However, many images are transmitted, analyzed and even displayed by transforming in to a one-dimensional form. Consider television as one prominent example. However, two-dimensional structures do have different properties. We will discuss some methods for the modeling of random fields, which are particularly useful in the consideration of geometric objects, textures and other patterns that are naturally two-dimensional.

These notes are undergoing continuous development, with additions and changes each time this course is taught. They are not considered to be finished by any means. However, they are intended to gather the lecture record in some sense and be a point of departure for students in the course.

Chapter 2

Probability Modeling

2.1 Introduction

Probability models are used extensively to describe the behavior and performance of systems. We will be particularly interested in applications to imaging systems, but the techniques are widely useful in science and engineering.

Consider the problem of designing a photon-counting detector. A simple model of reality postulates a photon flux of Φ photons per square meter per second. Then one would expect to collect $q = \Phi AT$ photons with a detector placed behind an aperture of size A that is kept open for an observation of time T. Imagine doing the photon counting experiment for a sequence of values for the parameter AT. The results may produce a plot like that shown in Figure 2.1. We note that as AT is increased the observation (normalized by plotting q/AT) seems to settle around a value close to 4. There is a random nature to the observations, but we may still be able to draw useful conclusions from them. We may even be able to make useful predictions about the performance of an imaging system. This is but one example of a situation in which we would like the ability to analyze and understand random behavior. The tools for this are based on the concepts of probability models.

Probability modeling is an abstract tool that can be applied to many different tasks by interpretation of its symbolism in the domain of interest. There are several possible viewpoints of the abstract model. In this course we will adopt the viewpoint of modeling an experiment.

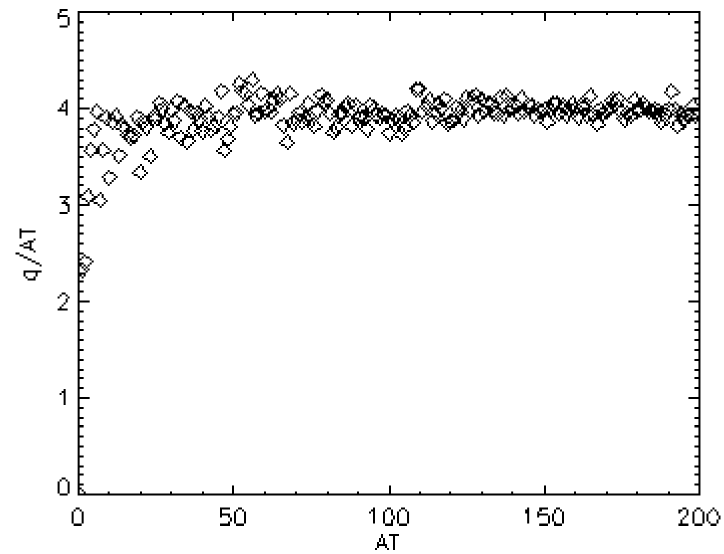

Figure 2.1: Observed values of photon counts divided by AT.

2.2 Structure of a Probability Model

An experiment is a process that can be done repeatedly and which will produce observable results. We will call each repetition a *trial*. The result of each trial may be different, even when the conditions are carefully controlled. Consider, for example, tossing a fair die and observing the top face. This is an experiment in which there are six possible results, corresponding to which of the faces is up. Note, however, that an outcome does not need to be a numerical quantity. Voting can be modeled as an experiment that selects between "yes" and "no," for example.

We will use the term *outcome* for each of the possible results. Each experiment has a set of possible outcomes, which we will denote by $\mathcal{U}$. For the die-tossing experiment the space of outcomes is $\mathcal{U} = \{e_i, i = 1, \ldots, 6\}$ corresponding to the six possible faces. On each trial exactly one element of $\mathcal{U}$ must occur.

It is common to refer to the set $\mathcal{U}$ as the *sample space* of the experiment. A trial is then equivalent to a selection from the sample space.

Suppose that a large number N trials of an experiment is to be conducted.

