Croatian Operational Research Review
CRORR **13**(2022), 227-237

On impact of statistical estimates on precision of stochastic optimization

Petr Volf[1]

[1] *Institute of Information Theory and Automation, AS CR*
Pod vodarenskou vezi 4, Praha 8, Czech Republic
E-mail: ⟨volf@utia.cas.cz⟩

Abstract. This paper studies the consequences of imperfect information for the precision of stochastic optimization. In particular, it is assumed that the stochastic characteristics of an optimization problem depend on unknown parameters estimated from available data. First, a theoretical result is presented, showing that consistent parameters estimation leads to consistent optimization. Further, a type of the studied models is specified; it is assumed that the random variables present in the optimization problem are influenced by covariates. This influence is expressed via a parametric regression model, whose parameters have to be estimated and used instead of the unknown correct parameters values. The objective is then to explore, with the aid of simulations, the imprecision of the optimization based on these estimates. Several types of regression models are recalled, the variability of estimates and the related precision of sub-optimal solutions is studied in detail on an example dealing with optimal maintenance. The impact of random right-censoring on the deterioration of precision is studied as well.

Keywords: stochastic optimization, regression model, random ensoring, statistical estimation, optimal maintenance

Received: July 31, 2022; accepted: December 6, 2022; available online: December 22, 2022

DOI: 10.17535/crorr.2022.0017

---

1. Introduction

This paper studies the impact of incomplete information, namely the consequence of statistically estimating model parameters, to the precision of stochastic optimization. The standard formulation of a stochastic optimization problem is given as a search for solution to

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

where φ is a cost function, v are input (control) variables 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) possessing this distribution function. An alternative formulation may ask for the optimization of a certain quantile of random criterion $\varphi(Y, v)$ instead of its expectation. A combination of several criteria may also be considered, which would lead to a multi-objective problem.

When the distribution function F is not known, it has to be estimated from observed data. This estimate is random (based on available data), hence the obtained solution based on an estimated distribution instead of the correct one is random as well. Either a parametric form of the underlying distribution can be assumed, or a non-parametric estimator is used. Both these instances were, for example, analyzed in [9] and [10], where the case of incomplete (censored) data has also been explored. In fact, the impact of using an empirical distribution instead of the exact one has been studied for a long time; see, e.g., [4].

This contribution considers just the parametric case. In Section 2, several assumptions concerning both the cost function φ and the distribution of Y are formulated in order to prove the proposition asserting the consistency of solutions based on consistent parameters estimates. This proposition provides a theoretical framework guaranteeing a successful use of the estimates instead of the unknown correct values of parameters.

In subsequent sections, we then focus on a related task, namely, what the precision of a solution is when the estimates are obtained from a finite data sample. In other words, we intend to explore the speed of convergence by comparing several data designs. Moreover, in the same setting of parametric models for the distribution of Y, it is assumed that Y depends on a set of covariates via a convenient parametric regression model. It means that the corresponding regression parameters should be estimated as well, which contributes to additional variability of the problem solution.

The contribution of this paper is thus two-fold: A theoretical result relating the convergence of parameters estimates to the convergence of related optimization problem solutions, and empirical exploration of this convergence speed.

The second problem is solved in the setting of regression models. Therefore, several types of parametric regression models are recalled in Section 3; for instance, a standard linear model, its transformation, and the Cox regression model, which is typically used in lifetime studies. In all of the presented instances, the consistent parameters estimates are available, and the convergence of the obtained sub-optimal solutions to the optimal results is therefore granted by Theorem 1. What remains is thus exploring the precision of solutions based on estimates from finite data samples. This is done in Section 4, which presents an example dealing with an optimal maintenance problem depending on a covariate (e.g. a load) via the log-linear regression model. A simulation study shows the behavior of sub-optimal solutions based on data samples of small and medium size. The last section then compares the results of the optimization based on censored data with the results obtained without censoring.

