6.434/16.391 Statistics for Engineers and Scientists

Lecture 16   11/21/2012

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Lecture 16 11/21/2012

NEYMAN-PEARSON CRITERION

Neyman-Pearson criterion

- This criterion is useful when it is difficult to assign costs. A typical application is Radar detection
- In general we would like P_{D} as large as possible and P_{F} as small as possible.
- Neyman-Pearson criterion:

$$\begin{aligned} &\text{maximize } P_{\mathrm{D}} \text{ (or minimize } P_{\mathrm{M}}) \\ &\text{subject to } P_{\mathrm{F}} \leq \alpha \end{aligned}$$

- This leads to a LRT

$$\Lambda(\boldsymbol{r}) = \frac{p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1)}{p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)} \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \lambda$$

Neyman-Pearson criterion

- Proof

$$P_{\mathrm{M}} = \int_{Z_0} p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1)d\boldsymbol{r}$$

$$P_{\mathrm{F}} = \int_{Z_1} p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)d\boldsymbol{r}$$

$$= 1 - \int_{Z_0} p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)d\boldsymbol{r}$$

NP test

$$\begin{aligned} &\text{minimize } P_{\mathrm{M}} \\ &\text{subject to } P_{\mathrm{F}} \leq \alpha \end{aligned} \qquad (*)$$

- Since we want to make a decision in any case $Z = Z_0 \cup Z_1$, the observation space Z must be partitioned in two regions.
  – First note that any test satisfy (*) must be such that $P_{\mathrm{F}} = \alpha$ since any region Z_0 for which $P_{\mathrm{F}} < \alpha$ can be improved by removing more points $\boldsymbol{r} \in \boldsymbol{R}$ from Z_0 since P_{D} and P_{F} increase, we can continue until $P_{\mathrm{F}} = \alpha$

Neyman-Pearson criterion

- Next which points are most efficient.
  Suppose Z_0 is chosen according to LRT, i.e.,

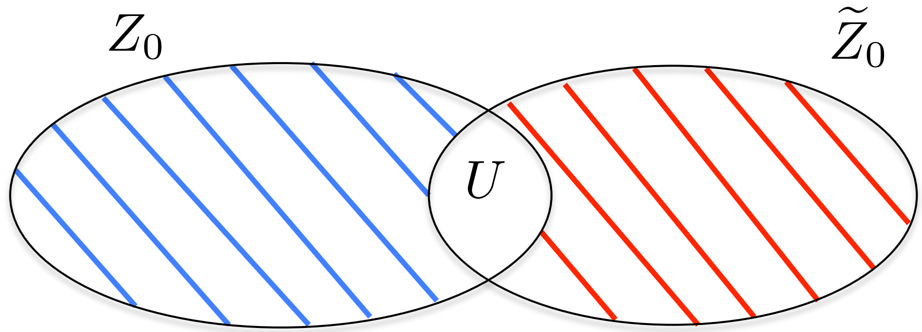
$$\frac{p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1)}{p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)} \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \lambda \qquad (*)$$

where λ satisfy

$$P_{\mathrm{F}} = \int_{\lambda}^{\infty} P_{\Lambda|\mathcal{H}_0}(x|\mathcal{H}_0)dx = \alpha$$

We will show that any other Z_0' with $P_{\mathrm{F}}' \leq P_{\mathrm{F}}$ is an inferior test.
That is $P_{\mathrm{M}} \leq P_{\mathrm{M}}'$

Neyman-Pearson criterion



$$I = \int_{Z_0 \setminus U} - \int_{\widetilde{Z}_0 \setminus U} \left[p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{R}|\mathcal{H}_1) - \lambda p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{R}|\mathcal{H}_0) \right] d\boldsymbol{r}$$

