Statistiska centralbyrån
Statistics Sweden
Journal of Official Statistics
30^{th} Anniversary, June 10-12, 2015
Stockholm, Sweden

Agreement between Observation and Theoretical Model: Anderson Darling Statistic

Lorentz JÄNTSCHI[1] and Sorana D. BOLBOACĂ[2]

MINISTRY OF EDUCATION AND RESEARCH
TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

[1] Technical University of Cluj-Napoca, 400641 Cluj-Napoca, Romania
[2] Iuliu Haţieganu University of Medicine and Pharmacy Cluj-Napoca, 400349 Cluj-Napoca, Romania
E-mails: lorentz.jantschi@gmail.com; sbolboaca@umfcluj.ro

1. Introduction

Application of any statistical test is done under certain assumptions and violation of these assumptions led to misleading interpretations and unreliable results. Anderson-Darling (AD) test [1,2] is one statistical test used to assess the distribution of data (H_0: Data follow the specified distribution vs. H_A (alternative hypothesis): Data do not follow the specified distribution). The interpretation of the AD test is done by comparison of statistic with the critical value for certain significance level (e.g. 20%, 10%, 5% and 1%), but the thresholds available are reported for specific distributions (see Table 1) and usually do not depend on sample size, both is simply not true for AD statistic.

Table 1. Anderson-Darling test: critical values according with significance level (α)

Distribution \ α	0.10	0.05	0.025	0.01
Normal & lognormal [3]	0.631	0.752	0.873	1.035
Weibull [4]	0.637	0.757	0.877	1.038

A more complex issue, using a single test to assess the distribution of a dataset is not proper since any distribution is dependent by at least two parameters and one test could not bring sufficient information [5] in regards of risk of being in error. It has been previously shown that existence of one outlier led to a full disarray among goodness-of-fit statistics [6].

Aim

Calculation of combined probability based on results obtained by several independent statistical tests as proposed by Fisher [7,8] becomes justified and necessary. As a part of a larger study concerning other statistics too, starting with this identified gap related with the use of Anderson-Darling (AD) statistic, the present study aimed to identify, assess and implement the probability (p-value) as an explicit function of the value of the statistic and sample size (n).

2. Method

A Monte Carlo type simulation study was conducted to determine the function able to estimate the probability associated to AD statistic. The following algorithm was used:

Step 1. Generate data sets of samples sizes from 2 to 46 using uniform continuous distribution [0,1]. Mersenne twister algorithm [9] was used in this step.

Step 2. Repeat Step 1 for $k\times u$ times (where k=number of thresholds, taken 1,000; u=number of repetitions, taken 9);

Step 3. Compute AD statistic for each replication in Step 2 and for each sample size;

Step 4. Order the AD value obtained in Step 3, and select a list of k values corresponding to success rates (algorithm in full available);

Step 5. Identify that function which best fit according to both sample size and the AD value.

The calculations were done with FreePascal, which use Mersenne Twister for randomization, and in order to reduce even more the noise of pseudo-randomization, the mantissa (4 significant digits) and the exponent (one significant digit) was separately randomized to obtain a better smoothing of the range [0..1].

Assessment

Demonstration of the performances of the identified solution was conducted with EasyFit Professional (v. 5.2 - MathWave) on six experimental data sets (expected to be normal distributed) with sample size from 18 to 165.

3. Results

The identified solution is estimation with a rational function with atan(log(1/p)) with coefficients as function of 1/n, $1/n^2$, $1=n^3$ and $(-1)^n$ with different values for two categories of p-values (lower p and higher p).

The calculation was implemented online and is available at:
http://l.academicdirect.org/Statistics/tests/AD/.

The input data are allowed with the sample size $2 \le n \le 1000$ and the value of Anderson-Darling statistic $0.1 \le AD \le 10$ while the output is represented by the probability to be observed a better agreement between the observed sample and the hypothetical distribution being tested (1-p).

The theoretical values of AD as function of probability to be in error (p) and the sample size (n) show that the AD values varies slowly with n and p (Figure 1).

Our experience shown that is extremely difficult to obtain very good approximations when the p-value varies in a large range (here 10^{-10} to $1\text{-}10^{-10}$).

