Contribution ID: **28** Type: **Poster**

Fitting of the Cumulative Function for a Distribution of Statistic from the One-Sample Anderson-Darling Test

The Anderson-Darling test uses a statistic:

$$W_n^2 = n\int_{-\infty}^{+\infty} \left(F_n(x) - F(x)\right)^2 \psi\left(F(x)\right) dF(x), \qquad (1)$$

where $u = F(x)$ is cumulative function of a known distribution; $F_n(x)$ is cumulative function of an empirical distribution; $\psi\left(u\right)$ is weight function; n is sample size [1],

to estimate the amount of an empirical distribution deviation from a given distribution.
The statistic (1) with weight function $\psi\left(u\right) \equiv 1$ for large samples has a distribution with a cumulative function, which is approximately expressed by the formula:

$$A_1(z) = \frac{1}{\pi\sqrt{z}} \sum_{j=0}^{\infty} \frac{\Gamma\left(j+\frac{1}{2}\right)}{j!\cdot\Gamma\left(\frac{1}{2}\right)} \cdot \sqrt{4j+1} \cdot \exp\left(-\frac{(4j+1)^2}{16z}\right) \cdot K\left(\frac{1}{4}, \frac{(4j+1)^2}{16z}\right), \qquad (2)$$

where $\Gamma(x)$ is gamma function; $K\left(\mu, x\right)$ is modified Bessel function of the second kind [1].
The complexity of calculating the significance points for a given significance level, the dependence of the expression of function (2) on the type of theoretical distribution when its parameters are unknown, led to the absence of the Anderson-Darling test, for example, in the Excel spreadsheet editor Maple math software. In practice, tabulated values of significance points are mostly used [2].
In these theses, it is shown that a cumulative function $A_1(z)$ can be fitting with satisfactory accuracy by a cumulative function belonging to a generalized family of cumulative functions as:

$$F(z) = (1+\nu)^{\omega} \cdot \left(\frac{G(z)}{\nu + G(z)}\right)^{\omega}, \qquad (3)$$

where $\nu > 0$, $\omega > 0$, $G(z)$ is cumulative function of a continuous distribution [3].
It is proposed to choose the cumulative function of Weibull distribution as the function $G(z)$:

$$G(z) = 1 - \exp\left(-\left(\frac{z}{\theta}\right)^{c}\right), \qquad (4)$$

where $z \geq 0$, $\theta > 0$, $c > 0$.
It is shown that at the values of the coefficients: $c = 1.1365$, $\theta = 0.2263$, $\nu = 0.0217$, $\omega = 20.5000$ the mean square error of approximating the cumulative distribution function (2) using functions (3) and (4) is $Q = 1.7359 \cdot 10^{-7}$. Calculations were performed with double precision in Maple software. The given numerical values are rounded to the fourth decimal place.
When using the proposed fitting, the significance points of the Anderson-Darling test coincide with those previously published up to and including the second decimal place.

1. D'Agostino R. B., Stephens M. A. (Eds.) (1986). Goodness-Of-Fit-Techniques. In Series: Statistics, textbooks and monographs. (Vol. 68). New York and Basel: Marcel Dekker, Ink.
2. Jäntschi L., Bolboacă S. D. (2018). Computation of Probability Associated with Anderson–Darling Statistic. Mathematics, 6, 88, 16 p. doi:10.3390/math6060088
3. Laxmi Prasad Sapkota, Nirajan Bam, Vijay Kumar. (2024). A New Exponential Family of Distribution with Applications to Engineering and Medical Data. Preprints. doi:10.21203/rs.3.rs-4522315/v1

Primary author: TULUCHENKO, Halyna (National Technical University "Kharkiv Polytechnic Institute")

Presenter: TULUCHENKO, Halyna (National Technical University "Kharkiv Polytechnic Institute")

Session Classification: MATHEMATICS

Track Classification: MATHEMATICS