On Precision of Optimization in the Case of Incomplete Information*

Petr Volf

Institute of Information Theory and Automation,
Pod vodárenskou věží 4, Prague 8
e-mail: volf@utia.cas.cz

Abstract. In the paper we study the impact of incomplete information to precision of results of stochastic optimization. It is assumed that the stochastic characteristics of optimized system are evaluated from the data. However, the observations can be incomplete, namely, we consider the random censoring frequently encountered in time-to-event or lifetime studies. The analysis of uncertainty of solution is based both on theoretical properties of estimated stochastic characteristics and on simulated examples.

Keywords: Optimization, censored data, Fisher information, product-limit estimate.

JEL classification: C41, J64
AMS classification: 62N02, 62P25

1. Introduction

In problems of optimization under uncertainty we often deal with a probabilistic model of optimized system. Then the optimization task consists in the search of a solution to

$$\inf_{\boldsymbol{v}} \phi_F(v) = \inf_{\boldsymbol{v}} E_F \varphi(Y, \boldsymbol{v}), \tag{1}$$

where φ is a cost function, $\boldsymbol{v}$ are input variables from certain feasibility set $\boldsymbol{V}$. Further, E_F stands for the expectation under distribution function F, and, finally, Y is a random variable (or vector) with distribution function F. If F is known, we deal with a "deterministic" optimization. However, our information on probability distributions governing the system could be non-complete. Either, known distribution type depends on unknown parameters. Then, as a rule, the estimates of parameters are plugged into objective function. Or, we have to employ nonparametric estimates, as is the empirical distribution function. Hence, our information on F is random and we have to analyze both possible bias and variability of obtained solution (compared to an ideal solution when F is known). Alternatively, we then can be interested in a kind of multi-objective optimization, minimizing simultaneously also variability (measured by variance, or certain inter-quantile range). Nevertheless, standard approach considers a solution of (1) and uses estimated characteristics instead of 'true'

*This research was supported by the Grant Agency of the Czech Republic under Grants 402/10/0956.

ones. An investigation of usage of empirical (estimated) characteristics in stochastic optimization problems started already in 70-ties. A number of papers has dealt with these problems, let us mention here just two: Dupačová and Wets (1984), and from more recent time Kaňková (2009) with a brief overview and a number of other references.

The situation is even more complicated if the data available for estimation are not complete. We shall consider one special type of incompleteness, the random censoring from the right side. It is quite frequent in the analysis of demographic, survival or insurance data. The lack of information leads to higher variability (and, sometimes, to a bias) of estimates and, consequently, to higher uncertainty of optimal solutions.

The approaches to statistical data analysis in cases when the data are censored or even truncated are provided by a number of authors. Let us mention here works of C. Huber (e. g. Huber, 2000), with classification of designs of censored and truncated data and with many references to papers dealing with specific methods of such data processing. The most of results were derived in the framework of statistical survival analysis and collected also in several monographs (cf. Kalbfleisch and Prentice, 2002).

The main objective of the present paper is to study the increase of uncertainty of results of optimization problem when the censoring is causing growing variability of estimates. We shall deal with both parametric and non-parametric cases. To this end, certain theoretical properties of estimates under random right censoring will be recalled. In the next two sections, we shall consider the product-limit estimate as a generalization of the empirical distribution function, and then the maximum likelihood estimates of parameters when random right censoring is present. We shall compare properties of estimates with and without censoring, in nonparametric case (in Section 2) as well as in the case of estimated parameters (Section 3). Section 4 contains the main contribution of the paper. We prove the convergence of optimal values corresponding to estimated distribution function or its estimated parameters to the optimum (1). Finally, in Section 5 a simple example will deal with optimal maintenance schedule, properties of obtained 'sub-optimal' solution will be illustrated with the aid of simulations. In the last section we discuss briefly another criterion based on quantiles of random objective function instead on the mean value.

2. Non-parametric case and product-limit estimate

In the present section we shall recall some useful results concerning the analysis of randomly right-censored data. They are collected in survival analysis literature, for instance in Kalbfleisch and Prentice (2002). Let us consider a continuous-type random variable Y characterizing for instance a random time to certain event. Let another continuous random variable Z be a censoring variable, both be positive, continuous and mutually independent. Further, let $f(y)$, $g(z)$, $F(y)$, $G(z)$, $\overline{F}(y) = 1 - F(y)$, $\overline{G}(z) = 1 - G(z)$ denote the density, distribution and survival functions of both variables. It is assumed that we observe just $X = \min(Y, Z)$ and $\delta = 1[Y \leq Z]$, i. e. δ indicates whether Y is observed or censored from right side. The data are then given as random sample (X_i, δ_i, $i = 1, \ldots, N$). Notice that the case without censoring is obtained when $G(t) \equiv 0$ on region where $F(t) < 1$. In the sequel we shall assume that $\sup\{t : G(t) < 1\} \geq \sup\{t : F(t) < 1\}$, so that Z does not cut off (with probability 1) the right tail of Y. Let us remark here that in some cases we can deal, for instance, with the logarithm of time. Then the domain of data can be the whole R_1.

A generalization of empirical distribution function is the well known Kaplan–Meier "Product Limit Estimate" (PLE) of survival function. Let us first sort (re-index) the data in increasing order, $X_1 \leq X_2 \leq \cdots \leq X_N$, then the PLE of $\overline{F}(t)$ has the form

$$\overline{F}_N(t) = \prod_{i=1}^{N} \left(\frac{N-i}{N-i+1} \right)^{\delta_i \cdot 1[X_i \leq t]}. \tag{2}$$

Again, notice that when all $\delta_i = 1$, we obtain the empirical survival function. The following proposition is due to Breslow and Crowley (1974):

Proposition 1. *Let $T > 0$ be such that still $\overline{F}(T) \cdot \overline{G}(T) > 0$. Then the random process*

$$V_N(t) = \sqrt{N} \left(\frac{\overline{F}_N(t)}{\overline{F}(t)} - 1 \right) = \sqrt{N}\, \frac{F(t) - F_N(t)}{\overline{F}(t)} \tag{3}$$

converges, on $[0, T]$, when $N \to \infty$, to Gaussian martingale with zero mean and variance function

$$C(t) = \int_0^t \frac{\mathrm{d}F(s)}{\overline{F}(s)^2\, \overline{G}(s)}. \tag{4}$$

Here, $F_N(t) = 1 - \overline{F}_N(t)$. In other words, $V_N(t)$ converges in distribution on $[0, T]$ to the process $W(C(t))$, where $W(\cdot)$ denotes the Wiener process. The asymptotic variance function can be estimated by its empirical version:

$$C_N(t) = \sum_{i=1}^{N} \frac{N\delta_i}{(N-i+1)^2} \cdot 1[X_i \leq t],$$

which is consistent in probability, uniformly w.r. to $t \in [0, T]$ (see again Breslow and Crowley, 1974).

