Using Approximate MLE for Testing Normality Based on Kullback-Leibler Information with Progressively Type-II Censored Data

A. Habibi Rad and F. Yousefzadeh

Department of Statistics, Ferdowsi University of Mashhad

Department of Statistics, Birjand University

We will use the joint entropy of progressively censored order statistics in terms of an incomplete integral of the hazard function, and provide a simple estimate of the joint entropy of progressively Type-II censored data, has been introduced by Balakrishnan et al. (2007). Then We construct a goodness-of-fit test statistic based on Kullback-Leibler information for Normal distribution by using approximate MLE. Finally, we used Monte Carlo simulations, the power of the test is estimated and compared against several alternatives under different progressive censoring schemes.

Keywords: Approximate Maximum Likelihood Estimate, Entropy, Goodness-of-fit test, Hazard function, Monte Carlo simulation, Progressively Type-II censored data.

1. Introduction

Suppose a random variable X has a distribution function $F(x)$ and a continuous density function $f(x)$. The differential entropy $H(f)$ of the random variable is defined in Shannon (1948), to be

$$H(f) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx. \tag{1}$$

The first time, the test normality performed based on sample entropy by Vasicek (1976) and the power compared with some leading test statistics for complete samples.
The entropy difference $H(f) - H(g)$ has been considered in Dudewicz et al. (1981) and Gokhale (1983) for establishing goodness-of-fit tests for the class of the maximum entropy distributions.

The Kullback-Leibler (KL) information in favor of $g(x)$ against $f(x)$ is defined in Kullback (1959) to be

$$I(g : f) = \int_{-\infty}^{\infty} g(x) \log \frac{g(x)}{f(x)} dx,$$

which is an extended concept of entropy.
Because $I(g : f)$ has the property that $I(g : f) \geq 0$, and the equality holds if $g = f$, the estimate of the KL information has also been considered as a goodness-of-fit test statistic by some authors including Arizono et al. (1989) and Ebrahimi et al. (1992), for complete

samples. Park (2005) and Balakrishnan et al. (2007), respectively, for Type-II censored data and progressively Type-II censored data.
Now, in this paper we will extend the goodness-of-fit test based on KL information with progressively Type-II censored data for Normal distribution.

The rest of the paper is organized as follows: In Section 2 as Preliminary, we introduce Type-II progressive censoring data, the joint entropy of progressively censored data in terms of the hazard function and the nonparametric estimate of the joint entropy. In Section 3, we define the KL information for progressively Type-II censored data and propose a goodness-of-fit test for Normality based on KL information, in Section 4. Finally, in Section 5 we use Monte Carlo simulations to evaluate the power under different Type-II progressive censoring schemes.

2. PRELIMINARY

2.1. *Progressively Type-II Censored Data*

Suppose n identical items are placed on a life-testing experiment. Assume that their lifetimes are independent and identically distributed with probability distribution function (cdf) $F(x;\underline{\theta})$ and probability density function (pdf) $f(x;\underline{\theta})$, where θ is a vector of parameters.

There are several scenarios in life-testing and reliability experiments in which units that are subject to test are lost or removed from the experiment before failure. Such units are usually called the censored unites. The two most common censoring schemes are termed as conventional Type-I and Type-II censoring schemes which are extensively studied in statistical and reliability literature, Balakrishnan and Cohen (1991). Briefly, they can be described as follows: Consider n items under observations in a particular experiment. In the conventional Type-I censoring scheme, the experiment continues up a pre-specified time T. The conventional Type-II censoring scheme requires the experiment to continue until a pre-specified number of failures $m(\leq n)$ occur.
One of the drawbacks of the conventional Type-I, Type-II censoring schemes is that they do not allow for removal of units at points other than the terminal point of the experiment. One censoring scheme known as Type-II progressive censoring scheme, which has this advantage, so it becomes very popular for the last few years. It can be described as follows: Consider n units in a study and suppose $m(\leq n)$ is fixed before the experiment. Moreover, m other integers, $R_1, \cdots R_m$ are also fixed before so that $R_1+\cdots+R_m+m = n$. At the time of the first failure, say $X_{1:m:n},\ R_1$ of the remaining units are randomly removed. Similarly, at the time of the second failure, say $X_{2:m:n},\ R_2$ of the remaining units are randomly removed and so on. Finally, at the time of the $m-th$ failure, say $X_{m:m:n}$, the rest of the R_m units are removed. Fore further details on Type-II progressive censoring, refer to Balakrishnan and Aggarwala (2000).

