ON PRECISION OF STOCHASTIC OPTIMIZATION BASED ON ESTIMATES FROM CENSORED DATA

PETR VOLF

In the framework of a stochastic optimization problem, it is assumed that the stochastic characteristics of optimized system are estimated from randomly right–censored data. Such a case is frequently encountered in time-to-event or lifetime studies. The analysis of precision of such a solution is based on corresponding theoretical properties of estimated stochastic characteristics. The main concern is to show consistency of optimal solution even in the random censoring case. Behavior of solutions for finite data sizes is studied with the aid of randomly generated example.

Keywords: optimization, censored data, product-limit estimate, consistency

Classification: 62N02, 62P25

1. INTRODUCTION

The problem of stochastic optimization is often formulated as to find a solution to

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

where φ is a cost function, v is an input "control" variable from certain feasibility set $\boldsymbol{V}$, E_F stands for the expectation under distribution function F, and, finally, Y is a random variable (or vector) with distribution function F. In the present paper it is assumed that F has to be estimated, in a non-parametric way, from observed data. Then, a standard approach considers a solution of (1) using estimated F instead of 'true' one. A number of papers has already dealt with the problem of optimization based on statistical estimates. The main concern was asymptotic optimality of solution (see for instance Dupačová and Wets [2]), as well as the rate of convergence. A nice overview of the history as well as of the most recent results in this area is provided in Houda and Kaňková [3].

The situation is even more complicated if the data available for estimation are not complete. We shall consider here 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 have been studied by a number of authors, mostly in the framework of statistical survival analysis (cf. Kalbfleisch and Prentice [4]). 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 and to formulate conditions ensuring asymptotic consistency of these results. This problem has been already examined by Volf [8] where the solution was based on the properties of the mean value estimated from censored data. Both parametric and non-parametric forms of unknown distribution function F were considered. The asymptotic optimality of solutions was proven, however without examination of convergence rates.

The approach of the present paper differs from that of Volf [8], though some starting points are the same. We concentrate to the non-parametric estimate of F. In the next section we recall the product-limit estimate (PLE) as a generalization of the empirical distribution function and some of its large sample properties. Then, in Section 3, we follow the approach proposed in Kaňková [5] and generalize it for the censored data case. It is the basis for the main results of the paper contained in Section 4. We prove the convergence of optimal solutions computed from estimated distribution function to the optimum (1), along with the rates of convergence. In Theorems 1 and 2 we also distinguish the cases with light and heavy tails. Finally, in Section 5 an example deals with a simple optimization problem, properties of obtained solutions are illustrated with the aid of simulations.

2. RANDOM CENSORING AND PRODUCT-LIMIT ESTIMATE

In the present section we shall recall some results concerning estimates from randomly right-censored data. They are collected in survival analysis literature, let us mention again Kalbfleisch and Prentice [4], and they shall be used in next parts of the paper.

Let Y be a continuous-type random variable characterizing, for instance, a random time to certain event. Let another continuous random variable Z be a censoring variable, Y, Z being 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 $(X_i, \delta_i, i = 1, \ldots, N)$. Notice that the case without censoring is obtained when $G(t) \equiv 0$ on the 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.

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

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

Again, notice that when all $\delta_i = 1$, we obtain the empirical survival function. The

following proposition is due to Breslow and Crowley [1]:

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

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

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

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

Here, $F_N(t) = 1 - \overline{F}_N(t)$. The asymptotic variance function can be estimated by its empirical version:

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

which is consistent in probability, uniformly w.r. to $t \in [0, T]$ (see again Breslow and Crowley [1]).

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 (2) it is also seen that the variance in the case with censoring (when $\overline{G}(t) \le 1$) is larger than without it (i.e. when $\overline{G}(t) = 1$ on whole $[0, T]$). It has been proved (see, for instance, Robbins and Siegmund [7]) that the approximate two-sided $1 - \alpha$ band for $|D_N(t)|$ on $(0, T)$ is based on $c_\alpha = \sqrt{-\ln(\frac{\alpha}{2})/2}$. Namely, in the case without censoring c_α is the approximate distribution-free critical value for the Kolmogorov–Smirnov test,

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

In the case with censoring, we have

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

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

The asymptotic properties of the PLE listed above concern just to any interval $(-\infty, T)$ with finite T. However, for our purposes the consistency on the whole line is needed. From a set of relevant results, we shall refer here to the paper of Rejto [6], where the following statement is proved.

