Chapter 18

A Comparative Analysis of Some Chi-Square Goodness-of-Fit Tests for Censored Data

In this chapter, we consider some chi-square-type goodness-of-fit tests for right censored samples, viz. the Nikulin–Rao–Robson (NRR) test and generalized Pearson–Fisher (GPF) test. We compare these tests in terms of power in the case of testing composite hypotheses by randomly censored and Type II censored samples. Various grouping methods are considered, including equifrequent grouping, grouping into intervals with equal expected numbers of failures as well as optimal grouping for some given pairs of competing hypotheses. In the case of Type II censored data, we compare by power chi-square goodness-of-fit tests with the Kolmogorov, Cramer–von Mises–Smirnov and Anderson–Darling tests.

18.1. Introduction

In this chapter, we consider the problem of testing the composite goodness-of-fit hypotheses $H_0 : F(t) \in \{F_0(t;\theta), \theta \in \Theta\}$, meaning that the distribution F of failure time T belongs to a certain parametric family $F_0(\cdot;\theta)$, where $\theta = (\theta_1, ..., \theta_q)^T \in \Theta \subset \mathbf{R}^q$ is unknown q-dimensional parameter. Let us denote by $\lambda_0(t;\theta)$ and $\Lambda_0(t;\theta)$ the hazard rate and cumulative hazard rate functions corresponding to the tested model, respectively. In reliability or survival studies, the observed data are usually of the form

$$\mathrm{X} = (X_1, \delta_1), (X_2, \delta_2), ..., (X_n, \delta_n),$$

Chapter written by Ekaterina CHIMITOVA and Boris LEMESHKO.

Statistical Models and Methods for Reliability and Survival Analysis
Edited by Vincent Couallier, Léo Gerville-Réache, Catherine Huber-Carol, Nikolaos Limnios and Mounir Mesbah

where $X_i = \min(T_i, C_i)$ is the observation value, T_i is the lifetime and C_i is a censoring time. The censoring indicator $\delta_i = I\{T_i \leq C_i\}$.

There are various forms of right censoring (for details, see [COH 91]):

– If individuals are observed up to some predetermined time c, then censoring is referred to as Type I censoring and in the case when $\delta_i = 0 : X_i = c$.

– If a life test is terminated whenever a specified number of failures (say, m) have occurred, it is referred to as Type II censoring and in this case when $\delta_i = 0$: $X_i = X_{(m)}$, where $X_{(m)}$ is the last observed failure time.

– Let the lifetime T and the censoring time C be independent random variables from the distribution functions $F(t)$ and $F^C(t)$, respectively. All lifetimes and censoring times are assumed mutually independent. Then, this form of censoring is referred to as independent random censoring.

In the case of complete data (without censored observations), a well-known modification of the classical chi-square tests is the NRR statistic, which was proposed for the first time in [NIK 73c] and further developed in [NIK 73b], [NIK 73a] and [RAO 74]. Later, it was studied in [LEM 01], [VOI 06] and [VOI 13]. The NRR statistic is based on the differences between two estimators of the probabilities to fall into grouping intervals: one estimator is based on the empirical distribution function and the other on the maximum likelihood estimators of unknown parameters of the tested model using initial non-grouped data. Habib and Thomas [HAB 86] and Hollander and Peña [HOL 92] considered natural modifications of the NRR statistic for the case of censored data. The idea of comparing observed and expected numbers of failures into grouping intervals in the case of randomly censored data was proposed by Akritas [AKR 88] and was developed independently by Hjort [HJO 90]. Bagdonavičius *et al.* [BAG 10] have developed this direction, considering the choice of random grouping intervals as data functions and writing simple formulas of test statistics for most applied families of lifetime distributions. The NRR chi-square goodness-of-fit test has been also developed for parametric accelerated failure time models [BAG 10, BAG 13].

Li and Doss [LI 93] have developed the generalization of the classical Pearson–Fisher chi-square goodness-of-fit test for any situation, for which there is available a non-parametric estimator $\hat{F}$ of F, such that $\sqrt{n}\left(\hat{F} - F\right) \xrightarrow{d} W$, where W is a continuous zero mean Gaussian process satisfying a mild regularity condition. The GPF statistic for censored data is based on the differences between two estimators of the probabilities to fall into grouping intervals: one estimator is based on the Kaplan–Meier estimator of lifetime distribution and the other on the minimum chi-square estimator of unknown parameters. This test was also developed for randomly censored data by Kim [KIM 93].

In this chapter, we present an empirical analysis of the NRR and GPF chi-square tests for right censored data. We compare these tests by means of power in the case of testing composite goodness-of-fit hypotheses by randomly censored and Type II censored samples. Various grouping methods are considered. The investigation is carried out through Monte Carlo simulations. From the obtained results, we draw some conclusions.

