4 Performance Limits of Hypothesis Testing

In this section we examine the fundamental performance limits of binary hypothesis testing. We begin by recalling that, thus far, we have (somewhat implicitly) restricted our attention to *deterministic* decision rules. However, as we suggested, there are situations in which what we will refer to as a *randomized* decision rule can perform better. To develop some preliminary insights, we examine such a scenario, in which the data are discrete-valued. We first make some preliminary observations about hypothesis tests with discrete data.

4.1 Tests with Discrete-Valued Observations

We separately consider the Bayesian and Neyman-Pearson formulations, beginning with the former.

4.1.1 Bayesian Tests

A couple of special issues arise in the case of a likelihood ratio test

$$L(\mathbf{y}) = \frac{p_{\mathsf{y}|H}(\mathbf{y}|H_1)}{p_{\mathsf{y}|H}(\mathbf{y}|H_0)} \underset{\hat{H}(\mathbf{y})=H_0}{\overset{\hat{H}(\mathbf{y})=H_1}{\gtrless}} \frac{P_0(C_{10}-C_{00})}{P_1(C_{01}-C_{11})} \triangleq \eta, \tag{1}$$

when the observations are discrete-valued. In particular, in this case the likelihood function $L(\mathbf{y})$ is also discrete-valued;[1] for future convenience, we let $\mathcal{L} = \{\eta_0, \eta_1, \dots\}$ denote the alphabet of values of $\mathsf{L} = L(\mathbf{y})$, where $0 \leq \eta_0 < \eta_1 < \eta_2 < \cdots$. A consequence of this property is that for many sets of costs and *a priori* probability assignments, the resulting η formed in (1) will often not coincide with one of the possible values of L, i.e., we will often have $\eta \neq \eta_i$ for all i. In such cases, the case of equality in the likelihood ratio test will not arise, and the minimum Bayes risk is achieved by a likelihood ratio test that corresponds, as usual, to a unique (P_D, P_F) point on the associated operating characteristic.

However, for some choices of the costs and priors, the resulting threshold η will satisfy $\eta = \eta_i$ for some particular i. This means that in this case equality in the likelihood ratio test will occur with nonzero probability. Nevertheless, it is easily verified that the Bayes risk is the same no matter how a decision is made in this event. Hence, when equality occurs, the decision can still be made arbitrarily, but these choices will correspond to different points on the operating characteristic.

[1] Of course, the likelihood function can also be discrete-valued even when the data are continuous-valued in some cases. Consider, for example, piecewise constant densities. In such cases, our observations developed in this section also apply.

4.1.2 The Operating Characteristic of the Likelihood Ratio Test, and Neyman-Pearson Tests

Let us more generally examine the form of the operating characteristic associated with the likelihood ratio test in the case of discrete-valued observations. In particular, let us begin by sweeping the threshold η in (1) from 0 to ∞ and examining the (P_D, P_F) values that are obtained. As our development in the last section revealed, in order to ensure that each threshold η maps to a unique (P_D, P_F), we need to choose an arbitrary but fixed convention for handling the case of equality in (1). For this purpose let us associate equality with the decision $\hat{H}(\mathbf{y}) = H_1$, expressing the likelihood ratio test in the form

$$L(\mathbf{y}) = \frac{p_{\mathsf{y}|H}(\mathbf{y}|H_1)}{p_{\mathsf{y}|H}(\mathbf{y}|H_0)} \mathop{\gtrless}_{\hat{H}(\mathbf{y})=H_0}^{\hat{H}(\mathbf{y})=H_1} \eta. \tag{2}$$

Before developing further results, let's explore a specific example to illustrate some of the key ideas.

Example 1. Consider an experiment where the firing pattern of neurons in the visual cortex of the brain (in, e.g., a rhesus monkey) are measured in response to the presence and absence of some particular visual stimulus (e.g., showing him a banana). Later a data record of the firing is being analyzed to determine whether or not the stimulus was present at the time of the recording. When the stimulus is absent, the firings pattern obeys a Poisson process with average arrival rate μ_0; when the stimulus is present, the rate is μ_1 with $\mu_1 > \mu_0$. We count the number of firings y in the data record and use this observed data to make a decision.

The likelihood functions for this decision problem are, for $m = 0, 1$,

$$p_{y|H}(y|H_m) = \begin{cases} \dfrac{\mu_m^y e^{-\mu_m}}{y!} & y = 0, 1, \dots \\ 0 & \text{otherwise.} \end{cases}$$

In this example the likelihood ratio (1) leads to the test

$$L(y) = \left(\frac{\mu_1}{\mu_0}\right)^y e^{-(\mu_1 - \mu_0)} \mathop{\gtrless}_{\hat{H}(y)=H_0}^{\hat{H}(y)=H_1} \eta,$$

which further simplifies to

$$y \mathop{\gtrless}_{\hat{H}(y)=H_0}^{\hat{H}(y)=H_1} \frac{\ln \eta + (\mu_1 - \mu_0)}{\ln(\mu_1/\mu_0)} \triangleq \gamma. \tag{3}$$

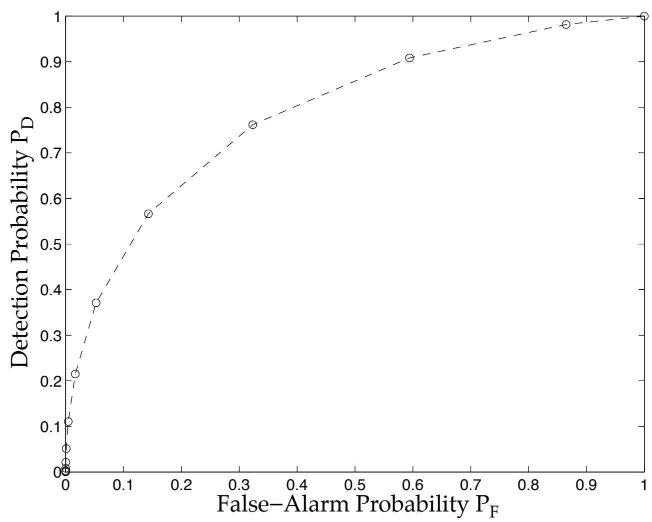

Figure 1: Operating characteristic associated with the likelihood ratio test for the Poisson stimulus detection problem. The neural firing rates under H_0 and H_1 are $\mu_0 = 2$ and $\mu_1 = 4$, respectively. The circles mark points achievable by likelihood ratio tests.

