NONPARAMETRIC GOODNESS-OF-FIT TESTS FOR CENSORED DATA

MIKHAIL NIKULIN
Université Victor Segalen Bordeaux 2
Bordeaux, France
email: mikhail.nikouline@u-bordeaux2.fr

BORIS LEMESHKO
Novosibirsk State Technical University
Novosibirsk, 630092, Russia
email: lemeshko@fpm.ami.nstu.ru

EKATERINA CHIMITOVA
Novosibirsk State Technical University
Novosibirsk, 630092, Russia
email: chim@mail.ru

ANGELIKA TSIVINSKAYA
Novosibirsk State Technical University
Novosibirsk, 630092, Russia
email: viento.acariciador@gmail.com

The problem of testing goodness-of-fit hypothesis for right censored data was widely discussed in the literature. The modifications of well known tests such as Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling tests were constructed. These modifications are based on the difference between non-parametric Kaplan-Meier estimate and hypothetical survival function. In this paper we investigate statistical properties of these modified tests. Also we consider the classical goodness-of-fit tests based on using the transformation of censored data to "complete" sample by means of randomization. Another approach bases on modification of the standard chi-square statistic of Pearson. In this paper we consider the modified Nikulin-Rao-Robson chi-square statistic for right censored data. By means of computer simulation methods we investigate statistic distributions and the power of these tests in case of close competing hypotheses.

Keywords: Goodness-of-fit tests, Kolmogorov test, Cramer-von Mises-Smirnov test, Anderson-Darling test, Nikulin-Rao-Robson chi-square test, censored sample)

1. Introduction

In reliability or survival studies lifetimes are typically right censored. The observed data are usually presented as $(t_1, \delta_1), ..., (t_n, \delta_n)$, where $\delta_i = 1$ if t_i is an observed lifetime, and $\delta_i = 0$ if t_i is a censoring time which means that lifetime of i-th individual is greater than t_i. There are various types of right-censoring mechanism:

- If individuals are observed at a predetermined time, then the censoring is called type I censoring.
- If a life test is terminated whenever a specified number of failures have occurred, it is called type II censoring.
- Let lifetime T and censoring time C are independent random variables from distribution functions $F(t)$ and $F^C(t)$ respectively. All lifetimes and censoring times are assumed mutually independent, and it is assumed that $F^C(t)$ does not depend on any of the parameters of $F(t)$. So, $t_i = \min(T_i, C_i)$ and $\delta_i = 1\{T_i \leq C_i\}$, it is called independent random censoring.

In this paper we consider the problem of testing simple hypotheses of the kind $H_0 : F(x) = F(x, \theta)$, where parameter $\theta = (\theta_1, ..., \theta_m)^T$ is known, and composite hypotheses $H_0 : F(x) \in \{F(x, \theta), \theta \in \Theta\}$. There are various goodness-of-fit tests for censored data. In [5], [3] and [14] the modification of classical Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling tests for I and II censoring types are given. The Renyi test [15] can also be used for type I and II censored samples. In case of randomly censored data these tests can be modified by using Kaplan-Maier estimate instead of empirical distribution function in the formulas of statistics (see, for example, [10], [12], [16], [7]).

Kim [9], Habib and Thomas [6], Hollander and Peña [8] considered natural modifications of the Nikulin-Rao-Robson statistic [13] to the case of censored data. These tests are also based on the differences between two estimators of the probabilities to fall into grouping intervals: one is based on the Kaplan-Meier estimator of the cumulative distribution function, other – on the maximum likelihood estimators of unknown parameters of the tested model using initial non-grouped censored data. The idea of comparing observed and expected numbers of failures in time intervals was discussed in [1] and was developed by Hjort [7]. In [2] this direction was developed considering the choice of random grouping intervals as data functions and writing simple formulas useful for computing test statistics for mostly applied classes of survival distributions.

Though a great number of papers are devoted to the problem of testing goodness-of-fit for censored samples, there are still a lot of questions concerning their application in practice. Is it correct to use the asymptotical results obtained in papers in case of limited sample sizes? How to simulate statistic distributions under true null hypothesis for randomly censored sample when the censoring distribution $F^c(t)$ is unknown? Which tests have a higher power for close competing hypotheses? So, the aim of this paper is to investigate the test statistic distributions under true null hypothesis and the test power for close competing hypotheses. The research is carried out with computer simulation technique (Monte-Carlo method).

