

Models of Statistic Distributions of Nonparametric Goodness-of-fit Tests in Composite Hypotheses Testing in Case of Generalized Weibull Law

Boris Yu. Lemeshko, Stanislav B. Lemeshko and Kseniya A. Akushkina

Department of applied mathematics and computer science, Novosibirsk State Technical University
Novosibirsk, Russia
Email: skyer@mail.ru

Abstract: In this paper there are presented results (tables of percentage points and statistic distribution models) for the nonparametric goodness-of-fit tests in testing composite hypotheses using the maximum likelihood method for parameters estimation for Generalized Weibull Distribution law. Statistic distributions of the nonparametric goodness-of-fit tests are investigated by the methods of statistical simulation.

Keywords: Goodness-of-fit test, Composite hypotheses, Kolmogorov test, Cramer-Von Mises-Smirnov test, Anderson-Darling test, Generalized Weibull distribution.

1 The family of Generalized Weibull Distribution in Reliability

Let us consider the sample $X = (X_1, X_2, \dots, X_n)^T$, we say that X_i follow the generalized Weibull distribution (GWD). The density function of the law is defined by:

$$f(x; \theta_0, \theta_1, \theta_2) = \frac{\theta_0}{\theta_1} \theta_2^{\theta_0} x^{\theta_0-1} \left(1 + \left(\frac{x}{\theta_2} \right)^{\theta_0} \right)^{\frac{1}{\theta_1}-1} e^{-\left(1 + \left(\frac{x}{\theta_2} \right)^{\theta_0} \right)^{\frac{1}{\theta_1}}}, \quad (1)$$

where $x \geq 0$, $\theta_0, \theta_1, \theta_2 > 0$. The family (1) defines a set of different laws. Special cases of Generalized Weibull Distribution are: $\theta_1 = 1$ – the family of Weibull distribution; $\theta_0 = 1, \theta_1 = 1$ – the family of exponential distribution.

The distribution function is

$$F(x; \theta_0, \theta_1, \theta_2) = 1 - e^{-\left(1 + \left(\frac{x}{\theta_2} \right)^{\theta_0} \right)^{\frac{1}{\theta_1}}}.$$

The hazard function of GWD can be monotone increase ($\theta_0 > 1, \theta_0 > \theta_1$ and $\theta_0 = 1, \theta_1 < 1$), monotone decrease ($0 < \theta_0 < 1, \theta_0 < \theta_1$ и $0 < \theta_0 < 1, \theta_0 = \theta_1$), \cap -shaped ($\theta_1 > \theta_0 > 1$), \cup -shaped ($0 < \theta_1 < \theta_0 < 1$) and we have:

$$\lambda(x) = \frac{\theta_0}{\theta_1} \theta_2^{\theta_1} x^{\theta_0-1} \left(1 + \left(\frac{x}{\theta_2} \right)^{\theta_0} \right)^{\frac{1}{\theta_1}-1}. \quad (2)$$

The Generalized Weibull distribution used in reliability and survival tasks along with lognormal distribution and inverse Gaussian distribution. As usual, in construction the models of laws for real observed random variables it is difficult to discriminate one laws from another and choose one of them. To certain degree these difficulties related with restricted facilities of application nonparametric goodness-of-fit tests with unknown statistic distribution for verification of composite hypotheses concerning GWD.

2 Nonparametric goodness-of-fit tests in verification single and composite hypotheses

One of the most popular statistic analysis problems in handling the results of experimental data is verification the agreement experimental distribution and theoretical one. There exist the verification of single hypothesis and composite hypothesis. Single hypothesis has the form $H_0: F(x) = F(x, \theta)$, $F(x, \theta)$ is probability distribution function, θ is known parameter value (θ is scalar parameter or vector parameter). In the case of single hypothesis marginal statistic distribution of nonparametric Kolmogorov, ω^2 Cramer-Von Mises-Smirnov, Ω^2 Anderson-Darling goodness-of-fit tests do not depend on view of observed distribution law and parameters values. These goodness-of-fit tests are “free from the distribution”

