Записки научных
семинаров ПОМИ
Том 408, 2012 г.

V. Bagdonavičius, R. Levuliene, M. S. Nikulin, Q. X. Tran

ON CHI-SQUARED TYPE TESTS AND THEIR APPLICATIONS IN SURVIVAL ANALYSIS AND RELIABILITY

Abstract. The famous chi-square test of Pearson is well known, but different modifications of this test are not so well known. The theory of the chi-squared tests is developed very actively till now, especially in accelerated trials. We shall discuss here some applications of the theory of chi-squared tests in reliability and survival analysis for parametric regression models with time depending covariates when data are right censored.

1. Introduction

In complete data case well known modification of the classical chi-squared tests is the NRR statistics which 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, other – on the maximum likelihood estimators of unknown parameters of the tested model using initial non-grouped data (see Nikulin [22–24], Rao and Robson [26], Greenwood and Nikulin [11], Drost [10], LeCam *et al.* [18], van der Vaart [27], Voinov and Nikulin [28], etc).

Habib and Thomas [12], Hollander and Peña [14], Zhang [30] considered natural modifications of the NRR statistics 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 proposed by Akritas [2] and was developed by Hjort [13]. We develop this direction considering choice of random grouping intervals

Key words and phrases: chi-squared test, complete data, composite hypothesis, Fisher's information, goodness-of-fit, maximum likelihood estimator, parametric hypothesis, random cells, regression models, reliability, right censored data, simple sample, survival analysis.

as data functions and writing simple formulas useful for computing of test statistics for mostly applied classes of survival distributions. More about these problems one can see in Bagdonavičius, Kruopis and Nikulin [6].

We give chi-squared type goodness-of fit tests for general hypothesis H_0 with possibly time dependent covariates. Choice of random grouping intervals as data functions is considered.

2. Estimation and Pearson's chi-squared tests for complete data

In this section we discuss some applications of the chi-squared type tests in classical situation for testing the *hypothesis* H_0 according to which the distribution of independent identically distributed random variables $X_1, X_2, \ldots, X_n$, $X_i \in \mathbf{R}^1$, belongs to the *parametric* family

$$\left\{\mathbf{P}_{\boldsymbol{\theta}}, \boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)^T \in \Theta \subset \mathbf{R}^s\right\}, \tag{1}$$

Θ is an open in $\mathbf{R}^s$. We assume that for each $\boldsymbol{\theta} \in \Theta$ the measure $\mathbf{P}_{\boldsymbol{\theta}}$ is absolutely continuous with respect to certain σ-finite measure μ, given on Borelian σ-algebra $\mathcal{B}$. We denote

$$f(x, \boldsymbol{\theta}) = \frac{d\mathbf{P}_{\boldsymbol{\theta}}}{d\mu}(x),$$

the density of the probability distribution $\mathbf{P}_{\boldsymbol{\theta}}$ with respect to the measure μ (in the continuous case we assume that μ is the measure of Lebesgue on $\mathcal{B}$ and in the discrete case μ is the counting measure on $\{0,1,2,\ldots\}$). We denote

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x, \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}, \tag{2}$$

the likelihood function of the *sample* $\boldsymbol{X} = (X_1, \ldots, X_n)^T$, which is called also the *simple sample*. We note here that in general in classical statistics we worked only with simple samples. Concerning the family $\{f(x, \boldsymbol{\theta})\}$ we assume that for sufficiently large n ($n \to \infty$) the conditions of LeCam of the *local asymptotic normality* (L.A.N.) and *asymptotic differentiability* of the likelihood function $L_n(\boldsymbol{\theta})$ hold (see, for example, Ibragimov and Has'minskii [15], Greenwood and Nikulin [11], Lawless [17]).

Let $\dot{\ell}(\boldsymbol{\theta}) = \text{grad } \ln L_n(\boldsymbol{\theta})$ be the *vector-informant*, based on the simple sample $\boldsymbol{X} = (X_1, \ldots, X_n)^T$, and let $\mathbf{1}_s = (1, \ldots, 1)^T$ be the *unit vector* in

$\mathbf{R}^s$, and $\mathbf{0}_s$ be the *zero-vector* in $\mathbf{R}^s$. Let denote

$$\boldsymbol{i}(\boldsymbol{\theta}) = \frac{1}{n}\mathbf{E}_{\boldsymbol{\theta}}\dot{\ell}(\boldsymbol{\theta})\dot{\ell}^T(\boldsymbol{\theta}) \tag{3}$$

is the *information matrix of Fisher*, corresponding to the observation X_1. We suppose that $\mathbf{i}(\cdot)$ is continuous on Θ and $\det \mathbf{i}(\boldsymbol{\theta}) > 0$. On the other hand we denote

$$\mathbf{I}(\boldsymbol{\theta}) = n\mathbf{i}(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}}\dot{\ell}(\boldsymbol{\theta})\dot{\ell}^T(\boldsymbol{\theta}), \tag{4}$$

the *information matrix of Fisher*, corresponding to the complete simple sample $\boldsymbol{X} = (X_1, \ldots, X_n)^T$.

We note also that under the LeCam's regularity conditions there exist asymptotically normal $\sqrt{n}$-consistent sequences of maximum likelihood estimators $\{\widehat{\boldsymbol{\theta}}_n\}$:

$$l(\widehat{\boldsymbol{\theta}}_n) = \max_{\boldsymbol{\theta}\in\Theta} l(\boldsymbol{\theta}), \tag{5}$$

which satisfy the likelihood equation

$$\dot{\ell}(\boldsymbol{\theta}) = \mathrm{grad}\ \ln L_n(\boldsymbol{\theta}) = \mathbf{0}_s \tag{6}$$

and for which it holds the relation

$$\mathcal{L}(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})) \to N(\mathbf{0}_s, \mathbf{i}^{-1}(\boldsymbol{\theta})), \tag{7}$$

where $\to$ denotes convergence in distribution.

