ESE 524: Detection and Estimation Theory

Recitation 5

Washington University in St. Louis

Outline

- Bayesian detection
- Neyman-Pearson detection
- ROC curves
- Likelihood ratio test

Useful Formulas

- Decision rule:

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \text{decide } \mathcal{H}_1, \\ 0, & \text{decide } \mathcal{H}_0. \end{cases}$$

Rule $\phi(\boldsymbol{x}) : \mathcal{X}_0 = \{\boldsymbol{x} : \phi(\boldsymbol{x}) = 0\}, \quad \mathcal{X}_1 = \{\boldsymbol{x} : \phi(\boldsymbol{x}) = 1\}.$

- False alarm probability:

$$P_{\mathrm{FA}} = \mathrm{E}_{\boldsymbol{x}|\theta}[\phi(\boldsymbol{x})|\theta] = \int_{\mathcal{X}_1} p(\boldsymbol{x}|\theta) d\boldsymbol{x} \quad \text{for } \theta \text{ in } \Theta_0$$

- Miss probability:

$$P_{\mathrm{M}} = \mathrm{E}_{\boldsymbol{x}|\theta}[1 - \phi(\boldsymbol{x})|\theta] = 1 - \int_{\mathcal{X}_1} p(\boldsymbol{x}|\theta) d\boldsymbol{x}$$

- Detection probability:

$$P_{\mathrm{D}} = 1 - P_{\mathrm{M}} = \int_{\mathcal{X}_1} p(\boldsymbol{x}|\theta) d\boldsymbol{x} \quad \text{for } \theta \text{ in } \Theta_1$$

Useful Formulas (Cont.)

- Likelihood ratio:

$$\underbrace{\Lambda(\boldsymbol{x})}_{\text{likelihood ratio}} = \frac{p(\boldsymbol{x}|\theta_1)}{p(\boldsymbol{x}|\theta_0)} \overset{\mathcal{H}_1}{\gtrless} \frac{\pi_0 L(1|0)}{\pi_1 L(0|1)} \equiv \tau$$

- Neyman-Pearson detection:

$$\text{Maximize } P_{\text{D}} = P[\boldsymbol{x} \in \mathcal{X}_1; \theta = \theta_0]$$

(equivalently, minimizes the miss probability P_M) under the constraint:

$$P_{\text{FA}} = P[\boldsymbol{x} \in \mathcal{X}_1; \theta = \theta_0] = \alpha' \leq \alpha$$

Frequentist/Neyman-Pearson Process

- Choose an upper bound on the probability of false alarm α, for example $\alpha = 5\% = 0.05$.
- Formulate Likelihood Ratio $\Lambda(\boldsymbol{x}) = \frac{p(\boldsymbol{x};\theta_1)}{p(\boldsymbol{x};\theta_2)} \gtreqless \lambda$.
- Find the critical value of the likelihood ratio λ by solving

$$\int_{\boldsymbol{x}:\Lambda(\boldsymbol{x})>\lambda} p(\boldsymbol{x};\theta_0)d\boldsymbol{x} = \alpha$$

  for lambda.
- This can be difficult (i.e. solving integral equations, or for complicated likelihood functions).
- It's usually easier to fix a λ and calculate the probabilities of false alarm and detection.

Receiver Operating Characteristic

- $P_{\mathrm{D}}(\lambda) = \int_{\boldsymbol{x}:\Lambda(\boldsymbol{x})>\lambda} p(\boldsymbol{x};\theta_1)d\boldsymbol{x}$ - Probability of detection is the y-axis.
- $P_{\mathrm{FA}}(\lambda) = \int_{\boldsymbol{x}:\Lambda(\boldsymbol{x})>\lambda} p(\boldsymbol{x};\theta_0)d\boldsymbol{x}$ - Probability of false alarm is the x-axis.
- For different λ's plot the pairs $(P_{\mathrm{FA}}, P_{\mathrm{D}})$
- This curve is called the Receiver Operating Characteristic, which comes from radar.

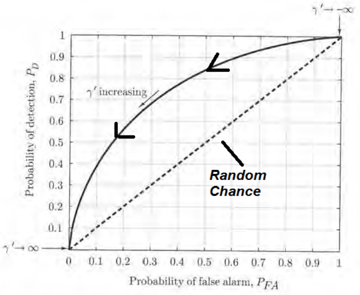

Figure 1: The ROC curve for the DC Level with Additive White Gaussian Noise Example. Note that as $\lambda = -\infty$, $P_{\mathrm{D}} = P_{\mathrm{FA}} = 1$.

Desirable ROCs

- The goal is to maximize P_{D} while minimizing P_{FA}.
- This means we want the ROC Curve to pull towards the top left of the plot.
- Curves below the random chance line have higher P_{FA} than P_{D}.

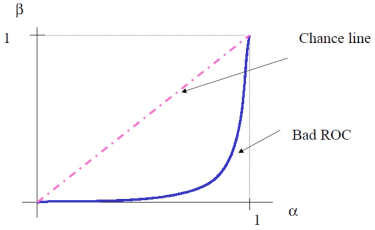

Figure 2: A poorly performing ROC curve.