Further, denote $D_N(t) = V_N(t)/(1 + C(t))$. For the case without censoring we obtain that $C(t) = F(t)/\overline{F}(t)$ and $D_N(t) = \sqrt{N}(F(t) - F_N(t))$ leading to standard Kolmogorov–Smirnov statistics. From (4) it is also seen that the variance in the case with censoring (when $\overline{G}(t) \leq 1$) is larger than without it (i.e. when $\overline{G}(t) = 1$ on whole $[0, T]$). Further, it has been proved (see, for instance, Robbins and Siegmund, 1970) that for c, $d > 0$ and sufficiently large T,

$$P\left(\sup_{0<t<T} |W(t)| < c + d \cdot t \right) \doteq 1 - 2\exp(-2cd).$$

Hence, if we take $c = d$ and "time" $C(t)$ instead t, we obtain that approximately

$$P\left(\sup_{t} |D_N(t)| > c \right) \doteq 2\exp(-2c^2).$$

In order to construct $1-\alpha$ band for $D_N(t)$ on $(0,T)$, we set $\alpha = 2\exp(-2c^2)$ and obtain critical value $c_\alpha = \sqrt{-\ln\frac{\alpha}{2}/2}$. In the case without censoring c_α is the distribution-free critical value for the Kolmogorov–Smirnov test, namely

$$P\left(\sup_t |F_N(t) - F(t)| \geq c_\alpha/\sqrt{N}\right) \doteq \alpha.$$

In the case with censoring, we have

$$D_N(t) = \sqrt{N}\left(F(t) - F_N(t)\right)/\overline{F}(t)/(1 + C(t)),$$

hence, corresponding $1-\alpha$ confidence band for $F(t)$ depends on both F and G and its width is increasing for larger t. Namely, $1-\alpha$ borders for $|F_N(t) - F(t)|$ on $[0,T]$ are given as $c_\alpha/\sqrt{N}\cdot\overline{F}(t)\cdot(1+C(t))$.

Example: Let us here, as an example, consider so called Koziol–Green model assuming that $\overline{G}(t) = \overline{F}(t)^a$, for some $a > 0$. Then

$$C(t) = \int_0^t \frac{\mathrm{d}F(s)}{\overline{F}(t)^{2+a}} = \frac{1}{\overline{F}(t)^{1+a}} - 1$$

and $\overline{F}(t)\cdot(1+C(t)) = 1/\overline{F}(t)^a$. It tends to infinity with increasing t because $\overline{F}(t)\to 0$. A more concrete example is presented in Part 5.

3. Parametric estimates under censoring

In the present part we shall study the influence of censoring to precision of estimated parameters. It means that we assume that the type of distribution $F(y,\theta)$ of random variable Y is known, unknown is the parameter $\theta\in\Theta$. The 'true' value of parameter, θ_0, is estimated as a rule by the method of maximum likelihood (MLE). Regarding the censoring distribution, except independence of Y and Z we must now assume that the distribution of Z does not depend on θ, censoring is 'non-informative'.

Desirable properties of the MLE estimates are connected with so called regularity conditions concerning distribution $F(y,\theta)$. Their formulation can be found elsewhere in statistical textbooks, for instance in Anděl (2005), Ch. 7.3. If they are fulfilled then there exists a consistent sequence of estimates, i.e. such that $\hat{\theta}_N \to \theta_0$ in probability when $N\to\infty$. Moreover, estimates are asymptotically normal, which means that the distribution of $\sqrt{N}(\hat{\theta}_N - \theta_0)$ tends to normal distribution with zero mean and finite variance.

The comparison of precision of estimates can be based on the Fisher information. It is defined as

$$I(\theta) = E\left(\frac{\mathrm{d}\ln L(\theta, X)}{\mathrm{d}\theta}\right)^2,$$

where $L(\theta, X)$ is the likelihood of θ based on random variable X. Regularity conditions ensure that $I(\theta) > 0$ exists for all $\theta\in\Theta$. Naturally, if θ is multi-dimensional, we consider a vector of partial derivatives and $I(\theta)$ is a matrix. What is important from our point of view, $I^{-1}(\theta_0)$ is the asymptotic variance of $\sqrt{N}(\hat{\theta}_N - \theta_0)$ mentioned above, where $\hat{\theta}_N$ is the consistent sequence of MLE-s of θ_0 from random sample of size N.

The form of likelihood and asymptotic properties of the MLE in the case of random right censoring is described for instance in Kalbfleisch and Prentice (2002), section 3.4. The log of likelihood (its part depending on θ), is then $\ln L(\theta, X) = \delta \cdot \ln f(X,\theta) + (1-\delta)\cdot \ln \bar{F}(X,\theta)$. If there exist the 1-st derivatives $f' = \mathrm{d}f/\mathrm{d}\theta$ and $\bar{F}' = \mathrm{d}\bar{F}/\mathrm{d}\theta$, then

$$E\left(\frac{\mathrm{d}\ln L(\theta,X)}{\mathrm{d}\theta}\right)^2 = \int_0^\infty \bar{G}(x)\left(\frac{f'(x)}{f(x)}\right)^2 f(x)\mathrm{d}x + \int_0^\infty \bar{F}(x)\left(\frac{\bar{F}'(x)}{\bar{F}(x)}\right)^2 g(x)\mathrm{d}x.$$

When the second integral is transformed with the aid of *per-partes*, we obtain that

$$I(\theta) = \int_0^\infty \bar{G}(x)\frac{(f'(x)\bar{F}(x) - f(x)\bar{F}'(x))^2}{f(x)\bar{F}((x)^2}\mathrm{d}x.$$

It is non-negative and is larger when $\bar{G}(x) \equiv 1$, i.e. when there is no censoring. A more concrete comparison is presented within the example in Section 5.