2. Consistency of optimum

Let us assume that the distribution function $F = F(y, \theta)$ is parameterized (by a parameter θ from a set $\Theta \subset R^m$) and is of continuous type, with a density function $f(y, \theta)$. The following statement on the consistency of optimal decision as a consequence of the consistency of the estimate of parameter θ is in fact a variant of well-known results (see again [4] or [7]), which are mostly based on the uniform convergence of estimated distribution functions. It will be shown that, in a parametric case, the convergence of parameter estimates is sufficient. Naturally, under appropriate assumptions. Let us first introduce some notation:

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

$$v_F^* = \arg\min_v \phi_F(v), \;\; \phi_F^* = \phi_F(v_F^*), \;\; v_N^* = \arg\min_v \phi(v, \hat{\theta}_N), \;\; \phi_N^* = \phi(v_N^*, \hat{\theta}_N). \tag{2}$$

Here, as in (1), φ is a cost function, $\phi(v, \theta)$ its expectation for the given values of control variable v and parameter θ. Further, θ_0 denotes the 'true' value of the parameter, while $\hat{\theta}_N$ its statistical estimate from data of size N. Then v_F^*, ϕ_F^* are optimal solutions with respect to θ_0, while v_N^*, ϕ_N^* are optimal with respect to $\hat{\theta}_N$.

Let us further assume that there exists a sequence of estimates $\hat{\theta}_N$ converging in probability to θ_0. If we use the Taylor expansion at θ_0, denoting $f'(y, \theta) = \frac{\partial f(y,\theta)}{\partial \theta}$, we obtain that

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

$$= \int_{-\infty}^{\infty} \varphi(y,v) f'(y, \overline{\theta}_N) \mathrm{d}y \cdot (\hat{\theta}_N - \theta_0), \tag{3}$$

where, for sufficiently large N, $\overline{\theta}_N$, lying between θ_0 and $\hat{\theta}_N$, is arbitrarily close to θ_0 (in the sense of convergence in probability).

Let us formulate several assumptions; first three concern general properties of the optimization criterion, they were already formulated in [7] in order to ensure the closeness of the solutions for sufficiently close (in the sense of L_1 norm) distribution functions. The other assumptions are inspired by (3).

A1. Values of variable v are from $\boldsymbol{V}$, where $\boldsymbol{V}$ is a compact subset in R^1.

A2. Functions $\varphi(y,v)$ are continuous in v on $\boldsymbol{V}$, uniformly with respect to $y \in R^1$.

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

A4. The distribution of r.v. Y is continuous, with density function $f(y;\theta)$ and its derivative $f'(y,\theta) = \partial f(y,\theta)/\partial\theta$ existing on Θ.

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

Theorem 1. Let $\hat{\theta}_N$ be a sequence of estimates of θ_0 consistent in probability. Further, let assumptions A1 through A5 hold. Then, for $N \to \infty$,

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

Proof: Let us divide the proof into several steps:

i) For a sequence of consistent estimates $\hat{\theta}_N$ it holds that each $\phi_N^* \le \phi_N(v^*)$ and that also $|\phi_N(v^*) - \phi_F^*| \to 0$ in P (due to A5). Then both $\liminf \phi_N^* \le \phi_F^*$ and $\limsup \phi_N^* \le \phi_F^*$; this assertion must hold a.s. Denote $\underline{\phi} = \liminf \phi_N^*$. We want to prove that $\underline{\phi} = \phi_F^*$ a.s. (notice that $\underline{\phi}$ is random, defined a.s., while ϕ_F^* is a constant).

ii) As each sequence converging in probability has a sub-sequence converging a.s. (see e.g. [5], Th. 3.4), there exists a sub-sequence of indices $\{N1\} \subset \{1, 2, ...\}$ such that $\underline{\phi} = \lim_{N1\to\infty} \phi_{N1}^*$ a.s., with the corresponding sub-sequence of solutions v_{N1}^*. Then, due to the compactness of $\boldsymbol{V}$, there is another sub-sequence with indices $\{N2\} \subset \{N1\}$ such that $\lim_{N2\to\infty} v_{N2}^* = \overline{v} \in \boldsymbol{V}$ exists a.s. Both $\underline{\phi}$ and $\overline{v}$ are defined a.s., to show this existence of theirs is the substance of this point.

iii) From A2 and ii) it follows that also $\lim_{N2\to\infty} \phi_{N2}(\overline{v}) = \underline{\phi}$, a.s. It follows that $\underline{\phi}$ cannot be smaller than the optimal value ϕ_F^*,

iv) Now we wish to prove that (at least in probability) $\lim_{N\to\infty} \phi_N(\overline{v}) = \phi_F(\overline{v})$. This fact follows directly from (3) and assumption A5.

v) When combining iii) and iv), we can see that $\phi_F(\overline{v}) = \underline{\phi}$ a.s. Simultaneously, $\underline{\phi} \le \phi_F^*$ holds true. It follows that $\underline{\phi} = \phi_F^*$ and $\overline{v}$ is also a solution, i.e., $\overline{v} \in \{\arg\min\, \phi_F(v)\}$ a.s. If the set $\{\arg\min\, \phi_F(v)\}$ contains just one point then $\overline{v}$ coincides with it a.s.