18.2. Chi-square goodness-of-fit tests for censored data

Chi-square-type tests require dividing an observed interval $[0, \tau]$ into k smaller intervals $I_j = (a_{j-1}, a_j]$, $0 = a_0 < a_1 < \ldots < a_{k-1} < a_k = \tau$.

18.2.1. *NRR* χ^2 *test*

For any $t > 0$, set

$$N(t) = \sum_{i=1}^{n} I\left\{X_i \leq t,\, \delta_i = 1\right\},$$

which counts the number of observed failures in $[0,\ t]$, and

$$Y(t) = \sum_{i=1}^{n} I\left\{X_i \geq t\right\},$$

which is the number of objects at risk just before time t.

Denote by $U_j = N(a_j) - N(a_{j-1})$ the number of observed failures and by

$$e_j = \int\limits_{a_{j-1}}^{a_j} \lambda_0\left(u;\hat{\theta}_n\right) Y(u)du$$

the "expected" number of failures in the interval I_j, $j = 1, ..., k$.

The NRR χ^2 test statistic can be written in the form [BAG 10]

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

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

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

$\hat{A}$ is the diagonal $k \times k$ matrix with diagonal elements $\hat{A}_j = U_j/n$

$$\hat{i} = \frac{1}{n}\int_0^\tau \frac{\partial}{\partial\theta}\ln\lambda_0\left(u;\hat{\theta}_n\right)\left(\frac{\partial}{\partial\theta}\ln\lambda_0\left(u;\hat{\theta}_n\right)\right)^T dN(u),$$

$$\hat{C}_j = \frac{1}{n}\int_{I_j} \frac{\partial}{\partial\theta}\ln\lambda_0\left(u;\hat{\theta}_n\right) dN(u),$$

where $\hat{\theta}_n$ is the maximum likelihood estimate of unknown parameters of the tested model using initial non-grouped censored data. The limiting distribution of the test statistic is χ_r^2, $r = rank(V^-)$. So, the hypothesis is rejected with approximate significance level α if $Y_n^2 > \chi_\alpha^2\,(r)$.

In [CHI 11a] and [CHI 11b], we have studied the distributions of statistic Y_n^2 and the power of NRR test, depending on the form and degree of censoring by means of Monte Carlo simulations.

18.2.2. *GPF χ^2 test*

Let $\hat{F}_n(t)$ be the Kaplan–Meier estimator of the lifetime distribution F. Define the empirical probabilities to fall into grouping intervals as

$$\hat{p}_j = \hat{F}_n(a_j) - \hat{F}_n(a_{j-1}), j = 1, ..., k,$$

where $\hat{p}_1$ and $\hat{p}_k$ must be defined as $\hat{p}_1 = \hat{F}_n(a_1)$ and $\hat{p}_k = 1 - \hat{F}_n(a_{k-1})$. Denote by

$$p_j(\theta) = F_0(a_j, \theta) - F_0(a_{j-1}, \theta)$$

the theoretical probabilities to fall into grouping intervals. Unknown parameters of the tested distribution are estimated as follows

$$\tilde{\theta}_n = Arg\min_\theta \xi_n^T(\theta)\xi_n(\theta),$$

where $\xi_n(\theta) = \left(\frac{n\hat{p}_1 - np_1(\theta)}{\sqrt{np_1(\theta)}}, ..., \frac{n\hat{p}_k - np_k(\theta)}{\sqrt{np_k(\theta)}}\right)^T$.

It is well known [LI 93] that $\xi_n(\tilde{\theta}_n) \xrightarrow{d} N(0, \Sigma)$, where

$$\Sigma = PDJ\Sigma^{(1)}J^TDP,$$

$$P = I - C\left(C^T C\right)^{-1} C^T, C = D\left(\frac{\partial}{\partial\theta} p(\theta)\right)^T, p(\theta) = (p_1(\theta), ..., p_k(\theta))^T,$$

$$D = \operatorname{diag}\left((p_1(\theta))^{-1/2}, ..., (p_k(\theta))^{-1/2}\right),$$

$$J = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & -1 \end{pmatrix}_{k\times(k-1)}.$$

The elements of matrix $\Sigma^{(1)}$ are calculated by the formula

$$\sigma_{ij}^{(1)} = (1 - F_0(a_i, \theta)) \cdot (1 - F_0(a_j, \theta)) \cdot v(\min(a_i, a_j)),$$

where the variance function $v(t) = \int\limits_0^t \frac{dF_0(u;\theta)}{(1-F(u;\theta))^2(1-F^C(u))}$. Here, the Kaplan–Meier estimator, obtained by the inverted sample, in which δ_i is replaced by $1 - \delta_i$, can be used as the distribution of censoring times $F^C(t)$. Let $\hat{\Sigma}$ be obtained by using the minimum chi-square estimator $\tilde{\theta}_n$ as θ in formulas for the matrix Σ, and $\hat{\Sigma}^+$ denotes the Moore–Penrose inverse of $\hat{\Sigma}$. Then, under H_0, the statistic of GPF test

$$X_n^2 = \xi_n^T\left(\tilde{\theta}_n\right)\hat{\Sigma}^+\xi_n\left(\tilde{\theta}_n\right)$$

has the χ^2-distribution with $k - q - 1$ degrees of freedom as $n \to \infty$.