Naturally, in practice the MLE is based on the data $\{X_i, \delta_i\}, i = 1, ..., N$ and maximizes the log-likelihood function

$$\ln L_N(\theta) = \sum_{i=1}^{N} \delta_i \cdot \ln f(X_i,\theta)) + (1-\delta_i)\cdot \ln \bar{F}(X_i,\theta))$$

over $\theta \in \Theta$. Consistent estimate of the Fisher information is then obtained (compare again Kalbfleisch and Prentice, 2002) as

$$I_N(\theta) = \frac{1}{N}\left(\frac{\mathrm{d}(\ln L_N(\theta))^2}{\mathrm{d}^2\theta}\right).$$

4. Consistency of optimum

In this section we are interested in the question whether (and under which assumptions) the consistency of estimate of F ensures already the consistency of optimal solution. Namely, whether optimal values (both ϕ and $\boldsymbol{v}$) of problem (1) obtained from a case with non-complete information on F converge to optimal ϕ^* and $\boldsymbol{v}^*$, respectively, obtained when F is known.

There are several ways how to address this problem. One of them can use the approach proposed in Kaňková (2009) supported by the results from Kaňková and Houda (2006). Let us recall here the theorem on which the other assertions are based:

Proposition 2 (Kaňková, 2009). *Let F, G be two distribution functions on R, let $\boldsymbol{v} \in \boldsymbol{V}$, $\boldsymbol{V}$ be a compact set. If for every $\boldsymbol{v} \in \boldsymbol{V}$:*

1. $\varphi(y, \boldsymbol{v})$ is a Lipschitz function of y on R, with the Lipschitz constant L not depending on $\boldsymbol{v}$,

2. finite $E_F\varphi(y,\boldsymbol{v})$, $E_G\varphi(y,\boldsymbol{v})$ exist,

3. $\varphi(y, \boldsymbol{v})$ *is a uniformly continuous function on* $R \times \boldsymbol{V}$*,*
then

$$|\inf_{\boldsymbol{v}} E_F\varphi(Y, \boldsymbol{v}) - \inf_{\boldsymbol{v}} E_G\varphi(Y, \boldsymbol{v})| \leq L \int_R |F(y) - G(y)|\mathrm{d}y.$$

It is seen that we need two sets of assumptions. The first concerning the properties of criterion function $\varphi(y, \boldsymbol{v})$, as for instance those required in Proposition 2 (though they can seem to be rather strong). Then, the second set of assumptions should concern the closeness of distributions, here expressed via the Wasserstein metric. Hence, we should derive such a closeness from the convergence of estimated distribution functions, which could be rather complicated problem. For the case of noncensored data it is solved, under specific conditions on function $\varphi(y, v)$, in Kaňková (2009).

4.1. Approach based on convergence of means

As the criterion (1) is expressed via the expectation, we shall focus our attention to the convergence of empirical means computed from censored data. Several such results are available in the framework of statistical survival analysis, e.g. Gill (1983), Akritas (2000) and others. We shall refer here to the paper of Volf (1987), where the following strong consistency result of Rejto (1983) is utilized.

Proposition 3 (Rejto, 1983). *Let us consider the random censoring model with distribution functions* F, G *continuous. Let there exist* $a, b \in (0, 1]$ *and a real* τ *such that* $a\overline{F}(t)^b \leq \overline{G}(t)$ *on* $[\tau, \infty)$*. Then almost surely*

$$\sup_{-\infty<t<\infty} |F_N(t) - F(t)| = \mathcal{O}\left(\left[\frac{\log N}{N}\right]^{\frac{1}{2+b}}\right). \tag{5}$$

Again, $F_N(t) = 1 - \overline{F}_N(t)$ and $\overline{F}_N(t)$ denotes the PLE of survival function $\overline{F}(t)$ of random variable Y.

The following proposition establishes the existence of strongly consistent estimate of the mean, generalizing slightly the result of Volf (1987).

Proposition 4. *Let* $h(y)$ *be an integrable function with finite mean* $\bar{h} = \int_{\infty}^{\infty} h(y)\mathrm{d}F(y)$*. Further, let the following hold:*

1. $h(y)$ *has bounded first derivative,* $|h'(y)| \leq C < \infty$*.*

2. Assumptions of Proposition 3 are fulfilled with some a, b, τ*.*

3. Let $A_N \to \infty$ *be a positive, increasing sequence such that for* $N \to \infty$

$$A_N \cdot \left(\frac{\log N}{N}\right)^{\frac{1}{2+b}} \to 0.$$

Then $\bar{h}_N = \int_{-A_N}^{A_N} h(y)\mathrm{d}F_N(y)$ *is the strongly consistent estimate of* $\bar{h}$*.*

Proof. Denote first $h_N = \int_{-A_N}^{A_N} h(y)\mathrm{d}F(y)$, hence $h_N \to \bar{h}$. Let, for given N, $X_{N,i}$ be ordered data, $X_{N,1} \le X_{N,2} \le X_{N,N}$. Further, denote K, L such indices that

$$K = \min\{i : X_{N,i} \ge -A_N\} - 1, \;\; L = \max\{i : X_{N,i} \le A_N\} + 1.$$

Denote, omitting index N, $T_K = -A_N$, $T_L = A_N$, $T_i = X_{N,i}$ for $i = K+1, ..., L-1$. Then we can rewrite the estimate

$$\bar{h}_N = \sum_{i=K+1}^{L} h(T_i)(F_N(T_i) - F_N(T_{i-1})) =$$

$$= h(A_N)F_N(A_N) - h(-A_N)F_N(-A_N) - \sum_{i=K+1}^{L} F_N(T_{i-1})(h(T_i) - h(T_{i-1})).$$

