Boston University

Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Problem Set No. 7

Fall 2016

Issued: Wednesday, Nov. 2, 2016 **Due:** Wednesday, Nov. 9, 2016

Problem 7.1

Consider the binary discrete-time communication system shown in Fig. 3. Assume that $M = 0$ and $M = 1$ are equally likely to occur. Under hypothesis H_M ($M = 0, 1$), the received signal $R(n)$ is given by

$$R(n) = y_M(n) + W(n) = s_M(n) * h(n) + W(n), \qquad \text{for all } n,$$

where "$*$" denotes convolution, and where the $W(n)$'s are zero-mean statistically independent Gaussian random variables with variance σ^2. The two signals, $s_0(n)$ and $s_1(n)$ are shown in Figs. 4 and 5, respectively. The parameter λ in Fig. 5 satisfies $0 \le \lambda \le 1$. The impulse response of the linear time invariant filter $h(n)$ is shown in Fig. 6.

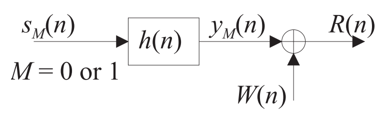


Figure 3

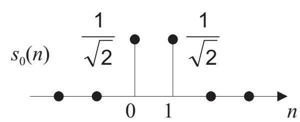


Figure 4

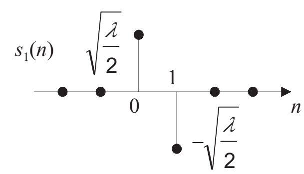


Figure 5

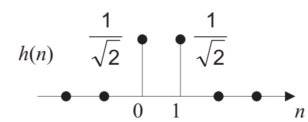


Figure 6

We wish to obtain a rule for making a decision about which hypothesis was used, based on the received sequence $R(n)$.

(a) Which samples of $R(n)$ provide information in making the decision? Justify your reasoning.

(b) Find the minimum probability of error decision rule, based on observation of $R(n)$. Simplify your processor as much as possible to minimize computation.

(c) Obtain an expression for the probability of error, $\Pr(\varepsilon)$, in terms of λ and $Q(\cdot)$, where:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\alpha^2/2} d\alpha \ .$$

Hint: Think in terms of values of the test statistic.

(d) Find the values of λ ($0 \le \lambda \le 1$) that minimize $\Pr(\varepsilon)$ for the detector obtained in part (c).

Problem 7.2

Consider a binary hypothesis testing problem where the density of the observation Y under each hypothesis is as given in Figure 7 for some a and b.

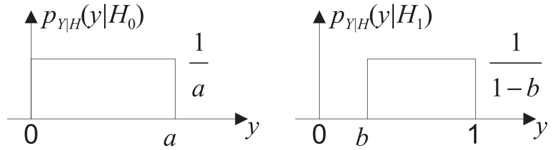


Figure 7

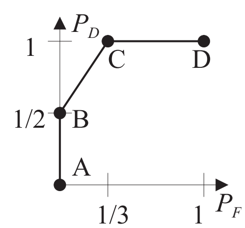


Figure 8

Figure 8 shows the possible (P_D, P_F) pairs as the threshold γ is varied from $-\infty$ to $+\infty$ (i.e. the ROC) for a decision rule (not necessarily the likelihood ratio test) of the form: $y \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$.

(a) What point on the ROC corresponds to $\gamma = b$? What point corresponds to $\gamma = a$? Explain.

(b) Find the values of a and b.

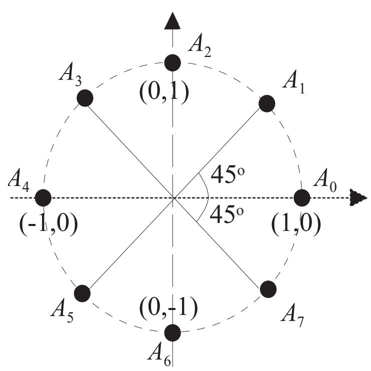
(c) If the hypotheses are equally likely a priori, what (P_D, P_F) pair on the ROC corresponds to the γ with minimum probability of error? What is the corresponding probability of error at this point?

(d) For this part consider the likelihood ratio test for this problem: $\mathcal{L}(y) = \dfrac{p_{Y|H}(y \mid H_1)}{p_{Y|H}(y \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$. Specify a value of η so that $P_D = 1$ and $P_F = 1/3$. Is this value of η unique.

