Reverse Engineering of Receiver Operating Characteristic Curves

by

Catherine Medlock

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of Masters of Engineering in Electrical Engineering and Computer Science
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2017

© Massachusetts Institute of Technology 2017. All rights reserved.
Reverse Engineering of Receiver Operating Characteristic Curves

by

Catherine Medlock

Submitted to the Department of Electrical Engineering and Computer Science on September 1, 2017, in partial fulfillment of the requirements for the degree of Masters of Engineering in Electrical Engineering and Computer Science

Abstract

Receiver operating characteristic (ROC) curves have played a crucial role in the design and evaluation of radar systems for many decades. More recently, their use has spread to a variety of fields including clinical decision-making and machine learning. The common thread in all of these fields is an interest in binary hypothesis testing problems, in which the objective is to use an observation of a random variable of interest, sometimes referred to as a score variable, to infer the answer to a yes-no question. The standard progression of a binary hypothesis testing system from an observation of a score variable, to a set of parameterized decision rules with a binary outputs, to an ROC curve that characterizes the performance of those decision rules is well-understood. Thus, it comes as no surprise that an ROC curve only contains partial information about the problem for which it was designed. In this thesis, a key objective is to find ways of “reverse engineering” ROC curves in order to infer as much information as possible about the underlying binary hypothesis testing problems. We focus specifically on ROC curves that were or could have been constructed using likelihood ratio tests on an actual score variable, which we refer to as LRT-consistent ROC curves. For example, a specific LRT-consistent ROC curve does not uniquely determine the conditional distributions of the score variable used to generate it. A main result is a method for starting with an LRT-consistent ROC curve and using it to construct the conditional distributions of an unlimited number of score variables that could have been used to produce it. One interpretation of the result is as a characterization of the family of score variables that lead to the same ROC curve. This approach is extended to the similar problem of characterizing the family of score variables that lead to the same a set of LRT decision rules.

Thesis Supervisor: Alan V. Oppenheim
Title: Ford Professor of Engineering
Acknowledgments

I would first like to express my immense gratitude to my research advisor, Alan Oppenheim. I feel extremely lucky to have taken the Discrete-Time Signal Processing course the last time he lectured it in Fall 2015, and even more lucky that an interesting project (that would eventually turn into the spark for this thesis) happened to be available at that time. Al has done far more for me than simply providing academic guidance and encouragement. He has been an incredible mentor who showed, and continues to show, me what it means to be a creative researcher and effective teacher. Our twice-weekly discussions about ROC curves were sometimes good-natured arguments, usually less than two-and-a-half hours, and always the highlight of my day. I admire Al’s ability to identify promising research directions and explain complex technical concepts in addition to his willingness to bend over backwards to help his students and the people around him. I can’t wait to “wander in the wilderness” with Al over the next few years.

I have also had the good fortune of working with many members of the Digital Signal Processing Group (in alphabetical order): Tom Baran, Petros Boufounos, Dan Dudgeon, Meir Feder, Hsin-Yu Jane Lai, Tarek Lahlou, Pablo Martinez-Nuevo, Maya Said, Guolong Su, and Jim Ward. Our discussions in the two group meetings related to this thesis led to important technical contributions. Three members of the group had a more direct impact on the thesis. To Guolong: Your identification of the constraints imposed on a score variable by a specific set of LRT decision regions was extremely helpful. Thank you also for explaining your method of designing a deterministic decision rule to achieve an arbitrary achievable operating point to me. To Tom: Your realization that Guolong’s constraints could be discretized and implemented as a linear program ultimately became a main result of the thesis. Thank you for taking the time to meet with me to discuss further ideas about the linear programming approach, and for giving me writing tips. To Jim: Your perspective as an expert in radar was valuable for the thesis and impacted the way I think about ROC curves. Thank you for taking the time to discuss the thesis with me throughout
the year, and especially for always caring about and asking how I’m doing, even
outside of 6.341 or DSPG.

I first became exposed to ROC curves as part of a summer internship with Digital
Cognition Technologies, Inc. (DCT). I am grateful to my direct supervisor, William
Souillard-Mandar, for being an attentive mentor, a good friend, and a constant source
of valuable suggestions for the project. I would also like to thank my other supervisors
at DCT (in alphabetical order): Randy Davis, Bruce Musicus, and Dana Penney for
their guidance and for helping me to understand the difficult process of taking an
idea from a research environment into the real world.

Last but not least, I would like to thank my family and friends for their support.
To my parents: Thank you for listening so patiently while I tried to work out ideas
about ROC curves on the fly, even though they didn’t usually make sense. Your
medical expertise provided a valuable perspective on binary hypothesis testing in the
clinical setting. To my mother: Thank you for all the times you surprised me with
blueberries, baloney sandwiches, M&M’s, and coffee right before a class, or in my
office, or sitting on a street corner that you know I pass between classes. It always
made my day. To my father: Thank you for giving me so many rides to and from
campus over the years, riding in the car with you has become one of my favorite
things to do. To Mariko, Kimiko, and Keiko: The fact that all three of you are ready
and willing to support me at any time, whether it be through a lunchtime walk along
the river or a late night phone call, means a lot. Thanks, and expect a lot more walks
and calls in the years to come! To Thomas: Thank you for telling so many goofy
jokes, for wearing your hat, for cooking me so many meals, and for letting me listen
to songs on repeat. To Miles: Gagu says thank you for our long and enlightening
conversations about “gucks.”
Contents

1 Introduction 15
   1.1 Outline and Objectives of the Thesis 18

2 Background 21
   2.1 Overview of Binary Hypothesis Testing 22
      2.1.1 Deterministic vs. Random Decision Rules 23
   2.2 Minimum Probability of Error (MPE) Decision Rules 24
   2.3 Neyman-Pearson Decision Rules 25
   2.4 ROC Curves Constructed Using LRTs 27
      2.4.1 Necessary Condition for LRT-Consistency 27
      2.4.2 Sufficient Condition for Equivalence of LRTs and Score Variable
            Threshold Tests (SVTs) 29
   2.5 The Ideal ROC Curve 30
   2.6 Area Under an ROC Curve 32
   2.7 Achievable Region of the $P_F-P_D$ Plane 35
      2.7.1 The Patient and The Two Doctors 39

3 Achievable Operating Points Using Deterministic Decision Rules 43
   3.1 “Addition” and “Subtraction” of Achievable Operating Points 45
   3.2 Method for Achieving an Arbitrary Operating Point in the Achievable
       Region 47
   3.3 Example 50
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Family of Score Variables Consistent with a Given ROC Curve</td>
<td>57</td>
</tr>
<tr>
<td>4.1</td>
<td>Sufficient Condition for LRT-Consistency</td>
<td>58</td>
</tr>
<tr>
<td>4.2</td>
<td>Construction of Score Variables Consistent with a Given ROC Curve</td>
<td>60</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Special Case: The Likelihood Ratio as a Score Variable</td>
<td>62</td>
</tr>
<tr>
<td>4.3</td>
<td>Alternatives to the AUC</td>
<td>63</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Conditional Expectations of the Likelihood Ratio</td>
<td>64</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Area Under a Geometrically Weighted ROC Curve</td>
<td>66</td>
</tr>
<tr>
<td>4.4</td>
<td>Examples</td>
<td>68</td>
</tr>
<tr>
<td>5</td>
<td>Family of Score Variables Consistent with a Given Set of LRT Decision Regions</td>
<td>73</td>
</tr>
<tr>
<td>5.1</td>
<td>Linear Programming Formulation</td>
<td>75</td>
</tr>
<tr>
<td>5.2</td>
<td>Examples</td>
<td>78</td>
</tr>
<tr>
<td>6</td>
<td>Summary and Future Work</td>
<td>81</td>
</tr>
</tbody>
</table>
List of Figures

1-1 The central component of a binary hypothesis testing system is a decision rule that infers the answer to a yes-no question. .................. 16

1-2 Each operating point on an ROC curve represents the rates of detection and false alarm achieved by a decision rule with a specific value of $\eta$. 16

2-1 Formalization of the components of a binary hypothesis testing system. 22

2-2 (a) Two conditional Gaussian PDFs with different means ($\mu_0 = 0$, $\mu_1 = 1$) and unit variance along with (b) the LRT ROC curve, which is identical to the SVT ROC curve. .................. 30

2-3 (a) Two conditional Gaussian PDFs with zero mean and different variances ($\sigma_0^2 = 1/2$, $\sigma_1^2 = 1$) along with (b) the LRT and SVT ROC curves. .................. 31

2-4 (a) Example of the ideal scenario in which the supports of the conditional PDFs of the score variable are non-overlapping. (b) All points on the ideal, step function curve can be achieved by including appropriate portions of the supports of $f_0(\cdot)$ and $f_1(\cdot)$ in the decision region $D_1$. (c) If only LRTs are used, there are only three achievable operating points. 32

2-5 Achievable region for the Gaussian conditional PDFs in Figure 2-3. 37
Empirical ROC curves of (a) Doctor A and (b) Doctor B. Each dot represents the relative frequencies of false alarm and detection achieved by SVTs on the Doctors’ respective score variables. The chosen operating points are marked by $\times$’s and are both located at the point $(P_F, P_D) = (0.273, 0.732)$. Doctor A achieved these relative frequencies using an SVT with threshold $\omega_A$. Doctor B achieved the relative frequencies of the operating points on the endpoints of the dashed line using SVTs with thresholds $\gamma_B$ and $\gamma_C$, but never achieved the relative frequencies of the desired operating point. If he tests Patient Z, he will use a randomization between the SVTs with thresholds $\gamma_B$ and $\gamma_C$.

Decision regions of Doctor A and Doctor B. If Patient Z’s heart rate falls between $\gamma_B$ and $\gamma_C$, then Doctor B flips a fair coin to decide whether to deem him healthy or ill, i.e., whether to include that heart rate in $\mathcal{D}_0$ or $\mathcal{D}_1$.

The unique line of slope 1 through a desired operating point ($\star$).

The unique pair of points on the curve that have the same change in both $P_F$ and $P_D$ as the points from Step 1 of the procedure.

(a) Conditional PDFs of the score variable used in this section and (b) the associated ROC curve. The ‘$\times$’ marks the desired operating point.

(a) Unique line of slope 1 through the desired operating point, as well as (b) the unique pair of points on the curve with changes in $P_F$ and $P_D$ both equal to $d$.

(a) Decision region $\mathcal{D}_1^*$ of the deterministic decision rule. (b) Empirical values of $P_F$ and $P_D$, each obtained over $10^6$ independent trials. As expected, they are consistent with the desired operating point to within a fraction of a percent.

The desired operating point could also be reached using randomization between the two deterministic operating points $(\mathcal{D}_1(\eta_0), \alpha_0, \beta_0)$ and $(\mathcal{D}_1(\eta_3), \alpha_3, \beta_3)$, as described in Section 2.7.1.
3-7  (a) Decision region $\mathcal{D}_1$ of the randomized decision rule. If the observed value of $S$ falls within the shaded region then the rule decides $\hat{H}_1$. If it falls in the cross-hatched region then the rule decides $\hat{H}_1$ with probability 0.495 and $\hat{H}_0$ with probability 0.505. (b) Empirical distribution of $P_F$ over $10^3$ trials. Both the mean value and standard deviation were rounded to three significant digits.

4-1  (a) Example of the ideal scenario in which the supports of the conditional PDFs of the score variable are non-overlapping. (b) All points on the ideal, step function curve can be achieved by including appropriate portions of the supports of $f_0(\cdot)$ and $f_1(\cdot)$ in the decision region $\mathcal{D}_1$.

4-2  Depiction of an ROC curve as the projection (dashed line) of a three-dimensional trajectory (solid line) onto the $P_F$-$P_D$ plane.

4-3  (a) Conditional PDFs of $S$ along with (b) the corresponding LRT-ROC curve.

4-4  (a) Conditional PDFs of the score variable $R$ as described in Section 4.1, along with (b) its associated LRT-consistent ROC curve. As expected, $R$’s ROC curve is identical to the original ROC curve in Figure 4-3.

4-5  (a) Conditional PDFs of the score variable $L$ as described in Section 4.2.1, along with (b) its associated LRT-consistent ROC curve. As expected, $L$’s ROC curve is identical to the original ROC curve in Figure 4-3.

4-6  Empirical validation that the conditional PDFs of the likelihood ratio are as shown in Figure 4-5.
4-7  (a) Five ROC curves, each constructed using LRTs on a score variable whose conditional PDFs are Gaussian distributions with means 0 and 1 and variance $\sigma^2$. (b) Conditional expectation of the likelihood ratio vs. AUC. The monotonic relationship between the two metrics shows consistency between them in the case of ROC curves that do not intersect. (c) Second ($\mu_2$) and third ($\mu_3$) moments of the derivative vs. AUC. .......................................................... 71

5-1  Abstraction of the implementation of the set of LRT decision rules for PDFs $f_0(\cdot)$ and $f_1(\cdot)$. .......................................................... 74

5-2  (a) Original conditional PDFs ($f_0(\cdot)$ in blue dots, $f_1(\cdot)$ in orange dots) used to establish the likelihood ratio constraint. (b) Solution of the linear program in Equation 5.9 ($\hat{f}_0(\cdot)$ in blue dots, $\hat{f}_1(\cdot)$ in orange dots). Each component of the vector $c$ was drawn from a uniform distribution between 0 and 1. The resulting vector was normalized to have unit length. .......................................................... 78

5-3  (a) Original conditional PDFs ($f_0(\cdot)$ in blue dots, $f_1(\cdot)$ in orange dots) used to establish the likelihood ratio constraint. (b) Solution of the linear program in Equation 5.9 ($\hat{f}_0(\cdot)$ in blue dots, $\hat{f}_1(\cdot)$ in orange dots). Each component of the vector $c$ was drawn from a uniform distribution between 0 and 1. The resulting vector was normalized to have unit length. .......................................................... 79
(a) Original conditional PDFs ($f_0(\cdot)$ in blue dots, $f_1(\cdot)$ in orange dots) used to establish the likelihood ratio constraint. (b) Solution of the linear program in Equation 5.9 ($\hat{f}_0(\cdot)$ in blue dots, $\hat{f}_1(\cdot)$ in orange dots). Each component of the vector $c$ was drawn from a uniform distribution between 0 and 1. The resulting vector was normalized to have unit length. The PDFs have been truncated to a range in which their ratio lies in a small range around 1, since we observed that very small or very large ratios led to discontinuous solutions owing to numerical accuracy issues.
Chapter 1

Introduction

Receiver operating characteristic (ROC) curves have a long and rich history, starting with their inception in the radar community during World War II. [13] In that context, each point on an ROC curve represents a possible pair of values for the rate of target detection and rate of false alarm of a radar receiver. The role of these curves has since expanded to a variety of fields, where they are used to visualize the performance of general decision-making devices or systems. [1, 2, 8, 9, 10, 12, 14]. For example, signal detection in a noisy environment, screening and diagnosis of an illness from clinical tests, and damage detection in the parts of a ship or an airplane are all areas that benefit from the use of ROC curves.

All of these examples fall into the class of problems referred to as binary hypothesis testing problems, in which the basic objective is to use a measurement of a phenomenon of interest to infer the answer to a yes-no question, such as “Is the signal present?” or “Is the patient ill?” The typical components of a binary hypothesis testing system are shown in Figure 1-1. At the center is an optimal, parameterized decision rule that takes the observation and a parameter value $\eta$ as input, with the output as a binary “yes” or “no” decision. The probabilities of detection and false alarm that it achieves, denoted by $P_D$ and $P_F$, are defined as the conditional probabilities that the rule outputs “yes” given that the true answer is “yes” or “no”, respectively.