ROC Curve, NP, and Bayesian Detection

- The Bayesian test sets $\lambda = \frac{P(H_0)}{P(H_1)}$
- The Neyman-Pearson Likelihood Ratio Test sets a specific P_{FA} and then finds λ.
- Generally, we choose the threshold that minimizes $(1 - P_{\mathrm{D}})^2 + P_{\mathrm{FA}}^2$.
- In the example of Additive Gaussian White Noise, the Bayesian approach increase its probability of detection at the expense of false positives, but the NP approach generally results in very low P_{D}.

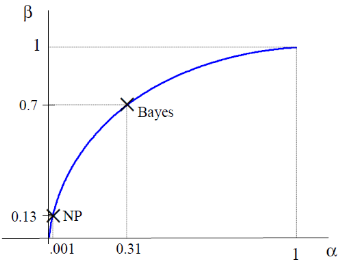

Figure 3: ROC Curve for Additive White Gaussian Noise example, here $\alpha = 0.001$

Likelihood Ratio for Linear Models

- Recall the linear model:

$$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{w}$$

- $\boldsymbol{w} = [w[0], \ldots, w[N-1]]^T$ is a vector of i.i.d. samples with joint pdf $\boldsymbol{w} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$
- $\boldsymbol{H}$ is the known model matrix.
- We want to test the simple hypotheses:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{0}$$

$$H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1, \text{ known}$$

- Here we assume we know $\boldsymbol{\theta_1}$, whereas before we had attempted to estimate it.

Likelihood Ratio

- The likelihood ratio is:

$$\frac{f(x|H_1)}{f(x|H_0)} = \frac{\frac{1}{(2\pi)^{N/2}\sqrt{\det(\sigma^2 \boldsymbol{I})}} \exp[-1/(2\sigma^2)(\boldsymbol{x}-\boldsymbol{H}\boldsymbol{\theta}_1)^T(\boldsymbol{x}-\boldsymbol{H}\boldsymbol{\theta}_1)]}{\frac{1}{(2\pi)^{N/2}\sqrt{\det(\sigma^2 \boldsymbol{I})}} \exp[-1/(2\sigma^2)\boldsymbol{x}^T\boldsymbol{x}]} \overset{H_1}{\gtrless} \lambda$$

- This simplifies, after some work to:

$$t(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{H}\boldsymbol{\theta}_1/\sigma^2 \overset{H_1}{\gtrless} \ln\lambda + \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H}\boldsymbol{\theta}_1}{2\sigma^2}$$

- $t(\boldsymbol{x})$ is a normal random variable, so for the NP test we need to find the means and variances under each hypothesis to get P_{D} and P_{FA}.

Mean and Variance Under H_0

- For H_0 use $\boldsymbol{x} = \boldsymbol{w}$:

$$\mathrm{E}(t(\boldsymbol{x})|H_0) = \mathrm{E}\left(\frac{\boldsymbol{w}^T \boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2}\right) = 0$$

$$\mathrm{var}(t(\boldsymbol{X})|H_0) = \mathrm{E}\left[\left(\frac{\boldsymbol{w}^T \boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2}\right)^2\right] = \mathrm{E}\left(\frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T}{\sigma^2}\boldsymbol{w}\boldsymbol{w}^{\boldsymbol{T}}\frac{\boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2}\right)$$

$$= \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T}{\sigma^2}\mathrm{E}(\boldsymbol{w}\boldsymbol{w}^T)\frac{\boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2} = \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T}{\sigma^2}\sigma^2\boldsymbol{I}\frac{\boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2} = \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2}$$

- So $f(t(x)|H_0) \sim \mathcal{N}(0, \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2})$

Probability of False Alarm

- Let $\lambda' = \ln \lambda + \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{\theta}_1}{2\sigma^2}$
- Then the probability of false alarm is:

$$\begin{aligned}
P_{\mathrm{FA}} &= P(t(\boldsymbol{x}) > \lambda' | H_0) \\
&= P\left(\underbrace{\frac{(t(\boldsymbol{x}) - 0)\sigma^2}{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{\theta}_1}}_{N(0,1)} > \frac{(\lambda' - 0)\sigma^2}{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{\theta}_1}\right) \\
&= 1 - \Phi\left(\frac{(\ln \lambda + \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{\theta}_1}{2\sigma^2})\sigma^2}{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{\theta}_1}\right)
\end{aligned}$$

- Φ is the cumulative distribution function for the standard normal distribution.