Is there any prediction that can be made about the observations? On the basis of experience or other information, we may at least have an expectation of the fraction of the results that correspond to each element of $\mathcal{U}$. In the example of the die-tossing experiment, let n_i be the number of times outcome e_i occurs in N trials. Then $n_1+n_2+n_3+\cdots+n_6 = N$. Upon dividing through by N we have

$$\frac{n_1}{N} + \frac{n_2}{N} + \frac{n_3}{N} + \cdots + \frac{n_6}{N} = 1 \tag{2.1}$$

Each of the terms is the fraction of the observations associated with the corresponding outcome. Denoting this fraction by $f_i = n_i/N$, we have the obvious result

$$f_1 + f_2 + f_3 + \cdots + f_6 = 1 \tag{2.2}$$

We expect that the fractions f_i are in some sense predictable when N is large enough. For a fair die, we expect each $f_i \approx 1/6$. The number that we expect the fraction to approach for large N is called the *probability*. We will denote the probability of the outcome e_i by p_i or $p(e_i)$.

A probability model contains the essential information needed to make predictions about expected observations of an experiment. The model consists of the sample space and the values of the corresponding probabilities. In this introduction we will assume that the sample space is finite. We will need to relax that restriction eventually, but we can go an amazing distance with the finite system.

Let L denote the number of elements in $\mathcal{U}$. The essential information about an experiment is contained in the values of the probabilities. These values must satisfy two requirements:

$$\begin{aligned} &p_i \geq 0 \text{ for all } i \\ &p_1 + \cdots + p_L = 1 \end{aligned} \tag{2.3}$$

The labels that we use for the outcomes are not particularly important. The real information is in the probabilities.

It is common to wonder whether a probability model is a good match to a real experiment. That depends on how well the chosen values of the probabilities match the observed outcome frequencies. However, except for uninteresting cases, impossible to *precisely* measure probabilities by observing outcome frequencies. The proper kind of question to ask is "If reality were to correspond to this model then what conclusions can be drawn?" We can use variations in the model to test the sensitivity of conclusions. The question of measuring probabilities is one of statistics. Here we take them as given. Our interest is in focused on developing techniques for their use in system analysis.

2.2.1 Events

We introduce the term *event* to enable us to talk about composite outcomes. In the die-tossing experiment, we may want to ask about the probability of an odd number on the upper face. An odd number is not a simple outcome. It consists of the outcomes $\{e_1, e_3, e_5\}$. If any of them turns up, then the odd-number event occurs. Simply put, an event is a set of possible outcomes. Any event $\mathcal{A}$ is then a subset of $\mathcal{U}$, and we can write $\mathcal{A} \subset \mathcal{U}$. The event $\mathcal{A}$ is said to occur if *any* of its outcomes occurs. The tools to talk about events are provided in a natural way by the language of sets.

In the die-tossing experiment, there is the odd set, the even set, the set of values less than four, and so on. The notion of events enables us to talk about more results that may be of interest in various situations. This train of thought brings up the question, "How many distinct sets are possible for an experiment with L outcomes?"

We can find out by a simple enumeration system. Construct a table with L columns, with column i headed by outcome e_i. Let each row correspond to a different event, with a 1 under each outcome that is included and a 0 for each outcome that is not included. Each row corresponds to an L-digit binary number. There are 2^L distinct L-digit binary numbers, so there are at most 2^L distinct events. Each binary pattern except the one with L zeros corresponds to a set of outcomes, and therefore is a legitimate event. For sake of completeness we introduce the null event ϕ to correspond to the all-zero pattern. Also note that the all-one pattern corresponds to the set of all outcomes, which is $\mathcal{U}$ itself. Commonly ϕ is called the impossible event and $\mathcal{U}$ is called the certain event.

The probability of any event $\mathcal{A} \subset \mathcal{U}$ is defined to be

$$p(\mathcal{A}) = \sum_{e \in \mathcal{A}} p(e) \tag{2.4}$$

This makes sense in terms of an experiment because $\mathcal{A}$ is the result if any one of the outcomes in the set occurs.

Example 2.2.1 *Determine all of the events and their probabilities for the experiment of tossing a fair die.*

Solution: There are $2^6 = 64$ *events as listed in Table 2.1. The probability of each event is the sum of the probabilities of each of the outcomes in that set. For a fair die each outcome is equally likely so that each event has a probability that is 1/6 times the number of outcomes in that set.*

2.2.2 Set Operations

It is often useful to have a language to describe compound events. If $\mathcal{A}$ and $\mathcal{B}$ are events, then it is convenient to be able to refer to the events such as: "either $\mathcal{A}$ or $\mathcal{B}$ occurred" or "$\mathcal{A}$ and $\mathcal{B}$ both occurred" or "$\mathcal{B}$ did not occur." Statements like these, and many that are much more complicated, can be stated accurately and succinctly by using set operations. The operations that are needed are the union, intersection and negation. Everything else can be constructed from these. Some examples are shown in Table 2.2

