Wickens, T. D. (2001). Elementary signal detection theory. Oxford University Press.

Chapter 1

The signal-detection model

Signal-detection theory provides a general framework to describe and study decisions that are made in uncertain or ambiguous situations. It is most widely applied in psychophysics—the domain of study that investigates the relationship between a physical stimulus and its subjective or psychological effect—but the theory has implications about how any type of decision under uncertainty is made. It is among the most successful of the quantitative investigations of human performance, with both theoretical and practical applications. This chapter introduces the basic concepts of signal-detection theory, in the context of simple yes or no decisions. These ideas are expanded and amplified in later chapters.

1.1 Some examples

A good place to begin the study of signal-detection theory is with several examples of detection situations. Consider a very common situation: an individual must decide whether or not some condition is present. Such decisions are easy to make when the alternatives are obvious and the evidence is clear. However, there are many tasks that are not so simple. Often, the alternatives are distinct, but the evidence on which the decision is to be based is ambiguous. Here are three examples:

- A doctor—physician, psychologist, or whoever—is examining a patient and trying to make a diagnosis. The patient shows a set of symptoms, and the doctor tries to decide whether a particular disorder is present or not. To complicate the problem, the symptoms are ambiguous, some of them pointing in the direction of the disorder, others pointing away from it; moreover, the patient is a bit confused and does not describe the symptoms clearly or consistently. The correct diagnosis is not obvious.

- A seismologist is trying to decide whether to predict that a large earthquake will occur during the next month. As it was for the doctor, the evidence is complex. Many bits of data can be brought to bear on the decision, but they are ambiguous and sometimes contradictory: records of earthquake activity in the distant past (some of which are not very accurate), seismographic records from recent times (whose form is clear, but whose interpretation is not), and so forth. The alternatives are obvious: either there will be a quake during the month or there will not. How to make the prediction is unclear.
- A witness to a crime is asked to identify a suspect. Was this person present at the time of the crime or not? The witness tries to remember the event, but the memory is unclear. It was dark, many things happened at once, and they were stressful, confusing, and poorly seen. Moreover, the crime occurred some time ago, and the witness has been interviewed repeatedly and has talked to others about the event. Again the alternatives are clear: either the person was present or not, but the information that the witness can bring to bear on the decision is ambiguous.

These three examples have an important characteristic in common. Although the basic decision is between simple alternatives—in each case, the response[1] is either YES or NO—the information on which that decision is based is incomplete, ambiguous, and frequently contradictory. These limitations make the decision difficult. No matter how assiduous the decision maker, errors will occasionally occur.

A full understanding of the decision process in situations such as these three is difficult, if not impossible. The decisions depend on many particulars: what the decider knows, his or her expectations and beliefs, how later information affects the interpretation of the original observations, and the like. An understanding of much domain-specific knowledge is needed—of medical diagnosis or seismology, for example. This base of knowledge must be coupled with a theory of information processing and problem solving. Fortunately, much can be said about the decision process without going into these details. Some characteristics of a yes-no decision transcend the particular situation in which it is made. The theory discussed here treats these common elements. This theory, known generally as *signal-detection theory*, is one of the greatest successes of mathematical psychology.

[1] Throughout this book, SMALL CAPITALS are used to denote overt responses.

Although signal-detection theory has implications for the type of complex decisions described above, it was developed for a much simpler situation: the detection of a weak signal occurring in a noisy environment. It is much easier to describe the theory in the simpler context. Consider two experiments:

- *Classical signal detection.* A person, referred to as the *observer*, is listening through earphones to white noise.[2] The observer also is watching a small light. At regular intervals the light goes on for one second. During this one-second interval one of two things happens: either the white noise continues as before, or a faint 523 Hz tone (middle C on the musical scale) is added to the noise. Whether the tone is added or not is decided randomly, so the observer has no way to tell in advance which event will occur. At the end of the one-second interval the stimulus returns to normal (if it had changed). The observer decides whether the tone was presented during the lighted interval and reports the choice by pressing one of two buttons, the left button if no tone was heard, the right button if a tone was heard. Because the tone is weak and the background noise loud, it is easy for the observer to make an error in detecting the tone.
- *Recognition memory.* The subject in this recognition-memory experiment is seated in a darkened room and views a series of slides projected on a screen. The slides are photographs of various outdoor scenes—trees, waterfalls, buildings, and so forth. Each slide is shown for three seconds, providing time to get a look at it but not to study it carefully. After 200 slides have been seen, the subject works on an unrelated task for half an hour, then returns to the room and is given a memory test. A new series of slides is shown. Half of these slides are old pictures, drawn from the 200 that were seen before. The other half are new pictures, generally similar to the old slides, but of novel scenes. For each of these slides the subject is asked to decide whether it had been seen in the first series or not.

If one overlooks the particular differences—perception or memory, audition or vision, meaningless tones or meaningful pictures—these two tasks are much alike. In each, a well-defined yes-no decision is made (tone present or absent in one case, picture old or new in the other), in each the signal that the observer is trying to detect is weak (the faint tone or the memory of a briefly seen picture), and in each the task is complicated by interfering information (the white noise or the competing memories of similar pictures

[2] This type of sound is a random mixture of a wide range of frequencies. It is called "white" by analogy with white light, which is a mixture of all colors. White noise sounds much like the hissing noise one hears from a radio that is not tuned to any station.

seen in the past). The common structure lets the two tasks be analyzed in similar ways.

1.2 Hits and false alarms

To start out, some terminology is needed. The observer in a detection experiment experiences two types of trials. On some trials only the random background environment is present, either the white noise or the random familiarity of a new picture. Because they contain no systematic component, these trials are called *noise trials*. On other trials, some sort of signal is added to the noise. In the examples, the signal is either the tone or the added familiarity associated with having recently seen a picture. These trials are called *signal plus noise trials* or, more simply, *signal trials*. When speaking of the recognition memory experiment, these possibilities correspond to *new items* and *old items*, respectively. In the discussion below, the signal-and-noise formulation will be used. Whichever terminology is used, the observer makes either the response YES or the response NO.

At first glance, it seems easy to score experiments of this type. Just measure how well the observer does by finding the proportion of times that the signal is detected. The event in question, saying YES to a signal, is known as a *hit*. The proportion of hits is calculated by dividing the frequency of hits by the frequency of signal trials to give the *hit rate*:

$$h = \frac{\text{Number of YES responses to signals}}{\text{Number of signal trials}}.$$

Good observers or observers presented with easy signals have high hit rates and poor observers or ones presented with difficult signals have low hit rates.

Unfortunately, the hit rate is incomplete as a summary of these experiments and is not a good way to indicate how well the signal is detected. It depends, in part, on aspects of the task other than the detectability. Consider one observer in two sessions of the tone detection experiment. During each session the observer is presented with 100 trials on which a pure noise stimulus is present and 100 trials on which the tone is added to the noise in an appropriate random sequence. The same signal is used in both sessions, so that the ability of the observer to detect it should be the same (at least as long as there are no effects of practice). The sessions differ in the instructions given to the observer. In the first session, the observer is told that it is very important to catch all the signals. To give incentive, the observer is paid 10 cents for every correctly detected signal. When the data for the session are analyzed, 82 hits have been made on 100 signal trials, so the proportion of hits is $h = 0.82$. The second session has different instructions. Now the observer is told that it is very important not to report a signal when there isn't one, a type of error known as a *false alarm*. To emphasize these instructions, the bonus for hits is removed, and the observer receives 10 cents for every correctly identified noise trial. There are far fewer hits in this session: only 55 of the 100 signal trials receive YES responses, giving $h = 0.55$. These results suggest that the proportion of hits is unsatisfactory, or at least incomplete, as a measure of the properties of the signal. Were it really a good measure, its value would be the same in both sessions—remember that the signal itself does not change. However, the proportion of hits drops from 0.82 to 0.55.

It is easy to see informally what happened here. The signals were weak and easily confused with the noise. On some trials this confusion made the observer unsure what to answer. In this state of uncertainty, the behavior of the observer in the two sessions is likely to differ. In the first session, to maximize the number of hits, the observer does best to say YES when unsure. After all, if that choice is correct, then the reward is received, while if it is wrong, nothing bad happens. In contrast, during the second session it is best to say NO when uncertain. With this response, if there really were no signal, then the reward would be received, while if there were a signal, nothing bad would happen. The detectability of the signal has not changed, but manipulating the payoff has made the observer change strategy, altering the proportion of hits. A better measure of detectability should not be affected by these strategic matters.

As this example shows, the problem with an analysis based only on the hit rate is that it neglects what happens on trials where the noise stimulus is presented. A complete picture requires attention to both possibilities. There are two types of trial, either noise or signal, and there are two responses, either NO or YES. Because each type of response can occur with each type of stimulus, there are four possible outcomes. These four possibilities are identified by name:

		Response	
		NO	YES
Trial Type	Noise	Correct Rejection	False Alarm
	Signal	Miss	Hit

Two of these possibilities, hits and correct rejections, are correct; the other two, false alarms and misses, are errors.

An idea of what is lost by looking only at the hits is seen by examining the complete data for the experiment described above. Suppose that in the two sessions the frequencies of the four events were

First Session

	NO	YES
Noise	54	46
Signal	18	82

and

Second session

	NO	YES
Noise	81	19
Signal	45	55

The decrease in hits between the first and the second session here is matched by a decrease in false alarms. Because the observer was rewarded for hits in the first session and for correct rejections (i.e., for avoiding false alarms) in the second session, this behavior is quite appropriate.

Parenthetically, note that this observer does not behave in a way that maximizes the earnings. There was no reward for correct rejections or penalty for false alarms in the first session, so the observer could earn the maximum possible amount money by saying YES on every trial. Likewise, the highest paying strategy in the second session is always to say NO. Had these strategies been used, the two tables of data would have appeared

First Session

	NO	YES
Noise	0	100
Signal	0	100

and

Second session

	NO	YES
Noise	100	0
Signal	100	0

However, most observers are loath to adopt such extreme strategies, and data of the type shown earlier, which are biased in the correct direction but less exaggerated, are more common.

Although there are four types of outcome here, one does not need four numbers to summarize the observer's behavior. In these detection experiments, the frequencies of the two types of trials—that is, of noise trials and signal trials—are determined by the experimenter. The observer's behavior governs the proportion of YES and NO responses on trials of a single type, not how many trials there are. When describing the observer's behavior, one does not want to summarize the data in a way that partially reflects the experimenter's behavior. So the results are usually converted to conditional proportions in each row. Conventionally in signal-detection theory, the two probabilities used are the *hit rate*,

$$h = \frac{\text{Number of hits}}{\text{Number of signal trials}},$$

and the *false-alarm rate*,

$$f = \frac{\text{Number of false alarms}}{\text{Number of noise trials}}.$$

One can also calculate a miss rate and a correct rejection rate, but these

quantities are redundant with the hit rate and the false-alarm rate:

$$\text{miss rate} = 1 - h,$$
$$\text{correct rejection rate} = 1 - f.$$

These two proportions add no new information to that provided by the hit rate and false-alarm rate, so it has become conventional to report only h and f. Using these two statistics, the data from the two sessions are summarized as

	h	f
First Session	0.82	0.46
Second Session	0.55	0.19

The way that the hit rate and the false-alarm rate increase or decrease together between the two sessions is quite apparent. This table contains the minimum information that must be reported to understand what is happening in a pair of two-alternative detection experiments.

The ideas expressed here have been developed in many domains, and with them other terminology. A more neutral nomenclature refers to hits and false alarms as *true positives* and *false positives*, and to correction rejections and misses as *true negatives* and *false negatives*. In the epidemiological literature, the *sensitivity* of a test is its hit rate, and the *specificity* is its correct-rejection rate.

1.3 The statistical decision representation

Although reporting both the hit rate and the false-alarm rate is much better than presenting one of these alone, even together they are not completely satisfactory. Together the two proportions give an idea of what the observer is doing, but neither number unambiguously measures the observer's ability to detect the signal. Neither the hit rate nor the false-alarm rate tells the whole story. A single number that represents the observer's sensitivity to the signal is better. The theories discussed in this book provide this measure.

To develop a measure of sensitivity, it is necessary to go beyond a simple description of the data. A measure that describes the detectability of a signal must be based on some idea of how the detection process works as a whole. A conceptual picture of the detection process, known broadly as *signal-detection theory*, has been developed to provide this structure. There are several ways that this broad picture can be made more specific and given a rigorous mathematical form. Each of these signal-detection *models*, as they will be called, leads to a different way to measure detectability. The simplest of these models describes an observer's performance by a pair of numerical quantities, one that measures the detectability of the signal and another that measures the observer's preference for YES or NO responses. Using this model, the data quantities h and f are translated into estimates of more meaningful theoretical quantities. This section describes one model for the two-alternative case; it is developed, extended, and modified in later chapters.

The basic model is drawn from statistical decision theory and is similar to the ideas that are used in statistical testing to make a decision between two hypotheses. This theory, as it is applied in signal-detection theory, is founded on three assumptions:

1. The evidence about the signal that the observer extracts from the stimulus can be represented by a single number.
2. The evidence that is extracted is subject to random variation.
3. The choice of response is made by applying a simple decision criterion to the magnitude of the evidence.

The next few paragraphs describe these assumptions in more detail.

First, consider the underlying dimension. The idea here is that the internal response of the observer to a stimulus, insofar as the detection decision is concerned, can be represented by a point on a single underlying continuous dimension. For example, when the signal is a pure tone of known frequency masked in white noise, this dimension might be the output of whatever neuron (or set of neurons) in the auditory system is maximally responsive to the appropriate signal frequency. For the recognition memory experiment, this dimension is some feeling of "familiarity" with the test word, perhaps as contrasted with a feeling for how familiar a new word would be. A more abstract definition is provided by the concept of a *likelihood ratio*, drawn from statistical theory, which is developed in Chapter 9.

In both examples a much more thorough analysis of specific properties of the detection system is possible. One could look (as many researchers have done) at the detailed neurology of tone detection or the various visual and semantic dimensions that influence recognition memory. However, this detail is unnecessary when seeking a detection measure. One of the powers of signal-detection theory is its ability to give useful quantitative answers without requiring one to delve into the specifics. Under the signal-detection model, the response depends on a single value, and all information obtained by the observer is summarized in this number. Thus, the theory can apply both to the complex medical or seismological examples at the start of this chapter and to the simpler perceptual or mnemonic studies.

Another important point drawn from statistical decision theory is embodied in the second assumption. Trials differ in their effect, even when the nominal stimulus is the same. Sometimes this variability has a physical interpretation. For the tone-detection experiment, the white noise is a random mixture of energy at all different frequencies; thus on a given noise trial the output of a physical detector tuned to 523 Hz (if that is the location of the signal) will be larger or smaller, depending on the accidental composition of the noise. On signal trials the added tone gives an increment to the output of the detector, but the variability attributable to the noise is still there. In other situations the variability, although no less present, lacks this clear physical interpretation. In the recognition memory experiment a given picture evokes a greater or lesser feeling of familiarity, depending on how often similar pictures have been seen in the past and under what circumstances, and, for old pictures, how attentive the subject was during the original presentation. Even if the physical basis for this variability is less apparent, its effect is no less true. Sometimes a new picture seems very familiar; sometimes an old one seems less so.

