

Simulation and optimization of complex systems reliability characteristics in grouped data structure

Ye. B. Tsoi, S. V. Tishkovskaya

Department of Applied Mathematics,

Novosibirsk State Technical University

20, K.Marx Prospekt, Novosibirsk, 630092, Russia.

Tel: (+7-3832) 46-03-01,46-05-11. Fax: +7-3832-46-02-09.

e-mail: ebcoi@nstu.nsk.su

Abstract

The present paper considers the problem of obtaining Bayesian estimates of the entity reliability characteristics. Bayesian estimates are calculated in the grouped data structure. Some properties and certain aspects of construction of Bayesian estimates in grouped data structure are studied.

Keywords

Complex systems, reliability characteristics, Bayesian estimation, grouped data.

1 INTRODUCTION

While studying or designing complex entities, there appear numerous problems requiring research of quantitative and qualitative laws of their functioning. To solve these problems, the entity considered is interpreted as a complex system, and mathematical simulation apparatus can be applied to its study. In constructing the mathematical model of the system studied, its major properties and laws of functioning are described by certain quantitative characteristics. Applying the mathematical simulation and optimization methods, these characteristics can be studied and, probably, improved.

The paper considers one of the characteristics of complex systems, its reliability. Reliability is the most important property of system's functioning quality, and the problem of its estimation reduces, in its turn, to the estimation problem of certain reliability quantitative properties such as average system lifetime, failure rate, etc. These properties have a probability character and are based on the assumption that system's lifetime duration is a random value. Note that everything concerning the system reliability and discussed above may be extended to its separate components.

2 PROBLEM STATEMENT

Let us denote by x the random value equal to the entity lifetime period, $F(x)$ is its probability distribution function. The reliability function $\bar{F}(x) = 1 - F(x)$ is considered as the major property of the entity reliability. In practice, function $F(x)$ and, consequently, $\bar{F}(x)$ is known as a rule to an accuracy of unknown parameter θ , i.e. $F(x) = F(x; \theta)$. (Values x and θ can be multivariate ones.) In this case the researcher faces a problem of unknown parameter estimation. The paper suggests to apply the Bayesian approach for obtaining estimates $\hat{\theta}$ of unknown parameters θ . Prior to considering directly the issues of Bayesian estimation, let us describe the data structure used for constructing estimates.

3 DATA STRUCTURE USED

According to the definition, the point estimate $\hat{\theta}$ is statistics, i.e. some measurable function of the sample used instead of the unknown parameter θ . Traditionally in statistics, almost everywhere it is assumed that the random sample, according to which the estimate $\hat{\theta}$ is constructed, consists of individually known observations. But in practice of statistical and experimental computations such a 'pure' data structure does not exist, as a rule. That is why development of methods for parameter estimation in the grouped data structure is of great practical value (Denisov, Lemesenko and Tsoi, 1993). Let us introduce the concept of grouped sample in the following way. Partition a set of X possible values of variable x into $k \geq 2$ non-intersecting intervals $R_i = (x_{(i-1)}, x_{(i)}]$, $i = 1, \dots, k$, $x_{(0)} = \inf(X)$, $x_{(k)} = \sup(X)$, $x_{(0)} < x_{(1)} < \dots < x_{(k-1)} < x_{(k)}$. Then the grouped data are a population composed of non-random partition for random value x changes' domain into k intervals R_i and discrete random values n_i , $i = 1, \dots, k$, which represent a number of observations belonging to the i th grouping interval. Thus we obtain a grouped sample $\bar{n} = (n_1, \dots, n_k)$.

A grouped sample is a special case of more general concept: partially grouped sample. Let x_{i1}, \dots, x_{in_i} are individual values of observations, belonging to the i th grouping interval. According to Kulldorf (Kulldorf, 1961), let us introduce the definition of partially grouped sample.

Definition 1. The sample is called a partially grouped one if the information available is connected with the multitude of non-intersecting intervals which divide the range of random values such that each interval belongs to one of the two types:

R^* , the i th interval belongs to the first type if the number n_i is known, but the individual values x_{ij} , $j = 1, \dots, n_i$ are unknown;

R^{**} , the i th interval belongs to the second type if not only the number n_i , but also all the individual values x_{ij} are known.

The concept of partially grouped sample combines non-grouped, grouped, and censored samples.

Let us note that some methods of estimation work only with data structures of a definite type. The Bayesian approach suggested in the paper do not impose any limitations on the data structure used. Thus, the Bayesian estimates can be constructed according to non-grouped, grouped, and partially grouped samples.

