

An Asymptotic Approach to Progressive Censoring

GLENN HOFMANN^{*}; *Universidad de Concepción, Chile*
ERHARD CRAMER[†]; *University of Oldenburg, Germany*
N. BALAKRISHNAN[‡]; *McMaster University, Canada*
GERD KUNERT[§]; *Technische Universität Chemnitz, Germany.*

October 11, 2002

Abstract

Progressive Type-II censoring was introduced by Cohen (1963) and has since been the topic of much research. The question stands whether it is sensible to use this sampling plan by design, instead of regular Type-II right censoring. We introduce an asymptotic progressive censoring model, and find optimal censoring schemes for location-scale families. Our optimality criterion is the determinant of the 2×2 covariance matrix of the asymptotic best linear unbiased estimators. We present an explicit expression for this criterion, and conditions for its boundedness. By means of numerical optimization, we determine optimal censoring schemes for the extreme value, the Weibull and the normal distributions. In many situations, it is shown that these progressive schemes significantly improve upon regular Type-II right censoring.

Key Words and Phrases: Order statistics, progressive censoring, best linear estimation, generalized variance, optimal censoring scheme, extreme value distribution, Weibull distribution, normal distribution.

Abbreviated title: Asymptotics for Progressive Censoring.

^{*}Corresponding author; glenn@gauss.cfm.udec.cl; phone: 011-56-41-203162, fax: 011-56-41-522051; Departamento de Estadística, Facultad de Ciencias Físicas y Matemáticas, Barrio Universitario, Casilla 160-C, Concepción, Chile. Supported in part by Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) grant # 1020479 of Chile.

[†]cramer@mathematik.uni-oldenburg.de; Department of Mathematics, University of Oldenburg, 26111 Oldenburg, Germany. Supported by the Deutsche Forschungsgemeinschaft (DFG).

[‡]bala@mcmaster.ca; Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1.

[§]gerd.kunert@mathematik.tu-chemnitz.de; Dept. of Mathematics, TU Chemnitz, 09107 Chemnitz, Germany

1 Introduction

The optimal selection of order statistics $X_{n_1:n} \leq \dots \leq X_{n_m:n}$ in a sample of size n has been addressed by many authors in the literature. First considerations date back to Mosteller (1946) and the pioneering work of Ogawa (1951,1962a,b). More recent surveys are provided by Balakrishnan and Cohen (1991, Ch. 7) and Ali and Umbach (1998). The aim of this approach is to find optimal choices for the numbers n_1, \dots, n_m when the number m of selected observations and the sample size n are prescribed. Commonly, a location-scale family of distributions and best linear unbiased estimation of the parameters are considered, based on a fixed selection of m order statistics. In the second step of the procedure, the optimal selection scheme is chosen. The usual optimality criteria are the variance of the best linear unbiased estimator (BLUE) in the single parameter case and a functional of the covariance matrix of the BLUEs (like trace or determinant) in the two-parameter case. However, in order to carry out this approach, moments of order statistics of the underlying distribution must be available. This generally leads to intractable problems and serious computational difficulties.

In order to avoid these problems appearing in small sample sizes, large sample approximations have been considered to find (nearly) optimal selections. These approaches are based on a limiting result due to Mosteller (1946) which states that appropriately normalized (central) order statistics $X_{n_1:n} \leq \dots \leq X_{n_m:n}$ based on an absolutely continuous distribution function are asymptotically normal if $n_i = [nq_i] + 1$, $1 \leq i \leq m$, where $0 < q_1 < \dots < q_m < 1$ are fixed values and $[z]$ denotes the largest integer less than or equal to z . The resulting mean vector and covariance matrix are explicitly available and are used to calculate the so-called asymptotically best linear unbiased estimators (ABLUEs) of the location and/or scale parameters. Since the covariance matrix of the resulting normal distribution has a nice structure, the ABLUEs can be calculated explicitly. This results in a relatively simple expression for the variance or the covariance matrix of the ABLUEs. Hence, the search for the optimal selection is possible for specific distributions such as exponential, Weibull, extreme value, normal, and logistic distributions (cf. Ali and Umbach 1998).

Here, we apply a similar approach for determining optimal selections of progressively Type-II censored order statistics. Moreover, we discuss the problem of finding optimal censoring schemes.

Progressive Type-II right censoring has been introduced in the context of life testing by Cohen (1963) in the following manner. Before conducting a life experiment, the experimenter fixes a sample size n , a number of complete observations m , and a censoring scheme (R_1, \dots, R_m) ($n = m + \sum_{i=1}^m R_i$). The n units are placed simultaneously on the life test. Immediately following the first failure, R_1 surviving units are randomly chosen and removed from the experiment. Immediately after the second failure, R_2 items are withdrawn, and so on. The procedure is continued until all R_m remaining units are removed after the m th failure. However, this model is often used as an approximation in case of missing data, i.e., items drop out for reasons not under study and, on that score, are not considered in the statistical analysis. For a detailed account and results on progressive censoring, we refer to the recent book by Balakrishnan and Aggarwala (2000).

If progressive Type-II censoring is to be used by design, it will naturally be of interest to know the optimal values of R_1, \dots, R_m for fixed m and n . Balakrishnan and Aggarwala (2000) considered this problem in the location-scale case by evaluating criteria based on the 2×2 variance-covariance matrix of the BLUEs. However, the moments of the progressively Type-II censored order statistics, and therefore the covariance matrix of the BLUEs, can only be obtained numerically. In order to find the best censoring scheme, one therefore needs to perform the calculations for all possible combinations of R_1, \dots, R_m . Even for moderate m , the computational effort in terms of time and cost is prohibitive. The small scale study ($m = 2, \dots, 6$) of Balakrishnan and Aggarwala (2000) for various distributions and the explicit results of Burkschat et al. (2002) for generalized Pareto distributions led us to suspect that in an optimal scheme only one of the R_i 's is non-zero. While it would be hard to find optimal patterns in the small sample case, we can present some results using an asymptotic approach.

In this paper, we tackle the problem of determining an optimal censoring scheme. We

introduce the asymptotic model in Section 2 and discuss some distributional properties in Section 3. Thereafter, we will focus on location-scale families and simultaneous estimation of both parameters. In Section 4, the ABLUEs are derived. The determinant of the variance-covariance matrix (the so-called generalized variance) of the ABLUEs will serve as criterion in Section 5 to obtain the optimal censoring scheme. It turns out that the generalized variance is bounded only for some specific distributions including extreme value, normal and particular Weibull distributions. Numerical results for these distributions are presented in Section 6. The criterion is unbounded for some important lifetime distributions, e.g., exponential, uniform and Pareto distributions. However, this is not surprising since the same observation has been reported in the literature for the exponential distribution (cf. Sarhan et al. 1963, Saleh and Ali 1966, and Ali and Umbach 1998). In these cases, the generalized variance is unbounded if q_1 tends to 0. Moreover, the asymptotic result of Mosteller (1946) assumes that all q_i are positive and the limiting result fails if $q_1 = 0$. The same comment applies to our limiting result presented in Theorem 3.2. This behavior of the generalized variance leads to an improper solution of the problem for some distributions. In these cases, the minimum of the observations seems to be important for the inference (may even be a sufficient statistic like in the exponential case). In order to avoid this difficulty, Saleh and Ali (1966) proposed to use $q_1 = 1/(n + 1/2)$ for arbitrary m , which means that the minimum of the sample is included in the estimation process. However, due to these problems we exclude these distributions and consider only those leading to a bounded generalized variance. Criteria to check boundedness are provided in Theorem 5.1. In Example 5.2, we establish the boundedness of generalized variance for extreme value, normal and particular Weibull distributions. Results in the case of order statistics in these setups are given by Chan and Mead (1971) (extreme value), Chan and Chan (1973) (normal), and Hassanein (1971) (Weibull). Note that optimality criteria other than the generalized variance have also been applied to find optimal selection schemes. For instance, Hassanein (1969) used the trace of the covariance matrix (the total variance) for an underlying extreme value distribution. For more examples, we refer the readers to Ali and Umbach (1998).

In the unrestricted case, we will see that our problem of finding the asymptotically optimal censoring scheme is equivalent to choosing the best asymptotic order statistics. This result leads us to introduce a restriction on the proportion of items which should not be destroyed in the life test. Namely, we introduce a number $\tau \in (0, 1]$ such that the proportion of progressively censored units is greater than $1 - \tau$. However, it could be interpreted in the sense that at most a proportion of τ items is observed. It turns out that true progressive censoring schemes emerge as being optimal, resulting in significantly smaller estimator variances than could be achieved with regular Type-II right censoring.

2 The asymptotic model

A total number of n independent units are placed on a life test. Let the corresponding failure times X_1, \dots, X_n be identically distributed with absolutely continuous distribution function F and density function f . Before the start of the experiment, R_0 units are removed, meaning that we do allow for a reduced sample of size $n - R_0$ if it could provide an optimal inference compared to the inference based on progressively censored samples. Then \bar{R}_1 failures (as first block) are observed. Immediately after the \bar{R}_1 -th failure, R_1 surviving units are removed. This process continues until all remaining R_m units are removed after the m -th observed block, i. e., after $\bar{R}_1 + \dots + \bar{R}_m$ failures. Therefore, $R_0 + \sum_{i=1}^m (\bar{R}_i + R_i) = n$. Censoring of R_i items takes place at the end of the i -th block of \bar{R}_i observed failures. Note that we allow a reduction of the sample size to $n - R_0$, even before the start of the experiment.

