Testing Exponentiality Based on the Kullback-Leibler Information With the Type II Censored Data

Sangun Park

***Abstract*—We express the joint entropy of order statistics in terms of an incomplete integral of the hazard function, and provide a simple estimate of the joint entropy of the type II censored data. Then we establish a goodness of fit test statistic based on the Kullback-Leibler information with the type II censored data, and compare its performance with some leading test statistics. A Monte Carlo simulation study shows that the proposed test statistic shows better powers than some leading test statistics against the alternatives with monotone increasing hazard functions.**

***Index Terms*—Entropy, hazard function, maximum entropy, Monte-Carlo simulation, order statistics.**

Notations

n	sample size.
$h(x)$	the hazard function (rate), $f(x)/(1-F(x))$.
$X_{(r;n)}$	the rth order statistic of an independently identically distributed (i.i.d.) sample of size n from $F(x)$.
$U_{(r;n)}$	the rth order statistic of a sample of size n from the uniform distribution.
$F_{1\cdots r;n}$	the joint cumulative distribution function (p.d.f) of $X_{(1;n)},\cdots,X_{(r;n)}$.
$f_{1\cdots r;n}$	p.d.f. of $F_{1\cdots r;n}$.
$H_{1\cdots r;n}$	the joint entropy of $X_{(1;n)},\cdots,X_{(r;n)}$.
$I_{1\cdots r;n}(g:f)$	the Kullback-Leibler information of $X_{(1;n)},\cdots,X_{(r;n)}$.
$f_{r\|r-1;n}$	the conditional p.d.f. of $X_{(r;n)}$ given $X_{(r-1;n)}=x_{(r-1;n)}$.
$H_{r\|r-1;n}(x_{(r-1;n)})$	the conditional entropy of $X_{(r;n)}$ given $X_{(r-1;n)}=x_{(r-1;n)}$.
$H_{r\|r-1;n}$	the expectation of $H_{r\|r-1;n}(x_{(r-1;n)})$ about $X_{(r-1;n)}$.

I. Introduction

SUPPOSE that a random variable X has a distribution function $F(x)$, with a continuous density function $f(x)$. The differential entropy $H(f)$ of the random variable is defined by [15] to be

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

Manuscript received June 18, 2001. This work was supported by Yonsei University Research Fund of 2003. Associate Editor: J.-C. Lu.

The author is with the Department of Applied Statistics, Yonsei University, Shinchon Dong 134, Seoul, Korea (e-mail: sangun@yonsei.ac.kr).

Digital Object Identifier 10.1109/TR.2004.837314

The entropy difference $H(f)-H(g)$ has been considered in establishing goodness of fit tests for the class of the maximum entropy distributions [5], [9].

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

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

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 been also considered as a goodness of fit test statistic by some authors including [1], [6]. It has been shown in the aforementioned papers that the test statistics based on the KL information perform very well for exponentiality [6], and s-normality [17] in terms of powers compared with some leading test statistics for complete samples. Thus, we may be interested in whether test statistics based on KL information also perform very well for the type II censored data. However, we can instantly find that the extension to the type II censored data is not so straightforward, because we need to estimate the joint entropy of the type II censored data, which is an r multiple integral.

The joint entropy of $X_{(1:n)},\cdots,X_{(r:n)}$ is simply defined to be

$$H_{1\cdots r:n}=-\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{x_{(2:n)}} f_{1\cdots r:n}\log f_{1\cdots r:n}dx_{(1:n)}\cdots dx_{(r:n)}$$

where $f_{1\cdots r:n}$ is the joint probability density function (p.d.f) of $X_{(1:n)},\cdots,X_{(r:n)}$.

However, $H_{1\cdots r:n}$ is an r multiple integral, and we need to simplify the multiple integral. The simple calculation of the entropy of single and consecutive order statistics has been studied in [13], [18]. In Section II we first provide a single-integral representation of $H_{1\cdots r:n}$ in terms of the hazard function, $h(x)$, as

$$H_{1\cdots r:n}=-(\log n+\cdots+\log(n-r+1))+r-n\times\int_{-\infty}^{\infty}(1-F_{r:n-1}(x))\,f(x)\log h(x)dx. \tag{2}$$

Then we provide an estimate of (2), and establish a goodness of fit test based on the KL information.

