Mashhad R. J. Math. Sci., Vol. 2(1)(2009) 61-68

Asymptotic fisher information in order statistics of geometric distribution*

M. Roozbeh†(⊠) and S.M.M. Tabatabaey
Department of Statistics, Faculty of Mathematical Sciences
Ferdowsi University of Mashhad.

Abstract

In this paper, the geometric distribution is considered. The means, variances, and covariances of its order statistics are derived. The Fisher information in any set of order statistics in any distribution can be represented as a sum of Fisher information in at most two order statistics. It is shown that, for the geometric distribution, it can be further simplified to a sum of Fisher information in a single order statistic. Then, we derived the asymptotic Fisher information in any set of order statistics.

Keywords and phrases: Fisher information, geometric distribution, order statistics, percentile, Quantile.
AMS Subject Classification 2000: Primary 94A15; Secondary 41A60.

1 Introduction

The geometric distribution with parameter θ is given by the probability mass function (pmf)

$$f(x;\theta) = (1-\theta)\theta^x,\ x = 0,1,2,\dots\ ,\ 0<\theta<1. \tag{1}$$

*Received: 17 November 2008; Revised: 12 June 2009
†e-mail: m.roozbeh.stat@gmail.com

In general, the distribution theory for order statistics is complex when the parent distribution is discrete. However, order statistics from a geometric distribution exhibit some interesting properties. The geometric distribution possesses several properties (like lack of memory) of exponential distribution. Due to the relationship between the geometric and the exponential distributions, there also exists a close relationship between the dependence structure of order statistics from a geometric distribution and those from the exponential distribution. To this end, we may first note that when Y has an exponential distribution, i.e., it's probability density function is given by

$$f(y;\lambda)=\frac{1}{\lambda}e^{-\frac{1}{\lambda}y},\ y>0,\ \lambda>0, \tag{2}$$

then $X=[Y]$, when [.] stands for integer part, is distributed as geometric with parameter $\theta=1-e^{-1/\lambda}$. Also, the geometric distribution is the only discrete distribution for which the first order statistic and the sample range are independent [2].

The Fisher information plays an important role in statistical inference in connection with estimation, sufficiency and properties of variance of estimators. It is well known that Fisher information serves as a valuable tool for derivation of variance in the asymptotic distribution of maximum likelihood estimators (MLE). For a random variable X, discrete or continuous, which pmf or pdf is $f(x;\theta)$, where $\theta\in\Theta$ is a real value and Θ is the space parameter, the exact Fisher information contained in X is defined as

$$I_X(\theta)=E\Big(\frac{\partial\log f(x;\theta)}{\partial\theta}\Big)^2=-E\Big(\frac{\partial^2\log f(x;\theta)}{\partial^2\theta}\Big)\ , \tag{3}$$

under certain regularity condition (see [7]). Let $X_i,\ i=1,...,n$ be a sample from F_θ, the exact Fisher information about θ in any k order statistics, $X_{r_1:n}\leq X_{r_2:n}\leq...X_{r_k:n},\ 1\leq r_1<r_2<...<r_k\leq n$, is defined as

$$I_{r_1r_2...r_k:n}(\theta)=E\Big\{\frac{\partial}{\partial\theta}\log f_{r_1r_2...r_k:n}(\theta)\Big\}^2, \tag{4}$$

where $f_{r_1r_2...r_k:n}(\theta)$ is joint pmf or pdf of $(X_{r_1:n},X_{r_2:n},...X_{r_k:n})$. The problem of obtaining the Fisher information in order statistics was described in [2] with the words: "while the recipe for $I_Y(\theta)$is simple, the details are messy in most cases" where Y is an arbitary collection of order statistics. Several results have been published in this direction in recent years. For example, Mehtoria et al. [8] presented the Fisher information in the first r order statistics. Park [9] used an indirect approach to obtain the Fisher information in r order statistics, and presented very information plots to demonstrate which order statistics have more Fisher information. Zheng and Gastwirth [14] calculated the Fisher information contained in any collection of order statistics. Abo-Eleneen and Nagaraja [1] studied the Fisher information in collections of order statistics and their concomitants from bivariate samples. Park and Zheng [12] derived a necessary and sufficient condition under which two distribution have equal Fisher information in any set of order statistics. Hofman et al. [6] used the Fisher information in minima and upper record values for characterization of hazard function. Park [10] considered the optimal spacing based on the Fisher information. Park and Kim [11] considered the Fisher information in exponential distribution and simplified the Fisher information in any set of order statistics to a sum of single integrals. In other application, such as life testing surveys (see [3]) and optimal spacing (see [4] and [10]), the asymptotic Fisher information is used.