Let the number of blocks m be fixed, and $n \rightarrow \infty$. We denote the proportional block sizes and the proportional censoring sizes by

$$\bar{\lambda}_i = \lim_{n \rightarrow \infty} \frac{\bar{R}_i}{n}, \quad i = 1, \dots, m; \quad \lambda_i = \lim_{n \rightarrow \infty} \frac{R_i}{n}, \quad i = 0, \dots, m,$$

respectively. We assume that $\lambda_i \in [0, 1)$ and that $\bar{\lambda}_i \in (0, 1)$ for $1 \leq i \leq m$. Clearly, $\lambda_0 + \sum_{i=1}^m (\bar{\lambda}_i + \lambda_i) = 1$. In order to get a nondegenerate distribution, we have to ensure that the last included observation is a central (progressively Type-II censored) order statistic. Hence, we assume that a positive proportion of units is right censored at the end of the

experiment, i. e., $\lambda_m > 0$. Further, let

$$\begin{aligned}
S_i &= \sum_{j=i}^m (\bar{R}_j + R_j), \quad \delta_i = \sum_{j=i}^m (\bar{\lambda}_j + \lambda_j), \quad i = 1, \dots, m; \\
p_i &= 1 - \frac{\bar{\lambda}_i}{\delta_i}, \quad i = 1, \dots, m; \\
t_0 &= 1 - \lambda_0, \quad t_i = 1 - \frac{\lambda_i}{\delta_i - \bar{\lambda}_i}, \quad i = 1, \dots, m - 1.
\end{aligned} \tag{1}$$

Therefore, $p_i \in (0, 1)$, $t_i \in (0, 1]$ for all i , without additional restrictions. The proportion of remaining units still in the experiment before the i -th block of failures is represented by δ_i . The proportion of the i -th block of failures when compared to all remaining observations at this point is denoted by p_i . Similarly, t_i stands for the proportion of remaining items being censored after the i -th block.

It should be mentioned that if no progressive censoring and no reduction of the sample size takes place, $R_i = \lambda_i = 0$ for $i = 0, \dots, m - 1$. This corresponds to the setting $t_0 = \dots = t_{m-1} = 1$, which will be important later on.

Let $Y_{i:m:n}$ be the last failure time observed in the i -th block. We consider the data consisting of the progressive block Type-II censored sample $(Y_{1:m:n}, \dots, Y_{m:m:n})$. It is important to note here that for the estimation, we are utilizing only the largest observed failure time within each block and not all the failures in each block.

3 Distribution theory

Under the scheme defined above, the joint density of $Y_{1:m:n}, \dots, Y_{m:m:n}$ is given by

$$\begin{aligned}
f_{Y_{1:m:n}, \dots, Y_{m:m:n}}(y_1, \dots, y_m) &= c \prod_{i=1}^m (F(y_i) - F(y_{i-1}))^{\bar{R}_i - 1} (1 - F(y_i))^{R_i} f(y_i), \\
& y_1 \leq y_2 \leq \dots \leq y_m, \tag{2}
\end{aligned}$$

where $F(y_0) = 0$, $c = \prod_{i=1}^m \frac{S_i!}{(\bar{R}_i - 1)!(S_i - \bar{R}_i)!}$. Through integration of (2) over y_m, \dots, y_{i+1} , it can be seen that the marginal distribution of $Y_{i:m:n}$ is free of $\bar{R}_{i+1}, R_{i+1}, \dots, \bar{R}_m, R_m$.

In equivalence to regular progressive Type-II censoring, $(Y_{1:m:n}, \dots, Y_{m:m:n})$ is a Markov sequence (see Balakrishnan and Aggarwala, 2000, ch. 2). Further, we have the following result.

Theorem 3.1 *Given $Y_{i:m:n} = y$, the subsample $(Y_{i+1:m:n}, \dots, Y_{i+j:m:n})$ forms a progressive block Type-II censored sample with censoring scheme $(R'_0 = n - S_{i+1}, \bar{R}'_1 = \bar{R}_{i+1}, R'_1 = R_{i+1}, \dots, \bar{R}'_j = \bar{R}_{i+j}, R'_j = S_{i+j} - \bar{R}_{i+j})$ from the left truncated distribution $\frac{F(x)}{1-F(y)}$, $x > y$.*

Proof: See Balakrishnan and Aggarwala (2000), Theorems 2.3 and 2.5.

First, we will consider the case when the underlying distribution is exponential. Let $F^*(x) = 1 - e^{-x}$, $x > 0$. We denote the resulting progressive block Type-II censored sample by $(Y_{1:m:n}^*, \dots, Y_{m:m:n}^*)$. Let $V_1 = Y_{1:m:n}^*$, $V_2 = Y_{2:m:n}^* - Y_{1:m:n}^*$, \dots , $V_m = Y_{m:m:n}^* - Y_{m-1:m:n}^*$ be the spacings. Applying a density transformation to (2), we obtain the joint density of the V_i 's as

$$f_{V_1, \dots, V_m}(v_1, \dots, v_m) = \prod_{i=1}^m \frac{S_i!}{(\bar{R}_i - 1)!(S_i - \bar{R}_i)!} (F^*(v_i))^{\bar{R}_i - 1} (1 - F^*(v_i))^{S_i - \bar{R}_i},$$

where $v_1, \dots, v_m \geq 0$. Hence, the spacings are independent and V_i is equal in distribution to the \bar{R}_i -th order statistic in a sample of size S_i from a standard exponential distribution. It follows (see, e.g., Arnold et al., 1992) that

$$V_i \stackrel{d}{=} \frac{Z_{\rho_{i-1}+1}}{S_i} + \dots + \frac{Z_{\rho_i}}{S_i - \bar{R}_i + 1}, \quad (3)$$

where $\rho_{i-1} = \sum_{j=1}^{i-1} \bar{R}_j$, $1 \leq i \leq m$, and Z_1, \dots, Z_{ρ_m} are iid standard exponential random variables.

As $n \rightarrow \infty$, recall that $\bar{R}_i \sim n\bar{\lambda}_i$ and $S_i \sim n\delta_i$. Since

$$\int_{S_i - \bar{R}_i + 1}^{S_i + 1} \frac{dx}{x} \leq EV_i \leq \int_{S_i - \bar{R}_i}^{S_i} \frac{dx}{x},$$

and the limits of both bounds are equal, we obtain

$$\lim_{n \rightarrow \infty} EV_i = -\log \left(1 - \frac{\bar{\lambda}_i}{\delta_i} \right) = -\log p_i = e_i. \quad (4)$$

Similarly, we can derive

$$\lim_{n \rightarrow \infty} n \text{Var}(V_i) = \frac{1}{\delta_i - \bar{\lambda}_i} - \frac{1}{\delta_i} = v_i. \quad (5)$$

We obtain the asymptotic joint distribution of the appropriately normalized progressive block Type-II censored order statistics $Y_{1:m:n}, \dots, Y_{m:m:n}$ by applying the central limit theorem and the δ -method. Subsequently, vectors are denoted by the respective bold letters, e.g., $\mathbf{Y} = (Y_{1:m:n}, \dots, Y_{m:m:n})'$. Furthermore, $\Sigma = \text{diag}(v_1, \dots, v_m)$, $I_m = \text{diag}(1, \dots, 1)$, and $D = (d_{ij})_{i,j}$ is a lower triangular matrix such that $d_{ij} = 1$, $1 \leq i \leq j \leq m$, and $d_{ij} = 0$ otherwise.

Theorem 3.2 *Let $(Y_{1:m:n}, \dots, Y_{m:m:n})$ be a progressive block Type-II censored sample drawn from an absolutely continuous differentiable distribution function F with density function f . Moreover, let $u_i = F^{-1}(1 - \prod_{j=1}^i p_j)$ and $\Delta = \text{diag}(\Delta_1, \dots, \Delta_m)$ with $\Delta_i^{-1} = f(u_i) / \prod_{j=1}^i p_j$, $1 \leq i \leq m$. Suppose that $f(u_i) \neq 0$, $1 \leq i \leq m$, and that the following condition holds:*

$$\sqrt{n} \left(-\log p_i - \sum_{j=1}^{\bar{R}_i} \frac{1}{S_i - j + 1} \right) \xrightarrow{n \rightarrow \infty} 0, \quad 1 \leq i \leq m. \quad (6)$$

Then,

$$\sqrt{n}(\mathbf{Y} - \mathbf{u}) \xrightarrow{\mathcal{L}} \mathcal{N}_m(\mathbf{0}, \Delta D \Sigma D' \Delta).$$