II. Entropy of the Censored Data in Terms of the Hazard Function

A. Entropy Representation

[16] derived another expression of (1) in terms of the hazard function as

$$H_{1:1} = 1 - \int_{-\infty}^{\infty} f(x) \log h(x) dx. \tag{3}$$

We first note that (3) gives a simple expression of $H_{1:n}$ as

$$H_{1:n} = -\log n + 1 - \int_{-\infty}^{\infty} f_{1:n}(x) \log h(x) dx. \tag{4}$$

Theorem 2.1 says that the multiple integral $H_{1\cdots r:n}$ can be simplified to a single integral.

Theorem 2.1:

$$H_{1\cdots r:n} = -(\log n + \cdots + \log(n-r+1)) + n\bar{H}_{1\cdots r:n}$$

where

$$\bar{H}_{1\cdots r:n} = \frac{r}{n} - \int_{-\infty}^{\infty} (1 - F_{r:n-1}(x)) f(x) \log h(x) dx$$

Proof: By the decomposition property of the entropy measure [13], we have

$$H_{1\cdots r:n} = H_{1:n} + H_{2|1:n} + \cdots + H_{r|r-1:n}.$$

Because $f_{r|r-1:n}$ can be interpreted as the density of the first order statistic among an $(n-r+1)$ sample from $f(x)/(1 - F(x_{(r:n)}))$, we can derive by using (4)

$$H_{r|r-1} = -\log(n-r+1) + 1 - \int_{-\infty}^{\infty} f_{r:n}(x) \log h(x) dx.$$

Because

$$\begin{aligned} f_{1:n}(x) + \cdots + f_{r:n}(x) &= nf(x) \sum_{i=0}^{r-1} C_{n-1,i} F^i(x) \\ &\quad \times (1 - F(x))^{n-i-1} \\ &= nf(x)\,(1 - F_{r:n-1}(x)) \end{aligned}$$

the result follows. □

$\bar{H}_{1\cdots r:n}$ in Theorem 2.1 can be written in terms of $\log f(x)$ as follows.

Lemma 2.1:

$$\begin{aligned} \bar{H}_{1\cdots r:n} = &- \int_{-\infty}^{\infty} (1 - F_{r:n-1}(x))\, f(x) \log f(x) dx - E \\ &\times \left(\left(1 - U_{(r:n-1)}\right) \log \left(1 - U_{(r:n-1)}\right) \right) \end{aligned}$$

where $U_{(r:n-1)}$ is the rth uniform order statistic from a sample of size $n-1$.

The result follows instantly by the integral-by-part.

B. Nonparametric Entropy Estimate and Test Statistic

Some nonparametric estimates of (1) have been proposed by [3], [7], [17]. [17] expressed (1) in the form,

$$H = \int_0^1 \log \left(\frac{d}{dp} F^{-1}(p) \right) dp,$$

and provided its estimate as

$$H(m,n) = \frac{1}{n} \sum_{i=1}^{n} \log \left(\frac{n}{2m} \left(x_{(i+m:n)} - x_{(i-m:n)} \right) \right) \tag{5}$$

where the window size m is a positive integer, which is less than $n/2$; and $x_{(i:n)} = x_{(1:n)}$ for $i < 1$, and $x_{(i:n)} = x_{(n:n)}$ for $i > n$.

Because

$$\begin{aligned} &- \int_{-\infty}^{\infty} (1 - F_{r:n-1}(x))\, f(x) \log f(x) dx \\ &\qquad = -E \left(\int_0^{U_{(r:n-1)}} \log \left(\frac{d}{dp} F^{-1}(p) \right) dp \right) \end{aligned}$$

if we consider the derivation of (5), we can obtain the estimate of $\bar{H}_{1\cdots r:n}$ as

$$\begin{aligned} H(m,n,r) = &\frac{1}{n} \sum_{i=1}^{r} \log \left\{ \frac{n}{2m} \left(x_{(i+m:n)} - x_{(i-m:n)} \right) \right\} \\ &- \left(1 - \frac{r}{n}\right) \log \left(1 - \frac{r}{n}\right). \end{aligned}$$

