where

\[ \hat{G}_d(x) = \begin{cases} 0 & \text{if } x < a; \\ \frac{\Phi(x) - \Phi(a)}{1 - \Phi(a)} & \text{if } x \geq a. \end{cases} \]

and \( 0 < a < \infty \).

LITERATURE CITED


DISTRIBUTION OF A \( \chi^2 \) STATISTIC WITH ADDITIONAL OBSERVATIONS USED TO ESTIMATE THE UNKNOWN PARAMETER

M. S. Nikulin and I. S. Yusas

An explicit form of the Kambhampati statistic is given, which is a \( \chi^2 \) statistic with additional observations used to estimate the unknown parameter.

This note is directly related to [1] and [2], and as far as possible we use the same notation. In [1] the behavior of a \( \chi^2 \) statistic was investigated for testing the hypothesis \( H_0 \) that independent identically distributed random variables \( \xi_1, \ldots, \xi_n \) follow a distribution from the family of distribution functions \( f(x, \xi), x \in \mathbb{R}^l, \xi = (\xi_1, \ldots, \xi_m) \in A \subset \mathbb{R}^s \), where \( A \) is an open set. The problem of testing the hypothesis \( H_0 \) was considered in a nonstandard setting since the vector of frequencies \( \psi^{(n)}(\xi_1, \ldots, \xi_m) \) was obtained from the observations \( \xi_1, \ldots, \xi_n \), whereas the unknown parameter \( \alpha \) was estimated in [1] using \( n = m \psi(\cdot) \) additional independent observations \( \xi_{m+1}, \ldots, \xi_{m+n} \), from the same distribution as the random variables \( \xi_1, \ldots, \xi_n \); here \( \psi(\cdot) \) is some monotone increasing function such that for \( n \to \infty \), we have \( n/\psi(\cdot) \to \gamma, 0 < \gamma < 1 \), \( n = n_1 + m \).

Because of the additional observations used for estimation of the parameter \( \alpha \), the standard \( \chi^2 \) statistic incorporated in the test of the hypothesis \( H_0 \) is highly perturbed and the limit behavior of the test statistic is distorted. It is shown in [1] that the following theorem holds when the parameter \( \alpha \) is estimated by the maximum likelihood method.

**THEOREM 1.** If the hypothesis \( H_0 \) is true, then the statistic

\[ X^2 = \sum_{i=1}^{n} \frac{\left( \frac{\psi^{(n)}(\cdot)}{n}, \xi_i \right)}{\psi_{i1}}, \]

is distributed in the limit for $N \to \infty$ as

$$\xi_1^2, \ldots, \xi_{\gamma-1}^2 + \sum_{i=1}^\delta (1 - \gamma_i) \eta_i^2,$$

where $\xi_1, \ldots, \xi_{\gamma-1}, \eta_1, \ldots, \eta_\delta$ are independent standard normal deviates, and the numbers $\gamma_i$ are between 0 and 1 and in general depend on the unknown value $\delta_i$ of the parameter $\delta$.

This theorem shows that it is not easy to use the standard Pearson statistic to test the hypothesis $H_0$ in this setting because of the complicated limit distribution of the statistic $X^2$. It is shown in [2, 3] that the Pearson statistic can be modified so that the modified test statistic for the hypothesis $H_0$ has a $\chi^2$ distribution in the limit. In this article we give explicit expressions for this modified statistic for the case when the parameter $\delta$ is estimated by the least squares and the maximum likelihood method.

The problem of testing the hypothesis $H_0$ with additional observations used for parameter estimation may arise in the following situation. Suppose that the entire output of some enterprise is characterized by a scalar variable whose numerical values are interpreted as sample values of independent identically distributed random variables $\xi, \ldots, \xi, \xi, \ldots, \xi, N = n+m$. These values are used by production control to estimate the unknown parameter $\delta$ of the probability distribution $F(\xi, \delta)$ of the random variables $\xi$. If the consumer purchases only part of the production output, he will find himself in a situation considered at the beginning of this note if he applies the $\chi^2$ statistic to test the hypothesis $H_0$ using the distribution parameter as estimated by the enterprise.

1. First consider the case when the parameter $\delta$ is estimated by the minimum $\chi^2$ method. Partition $R^1$ into $r$ intervals $(-\infty, \xi], (\xi, \xi], \ldots, (\xi, \infty)$ so that the probabilities $p_i(\delta) = P\{\xi \in (\xi, \xi]\} (i=1, \ldots, r)$ satisfy the Cramer conditions (1)-(4) formulated in [1]. Let $\delta^{(m)}$ be the estimator of the modified minimum $\chi^2$ method computed from the grouped observations $\xi, \ldots, \xi, \xi$, and let the vector $V = (\xi^{(m)}, \ldots, \xi^{(m)})^T$ be the result of grouping the first $n$ observations by the interval classes. Define the vector $V = (\xi^{(m)}, \ldots, \xi^{(m)})^T$ with the components

$$v_i = \frac{\xi_i - \delta^{(m)} p_i(\delta^{(m)})}{\sqrt{n \delta^{(m)} p_i(\delta^{(m)})}}.$$

It is shown in [1] that under the hypothesis $H_0$ the vector $V$ is asymptotically normally distributed for $N \to \infty$, and

$$E V = 0, \quad E VV^T = \sum (\delta^2) = I - p p^T - \delta B (B^T B)^{-1} B^T,$$

where $I$ is the unit matrix, $p = (p_1(\delta), \ldots, p_r(\delta))^T$, $B = \|I\| I$, $I_{ij} = \frac{I}{p_i(\delta)} \left( \frac{\delta}{\delta_j} \right)$, $i=1, \ldots, r$; $\delta$ is the unknown true value of the parameter $\delta$. Let $\Sigma(\delta')$ be the generalized inverse of the matrix $\Sigma(\delta)$. Following [2], we form the Kambhampati quadratic form

$$Q(\delta^{(m)}, \delta') = V^T \Sigma(\delta') V,$$