In Table 2.1 we see that $\mathcal{A}_7 = \{e_1, e_2, e_3\}$ and $\mathcal{A}_{21} = \{e_1, e_3, e_5\}$ so that $\mathcal{A}_7$ occurs when a number of 1,2 or 3 is up and $\mathcal{A}_{21}$ occurs when an odd number is up. The compound event $\mathcal{A}_7 \cup \mathcal{A}_{21} = \{e_1, e_2, e_3, e_5\}$ occurs when either an odd number or a low number appears. From the table we can find that $\mathcal{A}_7 \cup \mathcal{A}_{21} = \mathcal{A}_{23}$. Similarly, $\mathcal{A}_7 \cap \mathcal{A}_{21} = \{e_1, e_3\} = \mathcal{A}_5$ corresponds to the appearance of a low odd number.

The algebra of sets provides a very useful set of analytical tools for probability modeling. Below are listed some of the basic definitions and expressions.

Definition 2.2.1 *An* experiment *consists of a sample space,* $\mathcal{U} = \{e_i, i = 1, \ldots, L\}$ *of all possible outcomes. One and only one of the outcomes will occur on each trial of the experiment. Associated with each outcome is a number called the probability* $p(\dot{e}_i)$ *that satisfies* $p(\dot{e}_i) \geq 0$*, and* $\sum_{i=1}^{n} p(e_i) = 1$.

In applications the probabilities are associated with meaningful quantities, such as our expectation that each of the outcomes will occur. However, the definition in a purely mathematical sense does not require an association with any real experiment. The art of our analysis is to associate real and

n	$P(\mathcal{A}_n)$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_6$	**n**	$P(\mathcal{A}_n)$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_6$
0	0	0	0	0	0	0	0	32	1/6	0	0	0	0	0	1
1	1/6	1	0	0	0	0	0	33	2/6	1	0	0	0	0	1
2	1/6	0	1	0	0	0	0	34	2/6	0	1	0	0	0	1
3	2/6	1	1	0	0	0	0	35	3/6	1	1	0	0	0	1
4	1/6	0	0	1	0	0	0	36	2/6	0	0	1	0	0	1
5	2/6	1	0	1	0	0	0	37	3/6	1	0	1	0	0	1
6	2/6	0	1	1	0	0	0	38	3/6	0	1	1	0	0	1
7	3/6	1	1	1	0	0	0	39	4/6	1	1	1	0	0	1
8	1/6	0	0	0	1	0	0	40	2/6	0	0	0	1	0	1
9	2/6	1	0	0	1	0	0	41	3/6	1	0	0	1	0	1
10	2/6	0	1	0	1	0	0	42	3/6	0	1	0	1	0	1
11	3/6	1	1	0	1	0	0	43	4/6	1	1	0	1	0	1
12	2/6	0	0	1	1	0	0	44	3/6	0	0	1	1	0	1
13	3/6	1	0	1	1	0	0	45	4/6	1	0	1	1	0	1
14	3/6	0	1	1	1	0	0	46	4/6	0	1	1	1	0	1
15	4/6	1	1	1	1	0	0	47	5/6	1	1	1	1	0	1
16	1/6	0	0	0	0	1	0	48	2/6	0	0	0	0	1	1
17	2/6	1	0	0	0	1	0	49	3/6	1	0	0	0	1	1
18	2/6	0	1	0	0	1	0	50	3/6	0	1	0	0	1	1
19	3/6	1	1	0	0	1	0	51	4/6	1	1	0	0	1	1
20	2/6	0	0	1	0	1	0	52	3/6	0	0	1	0	1	1
21	3/6	1	0	1	0	1	0	53	4/6	1	0	1	0	1	1
22	3/6	0	1	1	0	1	0	54	4/6	0	1	1	0	1	1
23	4/6	1	1	1	0	1	0	55	5/6	1	1	1	0	1	1
24	2/6	0	0	0	1	1	0	56	3/6	0	0	0	1	1	1
25	3/6	1	0	0	1	1	0	57	4/6	1	0	0	1	1	1
26	3/6	0	1	0	1	1	0	58	4/6	0	1	0	1	1	1
27	4/6	1	1	0	1	1	0	59	5/6	1	1	0	1	1	1
28	3/6	0	0	1	1	1	0	60	4/6	0	0	1	1	1	1
29	4/6	1	0	1	1	1	0	61	5/6	1	0	1	1	1	1
30	4/6	0	1	1	1	1	0	62	5/6	0	1	1	1	1	1
31	5/6	1	1	1	1	1	0	63	1	1	1	1	1	1	1

Table 2.1: A table of the possible events in the die-tossing experiment. Event 0 is the impossible event ϕ and event 63 is the certain event $\mathcal{U}$. Event 28, for example, occurs whenever the top face shows 3, 4, or 5.

