ECE 7251: Signal Detection and Estimation

Spring 2002

Prof. Aaron Lanterman
Georgia Institute of Technology

Lecture 19, 2/20/02:
Introduction to Detection Theory
(Including Bayesian and MinMax Tests)

The Setup

- Parametric data model $p(y;q)$
- In "point" estimation theory, we tried to find the parameter $q \in \mathcal{S}$ given the data y
- Suppose $\mathcal{S}_0$ and $\mathcal{S}_1$ form a partition of $\mathcal{S}$

$$\mathcal{S}_0 \cup \mathcal{S}_1 = \mathcal{S}, \quad \mathcal{S}_0 \cap \mathcal{S}_1 = \emptyset$$

- In detection theory, we try to identify which hypothesis is true:

$$H_0 : q \in \mathcal{S}_0 \quad \text{(null hypothesis)}$$
$$H_1 : q \in \mathcal{S}_1 \quad \text{(alternative hypothesis)}$$

Terminology

- If q can only take on two values,

$$\mathcal{S} = \{q_0, q_1\}, \ \mathcal{S}_0 = \{q_0\}, \ \mathcal{S}_1 = \{q_1\}$$

  we say the hypotheses are simple.

- Otherwise, we say they are composite.
  - Example:

$$H_0 : q \in 0$$
$$H_1 : q \in (0, \infty)$$

The Decision Rule

- Design a decision rule (function) $f : \mathcal{Y} \to \{0,1\}$

$$f(y) = \begin{cases} 1, & \text{decide } H_1 \\ 0, & \text{decide } H_0 \end{cases}$$

- f partitions data space into decision regions:

$$\mathcal{Y}_0 = \{y : f(y) = 0\}, \ \mathcal{Y}_1 = \{y : f(y) = 1\}$$

- Note if hypothesis are simple, it would be very natural to write $\hat{q}$ instead of f

Probabilities of Bad and Good Things

- Probabilities of False Alarm and Miss

$$P_{FA}(q) = E_q[f] = \int_{\mathcal{Y}_1} p(y;q)dy \qquad \text{for } q \in \mathcal{S}_0$$

$$P_M = E_q[1-f] = 1 - \int_{\mathcal{Y}_1} p(y;q)dy \qquad \text{for } q \in \mathcal{S}_1$$

$$= \int_{\mathcal{Y}_0} p(y;q)dy$$

- Probability of Detection (correctly deciding H_1)

$$P_D(q) = 1 - P_M(q) = E_q[f] = \int_{\mathcal{Y}_1} p(y;q)dy \qquad \text{for } q \in \mathcal{S}_0$$

Weird Mathematical Terms

- Statisticians use the terms:
  - False Alarm = "Type I error"
  - Miss = "Type II error"
  - Probability of detection = "power"
  - Probability of false alarm = "significance level"

- Poor calls $\mathcal{S}_1$ the critical region or the rejection region (as in "reject the null hypothesis"); that terminology is a bit confusing, it would probably be better to call it the acceptance region (as in "accept in the alternative hypothesis")

Bayesian Detection Theory

- Assume we have a prior on q
- Compute prior probabilities on H_0, H_1

$$\Pr[H_k] = \Pr[\Theta \in \mathcal{S}_k] = \int_{\mathcal{S}_k} p(q)\,dq, \quad k = 0,1$$

- Compute likelihood probabilities

$$p(y \mid H_k) = \frac{\int_{\mathcal{S}_k} p(x \mid q)\, p(q)\,dq}{\Pr[H_k]}, \quad k = 1,2$$

- Note composite detection problem has been reduced to a simple detection problem

Minimize Average Risk

- Assign a cost function *c(declared,true)*
  - $c(1,0)$ is the cost of a false alarm
  - $c(0,1)$ is the cost of a miss
  - $c(1,1)$ and $c(0,0)$ (costs of correct decisions) always set to zero in real life
- Select f to minimize the average risk

$$\begin{aligned} R = E[C] &= c(1,0)\Pr(\text{say } H_1 \mid H_0)\Pr(H_0) \\ &\quad + c(0,1)\Pr(\text{say } H_0 \mid H_1)\Pr(H_1) \\ &= c(1,0)\Pr(H_0)P_{FA} + c(0,1)\Pr(H_1)P_M \end{aligned}$$

Minimizing the Average Risk

$$\begin{aligned} R &= c(1,0)\Pr(H_0)\int_{\mathcal{Y}_1} p(y \mid H_0)\,dy \\ &\quad + c(0,1)\Pr(H_1)[1 - \int_{\mathcal{Y}_1} p(y \mid H_1)\,dy] \\ &= \int_{\mathcal{Y}_1} c(1,0)\Pr(H_0)p(y \mid H_0) \\ &\quad - c(0,1)\Pr(H_1)p(y \mid H_1)\,dy + c(0,1)\Pr(H_1) \end{aligned}$$

