Chapter 11 of 3rd ed.

Chapter 8
Hypothesis Testing

Chien-Kang Huang (黃乾綱)

臺大工科海洋系

Outline

8.1 Significance Testing

8.2 Binary Hypothesis Testing (11.1)

8.3 Multiple Hypothesis Test (11.2)

- Some of the most important applications of probability theory involve reasoning in the presence of uncertainty. In these applications, we analyze the observations of an experiment in order to arrive at a conclusion. When the conclusion is based on the properties of random variables, the reasoning is referred to as **statistical inference**.
- Chapter 7: two types of statistical inference for model parameters:
  - Point estimation
  - Confidence interval estimation

- Two more categories of statistical inference
- **Significance Testing**
  - **Conclusion** Accept or reject the hypothesis that the observation result from a certain probability model H_0
  - **Accuracy Measure** Probability of rejecting the hypothesis when it is true
- **Hypothesis Testing**
  - **Conclusion** The observations result from one of M hypothetical probability models: $H_0, H_1, \ldots, H_{M-1}$
  - **Accuracy Measure** Probability that the conclusion is H_i when the true model is H_j for $i, j = 0, 1, \ldots, M-1$.

Example 8.1

Suppose $X_1, \ldots, X_n$ are iid samples of an exponential (λ) random variable X with unknown parameter λ. Using the observations $X_1, \ldots, X_n$, each of the statistical inference methods can answer questions regarding the unknown λ. For each of the methods, we state the underlying assumptions of the method and a question that can be addressed by the method.

- **Significance Test** Assuming λ is a constant, should we accept or reject the hypothesis that $\lambda = 3.5$?

- **Hypothesis Test** Assuming λ is a constant, does λ equal 2.5, 3.5, or 4.5?

8.1 Significance Testing

- A significance test begins with the hypothesis, H_0, that a certain probability model describes the observations of an experiment. The question addressed by the test has two possible answers: **accept** the hypothesis or **reject** it.
- The significance level of the test is defined as the probability of rejecting the hypothesis if it is true. The test divides S, the sample space of the experiment , into an event space consisting of an acceptance set A and a rejection set $R = A^C$, if the observation $s \in A$ we accept H_0. If $s \in R$, we reject the hypothesis. Therefore the significance level is

$$\alpha = P[s \in R]$$

- To design a significance test, we start with a value of α and then determine a set R that satisfies the equation.

- In many application, H_0 is referred to as the **null hypothesis**.
- In these applications there is a known probability model for an experiment. Then the conditions of the experiment change and a significance test is performed to determine whether the original probability model remains valid. The null hypothesis states that the changes in the experiment have no effect on the probability model.

Example 8.2 Problem

Suppose that on Thursdays between 9:00 and 9:30 at night, the number of call attempts N at a telephone switching office is a Poisson random variable with expected value 1000. Next Thursday, the President will deliver a speech at 9:00 that will be broadcast by all radio and television networks. The null hypothesis, H_0, is that the speech does not affect the probability model of telephone calls. In other words, H_0 states that on the night of the speech, N is a Poisson random variable with expected value 1000. Design a significance test for hypothesis H_0 at a significance level of $\alpha = 0.05$.

Example 8.2 Solution

The experiment involves counting the call requests, N, between 9:00 and 9:30 on the night of the speech. To design the test, we need to specify a rejection set, R, such that $P[N \in R] = 0.05$. There are many sets R that meet this condition. We do not know whether the President's speech will increase the number of phone calls (by people deprived of their Thursday programs) or decrease the number of calls (because many people who normally call listen to the speech). Therefore, we choose R to be a symmetrical set $\{n : |n - 1000| \geq c\}$. The remaining task is to choose c to satisfy Equation (8.1). Under hypothesis H_0, $E[N] = \mathrm{Var}[N] = 1000$. The significance level is

$$\alpha = P[|N - 1000| \geq c] = P\left[\left|\frac{N - E[N]}{\sigma_N}\right| \geq \frac{c}{\sigma_N}\right].$$

[Continued]

Example 8.2 Solution (continued)

Since $E[N]$ is large, we can use the central limit theorem and approximate $(N - E[N])/\sigma_N$ by the standard Gaussian random variable Z so that

$$\alpha \approx P\left[|Z| \geq \frac{c}{\sqrt{1000}}\right] = 2\left[1 - \Phi\left(\frac{c}{\sqrt{1000}}\right)\right] = 0.05.$$

In this case, $\Phi(c/\sqrt{1000}) = 0.975$ and $c = 1.95\sqrt{1000} = 61.7$. Therefore, if we observe more than $1000 + 61$ calls or fewer than $1000 - 61$ calls, we reject the null hypothesis at significance level 0.05.

- In Experiment 8.2, we implicitly assume that the alternative to the null hypothesis is a probability model with an expected value that is either higher than 1000 or lower than 1000.

- In a significance test, two kinds of errors are possible, Statistician refer to them as *Type I error* and *Type II error* with the following definitions:
- **Type I Error** *False Rejection*: Reject H_0, when H_0 is true.
- **Type II Error** *False Acceptance*: Accept H_0, when H_0 is false.
- The hypothesis specified in a significance test makes it possible to calculate the probability of a Type I error, $\alpha = P[s \in R]$.
- In the absence of a probability model for the condition "H_0 false," there is no way to calculate the probability of a Type II error.

- A binary hypothesis test described in Section 8.2 includes an alternative hypothesis H_1. Then it is possible to use the probability model given by H_1 to calculate the probability of a Type II error. Which is $\alpha = P[s \in A | H_1]$.
- Although a significance test does not specify a complete probability model as an alternative to the null hypothesis, the nature of the experiment influences the choice of the rejection set, R.

	Null Hypothesis (H0) is true	Alternative Hypothesis (H1) is true
Fail to Reject Null Hypothesis	Right decision	Wrong decision Type II Error False Negative
Reject Null Hypothesis	Wrong decision Type I Error False Positive	Right decision

http://en.wikipedia.org/wiki/Type_I_and_type_II_errors

Example 8.3 Problem

Before releasing a diet pill to the public, a drug company runs a test on a group of 64 people. Before testing the pill, the probability model for the weight of the people measured in pounds, is a Gaussian $(190, 24)$ random variable W. Design a test based on the sample mean of the weight of the population to determine whether the pill has a significant effect. The significance level is $\alpha = 0.01$.

Example 8.3 Solution

Under the null hypothesis, H_0, the probability model after the people take the diet pill, is a Gaussian $(190, 24)$, the same as before taking the pill. The sample mean, $M_{64}(X)$, is a Gaussian random variable with expected value 190 and standard deviation $24/\sqrt{64} = 3$. To design the significance test, it is necessary to find R such that $P[M_{64}(X) \in R] = 0.01$. If we reject the null hypothesis, we will decide that the pill is effective and release it to the public.

In this example, we want to know whether the pill has caused people to lose weight. If they gain weight, we certainly do not want to declare the pill effective. Therefore, we choose the rejection set R to consist entirely of weights below the original expected value: $R = \{M_{64}(X) \leq r_0\}$. We choose r_0 so that the probability that we reject the null hypothesis is 0.01:

$$P\left[M_{64}(X) \in R\right] = P\left[M_{64}(X) \leq r_0\right] = \Phi\left(\frac{r_0 - 190}{3}\right) = 0.01.$$

Since $\Phi(-2.33) = Q(2.33) = 0.01$, it follows that $(r_0 - 190)/3 = -2.33$, or $r_0 = 183.01$. Thus we will reject the null hypothesis and accept that the diet pill is effective at significance level 0.01 if the sample mean of the population weight drops to 183.01 pounds or less.

- Note the difference between the symmetrical reject set in Example 8.2 and the one-sided rejection set in Example 8.3, We selected these set on the basis of the application of the results of the test. In the language of statistical inference, the symmetrical set is part of a **two-tail significance test**, and the one-sided rejection set is part of a **one-tail significance test**.

Quiz 8.1

Under hypothesis H_0, the interarrival times between phone calls are independent and identically distributed exponential (1) random variables. Given X, the maximum among 15 independent interarrival time samples $X_1, \ldots, X_{15}$, design a significance test for hypothesis H_0 at a level of $\alpha = 0.01$.

