

Kullback-Leibler Information of Consecutive Order Statistics

Ilmun Kim^{1,a}, Sangun Park^b

^aDepartment of Statistics, Carnegie Mellon University, USA

^bDepartment of Applied Statistics, Yonsei University, Korea;

Abstract

A calculation of the Kullback-Leibler information of consecutive order statistics is complicated because it depends on a multi-dimensional integral. Park (2014) discussed a representation of the Kullback-Leibler information of the first r order statistics in terms of the hazard function and simplified the r -fold integral to a single integral. In this paper, we first express the Kullback-Leibler information in terms of the reversed hazard function. Then we establish a generalized result of Park (2014) to an arbitrary consecutive order statistics. We derive a single integral form of the Kullback-Leibler information of an arbitrary block of order statistics; in addition, its relation to the Fisher information of order statistics is discussed with numerical examples provided.

Keywords: cross entropy, entropy, Fisher information, Kullback-Leibler information, left-censoring, order statistics, reversed hazard function

1. Introduction

Suppose that a random variable X has the distribution function $F(x)$ and continuous density function $f(x)$. The differential entropy of X is defined as the expected value of $-\log f(x)$ and plays a vital role to measure the uncertainty in a random variable. Teitler *et al.* (1986) expressed the entropy in terms of the hazard function as

$$\mathcal{E}(X) = 1 - \int_{-\infty}^{\infty} f(x) \log h(x) dx. \quad (1.1)$$

The entropy of order statistics has been carefully studied by many authors including Wong and Chen (1990), Park (1995) and Ebrahimi *et al.* (2004). Park (2005) investigated a single integral representation of the entropy of the first r order statistics in terms of the hazard function as

$$\mathcal{E}_{1 \dots r:n} = r - \sum_{i=1}^r \log(n - i + 1) - \int_{-\infty}^{\infty} \log h(x) \sum_{i=1}^r f_{i:n}(x) dx. \quad (1.2)$$

The Kullback-Leibler (KL) information as a measure of discrepancy between two density functions is defined as

$$\text{KL}(f : g) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} dx, \quad (1.3)$$

¹ Corresponding author: Department of Statistics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, USA.
E-mail: ilmunk@andrew.cmu.edu

which can be represented in terms of the relative risk (Park and Shin 2014). Based on this, Park (2014) expressed the KL information of the first r order statistics in terms of the relative risk and studied its relation with the Fisher information of order statistics.

The main goal of this article is to provide a general extension of the result in Park (2014) to an arbitrary block of order sequence, namely $X_{r:n} \leq \cdots \leq X_{s:n}$. We first express the entropy of order statistics in terms of the (reversed) hazard function and extend this result to the KL information of an arbitrary consecutive order sequence. We also explore its relation with the Fisher information of order statistics. Finally, some numerical examples are provided in Section 5.

2. Entropy in Order Statistics

If we consider the reversed hazard function $\lambda(x) = f(x)/F(x)$ rather than the hazard function in (1.1), the entropy of a single random variable can be represented as

$$\mathcal{E}_{1:1} = 1 - \int_{-\infty}^{\infty} f(x) \log \lambda(x) dx. \quad (2.1)$$

Thus, we can obtain a simple representation of the entropy of the last order statistic as

$$\mathcal{E}_{n:n} = -\log n + 1 - \int_{-\infty}^{\infty} f_{n:n}(x) \log \lambda(x) dx. \quad (2.2)$$

Suppose we are interested in the entropy of the sequence of order statistics $X_{1:n} \leq \cdots \leq X_{n:n}$. Then based on the Theorem 1 in Wong and Chen (1990), the entropy of $X_{1:n} \leq \cdots \leq X_{n:n}$ is simplified as

$$\mathcal{E}_{1 \dots n:n} = -\log n! + n\mathcal{E}_{1:1}. \quad (2.3)$$

Park (2005) derived a single integral representation of $\mathcal{E}_{1 \dots r:n}$, which is a general result of (2.3). Similarly, we have the identity of the entropy of the last $s+1$ order statistics as

Lemma 1.

$$\mathcal{E}_{s \dots n:n} = (n-s+1) - \sum_{i=s}^n \log i - \int_{-\infty}^{\infty} \log \lambda(x) \sum_{i=s}^n f_{i:n}(x) dx. \quad (2.4)$$