Note that the word “optimal” must be defined with respect to an appropriate
performance criterion that usually depends on $P_D$ and $P_F$. Ideally, we would have $P_D = 1$ and $P_F = 0$. However, there is more typically a tradeoff between the two that is controlled by the value of $\eta$. An ROC curve is a plot of $P_D$ vs. $P_F$ for all possible values of $\eta$, as illustrated in Figure 1-2. Each point on the curve is referred to as an operating point.

\[ \text{Measurement (s) \rightarrow Decision rule \rightarrow Decision (Yes/No)} \]
\[ \text{Parameter (\eta)} \]

Figure 1-1: The central component of a binary hypothesis testing system is a decision rule that infers the answer to a yes-no question.

As an example, consider the case in which the phenomenon of interest is a patient who is either healthy or ill. The observation is some type of diagnostic measurement of the patient, and the decision rule might be a test that compares the measurement to a predetermined threshold value. If the measurement exceeds the threshold, then the patient is deemed ill. Otherwise the patient is deemed healthy. Each point on the associated ROC curve represents the rate of disease detection and rate of false alarm achieved by the test for a specific threshold value. More generally, the decision rule is not restricted to be a simple threshold test on the measurement.

\[ \text{Prob. of detection (P_D)} \]
\[ \text{Prob. of false alarm (P_F)} \]

Figure 1-2: Each operating point on an ROC curve represents the rates of detection and false alarm achieved by a decision rule with a specific value of $\eta$. 
The techniques used to analyze and compare ROC curves are fairly field-specific. For instance, researchers in the radar community often model ROC curves parametrically. The parameter values are frequently related to the signal-to-noise (SNR) level of the received signal when a target is present. [11] It is then common practice to plot families of ROC curves representing different SNR levels. When deciding where and on which curve to operate, practical concerns involving the limitations of the equipment and the amount of resources available to support actions caused by detection of a target are of prime importance. Note that in this case, each operating point has two separate parameters: the SNR of the corresponding curve and the parameter value of the decision rule associated with that point. Different points on the same curve share the same SNR, but not the same parameter value.

ROC curves may also be used in the radar setting to compare the performance of different receiver configurations. A typical criterion used in this case, referred to as the Neyman-Pearson criterion and discussed further in Section 2.3, is the highest achievable $P_D$ subject to an upper bound on $P_F$. The bound would typically represent the highest rate of false alarm that is tolerable in practice. An important point is that if the value of $P_F$ that will be used in practice is fixed (for instance, to be equal to its upper bound), then each ROC curve is characterized by the single operating point with that value of $P_F$.

Researchers in the clinical decision-making and radiology communities, on the other hand, are often interested in characterizing and comparing entire ROC curves as opposed to single points with a fixed rate of false alarm. A common objective is to compare the performance of different observation types in detecting the presence of an illness. In the field of machine learning, the comparison is usually between different classification algorithms instead of observation types. A popular metric in all of these fields is the area under an ROC curve (AUC). The AUC has an intuitively plausible interpretation and has proven useful in many situations. However, the underpinnings of its validity have been called into question. [4, 7]. Section 2.6 contains a more complete discussion of this topic.
1.1 Outline and Objectives of the Thesis

The debate about the AUC is just one piece of the larger issue of the information loss that occurs when condensing an entire ROC curve into a single number for the purpose of comparing it to other ROC curves. More generally, every step in the solution and analysis of a binary hypothesis testing problem results in a loss of information – from phenomenon of interest to observation, observation to set of parameterized decision rules, or set of decision rules to ROC curve. While this forward progression has been studied extensively and is very well-understood, to our knowledge, little attention has been paid to the progression in the opposite direction. The “reverse engineering” of a binary hypothesis testing system is the approach we take in this thesis.

More specifically, a key objective is to characterize the information that is lost at each step in the forward progression. The discussion in Chapter 4, for instance, is centered on the fact that different observation types (later termed score variables) can lead to different sets of optimal decision rules but the same ROC curve. The focus of Chapter 5 is how different score variables can lead to the same set of optimal decision rules but different ROC curves.

We start in Chapter 2 by reviewing well-established topics including Neyman-Pearson decision rules, properties of optimal Neyman-Pearson ROC curves, and the meaning of the AUC. The use of randomization to achieve any operating point on the straight line between two other operating points is also discussed. This leads to the concept of the region of achievable operating points for a given ROC curve. In Chapter 3, we show that under mild assumptions on the ROC curve, every operating point in the achievable region can also be achieved without randomization, i.e., with a deterministic decision rule. We present a procedure for designing a deterministic rule that achieves an arbitrary operating point in the achievable region and demonstrate using simulated data.

The theme of Chapters 4 and 5 is the “reverse engineering” objective described above. The main result of Chapter 4 is a method for taking an ROC curve and generating score variables that could have been used to produce it. We show in Chapter
5 that a linear program can be used to generate score variables that are consistent with a given set of optimal decision rules. With both of these chapters, it will be important to keep in mind that the ROC curves and decision rules considered are not required to be associated with real binary hypothesis testing problems. Rather, they simply need to satisfy certain properties (to be defined in Chapter 2) so that they could have been designed for real problems. A summary of the thesis and potential extensions are given in Chapter 6.
Chapter 2

Background

In this chapter, we introduce the notation and terminology used in the rest of the thesis and establish our perspective on several topics that play a fundamental role in later chapters. Among the most important is the topic of likelihood ratio tests (LRTs), which form the optimal family of decision rules with respect to the minimum probability of error (MPE) criterion. They are also optimal with respect to the Neyman-Pearson criterion, where the objective is to maximize the rate of detection ($P_D$) of a decision rule subject to an upper bound on the rate of false alarm ($P_F$).

Following a general review of binary hypothesis testing problems in Section 2.1, we discuss the MPE and Neyman-Pearson criteria in Sections 2.2 and 2.3, respectively. In Section 2.4 we derive the important fact that under certain reasonable assumptions, ROC curves constructed using LRTs are necessarily strictly concave. The distinction is then made between LRTs and threshold tests on an observation. We emphasize that, in general, these two types of decision rules are not equivalent. The ideal scenario in which perfect probabilities of detection and false alarm can be achieved is discussed along with the associated ideal ROC curve in Section 2.5. The focus of Section 2.6 is the area under an ROC curve (AUC), which is a common metric used to compare different ROC curves. We discuss its interpretation in addition to the debate around its validity. The chapter concludes in Section 2.7 with a discussion of the achievable region of the $P_F$-$P_D$ plane using randomization between deterministic decision rules.
2.1 Overview of Binary Hypothesis Testing

The basic components of a binary hypothesis testing system are summarized in Figure 2-1, which is essentially a formalization of Figure 1-1. The first component is a hidden random variable $H$ whose two possible values represent the possible states of the phenomenon of interest. The prior probabilities, denoted by

\[ P(H = H_0) = P_0 \]
\[ P(H = H_1) = P_1, \]

define the distribution of $H$ in the absence of any other information.

\[ S = \begin{cases} S_0 & \text{if } H = H_0 \\ S_1 & \text{if } H = H_1 \end{cases} \]


Figure 2-1: Formalization of the components of a binary hypothesis testing system.

The goal is to use an observation of another random variable which we refer to as the score variable, denoted by $S$ in Figure 2-1, to infer the value of $H$. All score variables are assumed to be continuous, but the results in subsequent chapters can be extended to the discrete case.

In the context of ROC curves, the most important features of a score variable are its distributions conditioned on the two possible values of $H$. An arbitrary score variable $S$ with conditional PDFs $f_0(\cdot)$ and $f_1(\cdot)$, defined by

\[ f_0(\cdot) = f_{S|H}(\cdot \mid H_0) \]
\[ f_1(\cdot) = f_{S|H}(\cdot \mid H_1), \]

will be referred to using the notation

\[ S \sim f_0(\cdot), f_1(\cdot). \]

(2.3)
Thus, the phrase “the score variable $S \sim f_0(\cdot), f_1(\cdot)$” is simply a convenient shorthand for “the random variable $S$ whose conditional PDFs under $H_0$ and $H_1$ are $f_0(\cdot)$ and $f_1(\cdot)$, respectively.” We will use $S \sim f_0(\cdot), f_1(\cdot)$ as a prototypical score variable for the remainder of the chapter.

Note that in this thesis, two score variables are considered to be distinct only if their conditional PDFs are different. A potentially confusing implication of this choice of terminology is that two distinct score variables, say $S \sim f_0(\cdot), f_1(\cdot)$ and $T \sim g_0(\cdot), g_1(\cdot)$, could represent the same physical quantity. For example, in a clinical setting the measurement of interest might be the patient’s heart rate. The score variable $S \sim f_0(\cdot), f_1(\cdot)$ might represent the heart rate of healthy or ill patients between 0 and 5 years old, while the score variable $T \sim g_0(\cdot), g_1(\cdot)$ might represent the heart rate of healthy or ill patients between 90 and 95 years old. In this case the random variables $S$ and $T$ would both have units of beats per minute, but their conditional PDFs would be specific to the corresponding demographic.

2.1.1 Deterministic vs. Random Decision Rules

The central component of the system in Figure 2-1 is a parameterized decision rule that is used to process the observed value of the score variable. We will focus mainly on deterministic decision rules, which partition the real line into two decision regions, denoted by $\mathcal{D}_0$ and $\mathcal{D}_1$. If the observed value of the score variable is in $\mathcal{D}_0$ then the rule decides $\hat{H}_0$, otherwise it decides $\hat{H}_1$. Note that the $\hat{\cdot}$ symbol denotes an inferred value of $H$ as opposed to its true, hidden value. Also note that each region may consist of multiple, disjoint intervals. They are not necessarily half-lines with a single shared boundary. We will return to this point in Section 2.4.2. A simple but significant remark is that if a specific value of $S$ is observed multiple times, the output of a deterministic decision rule will be the same every time. This differs from a random decision rule that maps every possible value of the score variable to a probability of inferring $\hat{H}_1$.

Nearly all of the discussion in subsequent chapters centers on the region $\mathcal{D}_1$ as opposed to $\mathcal{D}_0$. We say that a decision rule with decision regions $\mathcal{D}_0$ and $\mathcal{D}_1$ obtains,
or achieves, the operating point \((P_F, P_D) = (\alpha, \beta)\) if

\[
\alpha = \mathbb{P}(\hat{H}_1 | H_0) = \int_{\mathcal{D}_1} ds \cdot f_0(s) \tag{2.4a}
\]

\[
\beta = \mathbb{P}(\hat{H}_1 | H_1) = \int_{\mathcal{D}_1} ds \cdot f_1(s). \tag{2.4b}
\]

Of course, focusing only on \(\mathcal{D}_1\) is not restrictive since only one of the decision regions is needed to fully specify a decision rule.

### 2.2 Minimum Probability of Error (MPE) Decision Rules

In this section and the next, we discuss two common methods of identifying the “optimal” operating point on an ROC curve, the MPE criterion and the Neyman-Pearson criterion. One natural choice for an “optimal” decision rule is one that minimizes the probability of an incorrect decision for each possible value of \(S\), \(\mathbb{P}\)\(\text{(Error} \mid S = s)\). The optimal rule chooses the hypothesis that has the maximum a posteriori (MAP) probability conditioned on the observation \(S = s\),

\[
\mathbb{P}(H = H_1 \mid S = s) \overset{\hat{H}_1}{\underset{\hat{H}_0}{\gtrless}} \mathbb{P}(H = H_0 \mid S = s). \tag{2.5}
\]

Applying Bayes’ rule to both sides leads to the simpler form

\[
\frac{f_1(s)}{f_0(s)} \overset{\hat{H}_1}{\underset{\hat{H}_0}{\gtrless}} \frac{P_0}{P_1}. \tag{2.6}
\]

The quantity on the left-hand side of the inequality is referred to as the likelihood ratio associated with the value \(s\). A decision rule that performs a threshold test on the likelihood ratio is called a likelihood ratio test (LRT).

It makes intuitive sense and is straightforward to show that minimizing the probability of error for each individual value of \(S\) also minimizes the expected value of
\( \mathbb{P}(\text{Error} \mid S = s) \) over all values of \( S \), \( P(\text{Error}) \). This is because

\[
\mathbb{P}(\text{Error}) = \int ds \cdot p_S(s) \cdot \mathbb{P}(\text{Error} \mid S = s)
\]  

(2.7)

where

\[
p_S(s) = P_0 \cdot f_0(s) + P_1 \cdot f_1(s).
\]  

(2.8)

The probability density \( p_S(s) \) is non-negative everywhere by definition, so minimizing \( \mathbb{P}(\text{Error} \mid S = s) \) for every \( s \) minimizes the entire integral in Equation 2.7. Therefore, a decision rule that uses the MAP strategy for every \( s \) achieves minimum probability of error.

In summary, the family of optimal decision rules with respect to the MPE criterion is the family of LRTs on \( S \) with thresholds ranging from 0 (optimal for \( P_0 = 0 \)) to \( +\infty \) (optimal for \( P_0 = 1 \)). We now show that this family is also optimal for the Neyman-Pearson criterion.

### 2.3 Neyman-Pearson Decision Rules

While the MPE criterion is often an appropriate choice of criterion, it is impractical if the priors are unknown and difficult to estimate. In addition, in some situations it is useful or even necessary to bound either \( P_F \) or \( P_D \). For instance, in the radar community \( P_F \) is often constrained to be below \( 10^{-6} \) since false detection of a target can trigger costly actions and a waste of expensive resources. A reasonable alternative is to maximize \( P_D \) subject to a given upper bound on \( P_F \). This is the Neyman-Pearson criterion. The associated optimal decision rule is an LRT,

\[
\frac{f_1(s)}{f_0(s)} \begin{cases} \overset{\tilde{H}_1}{\geq} \eta \\
\overset{\tilde{H}_0}{\leq} \end{cases}
\]  

(2.9)

where the threshold \( \eta \geq 0 \) is chosen so that \( P_F \) is exactly equal to its upper bound. [5] The decision regions of an LRT with threshold \( \eta \) will be denoted by \( \mathcal{D}_0(\eta) \) and
\[ \mathcal{D}_1(\eta), \]
\[ \mathcal{D}_0(\eta) = \{ s \mid \frac{f_1(s)}{f_0(s)} < \eta \} \]  
(2.10a)
\[ \mathcal{D}_1(\eta) = \{ s \mid \frac{f_1(s)}{f_0(s)} \geq \eta \}. \]  
(2.10b)

The associated probabilities of false alarm and detection will be denoted by \( \phi_F(\eta) \) and \( \phi_D(\eta) \),
\[ P_F = \phi_F(\eta) = \int_{\mathcal{D}_1(\eta)} ds \cdot f_0(s) \]  
(2.11a)
\[ P_D = \phi_D(\eta) = \int_{\mathcal{D}_1(\eta)} ds \cdot f_1(s). \]  
(2.11b)

Note from Equation 2.10 that smaller values of \( \eta \) correspond to larger regions \( \mathcal{D}_1(\eta) \). Therefore, from Equation 2.11, they also correspond to larger values of \( \phi_F(\eta) \) and \( \phi_D(\eta) \). We emphasize that \( \phi_F(\cdot) \) and \( \phi_D(\cdot) \) are specific to decision rules of the form of Equation 2.9, i.e., LRTs. They will not be used to denote the values of \( P_F \) and \( P_D \) for decision rules that are, for example, a threshold test on the score variable (discussed further in Section 2.4.2).

