2011

Evaluating Fisher Information in Order Statistics

Mary F. Dorn
mfdorn@siu.edu

EVALUATING FISHER INFORMATION

IN ORDER STATISTICS

by

Mary Frances Dorn

B.A., University of Chicago, 2009

A Research Paper
Submitted in Partial Fulfillment of the Requirements for the
Master of Science Degree

Department of Mathematics
in the Graduate School
Southern Illinois University Carbondale
July 2011

RESEARCH PAPER APPROVAL

EVALUATING FISHER INFORMATION IN ORDER STATISTICS

By

Mary Frances Dorn

A Research Paper Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Master of Science

in the field of Mathematics

Approved by:

Dr. Sakthivel Jeyaratnam, Chair

Dr. David Olive

Dr. Kathleen Pericak-Spector

Graduate School
Southern Illinois University Carbondale
July 14, 2011

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my advisor, Dr. Sakthivel Jeyaratnam, for his help and guidance throughout the development of this paper. Special thanks go to Dr. Olive and Dr. Pericak-Spector for serving on my committee. I would also like to thank the Department of Mathematics at SIUC for providing many wonderful learning experiences.

Finally, none of this would have been possible without the continuous support and encouragement from my parents. Thank you for your unconditional love and trust.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

R.A. Fisher [5] introduced the concept of Fisher information in 1925 as a means to compare statistics by what he called "the intrinsic accuracy" of their random sampling distributions. He studied the loss of accuracy with the use of different estimates for an unknown parameter, such as the median or the maximum likelihood estimate. Fisher information appears in the Cramér-Rao Inequality, which provides a lower bound for the variance of an unbiased estimator, and is associated with the asymptotic variance of maximum likelihood estimates.

Order statistics also play an important role in statistical inference. The amount of information we obtain from a random sample is simply the sum of the information we obtain from each independent observation. However, if the independent observations are ranked in order of their numerical value, they are neither independent nor identically distributed. Just a few of these order statistics may tell us more about the mean, for example, than twice as many unordered observations. A natural question is then, "Which part of the ordered sample contains the most information?" This problem was first discussed by John Tukey [14] in 1965 in terms of linear sensitivity measure. He studied the sensitivity of asymptotic efficiency in inference using blocks of consecutive order statistics, when one order statistic is added adjacent to the block. He noted that the linear sensitivity of any estimate is bounded above by the Fisher information through the Cramér-Rao Inequality.

Fisher information in order statistics presents a means of finding and comparing

relatively efficient estimators. Regarding unbiased estimators, Fisher information is associated with efficiency through the Cramér-Rao Inequality. When the estimator is unbiased, the Cramér-Rao lower bound is exactly the reciprocal of the Fisher information. The order statistics that minimize the variance of the linear unbiased estimate are precisely those that contain the most Fisher information about the unknown parameter [18].

Finding the Fisher information contained in a random sample is simple. However, once order statistics are introduced, although the definition of Fisher information is straightforward, the calculation is very complicated because they are not identically distributed. Much research has been done in order to simplify the calculation of exact Fisher information in ordered data. For example, Type-I and Type-II censored data, and, more recently, experiments under various hybrid, random, and progressive censoring schemes have been studied. Research has also been extended to other situations, including record data, truncated samples, weighted samples, and order statistics and their concomitants from bivariate samples. Asymptotic Fisher information in order statistics has been considered for many common distributions [18].

In this paper, we will concentrate on the exact Fisher information contained in certain subsets of the order statistics. Chapter 1 presents some well known results about order statistics and Fisher information. It also discusses the determination of Fisher information contained in a single order statistic. Chapter 2 is a survey of the different approaches that have been developed to find the Fisher information in

collections of order statistics. Specifically, we look at Type-II and multiple Type-II censored data. We also study methods of finding the Fisher information in scattered collections of order statistics. Chapter 3 applies the results of Chapter 2 to a few common distributions in order to calculate the exact Fisher information in consecutive and scattered collections of order statistics. For the exponential distribution, we compute the amount of information about the scale parameter contained in consecutive order statistics. For the normal and logistic distributions, we compute the information about the location parameter in consecutive order statistics. For the logistic distribution, we also derive a simple expression for the information contained in pairs of order statistics, and find the optimal selection of order statistics.

CHAPTER 1

BACKGROUND

Suppose X is a continuous random variable with probability density function $f(x;\theta)$, where $\theta \in \Omega \subseteq \mathbb{R}$ is an unknown scalar. Under certain regularity conditions (e.g., Rao [13], p.329), the *Fisher information* contained in X about θ is defined as

$$I_X(\theta) = E\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2, \tag{1.1}$$

where log represents the natural log. If $\log f(x;\theta)$ is twice differentiable and $f(x;\theta)$ also satisfies

$$\frac{d}{d\theta}E\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) = \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) f(x;\theta)\right] dx,$$

then we have

$$I_X(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right). \tag{1.2}$$

This holds if X is a discrete random variable with probability mass function $f(x;\theta)$, under the modified assumption that summation and differentiation with respect to θ are interchangeable. By considering θ to be a vector of parameters, results can be generalized to compute the Fisher information matrix, which is used to calculate the covariance matrix associated with maximum likelihood estimates. However, for the remainder of the paper we will assume that X is absolutely continuous and that θ is a scalar.

1.1 FISHER INFORMATION IN THE SAMPLE

One basic property of Fisher information is its additivity. Suppose X and Y are independent variables with probability density functions $f_X(x;\theta)$ and $f_Y(y;\theta)$, respectively. Let $I_X(\theta)$ and $I_Y(\theta)$ be the informations contained in X and Y, respectively. Then the joint density of (X,Y) is the product $f_X(x;\theta)f_Y(y;\theta)$, and the information contained in (X,Y) is

$$\begin{aligned} I_{X,Y}(\theta) &= E\left(\frac{\partial}{\partial\theta}\log f(X;\theta)f(Y;\theta)\right)^2 \\ &= E\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)^2 + E\left(\frac{\partial \log f(Y;\theta)}{\partial\theta}\right)^2 \\ &\quad + 2E\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\cdot\frac{\partial \log f(Y;\theta)}{\partial\theta}\right) \\ &= I_X(\theta) + I_Y(\theta), \end{aligned}$$

since

$$E\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\cdot\frac{\partial \log f(Y;\theta)}{\partial\theta}\right) = E\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)E\left(\frac{\partial \log f(Y;\theta)}{\partial\theta}\right) = 0$$

by the independence of X and Y. If, on the other hand, X and Y are not independent, then it has been noted (e.g., Abo-Eleneen and Nagaraja [1]) that Fisher information has the subadditivity property, $I_{X,Y}(\theta) < I_X(\theta) + I_Y(\theta)$.

Now if we consider a random sample $X_1, ..., X_n$ from a density $f(x_1, ..., x_n;\theta)$, it follows that $I_{1,\ldots,n}(\theta) = nI_X(\theta)$, where $I_{1,\ldots,n}(\theta)$ is the information in the entire sample and $I_X(\theta)$ is the information contained in a single observation. Let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ denote the order statistics from this sample. The joint pdf of all order statistics is given by (e.g., Arnold et al. [2], p.10)

$$f_{1\cdots n:n}(x_1, ..., x_n) = n!f(x_1)\cdots f(x_n), \quad \text{for } x_1 < \cdots < x_n,$$

which is clear since there are $n!$ orderings of the x_i. It follows from

$$\sum_{i=1}^{n} \frac{\partial^2 \log f(X_{i:n};\theta)}{\partial\theta^2} = \sum_{i=1}^{n} \frac{\partial^2 \log f(X_i;\theta)}{\partial\theta^2}$$

that the information contained in the vector $\mathbf{X} = (X_{1:n}, ..., X_{n:n})$ of all order statistics is the same as the information contained in the entire sample.

It is also known that the Fisher information contained in a sufficient statistic is equal to the information in the sample. The factorization theorem (e.g., Casella and Berger [4], p.276) states that a statistic $T(\mathbf{X})$ is sufficient for θ if and only if there exist functions $g(t;\theta)$ and $h(\mathbf{x})$ such that

$$f(\mathbf{x};\theta) = g(T(\mathbf{x});\theta)h(\mathbf{x}).$$

Then, since $h(\mathbf{x})$ does not depend on θ,

$$\frac{\partial}{\partial\theta} \log f(\mathbf{x};\theta) = \frac{\partial}{\partial\theta} \log g(T(\mathbf{x});\theta),$$

which gives the equality of information.

Exponential Distribution

Let $X_1, ...X_n$ be a random sample from an exponential distribution, i.e., $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$ for $x \geq 0$. Then,

$$I_X(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log(\frac{1}{\theta}e^{-x/\theta})\right) = -E\left(\frac{\partial^2}{\partial\theta^2}(-\log\theta - \frac{x}{\theta})\right) = -\frac{1}{\theta^2} + \frac{2E(x)}{\theta^3} = \frac{1}{\theta^2}.$$

Thus, the information contained in $\mathbf{X} = (X_{1:n}, ...X_{n:n})$ about θ is $I_{\mathbf{X}} = n/\theta^2$.

Normal Distribution

Let X be $N(\theta, \sigma^2)$, where σ^2 is given and θ is unknown. Then $I_X(\theta) = 1/\sigma^2$. Thus, the information contained in $\mathbf{X} = (X_{1:n}, ...X_{n:n})$ about the mean is $I_{\mathbf{X}}(\theta) =$

n/σ^2. In particular, if X is $N(\theta, 1)$, then $I_{\mathbf{X}}(\theta) = n$. Similarly, if σ^2 is unknown, one can find the information contained in the sample about the scale parameter to be $I_{\mathbf{X}}(\sigma^2) = n/(2\sigma^4)$.

1.2 FISHER INFORMATION IN A SINGLE ORDER STATISTIC

There are many situations where it is practical to consider an order statistic instead of the unordered random observations. To reduce costs, rather than testing the lifetime of each component in a k-out-of-m system, a company may test only the lifetime of the system, which is the $(m-k+1)^{th}$ order statistic. In ballistic experiments, where several projectiles are fired at a target, often only the worst shot is used for further analysis. In this example, it may be important to understand the amount of information contained in the observation that was the greatest distance from the target, or the sample maximum [7].

Consider a continuous population with pdf $f(x;\theta)$, and let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ denote the order statistics from a sample of size n. While it is straightforward to compute the information contained in a single observation or the whole sample, determining the information contained in a subset of the order statistics is much more involved. To find the information in the r^{th} order statistic, we use the definition given in equation (1.2), where the pdf is of $X_{r:n}$ is given by

$$f_{r:n}(x;\theta) = \frac{n!}{(r-1)!\,(n-r)!}\, F(x)^{r-1}\,(1-F(x))^{n-r}\, f(x).$$

Nagaraja [9] showed that the regularity conditions necessary to define $I_X(\theta)$ are enough to define $I_{r:n}(\theta)$.

Since the order statistics are not identically distributed, there is no reason to expect the Fisher information in the r^{th} order statistic to equal that in a single observation. An interesting question is whether $I_{k:n}(\theta)$ is always greater than $I_X(\theta)$, the information in a single observation. Iyengar et al.[7] considered this problem by studying the Fisher information in a sample from a weighted distribution. Suppose a random sample $X_1, ..., X_n$ comes from a pdf that belongs to the exponential family of distributions

$$f(x;\theta) = a(x)e^{\theta T(x) - C(\theta)}. \tag{1.3}$$

In this case, the information in a single observation is given by $I_X(\theta) = C''(\theta)$, and the information in the sample is $nC''(\theta)$. Let Y have the weighted distribution with pdf

$$f^w(y;\theta) = \frac{w(y;\theta)f(y;\theta)}{\mathrm{E}\{w(X;\theta)\}}. \tag{1.4}$$

Using the definition in (1.2), they show that the information contained in Y is

$$I_Y(\theta) = I_X(\theta) + \frac{\partial^2}{\partial\theta^2}\log \mathrm{E}\{w(X;\theta)\} - \mathrm{E}\left\{\frac{\partial^2}{\partial\theta^2}\log w(Y;\theta)\right\}. \tag{1.5}$$

Iyengar et al. [7] noted that the density a single order statistic $X_{r:n}$ is a weighted distribution, where the weight function is $w(x_r;\theta) = F(w_r;\theta)^{r-1}(1 - F(x_r;\theta))^{n-r}$ and $\mathrm{E}\{w(X;\theta)\}$ is independent of θ. Hence, from (1.5),

$$\begin{aligned} I_{r:n}(\theta) &= I_X(\theta) - \mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\log w(X_{r:n};\theta)\right] \\ &= I_X(\theta) - \mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\{(r-1)\log F(X_{r:n};\theta) + (n-r)\log(1 - F(X_{r:n};\theta))\}\right]. \end{aligned} \tag{1.6}$$

Using this result they proved that the Fisher information in $X_{r:n}$ is greater(less) than or equal to that in a random observation X from the density $f(x;\theta)$ if

$$(r-1)\log F(x_r;\theta) + (n-r)\log(1-F(x_r;\theta)) \tag{1.7}$$

is a concave(convex) function of θ for every x_r. Furthermore, they showed that the inequality if strict if the function (1.7) is strictly concave(convex) for every θ.

