Lecture 1: Introduction to Statistical Inference

1 Introduction

We are interested in processing of information-bearing signals to extract information. There are two types of problems of fundamental interest.

1. Detection: Finite number of possible situations

2. Estimation: "Nearest" to the possible situation

We look at three examples of interest.

Example 1.1 (Communication System). We need to estimate unknown analog signal from the received signal that is distorted and corrupted.
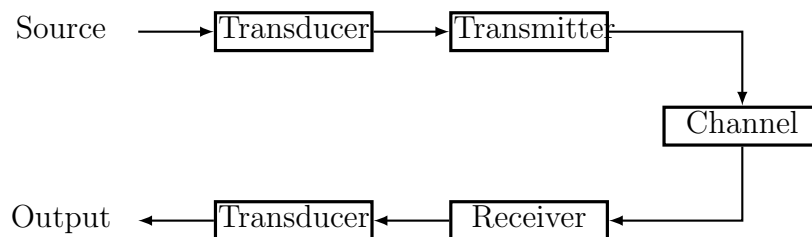

Figure 1: Block diagram for analog communication.

Example 1.2 (Radar communication). Pulse electromagnetic waves are sent and received after possible reflection from a target if it exists. Target could be an aircraft, ship, spacecraft, missile etc. If the target is detected, then one is interested in estimating range, angle, and velocity of the target. One may be interested in tracking the mobile target trajectory or even controlling it.

Example 1.3 (Automatic Control). In automatic control problem, given a linear time invariant system H, we need to design a controller C for achieving a desired response through output signal $y(t)$. Typically, reference signal $x(t)$ is unknown in such systems, and only a noisy version of the state may be observable.
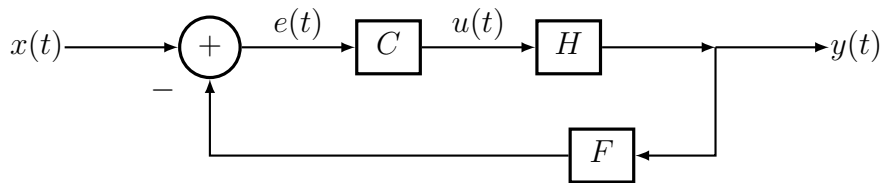
Figure 2: Block diagram for automatic control.

Other applications of estimation and detection theory are in seismology, radio astronomy, sonar, speech, signal, and image processing, biomedical signal processing, optimal communications etc.

2 Probability Review

We denote observation space by Γ equipped with a σ-algebra $\mathcal{G}$, that is a measurable collection of sets. Further, for all elements of $A \in \mathcal{G}$, we have a non-negative set function $P : \Gamma \to [0, 1]$ that satisfies the following axioms of probability:

1. $P(\Gamma) = 1$,

2. for any disjoint countable collection of sets $\{A_n : n \in \mathbb{N}\}$, we have $P(\cup_n A_n) = \sum_n P(A_n)$.

Example 2.1 (Finite Observations). When observation space Γ has finitely many elements, we can take $\mathcal{G} = \mathcal{P}(\Gamma)$. Further, specifying $P(\{\gamma\})$ for all $\gamma \in \Gamma$ completely specifies the probability set function.

Example 2.2 (Euclidean Space). For the case when observation space $\Gamma = \mathbb{R}^n$, we take $\mathcal{G} = \mathcal{B}^n$, Borel σ-algebra on $\mathbb{R}^n$. For this case, it suffices to specify the set function $P(A)$ for sets $A \in \mathcal{G}$ of the form $\{\gamma \in \Gamma : \gamma_i \leq x_i, i \in [n]\}$.

Definition 2.3 (Expectation). For a real valued function $g : \Gamma \to \mathbb{R}$, we denote its expectation by $\mathbb{E}[g(Y)]$ and define it as

$$\mathbb{E}[g(Y)] = \int_{y \in \Gamma} g(y) dP(y).$$

3 Hypothesis Testing

Definition 3.1. A **hypothesis** is a statement about a population parameter.

Definition 3.2. The two complementary hypotheses in a hypothesis testing problem are called the **null hypothesis** and the **alternative hypothesis**, and denoted by $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively.

We assume that observation is a random variable $Y \in \Gamma$ distributed with probability set function P_i when true hypothesis is $\mathcal{H}_i$ for $i \in \{0, 1\}$.

Definition 3.3. A **hypothesis test** is a rule $\delta : \Gamma \rightarrow \{0, 1\}$ that specifies for all values of $y \in \Gamma$, index of the accepted true hypothesis $\mathcal{H}_{\delta(y)}$.

Definition 3.4. Region $\Gamma_1 = \{y \in \Gamma : \delta(y) = 1\}$ is called the **rejection region**, and $\Gamma_0 = \Gamma_1^c$ is called the **acceptance region**.