Notice that the last term equals sum of integrals

$$\int_{T_{i-1}}^{T_i} F_N(y) \cdot h'(y)\mathrm{d}y,$$

because $F_N(y) = F_N(T_{i-1})$ on $[T_{i-1}, T_i)$. Further, by *per-partes* integration we obtain

$$\int_{-A_N}^{A_N} h(y)\mathrm{d}F(y) = F(A_N)h(A_N) - F(-A_N)h(-A_N) - \int_{-A_N}^{A_N} h'(y)F(y)\mathrm{d}y.$$

Then $\bar{h}_N - h_N = h(A_N)(F_N(A_N) - F(A_N)) - h(-A_N)(F_N(-A_N) - F(-A_N)) -$

$$- \int_{-A_N}^{A_N} h'(y)(F_N(y) - F(y))\mathrm{d}y.$$

From bounded derivative of $h(y)$ it follows that both $|h(-A_N)|$, $|h(A_N)| \le C\cdot(A_N + D)$, where D is some finite constant, for instance $|h(0)|$.

Therefore we may conclude that

$$|\bar{h}_N - h_N| \le \{2C(A_N + D) + 2CA_N\} \cdot \sup_y |F_N(y) - F(y)|,$$

which tends to zero a.s. due to assumptions 2 and 3.

$\square$

Remark. The proposition says that when considering the mean 'theoretically', we should have in mind such a truncated form. Any sequence A_N fulfilling assumption 3 may be considered. In practice, as we always deal with finite N, this means no restriction.

It is also seen that the condition of bounded derivative of $h(y)$ can be relaxed, that it suffices to assume that

$$A_N \cdot \sup_{-A_n \le y \le A_N} |h'(y)| \cdot \sup_y |F_N(y) - F(y)| \;\to\; 0.$$

4.2. Main result

In the follow-up we shall denote $\phi_F(v) = E_F\varphi(Y, v)$, assuming that it exists and is finite, further $\phi_N(v) = E_{F_N}\varphi(Y, v)$, the estimate of the former based on estimated distribution function (the product limit estimate) $F_N(y)$. As we want to use results of Proposition 4, we have to consider truncation by an appropriate sequence A_N. As it depends just on F and G, it can be the same for all 'empirical means' considered here. The goal is to show that the convergence concerns also the infima of ϕ_N, ϕ_F and, eventually, the convergence of optimal v. First, let us formulate several assumptions:

A1. Variable $v \in \boldsymbol{V}$, where $\boldsymbol{V}$ is a compact set in R_1.

A2. Functions $\varphi(y, v)$ are continuous in v on $\boldsymbol{V}$, uniformly w.r. to $y \in R_1$.

A3. The assumptions of Proposition 2 are fulfilled for some a, b, τ.

A4. Functions $\varphi(y, v)$ have bounded first derivative w.r. to y, $|\frac{\partial\varphi(y,v)}{\partial y}| \leq C < \infty$, for all $v \in \boldsymbol{V}$.

The assumptions imply some rather sharp properties. Thus, A4 supposes that there exist linear majorizing functions to $\varphi(y, v)$. Further, from A1 it follows that we deal with minimum instead of infimum, and there always exists at least one solution in $\boldsymbol{V}$. Therefore, let us denote

$v_F^* = \arg\min_v \phi_F(v)$, $\phi_F^* = \phi_F(v_F^*)$, $v_N^* = \arg\min_v \phi_N(v)$, $\phi_N^* = \phi_N(v_N^*)$.

Lemma 1. *From A2 it follows that functions $\phi_H(v)$ are continuous in $v \in \boldsymbol{V}$, uniformly for all distribution functions $H(y)$.*

Proof. Consider $v_0 \in \boldsymbol{V}$, $\varepsilon > 0$ and a distribution function $H(y)$. Than, due to A2, we can select such $\delta > 0$ that for each $v :\ |v - v_0| \leq \delta$ is $|\varphi(y, v) - \varphi(y, v_0)| \leq \varepsilon$, for each $y \in R_1$. Then

$$|\phi_H(v) - \phi_H(v_0)| \leq \int_0^\infty |\varphi(y, v) - \varphi(y, v_0)|\, \mathrm{d}H(y) \leq \varepsilon.$$

□

Theorem 1. *Let assumptions A1 – A4 hold. Then, for $N \to \infty$,*

1. *$\phi_F^* = lim\, \phi_N^*$ almost surely,*

2. *There exists a sub-sequence $v_{N,k}^* \subset \{v_N^*\}$, $k = 1, 2, ...$ such that it converges almost surely, when $k \to \infty$, to some $v_0^* \in \boldsymbol{V}$ such that $v_0^* \in \{arg\, min\, \phi_F(v)\}$.*

Proof. As each $\phi_N^* \leq \phi_N(v^*)$ and $|\phi_N(v^*) - \phi_F^*| \to 0$ a.s. (due to Proposition 4 and A4), then both $\liminf \phi_N^* \leq \phi_F^*$ and $\limsup \phi_N^* \leq \phi_F^*$, a.s. Denote $\underline{\phi} = \liminf \phi_N^*$. We want to prove that $\underline{\phi} = \phi_F^*$ a.s.

i) First, there exists (a.s.) a sub-sequence of indices $\{N1\} \subset \{1, 2, ...\}$ such that $\underline{\phi} = \lim_{N1\to\infty} \phi_{N1}^*$, with corresponding sequence of solutions v_{N1}^*. Then, due compactness of $\boldsymbol{V}$, there is another sequence $\{N2\} \subset \{N1\}$ such that there a.s. exists $\lim_{N2\to\infty} v_{N2}^* = \overline{v} \in \boldsymbol{V}$.

ii) From A2 it follows that also $\lim_{N2\to\infty} \phi_{N2}(\overline{v}) = \underline{\phi}$ a.s.

iii) It holds that $\underline{\phi} \leq \phi_F^* \leq \phi_F(\overline{v})$, a.s. On the other hand, from Proposition 3 it follows that (also a.s.) $\lim_{N\to\infty} \phi_N(\overline{v}) = \phi_F(\overline{v})$. When we compare it with ii), we conclude that a.s. $\underline{\phi} = \phi_F^* = \phi_F(\overline{v})$.