Proof: We start with a progressive block Type-II censored sample \mathbf{Y}^* from an exponential sample. Applying the transformation $\mathbf{V} = D^{-1}\mathbf{Y}^*$, we conclude from the spacings property in (3) that the asymptotic distribution of \mathbf{V} is the product measure of the asymptotic marginal distributions. Therefore, it is sufficient to derive the asymptotic distribution of V_i , $1 \leq i \leq m$. Applying Lyapunov's condition to the centered random variable $V_i - EV_i$, we obtain from (3) that

$$\begin{aligned} L &= \frac{1}{(\text{Var}(V_i))^{3/2}} \sum_{j=1}^{\bar{R}_i} E \left| \frac{Z_{\rho_{i-1+j}}}{S_i - j + 1} - E \frac{Z_{\rho_{i-1+j}}}{S_i - j + 1} \right|^3 \\ &= \frac{E|Z_1 - 1|^3}{(\text{Var}(V_i))^{3/2}} \sum_{j=1}^{\bar{R}_i} \frac{1}{(S_i - j + 1)^3}. \end{aligned} \quad (7)$$

Since the ratios $1/(S_i - j + 1)^3$ are decreasing in j , the sum is bounded from above by $\bar{R}_i/(S_i - \bar{R}_i + 1)^3$. Similarly, the variance $Var(V_i) = \sum_{j=1}^{\bar{R}_i} 1/(S_i - j + 1)^2$ is bounded from below by \bar{R}_i/S_i^2 . Hence, L tends to 0

$$0 \leq L \leq \frac{1}{\sqrt{\bar{R}_i}} \frac{S_i^3}{(S_i - \bar{R}_i + 1)^3} \xrightarrow{n \rightarrow \infty} 0.$$

Lyapunov's condition is sufficient for the central limit theorem to hold, and therefore $(V_i - EV_i)/\sqrt{Var(V_i)}$ is asymptotically standard normal. Using Slutsky's lemma and (6), the expectation and the variance can be replaced by their asymptotic equivalents such that $\sqrt{n}(V_i - e_i)/\sqrt{v_i} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$. Combining the previous results, we have shown that $\sqrt{n}\Sigma^{-1/2}(\mathbf{V} - \mathbf{e}) \xrightarrow{\mathcal{L}} \mathcal{N}_m(\mathbf{0}, I_m)$ which readily implies $\sqrt{n}(\mathbf{Y}^* - D\mathbf{e}) \xrightarrow{\mathcal{L}} \mathcal{N}_m(\mathbf{0}, D\Sigma D')$. Let us now introduce the functions $G(t) = F^{-1}(1 - \exp(-t))$, $t \geq 0$, and $\mathbf{G} : \mathfrak{R}^m \rightarrow \mathfrak{R}^m$ with $\mathbf{G}_i(\mathbf{x}) = G(x_i)$, $1 \leq i \leq m$. Then, $\mathbf{Y} = \mathbf{G}(\mathbf{Y}^*)$. Applying the δ -method (see, e.g., Sen and Singer 1993), we find that

$$\sqrt{n}(\mathbf{G}(\mathbf{Y}^*) - \mathbf{G}(D\mathbf{e})) \xrightarrow{\mathcal{L}} \mathcal{N}_m(\mathbf{0}, \Delta D\Sigma D'\Delta),$$

where $\Delta = \left(\frac{\partial}{\partial x_j} \mathbf{G}_i(D\mathbf{e}) \right)_{i,j}$. From the definition of \mathbf{G} , it follows that the partial derivatives are 0 if $i \neq j$. Otherwise, they are given by $\frac{\partial}{\partial x_j} \mathbf{G}_j(\mathbf{x}) = G'_j(x_j) = e^{-x_j}/f(F^{-1}(1 - e^{-x_j}))$. It remains to be shown that u_i and Δ_i have the claimed representation. However, since the i -th component of the vector $D\mathbf{e}$ is given by $\sum_{j=1}^i -\log p_j = -\log \prod_{j=1}^i p_j$, the assertion is obvious. \circ

Remark 3.1 *Condition (6) could be replaced by*

$$\sqrt{n} (\log(1 - \bar{R}_i/S_i) - \log p_i) \xrightarrow{n \rightarrow \infty} 0.$$

If we assume that $\bar{R}_i = [n\bar{\lambda}_i] + 1$, $R_i = [n\lambda_i] + 1$, $1 \leq i \leq m$, for given values of λ_i and $\bar{\lambda}_i$, respectively, the previous assertion holds. This could be carried out by using the approximation $[n\gamma] + 1 = n\gamma + \mathcal{O}(1)$, $\gamma > 0$.

Remark 3.2 *If $\lambda_1 = \dots = \lambda_{m-1} = 0$, i.e., no progressive censoring takes place, it is easy to show that $\prod_{j=1}^i p_j = \sum_{j=i+1}^m \bar{\lambda}_j + \lambda_m = \delta_i = 1 - q_i$, say, $1 \leq i \leq m-1$, and $\delta_m = \lambda_m > 0$,*

$q_m = 1 - \lambda_m$. Hence, the preceding result yields

$$u_i = F^{-1} \left(1 - \prod_{j=1}^i p_j \right) = F^{-1}(q_i) \quad \text{and} \quad \Delta_i = \frac{1 - q_i}{f(F^{-1}(q_i))}, \quad 1 \leq i \leq m.$$

The variances v_i are given by

$$v_i = \frac{1}{1 - q_i} - \frac{1}{1 - q_{i-1}}, \quad 1 \leq i \leq m, \quad q_0 = 0.$$

Combining these results, according to the limiting distribution given in Theorem 3.2, we arrive at Mosteller's (1946) expressions for the mean and covariance matrix.

4 Asymptotic BLUE's

We now apply the result from the last section to derive the asymptotic BLUEs (ABLUEs) within a location-scale family for a given distribution function F :

$$\mathcal{F} = \left\{ F_{\mu, \sigma} = F \left(\frac{\cdot - \mu}{\sigma} \right) \mid \mu \in R, \sigma > 0 \right\}.$$

Let \mathbf{Y} be a progressive block Type-II censored sample from an absolutely continuous differentiable distribution function $F_{\mu, \sigma}$ with location parameter μ and scale parameter σ . Then \mathbf{Y} can be expressed as $\mathbf{Y} = \mu \mathbf{1} + \sigma \tilde{\mathbf{Y}}$, where $\tilde{\mathbf{Y}}$ is a progressive block Type-II censored sample from F . Equivalently, we can represent \mathbf{Y} using the linear model $\mathbf{Y} = \mu \mathbf{1} + \sigma E \tilde{\mathbf{Y}} + \varepsilon$, where $E\varepsilon = \mathbf{0}$ and $Cov(\varepsilon) = \sigma^2 Cov(\tilde{\mathbf{Y}})$.

We now replace the exact expressions by the asymptotic ones, and obtain

$$\mathbf{Y} = \mu \mathbf{1} + \sigma \mathbf{u} + \tilde{\varepsilon},$$

where $E\tilde{\varepsilon} = \mathbf{0}$ and $Cov(\tilde{\varepsilon}) = \sigma^2 W = \frac{1}{n} \sigma^2 \Delta D \Sigma D' \Delta$. Based on this approximation, it follows directly from the Gauss-Markov theorem that the ABLUE has the form

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma} \end{pmatrix} = (B'W^{-1}B)^{-1} B'W^{-1}\mathbf{Y},$$

with covariance matrix $\frac{\sigma^2}{n} (B'W^{-1}B)^{-1}$, $B = (\mathbf{1}, \mathbf{u})$. Since all matrices in the product $\Delta D \Sigma D' \Delta$ are regular, $W^{-1} = \Delta^{-1} D^{-1} \Sigma^{-1} D^{-1} \Delta^{-1}$. D and Σ are diagonal matrices and

$D^{-1} = (d^{ij})_{i,j}$ is given by $d^{ii} = 1$, $1 \leq i \leq m$, $d^{i+1,i} = -1$, $1 \leq i \leq m-1$, and $d^{ij} = 0$ otherwise. Straightforward calculations yield that W^{-1} is a symmetric tridiagonal matrix with elements

$$\begin{aligned} w^{i,i} &= \Delta_i^{-2}(v_i^{-1} + v_{i+1}^{-1}), \quad 1 \leq i \leq m-1, & w^{m,m} &= \Delta_m^{-2}v_m^{-1}, \\ w^{i,i+1} = w^{i+1,i} &= -\Delta_i^{-1}\Delta_{i+1}^{-1}v_{i+1}^{-1}, \quad 1 \leq i \leq m-1, \\ w^{i,j} &= 0, \quad \text{otherwise.} \end{aligned}$$

For any $\mathbf{x}, \mathbf{y} \in \mathfrak{R}^m$, it follows that

$$\mathbf{x}'W^{-1}\mathbf{y} = \sum_{j=1}^m v_j^{-1}(x_j\Delta_j^{-1} - x_{j-1}\Delta_{j-1}^{-1})(y_j\Delta_j^{-1} - y_{j-1}\Delta_{j-1}^{-1}), \quad \Delta_0^{-1} = 0.$$

By choosing the particular values $\mathbf{1}, \mathbf{u}$ for \mathbf{x}, \mathbf{y} , we readily have the components of $B'W^{-1}B$ as

$$\begin{aligned} K_1 &= \mathbf{1}'W^{-1}\mathbf{1} = \sum_{j=1}^m v_j^{-1}(\Delta_j^{-1} - \Delta_{j-1}^{-1})^2, \\ K_2 &= \mathbf{u}'W^{-1}\mathbf{u} = \sum_{j=1}^m v_j^{-1}(u_j\Delta_j^{-1} - u_{j-1}\Delta_{j-1}^{-1})^2, \\ K_3 &= \mathbf{1}'W^{-1}\mathbf{u} = \sum_{j=1}^m v_j^{-1}(\Delta_j^{-1} - \Delta_{j-1}^{-1})(u_j\Delta_j^{-1} - u_{j-1}\Delta_{j-1}^{-1}). \end{aligned} \tag{8}$$

This leads to the expression

$$\frac{\sigma^2}{n}(B'W^{-1}B)^{-1} = \frac{\sigma^2}{n} \frac{1}{K_1K_2 - K_3^2} \begin{pmatrix} K_2 & -K_3 \\ -K_3 & K_1 \end{pmatrix}$$

for the covariance matrix of the ABLUE's. Hence, the generalized variance of the ABLUE's (the determinant of their covariance matrix) is given by

$$\left| \text{Var} \begin{pmatrix} \hat{\mu} \\ \hat{\sigma} \end{pmatrix} \right| = \frac{\sigma^2}{n} (K_1K_2 - K_3^2)^{-1}.$$

It should be mentioned that in the case of order statistics, similar quantities K_j arise (see Ali and Umbach 1998, p. 188). These appeared first in Ogawa (1951, p. 190–191).