Proof: A slight change in the proof of the Theorem 2.1 in Park (2005) gives the statement. In detail, the Markov chain property of order statistics provide the identity $\mathcal{E}_{s \dots n:n} = \mathcal{E}_{n:n} + \mathcal{E}_{n-1|n:n} + \cdots + \mathcal{E}_{s|s+1:n}$. In addition, since $f_{s|s+1:n}$ can be considered as the pdf of the last order statistic among s sample as

$$f_{s|s+1:n}(x_s|x_{s+1}) = s \frac{f(x_s)}{F(x_{s+1})} \left[\frac{F(x_s)}{F(x_{s+1})} \right]^{s-1}, \quad (2.5)$$

we have from (2.2) that

$$\mathcal{E}_{s|s+1:n} = -\log s + 1 - \int_{-\infty}^{\infty} f_{s:n}(x) \log \lambda(x) dx. \quad (2.6)$$

Hence, we can obtain the required result. \square

The conditional entropy $\mathcal{E}_{r+1 \dots n|r:n}$ is equivalent to $\mathcal{E}_{1 \dots n:n} - \mathcal{E}_{1 \dots r:n}$ as noted in Park (1995). In analogy to this, we can obtain the identity $\mathcal{E}_{1 \dots s-1|s:n} = \mathcal{E}_{1 \dots n:n} - \mathcal{E}_{s \dots n:n}$ and we then have the representation of the conditional entropy in terms of the (reversed) hazard function as follows.

Lemma 2.

$$\mathcal{E}_{1 \dots s-1|s;n} = (s - 1) - \sum_{i=1}^{s-1} \log i - \int_{-\infty}^{\infty} \log \lambda(x) \sum_{i=1}^{s-1} f_{i:n}(x) dx, \tag{2.7}$$

$$\mathcal{E}_{r+1 \dots n|r;n} = (n - r) - \sum_{i=r+1}^n \log(n - i + 1) - \int_{-\infty}^{\infty} \log h(x) \sum_{i=r+1}^n f_{i:n}(x) dx. \tag{2.8}$$

Proof: Since $\mathcal{E}_{1 \dots n;n} = n\mathcal{E}_{1;1} - \log n!$ and $\sum_{i=1}^n f_{i:n}(x) = nf(x)$ noted in Wong and Chen (1990), we obtain the relations by considering (2.2) and (2.4). \square

From the identity $\mathcal{E}_{s \dots r;n} = \mathcal{E}_{1 \dots r;n} + \mathcal{E}_{s \dots n;n} - \mathcal{E}_{1 \dots n;n}$ in Park (1995), we can express the entropy of an arbitrary block of order statistics $\mathcal{E}_{r \dots s;n}$ in terms of the (reversed) hazard function as Lemma 3.

Lemma 3.

$$\mathcal{E}_{s \dots r;n} = C_{n,r,s} + n \int_{-\infty}^{\infty} f(x) F_{r;n-1}(x) \log h(x) dx - n \int_{-\infty}^{\infty} f(x) F_{s-1;n-1}(x) \log \lambda(x) dx, \tag{2.9}$$

where $C_{n,r,s} = (r - s + 1) - \sum_{i=1}^r \log(n - i + 1) - \sum_{i=s}^n \log i + \log n!$.

Proof: If we consider the relations $\sum_{i=r+1}^n f_{i:n}(x) = nf(x)F_{r;n-1}(x)$ and $\sum_{i=1}^r f_{i:n}(x) = nf(x)(1 - F_{r;n-1}(x))$, the result follows. \square

3. KL Information in Terms of the Reversed Hazard Rate

Park and Shin (2014) investigated a representation of the KL information in terms of the relative risk as

$$KL(f : g) = \int_{-\infty}^{\infty} f(x) \left\{ \frac{h_g(x)}{h_f(x)} - \log \frac{h_g(x)}{h_f(x)} - 1 \right\} dx \tag{3.1}$$

and applied this result to a goodness-of-fit test in Type I censored case. In many practical situations, such as when analyzing left censored data, one may be interested in the reversed relative risk (See, Gupta *et al.* (2004) and the references therein). Hence, we shall express the KL information in terms of the reversed relative risk as follows.

