CS/CNS/EE/IDS 165 Winter 2021
Instructor: Prof. A. Anandkumar Caltech

Recitation 3: Bayesian Detection and Neyman Pearson Detection

1 **(Biased coin)** This problem generalizes the example given in the class. Consider two biased coins $\mathcal{H}_0$ and $\mathcal{H}_1$. Coin $\mathcal{H}_0$ has probability θ_0 to show "head", and coin $\mathcal{H}_1$ has probability $1-\theta_1$ to show "head". Modelling the coin flipping as a Bernuoulii trial with $Y=1$ indicating head and $Y=0$ tail, we consider the following binary hypotheses

$$\mathcal{H}_0 : p(y;\theta_0) = \begin{cases} 1-\theta_0 & y=0 \\ \theta_0 & y=1 \end{cases} \qquad \mathcal{H}_1 : p(y;\theta_1) = \begin{cases} \theta_1 & y=0 \\ 1-\theta_1 & y=1 \end{cases}$$

where θ_i are known. We assume that the prior probability that $\mathcal{H}_0$ is chosen is π, and we impose the uniform cost, *i.e.*, $C_{ii}=0$ for $i=0,1$, and $C_{ij}=1$ for $i \neq j$.

(a) Suppose that a single coin flip is performed with observation $Y=y$. Find the Bayesian detector $\delta(y)$ as a function of π, θ_i.

(b) Plot the minimum Bayesian risk $V(\pi)$ as the function of π for $\theta_0 = \frac{1}{4}$ and $\theta_1 = \frac{1}{3}$. Illustrate that the curve is concave.

(c) Show that, in general, the function $V(\pi)$ is always concave.

Solution: (a) From the simple binary hypothesis test:

$$\delta(y=1) = \begin{cases} 1 & \text{if } \frac{p(y=1|\theta_1)}{p(y=1|\theta_0)} = \frac{1-\theta_1}{\theta_0} \geq \frac{\pi}{1-\pi} \\ 0 & \text{otherwise} \end{cases}$$

$$\delta(y=0) = \begin{cases} 1 & \text{if } \frac{p(y=0|\theta_1)}{p(y=0|\theta_0)} = \frac{\theta_1}{1-\theta_0} \geq \frac{\pi}{1-\pi} \\ 0 & \text{otherwise} \end{cases}$$

(b) For $\theta_0 = 1/4$ and $\theta_1 = 1/3$, the detector is

$$\delta(y=1) = \begin{cases} 1 & \text{if } \pi \leq 8/11 \\ 0 & \text{otherwise} \end{cases}$$

$$\delta(y=0) = \begin{cases} 1 & \text{if } \pi \leq 4/13 \\ 0 & \text{otherwise} \end{cases}$$

Case 1: $\pi \leq 4/13$. We have $\delta(y)=1$. Therefore

$$\begin{aligned} V(\pi) &= \pi P_r(\delta(y)=1|H_0) + (1-\pi)P_r(\delta(y)=0|H_1) \\ &= \pi. \end{aligned}$$

Case 2: $4/13 < \pi \le 8/11$. We have $\delta(y = 1) = 1$ and $\delta(y = 0) = 0$. Therefore

$$\begin{aligned} V(\pi) &= \pi P_r(\delta(y) = 1|H_0) + (1-\pi)P_r(\delta(y) = 0|H_1) \\ &= \pi\theta_0 + (1-\pi)\theta_1 \\ &= 1/3 - \pi/12. \end{aligned}$$

Case 3: $\pi > 8/11$. We have $\delta(y) = 0$. Therefore

$$\begin{aligned} V(\pi) &= \pi P_r(\delta(y) = 1|H_0) + (1-\pi)P_r(\delta(y) = 0|H_1) \\ &= 1 - \pi. \end{aligned}$$

