ELE 530 April 21, 2011
Theory of Detection and Estimation Handout #11

Homework #6
Due April 27

Notice: Please bring your completed homework with you to your appointment.

1. *Deciding between Laplace and Gauss.* Consider a simple binary decision between two zero mean, unit variance, distributions based on a single sample. The null hypothesis is the Gaussian distribution, and the alternative hypothesis is the Laplace distribution, given as follows:

$$\mathcal{H}_0 \;:\; p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$
$$\mathcal{H}_1 \;:\; p(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}.$$

a. What is the form of the critical (rejection) regions for Neyman-Pearson detectors?

b. What is the ROC for this detection problem? This is characterized most simply by expressing α as a function of β for the Pereto optimal performance of detection. The hyperbolic sin function might come in handy:

$$\sinh(x) = \frac{e^x - e^{-x}}{2}.$$

c. Now find the minimax performance. (Just kidding, don't do it. But how would you if you cared enough?)

d. What is the optimal decision rule to minimize the Bayes probability of error if the null hypothesis (Gaussian) is π times more likely than the alternative hypothesis (Laplace)?

e. Would an optimal decision to minimize the Bayes probability of error ever ignore the data X? If so, what would the prior distribution have to be and what does this say about the shape of the ROC?

f. What is the minimum Bayes probability of error if the two hypotheses are equally likely?

g. Finally, suppose that you get N independent samples from which to make a decision of whether they all came from the Gaussian distribution or the Laplace distribution. Notice that the first, second, and third moments of the two distributions are the same, but the fourth moment (kurtosis) is larger for the Laplace distribution. One method to test between the two distributions is to measure a sample average of the fourth moment, given as:

$$T(X) = \frac{1}{N} \sum_{i=1}^{N} X_i^4.$$

Compare this to the test statistic that would be used for an optimal detector.

2. *Maximum Likelihood.* (From Poor.II.21): Suppose that $X_1, X_2, ..., X_n$ is a sequence of random observations, each taking the values 0 and 1 with probabilities 1/2. Consider the following two hypotheses concerning $X_1, X_2, ..., X_n$:

$$\begin{aligned} \mathcal{H}_0 &: X_1, X_2, ..., X_n \text{ are independent}, \\ \mathcal{H}_1 &: p(x_k|x_1, x_2, ..., x_{k-1}) = \begin{cases} 3/4 & \text{if } x_k = x_{k-1} \\ 1/4 & \text{if } x_k \neq x_{k-1} \end{cases} \quad \text{for } k = 2, 3, ..., n, \end{aligned}$$

where $p(x_k|x_1, x_2, ..., x_{k-1})$ denotes the conditional probability that $X_k = x_k$ given that $X_1 = x_1, X_2 = x_2, ..., X_{k-1} = x_{k-1}$. Find the Bayes decision rule for testing $\mathcal{H}_0$ versus $\mathcal{H}_1$ under the assumption of uniform costs and equal priors.

3. *Asymptotics of the Gaussian location test.* Consider a simple binary decision between two unit variance Gaussian distributions, one with zero mean, and the other with mean A, based on n i.i.d. samples.

$$\begin{aligned} \mathcal{H}_0 &: p_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \\ \mathcal{H}_1 &: p_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-A)^2}{2}}. \end{aligned}$$

   a. What decision rule gives the minimax probability of error, and what is the probability of error? What is the first-order term in the exponent of the minimax probability of error? (This can be done by using an approximation to the Q function for large arguments.)
   b. What is the best error exponent you can achieve for β while keeping α arbitrarily small? Please use a similar approach to the previous step for calculating β (don't use the KL-divergence method).
   c. What is $D(p_0||p_1)$?

4. *Asymptotics of the uniform location test.* Similarly to the previous problem, we will consider a simple binary decision between two unit variance distributions, one with zero mean, and the other with mean A, based on n i.i.d. samples. However, this time the distributions are uniform.

$$\begin{aligned} \mathcal{H}_0 &: X \sim Unif[-\sqrt{3}, \sqrt{3}], \\ \mathcal{H}_1 &: X \sim Unif[A - \sqrt{3}, A + \sqrt{3}]. \end{aligned}$$

   a. What decision rule gives the minimax probability of error, and what is the probability of error?
   b. What is the best error exponent you can achieve for β while keeping $\alpha < 1/2$?
   c. What is $D(p_0||p_1)$?
   d. For any n, there are only two optimal deterministic decision rules. What are they, and how does the ROC look?
   e. For any n, it is not possible to achieve $\alpha = 0$ and $\beta = 0$ simultaneously. However, if we use a sequential decision rule then we can achieve zero error. What is the expected number of samples taken (under each hypothesis) for the optimal sequential decision

rule that achieves zero error?

f. Notice that there are only three sequential probability ratio tests (SPRT) for this setting. What are they, and how do they perform in terms of α, β, and the expected number of samples taken? (Optional) Can you verify that the performance of the optimal minimax decision for deterministic stopping time n (from the first part of this problem) does not outperform the convex hull of these three SPRT's?

