Reference link: https://engineering.purdue.edu/ChanGroup/ECE645.html

ECE 645: Estimation Theory
Spring 2015
Instructor: Prof. Stanley H. Chan

Lecture Note 2: Neyman Pearson Testing

(LaTeX prepared by Jun Young Jeong)
March 31, 2015

This lecture note is based on ECE 645(Spring 2015) by Prof. Stanley H. Chan in the School of Electrical and Computer Engineering at Purdue University.

1 Trade off between False alarm and miss

In Lecture 1, we have looked at how the Bayes decision rule is applied to make a decision for binary and M-ary Hypothesis Testing given observation y. The basic idea of Bayes decision rule is to minimize Bayes risk defined as

$$R(\delta) = \mathbb{E}_{Y,\Theta}[C(\delta(Y), \theta)], \tag{1}$$

of which the optimal decision for binary hypothesis testing is

$$\frac{f_1(y)}{f_0(y)} \underset{H_1}{\overset{H_0}{\gtrless}} \frac{(c_{11} - c_{00})\pi_0}{(c_{01} - c_{11})\pi_1}. \tag{2}$$

By defining the likelihood ratio as $L(y) \overset{\text{def}}{=} \frac{f_1(y)}{f_0(y)}$, we observe that (2) can be completely characterized by three factors: $L(y)$, $C(i,j)$, and π_i. However, how should we characterize the *performance* of such a detection rule?

In order to characterize the performance of the detection rule, we first partition the observation Γ into to two subsets Γ_0 and Γ_1 over which we select H_0 and H_1. More precisely, we define the followings:

$$\Gamma_0 \overset{\text{def}}{=} \{y \in \Gamma \mid \delta(y) = 0\} \tag{3}$$

$$\Gamma_1 \overset{\text{def}}{=} \{y \in \Gamma \mid \delta(y) = 1\}. \tag{4}$$

Clearly, it holds that $\Gamma_0 \cup \Gamma_1 = \Gamma$ and $\Gamma_0 \cap \Gamma_1 =$. Given Γ_0 and Γ_1, we can now define two performance metric as follows.

Definition 1.
The **Probability of detection** (also known as the power of δ) is defined as

$$P_D(\delta) \overset{\text{def}}{=} \int_{\Gamma_1} f_1(y)\mathrm{d}y = P(\text{claim } H_1 \mid H_1). \tag{5}$$

Definition 2.
The **Probability of false alarm** (also known as the Type I error) is defined as

$$P_F(\delta) \overset{\text{def}}{=} \int_{\Gamma_1} f_0(y)\mathrm{d}y = P(\text{claim } H_1 \mid H_0). \tag{6}$$

Definition 3.
The **Probability of miss** (also known as the Type II error) is defined as

$$P_M(\delta) \stackrel{\text{def}}{=} \int_{\Gamma_0} f_1(y)\mathrm{d}y = P(\text{claim } H_0 \mid H_1). \tag{7}$$

Figure 1 illustrates how the above quantities are defined for binary hypothesis testing with a threshold τ. In the case of Bayes decision rule, τ is the right hand side of (2). For general cases, τ can take arbitrary values.

What we want to do next is to derive the Bayesian risk in terms of P_F and P_D.

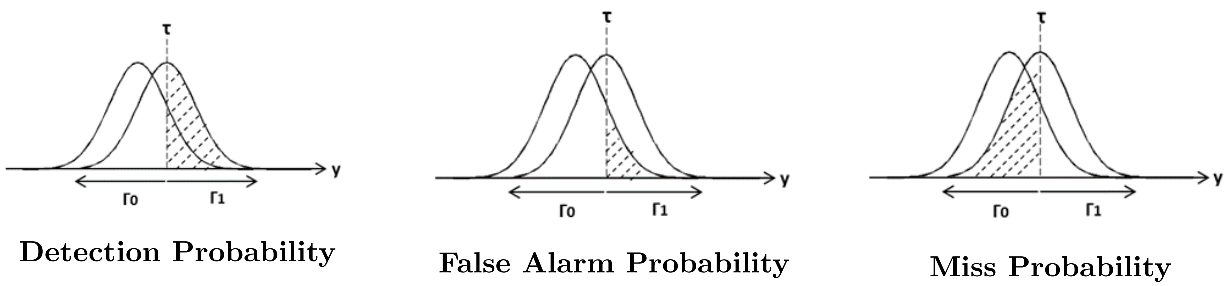

Figure 1: Region P_D, P_F and P_M at binary hypothesis testing with threshold τ

Proposition 1.
For any decision rule δ,

$$R(\delta) = \pi_0 C_{00} + \pi_1 C_{01} + \pi_0(C_{10} - C_{00})P_F(\delta) + \pi_1(C_{11} - C_{01})P_D(\delta). \tag{8}$$