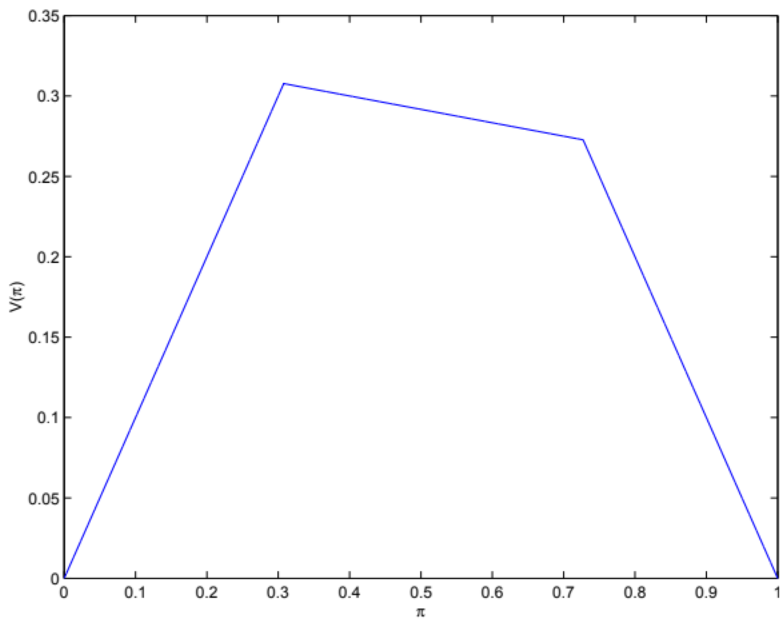

Figure 1: Question 1(b).

As shown in Fig. 1, the curve is concave.

(c) As shown in the three cases of (b), the Bayesian detector may be different for different π. For a given π, denote $\delta_\pi(y)$ the Bayesian detector. The Bayesian risk is given by

$$V(\pi) = \pi P_r(\delta_\pi(y) = 1|H_0) + (1-\pi)P_r(\delta_\pi(y) = 0|H_1).$$

Since the Bayesian detector δ_π gives the minimum Bayesian risk $V(\pi)$, we have for all π'

$$V(\pi) \le \pi P_r(\delta_{\pi'}(y) = 1|H_0) + (1-\pi)P_r(\delta_{\pi'}(y) = 0|H_1) \tag{1}$$

where the right hand side is the risk associated with detector $\delta_{\pi'}$. Therefore, for $\pi_1 < \pi_0 < \pi_2$,

$$
\begin{aligned}
&\frac{\pi_0 - \pi_1}{\pi_2 - \pi_1} V(\pi_2) + \frac{\pi_2 - \pi_0}{\pi_2 - \pi_1} V(\pi_1) \\
&\leq \frac{\pi_0 - \pi_1}{\pi_2 - \pi_1} \Big(\pi_2 P_r(\delta_{\pi_0}(y) = 1 | H_0) + (1 - \pi_2) P_r(\delta_{\pi_0}(y) = 0 | H_1) \Big) \\
&\quad + \frac{\pi_2 - \pi_0}{\pi_2 - \pi_1} \Big(\pi_1 P_r(\delta_{\pi_0}(y) = 1 | H_0) + (1 - \pi_1) P_r(\delta_{\pi_0}(y) = 0 | H_1) \Big) \\
&= \pi_0 P_r(\delta_{\pi_0}(y) = 1 | H_0) + (1 - \pi_0) P_r(\delta_{\pi_0}(y) = 0 | H_1) \\
&= V(\pi_0).
\end{aligned}
$$

The above inequality indicates that $V(\pi)$ is concave.

2 **(Linear detector)** Consider binary hypotheses

$$
\mathcal{H}_0 : \; Y_i \sim p_0(y) = N_i \quad \text{vs.} \quad \mathcal{H}_1 : \; Y_i = (-1)^i + N_i
$$

where $i = 1, \cdots, n$, and $N_i \sim \mathcal{N}(0, 1)$ is an i.i.d. Gaussian sequence. For a sequence of observations $Y_i = y_i$, a linear detector performs the test

$$
\sum_{i=1}^{n} h_i y_i \gtrless \tau
$$

where h_i and τ are coefficients to be determined. For any cost and prior, is the optimal Bayesian detector linear? Give justifications.

Solution: The likelihood ratio is given by

$$
\begin{aligned}
L(y^n) &= \frac{\frac{1}{(2\pi)^{n/2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (y_i - (-1)^i)^2 \right\}}{\frac{1}{(2\pi)^{n/2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} y_i^2 \right\}} \\
&= \exp\left\{ -\frac{1}{2} \left(\sum_{i=1}^{n} (y_i - (-1)^i)^2 - \sum_{i=1}^{n} y_i^2 \right) \right\} \\
&= \exp\left\{ \sum_{i=1}^{n} (-1)^i y_i - \frac{n}{2} \right\}.
\end{aligned}
$$