Exponential Distribution

For the exponential distribution, expressions for $I_{r:n}(\theta)$ were derived by Nagaraja [9] in terms of moments of order statistics. For $r = 1$, consider the distribution of $X_{1:n}$,

$$\begin{aligned} F_{1:n}(x) &= 1 - \{1 - F(x)\}^n \\ &= 1 - \{1 - (1 - e^{-x/\theta})\}^n \\ &= 1 - \{e^{-x/\theta}\}^n \\ &= 1 - e^{-x/(\theta/n)}, \end{aligned}$$

given $x > 0$. Thus, the sample minimum is also exponentially distributed, with scale parameter θ/n. Hence the Fisher information about θ in the smallest order statistic is $I_{1:n}(\theta) = 1/\theta^2$. Since

$$f_{r:n}(x;\theta) = \frac{n!}{(r-1)!(n-r)!}\frac{1}{\theta}(1-e^{-x/\theta})^{r-1}e^{-(n-r+1)x/\theta}$$

and noting that

$$\frac{\partial}{\partial\theta}\left(\frac{e^{-x/\theta}}{1-e^{-x/\theta}}\right) = \frac{x}{\theta^2}\frac{e^{-x/\theta}}{(1-e^{-x/\theta})^2},$$

we find that

$$\frac{\partial^2 \log f_{r:n}(x;\theta)}{\partial\theta^2} = \frac{1}{\theta^2} - \frac{2(n-r+1)}{\theta^2}\frac{x}{\theta} + \frac{2(r-1)}{\theta^2}\frac{x}{\theta}\frac{e^{-x/\theta}}{1-e^{-x/\theta}} - \frac{(r-1)}{\theta^2}\left(\frac{x}{\theta}\right)^2\frac{e^{-x/\theta}}{(1-e^{-x/\theta})^2}.$$

Taking the expectation, for $r = 2$,

$$\begin{aligned} I_{2:n}(\theta) &= \frac{1}{\theta^2} + \frac{2n(n-1)}{\theta^2}\sum_{j=0}^{\infty}\frac{1}{(n+j)^3} \\ &\simeq \frac{1}{\theta^2} + \frac{2n(n-1)}{\theta^2}\int_{x=0}^{\infty}\frac{1}{(n+x)^3}dx \\ &= \frac{1}{\theta^2} + \frac{(n-1)}{n\theta^2}, \end{aligned}$$

for large n. For $r \geq 3$,

$$\begin{aligned} I_{r:n}(\theta) = &-\frac{1}{\theta^2} + \frac{2(n-r+1)}{\theta^2}\mathrm{E}\left(X_{r:n}\right) - \frac{2(n-r+1)}{\theta^2}\mathrm{E}\left(X_{r-1:n}\right) \\ &+ \frac{n(n-r+1)}{(r-2)\theta^2}\mathrm{E}\left(X^2_{r-2:n-1}\right). \end{aligned} \tag{1.8}$$

The following expressions for the mean and variance of exponential order statistics (e.g., Arnold et al. [2], p.73)

$$\begin{aligned} \mu_{r:n} &= \mathrm{E}\left(X_{r:n}\right) = \sum_{i=1}^{r}\frac{1}{n-i+1}, \\ \sigma^2_{r:n} &= Var(X_{r:n}) = \sum_{i=1}^{r}\frac{1}{(n-i+1)^2}, \end{aligned} \tag{1.9}$$

allow us to write $\mathrm{E}\left(X_{r:n}\right) - \mathrm{E}\left(X_{r-1:n}\right) = 1/(n-r+1)$ and $\mathrm{E}\left(X^2_{r-2:n-1}\right) = Var(X_{r-2:n-1}) + \left(\mathrm{E}\left(X_{r-2:n-1}\right)\right)^2$. Thus, equation (1.8) simplifies to

$$\begin{aligned} I_{r:n}(\theta) &= \frac{1}{\theta^2} + \frac{1}{\theta^2}\frac{n(n-r+1)}{r-2}(\mu^2_{r-2:n-1} + \sigma^2_{r-2:n-1}) \\ &= \frac{1}{\theta^2} + \frac{1}{\theta^2}\frac{n(n-r+1)}{r-2}\left(\left(\sum_{i=1}^{r-2}\frac{1}{n-i}\right)^2 + \sum_{i=1}^{r-2}\frac{1}{(n-i)^2}\right). \end{aligned}$$

r	1	2	3	4	5	6	7	8	9	10
$\theta^2 I_{r:10}$	1	1.9945	2.9753	3.9302	4.8399	5.6734	6.3770	6.8497	6.8741	5.8561

Table 1.1. $I_{r:10}(\theta)$ – Fisher information the r^{th} order statistics from $Exp(\theta)$

For large n, the information in the r^{th} order statistic can be approximated for $r \geq 3$ as

$$I_{r:n}(\theta) \approx \frac{2}{\theta^2} + \frac{n}{\theta^2}\frac{(1-p)}{p}\{\log(1-p)\}^2,$$

where $p = r/n$, $0 < p < 1$. Nagaraja observed that the function $g(p) = (1-p)\{\log(1-p)\}^2$ is monotonically increasing in the interval $(0, p_0)$ and monotonically decreasing in $(p_0, 1)$, where p_0 is approximately 0.7968. Hence, for sufficiently large n, $I_{r:n}(\theta)$ increases as r increases up to roughly $0.8n$ and then decreases. Thus, the 80^{th} sample percentile contains the maximum information about the parameter θ, with about $(2 + 0.65n)$ times the information contained in a single observation.

Weibull Distribution

The exponential distribution is a special case of the Weibull distribution with shape parameter $\alpha = 1$. More generally, Iyengar et al. consider a random sample from a Weibull distribution with density $f(x;\beta) = \alpha\beta x^{\alpha-1}e^{-\beta x^\alpha}$ for $x \geq 0$, where $\alpha > 0$ is known and $\beta > 0$ is the unknown scale parameter. It is easy to see that this pdf belong to the exponential family in (1.3) when it is rewritten in the form

$$f(x;\beta) = \alpha x^{\alpha-1}e^{-\beta x^\alpha + \log(\beta)}.$$

Thus the information about β in a single observation is

$$I_X(\beta) = C''(\beta) = \frac{\partial^2}{\partial \beta^2}\left[-\log \beta\right] = \frac{1}{\beta^2},$$

which is independent of the value of α. From equation (1.6), Iyengar et al. found the Fisher information contained in $X_{r:n}$ about β as

$$\begin{aligned} I_{r:n}(\beta) &= \frac{1}{\beta^2} - \mathrm{E}\left[\frac{\partial^2}{\partial \beta^2}\{(r-1)\log(1-e^{-\beta X_{r:n}^{\alpha}}) + (n-r)\log e^{-\beta X_{r:n}^{\alpha}}\}\right] \\ &= \frac{1}{\beta^2} - \mathrm{E}\left[\frac{\partial}{\partial \beta}\{(r-1)\frac{X_{r:n}^{\alpha} e^{-\beta X_{r:n}^{\alpha}}}{1-e^{-\beta X_{r:n}^{\alpha}}} - (n-r)X_{r:n}^{\alpha}\}\right] \\ &= \frac{1}{\beta^2} - (r-1)\ \mathrm{E}\left[\frac{\partial}{\partial \beta}\frac{X_{r:n}^{\alpha}}{e^{\beta X_{r:n}^{\alpha}} - 1}\right]. \end{aligned}$$

Since $\frac{1}{exp\{\beta x^{\alpha}\}-1}$ is a decreasing function of β for all x, the expression given in (1.7) is concave, and thus $I_{r:n}(\beta) \geq I_X(\beta)$ for all β and α. In fact, Iyengar et al. found that for all β and α,

$$I_{1:n}(\beta) = I_X(\beta)$$

$$I_{r:n}(\beta) > I_X(\beta),\ 1 < r \leq n.$$

Normal Distribution

For the normal distribution, Nagaraja [9] showed that the information contained in $X_{r:n}$ about the mean can be expressed in terms of expectations of some functions of the standard normal pdf ϕ and cdf Φ. If we have a sample from $N(\theta, 1)$,

then the pdf if $\phi(x-\theta)$ and the cdf is $\Phi(x-\theta)$. Since

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2}\log f_{r:n}(x;\theta) &= \frac{\partial^2}{\partial\theta^2}\log\left\{\frac{n!}{(r-1)!\,(n-r)!}\,\Phi^{r-1}(x-\theta)\right.\\ &\qquad\left.\times\,(1-\Phi(x-\theta))^{n-r}\,\phi(x-\theta)\right\}\\ &= 1-(r-1)\left\{\frac{(x-\theta)\phi(x-\theta)}{\Phi(x-\theta)}+\frac{\phi^2(x-\theta)}{\Phi^2(x-\theta)}\right\}\\ &\qquad+(n-r)\left\{\frac{(x-\theta)\phi(x-\theta)}{1-\Phi(x-\theta)}-\frac{\phi^2(x-\theta)}{[1-\Phi(x-\theta)]^2}\right\},\end{aligned}$$

substituting $z=x-\theta$, we have

$$\begin{aligned}I_{r:n}(\theta) &= \mathrm{E}\left(-\frac{\partial^2}{\partial\theta^2}\log\,f_{r:n}(x;\theta)\right)\\ &= 1+\int_{-\infty}^{\infty}(r-1)\frac{z\phi(z)}{\Phi(z)}f_{r:n}(x)dz-\int_{-\infty}^{\infty}(n-r)\frac{z\phi(z)}{1-\Phi(z)}f_{r:n}(x)dz\\ &\quad+\int_{-\infty}^{\infty}(r-1)\frac{\phi^2(z)}{\Phi^2(z)}f_{r:n}(x)dz+\int_{-\infty}^{\infty}(n-r)\frac{\phi^2(z)}{[1-\Phi(z)]^2}f_{r:n}(x)dz\\ &= 1+\int_{-\infty}^{\infty}nz\phi(z)\frac{(n-1)!}{(r-2)!\,(n-r-2)!}\,\Phi^{r-2}(z)\,(1-\Phi(z))^{n-r-2}\,\phi(z)dz\\ &\quad-\int_{-\infty}^{\infty}nz\phi(z)\frac{(n-1)!}{(r-1)!\,(n-r-1)!}\,\Phi^{r-1}(z)\,(1-\Phi(z))^{n-r-1}\,\phi(z)dz\\ &\quad+\int_{-\infty}^{\infty}\frac{n(n-1)}{r-2}\phi^2(z)\frac{(n-2)!}{(r-3)!\,(n-r-4)!}\,\Phi^{r-3}(z)\,(1-\Phi(z))^{n-r-4}\,\phi(z)dz\\ &\quad+\int_{-\infty}^{\infty}\frac{n(n-1)}{n-r-1}\phi^2(z)\frac{(n-2)!}{(r-1)!\,(n-r-2)!}\,\Phi^{r-1}(z)\,(1-\Phi(z))^{n-r-2}\,\phi(z)dz\\ &= 1+\int_{-\infty}^{\infty}nz\phi(z)f_{r-1:n-1}(z)dz-\int_{-\infty}^{\infty}nz\phi(z)f_{r:n-1}(z)dz\\ &\quad+\int_{-\infty}^{\infty}\frac{n(n-1)}{r-2}\phi^2(z)f_{r-2:n-2}(z)dz+\int_{-\infty}^{\infty}\frac{n(n-1)}{n-r-1}\phi^2(z)f_{r:n-2}(z)dz.\end{aligned}$$

Thus, for $3\le r\le n-2$ and $n\ge 5$,

$$\begin{aligned}I_{r:n}(\theta) &= n\mathrm{E}_{\,r-1:n-1}\{Z\phi(Z)\}-n\mathrm{E}_{\,r:n-1}\{Z\phi(Z)\}+\frac{n(n-1)}{n-2}\mathrm{E}_{\,r-2:n-2}\{\phi^2(Z)\}\\ &\quad+\frac{n(n-1)}{n-r-1}\mathrm{E}_{\,r:n-2}\{\phi^2(Z)\}+1,\end{aligned}$$

where the expectation $\mathrm{E}_{i:j}$ is taken with respect to the distribution of $Z_{i:j}$, an order statistic from the standard normal distribution.