The joint probability density function (pdf) of all m progressively Type-II censored order statistics $(X_{1:m:n}, \cdots X_{m:m:n})$ which is define in Balakrishnan (2000) to be

$$f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1,x_2,\cdots,x_m) = c\prod_{i=1}^{m} f(x_i)\{1-F(x_i)\}^{R_i},$$

$$x_1 < x_2 < \cdots < x_m,$$

where

$$c = n(n - R_1 - 1) \cdots (n - R_1 - R_2 - \cdots - R_{m-1} - m + 1).$$

2.2. *Entropy of Progressively Censored Data in Terms of the Hazard Function*

The joint entropy of $X_{1:m:n}, \cdots, X_{m:m:n}$ defined in literature (Park, 2005), to be

$$H_{1\cdots m:m:n} = -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{x_{2:m:n}} f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m) \times \log f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m) dx_{1:m:n} \cdots dx_{m:m:n},$$

where $f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m)$ is the joint pdf of all m progressively Type-II censored order statistics.

$H_{1\cdots m:m:n}$ is an m-dimensional integral, and we need to simplify this multiple integral.

The simple calculation of the entropy of the usual single and consecutive order statistics has been studied in Wong et al. (1990) and Park (1995). The multiple integral of the entopy for Type-II censored data be simplified to a single-integral by Park (2005) and the joint entropy of progressively Type-II censored order statistics in terms of an incomplete integral of the hazard function , $h(x)$, has been simplified by Balakrishnan et al. (2007),

$$H_{1\cdots m:m:n} = -\log c + n\overline{H}_{1\cdots m:m:n},$$

where

$$\overline{H}_{1\cdots m:m:n} = \frac{m}{n} - \frac{1}{n} \int_{-\infty}^{\infty} \sum_{i=1}^{m} f_{X_{i:m:n}}(x) \log h(x) dx.$$

2.3. *Nonparametric Entropy Estimate*

The nonparametric estimate of the joint entropy ($H_{1\cdots m:m:n}$) was obtained, as

$$H_{1\cdots m:m:n}(w, n, m) = -\log c + nH(w, n, m),$$

where

$$H(w, n, m) = \frac{1}{n} \sum_{i=1}^{m} \log \left(\frac{(x_{i+w:m:n} - x_{i-w:m:n})}{E(U_{i+w:m:n}) - E(U_{i-w:m:n})} \right) - (1 - \frac{m}{n}) \log(1 - \frac{m}{n}).$$

(Balakrishnan et al., 2007).

3. Goodness-of-fit Test Based on the Kullback-Leibler Information

For a null density function $f^0(x;\theta)$, the KL information from a progressively Type-II censored data is given by

$$I_{1\cdots m:m:n}(f : f^0) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{x_{2:m:n}} f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m; \theta) \times \log \frac{f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m; \theta)}{f^0_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1, x_2, \cdots, x_m; \theta)} dx_1 \cdots dx_m,$$

where $f_{X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n}}(x_1,x_2,\cdots,x_m)$ is the joint pdf of all m progressively Type-II censored order statistics.

The KL information can be estimated by

$$I_{1\cdots m:m:n}(f:f^0) = -n\overline{H}_{1\cdots m:m:n} - \sum_{i=1}^{m}\log f^0(x_i;\theta) - \sum_{i=1}^{m} R_i \log\left(1-F^0(x_i;\theta)\right). \quad (2)$$

Thus, the test statistic based on $\frac{1}{n}I_{1\cdots m:m:n}(f:f^0)$ is given by

$$T(w,n,m) = -H(w,n,m) - \frac{1}{n}\left[\sum_{i=1}^{m}\log f^0(x_i;\widehat{\theta}) + \sum_{i=1}^{m} R_i\log(1-F^0(x_i;\widehat{\theta}))\right], \quad (3)$$

where $\widehat{\theta}$ is an estimation of θ.