Theorem 1 concerns the case when the estimated parameters are used instead of the correct parameters values. It shows that $\phi^*_{N,k} = \phi(v^*_{N,k}, \hat{\theta}_{N,k}) \to \phi^*_F$ in probability. However, in practice, we derive v^*_N optimal with respect to estimated parameter $\hat{\theta}_N$ and then we use it in the setting governed by the underlying (unknown) correct distribution F. That is why the result is in fact $\phi(v^*_N, \theta_0)$. The next proposition therefore proves the convergence of these sub-optimal solutions to the optimum.

Corollary. Under the assumptions and with the notation of Theorem 1,

$$\lim_{k\to\infty} \phi(v^*_{N,k}, \theta_0) \;=\; \phi^*_F.$$

holds in probability.

Proof: From (3) and A5 it follows that $|\phi(v^*_{N,k}, \hat{\theta}_{N,k}| - \phi(v^*_{N,k}, \theta_0)) \le K \cdot |\hat{\theta}_{N,k} - \theta_0|$. As A5 holds uniformly for each $v \in \boldsymbol{V}$ and $\hat{\theta}_{N,k} \to \theta_0$ in probability, Theorem 1 simultaneously claims that $\phi(v^*_{N,k}, \hat{\theta}_{N,k}) \to \phi^*_F$ as well, the proposition is proved.

We will assume below that the optimization problem depends on a set of covariates. Let us denote the space of covariates $\boldsymbol{Z} \subset R^k$, and their values $z \in \boldsymbol{Z}$. The impact of covariates is then expressed via a convenient regression model. Such a case may correspond to a number of real problems. The final part of this paper deals with one of them, namely the problem of optimal maintenance depending on the load affecting the maintained device.

In the derivation of the main result above, the value of covariate z was not included. In the context of the following setting dealing with (parametric) regression models, we can imagine that the covariate is taken as fixed. Nevertheless, the regression parameters are also subject to statistical estimation. In other words, the statements hold for any consistent sequence of model parameters estimates $\hat{\theta}_N$ and for such $z \in \boldsymbol{Z}$ that the assumptions listed above are fulfilled.

3. Examples of regression models and their parameters

A parametric regression model means that the type of distribution $F(y; z, \theta)$ of random variable Y for the given value of covariate z is known, while the parameters θ are unknown; their values lie in a set $\Theta \subset R^m$, containing both the parameters of a "baseline" distribution and the parameters characterizing the regression. Let us mention here several examples often used in practice:

1. Standard linear regression model, in which

$$Y = \alpha + \beta \cdot z + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2). \tag{4}$$

   Here the "baseline" parameters are α, σ, all parameters are $\theta = (\alpha, \sigma, \beta)$.

   Details on the analysis of the linear regression model and the corresponding confidence intervals and bands can be found elsewhere, e.g. in Lehmann and Romano [8], Ch. 7 and 9.

2. Let us consider Example 1 described above, however, with a transformation $T = \exp(Y) = \exp(\beta \cdot z) \cdot \epsilon$, where the random variable ϵ now has the lognormal distribution with parameters (α, σ). Let us denote the distribution function of ϵ by $F_e(.)$. The distribution function of T, for given covariate z, then equals

$$F_T(t; z) = P(T \le t; z) = P(\epsilon \le \exp(-\beta \cdot z) \cdot t) = F_e(\exp(-\beta \cdot z) \cdot t). \tag{5}$$

We can see that this instance leads to the so-called Accelerated Failure Time (AFT) regression model, which is often used in statistical reliability analysis to describe the consequences of a load or degradation (expressed via covariates) to the lifetime of a device. For more information see, e.g., [1].