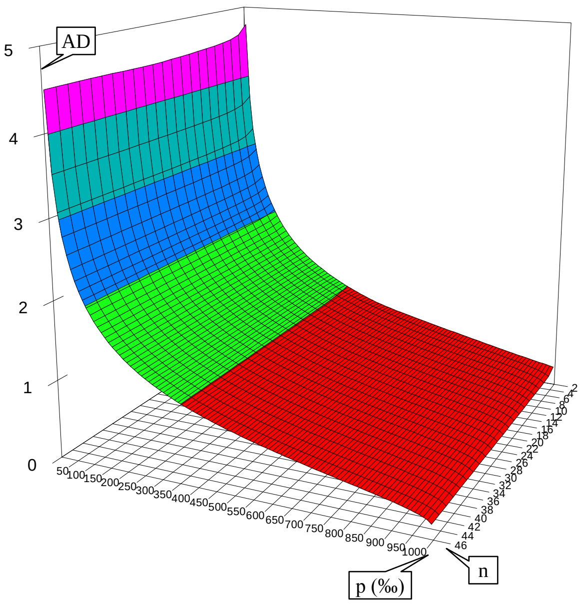

Fig. 1. AD depending on p and n

The values calculated by the implemented program for estimation are concordant with the values obtained by Monte Carlo simulation for $10^{-9} \le p \le 1\text{-}10^{-9}$ with relative standard deviation in average per pooled sample containing whole data for $2 \le n \le 46$ less than 0.1‰.

Table 2 presents the AD values, critical values give by EasyFit software and decision of the test (H_0: Data are normally distributed vs. H_A: Data are not normally distributed), and the associated p-values for each investigated data set and significance level (values in blue) as well as the p-value for the AD and n (last column of the table).

Table 2. Evaluation on different samples: EasyFit (critical values & decision (Reject H_0? - yes/no)) vs. identified solution (p-values – values in blue)

	Crit.	**1.3749**		**1.9286**		**2.5018**		**3.2892**		**3.9074**		**IS**
n	**AD**	**α = 0.2**		**α = 0.1**		**α = 0.05**		**α = 0.02**		**α = 0.01**		**p-value**
165	1.7777	yes	0.20866	no	0.10051	no	0.049147	no	0.01922	no	0.0094567	**0.1221**
60	2.1597	yes	0.20865	yes	0.10068	no	0.049296	no	0.01930	no	0.0095067	**0.0752**
79	1.4298	yes	0.20866	no	0.10062	no	0.049240	no	0.01927	no	0.0094879	**0.1936**
29	2.7217	yes	0.20864	yes	0.10096	yes	0.049544	no	0.01945	no	0.0095904	**0.0380**
18	0.6114	no	0.20862	no	0.10129	no	0.049832	no	0.01961	no	0.0096886	**0.6340**
27	4.2884	yes	0.20864	yes	0.10100	yes	0.049579	yes	0.01947	yes	0.0096023	**0.0063**

Crit. = critical value; n = sample size; AD = Anderson-Darling statistic; α = significance level;
IS p-value = p-value calculated with the implemented solution

References

1. Anderson TW, Darling DA. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Ann Math Stat 1952;23:193-212.
2. Anderson TW, Darling DA. A Test of Goodness-of-Fit. J Am Statist Assoc 1954;49:765-769.
3. D'Agostino RB, Stephes MA (Eds). Goodness-of-fit Techniques. Marcel-Dekker, New York. 1986; pp. 123 & 146.
4. Shin H, Jung Y, Jeong C, Heo J-H. Assessment of modified Anderson–Darling test statistics for the generalized extreme value and generalized logistic distributions. Stoch Environ Res Risk Assess 2012;26:105-114.
5. Fisher RA. On the Mathematical Foundations of Theoretical Statistics. Phil Trans R Soc Lond A 1922;222:309-368.
6. Jäntschi L, Bolboacă SD. Distribution Fitting 2. Pearson-Fisher, Kolmogorov-Smirnov, Anderson-Darling, Wilks-Shapiro, Kramer-von-Misses and Jarque-Bera statistics. Bulletin UASVM Cluj-Napoca. Horticulture 2009;66(2):691-697.
7. Fisher RA. Combining independent tests of significance. Am Stat 1948;2:30.
8. Bolboacă SD, Jäntschi L, Sestraş AF, Sestraş RE, Pamfil DC. Pearson-Fisher Chi-Square Statistic Revisited (Supplementary Material). Information 2011;http://www.mdpi.com/2078-2489/2/3/528/s1
9. Matsumoto M, Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM T Model Comput S 1998;8(1):3-30.

Dedicated in the honor of Sir Ronald Aylmer FISHER