The interpretation of (8) is that we can write $R(\delta)$ in terms of C_{ij}, π_j, P_F and P_D. In other words, $(P_F(\delta), P_D(\delta))$ completely characterizes the performance of the decision rule δ.

Proof.
By using iterative expectation and total expectation, we can decompose Bayes risk as

$$\begin{aligned} R(\delta) &= \mathbb{E}_{Y,\Theta}[C(\delta(Y),\theta)] \\ &= \mathbb{E}_{\Theta}[\mathbb{E}_{Y|\Theta}[C(\delta(Y),\theta)|\Theta=\theta]] \\ &= \mathbb{E}_{Y|\Theta}[C(\delta(Y),0)|H_0]\pi_0 + \mathbb{E}_{Y|\Theta}[C(\delta(Y),1)|H_1]\pi_1 \\ &= R_0(\delta)\pi_0 + R_1(\delta)\pi_1, \end{aligned} \tag{9}$$

where we defined

$$\begin{aligned} R_0(\delta) &\stackrel{\text{def}}{=} \mathbb{E}_{Y|\Theta}[C(\delta(Y),0)|H_0] = \int_\Gamma C(\delta(y),0))f_0(y)\mathrm{d}y, \\ R_1(\delta) &\stackrel{\text{def}}{=} \mathbb{E}_{Y|\Theta}[C(\delta(Y),1)|H_1] = \int_\Gamma C(\delta(y),1))f_1(y)\mathrm{d}y. \end{aligned}$$

It remains to determine $R_0(\delta)$ and $R_1(\delta)$:

$$
\begin{aligned}
R_0(\delta) &= \int_{\Gamma_0} C(\delta(y),0) f_0(y)\mathrm{d}y + \int_{\Gamma_1} C(\delta(y),0) f_0(y)\mathrm{d}y \\
&= \int_{\Gamma_0} C(0,0) f_0(y)\mathrm{d}y + \int_{\Gamma_1} C(1,0) f_0(y)\mathrm{d}y \\
&= C_{00}(1 - P_F(\delta)) + C_{10} P_F(\delta). \qquad (10)
\end{aligned}
$$

Similarly, we we can express $R_1(\delta)$ as

$$
\begin{aligned}
R_1(\delta) &= \int_{\Gamma_0} C(0,1) f_1(y)\mathrm{d}y + \int_{\Gamma_1} C(1,1) f_1(y)\mathrm{d}y \\
&= C_{01}(1 - P_D(\delta)) + C_{11} P_D(\delta). \qquad (11)
\end{aligned}
$$

Substituting (10) and (11) into (9) completes the proof.

□

Remark: In case of uniform cost where $C_{00} = C_{11} = 1$ and $C_{01} = C_{10} = 0$, (9) can be simplified as

$$R(\delta) = 1 - P_F(\delta) + P_D(\delta).$$

2 Receiver Operating Characteristic (ROC)

We now take a closer look at the space defined $(P_F(\delta),\ P_D(\delta))$. First of all, we define the following.

$$\mathcal{D} \stackrel{\text{def}}{=} \{(P_F(\delta), P_D(\delta)) \mid \forall \delta\}.$$

There are some important properties about $\mathcal{D}$.

1. $(0,0)$ and $(1,1) \in \mathcal{D}$
   - $(0,0)$ corresponds to a decision rule that always says H_0.
   - $(1,1)$ corresponds to a decision rule that always says H_1.
   - These pairs are ideally achievable if the number of observations are infinite or H_0 and H_1 do not overlap.

