Systems and Signals 344

2025

Lecture 26

Hypothesis Testing

Selections from Chapter 11

Hypothesis Testing

- 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 make a decision.
    - When the decision is based on the properties of random variables, the reasoning is referred to as **statistical inference** (from previous lectures).
    - Now we introduce **hypothesis testing**, another type of inference.

Probability vs Likelihood

Probability and **Likelihood** are related concepts but they don't refer to the same thing.

Probability is the area under the PDF, which represents the probability model for a specific question related to the experiment.

$$P(event|model)$$

Likelihood is a measure of the extent to which a sample provides support for a specific probability model.

$$L(model|samples)$$

In short: Probability quantifies anticipation of outcomes. Likelihood quantifies trust in the model.

Hypothesis Testing

Similar to probability theory, a statistical inference method consists of three steps: Perform an experiment, observe an outcome, state a decision. The aim of the assignment is to achieve the highest possible accuracy.

- **Decision:** The observations result from one of M hypothetical probability models: $H_0, H_1, \ldots, H_{M-1}$.
- **Accuracy Measure:** Probability that the decision is H_i when the true model is H_j for $i, j = 0, 1, \ldots, M-1$.

The assignment of decisions to outcomes is based on probability theory. The accuracy measure associates a cost with each type of wrong decision.

Hypothesis Testing

Example 11.1

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.

Decision Assuming λ is a constant, does λ equal $\lambda_0 = 2.5$, $\lambda_1 = 3.5$, or $\lambda_2 = 4.5$?

Accuracy Measure Probability that the decision is $\lambda = \lambda_i$ when the true model is λ_j for $i, j = 0, 1, 2$.

Binary Hypothesis Testing

- In a **binary hypothesis test**, there are two hypothetical probability models, H_0 and H_1, and two possible decisions:
    - accept H_0 as the true model, or accept H_1.
    - H_0 is often called the **null hypothesis** and H_1 is called the **alternative hypothesis**.
    - The null hypothesis states the "status quo". This hypothesis is *assumed to be true* until there is evidence to suggest otherwise.
- There is also a probability model for H_0 and H_1, conveyed by the numbers $P(H_0)$ and $P(H_1) = 1 - P(H_0)$ (if mutually exclusive)
- These numbers are referred to as the a **priori probabilities**.
    - They reflect the state of knowledge about the probability model before an outcome is observed.

Binary Hypothesis Testing

- The complete experiment for a binary hypothesis test consists of two subexperiments.
  - 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 sample space, S.
  - 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**.
    - This produces $P_{\mathbf{X}|H_0}(\mathbf{x})$ and $P_{\mathbf{X}|H_1}(\mathbf{x})$ (*discrete*) or $f_{\mathbf{X}|H_0}(\mathbf{x})$ and $f_{\mathbf{X}|H_1}(\mathbf{x})$ (*continuous*)
  - $\mathbf{X}$ is the evidence that helps us make a choice between the hypotheses.

Binary Hypothesis Testing

In the terminology of statistical inference, these functions are referred to as **likelihood functions**. For example, $f_{\mathbf{X}|H_0}(\mathbf{x})$ is the **likelihood** of $\mathbf{x}$ given H_0.

We call it *likelihood* since the main question we are asking is related to the model, and not the event. In other words, the how likely it is to see the observed data if the model is true.

Binary Hypothesis Testing

- The test design divides S into two sets that form a *partition*, A_0 and $A_1 = A_0^c$.
- If the outcome $s \in A_0$, the decision is *accept* H_0. Otherwise, $s \in A_1$, the decision is *accept* H_1.
- However, since we *assume* that H_0 is true, we typically don't "accept" the null hypothesis, we therefore either say:
  - We reject the null hypothesis or
  - We fail to reject the null hypothesis.
- Note that this does not mean that we have *proven* the null hypothesis, there was just not strong enough evidence to reject it.

Binary Hypothesis Testing

One electrical engineering application of binary hypothesis testing relates to a radar system. The transmitter sends out a signal, and it is the job of the receiver to decide whether a target is present.

To make this decision, the receiver examines the received signal to determine whether it contains a reflected version of the transmitted signal. The hypothesis H_0 corresponds to the situation in which there is no target. H_1 corresponds to the presence of a target. In the terminology of radar, a **Type I error** (decide target present when there is no target) is referred to as a *false alarm* and a **Type II error** (decide no target when there is a target present) is referred to as a *miss*.