We will always assume that the function \( L(s) = \frac{f_1(s)}{f_0(s)} \) varies smoothly as a function of \( s \) and that it is never constant over any finite range. An important implication of this assumption is that the functions \( \phi_F(\cdot) \) and \( \phi_D(\cdot) \) are continuous, strictly decreasing functions of \( \eta \), and are therefore both invertible. We will show in Section 2.4 that the invertibility of \( \phi_F(\cdot) \) implies that the ROC curve obtained by plotting \( \phi_D(\eta) \) vs. \( \phi_F(\eta) \) for \( 0 \leq \eta < \infty \) is strictly concave. The results of subsequent chapters concerning strictly concave ROC curves can be extended to more general ROC curves, but the analysis is more complicated and gives little added insight.

Equation 2.9 says that the family of optimal Neyman-Pearson decision rules for a score variable \( S \sim f_0(\cdot), f_1(\cdot) \) is the family of LRTs on \( S \) with thresholds ranging from 0 to \(+\infty\), the same as the family of optimal MPE decision rules. We will refer to the curve of operating points achieved by these LRTs as the Neyman-Pearson ROC
curve or the LRT ROC curve of $S$. We will only be interested in the Neyman-Pearson criterion for the remainder of the thesis. The notation $\psi_D(\cdot)$ will be used to refer to the functional form of a specific LRT ROC curve,

$$P_D = \psi_D(P_F) = \phi_D(\phi_F^{-1}(P_F))$$  \hspace{1cm} (2.12)

since from Equation 2.11, $P_D = \phi_D(\eta)$ and $\eta = \phi_F^{-1}(P_F)$. This notation along with Equation 2.11 will be useful for distinguishing the functional form of an LRT ROC curve ($P_D$ as a function of $P_F$) from its implicit parameterized form ($P_F$ and $P_D$ as functions of $\eta$).

2.4 ROC Curves Constructed Using LRTs

The term LRT-consistent will be used to describe an ROC curve that was or could have been constructed using LRTs on an actual score variable, i.e., that is or could be the Neyman-Pearson ROC curve of an actual score variable. We show in Section 2.4.1 that (1) strict concavity and (2) endpoints at (0,0) and (1,1) are necessary conditions for an ROC curve to be LRT-consistent. The fact that they are also sufficient conditions is less well-known and is proven in Section 4.1. We address the difference between LRTs and threshold tests on a score variable (SVTs) in Section 2.4.2.

2.4.1 Necessary Condition for LRT-Consistency

The following statement asserts that all Neyman-Pearson ROC curves that were constructed using LRTs on an actual score variable, and that are therefore LRT-consistent by definition, necessarily have two important properties.

\textit{Statement 1}. The Neyman-Pearson ROC curve of an arbitrary score variable has endpoints at (0,0) and (1,1) and is strictly concave.

\textit{Proof}. Choose an arbitrary score variable and denote it by $S \sim f_0(\cdot), f_1(\cdot)$. Clearly, the operating points achieved by LRTs with thresholds $\eta = \infty$ and $\eta = 0$ will be at $(P_F, P_D) = (0,0)$ and $(P_F, P_D) = (1,1)$, respectively and will also be at the endpoints
of the Neyman-Pearson ROC curve of \( S \). The proof of strict concavity hinges on the following well-known\(^1\) property of Neyman-Pearson ROC curves: The slope of the curve at the point \((P_F, P_D) = (\phi_F(\eta), \phi_D(\eta))\) is equal to \( \eta \). To see why this is true, we start from Equation 2.11 and note that for fixed \( \eta = \eta^* \),

\[
\phi_D(\eta^*) = \int_{\mathcal{D}_1(\eta^*)} ds \cdot f_1(s) = \int_{\mathcal{D}_1(\eta^*)} ds \cdot L(s) \cdot f_0(s) = \int_{\eta^*}^{+\infty} d\ell \cdot g(\ell) \cdot \ell \cdot h_0(\ell)
\]

\[\text{(2.13a)}\]

\[
\phi_F(\eta^*) = \int_{\mathcal{D}_1(\eta^*)} ds \cdot f_0(s) = \int_{\eta^*}^{+\infty} d\ell \cdot g(\ell) \cdot h_0(\ell)
\]

\[\text{(2.13b)}\]

where \( L(s) = f_1(s)/f_0(s) \), \( g(\ell) \) is defined so that \( ds = g(\ell) \cdot d\ell \), and

\[
h_0(\ell) = \int_{s:L(s)=\ell} ds \cdot f_0(s).
\]

\[\text{(2.14)}\]

Evaluating the derivative of \( \psi_D(\cdot) \) at \( P_F^* = \phi_F(\eta^*) \),

\[
\psi'_D(P_F^*) = \phi'_D((\phi_F^{-1}(P_F^*)) \cdot ((\phi_F^{-1})'(P_F^*))
\]

\[\text{(2.15a)}\]

\[
= \frac{\phi'_D(\eta^*)}{\phi'_F(\eta^*)}
\]

\[\text{(2.15b)}\]

\[
= \frac{g(\eta^*) \cdot \eta^* \cdot h_0(\eta^*)}{g(\eta^*) \cdot h_0(\eta^*)}
\]

\[\text{(2.15c)}\]

\[
= \eta^*
\]

\[\text{(2.15d)}\]

as desired, where Equation 2.15a follows from the chain rule and Equation 2.15b follows from the fact that for an invertible function \( f(\cdot) \) such that \( y = f(x) \), \( (f^{-1})'(y) = 1/f'(f^{-1}(y)) \).

Since from Section 2.3 \( P_F = \phi_F(\eta) \) is a strictly decreasing function of \( \eta \), Equation 2.15 says that \( P_F \) is a strictly decreasing function of the slope of the ROC curve. Equivalently, the slope of the ROC curve is a strictly decreasing function of \( P_F \), which is the definition of strict concavity.

\[\Box\]

\(^1\)See, for example, [5].

28
2.4.2 Sufficient Condition for Equivalence of LRTs and Score Variable Threshold Tests (SVTs)

It is not unusual in some contexts to focus only on decision rules that have the form of threshold tests on the score variable (SVTs), i.e., for a score variable $S \sim f_0(\cdot), f_1(\cdot)$,

$$s \overset{\sim}{\gtrless} \gamma \quad (2.16)$$

for a constant $\gamma$. In order for an LRT with arbitrary threshold $\eta$ to be equivalent to this form for some $\gamma$, and vice versa, it is sufficient for the function $L(s) = f_1(s)/f_0(s)$ to be invertible. In particular, if $L(\cdot)$ is a strictly decreasing function, then the LRT decision regions from Equation 2.10 simplify to

$$\mathcal{D}_0(\eta) = \{s \mid L(s) < \eta\} = \{s \mid s < L^{-1}(\eta)\} \quad (2.17a)$$

$$\mathcal{D}_1(\eta) = \{s \mid L(s) \geq \eta\} = \{s \mid s \geq L^{-1}(\eta)\}. \quad (2.17b)$$

which are the same as the regions of an SVT with threshold $\gamma = L^{-1}(\eta)$. Therefore,

$$L(s) = \frac{f_1(s)}{f_0(s)} \overset{\sim}{\gtrless} \eta \quad \iff \quad s \overset{\sim}{\gtrless} L^{-1}(\eta). \quad (2.18)$$

If instead $L(\cdot)$ is a strictly increasing function, then the decision regions of the SVT are simply exchanged (i.e., as stated in Equation 2.19, the decision rule on the right infers $\hat{H}_0$ rather than $\hat{H}_1$ if $s$ exceeds the threshold),

$$L(s) = \frac{f_1(s)}{f_0(s)} \overset{\sim}{\gtrless} \eta \quad \iff \quad s \overset{\sim}{\gtrless} L^{-1}(\eta). \quad (2.19)$$

If the function $L(s)$ is only required to be monotonic, i.e., monotonically non-increasing or monotonically non-decreasing, then every LRT can be written as an equivalent SVT, but the reverse does not hold. There are SVTs that cannot be expressed as equivalent LRTs. The results in the rest of the thesis can be generalized to this case, but we will not consider it any further.
Figure 2-2: (a) Two conditional Gaussian PDFs with different means ($\mu_0 = 0$, $\mu_1 = 1$) and unit variance along with (b) the LRT ROC curve, which is identical to the SVT ROC curve.

Examples of scenarios in which $L(\cdot)$ is and is not invertible are shown in Figures 2-2 and 2-3. We continue to assume a score variable $S \sim f_0(\cdot), f_1(\cdot)$. In Figure 2-2, $f_0(\cdot)$ and $f_1(\cdot)$ are Gaussian PDFs with different means and unit variance. The dotted line superimposed on the plot in Figure 2-2(a) is the log of the likelihood ratio, $\log L(s) = (\log f_1(s) - \log f_0(s))$. Since for this case $\log L(s)$ is strictly decreasing in $s$, every LRT can be expressed as an equivalent SVT, and vice versa. Consequently, as shown in Figure 2-2(b), the LRT ROC curve is the same as the SVT curve. In Figure 2-3(a), on the other hand, $f_0(\cdot)$ and $f_1(\cdot)$ are Gaussian PDFs with equal mean and different variances. In this case the function $\log L(\cdot)$ is not invertible, so the LRT and SVT ROC curves are different.

### 2.5 The Ideal ROC Curve

We now consider the ideal situation in which we assume a decision rule that achieves $P_F = 0$ and $P_D = 1$. The associated ideal ROC curve will be important in discussions involving metrics used to compare multiple curves.

Assume we have a score variable $S \sim f_0(\cdot), f_1(\cdot)$. In order for it to be possible to achieve $P_F = 0$ and $P_D = 1$, $f_0(\cdot)$ and $f_1(\cdot)$ must never be non-zero for the same
value of $s$. An example is depicted in Figure 2-4(a). In this case, the operating point $(P_F, P_D) = (0, 1)$ can be achieved by defining $D_1$ to be all values of $s$ for which $f_1(\cdot)$ is non-zero, i.e., all values in the support of $f_1(\cdot)$. More generally, it is possible to reach any operating point $(P_F, P_D) = (0, \beta)$ by defining $D_1$ to be an appropriate portion of the support of $f_1(\cdot)$ (any portion over which $f_1(\cdot)$ integrates to $\beta$). Similarly, it is possible to reach any operating point $(P_F, P_D) = (\alpha, 1)$ by defining $D_1$ to be the entire support of $f_1(\cdot)$ in addition to a portion of the support of $f_0(\cdot)$. Values of $s$ for which $f_0(\cdot)$ and $f_1(\cdot)$ are both zero can be included in either $D_1$ or $D_0$, since they have no effect on the values of $P_F$ or $P_D$. By varying both $\alpha$ and $\beta$ between 0 and 1, we find that the ideal ROC curve is the step function shown in Figure 2-4(b). The term “ideal ROC curve” will be used to refer to this curve for the remainder of the thesis.

Note that the ideal ROC curve was not constructed using LRTs. This is because for conditional PDFs with non-overlapping supports, the likelihood ratio only takes on two values – it is 0 over the support of $f_1(\cdot)$, $\infty$ over the support of $f_0(\cdot)$, and undefined elsewhere. Therefore, from Equations 2.10 and 2.11, the only three possible operating points using LRTs are

1. $\eta = 0$, $(P_F, P_D) = (1, 1)$
Figure 2-4: (a) Example of the ideal scenario in which the supports of the conditional PDFs of the score variable are non-overlapping. (b) All points on the ideal, step function curve can be achieved by including appropriate portions of the supports of $f_0(\cdot)$ and $f_1(\cdot)$ in the decision region $\mathcal{D}_1$. (c) If only LRTs are used, there are only three achievable operating points.

2. $\eta = \infty$, $(P_F, P_D) = (0, 0)$

3. $0 < \eta < \infty$, $(P_F, P_D) = (0, 1)$

The associated ROC “curve” is the set of these points, as shown in Figure 2-4(c). The edge case in which the likelihood ratio is undefined for some values of $s$ is not considered any further in the thesis.

## 2.6 Area Under an ROC Curve

The area under an ROC curve (AUC) is commonly used to compare ROC curves corresponding to score variable types that represent different physical quantities (blood pressure vs. heart rate, for example) or different classification algorithms. Such a comparison is inherently difficult because of the fundamental difference between metrics used to compare individual decision rules and metrics used to compare entire ROC curves, which represent collections of decision rules. It is less clear how to interpret the latter in terms of realizable differences in performance, since ultimately
only a single rule can be used. Nevertheless, the AUC is widely used in the literature and in practice.

One motivation for the AUC is that curves that with larger AUCs are likely to be more similar to the ideal ROC curve described in Section 2.5. Indeed, if one ROC curve lies above another for every value of \( P_F \), the higher curve is always preferable with respect to both the MPE and Neyman-Pearson criteria, and also has a larger AUC. The most popular quantitative interpretation of the AUC is often stated in the context of a clinical test. Let \( H_0 \) and \( H_1 \) represent the absence and presence of an illness in a patient, respectively, with the score variable \( S \sim f_0(\cdot), f_1(\cdot) \) representing the patient’s blood pressure. Assume that the PDFs \( f_0(\cdot) \) and \( f_1(\cdot) \) are such that the function \( L(s) = f_1(s)/f_0(s) \) is strictly decreasing, implying that any LRT on \( S \) can be expressed as an equivalent SVT, and vice versa. In this case \( \phi_F(\cdot) \) and \( \phi_D(\cdot) \) are

\[
P_F = \phi_F(\eta) = \int_{L^{-1}(\eta)}^{\infty} ds \cdot f_0(s) \tag{2.20a}
\]

\[
P_D = \phi_D(\eta) = \int_{L^{-1}(\eta)}^{\infty} ds \cdot f_1(s) \tag{2.20b}
\]

in accordance with Equation 2.18. The AUC can therefore be expressed as

\[
AUC = \int_0^1 dP_F \cdot \psi_D(P_F) = \int_{-\infty}^{\infty} du \cdot f_0(u) \cdot \psi_D(\phi_F(u)) = \int_{-\infty}^{\infty} du \cdot f_0(u) \cdot \phi_D(u) \tag{2.21}
\]

where we have used Equation 2.20a to write

\[
\frac{dP_F}{ds} \bigg|_{s=u} = f_0(u) \tag{2.22}
\]

and Equation 2.12 to write

\[
\psi_D(\phi_F(s)) = \phi_D(\phi_F^{-1}(\phi_F(s))) = \phi_D(s). \tag{2.23}
\]
Finally, using Equation 2.20b to expand $\phi_D(s)$, we have

$$\text{AUC} = \int_{-\infty}^{\infty} du \cdot f_0(u) \int_u^{\infty} dv \cdot f_1(v). \tag{2.24}$$

The inner integral in Equation 2.24 is the probability that the blood pressure of a sick patient is higher than the value $u$. The AUC is the expectation of this quantity over all healthy patients,

$$\text{AUC} = \int_{-\infty}^{\infty} du \cdot f_0(u) \cdot P(\text{blood pressure of a sick patient} > u). \tag{2.25}$$

The following experiment, sometimes referred to as a two-alternative forced choice experiment, makes the interpretation more concrete.

1. From a large pool of healthy patients choose a patient at random such that every patient has an equal chance of being chosen. Denote the blood pressure of the chosen patient by $s_0$.