We note that $H(m,n,n) = H(m,n)$. Thus the estimate of $H_{1\cdots r:n}$ is in view of Lemma 2.1 written as

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

Theorem 2.2: Suppose that $X_{(r:n)}$ be the pth sample quantile. Then

$$\begin{aligned} &|H(m,n,r) - \bar{H}_{1\cdots r:n}| \to 0 \\ &\qquad \text{in probability as } n, m \to \infty \text{ and } \frac{m}{n} \to 0. \end{aligned}$$

Proof: Since $Pr(X_{(r:n)} > x)$ converges to $I(x < \xi_p)$, we have in view of Lemma 2.1

$$\bar{H}_{1\cdots r:n} \to - \int_{-\infty}^{\xi_p} f(x) \log f(x) dx - (1-p) \log(1-p).$$

We can also show by following the lines of [17] that, as $n, m \to \infty$ and $m/n \to 0$,

$$\begin{aligned} H(m,n,r) \to &- \int_{-\infty}^{\xi_p} f(x) \log f(x) dx \\ &-(1-p) \log(1-p) \text{ in probability} \end{aligned}$$

Thus the result follows. □

For a null distribution function $f^0(x:\theta)$, the KL information for the type II censored data is defined to be

$$I_{1\cdots r:n}(f; f^0) = \int_{-\infty}^{\infty} f_{1\cdots r:n}(x;\theta) \log \frac{f_{1\cdots r:n}(x;\theta)}{f^0_{1\cdots r:n}(x;\theta)} dx.$$

Then the KL information can be approximated with

$$\begin{aligned} I_{1\cdots r:n}(f; f^0) = &-n\bar{H}_{1\cdots r:n} - \sum_{i=1}^{r} \log f^0 \\ &\times \left(x_{(i)};\theta\right) - (n-r) \\ &\times \log \left(1 - F^0\left(x_{(r)};\theta\right)\right). \end{aligned} \tag{6}$$

TABLE I
VALUES OF THE WINDOW SIZE m WHICH GIVES MINIMUM CRITICAL VALUES OF α LESS THAN 0.1

r	m
5-7	2
8-15	3
16-30	4
31-40	5
41-50	6

TABLE II
MONTE CARLO ESTIMATE OF THE CRITICAL VALUES OF $T(n,m,r)$ WHERE m IS DETERMINED FROM TABLE I

n	r	$\alpha=.1$	$\alpha=.05$	$\alpha=.025$	$\alpha=.01$
10	5	.4373	.5315	.6170	.7251
	6	.4490	.5429	.6289	.7485
	7	.4664	.5587	.6486	.7587
	8	.4906	.5769	.6572	.7733
	9	.5220	.6097	.6950	.8014
20	10	.2497	.2923	.3324	.3835
	12	.2620	.3040	.3428	.3956
	14	.2760	.3188	.3635	.4128
	16	.2901	.3356	.3739	.4227
	18	.3056	.3520	.3955	.4428
30	15	.1888	.2162	.2419	.2749
	18	.1973	.2246	.2495	.2844
	21	.2068	.2363	.2645	.2978
	24	.2159	.2434	.2713	.3052
	27	.2255	.2571	.2845	.3237
40	20	.0069	.1760	.1939	.2181
	24	.1649	.1846	.2045	.2267
	28	.1713	.1925	.2107	.2355
	32	.1782	.1990	.2191	.2448
	36	.1850	.2053	.2247	.2531
50	25	.1396	.1551	.1720	.1949
	30	.1452	.1606	.1769	.1962
	35	.1504	.1672	.1826	.2036
	40	.1565	.1742	.1899	.2097
	45	.1605	.1780	.1966	.2205

Thus the test statistic based on $I_{1\cdots r:n}(f;f^0)/n$ can be written as

$$T(n,m,r) = -\bar{H}(n,m,r) - \frac{1}{n}\left(\sum_{i=1}^{r}\log f^0\left(x_{(i)};\hat{\theta}\right)+(n-r)\times\log\left(1-F^0\left(x_{(r)};\hat{\theta}\right)\right)\right) \quad (7)$$

where $\hat{\theta}$ is an estimator of θ.

