

Generalized Kullback-Leibler information and its extensions to censored and discrete cases[†]

Sangun Park¹

¹Department of Applied Statistics, Yonsei University

Received 11 October 2012, revised 29 October 2012, accepted 5 November 2012

Abstract

In this paper, we propose a generalized Kullback-Leibler (KL) information for measuring the distance between two distribution functions where the extension to the censored case is immediate. The generalized KL information has the nonnegativity and characterization properties, and its censored version has the additional property of monotonic increase. We also extend the discussion to the discrete case and propose a generalized censored measure which is comparable to Pearson's chi-square statistic.

Keywords: Censored variable, Fisher information, hazard function, relative risk.

1. Introduction

Suppose that a random variable X has a distribution function $F(x)$ with a continuous density function $f(x)$. The Kullback-Leibler (KL) information is a measure of distance between two continuous distribution functions $f(x)$ and $g(x)$ in terms of $g(x)$, which is defined as

$$KL(g : f) = \int_{-\infty}^{\infty} g(x) \log \frac{g(x)}{f(x)} dx. \quad (1.1)$$

(1.1) can be written again as

$$KL(g : f) = \int_{-\infty}^{\infty} g(x) \left(\frac{f(x)}{g(x)} - \log \frac{f(x)}{g(x)} - 1 \right) dx. \quad (1.2)$$

Since $u - \log u - 1 \geq 0$ for $u > 0$ and the equality holds iff $u = 1$, it can be easily shown that (1.2) has nonnegativity and characterization properties. Hence, we can propose a generalized KL information between two distribution functions, which keeps the nonnegativity and characterization properties, as

$$D(\alpha_G : \alpha_F) = \int_{-\infty}^{\infty} g(x) \left(\frac{\alpha_F(x)}{\alpha_G(x)} - \log \frac{\alpha_F(x)}{\alpha_G(x)} - 1 \right) dx$$

[†] This work was supported in part by the Fund for Supporting Basic Science Research at the Department of Applied Statistics of Yonsei University.

¹ Professor, Department of Applied Statistics, Yonsei University, Seoul 120-749.
E-mail: sangun@yonsei.ac.kr.

where $\alpha_F(x)$ is an arbitrary positive continuous function of $F(x)$.

$D(\alpha_G : \alpha_F)$ is easily extendable to the censored case with still keeping the nonnegativity and characterization properties. In this paper, we consider several examples of the proposed class of measures and compare them in terms of the second-order approximation. We also consider the censored case where the proposed measure has the instant extension. We also extend our discussion to the discrete case.

2. Generalized Kullback-Leibler information

Kullback-Leibler (KL) information is a measure of distance between two continuous distribution functions $f(x)$ and $g(x)$ in terms of $g(x)$, which is defined as

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

In view of (1.2), we can instantly propose a generalized KL information with nonnegativity and characterization properties as follows.

Theorem 2.1 Suppose that $\alpha_F(x)$ is an arbitrary positive continuous function of $F(x)$ such that $\alpha_F(x) = \alpha_G(x)$ iff $F(x) = G(x)$ almost everywhere.

1. $D(g : f)$, defined as

$$D(\alpha_G : \alpha_F) = \int_{-\infty}^{\infty} g(x) \left(\frac{\alpha_F(x)}{\alpha_G(x)} - \log \frac{\alpha_F(x)}{\alpha_G(x)} - 1 \right) dx,$$

has the nonnegativity and characterization properties.

2. Under the assumption that $\alpha_F(x; \theta)$ is second differentiable about θ ,

$$D(\alpha_{F_\theta} : \alpha_{F_{\theta+\Delta\theta}}) \approx \frac{1}{2}(\Delta\theta)^2 \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} \log \alpha_{F_\theta}(x) \right)^2 f_\theta(x) dx.$$

The proof of the first part has been already stated and the proof of the second part can be done by considering the second-order expansions of $\alpha_{F_{\theta+\Delta\theta}}(x)$.