2. Apply the same procedure to a large pool of sick patients, and denote the patient’s blood pressure by $s_1$.

3. Compare $s_0$ to $s_1$ and infer that the patient with the higher blood pressure is the sick patient and that the other is the healthy one. The AUC is the probability that this conclusion is correct.

Note that this experiment is not consistent with Figure 2-1. The role of a decision rule is to produce a conclusion for a single observation, not to compare two observations. So while a high AUC might seem desirable on the surface, it is difficult to relate directly to practical situations.

Much of the debate about the AUC can be traced back to the fact that it is an aggregate metric for a collection of decision rules, rather than a single metric for the decision rule to be used in practice. For instance, Lobo [7] pointed out that the value of the AUC incorporates sets of operating points that are entirely impractical: It is hard to imagine a situation where one would operate in the upper right portion of
an ROC curve.² Hand [4] showed that the AUC can be expressed as the expected minimum loss incurred over a distribution of possible misclassification costs. Loss was defined as

\[ L(P_0, C_F, C_M) \equiv P_0 \cdot C_F \cdot P_F + P_1 \cdot C_M \cdot (1 - P_D) \]  

(2.26)

where \( C_F \) and \( C_M \) are the costs associated with a false alarm and a missed detection, respectively. The issue raised was that the distribution over \( C_F \) and \( C_M \) that is implicit in the AUC depends on the conditional PDFs of the score variable. This is problematic because the misclassification costs should depend only on external factors of the problem. As phrased in [4], “It is as if one chose to compare the heights of two people using rulers in which the basic units of measurement themselves depended on the heights.” ([4], p. 105)

### 2.7 Achievable Region of the \( P_F-P_D \) Plane

We now consider the problem of finding all achievable operating points for a given ROC curve. The curve is assumed to be the Neyman-Pearson ROC curve for the score variable \( S \sim f_0(\cdot), f_1(\cdot) \) and is therefore strictly concave according to the discussion in Section 2.4.1. In this section, an achievable operating point refers to any \((P_F, P_D)\) pair that can be obtained using a deterministic or random decision rule. This definition is adjusted in Chapter 3 to include only deterministic rules. As a reminder, we say that a deterministic decision rule with decision region \( D_1 \) obtains, or achieves, the operating point \((P_F, P_D) = (\alpha, \beta)\) if

\[
\alpha = \int_{D_1} ds \cdot f_0(s) \]  

(2.27a)

\[
\beta = \int_{D_1} ds \cdot f_1(s). \]  

(2.27b)

²The partial AUC, which only integrates over a specific region of an ROC curve, sidesteps this issue. However, it is not as widely used and does not stand up to other arguments against the AUC.
A random decision rule obtains the same operating point if

$$\alpha = \int ds \cdot \mathbb{P}(\hat{H}_1 \mid S = s) \cdot f_0(s) = \mathbb{E}[\mathbb{P}(\hat{H}_1 \mid H = H_0)] = \mathbb{E}[P_F] \quad (2.28a)$$

$$\beta = \int ds \cdot \mathbb{P}(\hat{H}_1 \mid S = s) \cdot f_1(s) = \mathbb{E}[\mathbb{P}(\hat{H}_1 \mid H = H_1)] = \mathbb{E}[P_D] \quad (2.28b)$$

where $\mathbb{P}(\hat{H}_1 \mid S = s)$ is the probability that the rule decides $\hat{H}_1$ conditioned on the observation $S = s$. Note that in Equation 2.28, $P_F = \mathbb{P}(\hat{H}_1 \mid H = H_0)$ and $P_D = \mathbb{P}(\hat{H}_1 \mid H = H_1)$ are random variables with expected values $\alpha$ and $\beta$, respectively. This is in contrast to Equation 2.27, in which $P_F = \alpha$ and $P_D = \beta$ are deterministic quantities. One interpretation of this distinction is that a deterministic decision rule can obtain the operating point $(\alpha, \beta)$ deterministically, whereas a random decision rule can only obtain it “on average.” We will discuss this distinction further in Section 2.7.1.

Note that the operating point of perfect decision making, $(P_F, P_D) = (0, 1)$, is achievable only if the supports of $f_0(\cdot)$ and $f_1(\cdot)$ are non-overlapping, as described in Section 2.5. If the supports are non-overlapping, the deterministic decision rule that achieves this point is any rule for which $D_1$ includes the entire support of $f_1(\cdot)$ but none of the support of $f_0(\cdot)$. The operating points along the line $P_D = P_F$, on the other hand, are achievable for any $f_0(\cdot)$ and $f_1(\cdot)$ using a random decision rule. Specifically, simply flipping a coin whose probability of heads is equal to $b$, then deciding $\hat{H}_1$ if it comes up heads and $\hat{H}_0$ otherwise, will achieve the operating point $(P_F, P_D) = (b, b)$.

We will show that for a given ROC curve, the achievable region is the region bounded by the original curve, $P_D = \psi_D(P_F)$, and its complementary curve below the diagonal, $P_D = 1 - \psi_D(1 - P_F)$. This is illustrated in Figure 2-5 for the Gaussian PDFs in Figure 2-3. Note that this statement assumes that the function $\psi_D(\cdot)$ is known, but makes no assumptions about $f_0(\cdot)$ and $f_1(\cdot)$. In fact, for a given ROC curve, $f_0(\cdot)$ and $f_1(\cdot)$ do not affect the set of achievable operating points.

The derivation starts with the fact that given a deterministic decision rule with
Figure 2-5: Achievable region for the Gaussian conditional PDFs in Figure 2-3.

decision regions $\mathcal{D}_0$ and $\mathcal{D}_1$ that achieves the operating point $(\alpha, \beta)$, it is possible to construct a complementary rule that achieves the point $(1 - \alpha, 1 - \beta)$. Specifically, let the decision regions of the complementary rule be $\mathcal{D}_0' = \mathcal{D}_1$ and $\mathcal{D}_1' = \mathcal{D}_0$. Then

\[
\begin{align*}
    P'_F &= \int_{\mathcal{D}_0'} ds \cdot f_0(s) = \int_{\mathcal{D}_0} ds \cdot f_0(s) = 1 - \int_{\mathcal{D}_1} ds \cdot f_0(s) = 1 - \alpha \\
    P'_D &= \int_{\mathcal{D}_1'} ds \cdot f_1(s) = \int_{\mathcal{D}_0} ds \cdot f_1(s) = 1 - \int_{\mathcal{D}_0} ds \cdot f_1(s) = 1 - \beta
\end{align*}
\]

(2.29)
as desired. Extending this result to all points on the ROC curve described by $P_D = \psi_D(P_F)$ leads to the complementary curve $P_D = 1 - \psi_D(1 - P_F)$.

Next we show that it is always possible to achieve any operating point on the line segment connecting two other achievable operating points. [5] Assume the desired operating point, denoted by $(\alpha_1, \beta_1)$, is located on the line segment connecting two achievable operating points $(\alpha_0, \beta_0)$ and $(\alpha_2, \beta_2)$. Let the decision rules associated with the latter two points be Rule0 and Rule2, respectively. Further assume that the distance between $(\alpha_0, \beta_0)$ and $(\alpha_1, \beta_1)$ is a fraction $b$ of the distance between $(\alpha_0, \beta_0)$
and \((\alpha_2, \beta_2)\), which implies

\[
\alpha_1 = b \cdot \alpha_0 + (1 - b) \cdot \alpha_2 \quad (2.30a)
\]
\[
\beta_1 = b \cdot \beta_0 + (1 - b) \cdot \beta_2. \quad (2.30b)
\]

Then the desired operating point can be achieved as follows: When the score variable \(S\) is observed, flip a coin whose probability of heads is equal to \(b\). If the coin comes up heads then use \(Rule0\) to make the decision, otherwise use \(Rule2\). The expected probabilities of false alarm and detection are

\[
\mathbb{E}[P_F] = \mathbb{P}(H) \cdot \alpha_0 + \mathbb{P}(T) \cdot \alpha_2 = b \cdot \alpha_0 + (1 - b) \cdot \alpha_2 = \alpha_1 \quad (2.31a)
\]
\[
\mathbb{E}[P_D] = \mathbb{P}(H) \cdot \beta_0 + \mathbb{P}(T) \cdot \beta_2 = b \cdot \beta_0 + (1 - b) \cdot \beta_2 = \beta_1 \quad (2.31b)
\]

as desired, where \(\mathbb{P}(H)\) and \(\mathbb{P}(T)\) are the probabilities that the coin comes up heads and tails, respectively. The bias of the coin could be changed to achieve any point on the line between \((\alpha_0, \beta_0)\) and \((\alpha_2, \beta_2)\). Note that as in Equation 2.28, \(\alpha_1\) and \(\beta_1\) can be interpreted as the expected values of two random variables. In this case, the random variables are binary and take on the value \(\alpha_0\) (or \(\beta_0\)) with probability \(b\) and \(\alpha_1\) (or \(\beta_1\)) with probability \((1 - b)\). Again, the desired operating point is only achieved “on average.”

The conclusion about the achievable region now follows in a straightforward manner. First, note that all operating points above the Neyman-Pearson ROC curve are unachievable since by definition, every operating point is the maximum achievable \(P_D\) for some value of \(P_F\). It follows that all points below its complement are also unachievable. This is because if a point below the complementary curve were achievable, then the decision regions could be interchanged to achieve a point above the Neyman-Pearson curve. Randomization between pairs of decision rules can be used to reach all points in the region bounded by these two curves. One consequence of this result is that it is always possible to operate at any point on the concave hull of
multiple ROC curves. The concept of randomization is often addressed specifically to make this point.

2.7.1 The Patient and The Two Doctors

The concept of randomization between decision rules is fairly simple, but its legitimacy may be questioned were it used in the context of a serious diagnostic test. In this section we suggest a hypothetical scenario to illuminate some issues that may arise when randomization is used in practice. Of particular importance is the lack of reproducibility of the output of a random decision rule for a fixed input value, as mentioned in Section 2.1.1. In Chapter 3 we propose a procedure for constructing a deterministic decision rule that achieves an arbitrary operating point in the achievable region, thereby avoiding this problem.

A second important issue is the fact that no decision rule can have an output that is partially correct. For instance, consider a diagnostic test that correctly detected an illness in 99% of a large group of sick patients, i.e., its relative frequency of detection is 0.99. Regardless of this fact, the diagnosis for a single sick patient will either be 100% correct or 100% incorrect, it cannot be “99% correct.” The underlying problem is that the relative frequency of detection reflects the test’s performance on a large group of patients. It does not have a definitive interpretation in the context of a single instance of the test.

Suppose that Patient Z needs to be tested for Disease X, and that \( H_0 \) and \( H_1 \) correspond to the absence and presence of the disease, respectively. Furthermore, Patient Z has the option of being tested by either Doctor A or Doctor B. Doctor A’s test is an SVT on his patient’s blood pressure, while Doctor B’s test is an SVT on his patient’s heart rate. Both doctors constructed empirical ROC curves by performing their tests on a groups of \( 10^7 \) healthy patients and \( 10^7 \) diseased patients for a range of threshold values. The results are shown in Figure 2-6. Assume for simplicity that the conditional distributions of the blood pressure and heart rate of the patients satisfy the sufficiency condition described in Section 2.4.2, so that every LRT can be expressed as an SVT and vice versa.
Figure 2-6: Empirical ROC curves of (a) Doctor A and (b) Doctor B. Each dot represents the relative frequencies of false alarm and detection achieved by SVTs on the Doctors’ respective score variables. The chosen operating points are marked by ×’s and are both located at the point \((P_F, P_D) = (0.273, 0.732)\). Doctor A achieved these relative frequencies using an SVT with threshold \(\omega_A\). Doctor B achieved the relative frequencies of the operating points on the endpoints of the dashed line using SVTs with thresholds \(\gamma_B\) and \(\gamma_C\), but never achieved the relative frequencies of the desired operating point. If he tests Patient Z, he will use a randomization between the SVTs with thresholds \(\gamma_B\) and \(\gamma_C\).

The two ×’s mark the operating points the doctors will use to test Patient Z. They are both located at \((P_F, P_D) = (0.273, 0.732)\). Figure 2-7(a) depicts the decision regions of Doctor A’s chosen test, an SVT with threshold \(\omega_A\) on the patient’s blood pressure. This resulted in \(2.73 \cdot 10^6\) false alarms and \(7.32 \cdot 10^6\) correct detections when applied to the groups of healthy and diseased patients, respectively. In other words, the relative frequencies of false alarm and detection of Doctor A’s test are

\[
\begin{align*}
P_F^A &= \frac{2.73 \cdot 10^6}{10^7} = 0.273 \\
P_D^A &= \frac{7.32 \cdot 10^6}{10^7} = 0.732.
\end{align*}
\]

The decision regions of Doctor B’s chosen test are shown in Figure 2-7(b). The test is a randomization between two SVTs with thresholds \(\gamma_B\) and \(\gamma_C\) on the patient’s heart.

\(^3\)Note that these values are arbitrary and are not meant to be representative of a test that would be used by an actual doctor.
Figure 2-7: Decision regions of Doctor A and Doctor B. If Patient Z’s heart rate falls between $\gamma_B$ and $\gamma_C$, then Doctor B flips a fair coin to decide whether to deem him healthy or ill, i.e., whether to include that heart rate in $\mathcal{D}_0$ or $\mathcal{D}_1$.

rate. When Doctor B applied the SVT with threshold $\gamma_B$ to the groups of healthy and diseased patients, he obtained $5.20 \cdot 10^6$ false alarms and $9.80 \cdot 10^6$ correct detections. When he applied the SVT with threshold $\gamma_C$, he obtained $2.01 \cdot 10^5$ false alarms and $4.80 \cdot 10^6$ correct detections. Note that Doctor B never used randomization when performing these tests. However, if he tests Patient Z, Doctor B will flip a coin with probability of heads equal to 0.505 to decide which of these tests to use. If it comes up heads he will use the test with threshold $\gamma_B$, if it comes up tails he will use the one with threshold $\gamma_C$. Of course, if Patient Z’s heart rate is less than $\gamma_B$ or is greater than or equal to $\gamma_C$, both tests have the same output and there is no need to flip the coin. From Equation 2.31, the expected probabilities of false alarm and detection are

$$\mathbb{E}[P_F^B] = 0.495 \cdot 0.0201 + 0.505 \cdot 0.52 = 0.273$$

$$\mathbb{E}[P_D^B] = 0.495 \cdot 0.48 + 0.505 \cdot 0.98 = 0.732,$$

the same as the values in Equation 2.32. Equation 2.33 says that $P_F^B$ and $P_D^B$ are random variables whose expected values are $P_F^A$ and $P_D^A$, respectively.

The question is whether, given all of this information, one doctor is preferable to the other from Patient Z’s perspective. Ultimately both psychological and mathematical components come into play and there is no “correct” answer. However, it is still interesting to consider the issues that may arise when making the decision.

First note that by presenting their empirical ROC curves as characterizations of
their diagnostic tests, both doctors are implying that Patient Z is similar enough to the
groups of healthy/diseased patients (in ethnicity, age, and lifestyle, for example) that
the tests’ average performance on those groups is indicative of its future performance
on Patient Z. Since they used the same groups of patients when constructing their
respective ROC curves, this implication is equally valid (or invalid) for both and is
not a reason to choose one over the other.