2. if $(x,y) \in \mathcal{D}$, then $(1-x, 1-y)$ is also in $\mathcal{D}$
   - Geometrically, it means that if we have a decision rule δ, we can always flip the decision δ makes and call the flipped decision as another rule.

3. $\mathcal{D}$ is convex. To show that $\mathcal{D}$ is convex, we need to show $\forall (P_F, P_D) \in \mathcal{D}$, and $(P'_F, P'_D) \in \mathcal{D}$, the convex combination $(\alpha P_F + (1-\alpha)P'_F, \alpha P_D + (1-\alpha)P'_D)$ is also in $\mathcal{D}$ for any $0 \le \alpha \le 1$. In order to have this property, we need to relax the binary decision to a randomized decision by modifying

$$\delta : \Gamma \to \{0,1\}$$

$$\downarrow$$

$$\delta : \Gamma \to [0,1]$$

The range of the resulting decision rule is $[0,1]$. However, if we have to output a binary decision, we can set $\delta(y)=1$ with probability γ if $L(y)=\tau$. Once we have the new decision rule, it becomes easy to show the convexity of $\mathcal{D}$: if $P_F \in D$, $P_F' \in D$ then

$$\begin{aligned}\alpha P_F + (1-\alpha)P_F' &= \alpha \int \delta(y) f_0(y)\mathrm{d}y + (1-\alpha)\int \delta'(y) f_0(y)\mathrm{d}y \\ &= \int (\alpha\delta(y) + (1-\alpha)\delta'(y)) f_0(y)\mathrm{d}y.\end{aligned}$$

Since $\alpha\delta(y) + (1-\alpha)\delta'(y)$ is just another randomized decision rule, it must belong to $\mathcal{D}$. Similar argument holds for $\alpha P_D + (1-\alpha)P_D'$.

4. The concave curve at upper bound of D is called the *Receiver Operating Characteristic* (ROC) curve - $\forall$ (P_F,P_D) on ROC curve, $P_D \geq P_F$. See Figure 2.

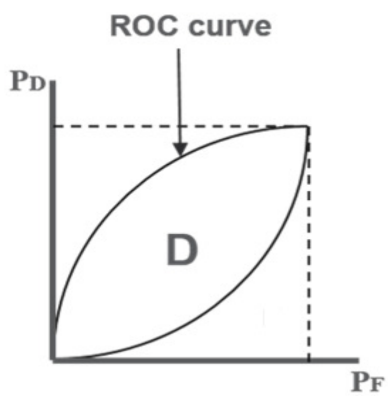

Figure 2: Domain of feasible tests and ROC for a binary hypothesis testing problem.

Example.(Drawing ROC curve)

Let Y be a Poisson random variable with parameters λ_0 and λ_1 under H_0 and H_1:

$$H_0 : P_0(y) = \frac{\lambda_0^y}{n!}e^{-\lambda_0}$$

$$H_1 : P_1(y) = \frac{\lambda_1^y}{n!}e^{-\lambda_1}$$

Then, the Binary HT must be in the form of the likelihood ratio

$$L(y) = \frac{P_1(y)}{P_0(y)} = (\frac{\lambda_1}{\lambda_0})^y e^{-(\lambda_1-\lambda_0)} \geq \tau$$

Take natural log on both sides, we have

$$\log L(y) = y[\log\lambda_1 - \log\lambda_0] - \lambda_1 + \lambda_0.$$

So, test becomes

$$y \underset{H_1}{\overset{H_0}{\gtrless}} \frac{\log(\tau) + \lambda_1 - \lambda_0}{\log\lambda_1 - \log\lambda_0} \stackrel{\text{def}}{=} \eta.$$

Note that Y takes integer values whereas the threshold η takes real values. Let $[\eta]$ denote the smallest non-negative integer greater than or equal to η. So, for a given threshold η, detection and false alarm probabilities can be expressed as following:

$$P_D = P_1(Y \geq [\eta]) = 1 - P_1(Y < [\eta]) = 1 - e^{\lambda_1} \sum_{k=0}^{[\eta]-1} \frac{\lambda_1^k}{k!}$$

$$P_F = P_0(Y \geq [\eta]) = 1 - P_0(Y < [\eta]) = 1 - e^{\lambda_0} \sum_{k=0}^{[\eta]-1} \frac{\lambda_0^k}{k!}$$

So, for various choices of τ, different P_D and P_F are easily computed.