To construct the Pearson chi-squared test for H_0 we divide the real line by the points

$$-\infty = x_0 < x_1 < \ldots < x_{k-1} < x_k = +\infty$$

into k $(k > s+2)$ intervals $(x_{i-1}, x_i]$ of grouping the data. Following Fisher and Cramer we suppose that

I) $$p_i(\boldsymbol{\theta}) = \mathbf{P}\{X_1 \in (x_{i-1}, x_i]\} > c > 0,\ i = 1, \ldots, k;$$
$$p_1(\boldsymbol{\theta}) + p_2(\boldsymbol{\theta}) + \ldots + p_k(\boldsymbol{\theta}) = 1,\ \boldsymbol{\theta} \in \Theta;$$

II) $\dfrac{\partial^2 p_i(\boldsymbol{\theta})}{\partial\theta_u \partial\theta_v}$ are continuous functions on Θ;

III) the matrix

$$\mathbf{B} = \left\| \frac{1}{\sqrt{p_i(\boldsymbol{\theta})}} \frac{\partial p_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \right\|_{k\times s} \tag{8}$$

has rank s. Further, let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)^T$ be the result of grouping the random variables $X_1, \ldots, X_n$ into the intervals $(x_0, x_1], \ldots, (x_{k-1}, x_k)$. As is well-known (see, for example, Cramer [9], Greenwood and Nikulin [11], Voinov *et al.* [29]), in the case when H_0 is true and the true value of the parameter $\boldsymbol{\theta}$ is known, *the standard Pearson statistics*

$$X_n^2(\boldsymbol{\theta}) = \sum_{i=1}^{k} \frac{[\nu_i - np_i(\boldsymbol{\theta})]^2}{np_i(\boldsymbol{\theta})} \tag{9}$$

has in the limit as $n \to \infty$ *chi-square distribution with* $(k-1)$ *degrees of freedom.*

On this fact is constructed the well-known chi-square criterion of Pearson, based on the statistics $X_n^2(\boldsymbol{\theta})$. According to this test

the simple hypothesis $H_0 : X_i \sim f(x, \boldsymbol{\theta})$ *is rejected if* $X_n^2 > c_\alpha$,

where $c_\alpha = \chi^2_{k-1,\alpha}$ is the α-upper quantile of the chi-square distribution with $(k-1)$ degrees of freedom.

If $\boldsymbol{\theta}$ is unknown, in this situation we need to estimate it using the data. Hence the limit distribution of the statistics $X_n^2(\boldsymbol{\theta}_n^*)$ depends on the asymptotical properties of estimator $\boldsymbol{\theta}_n^*$, which one puts in (9) instead of the unknown parameter $\boldsymbol{\theta}$.

Example 1. Minimum chi-square estimator of Fisher. Let consider the sequence $\{\widetilde{\boldsymbol{\theta}}_n\}$ of estimators $\widetilde{\boldsymbol{\theta}}_n$, which satisfy the condition

$$\widetilde{\boldsymbol{\theta}}_n = \operatorname*{argmin}_{\boldsymbol{\theta} \in \Theta} X_n^2(\boldsymbol{\theta}) \tag{10}$$

i.e.

$$X_n^2(\widetilde{\boldsymbol{\theta}}_n) = \min_{\boldsymbol{\theta} \in \Theta} X_n^2(\boldsymbol{\theta}), \tag{11}$$

obtained by the so called *minimum chi-square method.*

It is well-known that the sequences $\{\widetilde{\boldsymbol{\theta}}_n\}$ of estimators $\widetilde{\boldsymbol{\theta}}_n$ satisfy the condition

$$\mathcal{L}(\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})) \to N(\mathbf{0}_s, \mathbf{J}^{-1}), \tag{12}$$

where

$$n\mathbf{J} = n\mathbf{J}(\boldsymbol{\theta}) = n\mathbf{B}^T\mathbf{B}$$

is the *information matrix of Fisher* of the statistics $\boldsymbol{\nu} = (\nu_1, \dots, \nu_k)^T$.

Finally, it is easy to show that the *theorem of Fisher* holds

$$\lim_{n\to\infty} \mathbf{P}\left\{X_n^2(\widetilde{\boldsymbol{\theta}}_n) < x \mid H_0\right\} = \mathbf{P}\left\{\chi^2_{k-s-1} < x\right\}. \tag{13}$$

As is known (see, for example, Greenwood and Nikulin [11]) the estimator of Fisher $\widetilde{\boldsymbol{\theta}}_n$, obtained by the *minimum chi-square method*,

$$X_n^2(\widetilde{\boldsymbol{\theta}}_n) = \min_{\boldsymbol{\theta} \in \Theta} X_n^2(\boldsymbol{\theta}), \tag{14}$$

verifies the property (12).

Using these results Cramer considered the so called *multinomial maximum likelihood estimator* $\widetilde{\boldsymbol{\theta}}_n$ – the point of maximum of the likelihood function $L_n(\boldsymbol{\theta})$ of the statistics $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)^T$:

$$L_n(\widetilde{\boldsymbol{\theta}}_n) = \max_{\boldsymbol{\theta}\in\Theta} L_n(\boldsymbol{\theta}), \quad L_n(\boldsymbol{\theta}) = p_1^{\nu_1}(\boldsymbol{\theta}) p_2^{\nu_2}(\boldsymbol{\theta}) \cdots p_k^{\nu_k}(\boldsymbol{\theta}), \qquad (15)$$

also satisfies the relation (12). One can say that $\widetilde{\boldsymbol{\theta}}_n$ is the estimator for $\boldsymbol{\theta}$, based on censored data $(\nu_1, \ldots, \nu_k)^T$. The statistics $(\nu_1, \ldots, \nu_k)^T$ gives the trivial example of non-complete data, when we don't have the vector of observations $\boldsymbol{X} = (X_1, \ldots, X_n)^T$.

We note here that if $X_1, \ldots, X_n$ follow a continuous distribution $f(x, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$, then in this case the statistics $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)^T$ is *not sufficient* as the vector of observations $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ is itself, and hence in this case the matrix

$$n\mathbf{i}(\boldsymbol{\theta}) - n\mathbf{J}(\boldsymbol{\theta}) \qquad (16)$$

is positive definite, where $n\mathbf{i}(\boldsymbol{\theta})$ is the information matrix of the simple sample $\boldsymbol{X} = (X_1, \ldots, X_n)^T$.