Compound Event	**Expression**
either $\mathcal{A}$ or $\mathcal{B}$ occurred	$\mathcal{A} \cup \mathcal{B}$
$\mathcal{A}$ and $\mathcal{B}$ both occurred	$\mathcal{A} \cap \mathcal{B}$
$\mathcal{B}$ did not occur	$\mathcal{B}^c$

Table 2.2: Illustrations of the use of set expressions to express compound events.

conceptual experiments in a manner that will enable us to make useful predictions about reality. In this regard, the restriction of a sample space to a finite number of points is not a restriction. We will always be restricted to some finite set of possible outcomes in reality.

Definition 2.2.2 *An* event *$\mathcal{A}$ is a set of outcomes. Associated with $\mathcal{A}$ is a probability $p(\mathcal{A}) = \sum_{e \in \mathcal{A}} p(e)$ formed by summing the probabilities of the outcomes in the set $\mathcal{A}$.*

An event may contain zero, one or more outcomes. The event that contains zero outcomes is referred to with the symbol ϕ, which has probability $p(\phi) = 0$. It is called the *impossible event*, and is included in our language because it is helpful to have a term for it. The event $\mathcal{A} = \mathcal{U}$ is called the *certain event* and has probability $p(\mathcal{U}) = 1$.

Definition 2.2.3 *Two events $\mathcal{A}$ and $\mathcal{B}$ are equal, written $\mathcal{A} = \mathcal{B}$, when they consist of identical elements. Any $a \in \mathcal{A}$ is also in $\mathcal{B}$ and any $b \in \mathcal{B}$ is also in $\mathcal{A}$. If $\mathcal{A} = \mathcal{B}$ then $P(\mathcal{A}) = P(\mathcal{B})$.*

Definition 2.2.4 *An event $\mathcal{A}$ is a* subset *of $\mathcal{B}$, written $\mathcal{A} \subset \mathcal{B}$, if $a \in \mathcal{A}$ implies $a \in \mathcal{B}$. That is, every element of $\mathcal{A}$ is and element of $\mathcal{B}$.*

One can easily show that $\mathcal{A} \subset \mathcal{B}$ implies $P(\mathcal{A}) \leq P(\mathcal{B})$. Why must this be true?

Definition 2.2.5 Union: *The union $\mathcal{A} \cup \mathcal{B}$ is the event that consists of all outcomes in either $\mathcal{A}$ or $\mathcal{B}$ or both.*

Since $\mathcal{A} \subset \mathcal{A} \cup \mathcal{B}$, $P(\mathcal{A}) \leq P(\mathcal{A} \cup \mathcal{B})$.

Definition 2.2.6 Intersection: *The intersection $\mathcal{A} \cap \mathcal{B}$ is the event that consists of all outcomes in both $\mathcal{A}$ and $\mathcal{B}$.*

Since $\mathcal{A} \cap \mathcal{B} \subset \mathcal{A}$, $P(\mathcal{A} \cap \mathcal{B}) \leq P(\mathcal{A})$.

Definition 2.2.7 *If $\mathcal{A} \cap \mathcal{B} = \phi$ then $\mathcal{A}$ and $\mathcal{B}$ are said to be* disjoint *or* mutually exclusive*. In that case, $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$.*

When the sets are not disjoint then the probabilitiy $P(\mathcal{A} \cup \mathcal{B})$ must account for outcomes in both sets. Any element $s \in \mathcal{A} \cap \mathcal{B}$ contributes to both $P(\mathcal{A})$ and $P(\mathcal{B})$. Hence, we have to subtract the probability of those elements that are in the common set. Then for any sets $\mathcal{A}$ and $\mathcal{B}$, whether disjoint or not, we have the result

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B}) \tag{2.5}$$

The expression simplifies to $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$ when $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive.