Problem 7.3 (Shanmugan and Breipohl 6.19)
Consider an M-ary detection problem, where under hypothesis H_j the observation is:

$$H_j : \quad \underline{Y} = \underline{A}_j + \underline{W}, \quad j = 0, 1, \ldots, 7$$

where $\underline{A}_j$ is the vector signal value associated with the j-th hypothesis and $\underline{W} \sim N(\underline{0}, \Sigma_w)$ with $\Sigma_w = 0.1I$. The signal values $\underline{A}_j$ are shown in the figure below and it is known that $\Pr(H_j) = P_j = 1/8$.



(a) Find the decision boundaries in the observation space that lead to a minimum $\Pr[error]$.

(b) Explain graphically how you would calculate the probability of error using a figure like that above. Is this an easy calculation to make?

Problem 7.4 (Shanmugan and Breipohl 6.14)
The signaling waveforms used in a binary communications system are given by:

$$\begin{aligned} s_0(t) &= -4\sin(2\pi f_0 t), \qquad 0 \le t \le T, \quad T = 1\text{ms} \\ s_1(t) &= 4\sin(2\pi f_0 t), \qquad 0 \le t \le T \end{aligned}$$

where T is the duration of the signal and $f_0 = 10/T$. The observations under each hypothesis are given by:

$$\begin{aligned} H_0 : \quad y(t) &= s_0(t) + w(t) \\ H_1 : \quad y(t) &= s_1(t) + w(t) \end{aligned}$$

where $w(t)$ is zero mean white Gaussian noise with $S_{ww}(f) = 10^{-3}$W/Hz and $P_0 = P_1 = 1/2$.

(a) Find the decision that minimizes the probability of error $\Pr(\varepsilon)$.

(b) Find the corresponding $\Pr(\varepsilon)$.

(c) How would $\Pr(\varepsilon)$ change if the following signal set was used instead:

$$\begin{aligned} s_0(t) &= -\sqrt{8}, \quad 0 \le t \le T, \quad T = 1\text{ms} \\ s_1(t) &= \sqrt{8}, \quad 0 \le t \le T \end{aligned}$$

Note that the signal sets of (a)-(b) and (c) have the same energy, thus it is reasonable to ask which is better to use. You should find that the probability of error only depends on the energy in the difference signal when the signals all have equal energy.

Problem 7.5 (Old Exam Question)
Consider the following binary hypothesis testing problem on $0 \le t \le T$:

$$\begin{aligned} H_0 : y(t) &= v\, s(t) + w(t) \\ H_1 : y(t) &= (6 + v)\, s(t) + w(t) \end{aligned}$$

where $s(t)$ is a given deterministic waveform with $\int_0^T s^2(t)\, dt = 1$, $w(t)$ is a zero-mean, white Gaussian noise process with $R_{WW}(\tau) = \delta(\tau)$, and v is a zero-mean Gaussian random variable, independent of $w(t)$, with variance $E[v^2] = 1$. Suppose that the two hypothesis are a priori equally likely. We desire the minimum probability of error decision rule for deciding between H_0 and H_1 at $t = T$.

(a) What is the optimal threshold for this detection problem?

(b) What are a good choice of basis vectors for a Karhunen-Loéve expansion of $y(t)$ for this problem? What are the corresponding expansion coefficients for $y(t)$ with respect to this basis and their distributions under each hypothesis? Hint: They are Gaussian.

(c) Determine the minimum probability of error decision rule for this problem, i.e., specify the required processing of $y(t)$ and the subsequent threshold test.

(d) Determine the probability of error for the decision rule of part (c) in terms of $Q(\cdot)$. Recall: $1 - Q(-x) = Q(x)$

Computer Problems

Problem 7.6 Finding the ROC by Experiment

The aims of this laboratory are to learn how to evaluate the performance of decision rules through experimentally determined receiver operating characteristics,

In class we have focused on analytical expressions for detectors and the corresponding analytical expressions for P_D, P_F used in the receiver operating characteristic (ROC). The P_D vs P_F plot forming the basis of the ROC is really much more powerful and universally applicable than this might lead you to believe. In particular, one of the reasons that that ROC is so widely used for performance analysis of binary detection problems is that it may be used in situations where no analytic problem statement or P_D/P_F expressions exist. The aim of this computer exercise is to learn how to evaluate the performance of decision rules through experimentally determined receiver operating characteristics.

