The Korean Communications
in Statistics Vol. 6, No. 2, 1999
pp. 397-401

Recurrence Relations in the Fisher Information in Order Statistics[1]

Sangun Park[2]

Abstract

We first derive the Fisher information identity in order statistics in terms of the hazard rate by considering the Fisher information identity in terms of the hazard rate (Efron and Johnstone, 1990). Then we use the identity and show an interesting and useful result that some identities and recurrence relations for the Fisher information in order statistics can be directly obtained from those between the c.d.f.s of order statistics.

1. Introduction

Suppose that we have an independently and identically distributed (i.i.d.) sample of size n from a continuous probability density function (p.d.f.) $f(x;\theta)$ where θ is a real-valued scalar parameter. We will denote $X_{(r;n)}$ to be the t rh order statistic, $f_{r;n}$ to be its p.d.f. and $I_{1\cdots r;n}(\theta)$ to be the Fisher information about θ in a set of the first r order statistics. We assume that $f(x;\theta)$ satisfies some regularity conditions (Cox and Hinkley, 1974) such that the Fisher information per observation exists. We have many situations where we need to consider only a subset of consecutive order statistics (e.g. Type II censored data, trimmed sample), then how much we lose the information is an interesting and basic question. However, since $I_{1\cdots r;n}(\theta)$ is an r multiple integral, we had much difficulties in the calculation of $I_{1\cdots r;n}(\theta)$ in most cases.

Many authors have established lots of identities and recurrence relations between the moments of order statistics, which can be easily obtained from those between the c.d.f.s of order statistics. These results can be used in checking the accuracy of computation of the moments of order statistics and reducing the amount of direct computation of moments of order statistics. For the Fisher information in order statistics, Park (1996) recently derived some recurrence relations whose primary advantage, other than checking or reducing the computation, is that $I_{1\cdots r;n}(\theta)$ can be easily obtained as a linear combination of single integrals.

1) The author wishes to acknowledge the financial support of the Korea Research Foundation made in the program year of 1997.

2) Assistant Professor, Department of Applied Statistics, Yonsei University, Shinchon Dong 134, Seoul, Korea

In this paper, we provide a more general result that some relations between c.d.f.s of order statistics directly give some corresponding relations in the Fisher information in order statistics. Thus some recurrence relations in Park (1996) at last become some examples of this result. We provide some examples including those in Park (1996).

2. Main results

Efron and Johnstone (1990) studied an interesting identity,

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

where $h(x;\theta)$ is the hazard function.

Its direct application to the case of $X_{(1;n)}$ is given as

$$I_{1;n}(\theta)=\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\log h(x;\theta)\right)^2 dF_{1;n}(x;\theta).$$

We define $I_{r+1|r,n}(\theta|x_{(r,n)})$ to be the conditional information about θ in $X_{(r+1;n)}$ given $X_{(r,n)}=x_{(r,n)}$. Then we define $I_{r+1|r,n}(\theta)$ to be the average of the conditional information such that

$$I_{r+1|r,n}(\theta)=\int_{-\infty}^{\infty}I_{r+1|r,n}(\theta|x_{(r,n)})f_{r,n}dx_{(r,n)}.$$

We need the following lemma to prove our main results.

Lemma 2.1.

1. Markov property :

$$I_{r+1|r,n}(\theta)=I_{r+1|i\cdots r,n}(\theta)\ \text{ for all }\ i=1,\cdots r. \tag{2}$$

2. Decomposition of the Fisher information in order statistics :

$$I_{r,r+1;n}(\theta)=I_{r,n}(\theta)+I_{r+1|r,n}(\theta). \tag{3}$$

3. $I_{r+1|r,n}(\theta)$ in terms of the hazard rate :

$$I_{r+1|r,n}(\theta)=\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\log h(x;\theta)\right)^2 f_{r+1;n}(x;\theta)dx. \tag{4}$$

