Boston University

Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Problem Set No. 6 Solutions

Fall 2016

Issued: Wednesday, Oct. 26, 2016 **Due:** Wednesday, Nov. 2, 2016

Problem 6.1

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 $0 \leq \lambda \leq 1$. The impulse response of the linear time invariant filter $h(n)$ is shown in Fig. 6.

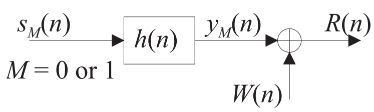


Figure 3

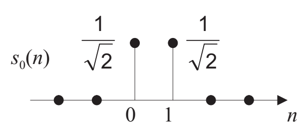


Figure 4

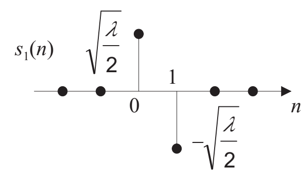


Figure 5

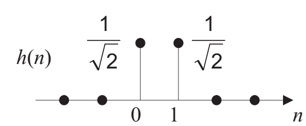


Figure 6

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 $(0 \leq \lambda \leq 1)$ that minimize $\Pr(\varepsilon)$ for the detector obtained in part (c).

Solution:

(a) First note that:

$$\begin{aligned} y_0(n) &= s_0(n) * h(n) = \frac{1}{2}\delta(n) + \delta(n-1) + \frac{1}{2}\delta(n-2) \\ y_1(n) &= s_1(n) * h(n) = \frac{\sqrt{\lambda}}{2}\delta(n) - \frac{\sqrt{\lambda}}{2}\delta(n-2) \end{aligned}$$

As $y_0(n) = y_1(n) = 0$ for $n \neq 0, 1, 2$ it follows that $R(n) = W(n)$ for $n \neq 0, 1, 2$ under either H_0 or H_1. Hence the useful samples of $R(n)$ are at $n = 0, 1, 2$. Also note that if $\lambda = 1$ the first two samples of y become the same under either hypothesis, and we would expect this sample to also provide no useful information. Thus whatever test statistic we arrive at we expect it to reflect this behavior, as we will see is the case.

(b) Define:

$$\underline{R} = \begin{bmatrix} R(0) \\ R(1) \\ R(2) \end{bmatrix} \qquad \underline{W} = \begin{bmatrix} W(0) \\ W(1) \\ W(2) \end{bmatrix} \qquad \underline{y}_0 = \begin{bmatrix} 1/2 \\ 1 \\ 1/2 \end{bmatrix} \qquad \underline{y}_1 = \begin{bmatrix} \sqrt{\lambda}/2 \\ 0 \\ -\sqrt{\lambda}/2 \end{bmatrix}$$

Now since $W(n)$ is white Gaussian noise, our observation vector $\underline{R}$ is Gaussian under each of the two hypotheses, as follows:

$$\begin{aligned} P_{\underline{R}|H}(\underline{r} \mid H_0) &= N(\underline{r}; \underline{y}_0, \sigma^2 I) \\ P_{\underline{R}|H}(\underline{r} \mid H_1) &= N(\underline{r}; \underline{y}_1, \sigma^2 I) \end{aligned}$$

Thus the MPE decision rule is given by:

$$\begin{aligned} \mathcal{L}(\underline{r}) = \frac{P_{\underline{R}|H}(\underline{r} \mid H_1)}{P_{\underline{R}|H}(\underline{r} \mid H_0)} &\underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0}{P_1} = 1 \\ \Longrightarrow \quad \exp\left[-\frac{(\underline{r}-\underline{y}_1)^T(\underline{r}-\underline{y}_1)}{2\sigma^2} + \frac{(\underline{r}-\underline{y}_0)^T(\underline{r}-\underline{y}_0)}{2\sigma^2}\right] &\underset{H_0}{\overset{H_1}{\gtrless}} 1 \\ \Longrightarrow \quad -\underline{r}^T\underline{r} + 2\underline{y}_1^T\underline{r} - \underline{y}_1^T\underline{y}_1 + \underline{r}^T\underline{r} - 2\underline{y}_0^T\underline{r} + \underline{y}_0^T\underline{y}_0 &\underset{H_0}{\overset{H_1}{\gtrless}} 0 \\ \Longrightarrow \quad \ell(\underline{r}) = \left(\underline{y}_1^T - \underline{y}_0^T\right)\underline{r} &\underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{2}\left(\underline{y}_1^T\underline{y}_1 - \underline{y}_0^T\underline{y}_0\right) = \frac{\lambda-3}{4} \\ \Longrightarrow \quad \begin{bmatrix} \sqrt{\lambda}-1 & -2 & -\sqrt{\lambda}-1 \end{bmatrix}\underline{r} &\underset{H_0}{\overset{H_1}{\gtrless}} \frac{\lambda-3}{2} \end{aligned}$$

Note that the vector on the left is composed of all negative entries, so we could multiple through by -1, but I won't bother.