Composite hypotheses has the form $H_0: F(x) \in \{F(x, \theta), \theta \in \Theta\}$. In this case the estimate of distribution parameter $\hat{\theta}$ is calculated by the same sample, the nonparametric Kolmogorov, ω^2 Cramer-Von Mises-Smirnov, Ω^2 Anderson-Darling goodness-of-fit tests lose the property called “free from the distribution”. In Kolmogorov goodness-of-fit test the value

$$D_n = \sup_{|x| < \infty} |F_n(x) - F(x, \theta)|,$$

where $F_n(x)$ is the empirical distribution function, n is the sample size, is used in Kolmogorov test characterized a distance between the empirical and theoretical laws. In testing of hypotheses used a statistic with Bolshev correction [1] in the form [2]

$$S_K = \frac{6nD_n + 1}{6\sqrt{n}}, \quad (3)$$

where $D_n = \max(D_n^+, D_n^-)$,

$$D_n^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F(x_i, \theta) \right\}, \quad D_n^- = \max_{1 \leq i \leq n} \left\{ F(x_i, \theta) - \frac{i-1}{n} \right\},$$

n is the sample size, x_1, x_2, \dots, x_n are sample values in increasing order is usually used. The distribution of statistic (3) obeys the Kolmogorov distribution law $K(S)$ [1] in testing simple hypotheses.

For verification of ω^2 Cramer-Von Mises-Smirnov goodness-of-fit test is used a statistic of the form

$$S_\omega = n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left\{ F(x_i, \theta) - \frac{2i-1}{2n} \right\}^2, \quad (4)$$

and in test of Ω^2 Anderson-Darling type [3, 4] the statistic of the form

$$S_\Omega = -n - 2 \sum_{i=1}^n \left\{ \frac{2i-1}{2n} \ln F(x_i, \theta) + \left(1 - \frac{2i-1}{2n} \right) \ln(1 - F(x_i, \theta)) \right\}. \quad (5)$$

In verification a simple hypothesis statistic (4) obeys the $a1(S)$ distribution and statistic (5) obeys the $a2(S)$ distribution (see [1]).

In verification of composite hypotheses the conditional distribution law of the statistic $G(S|H_0)$ is affected by a number of factors: the form of the observed law $F(x, \theta)$ corresponding to the true hypothesis H_0 ; the type of the parameter estimation and the number of estimated parameters; sometimes it is a specific value of the parameter (e.g., in the case of gamma-distribution and beta-distribution families); the method of parameter estimation. The distinctions in the marginal distributions of the same statistics in testing simple and composite hypotheses are so significant that we cannot neglect them.

The paper [5] was one of the first in investigating statistic distributions of the nonparametric goodness-of-fit tests with composite hypotheses. Then, for the solution to this problem, various approaches were used [6, 7], [8-10], [11, 12], [13], [14], [15-16], [17], [18, 19], [20], [21, 22].

In our research [23-28] statistic distributions of the nonparametric goodness-of-fit tests are investigated by the methods of statistical simulating and for constructed empirical distributions approximate models of law were founded. Obtained results were used for developing of recommendations for standardization [29]. Precise models of statistics distributions and the tables of upper percentage points presented in [30-34]. The comparative analysis results of power of goodness-of-fit tests in verification a single and composite hypothesis are presented in the paper [35-37].

3 Distributions of statistics of the tests in the case of verification composite hypotheses concerning generalized Weibull distribution

In the case of verification composite hypotheses concerning Generalized Weibull Distribution the distributions of statistics $G(S|H_0)$ for nonparametric goodness-of-fit tests depend on specific values of shape parameters θ_0 .

In the fig. 1 you can see the behavior of statistics distribution S_Ω in testing composite hypotheses for family (1). In the case when two parameters are estimated by MLM (fig. 1) you can see the following: with the growing of values of parameters θ_1 the distribution $G(S|H_0)$ shifts to the right.

