On problem of optimization under incomplete information

Petr Volf[1]

Abstract. In the paper we study consequences of incomplete information to uncertainty of results of stochastic optimization. Stochastic characteristics of optimized system are evaluated from observed data, moreover, the data may be incomplete. Namely, we consider the random censoring of observations frequently encountered in time-to-event (of lifetime) studies. The analysis of uncertainty will be 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 rely on a probabilistic model of optimized system. Then the optimization task can, schematically, have the form

$$\varphi(F) = \inf_{\boldsymbol{y}} E_F \boldsymbol{C}(\boldsymbol{y}, \boldsymbol{v}), \tag{1}$$

where $\boldsymbol{C}$ is a cost function, $\boldsymbol{y}$ are input variables (observed), E_F stands for the expectation under distribution function F, and, finally, $\boldsymbol{v}$ 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 optimal solutions. 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' ones. An investigation of usage of empirical (estimated) characteristics in stochastic optimization problems started already in 70-ties. A set of papers is available, let us mention here just two: Dupačová and Wets (1984), and from more recent time Kaňková (2009) with 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 deigns 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

[1]Institute of Information Theory and Automation, Prague 8, e-mail: volf@utia.cas.cz

parametric and non-parametric cases. To this end, certain theoretical properties of estimates under random right censoring will be recalled. In the next two parts, 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 Part 2) as well as in the case of estimated parameters (Part 3). Finally, we shall study the consequence of incomplete data to a certain problem of optimization, in Part 4. An example will deal with optimal maintenance schedule, properties of obtained 'quasi-optimal' solution will be illustrated with the aid of simulations.

2 Non-parametric case and product-limit estimate

Let us first recall the scheme of randomly right-censored data. Consider a random variable Y characterizing a random time to certain event, another random variable Z is a censoring variable, both are positive, continuous and mutually independent. Further, denote $f(y)$, $g(z)$, $F(y)$, $G(z)$, $\overline{F}(y) = 1-F(y)$, $\overline{G}(z) = 1 - G(z)$ density, distribution and survival functions of both variables. It is assumed that we observe just $X = \min(Y, Z)$ and $\delta = 1[Y \le Z]$, i. e. δ indicates whether Y is observed or censored from right side. The data are then given as random sample $(X_i,\ \delta_i,\ i = 1, \dots, N)$. Notice that the case without censoring is obtained when $G(z) \equiv 0$ on region where $F(y) < 1$.

A generalization of empirical distribution function is well known Kaplan–Meier "Product Limit Estimate" of survival function. Let us first sort (re-index) the data in increasing order, $X_1 \le X_2 \le \dots \le X_N$, then this estimator has the form

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

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

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

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

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

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

It means that, in other words, $V_N(t)$ converges in distribution on $[0, T]$ to Brown process $\beta(C(t))$, or the process $D_N(t) = V_N(t)/(1+C(t))$ to Brownian bridge on $[0,\ C(T)/(1+C(T))]$. 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 uniformly on $[0, T]$.

For the case without censoring we obtain that $C(t) = F(t)/\overline{F}(t)$ and $D_N(t) = \sqrt{N}(F_N(t) - F(t))$ leading to standard Kolmogorov–Smirnov statistics. From (4) it is also seen that 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 $[0, T]$). Further, it has been proved (see, for instance Robbins and Siegmund, 1970) that for c, $d > 0$ and sufficiently large T,

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

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 $(0, C(T))$, 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

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

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 4.

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 and parameter θ is estimated by the method of maximum likelihood. The precision of estimation will 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. Naturally, if θ is multi-dimensional, we consider a vector of partial derivatives, What is important from our point of view, $I^{-1}(\theta)$ is also the asymptotic variance of $\sqrt{N}(\hat{\theta}_N - \theta)$, where $\hat{\theta}_N$ is the maximum likelihood estimate from random sample of extent N.

In the case of random right censoring, the log of likelihood (its part depending on θ), is $\ln L(\theta, X) = \delta \cdot \ln f(X) + (1-\delta) \cdot \ln \bar{F}(X)$. Hence, with notation $f' = \mathrm{d}f/\mathrm{d}\theta$ and $\bar{F}' = \mathrm{d}\bar{F}/\mathrm{d}\theta$,

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

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

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

which is positive and is larger when $\bar{G}(x) \equiv 1$, i.e. when there is no censoring. Again, a more concrete comparison is presented within the example in Part 4.

4 Example of optimization problem

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 (fixed) time from renewal to preventive repair.