Proposition 2. (Rejto [6]) 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 $a\overline{F}(t)^b \le \overline{G}(t)$ on $[\tau, \infty)$. Then almost surely

$$\sup_{-\infty < t < \infty} |F_N(t) - F(t)| = \mathcal{O} \left(\left[\frac{\ln N}{N} \right]^{\frac{1}{2+b}} \right). \tag{3}$$

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. Notice that the proposition contains no explicit assumption concerning the distributions tails. Just implicitly, it requires that the right tail of censoring distribution G should be heavier than the same tail of the distribution F. In the case without censoring ($\bar{G}(t) \equiv 1$) the proposition states a known fact that (3) holds for any $b > 0$, hence $N^\beta \sup_t |F_N(t) - F(t)| \to 0$ a.s. for any $\beta < \frac{1}{2}$).

3. PRELIMINARY RESULTS

In next two sections we shall investigate 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 v) of problem (1) obtained from a case with non-complete information on F converge to optimal ϕ^* and v^* obtained when F is known.

There are several ways how to address this problem. We shall follow here the approach starting from certain results summarized in Kaňková [5] and later on in Houda and Kaňková [3]. Let us recall the statement from which other assertions can be derived:

Proposition 3. (Kaňková [5], Proposition 2.1) Let F, G be two distribution functions on R, let $v \in \boldsymbol{V}$, $\boldsymbol{V}$ be a compact set. If:

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

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

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

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

It is seen that two sets of assumptions are needed, the first concerning the properties of criterion function $\varphi(y, v)$. The second set of assumptions ensures the closeness of distributions, here expressed via the Wasserstein metric. Hence, we should derive such a closeness from the convergence of estimated distribution function. As it has been said, for the case of non-censored data the solution of the problem, under specific conditions, has been reported in Kaňková ([5] Corollary 2.5 and 2.7). In the sequel we shall follow a similar approach for the case of randomly right-censored data and the PLE.

Proposition 4. Let the assumptions of Proposition 2 hold, with some $a, b \in (0, 1]$, $\tau > 0$. Further, let there exist $C, D, T > 0$ such that $f(x) \leq C \exp(-D|x|)$ on $(-\infty, -T) \cup (T, \infty)$. Then, for each $\beta < 1/(2 + b)$, almost surely

$$\lim_{N \to \infty} N^\beta \int_{-\infty}^{\infty} |F(x) - F_N(x)| \, \mathrm{d}x \; = \; 0.$$

P r o o f.   For $y, z > T$ we obtain that $\bar{F}(y) = \int_y^\infty f(x) \, \mathrm{d}x \leq C/D \cdot \exp(-Dy)$ and $\int_z^\infty \bar{F}(y) \, \mathrm{d}y \leq C/D^2 \cdot \exp(-Dz)$. Also $\int_{-\infty}^{-z} F(y) \, \mathrm{d}y \leq C/D^2 \exp(-Dz)$, it can be shown in the same way.

Let us now select a sequence of numbers $A_N \to \infty$ when $N \to \infty$ such that

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

Further, let us consider a version of the PLE F_N truncated in such a way that $F_N = 0$ on $(-\infty, -A_N)$, $F_N = 1$ on (A_N, ∞). Then, almost surely, when already $A_N > T$,

$$\int_{-\infty}^{\infty} |F(x) - F_N(x)| \, \mathrm{d}x = \int_{-\infty}^{-A_N} F(x) \, \mathrm{d}x + \int_{-A_N}^{A_N} |F(x) - F_N(x)| \, \mathrm{d}x + \int_{A_N}^{\infty} \overline{F}(x) \, \mathrm{d}x$$

$$\leq 2 \frac{C}{D^2} \exp(-D \cdot A_N) + 2 \cdot A_N \cdot \mathcal{O} \left(\left[\frac{\ln N}{N} \right]^{\frac{1}{2+b}} \right).$$

It is seen that if we take for instance $A_N = \ln(N)/D$, we obtain desired statement of the proposition. $\qquad\square$

Remark 1. In the proof we had to consider a truncated form of the PLE. In practice, as we always deal with finite N, this means no restriction.

