C. R. Acad. Sci. Paris, Ser. I 349 (2011) 219–223

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics/Probability Theory

Chi-squared tests for general composite hypotheses from censored samples

Tests chi carré pour des hypothèses composites généralisées des données censurées

Vilijandas Bagdonavičius[a], Mikhail Nikulin[b]

[a] *University of Vilnius, 24, Naugarduko, 03225 Vilnius, Lithuania*
[b] *Université Victor-Segalen, 33000 Bordeaux, France*

A R T I C L E   I N F O

Article history:
Received 3 November 2010
Accepted after revision 5 January 2011
Available online 20 January 2011

Presented by Paul Deheuvels

A B S T R A C T

We give chi-squared goodness-of-fit tests for parametric models including various regression models such as accelerated failure time, proportional hazards, generalized proportional hazards, frailty models, transformation models, models with cross-effects of survival functions. Choice of random grouping intervals as data functions is considered.

R É S U M É

On propose des tests d'ajustement du type chi carré pour des modèles paramétriques y compris les modèles de la vie accélerée, de Cox, de risques proportionnels généralisés, de transformations et d'autres. Le choix de limites des intervales de groupement comme fonctions des données est considéré.

Version française abrégée

Nous considérons des tests du type chi carré pour l'hypothèse générale composite H_0 disante que des fonctions de survie S_i, $i = 1, \dots, n$, de n objets sont des fonctions spécifiées $S_i(t, \theta)$ de temps t est de paramètre de dimension finie θ. Des tests pour les modèles de la vie accélerée, de Cox, de risques proportionnels généralisés, de transformations et d'autres sont des cas particuliers. Des données censurées à droite sont considérées.

Les tests sont basés sur le vecteur des différences entre les nombres des décès observés et les nombres des décès prédits par le modèle paramétrique dans les intervales de groupement. Les techniques des processus de comptage sont utilisées pour obtenir la loi limite de difference entre le processus des décès observés et le processus de décès prédits par le modèle paramétrique. La statistique du test est une somme de deux formes quadratiques et sa loi limite et chi deux. Le choix de limites des intervales de groupement comme fonctions des données est considéré.

1. Introduction

Suppose that n independent objects are observed. Let us consider the hypothesis H_0 stating that the survival functions of these objects S_i, $i = 1, \dots, n$, are specified functions $S_i(t, \theta)$ of time t and finite-dimensional parameter $\theta \in \Theta \subset \mathbf{R}^s$.

E-mail addresses: Vilijandas.Bagdonavicius@mif.vu.lt (V. Bagdonavičius), mikhail.nikouline@u-bordeaux2.fr (M. Nikulin).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2011.01.007

The hypothesis H_0 can be also formulated in terms of the hazard functions $\lambda_i(t,\theta) = -S_i'(t,\theta)/S_i(t,\theta)$ or the cumulative hazard functions $\Lambda_i(t,\theta) = \int_0^t \lambda_i(u,\theta)\,du$.

Let us consider examples of such hypotheses (see [1,4,6,5,11]):

(1) Composite hypothesis H_0: $S_i(t) = S(t,\theta)$, where $S(t,\theta)$ is the survival function of a specified parametric class of survival distributions, for example, a class of Weibull, loglogistic, lognormal, Gompertz distributions. The distribution is the same for any object.

(2) Parametric accelerated failure time (AFT) model: $S_i(t) = S_0(\int_0^t e^{-\beta^T z_i(u)}\,du;\gamma)$, where $z_i(t) = (1, z_{i1}(t), \ldots, z_{im}(t))^T$ is a vector of possibly time dependent covariates, $\beta = (\beta_0, \ldots, \beta_m)^T$ 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: $S_0(t,\gamma)$, $\gamma = (\gamma_1, \ldots, \gamma_q)^T \in G \subset R^q$. If explanatory variables are constant over time then the parametric AFT model has the form $S_i(t) = S_0(e^{-\beta^T z_i} t;\gamma)$.

(3) Parametric proportional hazards (Cox) model: $\lambda_i(t) = e^{\beta^T z_i(t)} \lambda(t,\gamma)$, where $\lambda(t,\gamma)$ is a hazard function of specified parametric form.

(4) Parametric generalized proportional hazards models (including parametric frailty and linear transformations models): $h(\Lambda_i(t),\gamma) = \int_0^t e^{\beta^T z_i(u)} \lambda(u,\nu)\,du$, where the function $h(x,\gamma)$ and the hazard function $\lambda(t,\nu)$ have specified forms. In particular, if $h(x,\gamma) = [(1+x)^\gamma - 1]/\gamma$, $h(x,\gamma) = [1 - e^{-\gamma x}]/\gamma$, and $h(x,\gamma) = x + \gamma x^2/2$, we have respectively generalizations of positive stable, gamma, and inverse gaussian frailty models with explanatory variables.