The variation of the stimulus effect is well illustrated by a diagram that shows the distributions of evidence under the two alternatives (Figure 1.1). The horizontal axis is the single dimension on which the internal response to a stimulus is measured, and the vertical height of the line indicates how likely that value of the evidence is to occur. The top curve, or distribution of evidence, describes the internal response when a noise-only trial is given, and the bottom curve describes the response to the signal-plus-noise stimulus. The two distributions are not identical, indicating that the observations are, to some degree at least, sensitive to the signal—if they were identical, then there would be no way to tell the two events apart. The presence of a signal changes both the center and the spread of the distribution, both its mean and its standard deviation in statistical terms. Most important, the signal distribution is shifted to the right relative to the noise distribution. On average, larger values are observed when a signal is present than when it is absent. However, because the distributions overlap, some noise trials produce a larger observation than some signal trials. For example, both the noise distribution and the signal distribution have positive values at the point marked x in Figure 1.1, implying that there is some chance that evidence of that strength is observed, whichever type of stimulus is presented. This ambiguity means that no response rule based on this dimension can produce completely error-free performance.

The two distributions drawn in Figure 1.1 differ in other ways than in the mean—here the distribution when the signal is present is more spread out then when it is absent. This difference is not a necessary part of the signal-detection model. In some versions of the theory, such as that discussed

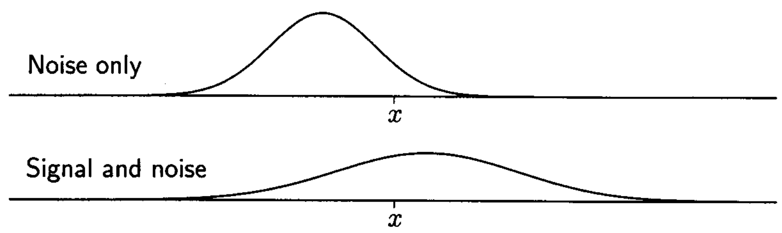


Figure 1.1: Probability distributions on the evidence dimension for a noise stimulus (upper distribution) and for a signal-plus-noise stimulus (lower distribution). The value x represents a particular observation, which could have come from either distribution.

in Chapter 2, the two distributions differ only in their center. In other version, such as those discussed later in this book, variability differences are also possible. Methods for testing which possibility is most appropriate are described in Chapter 3.

The third assumption of signal-detection theory links the abstract dimension of the internal response to the observer's overt dichotomous response. The linking is very simple: on any trial, the observer says YES when the amount of evidence for the signal is larger than some value known as the *criterion*, and NO when it is smaller than this value. The observer's *decision rule* to determine the response is

$$\begin{cases} \text{If evidence} > \text{criterion, then respond YES,} \\ \text{If evidence} < \text{criterion, then respond NO.} \end{cases}$$

When the evidence is assumed to be a continuous variable, there is no need to worry about whether the evidence exactly equals the criterion, as the probability of that event is negligible, and it can be assigned to either the YES or NO category without changing the properties of the rule. This decision rule is pictured in Figure 1.2. The point marked λ (the lowercase Greek letter lambda) on the abscissa is the criterion that divides the response types. For any value of X above λ, a response of YES is made; for any value of X below λ, a response of NO is made.

For compactness, Figure 1.2 (unlike Figure 1.1) has been drawn with both distributions on the same axis. This condensed diagram is the conventional way to present the signal-detection model. However, one should remember that each actual event is drawn from one or another of these distributions. On any trial, only one of them applies.

Some mathematical notation is needed to take the analysis further. Denote values on the abscissa of Figure 1.2 by the letter x. Under the signal-

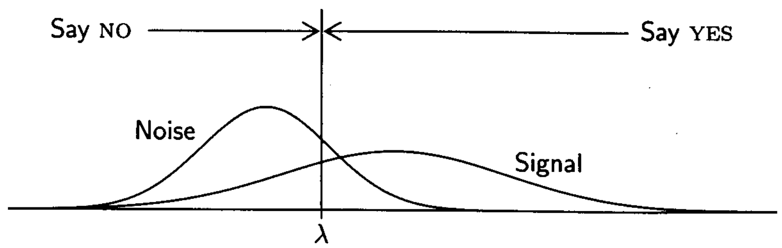


Figure 1.2: The signal and noise distributions of Figure 1.1 shown on a single axis with a decision criterion at the value λ.

detection model, the amount of evidence observed on a single trial is not a fixed number but a random variable (see Appendix Section A.2). Denote this random variable by X_n for noise trials and X_s for signal trials. Let $f_n(x)$ and $f_s(x)$ be the density functions of these two random variables—these are the curves pictured in Figures 1.1 and 1.2—and let $F_n(x)$ and $F_s(x)$ be their cumulative distribution functions. Using this notation, the probabilities of a hit or a false alarm, given the appropriate stimulus, are found by calculating the area under the density functions above the value λ. In mathematical nomenclature, this area is the *integral* of the function.[3] The false-alarm rate P_F is the probability that an observation from X_n exceeds λ, which can be written in various ways as

$$\begin{aligned} P_F &= P(\text{YES}|\text{noise}) \\ &= P(X > \lambda|\text{noise}) \\ &= P(X_n > \lambda) \\ &= \int_\lambda^\infty f_n(x)\,dx \\ &= 1 - F_n(\lambda). \end{aligned} \tag{1.1}$$

The shaded area in the upper distribution in Figure 1.3 corresponds to this probability. The hit rate P_H is defined similarly, using the distribution of

[3]This book uses the integral symbol from calculus to denote area. The area under the function $f(x)$ between $x = a$ and $x = b$ is written $\int_a^b f(x)\,dx$. The important parts of the expression are the function to be integrated $f(x)$ and the limits a and b; the rest is, in effect, conventional notation. When $a = -\infty$, the integral includes all the area below b, and when $b = \infty$, it includes all the area above a.

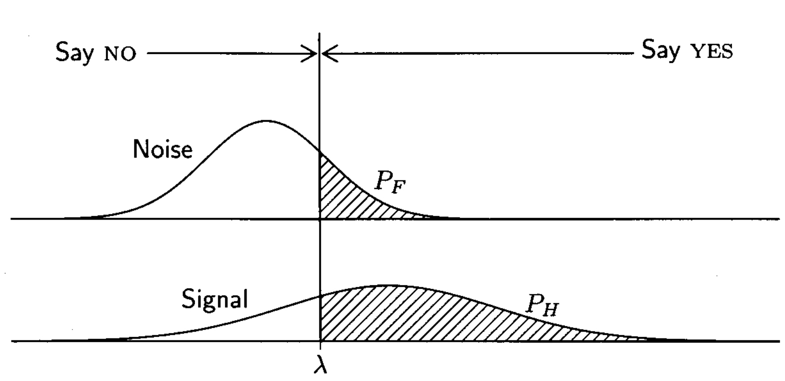


Figure 1.3: Probabilities of a false alarm and of a hit shown as shaded areas on the distributions of X_n and X_s from Figures 1.1 and 1.2.

evidence associated with the signal (Figure 1.3, bottom):

$$P_H = P(X_s > \lambda) = \int_\lambda^\infty f_s(x)\,dx = 1 - F_s(\lambda). \qquad (1.2)$$

The probabilities P_F and P_H are the theoretical counterparts of the proportions f and h calculated from data.

The definition of the hit rate and the false-alarm rate in terms of the distributions of X_s and X_n allows predictions of these values to be made once the forms of the random variables are specified. Before turning to these calculations in the next chapter, note that two aspects of the signal-detection model control the particular values of P_H and P_F:

- The overlap in the distributions. If $f_n(x)$ and $f_s(x)$ do not overlap much, then the hit rate can be high and the false-alarm rate can be low at the same time. If the distributions are nearly the same, then the hit rate and the false-alarm rate are similar.
- The placement of the criterion. If λ lies toward the left of the distributions, then both the hit rate and the false-alarm rate are large. If λ is lies toward the right, then both rates are small.

The importance of the signal-detection model is that it allows these two aspects of the detection situation to be measured separately. The detectability of the signal is expressed by the position of the distributions and their degree of overlap, while the observer's strategy is expressed by the criterion. The shape and position of the distributions in a particular detection task may be largely determined by the way that the stimulus is generated and the physiology of the detection process. The observer has

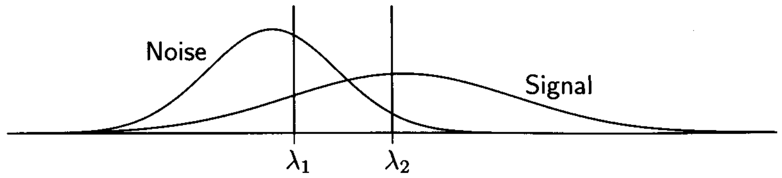


Figure 1.4: Signal and noise distributions with two criteria λ_1 and λ_2 representing the differences induced by instructions.

little or no control over these aspects. However, the observer can vary his or her propensity to say YES by changing the position of the criterion. Such changes can explain the differences in hit rate and false-alarm rate between the two sessions in the experiment described above. The criterion λ_1 shown in Figure 1.4 yields large hit rates and false-alarm rates, such as those in session 1, and the criterion λ_2 yields smaller rates such as those in session 2. Of course, one can only tell whether criterion differences are enough to explain the differences between the conditions by fitting the data quantitatively, which is the topic of the next two chapters.

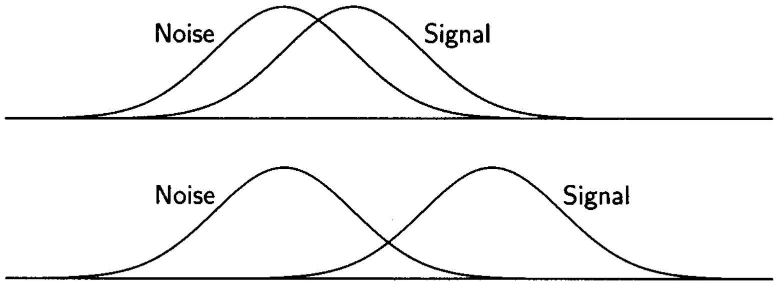
Reference notes

Signal-detection theory, as discussed here, was originally developed in work by Birdsall, Swets, and Tanner, although its roots go back at least to Gustav Fechner's *Elemente der Psychophysik* in 1860 (see Link, 1992). Many of the early articles relating to this development are collected in Swets (1964). The primary source for signal-detection theory is the book by Green and Swets (1966), which remains an essential (but sometimes difficult) reference. There are many secondary treatments of signal-detection theory directed toward different research domains. Among those oriented to psychophysics, the introduction by McNicol (1972) and the more mathematical treatment by Falmagne (1985) have influenced the present writing. Macmillan and Creelman (1991) provide a useful and more detailed treatment. John Swets has written a number of useful articles on the procedures, many of which are collected in Swets (1996). Briefer discussions of signal-detection theory appear in many texts on perception, cognitive science, and so forth. Articles that apply signal-detection theory are too numerous to cite.

Exercises

1.1. Describe two situations, other than the examples in Section 1.1, in which a decision between two clear alternatives must be made based on incomplete or ambiguous information. Identify the types of errors that could be made.

1.2. For each of the following pairs of hit and false-alarm rates, choose the pair of distributions that best describes it and locate a criterion that gives approximately those values. Remember that, by definition, the complete area under the distributions equals 1.0.



a. $P_F = 0.50$, $P_H = 0.84$.
b. $P_F = 0.20$, $P_H = 0.56$.
c. $P_F = 0.07$, $P_H = 0.93$.

1.3. Two screening tests are available that predict the appearance of a set of psychological symptoms. In a study of the first test, it is given to a group of 200 people. On a follow-up one year later, the test is found to have identified the 25 of the 30 people who developed the symptoms during the year. A comparable study using the second test looked at a different set of 150 people and identified 15 persons out of 28 who developed the symptoms. The rate of identification is greater for the first test than for the second (25/30 = 83% and 15/28 = 54%, respectively). Explain why these data do not, by themselves, tell which test is best at identifying the target people. What information is missing?

Chapter 2

The equal-variance Gaussian model

To make predictions from the signal-detection model described in the last chapter, the form of the distributions of X_s and X_n must be specified. One of the simplest and most natural choices is the conventional "bell-shaped" *normal distribution* of statistics, more commonly known in signal-detection applications as the *Gaussian distribution*. The next three chapters describe various uses of the Gaussian model, and it is central to much of the remainder of this book. The properties of this distribution, including methods used to find probabilities from it, are described in the appendix, starting on page 237. This material should be reviewed, either now or as required while reading this chapter.

The simplest Gaussian model is one in which the distributions for both signal and noise stimuli have the same shape, with the signal distribution being shifted to the right of the noise distribution, but otherwise identical to it. This representation is particularly useful as a description of a single detection condition and is the basis of the most commonly used detection statistics. It is the topic of this chapter.

2.1 The Gaussian detection model

In its most general form, the Gaussian signal-detection model assigns arbitrary Gaussian distributions with (potentially) different means and variances for the random variables representing the two stimuli:

$$X_n \sim \mathcal{N}(\mu_n, \sigma_n^2) \qquad \text{and} \qquad X_s \sim \mathcal{N}(\mu_s, \sigma_s^2).$$

The result is a model that depends on the values of five real-number parameters instead of on two arbitrary distributions. The two parameters μ_n and μ_s locate the centers of the distributions, the two parameters σ_n^2 and σ_s^2 give their variances, and the parameter λ specifies the response criterion. Once these five values are assigned, predictions of P_H and P_F can be made.

The signal-detection model does not use all five of these parameters. The values of the random variables X_n and X_s are unobservable. Only the shape and overlap of the distributions are important; the picture is unspecified as to its origin and scale. There is no way to determine the absolute size of X_n and X_s, only how they compare to each other. To illustrate this indeterminateness, consider Figure 1.2 on page 13. A zero point can be placed anywhere on the abscissa and scale values (1, 2, etc.) marked off at any spacing along the axis without changing the essential picture. The overlap of the distributions and the relative placement of the criterion are the same for any scaling. Thus, a set of detection data does not give enough information to assign unique values to the five parameters.[1]

To begin to remove the ambiguity, the values of any two of the five parameters (except the two variances) can be chosen arbitrarily. Fixing these parameters does not make the model less general or reduce its range of predictions; it only helps to give the other parameters unique values. One conventional way to fix two parameters is tô assign a standard Gaussian distribution to the noise distribution, by setting μ_n to zero and σ_n^2 to one, giving the pair of distributions

$$X_n \sim \mathcal{N}(0,1) \qquad \text{and} \qquad X_s \sim \mathcal{N}(\mu_s, \sigma_s^2). \tag{2.1}$$

With this restriction, the model depends on three parameters, μ_s, σ_s^2, and λ.