18.3. The choice of grouping intervals

When using chi-square goodness-of-fit tests, the problem of choosing boundary points and the number of grouping intervals is always important, as the power of these tests considerably depends on the grouping method used. In the case of complete samples (without censored observations), this problem was investigated in [VOI 09], [DEN 89], [LEM 98], [LEM 00], [LEM 01] and [DEN 04]. In particular, in [LEM 00], the investigation of the power of the Pearson and NRR tests for complete samples has been carried out for various numbers of intervals and grouping methods. The partition of the real line into equiprobable intervals (EPG) is not an optimal grouping method, as a rule. In [DEN 79], it was shown for the first time that asymptotically optimal grouping, for which the loss of the Fisher information from grouping is minimized, enables us to maximize the power of the Pearson test against close competing hypotheses. For example, it is possible to maximize the determinant of the Fisher information matrix for grouped data

$$\mathbf{J}_G(\theta) = \sum_{j=1}^{k} \frac{\frac{\partial}{\partial\theta} p_j(\theta) \left(\frac{\partial}{\partial\theta} p_j(\theta)\right)^T}{p_j(\theta)},$$

i.e. to solve the problem of D-optimal grouping

$$\max_{a_1<\ldots<a_{k-1}} \det\left(\mathbf{J}_G(\theta)\right). \qquad [18.1]$$

In the case of an A-optimality criterion, the trace of the information matrix $\mathbf{J}_G(\theta)$ is maximized by the boundary points

$$\max_{a_1<\ldots<a_{k-1}} \mathrm{Tr}\left(\mathbf{J}_G(\theta)\right), \qquad [18.2]$$

and an E-optimality criterion maximizes the minimum eigenvalue of the information matrix

$$\max_{a_1<\ldots<a_{k-1}} \min_{l=\overline{1,q}} \lambda_l\left(\mathbf{J}_G(\theta)\right). \qquad [18.3]$$

The problem of asymptotically optimal grouping by the A- and E-optimality criteria has been solved for certain distribution families, and the tables of A-optimal grouping are given in [LEM 11]. The versions of asymptotically optimal grouping maximize the test power relative to a set of close competing hypotheses, but they do not ensure the highest power against some given competing hypothesis. For the given competing hypothesis H_1, it is possible to construct the χ^2 test, which has the highest power for testing hypothesis H_0 against H_1. For example, in the case of χ^2 Pearson test, it is possible to maximize the non-centrality parameter for the given number of intervals k:

$$\max_{a_1<a_2<\ldots<a_{k-1}} \left(n \sum_{j=1}^{k} \frac{\left(p_j^1\left(\theta^1\right) - p_j^0\left(\theta^0\right)\right)^2}{p_j^0\left(\theta^0\right)} \right), \qquad [18.4]$$

where

$$p_j^0\left(\theta^0\right) = \int_{a_{j-1}}^{a_j} f_0\left(u;\theta^0\right) du \quad \text{and} \quad p_j^1\left(\theta^1\right) = \int_{a_{j-1}}^{a_j} f_1\left(u;\theta^1\right) du$$

are the probabilities to fall into jth interval according to the hypotheses H_0 and H_1, respectively. Let us refer to this grouping method as optimal grouping.

Asymptotically optimal boundary points, corresponding to different optimality criteria, as well as the optimal points, corresponding to [18.4], are considerably different from each other. For example, the boundary points maximizing criteria [18.1]–[18.4] for the following pair of competing hypotheses are given in Table 18.1. The null hypothesis H_0 is the normal distribution with density function $f(t) = \frac{1}{\theta_2\sqrt{2\pi}} \exp\left\{-\frac{(t-\theta_1)^2}{2\theta_2^2}\right\}$, and parameters $\theta_1 = 0$, $\theta_2 = 1$, and the competing hypothesis H_1 is the logistic distribution with density function $f(t) = \frac{\pi}{\theta_2\sqrt{3}} \exp\left\{-\frac{\pi(t-\theta_1)}{\theta_2\sqrt{3}}\right\} \Big/ \left[1 + \exp\left\{-\frac{\pi(t-\theta_1)}{\theta_2\sqrt{3}}\right\}\right]^2$, and parameters $\theta_1 = 0$, $\theta_2 = 1$.