Example 2.1 If we take $\alpha_F(x) = f(x)$, we have

$$D(g : f) = \int_{-\infty}^{\infty} g(x) \left(\frac{f(x)}{g(x)} - \log \frac{f(x)}{g(x)} - 1 \right) dx,$$

which is equal to the KL information. It is well-known (Kullback, 1959) that

$$D(f_\theta : f_{\theta+\Delta\theta}) \approx \frac{1}{2}(\Delta\theta)^2 \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} \log f(x; \theta) \right)^2 f(x; \theta) dx,$$

Example 2.2 If we take $\alpha_F(x)$ to be the hazard function, $h_F(x) = f(x)/(1 - F(x))$, we have

$$D(h_G : h_F) = \int_{-\infty}^{\infty} g(x) \left(\frac{h_F(x)}{h_G(x)} - \log \frac{h_F(x)}{h_G(x)} - 1 \right) dx,$$

which has been shown to be equal to the KL information (Park and Shin, 2012).

We also have

$$D(f_\theta; f_{\theta+\Delta\theta}) \approx \frac{1}{2}(\Delta\theta)^2 \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} \log h_F(x; \theta) \right)^2 f(x; \theta) dx.$$

Hence, it is expected that

$$\int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} \log h_F(x; \theta) \right)^2 f(x; \theta) dx = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} \log f(x; \theta) \right)^2 f(x; \theta) dx,$$

which has been shown by Efron and Johnstone (1990).

Example 2.3 If we take $\alpha_F(x)$ to be the reverse hazard function denoted by $\bar{h}_F = f(x)/F(x)$, we have

$$D(g : f) = \int_{-\infty}^{\infty} g(x) \left(\frac{f(x)/F(x)}{g(x)/G(x)} - \log \frac{f(x)/F(x)}{g(x)/G(x)} - 1 \right) dx$$

and we can show that $D(g : f)$ is equal to the KL information since we have, by integration by parts,

$$\int_{-\infty}^{\infty} g(x) \log \frac{F(x)}{G(x)} dx = \int_{-\infty}^{\infty} g(x) \left(\frac{f(x)/F(x)}{g(x)/G(x)} - 1 \right) dx.$$

Since we have

$$D(\bar{h}_{F_\theta}; \bar{h}_{F_{\theta+\Delta\theta}}) \approx \frac{1}{2}(\Delta\theta)^2 \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} \log \frac{f(x; \theta)}{F(x; \theta)} \right)^2 f(x; \theta) dx,$$

we can show

$$\int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} \log \frac{f(x; \theta)}{F(x; \theta)} \right)^2 f(x; \theta) dx = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} \log f(x; \theta) \right)^2 f(x; \theta) dx.$$

3. Extension to the censored case

Now we discuss the censored measure of distance, which is extension of $D(\alpha_G : \alpha_F)$ to the censored case. The following lemma without further proof says that the censored measure of distance with nonnegativity and characterization properties is instantly obtained and that it has the additional property of monotone increase.

Lemma 3.1 Suppose that $\alpha_F(x)$ is an arbitrary continuous function of $F(x)$ such that $\alpha_F(x) = \alpha_G(x)$ iff $F(x) = G(x)$ almost everywhere.

1. $D_{(-\infty, C)}(g : f)$, defined as

$$D_{(-\infty, C)}(\alpha_G : \alpha_F) = \int_{-\infty}^C g(x) \left(\frac{\alpha_F(x)}{\alpha_G(x)} - \log \frac{\alpha_F(x)}{\alpha_G(x)} - 1 \right) dx,$$

has the nonnegativity, characterization and monotone increase properties.

2. Under the assumption that $\alpha_F(x; \theta)$ is second differentiable about θ ,

$$D_{(-\infty, C)}(\alpha_{F_\theta}; \alpha_{F_{\theta+\Delta\theta}}) \approx \frac{1}{2}(\Delta\theta)^2 \int_{-\infty}^C \left(\frac{\partial}{\partial\theta} \log \alpha_F(x; \theta) \right)^2 f(x; \theta) dx.$$

Example 3.1 We can obtain the following censored measure of information by taking $\alpha_F(x) = f(x)$,

$$D_{(-\infty, C)}(g : f) = \int_{-\infty}^C g(x) \left(\frac{f(x)}{g(x)} - \log \frac{f(x)}{g(x)} - 1 \right) dx,$$

which is equal to $I(g : f, C)$ in Lim and Park (2007).