Binary Hypothesis Testing

Confusion Matrix

		Truth: Positive (H_0)	Truth: Negative (H_a)	
Test	Positive	True Positive TP	False Positive FP or Type I Error	Total Testing Positive
	Negative	False Negative FN or Type II Error	True Negative TN	Total Testing Negative
		Total Truly Positive	Total Truly Negative	Total Samples

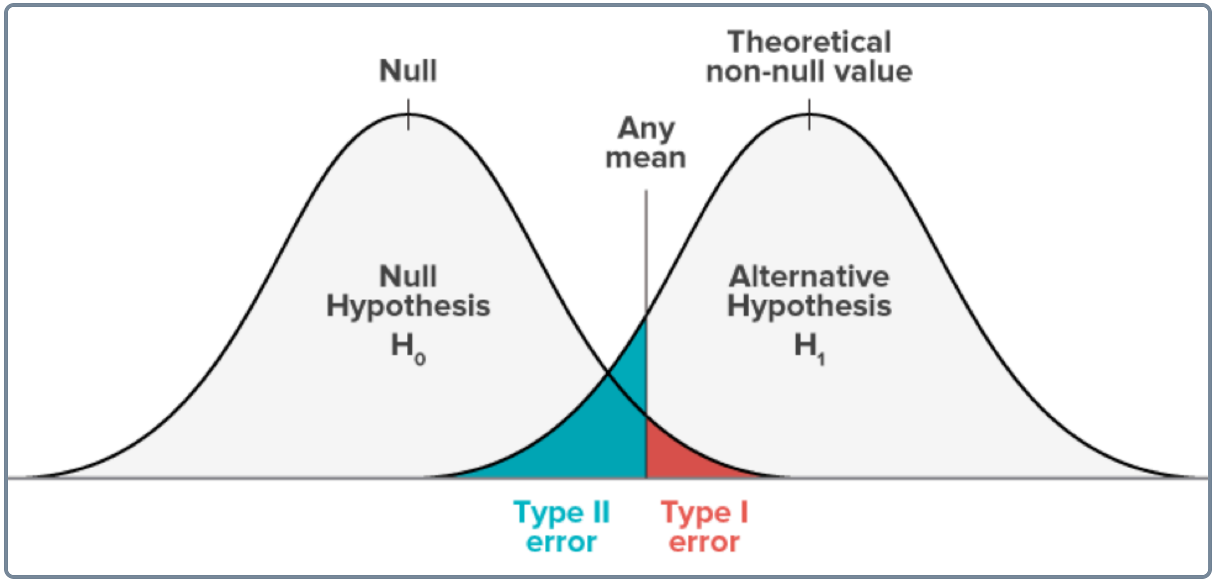
Binary Hypothesis Testing

For interest

		Predicted Class		
		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	**Negative Predictive Value** $\frac{TN}{(TN + FN)}$	**Accuracy** $\frac{TP + TN}{(TP + TN + FP + FN)}$

Binary Hypothesis Testing

- In a binary hypothesis test, two kinds of errors are possible. Statisticians refer to them as **Type I errors** and **Type II errors** 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.



Binary Hypothesis Testing

The accuracy measure of the test consists of the corresponding two error probabilities.

- $P(A_1|H_0)$ is the probability of a **Type I error**.
  - It is the probability of accepting H_1 when H_0 is the true probability model.
- $P(A_0|H_1)$ is the probability of a **Type II error**.
  - It is the probability of accepting H_0 when H_1 is the true probability model.

Binary Hypothesis Testing

- The design of a binary hypothesis test represents a trade-off between the two error probabilities,
  - $P(\text{False Alarm}) = P_{FP} = P(A_1|H_0)$ - a false positive
  - $P(\text{Miss}) = P_{FN} = P(A_0|H_1)$ - a false negative
- To understand the trade-off, consider an extreme design in which $A_0 = S$ and $A_1 = \emptyset$.
  - Then $P_{FP} = 0$ and $P_{FN} = 1$.
- Now let A_1 expand to include an increasing proportion of the outcomes in S. As A_1 expands, P_{FP} increases and P_{FN} decreases.

Detection Error Tradeoff (DET) curve

A graph representing the possible values A_0 and A_1 of $P_{\mathrm{FA}} = P_{FP}$ and $P_{\mathrm{MISS}} = P_{FN}$ is referred to as a **Detection Error Tradeoff (DET)** curve.