For other values of r and n,

$$I_{r:n}(\theta) = \begin{cases} 1,\ r=1,\ n=1 \\ 1+2\mathrm{E}\,\{\frac{\phi^2(Z)}{\Phi(Z)}\},\ r=1,\ n=2 \\ 1-n\mathrm{E}_{\,1:n}\{Z\phi(Z)\}+\frac{n(n-1)}{n-2}\mathrm{E}_{\,1:n-2}\{\phi^2(Z)\},\ r=1,\ n\geq 3 \\ 1+6\mathrm{E}_{\,1:2}\{\frac{\phi^2(Z)}{\Phi(Z)}\},\ r=2,\ n=3 \\ 1+n\mathrm{E}_{\,1:n-1}\{Z\phi(Z)\}-n\mathrm{E}_{\,2:n-1}\{Z\phi(Z)\}+n\mathrm{E}_{\,1:n-1}\{\frac{\phi^2(Z)}{\Phi(Z)}\} \\ \qquad +\frac{n(n-1)}{n-3}\mathrm{E}_{\,2:n-2}\{\phi^2(Z)\},\ r=2,\ n\geq 4. \end{cases}$$

For a symmetric population, it is known that $X_{i:n}$ and $-X_{n-i+1:n}$ have the same distribution (e.g., Arnold et al. [2], p.26). This follows from the fact that $f(-x+\theta) = f(x+\theta)$ and $F(-x+\theta) = 1-F(x+\theta)$ for a distribution that is symmetric about θ. Hence $I_{r:n}(\theta) = I_{n-r+1:n}(\theta)$. The remaining values of $I_{r:n}(\theta)$ can be found using this equality. The amount of information about θ in a single order statistic was given by Nagaraja [9] for sample size $n \leq 10$. Using *Mathematica* for numerical integration, these values are extended in Table 1.2 to include values for n up to 20. The median order statistics contain the most information about θ. For even n the median order statistics $X_{(n/2):n}$ and $X_{(n/2+1):n}$ both contain the same amount of information by the symmetry of the normal distribution. For $n = 10$, the $X_{5:10}$ and $X_{6:10}$ each contain 0.6622 times the total information in the sample. For $n = 20$, the proportion of information contained in the median statistic is 0.6498.

Iyengar et al. [7] compared the Fisher information in a single order statistic

with that in a single observation for the case of a $Normal(\theta, 1)$ distribution. In order to show that the expression given in (1.7) is concave, they write it in terms of the standard normal pdf ϕ and cdf Φ. Its second derivative is

$$\begin{aligned}&\frac{\partial^2}{\partial\theta^2}\{(r-1)\log\Phi(x_r-\theta)+(n-r)\log(1-\Phi(x_r-\theta))\}\\&=(r-1)\frac{\partial}{\partial\theta}\left\{\frac{\phi(x_r-\theta)}{\Phi(x_r-\theta)}\right\}-(n-r)\frac{\partial}{\partial\theta}\left\{\frac{\phi(x_r-\theta)}{1-\Phi(x_r-\theta)}\right\}.\end{aligned}$$

Here, $h(x)=\phi(x)/(1-\Phi(x))$ is the usual hazard rate function, and its reciprocal $M(x)=(1-\Phi(x))/\phi(x)$ is known as the Mills ratio. Iyengar et al. used the fact that $M(x)$ is a strictly decreasing function for all x, hence $h(x_r-\theta)$ is strictly increasing. Also, they note that by the symmetry of the normal density,

$$\frac{\phi(x)}{\Phi(x)}=\frac{\phi(-x)}{1-\Phi(-x)},$$

and therefore $\phi(x_r-\theta)/\Phi(x_r-\theta)$ is strictly decreasing. Thus, since the second derivative above is negative for all x_r, the r^{th} order statistic of a normal sample of size $n>1$ for $1\leq r\leq n$ always contains more Fisher information about the mean than a single observation. As noted by Zheng et al.[18], it follows that the sum of the Fisher information contained in each order statistic is strictly greater than the information in the whole sample, i.e., $\sum_{r=1}^{n} I_{r:n}(\theta) > I_{1\cdots n:n}(\theta)$. This is also the subadditivity property of Fisher information.

$n\|r$	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.7403									
3	1.9648	2.7806								
4	2.1914	2.7806								
5	2.3848	3.2286	3.4869							
6	2.5568	3.6102	4.0652							
7	2.7127	3.9448	4.5619	4.7524						
8	2.8555	4.2437	5.0006	5.3448						
9	2.9874	4.5145	5.3952	5.8701	6.0210					
10	3.1099	4.7625	5.7550	6.3446	6.6220					
11	3.2243	4.9914	6.0863	6.7790	7.1662	7.2911				
12	3.3317	5.2043	6.3938	7.1803	7.6655	7.8979				
13	3.4328	5.4033	6.6810	7.5542	8.1281	8.4556	8.5622			
14	3.5284	5.5903	6.9507	7.9044	8.5599	8.9730	9.1732			
15	3.6189	5.7667	7.2050	8.2342	8.9654	9.4569	9.7408	9.8338		
16	3.7051	5.9338	7.4458	8.5462	9.3481	9.9120	10.2722	10.4479		
17	3.7871	6.0925	7.6745	8.8423	9.7108	10.3422	10.7726	11.0233	11.1057	
18	3.8655	6.2437	7.8924	9.1242	10.0557	10.7504	11.2461	11.5657	11.7224	
19	3.9405	6.3881	8.1006	9.3934	10.3847	11.1393	11.6961	12.0794	12.3040	12.3780
20	4.0125	6.5262	8.2998	9.6511	10.6994	11.5107	12.1251	12.5679	12.8552	12.9966

Table 1.2. $I_{r:n}(\theta)$ – Fisher information the r^{th} order statistics from $N(\theta, 1)$

CHAPTER 2

FISHER INFORMATION IN A COLLECTION OF ORDER STATISTICS

In this chapter, we look at the methods developed to find the Fisher Information in various collections of order statistics. Although the recipe is simple, having to evaluate a multiple integral makes finding the Fisher information in order statistics very complicated. Several papers have aimed to find alternative expressions that simplify the detailed calculation of exact information.

We first consider the Fisher information in the first r order statistics. This is referred to as a Type-II censored, or right censored, sample. In Type-II censoring, n items are tested, but only the first $r < n$ of them are observed. Type-II censoring has many applications in life testing. For example, to reduce the cost and time of testing, the observer may wait only until the r^{th} failure and decide to censor the remaining observations. We also consider corresponding results for left censored samples, where only the last $n - r + 1$ observations are recorded.

Mehrotra, Johnson, and Bhattacharyya [8] defined three extended hazard rate functions, and derived the Fisher information in a right censored sample in terms of these functions. Park [10, 11] took an indirect approach, and used recurrence relations to compute the Fisher information in consecutive order statistics. Using these relations, he derived the information in right and left censored order statistics in terms of the information contained in the minima and maxima, respectively,

from samples of size up to n. These results aid in the computation of exact Fisher information in consecutive order statistics, which can be used, for example, to determine the number of order statistics to censor in order to achieve a certain level of efficiency. However, the k order statistics that provide the most information about a parameter may not be the first k order statistics. Thus, it is necessary to consider the information in scattered blocks.

Zheng and Gastwirth [17] followed the approach used by Mehrotra et al. to obtain results that allow for the calculation of information under multiple Type-II censoring, when disjoint blocks of order statistics are removed from the sample. In multiple Type-II censored data, a number of blocks of consecutive order statistics are available. Right censoring is the special case where a single block is removed from the right of the sample. The Fisher information contained in scattered blocks of order statistics has also been studied by Park [12], who derived an expression that depends only on the information in pairs of order statistics. Both Zheng and Gastwirth and Park reduced the computation of FI in any collection of order statistics to at most a double integral.

2.1 RIGHT CENSORED SAMPLE

The Fisher information contained in the first r order statistics from a sample of size n is defined to be

$$I_{1\cdots r:n}(\theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{x_{2:n}} \left(\frac{\partial}{\partial \theta} \log f_{1\cdots r:n} \right)^2 dF_{1\cdots r:n}, \tag{2.1}$$

where $f_{1\cdots r:n}$ is the joint density of $X_{1:n}, \ldots, X_{r:n}$ and satisfies certain regularity conditions. The direct computation of Fisher information from (2.1) is tedious and requires the calculation of multiple integrals. Two approaches simplify the necessary calculations.

The first approach was detailed by Mehrotra et al. [8], who decomposed $I_{1\cdots r:n}(\theta)$ as a function of moments of the following three extended hazard rate functions

$$\begin{aligned} K_1(X_{j:n}) &= -\frac{F'(X_{j:n};\theta)}{1-F(X_{j:n};\theta)}, \\ K_2(X_{i:n}) &= \frac{F'(X_{i:n};\theta)}{F(X_{i:n};\theta)}, \quad \text{and} \\ K_3(X_{i:n}, X_{j:n}) &= \frac{F'(X_{j:n};\theta) - F'(X_{i:n};\theta)}{F(X_{j:n};\theta) - F(X_{i:n};\theta)}, \end{aligned} \tag{2.2}$$

where $F'(x;\theta) = \partial F(x;\theta)/\partial\theta$. When θ is scalar, K_1 is the standard hazard rate function and K_2 is the reversed hazard rate function. Let $\psi_\theta(x) = \partial \log f(x;\theta)/\partial\theta$.

First, using moment relations, they showed that the partial derivative of the log likelihood of $(X_{1:n}, ..., X_{r:n})$ is a linear function of K_1, K_2, K_3, and ψ_θ. For the right-censored case,

$$\begin{aligned} \frac{\partial}{\partial\theta}\log f_{1\cdots r:n}(x_1, ...x_r;\theta) &= \sum_{i=1}^{r}\frac{\partial \log f(x_i;\theta)}{\partial\theta} - (n-r)\frac{F'(x_r;\theta)}{1-F(x_r;\theta)} \\ &= \sum_{i=1}^{r}\psi_\theta(x_i) - (n-r)K_1(x_r). \end{aligned}$$

Thus, the Fisher information can be expressed as a function of the moments of the extended hazard rate functions and score function $\psi_\theta(x)$.

Mehrotra et al. showed that these moments can in turn be expressed as a linear combination of

$$\tau_{ij} = E(\psi_\theta(X_{i:n})\ \psi_\theta(X_{j:n})). \tag{2.3}$$

For example,

$$(n-r)\mathrm{E}\left\{\psi(X_{i:n})K_1(X_{r:n}\right\} = \sum_{j=r+1}^{n} \tau_{ij},$$

$$(n-r)\mathrm{E}\left\{K_1^2(X_{r:n})\right\} = \frac{2}{n-r-1}\sum_{i=r+1}^{n-1}\sum_{j=i+1}^{n} \tau_{ij}.$$

These moment relations allow $I_{1\cdots r:n}(\theta)$ for $r < n-1$ to be expressed only in terms of τ_{ij} in the following manner

$$I_{1\cdots r:n}(\theta) = I_{1\cdots n:n}(\theta) - \left[\sum_{i=r+1}^{n} \tau_{ii} - \frac{2}{n-r-1}\sum_{i=r+1}^{n-1}\sum_{j=i+1}^{n} \tau_{ij}\right]. \tag{2.4}$$

This expression is particularly interesting because it lets us see how Fisher information is reduced from the total information in the sample.

