Testing Goodness-of-Fit of Parametric Models for Censored Data

Peer-reviewed author version

NYSEN, Ruth; AERTS, Marc & FAES, Christel (2011) Testing Goodness-of-Fit of Parametric Models for Censored Data. In: Conesa, David; Forte, Anabel; López-Quílez, Antonio; Muñoz, Facundo (Ed.). Proceedings of the 26th International Workshop on Statisical Modelling, p. 441-444.

Handle: http://hdl.handle.net/1942/16124

Testing Goodness-of-Fit of Parametric Models for Censored Data

R. Nysen[1], M. Aerts[1], C. Faes[1]

[1] Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Center for Statistics, Universiteit Hasselt. Agoralaan 1, B-3590 Diepenbeek, Belgium, ruth.nysen@uhasselt.be, marc.aerts@uhasselt.be, christel.faes@uhasselt.be

Abstract: A goodness-of-fit test for left-, right- and interval-censored data, assuming random censorship is proposed and studied. In the first step of the test, the null model is extended to a series of nested alternative models for censored data as in Zhang and Davidian (2008). Then a modified AIC model selection is used to select the best model to describe the data. If a model with one or more extra parameters is selected, then the null hypothesis is rejected. This new goodness-of-fit test procedure is based on the order selection test as described in Aerts, Claeskens and Hart (1999). The applicability of the test is illustrated in the context of microbial agents, and its performance characteristics are demonstrated through simulation studies.

Keywords: Goodness-of-fit test; Censored data; SNP estimator; Order selection test

1 Introduction

Censored data are often encountered in medical and public health studies. In survival studies, time to death can be right censored due to end-of-study or loss to follow-up. In infectious diseases, seroconversion time might only be known to fall in some interval, leading to interval-censored data. Within the framework of chemical risk assessment, the handling of concentration data reported to be below the limit of detection (left-censored) or between the limit of detection and the limit of quantification (interval-censored) present challenges to the statistical analysis of chemical occurrence data. When using parametric models, the choice of the distribution for such censored data is an important step in the analysis.

Goodness-of-fit tests for censored data have not been studied extensively. Hollander and Proschan (1979) present a test for a simple null hypothesis for right-censored data. This test can be applied for left-censored data by reversing the order of the observations. A test for interval censored data, based on the Cramér-von Mises statistic and a leveraged bootstrap, was introduced by Ren (2003). Bayesian tests were proposed by Yin (2009), Cao et al. (2010) and Calle and Gómez (2008, Chap. 21).

In this paper we propose and study a new goodness-of-fit test for left-, right- and interval-censored data, assuming random censorship. The test is based on the order selection test as described by Aerts, Claeskens and Hart (1999), which requires a series of nested alternative models in which the null model is nested. For censored data, such a family of densities can be described by the SNP (SemiNonParametric) representation of Zhang and Davidian (2008). The combination of the order selection test and the SNP representation results in a goodness-of-fit test for censored data.

2 Methodology

The test is based on the order selection test as described by Aerts, Claeskens and Hart (1999). They use a modified AIC criterion (MAIC) and accept the null hypothesis if and only if the prescribed distribution is chosen by the criterion

$$\text{MAIC}(r; C_n) = 2(\mathcal{L}_r - \mathcal{L}_0) - C_n r, \quad r = 0, 1, \ldots,$$

where C_n is some constant larger than 1. By appropriate choice of C_n, the asymptotic type I error probability of the test can be any number between 0 and 1. To determine C_n, a statistic T_n is defined as

$$T_n = \max_{1 \le r \le R_n} \{2(\mathcal{L}_r - \mathcal{L}_0)/r\},$$

for which the asymptotic distribution is known. The rejection of the null hypothesis is equivalent to $T_n > C_n$. For example, a test of asymptotic level .05 is obtained by $C_n = 4.18$. The P-value corresponding to an observed T_n can also be approximated by a bootstrap.
The procedure of Aerts, Claeskens and Hart (1999) requires that the null model is nested within the family of alternative models, which in turn form a sequence of nested models having more and more parameters. For censored data, such a broad class of densities can be described by the SNP (SemiNonParametric) representation of Zhang and Davidian (2008). In this representation, the density function under the null hypothesis is extended by multiplying with a polynomial of fixed degree r, introducing r new parameters in the model:

$$\gamma_r(z) = P_r^2(z)\psi(z),$$

where $P_r(z) = a_0 + a_1 z + \cdots + a_r z^r$ and $\int \gamma_r(z)dz = 1$. In case the null hypothesis states that data come from a lognormal distribution, the logarithm of the data can be written as $\log(T_0) = \theta_1 + \theta_2 Z$, where Z follows the standard normal distribution. The density $\psi(x)$ of the standard normal distribution is then multiplied with the square of a polynomial of degree r, such that r parameters are added to the model. The proposed goodness-of-fit test will reject the null hypothesis if the MAIC criterion selects a model

with $r > 0$. However, it is not guaranteed that the limiting distribution of T_n still holds when using censored data. Therefore the bootstrap offers an alternative.
The test can be used for different types of censoring. However, in the data analysis and the simulation study we focus on left- and interval-censored data.