3. Let the baseline distribution be the Weibull one, with distribution function $F_0(y) = 1 - \exp\{-(y/a)^b\}$ defined for $y > 0$, and with parameters, both positive, a, b. The dependence on a covariate z can be described via the Cox (proportional hazard) regression model (cf. [6]):

$$F(y; z) = 1 - \exp\{-(y/a)^b \cdot \exp(\beta\, z)\} = 1 - \exp\{-(\frac{y}{a(z)})^b\}, \tag{6}$$

where $a(z) = a \cdot \exp\{-\beta\, z/b\}$. The scale parameter $a(z)$ thus depends on the covariate, while the shape parameter remains b. We can formally write that $\theta = (a, b, \beta)$ is the set of the parameters we wish to estimate.

Parameters θ are estimated, as a rule, with the aid of the maximum likelihood estimation (MLE) method. Good properties of the MLE stem out of the so-called regularity conditions concerning the distributions $F(y; z, \theta)$. The formulation of these conditions can be found in statistical textbooks, for instance in [3], Ch. 7.3. If they are fulfilled then there exists a consistent sequence of the estimates, i.e. such that $\hat{\theta}_N \to \theta_0$ in probability when the data size $N \to \infty$. Here, θ_0 again denotes the 'true' value of the parameter. Further, the estimates are asymptotically normal, which means that the distribution of $\sqrt{N}(\hat{\theta}_N - \theta_0)$ converges to the normal distribution with zero mean and finite variance given by the inversion of the Fisher information matrix, $I^{-1}(\theta_0)$. In practice, the MLE is based on the data $\{y_i, z_i\}, i = 1, ..., N$ and maximizes the log-likelihood function

$$L_N(\theta) = \sum_{i=1}^{N} \ln f(y_i; z_i, \theta)) \tag{7}$$

over Θ. A consistent estimate of the Fisher information is then obtained as

$$I_N(\theta) = -\frac{1}{N}\left(\frac{\mathrm{d}(L_N(\theta))^2}{\mathrm{d}^2\theta}\right). \tag{8}$$

In the preceding part, it has been shown that the consistency of an estimate of θ ensures the consistency of the solutions computed as optimal with respect to this estimate. However, the estimates of parameters are in reality computed from a limited data sample $\{y_i,\, z_i,\, i = 1, ..., N\}$ providing just limited and random information for the identification of the system to be optimized. That is why we will now study the precision of the estimates and, consequently, of the solutions based on these estimates. We will proceed with the aid of randomly generated data from a selected type of the regression model.

The setting outlined above can also be interpreted as the regression parameter being an additional parameter; estimating it increases the uncertainty of the solution. It should be mentioned here that the confidence of the estimates depends also on the covariate design, not only on the data size. The impact of the covariate design has already been explored sufficiently elsewhere, cf. again remarks on the design of experiments in [8].

4. Example

Let us consider the following rather simple example of a stochastic optimization problem: A component of a machine has its time to failure T given (modeled) by a continuous-type probability distribution with distribution function, density, and 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, $0 < C_2 < C_1$.

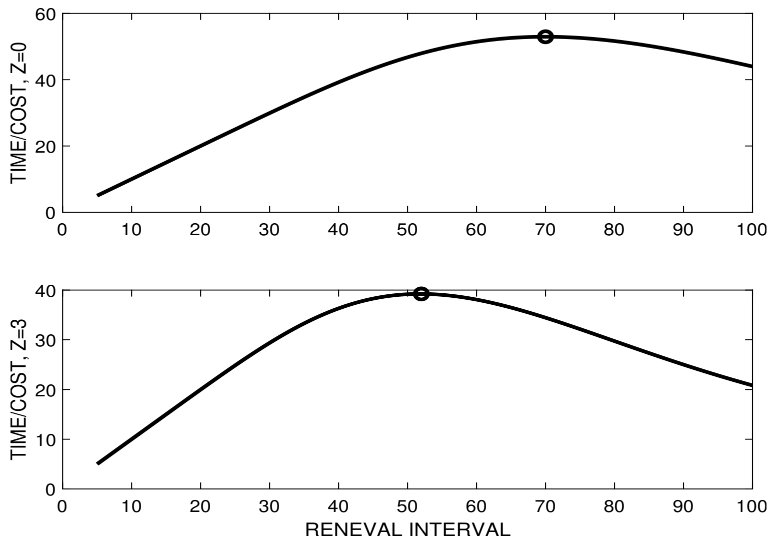

Figure 1: *Functions $\phi_N(\tau)$ and their maximal values at optimal τ^*, for $z = 0$ (above) and $z = 3$ (below).*