Then, the mean time between renewals is $ET(\tau) = \tau \cdot P(Y > \tau) + \int_0^\tau t f(t)\mathrm{d}t = \int_0^\tau \bar{F}(t)\mathrm{d}t$, while the mean cost of one renewal equals $EC(\tau) = C_1 \cdot F(\tau) + C_2 \cdot \bar{F}(\tau)$. Our task is to find optimal τ minimizing

$$\min_\tau \phi_F(\tau) = \min_\tau EC(\tau)/ET(\tau). \tag{5}$$

In the sequel the lifetime distribution will be specified and we shall compare the deterministic solution provided F is known, and the variability of 'quasi-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 $\overline{F}(t) = \exp\left(-\left(\frac{t}{a}\right)^b\right)$, numerical characteristics are $EY \sim 89$, $std(Y) \sim 46$. Costs of repairs were fixed as $C_1 = 10$, $C_2 = 1$. When the distribution function F is known, there exists an unique optimal solution with $\tau^* = 33.64$ and minimal costs per time unit $\phi(\tau^*) = 0.061$.

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 from four cases (parametric or nonparametric case, without or with censoring) 100 samples of 100 observation 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 % $\sim EY/250$.

4.1 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 1, left plot shows estimates from non-censored 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 of (5), τ_m and $\phi(\tau_m)$, $m = 1, 2, ..., 100$, was computed. They are shown in Figure 2, again for non-censored (left) and censored cases (right). Table 1 contains also sample means and standard deviations of those 'optimal' τ_m and $\phi(\tau_m)$.

	non-censored: I	I-std	sample mean	sample std	censored: I	I-std	sample mean	sample std
a	$3.80 \cdot 10^{-4}$	5.13	100.3849	5.5523	$2.64 \cdot 10^{-4}$	6.15	100.0807	6.9541
b	0,409	0,154	2.0147	0.1585	0.287	0.187	2.0074	0.1680
τ^*			33.9970	1.9425			33.9070	2.3610
$\phi(\tau^*)$			0.0606	0.0071			0.0611	0.0070

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

4.2 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 3 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. 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 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 3 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.

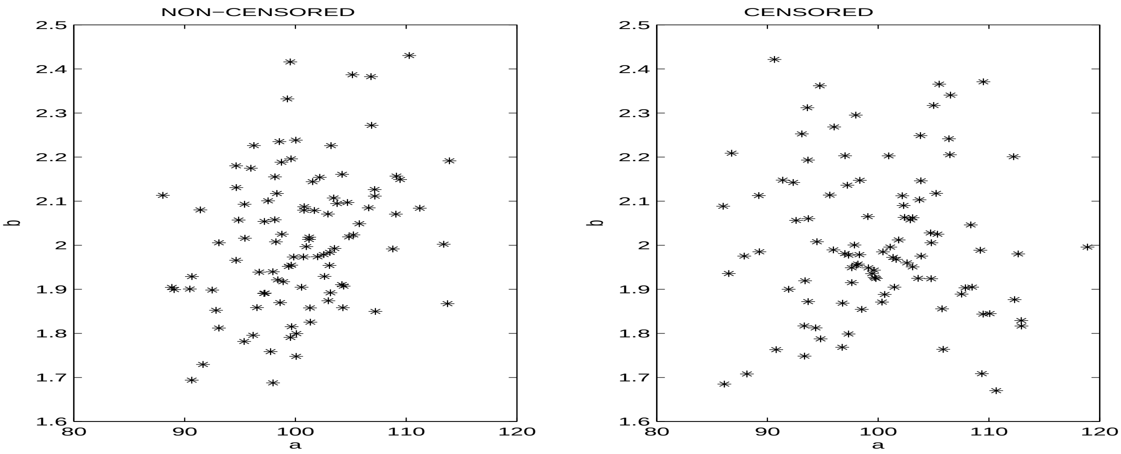

Figure 1: Maximum likelihood estimates of parameters (a_m, b_m) from 100 samples of non-censored (left) and censored data (right)

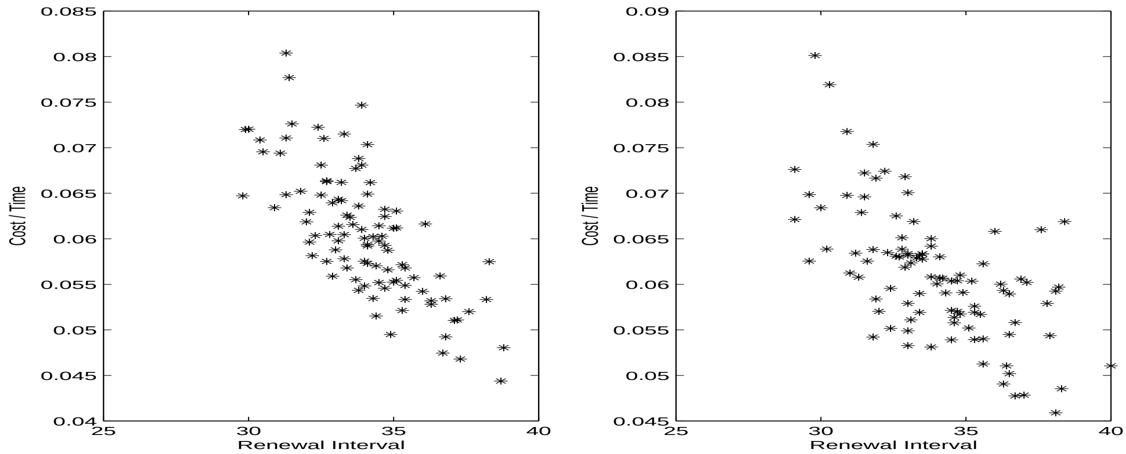

Figure 2: Optimal solutions $(\tau_m, \phi(\tau_m))$ based on estimates displayed in Figure 1, for non-censored (left) and censored (right) cases

Figure 4 displays optimal solutions τ_m^* and $\phi(\tau_m^*)$, each obtained as the solution of (5) with m-th estimate of F, $m = 1, 2, ..., 100$. Again, 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.