The discrete nature of the this hypothesis testing problem means that the operating characteristic associated with the likelihood ratio test (3) is a discrete collection of points rather than a continuous curve of the type we encountered in an earlier example involving Gaussian data. Indeed, while the left-hand side of (3) is integer-valued, the right side is not in general. As a result, we have that P_D and P_F are given in terms of γ by the expressions[2]

$$P_\mathrm{D} = \mathbb{P}\left(y \geq \gamma \mid \mathsf{H} = H_1\right) = \mathbb{P}\left(y \geq \lceil \gamma \rceil \mid \mathsf{H} = H_1\right) = \sum_{y \geq \lceil \gamma \rceil} \frac{m_1^y e^{-m_1}}{y!} \tag{4a}$$

$$P_\mathrm{F} = \mathbb{P}\left(y \geq \gamma \mid \mathsf{H} = H_0\right) = \mathbb{P}\left(y \geq \lceil \gamma \rceil \mid \mathsf{H} = H_0\right) = \sum_{y \geq \lceil \gamma \rceil} \frac{m_0^y e^{-m_0}}{y!}. \tag{4b}$$

The resulting operating characteristic is depicted in Fig. 1. In Figure 1, only the isolated $(P_\mathrm{D}, P_\mathrm{F})$ points indicated by circles are achievable by likelihood ratio tests. For example, the uppermost point is achieved for all $\gamma \leq 0$, the next highest point for all $0 < \gamma \leq 1$, the next for $1 < \gamma \leq 2$, and so on. This behavior is representative of such discrete decision problems.

As we discussed in Section 4.1.1, for Bayesian problems the specific cost and *a priori* probability assignments determine a threshold η in (2), which in turn corresponds to one of the isolated points that form the operating characteristic.

[2]We use $\lceil \cdot \rceil$ to denote the ceiling function, whose value is the smallest integer that is at least as large as its argument.

For Neyman-Pearson problems with such discrete observations, then, the threshold η in (1) is chosen so as to achieve the largest P_{D} subject to the constraint on the maximum allowable P_{F} (i.e., α). When α corresponds to at least one of the discrete points of the operating characteristic, then the corresponding P_{D} indicates the achievable detection probability. More typically, however, α will lie strictly between the P_{F} values of points on the operating characteristic. In this case, the appropriate operating point corresponds to the $(P_{\mathrm{D}}, P_{\mathrm{F}})$ whose P_{F} is the largest P_{F} that is smaller than α. This operating point then uniquely specifies the decision rule.

4.1.3 Improving Neyman-Pearson Decisions by Time-Sharing

There is a simple way to improve the average performance of such Neyman-Pearson tests in cases where the experiment is repeated many times, as we now develop. Since we know that the likelihood ratio test is the best deterministic test, it follows that the key to improving performance is to employ randomization. The particular randomization we describe now can be thought of as time-sharing.

We begin by noting that the likelihood ratio test is the best decision rule of the form

$$\hat{H}(\mathbf{y}) = \begin{cases} H_0 & \mathbf{y} \in \mathcal{Y}_0 \\ H_1 & \mathbf{y} \in \mathcal{Y}_1 \end{cases}$$

where $\mathcal{Y}_0$ and $\mathcal{Y}_1$ are the decision regions partitioning $\mathcal{Y}$. For these rules, each observation $\mathbf{y}$ maps to a unique decision $\hat{H}(\mathbf{y})$.

In contrast, with a randomized decision rule, a particular observation does not always produce the same decision. To see how this can help, consider the sequence of thresholds $\eta_0, \eta_1, \dots$ that correspond to values that the likelihood ratio can take on, and denote the corresponding operating points by $(P_{\mathrm{D}}(\eta_i), P_{\mathrm{F}}(\eta_i))$ for $i = 0, 1, \dots$. Determine $\hat{\imath}$ such that $\eta_{\hat{\imath}}$ is the threshold value that results in the likelihood ratio test with the smallest false-alarm probability that is greater than α. Then as illustrated in Fig. 2, $P_{\mathrm{F}}(\eta_{\hat{\imath}})$ and $P_{\mathrm{F}}(\eta_{\hat{\imath}+1})$ "bracket" α: using $\eta_{\hat{\imath}+1}$ results in test with the largest false-alarm probability that is less than α. Now consider the following randomized decision rule. We flip a biased coin for which the probability of "heads" is p and that of "tails" is $1-p$. If "heads" turns up, we use the likelihood ratio test with threshold $\eta_{\hat{\imath}}$; if "tails" turns up, we use the likelihood ratio test with threshold $\eta_{\hat{\imath}+1}$. This is referred to as "time-sharing," because over the course of many repeated experiments, one test is used a fraction p of the time, and the other a fraction $1-p$ of the time. Such a randomized test therefore achieves

$$\begin{aligned} P_{\mathrm{D}} &= pP_{\mathrm{D}}(\eta_{\hat{\imath}}) + (1-p)P_{\mathrm{D}}(\eta_{\hat{\imath}+1}) \\ P_{\mathrm{F}} &= pP_{\mathrm{F}}(\eta_{\hat{\imath}}) + (1-p)P_{\mathrm{F}}(\eta_{\hat{\imath}+1}), \end{aligned} \tag{5}$$

and corresponds to a point on the line segment connecting $(P_{\mathrm{D}}(\eta_{\hat{\imath}}), P_{\mathrm{F}}(\eta_{\hat{\imath}}))$ and $(P_{\mathrm{D}}(\eta_{\hat{\imath}+1}), P_{\mathrm{F}}(\eta_{\hat{\imath}+1}))$. This line segment is indicated with the dashed line segment in Fig. 2. In particular, as p is varied from 0 to 1, the operating point moves from