2. Modified Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling goodness-of-fit tests

The value

$$D_n = \sup_{t<\infty} \left| \hat{F}_n(t) - F(t;\theta) \right|, \tag{1}$$

where $\hat{F}_n(t)$ is the Kaplan-Meier estimator for cumulative distribution function $F(t)$, is used in the modified Kolmogorov test statistic for right censored data. In testing hypotheses, the statistic is usually used with Bolshev's correction [4] of the form

$$S_K^c = \frac{6nD_n + 1}{6\sqrt{n}}, \tag{2}$$

where $D_n = \max\left(D_n^+, D_n^-\right)$, $D_n^+ = \max\limits_{i:\ \delta_i=1} \left\{ \hat{F}_n\left(t_{(i)}\right) - F\left(t_{(i)}, \theta\right) \right\}$, $D_n^- = \max\limits_{i:\ \delta_i=1} \left\{ F\left(t_{(i)}, \theta\right) - \hat{F}_n\left(t_{(i-1)}\right) \right\}$.

Modified Cramer-von Mises-Smirnov test statistic

$$S_\omega^c = \int\limits_{-\infty}^{\infty} \left(\hat{F}_n(t) - F(t;\theta) \right)^2 dF(t;\theta) \tag{3}$$

can be calculated by the following formula

$$\begin{aligned} S_\omega^c = n_r \sum_{j:\,\delta_j=1} \Big\{ & \hat{F}_n^2\left(t_{(j)}\right)\left(F\left(t_{(j+1)};\theta\right) - F\left(t_{(j)};\theta\right)\right) - \\ & - \hat{F}_n\left(t_{(j-1)}\right)\left(F^2\left(t_{(j+1)};\theta\right) - F^2\left(t_{(j)};\theta\right)\right) \Big\} + \frac{n_r}{3}, \end{aligned}$$

where n_r is the number of uncensored observations.

Modified Anderson-Darling test statistic

$$S_{\Omega}^{c} = \int_{-\infty}^{\infty} \left(\hat{F}_n(t) - F(t;\theta)\right)^2 \frac{dF(t;\theta)}{F(t;\theta)(1 - F(t;\theta))}, \tag{4}$$

can be calculated as following

$$S_{\Omega}^{c} = -n_r + n_r \sum_{j:\,\delta_j=1} \Big\{ \left(\hat{F}_n^2\left(t_{(j-1)}\right) - \hat{F}_n^2\left(t_{(j)}\right)\right) \log F\left(t_{(j)};\theta\right) -$$
$$- \left(\left(1 - \hat{F}_n\left(t_{(j-1)}\right)\right)^2 - \left(1 - \hat{F}_n\left(t_{(j)}\right)\right)^2\right) \log\left(1 - F\left(t_{(j)};\theta\right)\right) \Big\}.$$

2.1. *Type I and II censored samples*

When testing a simple hypothesis by type I and II censored samples there is the limiting statistic distribution of S_K^c [3] which is defined as

$$K_a^c(S) = \sum_{i=-\infty}^{+\infty} (-1)^i \exp(-2i^2 S^2) \cdot P\left\{\left|X - 2iS\sqrt{\frac{a}{1-a}}\right| < \frac{S}{\sqrt{a-a^2}}\right\}, \tag{5}$$

where X is a standard normal random variable and a is the censoring degree. When $a = 0$ the limiting distribution of S_K^c coincides with the Kolmogorov distribution law.

For the Cramer-von Mises-Smirnov and Anderson-Darling type tests the tables of upper percentage points for statistic distributions under true null hypothesis were obtained in [14].

By means of computer simulation technique we have investigated considering statistic distributions for various censoring degrees and samples sizes. In case of simple hypotheses sufficient goodness-of-fit of the empirical distributions $G(S_K^c|H_0)$ to the limiting law $K_c^a(S)$ has been shown for the sample size beginning from $n = 30$, with a censoring degree being less than 0.5. When a censoring degree was increased to 0.95 a sufficient goodness-of-fit of $G(S_K^c|H_0)$ to $K_c^a(S)$ was observed only for $n \geqslant 500$. Similar regularities for the statistic distributions were observed for the Cramer-von Mises-Smirnov test and the Anderson-Darling test. For these criteria we compared sample quantiles obtained from the empirical distributions of the test statistics with the upper percentage points given in [14] depending on the sample size and the censoring degree.

When testing a composite hypothesis there are no principal problems to simulate test statistic distribution for given parametric distribution under null hypothesis, sample size and censoring scheme (for censoring of type I or II). Using simulated statistic distribution one can estimate the p-value in testing hypothesis H_0.

