Journal of the Korean Data & Information Science Society
2013, **24**(6), 1497–1501

http://dx.doi.org/10.7465/jkdi.2013.24.6.1497
한국데이터정보과학회지

On scaled cumulative residual Kullback-Leibler information†

Insung Hwang[1] · Sangun Park[2]

[12]Department of Applied Statistics, Yonsei University

Received 7 August 2013, revised 2 October 2013, accepted 7 October 2013

Abstract

Cumulative residual Kullback-Leibler (CRKL) information is well defined on the empirical distribution function (EDF) and allows us to construct a EDF-based goodness of fit test statistic. However, we need to consider a scaled CRKL because CRKL is not scale invariant. In this paper, we consider several criterions for estimating the scale parameter in the scale CRKL and compare the performances of the estimated CRKL in terms of both power and unbiasedness.

Keywords: Empirical distribution, exponential distribution, goodness of fit test.

1. Introduction

Kullback-Leibler (KL) information is defined for $f(x)$ and $g(x)$ being the reference distribution as

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

KL information is nonnegative and the equality to zero holds iff $f(x) = g(x)$. The sample estimate of $K(g : f)$ can be simply obtained as $K(g_n : f_n)$, but are not attainable for $f_n = dF_n$ and $g_n = dG_n$ where F_n and G_n are the empirical distribution functions.

Rao *et al.* (2004) introduced a cumulative residual entropy (CRE) as

$$CRE(F) = -\int_{-\infty}^{\infty} \bar{F}(x) \log \bar{F}(x) dx$$

where $\bar{F}$ is the survival function.

† This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2011-0029104).

[1] MS student, Department of Applied Statistics, Yonsei University, Seoul 120-749, Korea.

[2] Corresponding author : Professor,Department of Applied Statistics, Yonsei University, Seoul 120-749, Korea. E-mail : sangun@yonsei.ac.kr

Barapour and Rad (2012) suggested the cumulative residual KL information (CRKL) for the nonnegative random variable, which is well-defined on the empirical distribution function, as

$$CRKL(G:F) = \int_0^\infty \bar{G}(x)\log\frac{\bar{G}(x)}{\bar{F}(x)}dx + E(X) - E(Y).$$

CRKL is also nonnegative and has a characteristic property that the equality to 0 holds iff $F(x) = G(x)$. If we take $G(x)$ and $F(x)$ to be $F_n(x)$ and $F_\theta(x)$, respectively, where $F_n(x)$ is the empirical distribution function and $F_\theta(x)$ is an assumed parametric distribution, the CRKL can be written as follows:

$$CRKL_\theta = -\int_0^\infty \bar{F}_n(x)\log\bar{F}_\theta(x)dx + E_\theta(X) - \bar{x} - CRE(F_n)$$

where

$$CRE(F_n) = -\sum_{i=0}^{n}\frac{n-i}{n}\log\frac{n-i}{n}(x_{i+1:n} - x_{i:n}).$$

While KL information is location and scale invariant, but $CRKL_\theta$ is not scale invariant. Hence, we need to consider the scaled CRKL. In estimating the parameters in KL information, the minimum discrimination information (MDI) criterion (see Soofi, 2000) has been considered along with maximum likelihood estimator. However, concerning the CRKL, we need to first define the scaled CRKL and determine how to choose the parameter estimator.

In this paper, we introduce the scale adjustment parameter and consider the scaled CRKL as

$$sCRKL_{\lambda,\theta} = \frac{1}{\lambda}CRKL_\theta,$$

where λ is a scale adjustment parameter and θ is a scale parameter.

$sCRKL_{\hat{\lambda},\hat{\theta}}$ can be considered as a goodness of fit test statistic where $\hat{\lambda}$ and $\hat{\theta}$ are appropriate parameter estimators. We consider combinations of several estimators and compare the performances of the corresponding scaled CRKL's with Monte Carlo simulation.