Lemma 4.

$$KL(f : g) = \int_{-\infty}^{\infty} f(x) \left\{ \frac{\lambda_g(x)}{\lambda_f(x)} - \log \frac{\lambda_g(x)}{\lambda_f(x)} - 1 \right\} dx. \tag{3.2}$$

Proof: By using the integration by parts, we have

$$\int_{-\infty}^{\infty} f(x) \log g(x) dx = \int_{-\infty}^{\infty} f(x) \log \frac{g(x)}{G(x)} dx + \int_{-\infty}^{\infty} f(x) \log G(x) dx \tag{3.3}$$

$$= \int_{-\infty}^{\infty} f(x) \log \lambda_g(x) dx - \int_{-\infty}^{\infty} f(x) \frac{\lambda_g(x)}{\lambda_f(x)} dx. \tag{3.4}$$

Since, $KL(f : g) = \int_{-\infty}^{\infty} f(x) \log f(x) dx - \int_{-\infty}^{\infty} f(x) \log g(x) dx$, the result follows by considering (2.1) and (3.4). \square

Park and Shin (2014) also derived the Fisher information expression in Type I censored variable $\min(X, C)$ based on the relation between the Fisher information and the KL information. Followed by this result, the Fisher information in leftly censored variable can be discussed. If we consider the leftly censored variable $\max(C, X)$ where C is the censoring point, the KL information can be adjusted as

$$\text{KL}_C(f : g) = F(C) \log \frac{F(C)}{G(C)} + \int_C^\infty f(x) \log \frac{f(x)}{g(x)} dx \quad (3.5)$$

$$= \int_C^\infty f(x) \left\{ \frac{\lambda_g(x)}{\lambda_f(x)} - \log \frac{\lambda_g(x)}{\lambda_f(x)} - 1 \right\} dx. \quad (3.6)$$

Since the Fisher information represents the curvature of the KL information, we obtain the leftly censored Fisher information by using the second order expansion of the KL information as follows.

$$\text{KL}_C(f(x; \theta) : f(x; \theta + \Delta\theta)) \approx \frac{1}{2}(\Delta\theta)^2 I_C(\theta), \quad (3.7)$$

where $I_C(\theta) = \int_C^\infty (\partial/\partial\theta \log \lambda(x; \theta))^2 f(x; \theta) dx$. Therefore, we can derive the same result in Gupta *et al.* (2004) for $C = -\infty$.

4. KL Information of Consecutive Order Statistics

If we consider the discrepancy between $f_{1:n}(x)$ and $g_{1:n}(x)$, then from the result of (3.1), we have

$$\text{KL}_{1:n}(f : g) = E_{f_{1:n}} \left[\frac{h_g(x)}{h_f(x)} - \log \frac{h_g(x)}{h_f(x)} - 1 \right]. \quad (4.1)$$

If we are interested in the difference between $f_{n:n}(x)$ and $g_{n:n}(x)$, we can derive a simple expression of $\text{KL}_{n:n}(f : g)$ using (3.2) as

$$\text{KL}_{n:n}(f : g) = E_{f_{n:n}} \left[\frac{\lambda_g(x)}{\lambda_f(x)} - \log \frac{\lambda_g(x)}{\lambda_f(x)} - 1 \right]. \quad (4.2)$$

Park (2014) provided the representation of the KL information of the first r order statistics so a general result of (4.1) as

$$\text{KL}_{1 \dots r:n}(f : g) = \int_{-\infty}^\infty \left\{ \frac{h_g(x)}{h_f(x)} - \log \frac{h_g(x)}{h_f(x)} - 1 \right\} \sum_{i=1}^r f_{i:n}(x) dx. \quad (4.3)$$

It is now worth finding a representation of the KL information of the last order sequence. First, we provide the following identities.

Lemma 5.

$$\mathcal{CE}_{s \dots n:n} = - \int_{-\infty}^\infty f_{s \dots n:n}(x) \log g_{s \dots n:n}(x) dx \quad (4.4)$$

$$= - \sum_{i=s}^n \log i - \int_{-\infty}^\infty \left\{ \log \lambda_g(x) - \frac{\lambda_g(x)}{\lambda_f(x)} \right\} \sum_{i=s}^n f_{i:n}(x) dx, \quad (4.5)$$

$$\mathcal{CE}_{1 \dots r:n} = - \int_{-\infty}^\infty f_{1 \dots r:n}(x) \log g_{1 \dots r:n}(x) dx \quad (4.6)$$

$$= - \sum_{i=1}^r \log(n - i + 1) - \int_{-\infty}^\infty \left\{ \log h_g(x) - \frac{h_g(x)}{h_f(x)} \right\} \sum_{i=1}^r f_{i:n}(x) dx. \quad (4.7)$$