   Thus, the first point of the theorem is proved. Notice that as ϕ_F^* is not random, the others are deterministic (with probability 1), too.

iv) Finally, it is seen that $\overline{v}$ is desired limit solution, i.e. that $\overline{v} \in \{\arg\min \phi_F(v)\}$ a.s. If the set $\{\arg\min \phi_F(v)\}$ is just one point, that also $\overline{v}$ coincides with it.

□

We have proved just the convergence of optimal solutions, not establishing the rate of such a convergence. It would require a more sophisticated analysis of the right tail of both distributions F and G, similarly like it is carried out in Kaňková (2009). Notice that for the present result no explicit assumption concerning the distributions tails was needed. Implicitly, certain conditions concerning tails are hidden in the assumption of existence of the means and in assumption borrowed from Rejto (1983). It indicates that the right tail of censoring distribution G should be heavier then the same tail of the distribution F.

4.3. Consistency in parametric case

In the present part we shall refer to discussion and assumptions concerning the distributions considered in Part 3. We assume that the distribution of Y fulfills the regularity conditions and that there exists a sequence of strongly consistent MLE $\hat{\theta}_N$. We are interested in the convergence of corresponding criterion functions. Let us first introduce some notation:

$\phi(v, \theta) = \int_{-\infty}^{\infty} \varphi(y, v) f(y, \theta) \mathrm{d}y$, $\phi_N(v) = \phi(v, \theta_N)$, $\phi_F(v) = \phi(v, \theta_0)$, further $v_F^* = \arg\min_v \phi_F(v)$, $\phi_F^* = \phi_F(v_F^*)$, and $v_N^* = \arg\min_v \phi(v, \hat{\theta}_N)$, $\phi_N^* = \phi(v_N^*, \hat{\theta}_N)$.

If we use Taylor expansion at the 'true' value θ_0, we obtain that

$$\phi(v, \hat{\theta}_N) - \phi(v, \theta_0) = \int_{-\infty}^{\infty} \varphi(y, v) f(y, \hat{\theta}_N) \mathrm{d}y - \int_{-\infty}^{\infty} \varphi(y, v) f(y, \theta_0) \mathrm{d}y =$$

$$= \int_{-\infty}^{\infty} \varphi(y, v) f'(y, \overline{\theta}_N) \mathrm{d}y \cdot (\hat{\theta}_N - \theta_0), \tag{6}$$

where, for sufficiently large N, $\overline{\theta}_N$ is arbitrarily close to θ_0, $f'(y, \theta) = \frac{\partial f(y,\theta)}{\partial \theta}$. Let us formulate the following assumption inspired by (6):

A5. There exists a compact neighborhood $\boldsymbol{O}$ of θ_0 and a positive number $K < \infty$ such that $|\phi'(v, \theta)| = |\int_{-\infty}^{\infty} \varphi(y, v) f'(y, \theta) dy| \leq K$, for each $v \in \boldsymbol{V}$ and $\theta \in \boldsymbol{O}$.

Theorem 2. *Let $\hat{\theta}_N$ be a strongly consistent sequence of estimates of θ_0, further let assumptions A1, A2, A5 hold. Then, for $N \to \infty$,*

1. *$\phi_F^* = lim\, \phi_N^*$ almost surely,*
2. *There exists a sub-sequence $v_{N,k}^* \subset \{v_N^*\}$, $k = 1, 2, ...$ such that it converges almost surely when $k \to \infty$, $lim\, v_{N,k}^* = v_0^*$ and $v_0^* \in \{arg\, min\, \phi_F(v)\}$.*

Proof. The proof can follow, step by step, the proof of Theorem 1, except point iii). The aim of iii) was to prove that a.s. $\lim_{N\to\infty} \phi_N(\overline{v}) = \phi_F(\overline{v})$. Here the same result follows directly from (6) and assumption A5.

□

5. Example

Let us consider the following rather simple example of optimization problem: A component of a machine has its time to failure Y given (modeled) by a continuous-type probability distribution with distribution function, density, survival function F, f, $\bar{F} = 1 - F$, respectively. The cost of repair after failure is C_1, the cost of preventive repair is $C_2 < C_1$. For the simplicity we assume that only complete repairs, 'renewals', are provided, i.e. after each repair the component is new (exchanged) or as new. Let τ be the time from renewal to preventive repair, we wish to select an optimal value of τ.

Let us, as a criterion function, consider the mean time of component availability to the unit of cost, namely

$$\varphi(y, \tau) = \frac{y}{C_1} \quad \text{if} \quad y \leq \tau, \quad \varphi(y, \tau) = \frac{\tau}{C_2} \quad \text{if} \quad y > \tau.$$

Our task is to find optimal τ from a reasonable closed interval $\boldsymbol{T}$, i.e. to maximize

$$\phi_F(\tau) = E_F \varphi(Y, \tau) = \int_0^\tau \frac{y}{C_1} \mathrm{d}F(y) + \frac{\tau}{C_2} \overline{F}(\tau). \tag{7}$$

In such a simple case the optimal solution can be found directly, by solving equation $\mathrm{d}\phi_F(\tau)/\mathrm{d}\tau = 0$. In our case

$$\frac{\mathrm{d}\phi_F(\tau)}{\mathrm{d}\tau} = \frac{\tau}{C_1} f(\tau) + \frac{1}{C_2}\Big(\overline{F}(\tau) - \tau f(\tau)\Big).$$

In the sequel the lifetime distribution will be specified and we shall compare the deterministic solution provided F is known, and the variability of 'sub-solutions' in cases when lifetime distribution is estimated, in parametric or non-parametric setting, from censored and non-censored data. Namely, let the distribution of Y be Weibull, with parameters $a = 100, b = 2$, i.e. its survival function is $\overline{F}(t) = \exp\Big(-(\frac{t}{a})^b\Big)$, numerical characteristics are $\mathrm{E}Y \sim 89$, $\mathrm{sd}(Y) \sim 46$. Costs of repairs

