Statistics

List 4

Let $X_1, \ldots, X_m$ be the independent identically distributed random variables coming from the population with the continuous cumulative distribution function F. Let $Y_1, \ldots, Y_n$ be the independent identically distributed random variables coming from the population with the continuous cumulative distribution function G. We assume that all the random variables are independent. We consider testing the hypothesis

$$H_0 : F = G \quad \text{against the alternative} \quad H_1 : F \neq G \tag{1}$$

at the significance level $\alpha = 0.05$.

Let $N = m + n$. Set $\boldsymbol{Z} = (Z_1, \ldots, Z_N) = (X_1, \ldots, X_m; Y_1, \ldots, Y_n)$. Let R_i be the rank of Z_i in the sample $\boldsymbol{Z}$, $i = 1, \ldots, N$. The classical linear rank statistic related to a score function $\varphi \in L_2(0,1)$ has the form

$$T_\varphi = \sqrt{\frac{mn}{N}} \left\{ \frac{1}{m} \sum_{i=1}^{m} \varphi\Big(\frac{R_i - 0.5}{N}\Big) - \frac{1}{n} \sum_{i=m+1}^{N} \varphi\Big(\frac{R_i - 0.5}{N}\Big) \right\}, \tag{2}$$

while a selection of the function φ determines sensitivity of the related test based on T_φ. If $\varphi(u) = \varphi_1(u) = \sqrt{3}(2u - 1)$, we obtain the Wilcoxon statistic. Selection of $\varphi(u) = \varphi_2(u) = \sqrt{48}(0.25 - |u - 0.5|)$ leads to the Ansari-Bradley statistic. If $\int_0^1 \varphi(u)du = 0$ and $\int_0^1 \varphi^2(u)du = 1$, then, under the null hypothesis, the statistic T_φ has an asymptotic standard normal distribution. Furthermore, we reject H_0 in favour of H_1 for large values of $|T_\varphi|$.

Another classical solution of the above testing problem is, for instance, the Kolmogorov-Smirnov test rejecting H_0 for large values of the statistic

$$KS = \sqrt{\frac{mn}{N}} \sup_{x \in \mathbb{R}} |F_m(x) - G_n(x)|, \tag{3}$$

where F_m and G_n are the empirical cumulative distribution functions in the samples of Xs and Ys, respectively.

The goal of the lab is an investigation of the behaviour of the power functions of the selected solutions of the problem (1). Specifically, we will examine

(i) the Wilcoxon test based on the statistic $W = T_{\varphi_1}^2$,

(ii) the Ansari-Bradley test based on the statistic $AB = T_{\varphi_2}^2$,

(iii) the Lepage test based on the statistic $L = W + AB$,

(iv) the Kolmogorov-Smirnov test based on the statistic KS.

Exercise 1.

Generate $m = n = 20$ observations from the $N(0,1)$ distribution. Calculate the value of the statistics W, AB, L, and KS. Repeat the experiment 10 000 times. Find the critical values of the tests. Is such a method of finding the critical values correct?

Exercise 2.

Generate $m = n = 20$ observations from

(a) a normal distribution with the respective shift and scale parameters

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.2$, $\sigma_2 = 1$,
(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.4$, $\sigma_2 = 1$,
(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.6$, $\sigma_2 = 1$,
(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.8$, $\sigma_2 = 1$,
(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 1$,
(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.2$, $\sigma_2 = 1$,
(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.4$, $\sigma_2 = 1$,

(b) a logistic distribution with the respective shift and scale parameters

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.2$, $\sigma_2 = 1$,
(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.4$, $\sigma_2 = 1$,
(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.6$, $\sigma_2 = 1$,
(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.8$, $\sigma_2 = 1$,
(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 1$,
(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.2$, $\sigma_2 = 1$,
(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.4$, $\sigma_2 = 1$,

(c) a Cauchy distribution with the respective shift and scale parameters

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.0$, $\sigma_2 = 1$,
(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.5$, $\sigma_2 = 1$,
(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 1$,
(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.5$, $\sigma_2 = 1$,
(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 2.0$, $\sigma_2 = 1$,
(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 2.5$, $\sigma_2 = 1$,
(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 3.0$, $\sigma_2 = 1$.

Calculate the value of the statistics W, AB, L, and KS. Repeat the experiment 10 000 times. Estimate the values of the power functions of the tests under consideration. Draw them as functions of the parameter μ_2. Discuss the outcomes.

Exercise 3.

Generate $m = n = 20$ observations from

(a) a normal distribution with the respective shift and scale parameters

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 1.0$,
(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 1.5$,
(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 2.0$,
(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 2.5$,
(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 3.0$,
(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 3.5$,
(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 4.0$,

(b) a logistic distribution with the respective shift and scale parameters

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 1.0$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 1.5$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 2.0$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 2.5$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 3.0$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 3.5$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 4.0$,

(c) a Cauchy distribution with the respective shift and scale parameters

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 1.0$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 2.0$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 3.0$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 4.0$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 5.0$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 6.0$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0$, $\sigma_2 = 7.0$.

Calculate the value of the statistics W, AB, L, and KS. Repeat the experiment 10 000 times. Estimate the values of the power functions of the tests under consideration. Draw them as functions of the parameter σ_2. Discuss the outcomes.

Exercise 4.

Generate $m = n = 20$ observations from

(a) a normal distribution with the respective shift and scale parameters

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.2$, $\sigma_2 = 1.0$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.4$, $\sigma_2 = 1.5$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.6$, $\sigma_2 = 2.0$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.8$, $\sigma_2 = 2.5$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 3.0$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.2$, $\sigma_2 = 3.5$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.4$, $\sigma_2 = 4.0$,

(b) a logistic distribution with the respective shift and scale parameters

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.2$, $\sigma_2 = 1.0$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.4$, $\sigma_2 = 1.5$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.6$, $\sigma_2 = 2.0$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.8$, $\sigma_2 = 2.5$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 3.0$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.2$, $\sigma_2 = 3.5$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.4$, $\sigma_2 = 4.0$,

(c) a Cauchy distribution with the respective shift and scale parameters

(i) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.0$, $\sigma_2 = 1.0$,

(ii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 0.5$, $\sigma_2 = 2.0$,

(iii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.0$, $\sigma_2 = 3.0$,

(iv) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 1.5$, $\sigma_2 = 4.0$,

(v) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 2.0$, $\sigma_2 = 5.0$,

(vi) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 2.5$, $\sigma_2 = 6.0$,

(vii) $\mu_1 = 0$, $\sigma_1 = 1$; $\mu_2 = 3.0$, $\sigma_2 = 7.0$.

Calculate the value of the statistics W, AB, L, and KS. Repeat the experiment 10 000 times. Estimate the values of the power functions of the tests under consideration. Draw them as functions of the vector of the parameters (μ_2, σ_2). Discuss the outcomes.

Exercise 5.

Generate $m = n = 50$ observations from the $U(0, 1)$ distribution. Calculate the value of the statistics W, AB, L, and KS. Repeat the experiment 10 000 times. Find the critical values of the tests.

Exercise 6.

Generate $m = n = 50$ observations from

(a) a normal distribution with the respective shift and scale parameters,

(b) a logistic distribution with the respective shift and scale parameters,

(c) a Cauchy distribution with the respective shift and scale parameters.

Select the parameters μ_1 and μ_2, as well as σ_1 and σ_2, just as in Exercises 2, 3, 4, in order to obtain powers in the full range. Draw the power functions as the functions of the parameters: μ_2, σ_2, and (μ_2, σ_2), respectively. Discuss the outcomes.