5. *Difference of error exponents.* In the discussion of error exponents for simple binary hypothesis tests when given n samples (Sanov's Theorem, Chernoff information, etc.), we discussed that if we let the threshold τ for a log-likelihood ratio test grow linearly with n, $\tau = n\bar{\tau}$, then the exponents for α and β will be different. Furthermore, for each $\bar{\tau}$ there is an associated $\gamma \in [0, 1]$ for which the error exponents satisfy

$$\begin{aligned} \alpha &\doteq e^{-nD(p_\gamma || p_0)}, \\ \beta &\doteq e^{-nD(p_\gamma || p_1)}, \end{aligned}$$

and p_γ has the form

$$p_\gamma = \frac{p_0^\gamma p_1^{1-\gamma}}{Z_\gamma},$$

where Z_γ is a normalizing constant. We also know that error events, though rare, will look empirically like samples from p_γ.

The relationship between $\bar{\tau}$ and γ is not straightforward but is important. From first principles we can deduce that samples from p_γ should straddle the decision threshold. In other words,

$$\mathbf{E}_{p_\gamma} l(X) = \bar{\tau}.$$

a. Using the facts above, derive the difference between the error exponents for α and β as a function of the threshold slope $\bar{\tau}$. We've come across some special cases of this already.

b. Consider a hypothesis test between two Laplace distributions.

$$\begin{aligned} p_0(x) &= \frac{1}{2} e^{-|x|}, \\ p_1(x) &= \frac{1}{2} e^{-|x-1|}. \end{aligned}$$

For large n and an arbitrary $\bar{\tau}$, when a decision error occurs, what will the data look like? That is, specify a distribution that the data will appear to have been sampled from using an arbitrary γ, and don't bother to relate γ to $\bar{\tau}$ explicitly. Sketch the distribution for a value of $\gamma \in (0, 1)$.

6. *Wald's Approximation.* Here we will analyze a sequential test for deciding whether a coin is biased toward heads or tails. The two hypotheses are:

$$\mathcal{H}_0 \; : \; X_i \; i.i.d. \sim p_0(x) = \begin{cases} p, & x = -1 \text{ (tails)}, \\ 1-p, & x = 1 \text{ (heads)}, \end{cases}$$
$$\mathcal{H}_1 \; : \; X_i \; i.i.d. \sim p_1(x) = \begin{cases} 1-p, & x = -1 \text{ (tails)}, \\ p, & x = 1 \text{ (heads)}, \end{cases}$$

where $p > 1/2$ is a known parameter.

a. Consider the sequential decision rule that compares the number of heads to the number of tails (where $X = 1$ represents heads and $X = 0$ represents tails) and stops when the number of heads gets ahead by k_1 or the number of tails get ahead by k_0. In other words, the decision rule stops and decides $\mathcal{H}_0$ when $\sum_{i=1}^{k} X_i \leq -k_0$, or it stops and decides $\mathcal{H}_1$ when $\sum_{i=1}^{k} X_i \geq k_1$, whichever comes first. Show that this is a sequential probability ratio test, and calculate α and β (the probabilities or error under each hypothesis). (Wald's approximations are found on page 104 of Poor.)

b. For the decision rule in part (a), we can calculate the expected number of samples taken using Martingales. The quantity $Z_k = \sum_{i=1}^{k}(X_i - \mathbf{E}X_i))$ is a Martingale. The optional stopping theorem for Martingales tells us that when the decision is made at time N, the expected value of the stopped Martingale Z_N is 0. From that we derive:

$$\mathbf{E}_{\mathcal{H}_0} N = \frac{1}{1-2p} \mathbf{E}_{\mathcal{H}_0} \sum_{i=1}^{N} X_i.$$
$$\mathbf{E}_{\mathcal{H}_1} N = \frac{1}{2p-1} \mathbf{E}_{\mathcal{H}_1} \sum_{i=1}^{N} X_i.$$

The right-hand-side of both equations above are easy to calculate in terms of α and β. What are $\mathbf{E}_{\mathcal{H}_0} N$ and $\mathbf{E}_{\mathcal{H}_1} N$?

c. If we assume that the errors are small (α and β are small which means that k_0 and k_1 are large), then the expressions of parts (a) and (b) simplify. Using these simplified expressions, derive the exponent for the probabilities of error, to first order with respect to the expected number of samples. In other words:

$$\alpha \doteq e^{-C_0 \mathbf{E}_{\mathcal{H}_1} N},$$
$$\beta \doteq e^{-C_1 \mathbf{E}_{\mathcal{H}_0} N}.$$

Find C_0 and C_1.

d. How do these compare to the error exponents when the decision is based on a predetermined fixed number of samples n? It will help to express C_0 and C_1 in terms of $D(p_0||p_1)$.