Similarly like Kaňková [5] in Corollary 2.7 we shall investigate also the case with heavier tails of Pareto type.

Proposition 5. Let the assumptions of Proposition 2 hold, with some $a, b \in (0, 1]$, $\tau > 0$. Further, let there exist $C, T > 0$ and $\alpha > 1$ such that $f(x) \leq C \cdot |x|^{-\alpha-1}$ on $(-\infty, -T) \cup (T, \infty)$. Then, for each $\beta < \frac{1}{(2+b)} \cdot (1 - \frac{1}{\alpha})$, almost surely

$$\lim_{N \to \infty} N^{\beta} \int_{-\infty}^{\infty} |F(x) - F_N(x)| \, \mathrm{d}x = 0.$$

P r o o f.   Now, for $y, z > T$ we obtain that $\overline{F}(y) = \int_y^{\infty} f(x) \, \mathrm{d}x \leq C/\alpha \cdot y^{-\alpha}$ and $\int_z^{\infty} \overline{F}(y) \, \mathrm{d}y \leq C/(\alpha(\alpha - 1)) \cdot z^{-\alpha+1}$. Again, the same bound will be obtained for $\int_{-\infty}^{-z} F(y) \, \mathrm{d}y$.

Let us also consider a sequence of numbers $A_N \to \infty$ when $N \to \infty$ and a truncated version of the PLE such that $F_N = 0$ on $(-\infty, -A_N)$, $F_N = 1$ on (A_N, ∞). Then almost surely, when $A_N > T$,

$$\int_{-\infty}^{\infty} |F(x) - F_N(x)| \, \mathrm{d}x = \int_{-\infty}^{-A_N} F(x) \, \mathrm{d}x + \int_{-A_N}^{A_N} |F(x) - F_N(x)| \, \mathrm{d}x + \int_{A_N}^{\infty} \overline{F}(x) \, \mathrm{d}x$$

$$\leq 2 \frac{C}{\alpha(\alpha - 1)} \cdot A_N^{-\alpha+1} + 2 \cdot A_N \cdot \mathcal{O} \left(\left[\frac{\ln N}{N} \right]^{\frac{1}{2+b}} \right).$$

Let us now take $A_N = N^{\gamma}$ and find an optimal $\gamma > 0$ guaranteeing maximal speed of convergence. The last expression is then of the order

$$N^{\gamma(1-\alpha)} + (\ln N)^{\frac{1}{2+b}} \cdot N^{\gamma - \frac{1}{2+b}}.$$

Hence, optimal γ is such that $\gamma(1-\alpha) = \gamma - \frac{1}{2+b}$, i.e. $\gamma = \frac{1}{(2+b)\alpha}$. Therefore, with this γ, almost surely

$$\int_{-\infty}^{\infty} |F(x) - F_N(x)|\, \mathrm{d}x \leq \mathcal{O}\left((\ln N)^{\frac{1}{2+b}} \cdot N^{\frac{1}{2+b}(\frac{1}{\alpha}-1)}\right)$$

and the statement is proven. $\square$

4. MAIN RESULTS

In this part we wish to show consistency of solutions computed as optimal w.r. to the PLE of distribution function F. Therefore, we shall combine Proposition 3 with Propositions 4 or 5, respectively. Let us first summarize relevant assumptions.

A1. Variable $v \in V$, where V is a compact set (it also means that in (1) *infimum* over V becomes *minimum*).

A2. Function $\varphi(y, v)$ is a Lipschitz function of y on R, with Lipschitz constant C not depending on v.

A3. $E_F \varphi(Y, v)$ are finite for all $v \in V$.

A4. Function $\varphi(y, v)$ is uniformly continuous on $R \times V$.

A5. Assumptions of Proposition 2 hold, with some $a, b \in (0, 1]$, $\tau > 0$.

4.1. Convergence of optimal costs ϕ

Let us, for a distribution function H, denote $\phi^*(H) = \min_v E_H \varphi(Y, v)$ (provided it exists). The following two theorems are direct consequences of Proposition 3 and Propositions 4 or 5, respectively.