The simplest case is to find the Fisher information in all n order statistics $(X_{1:n}, ..., X_{n:n})$, which is given by

$$\begin{aligned} I_{1\cdots n:n}(\theta) &= n\int_{-\infty}^{\infty} \{\frac{\partial}{\partial\theta}\log f(x)\}^2 dF(x) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \tau_{ij} = \sum_{i=1}^{n} \tau_{ii}. \end{aligned}$$

Substituting this expression for the information in the sample,

$$I_{1\cdots r:n}(\theta) = \sum_{i=1}^{r} \tau_{ii} + \frac{2}{n-r-1}\sum_{i=r+1}^{n-1}\sum_{j=i+1}^{n} \tau_{ij}, \tag{2.5}$$

for $r < n-1$. Once τ_{ij} is computed the FI contained in the first r order statistics can be calculated easily.

The second approach was developed by Park [11], and uses the conditional Fisher information, which follows from the Markov chain property of order statistics (e.g., David and Nagaraja [6], p.17). Since

$$\frac{\partial}{\partial\theta}\log f_{1\cdots n:n} = \frac{\partial}{\partial\theta}\log f_{1\cdots r:n} + \frac{\partial}{\partial\theta}\log f_{r+1\cdots n|r:n},$$

where $f_{r+1\cdots n|r:n}$ is the conditional joint distribution of $X_{r+1:n}, \ldots X_{n:n}$ given $X_{r:n} = x_{r:n}$, we have the decomposition of information

$$I_{1\cdots n:n}(\theta) = I_{1\cdots r:n}(\theta) + I_{r+1\cdots n|r:n}(\theta), \tag{2.6}$$

where $I_{r+1\cdots n|r:n}(\theta)$ is the average of the conditional information in $X_{r+1:n}, \ldots X_{n:n}$ given $X_{r:n} = x_{r:n}$. The calculation of Fisher information in a Type-II censored sample is reduced to the calculation of a double integral by writing

$$I_{r+1\cdots n|r:n}(\theta) = (n-r)\int_{\infty}^{\infty} g(w;\theta) f_{r:n}(w;\theta) dw,$$

where

$$g(w;\theta) = \int_{w}^{\infty} \left\{ \frac{\partial}{\partial\theta} \log \frac{f(x;\theta)}{1-F(w;\theta)} \right\}^2 \frac{f(x;\theta)}{1-F(w;\theta)} dx.$$

The recurrence relation between the cdf's of order statistics, $F_{r:n-1} = \frac{n-r}{n} F_{r:n} + \frac{r}{n} F_{r+1:n}$ yields the decomposition of information $nI_{1\cdots r:n-1}(\theta) = (n-r-1)I_{1\cdots r:n}(\theta) + rI_{1\cdots r+1:n}(\theta)$. A corresponding decompositions for a left censored sample is given by Park. Using this and the relation,

$$f_{r:n}(x;\theta) = \sum_{i=n-r+1}^{n} (-1)^{i-n+r-1} \binom{i-1}{n-r} \binom{n}{i} f_{1:i}(x;\theta),$$

the expression for the information in the first r order statistics can be written as a sum of r single integrals. In particular, Park showed that the Fisher information in Type-II censored data is determined by the information in the smallest order statistic in samples of size up to n,

$$I_{1\cdots r:n}(\theta) = \sum_{i=n-r+1}^{n} \binom{i-2}{n-r-1} \binom{n}{i} (-1)^{i-n+r-1} I_{1:i}(\theta), \tag{2.7}$$

for $1 \leq r < n$.

The Fisher information in the smallest order statistic, $I_{1:n}(\theta)$, can be written as

$$I_{1:n}(\theta) = \int_{-\infty}^{\infty} \{\frac{\partial}{\partial\theta} \log f(x;\theta) + (n-1)\frac{\partial}{\partial\theta} \log(1-F(x;\theta))\}^2 n f(x;\theta)(1-F(x;\theta))^{n-1} dx, \tag{2.8}$$

since it has the pdf $f_{1:n}(x;\theta) = nf(x;\theta)(1-F(x;\theta))^{n-1}$.

Park simplified these expressions using a result from Efron and Johnstone [3], who found the Fisher information in a random sample in terms of the hazard function $h(x;\theta) = \frac{f(x;\theta)}{1-F(x;\theta)}$ as

$$\begin{aligned} I_{1:1}(\theta) &= \int_{-\infty}^{\infty} \left\{ \frac{\partial}{\partial\theta} \log f(x;\theta) \right\}^2 dF(x;\theta) \\ &= \int_{-\infty}^{\infty} \left\{ \frac{\partial}{\partial\theta} \log h(x;\theta) \right\}^2 dF(x;\theta). \end{aligned}$$

Since the hazard function of $X_{1:n}$ is n times that of X,

$$I_{1:n}(\theta) = \int_{-\infty}^{\infty} \left\{ \frac{\partial}{\partial\theta} \log h(x;\theta) \right\}^2 dF_{1:n}(x;\theta). \tag{2.9}$$

Park derived corresponding results for left censored samples. The Fisher information in the order statistics $(X_s, \ldots, X_n)$ can be written in terms of the information contained in the greatest order statistic in samples of size up to n,

$$I_{s\cdots n:n}(\theta) = \sum_{i=s}^{n} \binom{i-2}{s-2} \binom{n}{i} (-1)^{i-s} I_{i:i}(\theta), \tag{2.10}$$

for $1 < s \leq n$. The information in the largest order statistic is

$$I_{n:n}(\theta) = \int_{-\infty}^{\infty} \{\frac{\partial}{\partial\theta} \log f(x;\theta) + (n-1)\frac{\partial}{\partial\theta} \log F(x;\theta)\}^2 n f(x;\theta)(F(x;\theta))^{n-1} dx, \tag{2.11}$$

since it has the pdf $f_{n:n}(x;\theta) = nf(x;\theta)(F(x;\theta))^{n-1}$. To find an alternative expression, Park considered a random variable Y whose pdf is the mirror image of f about $x = 0$, and which thus has the hazard function $\frac{f(x;\theta)}{F(x;\theta)}$. On noting that the hazard function of $X_{n:n}$ is n times $\frac{f(x;\theta)}{F(x;\theta)}$,

$$I_{n:n}(\theta) = \int_{-\infty}^{\infty} \left\{ \frac{\partial}{\partial\theta} \log \frac{f(x;\theta)}{F(x;\theta)} \right\}^2 dF_{n:n}(x;\theta). \tag{2.12}$$

Park derived simplified expressions for $I_{1:n}(\theta)$ and $I_{n:n}(\theta)$ for several well-known distributions. Then, using (2.7) and (2.10), we can directly calculate the information contained in right and left censored samples about the parameter θ. However, as Park notes, we must consider the accumulation of rounding errors.

The decomposition based on conditional information in (2.6) can also be applied to finding the information in the first r outcomes from a Type-II progressive censoring scheme. Type-II progressive censoring, where a different number of surviving items are randomly removed from the test after each observation, has many applications in life testing. This decomposition can also be used to obtain the the asymptotic Fisher information contained in the lower p^{th} percentile of the distribution [18].

2.2 MULTIPLY CENSORED SAMPLE

Consider the k order statistics $\mathbf{X} = (X_{r_1:n}, \ldots, X_{r_k:n})$ from a sample of size n. Let $f_{r_1\cdots r_k:n}$ denote their joint density with parameter θ. Under certain regularity conditions, the Fisher information about θ contained in $\mathbf{X}$ is given by

$$I_{r_1\cdots r_k:n}(\theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{x_{r_2:n}} \left(\frac{\partial}{\partial\theta} \log f_{r_1\cdots r_k:n} \right)^2 dF_{r_1\cdots r_k:n}. \tag{2.13}$$

Zheng and Gastwirth [17] followed the approach used by Mehrotra et al. for right censored samples in order to simplify the computation of equation (2.13). Their idea was to consider an arbitrary collection of order statistics as a set of scattered blocks of consecutive order statistics or, in other words, a multiply censored sample. They approached the problem of calculating the information in multiply censored data by censoring the complete sample in several steps, and to study the loss of information at each stage.

First they consider the case where two arbitrary blocks of order statistics are available from a sample of size n. For $1 \leq r \leq u \leq v \leq w \leq n$, $I_{r\cdots uv\cdots w:n}(\theta)$ is found in three steps:

1. $I_{1\cdots w:n}(\theta) = I_{1\cdots n:n}(\theta) - I_R$ (right censoring)
2. $I_{r\cdots w:n}(\theta) = I_{1\cdots w:n}(\theta) - I_L$ (left censoring)
3. $I_{r\cdots uv\cdots w:n}(\theta) = I_{r\cdots w:n}(\theta) - I_M$ (middle censoring).

Using the Markov property of order statistics, Zheng and Gastwirth show that the change in Fisher information when a block of order statistics is removed from the left, the middle, or the right is the same regardless of previous censoring patterns. This means that

$$I_L = I_{1\cdots n:n}(\theta) - I_{r\cdots n:n}(\theta) = I_{1\cdots rj_1j_2\cdots j_t:n}(\theta) - I_{rj_1j_2\cdots j_t:n}(\theta), \tag{2.14}$$

$$I_M = I_{1\cdots n:n}(\theta) - I_{1\cdots uv\cdots n:n}(\theta) = I_{i_1i_2\cdots i_su\cdots vj_1j_2\cdots j_t:n}(\theta) - I_{i_1i_2\cdots i_suvj_1j_2\cdots j_t:n}(\theta), \tag{2.15}$$

$$I_R = I_{1\cdots n:n}(\theta) - I_{1\cdots w:n}(\theta) = I_{i_1i_2\cdots i_sw\cdots n:n}(\theta) - I_{i_1i_2\cdots i_sw:n}(\theta). \tag{2.16}$$

For example, we have equation (2.15) because I_M is precisely the change in informa-

tion when the block $(X_{u+1:n}, ..., X_{v-1:n})$ is removed from the middle of the sample. Putting everything together,

$$I_{r\cdots uv\cdots w:n}(\theta) = I_{1\cdots n:n}(\theta) - I_R - I_L - I_M \tag{2.17}$$

where I_R, I_L, and I_M depend only on where the censored blocks begin and end.