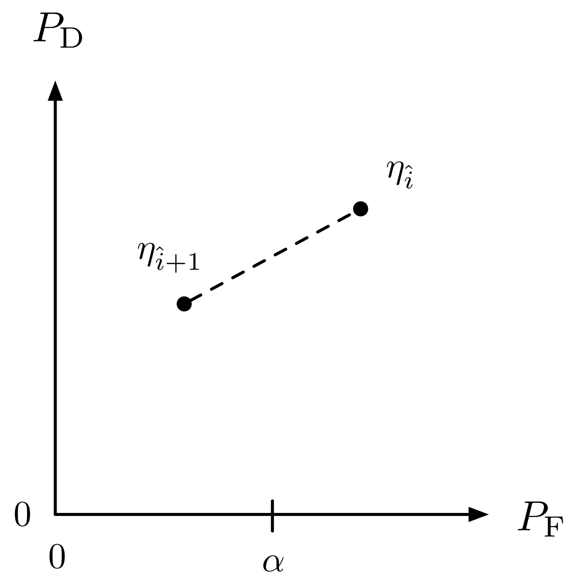

Figure 2: Achievable operating points for a randomization (time-sharing) between two likelihood ratio tests, one with threshold $\eta_{\hat{\imath}}$ and the other with threshold $\eta_{\hat{\imath}+1}$.

$(P_{\mathrm{D}}(\eta_{\hat{\imath}+1}), P_{\mathrm{F}}(\eta_{\hat{\imath}+1}))$ to $(P_{\mathrm{D}}(\eta_{\hat{\imath}}), P_{\mathrm{F}}(\eta_{\hat{\imath}}))$. Hence, by choosing p appropriately, we can achieve the false-alarm probability constraint with equality, i.e., there exists a $\hat{p}$ such that

$$P_{\mathrm{F}} = \hat{p} P_{\mathrm{F}}(\eta_{\hat{\imath}}) + (1-\hat{p}) P_{\mathrm{F}}(\eta_{\hat{\imath}+1}) = \alpha.$$

Since the operating characteristic is a monotonically increasing function, it follows that $P_{\mathrm{D}}(\eta_{\hat{\imath}}) > P_{\mathrm{D}}(\eta_{\hat{\imath}+1})$, which in turn implies that $\hat{p}$ is the value of p yielding the largest possible P_{D}.

Our randomization argument allows us to conclude that all points on the operating characteristic as well as all points on the line segments connecting operating characteristic points (adjacent or not) can be achieved via randomized likelihood ratio tests based on simple time-sharing. Referring to Example 1 in particular, this means that all operating points on the piecewise linear dashed line in Fig. 1 can be achieved, allowing any false alarm probability constraint α to be met with equality.

It is worth noting that such time-sharing strategies can be implemented in other ways that can sometimes be more convenient. To see this, note that for the randomized rule above, both likelihood ratio tests lead to the decision H_1 when $L(\mathbf{y}) \geq \eta_{\hat{\imath}+1}$. Hence, regardless of the outcome of the coin flip, the decision is H_1 for the associated values of $\mathbf{y}$. Similarly, if $L(\mathbf{y}) \leq \eta_{\hat{\imath}-1}$ then both tests lead to the decision H_0, so the decision is H_0 regardless of the outcome of the coin flip. Hence, only when $L(\mathbf{y}) = \eta_{\hat{\imath}}$ does the randomization play a role. In this case, if "heads" comes up, the decision will be H_1, while if "tails" comes up, the decision will be H_0. We can summarize this implementation of the test in the following form:

$$\mathbb{P}\left(\hat{H}(\mathbf{y}) = H_1 \;\middle|\; \mathbf{y} = \mathbf{y}\right) = p_{\hat{H}|\mathbf{y}}(H_1|\mathbf{y}) = \begin{cases} 1 & L(\mathbf{y}) \geq \eta_{\hat{\imath}+1} \\ \hat{p} & L(\mathbf{y}) = \eta_{\hat{\imath}} \\ 0 & L(\mathbf{y}) \leq \eta_{\hat{\imath}-1}. \end{cases} \tag{6}$$

Our results in this section have suggested that at least in some problems involving discrete-valued data and the Neyman-Pearson criterion a randomized test can lead to better (expected) performance than a deterministic test. This observation, in turn, raises some natural questions. For example, our time-sharing strategy is a very particular class of randomized tests—specifically, a simple random choice between the outcomes of two (deterministic) likelihood ratio tests. Would some other type of randomized test be able to perform better still? And could some more general form of randomized test be able to improve performance in the case of continuous-valued data with Bayesian or Neyman-Pearson criteria? To answer these questions, in the next section we develop decision rules optimized over of a broad class of randomized tests.

4.2 Randomized Tests

For a randomized test, the decision rule is a *random function* of the data, which we denote using $\hat{H}(\cdot)$. Hence, even for a deterministic argument $\mathbf{y}$, the decision $\hat{H}(\mathbf{y})$ is a random quantity. However, $\hat{H}(\mathbf{y})$ must have the Markov property that conditioned on knowledge of $\mathbf{y}$, the function is independent of the hypothesis H, which we express via the following Markov chain notation

$$H \leftrightarrow \mathbf{y} \leftrightarrow \hat{H}. \tag{7}$$

Such a test is fully described by the probabilities

$$p_{\hat{H}|\mathbf{y}}(H_m|\mathbf{y}) = p_{\hat{H}|\mathbf{y},H}(H_m|\mathbf{y}, H_i) \tag{8}$$

for $m = 0, 1$ (and $i = 0, 1$).