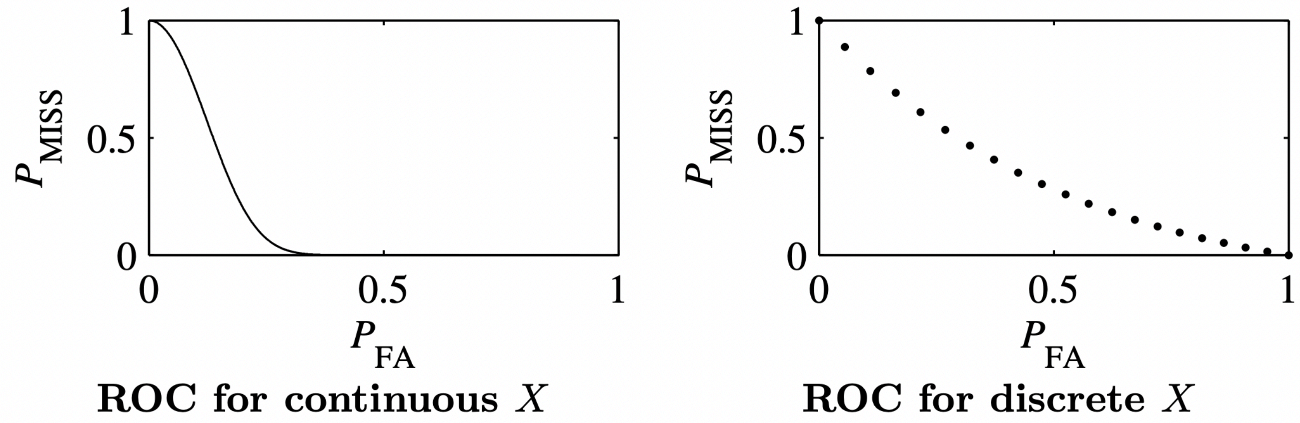
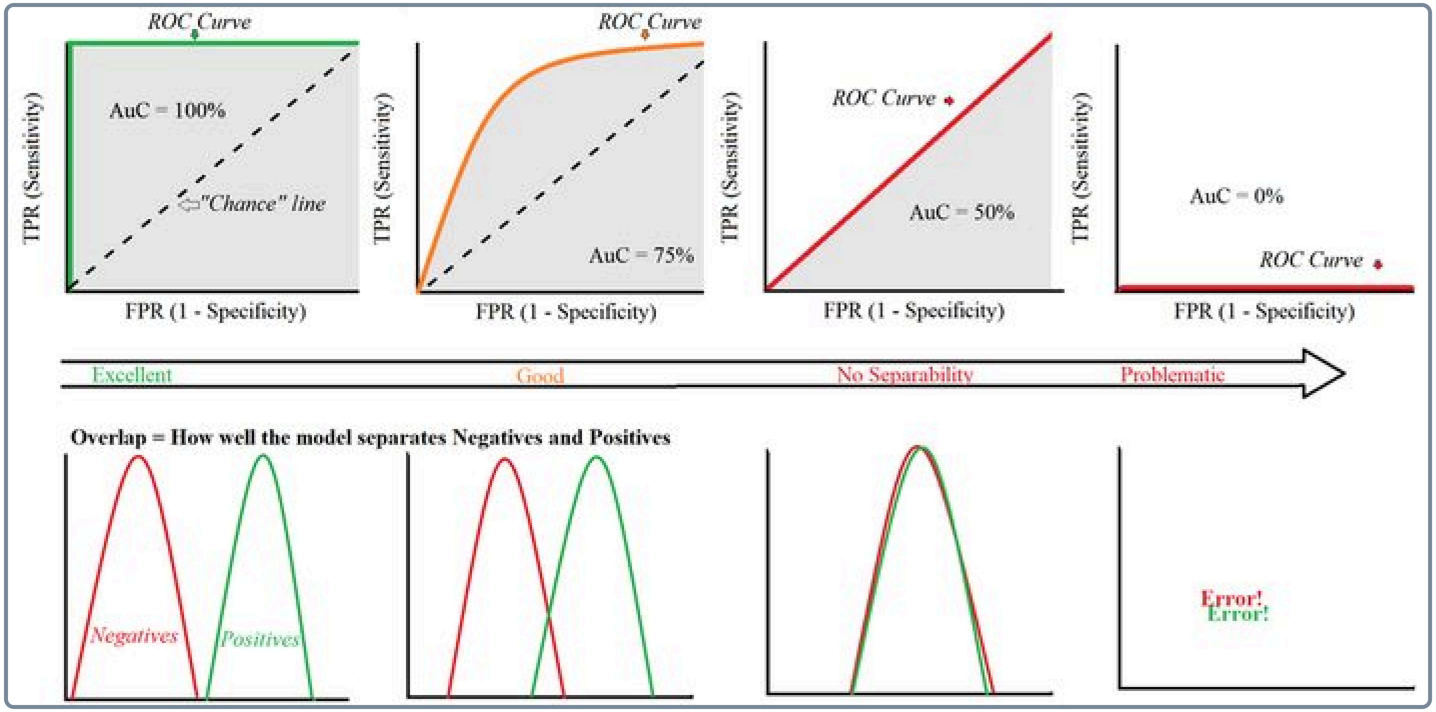