In the case when three parameters are estimated by MLM you can see the next: when the values of shape parameter are grow up till $\theta_1 \approx 2$ the distribution $G(S|H_0)$ is shift to the left. With the following growth of values of shape parameter the distribution $G(S|H_0)$ shifts to the opposite direction, to the right.

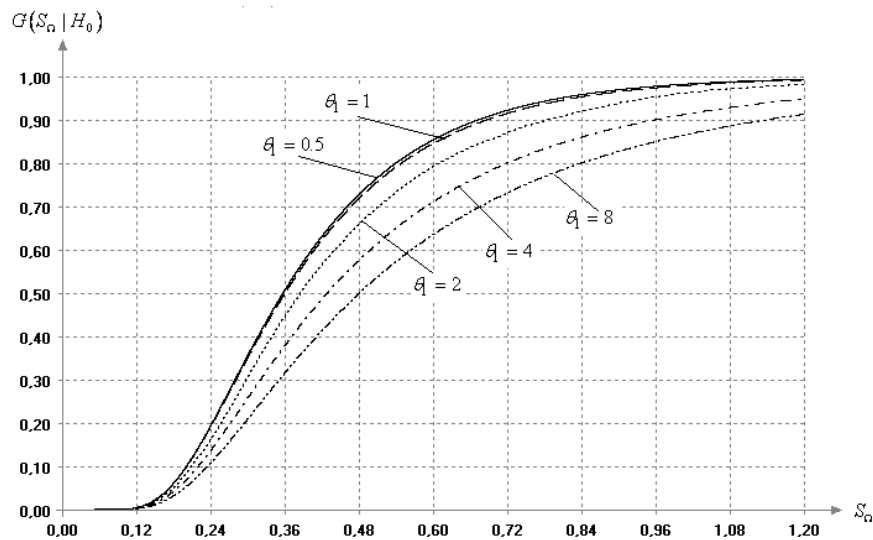


Fig. 1. Statistic distributions (5) of Anderson-Darling goodness-of-fit tests in composite hypotheses testing concerning family (1). MLM is used for estimation two (θ_0 and θ_1) parameters.

Percentage points obtained by statistic modeling and the models of marginal statistic distributions of Kolmogorov, Cramer-Von Mises-Smirnov and Anderson-Darling tests were computed for the values of shape parameter $\theta_1 = 1.0$ with MLM

used for parameter estimation are presented in table 1. Similar tables are constructed for the values of shape parameter $\theta_1 = 0.5, 2, 3, 4, 5, 6, 7, 8$.

Distributions $G(S|H_0)$ of the Kolmogorov, Cramer-Von Mises-Smirnov and the Anderson-Darling statistics are best approximated by the family of the III type beta-distributions with the density function

$$B_3(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4) = \frac{\theta_2^{\theta_0}}{\theta_3 B(\theta_0, \theta_1)} \frac{\left(\frac{x-\theta_4}{\theta_3}\right)^{\theta_0-1} \left(1-\frac{x-\theta_4}{\theta_3}\right)^{\theta_1-1}}{\left[1 + (\theta_2 - 1) \frac{x-\theta_4}{\theta_3}\right]^{\theta_0+\theta_1}},$$

or by the family of the *Sb*-Johnson distributions

$$Sb(\theta_0, \theta_1, \theta_2, \theta_3) = \frac{\theta_1 \theta_2}{(x-\theta_3)(\theta_2 + \theta_3 - x)} \exp\left\{-\frac{1}{2}\left(\theta_0 - \theta_1 \ln \frac{x-\theta_3}{\theta_2 + \theta_3 - x}\right)^2\right\}.$$

The tables of percentage points and statistic distributions models were constructed by modeled statistic samples with the size $N = 10^6$. In the case when $N = 10^6$ the deviation the empirical p.d.f. $G_N(S|H_0)$ from the theoretical one is less than 10^{-3} . In this case the values of statistics of goodness-of-fit- tests were calculated using a samples of pseudorandom variables which belong to $F(x, \theta)$ with sample size $n = 10^3$. For the such value of n statistic p.d.f. $G(S_n|H_0)$ almost coincides with the marginal p.d.f. $G(S|H_0)$.