Suppose that λ_0 is fixed as 2 and λ_1 has three different values: 3, 5 and 7. Then, the resulting ROC curves at different λ is shown in Figure 3. Since Y is a discrete, the ROC curve is discontinuous. To generate a continuous ROC curve like Figure 4, we must connect the points by randomization. For example, the segment connecting the points $(P_F(y+1)$,$P_D(y+1))$ and $(P_F(y)$,$P_D(y))$ can be obtained by using the probability γ, a test with $\eta = y+1$ and with probability $1-\gamma$, a test with $\eta = y$.

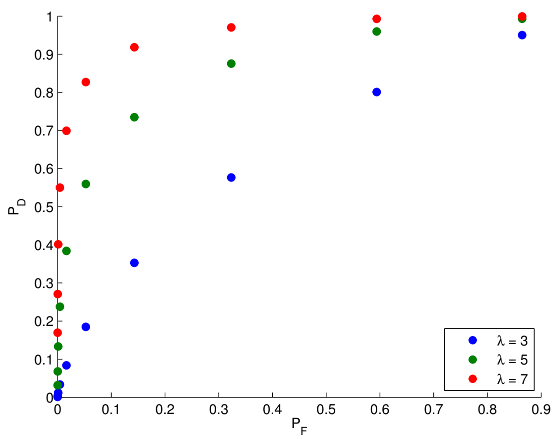

Figure 3: Discrete ROC curve

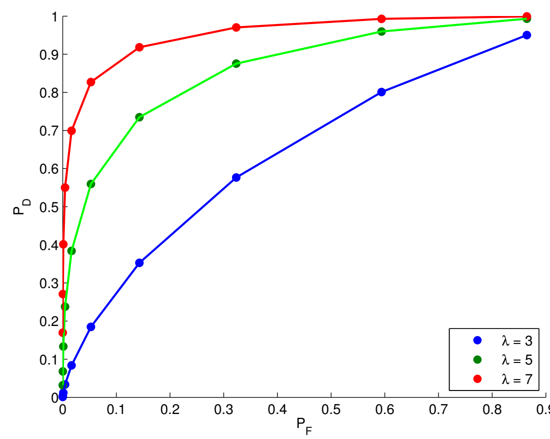

Figure 4: ROC curve with randomization

3 Neyman-Pearson Testing

3.1 Neyman-Pearson Problem

While Bayes decision rule works well for many problems, it has some fundamental limitations. The most important issue is the prior probability, which may not be available in some problems. To deal with this case, we should use alternative design which does not require the prior. The method is called Neyman-Pearson testing.

Neyman Pearson test defines the binary hypothesis testing problem by selecting the decision δ which maximizes the detection probability $P_D(\delta)$ while keeping false alarm probability $P_F(\delta)$ under certain threshold α (called the significance level of the NP-test). Thus the goal of Neyman Pearson testing is to find the most powerful α level test between H_0 vs H_1.

Definition 4. NEYMAN-PEARSON TESTING
The **Neyman-Pearson Testing** is defined as

$$\delta_{NP}(y) \stackrel{\text{def}}{=} \underset{\delta}{\operatorname{argmax}} \;\; P_D(\delta) \;\; \text{subject to} \;\; P_F(\delta) \leq \alpha. \tag{12}$$

The optimization posed in Neyman-Pearson testing turns out to be quite natural in many daily situations. Assume that someone plans to buy a number of cars with maximum budget α dollars. Consider two functions (they do not need to be probability distributions):

- Let $F(y)$ be the cost of a car y;
- Let $G(y)$ be the happiness of buying the car y.

Let $\mathcal{C}$ be the list of candidate cars to buy. Then, the best choice is to buy cars $y \in \mathcal{C}$ such that the cost is limited to α, i.e., $\sum_{y\in\mathcal{C}} F(y) \leq \alpha$, while the happiness is maximized, i.e., $\max \sum_{y\in\mathcal{C}} G(y)$.