(5) Models with cross effects of survival functions: $\lambda_i(t) = g(z_i, \beta, \gamma, \Lambda(t,\nu))$, where the cumulative hazard $\Lambda(t,\nu)$ has a specified form and the function $g(z_i,\beta,\gamma,x)$ has one of the following forms:

$$g(z_i,\beta,\gamma,x) = e^{\beta^T z_i}\left[1 + e^{(\beta+\gamma)^T z_i} x\right]^{e^{-\gamma^T z_i} - 1}, \qquad g(z_i,\beta,\gamma,x) = \frac{e^{\beta^T z_i + x e^{\gamma^T z_i}}}{1 + e^{(\beta+\gamma)^T z_i}[e^{x e^{\gamma^T z_i}} - 1]}.$$

In complete data case well-known modification of the classical chi-squared tests is the Nikulin–Rao–Robson statistic 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 (Nikulin [17], Rao and Robson [19], Greenwood and Nikulin [9], LeCam, Mahan, Singh [14]). Goodness-of-fit tests for linear regression have been studied by Mukantseva [16], Pierce and Kopecky [18], Loynes [15], Koul [13].

Habib and Thomas [10], Hollander and Peña [12] considered natural modifications of the NRR statistic 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 is dew to Akritas [2] and was developed by Hjort [11]. In censored data case Hjort [11] considered goodness-of-fit for parametric Cox models, Gray and Pierce [8], Akritas and Torbeyns [3] – for linear regression models.

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

2. The idea of chi-squared test construction

We shall give chi-squared tests for the hypothesis H_0 from right censored $(X_1,\delta_1,z_1), \ldots, (X_n,\delta_n,z_n)$, where $X_i = T_i \wedge C_i$, $\delta_i = \mathbf{1}_{\{T_i \leqslant C_i\}}$, T_i being failure times, C_i – censoring times, and $z_i = (1, z_{i1}, \ldots, z_{im})^T$ – the covariates (this third component is absent in the case of the first example).

Set $N_i(t) = \mathbf{1}_{\{X_i \leqslant t, \delta_i = 1\}}$, $Y_i(t) = \mathbf{1}_{\{X_i \geqslant t\}}$, $N(t) = \sum_{i=1}^n N_i(t)$, $Y(t) = \sum_{i=1}^n Y_i(t)$.

Suppose that the processes N_i, Y_i, z_i are observed finite time τ and censoring is non-informative and the multiplicative intensities model is verified: the compensators of the counting processes N_i with respect to the history of the observed processes are $\int Y_i \lambda_i\,du$.

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 $U_j = N(a_j) - N(a_{j-1})$ the number of observed failures in the j-th interval, $j = 1, 2, \ldots, k$.

What is "expected" number of observed failures in the interval I_j under a specified AFT model? Under H_0 and regularity conditions the equality $EN_i(t) = E\int_0^t \lambda_i(u,\theta) Y_i(u)\,du$ holds. It suggests that we can "expect" to observe $e_j = \sum_{i=1}^n \int_{I_j} \lambda_i(u,\hat\theta) Y_i(u)\,du$ failures in the interval I_j; here $\hat\theta$ is the ML estimator of θ under H_0.

So a test can be based on the vector

$$Z = (Z_1, \ldots, Z_k)^T, \qquad Z_j = \frac{1}{\sqrt{n}}(U_j - e_j), \quad j = 1, \ldots, k,$$

of differences between the numbers of observed and "expected" failures in the intervals $I_1, \ldots, I_k$.

3. Asymptotic properties of the test statistics

To investigate the properties of the statistic Z we need properties of the stochastic process

$$H_n(t) = \frac{1}{\sqrt{n}}\left(N(t) - \sum_{i=1}^n \int_0^t \lambda_i(u,\hat\theta)Y_i(u)\,\mathrm{d}u\right).$$

To obtain these properties we use the properties of the ML estimators which are well known:

Conditions A. $\hat\theta \xrightarrow{P} \theta_0$, $\frac{1}{\sqrt{n}}\dot\ell(\theta_0) \xrightarrow{d} N_m(0, i(\theta_0))$, $-\frac{1}{n}\ddot\ell(\theta_0) \xrightarrow{P} i(\theta_0)$, and $\sqrt{n}(\hat\theta - \theta_0) = i^{-1}(\theta_0)\frac{1}{\sqrt{n}}\dot\ell(\theta_0) + o_P(1)$, where

$$\dot\ell(\theta) = \sum_{i=1}^n \int_0^\tau \frac{\partial}{\partial\theta}\ln\lambda_i(u,\theta)\left\{\mathrm{d}N_i(u) - Y_i(u)\lambda_i(u,\theta)\,\mathrm{d}u\right\}$$

is the score function.