- Note that
  - The integrand is negative over the region $Z_0 \setminus U$ since $(*)$ holds and on Z_0 we choose $\mathcal{D}_0$
  - The integrand is non-negative over the region $\widetilde{Z}_0 \setminus U$ since U is a region of commonality with Z_0 and hence it contains all points of $\widetilde{Z}_0$ with negative integrand

$$\Rightarrow I < 0$$

Neyman-Pearson criterion

- Adding and subtracting $\int_U \left[p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{R}|\mathcal{H}_1) - \lambda p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{R}|\mathcal{H}_0) \right] d\boldsymbol{r}$

$$\int_{Z_0} - \int_{\widetilde{Z}_0} p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{R}|\mathcal{H}_1) d\boldsymbol{r} < \lambda \int_{Z_0} - \int_{\widetilde{Z}_0} p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{R}|\mathcal{H}_0) d\boldsymbol{r}$$

$$\begin{aligned} P_{\mathrm{M}} - \widetilde{P}_{\mathrm{M}} &\leq \lambda \left[(1 - P_{\mathrm{F}}) - (1 - \widetilde{P}_{\mathrm{F}}) \right] \\ &= \lambda [\widetilde{P}_{\mathrm{F}} - P_{\mathrm{F}}] \leq 0 \end{aligned}$$

$$P_{\mathrm{M}} \leq \widetilde{P}_{\mathrm{M}}$$

Finding the threshold in an NP test

- Note that $\Lambda(\boldsymbol{R}) \in \mathbb{R}$ is a r.v. with p.d.f.

$$p_{\Lambda}(x) \text{ unconditional}$$

$$p_{\Lambda|\mathcal{H}_i}(x|\mathcal{H}_i) \text{ conditional to } \mathcal{H}_i$$

It follows that we can write

$$\begin{aligned} P_{\mathrm{F}} &= \alpha' \\ &= \mathbb{P}\{\Lambda > \lambda|\mathcal{H}_0\} \\ &= 1 - \int_{\lambda}^{\infty} p_{\lambda|\mathcal{H}_0}(x|\mathcal{H}_0)dx \end{aligned}$$

- Solving this gives λ.

Finding the threshold in an NP test

- Thus for the NPT the steps are:

  A) choose the smallest possible λ for $P_{\mathrm{F}} = \alpha' \leq \alpha$
  (if P_{F} is a continuous function of λ we can reach $P_{\mathrm{F}} = \alpha$)

  B) perform the LRT as $\Lambda(\boldsymbol{r}) \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \lambda$

Summary for simple binary hypotheses testing

- Bayes Criterion (if everything is known)
  - Minimize $\mathcal{R}$

$$\text{test} \qquad \Lambda(\boldsymbol{r}) \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \eta = \frac{P_0}{P_1} \frac{C_{10} - C_{00}}{C_{01} - C_{11}}$$