Example 2.1: Suppose that $X_n \sim \mathcal{N}(0,1)$, that $X_s \sim \mathcal{N}(1.5, 4.0)$, and that $\lambda = 1.2$. What are P_H and P_F?

Solution: The noise distribution has standard form, so the false-alarm rate is obtained by looking up the position of the criterion in a Gaussian distribution table. Using Equation 1.1 and the table of $\Phi(z)$ on page 249,

$$\begin{aligned} P_F &= P(X_n > \lambda) \\ &= 1 - P(X_n \le \lambda) \\ &= 1 - \Phi(\lambda) \\ &= 1 - \Phi(1.2) \\ &= 1 - 0.885 = 0.115. \end{aligned}$$

[1]In technical terms, the fully parameterized model is not *identifiable*. Many different combinations of μ_n, σ_n^2, μ_s, σ_s^2, and λ lead to the same values of P_H and P_F; thus, there is no way to identify which combination of parameter values created a particular pair of probabilities.

The signal distribution does not have standard form, so to calculate the hit rate (i.e., to evaluate Equation 1.2), both the mean and the standard distribution must be rescaled using Equation A.46:

$$\begin{aligned} P_H &= P(X_s > \lambda) \\ &= 1 - P(X_s \le \lambda) \\ &= 1 - \Phi\left(\frac{\lambda - \mu_s}{\sigma_s}\right) \\ &= 1 - \Phi\left(\frac{1.2 - 1.5}{2.0}\right) \\ &= 1 - \Phi(-0.15) \\ &= 1 - 0.440 = 0.560. \end{aligned}$$

Setting the noise distribution to standard form is but one way to constrain the signal-detection model and give the parameters unique values. Another possibility is to put the origin exactly between the two distributions and equate their variances, so that

$$X_n \sim \mathcal{N}(-\mu, \sigma^2) \qquad \text{and} \qquad X_s \sim \mathcal{N}(\mu, \sigma^2).$$

This representation is more natural in situations where neither stimulus event can be singled out to provide a baseline, as was done with the noise distribution in simple detection. This form of model will be important in the descriptions of forced-choice and discrimination studies in Chapters 6 and 7, and it is a natural consequence of the likelihood-ratio approach described in Chapter 9.

One might ask at this point why Gaussian distributions have been used for X_n and X_s, when other distributions could work as well. One reason is familiarity: the Gaussian is the most commonly used distribution in signal-detection theory models, and in some treatments is presented as if it were the only signal-detection model. The choice of a Gaussian distribution can be justified on either empirical or theoretical grounds. Empirically, it is often supported by data, using the methods that are described in Section 3.4. Theoretically, one can argue that when the evidence being incorporated in the decision derives from many similar sources, the central limit theorem (described on page 241) implies that their combination has approximately a Gaussian form. Such a situation might obtain, for example, if X is formed by summing the outputs of a set of detectors, each of which responds to some aspect of the signal, or by pooling many nearly ambiguous pieces of evidence.

Notwithstanding these arguments, there are some situations in which distributions that are not Gaussian are necessary. The most important of these cases occur when a non-Gaussian form for the distribution arises out of a theoretical model for the detection task. Distributions very much unlike the Gaussian form occur in Chapter 8 (see Figure 8.2 on page 136) and Section 9.4 (see Figure 9.5 on page 168). In such cases, of course, it is essential to use the correct distribution.

Beyond these specific situations, one sometimes finds treatments of signal-detection theory in which X_n and X_s have what is known as a *logistic distribution* (page 241). This distribution, like the Gaussian, is unimodal, symmetrical, and bell-shaped. Its shape is so similar to that of the Gaussian distribution that it is indistinguishable from it in practical applications. The logistic distribution arises naturally out of certain axiomatic treatments of choice behavior, and when working with these theories, it is the natural form for X_n and X_s. Another reason to use it is mathematical tractability: the cumulative logistic distribution function can be written as an algebraic expression, while the cumulative Gaussian distribution function $\Phi(z)$ must be approximated or taken from tables. This difference has no practical consequence for someone applying the signal-detection model, particularly when the calculations are done by a computer. In view of its wider use, the Gaussian model is emphasized in this book.

2.2 The equal-variance model

The most common application of signal-detection theory is to a single detection condition. The data obtained from a two-alternative detection experiment consist of one hit rate h and one false-alarm rate f. By themselves, these two numbers cannot determine unique values for three parameters μ_s, σ_s^2, and λ. One more constraint is needed to fit the signal-detection model.

The variance σ_s^2 is the parameter with the least obvious interpretation, so the most natural way to restrict the model is to constrain its value. The conventional assumption is that the variance of the signal distribution is the same as the variance of the noise distribution. Setting $\sigma_s^2 = \sigma_n^2 = 1$ gives the *equal-variance Gaussian model*. The standard notation for this model denotes the mean of the signal distribution by the symbol d' rather than μ_s. Thus, the two random variables have the distributions

$$X_n \sim \mathcal{N}(0,1) \quad \text{and} \quad X_s \sim \mathcal{N}(d',1).$$

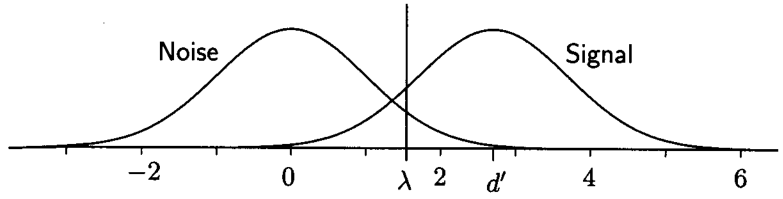


Figure 2.1: The distribution of noise and signal for the equal-variance Gaussian signal-detection model.

These distributions are shown in Figure 2.1. Using Equations 1.1 and 1.2 with the Gaussian distribution function, the probabilities of a false alarm and a hit are

$$P_F = 1 - \Phi(\lambda) \quad \text{and} \quad P_H = 1 - \Phi(\lambda - d'). \tag{2.2}$$

This model depends on two parameters, d' and λ, which can be estimated from observations of f and h.

The two parameters of the equal-variance Gaussian detection model have a simple and very important interpretation. The parameter d' describes the relationship of the noise and signal distributions to each other. When d' is near zero, the distributions are nearly identical; when it is large, they are widely separated. Thus, d' measures how readily the signal can be detected. The parameter λ describes the position of the decision criterion adopted by the observer. It is influenced by any propensity to say YES or NO, although it is not as satisfactory a measure of response bias as the quantities that will be described in Section 2.4.

Before going on to describe how f and h are converted to estimates of d' and λ, it is worth pausing to reconsider the restrictions that have been imposed on the general signal-detection model. There are three of these, each with a somewhat different status. The first is the use of the Gaussian distribution for X_n and X_s. As discussed at the end of the last section, this distribution is a reasonable choice for most sets of detection data. In practice, one rarely worries much about this assumption.

The second restriction fixes the parameters of the noise distribution, giving it standard form, with $\mu_n = 0$ and $\sigma_n^2 = 1$. This assumption has a different status from the others. It has no real content, in that it does not restrict the range of predictions that the model can make. The unobservable variables X_n and X_s are hypothetical, and they can be given any center and spread as long as they maintain the same relationship to each other. Thus, this restriction is made purely for numerical convenience. Either this assumption or an equivalent one is necessary if well-defined parameter estimates are to be made, but it cannot be tested by any data that could be collected.

The third assumption assigns equal variance to the two distributions, putting $\sigma_s^2 = \sigma_n^2$. This assumption is the most vulnerable of the three and the least likely to be correct. It can be tested, although not with the results of a single two-alternative detection experiment—the two data values from this experiment can be fitted perfectly by an appropriate choice of d' and λ. Unlike the standardization of X_n, the equal-variance assumption has implications for data in more elaborate experiments, a topic that will be discussed in the next chapters.

2.3 Estimating d' and λ

In order to use the detection parameters d' and λ to describe an observer's performance, their values must be calculated from the observed quantities h and f, a process known as *estimation*. Estimates of d' and λ should be chosen that match the model's predictions to the observed data. The derived values are known as *estimates*, not parameters, and are denoted here by placing a circumflex or "hat" on the symbol that denotes the theoretical parameter. For the equal-variance Gaussian model, the parameter estimates are $\widehat{d'}$ and $\widehat{\lambda}$. The hat serves to distinguish the estimates from their theoretical counterparts.

To find $\widehat{d'}$ and $\widehat{\lambda}$, one uses either a table of the Gaussian distribution or a program that calculates areas under this distribution. Essentially, these tables translate the observed rates of false alarms and hits to areas in a standard Gaussian distribution. The areas are then converted to values of $\widehat{d'}$ and $\widehat{\lambda}$. The calculation is easiest to do in three steps. Consider the data from the first session of the example experiment described in Section 1.2. The data from this experiment were

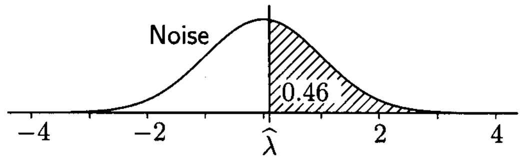
	NO	YES	
Noise	54	46	100
Signal	18	82	100

From these frequencies, the hit rate and the false-alarm rate are found:

$$h = \frac{82}{100} = 0.82 \quad \text{and} \quad f = \frac{46}{100} = 0.46.$$

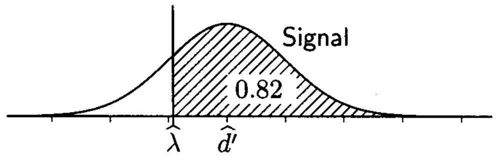
From here, the calculation proceeds as follows.

1. The false-alarm rate and the noise distribution are used to estimate the criterion λ. It is always helpful to illustrate the calculations with a picture. The noise distribution, with the criterion marked and the false-alarm rate shaded, is



This picture shows that $\widehat{\lambda}$ is the point on a standard normal distribution above which 46% of the probability falls, or, equivalently, below which 54% of the probability falls. The latter cumulated probability is the value tabulated in most normal-distribution tables, including the one on page 250. Looking up $Z(0.540)$ in this table shows that $\widehat{\lambda} = 0.100$.

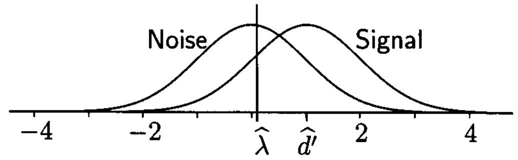
2. The hit rate is used to find the distance between d' and the criterion. A hit rate of 0.82 indicates that 82% of a Gaussian distribution centered at $\widehat{d'}$ falls above $\widehat{\lambda}$:



This distance is the difference between λ and d'. From the same table of $Z(p)$, 82% of the standard Gaussian distribution falls below $z = 0.915$. Converting this value to an upper-tail area, as the picture shows, $\widehat{\lambda}$ must be 0.915 units below $\widehat{d'}$:

$$\widehat{\lambda} - \widehat{d'} = -0.92.$$

3. The final step combines these two results to solve for $\widehat{d'}$. Graphically, the two diagrams are superimposed, to make the two values of $\widehat{\lambda}$ agree:



Numerically, the distance from μ_n up to $\widehat{\lambda}$ equals 0.10 and that from $\widehat{\lambda}$ up to $\widehat{d'}$ equals 0.92, so the center of the signal distribution lies $0.10 + 0.92 = 1.02$ units above the center of the noise distribution. Algebraically, the combination of steps 1 and 2 is

$$\begin{aligned}\widehat{d'} &= \widehat{\lambda} - (\widehat{\lambda} - \widehat{d'}) \\ &= 0.10 - (-0.92) = 1.02.\end{aligned}$$

From these calculations, detection performance in the first session is described by the parameter values $\widehat{d'} = 1.02$ and $\widehat{\lambda} = 0.10$.

To work in general, one must write these steps as equations. Because $Z(p)$ is the inverse of the cumulative Gaussian distribution, that is, the point for which $p = \Phi(z)$, the first step says that

$$Z(1-f) = \widehat{\lambda}.$$

The symmetry of the Gaussian distribution means that $Z(1-f) = -Z(f)$ (draw a picture), and that a simpler way to relate the false-alarm rate to the criterion is

$$\widehat{\lambda} = -Z(f). \tag{2.3}$$

Using the same two steps, the relationship between the hit rate and the estimates is

$$Z(1-h) = \widehat{\lambda} - \widehat{d'},$$

$$Z(h) = \widehat{d'} - \widehat{\lambda}.$$

Now use Equation 2.3 to replace $\widehat{\lambda}$ by $-Z(f)$ and solve for $\widehat{d'}$ to get

$$\widehat{d'} = Z(h) - Z(f). \tag{2.4}$$

When using Equations 2.3 and 2.4, it is helpful to verify that the criteria have been placed correctly and the right areas have been used by drawing a picture of the distributions, as was done above. Without seeing a picture, it is easy to drop the sign from $Z(h)$ or get the order of the subtracted terms wrong. Errors of this type are much easier to catch in a picture than in a formula.

Example 2.2: The results for the two sessions of the detection experiment described in Section 1.2 were

	h	f
First Session	0.82	0.46
Second Session	0.55	0.19

Show that the sessions differ in the criterion that was used more than in the detectability.

Solution: The signal-detection parameters for the first section were calculated in the three steps above. For the second session, Equations 2.3 and 2.4 give

$$\widehat{\lambda} = -Z(f) = -Z(0.19) = 0.88,$$

$$\widehat{d'} = Z(h) - Z(f) = Z(0.55) - Z(0.19) = 0.12 - (-0.88) = 1.00.$$

Summarizing these calculations,

	$\widehat{d'}$	$\widehat{\lambda}$
First Session	1.02	0.10
Second Session	1.00	0.88

It is immediately clear from this table that the criterion was shifted by the change in instructions, but the detectability of the signal remained the same. By converting from h and f to $\widehat{d'}$ and $\widehat{\lambda}$, the stability of detection, which was only suggested by the original proportions, becomes quantitatively clear. A formal statistical test of these conclusions, using the sampling model that will be developed in Chapter 11, is given in Example 11.6.

The two pairs of observations, (h, f) and $(\widehat{d'}, \widehat{\lambda})$, give different ways to view the outcome of a detection experiment. Each pair of numbers fully describes the results of the experiment. Although one pair can be converted to the other, each pair conveys information that is not so immediately obtained from the other pair. The signal-detection analysis translates the information provided by h and f into quantities that are much more descriptive of the detection process. Because the estimates $\widehat{d'}$ and $\widehat{\lambda}$ derive from the theoretical parameters of the signal-detection model, they separate the detection and the decision aspects of the observer's behavior in a way that h and f do not. Both types of information—the hit and false-alarm rates and the detection statistics—should be reported when describing the results of a detection task.