Definition 2.2.8 *The* complement $\mathcal{B}^c$ *of a set $\mathcal{B}$ is the set of all elements not in $\mathcal{B}$.*

For any set $\mathcal{B}$, $\mathcal{B}$ and $\mathcal{B}^c$ are mutually exclusive. Moreover, every outcome is in either $\mathcal{B}$ or $\mathcal{B}^c$ so that

$$\mathcal{B} \cup \mathcal{B}^c = \mathcal{U} \tag{2.6}$$

Since $P(\mathcal{U}) = 1$, we always have

$$P(\mathcal{B}^c) = 1 - P(\mathcal{B}) \tag{2.7}$$

This expression says that the probability of an even not happening is one minus the probability that it does happen on a trial of an experiment. That makes good sense.

A very useful pair of results from set theory are called *De Morgan's rules*. If $\mathcal{A}$ and $\mathcal{B}$ are arbitrary events then

$$(\mathcal{A} \cap \mathcal{B})^c = \mathcal{A}^c \cup \mathcal{B}^c \tag{2.8}$$

$$(\mathcal{A} \cup \mathcal{B})^c = \mathcal{A}^c \cap \mathcal{B}^c \tag{2.9}$$

It is interesting to note that (2.9) can be obtained from (2.8) simply by exchanging the operations of union and intersection. This turns out to be a particular case of a general rule that follows from De Morgan's rules, called the *principle of duality.* Any true identity formed of unions, intersections and complements of events remains true if in it the symbols $\cup$, $\cap$, $\subset$, $\mathcal{U}$ and ϕ are replaced by $\cap$, $\cup$, $\supset$, ϕ, and $\mathcal{U}$. The operations of equality and complementation are left unchanged.

Definition 2.2.9 Partition: *It is often useful to split $\mathcal{U}$ into a number of disjoint subsets. We refer to a collection of sets $\mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_n$ as a partition if they are disjoint and cover the space.*

$$\mathcal{U} = \bigcup_{i=1}^{n} \mathcal{A}_i \tag{2.10}$$

$$\mathcal{A}_i \cap \mathcal{A}_j = \phi \quad \text{if } i \neq j \tag{2.11}$$

In the exercises you are asked to show that the above partition of $\mathcal{U}$ is useful in partitioning any set $\mathcal{B}$. If $\mathcal{B}_i = \mathcal{B} \cap \mathcal{A}_i$ then $\mathcal{B} = \bigcup_{i=1}^{n} \mathcal{B}_i$ and $\mathcal{B}_i \cap \mathcal{B}_j = \phi$ if $i \neq j$.

2.3 Conditional Probability

In an experiment we often want to describe the relationships between events. If you know that an event $\mathcal{A}$ has occurred then what do you know about the possibility that $\mathcal{B}$ has also occurred? How do you make use of partial information? The essential relationships are expressed in terms of conditional probabilities.

If one knows that an event $\mathcal{A}$ has occurred then the set of possible outcomes has been changed from $\mathcal{U}$ to $\mathcal{A}$. As shown in Figure 2.2, the outcomes that are now associated with the occurrence of $\mathcal{B}$ are changed to those that are in the intersection, $\mathcal{A} \cap \mathcal{B}$. The probability of $\mathcal{B}$ conditioned on the hypothesis of $\mathcal{A}$ is the ratio of the probability of the outcomes in $\mathcal{A} \cap \mathcal{B}$ to the outcomes in all of $\mathcal{A}$.

$$P(\mathcal{B}|\mathcal{A}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{A})} \tag{2.12}$$

provided $P(\mathcal{A}) > 0$.

As an example, let $\mathcal{A}$ correspond to the event that an even-numbered face shows on the toss of a fair die, and let $\mathcal{B}$ be the event that the number 4, 5 or 6 shows. Then $P(\mathcal{A}) = 1/2$ and $P(B) = 1/2$. The joint event $\mathcal{A} \cap \mathcal{B}$ is the event that 4 or 6 shows, so that $P(\mathcal{A} \cap \mathcal{B}) = 1/3$. The probability of an even number given $\mathcal{A}$ is the probability that the face was 4 or 6 given that it was 4 or 5 o4 6. For a fair die, this must be 2/3. From (2.12) we calculate

$$P(\mathcal{B}|\mathcal{A}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{A})} = \frac{1/3}{1/2} = \frac{2}{3}$$

