한국통계학회 논문집 제1권1호
1994년 12월 pp. 52-56

On Characterizing Distributions by Some Properties of the Distribution Truncated at the rth order Statistic

Sangun Park[1)]

Abstract

When we have an i.i.d. sample of size n from a continuous distribution, the distribution truncated on the left at the rth order statistic plays an important role in the theoretical analysis of the Type 2 censored data. The charaterization of distributions by the average of the conditional expectation and the average of the conditional information concerning the truncated distribution is studied here.

1. Introduction

Suppose that $X_{1:n}, \cdots, X_{n:n}$ be the order statistics of an i.i.d. sample of size n from a distribution functioin $F(x)$. We denote $\mu_{r:n}$ be the expectation of $X_{r:n}$ and $I_{r:n}(\theta)$ be the Fisher information in $X_{r:n}$ about θ. Let $F_{(x_{r:n},\infty)}$ be the truncated distribution on the left at $x_{r:n}$, i.e,$f(x)/(1-F(x_{r:n}))$. $F_{(x_{r:n},\infty)}$ has its importance in statistics, since it is closely related with the lost likelihood in the Type 2 censored sampling where only first r order statistics are observed. By Wald's principle, the lost likelihood of the Type 2 censored data can be considered as the likelihood of an i.i.d. sample of size $n-r$ from $F_{(x_{r:n},\infty)}$. This interpretation provides us a simple and straightforward way for some asymptotics based on the censored data given $r=np+o(n)$, $0<p<1$ and the derivation of the exact Fisher information in the censored data.(Park,1994 A B)

The characterization of distributions by moments has been studied by many authors. The classical type is the Hausdorff moment problem. Hoeffding (1953) considered the moments of order statistics and showed that under mild conditions the set $\bigcup_{n=1}^{\infty} M_n$ completely determines the parent distribution, where $M_n=\{\mu_{r:m}: r=1,\cdots,m;\ m=1,\cdots,n\}$. Chan (1967) transformed this problem to the Hausdorff moment problem relating the moments of the inverse function of F where $F^{-1}(u)=\inf\{x \mid F(x)\geq u\}$. The recurrence relation between moments of order statistics says that M_n can be replaced by its proper subset, as discussed by some authors including Kadane (1971), Arnold and Meeden (1975). We consider here the characterization of

1) c/o Prof. Yoon, Ki Jung's Office, Department of Applied Statistics, Yonsei University, Shinchon Dong 134, Seoul, KOREA

distributions by the average of the conditional expectation concerning $F_{(x_{r:n},\infty)}$. We also study the characterization of distributions by the average of the conditional information concerning the parametrized $F_{(x_{r:n},\infty)}$.

2. Main results

Let $\mu_{(x_{r:n},\infty)}$ be the average of the conditional expectation, where the conditional expectation, $E(X|X>x_{r:n})$, is

$$E(X|X>x_{r:n})=\int_{x_{r:n}}^{\infty} x\frac{f(x)}{1-F(x_{r:n})}dx. \tag{1}$$

We first consider the set $\{\mu_{(x_{1:n},\infty)};n=2,\cdots,\infty\}$.

Theorem 2.1 *If f is positive and continuous, the sequence $\{\mu_{(x_{1:n},\infty)};n=2,\cdots,\infty\}$ uniquely determines F.*

Proof. Since $\mu_{(x_{r:n},\infty)}$ is the average of (1) about $X_{r:n}$,

$$\mu_{(x_{1:n},\infty)}=\int_{-\infty}^{\infty}\{\int_{x_{1:n}}^{\infty}xf(x)dx\}nf(x_{1:n})(1-F(x_{1:n}))^{n-2}dx_{1:n},$$

by a probablity integral transformation, letting $F(x_{1:n})$ be u,

$$\mu_{(x_{1:n},\infty)}=\int_0^1\{\int_{F^{-1}(u)}^{\infty}xf(x)dx\}n(1-u)^{n-2}du.$$