Optimality criterion	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
A-optimum	−2.3758	−1.6915	−1.1047	−0.4667	0.4667	1.1047	1.6915	2.3758
D-optimum	−2.3188	−1.6218	−1.0223	−0.3828	0.3828	1.0223	1.6218	2.3188
E-optimum	−1.8638	−1.1965	−0.6805	−0.2216	0.2216	0.6805	1.1965	1.8638
Optimal grouping	−3.1616	−2.0856	−1.2676	−0.4601	0.4601	1.2676	2.0856	3.1616

Table 18.1. *Optimal boundary points for* $k = 9$

Moreover, in the case of the given competing hypothesis, we can use the so-called Neyman–Pearson classes [GRE 96], for which the random variable domain is partitioned into intervals of two types, according to the inequalities $f_0(t) < f_1(t)$ and $f_0(t) > f_1(t)$, where $f_0(t)$ and $f_1(t)$ are the density functions, corresponding to the competing hypotheses. For H_0 and H_1 from our example, we have the first-type intervals

$$(-\infty; -2.3747], (-0.6828; 0.6828], (2.3747; \infty),$$

and the second-type intervals

$$(-2.3747; -0.6828], (0.6828; 2.3747].$$

Figures 18.1 and 18.2 illustrate the power of the Pearson χ^2 test for the hypotheses H_0 and H_1 of our example in the case of different grouping methods, depending on the number of intervals ($\alpha = 0.1$, $n = 500$). The powers of the well-known non-parametric Kolmogorov, Cramer-von Mises-Smirnov and Anderson–Darling goodness-of-fit tests are given for the comparison.

It is obvious that the power of the χ^2 tests for censored samples is also influenced by some other factors, associated with the form of censoring (the type and degree of censoring, as well as the distribution of censoring times). The usage of grouping methods, mentioned above, for censored samples may result in some difficulties, such as the fact that not all boundary points may belong to the observed interval

$[0, \tau]$, or some intervals may not contain complete observations, which does not allow calculation of the NRR statistic. In such situations, it is necessary to combine intervals or change boundary points of intervals. In any case, the boundary points of grouping intervals should be chosen, taking into account the form of censoring. In this chapter, we consider three such grouping methods.

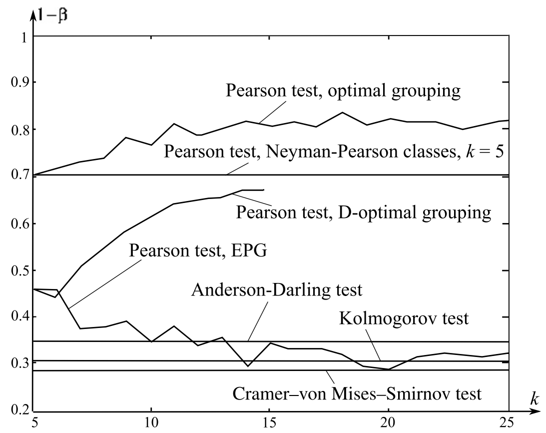

Figure 18.1. *The power of the χ^2 Pearson test when testing simple hypothesis*

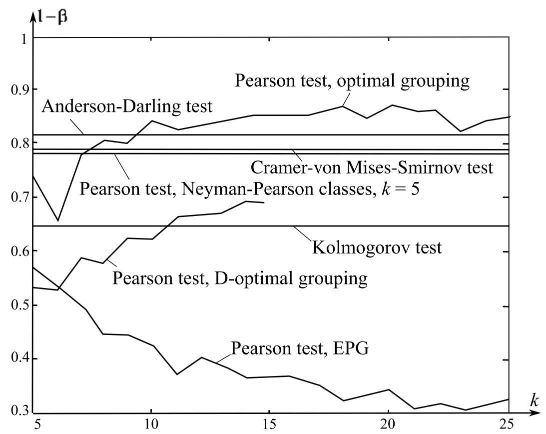

Figure 18.2. *The power of the χ^2 Pearson test when testing composite hypothesis*

18.3.1. *Equifrequent grouping (EFG)*

In this grouping method, the boundary points are chosen in the interval $[0, \tau]$ from the condition of equal (or almost equal) numbers of complete observations, falling into grouping intervals.

18.3.2. *Intervals with equal expected numbers of failures (EENFG)*

In [BAG 10], it is recommended to take a_j as random data functions, dividing the interval $[0, \tau]$ into k intervals with equal expected numbers of complete observations (which are not necessary integer). In this case, a_j are calculated as follows. Define

$$E_k = \int\limits_0^\tau \lambda_0 \left(u; \hat{\theta}_n\right) Y(u) du,$$

$E_j = \frac{j}{k} E_k$, $j = 1, ..., k$. Denote by $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ the ordered sample from $X_1, ..., X_n$, $X_{(0)} = 0$. Set

$$b_i = (n - i)\, \Lambda_0 \left(X_{(i)}; \hat{\theta}_n\right) + \sum_{l=1}^{i} \Lambda_0 \left(X_{(l)}; \hat{\theta}_n\right).$$

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

$$\hat{a}_j = \Lambda_0^{-1} \left(\left(E_j - \sum_{l=1}^{i-1} \Lambda_0 \left(X_{(l)}; \hat{\theta}_n\right) \right) \Big/ (n - i + 1); \hat{\theta}_n \right),$$

$\hat{a}_k = X_{(n)}$, where $\Lambda_0^{-1}(\cdot)$ is the inverse of the cumulative hazard rate function.