A problem sometime arises when one attempts to estimate detection parameters from a small sample of data. Although there is always some chance that the observer makes an error under the Gaussian detection model, with a small number of trials these events may never happen. The observer may respond YES on all the signal trials or NO on all the noise trials. These zero frequencies make either $h = 1$ or $f = 0$, so either $Z(h)$ or $Z(f)$ is undefined. Estimation of d' using Equation 2.4, which requires these values, does not work. Figure 2.2 illustrates the difficulty. The distributions are widely spaced, and either almost all of the signal distribution falls above the criterion or almost all of the noise distribution falls below it. However, unless some observations from each distribution fall on both sides of the criterion, it is impossible to tell how far apart to place the means. The distributions could be as shown in Figure 2.2, but they could as well lie somewhat closer or much farther apart.

How to deal with such data poses a dilemma. One possibility is to arbitrarily assign a small value to the empty category and proceed normally.

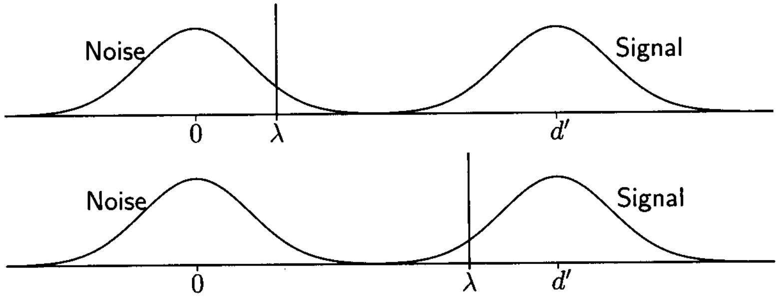


Figure 2.2: Distributions and criteria that represent data for which $h = 1$ (upper panel) or $f = 0$ (lower panel).

For example, a frequency of 1 observations, or $1/2$ observation, or $1/10$ observation might be assigned to an otherwise empty false-alarm category. If the noise stimulus has been presented on N_n trials, these values correspond to values of f of $1/(N_n + 1)$, $1/(2N_n + 1)$, or $1/(10N_n + 1)$, respectively. Giving a value to f lets Equations 2.3 and 2.4 be used to estimate λ and d'. Similarly, when $h = 1$, one can deduct 1, $1/2$, or $1/10$ observation from the number of hits to allow estimates to be made. The difficulty with these solutions is that even though the value that is substituted is arbitrary, it has an appreciable effect on the estimate. For example, suppose that in some study every one of 100 noise trials were correctly identified and 80 of 100 signal trials were hits. The three alternatives for the false-alarm rate give estimates of d' of 3.17, 3.42, and 3.93. No one of these numbers is more correct than another. It is important to recognize the arbitrariness of these values and to avoid drawing any conclusions that depend on the specific substitution.

An alternative to assigning a specific value to f or h is simply to acknowledge the indeterminateness and to treat the condition as implying a "large" detectability of the signal, without attempting to assign it an exact numerical value. In particular, the unknown estimate is larger than that of any comparable condition in which all the frequencies are positive. This problem will be revisited in Section 4.6.

2.4 Measuring bias

Although the parameter λ is the most direct way to express the placement of the observer's criterion, it is not the best way to measure the bias. Its value depends on the false-alarm rate, but not on the hit rate. Yet, how one interprets a particular criterion needs to take the detectability of the

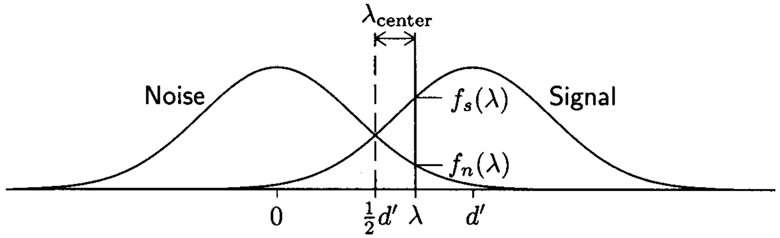


Figure 2.3: Measures of bias. The centered criterion λ_{center} is measured from the point between the signal and noise means. The heights of the density functions used to calculate the likelihood ratio are marked.

signal into account. For example, when $d' = 0.2$, the criterion $\lambda = 0.5$ represents a bias toward NO responses, but when $d' = 2.0$, it represents a bias toward YES responses—draw pictures of the distributions to see how the situations differ. A better descriptive measure of the bias takes account of the distributions of both signal and noise. There are several ways to combine both distributions in a bias measure, two of which are considered here.

One way to create a better measure of the bias is to express the position of the criterion relative to a point halfway between the signal and the noise distribution (Figure 2.3). A value of 0 for this measure indicates that the criterion is placed exactly between the distributions, implying that there is no preference for YES or NO responses and that the probabilities of misses and false alarms are the same. A negative displacement of the criterion relative to the center puts it nearer to the noise distribution and indicates a bias to say YES, and a positive value indicates a criterion nearer to the signal distribution and indicates a bias to say NO. The point between the distributions is $d'/2$ units above the mean of the noise distribution, so the placement of the criterion relative to this center is found by subtracting this midpoint from λ. Use the subscript $_{\text{center}}$ to distinguish the criterion measured this way from when it is measured from the noise mean:[2]

$$\lambda_{\text{center}} = \lambda - \tfrac{1}{2}d'. \tag{2.5}$$

Both λ and λ_{center} refer to the same criterion; they differ in the origin from which this criterion is measured. The use of a centered criterion is particularly natural when the origin of the decision axis is placed between the distributions, as described at the end of Section 2.1.

An estimate of λ_{center} can be found from $\widehat{d'}$ and $\widehat{\lambda}$, or it can be found directly from f and h. Substituting the estimates of λ and d' from Equations

[2]Macmillan and Creelman (1991) denote the measure λ_{center} by c.

2.3 and 2.4 into Equation 2.5 gives an estimate in terms of the response probabilities:

$$\widehat{\lambda}_{\text{center}} = \widehat{\lambda} - \tfrac{1}{2}\widehat{d'} = -\tfrac{1}{2}[Z(f) + Z(h)]. \qquad (2.6)$$

A second measure of the bias expresses the propensity to say YES or NO by the relative heights of the two distribution functions at the criterion. Denote this ratio by β (the lowercase Greek letter beta):

$$\beta = \frac{f_s(\lambda)}{f_n(\lambda)}. \qquad (2.7)$$

The two heights are marked in Figure 2.3 on page 27. In this figure, $\lambda = 1.80$ and $d' = 2.55$, and the height of the density functions at λ (using Equation A.43) are

$$f_n(\lambda) = \varphi(\lambda) = \frac{1}{\sqrt{2\pi}}e^{-\lambda^2/2} = \frac{1}{\sqrt{2\pi}}e^{-1.80^2/2} = 0.79,$$

and $f_s(\lambda) = \varphi(\lambda - d') = \varphi(-0.75) = 0.301$. The bias ratio is

$$\beta = \frac{f_s(\lambda)}{f_n(\lambda)} = \frac{\varphi(\lambda - d')}{\varphi(\lambda)} = \frac{\varphi(-0.75)}{\varphi(1.80)} = \frac{0.301}{0.079} = 3.81.$$

The ratio β is an instance of a *likelihood ratio*, which will play an important role in the general development of detection theory in Chapter 9.

To understand this likelihood-ratio bias measure, it helps to see how it varies with the position of the criterion. Figure 2.4 shows the two distributions of the equal-variance model and the ratio of their heights. Directly between the two means, the distributions are of equal height and the ratio is one; to the left of this center, the noise distribution is higher and the ratio is less than one; to the right of it, the signal distribution is higher and the ratio is greater than one.

Figure 2.5 shows noise and signal distributions with the criteria that might be adopted by three different observers. The observer with the criterion λ_1 is most biased toward saying YES. That observer makes both many hits and many false alarms. For this observer, $f_n(\lambda_1) > f_s(\lambda_1)$ and β is less than one. The observer with criterion λ_3 is just the opposite. Here the bias is such as to make few false alarms and, in consequence, a lower hit rate than in the other condition. Now $f_n(\lambda_3) < f_s(\lambda_3)$ and β exceeds one. The criterion λ_2 has an intermediate placement, for which $f_n(\lambda_2) = f_s(\lambda_2)$ and $\beta = 1$.

The major defect with β as a bias measure is that it is asymmetrical. As Figure 2.4 shows, bias favoring YES responses is indicated by values in the

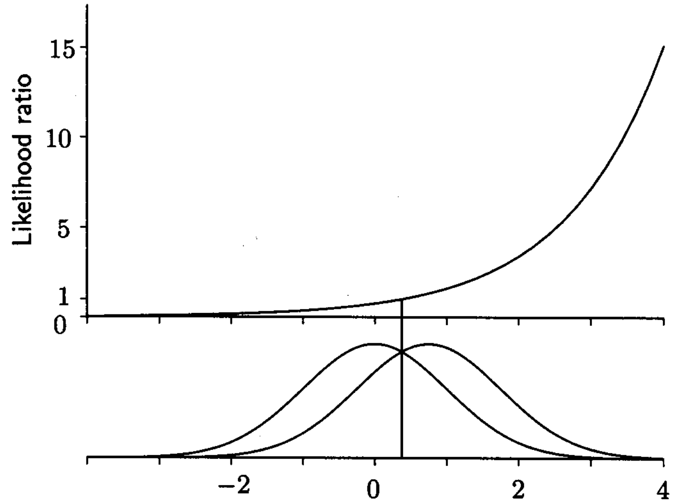


Figure 2.4: The ratio of the heights of two unit-variance Gaussian density functions. The means of the distributions differ by $d' = 0.75$. The vertical line marks the point where both distributions have equal height.

narrow range between 0 and 1, while bias favoring NO responses is indicated by values ranging from 1 all the way to ∞. Much more room is available on the NO side than on the YES side. This asymmetry is removed by taking the natural logarithm[3] of the likelihood ratio:

$$\log\beta = \log\left[\frac{f_s(\lambda)}{f_n(\lambda)}\right] = \log f_s(\lambda) - \log f_n(\lambda). \qquad (2.8)$$

The placement of the criterion in Figure 2.3 gives $\log\beta = \log(3.81) = 1.34$.

For the equal-variance Gaussian model, the distribution functions can be substituted into Equation 2.8 to express $\log\beta$ in terms of the other parameters. Substituting the Gaussian density functions with the appropriate means and variances (Equation A.42) gives

$$\log\beta = \log\left\{\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}(\lambda - d')^2\right]\right\} - \log\left\{\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}\lambda^2\right]\right\}$$

[3]Here and throughout this book the expression $\log(x)$ refers to the natural logarithm, that is, to the logarithm to the base $e = 2.71828$. This type of logarithm is sometimes denoted by $\ln(x)$, particularly on calculators. Common logarithms (to the base 10) are not used in this book. One reason for the importance of the natural logarithm is that it is the inverse of exponentiation. You can undo one by doing the other: $\log[\exp(x)] = x$ and $\exp[\log(x)] = x$, as in derivations such as that leading to Equation 2.9 below.

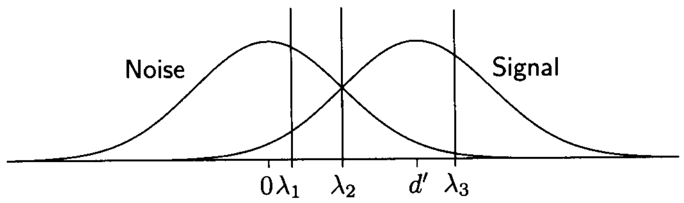


Figure 2.5: Distributions of signal and noise with three decision criteria.

$$= \log \frac{1}{\sqrt{2\pi}} + \log\left\{\exp[-\tfrac{1}{2}(\lambda - d')^2]\right\} - \log \frac{1}{\sqrt{2\pi}} - \log\left\{\exp[-\tfrac{1}{2}\lambda^2]\right\}$$
$$= -\tfrac{1}{2}(\lambda - d')^2 + \tfrac{1}{2}\lambda^2$$
$$= d'(\lambda - \tfrac{1}{2}d'). \qquad (2.9)$$

The value of $\log\beta$ is zero at the point where the density functions cross, which for this model occurs when the criterion is at $d'/2$, exactly between the distributions. It is negative for criteria lying to left of this midpoint, favoring YES responses, and positive for criteria lying to the right, favoring NO responses. As the form of Equation 2.9 shows, the log likelihood ratio is a linear function of λ, so would be a straight line if plotted on Figure 2.4.

An estimate of the bias is found by substituting estimates of d' and λ into Equation 2.9:

$$\widehat{\log\beta} = \widehat{d'}(\widehat{\lambda} - \tfrac{1}{2}\widehat{d'}). \qquad (2.10)$$

Substituting the estimates $\widehat{d'}$ and $\widehat{\lambda}$ (Equations 2.3 and 2.4) into this formula shows that $\widehat{\log\beta}$ is half the difference of the squares of the Z transforms of f and h:[4]

$$\widehat{\log\beta} = \tfrac{1}{2}[Z^2(f) - Z^2(h)]. \qquad (2.11)$$

The two measures λ_{center} and $\log\beta$ are quite similar, as a comparison of Equations 2.5 and 2.9 shows. They differ in that $\log\beta$ is scaled by the detectability relative to λ_{center}:

$$\log\beta = d'\lambda_{\text{center}} \qquad \text{and} \qquad \lambda_{\text{center}} = \log\beta/d'. \qquad (2.12)$$

Both measures are satisfactory indicators of response bias, and it is hard to choose one over the other. When comparing the bias in conditions in which the signals are equally detectable, the choice is inconsequential; as Equations 2.12 indicate, both measures order the conditions in the same way.

[4]A squared function $f^2(x)$ is a shorthand for the square of the function, $[f(x)]^2$.

However, when the conditions differ in detectability, the two measures can lead to different conclusions. With large differences in d', the measures can even order two conditions differently. It would be nice to be able to choose either $\log\beta$ or λ_{center} based on some fundamental property of the signal-detection model. The choice would be based on a clear understanding of how the observer reached the decision. However, theories of the observer's performance are not sufficiently developed to resolve the issue. It is likely that real behavior is sufficiently variable and influenced by experimental conditions to preclude a single choice. In the end, it is risky to make quantitative comparisons of bias across conditions with widely varying detection parameters.

This book emphasizes the likelihood-ratio measure $\log\beta$ for two plausible, but not definitive, reasons. First, the measure generalizes more easily than λ_{center} to situations in which the distributions of signal and noise events have different variance, are non-Gaussian, or are multivariate. Second, the likelihood-ratio definition seems somewhat closer to what is intended by the word "bias." An observer who is biased toward a particular alternative will choose it even when the evidence on which the choice was made is more likely to have occurred were the other alternative true. This sense of bias is more similar to a likelihood ratio than a criterion placement. In what follows the unqualified term *bias* refers to $\log\beta$.