Proof. The proofs of the first two are in Park (1996), and the proof of the third one is as follows. Since $f_{r+1|r,n}$ is the density of the first order statistic of an \$n-r\$ i.i.d. sample from $f(x;\theta)/(1-F(x_{(r,n)};\theta))$, $x\geq x_{(r,n)}$, whose hazard function is $(n-r)h(x;\theta)$, we have, in view of (1),

$$I_{r+1|r:n}(\theta|x_{(r:n)}) = \int_{x_{(r:n)}}^{\infty} (\frac{\partial}{\partial\theta} \log h(x;\theta))^2 f_{r+1|r:n}(x|x_{(r:n)};\theta)dx.$$

Then $I_{r+1|r:n}(\theta)$ can be obtained as

$$I_{r+1|r:n}(\theta) = \int_{-\infty}^{\infty} (\frac{\partial}{\partial\theta} \log h(x;\theta))^2 f_{r+1:n}(x;\theta)dx.$$

Thus we have the result. □

Remark 2.1 Mehrotra et al (1979) have first considered $I_{1\cdots r:n}(\theta)$ where the direct expression of $I_{1\cdots r:n}(\theta)$ contains the score functions and hazard functions. They tried to remove the hazard functions in the direct expression and succeeded in expressing $I_{1\cdots r:n}(\theta)$ as sums of double integrals concerning only score functions. However, Lemma 2.1 says that $I_{1\cdots r:n}(\theta)$ can be expressed as sums of single integrals concerning only the hazard function as follows.

$$I_{1\cdots r:n}(\theta) = \int_{-\infty}^{\infty} (\frac{\partial}{\partial\theta} \log h(x;\theta))^2 (f_{1:n}(x;\theta) + \cdots + f_{r:n}(x;\theta))dx. \ \square$$

The following result is concerning relations for the Fisher information and those between the c.d.f.s of order statistics.

Theorem 2.1.
If

$$\sum c_{r,m} f_{r,m}(x;\theta) = 0,$$

where the summation is taken over all subsets m of $\{1,2,\cdots,n\}$ and over r, then

$$\sum c_{r,m} I_{r|r-1:m}(\theta) = 0.$$

Proof. If we let $g(X_{(r:n)})$ be $(\partial/\partial\theta)\log h(X_{(r:n)};\theta)$, then $I_{r|r-1:n}(\theta)$ is $E(g(X_{(r:n)}))$ by (4). Thus the relation between c.d.f.s of order statistics directly holds for $I_{r|r-1:n}(\theta)$. □

The following corollary can be obtained by using the dual principle for the Fisher information in (Park, 1996) and that between the c.d.f.s of order statistics (Balasubramanian and Balakrishnan, 1993).

Corollary 2.1.
If

$$\sum c_{r,m} f_{r,m}(x;\theta) = 0,$$

where the summation is taken over all subsets m of $\{1,2,\cdots,n\}$ and over r, then

$$\sum c_{r,m} I_{r|r+1:m}(\theta) = 0.$$

3. Some examples

For example, it is well known that

$$\sum_{i=1}^{n} f_{i:n}(x;\theta) = nf_{1:1}(x;\theta).$$

Then we directly have, by using Theorem 2.1,

$$\sum_{i=1}^{n} I_{i|i-1:n}(\theta) = nI_{1:1}(\theta),$$

which means the simple fact, $I_{1\cdots n:n}(\theta) = nI_{1:1}(\theta)$. We provide the relations in the Fisher information in order statistics corresponding to some relations between c.d.f.s of order statistics.

Relation 3.1.

$$\sum_{r=1}^{n-1} I_{1\cdots r:n}(\theta) = \frac{1}{2}\, n(n-1) I_{1:2}(\theta)$$

This relation can be obtained from the relation,

$$\sum_{r=1}^{n-1} (f_{1:n}(x;\theta) + \cdots + f_{r:n}(x;\theta)) = \frac{1}{2}\, n(n-1) f_{1:2}(x;\theta).$$

The following two relations, Relation 3.2 and 3.5, have been derived in Park (1996), while the proof is different.