For the sake of 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. Further, let us consider the proportion of time of component availability to the unit of cost as a random criterion function,

$$\varphi(T,\tau) = \frac{T}{C_1} \quad \text{if} \quad T \le \tau, \quad \varphi(T,\tau) = \frac{\tau}{C_2} \quad \text{if} \quad T > \tau. \tag{9}$$

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

$$\phi_F(\tau) = E_F\varphi(T,\tau) = \int_0^\tau \frac{t}{C_1}\mathrm{d}F(t) + \frac{\tau}{C_2}\overline{F}(\tau). \tag{10}$$

Note that this formulation corresponds to scheme (1) with control variable τ, our goal here is to reach the maximum of (10).

In this example, the optimal solution τ^* can be found by solving the equation $\mathrm{d}\phi_F(\tau)/\mathrm{d}\tau = 0$. Namely, we solve

$$\frac{\mathrm{d}\phi_F(\tau)}{\mathrm{d}\tau} = \frac{\tau}{C_1} f(\tau) + \frac{1}{C_2}\Big(\overline{F}(\tau) - \tau f(\tau)\Big) = 0. \tag{11}$$

After some re-computation, we obtain the optimal τ^*, which fulfills

$$\tau^* \ h(\tau^*) = \frac{C_1}{C_1 - C_2}, \tag{12}$$

where $h(t) = f(t)/\overline{F}(t)$ denotes the hazard rate of random variable T. We can see that the right-hand side in (12) does not depend on the covariate. Let us also recall that, in the case of Cox regression model the hazard rate equals $h(t;z) = h(t;0) \cdot \exp(\beta\, z)$, while in the AFT model it equals $h(t;z) = h(t \cdot \exp(\beta\, z); 0) \cdot \exp(\beta\, z)$.

4.1. Example specification

The lifetime distribution will be specified below and we will compare the deterministic solution, in which all of the parameters are known, with the variability of 'sub-solutions', for which the

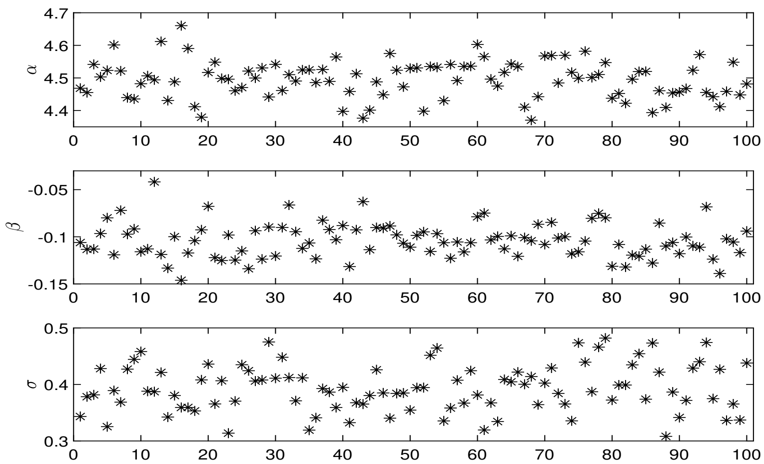

Figure 2: *Estimates of parameters from 100 data sets of size $N = 50$. Above α, then β, below σ.*

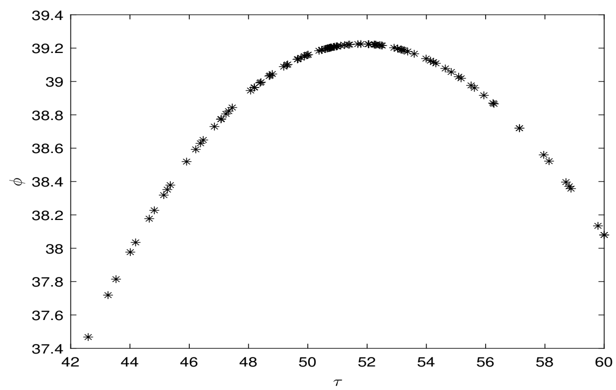

Figure 3: *Values of τ and ϕ achieved for parameters estimated from data of size $N = 50$, repeated 100 times.*

parameters are estimated. Namely, let us consider a case outlined in Example 2 of Section 3: Let the distribution of T be lognormal, fulfilling the relation