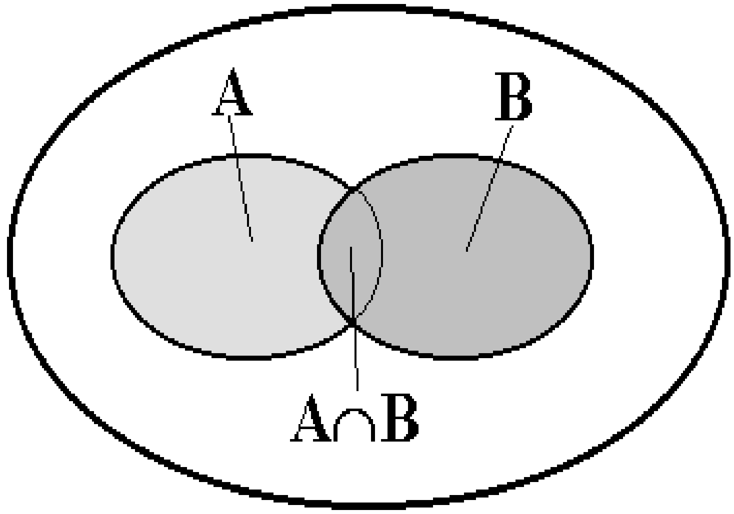

Figure 2.2: A Venn diagram showing sets $\mathcal{A}$ and $\mathcal{B}$ that are involved in the conditional probability $P(\mathcal{B}|\mathcal{A})$.

Equation (2.12) can be written in the form $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B}|\mathcal{A})$. Since the left side is symmetric in $\mathcal{A}$ and $\mathcal{B}$, it must be the case that they can be interchanged on the right side. Hence we get the pair of expressions

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B}|\mathcal{A}) = P(\mathcal{B})P(\mathcal{A}|\mathcal{B}) \tag{2.13}$$

This provides a useful set of relationships that can be used to compute one of the terms when the others are known. They are widely used in probability analysis.

Theorem 2.3.1 *Suppose events $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ partition a sample space $\mathcal{U}$, and that $P(\mathcal{A}_i) > 0$ for all $i = 1, 2, \ldots, n$. Then for any event $\mathcal{B}$ in $\mathcal{U}$*

$$P(\mathcal{B}) = \sum_{i=1}^{n} P(\mathcal{B}|\mathcal{A}_i)P(\mathcal{A}_i) \tag{2.14}$$

The proof of the theorem is left as an exercise. This result leads to the famous Bayes' rule[1].

[1] Named after the English philosopher Reverend Thomas Bayses (1702-1764).

Theorem 2.3.2 *Let $\mathcal{B}$ be an event in sample space $\mathcal{U}$. Suppose that the events $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ partition $\mathcal{U}$ and that $P(\mathcal{A}_i) > 0$ for all $i = 1, 2, \ldots, n$. Then*

$$P(\mathcal{A}_j|\mathcal{B}) = \frac{P(\mathcal{B}|\mathcal{A}_j)P(\mathcal{A}_j)}{\sum_{i=1}^{n} P(\mathcal{B}|\mathcal{A}_i)P(\mathcal{A}_i)} \tag{2.15}$$

The proof, which follows directly from (2.13) and (2.14) is left as an exercise. Bayes' rule is one of the most frequently-used analytical tools. It is particularly useful in analyzing the probability of a cause given the observation of a given effect.

Example 2.3.1 Observation in additive noise: *Suppose that you have an instrument that can measure brightness in steps of one unit from a level of 20 to 200 units. Suppose that any reading you make can be in error by as much as ± 2 units, with the error distribution as shown below.*

e	-2	-1	0	1	2
$P(e)$	1/5	1/5	1/5	1/5	1/5

A source of brightness b will be observed as a number $x = b + e$. Suppose that you are looking at a field of point sources that have two different brightness levels, b_1 and b_2. You can focus on each source and read a number x, and must decide whether it is of type 1 or type 2. Under the hypothesis b_i the observation will have the distribution $P(x|b_i)$. In this example let us take $b_1 = 99$ and $b_2 = 101$. Then we can form the following table.

x	97	98	99	100	101	102	103
$P(x|b_1)$	1/5	1/5	1/5	1/5	1/5	0	0
$P(x|b_2)$	0	0	1/5	1/5	1/5	1/5	1/5