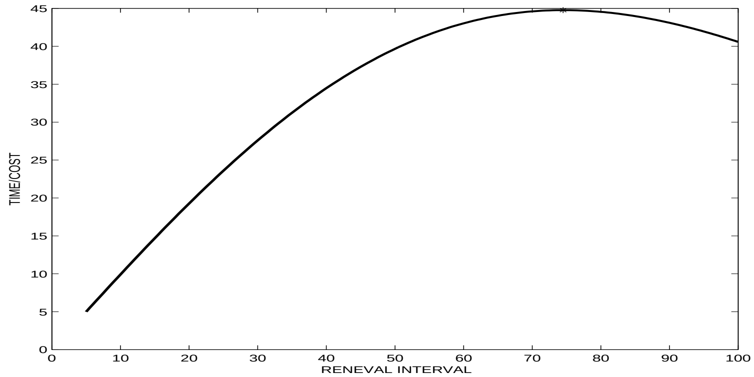

Figure 1: $\phi_F(\tau)$ vers. τ, with optimal point

were fixed as $C_1 = 10$, $C_2 = 1$. When the distribution function F is known, there exists an unique optimal solution with

$$\tau^* = a\Big(\frac{C_1}{(C_1 - C_2)b}\Big)^{1/b} = 74.5356 \tag{8}$$

and maximal working time per cost unit $\phi_F(\tau^*) = 44.7644$. Figure 1 displays the graph of $\phi_F(\tau)$.

In the next parts we provide a numerical study where it is assumed that the distribution of variable Y is estimated from data. In all four cases (parametric or nonparametric case, without or with censoring) 100 samples of 100 observations Y_i are generated from the Weibull distribution mentioned above. In cases with censoring, they are censored by the censoring variables Z_i having uniform distribution on $[0, 250]$, hence with survival function $\overline{G}(z) = (250 - z)/250$ (value 250 corresponds roughly to 0.998 quantile of distribution of Y). The rate of censoring is then about $36\,\% \sim EY/250$.

	non-cens.: I	I-std	sample mean	sample std	censored: I	I-std	sample mean	sample std
a	$3.80 \cdot 10^{-4}$	5.13	100.050	5.904	$2.64 \cdot 10^{-4}$	6.15	99.683	6.211
b	0.409	0.154	2.003	0.155	0.287	0.187	2.004	0.187
τ_m			74.734	4.185			74.580	5.009
$\phi(\tau_m)$			44.632	0.195			44.576	0.275

Table 1: Comparison of theoretical and sample-based characteristics of estimates and optimal solutions

5.1. Parametric case

In the first part of the study the Weibull-type distribution was taken for granted, its parameters were estimated from generated samples of data, by the maximum likelihood method. Hence, 100 couples of estimates $a_m, b_m, m = 1, 2, ..., 100$ were obtained. They are displayed in Figure 2, left plot shows estimates from non-censored

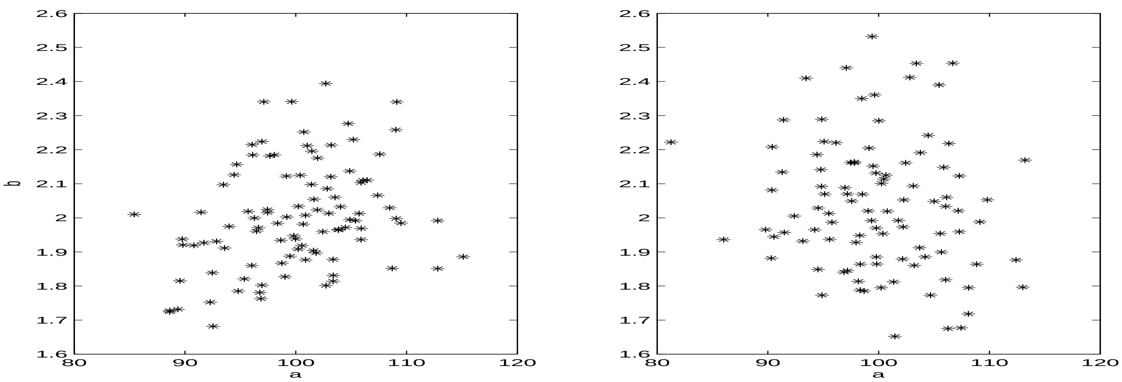

Figure 2: Maximum likelihood estimates of parameters (a_m, b_m) from 100 samples of non-censored (left) and censored data (right)

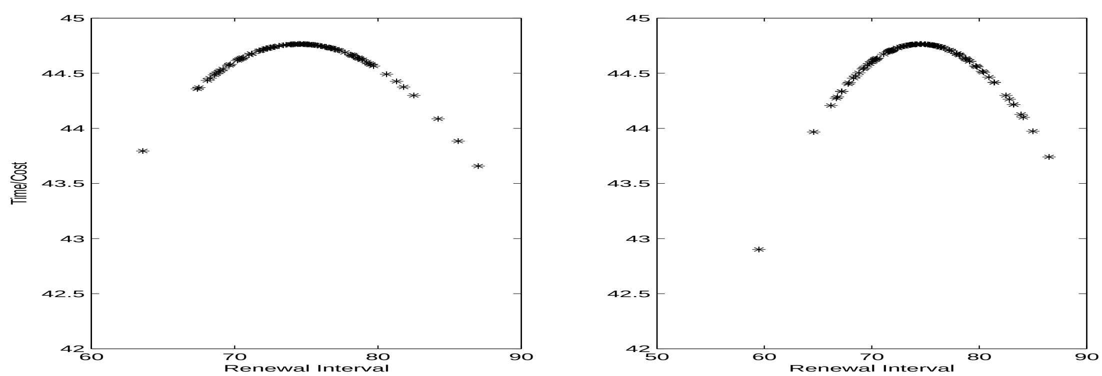

Figure 3: Optimal solutions τ_m and corresponding $\phi_F(\tau_m)$ based on estimates displayed in Figure 2, for non-censored (left) and censored (right) cases