We have

$$D_{(-\infty, C)}(f_\theta; f_{\theta+\Delta\theta}) \approx \frac{1}{2}(\Delta\theta)^2 \int_{-\infty}^C \left(\frac{\partial}{\partial\theta} \log f(x; \theta) \right)^2 f(x; \theta) dx,$$

which does not represent the relation with Fisher information for the censored case.

Example 3.2 If we take $\alpha_F(x)$ to be the hazard function $h_F(x)$, we have

$$D_{(-\infty, C)}(h_G : h_F) = \int_{-\infty}^C g(x) \left(\frac{h_F(x)}{h_G(x)} - \log \frac{h_F(x)}{h_G(x)} - 1 \right) dx,$$

which can be written as

$$D_{(-\infty, C)}(g : f) = \int_{-\infty}^C g(x) \log \frac{g(x)}{f(x)} dx - (1 - G(C)) \log \frac{1 - G(C)}{1 - F(C)}.$$

We also have

$$D_{(-\infty, C)}(h_{F_\theta}; h_{F_{\theta+\Delta\theta}}) \approx \frac{1}{2}(\Delta\theta)^2 \int_{-\infty}^C \left(\frac{\partial}{\partial\theta} \log h(x; \theta) \right)^2 f(x; \theta) dx.$$

We need to note that

$$\int_{-\infty}^C \left(\frac{\partial}{\partial\theta} \log h(x; \theta) \right)^2 f(x; \theta) dx = \int_{-\infty}^C \left(\frac{\partial}{\partial\theta} \log f(x; \theta) \right)^2 f(x; \theta) dx + \frac{\left(\frac{\partial}{\partial\theta} F(C; \theta) \right)^2}{1 - F(C; \theta)},$$

which is the Fisher information in a Type I censored variable. Hence, $D_{(-\infty, C)}(h_G : h_F)$ is more informative than $D_{(-\infty, C)}(g : f)$ in the sense of achieving the Fisher information.

Example 3.3 We can obtain another censored measure of information by taking $\alpha_F(x)$ to be the reverse hazard function $\bar{h}_F(x) = f(x)/F(x)$:

$$D_{(-\infty, C)}(g : f) = \int_{-\infty}^C g(x) \left(\frac{f(x)/F(x)}{g(x)/G(x)} - \log \frac{f(x)/F(x)}{g(x)/G(x)} - 1 \right) dx,$$

which can be written as

$$D_{(-\infty, C)}(g : f) = \int_{-\infty}^C g(x) \log \frac{g(x)}{f(x)} dx - G(C) \log \frac{G(C)}{F(C)}.$$

However, we have

$$D_{(-\infty, C)}(\bar{h}_{F\theta}; \bar{h}_{F\theta+\Delta\theta}) \approx \frac{1}{2}(\Delta\theta)^2 \int_{-\infty}^C \left(\frac{\partial}{\partial\theta} \log \frac{f(x; \theta)}{F(x; \theta)} \right)^2 f(x; \theta) dx,$$

which does not represent the relation with Fisher information for the censoring case. We need to note that

$$\int_{-\infty}^C \left(\frac{\partial}{\partial\theta} \log \frac{f(x; \theta)}{F(x; \theta)} \right)^2 f(x; \theta) dx < \int_{-\infty}^C \left(\frac{\partial}{\partial\theta} \log h(x; \theta) \right)^2 f(x; \theta) dx.$$

While $D(g : f)$, $D(h_G : h_F)$ and $D(\bar{h}_G : \bar{h}_F)$ are all equivalent on $(-\infty, \infty)$, $D_{(-\infty, C)}(h_G : h_F)$ is more informative than the other two. In Figure 3.1, we compare these three censored measures of information for exponential distribution and Weibull distribution by varying the shape parameter. We see that $D_{(-\infty, C)}(h_G : h_F)$ is most sensitive to the departure over the whole range.