The rest of the paper is organized as follows. In Section 2, the means, variances, and covariances of geometric order statistics are derived. We derived the asymptotic Fisher information about θ contained in the rth sample quantile ($X_{r:n}$) of geometric distribution in Section 3. In Section 4, we provide the simple method for obtaining the Fisher information and asymptotic Fisher information in any set of order statistics of geometric distribution.

2 Calculating means, variances, and covariances

Since the Fisher information is related to the variance-covariance matrix of the estimate of ϑ, being its inverse under certain conditions, we derive variances, and covariances of order statistics come from a geometric population.

Let $X_1, ..., X_n$ be a sample from (1) and denote the corresponding order statistics by $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$.

Lemma 2.1 *Let $\mu_{r:n}$, and $\sigma^2_{r:n}$, and $\sigma_{r,s:n}$ be mean of $X_{r:n}$, and variance of $X_{r:n}$, and covariance of $X_{r:n}$ and $X_{s:n}$, respectively. Then, we have, for $1 \leq r < s \leq n$*

$$\mu_{r:n} = \sum_{j=0}^{r} \frac{\theta^{n-j+1}}{1-\theta^{n-j+1}}, \tag{5}$$

$$\sigma^2_{r:n} = \sum_{j=0}^{r} \frac{\theta^{n-j+1}}{(1-\theta^{n-j+1})^2}, \tag{6}$$

and

$$\sigma_{r,s:n} = \sigma^2_{r:n} \ . \tag{7}$$

Proof. Under the transformation $Z_i = (n-i+1)(X_{i:n} - X_{i-1:n})$ for $i = 1, 2, ..., n$, one can see that the variables $Z_1, Z_2, ..., Z_n$ are independent random variables and pmf each of Z_i is given by (see [2])

$$f_{z_i}(z;\theta) = (1-\theta^{n-i+1})\theta^z, \ z = 0, \ n-i+1, \ 2(n-i+1), ... \ . \tag{8}$$

The equivalent transformation can be written as

$$X_{r:n} = \sum_{j=0}^{r} \frac{Z_j}{n-j+1}. \tag{9}$$

From (8) and (9) we immediately, conclude that

$$\mu_{r:n} = \sum_{j=0}^{r} \frac{\theta^{n-j+1}}{1-\theta^{n-j+1}},$$

$$\sigma^2_{r:n} = \sum_{j=0}^{r} \frac{\theta^{n-j+1}}{(1-\theta^{n-j+1})^2},$$

and

$$\sigma_{r,s:n} = \sum_{j=0}^{r} \frac{\theta^{n-j+1}}{(1-\theta^{n-j+1})^2}.$$

We may similarly derive the higher-order moments of $X_{r:n}$, if needed.

3 Asymptotic Fisher information in the rth order statistic

Definition ([13]) Assume $\frac{r_i}{n} \to p_i$ (for i=1, 2, ..., k) as $n \to \infty$, where $0 \le p_1 < p_2 < ... < p_k \le 1$. The asymptotic Fisher information about θ contained in k sample quantiles $(X_{r_1:n}, X_{r_2:n}, ...X_{r_k:n})$, denoted by $I_{p_1 p_2 ... p_k}(\theta)$, is defined as

$$I_{p_1 p_2 \dots p_k}(\theta) = \lim_{n\to\infty} \frac{1}{n} I_{r_1 r_2 \dots r_k : n}(\theta), \tag{10}$$

which can be written as ([13])

$$I_{p_1 p_2 \dots p_k}(\vartheta) = \sum_{i=0}^{k} \frac{1}{p_{i+1} - p_i} \left\{ \int_{\xi_{p_i}}^{\xi_{p_{i+1}}} \frac{\partial}{\partial \vartheta} f(x;\vartheta) dx \right\}^2 , \tag{11}$$

where $p_0 = 0, p_{k+1} = 1$, and $\xi_p = F^{-1}(p;\theta)$.

