Statistics

List 3

Let $X_1, \dots, X_n$ be the independent identically distributed random variables coming from the population with the continuous cumulative distribution function F. We consider testing the hypothesis

$$H_0 : F = F_0 \quad \text{against the alternative} \quad H_1 : F \neq F_0, \tag{1}$$

where F_0 is a known cumulative distribution function.

We define the new variables $U_1 = F_0(X_1), \dots, U_n = F_0(X_n)$. Then, the testing problem (H_0, H_1) is equivalent to verifying

$$H_0 : U_1 \sim U(0,1) \quad \text{against} \quad H_1 : U_1 \nsim U(0,1), \tag{2}$$

where $U(0,1)$ denotes the uniform distribution on $(0,1)$.

Let $A_1, \dots, A_k$ be a partition of the interval $(0,1)$, that is, $\cup_{j=1}^k A_j = (0,1)$ and $A_j \cap A_l = \emptyset$ for $j \neq l$, $j, l = 1, \dots, k$. Set $N_j = \#\{U_i \in A_j : i = 1, \dots, n\}$, and $p_j = P_0(U_1 \in A_j)$, $j = 1, \dots, k$.

The classical Pearson's chi-square test is based on the statistic

$$P_k = \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j}. \tag{3}$$

Under the null model, the statistic P_k has an asymptotic chi-square distribution with $k-1$ degrees of freedom. We reject the hypothesis H_0 for large values of the statistic P_k.

Let $\{b_j\}_{j \in \mathbb{N}}$ be the orthonormal system of the Legendre's polynomials in $L^2((0,1), du)$.

The Neyman's smooth test with the k components is based on the statistic

$$N_k = \sum_{j=1}^k \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n b_j(U_i) \right\}^2. \tag{4}$$

Under the null model, the statistic N_k has an asymptotic chi-square distribution with k degrees of freedom. We reject the hypothesis H_0 for large values of the statistic N_k.

The Kolmogorov-Smirnov test is based on the statistic

$$KS = \sqrt{n} \sup_{u \in (0,1)} |G_n(u) - u|, \tag{5}$$

where G_n is the empirical cumulative distribution function in the sample $U_1, \dots, U_n$. Under the null model, the statistic KS has the asymptotic Kolmogorov's distribution. We reject the hypothesis H_0 for large values of the statistic KS.

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 Pearson's chi-square test based on the statistic P_4 and P_8 with the uniform partition,

(ii) the Neyman's smooth test with 1, 4, and 8 components,

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

The significance level $\alpha = 0.05$.

Exercise 1.

Generate $n = 10$ observations from the $U(0,1)$ distribution. Calculate the values of the statistics P_4, P_8, N_1, N_4, N_8, and KS. Repeat the experiment 10 000 times. Find the critical values of the tests. Compare them with the 0.95-quantiles of the respective limiting distributions. Discuss the outcomes. Repeat the experiment for $n = 20, 30, \ldots, 100$.

Exercise 2.

Discuss the accept/reject von Neumann algorithm.

Exercise 3.

Generate $n = 10$ observations from the density $C_1(u, 0.4) = 1 + 0.4\cos(\pi u)$, $u \in (0,1)$. Calculate the values of the statistics P_4, P_8, N_1, N_4, N_8, and KS. Repeat the experiment 10 000 times. Estimate the values of the power functions of the tests under consideration. Repeat the experiment for $n = 20, 30, \ldots, 100$. Draw them as functions of the parameter n. Discuss the outcomes.

Exercise 4.

In Exercise 3, change $C_1(u, 0.4) = 1+0.4\cos(\pi u)$, $u \in (0,1)$ into $C_j(u, \rho) = 1+\rho\cos(j\pi u)$, $u \in (0,1)$. Repeat the numerical experiment from Exercise 3 under

(i) $j = 2$, $\rho = 0.5$,

(ii) $j = 3$, $\rho = 0.5$,

(iii) $j = 4$, $\rho = 0.6$,

(iv) $j = 5$, $\rho = 0.7$,

(v) $j = 6$, $\rho = 0.7$.

References

Rayner, J.C.W., Best, D.J. (1989). *Smooth Tests of Goodness of Fit.* Oxford University Press, New York.