Quiz 8.1 Solution

From the problem statement, each X_i has PDF and CDF

$$f_{X_i}(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad F_{X_i}(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geq 0 \end{cases}$$

Hence, the CDF of the maximum of $X_1, \ldots, X_{15}$ obeys

$$F_X(x) = P[X \leq x] = P[X_1 \leq x, X_2 \leq x, \cdots, X_{15} \leq x] = [P[X_i \leq x]]^{15}.$$

This implies that for $x \geq 0$,

$$F_X(x) = \left[F_{X_i}(x)\right]^{15} = \left[1 - e^{-x}\right]^{15}$$

To design a significance test, we must choose a rejection region for X. A reasonable choice is to reject the hypothesis if X is too small. That is, let $R = \{X \leq r\}$. For a significance level of $\alpha = 0.01$, we obtain

$$\alpha = P[X \leq r] = (1 - e^{-r})^{15} = 0.01$$

It is straightforward to show that

$$r = -\ln\left[1 - (0.01)^{1/15}\right] = 1.33$$

Hence, if we observe $X < 1.33$, then we reject the hypothesis.

8.2 Binary Hypothesis Testing

- In a binary hypothesis test, there are two hypothetical probability models, H_0 and H_1, and two possible conclusions: accept H_0 as the true model , and accept H_1.
- There is also a probability model for H_0 and H_1, conveyed by the numbers $P[H_0]$ and $P[H_1] = 1 - P[H_0]$. These numbers are referred to as the *a priori* probabilities or prior probabilities of H_0 and H_1.
- They reflect the state of knowledge about the probability model before an outcome is observed.

- The complete experiment for a binary hypothesis test consists of two subexperiments.

1. The first subexperiment chooses a probability model from sample space $S = \{H_0, H_1\}$. The probability models H_0 and H_1 have the same space, S.

2. The second subexperiment produces an observation corresponding to an outcome, $s \in S$. When the observation leads to a random vector $\mathbf{X}$, we call $\mathbf{X}$ the **decision statistic**. Often, the decision statistic is simply a random variable X.

- When the decision statistic $\mathbf{X}$ is discrete, the probability models are conditional probability mass function $P_{\mathbf{X}|H_0}(\mathbf{x})$ and $P_{\mathbf{X}|H_1}(\mathbf{x})$.
- When $\mathbf{X}$ is continuous, the probability models are conditional probability density function $f_{\mathbf{X}|H_0}(\mathbf{x})$ and $f_{\mathbf{X}|H_1}(\mathbf{x})$.
- In the terminology of statistical inference, theses functions are referred to as **likelihood function**. For example, $f_{\mathbf{X}|H_0}(\mathbf{x})$ is the likelihood of $\mathbf{x}$ given H_0.

- The test design divides S into two sets, A_0 and $A_1 = A_0{}^C$. If the outcomes $s \in A_0$, the conclusion is accept H_0. Otherwise, the conclusion is accept H_1.
  - The accuracy measure of the test consists of **two error probabilities**.
  - $P[A_1|H_0]$ is the probability of accepting H_1 when H_0 is the true probability model. It corresponds to the probability of a **Type I error**.
  - $P[A_0|H_1]$ is the probability of accepting H_0 when H_1 is the true probability model. It corresponds to the probability of a **Type II error**.

	Null Hypothesis (H0) is true	Alternative Hypothesis (H1) is true
Fail to Reject Null Hypothesis (A0)	Right decision	Wrong decision Type II Error, P[A0|H1] False Negative
Reject Null Hypothesis (A1)	Wrong decision Type I Error, P[A1|H0] False Positive	Right decision

Electrical engineering application: radar system

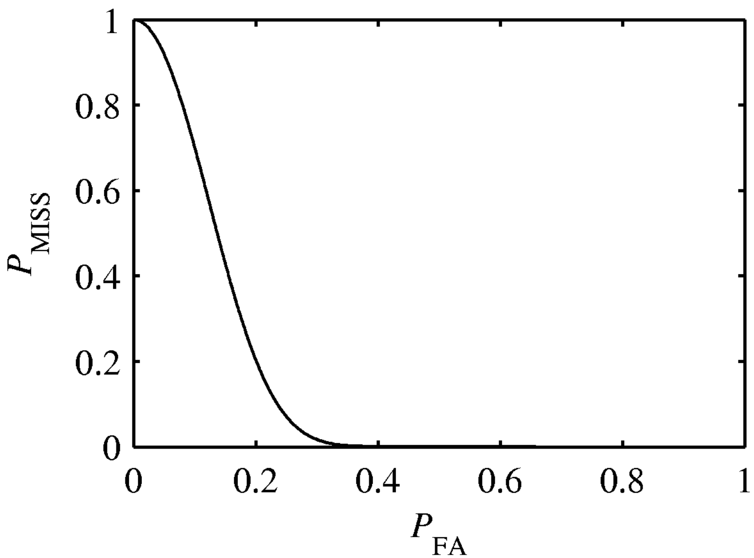
- Electrical engineering application: radar system
  - Transmitter sends out a signal, and receiver decides whether a target is present.
  - H_0: corresponds to the situation in which there is no target.
  - H_1: corresponds to the presence of a target.
  - Type I error (conclude target present when there is no target) is referred to as a false alarm.
  - Type II error (conclude no target when there is a target present) is referred to as a miss.

Trade-off of binary hypothesis test

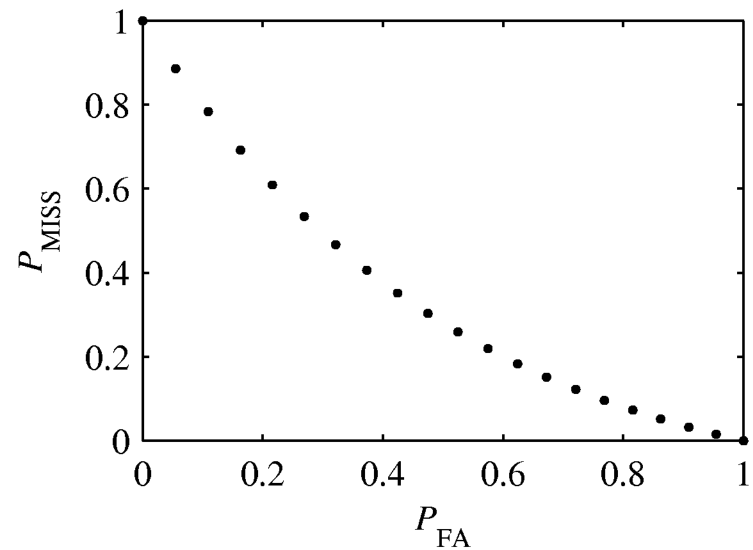
- The design of a binary hypothesis test represents a trade-off between the two error probabilities $P_{\mathrm{FA}} = P[A_1|H_0]$ and $P_{\mathrm{MISS}} = P[A_0|H_1]$.
- To understand the trade-off, consider two extreme designs:
  $A_0 = S$ and $A_1 = \phi \Rightarrow P_{\mathrm{FA}} = 0$ and $P_{\mathrm{MISS}} = 1$.
  $A_0 = \phi$ and $A_1 = S \Rightarrow P_{\mathrm{FA}} = 1$ and $P_{\mathrm{MISS}} = 0$.
- A graph representing the possible value of P_{FA} and P_{MISS} is referred to as a **receiver operating curve** (**ROC**). Figure 8.1.

Figure 8.1

Figure 8.1



ROC for continuous X



ROC for discrete X

Continuous and discrete examples of a receiver operating curve (ROC).

Example 8.4 Problem

The noise voltage in a radar detection system is a Gaussian $(0, 1)$ random variable, N. When a target is present, the received signal is $X = v + N$ volts with $v \geq 0$. Otherwise the received signal is $X = N$ volts. Periodically, the detector performs a binary hypothesis test with H_0 as the hypothesis *no target* and H_1 as the hypothesis *target present.* The acceptance sets for the test are $A_0 = \{X \leq x_0\}$ and $A_1 = \{X > x_0\}$. Draw the receiver operating curves of the radar system for the three target voltages $v = 0, 1, 2$ volts.