Example 2. Maximum likelihood estimator. Let consider $\sqrt{n}$-consistent sequences of maximum likelihood estimators $\{\widehat{\boldsymbol{\theta}}_n\}$, i.e.

$$\mathcal{L}(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})) \to N(\mathbf{0}_s, \mathbf{i}^{-1}), \qquad n \to \infty.$$

As was shown by Chernoff and Lehmann [8], (see also LeCam *et al.* [18]),

$$\lim_{n\to\infty} \mathbf{P}\left\{X_n^2(\widehat{\boldsymbol{\theta}}_n) < x \mid H_0\right\} = \mathbf{P}\left\{\chi^2_{k-s-1} + \lambda_1\xi_1^2 + \cdots + \lambda_s\xi_s^2 < x\right\}, \ (17)$$

where $\chi^2_{k-s-1}, \xi_1^2, \ldots, \xi_s^2$ are independent random variables, $\xi_i \sim N(0,1)$, and in general $\lambda_i = \lambda_i(\boldsymbol{\theta})$, $0 < \lambda_i < 1$, and λ_i are the roots of the equation

$$\mid (1-\lambda)\mathbf{i}(\boldsymbol{\theta}) - \mathbf{J}(\boldsymbol{\theta}) \mid = 0.$$

From this theorem it follows that in practice it is not easy to apply the best methods of estimation of unknown parameter for construction the Pearson chi-squared test. In general the limit distribution of the test statistics is rather very complicated! Only in 3 cases the Pearson statistics X_n^2 has in the limit chi-squared distribution with $k-1$ (in the case of Pearson), and $k-s-1$ degrees of freedom (in the cases of Fisher and Cramer).

These difficulties we have in very simples cases of parametric models. We want to test the statistical models used in survival analysis and reliability, when data are censored and we take into account the influence of covariates. It is evident that we have to modify the statistics of Pearson to do it more adaptive for solving our problems.

3. Statistics Y_n^2

Evidently if we want to use the maximum likelihood estimators for testing H_0 it is possible to construct the so called NRR statistics Y_n^2 (see, van der Vaart [27], Drost [10], Greenwood and Nikulin [11], Nikulin [22–24], Rao and Robson [26], Moore and Spruill [20], Lemeshko *et al.* [19], etc...), which is the sum of the standard statistics of Pearson and non-negative quadratic form Q:

$$Y_n^2 = Y_n^2(\widehat{\boldsymbol{\theta}}_n) = \mathbf{X}_n^*(\widehat{\boldsymbol{\theta}}_n)\mathbf{G}^-(\widehat{\boldsymbol{\theta}}_n)\mathbf{X}_n(\widehat{\boldsymbol{\theta}}_n) = X_n^2(\widehat{\boldsymbol{\theta}}_n) + Q, \qquad (18)$$

where the matrix $\mathbf{G}^-$ is a *generalized inverse* for the asymptotic matrix of covariance $\mathbf{G}$ of the statistics $\mathbf{X}_n(\widehat{\boldsymbol{\theta}}_n)$, where

$$\mathbf{X}_n(\boldsymbol{\theta}) = \left(\frac{\nu_1 - np_1(\boldsymbol{\theta})}{\sqrt{np_1(\boldsymbol{\theta})}}, \dots, \frac{\nu_k - np_k(\boldsymbol{\theta})}{\sqrt{np_k(\boldsymbol{\theta})}}\right)^T.$$

We note that the quadratic form Y_n^2 is *invariant* under choice of $\mathbf{G}^-$ in virtue of the specific Cramer's condition I) of the singularity of the matrix $\mathbf{G}$. As follows from the lemma of Chernoff-Lehmann

$$\lim_{n\to\infty} \mathbf{P}\left\{Y_n^2(\widehat{\boldsymbol{\theta}}_n) < x \mid H_0\right\} = \mathbf{P}\left\{\chi^2_{k-1} < x\right\}.$$

Some examples of the applications of statistics Y_n^2 can be found in the publications of Bagdonavičius *et al.* [6], Greenwood and Nikulin [11], Voinov *et al.* [29], Lemeshko *et al.* [19], Drost [10], etc.

The statistics $Y_n^2(\widehat{\boldsymbol{\theta}}_n)$ has a particularly convenient form when we construct a chi-square test with random cell boundaries for continuous distributions, which are used in survival analysis and reliability. It is evident that the one can construct the similar test based on the statistics Y_n^2 for testing H_0 if we use any square-root consistent estimator, by choosing the corresponding matrix of covariance G and its generalized inverse. For example, one can apply the method of moments estimator, etc. Such approach is very important in survival analysis and reliability, when often we have no the complete data, especially when data are right or/and left censored. We shall consider the situation with right censoring data.

4. Accelerated life models in reliability and survival analysis

Suppose that n independent failure time variables are observed. Let us consider the hypothesis H_0 stating that the survival function given the vector of explanatory variables (covariates),

$$z(t) = (z_0(t), z_1(t), \dots, z_m(t))^T, \quad z_0(t) \equiv 1,$$

has the form

$$S(t|z) = S_0(t; \boldsymbol{\theta}, z), \tag{1}$$

where S_0 is a specified functional of time t, finite-dimensional parameter $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_s)^T \in \boldsymbol{\Theta} \subset R^{s+1}$ and z. The hypothesis H_0 can be also formulated in terms of the hazard functions $\lambda(t|z) = -S'(t|z)/S(t|z)$ or the cumulative hazard functions $\Lambda(t|z) = -\ln S(t|z)$.

Let us consider examples of such hypotheses (see Aalen [1], Andersen *et al.* [3], Aven and Jensen [4], Bagdonavičius and Nikulin [5], Bagdonavičius *et al.* [6], Klein and Moeschberger [16], Billingsley [7], LeCam *et al.* [18], Nelson [21]):

1) Parametric accelerated failure time (AFT) model: $\boldsymbol{\theta} = (\beta^T, \gamma^T)^T$,

$$S(t|z) = S_0\Big(\int_0^t e^{-\beta^T z(u)} du; \gamma\Big), \tag{2}$$

where $\beta = (\beta_0, \dots, \beta_m)^T$ is a vector of unknown regression parameters, the function S_0 does not depend on z_i and belongs to a specified class of survival functions: $S_0(t, \gamma)$, $\gamma = (\gamma_1, \dots, \gamma_q)^T \in G \subset R^q$. If explanatory variables are constant over time then the parametric AFT model has the form $S(t|z) = S_0\left(e^{-\beta^T z} t; \gamma\right)$, and the logarithm of the failure time T under z may be written as

$$\ln\{T\} = \beta^T z + \varepsilon, \quad \varepsilon \sim S(t) = S_0(\ln t).$$

If ε is normally distributed random variable then the AFT model is standard multiple linear regression model.

Sometimes some specified functions of z instead of z may be used. In accelerated life testing transforms $\ln z$ (power rule model), $1/z$ (Arrhenius model), $\ln \frac{z}{1-z}$ (Meeker–Luvalle model), and others are used.

2) Parametric proportional hazards (PH or Cox) model: $\boldsymbol{\theta} = (\beta^T, \gamma^T)^T$,

$$\lambda(t|z) = e^{\beta^T z(t)} \lambda_0(t, \gamma), \tag{3}$$

where λ_0 belongs to a specified class of hazard functions $\lambda(t, \gamma)$, $\gamma \in G \subset R^q$.