Example 2.3: Estimate the bias of the observer in the two sessions in the example of Section 1.2 and Example 2.2.

Solution: Starting with the fitted detection model for the first of the two sessions and using Equations 2.10 and 2.6 gives the estimates

$$\widehat{\log\beta} = \widehat{d'}(\widehat{\lambda} - \tfrac{1}{2}\widehat{d'}) = 1.02(0.10 - 1.02/2) = -0.42,$$
$$\widehat{\lambda}_{\text{center}} = \widehat{\lambda} - \tfrac{1}{2}\widehat{d'} = 0.10/1.02/2 = -0.41.$$

The calculation can also start with the raw detection results and use Equations 2.11 and 2.6 to find the bias. For the second session:

$$\widehat{\log\beta} = \tfrac{1}{2}[Z^2(f) - Z^2(h)] = \tfrac{1}{2}[Z^2(0.19) - Z^2(0.55)]$$
$$= \tfrac{1}{2}[(-0.878)^2 - (0.126)^2] = 0.38,$$
$$\widehat{\lambda}_{\text{center}} = -\tfrac{1}{2}[Z(f) + Z(h)] = \tfrac{1}{2}[(-0.878) + 0.126] = 0.38.$$

The two measures $\widehat{\log\beta}$ and λ_{center} are almost identical here because d' is almost exactly unity. Moreover, the fact that d' is about the same in the two sessions means that either $\widehat{\log\beta}$ or $\widehat{\lambda}_{\text{center}}$ leads to the same conclusion about the relationship between the bias of the two sessions.

2.5 Ideal observers and optimal performance

One reason to construct a theoretical detection model is to determine how well the best possible observer could do. Such a hypothetical individual, who is able to make optimal use of the information available from the stimulus, is known as an *ideal observer*. Like a real observer, an ideal observer is limited by the intrinsic uncertainty of the stimulus events, so cannot attain perfect performance. The performance of the ideal observer indicates what is possible, given the limits imposed by the random character of the signals.

The ideal-observer model is not a description of real data. Real observers are not ideal, and they usually fall short of ideal behavior. The deviations between idea and real performance are interesting, because they tell something about the actual perceptual processes or decision-making behavior. The ideal observer gives the standard needed to make this comparison.

In complex tasks there can be several sorts of ideal observers, each optimally using different information about the stimulus. For example, an observer who knows exactly the form of the distribution of the signal and noise events can perform differently than an observer who knows less about the shape of these distributions. This section considers an observer in a yes/no detection task who is aware of the statistical properties of the two types of stimulus but has no specific control over these characteristics. This description is a good ideal representation for an observer who can modulate performance through changes in the decision criterion but who cannot control the nature of what is observed.

Specifically, consider an observer who has access to the stimulus only through the random variables X_n and X_s. This observer knows the forms of these distributions and the proportion of the trials on which a signal occurs. This observer can modulate the decision criterion λ (or select the likelihood ratio at the decision criterion) so that the largest possible number of correct responses are made. The signal-detection model lets this optimal value be determined.

The first step in determining the optimal criterion is to calculate the correct-response probability P_C as a function of the decision criterion λ. The observer can then choose a criterion $\lambda^\star$ that maximizes P_C with respect to λ. The first part of the calculation is very general. The probability of a correct response is the sum of the probability of a hit and the probability of a correct rejection. The probability of a signal trial is s and of a noise trial is $1-s$, and the probabilities of correct responses to signal and noise trials are $1-F_s(\lambda)$ and $F_n(\lambda)$, respectively. Combining these probabilities, the overall probability of a correct response is

$$P_C = P(\text{signal trial and YES}) + P(\text{noise trial and NO})$$

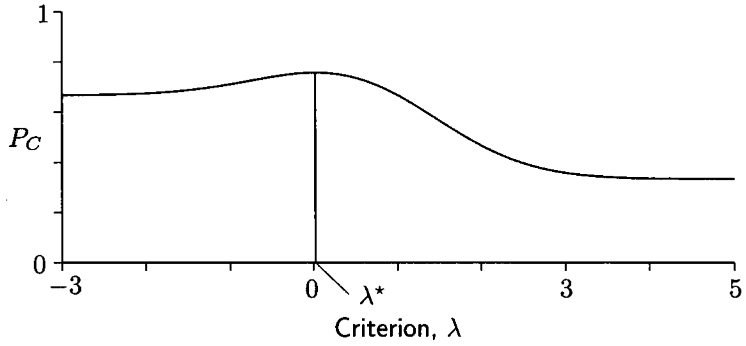


Figure 2.6: The probability of a correct response as a function of the criterion for signals with $d' = 1.2$ occurring on $s = 2/3$ of the trials. The value of P_C is greatest when the criterion is placed at $\lambda^\star$.

$$\begin{aligned} &= P(\text{signal trial})P(\text{YES}|\text{signal}) + P(\text{noise trial})P(\text{NO}|\text{noise}) \\ &= s[1-F_s(\lambda)] + (1-s)F_n(\lambda). \end{aligned} \tag{2.13}$$

Figure 2.6 shows this function for the equal-variance Gaussian model in a study with $s = 2/3$ signal trials. On the left, when all responses are YES, $P_C = s = 2/3$, and on the right, when all responses are NO, $P_C = 1 - s = 1/3$. In between, P_C rises to a maximum, then declines. The best performance is obtained by setting the criterion to hit this maximum point.

The value $\lambda^\star$ of the criterion for which P_C is greatest is found using calculus.[5] Differentiating Equation 2.13 with respect to λ gives

$$\begin{aligned} \frac{dP_C}{d\lambda} &= -s\frac{dF_s(\lambda)}{d\lambda} + (1-s)\frac{dF_n(\lambda)}{d\lambda} \\ &= -sf_s(\lambda) + (1-s)f_n(\lambda). \end{aligned}$$

By setting this quantity to zero and doing some algebraic manipulation, one finds that at the optimal criterion $\lambda^\star$, the likelihood ratio β is equal to the ratio of the probability of a noise trial to the probability of a signal trial, a quantity known as the odds (page 228):

$$\beta^\star = \frac{f_s(\lambda^\star)}{f_n(\lambda^\star)} = \frac{1-s}{s}. \tag{2.14}$$

[5]The extremum (minimum or maximum) of a smooth function occurs at the point where its derivative (or slope) is zero—the hill in Figure 2.6 is flat at the very top. Readers who are unfamiliar with the calculus can skip to the answer in Equation 2.14.

The optimal bias is minus the logarithm of this ratio, which is known as the logit (page 228):

$$\log \beta^\star = \log \frac{1-s}{s} = -\operatorname{logit}(s). \tag{2.15}$$

This result will be developed from a different background in Section 9.2.

When signals and noise events are equally likely (i.e., when $s = 1/2$), Equation 2.14 says that the criterion should be chosen so that the two densities are equal, that is, so that $f_n(\lambda^\star) = f_s(\lambda^\star)$. More simply, the optimal criterion lies at the point where the two distribution functions cross.[6] In Figure 2.5 the criterion λ_2 is located at this optimal point. Positioning the optimal criterion at the crossover makes good intuitive sense. To the left of the crossover, $f_n(x)$ is greater than $f_s(x)$, so as one moves the criterion in this direction, false alarms are being added more rapidly than hits are being increased, and P_C goes down. To the right of the crossover, $f_n(x) < f_s(x)$, so as one moves λ in that direction, hits decrease more rapidly than do false alarms, and P_C also goes down.

Unequally frequent signal and noise events displace the optimal point away from the crossover. When signals are more likely than noise events, the odds $s/(1-s)$ are larger than one, and the optimal criterion is at a point where $f_n(\lambda^\star) > f_s(\lambda^\star)$. This point lies to the left of the place where the distributions cross. When signals are rare, the optimal point shifts to the right, to a position at which evidence favoring noise events is given greater weight, so that $f_n(\lambda^\star) < f_s(\lambda^\star)$.

The analysis thus far has not used any information about the actual distribution of the variables X_n and X_s. Results such as Equations 2.14 and 2.15 apply to any pair of distributions. When the forms of the distribution functions are known, the position of the criterion can be worked out exactly. Under the equal-variance Gaussian model, Equation 2.9 gives the relationship among d', the optimal criterion $\lambda^\star$, and the optimal bias $\log \beta^\star$:

$$\log \beta^\star = d'(\lambda^\star - \tfrac{1}{2}d').$$

Substituting the optimal bias from Equation 2.15 and solving for optimal criterion gives

$$\lambda^\star = \tfrac{1}{2}d' - \frac{\operatorname{logit}(s)}{d'}. \tag{2.16}$$

[6] There are distributions for which Equation 2.14 has several solutions or where the solution to Equation 2.14 minimizes P_C. However, these configurations do not pose practical problems. When there are several solutions, the appropriate one is usually clear. Solutions that minimize P_C are primarily curiosities. More rigorous optimality criteria can be established when needed.

The point $d'/2$ is halfway between the centers of the distributions and is the optimal point when $s = 1/2$. The second term adjusts the optimal criterion for unequal stimulus frequencies, moving it downward when $s > 1/2$ and upward when $s < 1/2$. Expressed in terms of the centered criterion λ_{center}, the picture is even simpler. The optimal centered criterion is zero when $s = 1/2$, and is shifted to $-\operatorname{logit}(s)/d'$ when the frequencies are unequal.

Example 2.4: Suppose that in a difficult detection task $d' = 1.2$ and that signals occur on two thirds of the trials. Where should the criterion be placed under the equal-variance Gaussian model? What is the maximum probability of a correct response that an observer can attain? How do these results change when only one trial in 10 is a signal?

Solution: The logit of the signal probability is

$$\operatorname{logit}(s) = \log \frac{2/3}{1-2/3} = \log 2 = 0.693.$$

By Equation 2.16, the optimal criterion is

$$\lambda^\star = \tfrac{1}{2}d' - \frac{\operatorname{logit}(s)}{d'} = \frac{1.2}{2} - \frac{0.693}{1.2} = 0.600 - 0.5781 = 0.022.$$

The probability of a correct response using this optimal criterion (Equation 2.13) is

$$\begin{aligned} P_C &= s[1-\Phi(\lambda - d')] + (1-s)\Phi(\lambda) \\ &= \tfrac{2}{3}[1-\Phi(0.022-1.200)] + \tfrac{1}{3}\Phi(0.022) \\ &= \tfrac{2}{3}(1-0.120) + \tfrac{1}{3}(0.509) = 0.757. \end{aligned}$$

Figure 2.6 was drawn using this set of parameters. When signals are rare, with $s = 1/10$, the log-odds are

$$\operatorname{logit}(s) = \log \frac{1/10}{9/10} = -\log 9 = -2.197,$$

and the optimal criterion is

$$\lambda^\star = \frac{1.2}{2} - \frac{-2.197}{1.2} = 0.600 + 1.831 = 2.431.$$

This criterion actually lies on the far side of the mean of the signal distribution. The bias to say NO is very strong. Notwithstanding this asymmetry, the probability of a correct response is substantial:

$$P_C = \tfrac{1}{10}[1-\Phi(2.431-1.200)] + \tfrac{9}{10}\Phi(2.431) = 0.904.$$

This large probability reflects the fact that even before the trial starts, the observer can be fairly sure that no signal will occur. Only the strongest signals receive a YES response, and P_C is little greater than the value of 0.9 that could be obtained by unconditionally saying NO.

The probability P_C of a correct response is only one of the many quantities that an observer may chose to maximize. Where there are costs or values associated with the alternatives, the observer can adjust the criterion to minimize the cost or maximize the gain. If false alarms are expensive and misses are cheap, then an optimal observer will shift λ to reduce the false-alarm rate. Specifically, suppose the value of the various outcomes are given by $V(\text{hit})$, $V(\text{miss})$, and so forth—the first value is usually positive and the second negative. The average value of a trial is

$$\begin{aligned}\mathsf{E}(V) = {} & P(\text{signal and YES})V(\text{hit}) + P(\text{signal and NO})V(\text{miss}) \\ & + P(\text{noise and YES})V(\text{false alarm}) \\ & + P(\text{noise and NO})V(\text{correct rejection}).\end{aligned} \tag{2.17}$$

This equation can be expanded in the manner of Equation 2.13 and its maximum value found. It turns out the optimal criterion can be expressed in terms of the difference between the value of a correct response and the value of an error for each of the stimulus types, which are, in effect, the costs of the two types of errors:

$$\begin{aligned}c_m &= V(\text{hit}) - V(\text{miss}), \\ c_f &= V(\text{correct rejection}) - V(\text{false alarm}).\end{aligned}$$

The result (derived in Problem 2.8) is to add a second term to the optimal bias that was found in Equation 2.15:

$$\log \beta^\star = \log \frac{c_f}{c_m} - \operatorname{logit}(s). \tag{2.18}$$

Real observers make similar accommodations, albeit not always to the optimal point. An experimenter can easily manipulate an observer's criterion by changing the payoff for the various responses. It is instructive to consider how the real or perceived costs of a decision alter the bias in real-world detection tasks such as those mentioned at the start of Chapter 1.

Reference notes

The sources cited in the previous chapter apply here as well. Greater detail on some of the alternative approaches (logistic distributions, other bias measures, etc.) are given by Macmillan and Creelman (1991), who also summarize some empirical evidence. The choice theory mentioned in connection with the logistic distribution was originally formulated by Luce (1959) and is described in Atkinson, Bower, and Crothers (1965).

Exercises

2.1. Calculate P_H and P_F for the five-parameter detection model of Section 2.1 with $X_n \sim \mathcal{N}(2,4)$, $X_s \sim \mathcal{N}(7,9)$, and a criterion at 5. Show by calculation that the same values are obtained from the standard three-parameter model with $X_s \sim \mathcal{N}(2.5, 2.25)$ and $\lambda = 1.5$.

2.2. Suppose that, in the signal-detection model for a simple detection task, the means of the signal distribution and the noise distribution differ by four fifths of the standard deviation of the noise distribution and that the variance of the signal distribution is three times that of the noise distribution. What are the hit rate and the false-alarm rate for a decision criterion placed one half standard deviation of the noise distribution above the mean of the noise distribution? Start by drawing a picture, then do the calculation.

2.3. Suppose that an observer sets $\lambda = 0.8$ in a detection task with $d' = 1.4$. What are the predicted hit rate and the false-alarm rate?

2.4. What are β and $\log \beta$ in Problem 2.2?

2.5. Draw pictures illustrating the situation described in the first paragraph of Section 2.4. What values of $\log \beta$ are implied in the two cases?

2.6. In a detection experiment 1000 trials are run, 600 with the signal and 400 with noise alone. The resulting responses are

	NO	YES	
Noise	327	73	400
Signal	104	496	600

Estimate d', λ, $\log \beta$, and λ_{center}. Draw a picture illustrating the model.

2.7. When the distributions are as estimated in Problem 2.6, what value of λ produces the most correct responses? How close is the observed probability of a correct response to its optimal value?

