1.1 Administrative details

- Instructors: Aditya Gopalan (aditya@ece.iisc.ernet.in), Parimal Parag (parimal@ece.iisc.ernet.in)
- Teaching assistant: Raghava GD (raghavagd@ece.iisc.ernet.in)
- Grading policy: Homework assignments (including programming exercises): 30%, Scribing lectures: 10%, Midterm exam: 20%, Final exam: 40%
- Course webpage: www.ece.iisc.ernet.in/~parimal/estimation.html
- Scribing: Administered via Github, sign up via course webpage/email
- Course discussions: Administered via Slack, sign up via course webpage/email
- Textbooks: V. Poor [?], Casella & Berger [?] (also on the course webpage)

1.2 Introduction

Detection

Detection, broadly speaking, is the process of answering the question “Is \( X \) present (or not)\?”, based on observations or data. \( X \) is a property of interest.

Examples:

- (Communication systems) Is there a message that was sent (based on received data)\?
- (Radar) Is there an enemy airplane in the vicinity\?
- (Astronomy) Is there a planet/star/nebula in this part of space\?
- (Medicine) Is this patient sick? Is treatment \( T \) effective for disease \( D \)?
- (Physics) Is this body hot/massive/dense? Is the Higgs boson present? 
- (Economics) Is there a slowdown in economic growth since November 8, 2016 [?]?
Estimation

Estimation is the process of answering the question “What is the value $X$?”, based on observations or data. $X$ is typically a quantity of interest.

Examples:

- (Communication systems) What is the message that was sent? What was the temperature measurement from the sensor when it was transmitted to the processor?
- (GPS) What is my position/velocity/acceleration?
- (Astronomy) What is the size of the planet? What is its chemical composition?
- (Economics) What is India’s GDP/population size today? What was the typical waiting time to withdraw cash from ATMS on November 8, 2016?

1.2.1 Goals of this course and takeaways

1. Study the qualitative problems of detection and estimation in the framework of statistical inference.

2. Gain an understanding of, and develop the ability to design, automated systems for detection and estimation (these are often key subsystems of larger systems in real life).

1.3 Detection theory

We frame problems of detection in the language of statistical hypothesis testing [K. Pearson, W. S. Gosset, R. Fisher, J. Neyman, E. Pearson, pre- and post- World War II period].

1.3.1 The hypothesis testing framework

- $\Gamma$: Observation set
- $\mathcal{G}$: Class of subsets of $\Gamma$ that can be assigned probabilities. We will typically take this to be a $\sigma$-algebra, i.e., a class of subsets containing $\emptyset$ and $\Gamma$, and closed under complements and countable unions. Examples: (a) $\Gamma = \{1, 2, \ldots, n\} \equiv [n]$, $\mathcal{G} = 2^\Gamma$, the power set (all subsets) of $\Gamma$, (b) $\Gamma = \mathbb{R}$, $\mathcal{G} = B(\mathbb{R})$, the Borel $\sigma$-algebra on $\mathbb{R}$.

- Notation. $\mathbb{P}[A]$: the probability assigned to $A \subseteq \Gamma$, $\mathbb{1}_X$: the indicator function (0 if true, 1 if false) of $X$
M-ary hypothesis testing: Decide which of $M$ possible situations could describe the situation at hand, based on an observation $Y \in \Gamma$.

The case $M = 2$ is important and known as **binary hypothesis testing**, and will form the focus of our study for several classes.

1.3.2 Binary hypothesis testing

We assume that there are two possible hypotheses, $H_0$ and $H_1$, which correspond to two probability distributions $P_0$ and $P_1$, respectively, on $(\Gamma, \mathcal{G})$.

**Problem:** Decide $H_0 : Y \in \Gamma, Y \sim P_0$ (the “null” hypothesis) vs. $H_1 : Y \in \Gamma, Y \sim P_1$ (the “alternative” hypothesis)

**Definition 1.3.1.** A decision rule (or hypothesis test) for $H_0$ vs $H_1$ is a map $\delta : \Gamma \rightarrow \{0, 1\}$.

Equivalently, a decision rule is a partition of the observation space $\Gamma$ into two regions, corresponding to deciding $H_0$ or $H_1$ (Figure 1.1):

$$
\Gamma_0 := \{y \in \Gamma : \delta(y) = 0\}, \\
\Gamma_1 := \{y \in \Gamma : \delta(y) = 1\}.
$$

**Note:** We have the identity $\delta(y) = \mathbb{1}\{y \in \Gamma_1\} \forall y \in \Gamma$.

![Figure 1.1: A decision rule $\delta$ in terms of its partition of the observation space.](image)

**Cost:** We denote by $C_{i,j}$ the cost of deciding hypothesis $H_i$ when hypothesis $H_j$ was actually true, $i, j \in \{0, 1\}$. Typically, $C_{i,j} \approx 0$ if $i = j$ and $C_{i,j} \gg 0$ otherwise.

