E1–244 **Detection and Estimation Theory**

Discussion: 5:00 - 6:00 PM 23 February, 02 March 2015
Due: Tuesday 9:00 AM, 03 March 2015

Problem Set 3

Instructor: Rajesh Sundaresan *TA: TBA*

Problems:

1. Problem 3 in Section III.F
2. Problem 5 in Section III.F
3. Problem 6 in Section III.F.
4. Consider the composite hypothesis testing problem

$$\begin{aligned} H_0 &: Y_k = Z_k, \quad k = 1, \ldots, n \\ &\text{versus} \\ H_1 &: Y_k = \sqrt{\theta} S_k + Z_k, \quad k = 1, \ldots, n, \quad \theta > 0. \end{aligned}$$

   $Z \sim N(0, \sigma^2 I_n)$ and $S \sim N(0, \Sigma_S)$. Why does not a UMP exist? Show that the LMP test statistic may be taken as $T(y) = \frac{1}{n} y^T \Sigma_S y$ after scaling.

5. Continue with the previous problem. Suppose that $(\Sigma_S)_{k,l} = \rho_{k-l}$, i.e., the signal is wide sense stationary. With

$$\hat{\rho}_k := \frac{1}{n} \sum_{l=1}^{n-k} y_l y_{l+k}, \quad k = 0, \ldots, n-1,$$

   argue that $T(y)$ may be interpreted as a correlation of correlations, i.e.,

$$T(y) = \rho_0 \hat{\rho}_0 + 2 \sum_{k=1}^{n-1} \rho_k \hat{\rho}_k.$$

6. Consider the coherent detection problem

$$\begin{aligned} H_0 &: Y_k = Z_k, \quad k = 1, \ldots, n \\ &\text{versus} \\ H_1 &: Y_k = s_k + Z_k, \quad k = 1, \ldots, n \end{aligned}$$

   where $s_1, \ldots, s_n$ are known, and Z_k is iid Cauchy. What is the characteristic function of the log-likelihood ratio $T(y)$ under H_0 and under H_1? Using this (or otherwise), and with the critical region being $\Gamma_1 = \{y \mid T(y) > \tau\}$, find the false alarm and miss probabilities.

7. Suppose that the vectors Y and Θ are jointly Gaussian, i.e.,

$$\begin{pmatrix} Y \\ \Theta \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_Y \\ \mu_\Theta \end{pmatrix}, \begin{pmatrix} K_Y & K_{Y\Theta} \\ K_{\Theta Y} & K_\Theta \end{pmatrix} \right).$$

   (We must obviously have $K_{Y\Theta} = K_{\Theta Y}^t$. Prove that given $Y = y$, the vector Θ is also Gaussian with conditional mean $\hat{\mu}(y)$ and covariance $\hat{K}$ given by

$$\begin{aligned} \hat{\mu}(y) &= \mu_\Theta + K_{\Theta Y} K_Y^{-1} (y - \mu_Y) \\ \hat{K} &= K_\Theta - K_{\Theta Y} K_Y^{-1} K_{Y\Theta}. \end{aligned}$$

   Compare $\hat{K}$ and K_Θ and interpret.

8. Suppose $Y = H\Theta + Z$ where $\Theta \sim N(\mu_\Theta, K_\Theta)$ and $Z \sim N(0, K)$. Find the conditional mean and conditional variance of Θ given $Y = y$ using the above formula.

9. Verify the matrix inverse lemma (Sherman-Morrison-Woodbury formula):

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}.$$

where B and D are of appropriate sizes, and A and C are square matrices of dimension n and k respectively. What do you get when $k = 1$?

10. **Two-sample t-test**. Suppose that $X_1, X_2, \ldots, X_n$ are iid $N(\mu_1, \sigma^2)$, and that $Y_1, Y_2, \ldots, Y_m$ are iid $N(\mu_2, \sigma^2)$. The Y samples are independent of the X samples. The two distributions have the same but unknown variance. We would like to test

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ &\text{versus} \\ H_1 &: \mu_1 \neq \mu_2. \end{aligned}$$

(i) Show that the generalised likelihood ratio test is based on the following statistic:

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{S_p^2 \left(\frac{1}{n} + \frac{1}{m}\right)}},$$

where

$$S_p^2 = \frac{1}{(n+m-2)} \left(\sum_{i=1}^{n} (X_i - \overline{X})^2 + \sum_{j=1}^{m} (Y_j - \overline{Y})^2 \right),$$

the so-called *pooled variance estimate*, and $\overline{X}, \overline{Y}$ are the sample means. (Later we will see the reason for $n + m - 2$.)

(ii) Under H_0, what is the distribution of T?

(iii) Until I get a data set closer to home, here's a readily available data set that we will analyse. This old data set, gathered some years back, consists of miles per gallon for U.S. cars and Japanese cars. It's taken from the nist.gov website and is made available here for easy reference.

http://www.ece.iisc.ernet.in/~rajeshs/E1244/mileage_data_ps3.txt

Column 1 is miles per gallon for U.S. cars; Column 2 is the miles per gallon for Japanese cars. A value -999 in the second column informs us that fewer Japanese cars were sampled; ignore these values.

At 5% significance level, should we accept or reject the null hypothesis that the mean miles per gallon for U.S. and Japanese cars are equal?

(Note: If you haven't solved (ii), you could use a Monte Carlo method to generate an empirical cdf of the statistic under the null hypothesis. How robust is this cdf to the unknown variance?)