Figure 11.1 Continuous and discrete examples of a receiver operating curve (ROC).

(these are not ROC curves, but DET curves)

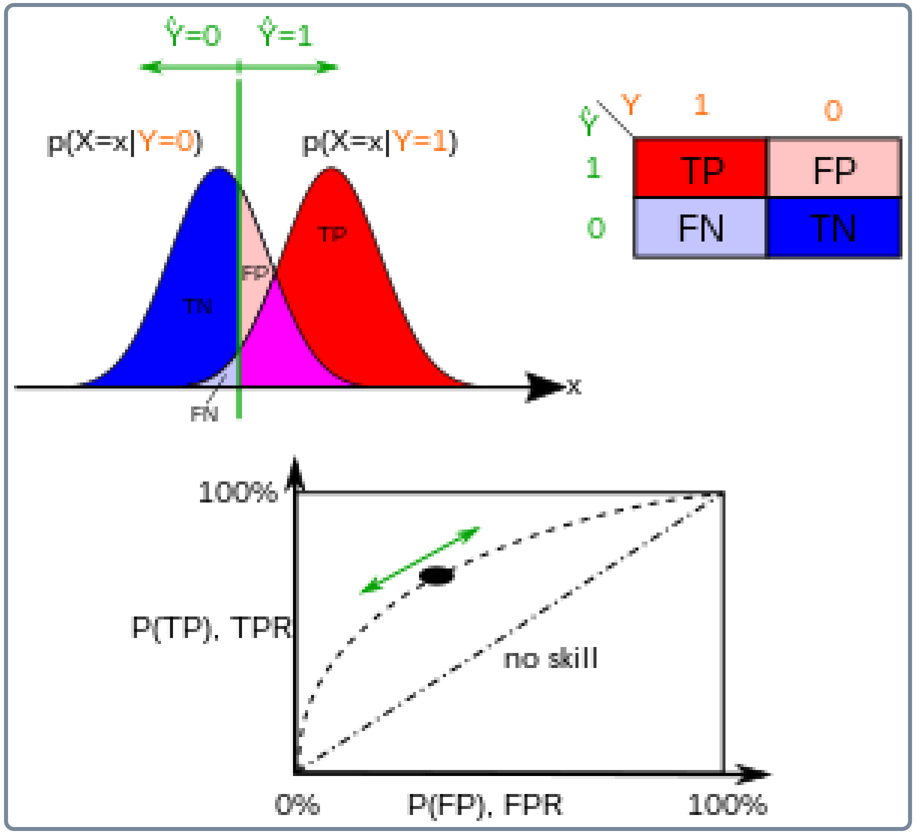
At the top left corner of the graph, the point $(0, 1)$ corresponds to $A_0 = S$ and $A_1 = \emptyset$, and $(1, 0)$ is $A_0 = \emptyset$ and $A_1 = S$

Receiver Operating Characteristic (ROC)



Receiver Operating Characteristic (ROC)

Here P_{TP} is on vertical axis and P_{FP} is on horizontal axis.