After substituting in the expressions for I_R, I_L, and I_M, the multiple censoring problem is reduced to censoring on the right, left, and middle. Now, we recall equation (2.5), which expressed the information in a right censored sample in terms of $\tau_{ij}(i,j)$. The left censored and middle censored cases can similarly be expressed as linear combinations of τ_{ij} by using the extended hazard rate functions K_2 and K_3 in (2.2). By symmetry, we have for $s > 2$

$$\begin{aligned} I_{s\cdots n:n}(\theta) &= I_{1\cdots n:n}(\theta) - \left[\sum_{i=1}^{s-1}\tau_{ii} - \frac{2}{s-2}\sum_{i=1}^{s-2}\sum_{j=i+1}^{s-1}\tau_{ij}\right] \\ &= \sum_{i=s}^{n}\tau_{ii} + \frac{2}{s-2}\sum_{i=1}^{s-2}\sum_{j=i+1}^{s-1}\tau_{ij}. \end{aligned} \tag{2.18}$$

Zheng and Gastwirth [17] derived the following expression for FI in a middle censored sample. If $v > u + 2$, then

$$I_{1\cdots uv\cdots n:n}(\theta) = I_{1\cdots n:n}(\theta) - \left[\sum_{i=u+1}^{v-1}\tau_{ii} - \frac{2}{v-u-2}\sum_{i=u+1}^{v-2}\sum_{j=i+1}^{v-1}\tau_{ij}\right]. \tag{2.19}$$

The multiply censored case can now be written in terms of $\tau_{ij}(i,j)$. Once these values are tabulated for a given n, $I_{r\cdots uv\cdots w:n}(\theta)$ can be computed directly. Equations (2.5), (2.18) and (2.19) allow us to see that the information in different collections of order statistics is not additive, although this is not obvious when studying the integral in (2.13). In particular, we see that $I_{1\cdots rs\cdots n:n}(\theta) = I_{1\cdots r}(\theta) + I_{s\cdots n:n}(\theta)$. [18]

This process was generalized by Zheng and Gastwirth to determine the Fisher information contained in data from which p disjoint blocks of order statistics are available. Let the collection of order statistics be defined as

$$\mathbf{X} = (X_{i_1:n}, \ldots, X_{i_1+k_1:n};\ \ldots\ ; X_{i_p:n}, \ldots, X_{i_p+k_p:n}), \tag{2.20}$$

where $(X_{i_m:n}, \ldots, X_{i_m+k_m:n})$ represents the m^{th} block of available order statistics. Then,

$$\begin{aligned} I_{\mathbf{X}}(\theta) = I_{1\cdots n:n}(\theta) - \sum_{j=0}^{p} \Bigg[& \sum_{u=i_j+k_j+1}^{i_{j+1}-1} \tau_{uu} \\ & - \frac{2}{i_{j+1} - i_j - k_j - 2} \sum_{u=i_j+k_j+1}^{i_{j+1}-2} \sum_{v=u+1}^{i_{j+1}-1} \tau_{uv} \Bigg], \end{aligned} \tag{2.21}$$

where $i_1 > 2$, $i_p + k_p < n - 1$, and $i_{j+1} - i_j - k_j > 2$, $j = 1, ..., p-1$, $i_0 = k_0 = 0$, $i_{p+1} = n + 1$. This expression gives the exact Fisher information in a multiply censored sample, where any number of blocks of order statistics are removed from the left, middle, and right of an ordered sample.

The advantage of Zheng and Gastwirth's approach is that once the τ_{ij}'s are tabulated the information in scattered order statistics can be found just as easily as in consecutive order statistics. However, equation (2.21) requires that censored blocks are of at least two order statistics. While $I_{3478:10}(\theta)$ and $I_{38:10}(\theta)$ are easily computed using this result, we have trouble with something like $I_{3579:10}(\theta)$, where the censored blocks contain only a single order statistic. Although this equation also assumes that $(X_{1:n}, X_{2:n})$ and $(X_{n-1:n}, X_{n:n})$ are censored, a minor adjustment to this process allow right and middle censored or left and middle censored to similarly be written in terms of τ_{ij}.

Zheng and Gastwirth also reformulate their result using matrices as follows. With $\mathbf{X}$ defined in (2.20), define the $n \times n$ symmetric matrix $\tau = (\tau_{ij})_{n\times n}$ as

$$I_{\mathbf{X}}(\theta) = \mathbf{1}' \left(\mathbf{W}_{1(i_1-1);\cdots;(i_p+k_p+1)n} \otimes \tau\right) \mathbf{1},$$

where $\mathbf{W} \otimes \tau$ is the entry-wise multiplication of the two $n \times n$ matrices, $\mathbf{1}$ is the $1 \times n$ vector of 1's, and

$$\mathbf{W}_{1\ (i_1-1);\cdots;(i_p+k_p+1)\ n} = \begin{bmatrix} \mathbf{C}_{i_1-2} & 0 & & \cdots & & 0 \\ 0 & \mathbf{I}_{k_1+1} & & & & \\ & & \mathbf{C}_{i_2-i_1-k_1-2} & & & \vdots \\ \vdots & & & \ddots & & \\ & & & & \mathbf{I}_{k_p+1} & 0 \\ 0 & & \cdots & & 0 & \mathbf{C}_{n-i_p-k_p-1} \end{bmatrix}.$$

$\mathbf{I}_a$ is the $a \times a$ identity matrix, $\mathbf{C}_b$ is a $(b+1) \times (b+1)$ matrix with 1 in off-diagonal entries and 0 in diagonal entries.

The description of this matrix formulation given by Zheng, Balakrishnan, and Park [18] for middle censored data provides a more intuitive way of understanding the computation of Fisher information using matrices. Suppose the smallest r and largest $(n-s+1)$ order statistics are available. We have three blocks of consecutive order statistics,

$$b_1 = (X_{1:n}, \ldots, X_{r:n}),$$

$$b_2 = (X_{r+1:n}, \ldots, X_{s-1:n}),$$

$$b_3 = (X_{s:n}, \ldots, X_{n:n}),$$

where b_1 and b_3 are the observed tails and b_2 is censored from the sample. To find the Fisher information $I_{1\cdots rs\cdots n:n}(\theta),\ r+2<s$, the matrix τ is partitioned into a 3×3 matrix $\tau = (\tau^{kk})_{3\times 3}$, where $(\tau^{11})_{r\times r}$, $(\tau^{22})_{(s-r-1)\times(s-r-1)}$, and $(\tau^{33})_{(n-s+1)\times(n-s+1)}$ are diagonal submatrices corresponding to b_1, b_2 and b_3, respectively.

$I_{1\cdots rs\cdots n:n}(\theta)$ is the sum of three parts, each corresponding to one of the blocks of order statistics. For each observed block b_k, we take the sum of the diagonal entries of the corresponding submatrix τ^k. For the censored block b_2, we take the sum of all off-diagonal entries of τ^2 multiplied by $1/(s-r-2)$, the reciprocal of one less than the size of the block. Thus,

$$I_{1\cdots rs\cdots n:n}(\theta) = \sum_{i=1}^{r}\tau_{ii} + \sum_{i=s}^{n}\tau_{ii} + \frac{2}{s-r-2}\sum_{i=r+1}^{s-2}\sum_{j=i+1}^{s-1}\tau_{ij},$$

which is the equation given in (2.19).

This can be generalized to a multiply censored case by partitioning τ differently, following the same process. In any case, the trace of the corresponding diagonal submatrix is calulated for each observed block, and the sum of the off-diagonal elements divided by some constant is calculated for each censored block.

Normal Distribution

Consider a sample of size n from $N(\theta,1)$, and suppose we want to find the amount of information about the mean contained in any set of order statistics. Then, $\tau_{ij} = \mathrm{E}\,(X_{i:n}X_{j:n})$ [18]. Several tables for the product moments of normal order statistics are printed in the literature. For example, for a sample size of $n=10$, the information contained in the first three and last three order statistics is

$I_{1\,2\,3\,8\,9\,10:10}(\theta) = 10 - 2(\tau_{44} + \tau_{55}) + 2(2\tau_{45} + 2\tau_{46} + \tau_{47} + \tau_{56})/3 = 9.4239$, since by the symmetry of the normal pdf, $\mathrm{E}\,(X_{6:10}X_{7:10}) = \mathrm{E}\,(X_{4:10}X_{5:10}) = \mathrm{E}\,(X_{5:10}X_{4:10})$ [18].

2.3 SCATTERED ORDER STATISTICS

Consider again the problem of finding the Fisher information contained in the k order statistics $\mathbf{X} = (X_{r_1:n}, \ldots, X_{r_k:n})$ given in equation (2.13). Park [12] takes a different approach, and decomposes $I_{r_1 \cdots r_k:n}(\theta)$ as a linear combination of $I_{ij}(\theta)$.

Let $r_1 < \cdots < r_k$. By the Markov property of order statistics, $I_{r_i|r_1 \cdots r_{i-1}:n}(\theta) = I_{r_i|r_{i-1}:n}(\theta)$. Then we have the decomposition of information

$$\begin{aligned} I_{r_1 \cdots r_k:n}(\theta) &= I_{r_1 \cdots r_{k-1}:n}(\theta) + I_{r_k|r_1 \cdots r_{k-1}:n}(\theta) \\ &= I_{r_1 \cdots r_{k-1}:n}(\theta) + I_{r_k|r_{k-1}:n}(\theta) \\ &= \ldots \\ &= I_{r_1:n}(\theta) + I_{r_2|r_1:n}(\theta) + \cdots + I_{r_k|r_{k-1}:n}(\theta). \end{aligned}$$

Since $I_{r_i|r_{i-1}:n}(\theta) = I_{r_i r_{i-1}:n}(\theta) - I_{r_i:n}(\theta)$, the information in the set of order statistics $(X_{r_1:n}, ..., X_{r_k:n})$ can be expressed a sum of single and double integrals,

$$\begin{aligned} I_{r_1 \cdots r_k:n}(\theta) &= I_{r_1:n}(\theta) + (I_{r_1 r_2:n}(\theta) - I_{r_2:n}(\theta)) + (I_{r_2 r_3:n}(\theta) - I_{r_3:n}(\theta)) \\ &\quad + \cdots + (I_{r_{k-1} r_k:n}(\theta) - I_{r_k:n}(\theta)) \\ &= \sum_{i=1}^{k-1} I_{r_i r_{i+1}:n}(\theta) - \sum_{i=2}^{k-1} I_{r_i:n}(\theta). \end{aligned} \tag{2.22}$$

In this way the Fisher information in any set of OS can be represented as a linear combination of the information in pairs of order statistics.

This result and the approach used by Zheng and Gastwirth reduce the problem of finding the information in arbitrary sets of order statistics to the calculation of

double integrals. Instead of needing $\tau(i,j)$, however, Park's approach requires computation of $I_{ij:n}(\theta)$. The advantage of this method is that all $I_{ij:n}(\theta)$'s can be found if $I_{1:1}(\theta)$ exists. For example, although $I_{1:1}(\theta)$ exists for the Cauchy distribution, Park notes that the τ_{ij}'s cannot be found for the extreme order statistics of this distribution.

Park also reformulates the concept described by Zheng and Gastwirth in equation (2.15) in a more general fashion. If $\mathbf{r}_1, ..., \mathbf{r}_k$ are sets of consecutive ordered integers such that $\mathbf{r}_i \cap \mathbf{r}_j \neq \emptyset$. If the elements of these sets are the ordered ranks of order statistics, then

$$I_{\mathbf{r}_1 \cup \cdots \cup \mathbf{r}_k:n}(\theta) = \sum_{i=1}^{k} I_{\mathbf{r}_i:n}(\theta) - \sum_{i=1}^{k-1} I_{\mathbf{r}_i \cap \mathbf{r}_{i+1}:n}(\theta). \tag{2.23}$$

Park gives the example of $I_{12345:5}(\theta)$, which can be decomposed as

$$\begin{aligned} I_{12345:5}(\theta) &= I_{123:5}(\theta) + I_{345:5}(\theta) - I_{3:5}(\theta) \\ &= I_{123:5}(\theta) + I_{2345:5}(\theta) - I_{23:5}(\theta) \\ &= I_{1234:5}(\theta) + I_{345:5}(\theta) - I_{34:5}(\theta) \\ &= \ldots. \end{aligned}$$

In the case where $k = 2$, (2.23) tells us that

$$I_{\mathbf{r}_1 \cup \mathbf{r}_2:n}(\theta) = I_{\mathbf{r}_1:n}(\theta) + I_{\mathbf{r}_2:n}(\theta) - I_{\mathbf{r}_1 \cap \mathbf{r}_2:n}(\theta),$$

or,

$$I_{\mathbf{r}_1 \cup \mathbf{r}_2:n}(\theta) - I_{\mathbf{r}_2:n}(\theta) = I_{\mathbf{r}_1:n}(\theta) - I_{\mathbf{r}_1 \cap \mathbf{r}_2:n}(\theta).$$

We have seen this expression before in terms of Zheng and Gastwirth's work on multiple censoring patterns. When the block $I_{\mathbf{r}_1 \cap \mathbf{r}_2:n}$ is removed from the left, middle,

or right of an ordered sample, the change in information is the same whether we started with the set of order statistics with ranks in $\mathbf{r}_1$ or those with ranks in $\mathbf{r}_1 \cup \mathbf{r}_2$.