$$\ln T = \alpha + \beta \cdot z + \varepsilon, \ \ \varepsilon \sim N(0, \sigma^2), \tag{13}$$

with parameters $\alpha = 4.5$, $\beta = -0.1$, $\sigma = 0.4$. Further, let a 1-dimensional covariate z have its values distributed uniformly in $(-5, 5)$. For $z = 0$ we have the "baseline" lognormal distribution with parameters α, σ; the approximate value of its expectation $\mathrm{E}T_0 = 97.51$, the standard deviation $\mathrm{std}T_0 = 40.62$,. For instance, when $z = 3$, these characteristics are $\mathrm{E}T_3 = 72.24$, $\mathrm{std}T_3 = 30.09$.

Further, let the costs be $C_1 = 10$ and $C_2 = 1$. When all parameters are known, optimal values of the inter-maintenance time τ and the achieved maximum of $\phi(\tau)$ are $\tau_0^* = 69.95$, $\phi_0^* = 52.94$ and $\tau_3^* = 51.82$, $\phi_3^* = 39.22$ for $z = 0$, $z = 3$, respectively. These solutions have been obtained from (10) and (11) and are shown in Figure 1.

The data of a size N=50 are then generated from the correct model, i.e., first the values of covariate z_i, $i = 1, ..., N$, are generated uniformly in $(-5, 5)$ and then one value of T_i is generated from (13) for each z_i. The parameters are now estimated from this data. Finally, we select one value of the covariate (namely, we put $z = 3$ to have a comparison with the corresponding correct τ_3^* and ϕ_3^* from above) and compute the optimal values of τ_3 according to (11), taking

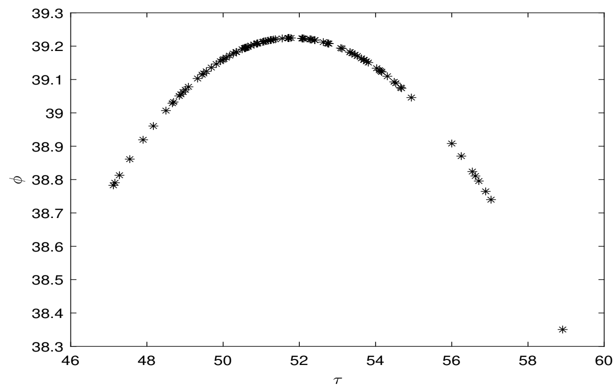

Figure 4: *Values of τ and ϕ achieved for parameters estimated from data of size $N = 200$, repeated 100 times.*

the estimated parameters instead of the 'true' ones. These sub-optimal τ values are inserted into (10) together with the correct F, i.e. given by the correct values of the parameters. The result is, naturally, smaller than the maximal possible value of ϕ^*. This procedure has been repeated 100 times in order to obtain a representation of the parameter estimates (this is shown in Figure 2) and the corresponding sub-optimal solutions, shown in Figure 3. It is seen how these values copy the curve from Figure 1 (for $z = 3$) and are more or less close to the optimal point.

The same procedure has been then performed for a larger data size, N=200. Table 1 lists the observed characteristics of the estimated parameters (from their 100 repetitions), for all of the instances reported here. We can see that the uncertainty of estimates (characterized by their standard deviations) for the case N=200 (second row of Table 1) is smaller than for N=50, i.e. the estimated values are closer to the correct values and, consequently, the sub-optimal solutions based on these estimates are closer to the correct solution. This observation follows from Figure 4 and its comparison with Figure 3.

Case	N	$\hat{\alpha}$: mean	std	$\hat{\beta}$: mean	std	$\hat{\sigma}$: mean	std
noncens.	50	4.5015	0.0539	-0.1038	0.0219	0.4019	0.0465
noncens.	200	4.5001	0.0264	-0.1002	0.0108	0.4014	0.0205
fixed z=3	50	$\hat{\mu}$=4.2051	0.0565	NA	NA	0.4089	0.0395
censored	200	4.4969	0.0383	-0.0991	0.0126	0.3953	0.0242

Table 1: *Sample means and standard deviations from 100 times repeated estimation of parameters.*