3.2 Neyman Pearson Solution

We now take a closer look at (12). In order to understand (12), we first note that $P_D(\delta)$ and $P_F(\delta)$ can be expressed as

$$\begin{aligned} P_F(\delta) &= \int_\Gamma \delta(y) f_0(y) \mathrm{d}y \\ P_D(\delta) &= \int_\Gamma \delta(y) f_1(y) \mathrm{d}y, \end{aligned}$$

where $\delta(y)$ is the **randomized decision rule** and Γ is the observation space of y. Then, we can show the following proposition.

Proposition 2.
The **solution** of (12) is given by

$$\delta_{NP}(y) = \begin{cases} 1 & \text{if } L(y) > \eta \\ \gamma & \text{if } L(y) = \eta \\ 0 & \text{if } L(y) < \eta \end{cases}, \tag{13}$$

where the parameter (η, γ) is chosen such that $P_F(\delta_{NP}) = \alpha$.

Proof.
Given significance level α, choose δ_{NP} such that

$$\begin{aligned} \alpha = P_F(\delta_{NP}) &= \int \delta_{NP}(y) f_0(y) \mathrm{d}y \\ &= \int_{L(y)>\eta} 1 \cdot f_0(y)\mathrm{d}y + \int_{L(y)=\eta} \gamma \cdot f_0(y)\mathrm{d}y + \int_{L(y)<\eta} 0 \cdot f_0(y)\mathrm{d}y \\ &= \int_{L(y)>\eta} f_0(y)\mathrm{d}y + \gamma \int_{L(y)=\eta} f_0(y)\mathrm{d}y \end{aligned} \tag{14}$$

Now consider an arbitrary decision rule δ such that $P_F(\delta) \le \alpha$

$$P_F(\delta) = \int_{L(y)>\eta} \delta(y) f_0(y)\mathrm{d}y + \int_{L(y)=\eta} \delta(y) f_0(y)\mathrm{d}y + \int_{L(y)<\eta} \delta(y) f_0(y)\mathrm{d}y \le \alpha \tag{15}$$

Since (14) $\ge$ (15), we have

$$\int_{L(y)>\eta} (1-\delta(y)) f_0(y)\mathrm{d}y + \int_{L(y)=\eta} (\gamma-\delta(y)) f_0(y)\mathrm{d}y - \int_{L(y)<\eta} \delta(y) f_0(y)\mathrm{d}y \ge 0 \tag{16}$$

Next, we consider $P_D(\delta_{NP})$ and $P_D(\delta)$. Following the same procedure as above, we can show that

$$\begin{aligned} P_D(\delta_{NP}) - P_D(\delta) &= \int \delta_{NP}(y) f_1(y)\mathrm{d}y - \int \delta(y) f_1(y)\mathrm{d}y \\ &= \int_{L(y)>\eta} (1-\delta(y)) f_1(y)\mathrm{d}y + \int_{L(y)=\eta} (\gamma-\delta(y)) f_1(y)\mathrm{d}y - \int_{L(y)<\eta} \delta(y) f_1(y)\mathrm{d}y \end{aligned} \tag{17}$$

Since $L(y) > \eta$, (16) becomes

$$\begin{aligned} P_D(\delta_{NP}) - P_D(\delta) &= \eta \int_{L(y)>\eta} (1-\delta(y)) f_0(y)\mathrm{d}y + \eta \int_{L(y)=\eta} (\gamma-\delta(y)) f_0(y)\mathrm{d}y - \eta \int_{L(y)<\eta} f_0(y)\mathrm{d}y \\ &= \eta(\text{ RHS of (16)}) \ge 0 \end{aligned} \tag{18}$$

Since (18) holds for all δ such that $P_F(\delta) \le \alpha$, we have $P_D(\delta_{NP}) \ge P_D(\delta)$ for $\forall \delta$ such that $P_F(\delta) \le \alpha$. Therefore, δ_{NP} is the maximizer of (12).

□

Choice of parameter γ and η

In proposition 2, we have to specify the parameter (γ, η) such that $P_F(\delta_{NP}) = \alpha$. We now discuss how to choose the parameters. First we know that

$$P_F = \int_{L(y)>\eta} f_0(y)\mathrm{d}y + \gamma \int_{L(y)=\eta} f_0(y)\mathrm{d}y.$$

Let us define

$$\Psi(\eta) \stackrel{\text{def}}{=} \int_{L(y)>\eta} f_0(y)\mathrm{d}y. \tag{19}$$

We can prove the following result.