CHAPTER 3

COMPUTATIONAL RESULTS

We divide the computational results discussed in this chapter into two categories: Fisher information in censored samples, and the optimal spacing problem. In the first category, we consider the Fisher information in blocks of consecutive order statistics. Many of these results can be found using more than one of the approaches considered in the previous chapter. For example, the Fisher information in the first r order statistics can be expressed as a linear combination of τ_{ij}, which was derived by Mehrotra et al. [8], or in terms of the information contained in the sample minimum, as discussed by Park [10, 11].

Zheng and Gastwirth [17] extended the use of τ_{ij} to express the Fisher information in multiply censored samples, where any number of blocks of size 2 or more are removed from the sample. However, two particular censoring patterns are of primary interest when looking at symmetric distributions. By studying the asymptotic Fisher information in consecutive order statistics, Zheng and Gastwirth found that the middle portion of data from normal, logistic, and Laplace distributions contains more information about the location parameter than any other interval of the same length, although this is not always true for order statistics from a Cauchy distribution. With this in mind, it makes sense to compare the amount of information about the location parameter that is contained in the median, the middle two order statistics, and so on, when studying a symmetric distribution.

To this end, we aim to find the information contained in the block of consecutive order statistics $(X_{r:n}, ..., X_{s:n})$. Using Zheng and Gastwirth's approach, this problem can be done in two steps by first censoring on the right and then the left. The equations in (2.15) describe the change in information when a block of order statistics is removed from the sample, and gives the special case, $I_{1\cdots s:n}(\theta) - I_{r\cdots s:n}(\theta) = I_{1\cdots n:n}(\theta) - I_{r\cdots n:n}(\theta)$, when the block $(X_{1:n}, ..., X_{r-1:n})$ is further censored. Thus, for $1 \leq r < s \leq n$, the Fisher information in the middle $s - r + 1$ order statistics can be written in terms of the information in right and left censored samples as

$$I_{r\cdots s:n}(\theta) = I_{1\cdots s:n}(\theta) + I_{r\cdots n:n}(\theta) - I_{1\cdots n:n}(\theta). \tag{3.1}$$

This in turn can be expressed in terms of τ_{ij}. Alternatively, using (3.1) in conjunction with Park's simplified equations for right and left censored samples given in (2.7) and (2.10), the Fisher information in any set of consecutive order statistics can easily be computed using only the information in a sample's minimum and maximum for a specified distribution. Equation (3.1) is also a special case of Park's decomposition of information in (2.23) with $\mathbf{r}_1 = \{1, ..., s\}$ and $\mathbf{r}_2 = \{r, ..., n\}$, $r < s$.

The equations in (2.15) also yield another method of computing the Fisher information in a single order statistic,

$$\begin{aligned} I_{m:n}(\theta) &= I_{1\cdots m:n}(\theta) + I_{m\cdots n:n}(\theta) - I_{1\cdots n:n}(\theta) \\ &= I_{r\cdots m:n}(\theta) + I_{m\cdots s:n}(\theta) - I_{r\cdots s:n}(\theta), \end{aligned}$$

for $1 \leq r \leq m \leq s \leq n$. In certain cases, the information in right or left censored data may be simpler to calculate, making these expressions particularly useful.

Zheng and Gastwirth found that a different censoring pattern contains more information about the scale parameter of a symmetric location-scale distribution. For the normal, logistic, Cauchy, and Laplace distributions, they studied the information contained in the two tails. Equation (2.19) allows the Fisher information in the two tails to be calculated as a linear combination of τ_{ij}. For example, they found that 50% of the data, 25% in each tail, contains more than 95% of the total Fisher information in the sample for the normal, logistic, and Laplace distributions. For the Cauchy distribution, the tails contain more than 80% of the information.

In a later paper, Zheng and Gastwirth [16] studied the Fisher information about the scale parameter contained in two symmetric portions of the order statistics, not necessarily the extreme tails. By studying the folded distribution for these symmetric distributions, Zheng and Gastwirth noted that a single quantile in the folded distribution gives the same amount of information as two symmetric quantiles in the original distribution. Using asymptotic formulas for Fisher information and maximizing the Fisher information in a single block of order statistics from the folded distribution, they determined the conditions for two symmetric blocks of order statistics to contain the most information about the scale parameter of each of these distributions. They found that for the normal, logistic, and Laplace distributions, the extreme tails are most informative, but that for the Cauchy distribution, more information is contained around the 25^{th} and 75^{th} percentiles.

The second category looks at the Fisher information in scattered order statistics. In the optimal spacing problem, the aim is to choose a fixed number of order

statistics from a given sample of size n in order to estimate an unknown parameter. The optimal spacing problem has been studied extensively from the point of view of minimizing the asymptotic variance of the estimators. Park [12] studied this with respect to the Fisher information for the logistic distribution. Following the definition used by Park, we define the optimal choice for the parameter θ of a particular distribution to be the subset of order statistics $(X_{r_1:n}, ..., X_{r_k:n})$ which maximizes the information $I_{r_1,...,r_k:n}(\theta)$.

In this chapter we study the exponential distribution, and the location parameter of the normal and logistic distributions for different sample sizes. We compute the Fisher information contained in consecutive order statistics for each of these distributions. For the logistic distribution, we derive the information in pairs of order statistics and find the optimal choice of scattered order statistics.

3.1 EXPONENTIAL DISTRIBUTION

Let $X_1, ..., X_n$ be a random sample from $Exp(\theta)$, and denote its order statistics $X_{1:n} < ... < X_{n:n}$. First we consider the information in Type-II censored data. To find this, Arnold et al. (p. 167) [2] noted that

$$T = \sum_{i=1}^{r} X_{i:n} + (n-r)X_{r:n}$$

is the sufficient statistic. A well-known result for exponential order statistics states that $Z_1, ..., Z_n$, where

$$Z_i = (n-i+1)(X_{i:n} - X_{i-1:n}), \quad i = 1, 2, ..., n,$$

are independent and exponentially distributed. Hence T, which can be expressed in the form

$$T = \sum_{i=1}^{r} (n-i+1)(X_{i:n} - X_{i-1:n}),$$

has a $\Gamma(r, \theta)$ distribution. Since the information contained in the sufficient statistic is precisely that in the sample, $I_{1\cdots r:n}(\theta) = r/\theta^2$. Park [11] found this result using the decomposition of information given in (2.6). By the lack of memory property for the exponential distribution, $I_{r+1\cdots n|r:n}(\theta) = (n-r)/\theta^2$. Hence, $I_{1\cdots r:n}(\theta) = I_{1\cdots n:n}(\theta) - I_{r+1\cdots n|r:n}(\theta) = r/\theta^2$.

In other words, the percentage of Fisher information in any Type-II censored exponential data is equal to the percentage of data observed. This is true for the Weibull family of distributions, which Zheng [15] characterized by the Fisher information in Type-II censored data and by the factorization of the hazard function. In particular, he showed that the hazard function can be factorized as $h(x) = u(x)v(\theta)$ for some positive functions $u(x)$ and $v(\theta)$ if and only if $I_{1\cdots r:n}(\theta) = rI_X(\theta)$.

In order to find the information in left censored data, Park's approach requires the calculation of information in the sample maximum. For $i > 2$, (1.8) gives

$$I_{i:i}(\theta) = \frac{1}{\theta^2} + \frac{1}{\theta^2}\frac{i}{i-2}\mu^{(2)}_{i-2:i-1}.$$

Alternatively, by using the following recurrence relations for the standard exponential distribution (e.g., Arnold et al. [2], p.74),

$$\mu^{(2)}_{r:n} = \mu^{(2)}_{r-1:n} + \frac{2}{n-r+1}\mu_{r:n},$$
$$\mu^{(2)}_{r:n} = \mu^{(2)}_{r-1:n-1} + \frac{2}{n}\mu_{r:n}, \quad 2 \le r \le n,$$

with $r = i - 1$ and $n = i - 1$, the information in the $X_{i:i}$ can be written in terms of the first and second moments of the sample maximum for different sample sizes,

$$\begin{aligned} I_{i:i}(\theta) &= \frac{1}{\theta^2} + \frac{1}{\theta^2}\frac{i}{i-2}\mu^{(2)}_{i-1:i-1} - \frac{2}{\theta^2}\frac{i}{i-2}\mu_{i-1:i-1} \\ &= \frac{1}{\theta^2} + \frac{1}{\theta^2}\frac{i}{i-2}\mu^{(2)}_{i-2:i-2} + \frac{1}{\theta^2}\frac{i}{i-2}\frac{2}{i-1}\mu_{i-1:i-1} - \frac{2}{\theta^2}\frac{i}{i-2}\mu_{i-1:i-1} \\ &= \frac{1}{\theta^2} + \frac{1}{\theta^2}\frac{i}{i-2}\mu^{(2)}_{i-2:i-2} - \frac{2}{\theta^2}\frac{i}{i-1}\mu_{i-1:i-1}. \end{aligned}$$

The moments of exponential order statistics are easily tabulated using the expressions for the mean and variance in (1.9). Then, the information in any left censored sample can be found using equation (2.10). The information in a block of consecutive order statistics $(X_{r:n}, X_{r+1:n}, ..., X_{s:n})$ follows from equation (3.1).

For the exponential case, however, the information any single block of consecutive order statistics can be found much more simply using only the information contained in a single order statistic. Since, as Zheng and Gastwirth [17] showed, the change in information when a block is removed from the right of the sample is the same regardless of previous censoring patterns,

$$\begin{aligned} I_{r\cdots n:n}(\theta) - I_{r:n}(\theta) &= I_{1\cdots n:n}(\theta) - I_{1\cdots r:n}(\theta) \\ &= \frac{n}{\theta^2} - \frac{r}{\theta^2} \end{aligned}$$

Hence, substituting this into (3.1),

$$\begin{aligned} I_{r\cdots s:n}(\theta) &= \frac{s}{\theta^2} + I_{r\cdots n:n}(\theta) - \frac{n}{\theta^2} \\ &= I_{r:n}(\theta) + \frac{s-r}{\theta^2} \end{aligned}$$

For example, $I_{4567:10}(\theta) = I_{4:10}(\theta) + 3/\theta^2 = 6.9302/\theta^2$. In other words, the middle four order statistics contain nearly 70 percent of the information in the sample about

θ. However, since the four greatest order statistics contain 0.9377 of the information, the right extreme data are clearly more informative. We can also see that the order statistics $X_{8:10}$ and $X_{9:10}$ each contains nearly 70 percent of the information in the sample. Furthermore, $(X_{9:10}, X_{10:10})$ is the block of size 2 that gives most information about θ, with $I_{9\,10:10}(\theta) = 0.7874$. The greatest three order statistics give 0.8850 of the total information in the sample. The Fisher information in any block of consecutive order statistics is given in Table 3.1.

$r\|s$	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10
2		1.9945	2.9945	3.9945	4.9945	5.9945	6.9945	7.9945	8.9945	9.9945
3			2.9753	3.9753	4.9753	5.9753	6.9753	7.9753	8.9753	9.9753
4				3.9302	4.9302	5.9302	6.9302	7.9302	8.9302	9.9302
5					4.8399	5.8399	6.8399	7.8399	8.8399	9.8399
6						5.6734	6.6734	7.6734	8.6734	9.6734
7							6.3770	7.3770	8.3770	9.3770
8								6.8497	7.8497	8.8497
9									6.8741	7.8741
10										5.8561

Table 3.1. $I_{r\cdots s:10}(\theta)$ – Fisher information in consecutive order statistics from $Exp(\theta)$

Suppose now that we want to calculate the Fisher information in scattered, rather than consecutive, order statistics. The first approach considered in Chap-

ter 2 assumes such a set of data to be a sample from which any number of blocks of order statistics have been censored. Then, the Fisher information can be expressed as a linear combination of $\tau_{ij} = \mathrm{E}\left[\phi(X_{i:n})\phi(X_{j:n})\right]$, where ϕ is the score function. The values of τ, can easily be computed using any statistical software. The upper diagonal entries of the symmetric matrix τ are given for $n = 10$ in Table 3.2. Using the expressions for Fisher information given in (2.19) and (2.21) the Fisher information in the two tails is found simply by counting. For example, $I_{1\,2\,3\,8\,9\,10:10}(\theta) = 10-(\tau_{44}+\tau_{55}+\tau_{66}+\tau_{77})+2(\tau_{45}+\tau_{46}+\tau_{47}+\tau_{56}+\tau_{57}+\tau_{67})/3 = 9.5739$. Similarly, for a multiply censored sample, $I_{3\,4\,7:10}(\theta) = (\tau_{33}+\tau_{44}+\tau_{77})+\tau_{12}+2(\tau_{67})+2(\tau_{89})+(\tau_{810}+\tau_{910}) = 6.8953$.