The asymptotic Fisher information in a single order statistic can be obtain rapidly by substituting $k = 1$ in (11). Thus, we get

$$I_p(\vartheta) \quad = \quad \frac{1}{p(1-p)} \left\{ \int_{-\infty}^{\xi_p} \frac{\partial}{\partial \vartheta} f(x;\vartheta) \right\}^2 . \tag{12}$$

In what follows, we find the asymptotic Fisher information in a single quantile of geometric distribution. From (1), we have $F(x;\theta) = (1 - \theta^{x+1})$ so $F^{-1}(p;\theta) = (\frac{\log(1-p)}{\log \theta} - 1)$. By (12), $I_p(\theta)$ for geometric distribution can be calculated as follows

$$\begin{aligned} I_p(\theta) &= \frac{1}{p(1-p)} \left\{ \sum_{x=0}^{\xi_p} \frac{\partial}{\partial \theta} f(x;\theta) \right\}^2 \\ &= \frac{1}{p(1-p)} \left\{ \sum_{x=0}^{\xi_p} x\theta^{x-1} - (x+1)\theta^x \right\}^2 \\ &= \frac{1}{p(1-p)} \left\{ \left([\frac{\log(1-p)}{\log \theta} - 1] + 1 \right)^2 \theta^{2[\frac{\log(1-p)}{\log \theta} - 1]} \right\} , \end{aligned} \tag{13}$$

where $[\cdot]$, denotes the integer part.

Remark 3.1 *By using (13) for geometric distribution, we can approximate the Fisher information contained in $X_{r:n}$ about θ by nothing that $I_{r:n}(\theta) \simeq nI_p(\theta)$ for large values of sample size and $r \le n$ as follows*

$$I_{r:n}(\theta) \simeq \frac{n^3}{r(n-r)} \left\{ \left([\frac{\log(n-r) - \log(r)}{\log\theta} - 1] + 1 \right)^2 \theta^{2[\frac{\log(n-r)-\log(r)}{\log\theta} - 1]} \right\}. \quad (14)$$

4 Asymptotic Fisher information in k order statistics

Park [10] has shown that the Fisher information in any set of order statistics can be written as

$$I_{r_1 r_2 \ldots r_k:n}(\theta) = \sum_{i=2}^{k} I_{r_{i-1} r_i:n}(\theta) - \sum_{i=2}^{k-1} I_{r_i:n}(\theta), \quad (15)$$

where $0 \leq r_1 < r_2 < \ldots < r_k \leq n$. We will show that it can be further simplified to a sum of Fisher information in a single order statistics while the parent distribution is geometric.

Theorem 4.1 *If the random sample comes from a geometric population, then*

$$I_{r_1 r_2 \ldots r_k:n}(\theta) = \sum_{i=1}^{k} I_{r_i - r_{i-1}:n-r_{i-1}}(\theta), \quad (16)$$

where $r_0 = 0$.

Proof. The proof follows by using the lack of memory property of the geometric distribution. As it has been shown in [2], $X_{r_i:n} - X_{r_{i-1}:n}$ is distributed as $X_{r_i - r_{i-1}:n-r_{i-1}}$ in geometric distribution, $(X_{r_1:n}, X_{r_2:n}, \ldots, X_{r_k:n})$ and $(X_{r_1:n}, X_{r_2:n} - X_{r_1:n}, \ldots, X_{r_k:n} - X_{r_{k-1}:n})$ are equivalent statistics and $(X_{r_1:n}, X_{r_2:n} - X_{r_1:n}, \ldots, X_{r_k:n} - X_{r_{k-1}:n})$ are independently and geometrically distributed, therefore the proof is completed.