With this notation, we see that deterministic rules are a special case, corresponding to

$$p_{\hat{H}|\mathbf{y}}(H_1|\mathbf{y}) = \begin{cases} 1 & \mathbf{y} \in \mathcal{Y}_1 \\ 0 & \mathbf{y} \in \mathcal{Y}_0. \end{cases} \tag{9}$$

Moreover, it also follows immediately that tests formed by a random choice among two likelihood ratio tests, such as were considered in Section 4.1.2 are also special cases. For example, the test described via (6) corresponds to

$$p_{\hat{H}|\mathbf{y}}(H_1|\mathbf{y}) = \begin{cases} 1 & L(\mathbf{y}) \geq \eta_{\hat{\imath}+1} \\ \hat{p} & L(\mathbf{y}) = \eta_{\hat{\imath}} \\ 0 & L(\mathbf{y}) \leq \eta_{\hat{\imath}-1}. \end{cases} \tag{10}$$

More generally, for a randomized test that corresponds to the random choice between two likelihood ratio tests with respective thresholds η_1 and η_2 such that $\eta_2 > \eta_1$, and where the probability of selecting the first test is p, it is straightforward to verify that

$$p_{\hat{H}|\mathbf{y}}(H_1|\mathbf{y}) = \begin{cases} 1 & L(\mathbf{y}) \geq \eta_2 \\ p & \eta_1 \leq L(\mathbf{y}) < \eta_2 \\ 0 & L(\mathbf{y}) < \eta_1. \end{cases} \tag{11}$$

From our general characterization (8), we see that specifying a randomized decision rule is equivalent to specifying $p_{\hat{H}|\mathbf{y}}(H_0|\cdot)$—or $p_{\hat{H}|\mathbf{y}}(H_1|\cdot) = 1 - p_{\hat{H}|\mathbf{y}}(H_0|\cdot)$. Hence, determining the optimum randomized test for a given performance criterion involves solving for the optimum mapping $p_{\hat{H}|\mathbf{y}}(H_0|\cdot)$. Using this approach, we develop the optimum randomized tests for Bayesian and Neyman-Pearson hypothesis testing problems in the next two sections, respectively. As we will see this allows us to draw some sharp conclusions about when randomized tests are—and are not—needed.

4.2.1 Bayesian Case

We begin by establishing a more general version of our Bayesian result for the case of randomized tests.

Claim 1. *A randomized test cannot achieve a lower Bayes' risk than the optimum likelihood ratio test in binary Bayesian hypothesis testing.*

Proof. We consider the case of continuous-valued data; an identical argument is used in the discrete case. We begin by writing our Bayes risk in the form

$$\varphi(p_{\hat{H}|\mathbf{y}}(H_0|\cdot)) = \int \tilde{\varphi}(\mathbf{y})\, p_{\mathbf{y}}(\mathbf{y})\, \mathrm{d}\mathbf{y},$$

where

$$\tilde{\varphi}(\mathbf{y}) = \mathbb{E}\left[C(H, \hat{H}(\mathbf{y})) \,\middle|\, \mathbf{y} = \mathbf{y} \right].$$

Again we see it suffices to minimize $\tilde{\varphi}(\mathbf{y})$ for each $\mathbf{y}$. Applying, in turn, (8) and Bayes' Rule, we can write $\tilde{\varphi}(\mathbf{y})$ in the form

$$\begin{aligned} \tilde{\varphi}(\mathbf{y}) &= \sum_{i,j} C_{ij}\, \mathbb{P}\left(H = H_j, \hat{H}(\mathbf{y}) = H_i \,\middle|\, \mathbf{y} = \mathbf{y} \right) \\ &= \sum_{i,j} C_{ij}\, \mathbb{P}\left(H = H_j \mid \mathbf{y} = \mathbf{y} \right) \mathbb{P}\left(\hat{H} = H_i \,\middle|\, \mathbf{y} = \mathbf{y} \right) \\ &= \sum_{i,j} C_{ij}\, p_{\hat{H}|\mathbf{y}}(H_i|\mathbf{y})\, \frac{P_j\, p_{\mathbf{y}|H}(\mathbf{y}|H_j)}{p_{\mathbf{y}}(\mathbf{y})}, \end{aligned} \tag{12}$$

from which, with

$$\Delta(\mathbf{y}) \triangleq C_{10} \frac{P_0 p_{\mathsf{y}|H}(\mathbf{y}|H_0)}{p_{\mathsf{y}}(\mathbf{y})} + C_{11} \frac{P_1 p_{\mathsf{y}|H}(\mathbf{y}|H_1)}{p_{\mathsf{y}}(\mathbf{y})}, \tag{13}$$

we obtain

$$\begin{aligned}\tilde{\varphi}(\mathbf{y}) &= \Delta(\mathbf{y}) + p_{\hat{H}|\mathsf{y}}(H_0|\mathbf{y}) \frac{p_{\mathsf{y}|H}(\mathbf{y}|H_0)}{p_{\mathsf{y}}(\mathbf{y})} P_1(C_{01} - C_{11}) \left[\frac{p_{\mathsf{y}|H}(\mathbf{y}|H_1)}{p_{\mathsf{y}|H}(\mathbf{y}|H_0)} - \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} - \right] \\ &= \Delta(\mathbf{y}) + p_{\hat{H}|\mathsf{y}}(H_0|\mathbf{y}) \frac{p_{\mathsf{y}|H}(\mathbf{y}|H_0)}{p_{\mathsf{y}}(\mathbf{y})} P_1(C_{01} - C_{11}) \left[L(\mathbf{y}) - \eta \right]\end{aligned} \tag{14}$$

using the notation defined in (1). From (14) we can immediately conclude that $\tilde{\varphi}(\mathbf{y})$ is minimized over $0 \leq p_{\hat{H}|\mathsf{y}}(H_0|\mathbf{y}) \leq 1$ by choosing $p_{\hat{H}|\mathsf{y}}(H_0|\mathbf{y}) = 0$ when the term in brackets is positive and $p_{\hat{H}|\mathsf{y}}(H_0|\mathbf{y}) = 1$ when the term in brackets is negative. But this is precisely the (deterministic) likelihood ratio test (1) we developed earlier. Hence, for Bayesian problems, we can conclude that a deterministic test will always suffice. Moreover, this conclusion is invariant to whether the data are continuous- or discrete-valued. $\square$

4.2.2 Neyman-Pearson Case

We next establish a more general version of our Neyman-Pearson result for the case of randomized tests. We begin with the case of continuous-valued data, where the densities for the data are bounded under each hypothesis.