(c) One way to solve this is to reflect the decision regions back into the data space – i.e. back into the space of $[R(0), R(1), R(2)]$ – and then integrate the 3-D vector Gaussian over the proper regions. The easier way to solve the problem in this case is to find the density of the test statistic and calculate P_D and P_F in the space of the test statistic. First note that:

$$\begin{gathered} \text{Under } H_0\text{: } \underline{R} \sim N(\underline{y}_0, \sigma^2 I) \\ \text{Under } H_1\text{: } \underline{R} \sim N(\underline{y}_1, \sigma^2 I) \end{gathered}$$

If $\ell(\underline{r}) \equiv \left(\underline{y}_1 - \underline{y}_0\right)^T \underline{r}$ then the test statistic $\ell(\underline{r})$ is Gaussian, and therefore completely characterized by its mean and variance. In particular, Under H_0 we have $\ell \sim N(\ell_0, \sigma_0^2)$ where:

$$\begin{aligned} \ell_0 &= E[\ell \mid H_0] = E\left[\left(\underline{y}_1 - \underline{y}_0\right)^T(\underline{y}_0 + \underline{w})\right] = \left(\underline{y}_1 - \underline{y}_0\right)^T \underline{y}_0 = -3/2 \\ \sigma_0^2 &= E[\ell\ell^T \mid H_0] - E[\ell \mid H_0]E[\ell \mid H_0]^T \\ &= E\left[\left(\underline{y}_1 - \underline{y}_0\right)^T(\underline{y}_0 + \underline{w})(\underline{y}_0 + \underline{w})^T\left(\underline{y}_1 - \underline{y}_0\right)\right] - \left[\left(\underline{y}_1 - \underline{y}_0\right)^T \underline{y}_0\underline{y}_0^T\left(\underline{y}_1 - \underline{y}_0\right)\right] \\ &= \left(\underline{y}_1 - \underline{y}_0\right)^T E\left[\underline{w}\underline{w}^T\right]\left(\underline{y}_1 - \underline{y}_0\right) = \left(\underline{y}_1 - \underline{y}_0\right)^T \sigma^2 I\left(\underline{y}_1 - \underline{y}_0\right) = \sigma^2\frac{\lambda+3}{2} \end{aligned}$$

Similarly, under H_1 we have that $\ell \sim N(\ell_1, \sigma_1^2)$ where a similar set of calculations provides:

$$\begin{aligned}\ell_1 &= \frac{\lambda}{2}\\ \sigma_0^2 &= \sigma^2\frac{\lambda+3}{2}\end{aligned}$$

Notice the means are different but the variances are the same. Now it is simple to calculate P_F and P_M using the expression for the decision rule above in terms of $\ell(\underline{r})$:

$$\begin{aligned}P_F &= \int_{\frac{\lambda-3}{4}}^{\infty} P(\ell \mid H_0)\, d\ell = \int_{\frac{\lambda-3}{4}}^{\infty} N\left(\ell; \frac{-3}{2}, \sigma^2\frac{\lambda+3}{2}\right) = Q\left(\frac{\sqrt{\lambda+3}}{\sigma 2\sqrt{2}}\right)\\ P_M &= \int_{-\infty}^{\frac{\lambda-3}{4}} P(\ell \mid H_1)\, d\ell = \int_{-\infty}^{\frac{\lambda-3}{4}} N\left(\ell; \frac{\lambda}{2}, \sigma^2\frac{\lambda+3}{2}\right) = 1 - Q\left(-\frac{\sqrt{\lambda+3}}{\sigma 2\sqrt{2}}\right) = Q\left(\frac{\sqrt{\lambda+3}}{\sigma 2\sqrt{2}}\right)\end{aligned}$$

Thus the $\Pr(\varepsilon)$ is given by:

$$\Pr(\varepsilon) = P_0 P_F + P_1 P_M = (1/2 + 1/2) Q\left(\frac{\sqrt{\lambda+3}}{\sigma 2\sqrt{2}}\right) = Q\left(\frac{\sqrt{\lambda+3}}{\sigma 2\sqrt{2}}\right)$$

(d) Since $Q(x)$ is a monotonically decreasing function of x, $\Pr(\varepsilon)$ is minimized when $\frac{\sqrt{\lambda+3}}{\sigma 2\sqrt{2}}$ is maximum. Given that $0 \le \lambda \le 1$, this occurs at $\lambda = 1$ and the corresponding error is given by $\Pr(\varepsilon)_{\min} = Q\left(\frac{1}{\sigma\sqrt{2}}\right)$.