Theorem 1. Let assumptions A1–A5 hold. Further, let there exist $C, D, T > 0$ such that $f(x) \leq C \exp(-D \cdot |x|)$ on $(-\infty, -T) \cup (T, \infty)$. Then, for each $\beta < 1/(2+b)$, almost surely

$$\lim_{N \to \infty} N^\beta \cdot |\phi^*(F) - \phi^*(F_N)| = 0.$$

Theorem 2. Let assumptions A1–A5 hold. Further, let there exist $C, T > 0$ and $\alpha > 1$ such that $f(x) \leq C \cdot |x|^{-\alpha-1}$ on $(-\infty, -T) \cup (T, \infty)$. Then, for each $\beta < \frac{1}{(2+b)} \cdot (1 - \frac{1}{\alpha})$, almost surely

$$\lim_{N \to \infty} N^\beta \cdot |\phi^*(F) - \phi^*(F_N)| = 0.$$

4.2. Convergence of optimal solutions v

As V is a compact set, there always exists at least one solution in V. Therefore, let us denote

$$v_F^* = \arg\min_v \phi_F(v), \ \ \phi^*(F) = \phi_F(v_F^*), \ \ v_N^* = \arg\min_v \phi_{F_N}(v), \ \ \phi^*(F_N) = \phi_{F_N}(v_N^*).$$

Let us first state several propositions which will help us to prove the main result of the present part.

Lemma 1. From A4 it follows that functions $\phi_H(v)$ are continuous in $v \in V$, uniformly for all distribution functions $H(y)$.

P r o o f.   Consider $\varepsilon > 0$ and a distribution function $H(y)$. Then, due to A4, we can select $\delta > 0$ such that for each $v_0, v \in V : |v - v_0| \leq \delta$ it holds that $|\varphi(y, v) - \varphi(y, v_0)| \leq \varepsilon$, for each $y \in R_1$. Then

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

$\square$

The following proposition is a variant of Proposition 4 of Volf [8]. As its proof needs the first derivative of a Lipschitz function, we shall refer here to the following property:

Remark 2. A Lipschitz function $h : R \to R$ is differentiable almost everywhere, i. e. at every point outside a set R_0 of Lebesgue measure zero. Corresponding derivative ($\underline{h}'$, say) is therefore defined on $R - R_0$ and bounded there by the Lipschitz constant of function h. Further, $h(b) - h(a) = \int_a^b \underline{h}'(x) \, \mathrm{d}x$ for every a, b. For convenience, we can set $\underline{h}'(x) \equiv 0$ on R_0.

Proposition 6. 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. $h(y)$ is a Lipschitz function with Lipschitz constant $C < \infty$.

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

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

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

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

P r o o f.   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, $X_{N,1} \leq X_{N,2} \leq X_{N,N}$. Further, denote K, L such indices that

$$K = \min\{i : X_{N,i} \geq -A_N\} - 1, \ \ L = \max\{i : X_{N,i} \leq A_N\} + 1.$$

Denote, omitting index N, $T_K = -A_N$, $T_L = A_N$, $T_i = X_{N,i}$ for $i = K+1, \ldots, L-1$. If we consider again a version of F_N truncated in such a way that $F_N = 0$ on $(-\infty, -A_N)$, $F_N = 1$ on (A_N, ∞), 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

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

because $F_N(y) = F_N(T_{i-1})$ on $[T_{i-1}, T_i)$. Here $\underline{h}'$ is the derivative of h in the sense of Remark 2.

Further, by *per-partes* integration we obtain

$$h_N = \int_{-A_N}^{A_N} h(y) \, dF(y) = F(A_N)h(A_N) - F(-A_N)h(-A_N) - \int_{-A_N}^{A_N} \underline{h}'(y)F(y) \, dy.$$

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

$$- \int_{-A_N}^{A_N} \underline{h}'(y)(F_N(y) - F(y)) \, dy.$$

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

Therefore we may conclude that

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

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

Theorem 3. Let either assumptions of Theorem 1 or assumptions of Theorem 2 hold. Then, for $N \to \infty$, there exists a subsequence $\{v_{N(k)}^*\} \subset \{v_N^*\}$, $k = 1, 2, \ldots$ such that it converges almost surely, when $k \to \infty$, to some $v_0 \in \{\arg\min_v \phi_F(v)\}$.