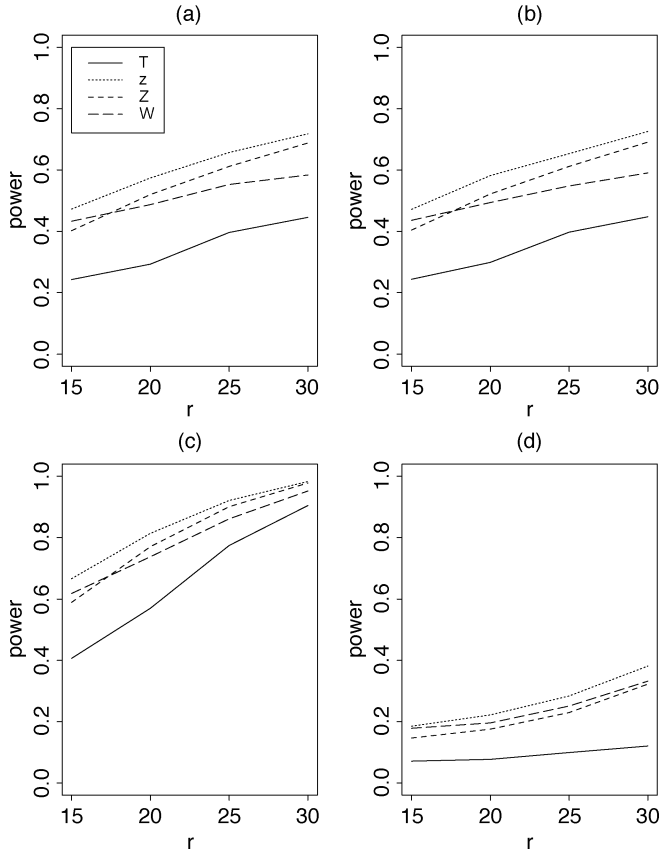

Fig. 1. Power comparison: monotone decreasing hazard alternatives at 10% when the sample size is 30. (a) Chi-square: df 1; (b) Gamma: shape 0.5; (c) Weibull: shape 0.5; (d) Weibull: shape 0.8.

III. TESTING EXPONENTIALITY BASED ON THE KULLBACK-LEIBLER INFORMATION

A. Test Statistic

Suppose that we are interested in a goodness of fit test for $H_0: f^0(x)=\exp(-x/\theta)/\theta$ vs $H_A: f^0(x)\neq\exp(-x/\theta)/\theta$ where θ is unknown. Then the KL information for the type II censored data can be approximated in view of (6) with

$$I_{1\cdots r:n}(f;f^0) = -n\bar{H}_{1\cdots r:n} + r\log\theta + \frac{1}{\theta}\left(\sum_{i=1}^{r}X_{(i:n)}+(n-r)X_{(r:n)}\right).$$

If we estimate the unknown θ with the maximum likelihood estimator, $(\sum_{i=1}^{r}X_{(i:n)}+(n-r)X_{(r:n)})/r$, then we have an estimate of $I_{1\cdots r:n}(f;f^0)/n$ as

$$T(m,n,r) = -H(m,n,r) + \frac{r}{n}\times\left(\log\left(\frac{1}{r}\left(\sum_{i=1}^{r}X_{(i:n)}+(n-r)X_{(r:n)}\right)\right)+1\right).$$

Under the null hypothesis, $T(n,m,r)$ will be close to 0. When $r=n$, $T(n,m,r)$ becomes the test statistic of [6].

B. Implementation of the Test

Because the sampling distribution of $T(n,m,r)$ is intractable, we determine the percentage points using 10 000 Monte Carlo samples from an exponential distribution. In

