ECE 830 Fall 2011 Statistical Signal Processing

instructor: R. Nowak

Lecture 5: Introduction to Detection Theory

1 Signal Detection and Hypothesis Testing

The simplest type of inference problem is the task of deciding which of two probability models best matches a set of data. Let X be a random variable and denote the two probability models as $p(x|H_0)$ and $p(x|H_1)$. Here H_0 and H_1 stand for "hypothesis" 0 and 1, respectively. In other words we have two hypotheses about how the data x might have been generated; as a random draw from $p(x|H_0)$ or from $p(x|H_1)$. To simplify the notation, we will write $p_i(x)$ to denote $p(x|H_i)$, from here on.The detection or *decision* problem is simply to decide which model is more appropriate.

Example 1 *Additive Gaussian White Noise (AWGN) communication channel. A bit,* 0 *or* 1*, is sent through a noisy channel. The receiver gets the bit plus noise, and the noise is modeled as a realization of a* $\mathcal{N}(0,1)$ *random variable. The receiver must decide between two hypotheses,* $H_0 : X \sim \mathcal{N}(0,1)$ *or* $H_1 : X \sim \mathcal{N}(1,1)$*.*

Example 2 *Radar signal processing. Suppose we measure a radar return signal and need to decide whether at target is present. Let* $p_0(x)$ *denote the probability distribution of the signal when the target is absent (usually called the "null" distribution), and let* $p_1(x)$ *be the distribution when the signal is present (often called the "alternative" distribution).*

Example 3 *Functional Magnetic Resonance Imaging (fMRI). MRI uses magnetic fields to map the density of hydrogen in the brain, and since different tissues have varying levels of hydrogen atoms, this produces an image of the brain's structure. fMRI is a dynamic version of MRI that essentially creates a movie of the brain. Neural activity is correlated with subtle variations in the measured signals, and by detecting these variations it is possible to obtain a map of the activity. Below on the right is the raw fMRI data for one cross-section of the brain and on the left is an image depicting a the value of a correlation test statistic. We can decide which pixels correspond to areas of neural activity by testing whether the pixel's correlation value is better modeled as* $\mathcal{N}(0,1)$ *or* $\mathcal{N}(\mu,1)$*, for some* $\mu > 0$*.*

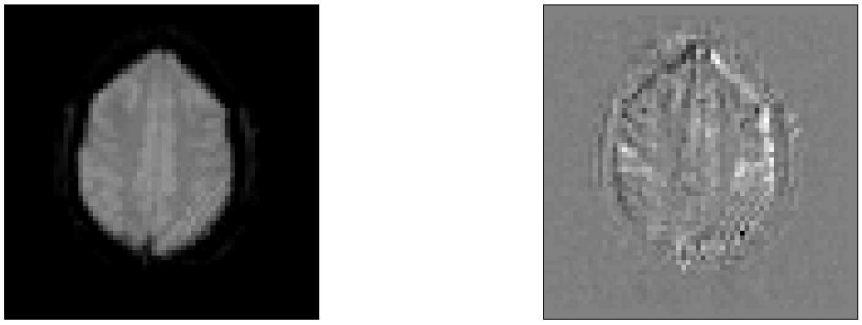

Figure 1: fMRI is used to map neural activity in the brain.

Example 4 *Gene microarrays provide a powerful tool for measuring the levels of gene expression (protein production). Different diseases produce different "expression profiles" across the genome, and by detecting these differences it is possible to discriminate between different disease types. The figure below depicts gene expression data. Suppose we are comparing two conditions, healthy vs. disease. To make a comparison we look at the difference between the expression level of each gene in the two conditions. The situation where there is no difference in gene expression can be modeled as* $\mathcal{N}(0,1)$*. If there is a difference, then the data can be modeled as* $\mathcal{N}(\mu,1)$ *for some* $\mu \neq 0$*.*

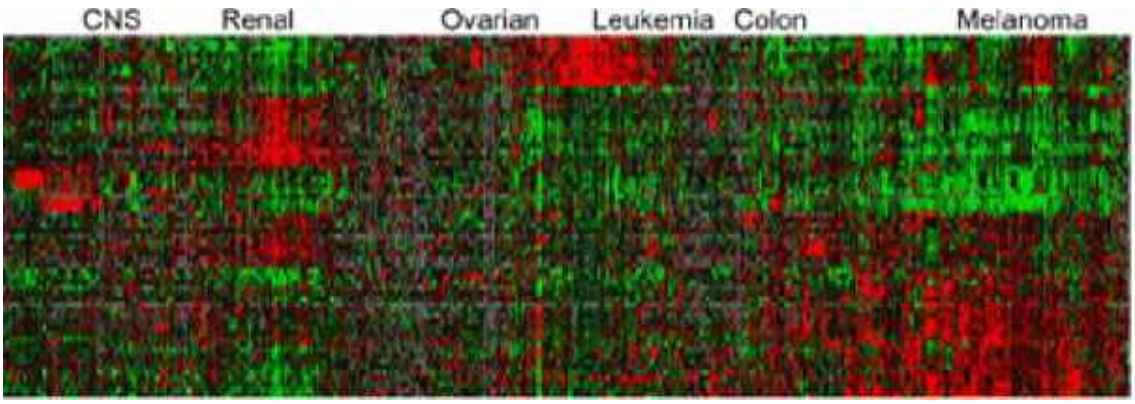

Figure 2: Gene expression data. Different rows and columns of microarrays correspond to different genes and cells. The value in each location indicates the level of expression (protein production) of a particular gene in cell. By testing for differences in the gene expression levels we can detect/classify different types of disease.