4. Test for Normality

Suppose we are interested in goodness-of-fit test for
$H_0: f^0 = (2\pi\sigma^2)^{\frac{-1}{2}}exp\{-(x-\mu)^2/2\sigma^2\}$ vs $H_A: f^0 \neq (2\pi\sigma^2)^{\frac{-1}{2}}exp\{-(x-\mu)^2/2\sigma^2\}$ where $\underline{\theta} = (\mu,\sigma^2)$ is unknown.
Then, the KL information for a progressively Type-II censored data can be approximated, by (3) and we estimate the unknown parameters (μ, σ^2) by the maximum likelihood estimate (MLE).
The MLE for progressively Type-II censored sample from a $Normal(\mu,\sigma^2)$ distribution obtain by solving the below equations, (Balakrishnan and Aggarwala, 2000)

$$\frac{\sum_{i=1}^{m} x_i}{m} = \overline{x} = \mu - \frac{\sigma}{m}\sum_{i=1}^{m} R_i Z_i,$$

$$\frac{\sum_{i=1}^{m}(x_i-\overline{x})^2}{m} = s^2 = \sigma^2\{1 - \frac{1}{m}\sum_{i=1}^{m} R_i\xi_i Z_i - (\frac{1}{m})^2\sum_{i=1}^{m}(R_i Zi)^2\},$$

where $Z_i = \frac{\varphi(\xi_i)}{1-\phi(\xi_i)}$ and $\varphi(.)$ is the probability density function of the standard normal distribution.
At the first we used a simple iterative procedure such as Newton's method for solving the above equations, but the MLE can not be obtained in explicit form so the next section we propose the approximate maximum likelihood estimates which have explicit forms.

4.1. *Approximate Maximum Likelihood Estimates for Normal Distribution*

In this section, we use the approximate maximum likelihood estimation method (AMLE) developed by Balakrishnan (1989 a,b, 1990 a,b,c) to estimate the scale and location parameters μ and σ.The likelihood function based on progressive Type-II censored sample $x_{1:m:n},...,x_{m:m:n}$ with censoring scheme $R_1,...,R_m$ can be written as

$$L(\mu,\sigma) = c\frac{1}{\sigma^m}\prod_{i=1}^{m} f(z_{i:m:n})(\bar{F}(z_{i:m:n}))^{R_i},$$

where $c = n(n-R_1-1)\cdots(n-R_1-R_2-\cdots-R_{m-1}-m+1)$, $z_{i:m:n} = \frac{x_{i:m:n}-\mu}{\sigma}$, $\overline{F}(.) = 1-F(.)$ and f, F are the probability density function(pdf) and cumulative distribution function (cdf) of Normal standard distribution, respectively.
Upon partial differentiation of the logarithm of the likelihood function with respect to μ and σ, the score equations to be solved for μ and σ in this case are given by

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{m} z_{i:m:n} + \frac{1}{\sigma}\sum_{i=1}^{m} R_i \frac{f(z_{i:m:n})}{\bar{F}(z_{i:m:n})} = 0 \tag{4}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{m}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{m} z_{i:m:n}^2 + \frac{1}{\sigma}\sum_{i=1}^{m} R_i z_{i:m:n} \frac{f(z_{i:m:n})}{\bar{F}(z_{i:m:n})} = 0, \tag{5}$$