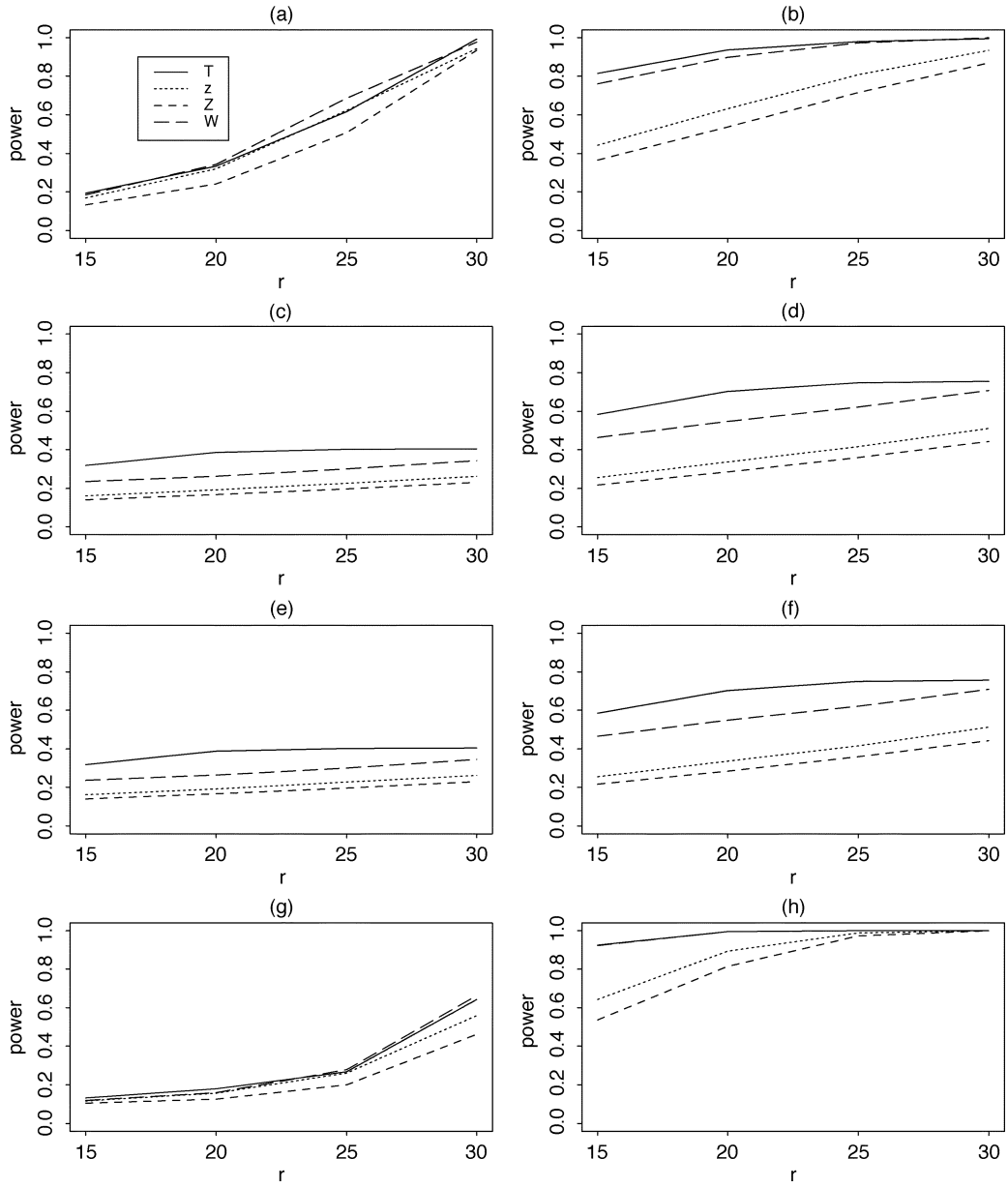

Fig. 2. Power comparison: monotone increasing hazard alternatives at 10% when the sample size is 30. (a) Uniform; (b) Weibull: shape 2; (c) Gamma: shape 1.5; (d) Gamma: shape 2; (e) Chi-square: df 3; (f) Chi-square: df 4; (g) Beta: shapes 1 and 2; (h) Beta: shapes 2 and 1.

determining the window size m which depends on n, r, and the α, we define the optimal window size m to be one which gives minimum critical points in the sense of [6]. However, we find from the simulated percentage points that the optimal window size m varies much according to r rather than n, and does not vary much according to α, if $\alpha \leq 0.1$. In view of these results, our recommended values of m for different r are listed in Table I, regardless of the sample sizes. We list the percentage points of the proposed statistic for $n = 10(10)50$, and some r under the null hypothesis.