To compare performances of different models, we often calculate the **area under the curve (AUC)**

Maximum A posteriori Probability (MAP) test

The **maximum a posteriori probability test** *minimizes* P_{ERR}, the total probability of error of a binary hypothesis test.

$$P_{\text{ERR}} = P(A_1|H_0)P(H_0) + P(A_1|H_1)P(H_1)$$

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

$$\boldsymbol{s} \in A_0 \textbf{ if } P(H_0|\boldsymbol{s}) \geq P(H_1|\boldsymbol{s}) \qquad \boldsymbol{s} \in A_1 \textbf{ otherwise}$$

$P(H_0|\boldsymbol{s})$ and $P(H_1|\boldsymbol{s})$ are referred to as the **a posteriori** probabilities of H_0 and H_1.

We must therefore accept the hypothesis with the higher a posterior probability to minimize P_{ERR}.

MAP test

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

Discrete

$$\mathbf{x} \in A_0 \text{ if } \frac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq \frac{P(H_1)}{P(H_0)} \quad \mathbf{x} \in A_1 \ \ \text{otherwise}$$

Continuous

$$\mathbf{x} \in A_0 \text{ if } \frac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \frac{P(H_1)}{P(H_0)} \quad \mathbf{x} \in A_1 \ \ \text{otherwise}$$

The ratio of conditional probabilities is referred to as a **likelihood ratio**.
The righthand side is a constant that we compare with.

MAP test

$$\mathbf{x} \in A_0 \text{ if } \frac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \frac{P(H_1)}{P(H_0)} \quad \mathbf{x} \in A_1 \ \text{ otherwise}$$

- We can view the likelihood ratio as the evidence, *based on an observation*, in favour of H_0.
- If the likelihood ratio is greater 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 favour of H_1.
- Therefore, the formula states that accepting H_0 is the better decision if the evidence in favour of H_0, based on the experiment, outweighs the prior evidence in favor of accepting H_1.

Example 11.7

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 $\mathrm{P}[H_1] = 0.1 = 1 - \mathrm{P}[H_0]$. 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}.

Minimum Cost Test

- The MAP test implicitly assumes that both types of errors (miss and false alarm) are equally serious. But this is typically not true.
- Consider an application in which $C = C_{10} = C_{FP}$ units is the cost of a false alarm (decide H_1 when H_0 is correct) and $C = C_{01} = C_{FN}$ 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

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

We then minimize $\mathbf{E}[C]$ as the goal of the **minimum cost hypothesis test**.

Minimum Cost Test

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

Discrete

$$\mathbf{x} \in A_0 \text{ if } \frac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq \frac{P(H_1)C_{01}}{P(H_0)C_{10}} \quad \mathbf{x} \in A_1 \;\; \text{otherwise}$$

Continuous

$$\mathbf{x} \in A_0 \text{ if } \frac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \frac{P(H_1)C_{01}}{P(H_0)C_{10}} \quad \mathbf{x} \in A_1 \;\; \text{otherwise}$$

Minimum Cost Test

$$\mathbf{x} \in A_0 \text{ if } \frac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \frac{P(H_1)C_{01}}{P(H_0)C_{10}} \quad \mathbf{x} \in A_1 \text{ otherwise}$$

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

Example 11.8

Continuing the disk drive test of Example 11.7, the factory produces 1000 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?

Based on the given facts, the cost C_{10} of a false alarm is 30 minutes (5000 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 1000 drives returned at \$200 per drive, the expected cost is \$200,000. The minimum cost test is

$$n \in A_0 \text{ if } \frac{P_{N|H_0}(n)}{P_{N|H_1}(n)} \geq \frac{\mathrm{P}[H_1]\, C_{01}}{\mathrm{P}[H_0]\, C_{10}}; \qquad n \in A_1 \text{ otherwise.} \tag{11.15}$$

Performing the same substitutions and simplifications as in Example 11.7 yields

$$n \in A_0 \text{ if } n \geq n^* = 1 + \frac{\ln\left(\frac{q_1\,\mathrm{P}[H_1]C_{01}}{q_0\,\mathrm{P}[H_0]C_{10}}\right)}{\ln\left(\frac{1-q_0}{1-q_1}\right)} = 58.92; \qquad n \in A_1 \text{ otherwise.} \tag{11.16}$$

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

$$P_{\mathsf{FA}} = \mathrm{P}[N \leq 58|H_0] = F_{N|H_0}(58) = 1 - (1 - 10^{-4})^{58} = 0.0058, \tag{11.17}$$

$$P_{\mathsf{MISS}} = \mathrm{P}[N \geq 59|H_1] = 1 - F_{N|H_1}(58) = (1 - 10^{-1})^{58} = 0.0022. \tag{11.18}$$