2.8. Suppose that costs are assigned to the two types of errors, with false alarms being charged a cost of c_f and misses being charged a cost of c_m.

a. Expand Equation 2.17 and insert probabilities to find the average cost associated with a criterion placed at position λ.

b. Maximize this expression and show that the optimal bias is given by Equation 2.18.

2.9. Consider an observer watching for a rare event, such as one that occurs on only one percent of the trials. The event by itself is reasonably detectable, with $d' = 2$. What would an observer minimizing errors do here? Using this criterion, what is the proportion of signals that are overlooked? Now suppose that this omission probability is too large. To induce the observer to reduce omissions, a reward of \$50 is given for each correct detection and \$1 for each correct rejection. How does the observer's performance change if it is adjusted to maximize monetary return? How does this change affect the false-alarm rate?

2.10. The theory of signal detectability has been extensively applied in radiology. Consider someone examining X-ray films for signs of a rare disorder. Discuss the various tradeoffs in performance and outcome inherent to the task, relating the performance to the situation described in Problem 2.9.

Chapter 3

Operating characteristics and the Gaussian model

In Section 1.2, data from two detection sessions with different biasing instructions were described. The analysis of these data showed that the sessions differed in the observer's bias toward YES or NO responses, but not appreciably in the detectability of the signal. In these analyses, each session's data were treated independently. A more comprehensive analysis should tie the two sets of observations together in a single model of the detection task. This chapter describes this integration. It also shows how to represent data from several detection conditions in a convenient, instructive, and widely used graphical form. This representation lets one investigate the adequacy of the assumptions that underlie the Gaussian signal-detection model and fit a model in which the variances of the signal and noise distributions are unequal.

3.1 The operating characteristic

Consider the results from the two sessions of Section 1.2. Fitting an equal-variance Gaussian model to these data, as in Examples 2.2 and 2.3, gives the parameter estimates at the top of Figure 3.1. As these numbers show, the estimates of d' are very similar, but those of λ and $\log \beta$ differ considerably. These values suggest that the proper representation of the full set of observations should use the same pair of distributions in the two sessions, but change the criterion. This representation is illustrated at the bottom of Figure 3.1. Using the same distributions for both sessions implies that the characteristics of the stimuli (which are controlled by the experimenter)

	h	f	$\widehat{d'}$	$\widehat{\lambda}$	$\widehat{\log\beta}$
First Session	0.82	0.46	1.02	0.10	−0.42
Second Session	0.55	0.19	1.00	0.88	0.38

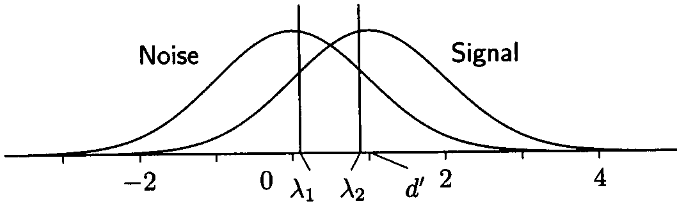


Figure 3.1: Data and parameter estimates for the two detection sessions of Section 1.2 and a representation using stimulus distributions separated by $d' = 1.01$ with specific criteria at $\lambda_1 = 0.10$ and $\lambda_2 = 0.88$.

do not change, while using session-specific criteria implies that the decision rules applied to these distributions are different.

The change in instructions that shifted the criteria in Figure 3.1 produced a trade-off between the two types of correct response, hits and correct rejections. The higher criterion produced many correct rejections but few hits, while a low criterion increased the hit rate at the cost of the correct-rejection rate. Put another way, changes in criterion cause the hit rate and the false-alarm rate to vary together. The trade-off is illustrated more directly by plotting the hit rate against the false-alarm rate, in the manner of a scatterplot, as shown in Figure 3.2. The false-alarm rate is plotted on the abscissa (horizontal axis) and the hit rate on the ordinate (vertical axis). Each session's results are represented by a point, S_1 for the first session and S_2 for the second session. Because the hit rate exceeds the false-alarm rate, these points lie above the diagonal of the square that connects the point (0, 0) to the point (1, 1). The outcome of any detection study corresponds to a point in this square, generally in the upper triangle.

The theoretical counterpart of Figure 3.2 plots the probabilities P_F and P_H generated by a particular signal-detection model. This plot is particularly valuable in illustrating how constraints on the parameters limit the predictions. Consider what is possible for an observer working according to the Gaussian model who can vary the criterion but who cannot alter d'. Figure 3.3 shows four points from such an observer, with the corresponding distributions in insets. Each point is based on the Gaussian model with $d' =$ 1.15—note that the relative positions of the distributions do not change. However, the criteria differ. Bias toward NO responses gives the point at the lower left, and bias toward YES responses gives the point at the upper right.

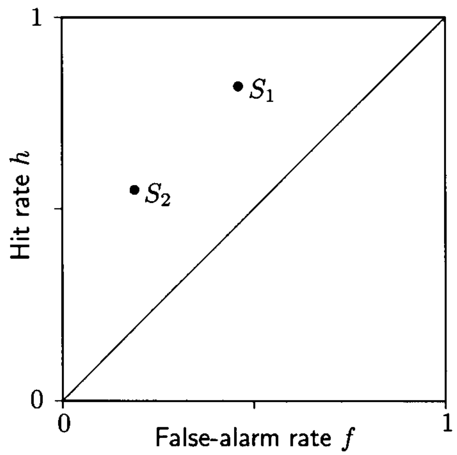


Figure 3.2: A plot of the false-alarm rate f and the hit rate h for the two detection sessions of Section 1.2.

The full range of performance available to this observer is shown by a line on this graph. When λ is very small, both P_F and P_H are large, leading to a point near the (1, 1) corner of the graph. As λ is increased, the area above the criterion decreases, as do P_F and P_H, and the point (P_F, P_H) traces out the curve shown in the figure. Eventually the prediction approaches (0, 0) when λ is so large that both distributions fall almost completely below it.

The solid line in Figure 3.3, which shows the possible range of performance as the bias is adjusted, is known as an *operating characteristic*. The original application of the statistical decision model to detection problems grew out of work on radio reception, where these curves were known as receiver operating characteristics. This name has stuck; the curve is often called a *receiver operating characteristic* or, more briefly, a *ROC curve*. Because it shows the performance possible for a fixed degree of discriminability, that is, a fixed sensitivity to the signal, the curve is also known as an *isosensitivity contour*. Both terms are used in this book.

Figure 3.3 shows the operating characteristic for a signal with detectability $d' = 1.15$. By varying the detectability of the signal—say, by changing its intensity—one obtains a family of curves. Figure 3.4 shows several members of this family. These are drawn for the equal-variance Gaussian model (using the methods described in the next section) and are labeled with values of d'. When $d' = 0$ the hit rate and the false-alarm rates are identical and the operating characteristic lies along the diagonal of the square. For positive values of d', the curves lie above this diagonal, moving close and closer to the upper left-hand corner as d' increases. When d' is

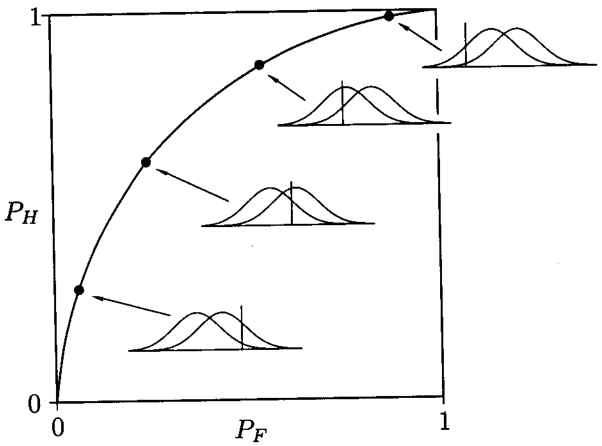


Figure 3.3: The operating characteristic or isosensitivity contour for a detection task derived from the equal-variance Gaussian model with $d' = 1.15$. The inset diagrams show the distributions of the random variables X_n and X_s and the criterion λ at four typical points.

very large, the operating characteristic is almost identical to the two lines at right angles that make up the left and top sides of the square. This operating characteristic describes performance when the signal is so strong that the observer can almost completely avoid making errors.

In an important sense, the operating characteristics are what describe the sensitivity of an observer to a particular signal. Together, the strength of that signal and the receptivity of the observer determine how detectable that signal will be. The combination selects the operating characteristics associated with that d', and the outcome of any study falls on that line. The point on that line that is actually observed is determined by the observer's bias.

3.2 Isocriterion and isobias contours

Aspects of the signal-detection model other than the bias can be varied to produce contours in the (P_F, P_H) space. *Isocriterion contours* are created by holding λ constant and varying the sensitivity. Because the criterion and the false-alarm rate determine each other (remember that P_F is the probability that $X_n > \lambda$), the isocriterion contours are vertical lines at constant values of P_F (Figure 3.5, top). Contours such as these might describe the results of a study in which signals of different strengths were

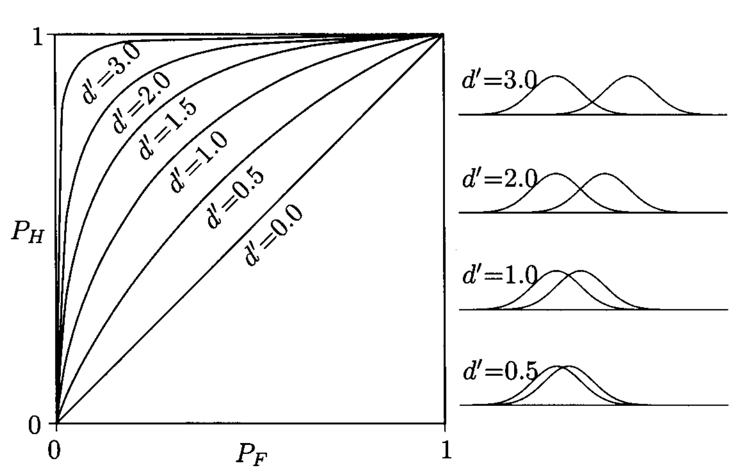


Figure 3.4: The family of isosensitivity contours determined by different values of d' for the equal-variance Gaussian model. At the side are shown the noise and signal distributions for four of the contours.

mixed with simple noise events, thus giving one false-alarm rate and several hit rates (one for each strength), with decisions based on a single criterion.

More interesting are the *isobias contours* created by holding one of the bias indices constant while varying the sensitivity. The lower two panels of Figure 3.5 show the lines of constant β or $\log \beta$ (left) and of constant (right). Both sets of isobias contours end in the upper left corner of the plot, a consequence of the fact that with sufficiently strong signals it is possible to attain near-perfect performance regardless of the bias. The curves otherwise have substantially different shapes, reflecting the different properties of the two measures. Each curve of constant likelihood ratio starts for very weak signals either at the point (0, 0) (for positive values of $\log \beta$) or the point (1, 1) (for negative values). In contrast, each curve of constant λ_{center} starts on a different point on the chance diagonal.

The isobias contours of either type cut across the isosensitivity contours of Figure 3.4. Thus, any point (P_F, P_H) in the space determines one level of sensitivity (a line from Figure 3.4) and one level of bias (a line from Figure 3.5). Similarly, any combination of sensitivity and bias determine, by the intersection of one line from each of the figures, a false-alarm rate and a hit rate.

Isocriterion and isobias curves are usually of less practical interest than are the isosensitivity curves. One reason is that it is easier to find situations in which one expects the sensitivity to remain constant while the bias changes than situations where one expects the criterion or the bias to remain constant while sensitivity changes. Another reason is that bias is a

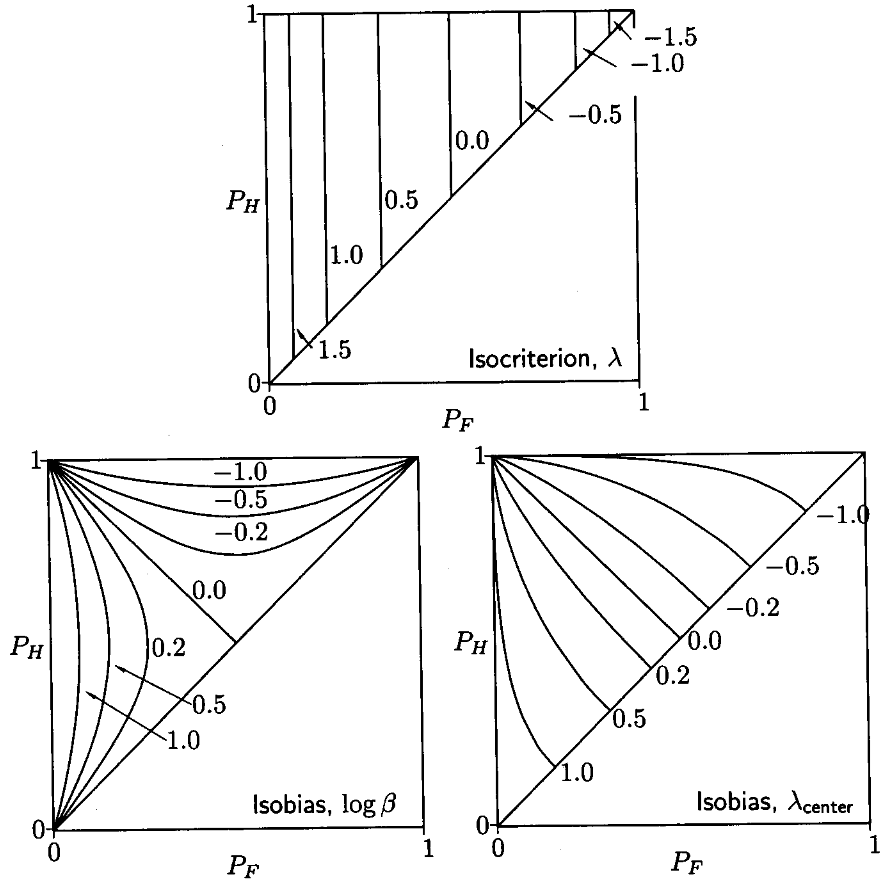


Figure 3.5: Families of isocriterion curves (top) and isobias curves (bottom) for the equal-variance Gaussian model. Isobias curves are shown for β or $\log \beta$ (left) and λ_{center} (right).

more complex psychological construct than is sensitivity. It is easy to understand how the sensitivity of an observer to a particular physical stimulus could remain constant, but it is much harder to assert that bias, by any particular definition, should not vary as the signal changes. Finally, most research that uses detection theory is concerned with factors that influence detectability, not with those affecting the decision process. Differences in response bias in these studies are a nuisance to be removed, not a process to be studied. Determining on which isosensitivity curve the performance falls is most important.