P r o o f .   From either Theorem 1 or Theorem 2 and compactness of $\boldsymbol{V}$ it follows that there exist (at least one of each) optimal solutions v_F^*, v_N^* and that a.s., when $n \to \infty$,

$$\phi_{F_N}(v_N^*) \to \phi_F(v_F^*) = \phi^*(F).$$

From compactness of V it also follows that there exists a sub-sequence $\{N(k), k = 1, 2, \ldots\} \subset \{1, 2, \ldots\}$ and $v_0 \in V$ such that a.s., for $k \to \infty$,

$$v_{N(k)}^* \to v_0, \quad \phi_{F_{N(k)}}(v_{N(k)}^*) \to \phi^*(F),$$

while, in general, $\phi_F(v_0) \geq \phi^*(F)$. Further, Lemma 1 implies that also $\phi_{F_{N(k)}}(v_0) \to \phi^*(F)$ a.s. for $k \to \infty$.

On the other hand, from Proposition 6 we have that simultaneously $\phi_{F_{N(k)}}(v_0) \to \phi_F(v_0)$. Therefore we may conclude that a.s. $\phi_F(v_0) = \phi^*(F)$ and $v_0 \in \{\arg\min_v \phi_F(v)\}$. $\square$

Notice that, in general, v_0 may be random, distributed on the set $\{\arg\min_v \phi_F(v)\}$, while v_F^*, $\phi^*(F)$ are not; when v_F^* is unique then $v_0 = v_F^*$ a.s.

From a practical point of view it is more appropriate to study behavior of $\phi_F(v_N^*)$ instead of $\phi_N(v_N^*)$, because v_N^* are computed optimally w.r. to estimated F_N, while real costs are given as the expectation w.r. to the 'true' (but unknown) distribution function F. As, in general, $\phi_F(v_N^*) \geq \phi_N(v_N^*)$, it seems to be rather difficult to establish the convergence of $\phi_F(v_N^*)$ without some additional conditions, not speaking about specification of convergence rate. Nevertheless, at least the following statement can be proven.

Corollary 1. Let the assumptions of Theorem 3 hold. Further, let $v_{N(k)}^*$ be a converging sub-sequence of optimal solutions specified in Theorem 3. Then $\phi_F(v_{N(k)}^*) \to \phi^*(F)$ almost surely when $k \to \infty$.

P r o o f . The proof follows directly from the convergence $v_{N(k)}^* \to v_0$, with $v_0 \in \{\arg\min_v \phi_F(v)\}$, and from Lemma 1 ensuring that then also $\phi_F(v_{N(k)}^*) \to \phi_F(v_0)$, which equals $\phi^*(F)$. $\square$

In the next part the behavior of $\phi_F(v_N^*)$ is studied with the aid of a simple artificial (randomly generated) example.

5. EXAMPLE

Let the time to failure Y of a machine component be modelled by a continuous-type probability distribution with distribution function, density, survival function F, f, $\overline{F} = 1 - F$, 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 the criterion function be the proportion of the time to repair to the cost of repair, 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 $\boldsymbol{T}$ maximizing the mean

$$\phi_F(\tau) = E_F \varphi(Y, \tau) = \int_0^\tau \frac{y}{C_1} \, \mathrm{d}F(y) + \frac{\tau}{C_2} \overline{F}(\tau). \tag{4}$$
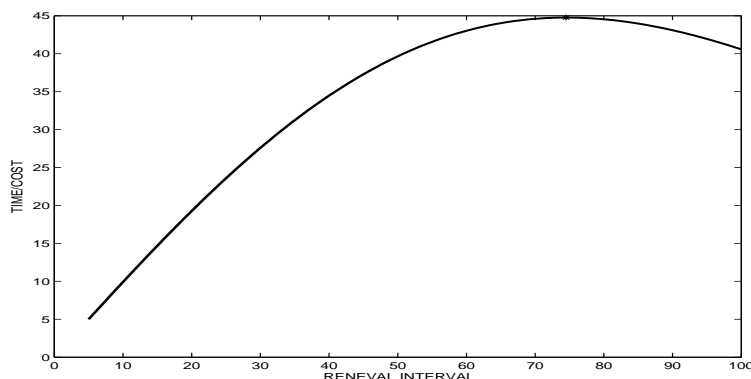
Fig. 1. $\phi_F(\tau)$ vers. τ, with optimal point.