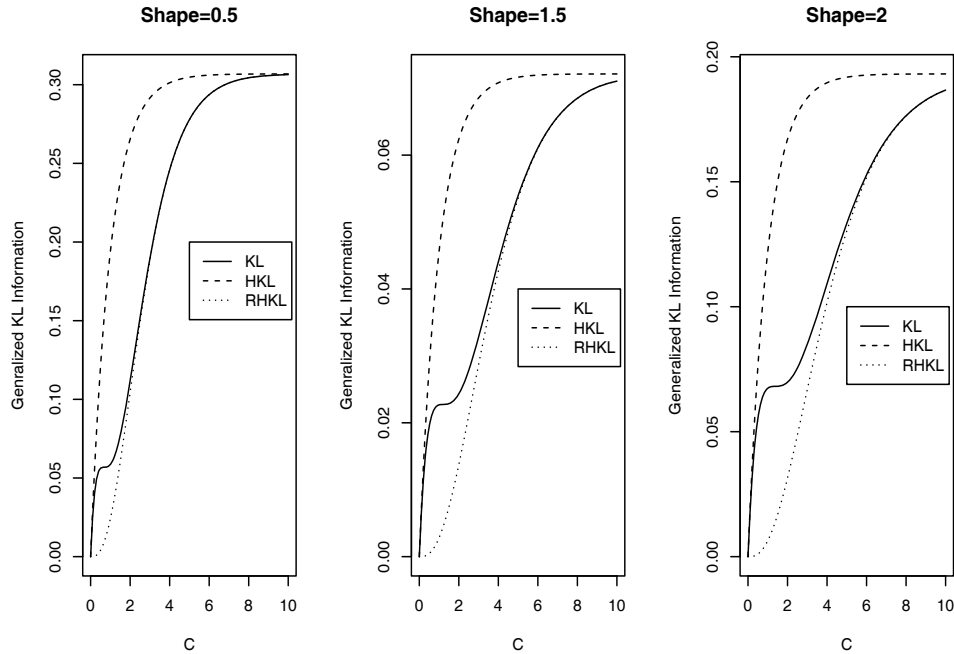


Figure 3.1 Generalized KL information : $KL(D_{(-\infty, C)}(g : f))$, $HKL(D_{(-\infty, C)}(h_G : h_F))$, $RHKL(D_{(-\infty, C)}(\bar{h}_G : \bar{h}_F))$

4. Discrete KL information

Suppose that we have two discrete random variables with $\mathbf{P} = (p_1, \dots, p_k)$ and $\mathbf{Q} = (q_1, \dots, q_k)$, respectively. Then the discrete KL information over whole categories is defined

as

$$K(\mathbf{P} : \mathbf{Q}) = \sum_{i=1}^k p_i \log \frac{p_i}{q_i}.$$

Its sample estimate,

$$2n\hat{K}(\mathbf{P} : \mathbf{Q}) = 2 \sum_{i=1}^k O_i \log \frac{O_i}{E_i},$$

is usually compared with Pearson's chi-square statistic $\sum_{i=1}^k (O_i - E_i)^2/E_i$ where O_i and E_i represent the numbers of observed and expected observations in i th category.

The amount of Pearson's chi-square statistic can be instantly decomposed into each category, respectively, with $(O_i - E_i)^2/E_i$. In this section, we present the similar decomposition of the KL information.

Theorem 2.1 can be extended to the discrete case as

$$D(\mathbf{P} : \mathbf{Q}) = \sum_{i=1}^k p_i \left(\frac{\alpha(q_i)}{\alpha(p_i)} - \log \frac{\alpha(q_i)}{\alpha(p_i)} - 1 \right). \quad (4.1)$$

We first consider $\alpha(p_i) = p_i$. Then $D(\mathbf{P} : \mathbf{Q})$ becomes equivalent to $K(\mathbf{P} : \mathbf{Q})$. The extension to any subcategory is given instantly as

$$K_{\mathbf{s}}(\mathbf{P} : \mathbf{Q}) = \sum_{i \in \mathbf{s}} \left(p_i \log \frac{p_i}{q_i} + q_i - p_i \right) \quad (4.2)$$

where \mathbf{s} is any subset of $\{1, 2, \dots, k\}$.