**Definition 1.3.2.** The **conditional risk** of a decision rule $\delta$ for hypothesis $j \in \{0, 1\}$ is its expected cost when hypothesis $j$ is actually true. We denote it by $R_j(\delta)$. Thus,

$$
R_j(\delta) := E_j \left[ C_{j,0} \mathbb{1}\{Y \in \Gamma_0\} + C_{j,1} \mathbb{1}\{Y \in \Gamma_1\} \right] \\
= C_{j,0} P_j[\Gamma_0] + C_{j,1} P_j[\Gamma_1].
$$
There are several ways of defining the performance of decision rules. The first one we will study is within a Bayesian setting.

### 1.3.3 Bayesian (binary) hypothesis testing

**Assumption 1.3.3.** There are known prior or a priori probabilities of the occurrences of hypotheses $H_0$ and $H_1$, denoted by $\pi_0$ and $\pi_1$, respectively (note: $\pi_0 + \pi_1 = 1$).

**Definition 1.3.4.** The Bayes risk of a decision rule $\delta$, denoted by $r(\delta)$, is the average cost incurred by $\delta$ under the prior distribution $(\pi_0, \pi_1)$, i.e., $r(\delta) := \pi_0 R_0(\delta) + \pi_1 R_1(\delta)$.

**Definition 1.3.5.** An optimal rule for Bayesian hypothesis testing is a rule $\delta$ that minimizes $r(\delta)$. We also call such a rule a Bayes rule (not to be confused by Bayes’ formula in probability!).

**Finding a Bayes rule**

For a decision rule $\delta$, we have

\[
r(\delta) := \pi_0 R_0(\delta) + \pi_1 R_1(\delta)
\]

\[
= \sum_{j=0}^{1} \pi_j R_j(\delta)
\]

\[
= \sum_{j=0}^{1} \pi_j \left( C_{j0} \mathbb{P}_j[\Gamma_0] + C_{j1} \mathbb{P}_j[\Gamma_1] \right)
\]

\[
= \sum_{j=0}^{1} \pi_j \left( C_{j0} \left(1 - \mathbb{P}_j[\Gamma_1]\right) + C_{j1} \mathbb{P}_j[\Gamma_1] \right)
\]

\[
= \sum_{j=0}^{1} \pi_j C_{j0} + \sum_{j=0}^{1} \pi_j \left( C_{j1} - C_{j0} \right) \mathbb{P}_j[\Gamma_1].
\]

Let us assume for concreteness that the probability distribution $\mathbb{P}_j$ has a probability density function (pdf) or probability mass function (pmf) $p_j$ (most cases we will encounter will have one of these properties). We can continue thus:

\[
r(\delta) = \sum_{j=0}^{1} \pi_j C_{j0} + \sum_{j=0}^{1} \pi_j \left( C_{j1} - C_{j0} \right) \int_{\Gamma_1} p_j(y)dy
\]

\[
= \sum_{j=0}^{1} \pi_j C_{j0} + \sum_{j=0}^{1} \pi_j \left( C_{j1} - C_{j0} \right) \int_{\Gamma} \delta(y) p_j(y)dy
\]

\[
= \sum_{j=0}^{1} \pi_j C_{j0} + \int_{\Gamma} \left[ \sum_{j=0}^{1} \pi_j \left( C_{j1} - C_{j0} \right) p_j(y) \right] \delta(y)dy.
\]
Therefore, the best choice of \( \delta : \Gamma \to \{0, 1\} \), to minimize \( r(\delta) \), is to set \( \delta(y) = 1 \) if and only if the integrand in square brackets in the above expression is nonpositive. In other words, the following rule \( \delta_B \) is a Bayes rule:

\[
\delta_B(y) = \begin{cases} 
1, & \text{if } \sum_{j=0}^{1} \pi_j (C_{j1} - C_{j0}) p_j(y) \leq 0 \\
0, & \text{if } \sum_{j=0}^{1} \pi_j (C_{j1} - C_{j0}) p_j(y) > 0.
\end{cases}
\]

(Note: Breaking ties either way does not change the Bayes risk and is thus inconsequential as far as Bayes optimality is concerned.)
Bibliography

