CS 8850: Advanced Machine Learning **Fall 2017**

Homework 4: Hypotheses Testing & Parameter Estimation

INSTRUCTOR: DANIEL L. PIMENTEL-ALARCÓN DUE 10/11/2017

Problem 4.1 (Oximetry)**.** Congratulations! You have just been hired to develop a new pulse oximetry device. Pulse oximetry is a technology based on light sensors that can non-invasively measure the oxygenation in a person's blood. The sensors are usually placed on a fingertip or earlobe, they transmit light, and measure how it bounces back. Because oxygenated and de-oxygenated hemoglobin have different colors, differences in the absorption of light indicate whether the blood is oxygenated or de-oxygenated.

The company is experimenting with two types of sensors:

- **Photon-counting sensor.** This sensor counts the number x of photons it detects, according to:

$$H_0 : \ x \ \sim \ \mathsf{P}_0(x = \mathrm{k}) = \frac{\lambda_0^{\mathrm{k}} e^{-\lambda_0}}{\mathrm{k}!} \Rightarrow \text{blood is de-oxygenated,}$$
$$H_1 : \ x \ \sim \ \mathsf{P}_1(x = \mathrm{k}) = \frac{\lambda_1^{\mathrm{k}} e^{-\lambda_1}}{\mathrm{k}!} \Rightarrow \text{blood is oxygenated,}$$

  where $\lambda_0 = 10$ and $\lambda_1 = 15$.

- **Light-intensity sensor.** This sensor measures the integrated light intensity x, according to:

$$H_0 : \ x \ \sim \ \mathsf{p}(\mathrm{x}) = \mathrm{x}^{\nu-1} \frac{\exp(-\mathrm{x}/\theta_0)}{\Gamma(\nu)\theta_0^{\nu}} \Rightarrow \text{blood is de-oxygenated,}$$
$$H_1 : \ x \ \sim \ \mathsf{p}(\mathrm{x}) = \mathrm{x}^{\nu-1} \frac{\exp(-\mathrm{x}/\theta_1)}{\Gamma(\nu)\theta_1^{\nu}} \Rightarrow \text{blood is oxygenated,}$$

  where $\nu = 10$, $\theta_0 = 1$, $\theta_1 = 1.5$, and Γ is the gamma function.

You want to compare these two sensors as follows.

(a) Determine the mean and variance of the sensor readings in each case. Which sensor seems more promising based on these calculations, and why?

(b) Determine and plot the ROC curves for the two sensors. Which sensor seems more promising based on these calculations, and why?

(c) Suppose we need to guarantee that the probability of deciding H_0 when in fact the data is distributed according to H_1 is less than 0.05. Determine the optimal detector for both sensors. Which sensor would you recommend given this constraint?

Problem 4.2 (Snapchat's delays)**.** Remember that girlfriend/boyfriend from homework 2? Recall that every time you sent her/him pictures, there was a time delay, which you modeled as i.i.d. realizations of a random variable

$$x \overset{iid}{\sim} \mathsf{p}(\mathrm{x}|\theta^\star) = \begin{cases} \frac{1}{\theta^\star} e^{-(\mathrm{x}-t_0)/\theta^\star} & \text{if } \mathrm{x} \geq t_0 \\ 0 & \text{otherwise.} \end{cases}$$

Suppose you measure the following ten delays values (in milliseconds):

$$\{22, 6, 5, 7, 6, 6, 10, 7, 14, 7\}.$$

What value of the parameter θ provides a reasonable fit of your model to the data? Explain your answer.

Problem 4.3 (Poisson counts)**.** Let x_1, x_2 be i.i.d. Poisson($\lambda^\star$) random variables modeling particle counts in two sensors.

(a) Find an unbiased estimator of $\lambda^\star$.

(b) Show that x_1 is an unbiased estimator of $\lambda^\star$, and compute its *mean squared error* (MSE).

(c) Find the *maximum likelihood estimator* (MLE) of $\lambda^\star$ and compute its MSE.