Clearly, (4) and (5) do not have explicit solutions. We expand the function $\frac{f(z_{i:m:n})}{\bar{F}(z_{i:m:n})}$ in Taylor series around the point $\xi_i = F^{-1}(p_i)$, where $p_i = 1 - q_i = 1 - \prod_{j=m-i+1}^{m} \alpha_j$. Balakrishnan and Aggarwala(2000) deduced that: if $U_{i:m:n}, \quad i = 1, \cdots, m$ denote a progressive Type-II censored sample from the uniform(0, 1) distribution obtained from a sample of size n with the censoring scheme $(R_1, \cdots, R_m)$, then $V_i, i = 1, \cdots, m$ are all independent random variables with $V_i = Beta(i + \sum\limits_{j=m-i+1}^{m} R_j, 1), \ i = 1, \cdots, m$, such that

$$U_{i:m:n} = 1 - \prod_{j=m-i+1}^{m} V_j, \quad i = 1, \cdots, m,$$

and

$$E(U_{i:m:n}) = 1 - \prod_{j=m-i+1}^{m} \alpha_j, \quad i = 1, \cdots, m,$$

where

$$\alpha_j = \frac{j + \sum\limits_{i=m-j+1}^{m} R_i}{1 + j + \sum\limits_{i=m-j+1}^{m} R_i}, \quad j = 1, \cdots, m.$$

Then we consider the following approximations

$$\frac{f(z_{i:m:n})}{\bar{F}(z_{i:m:n})} \simeq \alpha_i + \beta_i z_{i:m:n}, \tag{6}$$

where

$$\alpha_i = \frac{f(\xi_i)}{\bar{F}(\xi_i)} - \xi_i\left[-\xi_i\frac{f(\xi_i)}{\bar{F}(\xi_i)} + \left(\frac{f(\xi_i)}{\bar{F}(\xi_i)}\right)^2\right],$$

$$\beta_i = -\xi_i\frac{f(\xi_i)}{\bar{F}(\xi_i)} + \left(\frac{f(\xi_i)}{\bar{F}(\xi_i)}\right)^2.$$

Using the approximations (6) in (4) and (5), we obtain

$$\sum_{i=1}^{m} z_{i:m:n} + \sum_{i=1}^{m} R_i(\alpha_i + \beta_i z_{i:m:n}) = 0, \tag{7}$$

$$-m + \sum_{i=1}^{m} z_{i:m:n}^2 + \sum_{i=1}^{m} R_i z_{i:m:n}(\alpha_i + \beta_i z_{i:m:n}) = 0. \tag{8}$$

From (7) we obtain the AMLE of μ as

$$\hat{\mu} = B + \hat{\sigma}C,$$

where

$$B = \frac{m\bar{x} + \sum_{i=1}^{m} R_i\beta_i x_{i:m:n}}{m + \sum_{i=1}^{m} R_i\beta_i},$$

$$C = \frac{\sum_{i=1}^{m} R_i\alpha_i}{m + \sum_{i=1}^{m} R_i\beta_i}.$$

From (8), we obtain $\hat{\sigma}$ as a solution of the quadratic equation

$$A_1\sigma^2 + A_2\sigma + A_3 = 0,$$

where

$$A_1 = -m, A_2 = \sum_{i=1}^{m} R_i\alpha_i(x_{i:m:n} - B),$$

$$A_3 = \sum_{i=1}^{m} (1 + R_i\beta_i)(x_{i:m:n} - B)^2 > 0.$$

Therefore

$$\hat{\sigma} = \frac{-A_2 - \sqrt{A_2^2 - 4A_1A_3}}{2A_1},$$

is the only positive root.

5. Implementation of Test

Because the sampling distribution of $T(w, n, m)$ is intractable, we determine the percentage points using 10,000 Monte Carlo simulations from Normal distribution. In determining the window size w which depends on n, m and α, we define the optimal window size w to be one which gives minimum critical points. However, we find from the simulated percentage points that the optimal window size w varies much according to m rather than n, and does not vary much according to α, if $\alpha \leq 0.1$. In view of these observations, our recommended values of w for different m are as given in Ebrahimi (1992) and Park (2005).

To obtain the critical values, after deciding about the value of w, simulate the whole procedure by taking the observation from $Normal(0,1)$ distribution and calculate the value of $T(w,n,m)$, for about 10,000 times. Critical values can then be the percentage points of the thus derived (empirical) distribution of T.