Problem 6.2

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.

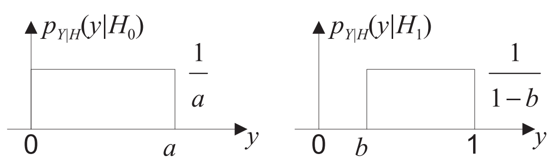


Figure 7

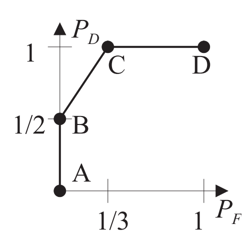


Figure 8

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.

Solution:

(a) From the shape of the ROC we know that $b < a$, $a, b > 0$ and $a, b < 1$. This follows from the fact that as γ is varied the ROC has a region where P_F decreases while $P_D = 1$, then has a region where both P_D and P_F decrease, then a region where $P_F = 0$ while P_D decreases. In particular, note that the two densities must overlap ($a > b$) or else we could achieve the point ($P_D = 1, P_F = 0$) in the upper right corner.

Thus when $\gamma = b$ $P_D = 1$ and P_F is something less than 1. This must be point C. Similarly, when $\gamma = a$, $P_F = 0$ and P_D is something less than 1. This must be point B. Thus we have:

$$\begin{array}{ll} \gamma = b \longrightarrow & \text{Point C} \\ \gamma = a \longrightarrow & \text{Point B} \end{array}$$

(b)

$$\begin{array}{ll} \text{At Point C} & \Longrightarrow \gamma = b \Longrightarrow P_F = \dfrac{1}{3} = \dfrac{a-b}{a} \Longrightarrow b = \dfrac{2}{3}a \\ \text{At Point B} & \Longrightarrow \gamma = a \Longrightarrow P_D = \dfrac{1}{2} = \dfrac{1-a}{1-b} \Longrightarrow 1 - \dfrac{2}{3}a = 2 - 2a \Longrightarrow \boxed{a = \dfrac{3}{4}} \Longrightarrow \boxed{b = \dfrac{1}{2}} \end{array}$$

where we have substituted the expression for b into the last expression to find a.

(c) $\Pr(\text{error}) = P_0 P_F + P_1(1 - P_D) = \frac{1}{2}(P_F + 1 - P_D)$. The point with the highest P_D and smallest P_F has the minimum probability of error. Clearly this must be on the line connecting B and C. On this line we have $P_D = \frac{1}{2} + \frac{3}{2}P_F$. Thus $\Pr(\text{error}) = \frac{1}{2}\left(\frac{2}{3} - \frac{1}{3}P_D\right)$. So we want P_D as large as possible. This means that it must be $\boxed{\text{point C}}$ that corresponds to the minimum probability of error. The probability of error at point C is $\boxed{\Pr(\text{error}) = \frac{1}{6}}$

(d) This (P_D, P_F) pair corresponds to point C. Whatever choice of η we make we can see the H_1 decision region must include the interval $[b, 1]$ so $P_D = 1$. In addition, we should exclude the interval $[0, b]$ from the H_1 region so that $P_F = 1/3$. Since the LRT can be reduced to:

$$p_{Y|H}(y \mid H_1) \underset{H_0}{\overset{H_1}{\gtrless}} \eta p_{Y|H}(y \mid H_0)$$

and $a > b$ we can see this will only be the case if

$$\frac{1}{1-b} > \eta \frac{1}{a} \quad \text{and} \quad \eta > 0$$

or equivalently, $\boxed{0 < \eta < \frac{a}{1-b} = \frac{3}{2}}$. The value of η is not unique, since many choices will work. This is because the densities are flat.

Problem 6.3 (Shanmugan and Breipohl 6.14)

The signaling waveforms used in a binary communications system are given by:

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

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.

Solution:

(a) First note that the problem is really the following: Observe $y(t)$, $0 \le t \le T$ then decide between:

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

where $w(t)$ is zero-mean white Gaussian noise with $R_{WW}(\tau) = q\delta(\tau)$, $\Pr(H_i) = 1/2$, and $C_{ij} = 1 - \delta_{ij}$ for MPE cost.

One way to solve this is to transform it into the equivalent 1-D problem: Observe $y'(t)$, $0 \le t \le T$ then decide between:

$$\begin{aligned} H_0 : &\quad y'(t) = w(t) \\ H_1 : &\quad y'(t) = s'(t) + w(t) \end{aligned}$$

where $y'(t) = y(t) - s_0(t)$, $s'(t) = s_1(t) - s_0(t)$, $w(t)$ is zero-mean white Gaussian noise with $R_{WW}(\tau) = q\delta(\tau)$, $\Pr(H_i) = 1/2$, and $C_{ij} = 1 - \delta_{ij}$ for MPE cost. Now let E' be the energy in the signal $s'(t)$, then:

$$E' = \int_0^T s'^2(t)\, dt = 4E = \int_0^T 8^2 \sin^2(2\pi f_0 t)\, dt = \frac{8^2 T}{2}$$

where we have used the fact that $s_1(t) = -s_0(t)$ so $s'(t) = 2s_1(t)$.