2.2. *Randomly censored samples*

Investigation of considering statistic distributions for various distributions of censoring times $F^c(t)$ has revealed an essential dependence of statistic distributions on $F^c(t)$. As an example in Figure 1 you can see the modified Kolmogorov statistic distributions when testing composite goodness-of-fit hypothesis with the Weibull distribution for different censoring degrees. In this example the distributions of censoring times were taken from the family of the Beta I distributions, where parameters were taken so that the censoring degree would be approximately equal to 0.2, 0.4, 0.6, 0.8.

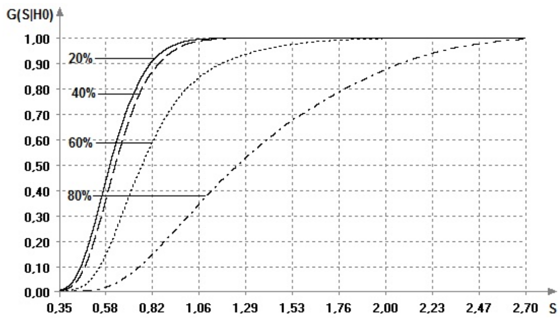

Figure 1. Kolmogorov test statistic distributions for different censoring degrees when testing the composite hypothesis of goodness-of-fit with the Weibull distribution, $n = 100$

Figure 2 illustrates the modified Kolmogorov statistic distributions when testing composite goodness-of-fit hypothesis with the Weibull distribution for two different distributions of censoring times. The first one is the Beta I distribution and the second one is the Weibull distribution. Parameters of these distributions were taken so that the censoring degree would be approximately equal to 0.6. As it is seen from the figure statistic distributions strongly depend not only on the number of censoring times, but also on that how they are distributed.

This result is a serious barrier for application of these modified tests in practice, because the distribution of censoring times $F^c(t)$ is usually unknown.

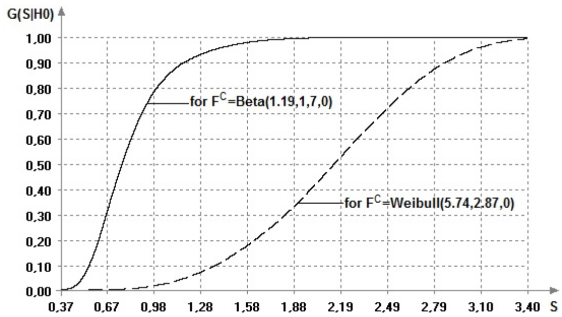

Figure 2. Kolmogorov test statistic distributions for different distributions of censoring times when testing the composite hypothesis of goodness-of-fit with the Weibull distribution, $n = 100$

2.3. *Application of classical tests for censored samples*

Here we would like to discuss an idea of transformation of censored sample into a "complete" sample in order to apply the classical goodness-of-fit tests. At first, let us consider a simple goodness-of-fit hypothesis H_0 : $F(x) = F(x, \theta)$. In the sample of observations $(t_1, \delta_1), (t_2, \delta_2), ..., (t_n, \delta_n)$ we replace all censored observations $(t_i, \delta_i = 0) = C_i$ by simulated times $\hat{T}_i = F^{-1}(\xi_i)$, where ξ_i is uniformly distributed at the interval $[F(C_i), 1)$. So, the classical Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling tests can be applied for transformed sample. In case of simple hypothesis testing nonparametric statistic distributions obtained for such transformed samples converge to their limiting distributions very quickly. For the sample size $n \geq 20$ one can use the corresponding limiting law (the Kolmogorov distribution, $a1(S)$ or $a2(S)$) for calculation of p-value without any risk of making a great mistake.

It is obvious that such procedure results in decrease of the test power. But if the censoring degree is not high (less than 30 percent) the losses in power are not large.

In the case of composite hypothesis testing the application of considered procedure is more complicated. Let $\hat{\theta}_n$ is the maximum likelihood estimate of unknown parameter calculated by the original censored sample. We replace all censored observations $(t_i, \delta_i = 0) = C_i$ by simulated times $\hat{T}_i = F^{-1}(\xi_i)$, where ξ_i is uniformly distributed at the interval $[F(C_i, \hat{\theta}_n), 1)$. After this transformation it is necessary to estimate unknown parameters of hypothetical distribution by obtained "complete" sample again. In case of small censoring degrees, when there is no significant bias of parameter estimates, the distributions of classical test statistics for transformed samples are close to the statistic distribution models for originally complete samples when testing the same composite hypothesis. So, if the censoring degree is not large it is possible to use the apparatus of testing composite hypotheses by complete samples. In [11] the approximations of the limiting statistic distributions for testing composite hypotheses were obtained for a wide range of distribution laws when using maximum likelihood estimates of unknown parameters.