4.2. Comparison with the case without covariates

To show an increase in the uncertainty of the solution caused by the presence of an unknown regression parameter, let us compare the instances studied above with a case in which all the data are related to just one covariate value, for instance, again to $z = 3$. Namely, let us have $N = 50$ realizations of times T_i from a simple model

$$Y_i = ln\, T_i = \mu + \varepsilon_i, \tag{14}$$

where again ε_i are the i.i.d. random variables from $N(0, \sigma^2)$ distribution, $\mu = \alpha + \beta \cdot 3$ and the values of the parameters are the same as above. Therefore, μ is estimated as the mean and σ as the standard deviation from the values of Y_i (while α, β are not identifiable). The 'true' value is $\mu_0 = 4.2$, and Figure 5 again shows 100 sub-optimal solutions obtained on the basis of 100 repetitions of randomly generating such data, with a data size $N = 50$. Notice that the variability of the optimal results is really smaller than in the preceding case, in which the values of the covariate varied and, therefore, additional estimation of the regression parameter β was necessary.

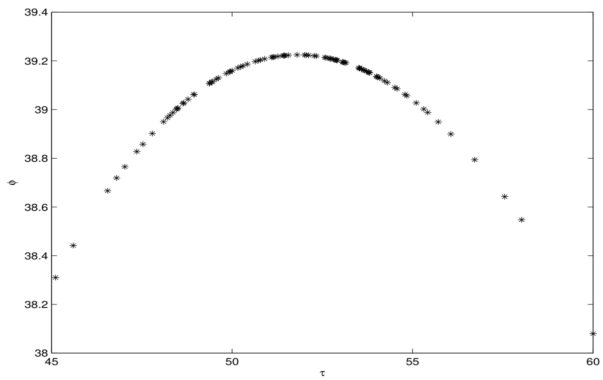

Figure 5: *Values of τ and ϕ achieved for parameters estimated from data of size $N = 50$, with fixed $z = 3$, repeated 100 times.*

5. Linear model with censoring

The phenomenon of censored data, a special case of incomplete data observation, is encountered quite often in lifetime data studies. Let us again consider model (4),

$$Y_i = \alpha + \beta \cdot z_i + \varepsilon_i, \quad i = 1, ..., N.$$

The random right-censoring means that there exists a random variable C, say, independent of Y and not providing us any information on the model parameters (the so-called 'uninformative censoring'). Then the variable $Y^* = \min(Y, C)$ is observed instead of Y. As a rule, it is known whether $Y \leq C$ or $Y > C$; this is denoted by an indicator $\delta_i = 1$ if $Y \leq C$, $\delta_i = 0$ if $Y > C$.

One of the first methods of analyzing the linear regression model with right-censoring has been proposed in Buckley and James [2]. However, a practical application of their iterative procedure is, in fact, not much effective; the slope is overestimated as a rule. There are some other, more direct options. Let us recall that the log-likelihood for right-censored data has the following form:

$$L_N(\theta) = \sum_{i=1}^{N} [\delta_i \cdot \ln(f(Y_i^*, z, \theta)) + (1 - \delta_i) \cdot \ln(1 - F(Y_i^*, z, \theta))]. \tag{15}$$

The MLE, via solving equation $dL_N(\theta)/d\theta = 0$, can easily be found (e.g. by the Newton-Raphson algorithm), for instance, for the Weibull distribution, while the case of the normal distribution has to be solved numerically, by a convenient search procedure. Such a method has also been used in the present paper. Hence, the solution is then just an approximation of the exact solution, however with an arbitrarily high precision level. Therefore this approach yields estimates which are still consistent in probability.

5.1. Example – continuation

Let us consider the same specification as above, in (13) in Subsection 4.1. Moreover, let the output values $Y = \ln T$ be censored from the right-hand side by realizations of random variable

$C \sim N(\mu_C = \alpha, \sigma_C = 3 \cdot \sigma)$. The rate of the censoring is then about 50%. $N = 200$ values have been generated, and again, the experiment has been repeated 100 times to obtain 100 estimates and, from them, 100 sub-optimal solutions to the optimal maintenance period problem. The results are shown in Figure 6, the characteristics of the estimated parameters are again listed in Table 1.

