Stochastic Processes, Session 15 — Group Work
Hypothesis Testing and Detection Theory

Carles Navarro Manchón and Troels Pedersen Aalborg University, October 2016

Session 15

Go through the exercises below. Allow yourself the time to reflect over your results and discuss them with other students! Use the book for inspiration and for further information.

15.1 Normal or Uniform?

Consider the following binary hypothesis testing problem:

$$\begin{array}{lll} h_0: \ Y \sim \mathcal{U}(-1,1) & f_{Y|H}(y|h_0) & \mathbb{P}(H=h_0)=p \\ h_1: \ Y \sim \mathcal{N}(0,1) & f_{Y|H}(y|h_1) & \mathbb{P}(H=h_1)=1-p \end{array}$$

a) What is the range of Y when $H = h_0$? And when $H = h_1$? Given this, what should your immediate decision be if an observation $Y = y$ is such that $|y| > 1$?

b) Write a program that plots in the same figure $f_{Y|H}(y|h_k)\mathbb{P}(H = h_k)$ as a function of y for $k = 0, 1$ and a set value p. Try different values of p and discuss how the MAP decision regions change with p. You should be able to identify three different situations.

c) Set $p = 0.4$. Calculate the decision regions for the MAP rule, i.e. find the value of $\hat{H}_{\mathrm{MAP}}(y)$ for all possible values of y.

15.2 Detection of Exponentially Distributed Variables

The conditional pdf's corresponding to two hypotheses are given:

$$\begin{aligned} f_{X|H}(x|h_0) &= \tfrac{1}{2}\exp\left(-\tfrac{x}{2}\right), \quad x \geq 0 \\ f_{X|H}(x|h_1) &= \tfrac{1}{4}\exp\left(-\tfrac{x}{4}\right), \quad x \geq 0. \end{aligned} \tag{1}$$

Suppose we want to test these hypotheses based on two independent samples X_1 and X_2, i.e. we observe the vector $\boldsymbol{x} = [x_1, x_2]^{\mathrm{T}}$ and we have to decide upon either h_0 or h_1. Assume equally likely a priori probabilities for the true hypothesis H.

a) Derive the MAP decision rule based on observing $\boldsymbol{x} = [x_1, x_2]^{\mathrm{T}}$.

b) Sketch the decision regions in the 2D plane with axes (x_1, x_2). How do the decision regions change if you now assume unequal probabilities for the two hypotheses $(\mathbb{P}(H = h_0) \neq \mathbb{P}(H = h_1))$?

c) How do the decision regions look like if you add a third hypothesis $f_{X|H}(x|h_2) = \frac{1}{8}\exp\left(-\frac{x}{8}\right), \quad x \geq 0$? (Consider again equal a priori probabilities for the hypotheses).

d) For the original problem (1) with equal a priori probabilities, implement a program that draws realizations of the vector $\boldsymbol{X} = [X_1, X_2]^{\mathrm{T}}$ and applies the MAP decision rule. Does it provide the correct decision often?

e) What is the probability that $\hat{H}_{\mathrm{MAP}}(\boldsymbol{X}) = h_0$ when $H = h_1$? And the opposite, $\mathbb{P}(\hat{H}_{\mathrm{MAP}}(\boldsymbol{X}) = h_1 | H = h_0)$? Calculate the decision error probability P_e.

f) To check that your calculations are correct, estimate the decision error probability by running Monte Carlo simulations of the program you implemented. Estimate as well the probability of each individual type of error.

15.3 Back to the Lab?

A group of students of Electrical Engineering at Aalborg University are doing a semester project on electronics. Among other tasks, they have performed some laboratory experiments by measuring the digitalized response of an electronic circuit excited with a voltage source of 12V. According to their model of the circuit, they expect the samples $\boldsymbol{x} = [x_1, x_2, x_3]^{\mathrm{T}}$ they have recorded to behave according to the stochastic model

$$\boldsymbol{X} = [X_1, X_2, X_3]^{\mathrm{T}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{2}$$

with

$$\boldsymbol{\mu} = [12, 12, 12]^{\mathrm{T}}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.7 & 0.49 \\ 0.7 & 1 & 0.7 \\ 0.49 & 0.7 & 1 \end{bmatrix}.$$

In the experiment they carried out, they obtained a measured response of $\boldsymbol{x} = [10.2, 9.9, 12.5]^{\mathrm{T}}$.

Weeks after the experiment, when preparing the project report, one of the students comes to the group room with grim news. He has been examining the voltage sources they had used for the experiment, and has found out that one of the 4 sources they have is malfunctioning. The faulty source only provides an output voltage of 10V, which would provide an output of the experiment distributed as in (2), but with a mean vector $\boldsymbol{\mu}' = [10, 10, 10]^{\mathrm{T}}$. Since they don't know which of the four sources they used, the students estimate that there is a probability of 1/4 that they used the faulty source in the experiment.

The students are in a tight spot. Repeating the experiment would take a full week of work, which would delay their report writing. However, if they used the faulty source in their experiments and don't correct it, they risk failing the project and having to repeat it next semester. Can you help them decide what to do?

a) Based on the output of the experiment, the students could guess whether they used or not the faulty source. Model the problem as a hypothesis testing problem and find the MAP hypothesis for the observed output $\boldsymbol{x} = [10.2, 9.9, 12.5]^{\mathrm{T}}$.

b) What is the probability that, given $\boldsymbol{x}$, the source used by the students was the faulty one?

The students are not yet convinced. They feel that the probability of having used the faulty source is too large to disregard. They would like you to factor in the costs of making the wrong decisions, so they estimate them in terms of time. They assume that making the right decision has no cost for them. If they mistakenly decide to perform the experiment again, they estimate a cost of 37 hours (1 week of work). On the other hand, if they decide not to repeat it and the results are incorrect, they estimate a cost of 600 hours (redoing a 20 ECTS project next semester).

c) Write up the cost matrix for this decision problem, and formulate the corresponding Bayes' decision rule.

d) Using Bayes' decision rule, what would the students decision be given the output of the experiment $\boldsymbol{x} = [10.2, 9.9, 12.5]^{\mathrm{T}}$?

(e) How small should the a priori probability of having used the faulty source be for the students to avoid repeating the experiment? (Hint: you can plot the cost of each decision given the data as a function of the a priori probability).