Claim 2. *Consider a binary hypothesis testing problem where the observations are continuous-valued with a bounded probability density function under each hypothesis. For a given P_F constraint, a randomized test cannot achieve a larger P_D than the optimum likelihood ratio test.*

Proof. As in the deterministic case, we follow a Lagrange multiplier approach, expressing our objective function as

$$\begin{aligned}\varphi(p_{\hat{H}|\mathsf{y}}(H_0|\cdot)) &= 1 - P_\mathrm{D} + \lambda(P_\mathrm{F} - \alpha') \\ &= \lambda(1-\alpha') + \mathbb{P}\left(\hat{H}(\mathbf{y}) = H_0 \;\middle|\; H = H_1\right) - \lambda\, \mathbb{P}\left(\hat{H}(\mathbf{y}) = H_0 \;\middle|\; H = H_0\right)\end{aligned} \tag{15}$$

for some $\alpha' \leq \alpha$. This time, though, (15) expands as

$$\varphi(p_{\hat{H}|\mathsf{y}}(H_0|\cdot)) = \lambda(1-\alpha') + \int p_{\hat{H}|\mathsf{y}}(H_0|\mathbf{y}) \left[p_{\mathsf{y}|H}(\mathbf{y}|H_1) - \lambda\, p_{\mathsf{y}|H}(\mathbf{y}|H_0) \right] \mathrm{d}\mathbf{y},$$

from which we obtain, using the definition of $L(\mathbf{y})$,

$$\begin{aligned}\varphi(p_{\hat{H}|\mathbf{y}}(H_0|\cdot)) &= \lambda(1-\alpha') + \int p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y}) \left[p_{\mathbf{y}|H}(\mathbf{y}|H_1) - \lambda p_{\mathbf{y}|H}(\mathbf{y}|H_0)\right] d\mathbf{y} \\ &= \lambda(1-\alpha') + \int p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y}) \left[L(\mathbf{y}) - \lambda\right] p_{\mathbf{y}|H}(\mathbf{y}|H_0)\,\mathrm{d}\mathbf{y}. \end{aligned} \tag{16}$$

Since $0 \le p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y}) \le 1$, the objective function (16) is minimized by setting $p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y}) = 0$ for all values of $\mathbf{y}$ such that the term in braces is positive, and $p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y}) = 1$ for all values of $\mathbf{y}$ such that the term in braces is negative; i.e.,

$$p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y}) = \begin{cases} 0 & L(\mathbf{y}) > \lambda \\ 1 & L(\mathbf{y}) < \lambda. \end{cases} \tag{17}$$

Hence, we can conclude that the optimum rule takes the form of time-sharing between likelihood ratio tests. In particular, at least except when $L(\mathbf{y}) = \lambda$ we have that

$$\hat{H}(\mathbf{y}) = \begin{cases} H_1 & L(\mathbf{y}) > \lambda \\ H_0 & L(\mathbf{y}) < \lambda. \end{cases} \tag{18}$$

It remains only to determine what the nature of the decision (i.e., $p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y})$) when $L(\mathbf{y}) = \lambda$ and the choice of λ. These quantities are determined by meeting the constraint $P_{\mathrm{F}} = \alpha'$. Specifically,

$$\begin{aligned} P_{\mathrm{F}} &= \int_{\mathcal{Y}} \left[1 - p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y})\right] p_{\mathbf{y}|H}(\mathbf{y}|H_0)\,\mathrm{d}\mathbf{y} \\ &= \int_{\{\mathbf{y}\in\mathcal{Y}\,:\,L(\mathbf{y})>\lambda\}} p_{\mathbf{y}|H}(\mathbf{y}|H_0)\,\mathrm{d}\mathbf{y} + \int_{\{\mathbf{y}\in\mathcal{Y}\,:\,L(\mathbf{y})=\lambda\}} \left[1 - p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y})\right] p_{\mathbf{y}|H}(\mathbf{y}|H_0)\,\mathrm{d}\mathbf{y} \\ &= \mathbb{P}\left[L(\mathbf{y}) > \lambda \mid H = H_0\right] + \int_{\{\mathbf{y}\in\mathcal{Y}\,:\,L(\mathbf{y})=\lambda\}} \left[1 - p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y})\right] p_{\mathbf{y}|H}(\mathbf{y}|H_0)\,\mathrm{d}\mathbf{y}, \end{aligned} \tag{19}$$

where the second equality follows from substituting for $p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y})$ using (17). Observe that $P_{\mathrm{F}} = 1$ if $\lambda < 0$, so it suffices to restrict our attention to $\lambda \ge 0$. Furthermore, note that the first term in (19) is a nonincreasing function of λ.

Now since $\mathbf{y}$ is a continuous-valued random variable, then the second term in (19) is zero and the remaining term, which is a continuous of λ, can be chosen so that it equals α'. In this case, it does not matter how we choose $p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y})$ when $L(\mathbf{y}) = \lambda$, so the optimum randomized rule degenerates to a deterministic likelihood ratio test again. It remains only to show that for optimum P_{D} we want $\alpha' = \alpha$. Given our preceding results, it suffices to exploit the fact that for likelihood ratio tests P_{D} is a monotonically nondecreasing function of P_{F}. $\square$

When the data are discrete-valued, the conclusion is different. In particular, the following establishes that the simple time-sharing strategy proposed at the end of Section 4.1.2 is optimal for discrete-valued data.