0.8200	0.7200	0.6075	0.4789	0.3289	0.1489	-0.0761	-0.3761	-0.8261	-1.7261
	0.6447	0.5461	0.4334	0.3019	0.1441	-0.0531	-0.3161	-0.7105	-1.4994
		0.4787	0.3839	0.2732	0.1405	-0.0255	-0.2468	-0.5788	-1.2426
			0.3299	0.2430	0.1388	0.0086	-0.1651	-0.4256	-0.9467
				0.2117	0.1409	0.0523	-0.0659	-0.2430	-0.5974
					0.1500	0.1114	0.0599	-0.0172	-0.1716
						0.1978	0.2297	0.2775	0.3731
							0.4838	0.6983	1.1272
								1.4127	2.3417
									5.2707

Table 3.2. Matrix τ_{ij} for $Exp(\theta)$

However, although this approach works for any multiple censoring pattern

where the censored blocks are of length greater than one, finding the information in scattered order statistics such as $(X_{1:10}, X_{3:10}, X_{5:10})$, where single order statistics are removed, requires a different approach. Park [12] reduced the computation of information to finding

$$\mathrm{E}\left[\frac{\partial}{\partial\theta}\log f_{ij:n}(x_i, x_j)dx\right]^2,$$

where $f_{ij:n}$ denotes the joint distribution of $X_{i:n}$ and $X_{j:n}$. Nevertheless, even this expression is quite messy.

It has been noted that the Fisher information in any pair of order statistics about the scale parameter of the exponential distribution is equal to that of the shape parameter of the Weibull distribution [12]. Hence, the optimal choice of order statistics of size k for the Weibull case is the same as that for the exponential distribution.

3.2 NORMAL DISTRIBUTION

Consider an ordered sample from a normal distribution with unknown location parameter θ and unit variance. We follow the approach taken by Park [10] in order to find the Fisher information in blocks of consecutive order statistics. Using the expression for the Fisher information in the smallest order statistic given in (2.9), Park derived the following for the normal location parameter in terms of the standard normal pdf ϕ and cdf Φ,

$$I_{1:n}(\theta) = \int_{-\infty}^{\infty}\left\{x - \frac{\phi(x)}{1-\Phi(x)}\right\}^2 n\phi(x)(1-\Phi(x))^{n-1}dx.$$

This is computed by *Mathematica* without difficulty, and we can use Park's recurrence relation in (2.7) to compute the information contained in the order statistics $(X_{1:n}, ..., X_{r:n})$. Park tabulated these values for $n = 10$. We supply the information about θ contained in right censored samples for $n = 20$ in Table 3.3.

Since the normal distribution is symmetric, there are several facts that reduce the required computations when calculating Fisher information (e.g., Arnold et al. [2], p.27). It is well known that for a symmetric distribution, $(X_{i:n}, X_{j:n}) \stackrel{d}{=} (-X_{n-j+1:n}, -X_{n-i+1:n})$. From this, it can be shown that

$$I_{ij:n}(\theta) = I_{(n-j+1)(n-i+1):n}(\theta).$$

In general, Park [12] notes that for symmetric distributions,

$$I_{r_1 \ldots r_k:n}(\theta) = I_{(n-r_k+1)\cdots(n-r_1+1):n}(\theta), \tag{3.2}$$

where $1 \leq r_1 < \cdots < r_k \leq n$. Using this result, $I_{1\cdots 11:20}(\theta)$, for example, can be read from the table below as it equals $I_{10\cdots 20:20}(\theta)$.

r	1	2	3	4	5	6	7	8	9	10
$I_{1\cdots r:20}$	4.0702	6.6885	8.6776	10.2898	11.6428	12.8032	13.8130	14.7007	15.4867	16.1862
$I_{r\cdots n:20}$	20	19.8383	19.6185	19.3712	19.0429	18.7177	18.3090	17.8669	17.3687	16.8105

Table 3.3. $I_{r\cdots s:20}(\theta)$ – Fisher information in consecutive order statistics from $N(\theta, 1)$

Now, using the relation (3.1), the FI in any block of order statistics can be calculated with the information in right and left censored samples. For example, $I_{9\cdots 12:20} = I_{1\cdots 12:20} + I_{9\cdots 20:20} - I_{1\cdots 20:20} = 2I_{9\cdots 20:20} - I_{1\cdots 20:20} = 14.7374$. In other

words, the middle twenty percent of the data contains approximately 0.7369 of the total information about the mean that is contained in the entire sample. In comparison, we may recall from Chapter 1 that the single order statistic with the most information is the median, $X_{10:20}$ and $X_{11:20}$, which have 0.6498 of the total information.

	% Information	
% Data	Park	Z&G
10%	0.6811	0.6810
20%	0.7369	0.7368
30%	0.7867	0.7867
40%	0.8309	0.8312
50%	0.8718	0.8708

Table 3.4. Fisher information in the middle portion of a sample from $N(\theta, 1)$

This approach for calculating the information using recurrence relations is quite convenient. However, as Park warns, the accumulation of rounding errors is a cause for concern. Table 3.4 compares the calculation of information in the middle portion of an ordered sample of size 20 using this method and the multiple censoring approach taken by Zheng and Gastwirth [17] using τ_{ij}. The values given in the far right column are taken from Zheng and Gastwirth, and the values in the middle column can be read from the previous table.

3.3 LOGISTIC DISTRIBUTION

Consider a sample of size n from the pdf

$$f(x;\theta,\beta) = \frac{1}{\beta}\frac{e^{-(x-\theta)/\beta}}{(1+e^{-(x-\theta)/\beta})^2},\ \beta > 0. \tag{3.3}$$

Assuming that $\beta > 0$ is known, direct computation from the definition in (1.2) yields the Fisher information contained in a single observation about the location parameter θ. Since the cdf is given by $F(x;\theta) = 1/(1+e^{-(x-\theta)/\beta})$,

$$\begin{aligned} I_X(\theta) &= \mathrm{E}\left\{-\frac{\partial^2}{\partial\theta^2}\log\frac{1}{\beta}\,\frac{e^{-(x-\theta)/\beta}}{(1+e^{-(x-\theta)/\beta})^2}\right\} \\ &= \frac{2}{\beta^2}\,\mathrm{E}\left\{\frac{e^{-(x-\theta)/\beta}}{(1+e^{-(x-\theta)/\beta})^2}\right\} \\ &= \frac{2}{\beta^2}\int_{-\infty}^{\infty}\{f(x;\theta)\}^2 dx. \end{aligned}$$

We note that $f(x;\theta) = F(x;\theta)[1-F(x;\theta)]$, and so

$$\begin{aligned} I_X(\theta) &= \frac{2}{\beta^2}\int_{-\infty}^{\infty}F(x;\theta)[1-F(x;\theta)]f(x;\theta)dx \\ &= \frac{2}{\beta^2}\int_0^1 t(1-t)dt \\ &= \frac{1}{\beta^2}\,\frac{1}{3}. \end{aligned}$$

Hence, the information in the entire sample is

$$I_{1\cdots n:n}(\theta) = nI_X(\theta) = \frac{1}{\beta^2}\,\frac{n}{3}.$$

In the interest of calculating the information about the location parameter contained in consecutive order statistics, we first reproduce a few results from Park[10]. Without loss of generality, let the scale parameter $\beta = 1$, and notice that the hazard

function is equal to the cdf, $h(x;\theta) = F(x;\theta) = 1/(1+e^{-(x-\theta)})$. Using (2.9), the expression for the information in the sample minimum simplifies to

$$\begin{aligned}
I_{1:n}(\theta) &= \int_{-\infty}^{\infty} \left\{ \frac{\partial}{\partial\theta} \log\left[\frac{1}{1+e^{-(x-\theta)}} \right] \right\}^2 dF_{1:n} \\
&= \int_{-\infty}^{\infty} \left\{ \frac{-e^{-(x-\theta)}}{1+e^{-(x-\theta)}} \right\}^2 n\{1-F(x;\theta)\}^{n-1} f(x;\theta)dx \\
&= n\int_{-\infty}^{\infty} \left\{ \frac{e^{-(x-\theta)}}{1+e^{-(x-\theta)}} \right\}^{n+1} \frac{e^{-(x-\theta)}}{1+e^{-(x-\theta)}} dx \\
&= n\int_0^1 u^{n+1} du \\
&= \frac{n}{n+2}.
\end{aligned}$$

Similarly, we can find the information in the sample maximum from (2.12) to be

$$\begin{aligned}
I_{n:n}(\theta) &= \int_{-\infty}^{\infty} \left\{ \frac{\partial}{\partial\theta} \log\left[\frac{e^{-(x-\theta)}}{1+e^{-(x-\theta)}} \right] \right\}^2 dF_{n:n}(x;\theta) \\
&= \int_{-\infty}^{\infty} \left\{ 1 - \frac{e^{-(x-\theta)}}{1+e^{-(x-\theta)}} \right\}^2 n\{F(x;\theta)\}^{n-1} f(x;\theta)dx \\
&= \frac{n}{n+2}.
\end{aligned}$$

The recurrence relations given in equations (2.7) and (2.10) give the information contained in right and left censored samples,

$$\begin{aligned}
I_{1\cdots s:n}(\theta) &= \sum_{i=n-s+1}^{n} \binom{i-2}{n-s-1} \binom{n}{i} (-1)^{i-n+s-1} \frac{i}{i+2}, \\
I_{r\cdots n:n}(\theta) &= \sum_{i=r}^{n} \binom{i-2}{r-2} \binom{n}{i} (-1)^{i-r} \frac{i}{i+2},
\end{aligned}$$

for $1 < r \leq n$ and $1 \leq s < n$. As noted in equation (3.2), since the logistic distribution is symmetric, symmetric selections of order statistics yield equal information.

Equation (3.1) allows us to find the Fisher information in a block of consecutive order statistics $(X_{r:n}, ..., X_{s:n})$. We tabulate these values using *Mathematica* for a

sample of size 10. These results are summarize in Table 3.5, which provides the values of $I_{r\cdots s:10}(\theta)$. Many of these values can be found by symmetry. The entries for $r = s$ give the information in a single order statistic. From this table, we see that the median order statistics $X_{5:10}$ and $X_{6:10}$ each contains 0.75 of the information in the entire sample. The two middle order statistics from a sample of size $n = 10$ together contain 0.82 of the information in the entire sample. The middle four order statistics contain 0.91 of the total information about the mean, while same order statistics contain only 0.74 of the information about the mean for the normal distribution.

$r\|s$	1	2	3	4	5	6	7	8	9	10
1	0.8333	1.5152	2.0606	2.4848	2.8030	3.0303	3.1818	3.2727	3.3182	3.3333
2		1.5000	2.0455	2.4697	2.7879	3.0152	3.1667	3.2576	3.3030	3.3182
3			2.0000	2.4242	2.7424	2.9697	3.1212	3.2121	3.2576	3.2727
4				2.3333	2.6515	2.8788	3.0303	3.1212	3.1667	3.1818
5					2.5000	2.7273	2.8788	2.9697	3.0152	3.0303
6						2.5000	2.6515	2.7424	2.7879	2.8030
7							2.3333	2.4242	2.4697	2.4848
8								2.0000	2.0455	2.0606
9									1.5000	1.5152
10										0.8333

Table 3.5. $I_{r\cdots s:10}(\theta)$ – Fisher information in consecutive order statistics from $Logistic(\theta, 1)$

For the logistic distribution, these results attest to the fact that the middle portion of the ordered data provides a large proportion of the total information in the sample. The question is, how much more information can we obtain by looking at scattered order statistics? We use the decomposition of information studied by Park [12] and given in equation (2.22), which reduces the Fisher information contained in any set of order statistics to a linear combination of the information contained in pairs of order statistics and that in a single order statistic. Since the $I_{r:n}(\theta)$ has already been considered in terms of left and right censoring, we need only to derive an expression for $I_{ij:n}(\theta)$.