5 Optimal censoring schemes and their relation to the equivalent order statistics problem

In this section, our goal is to determine optimal values for λ_i and $\bar{\lambda}_i$. One common measure of optimality in this case is minimizing the determinant of the variance-covariance matrix of the ABLUE's. In our setting, minimizing this determinant is equivalent to maximizing the generalized variance $\mathcal{D} = K_1 K_2 - K_3^2$. From (8), we can derive the following expression:

$$\mathcal{D} = \sum_{1 \leq j < k \leq m} v_j^{-1} v_k^{-1} \left((\Delta_j^{-1} - \Delta_{j-1}^{-1})(u_k \Delta_k^{-1} - u_{k-1} \Delta_{k-1}^{-1}) - (\Delta_k^{-1} - \Delta_{k-1}^{-1})(u_j \Delta_j^{-1} - u_{j-1} \Delta_{j-1}^{-1}) \right)^2,$$

where $\Delta_0^{-1} = 0$. By recalling that $\Delta_i^{-1} = f(u_i) / \prod_{j=1}^i p_j$ and noticing that v_i^{-1} can be written as

$$v_i^{-1} = \frac{p_i t_0}{1 - p_i} \prod_{j=1}^{i-1} (p_j t_j), \quad 1 \leq i \leq m,$$

we obtain the following representation:

$$\mathcal{D} = t_0^2 \sum_{1 \leq j < k \leq m} \frac{1}{(1 - p_j) p_j (1 - p_k) p_k} \prod_{i=1}^{j-1} (p_i^{-1} t_i) \prod_{i=1}^{k-1} (p_i^{-1} t_i) (U_{jk} - U_{kj})^2, \quad (9)$$

where $U_{jk} = [f(u_j) - p_j f(u_{j-1})][u_k f(u_k) - p_k u_{k-1} f(u_{k-1})]$, $1 \leq j < k \leq m$, and $f(u_0) = 0$.

Since the parameters $\lambda_0, \dots, \lambda_m, \bar{\lambda}_1, \dots, \bar{\lambda}_m$ ($\lambda_0 + \sum_{i=1}^m (\bar{\lambda}_i + \lambda_i) = 1$) are uniquely specified by $p_1, \dots, p_m, t_0, \dots, t_{m-1}$, i.e.,

$$\bar{\lambda}_i = t_0 (1 - p_i) \prod_{j=1}^{i-1} (p_j t_j), \quad \lambda_i = t_0 (1 - t_i) p_i \prod_{j=1}^{i-1} (p_j t_j), \quad 1 \leq i \leq m, \quad (10)$$

and vice versa (see (1)), the optimization can be carried out w.r.t. $p_1, \dots, p_m \in (0, 1)$ and $t_0, \dots, t_{m-1} \in (0, 1]$.

It is clear from (9) (and the fact that \mathcal{D} is the determinant of a positive definite matrix) that \mathcal{D} is positive on the considered region of the parameters. However, in some cases \mathcal{D} can be unbounded. The following theorem gives conditions on the underlying distribution function F for \mathcal{D} to be bounded.

Theorem 5.1 *Let $p_1, \dots, p_m \in (0, 1)$ and F be an absolutely continuous distribution function with support (s_f, S_F) , $-\infty \leq s_f < S_F \leq \infty$, and continuous density f on this support. For $t \in (0, 1]$, let*

$$g_t(x) = f(F^{-1}(1 - tx)), \quad h_t(x) = F^{-1}(1 - tx)f(F^{-1}(1 - tx)), \quad x \in (0, 1).$$

Further, let us introduce the following conditions:

$$S_F < \infty \text{ and } \lim_{x \rightarrow 0} \frac{g_1(x)}{\sqrt{x}} = a < \infty, \quad (\text{D1})$$

$$S_F = \infty \text{ and } \lim_{x \rightarrow 0} \frac{h_1(x)}{\sqrt{x}} = b < \infty, \quad (\text{D2})$$

$$s_f > -\infty \text{ and } \lim_{x \rightarrow 1} \frac{g_1(x)}{\sqrt{1-x}} = c < \infty, \quad (\text{D3})$$

$$s_f = -\infty \text{ and } \lim_{x \rightarrow 1} \frac{h_1(x)}{\sqrt{1-x}} = d < \infty, \quad (\text{D4})$$

$$\lim_{x \rightarrow 1} \frac{g_t(x) - g_t(1)}{\sqrt{1-x}} = 0 \text{ for all } t \in (0, 1), \quad (\text{D5})$$

$$\lim_{x \rightarrow 1} \frac{h_t(x) - h_t(1)}{\sqrt{1-x}} = 0 \text{ for all } t \in (0, 1). \quad (\text{D6})$$

If F satisfies either (D1) or (D2), and either (D3) or (D4), and both (D5) and (D6), then \mathcal{D} is bounded. If neither (D3) nor (D4) holds, then \mathcal{D} is unbounded.

Proof: Given the assumptions of the theorem, \mathcal{D} is a continuous function of $p_1, \dots, p_m \in (0, 1)$ and $t_0, \dots, t_{m-1} \in (0, 1]$. Hence, we have only to consider boundedness on the boundary w.r.t. p_1, \dots, p_m . Since \mathcal{D} is increasing in t_0, \dots, t_{m-1} , we can give the following upper bound that does not depend on the t 's:

$$0 \leq \mathcal{D} \leq \mathcal{D}_0 = \sum_{1 \leq j < k \leq m} \frac{1}{(1-p_j)p_j(1-p_k)p_k} \prod_{i=1}^{j-1} p_i^{-1} \prod_{i=1}^{k-1} p_i^{-1} (U_{jk} - U_{kj})^2.$$

By recalling the representation $U_{jk} = [f(u_j) - p_j f(u_{j-1})][u_k f(u_k) - p_k u_{k-1} f(u_{k-1})] = a_j b_k$,

we can express \mathcal{D}_0 as

$$\begin{aligned} \mathcal{D}_0 &= \left(\sum_{j=1}^m \frac{a_j^2}{p_j(1-p_j)} \prod_{i=1}^{j-1} p_i^{-1} \right) \left(\sum_{k=1}^m \frac{b_k^2}{p_k(1-p_k)} \prod_{i=1}^{k-1} p_i^{-1} \right) - \left(\sum_{j=1}^m \frac{a_j b_j}{p_j(1-p_j)} \prod_{i=1}^{j-1} p_i^{-1} \right)^2 \\ &\leq \left(\sum_{j=1}^m \frac{a_j^2}{p_j(1-p_j)} \prod_{i=1}^{j-1} p_i^{-1} \right) \left(\sum_{k=1}^m \frac{b_k^2}{p_k(1-p_k)} \prod_{i=1}^{k-1} p_i^{-1} \right). \end{aligned}$$

Hence, we can show boundedness of \mathcal{D} by proving that both factors of the product are bounded. In the following, we only consider the case of finite support. The remaining cases of an infinite left or right endpoint of the support can be derived similarly. Since both factors are composed of finite sums, we can handle each term of the sums separately.

It is obvious from the above sums that in the j th term only the variables p_1, \dots, p_j are involved. We will start by considering the first term, which only depends on p_1 .

It is given by $\frac{1}{p_1(1-p_1)} g_1^2(p_1)$. If $p_1 \in (0, 1)$ is fixed, nothing needs to be shown. Let $p_1 \rightarrow 0$. Since by assumption (D1) $\lim_{p_1 \rightarrow 0} \frac{g_1(p_1)}{\sqrt{p_1}} < \infty$, the assertion follows. For $p_1 \rightarrow 1$, we conclude directly from (D3) that the first term is bounded. The first term of the second factor is handled in the same way by noting that the endpoints of the support are finite. Hence, the first term is bounded in any case.

In order to analyze the convergence to the boundaries of the j th term in the sum of the first factor, for $j \in \{2, \dots, m\}$, we consider the variables (p_1, \dots, p_{j-1}) and p_j separately. Since the j th term depends on the variables p_1, \dots, p_{j-1} only via the product $\prod_{i=1}^{j-1} p_i$, it is obvious that the product tends to 0 if at least one of the p_i 's, $1 \leq i \leq j-1$, tends to 0. If all these p_i 's tend to 1, the product tends to 1. Otherwise, i.e., at least one and at most $j-2$ p_i 's tend to 1, the product converges to an inner point η of the interval $(0, 1)$. Hence, we could replace the product $\prod_{i=1}^{j-1} p_i$ by a variable s and consider the limits as $s \rightarrow 0$, $s \rightarrow 1$, and $s \rightarrow \eta \in (0, 1)$.