Claim 3. *Consider a binary hypothesis testing problem where the observations are discrete-valued. For a given P_{F} constraint, a randomized test can in general achieve a larger P_{D} than the optimum likelihood ratio test. Furthermore, the optimum randomized test corresponds to simple time-sharing between the two likelihood ratio tests achieving false-alarm probabilities closest to the target P_{F}.*

Proof. Proceeding as we did at the outset of Section 4.2.2, we obtain

$$\begin{aligned}\varphi(p_{\hat{H}|\mathsf{y}}(H_0|\cdot)) &= \lambda(1-\alpha') + \sum_{\mathbf{y}} p_{\hat{H}|\mathsf{y}}(H_0|\mathbf{y})\left[p_{\mathsf{y}|H}[\mathbf{y}|H_1] - \lambda p_{\mathsf{y}|H}[\mathbf{y}|H_0]\right] \\ &= \lambda(1-\alpha') + \sum_{\mathbf{y}} p_{\hat{H}|\mathsf{y}}(H_0|\mathbf{y})\left[L(\mathbf{y}) - \lambda\right] p_{\mathsf{y}|H}[\mathbf{y}|H_0]. \end{aligned} \tag{20}$$

from which we analogously conclude that for $L(\mathbf{y}) \neq \lambda$ we have (17). Hence, it remains only to determine λ and the decision when $L(\mathbf{y}) = \lambda$ from the false alarm constraint.

An important difference from the continuous case is that when y is discrete-valued, so that $L(\mathsf{y})$ is also discrete-valued, the first term in (19) is not a continuous function of λ, but is piecewise constant.

As before, let us denote the values that $L(\mathsf{y})$ takes on by $\eta_0, \eta_1, \eta_2, \dots,$ where $0 < \eta_0 < \eta_1 < \eta_2 < \cdots$. And let us choose $\hat{\imath}$ so that $\lambda = \eta_{\hat{\imath}}$ is the smallest threshold such that

$$\mathbb{P}\left[L(\mathsf{y}) > \lambda \mid H = H_0\right] = \mathbb{P}\left[L(\mathsf{y}) \geq \eta_{\hat{\imath}+1} \mid H = H_0\right] \leq \alpha'.$$

Then we can obtain $P_{\mathrm{F}} = \alpha'$ by choosing

$$p_{\hat{H}|\mathsf{y}}(H_1|\mathbf{y}) = 1 - p_{\hat{H}|\mathsf{y}}(H_0|\mathbf{y})$$

appropriately for all $\mathbf{y}$ such that $L(\mathbf{y}) = \eta_{\hat{\imath}}$. In particular, $0 < p_{\hat{H}|\mathsf{y}}(H_1|\cdot) < 1$ in this range must be chosen so that

$$\begin{aligned}\alpha' - \mathbb{P}\left(L(\mathsf{y}) \geq \eta_{\hat{\imath}+1} \mid H = H_0\right) &= \sum_{\{\mathbf{y} | L(\mathbf{y}) = \eta_{\hat{\imath}}\}} p_{\hat{H}|\mathsf{y}}(H_1|\mathbf{y})\, p_{\mathsf{y}|H}[\mathbf{y}|H_0] \\ &= q\; \mathbb{P}\left(L(\mathsf{y}) = \eta_{\hat{\imath}} \mid H = H_0\right), \end{aligned} \tag{21}$$

where

$$q \triangleq \mathbb{E}\left[p_{\hat{H}|\mathsf{y}}(H_1|\mathsf{y}) \mid L(\mathsf{y}) = \eta_{\hat{\imath}}\right]. \tag{22}$$

As we would expect, our decision probabilities $p_{\hat{H}|\mathsf{y}}(H_1|\cdot)$ when $L(\mathbf{y}) = \eta_{\hat{\imath}}$ appear in (21) only through q, so it suffices to appropriately select the latter.

For $q = 0$, the resulting decision rule is the deterministic test

$$L(\mathbf{y}) \underset{\hat{H}(\mathbf{y})=H_0}{\overset{\hat{H}(\mathbf{y})=H_1}{\gtrless}} \eta_{\hat{\imath}+1}. \tag{23}$$

More generally, when $q > 0$, we have that the decision rule involves a random choice between the decision rule (23) and

$$L(\mathbf{y}) \underset{\hat{H}(\mathbf{y})=H_0}{\overset{\hat{H}(\mathbf{y})=H_1}{\gtrless}} \eta_{\hat{\imath}}. \tag{24}$$

In particular, the test (24) is chosen with probability q, and the test (23) with probability $1-q$. In terms of the operating characteristic of the likelihood ratio test, this is a point on the line segment connecting the points corresponding to the two deterministic tests (23) and (24).

It remains only to verify that for optimum P_{D} we want $\alpha' = \alpha$ in the discrete case as well. However, for likelihood ratio tests involving discrete data, the P_{D} values form a nondecreasing sequence as a function of P_{F}. Then since the performance of randomized tests corresponds to points on the line segments connecting the performance points associated with deterministic tests, P_{D} is a nondecreasing function of P_{F} for our more general class of randomized tests as well. □

In summary, optimum Neyman-Pearson decision rules always take the form of either a deterministic rule in the form of a likelihood ratio test or a randomized rule in the form of simple time-sharing between two likelihood ratio tests. This insight allows us to draw a sharp conclusion about which $(P_{\mathrm{D}}, P_{\mathrm{F}})$ values are achievable by some decision rule and which cannot be achieved by any decision rule, as we develop in the sequel.

4.3 The Efficient Frontier of Operation for Hypothesis Testing

By exploiting the special role that the likelihood ratio test plays in optimum decisions rules, we can develop a number of key properties of the boundary of the region of achievable $(P_{\mathrm{D}}, P_{\mathrm{F}})$ operating points.

This boundary is the convex hull of the set of points achievable by likelihood ratio tests. That is, when the data are continuous-valued, it is the operating characteristic of the likelihood ratio test. When the data are discrete-valued, it is the set of points achievable by the likelihood ratio test, together with the straight line segments connect all adjacent pairs of such points.

This boundary is naturally interpreted as the efficient frontier of operating points, since any other achievable pair away from this boundary suffers either worse P_{D}, or worse P_{F}, or both, which cannot be preferable for any reasonable problem formulation.

We emphasize at the outset that the detailed shape of the frontier is determined by the measurement model for the data, since it is this information that is used to construct the likelihood ratio $L(\mathbf{y})$. However, all frontiers share some important characteristics in common, and it is these that we explore in this section. As a simple example, which was mentioned earlier, we have the following.

Fact 1. *The $(P_{\mathrm{D}}, P_{\mathrm{F}})$ points $(0,0)$ and $(1,1)$ always lie on the efficient frontier of operating points.*

Proof. It suffices to let $\eta \to \infty$ and $\eta \to 0$, respectively, in the likelihood ratio test. □

It is also straightforward to verify the following.