Proof: Let us note that (4.5) can be expressed based on the Markov chain property of order statistics as

$$- \sum_{i=s}^n \int_{-\infty}^{\infty} \int_{-\infty}^{x_{i+1}} f_{i|i+1:n}(x_i) \log g_{i|i+1:n}(x_i) dx_i f_{i+1:n}(x_{i+1}) dx_{i+1}. \tag{4.8}$$

Since, $f_{i|i+1:n}$ can be considered to be the pdf $f(x)/F(x_{i+1})$ of the last order statistics among i sample, in view of (3.4), we have

$$\int_{-\infty}^{x_{i+1}} f_{i|i+1:n}(x_i) \log g_{i|i+1:n}(x_i) dx_i = \int_{-\infty}^{x_{i+1}} f_{i|i+1:n}(x_i) \left\{ \log i + \log \lambda_g(x_i) - \frac{\lambda_g(x_i)}{\lambda_f(x_i)} \right\} dx_i. \tag{4.9}$$

By putting (4.9) into (4.8), we get (4.5). The result of (4.6) can be simply derived by considering the relation $\mathcal{CE}_{1 \dots r:n}(f : g) = \text{KL}_{1 \dots r:n}(f : g) + \mathcal{E}_{1 \dots r:n}(f)$. \square

Finally, the KL information of the last $s + 1$ order statistics can be represented as

Theorem 1.

$$\text{KL}_{s \dots n:n}(f : g) = \int_{-\infty}^{\infty} \left\{ \frac{\lambda_g(x)}{\lambda_f(x)} - \log \frac{\lambda_g(x)}{\lambda_f(x)} - 1 \right\} \sum_{i=s}^n f_{i:n}(x) dx. \tag{4.10}$$

Proof: The proof is done by considering $\text{KL}_{s \dots n:n}(f : g) = \mathcal{CE}_{s \dots n:n}(f : g) - \mathcal{E}_{s \dots n:n}(f)$. \square

The following identities of the cross entropy of order statistics can be obtained as the similar way to the identity of the entropy in Wong and Chen (1990) and Park (1995).

$$\mathcal{CE}_{1 \dots n:n} = n\mathcal{CE}_{1:1} - \log n!, \tag{4.11}$$

$$\mathcal{CE}_{s \dots r:n} = \mathcal{CE}_{1 \dots r:n} + \mathcal{CE}_{s \dots n:n} - \mathcal{CE}_{1 \dots n:n}. \tag{4.12}$$

Now, we present the joint cross entropy of consecutive order statistics $X_{s:n} \leq \dots \leq X_{r:n}$ in terms of the (reversed) hazard function.

Lemma 6.

$$\begin{aligned} \mathcal{CE}_{s \dots r:n} &= D_{n,r,s} + n \int_{-\infty}^{\infty} f(x) F_{r:n-1}(x) \left\{ \log h_g(x) - \frac{h_g(x)}{h_f(x)} \right\} dx \\ &\quad - n \int_{-\infty}^{\infty} f(x) F_{s-1:n-1}(x) \left\{ \log \lambda_g(x) - \frac{\lambda_g(x)}{\lambda_f(x)} \right\} dx, \end{aligned} \tag{4.13}$$

where $D_{n,r,s} = C_{n,r,s} - (r - s + 1) = - \sum_{i=1}^r \log(n - i + 1) - \sum_{i=s}^n \log i + \log n!$.

Proof: The decomposition of the cross entropy in (4.12) gives the solution. \square

Based on the previous lemma, we can provide the main result of this paper in the next theorem.

Theorem 2. *The single integral representation of the KL information between $f_{s \dots r:n}$ and $g_{s \dots r:n}$ can be expressed in terms of the (reversed) hazard function as follows.*

$$\begin{aligned} \text{KL}_{s \dots r:n}(f : g) &= n \int_{-\infty}^{\infty} f(x) F_{s-1:n-1}(x) \left\{ \frac{\lambda_g(x)}{\lambda_f(x)} - \log \frac{\lambda_g(x)}{\lambda_f(x)} - 1 \right\} dx \\ &\quad - n \int_{-\infty}^{\infty} f(x) F_{r:n-1}(x) \left\{ \frac{h_g(x)}{h_f(x)} - \log \frac{h_g(x)}{h_f(x)} - 1 \right\} dx. \end{aligned} \tag{4.14}$$