18.3.3. *Optimal grouping (OptG)*

In the case of testing hypothesis H_0 against a given competing hypothesis H_1 by randomly censored sample, the optimal boundary points of the grouping intervals can be obtained by solving the maximization problem, which is similar to [18.4]

$$\max_{a_1 < a_2 < \dots < a_{k-1}} \left(n \sum_{j=1}^{k} \frac{\left(p_j^1\left(\theta^1\right) - p_j^0\left(\theta^0\right)\right)^2}{p_j^0\left(\theta^0\right)} \right), \qquad [18.5]$$

where

$$p_j^0\left(\theta^0\right) = \int\limits_{a_{j-1}}^{a_j} \left(1 - F^C\left(u\right)\right) f_0\left(u; \theta^0\right) du$$

and

$$p_j^1\left(\theta^1\right) = \int\limits_{a_{j-1}}^{a_j} \left(1 - F^C\left(u\right)\right) f_1\left(u; \theta^1\right) du$$

are the probabilities to fall into jth interval according to hypotheses H_0 and H_1, respectively. If the distribution of censoring times $F^C(t)$ is unknown (which is a typical situation in practice), then we can use the Kaplan–Meier estimator as $F^C(t)$, which is obtained by the inverted sample, in which δ_i is replaced by $1 - \delta_i$.

REMARK 18.1.– The maximized function in [18.5] is multiextremal, so the global optimization methods are required to solve this problem.

REMARK 18.2.– Strictly speaking, the usage of EENFG and OptG does not guarantee that each grouping interval will contain complete observations, especially in the case of large degrees of censoring. In our simulation study, when such a situation occurs, we change boundary points so that each interval would contain at least one complete observation.

18.4. Empirical power study

In this chapter, we study the power of the considered χ^2 goodness-of-fit tests for two pairs of close competing hypotheses through Monte Carlo simulations. Three grouping methods: EFG, EENFG and OptG are used, and the number of intervals k is taken as equal to 3, 4, 5 and 6. Let us emphasize that all results presented below have been obtained in the case of testing composite hypotheses. The estimates of the power are calculated from the distributions of the test statistics $G\left(S|H_0\right)$ and $G\left(S|H_1\right)$, which are simulated on the basis of censored samples of size $n = 200$. The number of simulations used is $N = 10^5$. The values of the power of the tests are calculated with the nominal significance level $\alpha = 0.1$. Let us consider first the null hypothesis H_0 is the Weibull distribution with the density function $f\left(t; \theta\right) = \frac{\theta_2}{t}\left(\frac{t}{\theta_1}\right)^{\theta_2} \exp\left(-\left(\frac{t}{\theta_1}\right)^{\theta_2}\right)$ and the parameters $\theta_1 = 2$, $\theta_2 = 2$. The competing hypothesis H_1 is the gamma distribution with the density function $f\left(t; \theta\right) = \frac{1}{\theta_1 \cdot \Gamma(\theta_2)}\left(\frac{t}{\theta_1}\right)^{\theta_2 - 1} \exp\left(-\frac{t}{\theta_1}\right)$ and parameters $\theta_1 = 0.5577$, $\theta_2 = 3.1215$.

In the case of random censoring, we need to specify the distribution of censoring times. We have chosen two different families of distributions for censoring times: Type I beta-distribution family with the density function

$$f^C(t;\theta) = B_I(\theta_1, \theta_2, \theta_3) = \frac{1}{\theta_3 \mathrm{B}(\theta_1, \theta_2)} \left(\frac{t}{\theta_3}\right)^{\theta_1 - 1} \left(1 - \frac{t}{\theta_3}\right)^{\theta_1 - 1},$$

where $\mathrm{B}(a, b)$ is the beta function; and the Weibull distribution family. Censoring distributions are given in Table 18.2. The distribution parameters were adjusted so that the average degree of censoring under considered null hypothesis is equal to some given value. In the case of Type I beta distributions, censored observations appear in the variational series of a censored sample approximately uniformly, as opposed to the Weibull distributions, for which censored observations appear generally at the end of the variational series.

Average degree of censoring	Distribution of censoring times	
10%	$B_I(1.81, 1, 7)$	W(3.44, 6.88)
20%	$B_I(1.19, 1, 7)$	W(2.87, 5.74)
30%	$B_I(1, 1.24, 7)$	W(2.48, 4.96)
40%	$B_I(1, 1.83, 7)$	W(2.16, 4.32)
50%	$B_I(1, 2.58, 7)$	W(1.87, 3.74)

Table 18.2. *Distributions of censoring times in the case of the lifetime distribution W(2, 2)*

In Table 18.3, the powers of the considered χ^2 tests are presented for various degrees of censoring in the case of randomly censored samples with censoring times from the Weibull distributions with parameter values, given in Table 18.1. In parentheses, there is the number of intervals for which the corresponding estimate of the power was obtained. In Table 18.4, the powers of the considered χ^2 tests are presented in the case of Type I beta distributions of censoring times.