Utilizing this simplification, the j -th term ($j \geq 2$) reads

$$S_j = \frac{1}{p_j(1-p_j)} s^{-1} (g_1(p_j s) - p_j g_1(s))^2. \quad (11)$$

First, we consider the limits for $s \rightarrow 0$. Let $p_j \in (0, 1)$ be fixed. Then (D1) yields directly the boundedness. Suppose $p_j \rightarrow 0$. Then $p_j s \rightarrow 0$ and we obtain from (11) that

$$\lim_{p_j \rightarrow 0, s \rightarrow 0} S_j = \lim_{p_j \rightarrow 0, s \rightarrow 0} \frac{1}{1 - p_j} \left(\frac{g_1(p_j s)}{\sqrt{p_j s}} - \sqrt{p_j} \frac{g_1(s)}{\sqrt{s}} \right)^2 = a^2 < \infty$$

by (D1).

The case $p_j \rightarrow 1$ follows from (D5) by using the representation

$$\lim_{p_j \rightarrow 1} S_j = \lim_{p_j \rightarrow 1} \left(\frac{g_s(p_j) - g_s(1)}{\sqrt{(1 - p_j)s}} + \sqrt{\frac{1 - p_j}{s}} g_1(s) \right)^2 = 0 \quad \text{for all } s > 0. \quad (12)$$

Hence, the limit w.r.t. $s \rightarrow 0$ is obviously 0. On the other hand, the first term in (12) converges to $-\frac{\sqrt{1 - p_j}}{1 + \sqrt{p_j}} a$ for a positive p_j by (D1) if s tends to 0. This follows from

$$\frac{g_1(p_j s) - g_1(s)}{\sqrt{(1 - p_j)s}} = \frac{\sqrt{p_j}}{\sqrt{1 - p_j}} \frac{g_1(p_j s)}{\sqrt{p_j s}} - \frac{g_1(s)}{\sqrt{(1 - p_j)s}} \xrightarrow{s \rightarrow 0} \frac{\sqrt{p_j} - 1}{\sqrt{1 - p_j}} a = -\frac{\sqrt{1 - p_j}}{1 + \sqrt{p_j}} a.$$

Letting $p_j \rightarrow 1$, the limit turns out to be 0. The second term in (12) converges to 0 as well.

Suppose now that $s \rightarrow 1$. If p_j is fixed, the result is evident since $\lim_{s \rightarrow 1} g_1(s) = 0$ by (D3). The limits $\lim_{p_j \rightarrow 1} \lim_{s \rightarrow 1} S_j$, and $\lim_{p_j \rightarrow 0} \lim_{s \rightarrow 1} S_j$ reduce to the case $j = 1$, which yields a^2 as a limit. The limit for $p_j \rightarrow 0$ leads to

$$\lim_{p_j \rightarrow 0} S_j = \lim_{p_j \rightarrow 0} \left(\frac{g_1(p_j s)}{\sqrt{p_j s}} - \sqrt{\frac{p_1}{s}} g_1(s) \right)^2 = a^2 \quad \text{for all } s > 0$$

by (D1). For $p_j \rightarrow 1$, we apply (D5) and obtain

$$\lim_{p_j \rightarrow 1} S_j = s^{-1} \left(\lim_{p_j \rightarrow 1} \frac{g_s(p_j) - p_j g_s(1)}{\sqrt{1 - p_j}} \right)^2 = 0.$$

Taking limits w.r.t. $s \rightarrow 1$ yields the desired result.

It remains to consider the case $s \rightarrow \eta \in (0, 1)$. If p_j is fixed, nothing remains to be shown.

Let $p_j \rightarrow 0$. Then

$$\lim_{p_j \rightarrow 0} \lim_{s \rightarrow \eta} S_j = \eta^{-1} \lim_{p_j \rightarrow 0} \frac{1}{p_j(1 - p_j)} (g_\eta(p_j) - p_j g_\eta(1))^2 = a^2 = \lim_{s \rightarrow \eta} \lim_{p_j \rightarrow 0} S_j.$$

For $p_1 \rightarrow 1$, we obtain from (D5) that

$$\lim_{p_j \rightarrow 1} \lim_{s \rightarrow \eta} S_j = \eta^{-1} \lim_{p_j \rightarrow 1} \left(\frac{g_\eta(p_j) - p_j g_\eta(1)}{\sqrt{1 - p_j}} \right)^2 = 0.$$

Similarly, $\lim_{s \rightarrow \eta} \lim_{p_j \rightarrow 1} S_j = 0$. In any case, the terms are bounded. The j -th term of the second factor is treated in the same way using the fact that the right and left endpoints of the support are finite.

This proves the first part of the theorem.

Suppose now that $s_F > -\infty$ and that (D3) does not hold. Obviously, \mathcal{D} can be bounded from below by

$$\frac{t_0^2}{(1 - p_1)p_1^2(1 - p_2)p_2} (U_{12} - U_{21})^2 = \frac{t_0^2}{(1 - p_1)p_1^2(1 - p_2)p_2} (a_1 b_2 - a_2 b_1)^2. \quad (13)$$

The square is equal to $(g_{p_2}(p_1)g_1(p_1)(u_2 - u_1))^2$. In order to determine the limit for $p_1 \rightarrow 1$, we observe that $u_2 - u_1 \rightarrow F^{-1}(1 - p_2) - s_F$ which is finite and positive. Hence, the only relevant part of (13) is $\frac{g_1(p_1)}{\sqrt{1 - p_1}}$. But since (D3) does not hold this, the ratio tends to infinity for $p_1 \rightarrow 1$ proving that \mathcal{D} is unbounded. The proof for $s_F = -\infty$ is along the same lines. \circ

Remark 5.1 *Assuming that f is differentiable and positive on the support of F , assumptions (D5) and (D6) are fulfilled.*

Example 5.2 (i) *The exponential and uniform distributions lead to an unbounded \mathcal{D} .*

In both cases, (D3) does not hold. For the exponential distribution, we obtain $g_1(x) = f(F^{-1}(1 - x)) = x$ and for the uniform distribution, we have $g_1(x) = f(F^{-1}(1 - x)) = 1$, $x \in (0, 1)$.

Pareto distributions with $F(x) = 1 - \frac{1}{x^\alpha}$, $x \geq 1$, $\alpha > 0$, do not fulfil (D3) either, which is directly seen from $g_1(x) = \alpha x^{1+1/\alpha}$, $x \in (0, 1)$.

(ii) *For the Weibull distribution with $F(x) = 1 - \exp(-x^\beta)$, $x \geq 0$, $\beta > 0$, \mathcal{D} is bounded iff $\beta \geq 2$. It is easy to check that (D2) is satisfied for any $\beta > 0$ using $h_1(x) = -\beta x \ln(x)$.*

Considering (D3), we make use of the representation

$$\frac{g_1(x)}{\sqrt{1-x}} = \sqrt{\frac{\ln(x)}{x-1}} \beta x (-\ln(x))^{1/2-1/\beta}.$$

Obviously, the first factor tends to 1 for $x \rightarrow 1$. Since the last factor has the limits 0, 1, and ∞ for $\beta > 2$, $\beta = 2$, and $\beta < 2$, the result follows.

(iii) \mathcal{D} is bounded for the extreme value distribution with $f(x) = \exp(x - \exp(x))$, $x \in \mathfrak{R}$. According to the preceding remark, we have to check only (D2) and (D4). Since $F^{-1}(1-x) = \ln(-\ln(x))$, we obtain $g_1(x) = f(F^{-1}(1-x)) = -x \ln(x)$, $x \in (0, 1)$. Hence, we have to show that

$$\frac{h_1(x)}{\sqrt{x}} = -\sqrt{x} \ln(x) \ln(-\ln(x)) \rightarrow 0 \quad \text{as } x \rightarrow 0.$$

The substitution $x = e^{-y}$, $y \rightarrow \infty$, leads to the equivalent condition

$$y \ln(y) \exp\{-\frac{1}{2}y\} \rightarrow 0 \quad \text{as } y \rightarrow \infty,$$

which obviously holds. To determine the limit of

$$\frac{h_1(x)}{x} = v(x) = \frac{-\ln(x) \ln(-\ln(x))}{\sqrt{1-x}} \quad \text{as } x \rightarrow 1,$$

we derive from the well-known inequality $\ln(y) \leq y - 1$ the bounds

$$0 \leq v(x) \leq \frac{\ln^2(x) + \ln(x)}{\sqrt{1-x}}$$

which are valid in $(e^{-1}, 1)$. Applying l'Hospital's rule to $\frac{\ln(x)}{\sqrt{1-x}}$, it is seen that the upper bound converges to 0 if x tends to 1. This yields $\lim_{x \rightarrow 1} v(x) = 0$, too, so that (D4) is satisfied.