Next consider the precise meanings of Equations 2.32 and 2.33. The values in
Equation 2.32 are deterministic quantities that have been empirically validated and
are reproducible. If Doctor A were to perform his test on all of the healthy patients
and all of the diseased patients again, he would obtain exactly the same results. The
values in Equation 2.33, on the other hand, are the expected values of two random
variables. They were not empirically validated since Doctor B did not perform the
randomized test on all of the healthy and diseased patients. Even if he were to do
so, the observed empirical values of $P_{B}^{F}$ and $P_{B}^{D}$ would likely change from one trial
to the next. As mentioned in the beginning of the section, there is undeniably some
psychological value in the reproducibility of Doctor A’s diagnosis. It is unsettling to
think that if Patient Z’s blood pressure were equal to that of another patient (and in
between $\gamma_{B}$ and $\gamma_{C}$), only one of them might be sent for further treatment. A more
thorough discussion of the psychology of decision making is contained in the book
Thinking Fast and Slow by Daniel Kahneman.[6]
Chapter 3

Achievable Operating Points Using Deterministic Decision Rules

The use of randomization to achieve any operating point in the achievable region of a given ROC curve is widely known. [5] A natural question is which, if any, of the points within the achievable region can be obtained using deterministic decision rules. As a reminder, we say that a deterministic decision rule with decision regions \( \mathcal{D}_0 \) and \( \mathcal{D}_1 \) obtains, or achieves, the operating point \((P_F, P_D) = (\alpha, \beta)\) if

\[
\alpha = \int_{\mathcal{D}_1} ds \cdot f_0(s) \quad (3.1a)
\]
\[
\beta = \int_{\mathcal{D}_1} ds \cdot f_1(s). \quad (3.1b)
\]

A random decision rule achieves the same operating point if

\[
\alpha = \int ds \cdot \mathbb{P}(\hat{H}_1 \mid S = s) \cdot f_0(s) = \mathbb{E}[\mathbb{P}(\hat{H}_1 \mid H = H_0)] = \mathbb{E}[P_F] \quad (3.2a)
\]
\[
\beta = \int ds \cdot \mathbb{P}(\hat{H}_1 \mid S = s) \cdot f_1(s) = \mathbb{E}[\mathbb{P}(\hat{H}_1 \mid H = H_1)] = \mathbb{E}[P_D] \quad (3.2b)
\]

where \( \mathbb{P}(\hat{H}_1 \mid S = s) \) is the probability that the rule decides \( \hat{H}_1 \) conditioned on the observation \( S = s \). We emphasize that in Equation 3.1, the probabilities of false alarm
and detection are the deterministic quantities $P_F = \alpha$ and $P_D = \beta$. In Equation 3.2, on the other hand, $P_F$ and $P_D$ can be viewed as random variables whose expected values are $\alpha$ and $\beta$, respectively. As mentioned in Section 2.7, one interpretation of this distinction is that a deterministic rule can achieve the operating point $(\alpha, \beta)$ deterministically, whereas the random rule only achieves it “on average.”

In this chapter, we show that all of the points in the achievable region of an LRT-consistent, i.e., strictly concave, ROC curve are achievable using deterministic rules. The main result is a procedure for constructing a deterministic rule that achieves an arbitrary achievable operating point. The procedure is described in Section 3.2 and demonstrated in Section 3.3. It may be useful in situations where the reproducibility and/or consistency of decisions is of high importance. For instance, as mentioned in Section 2.7.1, a randomized rule in the clinical setting might diagnose two patients differently even if their associated measurements were equal. An even more worrisome possibility is that if two measurements of a single patient were taken in quick succession and resulted in the same value, the randomized rule might diagnose that patient as healthy at one moment and ill the next. Neither of these situations seems desirable, and both could be avoided with a deterministic decision rule.

Throughout the chapter, we assume an LRT ROC curve that was constructed using the score variable $S \sim f_0(\cdot), f_1(\cdot)$. It is important to note that this score variable is not unique, i.e., there are other random variables (that may or may not represent the same physical quantity) with different conditional PDFs that could be used to generate the same LRT ROC curve. This topic is the subject of Chapter 4 and will not be discussed further here, but it is worth keeping in mind when considering the results of this chapter.

We begin by introducing additional notation that will be helpful in the subsequent derivations. Note that the region $\mathcal{D}_1$ in Equation 3.1 is not necessarily unique. It is for this reason that, in this chapter, we will always specify the decision region associated with a particular operating point. Specifically, the notation $(\mathcal{V}_i, \alpha_i, \beta_i)$ for an integer $i$

---

1Recall from Chapter 2 that the phrase “the score variable $S \sim f_0(\cdot), f_1(\cdot)$” is used as a shorthand for “the random variable $S$ whose conditional PDFs under $H_0$ and $H_1$ are $f_0(\cdot)$ and $f_1(\cdot)$, respectively.”
will be used to refer to the operating point $(P_F, P_D) = (\alpha_i, \beta_i)$, achieved by a decision rule with decision region $D_1 = V_i$. In other words, the notation $(V_i, \alpha_i, \beta_i)$ implies that

\[ \alpha_i = \int_{V_i} ds \cdot f_0(s) \]  
\[ \beta_i = \int_{V_i} ds \cdot f_1(s). \]  

(3.3a)  
(3.3b)

When the discussion is specifically about LRT decision regions, the region $V_i$ will correspond to the region $D_1(\eta_i)$, defined as in Chapter 2 by

\[ D_1(\eta_i) = \{ s \mid \frac{f_1(s)}{f_0(s)} \geq \eta_i \}. \]  

(3.4)

### 3.1 “Addition” and “Subtraction” of Achievable Operating Points

All of the decision regions considered in this thesis are subsets of the real line. Thus, before discussing ways of combining multiple decision regions in Section 3.1, we summarize commonly used notation from set theory that we use for the union and difference of two sets.

The union of the sets $A$ and $B$ is defined as the set that contains all elements that are in one or both of $A$ and $B$. It is denoted by

\[ A \cup B. \]  

(3.5)

We will only be concerned with cases where $A$ and $B$ are disjoint, i.e., they do not have any common elements. In this case, Equation 3.5 can be viewed as the “addition” of $B$ to $A$.

Next consider two sets $C$ and $D$. Their difference is denoted by

\[ C \setminus D \]  

(3.6)
and is defined as the set that contains all elements that are in $\mathcal{C}$ but not in $\mathcal{D}$. We will focus on situations where $\mathcal{D}$ is a subset of $\mathcal{C}$, in which case Equation 3.6 can be interpreted as the “subtraction” of $\mathcal{D}$ from $\mathcal{C}$.

We now extend the notions of the “addition” and “subtraction” of two decision regions to the same operations on two operating points. The purpose is to provide a framework for using existing decision regions as building blocks for constructing new ones that obtain arbitrary achievable operating points.

Consider two operating points $(\mathcal{V}_0, \alpha_0, \beta_0)$ and $(\mathcal{V}_1, \alpha_1, \beta_1)$ where the decision regions $\mathcal{V}_0$ and $\mathcal{V}_1$ are disjoint. We can “add” them together to form the operating point

$$(\mathcal{V}_0 \cup \mathcal{V}_1, \alpha_0 + \alpha_1, \beta_0 + \beta_1) \quad (3.7)$$

since

$$\int_{\mathcal{V}_0 \cup \mathcal{V}_1} ds \cdot f_0(s) = \int_{\mathcal{V}_0} ds \cdot f_0(s) + \int_{\mathcal{V}_1} ds \cdot f_0(s) = \alpha_0 + \alpha_1 \quad (3.8a)$$

$$\int_{\mathcal{V}_0 \cup \mathcal{V}_1} ds \cdot f_1(s) = \int_{\mathcal{V}_0} ds \cdot f_1(s) + \int_{\mathcal{V}_1} ds \cdot f_1(s) = \beta_0 + \beta_1. \quad (3.8b)$$

Following similar reasoning, consider two operating points $(\mathcal{V}_2, \alpha_2, \beta_2)$ and $(\mathcal{V}_3, \alpha_3, \beta_3)$ where the region $\mathcal{V}_3$ is a subset of $\mathcal{V}_2$. Then “subtracting” the second point from the first yields the operating point

$$(\mathcal{V}_2 \setminus \mathcal{V}_3, \alpha_2 - \alpha_3, \beta_2 - \beta_3) \quad (3.9)$$

since

$$\int_{\mathcal{V}_2 \setminus \mathcal{V}_3} ds \cdot f_0(s) = \int_{\mathcal{V}_2} ds \cdot f_0(s) - \int_{\mathcal{V}_3} ds \cdot f_0(s) = \alpha_2 - \alpha_3 \quad (3.10a)$$

$$\int_{\mathcal{V}_2 \setminus \mathcal{V}_3} ds \cdot f_1(s) = \int_{\mathcal{V}_2} ds \cdot f_1(s) - \int_{\mathcal{V}_3} ds \cdot f_1(s) = \beta_2 - \beta_3. \quad (3.10b)$$
3.2 Method for Achieving an Arbitrary Operating Point in the Achievable Region

Using the results of Section 3.1, it is now possible to construct a decision rule that achieves an arbitrary operating point in the upper half (above the line $P_D = P_F$) of the achievable region of a given ROC curve. Any operating point in the lower half can be obtained by exchanging the decision regions of the decision rule that obtains the appropriate complementary operating point in the upper half, as discussed in Section 2.7. We denote the desired operating point by $(\mathcal{D}_1^*, \alpha^*, \beta^*)$, where $\mathcal{D}_1^*$ is the unknown region that we wish to construct, and propose the following procedure.

1. (Figure 3-1) Draw the unique line of slope 1 through the desired operating point. Since the ROC curve is strictly concave, the line will have two points of intersection with the ROC curve. Denote the leftmost one by $(\mathcal{D}_1(\eta_0), \alpha_0, \beta_0)$, and define

$$d \equiv \alpha^* - \alpha_0 = \beta^* - \beta_0.$$  \hspace{1cm} (3.11)

![Figure 3-1](image)

Figure 3-1: The unique line of slope 1 through a desired operating point ($\star$).

2. (Figure 3-2) Find the two points on the curve, denoted by $(\mathcal{D}_1(\eta_1), \alpha_1, \beta_1)$ and
that satisfy
\[ d = \alpha_2 - \alpha_1 = \beta_2 - \beta_1. \] (3.12)

Note from Figure 3-2 that \( \eta_0 > \eta_1 > \eta_2 \). This is because the likelihood ratio threshold associated with each point on a Neyman-Pearson ROC curve decreases from left to right along the curve, as discussed in Section 2.4.1.

Figure 3-2: The unique pair of points on the curve that have the same change in both \( P_F \) and \( P_D \) as the points from Step 1 of the procedure.

3. The region \( \mathcal{D}_1^* \) can be written as
\[ \mathcal{D}_1^* = \mathcal{D}_1(\eta_0) \cup (\mathcal{D}_1(\eta_2) \setminus \mathcal{D}_1(\eta_1)) = \{ s \mid \eta_2 \leq \frac{f_1(s)}{f_0(s)} < \eta_1 \text{ or } \eta_0 \leq \frac{f_1(s)}{f_0(s)} \} \] (3.13)

and satisfies
\[ \alpha^* = \int_{\mathcal{D}_1^*} ds \cdot f_0(s) \] (3.14a)
\[ \beta^* = \int_{\mathcal{D}_1^*} ds \cdot f_1(s). \] (3.14b)

as desired. (See below for a derivation of Equation 3.14.)

The reasoning leading to Equations 3.13 and 3.14 is as follows. From Equation
3.4, the regions $\mathcal{D}_1(\eta_1)$ and $\mathcal{D}_1(\eta_2)$ are defined as

$$\mathcal{D}_1(\eta_1) = \{s \mid \frac{f_1(s)}{f_0(s)} \geq \eta_1\} \quad (3.15a)$$

$$\mathcal{D}_1(\eta_2) = \{s \mid \frac{f_1(s)}{f_0(s)} \geq \eta_2\}. \quad (3.15b)$$

Since $\eta_1 > \eta_2$, we have that $\mathcal{D}_1(\eta_1)$ is a subset $\mathcal{D}_1(\eta_2)$. Thus from Section 3.1, the “subtraction” of $(\mathcal{D}_1(\eta_1), \alpha_1, \beta_1)$ from $(\mathcal{D}_1(\eta_2), \alpha_2, \beta_2)$, i.e.,

$$(\mathcal{D}_1(\eta_2) \setminus \mathcal{D}_1(\eta_1), \alpha_2 - \alpha_1, \beta_2 - \beta_1) \quad (3.16)$$

is a valid operating point, where

$$\mathcal{D}_1(\eta_2) \setminus \mathcal{D}_1(\eta_1) = \{s \mid \eta_2 \leq \frac{f_1(s)}{f_0(s)} < \eta_1\}. \quad (3.17)$$

Next note that, since $\eta_0 > \eta_1 > \eta_2$, the region

$$\mathcal{D}_1(\eta_0) = \{s \mid \frac{f_1(s)}{f_0(s)} \geq \eta_0\}, \quad (3.18)$$

is disjoint from the region $\mathcal{D}_1(\eta_2) \setminus \mathcal{D}_1(\eta_1)$. Therefore, the “addition” of the point $(\mathcal{D}_1(\eta_0), \alpha_0, \beta_0)$ to the point in Equation 3.16,

$$(\mathcal{D}_1(\eta_0) \cup (\mathcal{D}_1(\eta_2) \setminus \mathcal{D}_1(\eta_1)), \alpha_0 + \alpha_2 - \alpha_1, \beta_0 + \beta_2 - \beta_1), \quad (3.19)$$

is a valid operating point. But from Equations 3.11 and 3.12,

$$\alpha_0 + \alpha_2 - \alpha_1 = \alpha_0 + d = \alpha^* \quad (3.20a)$$

$$\beta_0 + \beta_2 - \beta_1 = \beta_0 + d = \beta^* \quad (3.20b)$$

so Equation 3.19 simplifies to

$$(\mathcal{D}_1(\eta_0) \cup (\mathcal{D}_1(\eta_2) \setminus \mathcal{D}_1(\eta_1)), \alpha^*, \beta^*) = (\mathcal{D}_1^*, \alpha^*, \beta^*) \quad (3.21)$$
as desired. Graphically, the procedure can be thought of as a way of taking the line segment between the operating points \((D_1(\eta_1), \alpha_1, \beta_1)\) and \((D_1(\eta_2), \alpha_2, \beta_2)\) and placing its left endpoint at the point \((D_1(\eta_0), \alpha_0, \beta_0)\). Its right endpoint is then the desired operating point. From this geometric perspective, it is straightforward to see that the significance of the line segment having slope 1 is that any operating point within the upper half of the achievable region (above the line \(P_D = P_F\)) can be obtained in this way. Indeed, the procedure could be used with line segments of other slopes but with some loss of generality, i.e., some operating points in the achievable region would not be obtainable.

### 3.3 Example

In this section we return to “The Patient and The Two Doctors” scenario of Section 2.7.1 to illustrate the procedure described above. The hypothetical scenario consisted of two possible diagnostic tests for the same disease, one of which was a randomization between two LRTs on a patient’s heart rate. We will use the procedure described in Section 3.2 to design a deterministic decision rule that achieves the same operating point as the randomized rule.