Conditions A for consistency and asymptotic normality of the ML estimator $\hat\theta$ hold, for example, if conditions VI.1.1 given in Andersen et al. [4], Billingsley [7] hold. Set

$$S^{(0)}(t,\theta) = \sum_{i=1}^n Y_i(t)\lambda_i(t,\theta), \qquad S^{(1)}(t,\theta) = \sum_{i=1}^n Y_i(t)\frac{\partial\ln\lambda_i(t,\theta)}{\partial\theta}\lambda_i(t,\theta),$$
$$S^{(2)}(t,\theta) = \sum_{i=1}^n Y_i(t)\frac{\partial^2\ln\lambda_i(t,\theta)}{\partial\theta^2}\lambda_i(t,\theta).$$

Conditions B. There exist a neighborhood θ of θ_0 and continuous bounded on $\theta \times [0,\tau]$ functions

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

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

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

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

$$\frac{1}{n}\sum_{i=1}^n \int_0^t \lambda_i(u,\theta_0)Y_i(u)\,\mathrm{d}u \xrightarrow{P} A(t), \qquad \frac{1}{n}\sum_{i=1}^n \int_0^t \dot\lambda_i(u,\theta_0)Y_i(u)\,\mathrm{d}u \xrightarrow{P} C(t),$$

where A and C are finite quantities.

Theorem 3.1. *Under Conditions A and B the following convergence holds*:

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

here H is zero mean Gaussian martingale such that for all $0 \leqslant s \leqslant t$

$$cov\big(H(s), H(t)\big) = A(s) - C^T(s)I^{-1}(\theta_0)C(t),$$

$D[0,\tau]$ is the space of cadlag functions with Skorokhod metric.

For $i = 1,\dots,s;\ j, j' = 1,\dots,k$ set

$$V_j = H(a_j) - H(a_{j-1}), \qquad v_{jj'} = cov(V_j, V_{j'}), \qquad A_j = A(a_j) - A(a_{j-1}),$$
$$C_{ij} = C_i(a_j) - C_i(a_{j-1}), \qquad C_j = (C_{1j},\dots,C_{s,j})^T, \qquad V = [v_{jj'}]_{k\times k}, \qquad C = [C_{ij}]_{s\times k},$$

and denote by A a $k\times k$ diagonal matrix with the diagonal elements $A_1,\dots,A_k$.

Theorem 3.2. *Under Conditions A and B*

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

where $V = A - C^T i^{-1}(\theta_0)C$.

Set $G = i - CA^{-1}C^T$. The formula $V^- = A^{-1} + A^{-1}C^T G^- C A^{-1}$ implies that we need to inverse only diagonal $k \times k$ matrix A and find the general inverse of the $s \times s$ matrix G.

Theorem 3.3. *Under Conditions A and B the following estimators of A_j, C_j, $I(\theta_0)$ and V are consistent*:

$$\hat{A}_j = U_j/n, \qquad \hat{C}_j = \frac{1}{n}\sum_{i=1}^{n}\int_{I_j}\frac{\partial}{\partial\theta}\ln\lambda_i(u,\hat{\theta})\,\mathrm{d}N_i(u),$$

and

$$\hat{i} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\frac{\partial \ln\lambda_i(u,\hat{\theta})}{\partial\theta}\left(\frac{\partial \ln\lambda_i(u,\hat{\theta})}{\partial\theta}\right)^T \mathrm{d}N_i(u), \qquad \hat{V} = \hat{A} - \hat{C}^T\hat{i}^{-1}\hat{C}.$$

4. Chi-squared goodness-of-fit test

Theorems 1 and 2 imply that a test for the hypothesis H_0 can be based on the statistic $Y^2 = Z^T\hat{V}^- Z$, where $\hat{V}^- = \hat{A}^{-1} + \hat{A}^{-1}\hat{C}^T\hat{G}^-\hat{C}\hat{A}^{-1}$, $\hat{G} = \hat{I} - \hat{C}\hat{A}^{-1}\hat{C}^T$. This statistic can be written in the form

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

where

$$U_j = \sum_{i:\ X_i \in I_j}\delta_i, \qquad e_j = \sum_{i:\ X_i > a_{j-1}}\left[\Lambda_0(a_j \wedge X_i;\hat{\gamma}) - \Lambda_0(a_{j-1};\hat{\gamma})\right], \qquad Q = W^T\hat{G}^- W,$$

$$W = \hat{C}\hat{A}^{-1}z = (W_1,\ldots,W_s)^T, \qquad \hat{G} = [\hat{g}_{ll'}]_{s\times s}, \quad \hat{g}_{ll'} = \hat{i}_{ll'} - \sum_{j=1}^{k}\hat{C}_{lj}\hat{C}_{l'j}\hat{A}_j^{-1},$$

$$\hat{C}_j = \frac{1}{n}\sum_{i:\ X_i\in I_j}\delta_i\frac{\partial}{\partial\theta}\ln\lambda_i(X_i,\hat{\theta}), \qquad \hat{i} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{\partial\ln\lambda_i(X_i,\hat{\theta})}{\partial\theta}\left(\frac{\partial\ln\lambda_i(X_i,\hat{\theta})}{\partial\theta}\right)^T,$$