3 Data analysis

The applicability of the test is illustrated through the analysis of some real data. The data under consideration consist of measurements of the cadmium level in some food category. 99 observations are available of which 42 are censored by the Limit of Detection (LOD). These limits of detection are in the range $[0.001, 0.01]$. The truly observed values are in between 0.0015 and 4.14. Some of the truly observed values are smaller than some of the LOD, because the data come from different laboratories, where different LOD's are applied.

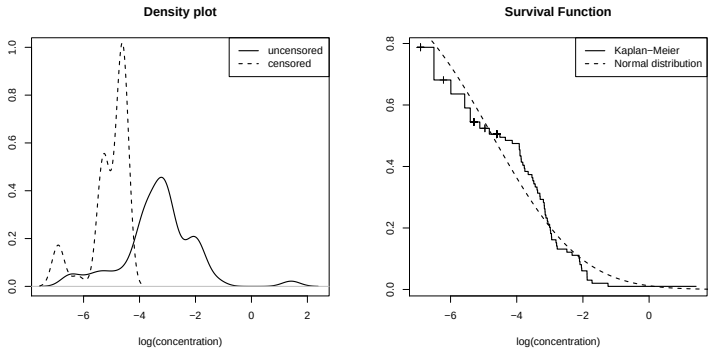

FIGURE 1. Cadmium data: Kernel density estimate and estimated survival functions of log(concentration).

A visual representation of the data is given in Figure 1. The left panel shows a kernel density of the logarithm of the concentrations. The right panel shows the Kaplan-Meier estimate, a frequently used estimator of the survival function in censored data. The LOD's are concentrated to the left and these values are denoted by a plus-sign (+) in the KM estimate. The fit for the normal distribution is represented by the dashed line.
In this situation we are interested in testing whether the concentrations are lognormally distribution. The proposed test was applied and the maximum MAIC was reached for $r = 3$, meaning that the null hypothesis of the lognormal distribution is rejected at significance level 5%.

4 Simulation Study and Discussion

The simulation study shows good performance of the test. Data are drawn from different distributions, with different sample sizes and different percentages of censoring. Focus is on left- and interval-censored data, where the null hypothesis states that the data come from the lognormal distribution.
Under the null hypothesis, the achieved percentage of rejected null hypotheses approximates the significance level. For example, a data set of size 100 is simulated with 12% left censoring, induced by five limits of detection. At 5% (respectively 10%) significance level, in 3% (respectively 8%) of the simulations, the hypothesis is rejected. The power of the test is high, especially for large sample sizes. For example, data are drawn from a mixture of two lognormal distributions. At 5% (10%) significance level, 55% (85%) of the tests are rejected.
A bootstrap is used to further investigate the distribution of the test statistic when data are censored. The bootstrapped sample is simulated under the null hypothesis and censoring is imposed by two different principles. Both principles try to resemble the censoring as good as possible. The p-values from both methods are close to the theoretical p-value.

References

Aerts, M., Claeskens, G. and Hart, J. (1999). Testing the fit of a parametric function. *Journal of the American Statistical Association*, **94**, 869-879.

Calle, M. L. and Gómez, G. (2008) *Statistical models and methods for biomedical and technical systems* Birkhäuser Boston

Cao, J., Moosman, A. and Johnson, V. E. (2010). A Bayesian Chi-Squared Goodness-of-Fit Test for Censored Data Models. *Biometrics*, **66**, 426-434.

Hollander, M. and Proschan, F. (1979). Testing to determine the underlying distribution using randomly censored data. *Biometrics*, **35**(2), 393-401.

Ren, J. (2003). Goodness of fit tests with interval censored data, *Scandinavian Journal of Statistics. Theory and Applications*, **30**(1), 211-226.

Yin, G. (2009). Bayesian goodness-of-fit test for censored data. *Journal of Statistical Planning and Inference*, **139**(4), 1474-1483.

Zhang, M., Davidian, M. (2008). Smooth semiparametric regression analysis for arbitrarily censored time-to-event data. *Biometrics*, **64**, 567-669.