ECE 830 Fall 2010 Statistical Signal Processing

instructor: R. Nowak , **scribe:** Kayvan Samimi

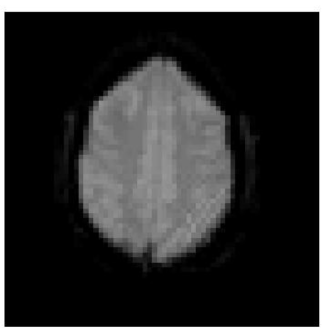
Lecture 5: Introduction to Detection Theory

The simplest form 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 $p(x|H_1)$. 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 a target is present. Let* $p(x|H_0)$ *denote the probability distribution of the signal when the target is absent (usually called the "null" distribution), and let* $p(x|H_1)$ *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 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* $\mu > 0$.



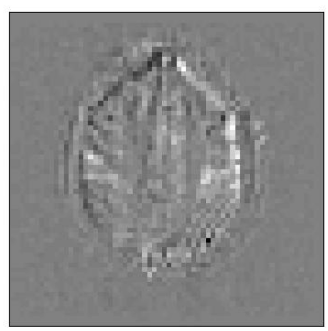


Figure 1: fMRI is used to map neural activity in the brain.

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*

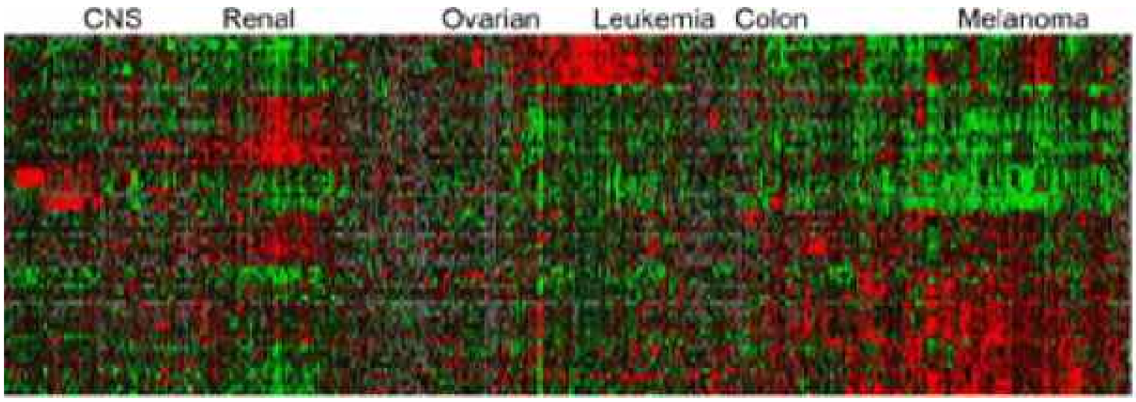


Figure 2: Gene expression data. Different rows and columns of microarrays correspond to different genes and cells. The value in each location indicates the level of expression (protein production) of a particular gene in cell. By testing for differences in the gene expression levels we can detect/classify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

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 $\mu \neq 0$.

Simple Binary Hypothesis Test

Outcomes of the random variable X are generated according to either $p(x|H_0)$ or $p(x|H_1)$. We need to design a test for x to make a decision. Suppose $x \in X, (e.g., X = \mathbb{R}^n)$, we define a decision boundary to partiton the space X into decision regions R_0 and R_1.

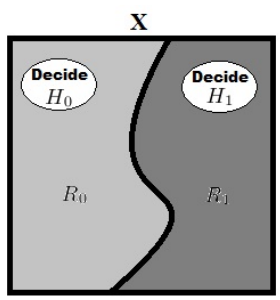


Figure 3: Decision boundary partitions the space into decision regions.

What are the possible outcomes of our testing? (Fig. 4)
Let's assign costs to each outcome: $c_{00}, c_{01}, c_{10}, c_{11}$. Assuming $P_0 = \mathbb{P}(H_0)$ and $P_1 = \mathbb{P}(H_1)$, the average cost will be

$$\begin{aligned}\bar{c} = \quad & c_{00}P_0\mathbb{P}(X \in R_0|H_0) + c_{01}P_1\mathbb{P}(X \in R_0|H_1) \\ & + c_{10}P_0\mathbb{P}(X \in R_1|H_0) + c_{11}P_1\mathbb{P}(X \in R_1|H_1)\end{aligned}$$

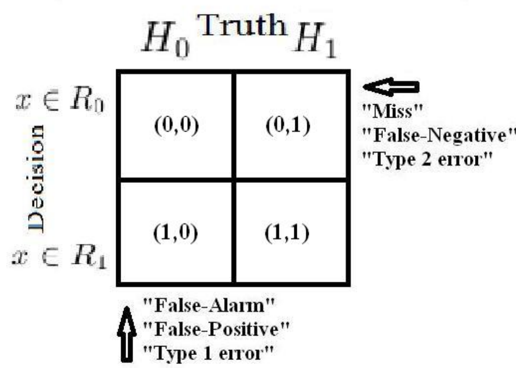


Figure 4: Possible outcomes of a decision-making test.

Given this choice of costs, choose regions R_0, R_1 to minimize $\bar{c}$. If we have probability densities $p(x|H_0)$ and $p(x|H_1)$ we can write:

$$\begin{aligned}\bar{c} = \quad & c_{00}P_0\int_{R_0} p(x|H_0)\mathrm{d}x + c_{01}P_1\int_{R_0} p(x|H_1)\mathrm{d}x \\ & + c_{10}P_0\int_{R_1} p(x|H_0)\mathrm{d}x + c_{11}P_1\int_{R_1} p(x|H_1)\mathrm{d}x \\ = \quad & \int_{R_0} \left[c_{00}P_0p(x|H_0) + c_{01}P_1p(x|H1)\right]\mathrm{d}x \\ & + \int_{R_1} \left[c_{10}P_0p(x|H_0) + c_{11}P_1p(x|H1)\right]\mathrm{d}x\end{aligned}$$

Since R_0 and R_1 form a partition of the space X, in order to minimize average cost, we define R_0 and R_1 in the following way

$$R_0 = \left\{x \in X : P_0c_{00}p(x|H_0) + P_1c_{01}p(x|H_1) < P_0c_{10}p(x|H_0) + P_1c_{11}p(x|H_1)\right\}$$
$$R_1 = \left\{x \in X : P_0c_{00}p(x|H_0) + P_1c_{01}p(x|H_1) > P_0c_{10}p(x|H_0) + P_1c_{11}p(x|H_1)\right\}$$

Therefore, for $x \in R_0$,

$$(P_0c_{00} - P_0c_{10})p(x|H_0) < (P_1c_{11} - P_1c_{01})p(x|H_1)$$

The Likelihood Ratio Test (LRT) for the detection problem is performed as follows:

$$\boxed{\frac{p(x|H_1)}{p(x|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0c_{00} - P_0c_{10}}{P_1c_{11} - P_1c_{01}}}$$