3. NRR χ^2 test for censored samples

Chi-squared type tests require dividing an observed interval $[0, \tau]$ into k smaller intervals $I_j = (a_{j-1}, a_j]$, $a_0 = 0$, $a_k = \tau$.

Denote by $U_j = \sum\limits_{i: X_i \in I_j}^{n} \delta_i$ the number of observed failures and by e_j an "expected" number of failures in the interval I_j, $j = 1, ..., k$.

The NRR χ^2 test statistic [2] can be written in the form

$$Y^2 = Z^T \hat{V}^- Z,$$

where

$$Z = (Z_1, ..., Z_k)^T, Z_j = \frac{1}{\sqrt{n}} (U_j - e_j),$$

$\hat{V}^-$ is the general inverse of the matrix $\hat{V}$,

$$\hat{V}^- = \hat{A}^{-1} + \hat{A}^{-1} \hat{C}^T \hat{G}^- \hat{C} \hat{A}^{-1},$$

$\hat{A}$ is the diagonal $k \times k$ matrix with diagonal elements

$$\hat{A}_j = \frac{U_j}{n}, j = 1, \ldots, k,$$

$$\hat{G} = [\hat{g}_{ll'}]_{m \times m}, \hat{g}_{ll'} = \hat{i}_{ll'} - \sum_{j=1}^{k} \hat{C}_{lj} \hat{C}_{l'j} \hat{A}_j^{-1},$$

$$\hat{i}_{ll'} = \frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{\partial \ln \lambda \left(t_i, \hat{\theta}\right)}{\partial \theta_l} \frac{\partial \ln \lambda \left(t_i, \hat{\theta}\right)}{\partial \theta_{l'}},$$

$$\hat{C}_{lj} = \frac{1}{n} \sum_{i: t_i \in I_j}^{n} \delta_i \frac{\partial}{\partial \theta_l} \ln \lambda \left(t_i, \hat{\theta}\right),$$

where $\lambda(t)$ is the hazard rate function and $\hat{\theta}$ is the maximum likelihood estimate of unknown parameter.

The choice of group intervals. We recommend to take a_j as random data functions so that to divide the interval $[0, \tau]$ into k intervals with equal expected numbers of failures. So a_j can be calculated as follows. Define

$$E_k = \sum_{i=1}^{n} \Lambda\left(t_i, \hat{\theta}\right), E_j = \frac{j}{k} E_k, j = 1, \ldots, k.$$

Set

$$b_i = (n-i)\,\Lambda\left(t_{(i)}, \hat{\theta}\right) + \sum_{l=1}^{i} \Lambda\left(t_{(i)}, \hat{\theta}\right), t_{(0)} = 0.$$

If i is the smallest natural number verifying $E_j \in [b_{i-1}, b_i]$, $j = 1, ..., k-1$, then

$$\hat{a}_j = \Lambda^{-1}\left(\left(E_j - \sum_{l=1}^{i} \Lambda\left(t_{(l)}, \hat{\theta}\right)\right) / \left(n - i + 1\right), \hat{\theta}\right), \hat{a}_k = t_{(n)},$$

where Λ^{-1} is the inverse of the cumulative hazard function Λ. We have

$$0 < \hat{a}_1 < \hat{a}_2 < ... < \hat{a}_k = \tau.$$

Under this choice of the intervals $e_j = E_k/k$ for any j.

The limit distribution of the test statistic is χ^2_r, $r = rank(V^-)$. So the hypothesis is rejected with approximate significance level α if $Y^2 > \chi^2_\alpha\,(r)$.

By means of computer simulation we investigate the NRR statistic distribution for various sample sizes, censoring degrees, number of intervals k. For example, in Figure 3 there are the NRR statistic distributions when testing the composite hypothesis of goodness-of-fit with the Weibull distribution by randomly censored samples of size $n = 100$ for different censoring degrees. The number of intervals $k = 5$. As you can see from the figure, obtained empirical distributions of considering statistic are rather close to the limiting χ^2_4-distribution.

With the sample size growth NRR statistic distributions converge to the corresponding χ^2_r distribution law. But the minimal sample size for which empirical distributions of the NRR statistic fit with the corresponding χ^2_r distribution law depends on the censoring degree.