Recall that the ROC is defined as a plot of $P_D = \Pr(\text{Choose } H_1 \mid H_1 \text{ True})$ vs $P_F = \Pr(\text{Choose } H_1 \mid H_0 \text{ True})$ for a given decision rule as a threshold Γ is varied. Certainly one way of generating these values is to combine these definitions with explicit expressions for the probability density functions and the decision rule involved to calculate *analytic formulas* for the probabilities involved. This approach, which we take in class, is based on knowledge of the underlying models and can be done in the absence of data. For many problems, however, either the underlying pdf expressions are complicated (or even unknown) or the detector structure itself is complicated (or unknown). Sometimes this is because the detector is actually the result of many smaller steps linked together, or sometimes this may just reflect the commercial fact that a company wants to hide its secret algorithm. In this problem, we learn how to estimate the ROC directly from data.

First, suppose we have a decision rule $\mathcal{D}(y,\Gamma)$ (not necessarily optimal) which takes as its inputs an observation y and a threshold value Γ and returns a decision H_0 or H_1. We can consider this decision rule to be a "black box" which produces an output for given inputs. We may have no knowledge of its internal structure, but can apply it to data – it is like a piece of compiled code. Now, instead of probabilistic observation *models* $p(y \mid H_i)$, suppose that we have available training data for which we know ground truth (i.e. we know which hypothesis corresponds to each observation). Then we may split this data into two sets: data set $\{y_i\}_0$ corresponds to data for which H_0 is true and data set $\{y_i\}_1$ corresponds to data for which H_1 is true. For a given value of Γ we may experimentally estimate $P_D(\Gamma)$ by actually applying the decision rule $\mathcal{D}(y,\Gamma)$ to the $\{y_i\}_1$ data and calculating the fraction of the $\{y_i\}_1$ data that yields an H_1 decision. For example, if there are 100 $\{y_i\}_1$ samples and the decision rule picks H_1 for 1/4 of them, then we can say that $P_D = .25$ for this decision rule at this value of Γ. We repeat this experiment for each value of Γ in the range of interest to obtain $P_D(\Gamma)$. Similarly, we may repeat this experiment on the $\{y_i\}_0$ data to obtain an estimate of $P_F(\Gamma)$: for each value of Γ the fraction of the $\{y_i\}_0$ data that produces an H_1 decision gives an estimate of $P_F(\Gamma)$.

In this way we may obtain an ROC without the need for detailed, analytic observation densities. Further, we may obtain an ROC for any decision rule $\mathcal{D}(y,\Gamma)$ depending on a threshold Γ without detailed information about the rule's structure. This approach is often how decision rules are compared in practice. For example, in the realm of automatic target recognition, the government will commission the creation of just such a "ground truth" data set, which is then used to evaluate competitors target recognition algorithms. An example of such data can be found at `http://www.mbvlab.wpafb.af.mil/public/sdms/datasets/index.htm`. Another example can be found in the in the realm of speech processing, where large "corpora" of ground-truth speech data are created for algorithm development and testing.

We will use the above idea as motivation to write a MATLAB program to estimate $P_D(\Gamma)$ and $P_F(\Gamma)$ for a given "black box" decision rule and range of Γ based on sets of training data. First, we will need to assume a standard detector interface. For the purposes of this laboratory we will assume that a detector is a MATLAB function with the following calling sequence: `D = detfunname(y,Gamma)`, where the detector function name is `detfunname.m`, its first input is a matrix of data `y` with each row a different (possibly vector) experimental observation and its second input is a vector of thresholds `Gamma`. Its output will be assumed to be a matrix `D` whose ij-th entry is a `0` or `1` corresponding to the decision associated to data point `y(i,:)` at threshold `Gamma(j)`. Note in particular that the result of evaluating the function on a single observation at a single threshold is the corresponding decision of the given decision rule.