Theorem 4.2 *If the random sample has geometric distribution, then*

$$I_{p_1 p_2 \ldots p_k}(\theta) = \sum_{i=0}^{k} \frac{1 - p_{i-1}}{(1-p_i)(p_i - p_{i-1})} \left\{ \left(\left[\frac{\log(\frac{1-p_i}{1-p_{i-1}})}{\log\theta} - 1 \right] + 1 \right)^2 \theta^{2[\frac{\log(\frac{1-p_i}{1-p_{i-1}})}{\log\theta} - 1]} \right\}. \quad (17)$$

Proof. By using (13), the asymptotic Fisher information of $I_{r_i - r_{i-1}:n-r_{i-1}}(\theta)$ can be written as $\frac{1}{1-p_{i-1}} I_{\frac{p_i - p_{i-1}}{1-p_{i-1}}}(\theta)$. Thus, the the proof is completed by considering Theorem 4.1.

Remark 4.1 *By using (17) for geometric distribution, we can approximate the Fisher information contained in* $(X_{r_1:n}, X_{r_2:n}, ...X_{r_k:n})$ *about* θ *by nothing that* $I_{r_1 r_2 ... r_k:n}(\theta) \simeq n I_{p_1 p_2 ... p_k}(\theta)$ *for large values of sample size and* r_i *(for* $i = 1, ..., k$*) as follows*

$$I_{r_1 r_2 ... r_k:n}(\theta) \simeq \sum_{i=0}^{k} \frac{n^2(n - r_{i-1})}{(n - r_i)(n - r_{i-1})} \left\{ \left(\left[\frac{\log(\frac{n-r_i}{n-r_{i-1}})}{\log \theta} - 1 \right] + 1 \right)^2 \theta^{2\left[\frac{\log(\frac{n-r_i}{n-r_{i-1}})}{\log \theta} - 1\right]} \right\}, \tag{18}$$

where $0 \leq r_1 < r_2 < ... < r_k \leq n$.

5 Acknowledgements

The authors would like to thank the anonymous referees and the Associate Editor for their constructed suggestions which significantly improved the presentation of the paper. Partial support from the "Ordered and Spatial Data Center of Excellence" of Ferdowsi University of Mashhad, Iran is acknowledged.

References

[1] Abo-Eleneen, Z.A. and Nagaraja, H.N., Fisher information in order statistics and its concomitant, *Ann. Inst. Statist. Math.* **54**(2002), 667–680.

[2] Arnold, B.C., Balakrishnan, N. and Nagaraga, H.N., *A First Course in Order Statistics*, John Wiley, New York, 1992.

[3] Battacharyya, G.K., The asymptotic of maximum likelihood and related estimators based on Type II censored data, *J. Amer. Statist. Assoc.* **80**(1985), 398–404.

[4] Cheng, S.W., A unified approach to choosing optimum quantiles for ABLE's. *J. Amer. Statist. Assoc.* **70**(1975), 155–159.

[5] David, H.A. and Nagaraja, H.N., *Order Statistics*, 3^{rd} Ed. John Wiley, New York, 2003.

[6] Hofmann, G., Balakrishnan, N. and Ahmadi, J., Characterization of hazard function factorization by Fisher information in minima and upper record values, *Statist. Probab. Lett.* **72**(2005), 51–57.

[7] Lehmann, E.L., Theory of Point Estimation, Second Edition, John Wiley, New York, 1998.

[8] Mehtoria, K.G., Johnson, R.A. and Battacharyya, G.K., Exact Fisher information for censored samples and the extended hazard rate function, *Comm. Statist. - Theory Methods* **15**(1979), 1493–1510.

[9] Park, S., Fisher information in order statistics, *J. Amer. Statist. Assoc.* **91**(1996), 385–390.

[10] Park, S., On simple calculation of the Fisher information in order statistics, *Statist. Papers* **46**(2005), 293–301.

[11] Park, S. and Kim, C.E., A note on the Fisher information in exponential distribution, *Comm. Statist. -Theory Methods* **35**(2006), 13–19.

[12] Park. S. and Zheng, G., Equal Fisher information in order statistics, *Sankhya Statist.* **66**(2004), 20–34.

[13] Zheng, G., On the rate of convergence of Fisher information in multiple type II censored data, *J. Japan Statist. Soc.* **30**(2000), 197–204.

[14] Zheng, G. and Gastwirth, J.L., Where is the Fisher information in the ordered sample?, *Statist. Sinica* **10**(2000), 1267–1280.