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

The limit distribution of the statistic Y^2 is chi-squared with $r = rank(V^-) = Tr(V^-V)$ 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 all examples (1)–(5) the rank $r = k - 1$ in the case of exponential, Weibull, Gompertz baseline distributions, and also for distribution with hyperbolic baseline hazard function, $r = k$ for lognormal, loglogistic baseline distributions. In the case of composite hypothesis of example (1) and exponential distribution the second quadratic form in the expression of the test statistic is equal to zero.

5. Choice of random grouping intervals

Let us consider choice of the limits of grouping intervals as random data functions. To take intervals of equal length may be not good solution. For example, if in accelerated life testing the AFT model is used and failure times of several groups tested under different stresses are observed then failures of different groups are concentrated around different means and if we take intervals of equal length then some empty or almost empty intervals are probable and the limit theorems cannot be applied. Moreover, under right censoring failure times of many units with long life are not observed and it can be other cause of empty intervals of equal length. Our method of choice of intervals with equal "expected numbers of failures" takes into account censoring via Y_i.

Define

$$E_k = \sum_{i=1}^{n} \int_0^{\tau} \lambda_i(u, \hat{\theta}) Y_i(u)\, \mathrm{d}u = \sum_{i=1}^{n} \Lambda_i(X_i, \hat{\theta}), \qquad E_j = \frac{j}{k} E_k, \quad j = 1, \ldots, k.$$

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

$$g(\hat{a}_j) = E_j, \qquad g(a) = \sum_{i=1}^{n} \int_0^{a} \lambda_i(t, \hat{\theta}) Y_i(u)\, \mathrm{d}u.$$

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

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

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

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

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

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

We have $0 < \hat{a}_1 < \hat{a}_2 < \cdots < \hat{a}_k = \tau$. Under this choice of the intervals $e_j = E_k/k$ for any j.

Theorem 5.1. *Under Conditions A and B and random choice of the endpoints of grouping intervals the limit distribution of the statistic Y^2 is chi-squared with r degrees of freedom.*

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] M.G. Akritas, A.F. Torbeyns, Pearson-type goodness-of-fit tests for regression, Canad. J. Statist. 25 (1997) 359–374.
[4] P.K. Andersen, O. Borgan, R.D. Gill, N. Keiding, Statistical Models Based on Counting Processes, Springer-Verlag, New York, 1993.
[5] V. Bagdonavičius, Testing the hyphothesis of the additive accumulation of damages, Probab. Theory Appl. 23 (1978) 403–408.
[6] V. Bagdonavičius, M. Nikulin, Accelerated Life Models, Chapman and Hall/CRC, Boca Raton, 2002.
[7] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[8] R.J. Gray, D.A. Pierce, Goodness of fit tests for censored survival data, Ann. Statist. 13 (1985) 552–563.
[9] P. Greenwood, M.S. Nikulin, A Guide to Chi-squared Testing, Wiley, New York, 1996.
[10] M.G. Habib, D.R. Thomas, Chi-squared goodness-of-fit tests for randomly censored data, Ann. Statist. 14 (1986) 759–765.
[11] N.L. Hjort, Goodness of fit tests in models for life history data based on cumulative hazard rates, Ann. Statist. 18 (1990) 1221–1258.
[12] M. Hollander, E. Peña, Chi-square goodness-of-fit test for randomly censored data, JASA 417 (1992) 458–463.
[13] H. Koul, Tests of goodness-of-fit in linear regression, Colloq. Math. Soc. János Bolyai 45 (1984) 279–315.
[14] L. LeCam, C. Mahan, A. Singh, An extension of a theorem of H. Chernoff and E.L. Lehmann, in: Recent Advances in Statistics, Academic Press, Orlando, 1983, pp. 303–332.
[15] R.M. Loynes, The empirical distribution function of residuals from generalized regression, Ann. Statist. 8 (1980) 285–298.
[16] L.A. Mukantseva, Testing normality in one-dimensional and multi-dimensional linear regression, Theory Probab. Appl. 22 (1977) 591–602.
[17] M.S. Nikulin, Chi-square test for continuous distribution with shift and scale parameters, Theory Probab. Appl. 19 (1973) 559–568.
[18] D. Pierce, K. Kopecky, Testing goodness of fit for the distribution of errors in regression models, Biometrika 66 (1979).
[19] K.C. Rao, D.S. Robson, A chi-square statistic for goodness-of-fit tests within the exponential family, Comm. Statist. 3 (1974) 1139–1153.