Relation 3.2.

$$I_{1\cdots r:n-1}(\theta) = \frac{n-r-1}{n} I_{1\cdots r:n}(\theta) + \frac{r}{n} I_{1\cdots r+1:n}(\theta)$$

This relation can be obtained from the relation, established by Cole (1951),

$$f_{r:n-1} = \frac{n-r}{n} f_{r:n} + \frac{r}{n} f_{r+1:n}.$$

Relation 3.3.

$$I_{1\cdots r:n}(\theta) = \sum_{i=n-r+1}^{n} C_{i-2,\,n-r-1} C_{n,\,i} (-1)^{i-n+r-1} I_{1:i}(\theta)$$

This relation can be obtained from the relation, established by Srikantan (1962),

$$f_{r:n} = \sum_{i=n-r+1}^{n} (-1)^{i-n+r-1} C_{i-1,\,n-r} C_{n,\,i} f_{1:i}.$$

Remark 3.1. Relation 3.1 follows automatically if any one of Relation 3.2 and 3.3 is applied. Similar argument for the moments of order statistics has been discussed by Balakrishnan and Malik (1986). □

Relation 3.4.

$$C_{n,m} I_{1\cdots r;m}(\theta) = \sum_{i=0}^{n-m} C_{n-r-i-1,m-r-1} C_{r+i-1,i} I_{1\cdots r+i;n}(\theta)$$

This relation can be obtained from the relation, established by Sillito (1964),

$$C_{n,m} f_{r;m}(\theta) = \sum_{i=0}^{n-m} C_{n-r-i,m-r} C_{r+i-1,i} f_{r+i;n}(\theta).$$

Relation 3.5.

$$\sum_{i=1}^{n} \frac{1}{i} (I_{1\cdots i;n}(\theta) - I_{1\cdots i-1;n}(\theta)) = \sum_{i=1}^{n} \frac{1}{i} I_{1;i}(\theta)$$

This relation can be obtained from the relation, established by Joshi (1973),

$$\sum_{i=1}^{n} \frac{1}{i} f_{i;n} = \sum_{i=1}^{n} \frac{1}{i} f_{1;i}.$$

Remark 3.2. By the dual principle for the Fisher information in order statistics, we can instantly derive some recurrence relations for $I_{s\cdots n;n}(\theta)$ from Relation 3.1-3.5. □

References

[1] Balasubramanian, K. and Balakrishnan, N(1993). Duality principle in order statistics, *Journal of Royal Statistical Society* B, 55, 687-691.

[2] Balakrishnan, N. and Malik, H. J. (1985). Some general identities involving order statistics, *Communications in Statistics. Theory and Methods*, 14(2), 333-339.

[3] Cole, R. H. (1951). Relations between moments of order statistics, *Annals of Mathematical Statistics*, 22, 308-310.

[4] Cox, D. R. and Hinkley D. V. (1974). *Theoretical Statistics*, Chapman and Hall

[5] Efron, B and Johnstone, I. (1990). Fisher information in terms of the hazard rate, *Annals of Statistics*, 18, 38-62.

[6] Joshi, P. C. (1973). Two identities involving order statistics, *Biometrika,* 60, 428-429.

[7] Mehrotra, K. G., Johnson R. A. and Bhattacharyya, G. K. (1979). Exact Fisher information for censored samples and the extended hazard rate functions. *Communications in Statistics*. A, 15, 1493-1510.

[8] Park, S. (1996). Fisher information in order statistics, *Journal of American Statistical Association*, 91, 385-390.

[9] Sillito, G. P. (1964). Some relations between expectations of order statistics in samples of different sizes, *Biometrika*, 51, 259-262.

[10] Srikantan, K. S. (1962). Recurrence relations between the pdf's of order statistics, and some applications, *Annals of Mathematical Statistics*, 33, 169-177.