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 when F is known with solutions in cases when lifetime distribution is estimated non-parametrically 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 $\overline{F}(t) = \exp\Big(-\big(\frac{t}{a}\big)^b\Big)$, numerical characteristics are $EY \sim 89$, $\mathrm{std}(Y) \sim 46$. Costs of repairs are fixed to $C_1 = 10$, $C_2 = 1$. When the distribution function F is known, along (4) there exists unique optimal solution with

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

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

In the present part it is assumed that the distribution of Y is not known and therefore is estimated from generated data non-parametrically, by the PLE. In all cases (without or with censoring) 100 samples of N observations Y_i were generated from the Weibull distribution mentioned above. In cases with censored data, censoring variables Z_i followed uniform distribution on $[0, 250]$. Hence, they have survival function $\overline{G}(z) = (250-z)/250$ (value 250 corresponds roughly to 0.998 quantile of distribution of Y). The rate of censoring was then about $36\% \sim EY/250$. For comparison, we generated data of sizes $N = 100, 200, 500, 1000$.

Figure 2 displays cloud of 100 non-parametric estimates of F obtained from 100 generated samples, the cases without censoring are plotted in the left subplot, the right subplot shows estimates obtained from censored data. It is well seen how the variability
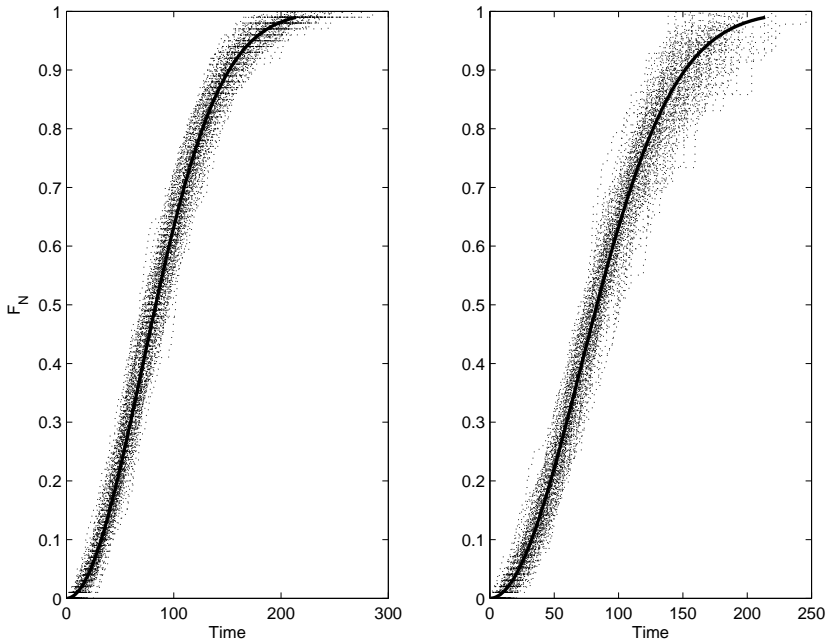
Fig. 2. Set of 100 estimates of distribution function, $F_N(t)$, from non-censored (left) and censored data (right), for $N = 100$. 'True' distribution function $F(t)$ is plotted by solid curve.

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.