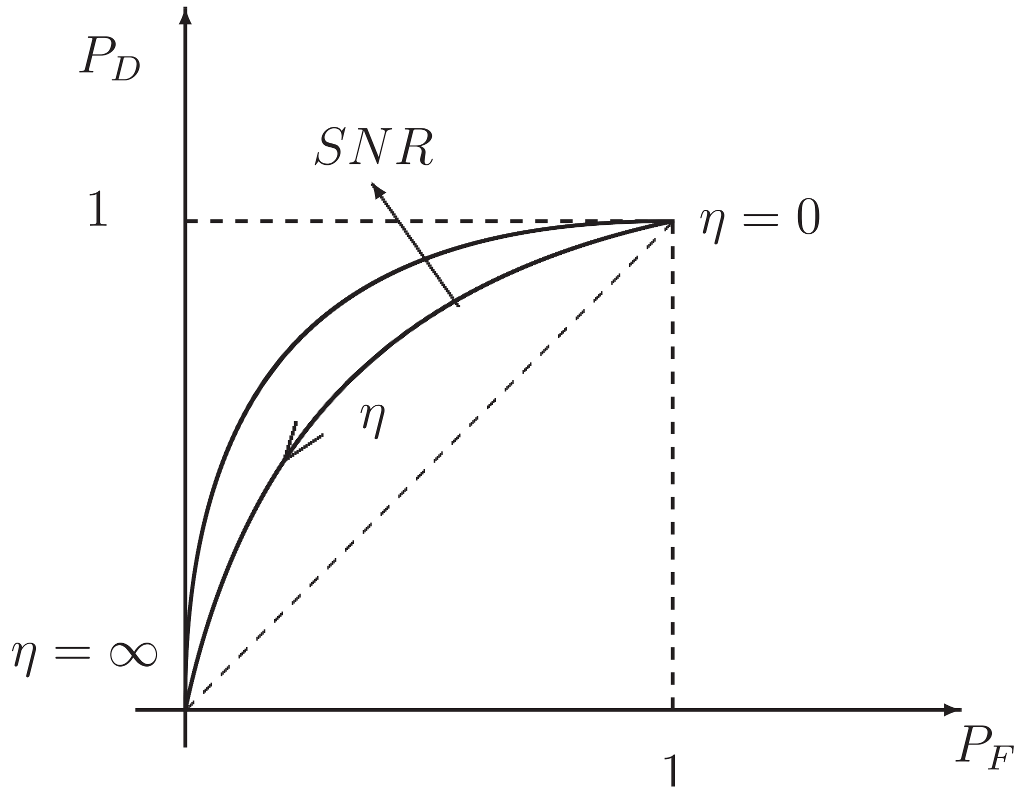
- Neyman-Pearson Criterion (if costs or a-priori probabilities are unknown)
  - Minimize P_{M} under the constraint $P_{\mathrm{F}} \leq \alpha$

$$\text{test} \qquad \Lambda(\boldsymbol{r}) \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \lambda = \lambda(\alpha)$$

- NOTE: ALL TESTS ARE LRT!

Receiver Operating Characteristics (ROC)

- The curves P_{D} as a function of P_{F} are called Receiver Operating Characteristics (ROC), and are used to compare tests.
- The ideal point for a test is $P_{\mathrm{D}} = 1, P_{\mathrm{F}} = 0$.



Properties of ROC

- All continuous LRT have ROCs that are concave (downward). If they were not, a randomized test would be better. This would contradict the fact that the LRT is optimum. (See 2.2.12 of Van Trees)
- All continuous LRT have ROCs that are above the $P_\mathrm{D} = P_\mathrm{F}$ line. (Follows from point 1, since all ROCs starts in $P_\mathrm{D} = P_\mathrm{F} = 0$ and arrive at $P_\mathrm{D} = P_\mathrm{F} = 1$).
- The slope of a curve in a ROC at a particular point is equal to the threshold η.

Properties of ROC

- Proof:

$$\frac{dP_{\mathrm{D}}}{dP_{\mathrm{F}}} = \frac{dP_{\mathrm{D}}/d\eta}{dP_{\mathrm{F}}/d\eta} = \frac{dP_{\mathrm{D}}/d\eta}{-p_{\Lambda|\mathcal{H}_0}(\eta|\mathcal{H}_0)}$$

$$\begin{aligned} P_{\mathrm{F}} &= \mathbb{P}\{\Lambda > \eta|\mathcal{H}_0\} \\ &= \int_{\eta}^{\infty} p_{\Lambda|\mathcal{H}_0}(x|\mathcal{H}_0)dx \end{aligned}$$

and therefore

$$\frac{dP_{\mathrm{F}}}{d\eta} = -p_{\Lambda|\mathcal{H}_0}(\eta|\mathcal{H}_0)$$

Properties of ROC

To find $dP_{\mathrm{D}}/d\eta$

$$Z_1 = Z_1(\eta) = \{\boldsymbol{r} : \Lambda(\boldsymbol{r}) \geq \eta\}$$

$$\begin{aligned} P_{\mathrm{D}} = P_{\mathrm{D}}(\eta) &= \int_{Z_1(\eta)} p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1)d\boldsymbol{r} \\ &= \int_{Z_1(\eta)} \Lambda(\boldsymbol{r})p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)d\boldsymbol{r} \\ &= \int_{\eta}^{\infty} x\, p_{\Lambda|\mathcal{H}_0}(x|\mathcal{H}_0)dx \end{aligned}$$

$$\frac{dP_{\mathrm{D}}}{d\eta} = -\eta p_{\Lambda|\mathcal{H}_0}(\eta|\mathcal{H}_0)$$