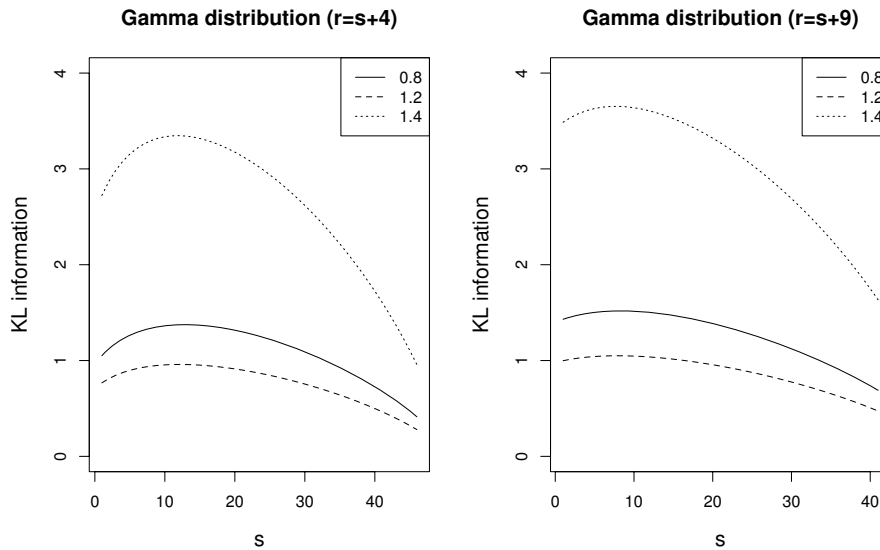


Figure 1: The Kullback-Leibler (KL) information plot for $\text{Gamma}(p, 1)$ and $\text{Exp}(1, 1)$.

Proof: Note that (2.9) can be denoted as

$$D_{n,r,s} + n \int_{-\infty}^{\infty} f(x)F_{r:n-1}(x) \{ \log h_f(x) - 1 \} dx - n \int_{-\infty}^{\infty} f(x)F_{s-1:n-1}(x) \{ \log \lambda_f(x) - 1 \} dx.$$

Accordingly, the definition of the KL information leads to the conclusion. □

Using the Theorem 2, the Fisher information in the block of order statistics can be represented in terms of the (reversed) hazard function. The result can be found in the following remark.

Remark 1.

$$\text{KL}_{s \dots r:n}(f(x; \theta) : f(x; \theta + \Delta\theta)) \approx \frac{1}{2}(\Delta\theta)^2 I_{s \dots r:n}(\theta), \tag{4.15}$$

where

$$I_{s \dots r:n}(\theta) = n \int_{-\infty}^{\infty} f(x)F_{s-1:n-1}(x) \left(\frac{\partial}{\partial \theta} \log \lambda_f(x) \right)^2 dx - n \int_{-\infty}^{\infty} f(x)F_{r:n-1}(x) \left(\frac{\partial}{\partial \theta} \log h_f(x) \right)^2 dx. \tag{4.16}$$

5. Numerical Examples

Due to limited amount of time or money, it is common to observe censored data in experiments and survival analysis. Especially when the number of (right) censored unit is predetermined but censoring time is random, we call this type of censoring as Type II censoring. More generally, let us assume we can only observe X_s, \dots, X_r consecutive sample out of n total items. Hence, the first $s - 1$ samples are left-censored and the last $n - r$ samples are right-censored. In this case, it might be an important issue to predetermine s and r before the experiment in order to enhance statistical power under a certain parametric model assumption. We provide simple numerical examples to illustrate this point based on the main result in the Theorem 2. Let us consider the $\text{gamma}(k, 1)$ and $\text{Weibull}(k, 1)$ distribution as