3) Parametric generalized proportional hazards (GPH) models (including parametric frailty and linear transformations models):
$\boldsymbol{\theta} = (\beta^T, \gamma^T, \nu^T)^T$,

$$h(\Lambda(t|z), \nu) = \int_0^t e^{\beta^T z(u)} \lambda_0(u, \gamma)\, du, \tag{4}$$

where the function $h(x, \nu)$ and the hazard function $\lambda_0(t, \gamma)$ belong to a specified parametric classes. In particular, if

$$h(x, \nu) = \frac{(1+x)^\nu - 1}{\nu}, \quad h(x, \nu) = \frac{1 - e^{-\nu x}}{\nu} \quad \text{or} \quad h(x, \nu) = x + \frac{\nu x^2}{2},$$

we have respectively parametric positive stable, gamma, and inverse Gaussian frailty models with explanatory variables.

4) Models with cross effects of survival functions:

$$\lambda(t|z) = g(z, \beta, \nu, \Lambda_0(t, \gamma)), \tag{5}$$

where the cumulative hazard Λ_0 has specified form and the function g has one of the following forms:

$$e^{\beta^T z} \left[1 + e^{(\beta+\nu)^T z} x\right]^{e^{-\nu^T z} - 1}; \quad \frac{e^{\beta^T z + x e^{\nu^T z}}}{1 + e^{(\beta+\nu)^T z}[e^{x e^{\nu^T z}} - 1]}. \tag{6}$$

5. Right censored data and time depending covariates

We shall give chi-squared tests for the hypothesis H_0 from right censored failure time regression data:

$$(X_1, \delta_1, z_1(s), 0 \leqslant s \leqslant X_1); \ldots; (X_n, \delta_n, z_n(s), 0 \leqslant s \leqslant X_n), \tag{1}$$

where:

$$X_i = T_i \wedge C_i, \quad \delta_i = \mathbf{1}_{\{T_i \leqslant C_i\}}, \quad i = 1, 2, \ldots, n,$$

T_i is being failure times; C_i is being censored times and

$$z_i(t) = (z_{i0}(t), z_{i1}(t), \ldots, z_{im}(t))^T, \quad z_{i0}(t) \equiv 1.$$

the possibly time depending covariates. the random variable δ_i is the indicator of the event $\{T_i \leqslant C_i\}$, and τ is the *finite time of the experiment.*

Denote by $\overline{G}_i$ the survival function of the censoring time C_i and $g_i(t)$ is the density of $\overline{G}_i$. Set:

$$N_i(t) = \mathbf{1}_{\{X_i \leqslant t\,, \delta_i = 1\}} = \begin{cases} 1, & \text{if} \quad X_i \leqslant t \quad \text{and} \quad \delta_i = 1, \\ 0, & \text{if} \quad X_i > t. \end{cases}$$

$N_i(t)$ is the number of failures of i-th item,

$$Y_i(t) = \mathbf{1}_{\{X_i \geqslant t\}} = \begin{cases} 1, & \text{if} \quad X_i \geqslant t, \\ 0, & \text{if} \quad X_i < t. \end{cases}$$

It is also clear that our data can be easily represented in terms of processes N_i and Y_i under the form following

$$(N_1(t),\, Y_1(t),\, t \geqslant 0), \ldots, (N_n(t),\, Y_n(t),\, t \geqslant 0)$$

and vice versa. Set

$$N(t) = \sum_{i=1}^{n} N_i(t), \quad Y(t) = \sum_{i=1}^{n} Y_i(t). \tag{2}$$

We note also that the representation

$$\{N_i(s),\, Y_i(s),\, 0 \leqslant s \leqslant t,\, i = 1, 2, \ldots, n\},$$

gives us the dynamics of the stories of failures and censoring up to time t. The very concept history is well formalized in terms of the concept Filtration of a random process (Bagdonavičius and Nikulin [5], Lawless [17], Andersen *et al.* [3], Klein and Moeschberger [16]).

Suppose that
1) The processes N_i, Y_i, z_i are observed up to the finite time τ;
2) Survival distributions of all n objects, given z_i, are absolutely continuous with the survival functions $S_i(t, \boldsymbol{\theta}) = S(t, \boldsymbol{\theta}, z_i)$ and the hazard rates $\lambda_i(t, \boldsymbol{\theta}) = \lambda(t, \boldsymbol{\theta}, z_i)$ of T_i under z_i;
3) Censoring is non informative and the multiplicative intensity model holds: the compensators of the counting processes N_i with respect to the history of the observed processes are $\int_0^t Y_i(u)\lambda_i(u, \boldsymbol{\theta})\, du$.

In this case for non-informative and independent censoring we may give the following expressions for the likelihood function:

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_i^{\delta_i}(X_i, \boldsymbol{\theta}) S_i^{1-\delta_i}(X_i, \boldsymbol{\theta}) . \overline{G}_i^{\delta_i}(X_i) g_i^{1-\delta_i}(X_i)\,; \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Since the problem is to estimate the parameter $\boldsymbol{\theta}$, we can skip the multipliers which do not depend on this parameter. So under non-informative censoring the likelihood function has the next form:

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_i^{\delta_i}(X_i, \boldsymbol{\theta}) S_i^{1-\delta_i}(X_i, \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Using the relation $f_i(t, \boldsymbol{\theta}) = \lambda_i(t, \boldsymbol{\theta}) S_i(t, \boldsymbol{\theta})$ the likelihood function can be written

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} \lambda_i^{\delta_i}(X_i, \boldsymbol{\theta}) S_i(X_i, \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

The log-likelihood functions are the form:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \{\delta_i \ln \lambda_i(X_i, \boldsymbol{\theta}) + \ln S_i(X_i, \boldsymbol{\theta})\}, \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

The log-likelihood function is maximized at the same point as the likelihood function. As before the estimator $\widehat{\boldsymbol{\theta}}$, maximizing the likelihood function $L_n(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, is called maximum likelihood estimator. We denote $\widehat{\boldsymbol{\theta}}$ the maximum likelihood estimator of $\boldsymbol{\theta}$ under H_0. We remind that $\boldsymbol{\theta} = (\beta^T, \gamma^T)^T$. If $\lambda_i(u, \boldsymbol{\theta})$ is sufficiently smooth function of the parameter $\boldsymbol{\theta}$ then the MLE $\widehat{\boldsymbol{\theta}}$ satisfies the equation:

$$\dot{l}(\widehat{\boldsymbol{\theta}}) = 0,$$

where $\dot{l}(\boldsymbol{\theta})$ is the score vector

$$\dot{l}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}) = \left(\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_0}, \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_1}, \cdots, \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_s} \right)^T.$$