Now we expand the signals in an orthonormal basis. Since the noise is white we don't have to worry about decorrelating it and can choose a basis that simply captures the signal space component of the data. Since the noise is white, the corresponding noise coefficients will be uncorrelated white noise for any orthonormal basis. Thus we choose as our basis:

$$\phi_1(t) = \frac{s'(t)}{\sqrt{E'}}$$

$$\left.\begin{array}{c} \phi_2(t) \\ \phi_3(t) \\ \vdots \end{array}\right\} \text{Any orthonormal set}$$

With this choice of basis functions, we obtain the following equivalent hypothesis test in terms of the coefficients of the expansion done with respect to the $\phi_i(t)$: Observe y_i', then decide between:

$$\begin{array}{lrcl|lrcl} H_0 : & y_1' &=& w_1 & H_1 : & y_1' &=& s_1' + w_1 \\ & y_2' &=& w_2 & & y_2' &=& w_2 \\ & y_3' &=& w_3 & & y_3' &=& w_3 \\ & \vdots & & & & \vdots & & \end{array}$$

where $y_i'(t) = \int_0^T y'(t)\phi_i(t)\,dt$, $w_i = \int_0^T w(t)\phi_i(t)\,dt$ are i.i.d. white Gaussian noise samples with $E[w_i w_j] = q\delta_{ij}$, $\Pr(H_i) = 1/2$, and $C_{ij} = 1 - \delta_{ij}$ for MPE cost. Since the noise coefficients are independent, only the coefficient y_1 will be useful in distinguishing between H_0 and H_1, and thus the corresponding problem has been reduced to a scalar binary hypothesis testing problem, which we know how to solve. In particular, for the given set of cost assignments and prior probabilities, we have that the LRT is given by:

$$
\begin{aligned}
y_1' \quad & \overset{H_1}{\underset{H_0}{\gtrless}} \quad \frac{\sqrt{E'}}{2} \\
\Longrightarrow \quad \int_0^T y'(t)s'(t)\,dt \quad & \overset{H_1}{\underset{H_0}{\gtrless}} \quad \frac{E'}{2} \\
\Longrightarrow \quad \int_0^T y(t)\left(s_1(t) - s_0(t)\right)\,dt \quad & \overset{H_1}{\underset{H_0}{\gtrless}} \quad \frac{E_1 - E_0}{2}
\end{aligned}
$$

where $E_i = \int_0^T s_i^2(t)\,dt$. Now note that $E_1 = E_0$, and $s_1(t) = -s_0(t) = 4\sin(2\pi f_0 t)$, thus we obtain for the final LRT:

$$
\int_0^T y(t)\sin(2\pi f_0 t)\,dt \quad \overset{H_1}{\underset{H_0}{\gtrless}} \quad 0
$$

Another way to get this result is to choose the basis functions to span the space spanned by the signal components $s_0(t)$, $s_1(t)$. Note for this problem that $s_1(t) = -s_0(t)$, so the space spanned by the signals is really one-dimensional! Thus *both* signals can be represented using only a single basis function $\phi_1(t) = s_1(t)/\sqrt{E}$. Thus we will use the following basis set:

$$
\begin{aligned}
&\phi_1(t) = \frac{s_1(t)}{\sqrt{E}} \\
&\left.\begin{array}{c} \phi_2(t) \\ \phi_3(t) \\ \vdots \end{array}\right\} \text{Any orthonormal set}
\end{aligned}
$$

where:

$$
E = \int_0^T s_1^2(t)\,dt = \int_0^T 16\sin^2(2\pi f_0 t)\,dt = 8T
$$

is the energy in the signal $s_1(t)$. Now with this basis we get for the expansion coefficients under each hypothesis:

$$
\begin{array}{rrcl|rrcl}
H_0: & y_1 & = & -s_1 + w_1 & H_1: & y_1 & = & s_1 + w_1 \\
& y_2 & = & w_2 & & y_2 & = & w_2 \\
& y_3 & = & w_3 & & y_3 & = & w_3 \\
& & \vdots & & & & \vdots &
\end{array}
$$

where $y_i(t) = \int_0^T y(t)\phi_i(t)dt$, $s_1 = \int_0^T s_1(t)\phi_1(t)dt = \sqrt{E}$, $w_i = \int_0^T w(t)\phi_i(t)dt$ are i.i.d. white Gaussian noise samples with $E[w_i w_j] = q\delta_{ij}$, $\Pr(H_i) = 1/2$, and $C_{ij} = 1 - \delta_{ij}$ for MPE cost.