(iv) The normal distribution leads to a bounded \mathcal{D} as well. Let Φ be the distribution function and φ be the density function of a standard normal distribution, respectively. We prove that

$$\lim_{x \rightarrow 0} \frac{\Phi^{-1}(1-x) \varphi(\Phi^{-1}(1-x))}{\sqrt{x}} = 0, \quad \lim_{x \rightarrow 1} \frac{\Phi^{-1}(1-x) \varphi(\Phi^{-1}(1-x))}{\sqrt{1-x}} = 0.$$

Introducing the substitution $1 - x = \Phi(y)$, we replace the preceding limits by

$$\lim_{y \rightarrow \infty} \frac{y\varphi(y)}{\sqrt{1 - \Phi(y)}}, \quad \text{and} \quad \lim_{y \rightarrow -\infty} \frac{y\varphi(y)}{\sqrt{\Phi(y)}}, \quad \text{respectively.}$$

Applying the well-known asymptotic relations $1 - \Phi(y) \sim \varphi(y)/y$ as $y \rightarrow \infty$, and $\Phi(y) \sim \varphi(y)/|y|$ as $y \rightarrow -\infty$, to these limits, we deduce the desired results:

$$\lim_{y \rightarrow \infty} \frac{y\varphi(y)}{\sqrt{1 - \Phi(y)}} = \lim_{y \rightarrow \infty} y^{3/2} \sqrt{\varphi(y)} = 0, \quad \lim_{y \rightarrow -\infty} \frac{y\varphi(y)}{\sqrt{\Phi(y)}} = \lim_{y \rightarrow -\infty} y \sqrt{|y|\varphi(y)} = 0.$$

Remark 5.2 If \mathcal{D} is unbounded, clearly the optimal censoring scheme needs to include the minimum (or maximum) of the sample, i. e., the first (last) failure. Since the minimum and maximum do not have an asymptotic normal distribution, we can not treat these cases in our setup. They will be excluded in the following.

Obviously, \mathcal{D} is increasing w.r.t. each t_i , $0 \leq i \leq m - 1$, so that the optimal solution for any given choice of p_1, \dots, p_m is given by $t_i = 1$, $0 \leq i \leq m - 1$. This implies $\lambda_i = 0$, $0 \leq i \leq m - 1$. Only λ_m is positive which corresponds to the usual Type-II right censoring, and so no progressive censoring is carried out. This implies that the optimal solution of the asymptotic progressive block Type-II censoring scheme is equivalent to the corresponding asymptotic order statistics problem. The latter is commonly defined as the determination of the quantiles $0 < \bar{\lambda}_1 < \bar{\lambda}_1 + \bar{\lambda}_2 < \dots < \bar{\lambda}_1 + \dots + \bar{\lambda}_m < 1$ such that the BLUE based on the corresponding order statistics $X_{\bar{R}_1:n}, X_{\bar{R}_1 + \bar{R}_2:n}, \dots, X_{\sum_{i=1}^m \bar{R}_i:n}$ has minimum variance (in our case, minimum generalized variance). This objective has been discussed extensively in the literature; see Ali and Umbach (1998) for a summary.

So far, we have not imposed any restriction on the total number of failures allowed to be observed. In regular progressive Type-II censoring, this number is fixed. In fact, savings in experimental units is one of the advantages of progressive censoring. In particular, the associated cost savings are often presented as a motivation for progressive Type-II censoring. It is therefore sensible to restrict the progressive block Type-II censoring scheme by imposing an upper bound on the total proportion of observed blocks of failures:

$$\sum_{i=1}^m \bar{\lambda}_i \leq \tau, \quad \tau \in (0, 1], \quad (14)$$

where $\tau = 1$ corresponds to the unrestricted case. Let us see how the optimal censoring scheme is modified by this condition. From (10), it follows that (14) is equivalent to

$$t_0 \sum_{i=1}^m (1 - p_i) \prod_{j=1}^{i-1} p_j t_j \leq \tau. \quad (15)$$

Theorem 5.3 *Let $(p_1, \dots, p_m, t_0, \dots, t_{m-1}) \in (0, 1)^m \times (0, 1]^m$ be an optimal solution to the problem of maximizing \mathcal{D} in (9) under condition (15). Then, either $t_0 = \dots = t_{m-1} = 1$ or equality holds in (15).*

Proof: Note that both \mathcal{D} and the left hand side of (15) are increasing in the t_i 's. Hence, if the optimal solution contains one $t_i < 1$ ($1 \leq i \leq m$) and (15) is strict, then replacing it by $t_i + \varepsilon$ ($\varepsilon > 0$, small) will increase \mathcal{D} while still satisfying (15). Therefore, the solution would not be optimal. \circ

6 Numerical results

We numerically solved the problem of maximization of \mathcal{D} under restriction (15) for the extreme value distribution, the Weibull distribution ($\beta = 3$) and the normal distribution.

For optimization of a nonlinear function with linear and nonlinear constraints, the standard method is *sequential quadratic programming (SQP)*; see, for example, Fletcher (1987) and Gill *et al.* (1981) for an overview. We numerically implemented it by making use of the routine `E04UCF` from the Fortran NAG library (<http://www.nag.co.uk/>). We found this algorithm preferable to a similar implementation in Matlab (routine `fmincon`) which fails to obey the interval restrictions $p_i \in (0, 1]$.

Although the optimization seems to be a standard problem at first sight, the following list gives a flavor of the difficulties encountered in our case:

- (i) The exact partial derivatives of \mathcal{D} are hard to obtain, and thus have to be approximated by finite differences. The built-in approximation of the routine `E04UCF` fails to provide acceptable results.

- (ii) For convergence, the starting point needs to be sufficiently close to an optimum (more critical for small τ and/or large m). Even then, the SQP method can only compute a *local* optimum. *Global* optimization is virtually impossible unless one performs a search over the whole range of variable combinations. This is limited to about 10 variables, corresponding to $m \leq 5$. In other cases, we ran the optimization routine with thousands of random start vectors to find the *global* optimum with sufficient certainty.
- (iii) The routine E04UCF employs several parameters that specify the *numerical accuracy* of the method. The smallest feasible values are of order 1E-8 (e.g. satisfaction of the bounds and the constraint). Accuracy issues become particularly critical for small τ , for large m . The Weibull distribution leads to the most ill-conditioned problem, with $\beta = 2$ (any m) and $\beta = 3$ ($m > 10$) being beyond computational feasibility.

Computing time on a 1.6 Ghz Pentium IV PC was between 17 hours and several days for one run of one distribution.

When considering the results, first note that for the solution of the unrestricted case ($\tau = 1$), (14) is strict. This follows from (15) since the left hand side of the inequality is bounded by

$$1 - \prod_{i=1}^m p_i < 1.$$

Let $\tau_c = \sum_{i=1}^m \bar{\lambda}_i = 1 - \prod_{i=1}^m p_i$, where the $\bar{\lambda}_i$'s (p_i 's) correspond to the optimal solution for $\tau = 1$. Obviously, the solution will not change as long as $\tau \in [\tau_c, 1]$. For $\tau < \tau_c$, our numerical evaluations always show the following behavior. If τ falls slightly below τ_c , the optimal solution for p_1, \dots, p_m changes. However, the solution for t_0, \dots, t_{m-1} remains unchanged so that we still have $\lambda_1 = \dots = \lambda_{m-1} = 0$, i.e., there is still no progressive censoring. It turns out that there exists a critical value $\tau_t < \tau_c$, such that the optimal solution gives $\lambda_1 = \dots = \lambda_{m-1} = 0$ for $\tau \in (\tau_t, 1]$, but not for $\tau < \tau_t$. The (numerically calculated) critical τ -values are presented in Tables 1–3 for different values of m .

For $\tau < \tau_t$, the optimal solutions consist in true progressive censoring (not just Type-

m	2	3	4	5	10	15	20
τ_c	.9262	.9671	.9789	.9878	.9978	.9993	.9997
τ_t	.8867	.9054	.9595	.9741	.9960	.9988	.9995

Table 1: Critical τ 's for an underlying extreme value distribution

m	2	3	4	5	10
τ_c	.8602	.8745	.9482	.9565	.9903
τ_t	.7202	.7163	.7212	.8357	.9485

Table 2: Critical τ 's for an underlying Weibull distribution ($\beta = 3$)

II right censoring). In all cases, exactly one intermediate censoring step is suggested. Therefore, such a censoring scheme would be easy to implement. Tables 4–6 show the optimal censoring schemes for $\tau = .01, .05, .1, .15, .2, .5, 1$. For $m = 2, 3, 4, 5$, all sizes $\bar{\lambda}_i$ of the observed blocks, and all non-zero censoring block sizes λ_i are given. For $m = 10, 15, 20$, we only provide the non-zero λ_i 's. The rows labelled ΔVar represent the percentage reduction in generalized variance of the optimal censoring scheme, when compared to the optimal Type-II right censoring scheme ($\lambda_1 = \dots = \lambda_{m-1} = 0$). Therefore,

$$\Delta Var = 1 - \frac{\left| Var \begin{pmatrix} \hat{\mu}_{opt} \\ \hat{\sigma}_{opt} \end{pmatrix} \right|}{\left| Var \begin{pmatrix} \hat{\mu}_{right} \\ \hat{\sigma}_{right} \end{pmatrix} \right|},$$

where $\begin{pmatrix} \hat{\mu}_{opt} \\ \hat{\sigma}_{opt} \end{pmatrix}$, $\begin{pmatrix} \hat{\mu}_{right} \\ \hat{\sigma}_{right} \end{pmatrix}$ are the ABLUE's of the overall optimal censoring scheme and the optimal right censoring scheme, respectively. Note that for fixed τ , the percentage reduction of the generalized variance by the use of optimal progressive censoring is quite stable over m . For small τ (heavy censoring), the reduction is as high as 75–90%, 30–50% and 50–70% for the extreme value distribution, the Weibull distribution and the normal distribution, respectively. In the extreme value case, the reduction remains substantial even for τ as large as .5 (35–50%), whereas for Weibull and normal it becomes quite small as τ increases. Our asymptotic results give some insights in the optimal choice of censoring schemes. We suspect that the advantages of well-chosen progressive censoring will carry over to the small sample case.