As before, the Fisher information matrix is $\boldsymbol{I}(\boldsymbol{\theta}) = -\mathbf{E}_{\boldsymbol{\theta}} \ddot{l}(\boldsymbol{\theta})$, where

$$\ddot{l}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln \lambda_i(X_i, \boldsymbol{\theta}) - \sum_{i=1}^{n} \delta_i \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \Lambda_i(X_i, \boldsymbol{\theta}).$$

With the sample (1), the parametric loglikelihood function is

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^{\infty} \{\ln \lambda_i(u, \boldsymbol{\theta}) dN_i(u) - Y_i(u) \lambda(u, \boldsymbol{\theta})\} \, du.$$

and the score function is

$$\dot{l}(\boldsymbol{\theta}) = \int_0^\infty \frac{\partial}{\partial \boldsymbol{\theta}} \ln \lambda_i(u, \boldsymbol{\theta})\, dM_i(u, \boldsymbol{\theta}),$$

where

$$M_i(t, \boldsymbol{\theta}) = N_i(t) - \int_0^t Y_i(u) \lambda_i(u, \boldsymbol{\theta})\, du$$

is the zero mean martingale with respect to the filtration generated by the data. The Fisher's information matrix is given by formula

$$\boldsymbol{I}(\boldsymbol{\theta}) = -\mathbf{E}_{\boldsymbol{\theta}} \ddot{l}(\boldsymbol{\theta}) =$$

$$\mathbf{E}_{\boldsymbol{\theta}} \sum_{i=1}^n \int_0^\infty \left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln \lambda_i(u, \boldsymbol{\theta}) \right) \left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln \lambda_i(u, \boldsymbol{\theta}) \right)^T \lambda_i(u, \boldsymbol{\theta}) Y_i(u)\, du.$$

We suppose that there exists the matrix

$$\boldsymbol{i}(\boldsymbol{\theta}_0) = \lim_{n \to \infty} \boldsymbol{I}(\boldsymbol{\theta}_0)/n,$$

where $\boldsymbol{\theta}_0$ is the true value of parameter $\boldsymbol{\theta}$.

6. Chi-squared type tests construction

Divide the interval $[0, \tau]$ into k smaller intervals $I_j = (a_{j-1}, a_j]$, $a_0 = 0$, $a_k = \tau$, and denote by:

$$U_j = N(a_j) - N(a_{j-1}), \tag{1}$$

the number of observed failures in the j-th interval, $j = 1, 2, \dots, k$.
Under regularity conditions the equality

$$\mathbf{E} N_i(t) = \mathbf{E} \int_0^t \lambda_i(u, \boldsymbol{\theta}) Y_i(u)\, du,$$

holds, where $\boldsymbol{\theta} = (\beta^T, \gamma^T)^T$ and $\lambda_i(t, \boldsymbol{\theta}) = \lambda(t, z_i, \boldsymbol{\theta})$ is the hazard function of T_i under z_i. It suggests that we can expect to observe

$$e_j = \sum_{i=1}^n \int_{a_{j-1}}^{a_j} \lambda_i(u, \widehat{\boldsymbol{\theta}}) Y_i(u) du, \tag{2}$$

failures in the interval I_j, here $\widehat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$ under H_0.

Let us consider the stochastic process

$$H_n(t) = \frac{1}{\sqrt{n}}(N(t) - \sum_{i=1}^{n} \int_0^t \lambda_i(u, \widehat{\boldsymbol{\theta}}) Y_i(u) du),$$

which characterizes the difference between observed and expected numbers of failures.

It looks very reasonable to construct the test for H_0, based on the vector $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_k)^T$, where

$$Z_j = H_n(a_j) - H_n(a_{j-1}) = \frac{1}{\sqrt{n}}(U_j - e_j),\ j = 1, \ldots, k. \tag{3}$$

7 Estimation and chi-squared type tests

To investigate the properties of the statistics Z we need properties of the stochastic process $H_n(t), t \geqslant 0$. To obtain these properties we use the properties of the ML estimators which are well known. We present here these properties:

Conditions A:

(1) $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$;

(2) $\frac{1}{\sqrt{n}} \dot{\ell}(\boldsymbol{\theta}_0) \xrightarrow{d} N_m(0, \boldsymbol{i}(\boldsymbol{\theta}_0))$;

(3) $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \boldsymbol{i}^{-1}(\boldsymbol{\theta}_0) . \frac{1}{\sqrt{n}} \dot{\ell}(\boldsymbol{\theta}_0) + o_P(1)$;

(4) $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_m(0, \boldsymbol{i}^{-1}(\boldsymbol{\theta}_0))$;

(5) $\frac{-1}{n} \ddot{\ell}(\boldsymbol{\theta}_0) \xrightarrow{P} \boldsymbol{i}(\boldsymbol{\theta}_0)$; $\quad \frac{-1}{n} \ddot{\ell}(\widehat{\boldsymbol{\theta}}) \xrightarrow{P} \boldsymbol{i}(\boldsymbol{\theta}_0)$,

where $\boldsymbol{\theta}_0$ is the true value of the parameter $\boldsymbol{\theta}$, and

$$\dot{l}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial}{\partial \boldsymbol{\theta}} \ln\left(\lambda_i(u, \boldsymbol{\theta})\right) \{dN_i(u) - Y_i(u) \lambda_i(u, \boldsymbol{\theta})\, du\},$$

is the score function.

These conditions mean consistency and asymptotic normality of the ML estimator $\widehat{\boldsymbol{\theta}}$. We suppose also that the Conditions VI.1.1 given in Andersen *et al.* ([3]) hold.