Since the noise coefficients are independent, only the coefficient y_1 will be useful in distinguishing between H_0 and H_1, and thus the corresponding problem has again been reduced to a scalar binary hypothesis testing problem, which we know how to solve. Thus we are trying to decide between two

Gaussian's with different means. For the given set of cost assignments and prior probabilities, we have that the LRT is given by:

$$
\begin{array}{rrcl}
 & y_1 & \underset{H_0}{\overset{H_1}{\gtrless}} & 0 \\
\Longrightarrow & \displaystyle\int_0^T y(t)s_1(t)\,dt & \underset{H_0}{\overset{H_1}{\gtrless}} & 0 \\
\Longrightarrow & \displaystyle\int_0^T y(t)\sin(2\pi f_0 t)\,dt & \underset{H_0}{\overset{H_1}{\gtrless}} & 0
\end{array}
$$

(b) First, lets work in terms of the first solution approach above. Using the notation of part (a) we can define $\ell'(y')$ and Γ' as follows:

$$
\begin{aligned}
\ell' &= \int_0^T y'(t)s'(t)\,dt = \sqrt{E'}y_1' \\
\Gamma' &= E'/2
\end{aligned}
$$

Then the test of part (a) is given by:

$$
\ell' \underset{H_0}{\overset{H_1}{\gtrless}} \Gamma'
$$

Now based on our work in (a) it is easy to see that ℓ' is conditionally Gaussian, in particular under each hypothesis we have:

$$
\begin{array}{ll}
H_0: & \ell' \sim N(0, qE') \\
H_1: & \ell' \sim N(E', qE')
\end{array}
$$

Notice that using this approach the problem reduces to deciding between two Gaussians with the same variance; one with zero mean and one with mean E'. It is straightforward to find P_D and P_F and thus the probability of error for this problem. Recall that $\Pr(\varepsilon) = P_F P_0 + P_M P_1$. Now:

$$
\begin{aligned}
P_F &= \Pr(\ell' > \Gamma' | H_0) = Q\left(\frac{E'/2}{\sqrt{qE'}}\right) = Q\left(\frac{1}{2}\sqrt{\frac{E'}{q}}\right) \\
P_D &= \Pr(\ell' > \Gamma' | H_1) = Q\left(\frac{\Gamma' - E'}{\sqrt{qE'}}\right) = Q\left(-\frac{1}{2}\sqrt{\frac{E'}{q}}\right) \\
P_M &= 1 - P_D = 1 - Q\left(-\frac{1}{2}\sqrt{\frac{E'}{q}}\right) = Q\left(\frac{1}{2}\sqrt{\frac{E'}{q}}\right)
\end{aligned}
$$

Putting these pieces together with our particular value of E' from above, $q = 10^{-3}$, $T = 10^{-3}$, and with the fact that $P_0 = P_1 = 1/2$ we have for the probability of error:

$$
\Pr(\varepsilon) = P_F P_0 + P_M P_1 = Q\left(\frac{1}{2}\sqrt{\frac{E'}{q}}\right) = Q\left(\frac{1}{2}\frac{8}{\sqrt{2}}\sqrt{\frac{T}{q}}\right) = Q\left(2\sqrt{2}\right)
$$

We can also get this result using the second approach to the problem developed above. We'll go through that argument for completeness. Again, let us start by using the notation of part (a) to define a sufficient statistic ℓ as follows:

$$\ell \;=\; y_1 = \int_0^T y(t)\frac{s_1(t)}{\sqrt{E}}\,dt$$

Then the test of part (a) is given by:

$$\ell \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

Now based on our work in (a) it is easy to see that ℓ is conditionally Gaussian, in particular under each hypothesis we have:

$$\begin{aligned} H_0 : &\quad \ell \sim N(-\sqrt{E}, q) \\ H_1 : &\quad \ell \sim N(+\sqrt{E}, q) \end{aligned}$$

Notice that using this second approach the problem reduces to deciding between two Gaussians with the same variance; one with mean $-\sqrt{E}$ and the other with mean $+\sqrt{E}$. Note that the sufficient statistic used in either case is different (and thus a sufficient is in general not unique).