cases, the right plot corresponds to censored cases. From those values their means and sample standard deviations were computed. Simultaneously, we computed theoretical Fisher information I for both parameters and approximate standard deviations of estimates I-std= $\sqrt{1/I/N}$, for extent of sampled data $N = 100$. All these characteristics are collected in Table 1. Further, to each estimated couple of parameters, an optimal solution τ_m was computed from (8). To do it, we computed the corresponding value $\phi_F(\tau_m)$, i.e. the cost at τ_m was computed w.r. to the 'true' F. They are shown in Figure 3, again for non-censored (left) and censored cases (right). Table 1 contains also sample means and standard deviations of those 'sub-optimal' τ_m and $\phi_F(\tau_m)$. It is seen that the variability of results increases with censoring, though not strongly.

5.2. Nonparametric estimate of distribution function

Let us now imagine that we do not know the type of distribution of Y and therefore we estimate it with the aid of the product-limit estimator (i.e. as the empirical distribution function when censoring is absent). Figure 4 displays cloud of 100 estimates obtained from 100 generated samples, the cases without censoring are plotted in the left subplot, the right subplot shows estimates obtained from censored data.

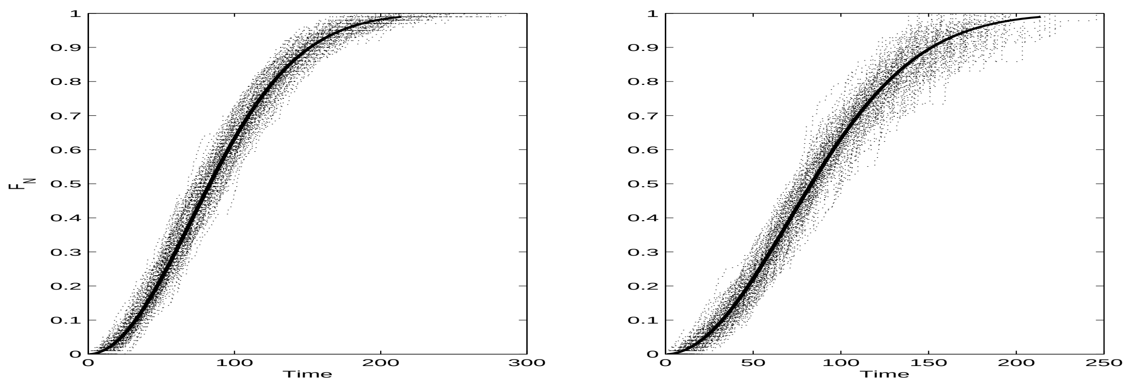

Figure 4: Set of 100 estimates of distribution function, $F_N(t)$, from non-censored (left) and censored data (right). 'True' distribution function $F(t)$ is plotted by solid curve

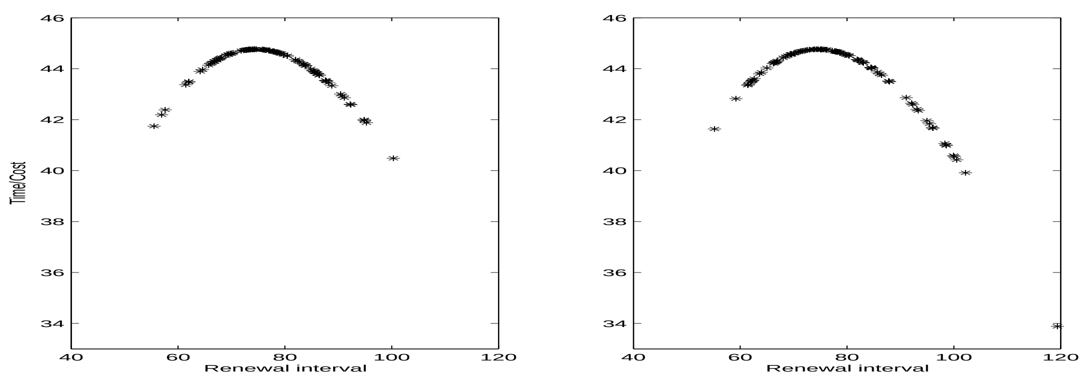

Figure 5: Optimal solutions τ_m and corresponding $\phi_F(\tau_m)$ based on nonparametric estimates displayed in Figure 4, for non-censored (left) and censored (right) cases

It is well seen how the variability in the right subplot increases for large times. Theoretically, if we take $\alpha = 0.05$ and $N = 100$, the half-width of 95% confidence band for 'true' distribution function, in the non-censored case, is given approximately (see Part 2) as $c_\alpha/\sqrt{N} = 0.136$. As regards the censored data, function $C(t)$ (here defined on $[0, 250)$) has no analytical form, nevertheless, we can compute it numerically. We have seen in Part 2 that, at a given t, the half-width of $1-\alpha$ band is given as $d_{N,\alpha}(t) = c_\alpha/\sqrt{N} \cdot \overline{F}(t) \cdot (1 + C(t))$. We computed it at three points corresponding roughly to three quartiles of utilized Weibull distribution. Namely, at points $t = 55$, 85, 120 we obtained $d_{N,\alpha}(t) = 0.142$, 0.158, 0.195, respectively.

non-cens.:	mean	std	cens.:	mean	std
τ_m	76.751	9.154		78.288	11.585
$\phi_F(\tau_m)$	44.123	0.839		43.763	1.553

Table 2: Sample means and standard deviations of τ_m and $\phi_F(\tau_m)$ corresponding to samples plotted in Figure 5

Figure 5 displays optimal solutions τ_m, each obtained as the solution of (8) with m-th estimate of F, $m = 1, 2, ..., 100$. Again, we then computed corresponding $\phi_F(\tau_m)$, i.e. w.r. to 'true' F. The left subplot shows the case without censoring, the right subplot then results from censored samples. Notice (expected) larger variability (i.e. uncertainty) in censored data cases. Table 2 contains sample characteristics of obtained 'sub-optimal' solutions.