Set

$$S^{(0)}(t, \boldsymbol{\theta}) = \sum_{i=1}^{n} Y_i(t) \lambda_i(t, \boldsymbol{\theta}), \quad S^{(1)}(t, \boldsymbol{\theta}) = \sum_{i=1}^{n} Y_i(t) \frac{\partial \ln \lambda_i(t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \lambda_i(t, \boldsymbol{\theta}),$$

$$S^{(2)}(t,\boldsymbol{\theta}) = \sum_{i=1}^{n} Y_i(t) \frac{\partial^2 \ln \lambda_i(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \lambda_i(t,\boldsymbol{\theta}).$$

Conditions B: There exist a neighborhood $\boldsymbol{\Theta}_0$ of $\boldsymbol{\theta}_0$ and continuous bounded on $\boldsymbol{\Theta}_0 \times [0,\tau]$ functions

$$s^{(0)}(t,\boldsymbol{\theta}), \quad s^{(1)}(t,\boldsymbol{\theta}) = \frac{\partial s^{(0)}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad s^{(2)}(t,\boldsymbol{\theta}) = \frac{\partial^2 s^{(0)}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2},$$

such that for $j = 0,1,2$

$$\sup_{t\in[0,\tau],\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} ||\frac{1}{n} S^{(j)}(t,\boldsymbol{\theta}) - s^{(j)}(t,\boldsymbol{\theta})|| \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty.$$

The conditions B imply that uniformly for $t \in [0,\tau]$

$$\frac{1}{n}\sum_{i=1}^{n} \int_0^t \lambda_i(u,\boldsymbol{\theta}_0) Y_i(u)\, du \xrightarrow{P} A(t), \quad \frac{1}{n}\sum_{i=1}^{n} \int_0^t \frac{\partial}{\partial \boldsymbol{\theta}} \lambda_i(u,\boldsymbol{\theta}_0) Y_i(u)\, du \xrightarrow{P} C(t),$$

where A and C are finite functions.

Lemma. *Under the conditions* A *and* B *the following convergence holds:*

$$H_n \xrightarrow{d} V \quad \text{on} \quad D[0,\tau];$$

here $D[0,\tau]$ is space of cadlag functions with Skorokhod metric, V is zero mean Gaussian martingale such that, for all $0 \leqslant u \leqslant v \leqslant \tau$

$$\operatorname{cov}(V(u),V(v)) = A(u) - C^T(u)\mathbf{i}^{-1}(\boldsymbol{\theta}_0) C(v).$$

Proof. By conditions A

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathbf{i}^{-1}(\boldsymbol{\theta}_0)\, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^\tau \frac{\partial}{\partial \boldsymbol{\theta}} \ln \lambda_i(u,\boldsymbol{\theta}_0)\, dM_i(u) + o_P(1),$$

$$dM_i(u) = dN_i - Y_i(u)\lambda_i(u,\boldsymbol{\theta}_0)\, du.$$

Using Conditions B and the Taylor formula we have:

$$H_n(t) = \frac{1}{\sqrt{n}} M(t) - \sqrt{n} \int_0^t [s^{(0)}(u,\widehat{\boldsymbol{\theta}}) - s^{(0)}(u,\boldsymbol{\theta}_0)]\, du + o_P(1)$$

$$= \frac{1}{\sqrt{n}} M(t) - \int_0^t [s^{(1)}(u,\boldsymbol{\theta}_0)]^T du \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_P(1)$$

$$= \frac{1}{\sqrt{n}} M(t) - C^T(t)\, \mathbf{i}^{-1}(\boldsymbol{\theta}_0) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial}{\partial \boldsymbol{\theta}} \ln \lambda_i(u, \boldsymbol{\theta}_0)\, dM_i(u) + o_P(1)$$

$$=: M_1^*(t) - C^T(t)\, \mathbf{i}^{-1}(\boldsymbol{\theta}_0) M_2^*(\tau).$$

The predictable variations and covariations are

$$\langle M_1^* \rangle(t) = \frac{1}{n} \int_0^t S^{(0)}(u, \boldsymbol{\theta})\, du \xrightarrow{P} A(t),$$

$$\langle M_1^*, M_2^* \rangle(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \dot{\lambda}_i(u, \boldsymbol{\theta}_0) Y_i(u)\, du \xrightarrow{P} C(t),$$

$$\langle M_2^* \rangle(\tau)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial}{\partial \boldsymbol{\theta}} \ln \lambda_i(u, \boldsymbol{\theta}) \left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln \lambda_i(u, \boldsymbol{\theta}) \right)^T \lambda_i(u, \boldsymbol{\theta}_0) Y_i(u) du \xrightarrow{P} \mathbf{i}(\boldsymbol{\theta}_0).$$

The result of the lemma is implied by the CLT for martingales because Conditions A and B imply that the Lindeberg condition (see Andersen *et al.* [3]) is obviously satisfied: for all $t \in [0, \tau]$

$$\left| \left\langle \frac{1}{\sqrt{n}} \int_0^t \mathbf{1}_{\{\frac{1}{\sqrt{n}} > \varepsilon\}}\, dM(u) \right\rangle \right| = \frac{1}{n} \mathbf{1}_{\{\frac{1}{\sqrt{n}} > \varepsilon\}} \sum_{i=1}^{n} \int_0^t \lambda_i(u, \boldsymbol{\theta}_0) Y_i(u)\, du \to 0,$$

$$\left| \left\langle \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial}{\partial \theta_j} \ln \lambda_i(u, \boldsymbol{\theta}_0) \mathbf{1}_{\{\frac{\partial}{\partial \theta_j} \ln \lambda_i(u, \boldsymbol{\theta}_0) > \varepsilon \sqrt{n}\}}\, dM_i(u) \right\rangle \right|$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} \left\{ \frac{\partial}{\partial \theta_j} \ln \lambda_i(u, \boldsymbol{\theta}_0) \right\}^2 \mathbf{1}_{\{\frac{\partial}{\partial \theta_j} \ln \lambda_i(u, \boldsymbol{\theta}_0) > \varepsilon \sqrt{n}\}} \lambda_i(u, \boldsymbol{\theta}_0) Y_i(u)\, du \to 0.$$

The proof is complete. □

Let us consider now

$$V_j = V(a_j) - V(a_{j-1}), \quad \sigma_{jj'} = cov(V_j, V_{j'}), \quad A_j = A(a_j) - A(a_{j-1}),$$

$$C_{ij} = C_i(a_j) - C_i(a_{j-1}), \quad \mathcal{C}_j = (C_{0j}, \ldots, C_{sj})^T, \Sigma = [\sigma_{jj'}]_{k \times k},$$

$$\mathcal{C} = [C_{ij}]_{(s+1) \times k},$$

for $i = 0, \dots, s$; $j, j' = 1, \dots, k$, and denote by $\mathcal{A}$ a $k \times k$ diagonal matrix with the diagonal elements $A_1, \dots, A_k$.

From Lemma one can obtain Theorem 1.