m	2	3	4	5	10	15	20
τ_c	.8666	.9167	.9551	.9701	.9941	.9979	.9990
τ_t	.7370	.7414	.8055	.8669	.9658	.9873	.9940

Table 3: Critical τ 's for an underlying normal distribution

Remark 6.1 *As mentioned before, for $\tau = 1$, the problem reduces to the order statistics case of Ogawa (1951) as summarized in Ali and Umbach (1998). Ogawa (1951) looked at the extreme value distribution and calculated optimal solutions for small m ($m = 2, 3, 4$) and our results agree with his. Hassanein (1971) considered the Weibull distribution. Numerical solution of his equations leads to the same results as presented in our tables, but his table seems to be in error. We suspect that this may be due to the considerable computational difficulties this task must have caused thirty years ago.*

References

- Ali, M. M. and Umbach, D. (1998) Optimal linear inference using selected order statistics in location-scale models. In N. Balakrishnan and C. R. Rao *Order Statistics: Applications*, Handbook of Statistics Vol. 17, 183-213. Elsevier, Amsterdam.
- Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992) *A First Course in Order Statistics*. John Wiley, New York.
- Balakrishnan, N. and Aggarwala, R. (2000) *Progressive Censoring: Theory, Methods, and Applications*. Birkhäuser, Boston.
- Balakrishnan, N. and Cohen, A. C. (1991) *Order Statistics and Inference: Estimation Methods*. Academic Press, San Diego.
- Burkschat, M., Cramer, E., and Kamps, U. (2002) On optimal schemes in progressive censoring. Submitted for publication.
- Cohen, A. C. (1963) Progressively censored samples in life testing. *Technometrics* 5, 327-329.
- Chan, L. K. and Chan, N. N. (1973) On the optimum best linear unbiased estimates of the parameters of the normal distribution based on selected order statistics. *Skand. Aktuarietid.* 120-128.
- Chan, L. K. and Mead, E. R. (1971) Linear estimation of the parameters of the extreme-value distribution based on suitably chosen order statistics. *IEEE Trans. Reliability* R-20, 74-83.
- Fletcher, R. (1987) *Practical Methods of Optimization* (2nd Edition). John Wiley, New York.
- Gill, P. E., Murray, W. and Wright, M. H. (1981) *Practical Optimization*. Academic Press, New York.

- Hassanein, K. M. (1969) Estimation of the parameters of the extreme value distribution by use of two or three order statistics. *Biometrika* 56, 429-436.
- Hassanein, K. M. (1971) Percentile estimators for the parameters of the Weibull distribution. *Biometrika* 58, 673-676.
- Mosteller, F. (1946) On some useful “inefficient” statistics. *Ann. Math. Statist.* 17, 377-408.
- Ogawa, J. (1951) Contributions to the theory of systematic statistics. I *Osaka Math. J.* 3, 175-213.
- Ogawa, J. (1962a) Determinations of optimum spacings in the case of normal distribution. In A. E. Sarhan and B. G. Greenberg (eds.) *Contributions to Order Statistics*, 272-276, John Wiley, New York.
- Ogawa, J. (1962b) Optimum spacing and grouping for the exponential distribution. In A. E. Sarhan and B. G. Greenberg (eds.) *Contributions to Order Statistics*, 371-381. John Wiley, New York.
- Saleh, A. K. Md. E. and Ali, M. M. (1966) Asymptotic optimum quantiles for the estimation of the parameters of the negative exponential distribution. *Ann. Math. Statist.* 37, 1717-1735.
- Sarhan, A. E., Greenberg, B. G., and Ogawa, J. (1963) Simplified estimates for the exponential distribution. *Ann. Math. Statist.* 34, 102-116.
- Sen, P. K. and Singer, J. M. (1993) *Large Sample Methods in Statistics*. Chapman & Hall, New York.

m	τ	.01	.05	.1	.15	.2	.5	1
2	$\bar{\lambda}_1$.3998E-2	.1831E-1	.3462E-1	.4988E-1	.6435E-1	.1398E+0	.2390E+0
	$\bar{\lambda}_2$.6002E-2	.3169E-1	.6538E-1	.1001E+0	.1356E+0	.3602E+0	.6872E+0
	λ_1	.9881E+0	.9422E+0	.8855E+0	.8291E+0	.7728E+0	.4360E+0	
	λ_2	.1869E-2	.7838E-2	.1454E-1	.2094E-1	.2718E-1	.6404E-1	.7378E-1
	ΔVar	90.8%	82.9%	76.4%	70.8%	65.6%	35.7%	
3	$\bar{\lambda}_1$.3960E-2	.1789E-1	.3339E-1	.4751E-1	.6050E-1	.1164E+0	.1902E+0
	$\bar{\lambda}_2$.4455E-2	.2411E-1	.5009E-1	.7679E-1	.1038E+0	.2547E+0	.5627E+0
	$\bar{\lambda}_3$.1586E-2	.7994E-2	.1651E-1	.2570E-1	.3566E-1	.1289E+0	.2142E+0
	λ_1	.9893E+0	.9472E+0	.8947E+0	.8421E+0	.7895E+0	.4683E+0	
	λ_3	.6662E-3	.2818E-2	.5336E-2	.7860E-2	.1046E-1	.3166E-1	.3292E-1
4	$\bar{\lambda}_1$.3923E-2	.1738E-1	.3094E-1	.4241E-1	.5269E-1	.9189E-1	.9918E-1
	$\bar{\lambda}_2$.3668E-2	.1863E-1	.3012E-1	.3959E-1	.5205E-1	.1383E+0	.3017E+0
	$\bar{\lambda}_3$.1653E-2	.9927E-2	.2844E-1	.5039E-1	.7107E-1	.2041E+0	.4581E+0
	$\bar{\lambda}_4$.7547E-3	.4069E-2	.1051E-1	.1761E-1	.2419E-1	.6570E-1	.1199E+0
	λ_1	.9897E+0	.9485E+0	.8964E+0	.8442E+0	.7923E+0	.4811E+0	
5	$\bar{\lambda}_1$.3884E-2	.3820E-2	.7234E-2	.1033E-1	.1328E-1	.2992E-1	.7433E-1
	$\bar{\lambda}_2$.2947E-2	.1461E-1	.2721E-1	.3831E-1	.4867E-1	.1033E+0	.2314E+0
	$\bar{\lambda}_3$.1795E-2	.1845E-1	.3480E-1	.4673E-1	.5905E-1	.1539E+0	.4011E+0
	$\bar{\lambda}_4$.9299E-3	.9268E-2	.2217E-1	.4000E-1	.5838E-1	.1601E+0	.2169E+0
	$\bar{\lambda}_5$.4448E-3	.3852E-2	.8595E-2	.1463E-1	.2062E-1	.5277E-1	.6418E-1
10	λ_1	.9898E+0						
	λ_2		.9486E+0	.8970E+0	.8451E+0	.7933E+0	.4845E+0	
	λ_5	.1934E-3	.1441E-2	.2995E-2	.4892E-2	.6695E-2	.1555E-1	.1216E-1
	ΔVar	89.3%	81.2%	74.9%	69.8%	65.2%	40.9%	
	15	λ_2	.9900E+0					
λ_3			.9497E+0	.8995E+0	.8493E+0	.7991E+0	.4980E+0	
λ_{10}		.4686E-4	.2730E-3	.5036E-3	.7196E-3	.9256E-3	.2047E-2	.2195E-2
ΔVar		89.5%	82.1%	76.5%	72.0%	67.9%	46.4%	
20	λ_4	.9900E+0	.9499E+0	.8998E+0	.8498E+0	.7997E+0	.4994E+0	
	λ_{15}	.2075E-4	.8390E-4	.1513E-3	.2132E-3	.2717E-3	.5872E-3	.7139E-3
	ΔVar	89.7%	82.4%	77.0%	72.6%	68.7%	47.8%	
20	λ_5	.9900E+0	.9500E+0	.8999E+0	.8499E+0			
	λ_6					.7999E+0	.4997E+0	
	λ_{20}	.8748E-5	.3496E-4	.6264E-4	.8784E-4	.1359E-3	.2889E-3	.3133E-3
	ΔVar	89.7%	82.6%	77.2%	72.8%	69.0%	48.5%	