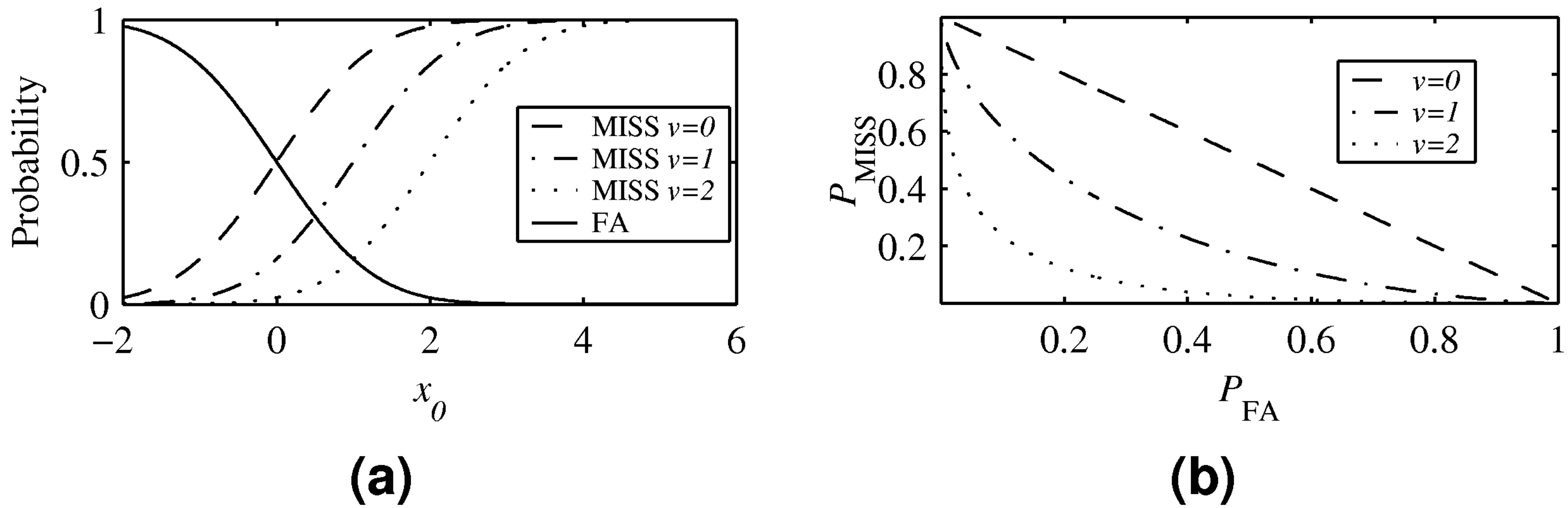
Example 8.4 Solution

To derive a receiver operating curve, it is necessary to find P_{MISS} and P_{FA} as functions of x_0. To perform the calculations, we observe that under hypothesis H_0, $X = N$ is a Gaussian $(0, \sigma)$ random variable. Under hypothesis H_1, $X = v + N$ is a Gaussian (v, σ) random variable. Therefore,

$$\begin{aligned} P_{\text{MISS}} &= P\left[A_0 | H_1\right] = P\left[X \leq x_0 | H_1\right] = \Phi(x_0 - v) \\ P_{\text{FA}} &= P\left[A_1 | H_0\right] = P\left[X > x_0 | H_0\right] = 1 - \Phi(x_0). \end{aligned}$$

Figure 8.2(a) shows P_{MISS} and P_{FA} as functions of x_0 for $v = 0$, $v = 1$, and $v = 2$ volts. Note that there is a single curve for P_{FA} since the probability of a false alarm does not depend on v. The same data also appears in the corresponding receiver operating curves of Figure 8.2(b). When $v = 0$, the received signal is the same regardless of whether or not a target is present. In this case, $P_{\text{MISS}} = 1 - P_{\text{FA}}$. As v increases, it is easier for the detector to distinguish between the two targets. We see that the ROC improves as v increases. That is, we can choose a value of x_0 such that both P_{MISS} and P_{FA} are lower for $v = 2$ than for $v = 1$.

Figure 8.2



(a) The probability of a miss and the probability of a false alarm as a function the threshold x_0 for Example 8.4. **(b)** The corresponding receiver operating curve for the system. We see that the ROC improves as v increases.

Trade-off between P_{FA} and P_{MISS}

- In a practical binary hypothesis test, it is necessary to adopt one test (a specific A_0) and a corresponding trade-off between P_{FA} and P_{MISS}. There are many approaches to selecting A_0.
- In the radar application, the cost of miss (ignoring a threatening target) could be far higher than the cost of a false alarm (causing the operator to take an unnecessary precaution).
  This suggest that the radar system should operate with a low value of x_0 to produce a low P_{MISS} even though this will produce a relatively high P_{FA}.

- The remainder of this section describes four methods of choosing A_0.

1. Maximum A posteriori Probability (MAP) test.
2. Minimum Cost Test
3. Neyman-Pearson Test
4. Maximum Likellihood Test.

Example 8.5

A modem transmits a binary signal to another modem. Based on a noisy measurement, the receiving modem must choose between hypothesis H_0 (the transmitter sent a 00 and hypothesis H_1 (the transmiiter sent a 1). A false alarm occurs when a 0 is sent but a 1 is detected at the receiver. A miss occurs when a 1 is sent but a 0 is detected. For both types of error, the cost is the same; one bit is detected incorrectly.

Maximum A posteriori Probability (MAP) test

- The maximum a posteriori probability test minimize P_{ERR}, the total probability of error of a binary hypothesis test. The law of total probability. Theorem 1.8, relates P_{MISS} to the a priori probabilities of H_0 and H_1 and to the two conditional error probabilities, $P_{\mathrm{FA}} = P[A_1|H_0]$ and $P_{\mathrm{MISS}} = P[A_0|H_1]$:

$$P_{\mathrm{ERR}} = P[A_1|H_0]\, P[H_0] + P[A_0|H_1]\, P[H_1]$$

- When the two types of errors have the same cost as in Example 8.5, minimizing P_{ERR} is a sensible strategy. The following theorem specifies the binary hypothesis test that produces the minimum possible P_{ERR}.

Theorem 8.1 Maximum A posteriori Probability (MAP) Binary Hypothesis Test

Given a binary hypothesis testing experiment with outcome s, the following rule leads to the lowest possible value of P_{ERR}:

$$s \in A_0 \text{ if } P\left[H_0|s\right] \geq P\left[H_1|s\right]; \qquad s \in A_1 \text{ otherwise.}$$

Proof: Theorem 8.1

To create the event space $\{A_0, A_1\}$, it is necessary to place every element $s \in S$ in either A_0 or A_1. Consider the effect of a specific value of s on the sum in Equation (8.7). Either s will contribute to the first (A_1) or second (A_0) term in the sum. By placing each s in the term that has the lower value for the specific outcome s, we create an event space that minimizes the entire sum. Thus we have the rule

$$s \in A_0 \text{ if } P\left[s|H_1\right] P\left[H_1\right] \leq P\left[s|H_0\right] P\left[H_0\right]; \qquad s \in A_1 \text{ otherwise.}$$

Applying Bayes theorem (Theorem 1.11), we see that the left side of the inequality is $P[H_1|s]P[s]$ and the right side of the inequality is $P[H_0|s]P[s]$. Therefore the inequality is identical to $P[H_0|s]P[s] \geq P[H_1|s]P[s]$, which is identical to the inequality in the theorem statement.