Test	Grouping method	Average degree of censoring					
		0%	10%	20%	30%	40%	50%
NRR	EFG	0.51 (5)	0.38 (5)	0.32 (5)	0.28 (5)	0.25 (5)	0.22 (5)
	EENFG	0.51 (5)	0.41 (5)	0.34 (5)	0.30 (5)	0.27 (5)	0.23 (5)
	OptG	0.61 (6)	0.46 (4)	0.41 (4)	0.37 (4)	0.32 (3)	0.29 (3)
GPF	EFG	0.20 (6)	0.17 (6)	0.15 (6)	0.15 (6)	0.14 (6)	0.13 (6)
	EENFG	0.26 (6)	0.22 (6)	0.19 (6)	0.18 (6)	0.16 (6)	0.16 (6)
	OptG	0.56 (4)	0.51 (4)	0.44 (4)	0.37 (4)	0.35 (4)	0.29 (4)

Table 18.3. *Power of the tests for the Weibull-gamma pair in the case of randomly censored data with Weibull distribution of censoring times*

As can be seen from Tables 18.3 and 18.4, the power of the considered tests depends on the distribution of censoring times: in the case of Type I beta distribution, the power of the tests is higher, than for the Weibull distribution of censoring times. The GPF χ^2 test is considerably worse in terms of the power than the NRR test, when using EFG and EENFG. But in the case of optimal grouping, it is impossible to give preference to one test. The power of both tests in the case of optimal grouping is much higher than in the case of equifrequent grouping and grouping into intervals with equal expected numbers of failures.

Test	Grouping method	Average degree of censoring					
		0%	10%	20%	30%	40%	50%
NRR	EFG	0.51 (5)	0.43 (5)	0.40 (5)	0.36 (5)	0.33 (5)	0.28 (5)
	EENFG	0.51 (6)	0.46 (6)	0.42 (6)	0.38 (6)	0.34 (6)	0.30 (6)
	OptG	0.61 (6)	0.54 (6)	0.51 (4)	0.48 (4)	0.44 (4)	0.40 (4)
GPF	EFG	0.20 (6)	0.18 (6)	0.17 (6)	0.17 (6)	0.15 (6)	0.15 (6)
	EENFG	0.26 (6)	0.24 (6)	0.23 (6)	0.22 (6)	0.21 (6)	0.19 (6)
	OptG	0.56 (4)	0.53 (4)	0.49 (4)	0.45 (4)	0.41 (4)	0.37 (4)

Table 18.4. *Power of the tests for the Weibull-gamma pair in the case of randomly censored data with Type I beta distribution of censoring times*

It is significant to compare the considered χ^2 tests with the modified Kolmogorov, Cramer–von Mises–Smirnov and Anderson–Darling tests for censored samples, which have been studied in detail through Monte Carlo simulations in [LEM 10a] for Type I and Type II censored data and in [LEM 13] for randomly censored data. In Table 18.5, the powers of the considered χ^2 tests are presented for the Weibull-gamma pair of competing hypotheses in the case of Type II censored samples, when using the optimal grouping. In parentheses, there is the number of intervals, for which the corresponding estimate of the power was obtained. The powers of modified Kolmogorov, Cramer-von Mises-Smirnov and Anderson–Darling tests for censored samples are given for comparison in the same table.

As can be seen from Table 18.5, the NRR χ^2 test has an advantage in power comparing not only with GPF test, but also with modified non-parametric goodness-of-fit tests.

In Table 18.6, the powers of the considered χ^2 tests, as well as the powers of modified non-parametric tests are presented in the case of Type II censored samples for the following pair of competing hypotheses. The null hypothesis H_0 is the exponential distribution with the density function $f\left(t;\theta\right) = \frac{1}{\theta_1}\exp\left(-\frac{t}{\theta_1}\right)$ and the parameter $\theta_1 = 1$. The competing hypothesis H_1 is the Weibull distribution with parameters $\theta_1 = 1$, $\theta_2 = 1.2$. The powers of χ^2 tests in Table 18.6 are given in the case of optimal grouping.