If we define e_i to be $p_i \log(p_i/q_i) + q_i - p_i$ in (4.2), then e_i has the nonnegativity and characterization properties since $e_i \geq 0$ and the equality holds iff $p_i = q_i$. Further we have $e_i \approx (p_i - q_i)^2/(2p_i)$. Hence, $2nK_c(\mathbf{P} : \mathbf{Q})$ can be taken to the decomposition of the KL information to c th category. (4.2) also gives us another example as follows.

Example 4.1 If we take $\alpha(p_i) = p_i^2$, then we have

$$\begin{aligned} K_{\mathbf{s}}^{(2)}(\mathbf{P} : \mathbf{Q}) &= \sum_{i \in \mathbf{s}} \left(2p_i \log \frac{p_i}{q_i} + \frac{q_i^2 - p_i^2}{p_i} \right) \\ &\approx 2 \sum_{i \in \mathbf{s}} \frac{(q_i - p_i)^2}{p_i} \end{aligned}$$

Hence, $nK_{\mathbf{s}}^{(2)}(\mathbf{P} : \mathbf{Q})/2$ is comparable to Pearson's chi-square statistic. However, we can show that $nK_c^{(2)}(\mathbf{P} : \mathbf{Q})/2$ is the average of $2nK_c(\mathbf{P} : \mathbf{Q})$ and Pearson chi-square statistic.

In Table 4.1, we take Mendel's famous experiment with peas (Cramer, 1955). We compare $2nK_c(\mathbf{P} : \mathbf{Q})$ and $nK_c^{(2)}(\mathbf{P} : \mathbf{Q})/2$ with Pearson's chi-square test statistic. $2nK_c(\mathbf{P} : \mathbf{Q})$ enables us to decompose the amount of KL information into subcategories and check the deviance in each category. We also note that $nK_c^{(2)}(\mathbf{P} : \mathbf{Q})/2$ always lies between $2nK_c(\mathbf{P} : \mathbf{Q})$ and Pearson's chi-square statistic.

Table 4.1 Comparison of $2nK_c(\mathbf{P} : \mathbf{Q})$'s and Pearson's χ^2 .

Phenotype	Round & Y.	Round & Gr.	Angular & Y.	Angular & Gr.	Sum
Observed #	315	108	101	32	556
Expected Pr.	9/16	3/16	3/16	1/16	1
Pearson χ^2	0.0162	0.1349	0.1013	0.2176	0.4700
$2nK_c(\mathbf{P} : \mathbf{Q})$	0.0161	0.1317	0.1035	0.2298	0.4812
$K_c^{(2)}(\mathbf{P} : \mathbf{Q})/2$	0.0161	0.1333	0.1024	0.2237	0.4756

5. Conclusions

We proposed a class of measures of distance which can be instantly extended to the censored case. All three measures of information with $\alpha_F(x)$'s, $f(x)$, $f(x)/(1 - F(x))$ and $f(x)/F(x)$, are all equivalent for $(-\infty, \infty)$, but the censored measure of information with $\alpha_F(x) = f(x)/(1 - F(x))$ is most informative for the right censored case in the sense of achieving the Fisher information. Finally, we discussed the extension to the discrete case and proposed appropriate censored measures which are comparable to Pearson's chi-square statistic.

References

- Cramer, H. (1955). *The elements of probability theory*, John Wiley and Sons, NY.
- David, H. A. and Nagaraja, H. N. (2003), *Order statistics*, 3rd Ed., John Wiley and Sons, NY.
- Efron, B. and Johnstone, I. (1990). Fisher information in terms of the hazard rate. *Annals of Statistics* **18**, 38–62.
- Kullback, S. (1959). *Information theory and statistics*, John Wiley and Sons, NY.
- Lim, J. and Park, S. (2007). Censored Kullback-Leibler information and goodness of fit test with Type II censored data. *Journal of Applied Statistics*, **34**, 1051-1064.
- Park, S. and Shin, M. (2012). Kullback-Leibler information of Type I censored variable and its application. Submitted.