Mean and Variance Under H_1

- For H_1 use $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{\theta}_1 + \boldsymbol{w}$:

$$\mathrm{E}(t(\boldsymbol{x})|H_1) = \mathrm{E}\left(\frac{(\boldsymbol{H}\boldsymbol{\theta}_1 + \boldsymbol{w})^T \boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2}\right) = \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2}$$

- The variance is a bit more complicated:

$$\begin{aligned}
\mathrm{var}(t(\boldsymbol{x})|H_1) &= \mathrm{E}\left\{\left[\frac{\boldsymbol{x}^T \boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2} - \mathrm{E}\left(\frac{\boldsymbol{x}^T \boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2}\right)\right]^2\right\} \\
&= \mathrm{E}\left\{\left([\boldsymbol{x} - \mathrm{E}(x)]^T \frac{\boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2}\right)^2\right\} \\
&= \mathrm{cov}(\boldsymbol{x})\left(\frac{\boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2}\right)^2 = \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2} = \mathrm{var}(t(\boldsymbol{x})|H_0)
\end{aligned}$$

- So under H_1 $t(\boldsymbol{x}) \sim \mathcal{N}\left(\frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2}, \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H}\boldsymbol{\theta}_1}{\sigma^2}\right)$

- The probability of detection is given by:

$$
\begin{aligned}
P_{\mathrm{D}} &= P(t(\boldsymbol{x}) > \lambda' | H_1) \\
&= P\left(\underbrace{\frac{\left(t(\boldsymbol{x}) - \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{\theta}_1}{\sigma^2} \right) \sigma^2}{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{\theta}_1}}_{N(0,1)} > \frac{\left(\lambda' - \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{\theta}_1}{\sigma^2} \right) \sigma^2}{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{\theta}_1} \right) \\
&= 1 - \Phi\left(\frac{\left(\ln\lambda - \frac{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{\theta}_1}{2\sigma^2} \right) \sigma^2}{\boldsymbol{\theta}_1^T \boldsymbol{H}^T \boldsymbol{H} \boldsymbol{\theta}_1} \right)
\end{aligned}
$$

- This is almost the same as P_{FA}, with just a subtraction instead of addition.
- Large λ leads to higher P_{D} and P_{FA}
- $P_{\mathrm{D}} = \Phi^{-1}(\Phi(P_{\mathrm{FA}}) - \sqrt{\frac{\boldsymbol{\theta}_1^T \boldsymbol{\theta}_1}{\sigma^2}})$ is another what to express P_{D} in this case - proof is in Kay CH. 3 and 4.

Example 1: Sum of Sinusoids, Kay ex. 4.9

- $H_0 : x[n] = w[n]$
- $H_1 : \boldsymbol{x}[n] = a\cos(2\pi f_0 n) + b\sin(2\pi f_0 n) + w[n]$
- $\boldsymbol{\theta}_1 = \begin{bmatrix} a \\ b \end{bmatrix}$
- $\boldsymbol{H} = \begin{bmatrix} 1 & 0 \\ \cos(2\pi f_0) & \sin(2\pi f_0) \\ \vdots & \vdots \\ \cos(2\pi f_0(N-1)) & \sin(2\pi f_0(N-1)) \end{bmatrix}$
- For this example, let $N = 2$.

Sum of Sinusoids, Kay ex. 4.9 (Cont.)

- The test statistic is given by:

$$\begin{aligned}
\frac{1}{\sigma^2}\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{H}\boldsymbol{\theta}_1 &= \frac{1}{\sigma^2}\begin{bmatrix} x[0] & x[1] \end{bmatrix}\begin{bmatrix} 1 & 0 \\ \cos(2\pi f_0) & \sin(2\pi f_0) \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} \\
&= \frac{1}{\sigma^2}\{ax[0] + x[1][a\cos(2\pi f_0) + b\sin(2\pi f_0)]\} \\
&= \frac{1}{\sigma^2}\{a[x[0] + x[1]\cos(2\pi f_0)] + b[\sin(2\pi f_0)]\} \\
&= a\hat{a} + b\hat{b}
\end{aligned}$$

- Here $\hat{a}$ and $\hat{b}$ are the estimators of a and b based on the Fourier series formula!
- Under H_0, $\hat{a}$ and $\hat{b}$ are very small, so $t(x)$ is small.
- Under H_1, $t(\boldsymbol{x}) \approx a^2 + b^2$. This is proportional to the signal power.

Comments

- What happens when $\boldsymbol{\theta}_1$ is unknown?
- We have to use the MLE estimator in something called the Generalized Likelihood Ratio Test.
- The hypotheses are slightly different:

$$H_0 : \boldsymbol{A}\boldsymbol{\theta} = \boldsymbol{b}$$

$$H_1 : \boldsymbol{A}\boldsymbol{\theta} \neq \boldsymbol{b}$$

- This hypothesis test whether or not $\boldsymbol{\theta}$ lives in a particular subspace (e.g. line, plane, hyperplane)
- $\boldsymbol{A}$, $\boldsymbol{b}$ are known and form a consistent (solvable) set of equations.
- Then you apply similar formulas, but the covariance matrix of t changes because $\hat{\boldsymbol{\theta}}$ is a function of $\boldsymbol{x}$ and therefore a random variable.
- P_{D} and P_{FA} become χ^2 distributions.