3.3 The equal-variance Gaussian operating characteristic

The actual shape of the operating characteristics is derived theoretically from the distributions of the random variables X_s and X_n. First consider the process in its most general form. Suppose that $f_s(x)$ and $f_n(x)$ are the density functions of these random variables. The hit rate and the false-alarm rate are areas under these curves above the criterion point and are calculated as integrals of the corresponding density functions above those points:

$$P_F = \int_\lambda^\infty f_n(x)\,dx \qquad \text{and} \qquad P_H = \int_\lambda^\infty f_s(x)\,dx \tag{3.1}$$

(Equations 1.1 and 1.2). Denote these quantities, as functions of λ, by $P_F = F(\lambda)$ and $P_H = H(\lambda)$, respectively—here F and H stand for false alarms and hits, respectively, and are not cumulative distribution functions. If one could write these integrals as simple expressions, then one could solve $F(\lambda)$ for λ and substitute it in $H(\lambda)$ to get the isosensitivity curve. Working purely formally:[1]

$$\begin{aligned} \lambda &= F^{-1}(P_F), \\ P_H &= H(\lambda) = H[F^{-1}(P_F)]. \end{aligned} \tag{3.2}$$

Although this procedure is simple in the abstract, a practical difficulty arises with the Gaussian distribution. For it, Equations 3.1 cannot be written as simple expressions, and they cannot be solved algebraically to give a formula for Equation 3.2. Gaussian operating characteristics, such as those in Figures 3.3 and 3.4 are constructed using tables or numerical methods.

The relationship between P_F and P_H implied by the operating characteristic is simplified by transforming the probabilities before they are plotted. The appropriate transformation here is the inverse Gaussian function $Z(p)$. After applying it, the operating characteristic is a straight line. In more detail, first recall (from Equations 2.2 on page 21) that the response probabilities under the Gaussian model with $\sigma_s^2 = 1$ are

$$\begin{aligned} P_F &= 1 - \Phi(\lambda) = \Phi(-\lambda), \\ P_H &= 1 - \Phi(\lambda - d') = \Phi(d' - \lambda). \end{aligned}$$

[1]A function raised to a negative power is the inverse function: if $y = f(x)$, then $x = f^{-1}(y)$. This use of the exponent differs from the notation for positive powers, for which $f^2(x) = [f(x)]^2$ (see footnote 4 on page 30).

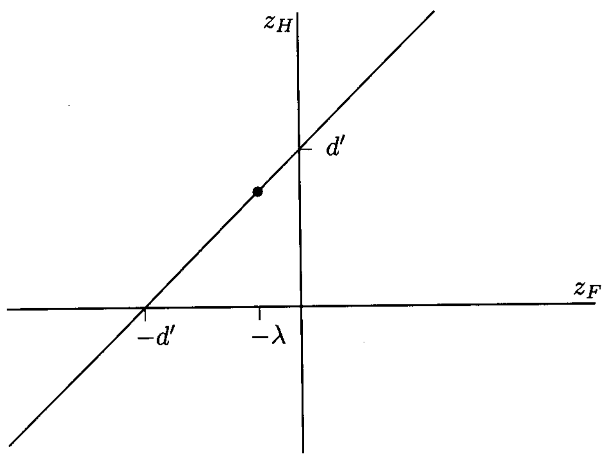


Figure 3.6: The isosensitivity function for the equal-variance Gaussian model plotted in Gaussian coordinates. The solid point corresponds to performance with a criterion of λ.

The formula for the hit rate takes the nonzero mean of X_s into account. Now transform these probabilities to $z_F = Z(P_F)$ and $z_H = Z(P_H)$. This transformation undoes the function $\Phi(z)$:

$$z_F = Z(P_F) = Z[\Phi(-\lambda)] = -\lambda,$$
$$z_H = Z(P_H) = Z[\Phi(d' - \lambda)] = d' - \lambda.$$

Eliminating λ from these equations gives

$$z_H = z_F + d'. \tag{3.3}$$

This equation describes the isosensitivity function when it is plotted on Gaussian transformed axes, that is, in *Gaussian coordinates*.

Figure 3.6 shows the operating characteristic plotted in these coordinates. Four things about it are important to notice:

- The function is linear. Plotted in Gaussian-tranformed coordinates, the isosensitivity function is a straight line. This fact is a consequence of the choice of the Gaussian form for X_n and X_s.
- The function has a 45° slope. This slope is a consequence of the equal-variance assumption. As will be seen in the next section, models with unequal variances have different slopes.
- The line crosses the axes at $-d'$ and d'. Thus, having drawn the line, one can easily read off the value of d' from either intercept or, having d', can easily draw the line.

Probabilities		Gaussian scores	
f	h	$Z(f)$	$Z(h)$
0.12	0.47	−1.18	−0.08
0.18	0.72	−0.92	0.58
0.20	0.58	−0.84	0.20
0.38	0.78	−0.31	0.77
0.51	0.77	0.02	0.74
0.66	0.92	0.41	1.40
0.77	0.96	0.74	1.75

Table 3.1: Hit and false-alarm data obtained by varying the bias at without changing the signal.

- Isocriterion contours are vertical lines at $-\lambda$. So the point corresponding to a model with a particular criterion λ lies at the point of the operating characteristic with a value of $-\lambda$ on the z_F axis.

Because the isosensitivity function in Gaussian coordinates is a straight line, the model is easy to apply to data from several bias conditions. First convert the observed hit and false-alarm proportions to Gaussian coordinates using either tables or a computer program. Then draw a straight line at 45° through these points. The intercept on the ordinate is an estimate of d'. This procedure is discussed in more detail in Section 3.5; for the moment an example will illustrate it.

Example 3.1: Table 3.1 shows the proportions that might be obtained from a seven-level manipulation of bias. Fit the equal-variance Gaussian model to these data and draw the operating characteristic.

Solution: The observed proportions f and h are plotted directly on the left in Figure 3.7. The points roughly trace out an isosensitivity curve, but there is sufficient scatter that an operating characteristic cannot be constructed by connecting them. The resulting line is neither smooth nor monotonic. The scatter is not surprising and is inevitable unless an extremely large number of observations have been obtained at each point, a condition fulfilled only by a few psychophysical experiments.

To fit the function the proportions are converted to Gaussian scores in the second part of Table 3.1 and plotted on the right in Figure 3.7. It is now easy to draw a line at 45° through the midst of them. The points are adequately fitted by this line—certainly they have no systematic curvature—and from the place where this line crosses the axes, $\widehat{d'}$

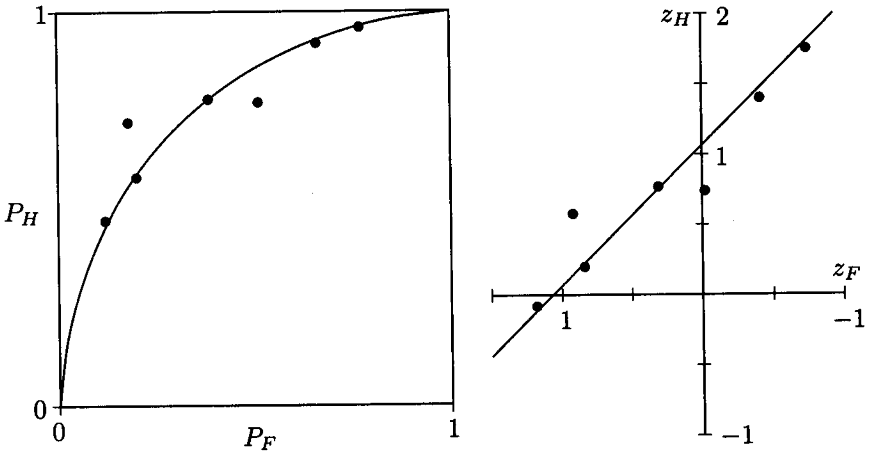


Figure 3.7: Plots of the data in Table 3.1 with fitted operating characteristics in probability coordinates (left) and Gaussian coordinates (right).

is apparently slightly greater than one. A more accurate estimate, obtained from a computer program, is $\widehat{d'} = 1.065$. To plot the isosensitivity function in its conventional form, points on the line in Gaussian coordinates are reconverted to probability coordinates using the transformation $p = \Phi(z)$ giving the theoretical line in the probability plot in Figure 3.7.

3.4 The unequal-variance Gaussian model

In the general Gaussian signal-detection model, the distributions of X_n and X_s can differ in their variance as well as their means. With the constraints placed on parameters of the noise distribution to give the model unique values (as discussed in Section 2.1), the distributions of the random variables associated with the two stimulus events are

$$X_n \sim \mathcal{N}(0,1) \quad \text{and} \quad X_s \sim \mathcal{N}(\mu_s, \sigma_s^2).$$

The resulting model is more flexible than the equal-variance version, and when data from several conditions are available, it often provides a superior fit. Figure 3.8 shows an example in which the signal distribution has greater variance than the noise distribution.

Unequal variability of the signal and noise events arises quite naturally in a number of situations. One explanation is based on the rules for the addition of random variables. Consider the detection of a pure tone of

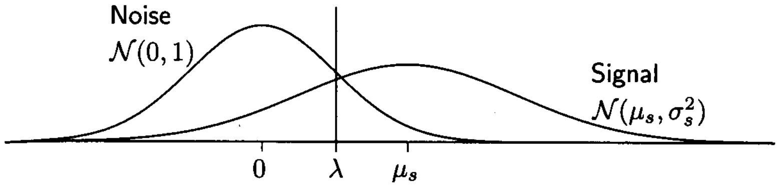


Figure 3.8: Distributions of signal and noise under a Gaussian model in which the standard deviation of the signal distribution is 1½ times that of the noise distribution.

known frequency ν in a white noise background. Suppose that the observer bases the detection on the output of a tuned detector responding to the intensity of the stimulation at the frequency ν. The "noise" here consists of the variation in the extent to which the white noise near the frequency ν excites the detector. This variation creates the variability of X_n. When the tone is added to the background, creating the signal-plus-noise condition, the background variability does not vanish. The added signal increases the response of the detector above that in the noise conditions, say, by an amount S:

$$X_s = X_n + S.$$

If S is a pure constant, without any variability itself, then the random variable X_s is simply a displaced version of X_n, and their variances are the same. However, if the signal has variability, then S is a random variable and its variation combines with that of the noise background. A reasonable assumption here is that S and X_n are independent. The variance of the signal-plus-noise distribution is now the sum of the variances of its parts (Equation A.33 on page 235):

$$\operatorname{var}(X_s) = \operatorname{var}(X_n) + \operatorname{var}(S) \geq \operatorname{var}(X_n).$$

Any variation in the signal or in its detection makes $\operatorname{var}(X_s) > \operatorname{var}(X_n)$. Equality of variance holds only when the signal is a fixed nonrandom quantity.

Another cause of differences in signal and noise variance is a direct consequence of the mechanism by which observations of X_n or X_s are generated. The mean and variance of many random processes are related, so that random variables with larger means also have larger variances. If the response to the stimulus is generated by such a process, then the larger mean of X_s also gives it a greater variance. As a specific example, suppose that X is observed by counting the number of discrete events (perhaps neural responses) during a fixed interval of time. The rate at which these events occur is larger when the signal is present than when it is not, so the count is, on average, bigger when the signal is present. *Counting processes* of this type have been widely studied. For many of them, the variance of the number of events is approximately proportional to the mean.[2] Although counting processes do not produce true Gaussian distributions, they are usually closely approximated by Gaussian distributions whenever the number of counts is large. Because of the relationship between the mean and the variance, the unequal-variance model must be used.

In other situations one might expect the noise variance to exceed that of the signal. Sometimes the signal event acts to reduce the diversity of an original distribution. Consider a word recognition experiment in which the subject identifies words that were presented in the first part of the experiment, and suppose that the subject does this by estimating how recently he or she has heard the target word. The old, or signal, words have all been presented recently and so have a relatively tight distribution of ages, but the new, or noise, words have a great range of ages. Some words have been seen or heard only a few hours ago, while other words are days, months, or years old. When the words are studied, values from this highly variable distribution are replaced by much more similar values that refer to the experimental presentation.

The unequal-variance Gaussian model depends on three parameters: the two distributional parameters, μ_s and σ_s^2, and the criterion λ. Thus, its parameters cannot be determined by a single yes/no detection study, which yields but two independent results (f and h). Fitting the unequal-variance model requires either several conditions varying in bias or the rating-scale experiment of Chapter 5.

The theoretical operating characteristic for the unequal-variance model is constructed by converting the response rates to Gaussian coordinates as in Section 3.2. As in the equal-variance model, the false-alarm rate depends directly on the criterion:

$$P_F = \Phi(-\lambda) \qquad \text{and} \qquad z_F = -\lambda.$$

The signal distribution now involves its variance. Using the rule for finding area under a normal distribution with nonunit variance (Equation A.46 on page 240),

$$P_H = P(X_s > \lambda) = 1 - \Phi\!\left(\frac{\lambda - \mu_s}{\sigma_s}\right) = \Phi\!\left(\frac{\mu_s - \lambda}{\sigma_s}\right).$$

[2]The simplest of the counting processes is the *Poisson process*, in which the counted events are postulated to occur independently at a constant rate. It gives rise (not surprisingly) to Poisson distributions for X_n and X_s. The mean and variance of a Poisson random variable are equal.

Inverting this relationship gives the ordinate of the operating characteristic:

$$z_H = Z(P_H) = \frac{\mu_s - \lambda}{\sigma_s}.$$

Finally, substituting $-z_F$ for λ gives the equation of the isosensitivity function:

$$z_H = \frac{1}{\sigma_s} z_F + \frac{\mu_s}{\sigma_s}. \tag{3.4}$$

This function is linear, with a slope that is the reciprocal of the signal standard deviation. In the case depicted in Figure 3.8, where $\sigma_s = 1.5\sigma_n$, the slope is $1/1.5 = 2/3$.

When the isosensitivity function for a model with unequal variance is translated from Gaussian coordinates back to probability coordinates, the resulting operating characteristic is not symmetric about the minor diagonal of the unit square (the line from lower right to upper left), as it was in the equal-variance case. If $\sigma_s > 1$, then the function has a form like that shown in Figure 3.9. The line rises sharply from (0, 0) as the criterion drops through the signal distribution without reaching much of the noise distribution, then turns more slowly toward the right as the noise distribution is passed while there is still appreciable area under the signal distribution. Eventually the curve approaches (1, 1) when the criterion is below both distributions.

The operating characteristic in Figure 3.9 has several disconcerting features. The most obvious of these is that it dips below the diagonal at the upper right. This dip suggests that for certain criterion positions the false-alarm rate exceeds the hit rate. More subtle is the fact that the slope of the operating characteristic goes from shallower to steeper in this region, a condition that is necessary for the dip to occur. Operating characteristics in which the slope changes nonmonotonically like this are said to be *improper*. Those with curves that progress from (0, 0) to (1, 1) with ever-decreasing slope are said to be *proper*. An improper operating characteristic is a sign that the observer is using a response rule that does not make optimal use of the available information. In the case of the unequal-variance Gaussian model, the improper operating characteristic indicates that a simple criterion applied to the axis is not the best way to make the response. A superior response rule will be described in Section 9.3.