Test	Degree of censoring					
	0%	10%	20%	30%	40%	50%
NRR	0.61 (6)	0.53 (5)	0.44 (5)	0.34 (5)	0.25 (3)	0.20 (3)
GPF	0.56 (4)	0.27 (5)	0.25 (5)	0.17 (5)	0.14 (5)	0.13 (5)
Kolmogorov	0.51	0.40	0.33	0.26	0.22	0.18
Cramer-von Mises-Smirnov	0.58	0.44	0.35	0.27	0.22	0.18
Anderson–Darling	0.41	0.34	0.28	0.23	0.20	0.17

Table 18.5. *Power of the tests for the Weibull-gamma pair in the case of Type II censored data*

Test	Degree of censoring					
	0%	10%	20%	30%	40%	50%
NRR	0.83 (3)	0.78 (3)	0.72 (3)	0.66 (3)	0.59 (3)	0.52 (3)
GPF	0.76 (3)	0.68 (5)	0.28 (4)	0.27 (4)	0.26 (4)	0.22 (4)
Kolmogorov	0.84	0.76	0.69	0.62	0.55	0.47
Cramer-von Mises-Smirnov	0.90	0.83	0.76	0.69	0.61	0.53
Anderson–Darling	0.91	0.85	0.79	0.72	0.65	0.56

Table 18.6. *Power of the tests for the exponential-Weibull pair in the case of Type II censored data*

As can be seen from Table 18.6, the NRR test here has a higher power than the GPF test. Nevertheless, the power of the NRR test turned out to be slightly less than the power of the Cramer–von Mises–Smirnov and Anderson–Darling tests.

18.5. Conclusions

The main advantage of the chi-square goodness-of-fit tests for censored data is that the limiting distribution of these statistics is the well-known χ^2-distribution. As has been shown in [LEM 09], in the case of testing simple hypotheses by complete samples (without censoring), the Pearson χ^2 test has a considerable advantage in power due to choosing optimal boundary points in the comparison with the non-parametric Kolmogorov, Cramer–von Mises–Smirnov and Anderson–Darling tests. As to the case of testing composite hypotheses, the non-parametric tests, as a rule, have higher power, as compared with the chi-square tests [LEM 10b]. The best properties of the chi-square tests, which enable them to compete with the Cramer–von Mises–Smirnov and Anderson–Darling tests, are shown up when using optimal grouping and choosing optimal number of intervals. As has been shown in this study, this result is also confirmed in the case of censored data.

It has been shown that in the case of a given pair of competing hypotheses, it is possible to increase essentially the power of the χ^2 tests due to the choice of optimal boundary points of grouping intervals. The power of both considered tests in the case of optimal grouping is much higher than in the case of equifrequent grouping and grouping into intervals with equal expected numbers of failures. The generalized χ^2 Pearson–Fisher test is considerably worse in terms of the power, than the NRR χ^2 test, when EFG and EENFG are used. But in the case of optimal grouping, it is impossible to give preference to one test. The power of the considered tests also depends on the distribution of censoring times: in the case when censored observations are "uniformly" distributed in the variational series, the power of the tests is higher as compared to when censored observations appear generally at the end of the variational series.

18.6. Acknowledgment

This work was supported by the Ministry of Education and Science of the Russian Federation within the state assignment (project 8.1274.2011) and the Federal target program "Research and Educational Personnel for Innovative Russia" (agreement # 14.B37.21.0860).

18.7. Bibliography

[AKR 88] AKRITAS M., "Pearson-type goodness-of-fit tests: the univariate case", *Journal of the American Statistical Association*, vol. 83, pp. 222–230, 1988.

[BAG 10] BAGDONAVIČIUS V., KRUOPIS J., NIKULIN M., *Nonparametric Tests for Censored Data*, John Wiley & Sons, Inc., New York, 2010.

[BAG 13] BAGDONAVIČIUS V., LEVULIENE R., NIKULIN M., "Chi-squared goodness-of-fit tests for parametric accelerated failure time models", *Communications in Statistics – Theory and Methods*, vol. 42, pp. 2768–2785, 2013.

[CHI 11a] CHIMITOVA E., NIKULIN M., LEMESHKO B., *et al.*, "Nonparametric goodness-of-fit tests for censored data", *Proceedings of the 7th International Conference on Mathematical Methods in Reliability: Theory, Methods and Applications*, Beijing, China, pp. 817–823, 2011.

[CHI 11b] CHIMITOVA E., TSIVINSKAYA A., "Simulation study for the NRR chi-square test of goodness-of-fit for censored data", *Proceedings of the International Workshop on Applied Methods of Statistical Analysis*, Novosibirsk, Russia, pp. 44–52, 2011.

[COH 91] COHEN A., *Truncated and Censored Samples: Theory and Applications*, Marcel Dekker, New York, 1991.

[DEN 79] DENISOV V., LEMESHKO B., "Optimal grouping in the analysis of experimental data", *Measuring Information Systems*, Novosibirsk, pp. 5–14, 1979. [In Russian]

[DEN 89] DENISOV V., LEMESHKO B., "Optimal grouping in estimation and tests of goodness-of-fit hypotheses", *Wissenschaftliche Schriftenreihe der Technishen universitat Karl-Marx-Stadt.*, vol. 10, pp. 63–81, 1989.