- Note that $P[H_0|s]$ and $P[H_1|s]$ are referred to as the a posteriori probabilities of H_0 and H_1. Just as the a priori probabilities $P[H_0]$ and $P[H_1]$ reflect out knowledge of H_0 and H_1 prior to performing an experiment, $P[H_0|s]$ and $P[H_1|s]$ reflect out knowledge after observing s.
- Theorem 8.1 states that in order to minimize P_{ERR} it is necessary to accept the hypothesis with the higher a posteriori probability. A test that follows this rule is a maximum a posteriori probability (MAP) hypothesis test.
- In such a test A_0 contains all outcomes s for which $P[H_0|s] > P[H_1|s]$, and A_1 contains all outcomes s for which $P[H_0|s] < P[H_1|s]$. If $P[H_0|s] = P[H_1|s]$, the assignment of s to either A_0 or A_1 does not affect P_{ERR}.

- In Theorem 8.1 we arbitrarily assign s to A_0 when the a posteriori probabilities are equal. We would have the same probability of error if we assign s to A_1 for all outcomes that produce equal a posteriori probabilities or if we assign some outcomes with equal a posteriori probabilities to A_0 and others to A_1.
- Equation in proof is another statement of the MAP decision rule. It contains the three probability models that are assumed to be known:
  - The a priori probability of the hypotheses: $P[H_0]$ and $P[H_1]$.
  - The likelihood function of H_0: $P[s|H_0]$.
  - The likelihood function of H_1: $P[s|H_1]$.

- When the outcomes of an experiment yield a random vector $\mathbf{X}$ as the decision statistic, we can express the MAP rule in terms of conditional PMFs and PDFs.
- If $\mathbf{X}$ is discrete, we can take $\mathbf{X} = \mathbf{x}_i$ to be the outcome of the experiment.
- If the sample space S of the experiment is continuous, we interpret the conditional probabilities by assuming that each outcome corresponds to the random vector $\mathbf{X}$ in the small volume $\mathbf{x} \leq \mathbf{X} < \mathbf{x} + d\mathbf{x}$ with probability $f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$. Section 4.9 demonstrates that the conditional probabilities are ratios of probability densities. Thus in terms of the random variable X, we have the following version of the MAP hypothesis test.

Theorem 8.2

For an experiment that produces a random vector $\mathbf{X}$, the MAP hypothesis test is

Discrete: $\mathbf{x} \in A_0$ if $\dfrac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]}{P[H_0]}$; $\mathbf{x} \in A_1$ otherwise

Continuous: $\mathbf{x} \in A_0$ if $\dfrac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]}{P[H_0]}$; $\mathbf{x} \in A_1$ otherwise.

Likelihood ratio (based on experiments)

Ratio of prior probability (prior to experiment)

- In these formulas, the ratio of conditional probabilities is referred to as a **likelihood ratio**. The formulas state that in order to perform a binary hypothesis test, we observe the outcome of an experiment, calculate the likelihood ratio on the left side of the formula, and compare it with a constant on the right side of the formula.
- We can view the likelihood ratio as the evidence, based on an observation, in favor of H_0. If the likelihood ratio is grater than 1, H_0 is more likely than H_1. The ratio of prior probabilities, on the right side, is the evidence, prior to performing the experiment, in favor of H_1. Therefore, Theorem 8.2 states that H_0 is the better conclusion if the evidence in favor of H_0, based on the **experiment**, outweighs the prior evidence in favor of H_1.
- In many practical hypothesis tests, it is convenient to compare the logarithms of the two ratios.

Example 8.6 Problem

With probability p, a digital communications system transmits a 0. It transmits a 1 with probability $1-p$. The received signal is either $X = -v + N$ volts, if the transmitted bit is 0; or $v + N$ volts, if the transmitted bit is 1. The voltage $\pm v$ is the information component of the received signal, and N, a Gaussian $(0, \sigma)$ random variable, is the noise component. Given the received signal X, what is the minimum probability of error rule for deciding whether 0 or 1 was sent?

Example 8.6 Solution

With 0 transmitted, X is the Gaussian $(-v, \sigma)$ random variable. With 1 transmitted, X is the Gaussian (v, σ) random variable. With H_i denoting the hypothesis that bit i was sent, the likelihood functions are

$$f_{X|H_0}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x+v)^2/2\sigma^2}, \qquad f_{X|H_1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-v)^2/2\sigma^2}.$$

Since $P[H_0] = p$, the likelihood ratio test of Theorem 8.2 becomes

$$x \in A_0 \text{ if } \frac{e^{-(x+v)^2/2\sigma^2}}{e^{-(x-v)^2/2\sigma^2}} \geq \frac{1-p}{p}; \qquad x \in A_1 \text{ otherwise.}$$

Taking the logarithm of both sides and simplifying yields

$$x \in A_0 \text{ if } x \leq x^* = \frac{\sigma^2}{2v} \ln\left(\frac{p}{1-p}\right); \qquad x \in A_1 \text{ otherwise.}$$

[Continued]

Example 8.6 Solution (continued)

When $p = 1/2$, the threshold $x^* = 0$ and the conclusion depends only on whether the evidence in the received signal favors 0 or 1, as indicated by the sign of x. When $p \neq 1/2$, the prior information shifts the decision threshold x^*. The shift favors 1 ($x^* < 0$) if $p < 1/2$. The shift favors 0 ($x^* > 0$) if $p > 1/2$. The influence of the prior information also depends on the signal-to-noise voltage ratio, $2v/\sigma$. When the ratio is relatively high, the information in the received signal is reliable and the received signal has relatively more influence than the prior information (x^* closer to 0). When $2v/\sigma$ is relatively low, the prior information has relatively more influence.

In Figure 8.3, the threshold x^* is the value of x for which the two likelihood functions, each multiplied by a prior probability, are equal. The probability of error is the sum of the shaded areas. Compared to all other decision rules, the threshold x^* produces the minimum possible P_{ERR}.

Example 8.7 Problem

Find the error probability of the communications system of Example 8.6.

Example 8.7 Solution

Applying Equation (8.7), we can write the probability of an error as

$$P_{\text{ERR}} = pP\left[X > x^*|H_0\right] + (1-p)P\left[X < x^*|H_1\right].$$

Given H_0, X is Gaussian $(-v, \sigma)$. Given H_1, X is Gaussian (v, σ). Consequently,

$$\begin{aligned} P_{\text{ERR}} &= pQ\left(\frac{x^*+v}{\sigma}\right) + (1-p)\Phi\left(\frac{x^*-v}{\sigma}\right) \\ &= pQ\left(\frac{\sigma}{2v}\ln\frac{p}{1-p} + \frac{v}{\sigma}\right) + (1-p)\Phi\left(\frac{\sigma}{2v}\ln\frac{p}{1-p} - \frac{v}{\sigma}\right). \end{aligned}$$

This equation shows how the prior information, represented by $\ln[(1-p)/p]$, and the power of the noise in the received signal, represented by σ, influence P_{ERR}.

Example 8.8 Problem

At a computer disk drive factory, the manufacturing failure rate is the probability that a randomly chosen new drive fails the first time it is powered up. Normally the production of drives is very reliable, with a failure rate $q_0 = 10^{-4}$. However, from time to time there is a production problem that causes the failure rate to jump to $q_1 = 10^{-1}$. Let H_i denote the hypothesis that the failure rate is q_i.

Every morning, an inspector chooses drives at random from the previous day's production and tests them. If a failure occurs too soon, the company stops production and checks the critical part of the process. Production problems occur at random once every ten days, so that $P[H_1] = 0.1 = 1 - P[H_0]$. Based on N, the number of drives tested up to and including the first failure, design a MAP hypothesis test. Calculate the conditional error probabilities P_{FA} and P_{MISS} and the total error probability P_{ERR}.

Example 8.8 Solution

Given a failure rate of q_i, N is a geometric random variable (see Example 2.11) with expected value $1/q_i$. That is, $P_{N|H_i}(n) = q_i(1-q_i)^{n-1}$ for $n = 1, 2, \ldots$ and $P_{N|H_i}(n) = 0$ otherwise. Therefore, by Theorem 8.2, the MAP design states

$$n \in A_0 \text{ if } \frac{P_{N|H_0}(n)}{P_{N|H_1}(n)} \geq \frac{P[H_1]}{P[H_0]}; \qquad n \in A_1 \text{ otherwise}$$

With some algebra, we find that the MAP design is:

$$n \in A_0 \text{ if } n \geq n^* = 1 + \frac{\ln\left(\frac{q_1 P[H_1]}{q_0 P[H_0]}\right)}{\ln\left(\frac{1-q_0}{1-q_1}\right)}; \qquad n \in A_1 \text{ otherwise.}$$

Substituting $q_0 = 10^{-4}$, $q_1 = 10^{-1}$, $P[H_0] = 0.9$, and $P[H_1] = 0.1$, we obtain $n^* = 45.8$. Therefore, in the MAP hypothesis test, $A_0 = \{n \geq 46\}$.