Table 1. Percentage points and models of limiting statistic distributions of the nonparametric goodness-of-fit test when MLM is used for parameter estimation (for $\theta_1 = 1$)

Parameter estimated	Percentage points			Model
	0.9	0.95	0.99	
for Kolmogorov's test				
θ_0	1.181	1.316	1.585	$B_3(6.9734, 4.8247, 5.3213, 2.3800, 0.2690)$
θ_1	1.083	1.196	1.425	$B_3(4.6425, 6.6688, 2.8491, 2.2246, 0.3200)$
θ_2	0.994	1.092	1.290	$B_3(6.2635, 7.1481, 3.2059, 2.0000, 0.2800)$
θ_0, θ_1	0.874	0.954	1.117	$Sb(2.4299, 1.8866, 1.7504, 0.2598)$
θ_0, θ_2	0.823	0.893	1.033	$B_3(5.8989, 7.5040, 2.4180, 1.3724, 0.2800)$
θ_1, θ_2	0.815	0.883	1.023	$Sb(2.4499, 1.9720, 1.6016, 0.2486)$
$\theta_0, \theta_1, \theta_2$	0.758	0.820	0.946	$Sb(2.3012, 1.9386, 1.3863, 0.2464)$

for Cramer-Von Mises-Smirnov's test				
θ_0	0.320	0.431	0.706	B_3 (2.2422, 2.2970, 16.4663, 1.6500, 0.0130)
θ_1	0.227	0.295	0.464	B_3 (5.3830, 2.6954, 40.5199, 1.6450, 0.0050)
θ_2	0.174	0.221	0.336	B_3 (3.6505, 3.2499, 16.5445, 1.0000, 0.0100)
θ_0, θ_1	0.117	0.144	0.209	Sb (3.8667, 1.4603, 0.7583, 0.0059)
θ_0, θ_2	0.102	0.123	0.174	B_3 (12.2776, 4.1107, 27.2069, 0.4875, 0.0000)
θ_1, θ_2	0.103	0.127	0.182	B_3 (4.7144, 4.6690, 10.8816, 0.5261, 0.0059)
$\theta_0, \theta_1, \theta_2$	0.080	0.097	0.135	Sb (4.1842, 1.6587, 0.4794, 0.0061)
for Anderson-Darling's test				
θ_0	1.724	2.280	3.639	B_3 (4.8106, 2.6855, 35.5593, 11.8700, 0.0500)
θ_1	1.275	1.617	2.468	B_3 (3.6999, 3.9108, 16.4841, 9.0300, 0.0740)
θ_2	1.056	1.314	1.953	B_3 (4.9871, 4.1479, 16.5432, 6.4500, 0.0600)
θ_0, θ_1	0.687	0.827	1.161	B_3 (4.6368, 6.6727, 7.1680, 3.6356, 0.0521)
θ_0, θ_2	0.633	0.753	1.037	B_3 (3.0467, 5.9239, 5.0944, 2.7870, 0.1000)
θ_1, θ_2	0.696	0.842	1.194	B_3 (6.9638, 4.5238, 17.7792, 3.8000, 0.0522)
$\theta_0, \theta_1, \theta_2$	0.494	0.582	0.786	Sb (3.9578, 1.6861, 2.5760, 0.0547)

4 Conclusions

The Generalized Weibull probability distribution plays an important role in a statistical analysis of lifetime or response data in reliability and survival studies. In certain parameters the function of Generalized Weibull Distribution agree with the Weibull distribution function and the exponential distribution function. In this work you can find how density of distribution depends on the parameters values of the law.

In this work are presented models of the statistic distributions of the nonparametric goodness-of-fit tests for testing composite hypotheses with the distributions family (1).