Definition 3.5. When the true underlying hypothesis is $\mathcal{H}_j$, the **cost** incurred on accepting hypothesis $\mathcal{H}_i$ is denoted by C_{ij}. Uniform cost is given by

$$C_{ij} = 1_{\{i \neq j\}}, i, j \in \{0, 1\}.$$

Definition 3.6. For each hypothesis $\mathcal{H}_i$, the **conditional risk** is denoted by $R_j(\delta)$ and defined as the expected cost incurred by the decision rule δ, when it is the underlying true hypothesis. That is,

$$R_j(\delta) = \mathbb{E}\left[\sum_i C_{ij} 1_{\{\delta(y)=i\}} | \mathcal{H}_j \text{ holds}\right] = \mathbb{E}_j\left[\sum_i C_{ij} 1_{\{\delta(y)=i\}}\right]$$
$$= C_{0j} P_j(\Gamma_0) + C_{1j} P_j(\Gamma_1).$$

Our objective is to design a decision rule δ that minimizes risk. Usually, costs of correct identification of the true hypothesis is low, and cost of incorrect identification is higher. Hence, minimizing risk for any hypothesis would be to ensure that probability $P_j(\Gamma_i)$ is low for $i \neq j$. One can't simultaneously decrease all decision regions $\{\Gamma_i\}$, since they form a partition of observation space Γ.

3.1 Bayesian Hypothesis Testing

In this approach, we assume a prior distribution π on hypotheses to be true. Specifically, let π_i denotes the prior probability of $\mathcal{H}_i$ being true.

Definition 3.7. We can defined unconditional **risk** as expected value of risk over all possible hypotheses, denoted by $r(\delta)$. That is,

$$r(\delta) = \mathbb{E}R_j(\delta) = \sum_j R_j(\delta)\pi_j.$$

Definition 3.8. For two distributions $P_1(y)$ and $P_0(y)$ we can define likelihood ratio as $L(y) = \frac{dP_1}{dP_0}(y)$. When two distributions admit density, it is ratio of their densities at y. For discrete distributions, it is ratio of their probability mass functions at y.

Theorem 3.9. *For a Bayesian hypothesis testing problem optimal decision rule that minimizes unconditional risk for a prior distribution π and costs $\{C_{ij}\}$ is a threshold based rule called likelihood ratio test . That is,*

$$\delta_B(y) = 1_{\{L(y) \geq \tau\}},$$

where likelihood ratio $L(y) = \frac{dP_1}{dP_0}(y)$ and threshold $\tau = \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})}$.

Proof. We can write unconditional risk as

$$r(\delta) = \sum_j \pi_j C_{0j} + \int_{y \in \Gamma_1} \sum_j \pi_j(C_{1j} - C_{0j})dP_j(y).$$

Minimizing unconditional risk is equivalent to selecting rejection region Γ_1 such that the integrand is negative. That is,

$$\Gamma_1 = \{y \in \Gamma : \sum_j \pi_j(C_{1j} - C_{0j})dP_j(y) \leq 0\}.$$

By the definition of likelihood ratio and the threshold as defined in the theorem hypothesis, theorem follows. $\square$

Remark 1. For uniform cost, threshold $\tau = \frac{\pi_0}{\pi_1}$ and conditional risk

$$r(\delta) = \pi_0 P_0(\Gamma_1) + \pi_1 P_1(\Gamma_0),$$

is probability of error in detection.

Definition 3.10. We can define **posterior** probability of hypothesis $\mathcal{H}_j$ being true conditioned on observation being y as

$$\pi_j(y) = \Pr\{\mathcal{H}_j \text{ is true } | Y = y\} = \frac{\pi_j dP_j(y)}{\pi_0 dP_0(y) + \pi_1 dP_1(y)}.$$

Remark 2. Observe that rejection region can be written in terms of posterior probabilities as

$$\Gamma_1 = \{y \in \Gamma : (C_{10} - C_{00})\pi_0(y) - (C_{01} - C_{11})\pi_1(y) \leq 0\}.$$

This is again a likelihood ratio test in term of ratio of posterior probabilities $L'(y) = \frac{\pi_1(y)}{\pi_0(y)}$ and threshold $\tau' = \frac{C_{10} - C_{00}}{C_{01} - C_{11}}$. That is, Bayes' decision rule is

$$\delta_B(y) = 1_{\{L'(y) \geq \tau'\}}.$$

Definition 3.11. The expected cost of choosing hypothesis $\mathcal{H}_i$ given observation y is called **posterior cost** and denoted as $R_i(y)$, where

$$R_i(y) = \mathbb{E}[\sum_{k,j} C_{kj} 1_{\{\delta(y)=i\}}|Y=y] = \sum_{j} C_{ij}\pi_j(y) = C_{i0}\pi_0(y) + C_{i1}\pi_1(y).$$

Remark 3. Alternatively, one can also write rejection region in terms of posterior costs as

$$\Gamma_1 = \{y \in \Gamma : R_0(y) \le R_1(y)\}.$$

Therefore, Bayes' decision rule can be interpreted as the one that minimizes the posterior cost of choosing a hypothesis when the observation is y. That is,

$$\delta_B(y) = 1_{\{\frac{R_1(y)}{R_0(y)} \ge 1\}}.$$

Remark 4. For uniform cost, Bayes' decision rule is likelihood ratio test for posterior probabilities when the threshold is unity, That is,

$$\delta_B(y) = 1_{\{L'(y) \ge 1\}}.$$

This is equivalent to maximizing a posterior probability of underlying hypothesis based on the observation. This is also called a MAP decision rule for binary hypothesis test.