[Continued]

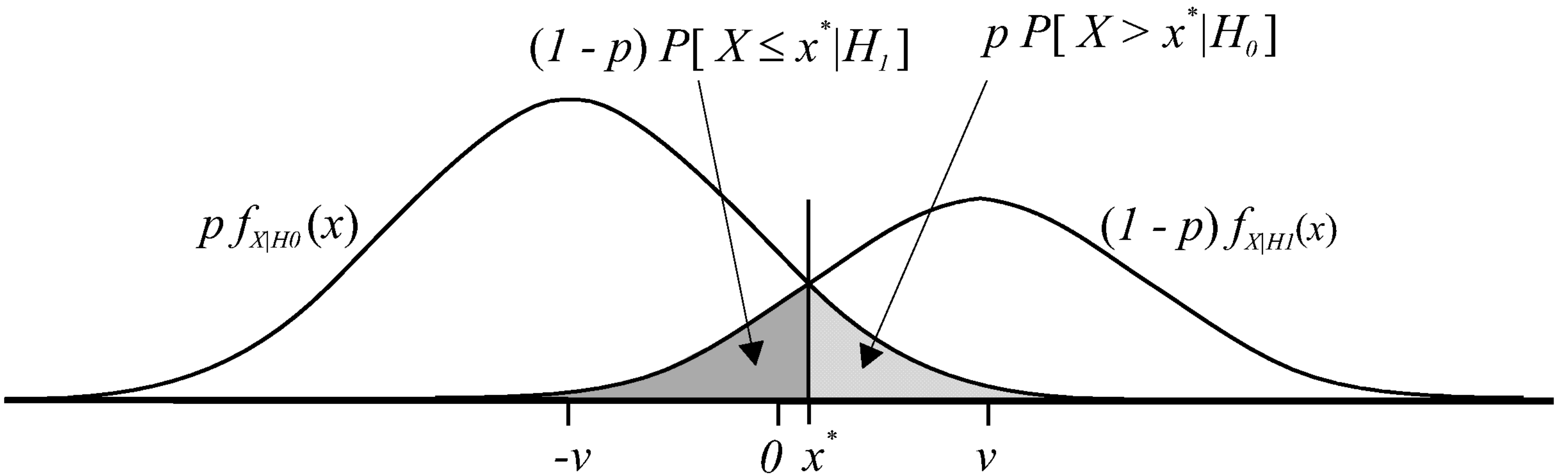
Example 8.8 Solution (continued)

This implies that the inspector tests at most 45 drives in order to reach a conclusion about the failure rate. If the first failure occurs before test 46, the company assumes that the failure rate is 10^{-2}. If the first 45 drives pass the test, then $N \geq 46$ and the company assumes that the failure rate is 10^{-4}. The error probabilities are:

$$P_{\text{FA}} = P\left[N \leq 45|H_0\right] = F_{N|H_0}(45) = 1 - (1 - 10^{-4})^{45} = 0.0045,$$
$$P_{\text{MISS}} = P\left[N > 45|H_1\right] = 1 - F_{N|H_1}(45) = (1 - 10^{-1})^{45} = 0.0087.$$

The total probability of error is $P_{\text{ERR}} = P[H_0]P_{\text{FA}} + P[H_1]P_{\text{MISS}} = 0.0049$.

Figure 8.3



Decision regions for Example 8.6.

Minimum Cost Test

- The MAP test implicitly assume that both types of errors (miss and false alarm) are equally serious. As discussed in connection with the radar application earlier in this section, this is not the case in many important situations. Consider application in which $C = C_{10}$ units is the **cost** of a false alarm (decide H_1 when H_0 is correct) and $C = C_{01}$ units is the cost of a miss (decide H_0 when H_1 is correct). In this situation the expected cost of test errors is

$$E[C] = P[A_1|H_0]\ P[H_0]\ C_{10} + P[A_0|H_1]\ P[H_1]\ C_{01}$$

- Minimizing $E[C]$ is the goal of the minim cost hypothesis test. When the decision statistic is a random vector $\mathbf{X}$, we have the following theorem.

Theorem 8.3 Minimum Cost Binary Hypothesis Test

For an experiment that produces a random vector $\mathbf{X}$, the minimum cost hypothesis test is

Discrete: $\mathbf{x} \in A_0$ if $\dfrac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]C_{01}}{P[H_0]C_{10}}$; $\mathbf{x} \in A_1$ otherwise

Continuous: $\mathbf{x} \in A_0$ if $\dfrac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]C_{01}}{P[H_0]C_{10}}$; $\mathbf{x} \in A_1$ otherwise.

Proof: Theorem 8.3

The function to be minimized, Equation (8.19), is identical to the function to be minimized in the MAP hypothesis test, Equation (8.7), except that $P[H_1]C_{01}$ appears in place of $P[H_1]$ and $P[H_0]C_{10}$ appears in place of $P[H_0]$. Thus the optimum hypothesis test is the test in Theorem 8.2 with $P[H_1]C_{01}$ replacing $P[H_1]$ and $P[H_0]C_{10}$ replacing $P[H_0]$.

- In this test we note that only the relative cost C_{01}/C_{10} influences that test, not the individual costs or the units in which cost is measured.
- A ratio > 1 implies that misses are more costly than false alarms. Therefore, a ratio > 1 expands A_1, the acceptance set for H_1, make it harder to miss H_1 when it is correct.
- On the other hand, the same ratio contacts H_0 and increases the false alarm probability, because a false alarm is less costly than a miss.

Example 8.9 Problem

Continuing the disk drive test of Example 8.8, the factory produces 1,000 disk drives per hour and 10,000 disk drives per day. The manufacturer sells each drive for \$100. However, each defective drive is returned to the factory and replaced by a new drive. The cost of replacing a drive is \$200, consisting of \$100 for the replacement drive and an additional \$100 for shipping, customer support, and claims processing. Further note that remedying a production problem results in 30 minutes of lost production. Based on the decision statistic N, the number of drives tested up to and including the first failure, what is the minimum cost test?

Example 8.9 Solution

Based on the given facts, the cost C_{10} of a false alarm is 30 minutes (5,000 drives) of lost production, or roughly \$50,000. On the other hand, the cost C_{01} of a miss is that 10% of the daily production will be returned for replacement. For 1,000 drives returned at \$200 per drive, The expected cost is 200,000 dollars. The minimum cost test is

$$n \in A_0 \text{ if } \frac{P_{N|H_0}(n)}{P_{N|H_1}(n)} \geq \frac{P[H_1]\,C_{01}}{P[H_0]\,C_{10}}; \qquad n \in A_1 \text{ otherwise.}$$

Performing the same substitutions and simplifications as in Example 8.8 yields

$$n \in A_0 \text{ if } n \geq n^* = 1 + \frac{\ln\left(\frac{q_1 P[H_1] C_{01}}{q_0 P[H_0] C_{10}}\right)}{\ln\left(\frac{1-q_0}{1-q_1}\right)} = 58.92; \qquad n \in A_1 \text{ otherwise.}$$

Therefore, in the minimum cost hypothesis test, $A_0 = \{n \geq 59\}$. An inspector tests at most 58 disk drives to reach a conclusion regarding the state of the factory. If 58 drives pass the test, then $N \geq 59$, and the failure rate is assumed to be 10^{-4}. The error probabilities are:

$$P_{\text{FA}} = P[N \leq 58|H_0] = F_{N|H_0}(58) = 1 - (1 - 10^{-4})^{58} = 0.0058,$$
$$P_{\text{MISS}} = P[N \geq 59|H_1] = 1 - F_{N|H_1}(58) = (1 - 10^{-1})^{58} = 0.0022.$$

[Continued]

Example 8.9 Solution (continued)

The average cost (in dollars) of this rule is

$$\begin{aligned} E[C] &= P[H_0]\, P_{\mathsf{FA}} C_{10} + P[H_1]\, P_{\mathsf{MISS}} C_{01} \\ &= (0.9)(0.0058)(50{,}000) + (0.1)(0.0022)(200{,}000) = 305. \end{aligned}$$

By comparison, the MAP test, which minimizes the probability of an error, rather than the expected cost, has an expected cost

$$E[C_{\mathsf{MAP}}] = (0.9)(0.0046)(50{,}000) + (0.1)(0.0079)(200{,}000) = 365.$$

A savings of \$60 may not seem very large. The reason is that both the MAP test and the minimum cost test work very well. By comparison, for a “no test” policy that skips testing altogether, each day that the failure rate is $q_1 = 0.1$ will result, on average, in 1,000 returned drives at an expected cost of \$200,000. Since such days will occur with probability $P[H_1] = 0.1$, the expected cost of a “no test” policy is \$20,000 per day.