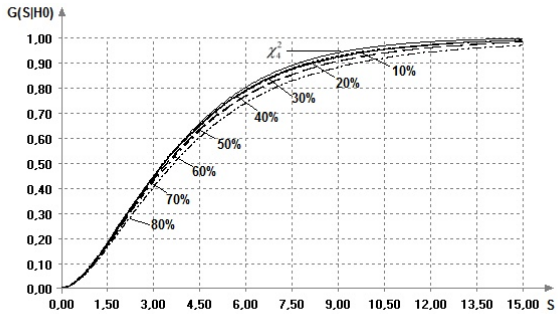

Figure 3. RRN χ^2 test statistic distributions for different censoring degrees when testing the composite hypothesis of goodness-of-fit with the Weibull distribution, $n = 100$, $k = 5$

4. Conclusions

There are no principal difficulties in the usage of modified Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling tests in case of type I and II censored data. One can use the limiting statistic distribution or tables of percentage points for these tests in simple hypothesis testing. While testing a composite hypothesis it is possible to simulate statistic distribution for considering distribution under test and given censoring scheme. But when we have randomly censored data the distributions of these statistics strongly depend on the distribution of censoring times. This fact doesn't enable to recommend using these tests for randomly censored samples.

There is a good possibility to use considered transformation of a censored sample to "complete" one for application of classical Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling tests. In case of small censoring degrees, when there is no significant bias of parameter estimates, it is possible to use the approximations of the limiting statistic distributions obtained in [11] for calculation of p-value while testing composite hypotheses. The loss of power of the test applied for transformed samples is not significant if the censoring degree is not high.

The NRR χ^2 test has a number of advantages comparing with the considered nonparametric tests. In particular, there is no significant dependence of the NRR statistic distributions (in case of limited sample sizes) on the distribution of censoring times. With the sample size growth NRR statistic distributions converge to the corresponding χ^2_r distribution law. It is possible to choose the number of grouping intervals and boundary points for considering pair of competing hypotheses to increase the test power.

5. Acknowledgments

These researches have been supported by the Federal targeted program of the Ministry of Education and Science of the Russian Federation "Scientific and scientific-educational personnel of innovative Russia" in 2009-2013.

References

[1] Akritas M.G. (1988) Pearson-type goodness-of-fit tests: the univariate case. J. Amer. Statist. Assoc. Vol. 83 - P. 222-230.

[2] Bagdonavicus V., Kruopis J., Nikulin M. (2010) Nonparametric Tests for Censored Data. John Wiley and Sons - 233 p.

[3] Barr D.M., Davidson T. (1973) A Kolmogorov-Smirnov test for censored samples. Technometrics. V.15. No. 4.

[4] Bolshev L.N., Smirnov N.V. (1983) Tables of Mathematical Statistics. Moscow: Science. (in Russian).

[5] D'Agostino R.B., Stephens M.A. (1986) Goodness of Fit Techniques. NY: Marcel Dekker.

[6] Habib M.G., Thomas D.R. (1986) Chi-square goodness-of-fit tests for randomly censored data. Annals of Statistics. vol. 14. P. 759-765.

[7] Hjort N.L. (1992) On Inference in Parametric Survival Data. International Statistical Review. Vol. 60, No. 3. - P. 355-387.

[8] Hollander M., Peña E.A. (1992) A chi-squared goodness-of-fit test for randomly censored data. J.Amer.Statist.Assoc. vol. 87, p.458-463.

[9] Kim J.H. (1993) Chi-Square Goodness-of-Fit Tests for Randomly Censored Data. The Annals of Mathematical Statistics. Vol. 21, No. 3. - P. 1621-1639.

[10] Koziol J.A., Green S.B. (1976) A Cramer-von Mises Statistic for Randomly Censored Data. Biometrika. Vol. 63, No. 3. - P. 465-474.

[11] Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N. (2010) Statistic Distribution Models for Some Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses. Communications in Statistics - Theory and Methods. Vol. 39, No. 3. - P. 460-471.

[12] Nair V. (1981) Plots and tests for goodness of fit with randomly censored data. Biometrika. Vol. 68. - P. 99-103.

[13] Nikulin M.S. (1973) On the chi-squared test for continuous distributions. Prabability theory and applications 1973. Vol. XVIII. No 3. - P.675-676.

[14] Pettitt A.N., Stephens M.A. (1976) Modified Cramer von Mises statistics for censored data. Biometrika. Vol. 63. N. 2.

[15] Renyi A. (1953) On the theory of order statistics. Acta Mathem. Acad. Sci. Hung. Vol. 4. - P. 191-232.

[16] Reineke D., Crown J. (2004) Estimation of Hazard, Density and Survival Functions for Randomly Censored Data. Journal of Applied Statistics. Vol. 31, No. 10. - P. 1211-1225.