C. Power Results

There are lots of test statistics for exponentiality concerning uncensored data [2], [8], [10]–[12], but only some of them can be extended to the censored data. We consider here the test statistics of [4], [14] among them. [4] proposed two test statistics as

$$z = \left(\frac{12}{r-2}\right)^{\frac{1}{2}} \frac{\sum_{i=1}^{r-1}\left(i-\frac{r}{2}\right)Y_{i+1}}{\sum_{i=1}^{r-1} Y_{i+1}}$$

$$Z = z^2 + \left(\frac{5}{4(r+1)(r-2)(r-3)}\right)^{\frac{1}{2}} \times \frac{12\sum_{i=1}^{r-1}\left(i-\frac{r}{2}\right)^2 Y_{i+1} - r(r-2)\sum_{i=1}^{r-1} Y_{i+1}}{\sum_{i=1}^{r-1} Y_{i+1}}$$

where $Y_1 = nX_{(1:n)}$, and $Y_i = (n-i+1)(X_{(i:n)} - X_{(i-1:n)})$, $i = 2, \cdots, r$; and show that z & Z perform better than other test statistics for the censored data.

[14] proposed a test statistic as

$$W = \frac{\left(\sum_{i=1}^{r} Y_i\right)^2}{r\sum_{i=2}^{r+1}\sum_{j=2}^{r+1}\frac{min(i,j)-1}{r-min(i,j)+2}Y_{i-1}Y_{j-1}},$$

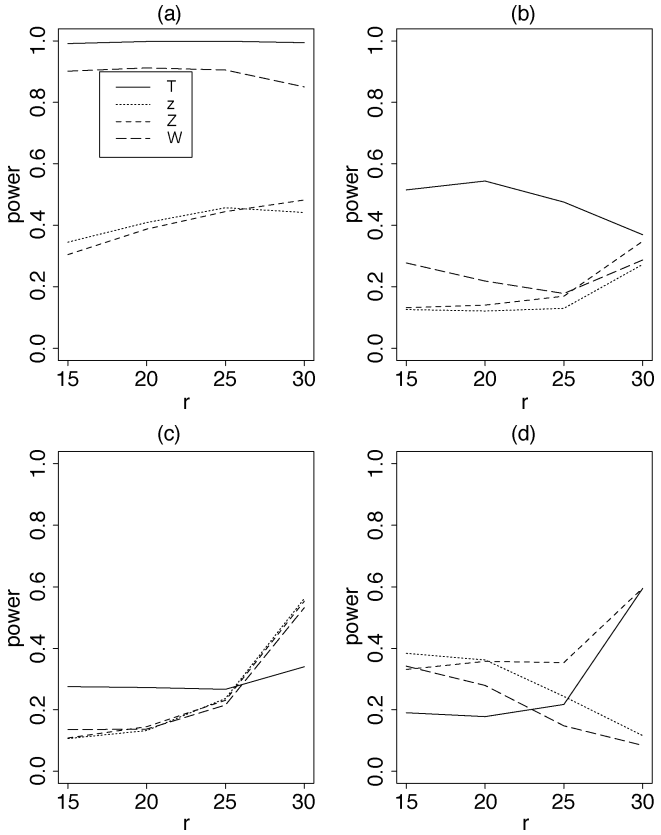

Fig. 3. Power comparison: nonmonotone hazard alternatives at 10% when the sample size is 30. (a) Log normal: shape 0.6; (b) Log normal: shape 1.0; (c) Log normal: shape 1.2; (d) Beta: shapes 0.5 and 1.

and showed that W has competing performance with z & Z for the censoreddata.

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

1) Monotone decreasing hazard: Gamma (shape parameter 0.5), Weibull (shape parameter: 0.5, 0.8), Chi-square (degree of freedom, 1)
2) Monotone increasing hazard: Uniform, Gamma (shape parameter: 1.5, 2), Weibull distribution (shape parameter 2), Chi-square (degree of freedom: 3, 4), Beta (shape parameters: 1 and 2, 2 and 1)
3) Non-monotone hazard: Log normal (shape parameter: 0.6, 1.0, 1.2), Beta (shape parameters: 0.5 and 1).