2. Scaled cumulative residual KL information

It is well-known that $KL(g:f)$ is nonnegative and has the characterization property that the equality to zero holds iff $g(x) = f(x)$ almost everywhere. Some extensions of KL information have been studied by some authors including Park (2012), Park and Shin (2013) for the Type I censored distribution and Balakrishnan *et al.* (2007) for the Type II progressively censored distribution. Barapour and Rad (2012) recently suggested a cumulative residual KL information, an extension of KL information to the survival function, as

$$CRKL(G:F) = \int_0^\infty \bar{G}(x)\left(\frac{\bar{F}(x)}{\bar{G}(x)} - \log\frac{\bar{F}(x)}{\bar{G}(x)}dx - 1\right)dx. \tag{2.1}$$

Since $u - \log u - 1 \geq 0$, we can see that CRKL is nonnegative and the equality to zero holds iff $F(x) = G(x)$ almost everywhere.

For $G(x) = F_n(x)$ and $F(x) = F_\theta(x)$ where $F_n(x)$ is an empirical distribution function and θ is a scale parameter of $F(x)$, the CRKL can be written as

$$CRKL_\theta = -\int_0^\infty \bar{F}_n(x) \log \bar{F}_\theta(x) dx + E_\theta(X) - \bar{x} - CRE(F_n).$$

However, because $CRKL_\theta$ is not scale invariant, we need to consider a scaled CRKL.

We define the scaled CRKL as

$$sCRKL_{\lambda,\theta} = \frac{1}{\lambda} CRKL_\theta,$$

where λ is the scale adjustment parameter and θ is the scale parameter.

We note that $sCRKL_{\theta,\theta}$ has been considered in Baratpour (2012) and Park *et al.* (2012) but they considered different parameter estimators. Park (2013) actually considered $sCRKL_{\bar{x},\bar{x}}$.

For an exponential distribution, $f_\theta(x) = \exp(-x/\theta)/\theta$, $sCRKL_{\lambda,\theta}$ can be written as

$$sCRKL_{\lambda,\theta} = \frac{1}{\lambda}\left(\frac{\sum_{i=1}^n x_i^2/(2n)}{\theta} - CRE(F_n) + \theta - \bar{x}\right).$$

Baratpour and Rad (2012) considered $sCRKL_{\hat{\theta}_1,\hat{\theta}_1}$ where

$$\hat{\theta}_1 = \frac{\sum_{i=1}^n x_i^2/(2n)}{\bar{x}},$$

which satisfies the moment constraint, $2E(X^2) = E(X)^2$.

Park *et al.* (2012) considered $sCRKL_{\hat{\theta}_2,\hat{\theta}_2}$ where

$$\hat{\theta}_2 = \frac{2\sum_{i=1}^n x_i^2/(2n)}{\bar{x} + CRE(F_n)},$$

which minimizes $sCRKL_{\theta,\theta}$ (the minimum discriminant information (MDI) estimator; Soofi, 2000).

Here we consider four estimators for θ as

1. $e_1 = \bar{x}$
2. $e_2 = \sqrt{\frac{\sum_{i=1}^n x_i^2}{2n}}$
3. $e_3 = \frac{\sum_{i=1}^n x_i^2/(2n)}{\bar{x}}$
4. $e_4 = \frac{2\sum_{i=1}^n x_i^2/(2n)}{\bar{x}+CRE(F_n)}$.

The first one is the sample mean which is known to the best estimator. The second one is the method of moment estimator based on the second moment, which also minimizes $CRKL_\theta$. The third one is the estimator considered in Baratpour and Rad (2012), and the fourth one is the estimator considered in Park *et al.* (2012).

In Table 1, we tabulate the bias and MSE of each estimator based on 100,000 Monte Carlo simulated samples of size 20. We can confirm that e_1 shows the best performance and e_3 shows the worst performance.