It should be stressed, that obtained percentage points and models guarantee proper implementation of the nonparametric goodness-of-fit tests in statistic analysis problems if MLM is used. These results can't be used with other estimations because statistic distributions of these tests are essential depend on estimation method [25].

The authors hope that release of the article will be conducive to decrease mistake amount, committed in statistic analysis problems if nonparametric goodness-of-fit tests are used [27].

Acknowledgments

This research was supported by the Russian Foundation for Basic Research (project № 09-01-00056a) and by Federal Education Agency of Russian Federation Ministry of Education and Science in the Analytical departmental purposeful program framework "Development of Higher School Potential" and in the framework of Federal task program "Scientific and scientific-educational staff of the innovation Russia".

References

1. Bolshev L.N., On the question on testing some composite statistical hypotheses. In *Theory of Probability and Mathematical Statistics*. Selected Works. Nauka, Moscow, 5-63 (1987).
2. Bolshev L.N., Smirnov N.V., Tables of Mathematical Statistics. Moscow: Science (1983). (in Russian)
3. Anderson T.W., and Darling D.A. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes, *Ann. Math. Statist.*, 23, 193-212 (1952).
4. Anderson T.W., and Darling D.A., A test of goodness of fit, *J. Amer. Statist. Assoc.*, 29, 765-769 (1954).
5. Kac M., Kiefer J., Wolfowitz J., On Tests of Normality and Other Tests of Goodness of Fit Based on Distance Methods. *Ann. Math. Stat.*, 26, 189-211 (1955).
6. Darling D.A. The Cramer-Smirnov test in the parametric case, *Ann. Math. Statist.*, 26, 1-20 (1955).
7. Darling D.A. The Cramer-Smirnov test in the parametric case, *Ann. Math. Statist.*, 28, 823-838 (1957).
8. Durbin J., Weak convergence of the sample distribution function when parameters are estimated, *Ann. Statist.*, 1, 279-290 (1973).
9. Durbin J., Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests of spacings, *Biometrika*, 62, 5-22 (1975).
10. Durbin J., Kolmogorov-Smirnov Test when Parameters are Estimated, *Lect. Notes Math.*, 566, 33-44 (1976).
11. Gihman I.I., Some remarks on the consistency criterion of A.N. Kolmogorov. *Dokl. Akad. Nauk SSSR*, 91, 4, 715-718 (1953).
12. Gihman I.I., On the empirical distribution function in the case of grouping data, In *Selected Translations in Mathematical Statistics and Probability*, 1, American Math. Soc., Providence, RI, 77-81, (1961).
13. Martinov G.V., Omega-Quadrat Tests. *Moscow: Science* (1978). (in Russian)