Example 1

- Test for Gaussian signals with different variances We have the following HT problem:

$$\begin{array}{llll} \mathcal{H}_0: & r_i = n_i & n_i \sim \mathcal{N}(0, \sigma_0^2) \;\; i.i.d. & i = 1, 2, \ldots, N \\ \mathcal{H}_1: & r_i = n_i & n_i \sim \mathcal{N}(0, \sigma_1^2) \;\; i.i.d. & i = 1, 2, \ldots, N \end{array}$$

Find the LRT.

- Solution: Let $\boldsymbol{R} = (R_1, R_2, \ldots, R_N)$

$$p_{\mathbf{R}|\mathcal{H}_0}(\mathbf{r}|\mathcal{H}_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{r_i^2}{2\sigma_0^2}\right)$$
$$p_{\mathbf{R}|\mathcal{H}_1}(\mathbf{r}|\mathcal{H}_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{r_i^2}{2\sigma_1^2}\right)$$

so the LRT becomes

$$\Lambda(\mathbf{r}) = \frac{p_{\mathbf{R}|\mathcal{H}_1}(\mathbf{r}|\mathcal{H}_1)}{p_{\mathbf{R}|\mathcal{H}_0}(\mathbf{r}|\mathcal{H}_0)} = \prod_{i=1}^{N} \frac{\sigma_0}{\sigma_1} \exp\left(-\frac{r_i^2}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\right) \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \eta$$

Example 1

- The LLRT becomes

$$\ln(\Lambda(\mathbf{r})) = N \ln\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{\sum_{i=1}^{N} r_i^2}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \ln(\eta)$$

Assuming also $\sigma_1^2 > \sigma_0^2$ we can write the test as

$$\sum_{i=1}^{N} r_i^2 \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} 2\left[\ln\eta - N\ln\left(\frac{\sigma_0}{\sigma_1}\right)\right]\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)^{-1} = \gamma$$

Detectors of the type

$$\sum_{i=1}^{N} r_i^2 \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \gamma$$

are called **energy detectors**.

The term $\sum_{i=1}^{N} r_i^2$ is a *sufficient statistic* for this test.

Example 2

- Radar second case: multiple samples decision

$$\begin{array}{llll} \mathcal{H}_0: & r_i & = & n_i & i = 1, 2, \dots, N \\ \mathcal{H}_1: & r_i & = & A + n_i & i = 1, 2, \dots, N \end{array}$$

where $A > 0$, and n_i's are independent, identically distributed (i.i.d.) random variable (r.v.), $n_i \sim \mathcal{N}(0, \sigma^2)$.

- Solution: Let $\boldsymbol{R} = (R_1, R_2, \dots, R_N)$

$$p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\sum_i r_i^2}{2\sigma^2}\right)$$

$$p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\sum_i (r_i - A)^2}{2\sigma^2}\right)$$

Example 2

- So the LRT becomes

$$\Lambda(\boldsymbol{r}) = \frac{p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1)}{p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)}$$

$$= \exp\left(\frac{2A\sum_i r_i}{2\sigma^2 - NA^2}\right)$$

$$\ln\Lambda(\boldsymbol{r}) = \frac{A}{\sigma^2}\sum_i r_i - \frac{NA^2}{2\sigma^2} \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \ln\eta$$

$\Rightarrow$ LRT is equivalent to (note that $A > 0$)

$$\sum_{i=1}^{N} r_i \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \frac{\sigma^2}{A}\ln\eta + \frac{NA}{2} \triangleq \xi_{th}$$