5 Conclusion

We have studied the impact of variability of statistical estimates to solutions in stochastic optimization problems. We have compared theoretical as well as empirical behavior of estimates in various situations, namely in the parametric or non-parametric setting, cases of fully observed or randomly right-censored data. Influence of variability of estimates to imprecision of optimal solutions has been studied on an example with a model case and randomly generated data.

Acknowledgements

The research is supported by the project of GA CR No 402/10/0956.

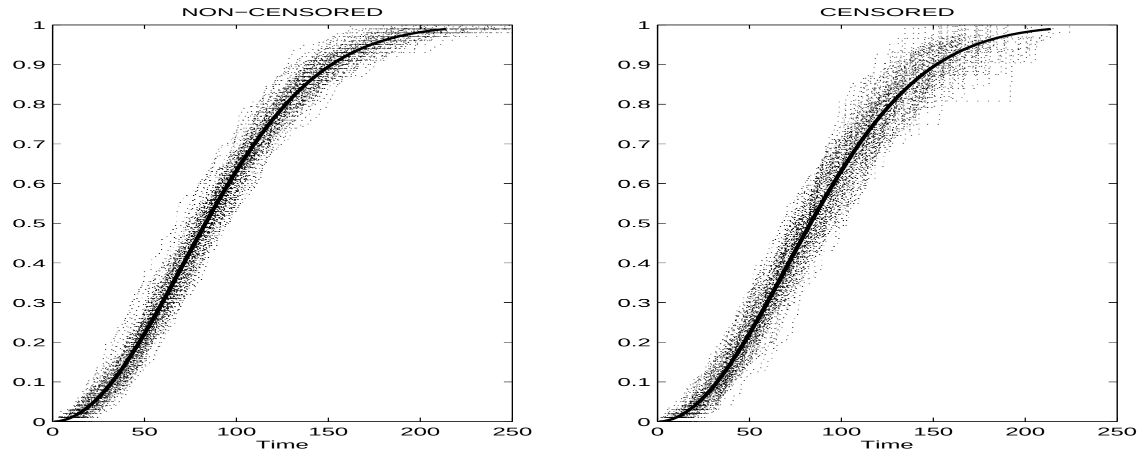

Figure 3: Set of 100 estimates of distribution function, $F_m(t)$, from non-censored (left) and censored data (right). 'True' distribution function $F(t)$ is plotted by solid curve

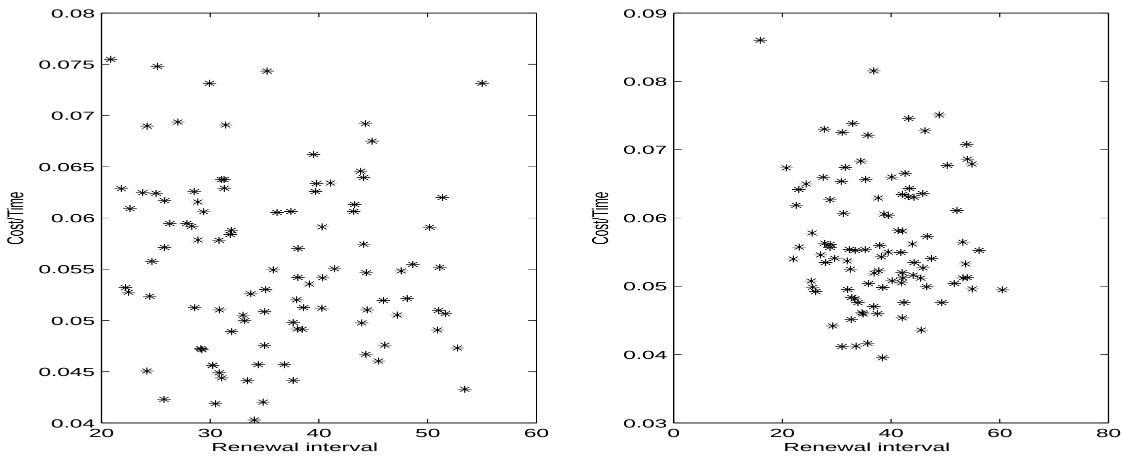

Figure 4: Optimal solutions $(\tau_m, \phi(\tau_m))$ based on nonparametric estimates displayed in Figure 3, for non-censored (left) and censored (right) cases

References

[1] 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.

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

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

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

[5] 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.

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