[DEN 04] DENISOV V., EGER K.-H., LEMESHKO B., *et al.*, *Design of Experiments and Statistical Analysis for Grouped Observations*, NSTU Publishing House, Novosibirisk, 2004.

[GRE 96] GREENWOOD P., NIKULIN M., *A Guide to Chi-Squared Testing*, John Wiley & Sons, New York, 1996.

[HAB 86] HABIB M., THOMAS D., "Chi-square goodness-of-fit tests for randomly censored data", *Annals of Statistics*, vol. 14, pp. 759–765, 1986.

[HJO 90] HJORT N., "Goodness of fit tests in models for life history data based on cumulative hazard rates", *The Annals of Statistics*, vol. 18, no. 3, pp. 1221–1258, 1990.

[HOL 92] HOLLANDER M., PEÑA E., "A chi-squared goodness-of-fit test for randomly censored data", *Journal of the American Statistical Association*, vol. 87, pp. 458–463, 1992.

[KIM 93] KIM J., "Chi-square goodness-of-fit tests for randomly censored data", *The Annals of Statistics*, vol. 21, no. 3, pp. 1621–1639, 1993.

[LEM 98] LEMESHKO B., "Asymptotically optimal grouping of observations in goodness-of-fit tests", *Industrial Laboratory*, vol. 64, no. 1, pp. 59–67, 1998 [In Russian].

[LEM 00] LEMESHKO B., CHIMITOVA E., "Maximization of the power of χ^2 tests", *Papers of Siberian Branch of Academy of Science of Higher Education*, vol. 2, pp. 53–61, 2000. [In Russian]

[LEM 01] LEMESHKO B., POSTOVALOV S., CHIMITOVA E., "On statistic distrbutions and the power of the Nikulin χ^2 test", *Industrial Laboratory*, vol. 67, no. 3, pp. 52–58, 2001. [in Russian]

[LEM 09] LEMESHKO B., LEMESHKO S., POSTOVALOV S., "Comparative analysis of the power of goodness-of-fit tests for near competing hypotheses. Part I. Verification of simple hypotheses", *Journal of Applied and Industrial Mathematics*, vol. 3, no. 4, pp. 462-475, 2009.

[LEM 10a] LEMESHKO B., CHIMITOVA E., PLESHKOVA T., "Testing simple and composite goodness-of-fit hypotheses by censored samples", *Nauchniy Vestnik NGTU*, vol. 41, no. 4, pp. 13–28, 2010. [In Russian]

[LEM 10b] LEMESHKO B., LEMESHKO S., POSTOVALOV S., "Comparative analysis of the power of goodness-of-fit tests for near competing hypotheses. Part II. Verification of complex hypotheses", *Journal of Applied and Industrial Mathematics*, vol. 4, no. 1, 2010.

[LEM 11] LEMESHKO B., LEMESHKO S., POSTOVALOV S., *et al.*, *Statistical Data Analysis, Simulation and Research of Probabilistic Regularities. Computer Approach*, NSTU Monographs, NSTU Publishing House, Novosibirsk, 2011. [In Russian]

[LEM 13] LEMESHKO B., CHIMITOVA E., VEDERNIKOVA M., "Modified Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling goodness-of-fit tests for randomly censored data", *Nauchniy Vestnik NSTU*, vol. 50, no. 1, pp. 3–16, 2013.

[LI 93] LI G., DOSS H., "Generalized Pearson-Fisher chi-square goodness-of-fit tests, with applications to models with life history data", *The Annals of Statistics*, vol. 21, pp. 772–797, 1993.

[NIK 73a] NIKULIN M., "Chi-square test for continuous distributions", *Theory of Probability and Its Applications*, vol. 18, pp. 638–639, 1973. [In Russian].

[NIK 73b] NIKULIN M., "Chi-square test for continuous distributions with shift and scale parameters", *Theory of Probability and Its Applications*, vol. 18, pp. 559–568, 1973. [In Russian]

[NIK 73c] NIKULIN M., "Chi-square test for normality", *Proceedings of the International Vilnius Conference on Probability Theory and Mathematical Statistics*, vol. 2, Vilnius, pp. 119–122, 1973. [in Russian].

[RAO 74] RAO K., ROBSON D., "A chi-squared statistic for goodness-of-fit tests within the exponential family", *Communications in Statistics*, vol. 3, pp. 1139–1153, 1974.

[VOI 06] VOINOV V., "On optimality of the Rao-Robson-Nikulin test", *Industrial Laboratory*, vol. 72, pp. 65–70, 2006. [In Russian]

[VOI 09] VOINOV V., PYA N., ALLOYAROVA R., "A comparative study of some modified chi-squared tests", *Communications in Statistics-Simulation and Computation*, vol. 38, pp. 355–367, 2009.

[VOI 13] VOINOV V., NIKULIN M., BALAKRISHNAN N., *Chi-Squared Goodness of Fit Tests with Applications*, Academic Press, 2013.