For a random sample of size n from a distribution $F(x;\theta)$ and density $f(x;\theta)$, the joint density of the order statistics $X_{i:n}$ and $X_{j:n}$ is given as (e.g., Arnold et al. [2], p.16)

$$\begin{aligned} f_{ij:n}(x_i, x_j; \theta) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}\, F(x_i)^{i-1}\, (F(x_j) - F(x_i))^{j-i-1} \\ \times (1 - F(x_j))^{n-j}\, f(x_i)\, f(x_j), \quad -\infty < x_i < x_j < \infty. \end{aligned} \tag{3.4}$$

Taking the derivative of the logarithm with respect to θ,

$$\frac{\partial}{\partial\theta} \log f_{ij:n}(x_i, x_j; \theta) = (n-i+1) - j\, \frac{e^{-(x_i-\theta)}}{1+e^{-(x_i-\theta)}} - (n-i+1)\, \frac{e^{-(x_j-\theta)}}{1+e^{-(x_j-\theta)}},$$

upon simplification. Then,

$$-\frac{\partial^2}{\partial\theta^2} \log f_{ij:n}(x_i, x_j; \theta) = j\, \frac{e^{-(x_i-\theta)}}{(1+e^{-(x_i-\theta)})^2} + (n-i+1)\, \frac{e^{-(x_j-\theta)}}{(1+e^{-(x_j-\theta)})^2}.$$

Taking expectations with respect to the corresponding order statistics,

$$I_{ij:n}(\theta) = j\, \mathrm{E}_{\,i:n} \left[\frac{e^{-(x_i-\theta)}}{(1+e^{-(x_i-\theta)})^2} \right] + (n-i+1)\, \mathrm{E}_{\,j:n} \left[\frac{e^{-(x_j-\theta)}}{(1+e^{-(x_j-\theta)})^2} \right].$$

Since

$$\begin{aligned}
\mathrm{E}_{\,i:n}\left[\frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}\right] &= \int_{-\infty}^{\infty} \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}\, f_{i:n}(x;\theta)\, dx \\
&= \int_{-\infty}^{\infty} \frac{n!}{(i-1)!(n-i)!} \left\{\frac{1}{1+e^{-(x-\theta)}}\right\}^{i-1} \left\{\frac{e^{-(x-\theta)}}{1+e^{-(x-\theta)}}\right\}^{n-j} \\
&\qquad \times \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}\, dx \\
&= \frac{i(n-i+1)}{(n+2)(n+1)} \int_{-\infty}^{\infty} f_{i+1:n+2}(x;\theta) dx \\
&= \frac{i(n-i+1)}{(n+2)(n+1)},
\end{aligned}$$

the Fisher information about θ contained in a pair of order statistics from a $logistic(\theta, 1)$ distribution simplifies to

$$I_{ij:n}(\theta) = \frac{j\,(n-i+1)(n-j+i+1)}{(n+2)(n+1)}. \tag{3.5}$$

Table 3.6 provides the information in (X_i, X_j) for a sample size of $n = 10$. As in previous tables, $I_{ii}(\theta)$ denotes the information contained in the single order statistic X_i. From this table, we see that the optimal choices of size 2 are $(X_{3:10}, X_{7:10})$, $(X_{4:10}, X_{7:10})$, and $(X_{4:10}, X_{8:10})$. Each choice contains 89% of the total information in the sample, compared to 82% in the two central order statistics.

Using the decomposition of information in equation (2.22), we can find the information in any collections of order statistics without further computation. For example, the information in the three arbitrary order statistics $X_{i:n} < X_{j:n} < X_{k:n}$ is given by

$$I_{ijk:n}(\theta) = I_{ij:n}(\theta) + I_{jk:n}(\theta) - I_{j:n}(\theta),$$

where the terms on the right hand side can be found in Table 3.6 for sample

size $n = 10$. Computing $I_{ijk:10}(\theta)$, we find the optimal choice of size 3 to be $(X_{2:10}, X_{5:10}, X_{8:10})$, $(X_{3:10}, X_{5:10}, X_{8:10})$, $(X_{3:10}, X_{6:10}, X_{8:10})$, and $(X_{3:10}, X_{6:10}, X_{9:10})$. Each of these collections has an information measure of 3.1364, or approximately 94 percent of the total information in the sample. These choices give more information about θ than the most informative three consecutive order statistics, $I_{4\,5\,6:10}(\theta) = I_{5\,6\,7:10}(\theta) = 2.8788$.

$i\|j$	1	2	3	4	5	6	7	8	9	10
1	0.8333	1.5152	2.0455	2.4242	2.6515	2.7273	2.6515	2.4242	2.0455	1.5152
2		1.5000	2.0455	2.4545	2.7273	2.8636	2.8636	2.7273	2.4545	2.0455
3			2.0000	2.4242	2.7273	2.9091	2.9697	2.9091	2.7273	2.4242
4				2.3333	2.6515	2.8636	2.9697	2.9697	2.8636	2.6515
5					2.5000	2.7273	2.8636	2.9091	2.8636	2.7273
6						2.5000	2.6515	2.7273	2.7273	2.6515
7							2.3333	2.4242	2.4545	2.4242
8								2.0000	2.0455	2.0455
9									1.5000	1.5152
10										0.8333

Table 3.6. $I_{ij:10}(\theta)$ – Fisher information in pairs of order statistics from $Logistic(\theta, 1)$

For the location parameter of the logistic distribution, these calculations are easily extended to larger sample sizes. Again, we use Park's recursive approach to consider a random sample of size $n = 25$. The information contained in the middle

k order statistics is given in Table 3.7 as a proportion of the total Fisher information contained in the sample. We can compare these results to the Fisher information in scattered order statistics. For example, Park [12] found the optimal choice of size 4 to be $(X_{5:25}, X_{10:25}, X_{16:25}, X_{21:25})$, with 0.9607 of the information in the sample. It is interesting to note that we would need to include the middle 15 order statistics, or 60% of the data, to get an equivalent amount of information about θ.

Although the calculations for consecutive order statistics and pairs of order statistics are easily extended to larger sample sizes, determining the optimal choice of size k poses a heavy computational burden. Even for a value as small as $k = 3$ from a sample of size 10, there are 120 possible combinations of three distinct order statistics to compare. For large sample sizes, it is necessary to consider asymptotic information, especially in regard to distributions for which $I_{ij:n}(\theta)$ is complicated.

Data	% FI
$(X_{13:25})$	0.7511
$(X_{12:25}, X_{13:25}, X_{14:25})$	0.8044
$(X_{11:25}, X_{12:25}, X_{13:25}, X_{14:25}, X_{15:25})$	0.8496
$(X_{10:25}, \ \ldots\ , X_{16:25})$	0.8872
$(X_{9:25}, \ \ldots\ , X_{17:25})$	0.9179
$(X_{8:25}, \ \ldots\ , X_{18:25})$	0.9426

Table 3.7. Fisher information in the middle portion of a sample from $Logistic(\theta, 1)$

The Fisher information about the scale parameter of a logistic distribution has

been studied in a similar manner. Zheng and Gastwirth [17] computed the Fisher information in the two tails for sample sizes of 15 and 20, and compared these results with those for normal, Cauchy, and Laplace distributions. Park [12] provided a table for $I_{ij:5}$, the information in scattered pairs from a sample of size 5. He also found the optimal choice of size 4 when $n = 25$ to be $(X_{1:25}, X_{5:25}, X_{21:25}, X_{25:25})$ with 86.77 percent of the total information.

3.4 CONCLUSION

In this paper we have studied the exact Fisher information contained in various collections of order statistics to determine which part of the ordered sample has the most information about an unknown parameter. These results may be utilized to determine censoring patterns or in the selection of efficient estimators. Some research has already been done to compare the efficiencies of the Best Linear Unbiased Estimator (BLUE) based on subsets of the order statistics and the BLUE based on the entire sample. Zheng and Gastwirth [16] studied the relative efficiency of the BLUE based on the the middle portion of the sample compared to the BLUE using the complete sample for the location parameter of certain symmetric distributions. For the normal, logistic, Cauchy, and Laplace distributions, they found that sample sizes of at least 25, 25, 85, and 30, respectively, are needed to get a relative efficiency of 90 percent. It may be interesting to compare the relative efficiency of the BLUE based on the optimal choice of k order statistics to that based on the most informative block of k consecutive order statistics.

REFERENCES

[1] Abo-Eleneen,Z.A., and Nagaraja, H.N., Fisher Information in an Order Statitic and Its Concomitant, *Annals of the Institute of Statistical Mathematics* **54** (2002), 667–680.

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

[3] Efron, B., and Johnstone, I., Fisher Information in Terms of the Hazard Rate, *Annals of Statistics* **18** (1990), 38-62.

[4] Casella, G., and Berger, R.L., *Statistical Inference*, Second Edition, Duxbury, California, 2002.

[5] Fisher, R.A., Theory of Statistical Estimation, *Proceedings of the Cambridge Philosphical Society* **22** (1925), 700–725.

[6] David, H.A., and Nagaraja, H.N., *Order Statistics*, Third Edition, John Wiley & Sons, New Jersey, 2003.

[7] Iyengar, S., Kvam, P., and Singh, H., Fisher Information in Weighted Distributions, *Canadian Journal of Statistics* **27** (1999), 833–841.

[8] Mehrotra, K.G., Johnson, R.A., and Bhattacharyya, G.K., Exact Fisher Information for Censored Samples and the Extended Hazard Rate Functions, *Communications in Statistics - Theory and Methods* **15** (1979), 1493–1510.

[9] Nagaraja, H.N., On the Information Contained in an Order Statistic, Technical Report No. 278, Department of Statistics, Ohio State University, 1983.

[10] Park, S., *Fisher Information in Order Statistics*, Ph.D. Dissertation, The University of Chicago, 1994.

[11] Park, S., Fisher Information in Order Statistics, *Journal of the American Statistical Association*, **91** (1996), 385–390.

[12] Park, S., On Calculating the Fisher Information in Order Statistics, *Statistical Papers*, **46** (2005a), 293–301.

[13] Rao, C.R., *Linear Statistical Inference and Its Applications*, Second Edition, John Wiley, New York, 1973.

[14] Tukey, J.W., Which Part of the Sample Contains the Information?, *Proceedings of the National Academy of Sciences*, USA, **53** (1965), 127–134.

[15] Zheng, G., A Characterization of the Factorization of Hazard Function by the Fisher Information Under Type II Censoring with Application to the Weibull Family, *Statistical Probability Letters* **52** (2001), 249-253.

[16] Zheng, G. and Gastwirth, J. L., Do Tails of Symmetric Distributions Contain More Fisher Information about the Scale Parameter?, *Sankyhā* Series B **64** (2002), 289–300.

[17] Zheng, G., and Gastwirth, J.L., Where Is the Fisher Information in an Ordered Sample?, *Statistica Sinica*, **10** (2000), 1267–1280.

[18] Zheng, G., Balakrishnan, N., and Park, S., Fisher Information in Ordered Data: A Review, *Statistics and Its Interface*, **2** (2009), 101–113.

VITA

Graduate School
Southern Illinois University

Mary Frances Dorn Date of Birth: January 11, 1988

911 South Valley Road, Carbondale, Illinois 62901

mfdorn@gmail.com

University of Chicago
Bachelor of Arts, Economics and Statistics, August 2009

Research Paper Title:
Evaluating Fisher Information in Order Statistics

Major Professor: Dr. S. Jeyaratnam