Table 4: Optimal block sizes for an underlying extreme value distribution, and percentage reduction (ΔVar) of the generalized variance when compared to right censoring ($\lambda_1 = \dots = \lambda_{m-1} = 0$)

m	τ	.01	.05	.1	.15	.2	.5	1
2	$\bar{\lambda}_1$.1896E-2	.7658E-2	.1324E-1	.1784E-1	.2177E-1	.3770E-1	.6186E-1
	$\bar{\lambda}_2$.8104E-2	.4234E-1	.8676E-1	.1322E+0	.1782E+0	.4623E+0	.7984E+0
	λ_1	.9831E+0	.9235E+0	.8521E+0	.7820E+0	.7125E+0	.3004E+0	
	λ_2	.6945E-2	.2655E-1	.4789E-1	.6801E-1	.8752E-1	.1996E+0	.1398E+0
	ΔVar	53.5%	42.9%	35.6%	30.1%	25.4%	6.15%	
3	$\bar{\lambda}_1$.1195E-3	.5351E-3	.1002E-2	.1439E-2	.1857E-2	.4158E-2	.1064E-1
	$\bar{\lambda}_2$.2310E-2	.1024E-1	.1896E-1	.2696E-1	.3446E-1	.7368E-1	.1709E+0
	$\bar{\lambda}_3$.7570E-2	.3923E-1	.8003E-1	.1216E+0	.1637E+0	.4222E+0	.6929E+0
	λ_2	.9831E+0	.9235E+0	.8522E+0	.7820E+0	.7124E+0	.2984E+0	
	λ_3	.6919E-2	.2647E-1	.4781E-1	.6798E-1	.8758E-1	.2016E+0	.1255E+0
ΔVar	47.8%	37.8%	31.0%	26.0%	21.9%	5.09%		
4	$\bar{\lambda}_1$.1557E-4	.7252E-4	.1404E-3	.2076E-3	.2752E-3	.7334E-3	.6500E-2
	$\bar{\lambda}_2$.3017E-3	.1403E-2	.2711E-2	.3999E-2	.5290E-2	.1391E-1	.1096E+0
	$\bar{\lambda}_3$.2356E-2	.1084E-1	.2070E-1	.3021E-1	.3957E-1	.9833E-1	.5886E+0
	$\bar{\lambda}_4$.7327E-2	.3769E-1	.7645E-1	.1156E+0	.1549E+0	.3870E+0	.2434E+0
	λ_3	.9831E+0	.9238E+0	.8528E+0	.7830E+0	.7138E+0	.3027E+0	
λ_4	.6865E-2	.2620E-1	.4720E-1	.6701E-1	.8621E-1	.1973E+0	.5179E-1	
ΔVar	44.9%	35.0%	28.5%	23.8%	19.9%	4.55%		
5	$\bar{\lambda}_1$.3033E-5	.7186E-4	.1364E-3	.1973E-3	.2554E-3	.5609E-3	.1600E-2
	$\bar{\lambda}_2$.5880E-4	.1390E-2	.2633E-2	.3801E-2	.4912E-2	.1069E-1	.2963E-1
	$\bar{\lambda}_3$.4602E-3	.1074E-1	.2010E-1	.2872E-1	.3677E-1	.7646E-1	.1906E+0
	$\bar{\lambda}_4$.2284E-2	.2417E-1	.4938E-1	.7477E-1	.1002E+0	.2498E+0	.5359E+0
	$\bar{\lambda}_5$.7194E-2	.1363E-1	.2774E-1	.4251E-1	.5791E-1	.1625E+0	.1988E+0
λ_3		.9386E+0	.8794E+0	.8205E+0	.7617E+0	.4068E+0		
λ_4	.9832E+0							
λ_5	.6826E-2	.1144E-1	.2063E-1	.2951E-1	.3832E-1	.9319E-1	.4348E-1	
ΔVar	43.1%	33.6%	27.7%	23.4%	19.8%	5.72%		
10	λ_6		.9458E+0	.8921E+0	.8385E+0	.7851E+0	.4684E+0	
	λ_7	.9882E+0						
	λ_{10}	.1765E-2	.4181E-2	.7869E-2	.1154E-1	.1486E-1	.3160E-1	.9670E-2
	ΔVar	41.9%	34.1%	29.1%	25.6%	22.7%	10.8%	

Table 5: Optimal block sizes for an underlying Weibull distribution ($\beta = 3$), and percentage reduction (ΔVar) of the generalized variance when compared to right censoring ($\lambda_1 = \dots = \lambda_{m-1} = 0$)

m	τ	.01	.05	.1	.15	.2	.5	1
2	$\bar{\lambda}_1$.3532E-2	.1514E-1	.2713E-1	.3745E-1	.4654E-1	.8561E-1	.1334E+0
	$\bar{\lambda}_2$.6468E-2	.3486E-1	.7287E-1	.1126E+0	0.1535E+0	0.4144E+0	0.7333E+0
	λ_1	.9811E+0	.9228E+0	.8538E+0	.7860E+0	.7187E+0	.3173E+0	
	λ_2	.8877E-2	.2716E-1	.4618E-1	.6401E-1	.8133E-1	.1827E+0	.1334E+0
	ΔVar	71.5%	56.4%	46.5%	39.2%	33.1%	8.44%	
3	$\bar{\lambda}_1$.7133E-3	.3245E-2	.6139E-2	.8875E-2	.1151E-1	.2643E-1	.8327E-1
	$\bar{\lambda}_2$.3201E-2	.1485E-1	.2830E-1	.4104E-1	.5331E-1	.1223E+0	.4167E+0
	$\bar{\lambda}_3$.6086E-2	.3190E-1	.6556E-1	.1001E+0	.1352E+0	.3513E+0	.4167E+0
	λ_2	.9813E+0	.9241E+0	.8563E+0	.7897E+0	.7235E+0	.3262E+0	
	λ_3	.8663E-2	.2594E-1	.4369E-1	.6029E-1	.7645E-1	.1738E+0	.8327E-1
ΔVar	66.4%	50.6%	40.7%	33.8%	28.2%	6.97%		
4	$\bar{\lambda}_1$.6949E-3	.3157E-2	.5883E-2	.8365E-2	.1066E-1	.2136E-1	.4493E-1
	$\bar{\lambda}_2$.3113E-2	.1440E-1	.2695E-1	.3833E-1	.4878E-1	.9616E-1	.2003E+0
	$\bar{\lambda}_3$.2391E-2	.1782E-1	.3816E-1	.5907E-1	.8020E-1	.2059E+0	.5094E+0
	$\bar{\lambda}_4$.3801E-2	.1463E-1	.2901E-1	.4424E-1	.6036E-1	.1766E+0	.2003E+0
	λ_2	.9825E+0	.9337E+0	.8742E+0	.8151E+0	.7560E+0	.3964E+0	
λ_4	.7460E-2	.1633E-1	.2577E-1	.3485E-1	.4396E-1	.1036E+0	.4493E-1	
ΔVar	64.4%	48.9%	39.5%	33.0%	27.9%	8.05%		
5	$\bar{\lambda}_1$.2454E-3	.1135E-2	.2157E-2	.3126E-2	.4061E-2	.9435E-2	.2994E-1
	$\bar{\lambda}_2$.1065E-2	.4964E-2	.9437E-2	.1366E-1	.1771E-1	.4067E-1	.1295E+0
	$\bar{\lambda}_3$.2619E-2	.1243E-1	.2381E-1	.3458E-1	.4492E-1	.1033E+0	.3406E+0
	$\bar{\lambda}_4$.2446E-2	.1752E-1	.3745E-1	.5804E-1	.7902E-1	.2099E+0	.3406E+0
	$\bar{\lambda}_5$.3624E-2	.1395E-1	.2714E-1	.4060E-1	.5428E-1	.1367E+0	.1295E+0
λ_3	.9827E+0	.9342E+0	.8754E+0	.8173E+0	.7594E+0	.4153E+0		
λ_5	.7270E-2	.1584E-1	.2459E-1	.3273E-1	.4058E-1	.8475E-1	.2994E-1	
ΔVar	63.7%	48.4%	39.3%	33.1%	28.2%	9.66%		
10	λ_5	.9839E+0	.9410E+0	.8890E+0	.8372E+0	.7855E+0	.4764E+0	
	λ_{10}	.6129E-2	.9044E-2	.1100E-1	.1280E-1	.1450E-1	.2364E-1	.5926E-2
	ΔVar	62.9%	48.6%	40.5%	35.1%	30.8%	14.3%	
15	λ_7		.9437E+0	.8939E+0	.8437E+0	.7934E+0	.4907E+0	
	λ_8	.9842E+0						
	λ_{15}	.5822E-2	.6329E-2	.6075E-2	.6276E-2	.6635E-2	.9269E-2	.2086E-2
	ΔVar	62.8%	48.8%	41.0%	35.8%	31.8%	15.9%	
20	λ_8		.9449E+0	.8961E+0	.8469E+0	.7969E+0	.4963E+0	
	λ_9							
	λ_{11}	.9843E+0						
	λ_{20}	.5685E-2	.5101E-2	.3936E-2	.3146E-2	.3088E-2	.3739E-2	.9606E-3
ΔVar	62.7%	48.9%	41.3%	36.2%	32.2%	16.7%		

Table 6: Optimal block sizes for an underlying normal distribution, and percentage reduction (ΔVar) of the generalized variance when compared to right censoring ($\lambda_1 = \dots = \lambda_{m-1} = 0$)