This table is formed by noting that if the brightness has a certain value then the observation must be that value plus the error, and we know the distribution of errors. Now consider how to use this table to classify an observation. When $x = 98$, for example, it appears that we should decide that the source was b_1 because that x would be impossible if the source was b_2. Using this reasoning leads us to a decision rule for $x = 97$, 98, 102, 103. But it is not clear what to do for $x = 99$, 100 and 101. To go further we need to use more information. Suppose that the field is known to contain 20% of sources of type 1 and 80% of type 2. That would bias our decision toward type 2 whenever there was any doubt. Bayes' rule gives us a way to do the analysis

in a systematic fashion.
We can first construct the joint probabilities $P(x \cap b_1)$ *and* $P(x \cap b_2)$ *by using (2.13) with the* a priori *probabilities* $P(b_1) = 0.2$ *and* $P(b_2) = 0.8$. *This enables to construct the following table:*

x	97	98	99	100	101	102	103
$P(x \cap b_1)$	.04	.04	.04	.04	.04	0	0
$P(x \cap b_2)$	0	0	.16	.16	.16	.16	.16
$P(x)$	.04	.04	.2	.2	.2	.16	.16

The bottom row is just the sum of the two rows above. This is an application of (2.14).
We can now apply Bayes' rule to find the probabilities $P(b_1|x)$ *and* $P(b_2|x)$. *This leads to the next table:*

x	97	98	99	100	101	102	103
$P(b_1|x)$	1	1	.2	.2	.2	0	0
$P(b_2|x)$	0	0	.8	.8	.8	1	1

We have now computed the probability of each of the possible causes given each of the possible observations. This is exactly what is needed to minimize the number of decision errors. The decision rule is given by the table, in which an entry of 1 corresponds to "true" and 0 corresponds to "false."

x	97	98	99	100	101	102	103
Decide b_1	1	1	0	0	0	0	0
Decide b_2	0	0	1	1	1	1	1

A correct decision is made whenever the source and observation actually correspond to the cells with a 1. The probability of a correct decision can be calculated by multiplying the cells of the decision table by the corresponding cells of the joint probability table and summing. The result for this example is

$$P(C) = .04 + .04 + .16 + .16 + .16 + .16 + .16 = 0.88$$

The probability of an erroneous decision is

$$P(E) = P(C^c) = 1 - P(C) = 0.12$$

By knowing the noise characteristics of the detector and assuming a distribution of sources we were able to construct a minimum-error decision rule and to compute the probability of decision error. You will repeat this example in the exercises with a different noise and source distribution.

The above example is an illustration of the usefulness of conditional probabilities. Among the primary uses is the ability to construct of decsion rules that are match to actual situations.

2.4 Statistical Independence

Conditional probability gives us a way to quantify the linkage between events. According to (2.9) the probability that any two events $\mathcal{A}$ and $\mathcal{B}$ both occur in the outcome of an experiment is $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B}|\mathcal{A})$. The probability $P(\mathcal{B}|\mathcal{A})$ is the probability of event $\mathcal{B}$ if one already has observed $\mathcal{A}$, and it may be greater than, less than, or equal to $P(\mathcal{B})$. You are asked to probe this issue in Exercise 12. Here we will examine the case $P(\mathcal{B}|\mathcal{A}) = P(\mathcal{B})$. That means that the probability of event $\mathcal{B}$ does not change because of the observation of event $\mathcal{A}$. If $\mathcal{A}$ and $\mathcal{B}$ are statistically independent then $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$. The concept of statistical independence is sufficiently important to warrant a definition. Because of the elegant symmetry in the equation above it is common to use it rather than the conditional probability in the definition. Of course, they are completely equivalent.

Definition 2.4.1 *Two events $\mathcal{A}$ and $\mathcal{B}$ are statistically independent if*

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B}) \tag{2.16}$$

The concept of statistical independence is of great importance in experiments which consist of a number of sub-experiments. Suppose that a number of photons are impinging on a detector array. If you knew the fate of one of the photons would you have useful information about the fate of any of the others? Suppose an experiment consists of the tossing of two fair dice. Does knowledge of the top face of the first die tell you anything about the top face of the second? Suppose that you are scanning a page to be transmitted by a fax machine. Does that event that a cell is white imply anything about the color of the next cell? If so, how can it be exploited (leading to compression of the binary code)? We will often use the tactic of constructing an experiment from sub-experiments. We will find that the case of statistical independence to have quite different properties from the case with statistical dependence.