It is straightforward to find P_D and P_F and thus the probability of error for this problem. Now:

$$\begin{aligned} P_F &= \Pr(\ell > 0|H_0) = Q\left(\frac{0+\sqrt{E}}{\sqrt{q}}\right) = Q\left(\sqrt{\frac{E}{q}}\right) \\ P_D &= \Pr(\ell > 0|H_1) = Q\left(\frac{0-\sqrt{E}}{\sqrt{q}}\right) = Q\left(-\sqrt{\frac{E}{q}}\right) \\ P_M &= 1 - P_D = 1 - Q\left(-\sqrt{\frac{E}{q}}\right) = Q\left(\sqrt{\frac{E}{q}}\right) \end{aligned}$$

Putting these pieces together with our particular value of E from above, $q = 10^{-3}$, $T = 10^{-3}$, and with the fact that $P_0 = P_1 = 1/2$ we have for the probability of error:

$$\Pr(\varepsilon) = P_F P_0 + P_M P_1 = Q\left(\sqrt{\frac{E}{q}}\right) = Q\left(2\sqrt{2}\sqrt{\frac{T}{q}}\right) = Q\left(2\sqrt{2}\right)$$

This solution agrees with our earlier, as it should.

(c) Note that $\Pr(\varepsilon) = Q\left(\frac{1}{2}\sqrt{\frac{E'}{q}}\right)$ where E' is the energy in the difference signal, as shown in class. Thus the probability of error only depends on the energy in this difference signal and not on the details of the signal structure. In particular, given a choice we want the signals as far apart as possible so the energy in this difference vector are as large as possible. Thus the best we can do is choose $s_1 = -s_0$. Note in particular, that both the signal sets satisfy this. Further, they both have the same energy in their difference signal. This is easy to see since they both have the same energy, so both lie on a sphere of the same radius (i.e. the energy) and are both the opposites of one another. Thus they both have the same probability of error.

Problem 6.4 (Old Exam Question)

Consider the following binary hypothesis testing problem on $0 \leq t \leq T$:

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

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 apriori 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: $1-Q(-x) = Q(x)$

Solution:

(a) Since we desire MPE and the prior probabilities are the same we really want the ML rule, so that the optimal threshold is just $\eta = 1$

(b) As we saw in class, the thing to do is to choose $\phi_1(t) = s(t)$, and the rest of the basis vectors to be any orthonormal set that fills out the basis. Note that $s(t)$ is already normalized to have unit energy. Let $y_i = \int_0^T y(t)\phi_i(t)\,dy$, with $\phi_1(t) = s(t)$. Then under H_i we have:

$$\begin{array}{lll} H_0: & y_1 = v + w_1 & \sim N(0,2) \\ & y_i = w_i & \sim N(0,1) \quad i \geq 2 \end{array}$$

$$\begin{array}{lll} H_1: & y_1 = (6+v) + w_1 & \sim N(6,2) \\ & y_i = w_i & \sim N(0,1) \quad i \geq 2 \end{array}$$

(c) Note that only y_1 is useful in making the decision. Thus we just have a scalar Gaussian binary hypothesis test in terms of the sufficient statistic y_1. Further this sufficient statistic is Gaussian with different means but the same variances under each hypothesis. We know that the optimal ML test for a Gaussian BHT is given by (from any number of sources):

$$y_1 \underset{H_0}{\overset{H_1}{\gtrless}} (m_0 + m_1)/2 = 3$$

Thus in terms of the original observation $y(t)$ we have:

$$\int_0^T y(t)s(t)\,dt \underset{H_0}{\overset{H_1}{\gtrless}} 3$$

(d) The probability of error calculation is easy.

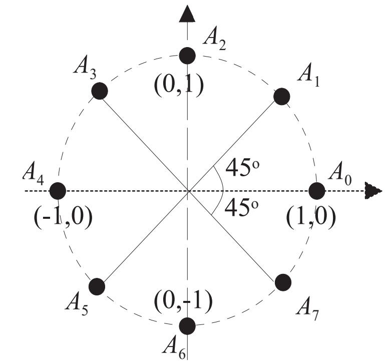
$$\begin{aligned} \Pr(\varepsilon) &= P_0 \Pr(y_1 > 3|H_0) + P_1 \Pr(y_1 < 3|H_1) \\ &= \frac{1}{2}\left[Q\left(\frac{3}{\sqrt{2}}\right) + 1 - Q\left(\frac{3-6}{\sqrt{2}}\right)\right] = \frac{1}{2}\left[Q\left(\frac{3}{\sqrt{2}}\right) + Q\left(\frac{3}{\sqrt{2}}\right)\right] \\ &= Q\left(\frac{3}{\sqrt{2}}\right) \end{aligned}$$

Problem 6.5 (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 $\Pr(H_j) = P_j = 1/8$.