Theorem 1. *Under conditions* A *and* B

$$Z \xrightarrow{d} Y \sim N_k(0, \Sigma) \quad as \quad n \to \infty,$$

where

$$\Sigma = \mathcal{A} - \mathcal{C}^T \mathbf{i}^{-1}(\boldsymbol{\theta}_0)\mathcal{C}.$$

Remark. Set $G = \mathbf{i} - CA^{-1}C^T$. The formula

$$\Sigma^- = \mathcal{A}^{-1} + \mathcal{A}^{-1}\mathcal{C}^T G^- \mathcal{C}\mathcal{A}^{-1}$$

implies that we need to invert only diagonal $k \times k$ matrix A and find the general inverse of the $s \times s$ matrix G.

From Theorem 1 it follows that under conditions A and B the following estimators of A_j, $\mathcal{C}_j$, Σ and $\mathbf{i}(\boldsymbol{\theta}_0)$ are consistent:

$$\widehat{A}_j = U_j/n, \quad \widehat{\mathcal{C}}_j = \frac{1}{n}\sum_{i=1}^{n} \int_{I_j} \frac{\partial}{\partial \boldsymbol{\theta}} \ln \lambda_i(u, \widehat{\boldsymbol{\theta}}) dN_i(u), \quad \widehat{\Sigma} = \widehat{\mathcal{A}} - \widehat{\mathcal{C}}^T \widehat{\mathbf{i}}^{-1} \widehat{\mathcal{C}};$$

and

$$\widehat{\mathbf{i}} = \frac{1}{n}\sum_{i=1}^{n} \int_0^{\tau} \frac{\partial \ln \lambda_i(u, \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \left(\frac{\partial \ln \lambda_i(u, \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \right)^T dN_i(u).$$

8. Test statistics

The theorems 1 implies that a test for the hypothesis H_0 can be based on the statistics $Y^2 = Z^T \widehat{\Sigma}^- Z$, where

$$\widehat{\Sigma}^- = \widehat{\mathcal{A}}^{-1} + \widehat{\mathcal{C}}^{-1} \widehat{\mathcal{A}}^T \widehat{G}^- \widehat{\mathcal{C}} \widehat{\mathcal{A}}^{-1}, \quad \widehat{G} = \widehat{\mathbf{i}} - \widehat{\mathcal{C}} \widehat{\mathcal{A}}^{-1} \widehat{\mathcal{C}}^T.$$

This statistics can be written in the form

$$Y^2 = \sum_{j=1}^{k} \frac{(U_j - e_j)^2}{U_j} + Q,$$

where

$$U_j = \sum_{i: X_i \in I_j} \delta_i, \quad e_j = \sum_{i=1}^{n} \int_{I_j} \lambda_i(u, \widehat{\boldsymbol{\theta}}) Y_i(u) du, \quad Q = W^T \widehat{G}^- W,$$

$$W = (W_0, \dots, W_s)^T, \quad \widehat{G} = [\widehat{g}_{ll'}]_{s\times s}, \quad \widehat{g}_{ll'} = \widehat{i}_{ll'} - \sum_{j=1}^{k} \widehat{C}_{lj}\widehat{C}_{l'j}\widehat{A}_j^{-1},$$

$$\widehat{C}_{lj} = \frac{1}{n}\sum_{i:X_i\in I_j} \delta_i \frac{\partial}{\partial\theta_l} \ln\lambda_i(X_i, \widehat{\boldsymbol{\theta}}),$$

$$\widehat{i}_{ll'} = \frac{1}{n}\sum_{i=1}^{n} \delta_i \frac{\partial \ln\lambda_i(X_i, \widehat{\boldsymbol{\theta}})}{\partial\theta_l} \frac{\partial \ln\lambda_i(X_i, \widehat{\boldsymbol{\theta}})}{\partial\theta_{l'}},$$

$$W_l = \sum_{j=1}^{k} \widehat{C}_{lj}\widehat{A}_j^{-1} Z_j, \quad l, l' = 0, \dots, s.$$

The limit distribution of the statistics Y^2 is chi square with $r = \mathrm{rank}(V^-) = \mathrm{Tr}(V^-V)$ degrees of freedom. If the matrix G is non-degenerate then $r = k$.

Test for the hypothesis H_0: the hypothesis is rejected with approximate significance level α if $Y^2 > \chi^2_\alpha(r)$.

Note that for there are many examples in Bagdonavičius *et al.* [6], Voinov *et al.* [29] related with some models very important for reliability and survival analysis.

Remark. On the choice of random grouping intervals.

An usual experiment plan in accelerated life testing is to test several groups of units under different higher stress conditions. In such experiment it is possible that the failures of units from different groups are mostly concentrated in different non-intersecting intervals. So using common idea of constructing chi square test by division of the interval $[0, \tau)$ into smaller intervals and comparing observed and expected numbers of failures the choice of the ends of the intervals is very important because dividing into intervals of equal length may give intervals where the numbers of observed failures are zero or very small.

Let us consider the choice of the limits of grouping intervals as random data functions.

Define for $j = 1, \dots, k$

$$E_k = \sum_{i=1}^{n} \int_0^{\tau} \lambda_i(u, \widehat{\boldsymbol{\theta}}) Y_i(u)du = \sum_{i=1}^{n} \Lambda_i(X_i, \widehat{\boldsymbol{\theta}}), \quad E_j = \frac{j}{k}E_k. \tag{10}$$

So we seek $\widehat{a}_j$ to have equal numbers of expected failures (not necessary an integer) in all intervals. So $\widehat{a}_j$ satisfy the equalities

$$g(\widehat{a}_j) = E_j, \quad g(a) = \sum_{i=1}^{n} \int_0^a \lambda_i(t, \widehat{\theta}) Y_i(u)\, du.$$

Denote by $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ the ordered sample from $X_1, \ldots, X_n$. Note that the function

$$\begin{aligned} g(a) &= \sum_{i=1}^{n} \Lambda_i(X_i \wedge a, \widehat{\boldsymbol{\theta}}) \\ &= \sum_{i=1}^{n} \left[\sum_{l=i}^{n} \Lambda_{(l)}(a, \widehat{\boldsymbol{\theta}}) + \sum_{l=1}^{i-1} \Lambda_{(l)}(X_{(l)}, \widehat{\boldsymbol{\theta}}) \right] \mathbf{1}_{[X_{(i-1)}, X_{(i)}]}(a) \end{aligned}$$

is continuous and increasing on $[0, \tau]$; here $X_{(0)} = 0$, and we understand $\sum_{l=1}^{0} c_l = 0$. Set

$$b_i = \sum_{l=i+1}^{n} \Lambda_{(l)}(X_{(i)}, \widehat{\boldsymbol{\theta}}) + \sum_{l=1}^{i} \Lambda_{(l)}(X_{(l)}, \widehat{\boldsymbol{\theta}}). \tag{11}$$

If $E_j \in [b_{i-1}, b_i]$ then $\widehat{a}_j$ is the unique solution of the equation

$$\sum_{l=i}^{n} \Lambda_{(l)}(\widehat{a}_j, \widehat{\boldsymbol{\theta}}) + \sum_{l=1}^{i-1} \Lambda_{(l)}(X_{(l)}, \widehat{\boldsymbol{\theta}}) = E_j. \tag{12}$$

We have $0 < \widehat{a}_1 < \widehat{a}_2 \ldots < \widehat{a}_k = \tau$.