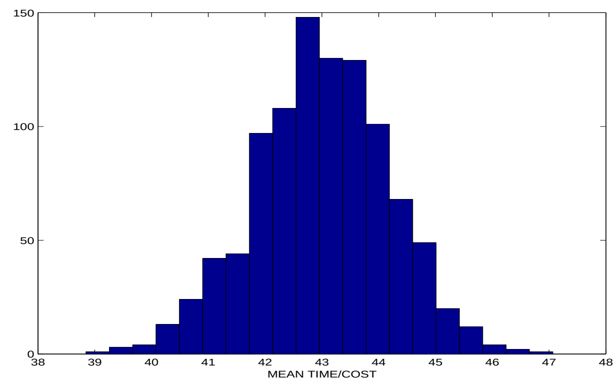

Figure 6: Empirical distribution of average of proportions (time-to-renewal / cost-of-renewal), from 500 renewals, for $\tau = 60$, – histogram from 1000 replications

6. Criterion based on quantiles

Criterion based on averaging, as $\max_\tau E_F\varphi(Y, \tau)$ in the preceding example, does not take in account possible variability of r.v. Y and is actually reasonable for optimizing over long time period. Even then the variability can be large. It is here shown just with the aid of simulation. We generated a series of 500 renewals according the preceding example, i.e. from Weibull($a = 100, b = 2$) distribution, for fixed $\tau = 60$. We computed always the relation of time to renewal to its cost and computed the mean from these 500 renewals. Such sequence was replicated 1000 times. The result is displayed in Figure 6.

We can as well compute directly distribution of values of function $\varphi(Y, \tau)$, for each fixed τ. It is easy if $\varphi(Y, \tau)$ is monotone w.r. to Y. Then we can try to optimize certain quantiles of this distribution.

In our example, we consider random variable $Z_\tau = \varphi(Y, \tau)$, with property $Z_\tau = \frac{Y}{C_1}$ if $Y \le \tau$ and $Z(\tau) = \frac{\tau}{C_2}$ if $Y > \tau$. If $Y \sim$ Weibull(a, b) and certain τ is selected, the conditional distribution of random variable $Z_\tau | Z_\tau \le \frac{\tau}{C_1}$ has distribution function

$$F_Z(z) = \frac{1 - \exp(-(\frac{z \cdot C_1}{a})^b)}{(1 - \exp(-(\frac{\tau}{a})^b)},$$

i.e. Z_τ has on interval $(0, \tau/C_1)$ Weibull distribution with parameters $(a/C_1, b)$ and $Z_\tau = \tau/C_2$ with probability $P(Y > \tau)$.

Thus, if we wish to maximize certain α-quantile of Z_τ (and still $C_1 > C_2$), optimal $\tau^*(\alpha)$ should be such that $P(Y > \tau^*(\alpha)) = 1 - \alpha$, i.e. $\tau^*(\alpha)$ is the α quantile of the distribution of Y. Guaranteed value reached by Z with probability $1 - \alpha$ is then $\tau^*(\alpha)/C_2$. It is a consequence of the form of function $\varphi(y, \tau)$.

For instance for $\alpha = 0.1$ and $Y \sim$ Weibull(100,2) we obtain that $\tau^*(\alpha) = 32.4593$ and 90% guaranteed value of Z is $\tau^*(\alpha)/C_2 = 32.4593$, too, because $C_2 = 1$.

7. Conclusion

We have studied the impact of randomly right-censored data to the increase of variability of statistical estimates and, consequently, to the imprecision of solution in a stochastic optimization problem. We have compared theoretical as well as empirical behavior of estimates in the parametric or non-parametric setting, in the cases of fully observed or censored data. We have proved asymptotic consistency of solutions based on censored data. Further, the influence of data incompleteness to optimal solutions has been studied on randomly generated examples.

References

[1] Akritas, M. G.: The central limit theorem under censoring. *Bernoulli* **6** (2000), 1109–1120.

[2] Anděl, J.: *Foundantions of Mathematical Statistics (in Czech: Základy matematické statistiky).* Matfyzpress, Praha, 2005.

[3] Breslow, N. and Crowley, J. E.: A large sample study of the life table and product limit estimates under random censorship, *Ann. Statist.* **2** (1974), 437–453.

[4] Dupačová, J. and Wets, R.: Asymptotic behaviour of statistical estimates and optimal solutions of stochastic optimization problems, *Ann. Statist.* **16** (1984), 1517–1549.

[5] Gill, R.: Large sample behaviour of the product-limit estimator on the whole line, *Annals Statist.* **11** (1983), 49-58.

[6] Huber, C.: Censored and truncated lifetime data. In *Recent Advances in Reliability Theory (N. Limnios and M. Nikulin, eds.)*, Birkhauser, Boston, 2000, 291–305.

[7] Kalbfleisch, J. D. and Prentice, R. L.: *The Statistical Analysis of Failure Time Data (2-nd Edition)*, Wiley, New York, 2002.

[8] Kaňková, V.: A remark of empirical estimates via economic problems. In *Proceedings of MME 2009 (H. Brožová ed.)*, Czech Univ. of Life Sciences, Prague, 2009.

[9] Kaňková, V. and Houda, M.: Empirical estimates in stochastic programming. In: Proceedings of Prague Stochastics 2006 (M. Hušková and M. Janžura, eds.), MATFYZPRESS, Prague, 2006, 426-436.

[10] Rejto, L.: On fixed censoring model and consequences for the stochastic case. In *Transactions from the 9-th Prague Conference on Stochastic Decision Functions, 1982,* Academia, Prague, 1983, 141-147.

[11] Robbins, H. and Siegmund, D.: Boundary crossing probabilities for a Wiener process and sample sums, *Ann. Math. Statist.* **41** (1970), 14101429.

[12] Volf, P.: Estimation of the mean from randomly censored data, *Problems of Control and Information Theory* **16** (1987), 233-241.