The score variable, i.e., the patient’s heart rate, is denoted by \(S \sim f_0(\cdot), f_1(\cdot)\). The conditional PDFs \(f_0(\cdot)\) and \(f_1(\cdot)\) are Gaussian distributions with different means and a common variance, as shown in Figure 3-3(a),

\[
f_0(s) = \frac{1}{\sqrt{2\pi \cdot 0.5^2}} \cdot \exp \left[ \frac{-s^2}{2 \cdot 0.5^2} \right],
\]

(3.22a)

\[
f_1(s) = \frac{1}{\sqrt{2\pi \cdot 0.5^2}} \cdot \exp \left[ \frac{-(s - 1)^2}{2 \cdot 0.5^2} \right].
\]

(3.22b)
Note that the function $L(s) = f_1(s)/f_0(s)$ is invertible,

$$L(s) = \frac{f_1(s)}{f_0(s)} = \exp \left[ \frac{2s - 1}{2 \cdot 0.5^2} \right]$$

$$L^{-1}(s) = 0.5^2 \cdot \ln s + 0.5,$$

so according to the discussion in Section 2.4.2, every LRT on $S$ can be expressed as an SVT and vice versa. Thus without loss of generality, all decision rules in this section were implemented as SVTs.

The LRT ROC curve of $S$ is depicted in Figure 3-3(b). It was generated using 1000 SVTs with thresholds evenly spaced between $-2.5$ and $3.5$. Note that the resulting operating points include the ones depicted in Figure 2-6(b) in Section 2.7.1. The desired operating point is $(\alpha^*, \beta^*) = (0.273, 0.732)$ and is marked by the symbol ‘×’.

![Figure 3-3: (a) Conditional PDFs of the score variable used in this section and (b) the associated ROC curve. The ‘×’ marks the desired operating point.](image)

The first step in the procedure, shown in Figure 3-4(a), is to draw the unique line of slope 1 through the desired operating point. We found the operating point $(D_1(\eta_0), \alpha_0, \beta_0)$ and the associated value of $d$ from Equation 3.11 to be

$$(D_1(\eta_0), \alpha_0, \beta_0) = (D_1(8.183), 0.020, 0.480)$$

$$d = 0.273 - 0.020 = 0.253.$$
Figure 3-4(b) depicts the unique pair of operating points from the second step in the procedure, which satisfy $d = (\alpha_2 - \alpha_1) = (\beta_2 - \beta_1)$,

\[
(D_1(\eta_1), \alpha_1, \beta_1) = (D_1(2.844), 0.064, 0.683) \quad (3.25a)
\]

\[
(D_1(\eta_2), \alpha_2, \beta_2) = (D_1(0.352), 0.317, 0.936). \quad (3.25b)
\]

As expected, we have $\eta_0 > \eta_1 > \eta_2$.

Finally, according to Equation 3.13, the decision rule with decision region

\[
D_1^* = \{s \mid 0.352 \leq \frac{f_1(s)}{f_0(s)} < 2.844 \text{ or } 8.183 \leq \frac{f_1(s)}{f_0(s)} \}
\]

\[
= \{s \mid 0.239 \leq s < 0.761 \text{ or } 1.026 \leq s \}
\]

achieves the operating point $(\alpha^*, \beta^*)$. Note that in the second line we have used the invertibility of $L(\cdot) = f_1(\cdot)/f_0(\cdot)$ to write the region in a simpler form. $D_1^*$ is depicted in Figure 3-5(a).

![Figure 3-4](image)

Figure 3-4: (a) Unique line of slope 1 through the desired operating point, as well as (b) the unique pair of points on the curve with changes in $P_F$ and $P_D$ both equal to $d$.

To confirm that the decision rule with $D_1 = D_1^*$ achieves the desired operating point, we drew $10^6$ values of $S$ according to each of $f_0(\cdot)$ and $f_1(\cdot)$ and counted the number of times the value fell within $D_1^*$. The empirical probabilities of false alarm
and detection are listed in Figure 3-5(b). They are equal to the desired values \( \alpha^* \) and \( \beta^* \), respectively, to within a fraction of a percent. The discrepancy can be attributed to errors in our estimations of \( \eta_0, \eta_1, \) and \( \eta_2 \) and to the finite number of trials.

Alternatively, we could have achieved the desired operating point using randomization between the operating points \((D_1(\eta_0), \alpha_0, \beta_0)\) and \((D_1(\eta_3), \alpha_3, \beta_3)\), shown in Figure 3-6, where

\[
(D_1(\eta_0), \alpha_0, \beta_0) = (D_1(8.183), 0.020, 0.480) \tag{3.27a}
\]
\[
(D_1(\eta_3), \alpha_3, \beta_3) = (D_1(0.122), 0.520, 0.980). \tag{3.27b}
\]

Flipping a coin whose probability of heads is 0.505 to decide whether to use the conclusion of the rule with \( D_1 = D_1(\eta_0) \) (heads) or the one with \( D_1 = D_1(\eta_3) \) (tails) results in expected values of \( P_F \) and \( P_D \) that are consistent with the desired operating point,

\[
\mathbb{E}[P_F] = 0.495 \cdot 0.020 + 0.505 \cdot 0.52 = 0.273 \tag{3.28a}
\]
\[
\mathbb{E}[P_D] = 0.495 \cdot 0.48 + 0.505 \cdot 0.98 = 0.732. \tag{3.28b}
\]
The associated decision region $\mathcal{D}_1$ is shown in Figure 3-7(a).

We used the randomized rule to compute the empirical values of $P_F$ and $P_D$ in $10^3$ independent trials, each time with the same values of $S$ as the ones used as input to the deterministic rule. Figure 3-7(b) illustrates the distribution of the empirical value of $P_F$ over these trials. The distribution of the empirical value of $P_D$ is similar and is not shown for brevity. In both cases, the mean values are consistent with Equation 3.28.

Figure 3-6: The desired operating point could also be reached using randomization between the two deterministic operating points $(\mathcal{D}_1(\eta_0), \alpha_0, \beta_0)$ and $(\mathcal{D}_1(\eta_3), \alpha_3, \beta_3)$, as described in Section 2.7.1.
Figure 3-7: (a) Decision region $\mathcal{D}_1$ of the randomized decision rule. If the observed value of $S$ falls within the shaded region then the rule decides $\hat{H}_1$. If it falls in the cross-hatched region then the rule decides $\hat{H}_1$ with probability 0.495 and $\hat{H}_0$ with probability 0.505. (b) Empirical distribution of $P_F$ over $10^3$ trials. Both the mean value and standard deviation were rounded to three significant digits.
Chapter 4

Family of Score Variables Consistent with a Given ROC Curve

In this chapter and the next, the focus is on the “reverse engineering” process described in Section 1.1. The current objective is to characterize the information lost when condensing a score variable and its conditional PDFs into an ROC curve. We say that a score variable is consistent with a specific ROC curve, and vice versa, if the curve is the LRT ROC curve of that score variable. Obviously, the ROC curve must be LRT-consistent in order for this to be possible. Using this terminology, there are an unlimited number of score variables that are consistent with a specific LRT-consistent ROC curve. In this chapter, we show that the unifying characteristic of the score variables in this family is the conditional PDFs of the likelihood ratio random variable, i.e., the conditional distributions of $L = f_1(S)/f_0(S)$ given that $H = H_0$ or $H = H_1$ for any score variable $S \sim f_0(\cdot), f_1(\cdot)$ in the family.

We emphasize that, as mentioned in Section 2.1, two score variables are considered distinct only if their conditional PDFs are different. No assumptions are made about the units of any particular score variable. The phrase “the score variable $S \sim f_0(\cdot), f_1(\cdot)$” is thus used as a convenient shorthand for “the random variable $S$ whose conditional PDFs under $H_0$ and $H_1$ are $f_0(\cdot)$ and $f_1(\cdot)$, respectively.”

The discussion is organized as follows. In Chapter 2 we showed that an ROC curve that was constructed using LRTs on an actual score variable is necessarily
strictly concave. We extend this result in Section 4.1 by proving that strict concavity is also a sufficient condition for an ROC curve to be consistent with this method of construction. More specifically, we show that a strictly concave ROC curve described by $P_D = \psi_D(P_F)$ is consistent with the score variable $R \sim q_0(\cdot), q_1(\cdot)$, where $q_0(r) = 1$ and $q_1(r) = \psi'_D(r)$ for $0 \leq r \leq 1$.

In Section 4.2 we describe a way of generating other score variables that are consistent with the same ROC curve using invertible transformations on $R$. The requirement of monotonicity restricts the score variables that are generated to ones for which any LRT can be expressed as an equivalent SVT and vice versa, but the discussion still demonstrates the important concepts. In Section 4.2.1 we describe the special case in which the transformation produces the conditional PDFs of the likelihood ratio random variable. This leads to a discussion of potential alternatives to AUC in Section 4.3. Examples using simulated data are given in Section 4.4.

### 4.1 Sufficient Condition for LRT-Consistency

As mentioned in Section 2.4, the term LRT-consistent will be used to describe an ROC curve that was or could have been constructed using LRTs of an actual score variable. In Section 2.4.1 we showed that strict concavity is a necessary condition for LRT-consistency. The following statement asserts that it is also a sufficient condition.\(^1\)

*Statement 2.* An ROC curve is LRT-consistent if it has endpoints at $(0,0)$ and $(1,1)$ and is strictly concave.

*Proof.* Assume that the functional form of an arbitrary ROC curve, i.e., one that is not assumed to have been constructed using LRTs on a specific score variable, is $P_D = \psi_D(P_F)$, where $\psi_D(\cdot)$ satisfies the endpoint conditions and is strictly concave. Then it is possible to construct a score variable $R \sim q_0(\cdot), q_1(\cdot)$ such that the ROC curve of $q_0(\cdot)$ and $q_1(\cdot)$ has the same functional form. The first step is to write any

\(^1\)We first became aware of this result through an elegant exam problem written by Professor G. Wornell for the MIT course *Inference and Information.*
point \((\alpha, \psi_D(\alpha))\) on the curve as

\[
\alpha = \int_0^\alpha dr \cdot 1 = \int_0^\alpha dr \cdot q_0(r) \tag{4.1a}
\]

\[
\psi_D(\alpha) = \psi_D(0) + \int_0^\alpha dr \cdot \psi'_D(r) = \int_0^\alpha dr \cdot \psi'_D(r) = \int_0^\alpha dr \cdot q_1(r) \tag{4.1b}
\]

where

\[
q_0(r) = 1, \quad 0 \leq r \leq 1 \tag{4.2a}
\]

\[
q_1(r) = \psi'_D(r), \quad 0 \leq r \leq 1. \tag{4.2b}
\]

To verify that \(q_0(\cdot)\) and \(q_1(\cdot)\) are valid PDFs, note that \(q_0(\cdot)\) is simply the uniform PDF on the interval \([0, 1]\). \(q_1(\cdot)\) is non-negative by assumption, since a curve that extends from the origin to \((1, 1)\), never exceeds 1, and is strictly concave cannot have negative slope anywhere. It integrates to 1 because it is the derivative of a curve that extends from the origin to \((1, 1)\).

Equation 4.1 can now be interpreted as saying that any point \((\alpha, \psi_D(\alpha))\) on the given ROC curve is achievable by a threshold test on \(R\), or an SVT, with threshold \(\alpha\). Thus, the curve is the SVT ROC curve of \(R\). Furthermore, since the function \(q_1(r)/q_0(r) = \psi'_D(r)\) is strictly decreasing according to the concavity assumption, it follows from Section 2.4 that any SVT on \(R\) can be expressed as an equivalent LRT, and vice versa. Specifically, an SVT on \(R\) with threshold \(r^*\) is equivalent to an LRT on \(R\) with threshold \(\psi'_D(r^*)\). This implies that the LRT ROC curve of \(R\) is identical to the SVT ROC curve, and is therefore identical to the given LRT-consistent ROC curve.
4.2 Construction of Score Variables Consistent with a Given ROC Curve

The score variable \( R \sim q_0(\cdot), q_1(\cdot) \) defined in Section 4.1 is not the only one that is consistent with the ROC curve described by \( P_D = \psi_D(P_F) \). We now show that applying an invertible transformation to \( R \) generates another score variable consistent with the same ROC curve. It is important to note that this method can only generate score variables for which any LRT can be expressed as an equivalent SVT, and vice versa. There are other score variables consistent with the same curve that do not satisfy this property. However, the basic rationale still leads to interesting insights.

**Statement 3.** Consider the score variable \( R \sim q_0(\cdot), q_1(\cdot) \) described in Section 4.1, whose LRT ROC curve is described by \( P_D = \psi_D(P_F) \). The score variable \( U = M(R) \), where \( M(\cdot) \) is an invertible function on the interval \([0, 1]\), has the same LRT ROC curve.

**Proof.** Assume that \( M(\cdot) \) is a strictly increasing function on the interval \([0, 1]\). The proof for a strictly decreasing function is essentially identical. The objective is to show that any operating point on the LRT ROC curve of \( R \) can be achieved by an LRT on \( U = M(R) \), and vice versa. The conditional PDFs of \( U \) are

\[
\begin{align*}
    w_0(u) &= \frac{q_0(M^{-1}(u))}{M'(M^{-1}(u))}, \quad M(0) \leq u < M(1) \quad (4.3a) \\
    w_1(u) &= \frac{q_1(M^{-1}(u))}{M'(M^{-1}(u))}, \quad M(0) \leq u < M(1). \quad (4.3b)
\end{align*}
\]

Therefore, an SVT on \( U \) with threshold \( u^* \) achieves the operating point

\[
\begin{align*}
    P_F &= \int_{M(0)}^{u^*} du \cdot \frac{q_0(M^{-1}(u))}{M'(M^{-1}(u))} = \int_0^{M^{-1}(u^*)} dv \cdot M'(v) \cdot \frac{q_0(v)}{M'(v)} \quad (4.4a) \\
    P_D &= \int_{M(0)}^{u^*} du \cdot \frac{q_1(M^{-1}(u))}{M'(M^{-1}(u))} = \int_0^{M^{-1}(u^*)} dv \cdot M'(v) \cdot \frac{q_1(v)}{M'(v)} \quad (4.4b)
\end{align*}
\]

where we have used the substitution \( v = M^{-1}(u) \), so that \( dv/du = (M^{-1})'(u) = 1/M'(v) \).
Cancelling the factors of $M'(v)$ yields

\[
P_F = \int_0^{M^{-1}(u^*)} dv \cdot q_0(v) \tag{4.5a}
\]

\[
P_D = \int_0^{M^{-1}(u^*)} dv \cdot q_1(v) = \psi_D(P_F). \tag{4.5b}
\]

To see why the second equality in Equation 4.5b is true, note that the function $\psi_D(\cdot)$ can be viewed as the function whose input is an integral (starting from 0) over $q_0(\cdot)$ and whose output is the same integral over $q_1(\cdot)$. Equation 4.5 therefore says that an SVT on $U$ with threshold $u^*$ achieves the same operating point as an SVT on $R$ with threshold $M^{-1}(u^*)$. From Section 4.1, an SVT on $R$ with threshold $M^{-1}(u^*)$ is equivalent to an LRT on $R$ with threshold $\psi_D'(M^{-1}(u^*))$.