(a) Given the above MATLAB detector format you will write a MATLAB function `roc.m` to estimate $P_D(\Gamma)$ and $P_F(\Gamma)$ for any detector. The function will generate decisions for validated test data at a variety of thresholds and use the fractional outcomes to estimate the detection and false alarm probabilities. The function call of your program will be the following: `[pd,pf] = roc(detfunname,Gamma,y0,y1)`, where `detfunname` is *a string* containing the name of the detection function under investigation[1] which obeys the above standard format, `Gamma` is a vector of threshold values, `y0` is a vector of data observations obtained when the H_0 hypothesis is true, and `y1` is a vector of data observations obtained when the H_1 hypothesis is true. It produces as output the two vectors `pd`, `pf` containing the estimated P_D,P_F values for each value in `gamma`. Each step below will form a line of the program:

(i) The first step is to call the detector function for each of the data sets `y0` and `y1` and all the values in `Gamma` to produce experimental decisions. The decisions for the `y0` data are stored in `D0` while the decisions for the `y1` data are stored in `D1`. Recall that `detfunname` is a string containing the name of the detection function.

```
eval(['D0 = ',detfunname,'(y0,Gamma);'])
eval(['D1 = ',detfunname,'(y1,Gamma);'])
```

[1]For example, if the detector was the m-file `foo.m`, the function call would be `roc('foo',Gamma,y0,y1)`.

(ii) Given the set of decisions in `D0` and `D1` for each ground truth, we can now estimate `pd` and `pf` as the fraction of H_1 decisions under each hypothesis (i.e. 1's in `D0` or `D1`). Since we are just counting 1's for each choice of Γ, we can do this easily in MATLAB.

```
pf = sum(D0)/length(y0);
pd = sum(D1)/length(y1);
```

Add these steps together to create your program. We now have a way to find the ROC for any detection problem and any decision rule given ground truth data and a black box implementation of the rule.

(b) Consider the following detection problem:

Problem A: Detect the presence of a constant in Gaussian noise:

$$\begin{array}{lcll} H_0 : y & = & w, & w \sim N(0,1) \\ H_1 : y & = & m + w, & w \sim N(0,1) \end{array}$$

where $m = 2$. Generate $N = 5000$ data samples `y0` under hypothesis H_0 and $N = 5000$ data samples `y1` under hypothesis H_1. These are our ground truth data.

On the class web site there are 4 detectors for scalar data: `det1.m`, `det2.m`, `det3.m`, and `det4.m`. The detector `det1.m` just compares the observation to a threshold, the detector `det2.m` compares the *power* in the observation to a threshold, the detector `det3.m` compares `tan(y)` to a threshold, and the last detector `det4.m` does not use the observation, but rather compares a randomly chosen number to the threshold.

Find the ROC for each of these four detectors for this problem. Plot these ROCs on a common axis. Which detector corresponds to the optimal likelihood ratio test for this problem? This detector should have the highest P_D for a given P_F of all the detectors. Do your plots reflect this?

(c) Now consider the following detection problem:

Problem B: Decide which of two zero-mean Gaussians an observation comes from:

$$\begin{array}{lcll} H_0 : y & = & w_0, & w_0 \sim N(0,1) \\ H_1 : y & = & w_1, & w_1 \sim N(0,4) \end{array}$$

Generate $N = 5000$ data samples `y0` under hypothesis H_0 and $N = 5000$ data samples `y1` under hypothesis H_1. These are our ground truth data.

Find the ROC for each of the four detectors for this problem. Plot these ROCs on a common axis. Which detector corresponds to the optimal likelihood ratio test for this problem? Do your plots reflect this?

(d) ROCs are useful not just for comparing different detectors for a given problem, but also for understanding the effect of problem changes on performance of a given detector structure. To this end, let us focus on the amplitude detector `det1.m` and examine what happens as we change elements of the problem. Compare the ROCs for this detector applied to Problem A as the mean m (i.e. the constant) under H_1 is varied from 0 to 4. From your graph, how large does the difference in means have to be to achieve $P_D \approx .9$ at a $P_F = 0.2$? What happens to the ROC as the noise level is varied from 0 to 4?

Suppose we now want to understand what happens to the performance of our detector `det1.m` if we are wrong about the noise model. In particular, suppose that instead of additive Gaussian noise we believe the noise is additive Cauchy noise. You will find a MATLAB function `randcau.m` on the web site to generate such noise. Plot the corresponding ROC for the Cauchy noise case as m is varied from 0 to 4. How large does m have to be to achieve $P_D \approx .9$ at a $P_F = 0.2$ in the presence of Cauchy noise?