École Polytechnique Fédérale de Lausanne

School of Computer and Communication Sciences

Handout 6 Principles of Digital Communication

Binary Hypothesis testing Feb. 28, 2018

1 Hypothesis testing

Consider the problem of deciding which of two hypotheses, hypothesis 0 or hypothesis 1, is true, based on an observation Y. For simplicity, assume that the observation Y is a random variable taking values in an alphabet $\mathcal{Y}$ — a finite set of $K = |\mathcal{Y}|$ letters — and under hypothesis j it has distribution P_j. To avoid trivial cases we will assume that for each $y \in \mathcal{Y}$ both $P_0(y)$ and $P_1(y)$ are strictly positive. Otherwise, if we observe a y with, say, $P_0(y) = 0$, we would know for sure that hypothesis 1 is true.

A deterministic decision rule associates to each $y \in \mathcal{Y}$ a binary value — i.e., the rule is a function $\phi : \mathcal{Y} \to \{0, 1\}$ — and we decide in favor of hypothesis $\phi(y)$ if the observation Y equals y. In general, we will allow for randomized decision rules: such a rule is characterized by a function $\phi : \mathcal{Y} \to [0, 1]$ that associates to each $y \in \mathcal{Y}$ a value in the *interval* $[0, 1]$, that gives the probability of deciding in favor of hypothesis 1. If our observation Y equals y, we flip a coin that comes heads with probability $\phi(y)$ and tails with probability $1 - \phi(y)$, and decide accordingly: 1 if heads, 0 if tails. We will identify a decision rule with the function ϕ.

In this set up there are two kinds of error: deciding 1 when the true hypothesis is 0, and deciding 0 when the true hypothesis is 1. Letting $\pi_\phi(i|j)$ for rule ϕ denote the probability of deciding i when the truth is j, we see that

$$\pi_\phi(0|1) = \sum_y P_1(y)[1 - \phi(y)], \quad \pi_\phi(1|0) = \sum_y P_0(y)\phi(y).$$

Given P_0 and P_1 and a positive real number $\eta > 0$, let Φ_η to be the set of decision rules ϕ of the form

$$\phi(y) = \begin{cases} 1 & \text{if } P_1(y) > \eta P_0(y) \\ 0 & \text{if } P_1(y) < \eta P_0(y). \end{cases} \tag{1}$$

Note that if there is no y for which $P_1(y) = \eta P_0(y)$, the test ϕ is uniquely specified and Φ_η contains only this test.

Lemma 1. The rules in Φ_η are minimizers of $\pi(0|1) + \eta\pi(1|0)$.

Proof. For any rule $\phi \in \Phi_\eta$, as a consequence of (1), for every $y \in \mathcal{Y}$

$$P_1(y)[1 - \phi(y)] + \eta P_0(y)\phi(y) = \min\{P_1(y), \eta P_0(y)\}.$$

Thus for any rule $\phi \in \Phi_\eta$

$$\pi_\phi(0|1) + \eta\pi_\phi(1|0) = \sum_y P_1(y)[1 - \phi(y)] + \eta P_0(y)\phi(y) = \sum_y \min\{P_1(y), \eta P_0(y)\}.$$

Suppose now ψ is any decision rule. The lemma follows by noting that

$$\pi_\psi(0|1) + \eta\pi_\psi(1|0) = \sum_y P_1(y)[1 - \psi(y)] + \eta P_0(y)\psi(y) \geq \sum_y \min\{P_1(y), \eta P_0(y)\}. \quad \square$$

THEOREM (NEYMAN–PEARSON, 1933). For any $\alpha \in [0,1]$, (i) there is a rule ϕ of the form (1) such that $\pi_\phi(1|0) = \alpha$, and (ii) for any decision rule ψ either $\pi_\psi(0|1) \geq \pi_\phi(0|1)$ or $\pi_\psi(1|0) \geq \pi_\phi(1|0)$.

Proof. Assertion (ii) follows from the lemma above: a ψ that violates both the inequalities would contradict the lemma. It thus suffices to prove (i), the existence of a rule ϕ of the form (1) with $\pi_\phi(1|0) = \alpha$. To that end, define $\Lambda(y) = P_1(y)/P_0(y)$, and label the elements of $\mathcal{Y}$ as $\mathcal{Y} = \{y_1, \dots, y_K\}$ such that $\Lambda(y_1) \geq \Lambda(y_2) \geq \cdots \geq \Lambda(y_K)$. Now define, $a_i = \sum_{j=1}^{i} P_0(y_j)$ for $i = 0, \dots, K$. We then have $0 = a_0 < a_1 < \cdots < a_K = 1$. Given $0 \leq \alpha \leq 1$, we can find $1 \leq i \leq K$ for which $a_{i-1} \leq \alpha \leq a_i$, so that $\alpha = (1-\rho)a_{i-1} + \rho a_i$ for some $\rho \in [0,1]$. Observe that $\alpha = \sum_{j=1}^{i-1} P_0(y_j) + \rho P_0(y_i)$, and that the rule

$$\phi(y) = \begin{cases} 1 & y \in \{y_1, \dots, y_{i-1}\} \\ \rho & y = y_i \\ 0 & y \in \{y_{i+1}, \dots, y_K\} \end{cases}$$

is of the form (1) with $\eta = \Lambda(y_i)$, and $\pi_\phi(1|0) = \alpha$. □

Rules of the form (1) are based on a *likelihood ratio test*: they compare the likelihood ratio $\Lambda(y) = P_1(y)/P_0(y)$ to a threshold η to make a decision. If the likelihood ratio is larger than the threshold, decide 1; if less, decide 0. Equivalently one may compare the *log likelihood ratio*, $\log(P_1(y)/P_0(y))$ to the threshold $\log \eta$.

The theorem stated just above shows the dominant nature of likelihood ratio tests in making decisions: given any decision rule ψ, we can find a (log) likelihood ratio test ϕ which is 'as good or better' — in the sense that the two error probabilities satisfy $\pi_\phi(0|1) \leq \pi_\psi(0|1)$ and $\pi_\phi(1|0) \leq \pi_\psi(1|0)$.