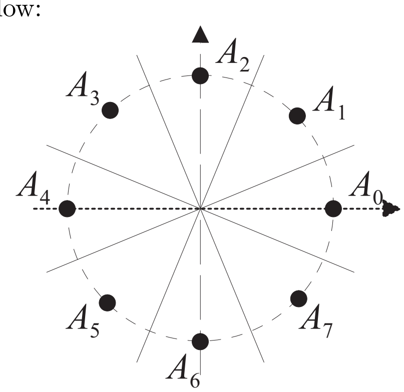


(a) Find the decision boundaries in the observation space that lead to a minimum $\Pr[error]$.

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

Solution:

(a) For the MPE decision rule we have the cost structure $C_{ij} = 1 - \delta_{ij}$ which leads to the MAP rule. In addition, we have that the hypotheses are a priori equally likely, so this simplifies further to the ML rule. Now this is a Gaussian problem with same covariances and different means under the different hypotheses, so the ML decision rule is the <u>"minimum distance receiver"</u> or nearest neighbor rule discussed in class. Finally, note that the signals under all the hypotheses have the same energy and lie on a circle. Thus the decision rule reduces to the <u>"nearest angle"</u> classifer. Thus the decision boundaries are as depicted below:



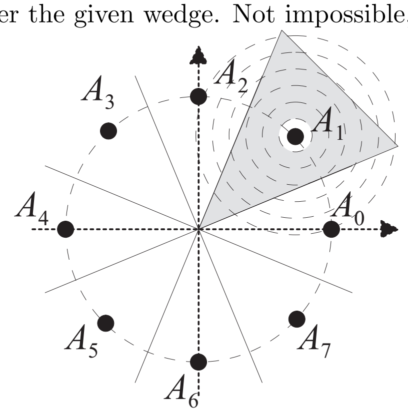
(b) To calculate P_e we need to find

$$\begin{aligned}
&P_e \\
&= \int_{\{y|\text{Don't Say } H_0\}} P(y|H_0)P_0 \, dy + \int_{\{y|\text{Don't Say } H_1\}} P(y|H_1)P_1 \, dy + \cdots + \int_{\{y|\text{Don't Say } H_7\}} P(y|H_7)P_7 \, dy \\
&= 1/8 \sum_{i=0}^{7} \left(\int_{\{y|\text{Don't Say } H_i\}} P(y|H_i) \, dy \right) = 1/8 \sum_{i=0}^{7} \left(1 - \int_{\{y|\text{Say } H_i\}} P(y|H_i) \, dy \right)
\end{aligned}$$

where we have used the fact that the prior probabilities are the same. Note that by symmetry this is equal to:

$$P_e = 8(1/8)\left(1 - \int_{\substack{\text{Shaded}\\\text{Area}}} P(y|H_1)\,dy\right)$$

where the shaded area is indicated in the figure. The basic calculation is thus the integral of a 2-D Gaussian of non-zero mean over the given wedge. Not impossible, but not fun either.



Computer Problems

Problem 6.6 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);'])
```

  (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:

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

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

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:

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

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?

Solution:

(a) The `MATLAB` code for the function `roc.m` is as follows :

```
function [pd,pf] = roc(detfunname,Gamma,y0,y1)
%  [pd,pf] = roc(detfunname,Gamma,y0,y1)
%
% detfunname : string containing the name of the detector function
% Gamma      : vector of threshold values
% y0         : vector of data observations when hypothesis H0 is true
% y1         : vector of data observations when hypothesis H1 is true
% pd         : vector of estimated probability of detection for each Gamma
% pf         : vector of estimated probability of false alarm for each Gamma
%

