Mathematical Statistics - Section 1 - NYU Spring 2019 - Homework 7

Testing the parameter of an exponential distribution Let $X_1, \ldots, X_n$ be independent random variables distributed according to the exponential distribution $f(x;\theta) := \theta e^{-\theta x}$. We believe that $\theta \geq 1$, this is our null hypothesis H_0, and the alternative hypothesis is $H_1 : \{\theta < 1\}$.

Design a test of size 0.1. That is: build a test statistic, and an appropriate rejection region, and justify what you are doing.

Here is a data set with 15 data points. Do you keep or reject H_0? Give the p-value, and any comment that you consider relevant about the methodology.

$$0.52, 0.33, 0.30, 3.27, 0.66, 0.20, 1.42, 0.49, 0.07, 0.09, 0.30, 0.84, 0.41, 0.62, 1.44$$

Noise in the dictatorship Let us consider the "dictatorial plus noise" situation, where we have couples (feature X is a real number, label Y is a real number) and where the distribution of Y knowing X is given by

$$Y = d + \varepsilon \mathcal{N}(0,1),$$

where d is the "dictatorial constant", $\varepsilon > 0$ and $\mathcal{N}(0,1)$ represents a standard normal random variable

1. For any real numbers a, and b, let $Z_{a,b}$ be a random variable with distribution $\mathcal{N}(a, b^2)$. Compute the following quantity
$$\mathbb{E}\left[|Z_{a,b}|^2\right]$$
2. Recall the two rules that we are using when we write
$$a + b\mathcal{N}(0,1) = \mathcal{N}(a, b^2)$$
3. We choose the cost function c as $c(x,y) = |x-y|^2$. Given a predictor $f : \mathbb{R} \to \mathbb{R}$, compute the risk
$$\mathbb{E}\left[c(f(X), Y)\right],$$
and show that it is minimal when f is the constant function equal to the "dictatorial constant" d.
4. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a data set. Our learning rule will be to use the data set to get an estimator $\hat{d}_n$ of d, and then to predict $\hat{d}_n$ all the time. We take for $\hat{d}_n$ the empirical mean. Compute the average risk, i.e.
$$\mathbb{E}\left[c(\hat{d}_n, Y)\right],$$
and compare it to the result of question 3.