Solution: Let $\mathcal{Y}_1 =$

$$\{y : c(0,1)\Pr(H_1)p(y \mid H_1) > c(1,0)\Pr(H_0)p(y \mid H_0)\}$$

The Bayesian Likelihood Ratio Test

$$c(0,1)\Pr(H_1)p(y \mid H_1) \underset{H_0}{\overset{H_1}{\gtrless}} c(1,0)\Pr(H_0)p(y \mid H_0)$$

$$\underbrace{\Lambda(y)}_{\text{likelihood ratio}} \equiv \frac{p(y \mid H_1)}{p(y \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{c(1,0)\Pr(H_0)}{c(0,1)\Pr(H_1)} \equiv \underbrace{t}_{\text{threshold}}$$

- Equivalently:

$$\log \Lambda(y) = \ln p(y \mid H_1) - \ln p(y \mid H_0) \underset{H_0}{\overset{H_1}{\gtrless}} \ln t \equiv t'$$

- Recall from Lecture 2 that $\Lambda(y)$ is a suff. stat.

Minimum Prob. of Error Test

- In the special case $C(1,0)=C(0,1)=1$,

$$R = \Pr(H_0)P_{FA} + \Pr(H_1)P_M = P_e$$

$$\frac{p(y \mid H_1)}{p(y \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\Pr(H_0)}{\Pr(H_1)}$$

$$\frac{p(H_1 \mid y)}{p(H_0 \mid y)} = \frac{p(y \mid H_1)\Pr(H_1)}{p(y \mid H_0)\Pr(H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

"Maximum a posteriori" test

MinMax Tests

- What if we don't know $\Pr[H_0]$, $\Pr[H_0]$?
- Define the <u>conditional risks</u>

$$R_0 = E[C \mid H_0] = c(1,0)\Pr[\text{say } H_1 \mid H_0]$$

$$R_1 = E[C \mid H_1] = c(0,1)\Pr[\text{say } H_0 \mid H_1]$$

$$R = R_0 \Pr[H_0] + R_1 \Pr[H_1]$$

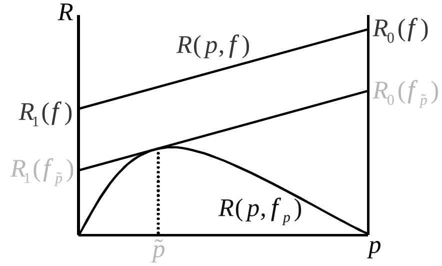
- Pessimistic approach: Choose f which minimizes

$$\max\{R_0(f), R_1(f)\}$$

Graphical Description

Let $p = \Pr[H_0]$

$$R(p,f) = pR_0(f) + (1-p)R_1(f)$$



What Does That Graph Show?

- Graph shows that choosing the decision rule f which minimizes

$$\max\{R_0(f), R_1(f)\}$$

is equivalent to finding the f which minimizes

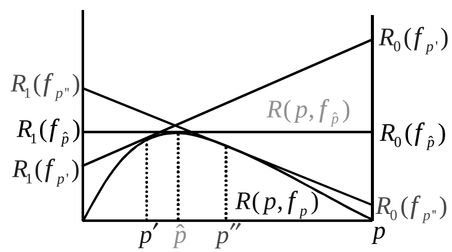
$$\max_{p\in[0,1]} R(p,f)$$

- Note the MinMax test is in fact a Bayesian test

The Equalizer Rule

- MinMax test satisfies the equalizer rule:

$$R_0(f_{\hat{p}}) = R_1(f_{\hat{p}})$$



- $\hat{p}$ is called the least-favorable prior

MinMax Prob. of Error Test

- In the special case $C(1,0)=C(0,1)=1$,

$$R_0(f_{\hat{p}}) = R_1(f_{\hat{p}})$$

reduces to

$$P_{FA}(f_{\hat{p}}) = P_M(f_{\hat{p}})$$

- If $\Pr[\Lambda(Y)=t \mid \bullet] \neq 0$, just find the threshold t which satisfies

$$\Pr[\Lambda(Y) > t \mid H_0] = \Pr[\Lambda(Y) < t \mid H_1]$$

Classification

- Previous slides on Bayesian approach extend naturally to multiple (>2) hypotheses
- Minimum Probability of Error seems to be the most common criterion
- Choose the hypothesis k for which

$$p(H_k \mid y) = p(y \mid H_k)\Pr[H_k]$$

is the greatest
- If $P[H_k]$ are all equal, sometimes called a "maximum likelihood" test