We consider here the sample size to be 30, and draw conclusions, but the similar conclusions can be made for different sample sizes if we consider r/n. We made 10 000 Monte Carlo simulations for $n = 30$ to estimate the powers of our proposed test statistic, and the competing test statistics. The simulation results are summarized in Figs. 1, 2, and 3.

We can see from the figures that any test statistic does not beat others against all alternatives, but it is notable that the proposed test statistic shows better powers than the competing test statistics against the alternatives with monotone increasing hazard functions, which applies to many real-life applications.

ACKNOWLEDGMENT

The author is pleased to thank an anonymous referee and the Associate Editor, Professor J.C. Lu, for their comments which improved the earlier version of this paper.

REFERENCES

[1] I. Arizono and H. Ohta, "A test for normality based on Kullback-Leibler information," *American Statistician*, vol. 43, pp. 20–23, 1989.
[2] S. Ascher, "A survey of tests for exponentiality," *Communications in Statistics. Theory and Methods*, vol. 19, pp. 1811–1825, 1990.
[3] A. W. Bowman, "Density based tests for goodness-of-fit," *J. Statistical Computation and Simulation*, vol. 40, pp. 1–13, 1992.
[4] C. W. Brain and S. S. Shapiro, "A regression test for exponentiality: censored, complete samples," *Technometrics*, vol. 25, pp. 69–76, 1983.
[5] E. J. Dudewicz and E. C. van der Meulen, "Entropy-based tests of uniformity," *J. Amer. Statist. Assoc.*, vol. 76, pp. 967–974, 1981.
[6] N. Ebrahimi *et al.*, "Testing exponentiality based on Kullback-Leibler information," *J. Royal Statist. Soc. B*, vol. 54, pp. 739–748, 1992.
[7] A. Foldes, L. Rejto, and B. B. Windter, "Strong consistency of nonparametric estimators for randomly censored data," in *Handbook of Statistics*: North Holland, 1980.
[8] F. F. Gan and K. J. Koehler, "Goodness-of-fit tests based on P-P probability plots," *Tecnometrics*, vol. 32, pp. 289–303, 1990.
[9] D. V. Gokhale, "On entropy-based goodness-of-fit tests," *Computational Statistics and Data Analysis*, vol. 1, pp. 157–165, 1983.
[10] N. Henze, "A new flexible class of omnibus tests for exponentiality," *Communications in Statistics Theory and Methods*, pp. 115–133, 1993.
[11] W. C. M. Kallenberg and T. Ledwina, "Data driven smooth tests for composite hypothesis: comparisons of powers," *J. Statist. Comput. Simul.*, vol. 59, pp. 101–121, 1997.
[12] V. LaRiccia, "Smooth goodness of fit tests: a quantile function approach," *J. Amer. Statist. Assoc.*, vol. 86, pp. 427–431, 1991.
[13] S. Park, "The entropy of consecutive order statistics," *IEEE Trans. Inform. Theory*, vol. 41, pp. 2003–2007, 1995.
[14] M. Samanta and C. J. Schwarz, "The Shapiro-Wilk test for exponentiality based on censored data," *J. Amer. Statist. Assoc.*, vol. 83, pp. 528–531, 1988.
[15] C. E. Shannon, "A mathematical theory of communications," *Bell System Tech. J.*, vol. 27, pp. 379–423, 1948.
[16] S. Teitler *et al.*, "Maximum entropy, reliability distributions," *IEEE Trans. Rel.*, vol. 35, pp. 391–395, 1986.
[17] O. Vasicek, "A test for normality based on sample entropy," *J. Royal Statist. Soc. B*, vol. 38, pp. 730–737, 1976.
[18] K. M. Wong and S. Chan, "The entropy of ordered sequences and order statistics," *IEEE Trans. Inform. Theory*, vol. 36, pp. 276–284, 1990.

Sangun Park is Associate Professor in the Department of Applied Statistics, Yonsei University, South Korea. He received the M.S. in 1990 in Statistics from Iowa State University, Ames, and the Ph.D. in 1994 in Statistics from the University of Chicago. He is a member of the American Statistical Association, and the Institute of Mathematical Statistics. His research interests include statistical information theory, order statistics, censored models, and nonparametric entropy estimation.