Finally, note that since $M^{-1}(\cdot)$ and $\psi_D'(\cdot)$ are both strictly decreasing by assumption, the function

\[
\frac{w_1(u)}{w_0(u)} = \frac{q_1(M^{-1}(u))}{q_0(M^{-1}(u))} = \psi_D'(M^{-1}(u)) \tag{4.6}
\]

is strictly increasing. Therefore, any SVT on $U$ can be reduced to an LRT, and vice versa. In particular, an SVT on $U$ with threshold $u^*$ is equivalent to an LRT on $U$ with threshold $\psi_D'(M^{-1}(u^*))$. In summary, the following four tests achieve the same operating point:

1. an SVT on $R$ with threshold $r^*$, $0 \leq r^* \leq 1$

2. an LRT on $R$ with threshold $\psi_D'(r^*)$, $\psi_D'(1) \leq \psi_D'(r^*) < \psi_D'(0)$

3. an SVT on $U$ with threshold $M(r^*)$, $M(0) < M(r^*) < M(1)$

4. an LRT on $U$ with threshold $\psi_D'(r^*)$, $\psi_D'(1) \leq \psi_D'(r^*) < \psi_D'(0)$

Varying the above thresholds over their entire ranges will yield four identical, LRT-consistent ROC curves.
4.2.1 Special Case: The Likelihood Ratio as a Score Variable

An important special case of the above discussion is

$$U = L = \phi_F^{-1}(R).$$  

(4.7)

The random variable $L$ can be interpreted as the likelihood ratio random variable, i.e., its conditional PDFs are identical to the PDFs of the random variable $Z = f_1(S)/f_0(S)$, for any $S \sim f_0(\cdot), f_1(\cdot)$ consistent with the ROC curve described by $P_D = \psi_D(P_F)$.

The PDFs are

$$w_0(\ell) = \frac{q_0(M^{-1}(\ell))}{-M'(M^{-1}(\ell))} = \frac{1}{-\phi_F^{-1}(\phi_F(\ell))}$$  

(4.8a)

$$= -\phi_F'(\ell), \quad \phi_F^{-1}(1) \leq \ell < \phi_F^{-1}(0)$$  

(4.8b)

$$w_1(\ell) = \frac{q_1(M^{-1}(\ell))}{-M'(M^{-1}(\ell))} = \frac{\psi_F'(\phi_F(\ell))}{-\phi_F^{-1}(\phi_F(\ell))} = \frac{\phi_D'(\ell)}{\phi_F'(\ell)} \cdot -\phi_F'(\ell)$$  

(4.8c)

$$= -\phi_D'(\ell), \quad \phi_F^{-1}(1) \leq \ell < \phi_F^{-1}(0).$$  

(4.8d)

Note that the negative signs in the denominators of Equations 4.8a and 4.8c came from the fact that $M(\cdot)$ is a decreasing function, so its derivative is negative. We also used the fact that for an invertible function $f(\cdot)$ such that $y = f(x), (f^{-1})'(y) = 1/f'(f^{-1}(y))$.

Next note that, by definition,

$$1 - \phi_F(\eta) = \mathbb{P}(L \leq \eta \mid H = H_0)$$

$$1 - \phi_D(\eta) = \mathbb{P}(L \leq \eta \mid H = H_1).$$  

(4.9)

where $L$ is the likelihood ratio random variable. Equation 4.9 says that $(1 - \phi_F(\eta))$ and $(1 - \phi_D(\eta))$ are the CDFs of $L$ under $H_0$ and $H_1$, respectively. Their derivatives
are therefore its conditional PDFs,

\[ f_{L \mid H}(\eta \mid H_0) = -\phi_F'(\eta) = w_0(\eta) \]  
\[ f_{L \mid H}(\eta \mid H_1) = -\phi_D'(\eta) = w_1(\eta). \]  

(4.10a)  

(4.10b)

The fact that this conclusion depends only on the ROC curve itself, not on a particular score variable consistent with the curve, implies that it holds for all such score variables.

As an aside, note that

\[ \frac{w_1(\ell)}{w_0(\ell)} = \frac{\psi_D'(\ell)}{\psi_F'(\ell)} = \psi_D(\phi_F(\ell)) = \ell, \]  

(4.11)

so an LRT on \( L \) with threshold \( \ell^* \) is equivalent to an SVT on \( L \) with threshold \( \ell^* \). In this sense, Equation 4.11 can be interpreted as a very literal way of saying that an LRT is, in fact, a threshold test on the likelihood ratio.

### 4.3 Alternatives to the AUC

In Section 4.2, we introduced a way of characterizing the family of score variables consistent with a given ROC curve: Given any score variable \( S \sim f_0(\cdot), f_1(\cdot) \) in the family, the likelihood ratio random variable \( L = f_1(S)/f_0(S) \) always has the same conditional PDFs. A natural question is whether or not this unifying characteristic can be used to create new metrics with which to compare ROC curves. In Section 4.3.1, we suggest viewing ROC curves as trajectories in \( P_F-P_D-\eta \) space and show that the conditional expectations of \( L \) under \( H_0 \) and \( H_1 \) are directly related to the integrals of the projections of the trajectory onto the \( P_F-\eta \) and \( P_D-\eta \) planes. This conclusion is reminiscent of the fact that the AUC is the integral of the projection of the trajectory onto the \( P_F-P_D \) plane. We propose another metric, the area under an ROC curve that has been weighted by a factor of the form \( m \cdot P_F^{m-1} \) for integer \( m \), in Section 4.3.2.
4.3.1 Conditional Expectations of the Likelihood Ratio

It is intuitively desirable for the likelihood ratio to be either very small or very large for each value of $S$. A very small value means that $f_0(s) \gg f_1(s)$, so we can infer $\hat{H}_0$ with high confidence, and vice versa for very large values. Note that with this line of reasoning, any value of the ratio that is less than 1 is just as desirable as its reciprocal, since both correspond to the same level of confidence in the decision (but different inferred values of $H$). Thus, a reasonable metric to consider is the conditional expectation of $L$ given that $L \leq 1$ and $H = H_0$. “Better” ROC curves presumably have lower conditional expectations and, therefore, higher expected levels of confidence in the decision $\hat{H}_0$.

Direct calculation using the results of Section 4.2.1 yields

$$E[L \mid (L \leq 1, H = H_0)] = \int_0^1 d\ell \cdot \ell \cdot \frac{w_0(\ell)}{P(L \leq 1 \mid H_0)} \quad (4.12a)$$

$$= \int_0^1 d\ell \cdot \ell \cdot \frac{-\phi_F'(\ell)}{1 - \phi_F(\ell)} \quad (4.12b)$$

$$= \frac{1}{1 - \phi_F(\ell)} \left[ -\ell \cdot \phi_F(\ell) \right]_0^1 + \frac{1}{1 - \phi_F(\ell)} \int_0^1 d\ell \cdot \phi_F(\ell) \quad (4.12c)$$

$$= \frac{\phi_F(1) - \int_0^1 d\ell \cdot \phi_F(\ell)}{\phi_F(1) - 1} \quad (4.12d)$$

where Equation 4.12c follows from integration by parts. For an ideal pair of conditional PDFs such as those in Figure 2-4 (reproduced in Figure 4-1 for convenience), $E[L \mid (L \leq 1, H = H_0)]$ takes on its lowest possible value. Indeed, it can be seen by inspection of Figure 4-1 that

$$E[L \mid (L \leq 1, H = H_0)] = 0 \quad \text{(ideal).} \quad (4.13)$$
Figure 4-1: (a) Example of the ideal scenario in which the supports of the conditional PDFs of the score variable are non-overlapping. (b) All points on the ideal, step function curve can be achieved by including appropriate portions of the supports of $f_0(\cdot)$ and $f_1(\cdot)$ in the decision region $D_1$.

Graphically, Equation 4.12 can be interpreted as follows. Each point on an ROC curve is defined by the values $(\eta, P_F, P_D) = (\eta, \phi_F(\eta), \phi_D(\eta))$. An ROC curve can therefore equivalently be thought of as a trajectory in three-dimensional space, projected onto the $P_F-P_D$ plane. This is depicted in Figure 4-2 for two Gaussian PDFs with different means and a common variance. The trajectory was computed using SVTs instead of LRTs, so the third axis represents a threshold on the score variable as opposed to a threshold on the likelihood ratio. In the same way that the AUC is the integral of the projection onto the $P_F-P_D$ plane, the integral in Equations 4.12 is an integral over the projection onto the $P_F-\eta$ plane.

As a final remark, note that we chose to condition on the value $H = H_0$ arbitrarily. The choice $H = H_1$ would have led to the very similar metric $\mathbb{E}[L \mid (L \geq 1, H = H_1)]$, with

$$\mathbb{E}[L \mid (L \geq 1, H = H_1)] = +\infty \quad \text{(ideal).} \quad (4.14)$$

Unfortunately, neither metric possesses more practical interpretation, i.e., the two-alternative forced choice experiment, of the AUC.
4.3.2 Area Under a Geometrically Weighted ROC Curve

In this section, we return to the random variable $R$ of Section 4.1 and use it as motivation for another possible ROC curve metric: the area under the function

$$
\Lambda_m(P_F) = m \cdot P_F^{m-1} \cdot \psi_D(P_F), \quad 0 \leq P_F \leq 1,
$$

for an LRT-consistent ROC curve described by $P_D = \psi_D(P_F)$. As a reminder, the conditional PDFs of $R$ are

$$
q_0(r) = 1, \quad 0 \leq r \leq 1 \quad (4.16a)
$$

$$
q_1(r) = \psi'_D(r), \quad 0 \leq r \leq 1 \quad (4.16b)
$$

and we can regenerate the original curve by performing LRTs on $R$. Equivalently, since the function $q_1(\cdot)/q_0(\cdot)$ is strictly decreasing by assumption, we can regenerate the original curve by performing SVTs on $R$.

Now consider a score variable $V \sim y_0(\cdot), y_1(\cdot)$ where $y_0(\cdot)$ is required to be the uniform PDF over the interval $[0, 1]$, and consider the problem of finding the “ideal” choice for $y_1(\cdot)$. The ideal choice should make perfect detection on $V$, i.e., $(P_F, P_D) = (0, 1)$,
possible. One option that fits this criterion is

\[ y_0(v) = 1, \quad 0 \leq v \leq 1 \quad (4.17a) \]
\[ y_1(v) = \delta(v), \quad 0 \leq v \leq 1. \quad (4.17b) \]

For the PDFs in Equation 4.17, the deterministic decision rule with \( D_1 = \{0\} \) achieves the operating point \((P_F, P_D) = (0, 1)\).\(^2\) It is also straightforward to see that performing SVTs on \( V \) generates the ideal ROC curve described in Section 2.5.

Equation 4.17 says that the ideal choice for \( y_1(\cdot) \), or equivalently the ideal choice for the derivative of an ROC curve, is an impulse centered at the origin. Motivated by this fact, a possible set of metrics are the moments of the derivative of an ROC curve. The smaller the moments, the more concentrated the derivative is to the origin and the closer it is to an impulse. For an ROC curve described by \( P_D = \psi_D(P_F) \), we denote the \( m \)th moment of the derivative of an ROC by \( \mu_m \) and find that

\[ \mu_m = \int_0^1 dP_F \cdot P_F^m \cdot \psi_D'(P_F) \quad (4.18a) \]
\[ = [P_F^m \cdot \psi_D(P_F)]_0^1 - \int_0^1 dP_F \cdot m \cdot P_F^m \cdot \psi_D(P_F) \quad (4.18b) \]
\[ = 1 - \int_0^1 dP_F \cdot m \cdot P_F^m \cdot \psi_D(P_F). \quad (4.18c) \]

The integral in Equation 4.18c is the area under the original curve weighted by \( m \cdot P_F^{m-1} \). Note that the AUC is directly related to the first moment.

\[ \mu_1 = 1 - \int_0^1 dP_F \cdot \psi_D(P_F) = 1 - \text{AUC}. \quad (4.19) \]

\(^2\)This is not a mathematically rigorous statement, but our objective is simply to provide an intuitive argument for the proposed metric.
4.4 Examples

We illustrate the results of this chapter with several examples. We start by constructing an LRT-consistent ROC curve using a score variable $S \sim f_0(\cdot), f_1(\cdot)$. The PDFs $f_0(\cdot)$ and $f_1(\cdot)$ are chosen to ensure that every SVT on $S$ can be expressed as an equivalent LRT, and vice versa. We then derive the conditional PDFs of the random variables $R$ and $L$ from Sections 4.1 and Section 4.2.1, respectively. We confirm that the corresponding LRT ROC curves are identical to the LRT ROC curve of $S$. Since it is true for $R$, $L$, and $S$ that every LRT can be expressed as an equivalent SRT, and vice versa, all decision rules in this section were implemented as SVTs.

Figures 4-3(a) and (b) depict the PDFs $f_0(\cdot)$ and $f_1(\cdot)$ and the LRT ROC curve of $S$, respectively. The curve was generated using 1000 SVTs with thresholds evenly spaced between $-2.5$ and $3.5$. Thus, each point in Figure 4-3(b) has three associated parameters – the SVT threshold (easily mapped to an equivalent LRT threshold), the corresponding probability of false alarm, and the corresponding probability of detection. The functional form of the curve is denoted by $P_D = \psi_D(P_F)$. 

Figure 4-3: (a) Conditional PDFs of $S$ along with (b) the corresponding LRT-ROC curve.
Figure 4-4: (a) Conditional PDFs of the score variable $R$ as described in Section 4.1, along with (b) its associated LRT-consistent ROC curve. As expected, $R$’s ROC curve is identical to the original ROC curve in Figure 4-3.

The conditional PDFs of the random variable $R$ are shown in Figure 4-4(a),

$$q_0(r) = 1, \quad 0 \leq r \leq 1 \quad (4.20a)$$
$$q_1(r) = \psi_D'(r), \quad 0 \leq r \leq 1. \quad (4.20b)$$

$q_1(\cdot)$ was estimated using a simple forward difference method on the 1000 $(P_F, P_D)$ pairs of the original ROC curve in Figure 4-3(b). The LRT ROC curve of $R$, shown in Figure 4-4(b), was generated by computing the cumulative distribution functions (CDFs) of $q_0(\cdot)$ and $q_1(\cdot)^3$, or equivalently by performing SVTs on $R$. As expected, it is identical to the LRT ROC curve of $S$.

The conditional PDFs of the likelihood ratio random variable, $L$, are shown in Figure 4-5(a),

$$w_0(\ell) = -\phi_F'(\ell) \quad (4.21a)$$
$$w_1(\ell) = -\phi_D'(\ell). \quad (4.21b)$$

Again, the derivatives were estimated using a simple forward difference method on

\[\text{Technically these were cumulative mass functions since the approximations of } q_0(\cdot) \text{ and } q_1(\cdot) \text{ were discrete.}\]
Figure 4-5: (a) Conditional PDFs of the score variable $L$ as described in Section 4.2.1, along with (b) its associated LRT-consistent ROC curve. As expected, $L$’s ROC curve is identical to the original ROC curve in Figure 4-3.

The 1000 ($\ell, P_F$) or ($\ell, P_D$) pairs from the original ROC curve in Figure 4-3(b), where $\ell$ is the LRT threshold associated with a particular point. The corresponding LRT ROC curve was generated by performing SVTs on $L$ and is shown in Figure 4-5(b). Again, the result confirms the theory presented in this chapter since the ROC curve is the same as the LRT ROC curve of $S$.

We also confirmed that the conditional PDFs of $L$ are the distributions of the likelihood ratio viewed as the random variable $L = f_1(S)/f_0(S)$, by drawing $10^7$ values of $S$ from each of $f_0(\cdot)$ and $f_1(\cdot)$ and computing the ratio $f_1(s)/f_0(s)$ for each one. The resulting distributions are shown in Figure 4-6. They are consistent with the PDFs in Figure 4-5.