2 The Likelihood Ratio Test

Assume that we observe a random variable distributed according to one of two distributions.

$$\begin{aligned} H_0 : X &\sim p_0 \\ H_1 : X &\sim p_1 \end{aligned}$$

Deciding which of the two best fits an observation of X is called a *simple* binary hypothesis test, simple because the two distributions are known precisely (i.e., without unknown parameters or other uncertainties). A decision is made by partitioning the range of X into two disjoint regions. Let us denote the regions by R_0 and R_1. Let x denote the observed value of X. If $x \in R_i$, then we decide that H_i is the best match to the data; i.e., we decide that the data were distributed according to p_i. The key question is how to design the *decision regions* R_0 and R_1. Note that since $R_0 \bigcup R_1$ is assumed to be the entire range of X, R_1 is simply the complement of R_0, and so the choice of either region determines the other.

There are four possible outcomes in a test of this form, depending on the decision we make (H_0 or H_1), and the true distribution of the data (also H_0 or H_1). Let us denote these as $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$, where the first argument denotes the decision based on the regions R_0 and R_1 and the second denotes the true distribution that generated the data. Note that the outcomes $(0,1)$ and $(1,0)$ are mistakes or *errors*. The test made the wrontg decision about which distribution generated the data.

In order to optimize the choice of decision regions, we can specify a *cost* for incorrect (and correct, if we wish) decisions. Without loss of generality, let's assume the costs are non-negative. Let $c_{i,j}$ be the cost associated with outcome (i,j), $i,j \in \{0,1\}$. The costs reflect the relative importance of correct and incorrect decisions. Since our aim is to design a test that makes few mistakes, it is reasonable to assume that $c_{1,0}$ and $c_{0,1}$ are larger than $c_{0,0}$ and $c_{1,1}$; in fact often it is reasonable to assign a zero cost to correct decisions. The

costs $c_{0,1}$ and $c_{1,0}$ may be different. For example, it may be that one type of error is less desirable than the other.

The overall cost associated with a test (i.e., with decision regions R_0 and R_1) is usually called the *Bayes Cost*, and it is defined as follows.

$$C = \sum_{i,j=0}^{1} c_{i,j}\pi_j \, \mathbb{P}(\text{decide } H_i \,|\, H_j \text{ is true})$$

where $\pi_j := \mathbb{P}(H_j \text{ is true})$, $j = 0, 1$, is called the prior probability of H_j. This is just the probability that an observation will be generated according to p_j. The prior probabilities sum to 1, since we assume the data are generated either according to p_0 or p_1, but they need not be equal. One distribution may be more probable than the other (e.g., more people are healthy than have a disease). Our goal is to design the decision regions in order to minimize the Bayes Cost.

The Bayes Cost can be expressed directly in terms of the decision regions as follows. We will assume that p_0 and p_1 are continuous densities, but an analogous representation exists when they are discrete probability mass functions (i.e., replace integrals with sums in expressions below).

$$\begin{aligned} C &= \sum_{i,j=0}^{1} c_{i,j}\pi_j \, \mathbb{P}(\text{decide } H_i \,|\, H_j \text{ is true}) \\ &= \sum_{i,j=0}^{1} c_{i,j}\pi_j \, \mathbb{P}(X \in R_i \,|\, H_j \text{ is true}) \\ &= \sum_{i,j=0}^{1} c_{i,j}\pi_j \int_{R_i} p_j(x)\, dx \end{aligned}$$

The choice of R_0 and R_1 that minimizes the cost C becomes obvious if we expand the sum above.

$$\begin{aligned} C &= \sum_{i,j=0}^{1} c_{i,j}\pi_j \int_{R_i} p_j(x)\, dx \\ C &= \int_{R_0} (c_{0,0}\pi_0\, p_0(x) + c_{0,1}\pi_1\, p_1(x))\; dx \;+\; \int_{R_1} (c_{1,0}\pi_0\, p_0(x) + c_{1,1}\pi_1\, p_1(x))\; dx \end{aligned}$$

The integrands are non-negative, so it follows that we should let R_0 be the set of x for which the first integrand is smaller than the second. That is,

$$\begin{aligned} R_0 &:= \{x : c_{0,0}\pi_0\, p_0(x) + c_{0,1}\pi_1\, p_1(x) < c_{1,0}\pi_0\, p_0(x) + c_{1,1}\pi_1\, p_1(x)\} \\ R_1 &:= \{x : c_{0,0}\pi_0\, p_0(x) + c_{0,1}\pi_1\, p_1(x) > c_{1,0}\pi_0\, p_0(x) + c_{1,1}\pi_1\, p_1(x)\} \end{aligned}$$

Therefore, the optimal test (relative to the assigned costs) takes the following simple form:

$$\frac{p_1(x)}{p_0(x)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0(c_{1,0} - c_{0,0})}{\pi_1(c_{0,1} - c_{1,1})} \,.$$

Note that the term on the right hand side is a constant that depends on the prior probabilities and the costs (i.e., it does not depend on the data x). The term of the left hand side is a ratio of probability densities evaluated at x. The value of a probability density at the observed x is called the *likelihood* of x under that model. Thus, $\frac{p_1(x)}{p_0(x)}$ is called the *likelihood ratio* and the test is called the *likelihood ratio test* (LRT). No matter what the prior probabilities are or how costs are assigned, the optimal test *always* has takes the form

$$\frac{p_1(x)}{p_0(x)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \,,$$

where $\gamma > 0$ is the *threshold* of the test. We have shown is that the LRT, with an appropriate threshold, is optimal. This implies that the likelihood ratio is the minimal sufficient statistic for deciding between two simple hypotheses.