Since we are dealing with a simple binary hypothesis, the optimal Bayesian detector is

$$
\delta(y^n) = \begin{cases} 1 & \text{if } L(y^n) > \frac{\pi_0 (C_{10} - C_{00})}{\pi_1 (C_{01} - C_{11})} \iff \sum_{i=1}^{n} (-1)^i y_i > \ln\left(\frac{\pi_0 (C_{10} - C_{00})}{\pi_1 (C_{01} - C_{11})} \right) + \frac{n}{2} \\ 0 & \text{o.w.} \end{cases}
$$

Therefore, the optimal Bayesian detector is linear for any cost and prior, where $h_i = (-1)^i$ and

$$\tau = \ln\left(\frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})}\right) + \frac{n}{2}.$$

3 (**Gaussian vs. uniform**) Consider the following simple binary hypotheses:

$$\mathcal{H}_0 : Y \sim \mathcal{N}(0, \sigma^2) \quad \text{vs.} \quad \mathcal{H}_1 : Y \sim \mathcal{U}(-\sqrt{3}\sigma, \sqrt{3}\sigma)$$

The two distributions both have zero mean and variance σ^2.

(a) First think intuitively what the detection rule should be like. For example, how does it partition the space of the observation? Derive the Neyman-Pearson detector for given α. Verify your intuition.

(b) Plot the ROC curve (for $\sigma = 1$). How does the curve change when σ^2 increases?

Solution:

(a) The likelihood ratio will be 0 for $|y| > \sqrt{3}\sigma$, because of the uniform distribution. Inside its nonzero region, it will peak at the edges (i.e. at $y = \pm\sqrt{3}\sigma$) and droop in the middle, as the Gaussian distribution peaks at $y = 0$. Thus, putting a threshold on the likelihood ratio cuts the line into five regions: two symmetrical intervals above the threshold that end at $\pm\sqrt{3}\sigma$, and the other three intervals below the threshold, two of which are the infinite intervals beyond $\pm\sqrt{3}\sigma$ and the other an interval centered at $y = 0$. Hence, the Neyman-Pearson detector has the form

$$\delta y = \begin{cases} 1 & \text{if } \tau < |y| < \sqrt{3}\sigma \\ 0 & \text{otherwise.} \end{cases}$$

To find τ in terms of the size α, we first find the probability of false alarm for this detector:

$$P_F(\delta) = \Pr\left(\tau < |y| < \sqrt{3}\sigma; \mathcal{H}_0\right) = 2\left(Q\left(\frac{\tau}{\sigma}\right) - Q(\sqrt{3})\right).$$

Setting this to α yields

$$\tau = \sigma Q^{-1}\left(Q(\sqrt{3}) + \frac{\alpha}{2}\right).$$

(b) The detection probability for this detector is

$$P_D(\delta) = \Pr\left(\tau < |y| < \sqrt{3}\sigma; \mathcal{H}_1\right) = \frac{\sqrt{3}\sigma - \tau}{\sqrt{3}\sigma} = 1 - \frac{1}{\sqrt{3}} Q^{-1}\left(Q(\sqrt{3}) + \frac{\alpha}{2}\right). \tag{2}$$

The ROC curve is shown in Figure 2. The curve does not change when σ^2 increases, since (2) is independent of σ.

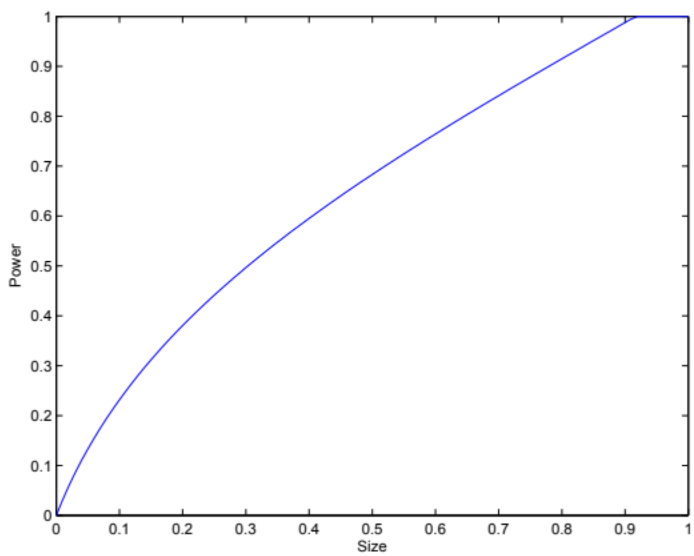

Figure 2: ROC curve for the Gaussian vs Uniform detector.