Neyman-Pearson Test

- Given an observation, the MAP test minimizes the probability of accepting the wrong hypothesis and minimum cost test minimizes the cost of errors.
- However, the MAP test requires that we know the a priori probabilities $P[H_i]$ of the competing hypotheses, and the minimum cost test required that we know in addition the relative costs of the two types of errors.
- In many situations, these costs and a priori probabilities are difficult or even impossible to specify. In this case an alternate approach would be specify a **tolerable level** for either the false alarm or miss probability. This idea is the basis for the **Neyman-Pearson test**.

- The Neyman-Pearson test minimizes P_{MISS} subject to the false alarm probability constraint $P_{\mathrm{FA}} = \alpha$, where α is a constant that indicates out tolerance of false alarm Because $P_{\mathrm{FA}} = P[A_1|H_0]$ and $P_{\mathrm{MISS}} = P[A_0|H_1]$ are conditional probabilities, the test does not require the a priori probabilities $P[H_0]$ and $P[H_1]$. We first describe the Neyman-Pearson test when the decision statistic is a continuous random vector $\mathbf{X}$.

Theorem 8.4 Neyman-Pearson Binary Hypothesis Test

Based on the decision statistic $\mathbf{X}$, a continuous random vector, the decision rule that minimizes P_{MISS}, subject to the constraint $P_{\text{FA}} = \alpha$, is

$$\mathbf{x} \in A_0 \text{ if } L(\mathbf{x}) = \frac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \gamma; \qquad \mathbf{x} \in A_1 \text{ otherwise,}$$

where γ is chosen so that $\int_{L(\mathbf{x})<\gamma} f_{\mathbf{X}|H_0}(\mathbf{x})\, d\mathbf{x} = \alpha$.

Proof: Theorem 8.4

Using the Lagrange multiplier method, we define the Lagrange multiplier λ and the function

$$
\begin{aligned}
G &= P_{\text{MISS}} + \lambda(P_{\text{FA}} - \alpha) \\
&= \int_{A_0} f_{\mathbf{X}|H_1}(\mathbf{x})\, d\mathbf{x} + \lambda \left(1 - \int_{A_0} f_{\mathbf{X}|H_0}(\mathbf{x})\, d\mathbf{x} - \alpha\right) \\
&= \int_{A_0} \left(f_{\mathbf{X}|H_1}(\mathbf{x}) - \lambda f_{\mathbf{X}|H_0}(\mathbf{x})\right) d\mathbf{x} + \lambda(1-\alpha) \qquad (8.29)
\end{aligned}
$$

For a given λ and α, we see that G is minimized if A_0 includes all $\mathbf{x}$ satisfying

$$
f_{\mathbf{X}|H_1}(\mathbf{x}) - \lambda f_{\mathbf{X}|H_0}(\mathbf{x}) \leq 0. \qquad (8.30)
$$

Note that λ is found from the constraint $P_{\text{FA}} = \alpha$. Moreover, we observe that Equation (8.29) implies $\lambda > 0$; otherwise, $f_{\mathbf{X}|H_0}(\mathbf{x}) - \lambda f_{\mathbf{X}|H_1}(\mathbf{x}) > 0$ for all $\mathbf{x}$ and $A_0 = \phi$, the empty set, would minimize G. In this case, $P_{\text{FA}} = 1$, which would violate the constraint that $P_{\text{FA}} = \alpha$. Since $\lambda > 0$, we can rewrite the inequality (8.30) as $L(\mathbf{x}) \geq 1/\lambda = \gamma$.

- In the radar system of Example 8.4, the decision statistic was a random variable $\mathbf{X}$ and the receiver operating curves (ROCs) of figure 8.2 were generated by adjusting a threshold x_0 that specified the sets $A_0 = \{X \le x_0\}$ and $A_1 = \{X > x_0\}$. Example 8.4 did not question whether this rule finds the best ROC. For each specified value of $P_{\mathrm{FA}} = \alpha$, the Neyman-Pearson test identifies the decision rule that minimizes P_{MISS} .
- In the Neyman-Pearson test, an increase in γ decreases P_{MISS} but increase P_{FA}. When the decision statistic $\mathbf{X}$ is a continuous random vector, we can choose γ so that false alarm probability is exactly α. This may not be possible when $\mathbf{X}$ is discrete, In discrete case, we have the following version of the Neyman-Pearson test.

Theorem 8.5 Discrete Neyman-Pearson Test

Based on the decision statistic $\mathbf{X}$, a decision random vector, the decision rule that minimizes P_{MISS}, subject to the constraint $P_{\text{FA}} \leq \alpha$, is

$$\mathbf{x} \in A_0 \text{ if } L(\mathbf{x}) = \frac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq \gamma; \qquad \mathbf{x} \in A_1 \text{ otherwise,}$$

where γ is the largest possible value such that $\sum_{L(\mathbf{x})<\gamma} P_{\mathbf{X}|H_0}(\mathbf{x})\, d\mathbf{x} \leq \alpha$.

Example 8.10 Problem

Continuing the disk drive factory test of Example 8.8, design a Neyman-Pearson test such that the false alarm probability satisfies $P_{\text{FA}} \leq \alpha = 0.01$. Calculate the resulting miss and false alarm probabilities.

Example 8.10 Solution

The Neyman-Pearson test is

$$n \in A_0 \text{ if } L(n) = \frac{P_{N|H_0}(n)}{P_{N|H_1}(n)} \geq \gamma; \qquad n \in A_1 \text{ otherwise.}$$

We see from Equation (8.15) that this is the same as the MAP test with $P[H_1]/P[H_0]$ replaced by γ. Thus, just like the MAP test, the Neyman-Pearson test must be a threshold test of the form

$$n \in A_0 \text{ if } n \geq n^*; \qquad n \in A_1 \text{ otherwise.}$$

Some algebra would allow us to find the threshold n^* in terms of the parameter γ. However, this is unnecessary. It is simpler to choose n^* directly so that the test meets the false alarm probability constraint

$$P_{\text{FA}} = P\left[N \leq n^* - 1|H_0\right] = F_{N|H_0}\left(n^* - 1\right) = 1 - (1 - q_0)^{n^*-1} \leq \alpha.$$

This implies

$$n^* \leq 1 + \frac{\ln(1-\alpha)}{\ln(1-q_0)} = 1 + \frac{\ln(0.99)}{\ln(0.9)} = 101.49.$$

Thus, we can choose $n^* = 101$ and still meet the false alarm probability constraint.

[Continued]

Example 8.10 Solution (continued)

The error probabilities are:

$$P_{\text{FA}} = P\left[N \leq 100|H_0\right] = 1 - (1 - 10^{-4})^{100} = 0.00995,$$

$$P_{\text{MISS}} = P\left[N \geq 101|H_1\right] = (1 - 10^{-1})^{100} = 2.66 \cdot 10^{-5}.$$

We see that a one percent false alarm probability yields a dramatic reduction in the probability of a miss. Although the Neyman-Pearson test minimizes neither the overall probability of a test error nor the expected cost $E[C]$, it may be preferable to either the MAP test or the minimum cost test. In particular, customers will judge the quality of the disk drives and the reputation of the factory based on the number of defective drives that are shipped. Compared to the other tests, the Neyman-Pearson test results in a much lower miss probability and far fewer defective drives being shipped.