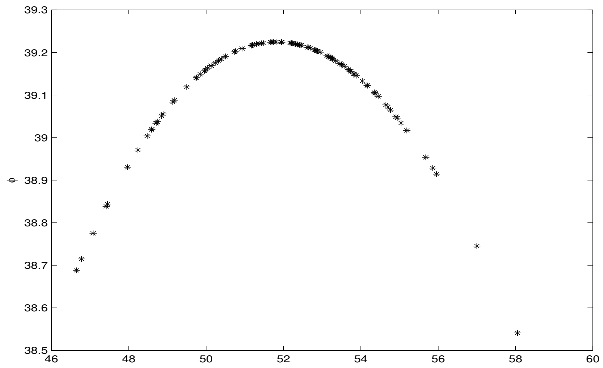

Figure 6: *Values of τ and ϕ achieved for parameters estimated from censored data of size $N = 200$, again repeated 100 times.*

As expected, the variation of the estimates as well as of the sub-optimal solutions is higher than those achieved in the case without censoring and the same N.

6. Concluding remarks

We have studied the impact of the variability of statistical estimates on the precision of solutions in a stochastic optimization problem. First, in the setting of parametric models with consistent estimates of parameters, the consistency of corresponding solutions of optimization has been proved. Further, the influence of empirical estimates on optimal solutions has been studied with the aid of randomly generated examples. This study has been performed in a situation when the random element of the optimization problem depends on covariates, via a parametric regression model. Finally, the analysis has been extended to the case of randomly right-censored data. Another extension to semi-parametric regression models, where the baseline distribution is not specified, should be a further step. Such a generalization may concern both the Cox and AFT models mentioned in the present paper.

As regards the example presented in Section 4, dealing with the optimal maintenance problem, another interesting point is worth exploring, namely, the proportion of failures allowed by optimal renewal interval. It theoretically equals $P(T_z \leq \tau_z^*) = F_z(\tau_z^*)$. It can be shown, and it also follows from (12), that in the AFT model this proportion does not depend on z because the accelerating term $\exp(\beta \cdot z)$ just changes the scale of time. An equality $\tau_z^* = \tau_0^* \cdot \exp(\beta\, z)$ also holds true. Naturally, this proportion in general depends on covariates as well. It simultaneously depends on costs C_1, C_2 because the optimal value of τ^* also depends on them.

Under the specification given in (13), the proportion of admissible failures equals $P(T_z \leq \tau_z^*) = P(Y_z < ln\, \tau_z^*)$, where Y_z has normal distribution $N(\mu = \alpha + \beta \cdot z, \sigma^2)$. For instance, this proportion equals 0.264 for costs $C_1 = 10$, $C_2 = 1$; however, when the costs are $C_1 = 5$, $C_2 = 2$, then for instance the optimal value is $\tau_0^* = 82.62$, $\phi_0^* = 29.42$ for $z = 0$, and the proportion $P(T_z \leq \tau_z^*)$ changes to 0.415.

References

[1] Bagdonavicius, V. and Nikulin, M. (2002). Accelerated Life Models. Modeling and Statistical Analysis. New York: Chapman&Hall/CRC.

[2] Buckley, J. and James, I. (1979). Linear regression with censored data. Biometrika, 66, 429-436. doi:10.1093/biomet/66.3.429

[3] Cox, D.R. and Hinkley, D.V. (1974). Theoretical Statistics. London: Chapman & Hall.

[4] Dupacová, J. and Wets, R. (1988). Asymptotic behaviour of statistical estimates and optimal solutions of stochastic optimization problems. Ann. Statist., 16, 1517-1549. doi:10.1214/aos/1176351052

[5] Gut, A. (2005). Probability: A graduate course. New York: Springer.

[6] Kalbfleisch, J.D. and Prentice, R.L. (2002). The Statistical Analysis of Failure Time Data (2nd edition). Hoboken, NJ: Wiley Series in Probability and Statistics.

[7] Kaňková, V. (2009). A remark of empirical estimates via economic problems. In Brožová, H. (Ed.), Proceedings of the Mathematic Methods in Economics 2009 (pp. 169-173). Prague: Czech Univ. of Life Sciences. doi: 10.1007/978-3-642-17022-5_49

[8] Lehmann, E.L. and Romano, J.P. (2005). Testing Statistical Hypotheses (3rd edition). New York: Springer.

[9] Volf, P. (2014). On precision of stochastic optimization based on estimates from censored data. Kybernetika, 50, 297-309. doi:/10.14736/kyb-2014-3-0297

[10] Volf, P. (2018). On Quantile Optimization Problem with Censored Data. Kybernetika, 54, 1156-1166. doi: 10.14736/kyb-2018-6-1156