Fact 2. *On the efficient frontier, $P_{\mathrm{D}} \geq P_{\mathrm{F}}$, i.e., that the efficient frontier lies above the diagonal in the P_{D}–P_{F} plane.*

Proof. This can be verified using a randomization argument. In particular, suppose our decision rule ignores the data $\mathbf{y}$ and bases its decision solely on the outcome of a biased coin flip, where the probability of "heads" is p. If the coin comes up "heads" we make the decision $\hat{H}(\mathbf{y}) = H_1$, while if it comes up "tails" we make the decision $\hat{H}(\mathbf{y}) = H_0$. Then for this rule we have

$$\begin{aligned} P_{\mathrm{D}} &= \mathbb{P}\left(\hat{H}(\mathbf{y}) = H_1 \;\middle|\; \mathsf{H} = H_1\right) = \mathbb{P}\left(\hat{H}(\mathbf{y}) = H_1\right) = p \\ P_{\mathrm{F}} &= \mathbb{P}\left(\hat{H}(\mathbf{y}) = H_1 \;\middle|\; \mathsf{H} = H_0\right) = \mathbb{P}\left(\hat{H}(\mathbf{y}) = H_1\right) = p, \end{aligned}$$

which corresponds to a point on the diagonal in the P_{D}–P_{F} plane. Hence, we can achieve all points on the diagonal by varying the bias p in the coin. However, for a given P_{F} this decision rule cannot provide a better P_{D} than the corresponding point on the efficient frontier; otherwise, this would contradict our Neyman-Pearson result that this frontier defines the best achievable P_{D} for a given P_{F}.[3] Hence, we conclude $P_{\mathrm{D}} \geq P_{\mathrm{F}}$. □

Randomization arguments play an important role in establishing other properties of likelihood ratio test operating characteristics as well. For example, we have the following.

Fact 3. *The efficient frontier of operating points is a concave function.*

[3] Indeed, we wouldn't expect to obtain better performance by ignoring the data $\mathbf{y}$.

Proof. Let $(P_\mathrm{D}(\eta_1), P_\mathrm{F}(\eta_1))$ and $(P_\mathrm{D}(\eta_2), P_\mathrm{F}(\eta_2))$ be points on the operating characteristic of the likelihood ratio test, corresponding to two arbitrary thresholds η_1 and η_2, respectively.

Let us argue our result by contradiction. Suppose the frontier is not concave. Then for some choice of η_1 and η_2, the straight line segment joining these two points lies above the frontier. However, the points on this straight line segment can be parameterized according to

$$(P_\mathrm{D}, P_\mathrm{F}) = (pP_\mathrm{D}(\eta_1) + (1-p)P_\mathrm{D}(\eta_2), pP_\mathrm{F}(\eta_1) + (1-p)P_\mathrm{F}(\eta_2)), \tag{25}$$

where $0 \leq p \leq 1$ is the parameter. Moreover, the points (25) can be achieved via the following randomized test: a biased coin is flipped, and if it turns up "heads," the likelihood ratio test with threshold η_1 is used; otherwise, the likelihood ratio test with threshold η_2 is used. Again by varying the bias p in our coin, we can achieve all the points on the line segment. But then this randomized test achieves a better P_D value for a given P_F value than the optimum Neyman-Pearson test, which achieves points on the frontier, which is a contradiction. Thus, the frontier must be concave. □

Let's consider one final property.

Fact 4. *For a binary hypothesis test with continuous-valued data, at those points where it is defined, the slope of the frontier is equal to the corresponding threshold η, i.e.,*

$$\frac{dP_\mathrm{D}}{dP_\mathrm{F}} = \eta. \tag{26}$$

Note that since $\eta \geq 0$, this is another way of verifying that the frontier is nondecreasing.

Proof. With

$$\mathcal{Y}_1(\eta) = \{\mathbf{y} \in \mathcal{Y} \,:\, L(\mathbf{y}) > \eta\} \tag{27}$$

we have

$$P_\mathrm{D}(\eta) = \int_{\mathcal{Y}_1(\eta)} p_{\mathbf{y}|H}(\mathbf{y}|H_1)\,\mathrm{d}\mathbf{y},$$

which after applying the definition of the likelihood function (1) yields

$$P_\mathrm{D}(\eta) = \int_{\mathcal{Y}_1(\eta)} L(\mathbf{y})\, p_{\mathbf{y}|H}(\mathbf{y}|H_0)\,\mathrm{d}\mathbf{y}. \tag{28}$$

Next, note that using the following convenient variant of our Kronecker notation

$$\mathbb{1}_+(x) \triangleq \begin{cases} 1 & x > 0 \\ 0 & \text{otherwise} \end{cases},$$

we have

$$\mathbb{1}_+(L(\mathbf{y}) - \eta) \triangleq \begin{cases} 1 & \mathbf{y} \in \mathcal{Y}_1(\eta) \\ 0 & \text{otherwise} \end{cases}. \tag{29}$$

In turn, using the result (29) in (28) we obtain

$$\begin{aligned} P_{\mathrm{D}}(\eta) &= \mathbb{E}\left[\mathbb{1}_+(L(\mathsf{y}) - \eta)\, L(\mathsf{y}) \mid \mathsf{H} = H_0\right] \\ &= \mathbb{E}\left[\mathbb{1}_+(\mathsf{L} - \eta)\, \mathsf{L} \mid \mathsf{H} = H_0\right] \\ &= \int_\eta^\infty L\, p_{\mathsf{L}|\mathsf{H}}(L|H_0)\, \mathrm{d}L \end{aligned} \tag{30}$$

Finally, differentiating (30) with respect to η then yields

$$\frac{dP_{\mathrm{D}}}{d\eta} = -\eta\, p_{\mathsf{L}|\mathsf{H}}(\eta|H_0). \tag{31}$$