Maximum Likelihood Test

- Similar to the Neyman-Pearson test, the **maximum likelihood** (**ML**) **test** is another method that avoids the need for a priori probabilities.
- Under the ML approach, we treat the hypothesis as some sort of "unknown" and choose a hypothesis H_i for which $P[s|H_i]$, the conditional probability of the outcome s given the hypothesis H_i is largest. The idea behind choosing hypothesis to maximize the probability of the observation is to avoid making assumptions about the a priori probabilities $P[H_i]$. The resulting decision rule, called maximum likelihood (ML) rule, can be written mathematically as:

Definition 8.1 Rule

For a binary hypothesis test based on the experimental outcome $s \in S$, *the maximum likelihood (ML) decision rule is*

$$s \in A_0 \text{ if } P\left[s|H_0\right] \geq P\left[s|H_1\right]; \qquad s \in A_1 \text{ otherwise.}$$

$$\text{MAP}: s \in A_0 \text{ if } P\left[H_0 \mid s\right] \geq P\left[H_1 \mid s\right]; \; s \in A_1 \text{ otherwise.}$$

- Comparison Theorem 8.1 and Definition 8.1, we see that in the absence of information about the a priori probabilities $P[H_i]$, we have adopted a maximum likelihood decision rule that is the same as the MAP rule under the assumption that hypotheses H_0 and H_1 occur with equal probability. In essence, in the absence of a priori information, the ML rule assumes that all hypotheses are equally likely. By comparing the likelihood ratio to a threshold equal to 1, the ML hypothesis test is neutral about whether H_0 has a higher probability than H_1 or vice versa.
- When the decision statistic of the experiment is a random vector $\mathbf{X}$, we can express the ML rule in terms of conditional PMFs or PDFs, just as we did for the MAP rule.

Theorem 8.6

If an experiment produces a random vector $\mathbf{X}$, the ML decision rule states

Discrete: $\mathbf{x} \in A_0$ if $\dfrac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq 1;$ $\mathbf{x} \in A_1$ otherwise

Continuous: $\mathbf{x} \in A_0$ if $\dfrac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq 1;$ $\mathbf{x} \in A_1$ otherwise.

- Comparing Theorem 8.6 to Theorem 8.4, when $\mathbf{X}$ is continuous, or Theorem 8.5, when $\mathbf{X}$ is discrete, we see that the maximum likelihood test is the same as the Neyman-Pearson test with parameter $\gamma = 1$. This guarantees that the maximum likelihood test is optimal in the limited sense that no other test can reduce P_{MISS} for the same P_{FA}.
- In practice, we use a ML hypothesis test in many applications. It is almost as effective as the MAP hypothesis test when the experiment that produces outcomes s is reliable in the sense that P_{ERR} for the ML test is low. To see why this is true, examine the decision rule in Example 8.6. When the signal-to-noise ratio $2v/\sigma$ is high, the threshold (of the log-likelihood ratio) is close to 0, which means that the result of the MAP hypothesis test is close to the result of a ML hypothesis test, regardless of the prior probability p.

Example 8.11 Problem

Continuing the disk drive test of Example 8.8, design the maximum likelihood test for the factory state based on the decision statistic N, the number of drives tested up to and including the first failure.

Example 8.11 Solution

The ML hypothesis test corresponds to the MAP test with $P[H_0] = P[H_1] = 0.5$. In ths icase, Equation (8.16) implies $n^* = 66.62$ or $A_0 = \{n \geq 67\}$. The conditional error probabilities under the ML rule are

$$P_{\text{FA}} = P[N \leq 66|H_0] = 1 - (1 - 10^{-4})^{66} = 0.0066,$$

$$P_{\text{MISS}} = P[N \geq 67|H_1] = (1 - 10^{-1})^{66} = 9.55 \cdot 10^{-4}.$$

For the ML test, $P_{\text{ERR}} = 0.0060$. Comparing the MAP rule with the ML rule, we see that the prior information used in the MAP rule makes it more difficult to reject the null hypothesis. We need only 46 good drives in the MAP test to accept H_0, while in the ML test, the first 66 drives have to pass. The ML design, which does not take into account the fact that the failure rate is usually low, is more susceptible to false alarms than the MAP test. Even though the error probability is higher for the ML test, it might be a good idea to use this test in the drive company because the miss probability is very low. The consequence of a false alarm is likely to be an examination of the manufacturing process to find out if something is wrong. A miss, on the other hand (deciding the failure rate is 10^{-4} when it is really 10^{-1}), would cause the company to ship an excessive number of defective drives.

Quiz 8.2

In an optical communications system, the photodetector output is a Poisson random variable K either with an expected value of 10,000 photons (hypothesis H_0) or with an expected value of 1,000,000 photons (hypothesis H_1). Given that both hypotheses are equally likely, design a MAP hypothesis test using observed values of random variable K.

Quiz 8.2 Solution

From the problem statement, the conditional PMFs of K are

$$P_{K|H_0}(k) = \begin{cases} \frac{10^{4k} e^{-10^4}}{k!} & k = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

$$P_{K|H_1}(k) = \begin{cases} \frac{10^{6k} e^{-10^6}}{k!} & k = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Since the two hypotheses are equally likely, the MAP and ML tests are the same. From Theorem 8.6, the ML hypothesis rule is

$$k \in A_0 \text{ if } P_{K|H_0}(k) \geq P_{K|H_1}(k); \qquad k \in A_1 \text{ otherwise.}$$

This rule simplifies to

$$k \in A_0 \text{ if } k \leq k^* = \frac{10^6 - 10^4}{\ln 100} = 214,975.7; \qquad k \in A_1 \text{ otherwise.}$$

Thus if we observe at least $214,976$ photons, then we accept hypothesis H_1.

8.3 Multiple Hypothesis Test

- There are many application in which an experiment can conform to more than two known probability models, all with the same sample space S. A multiple hypothesis test is a generalization of a binary hypothesis test.
- There are M hypothetical probability models: $H_0, H_1, \ldots, H_{M-1}$. We perform an experiment and based on the outcome, we come to the conclusion that a certain H_m is the true probability model.
- The design of the experiment consists of dividing S into an event space consisting of mutually exclusive, collectively exhaustive sets, $A_0, A_1, \ldots, A_{M-1}$, such tat the conclusion is accept H_i if $s \in A_i$, The accuracy measure of the experiment consists of M^2 conditional probabilities, $P[A_i|H_j]$, $i, j = 0, 1, 2, \ldots, M-1$. The M probabilities, $P[A_i|H_i]$, $i = 0, 1, 2, \ldots, M-1$ are probabilities of correct decisions. The remaining probabilities are error probabilities.

Example 8.12

A computer modem is capable of transmitting 16 different signals. Each signal represents a sequence of four bits in the digital bit stream at the input to the modem. The modem receiver examines the received signal and produces four bits in the bit stream at the output of the modem. The design of the modem considers the task of the receiver to be a test of 16 hypotheses $H_0, H_1, \ldots, H_{15}$, where H_0 represents 0000, H_1 represents 0001, $\cdots$ and H_{15} represents 1111. The sample space of the experiment is an ensemble of possible received signals. The test design places each outcome s in a set A_i such that the event $s \in A_i$ leads to the output of the four-bit sequence corresponding to H_i.

- For a multiple hypothesis test, the MAP hypothesis test and the ML hypothesis test are generalizations of the tests in Theorem 8.1 and Definition 8.1. Minimizing the probability of error corresponds to maximizing the probability of a correct decision,

$$P_{\text{CORRECT}} = \sum_{i=0}^{M-1} P[A_i \mid H_i] P[H_i]$$

Theorem 8.7 MAP Multiple Hypothesis Test

maximum a posteriori probabilityGiven a multiple hypothesis testing experiment with outcome s, the following rule leads to the highest possible value of P_{CORRECT}:

$$s \in A_m \text{ if } P\left[H_m|s\right] \geq P\left[H_j|s\right] \text{ for all } j = 0, 1, 2, \ldots, M-1.$$

- As in binary hypothesis testing, we can apply Bayes’ theorem to derive a decision rule based on the probability models (likelihood functions) corresponding to the hypotheses and the a priori probabilities of the hypotheses.
- Therefore, corresponding to Theorem 8.2 we have the following generalization of the MAP binary hypothesis test.