We confirmed that the proposed metrics of Sections 4.3 are consistent with the AUC in the sense that if one ROC curve always lies above another, then it has a higher AUC and is also superior with respect to our proposed metrics. This is illustrated in Figure 4-7(b) and (c) for several different ROC curves. Each one is the LRT ROC curve of a score variable whose conditional PDFs are Gaussian distributions with means 0 and 1 and equal variances. Clearly, the curves with higher AUCs have lower values of $E[L \mid (L \leq 1, H = H_0)]$, $\mu_2$, and $\mu_3$. The results were similar for $\mu_m$, $m > 3$. 

70
Figure 4-6: Empirical validation that the conditional PDFs of the likelihood ratio are as shown in Figure 4-5.

Figure 4-7: (a) Five ROC curves, each constructed using LRTs on a score variable whose conditional PDFs are Gaussian distributions with means 0 and 1 and variance $\sigma^2$. (b) Conditional expectation of the likelihood ratio vs. AUC. The monotonic relationship between the two metrics shows consistency between them in the case of ROC curves that do not intersect. (c) Second ($\mu_2$) and third ($\mu_3$) moments of the derivative vs. AUC.
Chapter 5

Family of Score Variables Consistent with a Given Set of LRT Decision Regions

In this chapter we continue to pursue the “reverse engineering” objective of Chapter 4, but with a different focus. A goal in Chapter 4 was to characterize the family of score variables that lead to the same LRT ROC curve. We now consider the family of score variables that lead to the same set of LRT decision regions. These are different questions in that two score variables could lead to the same LRT ROC curve even if the decision regions associated with a particular operating point on the curve (i.e., the decision regions of the LRT that achieves that operating point) are different. On the other hand, two score variables could lead to the same decision regions for an LRT with a specific threshold even if the operating point obtained by that LRT, and consequently the LRT ROC curves of the two score variables, are different.

A set of LRT decision regions is characterized by a pair of PDFs \( f_0(\cdot) \) and \( f_1(\cdot) \). Each decision region in the set corresponds to a specific value \( \eta \geq 0 \) and is defined as

\[
\mathcal{D}_1(\eta) = \{ s \mid \frac{f_1(s)}{f_0(s)} \geq \eta \}. \quad (5.1)
\]
The associated decision rule is

\[
\frac{f_1(s)}{f_0(s)} \stackrel{H_1}{\gtrless} \frac{\bar{H}_1}{\bar{H}_0} \eta. \tag{5.2}
\]

Figure 5-1: Abstraction of the implementation of the set of LRT decision rules for PDFs \(f_0(\cdot)\) and \(f_1(\cdot)\).

We emphasize that any two PDFs can be used to define a set of LRT decision regions. Although it may be useful to interpret them as the conditional PDFs of a hypothetical score variable, they are not required to correspond to the distributions of any measurable quantity. With this in mind, note that a set of LRT decision rules can be viewed as implementable using a system with two inputs \((s, \eta)\) and one output \((h(s, \eta))\) as shown in Figure 5-1. The system operates by computing \(f_1(s)/f_0(s)\) for two functions \(f_0(\cdot)\) and \(f_1(\cdot)\) that satisfy the requirements for valid PDFs, i.e., they are each non-negative and integrate to 1. The result is compared to the value of \(\eta\). If \(f_1(s)/f_0(s)\) exceeds \(\eta\), then the output is a 1, otherwise it is a 0,

\[
h(s, \eta) = \begin{cases} 
0 & \text{if } \frac{f_1(s)}{f_0(s)} < \eta \\
1 & \text{if } \frac{f_1(s)}{f_0(s)} \geq \eta.
\end{cases} \tag{5.3}
\]

If \(f_0(\cdot)\) and \(f_1(\cdot)\) are the conditional PDFs of an actual score variable \(S \sim f_0(\cdot), f_1(\cdot)\) in a binary hypothesis testing problem, then the system can be used to make optimal Neyman-Pearson decisions for observations of that score variable.

It is possible for multiple systems of this form to each be optimal according to the Neyman-Pearson criterion for a different (possibly hypothetical) score variable, but that all have the same input-output relationship. Therefore, all of these score variables effectively lead to the same set of LRT decision regions. From Equation 5.1,
the set of LRT decision regions of two arbitrary PDF pairs \( f_0(\cdot), f_1(\cdot) \) and \( g_0(\cdot), g_1(\cdot) \) are equivalent if and only if

\[
\{ s \mid \frac{f_1(s)}{f_0(s)} \geq \eta \} = \{ s \mid \frac{g_1(s)}{g_0(s)} \geq \eta \}, \quad 0 \leq \eta < \infty
\]  

or equivalently

\[
\frac{f_1(s)}{f_0(s)} = \frac{g_1(s)}{g_0(s)}, \quad -\infty < s < \infty.
\]  

In the remainder of the chapter, we consider the problem in which the input-output relationship of a system such as that in Figure 5-1 is specified, but the PDFs \( f_0(\cdot) \) and \( f_1(\cdot) \) are not. The objective is to generate other PDF pairs that are consistent with the same input-output relationship, i.e., with the same set of LRT decision regions. Since each of these PDF pairs can be viewed as the conditional distributions of a hypothetical score variable, the collection of PDF pairs consistent with a given set of LRT decision regions can equivalently be viewed as the collection of score variables consistent with a given set of LRT decision regions. We show in Section 5.1 that the problem can be formulated as a linear program, and demonstrate the technique in Section 5.2 using simulated data.

### 5.1 Linear Programming Formulation

We assume that we have a system of the type in Figure 5-1 that was designed using a specific but unknown PDF pair \( f_0(\cdot), f_1(\cdot) \). \( f_0(s) \) is assumed to be non-zero for all values of \( s \) so that the ratio \( f_1(s)/f_0(s) \) is always defined. We have control over the inputs \( s \) and \( \eta \) and access to its binary output \( h \). Our objective is to probe the system in order to determine \( f_1(s)/f_0(s) \) for arbitrary \( s \). The proposed procedure is:

1. Set the input \( s \) to a specific value, \( s = s_0 \).

2. Set the input \( \eta \) to 0 so that the output \( h(s_0, \eta) = h(s_0, 0) \) is guaranteed to be 1.

This follows from the fact that the ratio of two non-negative functions is always non-negative.
3. Continuously increase $\eta$. The smallest value of $\eta$ for which $h(s_0, \eta) = 0$ is equal to $f_1(s_0)/f_0(s_0)$,

\[ L(s_0) = \frac{f_1(s_0)}{f_0(s_0)} = \arg\min_{\eta} h(s_0, \eta) \text{ s.t. } h(s_0, \eta) = 0 \quad (5.6) \]

where $L(\cdot)$ is the ratio $f_1(\cdot)/f_0(\cdot)$.

Theoretically, the procedure above could be applied for a continuous range of values of $s$. However, we assume that it is only performed for a sufficiently dense set of $N$ values, denoted by $s_i$, $0 \leq i \leq (N - 1)$, in order to infer the behavior for a continuous range. The linear program to be proposed will therefore find two discrete probability mass functions (PMFs). These PMFs are denoted by $\hat{f}_0(\cdot)$ and $\hat{f}_1(\cdot)$.

In summary, the PMFs $\hat{f}_0(\cdot)$ and $\hat{f}_1(\cdot)$ must satisfy the following constraints.

\[ f_1(s_i) - f_0(s_i) \cdot L(s_i) = 0, \quad 0 \leq i \leq N - 1 \quad (5.7a) \]
\[ f_0(s_i) \geq 0, \quad 0 \leq i \leq N - 1 \quad (5.7b) \]
\[ f_1(s_i) \geq 0, \quad 0 \leq i \leq N - 1 \quad (5.7c) \]
\[ f_0(s_0) + \cdots + f_0(s_{N-1}) = 1 \quad (5.7d) \]
\[ f_1(s_0) + \cdots + f_1(s_{N-1}) = 1. \quad (5.7e) \]

The first line in Equation 5.7 comes from probing the system to find $L(s_i)$ for $0 \leq i \leq (N - 1)$. The remaining lines express the normalization and non-negativity constraints that must be satisfied in order for $\hat{f}_0(\cdot)$ and $\hat{f}_1(\cdot)$ to be valid PMFs.

In matrix form for $N = 3$, for example, the likelihood ratio and normalization
Equation 5.8 generalizes to larger values of \( N \) in a straightforward manner. Note that the first \( N \) rows of the matrix \( A \) correspond to the likelihood ratio constraints, while the last two rows correspond to the normalization constraints.

Any vector in the nullspace of \( A \) that also satisfies the non-negativity constraints represents a valid set of values for \( \hat{f}_0(s_i) \) and \( \hat{f}_1(s_i) \), \( 0 \leq i \leq (N - 1) \). A convex optimization package such as CVX offers a convenient way of finding such a vector.

[3] Formally, we can use the following linear program

\[
\min_{\hat{f}} \mathbf{c}^T \hat{f} \; \text{s.t.} \; A \hat{f} = 0 \text{ and } \hat{f} \geq 0 \tag{5.9}
\]

for some real vector \( \mathbf{c} \), where the notation \( \mathbf{c}^T \) denotes the transpose of \( \mathbf{c} \). In the next section, we present simulations of this technique for various choices of \( f_0(\cdot) \), \( f_1(\cdot) \), and \( N \).

It is worth pointing out that although the solution to Equation 5.9 will correspond to valid PMFs \( \hat{f}_0(\cdot) \) and \( \hat{f}_1(\cdot) \), it is not immediately clear whether one choice of the vector \( \mathbf{c} \) is “better” than any other in an appropriate sense. In fact, it is not even clear whether linear cost is an appropriate choice. Other possible options are minimization
Figure 5-2: (a) Original conditional PDFs ($f_0(\cdot)$ in blue dots, $f_1(\cdot)$ in orange dots) used to establish the likelihood ratio constraint. (b) Solution of the linear program in Equation 5.9 ($\hat{f}_0(\cdot)$ in blue dots, $\hat{f}_1(\cdot)$ in orange dots). Each component of the vector $c$ was drawn from a uniform distribution between 0 and 1. The resulting vector was normalized to have unit length.

of the Euclidean norm of the vector $\hat{f}$ (quadratic cost), or maximization of the entropy of either $\hat{f}_0(\cdot)$ or $\hat{f}_1(\cdot)$. These topics require further exploration.

5.2 Examples

We give several examples to demonstrate the results of this chapter. In each one, we chose two conditional PDFs $f_0(\cdot)$ and $f_1(\cdot)$ and computed their ratio on a set of 150 values of the corresponding score variable, denoted by $S \sim f_0(\cdot), f_1(\cdot)$. The ratio of $f_1(\cdot)$ to $f_0(\cdot)$ over this set was used to construct the matrix $A$, which was then used as input to the linear program in Equation 5.9. We experimented with linear, piecewise linear, and Gaussian PDF pairs. The results are shown in Figures 5-2 to 5-4. In general, the smoothness of the solution was highly sensitive to changes in the vector $c$, and that scenarios in which ratio of the PDFs took on very low or very high values led to numerical issues and discontinuous solutions.
Figure 5-3: (a) Original conditional PDFs \( f_0(\cdot) \) in blue dots, \( f_1(\cdot) \) in orange dots) used to establish the likelihood ratio constraint. (b) Solution of the linear program in Equation 5.9 \( \hat{f}_0(\cdot) \) in blue dots, \( \hat{f}_1(\cdot) \) in orange dots). Each component of the vector \( c \) was drawn from a uniform distribution between 0 and 1. The resulting vector was normalized to have unit length.

Figure 5-4: (a) Original conditional PDFs \( f_0(\cdot) \) in blue dots, \( f_1(\cdot) \) in blue dots) used to establish the likelihood ratio constraint. (b) Solution of the linear program in Equation 5.9 \( \hat{f}_0(\cdot) \) in black dots, \( \hat{f}_1(\cdot) \) in blue dots). Each component of the vector \( c \) was drawn from a uniform distribution between 0 and 1. The resulting vector was normalized to have unit length. The PDFs have been truncated to a range in which their ratio lies in a small range around 1, since we observed that very small or very large ratios led to discontinuous solutions owing to numerical accuracy issues.
Chapter 6

Summary and Future Work

In this thesis we have studied ROC curves in a broad context, with a focus on the information they contain about the binary hypothesis testing problems for which they were constructed. The forward progression in a binary hypothesis testing system from an observation of a score variable to a set of optimal decision rules to an ROC curve, summarized in Figure 6-1, is widely known and understood. The backward progression from a given set of decision rules or a given ROC curve to a score variable with a specific pair of conditional PDFs, on the other hand, has received less attention. To this end, one of our key objectives was to find ways of “reverse engineering” the components in Figure 6-1 in order to better understand the information lost when moving from one to another.

We started in Chapter 3 by providing an extension to a standard result about the achievable region of the $P_F$-$P_D$ plane for a given ROC curve. Specifically, the region of achievable operating points for an ROC curve described by $P_D = \psi_D(P_F)$ is the region between the curve itself and its complement below the line $P_D = P_F$, $P_D = 1 - \psi_D(1 - P_F)$. A well-known fact is that randomization between deterministic decision rules can be used to obtain any operating point in this region. We showed that for a strictly concave ROC curve, all points in that region are also achievable using a single deterministic rule. A procedure for designing such a rule for a given operating point was outlined and demonstrated through an example. Such a procedure might be valuable in scenarios, possibly in a clinical setting, in which the reproducibility of
Figure 6-1: Components of a binary hypothesis testing system.

A decision is crucial. We note, however, that the proposed procedure is not unique. In the future, it would be interesting to consider other methods of designing suboptimal (in the Neyman-Pearson sense) deterministic decision rules and the advantages they might offer.

The main result of Chapter 4 was a method of “reverse engineering” a given ROC curve, i.e., a method of constructing one or more score variables, each with a distinct set of conditional PDFs, that could have been used to generate it. Specifically, we showed how to construct a score variable, denoted by $R \sim q_0(\cdot), q_1(\cdot)$, from any strictly concave ROC curve such that the ROC curve of $R$ was identical to the original. Monotonic transformations of $R$ led to a family of score variables all having the same ROC curve. We pointed out that the unifying characteristic of the members of this family could be viewed in terms of the conditional PDFs of the likelihood ratio random variable, $L = f_1(S)/f_0(S)$ for any score variable $S \sim f_0(\cdot), f_1(\cdot)$ in the family. Finally, we proposed two sets of metrics that could be used to compare different ROC curves. The first was the conditional expectation of $L$, given that it was less than or equal to 1 and that $H = H_0$ or $H = H_1$. We related this metric to the interpretation of an ROC curve as a trajectory in $P_F-P_D-\eta$ space. Other three-dimensional perspectives exist in the literature [15], so it might be interesting to pursue this avenue further in search of useful metrics. The second set of metrics was the set of moments of the
derivative of an ROC curve, which should intuitively be smaller for curves closer to the ideal ROC curve. Other parameterizations of the derivative could be explored in future work. For both sets of metrics, we demonstrated that they correctly ranked a series of ROC curves that did not overlap (and thus had a clear order in terms of quality).

In Chapter 5 we focused on “reverse engineering” a set of LRT decision rules and showed that a linear program could be used to find at least one score variable that was consistent with that set. We successfully demonstrated this result by way of several examples, but pointed out that a linear cost function was a somewhat arbitrary choice. Extending this method to include quadratic or other types of cost functions may lead to valuable insights not explored in this thesis. A completely different approach would be to use an iterative algorithm instead of an optimization.
Bibliography