5.1. *Power Results for Normal Distribution*

As the proposed test statistic is related to the hazard function of the distribution, we consider the alternatives according to the type of hazard function as follows:

a) Monotone increasing hazard: Gamma and Weibull (shape parameter 2),
b) Monotone decreasing hazard: Gamma and Weibull (shape parameter 0.5),
c) Nonmonotone hazard: Center Beta (shape parameter 0.5),
Log-normal (shape parameter 1).

We used 10,000 Monte Carlo simulations for $n = 10, 20$, to estimate the power of our proposed test statistic. The simulation results are summarized in Tables 1 and 2.

We can see from Tables 1 and 2 that the scheme $(R_1 = 0, \cdots, R_{m-1} = 0, R_m = n-m)$ (the conventional Type-II censored data) shows better power than the other schemes when the alternative is monotone increasing hazard function. For the alternative with monotone decreasing hazard functions, the scheme $(R_1 = n - m, R_2 = 0, \cdots, R_m = 0)$ shows better power; finally, for the alternative with nonmonotone hazard function, sometimes the former censoring scheme gives higher power and sometimes the latter censoring scheme does.

Table 1: Power for different hazard alternatives at 10% significance level for several progressively censored samples when the sample size is $n = 10$.

		monotone increasing hazard alternatives		monotone decreasing hazard alternatives		nonmonotone hazard alternatives	
m	schemes $(R_1, \cdots, R_m)$	Gamma shape 2	Weibull shape 2	Gamma shape 0.5	Weibull shape 0.5	Center Beta shape 0.5	Log-normal shape 1
5	5,0,0,0,0	.198	.123	.555	.705	.341	.403
5	0,5,0,0,0	.196	.118	.563	.700	.387	.388
5	1,1,1,1,1	.165	.116	.516	.624	.417	.274
5	0,0,0,5,0	.123	.101	.435	.566	.312	.214
5	0,0,0,0,5	.173	.133	.464	.539	.422	.235
7	3,0,0,0,0,0,0	.272	.137	.709	.853	.420	.564
7	0,3,0,0,0,0,0	.272	.142	.711	.851	.426	.563
7	1,0,0,1,0,0,1	.262	.148	.704	.825	.507	.486
7	0,0,0,0,0,3,0	.165	.118	.567	.719	.395	.326
7	0,0,0,0,0,0,3	.262	.163	.681	.789	.561	.433
9	1,0,0,...,0,0,0	.319	.147	.808	.926	.501	.660
9	0,1,0,...,0,0,0	.325	.149	.809	.928	.505	.673
9	0,0,...,1,...,0,0	.288	.131	.798	.920	.490	.636
9	0,0,0,...,0,1,0	.249	.141	.735	.876	.506	.542
9	0,0,0,...,0,0,1	.355	.188	.834	.930	.596	.654

Table 2: Power for different hazard alternatives at 10% significance level for several progressively censored samples when the sample size is $n = 20$.

		monotone increasing hazard alternatives		monotone decreasing hazard alternatives		nonmonotone hazard alternatives	
m	schemes $(R_1, \cdots, R_m)$	Gamma shape 2	Weibull shape 2	Gamma shape 0.5	Weibull shape 0.5	Center Beta shape 0.5	LogNormal shape 1
5	15,0,0,0,0	.194	.114	.586	.727	.388	.401
5	0,15,0,0,0	.230	.136	.639	.762	.482	.429
5	3,3,3,3,3	.158	.121	.576	.644	.546	.221
5	0,0,0,15,0	.129	.102	.584	.671	.499	.223
5	0,0,0,0,15	.167	.144	.493	.535	.482	.196
10	10,0,0,...,0,0,0	.354	.148	.906	.973	.667	.745
10	0,10,0,...,0,0,0	.409	.181	.918	.977	.708	.769
10	1,1,1,...,1,1,1	.332	.175	.912	.960	.830	.595
10	0,0,0,...,0,10,0	.146	.123	.651	.788	.507	.265
10	0,0,0,...,0,0,10	.346	.224	.890	.930	.847	.515
15	5,0,0,...,0,0,0	.438	.182	.967	.995	.836	.856
15	0,5,0,...,0,0,0	.472	.204	.969	.996	.846	.869
15	1,1,...,1,...,1,1	.526	.263	.981	.996	.907	.851
15	0,0,0,...,0,5,0	.260	.182	.789	.918	.657	.468
15	0,0,0,...,0,0,5	.558	.302	.982	.996	.934	.833
18	2,0,0,...,0,0,0	.480	.209	.980	.998	.897	.904
18	0,2,0,...,0,0,0	.492	.210	.981	.998	.898	.902
18	1,0,0,...,0,0,1	.630	.287	.996	.999	.948	.940
18	0,0,0,...,0,2,0	.379	.212	.935	.987	.826	.740
18	0,0,0,...,0,0,2	.670	.324	.997	.999	.962	.945