Theorem 8.8

For an experiment that produces a random variable X, the MAP multiple hypothesis test is

Discrete: $x_i \in A_m$ if $P[H_m]P_{X|H_m}(x_i) \geq P[H_j]P_{X|H_j}(x_i)$ for all j

Continuous: $x \in A_m$ if $P[H_m]f_{X|H_m}(x) \geq P[H_j]f_{X|H_j}(x)$ for all j.

Maximum Likelihood (ML)

Definition 8.2 Multiple Hypothesis Test

A maximum likelihood test of multiple hypotheses has the decision rule

$$s \in A_m \text{ if } P[s|H_m] \geq P\left[s|H_j\right] \text{ for all } j.$$

- If information about the a priori probabilities of the hypotheses is not available, a maximum likelihood hypothesis test is appropriate.
- The ML hypothesis test corresponds to the MAP hypothesis test when all hypotheses H_i have equal probability.

Example 8.13 Problem

(QPSK)

In a quaternary phase shift keying () communications system, the transmitter sends one of four equally likely symbols $\{s_0, s_1, s_2, s_3\}$. Let H_i denote the hypothesis that the transmitted signal was s_i. When s_i is transmitted, a QPSK receiver produces the vector $\mathbf{X} = \begin{bmatrix} X_1 & X_2 \end{bmatrix}'$ such that

$$X_1 = \sqrt{E}\cos(i\pi/2 + \pi/4) + N_1, \qquad X_2 = \sqrt{E}\sin(i\pi/2 + \pi/4) + N_2,$$

where N_1 and N_2 are iid Gaussian $(0, \sigma)$ random variables that characterize the receiver noise and E is the average energy per symbol. Based on the receiver output $\mathbf{X}$, the receiver must decide which symbol was transmitted. Design a hypothesis test that maximizes the probability of correctly deciding which symbol was sent.

Example 8.13 Solution

Since the four hypotheses are equally likely, both the MAP and ML tests maximize the probability of a correct decision. To derive the ML hypothesis test, we need to calculate the conditional joint PDFs $f_{\mathbf{X}|H_i}(\mathbf{x})$. Given H_i, N_1 and N_2 are independent and thus X_1 and X_2 are independent. That is, using $\theta_i = i\pi/2 + \pi/4$, we can write

$$
\begin{aligned}
f_{\mathbf{X}|H_i}(\mathbf{x}) &= f_{X_1|H_i}(x_1)\, f_{X_2|H_i}(x_2) \\
&= \frac{1}{2\pi\sigma^2} e^{-(x_1-\sqrt{E}\cos\theta_i)^2/2\sigma^2} e^{-(x_2-\sqrt{E}\sin\theta_i)^2/2\sigma^2} \\
&= \frac{1}{2\pi\sigma^2} e^{-[(x_1-\sqrt{E}\cos\theta_i)^2+(x_2-\sqrt{E}\sin\theta_i)^2]/2\sigma^2}.
\end{aligned}
$$

We must assign each possible outcome $\mathbf{x}$ to an acceptance set A_i. From Definition 8.2, the acceptance sets A_i for the ML multiple hypothesis test must satisfy

$$\mathbf{x} \in A_i \text{ if } f_{\mathbf{X}|H_i}(\mathbf{x}) \geq f_{\mathbf{X}|H_j}(\mathbf{x}) \text{ for all } j.$$

[Continued]

Example 8.13 Solution (continued)

Equivalently, the ML acceptance sets are given by the rule that $\mathbf{x} \in A_i$ if for all j,

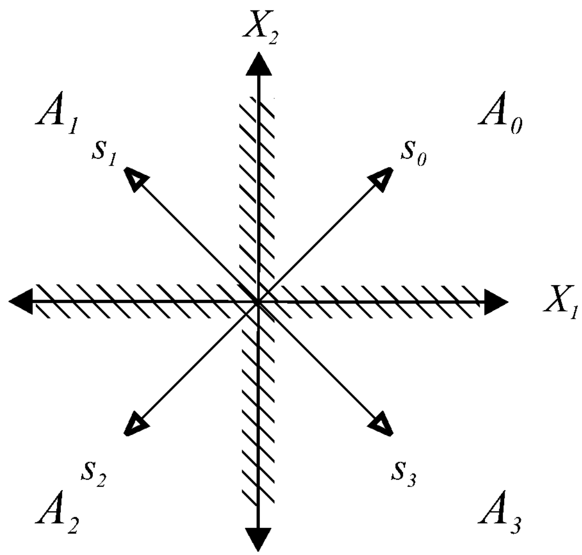
$$(x_1 - \sqrt{E}\cos\theta_i)^2 + (x_2 - \sqrt{E}\sin\theta_i)^2 \le (x_1 - \sqrt{E}\cos\theta_j)^2 + (x_2 - \sqrt{E}\sin\theta_j)^2.$$

Defining the signal vectors $\mathbf{s}_i = \left[\sqrt{E}\cos\theta_i \quad \sqrt{E}\sin\theta_i\right]'$, we can write the ML rule as

$$\mathbf{x} \in A_i \text{ if } \|\mathbf{x} - \mathbf{s}_i\|^2 \le \left\|\mathbf{x} - \mathbf{s}_j\right\|^2$$

where $\|\mathbf{u}\|^2 = u_1^2 + u_2^2$ denotes the square of the Euclidean length of two-dimensional vector $\mathbf{u}$. In short, the acceptance set A_i is the set of all vectors $\mathbf{x}$ that are closest to the vector $\mathbf{s}_i$. These acceptance sets are shown in Figure 8.4. In communications textbooks, the space of vectors $\mathbf{x}$ is called the *signal space*, the set of vectors $\{\mathbf{s}_1, \ldots, \mathbf{s}_4\}$ is called the *signal constellation*, and the acceptance sets A_i are called *decision regions*.

Figure 8.4



For the receiver of Example 8.13, the four quadrants (with boundaries marked by shaded bars) are the four acceptance sets $\{A_0, A_1, A_2, A_3\}$.

Quiz 8.3

For the communications system of Example 8.13, what is the probability that the receiver makes an error and decodes the wrong symbol?

Quiz 8.3 Solution

For the QPSK system, a symbol error occurs when s_i is transmitted but $(X_1, X_2) \in A_j$ for some $j \neq i$. For a QPSK system, it is easier to calculate the probability of a correct decision. Given H_0, the conditional probability of a correct decision is

$$P[C|H_0] = P[X_1 > 0, X_2 > 0|H_0] = P\left[\sqrt{E/2} + N_1 > 0, \sqrt{E/2} + N_2 > 0\right]$$

Because of the symmetry of the signals, $P[C|H_0] = P[C|H_i]$ for all i. This implies the probability of a correct decision is $P[C] = P[C|H_0]$. Since N_1 and N_2 are iid Gaussian $(0, \sigma)$ random variables, we have

$$\begin{aligned} P[C] = P[C|H_0] &= P\left[\sqrt{E/2} + N_1 > 0\right] P\left[\sqrt{E/2} + N_2 > 0\right] \\ &= \left(P\left[N_1 > -\sqrt{E/2}\right]\right)^2 \\ &= \left[1 - \Phi\left(\frac{-\sqrt{E/2}}{\sigma}\right)\right]^2 \end{aligned}$$

Since $\Phi(-x) = 1 - \Phi(x)$, we have $P[C] = \Phi^2(\sqrt{E/2\sigma^2})$. Equivalently, the probability of error is

$$P_{\text{ERR}} = 1 - P[C] = 1 - \Phi^2\left(\sqrt{\frac{E}{2\sigma^2}}\right)$$