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

Remark. One can verify Under conditions A and B and random choice of the endpoints of grouping intervals the limit distribution of the statistics Y^2 is chi-squared with r degrees of freedom.

So the hypothesis H_0 is rejected with approximate significance level α if $Y^2 > \chi^2_\alpha(r)$, the statistics Y^2 is computed using the formulas (10)–(12), *replacing a_j by $\widehat{a}_j$ in all formulas and taking*

$$e_1 = \cdots = e_k = \frac{1}{k} \sum_{i=1}^{n} \Lambda(X_i, \widehat{\boldsymbol{\theta}}, z_i).$$

References

1. O. Aalen, *Nonparametric inference for the family of counting processes.* — Ann. Statist. **6** (1978), 701–726.
2. M. G. Akritas, *Pearson-type goodness-of-fit tests: the univariate case.* — J. Amer. Statist. Assoc. **83** (1988), 222–230.
3. P. K. Andersen, O. Borgan, R. G. Gill, N. Keiding, *Statistical models based on counting processes.* Springer-Verlag, New York, 1993.
4. T. Aven, U. Jensen, *Stochastic models in reliability.* Springer, New York, 1999.
5. V. Bagdonavičius, M. Nikulin, *Accelerated life models.* Chapman and Hall/CRC, Boca-Raton, 2002.
6. V. Bagdonavičius, J. Kruopis, M. S. Nikulin, *Nonparametric tests for censored data.* ISTE and Wiley: London, 2011.
7. P. Billingsley, *Convergence of probability measures.* Wiley, New York, 1968.
8. H. Chernoff, E. L. Lehmann, *The use of maximum likelihood estimates in χ^2 tests for goodness of fit.* — Ann. Statist. **25** (1954), 579–586.
9. H. Cramer, *Mathematical methods of statistics.* Princeton University Press, Princeton, New York, 1946.
10. F. Drost, *Asymptotics for generalized chi-squared goodness-of-fit tests.* CWI Tracs **48**, Amsterdam, Centre for Mathematics and Computer Sciences, (1988).
11. P. Greenwood, M. S. Nikulin, *A guide to chi-squared testing.* Wiley, New York, 1996.
12. M. G. Habib, D. R. Thomas, *Chi-squared goodness-of-fit tests for randomly censored data.* — Ann. of Statist. **14** (1986), 759–765.
13. N. L. Hjort, *Goodness of fit tests in models for life history data based on cumulative hazard rates.* — Ann. Statist. **18** (1990), 1221–1258.
14. M. Hollander, E. Pena, *Chi-square goodness-of-fit test for randomly censored data.* — JASA, **87**, (1992), 458–463.
15. I. A. Ibragimov, R. Z. Hasminskii, *Statistical estimation. Asymptotic theory.* Springer-Verlag, Heidelberg, 1981.
16. J. P. Klein, M. L. Moeschberger, *Survival analysis,techniques for censored and truncated data.* Springer-Verlag, New York, 1997.
17. J. F. Lawless, *Statistical models and methods for lifetime data.* Wiley, 2nd ed., New York, 2003.
18. L. LeCam, C. Mahan, C. A. Singh, *An extension of a theorem of H. Chernoff and E. L. Lehmann.* — In: Recent advances in statistics, Academic Press, Orlando (1983), 303–332.
19. B. Y. Lemeshko, S. B. Lemeshko, S. N. Postovalov, E. V. Chimitova, *Statistical data analysis, simulation and study of probability regularities. Computer approach.* Novosibirsk, NSTU Publisher, 2011.
20. D. Moore, M. Spruill, *Unified large-sample theory of general chi-squared statistics for tests of fit.* — Ann. Statist. **3** (1975), 599–616.
21. W. Nelson, *Accelerated testing: statistical methods, tests, plans, and data analysis*, Wiley, New York, 1990.
22. M. S. Nikulin, *Chi-square test for normality.* — International Vilnius conference on probability theory and mathematical statistics, **2** (1973), 119–122.

23. M. S. Nikulin, *Chi-square test for continuous distribution with shift and scale parameters.* — Theory Probab. Appl. **18** (1973), 559–568.
24. M. S. Nikulin, *On a chi-square test for continuous distribution.* — Theory Probab. Appl., **18** (1973), 638–639.
25. E. Pena, *Smooth goodness-of-fit tests for composite hypothesis in hazard based models.* — Ann. Statist. **26** (1998), 1935–1971.
E. A. Pena, *Smooth goodness-of-fit tests for the baseline hazard in Cox's proportional hazards model.* — J. Amer. Statist. Assoc., **93**, (1998), 673–692.
26. K. C. Rao, D. S. Robson, *A chi-square statistics for goodness-of-fit tests within the exponential family.* — Commun. Statist. **3** (1974), 1139–1153.
27. A. W. van der Vaart, *Asymptotic Statistics.* Cambridge, UK, 1998.
28. V. Voinov, M. S. Nikulin, *Unbiased estimators and their applications.* Vol. 1, *Continuous case.* Kluwer Academic Publishers, Dordrecht, 1993.
29. V. G. Voinov, M. S. Nikulin, N. Balakrishnan, *Chi-squared goodness-of-tit tests with applications.* Academic Press (to appear), 2012.
30. B. Zhang, *A chi-squared goodness of-fit test for logistic regression models based on case-control data.* — Biometrika **86** (1999), 531–539.

Vilnius University,
Vilnius, Lithuania
E-mail: Vilijandas.Bagdonavicius@mif.vu.lt
Ruta.levuliene@mif.vu.lt

Поступило 15 октября 2012 г.

IMB, Victor Segalen University,
Bordeaux, France
E-mail: mikhail.nikouline@u-bordeaux2.fr