Statistical Inference and Learning- Ex-2

Yaniv Tenzer and Boaz Nadler

Due on Tuesday, May 14, 2019,
email submissions: `stat.wisdom@gmail.com`

Q1 Let $X_1, \dots, X_n$ be a random sample of size n from a normal distribution with unknown mean and variance. Obtain the likelihood ratio statistic for testing $H_0 : \mu = \mu_0$, versus $H_1 : \mu \neq \mu_0$, and show that it is equivalent to a two-sided T test with significance level α.

Q2 A user of a certain gauge of steel wire suspects that the standard deviation of its breaking strength, in newtons (N), is different from the value of 0.75, as specified by the manufacturer.

Consequently the user tests the breaking strength of each of a random sample of nine lengths of wire and obtains the following results: 72.1, 74.5, 72.8, 75.0, 73.4, 75.4, 76.1, 73.5, 74.1.

Assuming breaking strength to be normally distributed, test, at the 10% level of significance, the manufacturer's specification.

Q3 FDR versus FWER

Let V be the set of true-null hypotheses rejected by a statistical test. Recall that the Family Wise Error Rate (FWER) is defined as $FWER = \mathbb{P}(V > 0)$.

1. Prove that $FDR \leq FWER$
2. What control of the FDR offers the Bonferroni procedure?
3. Prove that the number of hypotheses rejected by the Bonferroni procedure is smaller than the number of hypotheses rejected by Benjamini-Hochberg procedure.

Q4 Sidak Correction: Recall that a procedure T is said to control the FWER at level α in the weak sense, if $FWER(T) \leq \alpha$, under the global null assumption. Assume that the hypotheses are independent and define $\alpha_{SID} = 1 - (1 - \alpha)^{1/m}$, where m is the number of hypotheses. Show that testing each hypothesis at α_{SID} level, controls the FWER at level α in the weak sense. Compare to Bonferroni correction in terms of power.

Q5 Let $X_1, \dots, X_n$ be n i.i.d. random variables with mean μ, variance σ^2. Let $S_n = \frac{1}{n} \sum X_i$. The CLT states that regardless of the distribution of X, as $n \to \infty$, $\sqrt{n}(S_n - \mu)/\sigma$ converges to a $N(0, 1)$ distribution (a single attracting distribution). Here we shall look at a different limiting law, of extreme events. Let $Y_n = \max X_i$. We are interested in limiting distribution of Y_n if it exists. The following examples show that there is more than one such possible limiting law.

1. Suppose $X_i \sim \exp(1)$, namely $\Pr[X > t] = \exp(-t)$. Prove that as $n \to \infty$, $\Pr[Y_n - \ln(n) < z] \to G(z) = e^{-e^{-z}}$. The function $G(z)$ is known as the Gumbel distribution.
2. Consider m independent p-values, $p_1, \dots, p_m$ all i.i.d. distributed as $U[0, 1]$. Consider the smallest p-value, $p_{(1)} = \min p_i$. Compute the average $\mathbb{E}[p_{(1)}]$ and show that it is approximately $1/m$. Further, show that as $m \to \infty$, $\Pr[p_{(1)} > t/m] \to \exp(-t)$.