On first being introduced to the unequal-variance Gaussian model, one is inclined to worry about the dip below the diagonal and the apparently perverse behavior that it implies. However, to fret much about it is to take the model too seriously. The model is an useful description of detection behavior, but cannot be taken as mathematical truth. There are many

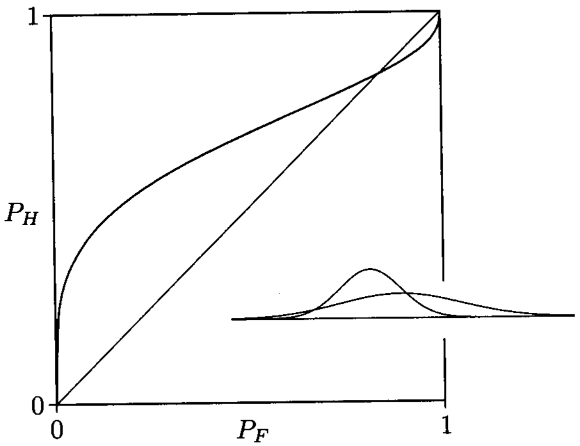


Figure 3.9: The isosensitivity function for the unequal-variance Gaussian model with $X_n \sim \mathcal{N}(0,1)$ and $X_s \sim \mathcal{N}(1,4)$. The inset shows the distributions of X_n and X_s.

circumstances in which the nonunit slope of a transformed operating characteristic implies that an unequal-variance representation is appropriate, but where no evidence of the dip is present. In fact, to observe the dip in real data is almost impossible. It would occur only for conditions where the observer was so biased toward YES responses that both the hit rate and the false-alarm rate were almost, but not quite, equal to one. An accurate determination of such an extreme point would require an enormous sample of data. It is doubtful that an observer could be induced to perform in a stable way in such extreme conditions for the requisite number of observations.

3.5 Fitting an empirical operating characteristic

Fitting the unequal-variance signal-detection model requires a set of conditions that differ in bias but not in the detectability of the signal. These conditions can be created by keeping the signal and noise events the same while inducing the observer to alter the criterion. There are several ways to produce this shift. Practiced observers can respond to instructions to change their detection performance in a way that increases or decreases the hit rate or false-alarm rate. Less trained observers will change the criterion in response to the proportion of signals. The analysis of the ideal observer in Section 2.5 indicated that bias should be shifted with the signal frequency

to maximize correct responses (Equation 2.15). Real observers shift their performance in this direction, although usually not to the optimal extent. When signals are rare, fewer YES responses are made, lowering both the hit rate and the false-alarm rate. When signals are common, the bias is shifted to increases the number of YES responses. Thus, running sessions with different proportions of signals will give different points on the operating characteristic. Bias can also be manipulated by assigning payoffs to the various types of responses, as in the example of Section 1.2. Each of these manipulations concerns the observer's response behavior only and can typically be made without substantially altering the sensitivity.

The first thing to do with such data is to look them over for anomalies or inconsistencies that could indicate something wrong with the study. Plot the hit rate against the false-alarm rate and verify that the general form of an operating characteristic is obtained. Problems are indicated by points for which the false-alarm rate exceeds the hit rate or by a failure of the order of the conditions on the operating characteristic to match the manipulation used to shift the bias. The sampling fluctuation associated with any set of empirical observations can lead to some reversals or to points where $h < f$, particularly when the amount of data is small, but large discrepancies are unlikely. When they happen, it is possible that the observer misunderstood the task or was behaving perversely, but that is unlikely in a well-run study. A more common cause of these irregularities is a mistake by the researcher in recording the data or in entering them into a computer program for analysis. Such errors should be ruled out before taking an irregular point too seriously.

When the data have been deemed satisfactory, turn to the analysis. It is helpful to organize the analysis in five steps:

1. Convert the data from hit and false-alarm rates h and f to the transformed values $Z(h)$ and $Z(f)$, using the table of the Gaussian distribution on page 250 or an equivalent computer program. Plot these points as an operating characteristic in Gaussian coordinates. Be accurate—it helps to use a full sheet of graph paper with closely spaced grid lines.
2. Evaluate the Gaussian model. If the points fall in a straight line (except for what can be deemed sampling error), then the use of this model is justified. If they curve, then consider a model based on another distribution. Under most circumstances a visual evaluation is sufficient (formal statistical tests are covered in Section 11.5). Usually, the Gaussian model will be adequate.
3. Fit a straight line to the set of points. Be sure to try the line at 45° that corresponds to the equal-variance model. When there is little

scatter in the data, the line can be drawn by eye without serious error. When the data are more scattered, some form of statistical fitting procedure is better, although rough estimates still can be made by eye. A transparent 45° drafting triangle is very useful here. Decide whether the 45° is adequate (some statistical procedures are discussed in Section 11.6). Otherwise conclude that the variances are unequal and that a line of some other slope is needed.

4. Estimate the parameters of the signal distribution from the fitted line. If the chosen line has a slope of one, then estimate d' by the intercept of the function, as described in Section 3.2. If the line does not lie at 45°, then find the slope b and intercept a of the equation

$$Z(h) = bZ(f) + a.$$

Match the slope and intercept to Equation 3.4 to estimate the Gaussian model's parameters:

$$\widehat{\mu}_s = a/b \quad \text{and} \quad \widehat{\sigma}_s = 1/b. \tag{3.5}$$

When working from a graph on which the line has been drawn, it is easier to forget about the slope and intercept and instead note the points x_0 and y_0 where the line crosses the horizontal and vertical axes, respectively. The parameters of the line are

$$a = y_0 \quad \text{and} \quad b = -y_0/x_0, \tag{3.6}$$

and those of the detection model are

$$\widehat{\mu}_s = -x_0 \quad \text{and} \quad \widehat{\sigma}_s = -x_0/y_0. \tag{3.7}$$

5. To estimate the criterion λ for a condition, the observed point must be translated to one on the fitted line. A full analysis here takes account of the accuracy with which each coordinate is observed. However, for most purposes it is sufficient to take the bias from the point on the operating characteristic that is closest to the observation. Frequently this can be done by sketching a line perpendicular to the line that goes through the point. Details and formulae are given in Section 4.5. Once the point is chosen, $\widehat{\lambda}$ is minus the abscissa of the point.

Example 3.2: Three carefully measured detection conditions give the pairs of false-alarm and hit rates (0.12, 0.43), (0.30, 0.76), and (0.43, 0.89). Does a Gaussian model fit these data? If it does, then estimate its parameters.

Solution: First sketch the data in probability coordinates (not shown) and look for irregularities. Here there are none. Then transform the probabilities to Gaussian coordinates:

f	h	$Z(f)$	$Z(h)$
0.12	0.43	−1.17	−0.18
0.30	0.76	−0.52	0.71
0.43	0.89	−0.18	1.23

The transformed points are plotted in Figure 3.10. A line at 45° (dashed in the figure) is clearly unsatisfactory, so the equal-variance Gaussian model does not apply. However, they lie so close to another straight line that it can be drawn through them by eye (the solid line in the figure). Evidently, the unequal-variance Gaussian model fits these data. Because the slope is greater than one, the signal distribution has a smaller variance than the noise distribution. Reading the graph carefully, the line crosses the horizontal axis at $x_0 = -1.04$ and the vertical axis at $y_0 = 1.46$. The slope and intercept of the line are found from these crossing points (Equations 3.6):

$$a = y_0 = 1.46 \quad \text{and} \quad b = \frac{y_0}{-x_0} = \frac{1.46}{1.04} = 1.40.$$

These values can be converted to estimates of the parameters of the signal distribution using Equations 3.5, or they can be found directly from the crossing points (Equations 3.7):

$$\widehat{\mu}_s = -x_0 = 1.04 \quad \text{and} \quad \widehat{\sigma}_s = \frac{-x_0}{y_0} = \frac{1.04}{1.46} = 0.71.$$

Finding the criteria associated with the conditions is simple here. The line fits so accurately that its closest approaches to the three points are at the abscissas already calculated as $Z(f)$. Accordingly, the criteria are 1.17, 0.52, and 0.18.

The problem of finding the best-fitting line deserves comment. The fact that one is fitting a straight line to a group of points makes it tempting to use the simple fitting equations of ordinary linear regression. One would write z_H as a linear function of z_F and calculate the slope and intercept from the usual regression equations. Unfortunately, this procedure gives the wrong line and systematically biased estimates of the parameters. The difficulty lies in the way that the regression line is defined. The model for linear regression treats one variable x as a predictor and the other variable y as an outcome to be predicted from x. The outcome variable y is represented theoretically as a linear function of the predictor plus a random error e:

$$y = \alpha + \beta x + e.$$

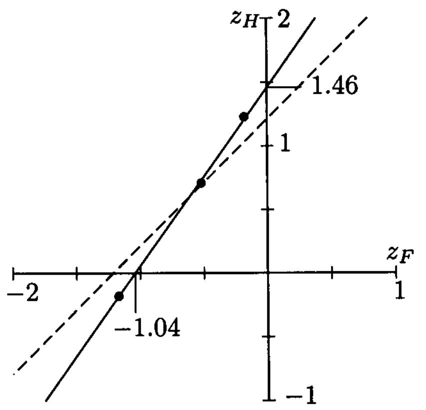


Figure 3.10: A three-point operating characteristic for Example 3.2 plotted in Gaussian coordinates. The dashed line at 45° is unsatisfactory, but the solid line from the unequal-variance model fits well.

In this regression model, x is an exact number, and all the sampling uncertainty is attributed to y. This model is a unsatisfactory representation of detection data, for which both $Z(h)$ and $Z(f)$ are subject to sampling error.

The use of the regression model here would not create problems if it did not bias the estimates of the detection parameters. The problem is a phenomenon known as *regression to the mean.* Geometrically, variability in the data acts to flatten the regression line, so that the best prediction of y is nearer (in standard deviation units) to the mean of that variable than the predictor x is to its mean. The greater the variability of the data, the more the line is flattened. Shrinking the predictions toward the mean is appropriate for the asymmetric measurement structure of regression analysis, but is incorrect for detection data. The line regressing $Z(h)$ on $Z(f)$ has a smaller slope than does a line that treats both $Z(h)$ and $Z(f)$ as subject to sampling error. Unless the data are almost error free, a regression analysis will give too small a value for b and consequently overestimate $\widehat{\sigma}_s$. When there is considerable scatter in the data, this bias can lead the equal-variance model to be rejected when in fact it is appropriate.

3.6 Computer programs

The amount of calculation involved in fitting the signal-detection theory model makes it a good candidate for computerized calculation. The calculations for a single detection condition are easy to implement. Algorithms

for the functions $Z(f)$ and $Z(h)$ are well established, and their values are directly available in some higher level languages. Using them, Equations 2.3 and 2.4 can be calculated directly.

When three or more conditions with different bias are to be fitted, the task is considerably harder. It is no longer possible to convert the proportions directly into parameter estimates, but a program must, in effect, find the operating characteristic that fits the data best, even though it may not pass exactly through any of the observed points. However, because the probabilistic structure of the signal-detection model is well defined, it can be fitted with the standard statistical technique known as maximum-likelihood estimation. Several programs that make these calculations have been published (see reference notes). These usually report most or all of the statistics that will be discussed in Chapter 4. One of the great advantages of the maximum-likelihood procedure is that it gives values for the standard errors of the parameter estimates. The use of these estimates will be discussed in Chapter 11.

When using one of these programs, it helps to know a little about what they are doing. There are no equations that directly give estimates of the parameters of the Gaussian model in terms of the data (e.g., as there are in multiple regression). Instead, the signal-detection model is fitted by an iterative algorithm. The programs start by making some guess at the parameter values—not necessarily a very good one. Then it adjusts these values to improve the fit. This process of adjustment, or iteration, is repeated (generally out of sight of the user) until the estimates cease to improve, at which time the results are reported and the values of any summary measures are calculated. Many programs report the number of iterations required to complete the process or the criterion of change used to decide when to stop.

For most sets of data, this procedure runs successfully and delivers good estimates. However, with certain very irregular sets of data, the algorithm can break down and fail to find a solution—a failure to converge, as it is called. Exactly how the program handles this contingency depends on the implementation, and some versions of the algorithm are more robust than others. Failure of the estimates to converge is uncommon with signal-detection data. It most often occurs when the observations are very much inconsistent with the signal-detection models, for example, with points lying very far from a line in Gaussian space. Thus, when a program fails to converge, it is advisable to review the data and see if they have been entered incorrectly or if they are sufficiently at odds with the signal-detection model that any parameters obtained by fitting that model will be meaningless. If it is necessary find estimates for such data nonetheless, it may be possible to revise the starting point for the search to one from which it will converge.

Fitting a line to the data by eye may give a good place to start. It is also worth trying another program, as minor differences in the way the program selects its starting point or implements the iterative algorithm give them different sensitivities.

Reference notes

Some approximations for calculating $\Phi(z)$ and $Z(p)$, sufficiently accurate for calculation of d' and the like, are given by Zelen and Severo (1964). The standard maximum-likelihood estimation algorithm used to fit the signal-detection model to several conditions was originally published by Dorfman and Alf (1968a, 1968b). The background to these methods are found are discussed in most advanced statistics texts. Several implementations of the procedure have been published, and a summary of programs is given in Swets (1996). I will make my version of these programs available on the web site mentioned in the preface.

Exercises

3.1. Suppose that $\mu_s = \sigma_s = 1.5$ in the Gaussian signal-detection model.

a. Draw the isosensitivity function in Gaussian coordinates.
b. Convert this function to an operating characteristic in probability coordinates.

3.2. Two detection conditions give estimates of $\widehat{d'_1} = 1.05$, $\widehat{\lambda}_1 = 1.03$ and $\widehat{d'_2} = 1.70$, $\widehat{\lambda}_2 = 0.38$. Use these results to fit an unequal-variance model to the pair of conditions. Draw the operating characteristic.

3.3. Six detection conditions (A to F) are run with the same signal and observer but with the bias manipulated. The response frequencies observed are

	Noise		Signal	
	YES	NO	YES	NO
A	264	36	294	6
B	168	132	273	27
C	102	198	252	48
D	30	270	198	102
E	17	283	171	129
F	2	298	108	192

a. Estimate d' and $\log\beta$ for each condition, based on the equal-variance Gaussian model. Is there a pattern to these values? Why is this analysis unsatisfactory?
b. Sketch (by eye) an isosensitivity function for these data in standard coordinates.
c. Plot the data in Gaussian coordinates and draw a straight line by eye through them.
d. Use your line to decide whether $\sigma_s^2 = \sigma_n^2$.
e. Estimate the parameters of whichever model you chose in the previous step.

3.4. Use a program to fit the data from Problem 3.3. Compare the results to the estimates made in this problem.

3.5. Suppose that the hit rate and false-alarm rate for detection of visual stimulus are collected from five different subjects. Explain why it would be inappropriate to use the five points to create an operating characteristic as described in this chapter.