Table 2.1 Bias and MSE for the chosen estimators under the standard exponential distribution ; n=20

$\hat{\theta}$	Bias	MSE
e_1	0.0009	0.0498
e_2	0.0295	0.0555
e_3	0.0492	0.0808
e_4	0.0041	0.0601

3. Power comparisons and unbiasedness

We will consider sixteen combinations, $sCRKL_{\hat{\lambda},\hat{\theta}}$, which include the test statistics in Barapour and Rad (2012), and Park *et al.* (2012). The critical values of sixteen test statistics for a sample of size 20 are obtained with the Monte Carlo simulation, where the simulation size is 100000, and are tabulated in Table 2. We note that the critical values of $sCRKL_{e_i,e_1}$ and $sCRKL_{e_i,e_3}$ for $i = 1, \cdots, 4$ are same because $CRKL_{e_1} = CRKL_{e_3}$. The critical value of $sCRKL_{e_i,\hat{\theta}}$ for given e_i is naturally minimized at $\hat{\theta} = e_2$ because e_2 minimizes $CRKL_\theta$. The critical value of $sCRKL_{\hat{\theta},\hat{\theta}}$ is naturally minimized at $\hat{\theta} = e_4$.

Table 3.1 Critical values of test statistics for $\alpha = 0.05$; n=20

$\hat{\lambda}/\hat{\theta}$	e_1	e_2	e_3	e_4
e_1	0.1398	0.1153	0.1398	0.1189
e_2	0.1484	0.1210	0.1484	0.1249
e_3	0.1612	0.1309	0.1612	0.1351
e_4	0.1396	0.1137	0.1396	0.1174

We obtain the powers against gamma and Weibull alternatives with the shape parameter values, 0.5 (decreasing failure rate) and 2 (increasing failure rate), lognormal alternative, and chi-square alternatives with df=1 and 4, for a sample of size 20. The results are summarized in Table 3. As we already noted, the powers of $sCRKL_{e_i,e_1}$ and $sCRKL_{e_i,e_3}$ for $i = 1, 2, 3, 4$ are same. We also need to note that the powers of $sCRKL_{e_i,e_j}$'s for $i, j = 2, 4$ are same because they are based on the same statistic $2\sum_{i=1}^n x_i^2/(2n)/(\bar{x} + CRE(F_n))$.

As we can see from Table 3, the powers vary much according to the combination of parameter estimators. $sCRKL_{e_1,e_1}$ shows best powers against alternatives with decreasing failure rate, while $sCRKL_{e_3,e_1}$ shows best powers against alternatives with increasing failure rate. Though e_3 show poor performances as a point estimator, $sCRKL$ estimated with e_3 shows good performances.

Table 3.2 Power estimates under 7 alternatives ($\alpha = 0.05$) ; n=20

Alternatives	(e_1, e_1)	(e_1, e_2)	(e_1, e_4)	(e_2, e_1)	(e_2, e_2)	(e_3, e_1)	(e_3, e_2)	(e_3, e_4)	(e_4, e_1)
W(0.5)	0.7819	0.7365	0.7339	0.6937	0.5999	0.5385	0.3392	0.3524	0.6979
W(2)	0.7675	0.6407	0.6487	0.8731	0.8056	0.9178	0.8804	0.8788	0.8759
G(0.5)	0.3538	0.3096	0.3067	0.2552	0.1903	0.1385	0.0641	0.0688	0.2581
G(2)	0.2110	0.1480	0.1521	0.3154	0.2565	0.3895	0.3392	0.3368	0.3188
LN(0,1)	0.3497	0.3074	0.3045	0.2531	0.1889	0.1372	0.0647	0.0693	0.2564
χ_1^2	0.2149	0.1505	0.1546	0.3184	0.2605	0.3912	0.3429	0.3405	0.3218
χ_4^2	0.1929	0.1923	0.1917	0.1677	0.1603	0.1309	0.1139	0.1159	0.1675

Next, we consider some local alternatives to check whether the test statistics are unbiased or not. We consider gamma and Weibull alternatives with shape parameter values 0.9 and