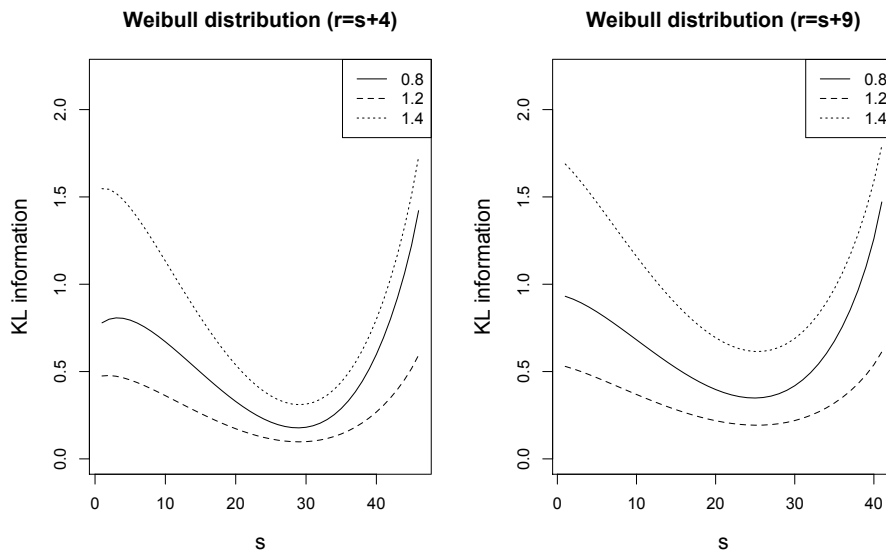


Figure 2: The Kullback-Leibler (KL) information plot for Weibull($p, 1$) and Exp(1, 1).

$f(x)$ and the standard exponential distribution as $g(x)$ in (4.14). We further consider the three different shape parameters in the gamma and Weibull distribution to be $k = 0.8, 1.2,$ and 1.4 . For the following censoring schemes with the length of the order statistics 5 and 10: $(s = 1, r = 5, n = 50), (s = 2, r = 6, n = 50), \dots, (s = 46, r = 50, n = 50)$ and $(s = 1, r = 10, n = 50), (s = 2, r = 11, n = 50), \dots, (s = 41, r = 50, n = 50)$, we present the KL information plots in Figures 1 and 2. From the results, we note that there is no monotonic change in the KL information while increasing the start point s in both cases. A starting point around 10 is suitable for the gamma distribution against the standard exponential distribution since it provides more information. However, we can get more information as s stays away from the middle point for the Weibull distribution. In addition, we can check a logical consequence that the KL information is increasing in the length of the order statistics and indicates that we have more information as the sample size increases.

6. Conclusion

In this paper, we provided a representation of the KL information of an arbitrary consecutive order statistics in terms of the (reversed) hazard function. Hence, we generalized the main result in Park (2014) who provided the expression of the KL information of the first r order statistics in terms of relative risk. We also discussed the representation of the Fisher information in order statistics in terms of (reversed) hazard function based on the relation between the KL information and the Fisher information. Afterwards, we considered the order statistics from the gamma and Weibull distribution to illustrate the major point of this paper.

We hope that some related topics can be dealt with in future studies. For example, up to date, there have been numerous studies about goodness-of-fit tests with censored data based on KL information (Balakrishnan *et al.*, 2007; Park, 2005; Park and Shin, 2014; Rad *et al.*, 2011). However, a goodness-of-fit test for data including both left and right censoring has not been done. We expect that this can be conducted using the results provided in this paper.

Acknowledgement

The authors are grateful to the editors and an associate editor and two anonymous referees for their careful and productive comments.

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**, 301–307.
- Ebrahimi, N., Soofi, E. S. and Zahedi, H. (2004). Information properties of order statistics and spacings, *IEEE Transactions on Information Theory*, **50**, 177–183.
- Gupta, R. D., Gupta, R. C. and Sankaran, P. G. (2004). Some characterization results based on factorization of the (reversed) hazard rate function, *Communications in Statistics - Theory and Methods*, **33**, 3009–3031.
- Park, S. (1995). The entropy of consecutive order statistic, *IEEE Transactions on Information Theory*, **41**, 2003–2007.
- Park, S. (2005). Testing exponentiality based on the Kullback-Leibler information with the type II censored data, *IEEE Transactions on Reliability*, **54**, 22–26.
- Park, S. (2014). On Kullback-Leibler information of order statistics in terms of the relative risk, *Metrika*, **77**, 609–616.
- Park, S. and Shin, M. (2014). Kullback-Leibler information of a censored variable and its applications, *Statistics*, **48**, 756–765.
- Rad, A. H., Yousefzadeh, F. and Balakrishnan, N. (2011). Goodness-of-fit test based on Kullback-Leibler information for progressively Type-II censored data, *IEEE Transactions on Reliability*, **60**, 570–579.
- Teitler, S., Rajagopal, A. K. and Ngai, K. L. (1986). Maximum entropy and reliability distributions, *IEEE Transactions on Reliability*, **35**, 391–395.
- Wong, K. M. and Chen, S. (1990). The entropy of ordered sequences and order statistics, *IEEE Transactions on Information Theory*, **36**, 276–284.

Received July 17, 2015; Revised August 20, 2015; Accepted August 20, 2015