However, since

$$P_{\mathrm{F}} = \int_\eta^\infty p_{\mathsf{L}|\mathsf{H}}(L|H_0)\, \mathrm{d}L$$

we know

$$\frac{dP_{\mathrm{F}}}{d\eta} = -p_{\mathsf{L}|\mathsf{H}}(\eta|H_0). \tag{32}$$

Dividing (31) by (32) we obtain (26). □

4.3.1 Achievable Operating Points

Having determined some of the structure of the efficient frontier, let us next explore more generally how to use this frontier to define what operating points can be achieved by any test—deterministic or randomized, and regardless of the criterion with respect to which it was designed—in the P_{D}–P_{F} plane. First, we note that every point between the frontier and the diagonal ($P_{\mathrm{D}} = P_{\mathrm{F}}$) can be achieved by a simple randomized test. To see this it suffices to recognize that every point in this region lies on some line connecting two points on the frontier. Hence, every such point can be achieved by a simple randomization between two likelihood ratio tests having the corresponding thresholds, using a coin with suitable bias p.

The test that achieves a given operating point in this region need not be unique, however. To illustrate this, let η_0 be the threshold corresponding to a particular point on the frontier. Then if for our decision rule we use a random choice (using a coin with bias p) between the outcome of this likelihood ratio test and that of the "pure guessing" rule that achieves an arbitrary point on the diagonal (using a different coin with bias q). Hence, by choosing $\eta_0 > 0$, $0 \leq p \leq 1$, and $0 \leq q \leq 1$, appropriately, our "doubly-randomized" test can also achieve any desired point in the region of interest.

Let us next consider which points below the diagonal are achievable. This can be addressed via a simple "rule-reversal" argument. For this we require the following

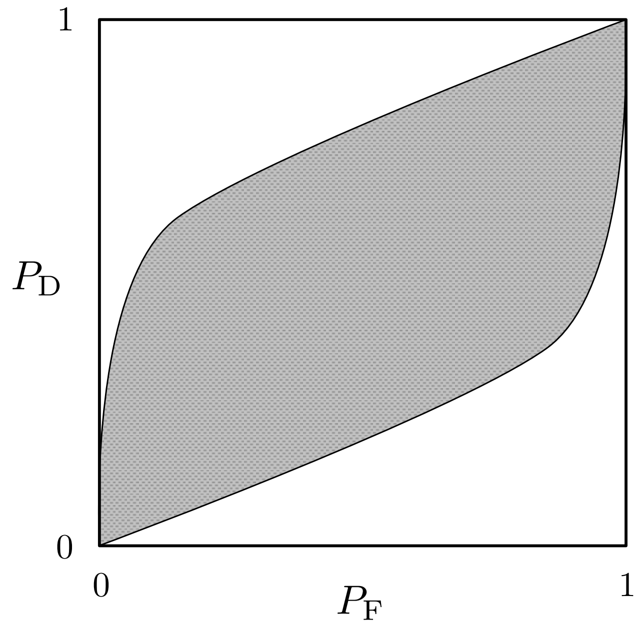

Figure 3: Achievable region in the P_{D}–P_{F} plane.

notion of a reversed test. If $\hat{H}(\cdot)$ describes a deterministic decision rule, then the corresponding reversed rule, which we denote using $\overline{\hat{H}}(\cdot)$, is simply one whose decisions are made as follows:

$$\overline{\hat{H}}(\mathbf{y}) = \begin{cases} H_0 & \hat{H}(\mathbf{y}) = H_1 \\ H_1 & \hat{H}(\mathbf{y}) = H_0 \end{cases}.$$

More generally, if $p_{\hat{H}|\mathbf{y}}(H_0|\cdot)$ describes a randomized decision rule, then the corresponding reversed decision rule, which we denote using $\bar{p}_{\hat{H}|\mathbf{y}}(H_0|\cdot)$, is defined via

$$\bar{p}_{\hat{H}|\mathbf{y}}(H_0|\cdot) = p_{\hat{H}|\mathbf{y}}(H_1|\mathbf{y}) = 1 - p_{\hat{H}|\mathbf{y}}(H_0|\mathbf{y}).$$

If a deterministic or randomized test achieves the operating point $(P_{\text{D}}, P_{\text{F}}) = (\beta, \alpha)$, then it is easy to verify that the corresponding reversed test achieves the operating point $(P_{\text{D}}, P_{\text{F}}) = (1 - \beta, 1 - \alpha)$, i.e.,

$$\begin{aligned} \mathbb{P}\left(\overline{\hat{H}}(\mathbf{y}) = H_1 \;\middle|\; H = H_1\right) &= \mathbb{P}\left(\hat{H}(\mathbf{y}) = H_0 \;\middle|\; H = H_1\right) = 1 - \beta \\ \mathbb{P}\left(\overline{\hat{H}}(\mathbf{y}) = H_1 \;\middle|\; H = H_0\right) &= \mathbb{P}\left(\hat{H}(\mathbf{y}) = H_0 \;\middle|\; H = H_0\right) = 1 - \alpha. \end{aligned}$$

Using this property, it follows that those points lying on the curve corresponding to the frontier reflected across the $P_{\text{D}} = 1/2$ and $P_{\text{F}} = 1/2$ lines are achievable by likelihood ratio tests whose decisions are reversed. In turn, all points between the diagonal and this "reflected efficient frontier" are achievable using a suitably designed randomized test. Hence, as illustrated in Fig. 3 we can conclude that all points within the region bounded by the efficient frontier and reversed operating characteristic curves can be achieved using a suitably designed decision rule.

We can also establish the converse: that no decision rule—deterministic or randomized—can achieve points outside this region. That no $(P_{\mathrm{D}}, P_{\mathrm{F}})$ point above the efficient frontier can be achieved follows from our Neyman-Pearson results. That no point below the reflected frontier can be achieved (including $(P_{\mathrm{D}}, P_{\mathrm{F}}) = (0, 1)$!) follows as well, using a proof-by-contradiction argument. In particular, if such a point could be achieved, then so could its reflection via a reversed test. However, this reflection would then lie above the efficient frontier, which would contradict the Neyman-Pearson optimality of the likelihood ratio test.