Figure 3 displays optimal solutions τ_m, each obtained as the maximizer of (4) with mth estimate F_N of F, $m = 1, 2, \ldots, 100$. We then computed corresponding $\phi_F(\tau_m)$, i.e. the mean costs under τ_m evaluated w.r. to 'true' Weibull 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. In order to compare convergence and variability of results for different data sizes N, the data of the same type were generated also with $N = 200, 500, 1000$. Table 1 contains sample characteristics of obtained 'sub-optimal' solutions, each computed from 100 repetitions of data of size N. The table contains also results observed for the higher rate of censoring. Namely, in the case denoted as "censoring 2" variables Z_i followed Weibull distribution with parameters $a_Z = 80, b_Z = 2$. Then the rate of censoring was about 60 %. As expected, imprecision of results is higher, in particular for larger N.
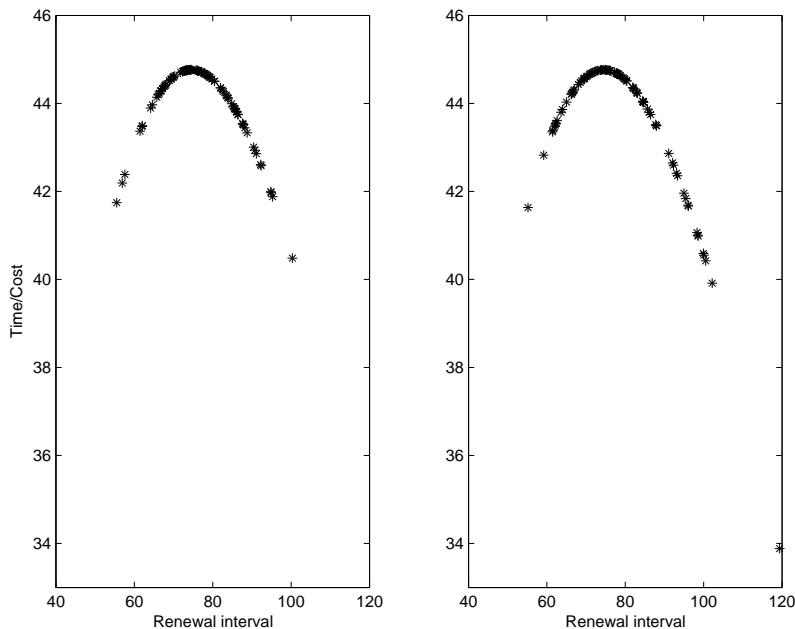
Fig. 3. Optimal solutions τ_m (on x axis) and corresponding $\phi_F(\tau_m)$
(on y axis) based on nonparametric estimates displayed in Figure 2,
for non-censored (left) and censored (right) cases, both for $N = 100$.

N	non-cens.:	mean	std	cens.1: mean	std	cens.2: mean	std
100	τ_m	76.751	9.154	78.288	11.585	76.453	11.437
	$\phi_F(\tau_m)$	44.123	0.839	43.763	1.553	43.821	1.352
200	τ_m	75.469	8.088	78.725	10.015	77.217	10.782
	$\phi_F(\tau_m)$	44.283	0.611	43.948	1.172	43.903	1.319
500	τ_m	75.706	5.492	75.239	6.897	76.570	8.508
	$\phi_F(\tau_m)$	44.530	0.289	44.409	0.535	44.222	0.915
1000	τ_m	74.590	4.647	75.517	4.691	76.365	6.647
	$\phi_F(\tau_m)$	44.597	0.203	44.596	0.256	44.418	0.530

Tab. 1. Sample means and standard deviations of τ_m and $\phi_F(\tau_m)$,
for different N and two different proportions of censorig. Case with
$N = 100$ and censoring 1 corresponds to samples plotted in Figures 2
and 3.

Observed (prevailing) decrease of means of τ_m with increasing N is probably caused
by a nonsymmetric shape of dependence of $\phi_F(\tau)$ on renewal interval τ, which is seen

both in Figure 1 and Figure 3. The mean of τ_m has therefore tendency to be greater than optimal τ^*, simultaneously converging to it.

6. CONCLUSION

The contribution was devoted to the study of impact of randomly right-censored data to the increase of variability of statistical estimates and, consequently, to the imprecision of solution of a stochastic optimization problem. To this end, relevant theoretical properties of nonparametric product-limit estimator of distribution function were recalled. They were utilized to proving asymptotic consistency of optimal solutions computed from censored data. The impact of data incompleteness to imprecision of solution was studied also with the aid of randomly generated example.

REFERENCES

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

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

[3] M. Houda and V. Kaňková: Empirical estimates in economic and financial optimization problems. Bull. Czech Econometr. Soc. *19* (2012), 50–69.

[4] J. D. Kalbfleisch and R. L. Prentice: The Statistical Analysis of Failure Time Data. Second edition. Wiley, New York 2002.

[5] V. Kaňková: Empirical estimates in stochastic optimization via distribution tails. Kybernetika *46* (2010), 459–471.

[6] L. Rejto: On fixed censoring model and consequences for the stochastic case. In: Trans. 9th Prague Conference on Stochastic Decision Functions 1982, Academia, Prague 1983, pp. 141–147.

[7] H. Robbins and D. Siegmund: Boundary crossing probabilities for a Wiener process and sample sums. Ann. Math. Statist. *41* (1970), 1410–1429.

[8] P. Volf: On precision of optimization in the case of incomplete information. Bull. Czech Econometr. Soc. *19* (2012), 170–184.

Petr Volf, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
    e-mail: volf@utia.cas.cz