% Find detector performance for data sets y0 and y1 for all gammas
eval(['D0 = ',detfunname,'(y0,Gamma);']);
eval(['D1 = ',detfunname,'(y1,Gamma);']);
```

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

(b) Figure 1 shows the ROC for the four detectors for the problem of detecting a constant in Gaussian noise. The ROCs were generated using 5000 data samples under each hypothesis and with the threshold values ranging between -100 and 100. For this problem we can show that `det1`, which compares the observation to a threshold, corresponds to the optimal likelihood ratio test detector. Hence in Figure 1 `det1` has the highest P_D for any given P_F.

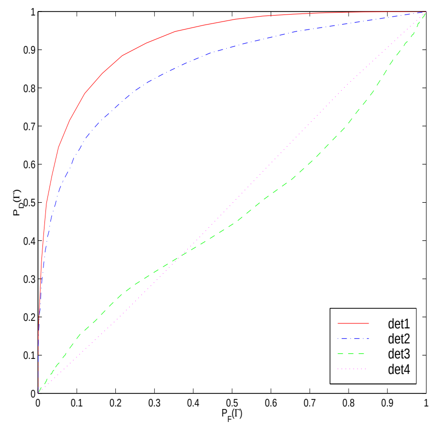


Figure 1: ROC for the four detectors `det1.m, det2.m, det3.m, det4.m` for the problem of detecting a constant ($m = 2$) in unit varaince Gaussain noise

(c) Figure 2 shows the ROC for the four detectors, for the problem of deciding between two zero mean Gaussians with different variances. We again used 5000 sample observations under each hypothesis, to evaluate the ROC curve. For this problem the optimal likelihood ratio test detector compares the power of the received the signal with a threshold to decide between the hypotheses. This corresponds to `det2`, which is seen to have the highest P_D for any given P_F in Figure 2

(d) If for the problem of detecting a constant in Gaussian noise, we vary the value of the constant (i.e. the mean m), the ROC of the detector changes. Figure 3 shows the change in the ROCs for `det1` (which is the optimal detector for this problem) when the constant is varied from 0 to 4. We require the differnce between the means to be 2 in order to acheive $P_D \approx 0.9$ at a $P_F = 0.2$.

If for the same problem we choose the mean to be a constant $m = 2$, and vary the variance of the noise instead, the ROC for `det1` again varies as shown in Figure 4. We note that the performance of the detector degrades with increasing noise levels.

Figure 5 shows the ROC for `det1` for Problem A when the true noise distribution is Cauchy instead of Gaussian. In this case `det1` no longer corresponds to the optimal likelihood ratio test detector. In order to acheive a $P_D \approx 0.9$ at a false alarm of $P_F = 0.2$ we now require the difference between the means to be around 4.

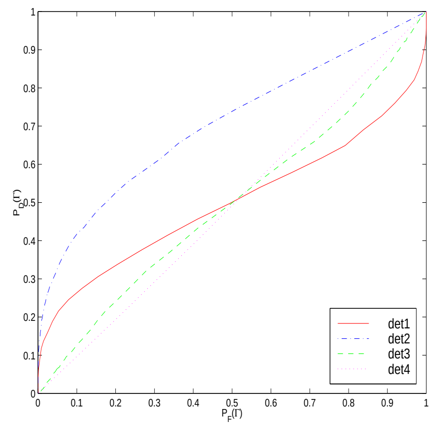


Figure 2: ROC for the four detectors `det1.m, det2.m, det3.m, det4.m` for the problem of deciding whether the observation was generated by a zero mean Gaussian distribution with variance 1 or a zero mean Gaussian distribution with variance 2

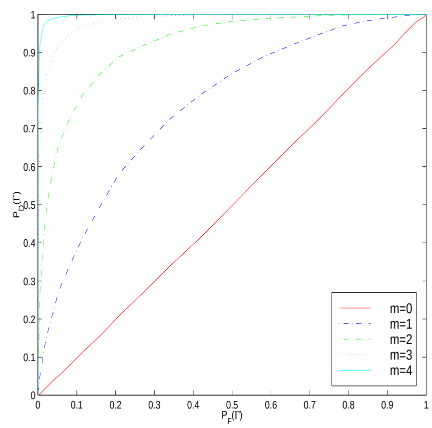


Figure 3: ROC for `det1` for the problem of detecting a constant in zero mean Gaussiance noise, when the constant in varied from 0 to 4

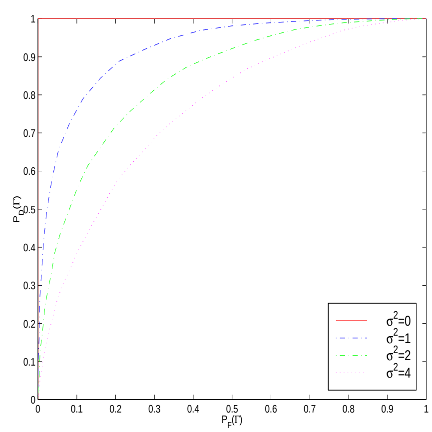


Figure 4: ROC for `det1` for the problem of detecting a constant ($m = 2$) in zero mean Gaussian noise, with variance varied from 0 to 4

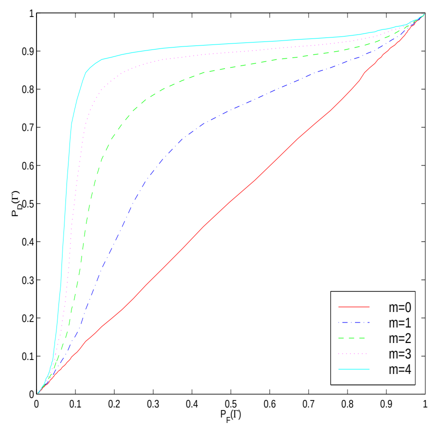


Figure 5: ROC for `det1` for the problem of detecting a constant in zero mean Cauchy noise, when the constant in varied from 0 to 4