1.1 and obtain the powers under the alternatives for a sample of size 20 where the simulation size is 100,000. If the estimated power is below the significance level 5 %, we can suspect the unbiasedness. The results are tabulated in Table 4. In Table 4, we can see that $sCRKL_{e_1,e_1}$, $sCRKL_{e_3,e_1}$ tend to be biased, while the p-values of $sCRKL_{e_2,e_1}$, $sCRKL_{e_2,e_2}$ and $sCRKL_{e_4,e_1}$ are above 5 % against all local alternatives. Among these three test statistics, $sCRKL_{e_4,e_1}$ shows best performance, though $sCRKL_{e_4,e_4}$ ($sCRKL_{e_2,e_2}$) is based on minimizing $sCRKL_{\theta,\theta}$. In comparing test statistics, the unbiasedness property should be considered along with the power, which has been neglected in lots of past works.

Table 3.3 Power estimates under some local alternatives ($\alpha = 0.05$) ; n=20

Alternatives	(e_1,e_1)	(e_1,e_2)	(e_1,e_4)	(e_2,e_1)	(e_2,e_2)	(e_3,e_1)	(e_3,e_2)	(e_3,e_4)	(e_4,e_1)
W(0.9)	0.0927	0.0884	0.0876	0.0685	0.0622	0.0437	0.0377	0.0383	0.0689
W(1.1)	0.0488	0.0449	0.0455	0.0697	0.0674	0.0869	0.0857	0.0851	0.0698
G(0.9)	0.0640	0.0634	0.0631	0.0526	0.0507	0.0407	0.0398	0.0399	0.0526
G(1.1)	0.0474	0.0459	0.0462	0.0567	0.0559	0.0646	0.0648	0.0644	0.0568

4. Conclusion

In this paper, we considered some estimates of the scaled CRKL and study the performance as a goodness of the fit test statistic for an exponential distribution, which include the test statistics in Barapour and Rad (2012), and Park *et al.* (2012). As a result, we found that the performance of the sCRKL test statistic varies much according to the chosen parameter estimators. The sCRKL test statistics plugged in with e_3 shows good performance, though e_3 shows poor performance as a point estimator. However, those test statistics tend to be biased against some local alternatives, which supports the usual convention that the MDI principle is considered in estimating the parameter in KL information. Then the similar discussion may be done for the Kolmogorov-Smirnov type goodness of fit test statistic.

Acknowledgements

The authors are grateful to three anonymous referees and the editor for providing some useful comments on an earlier version of this manuscript.

References

Balakrishnan, N., Rad, A. H. and Arghami, N. R. (2007). Testing exponentiality based on Kullback-Leibler information with progressively Type-II censored data. *IEEE Transactions on Reliability*, **56**, 349-356.

Barapour, S. and Rad, A. H. (2012). Testing goodness-of-fit for exponential distribution based on cumulative residual entropy. *Communications in Statistics-Theory and Methods*, 41, 1387-1396.

Park, S. (2012). Generalized Kullback-Leibler information and its extensions to censored and discrete cases. *Journal of the Korean Data & Information Science Society*, 23, 1223-1229.

Park, S. (2013). On censored cumulative residual Kullback-Leibler information and goodness-of-fit test with Type II censored data. Submitted to *Statistical Papers* (under 2nd revision).

Park, S., Rao, M. and Shin, D.W. (2012). On cumulative residual Kullback-Leibler information. *Statistics and Probability Letters*, 82, 2025-2032.

Park, S. and Shin, M. (2013). Kullback-Leibler information of Type I censored variable and its application. To appear in *Statistics*.

Rao, M., Chen, Y., Vemuri, B.C. and Wang, F. (2004). Cumulative residual entropy: A new measure of information. *IEEE Transactions on Information Theory*, 50, 1220-1228.

Soofi, E. S. (2000). Principal information theoretic approaches. *Journal of the American Statistical Association*, 95, 1349-1353.