Let $G(y)$ be another distribution. Then, if $\mu_{(x_{1:n},\infty)}=\mu_{(y_{1:n},\infty)}$, for $n=2,\cdots,\infty$,

$$\int_0^1\{\int_{F^{-1}(u)}^{\infty}xf(x)dx\}n(1-u)^{n-2}du=\int_0^1\{\int_{G^{-1}(u)}^{\infty}yg(y)dy\}n(1-u)^{n-2}du..$$

Thus, by the completeness of the Legendre polynomials on L_1 space,

$$\int_{F^{-1}(u)}^{\infty}xf(x)dx=\int_{G^{-1}(u)}^{\infty}yg(y)dy$$

If we differentiate both sides about u, we have

$$F^{-1}(u)=G^{-1}(u) \quad \text{for} \quad u\in(0,1).$$

Thus we have our result.

Corollary 2.1 *Lemma 2.1 is true for the set $\{\mu_{(x_{r:n},\infty)};n=r+1,\cdots,\infty\}$, where r is fixed.*

Proof. We prove this by showing the fact that if

$$\int_0^1 w(u)u^{r-1}(1-u)^{n-r-1} = \int_0^1 v(u)u^{r-1}(1-u)^{n-r-1}, \text{ for } n=r+1,\cdots,\infty, \tag{2}$$

then $w(u)=v(u)$ a.e.; (2) implies that

$$\int_0^1 w(u)u^{n} = \int_0^1 v(u)u^{n}, \text{ for } n=r+1,\cdots,\infty. \tag{3}$$

Muntz theorem says that $\{x^n; n=n_1,n_2,\cdots\}$ is L_1-complete,where n_i is a sequence of distinct positive real numbers, iff $\sum_{i=1}^{\infty} n_i=\infty$. Thus the result follows.

Example 2.1 The exponential distribution is the only one among the continuous distributions with a positive derivative on its support, whose $\mu_{(x_{1:n},\infty)}$ is $(n+1)/n$.

Ferguson (1967) found some continuous distributions for which the regression function, $E(X_{r:n}|X_{r+1:n})$, is linear. Wang and Srivastava (1980) found some distributions for which $E(Z_r|X_{r:n})$ is linear where $Z_r=\sum_{i=r+1}^{n}(X_{i:n}-X_{r:n})/(n-r)$. We find some distributions for which $\mu_{(x_{r:n},\infty)}$ is linear. Some parts in the proof of the following lemma comes from the detailed lines in the aforementioned papers.

Theorem 2.2 *If, for* F *which is continuous with a finite first moment,*

$$\mu_{(x_{r:n},\infty)}=\alpha+(\beta+1)\mu_{r:n}, \text{ for } n=r+1,\cdots,\infty,$$

then F *is exponential* $(\beta=0)$, *Pearson of type 1* $(0>\beta>-1)$,*and Pareto distribution* $(\beta>0)$.

Proof. By a simiar way as in the proof of Corollary 2.1, we have

$$E(X|X\geq x_{r:n})=\alpha+(\beta+1)x_{r:n}.$$

By (1),

$$\int_{x_{r:n}}^{\infty} xf(x)dx=(\alpha+(\beta+1)x_{r:n})(1-F(x_{r:n})). \tag{4}$$

The lefthand side is $\int_{x_{r:n}}^{\infty}\int_0^x dtf(x)dx$. Thus, by the Fubini's theorem, (4) can be written as

$$\int_{x_{r:n}}^{\infty}(1-F(t))dt=(\alpha+\beta x_{r:n})(1-F(x_{r:n})).$$

Then we can solve the differential equation to get our result.

Remark 2.1 This actually includes the result of Shanbhag (1970) and also that of Revankar et al (1974) who showed that the Pareto distribution is the only one with

$E(X|X>y)=\alpha+(\beta+1)y$ for $\beta>0$, which has an important meaning in the laws of income distribution in economics.