Proposition 3.

$$\int_{L(y)=\eta} f_0(y)\mathrm{d}y = \lim_{n\to\infty} \Psi(\eta - \frac{1}{n}) - \Psi(\eta)$$

Proof.

$$\begin{aligned} \Psi(\eta - \frac{1}{n}) - \Psi(\eta) &= \int_{L(y)>\eta-\frac{1}{n}} f_0(y)\mathrm{d}y - \int_{L(y)>\eta} f_0(y)\mathrm{d}y \\ &= \int_{\eta-\frac{1}{n}<L(y)\le\eta} f_0(y)\mathrm{d}y \end{aligned}$$

Since $(\eta - \frac{1}{n}, n] \overset{n\to\infty}{\longrightarrow} \eta$ it holds that

$$\int_{L(y)=\eta} f_0(y)\mathrm{d}y = \lim_{n\to\infty} \Psi(\eta - \frac{1}{n}) - \Psi(n)$$

□

The consequence of the proposition is the following procedure of determining the parameters.

1. if $\Psi(\eta)$ is continuous,

$$\int_{L(y)=\eta} f_0(y)\mathrm{d}y = 0$$

2. if $\Psi(\eta)$ is discontinuous at η, then we can let A and B

$$A = \lim_{n\to\infty} \Psi(\eta - \frac{1}{n}) - \Psi(\eta),$$
$$B = \Psi(\eta),$$

Hence

$$\int_{L(y)=\eta} f_0(y)\mathrm{d}y = A - B.$$

Thus, if $P_F = \alpha$, then we can derive γ as

$$\alpha = B + \gamma\,|A - B| \Rightarrow \gamma = \frac{\alpha - B}{|A - B|}.$$

3.3 Examples

Example 1: Z-channel

Find the NP test decision rule of below Z-channel. See Figure 5.

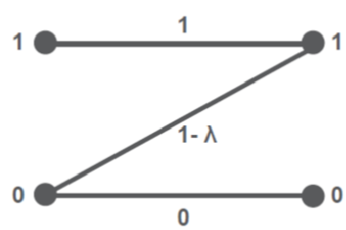

Figure 5: Z channel

The Z-channel problem can be modeled as a binary hypothesis testing problem with the following distributions:

$$f_0(y) = \begin{cases} 1 & \text{if } y = 0 \\ 0 & \text{if } y = 1 \end{cases}$$

$$f_1(y) = \begin{cases} \lambda & \text{if } y = 0 \\ 1-\lambda & \text{if } y = 1 \end{cases}$$

Solution

The Likelihood ratio becomes

$$L(y) = \frac{f_1(y)}{f_0(y)} = \begin{cases} \lambda & \text{if } y = 0 \\ \infty & \text{if } y = 1 \end{cases}$$

By Proposition (2), we can define $\delta_{NP}(y)$ as

$$\delta_{NP}(y) = \begin{cases} 1 & \text{if } L(y) > \eta \\ \gamma & \text{if } L(y) = \eta \\ 0 & \text{if } L(y) < \eta \end{cases}$$

So it remains to determine η. Define

$$\Psi(\eta) = \int_{L(y)>\eta} f_0(y))\mathrm{d}y.$$

Since $L(y)$ can take either λ and ∞, $\Psi(\eta)$ should look following:

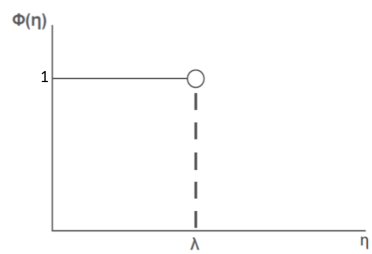

Figure 6: The function $\Psi(\eta)$.