14. Pearson E.S., Hartley H.O. *Biometrika Tables for Statistics*. Vol. 2. Cambridge: University Press (1972).
15. Stephens M.A., Use of Kolmogorov–Smirnov, Cramer – von Mises and Related Statistics – Without Extensive Table, *J. R. Stat. Soc.*, 32, 115-122 (1970).
16. Stephens M.A., EDF Statistics for Goodness of Fit and Some Comparisons, *J. Am. Statist. Assoc.*, 69, 730-737 (1974).
17. Chandra M., Singpurwalla N.D., Stephens M.A., Statistics for Test of Fit for the Extrem–Value and Weibull Distribution, *J. Am. Statist. Assoc.*, 76, 375, 729-731 (1981).
18. Tyurin Yu.N., On the Limiting Kolmogorov-Smirnov Statistic Distribution for Composite Hypothesis, *News of the AS USSR, Ser. Math.*, 48, 6, 1314-1343 (1984). (in Russian)
19. Tyurin Yu.N., Savvushkina N.E., Goodness-of-Fit Test for Weibull-Gnedenko Distribution, *News of the AS USSR, Ser. Techn. Cybernetics*, 3, 109-112 (1984). (in Russian)
20. Dzharparidze K.O., and Nikulin M.S., Probability distribution of the Kolmogorov and omega-square statistics for continuous distributions with shift and scale parameters, *J. Soviet Math.*, 20, 2147-2163 (1982).
21. Nikulin M.S., Gihman and goodness-of-fit tests for grouped data, *Mathematical Reports of the academy of Science of the Royal Society of Canada*, 14, 4, 151-156 (1992).
22. Nikulin M.S., A variant of the generalized omega-square statistic, *J. Soviet Math.*, 61, 4, 1896-1900 (1992).
23. Lemeshko B.Yu., Postovalov S.N., Statistical Distributions of Nonparametric Goodness-of-Fit Tests as Estimated by the Sample Parameters of Experimentally Observed Laws, *Industrial laboratory*, 64, 3, 197-208 (1998).
24. Lemeshko B.Yu., Postovalov S.N., Application of the Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses. *Optoelectronics, Instrumentation and Data Processing*, 37, 2, 76-88 (2001).
25. Lemeshko B.Yu., Postovalov S.N. About dependence of statistics distributions of the nonparametric tests and their power from a method of the parameters estimation, *Industrial laboratory. Diagnostics of materials*, 6, 77, 62-71 (2001). (in Russian)
26. Lemeshko B.Yu., Postovalov S.N., The nonparametric goodness-of-fit tests about fit with Johnson distributions in testing composite hypotheses. *News of the SB AS HS*, 1, 5, 65-74 (2002). (in Russian)
27. Lemeshko B.Yu., Errors when using nonparametric fitting criteria, *Measurement Techniques*, 47, 2, – P.134-142 (2004).
28. Lemeshko B.Yu., Maklakov A.A., Nonparametric Test in Testing Composite Hypotheses on Goodness of Fit Exponential Family Distributions, *Optoelectronics, Instrumentation and Data Processing*, 40, 3, 3-18 (2004).

29. R 50.1.037-2002. Recommendations for Standardization. Applied statistics. Rules of check of experimental and theoretical distribution of the consent. Part II. Nonparametric goodness-of-fit test. *Moscow: Publishing house of the standards*, (2002). (in Russian)
30. Lemeshko S.B., Lemeshko B.Yu., Statistic distributions of the nonparametric goodness-of-fit tests in testing hypotheses relative to beta-distributions, *News of the SB AS HS*, 2, 9, 6-16 (2007). (in Russian)
31. Lemeshko B.Yu., Lemeshko S.B., Distribution models for nonparametric tests for fit in verifying complicated hypotheses and maximum-likelihood estimators. Part 1, *Measurement Techniques*, 52, 6, 555-565 (2009).
32. Lemeshko B.Yu., Lemeshko S.B., Models for statistical distributions in nonparametric fitting tests on composite hypotheses based on maximum-likelihood estimators. Part II. *Measurement Techniques*, 52, 8, 799-812 (2009).
33. Lemeshko B.Yu., Lemeshko S.B., Models of Statistic Distributions of Nonparametric Goodness-of-fit Tests in Composite Hypotheses Testing in Case of Double Exponential Law, *The XIII International Conference "Applied Stochastic Models and Data Analysis"* (ASMDA-2009), June 30-July 3, Selected papers. Vilnius, Lithuania, 153-157 (2009).
34. Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N., Statistic Distribution Models for some nonparametric goodness-of-fit tests in testing composite hypothesis, *Communications in Statistics - Theory and Methods*, 39, 3, 460-471, (2010).
35. Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N., The power of goodness-of-fit tests for close alternatives, *Measurement Techniques*, 50, 2, 132-141 (2007).
36. Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N., Comparative Analysis of the Power of Goodness-of-Fit Tests for Near Competing Hypotheses. I. The Verification of Simple Hypotheses, *Journal of Applied and Industrial Mathematics*, 3, 4, 462-475 (2009).
37. Lemeshko B.Yu. Lemeshko S.B., Postovalov S.N., Comparative analysis of the power of goodness-of-fit tests for near competing hypotheses. II. Verification of complex hypotheses, *Journal of Applied and Industrial Mathematics*, 4, 1, 79-93 (2010).