References

Arizono, I. and Ohta, H. " A test for normality based on Kullback-Leibler information," *The American Statistician,* **43**, 20–23, (1989).

Balakrishnan, N. "Approximate MLE of the scale parameter of the Rayleigh distribution with censoring," *IEEE Trans. on Rel.,* **38**, 355–357, (1989a).

Balakrishnan, N. "Approximate maximum likelihood estimation of the mean and standard deviation of the normal distribution based on Type-II censored samples," *J. of Statistical computation and simulation,* **32**, 137-148, (1989b).

Balakrishnan, N. *"Maximum likelihood estimation based on complete and Type-II cencored samples, in the Logistic Distribution,"* (ed. N. Balakrishnan), Marcel Dekker, New York, (1990a).

Balakrishnan, N. "Approximate maximum likelihood estimation for a generalized Logistic distribution," *J. of Statistical Planning and Inference,* **26**, 221-236, (1990b).

Balakrishnan, N. "On the maximum likelihood estimation of the location and scale parameters of exponential distribution based on multiply Type-II censored samples," *J. App. Statist.,* **17**, 55-61, (1990c).

Balakrishnan, N. and Cohen, A. C. *"Order Statistics and Inference: Estimation Methods,"* Academic Press, San Diego, (1991).

Balakrishnan, N. and Aggarwala, R. *"Progressive Censoring: Theory, Methods, and Applications,"* Birkhauser, Boston, (2000).

Balakrishnan, N., Ng, H. K. T. and Kannan, N. "Goodness-of-fit tests based on spacings for progressively type-II censored data from a general location-scale distribution," *IEEE Trans. on Rel.,* **53**, 349–356, (200).

Balakrishnan, N., Habibi Rad, A. and Arghami, N. R. "Testing exponentiality based on Kullback-Leibler information with progressively type-II censored data," *IEEE Trans. on Rel.,* **53**, 349–356, (2007).

Dudewicz, E. J. and van der Meulen, E. C. "Entropy-based tests of uniformity," *J. Amer. Statist. Assoc.,* **76**, 967–974, (1981).

Ebrahimi, N. and Habibullah, M. "Testing exponentiality based on Kullback-Leibler information," *J. Royal Statist. Soc. B,* **54**, 739–748, (1992).

Gokhale, D. V. "On entropy-based goodness-of-fit tests," *Computational Statistics and Data Analysis,* **1**, 157–165, (1983).

Kullback, S. "*Information Theory and Statistics,"* New York: Wiley, (1959).

Park, S. "The entropy of consecutive order statistics," *IEEE Trans. Inform. Theory,* **41**, 2003–2007, (1995).

Park, S. "Testing exponentiality based on the Kullback-Leibler information with the type-II censored data," *IEEE Trans. on Rel.,* **54**, 22–26, (2005).

Shannon, C. E. "A mathematical theory of communications," *Bell System Tech. J.,* **27**, 379–423, (1948).

Vasicek, O. "A test for normality based on sample entropy," *J. Royal Statist. Soc. B,* **38**, 730–737, (1976).

Wong, K. M. and Chan, S. "The entropy of ordered sequences and order statistics," *IEEE Trans. Inform. Theory,* **36**, 276–284, (1990).