Therefore, $A = 1$ and $B = 0$. So by using the formula $\alpha = B + \gamma|A - B|$, we can show that

$$\gamma = \alpha.$$

Moreover, the figure indicates that there is a discontinuous point at $\eta = \lambda$. Therefore the NP rule becomes

$$\delta_{NP}(y) = \begin{cases} \gamma & \text{if } y = 1 \\ \alpha & \text{if } y = 0 \end{cases}$$

We can also compute P_D which is

$$\begin{aligned} P_D(\delta) &= \int_{L(y)>\lambda} f_1(y)\mathrm{d}y + \int_{L(y)=\lambda} \gamma f_1(y)\mathrm{d}y \\ &= (1 - \lambda) + \alpha(\lambda) \end{aligned}$$

Example 2
Consider a binary hypothesis testing with given pdfs

$$H_0 : Y = N$$
$$H_1 : Y = N + S$$

where N and S are independent random variables having pdf

$$f(x) = \begin{cases} \mathrm{e}^{-x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

(a) Find the likelihood ratio between H_0 and H_1.
(b) Find the threshold and detection probability of α-level NP testing of H_0 vs H_1.

Solution
Since N and S are independent, the density of $N+S$ is the convolution of N and S. So, $f_1(y)$ becomes:

$$\begin{aligned} f_1(y) &= \int_0^y f(x)f(y-x)\mathrm{d}x \\ &= \int_0^y e^{-x}e^{y-x}\mathrm{d} \\ &= \int_0^y e^{-y}\mathrm{d}x = ye^{-y} \text{ if } y \geq 0 \end{aligned}$$

So we have the hypothesis pair:

$$f_0(y) = \begin{cases} \mathrm{e}^{-y} & \text{if } y \geq 0 \\ 0 & \text{if } y < 0 \end{cases}$$

$$f_1(y) = \begin{cases} y\mathrm{e}^{-y} & \text{if } y \geq 0 \\ 0 & \text{if } y < 0 \end{cases}$$

Then the likelihood ratio becomes $L(y) = \frac{f_1(y)}{f_0(y)} = y$ for $y \geq 0$. Define $\delta_{NP}(y)$ as

$$\delta_{NP}(y) = \begin{cases} 1 & \text{if } L(y) > \eta \\ \gamma & \text{if } L(y) = \eta \\ 0 & \text{if } L(y) < \eta \end{cases}$$

To determine (γ,α), we find that

$$\Psi(\eta) = \int_{L(y)>\eta} f_0(y)\mathrm{d}y = \int_\eta^\infty e^{-y}\mathrm{d}y = e^{-\eta}.$$

Thus, $\Psi(\eta)$ looks like below:

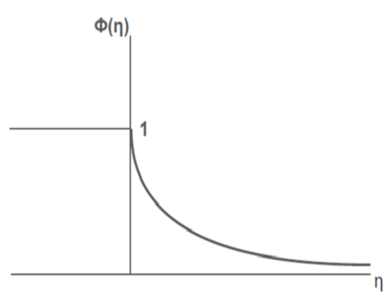

Figure 7: The function $\Psi(\eta)$.

Since $\Psi(\eta)$ is continuous, $\gamma = 0$. Therefore, the NP rule becomes

$$\delta_{NP}(y) = \begin{cases} 1 & \text{if } y > -\log(\alpha) \\ 0 & \text{if } y \leq -\log(\alpha) \end{cases}$$

Finally, the probability of detection is

$$
\begin{aligned}
P_D(\delta) = E_1(\delta_{NP}(Y)) &= \int_{-\log\alpha}^{\infty} xe^{-x}\mathrm{d}x \\
&= e^{\log\alpha}(-\log\alpha) + 1) \\
&= \alpha(-\log\alpha + 1) \\
&= \alpha(1 - \log\alpha).
\end{aligned}
$$

References

[1] H.V Poor. "Elements of Hypothesis Testing," in *An Introduction to Signal Detection and Estimation*, Springer, 1994, pp.5-12.

[2] B.C.Levy "Binary and M-ary Hypothesis Testing." in *Principles of Signal Detection and Parameter Estimation*, Springer, 2008, pp.27-32

[3] J.Krogmeier. ECE 645. Class Lecture, Topic:"Hypothesis Testing", Faculty of Engineering, Purdue University, United States, IN

[4] N.Vaswani. EE 527. Class Lecture, Topic:"Neyman Pearson test for simple binary hypotheses", Faculty of Engineering, Iowa State University, United States, IA

[5] R.Nowak. ECE 830. Class Lecture, Topic:"Neyman Pearson Detectors", Faculty of Engineering, University of Wisconsin, United States, WI