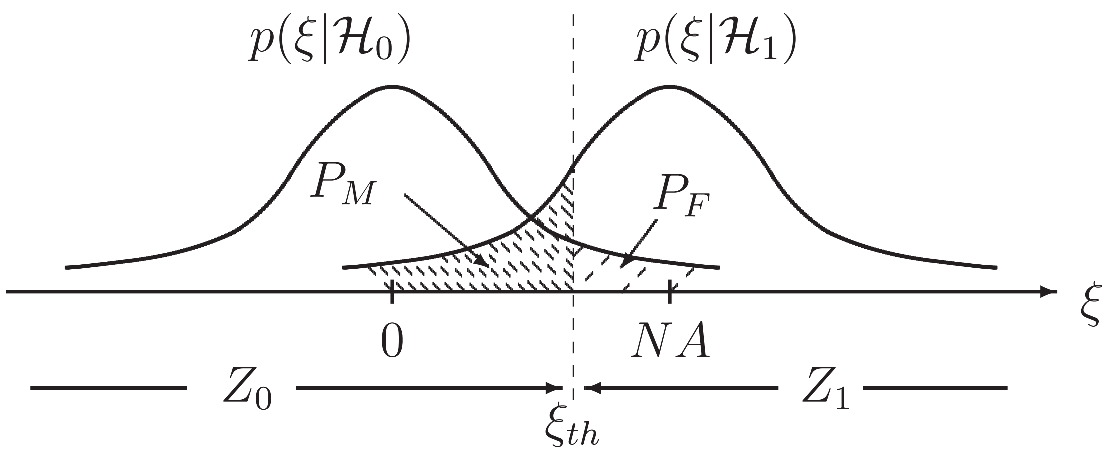
Example 2, the performance

- NOTE: $\xi = \sum_{i=1}^{N} r_i$ is a sufficient statistic for this test.

$$\begin{array}{lllll} \text{if} & \mathcal{H}_0 \Rightarrow R_i \sim \mathcal{N}(0,\sigma^2) & \text{i.i.d.} & \Rightarrow & \xi \sim \mathcal{N}(0, N\sigma^2) \\ \text{if} & \mathcal{H}_1 \Rightarrow R_i \sim \mathcal{N}(A,\sigma^2) & \text{i.i.d.} & \Rightarrow & \xi \sim \mathcal{N}(NA, N\sigma^2) \end{array}$$

$$p_{\xi|\mathcal{H}_0}(\xi|\mathcal{H}_0) = \frac{1}{\sqrt{2\pi N\sigma^2}} \exp\left(-\frac{\xi^2}{2N\sigma^2}\right)$$

$$p_{\xi|\mathcal{H}_1}(\xi|\mathcal{H}_1) = \frac{1}{\sqrt{2\pi N\sigma^2}} \exp\left(-\frac{(\xi - NA)^2}{2N\sigma^2}\right)$$



Example 2, the performance

$$P_{\mathrm{F}} = \int_{\xi_{th}}^{\infty} p_{\xi|\mathcal{H}_0}(\xi|\mathcal{H}_0)d\xi = Q\left(\frac{\xi_{th}}{\sqrt{N}\sigma}\right)$$

$$P_{\mathrm{M}} = \int_{-\infty}^{\xi_{th}} p_{\xi|\mathcal{H}_1}(\xi|\mathcal{H}_1)d\xi = 1 - Q\left(\frac{-NA+\xi_{th}}{\sqrt{N}\sigma}\right)$$

- If we define the signal-to-noise ration (SNR) as

$$SNR \triangleq \frac{NA^2}{\sigma^2}$$

Then

$$P_{\mathrm{F}} = Q\left(\frac{\ln\eta}{\sqrt{SNR}} + \frac{\sqrt{SNR}}{2}\right)$$

and

$$P_{\mathrm{D}} = 1 - P_{\mathrm{M}} = Q\left(\frac{\ln\eta}{\sqrt{SNR}} - \frac{\sqrt{SNR}}{2}\right)$$