We will denote the parametrized distribution as $F(x;\theta)$, which is assumed to satisfy some regularity conditions so that $I_{1;1}(\theta)$ exists. Let $I_{(x_{r;n},\infty)}$ be the average of the conditional information where the conditonal information, $I_{(x_{r;n},\infty)}(\theta|x_{r;n})$, is

$$I_{(x_{r;n},\infty)}(\theta|x_{r;n})=\int_{x_{r;n}}^{\infty}\left(\frac{\partial}{\partial\theta}\log\frac{f(x;\theta)}{1-F(x_{r;n};\theta)}\right)^2\frac{f(x;\theta)}{1-F(x_{r;n};\theta)}dx.$$

Theorem 2.3 *The set* $\{I_{1;1}(\theta), I_{(x_{r;n},\infty)}(\theta); n=2,\cdots\infty\}$, *where the integers* $r=r(n)$ *are allowed to vary with* $n(1\le r(n)<n)$, *specifies* $F(x;\theta)$ *in the same manner as* $\{I_{1;n}(\theta); n=1,\cdots,\infty\}$ *specifies* $F(x;\theta)$.

Proof. Since $f_{r+1\cdots n|r;n}$ can be interpreted as the joint density of an $n-r$ i.i.d sample from $F(x_{r;n},\infty)$,

$$(n-r)I_{(x_{r;n},\infty)}(\theta)=I_{r+1\cdots n|r;n}(\theta),$$

where $I_{r+1\cdots n|r;n}(\theta)$ is the average of the conditional information in $X_{r+1;n},\cdots,X_{n;n}$ given $X_{r;n}=x_{r;n}$. By the Markov chain property of order statistics,

$$I_{r+1\cdots n|r;n}(\theta)=nI_{1;1}(\theta)-I_{1\cdots r;n}(\theta). \tag{5}$$

Let $G(y;\theta)$ be another distribution. Then, if $I^X_{1;1}(\theta)=I^Y_{1;1}(\theta)$ and $I_{(x_{r;n},\infty)}(\theta)=I_{(y_{r;n},\infty)}(\theta)$ for $n=2,\cdots,\infty$, we have by (5),

$$I^X_{1\cdots r;n}(\theta)=I^Y_{1\cdots r;n}(\theta) \text{ for } n=2,\cdots,\infty. \tag{6}$$

Thus, the standard recurrence relation in the Fisher information of order statistics(Park, 1994 A),

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

says that (6) implies

$$I^X_{1;n}(\theta)=I^Y_{1;n}(\theta) \text{ for } n=1,\cdots,\infty.$$

References

[1] Arnold, B. C. and Meeden, G.(1975). Characterization of distributions by sets of moments of order statistics. *Annals of Statistics* **3**, 754-758.

[2] Chan, L. K. (1967). On a characterization of distributions by expected values of extreme order statistics. *American Mathematics Monthly* **74**, 950-951.

[3] Ferguson, T. S. (1967). On charaterizing distributions by properties of order statistics. *Sankhya* A **29**, 265-278.

[4] Hoeffding, W.(1953). On the distribution of the expected values of the order statistics. *Annals of Mathematical Statisitcs* **24**, 93-100.

[5] Kadane, J. B. (1971). A moment problems for order statistics. *Annals of Mathematical Statistics* **42**, 745-751.

[6] Park, S. (1994 A). Recurrence relations for the Fisher information in order statistics and information plot. *(submitted for publication)*

[7] Park, S. (1994 B). Asymptotic relation arising in the decomposition of likelihood of order statistics. *(submitted for publication)*

[8] Revankar, N., Hartley, M. and Pagano, M.(1974). A characterization of the Pareto distribution. *Annals of Statistics* **2**, 599-601.

[9] Shanbag, D. N. (1970). The characterizations for exponential and geometric distributions. *Journal of the American Statistical Association* **65**, 1256-1259.

[10] Wang, Y. H. and Srivastava, R. C.(1980). A characterization of the exponential and related distributions by linear regression. *Annals of Statistics* **8**, 217-220.