The average cost (in dollars) of this rule is

$$\begin{aligned} \mathrm{E}[C_{\mathsf{MC}}] &= \mathrm{P}[H_0]\, P_{\mathsf{FA}} C_{10} + \mathrm{P}[H_1]\, P_{\mathsf{MISS}} C_{01} \\ &= (0.9)(0.0058)(50{,}000) + (0.1)(0.0022)(200{,}000) = 305. \end{aligned} \tag{11.19}$$

By comparison, the MAP test, which minimizes the probability of an error rather than the expected cost, has error probabilities $P_{\mathsf{FA}} = 0.0045$ and $P_{\mathsf{MISS}} = 0.0087$ and the expected cost

$$\mathrm{E}\left[C_{\mathsf{MAP}}\right] = (0.9)(0.0045)(50{,}000) + (0.1)(0.0087)(200{,}000) = 376.50. \qquad (11.20)$$

The effect of the high cost of a miss has been to reduce the miss probability from 0.0087 to 0.0022. However, the false alarm probability rises from 0.0047 in the MAP test to 0.0058 in the minimum cost test. A savings of $\$376.50 - \$305 = \$71.50$ 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 1000 returned drives at an expected cost of $\$200{,}000$. Since such days occur with probability $\mathrm{P}[H_1] = 0.1$, the expected cost of a "no test" policy is $\$20{,}000$ per day.

Maximum Likelihood (ML) Test

- For the **Maximum Likelihood (ML) test**, for each outcome $\boldsymbol{s}$ we decide the hypothesis H_i for which $P(\boldsymbol{s}|H_i)$ is largest.
- The idea behind choosing a hypothesis to maximize the probability of the observation is to *avoid making assumptions* about costs and a priori probabilities $P[H_i]$.
- The resulting decision rule, called the maximum likelihood (ML) rule, can be written mathematically as:

$$\boldsymbol{s} \in A_0 \text{ if } P(\boldsymbol{s}|H_0) \geq P(\boldsymbol{s}|H_1) \quad \boldsymbol{s} \in A_1 \ \text{ otherwise}$$

This is the same as MAP, but now the priors are equal

$$P(H_0) = P(H_1)$$

Maximum Likelihood (ML) Test

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

Discrete

$$\mathbf{x} \in A_0 \text{ if } \frac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq 1 \quad \mathbf{x} \in A_1 \text{ otherwise}$$

Continuous

$$\mathbf{x} \in A_0 \text{ if } \frac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq 1 \quad \mathbf{x} \in A_1 \text{ otherwise}$$

Therefore, choose the model that maximizes the likelihood ratio.

Example 11.10

Continuing the disk drive test of Example 11.7, design the maximum likelihood test for the factory status based on the decision statistic N, the number of drives tested up to and including the first failure.

The ML hypothesis test corresponds to the MAP test with $\mathrm{P}[H_0] = \mathrm{P}[H_1] = 0.5$. In this case, Equation (11.11) implies $n^* = 66.62$ or $A_0 = \{n \geq 67\}$. The conditional error probabilities and the cost of the ML decision rule are

$$\begin{aligned}
P_{\mathsf{FA}} &= \mathrm{P}\left[N \leq 66 | H_0\right] = 1 - (1 - 10^{-4})^{66} = 0.0066, \\
P_{\mathsf{MISS}} &= \mathrm{P}\left[N \geq 67 | H_1\right] = (1 - 10^{-1})^{66} = 9.55 \cdot 10^{-4}, \\
\mathrm{E}\left[C_{\mathsf{ML}}\right] &= (0.9)(0.0066)(50{,}000) + (0.1)(9.55 \cdot 10^{-4})(200{,}000) = \$316.10.
\end{aligned}$$

For the ML test, $P_{\mathsf{ERR}} = 0.0060$. 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, the cost is lower because a costly miss occurs very infrequently (only once every four months). The cost of the ML test is only \$11.10 more than the minimum cost. This is because the a priori probabilities suggest avoiding false alarms because the factory functions correctly, while the costs suggest avoiding misses, because each one is very expensive. Because these two prior considerations balance each other, the ML test, which ignores both of them, is very similar to the minimum cost test.