4 BAYESIAN METHOD OF PARAMETER ESTIMATION

Let us describe the scheme of constructing a Bayesian estimate for unknown parameter $\theta \in \Theta$. Let $f(x; \theta)$ be the density function corresponding to the distribution function $F(x; \theta)$. Let us consider the general case and assume that data \mathbf{D} represent a partially grouped sample, i.e., the domain \mathbf{X} is partitioned into $k \geq 2$ non-intersecting intervals, and part of these intervals belongs to the type \mathbf{R}^* , and the remaining part to the type \mathbf{R}^{**} . Then the likelihood function will have the form

$$l(\theta|\mathbf{D}) = \gamma \prod_{(1)} \left[\int_{x_{(j-1)}}^{x_{(j)}} f(x; \theta) dx \right]^{n_i} \prod_{(2)} \prod_{j=1}^{n_i} f(x_{ij}; \theta)$$

where γ is a certain constant, indexes (1) and (2) indicate that multiplication is done correspondingly according to intervals of the first and second type, x_{ij} are individual values of observations belonging to the i th interval. Let $\pi(\theta)$ be an a priori density function θ . The a posteriori density function will have the following form

$$\pi(\theta|\mathbf{D}) = \frac{\pi(\theta)l(\theta|\mathbf{D})}{\int_{\Theta} \pi(\theta)l(\theta|\mathbf{D})d\theta}$$

We take as the Bayesian estimation of parameter θ with respect to a priori distribution $\pi(\theta)$ the estimate minimizing the a posteriori risk $\int_{\Theta} L(\theta, t(\mathbf{D})) \pi(\theta|\mathbf{D})d\theta$ where $L(\theta, t(\mathbf{D}))$ is the given loss function, $t(\mathbf{D})$ is the estimation of parameter θ constructed according to the data \mathbf{D} (Zacks, 1971). Let us denote by $\hat{\theta}(\mathbf{D}) = \hat{\theta}$ the Bayesian estimate according to partially grouped sample. Then

$$\int_{\Theta} L(\theta, \hat{\theta}(\mathbf{D})) \pi(\theta|\mathbf{D})d\theta = \inf_{t \in T} \int_{\Theta} L(\theta, t(\mathbf{D})) \pi(\theta|\mathbf{D})d\theta$$

where T is set of all possible estimations of parameter θ .

5 CONVERGENCE OF BAYESIAN ESTIMATES IN THE STRUCTURE OF GROUPED DATA

As sample data \mathbf{D} , let us consider a certain fixed non-grouped sample $\bar{x} = (x_1, \dots, x_N)$ from distribution $f(x; \theta)$. Let us denote correspondingly by $\hat{\theta}^{ng}$ and $\pi^{ng}(\theta|\bar{x})$ the Bayesian estimate and a posteriori density function constructed from \bar{x} . Let $\bar{n} = (n_1, \dots, n_k)$ be a grouped sample corresponding to \bar{x} and connected with grouping intervals boundary points $x_{(0)} = \inf(\mathbf{X}) < x_1 < \dots < x_{(k-1)} < x_{(k)} = \sup(\mathbf{X})$. Let us denote correspondingly by $\hat{\theta}^{gr}$ and $\pi^{gr}(\theta|\bar{n})$ the Bayesian estimate and the a posteriori density function constructed according to \bar{n} .

Let us note that there exists the following connection between the grouped data structure and non-grouped one. If $\Delta x_{\max} = \max_{i=2, \dots, k-1} (x_{(i)} - x_{(i-1)}) \rightarrow 0$, $x_{(1)} \rightarrow x_{(0)}$, $x_{(k-1)} \rightarrow x_{(k)}$, then the number of grouping intervals $k \rightarrow \infty$ and grouped sample $\bar{n} = (n_1, \dots, n_k)$ degenerates into non-grouped one $\bar{x} = (x_1, \dots, x_N)$. In this connection a question arises: how do the function $\pi^{gr}(\theta|\bar{n})$ and estimate $\hat{\theta}^{gr}$ as $k \rightarrow \infty$ behave with respect to function $\pi^{ng}(\theta|\bar{x})$ and estimate $\hat{\theta}^{ng}$ respectively?

This paper shows that under performing a number of conditions functions $\pi^{gr}(\theta|\bar{n})$ converge uniformly on the set Θ to the function $\pi^{ng}(\theta|\bar{x})$. Before formulating these conditions let us introduce some designations and definitions.

We consider the set \mathbf{A} of all possible partitions of the domain \mathbf{X} into $k \geq 2$ non-intersecting intervals. Elements of set \mathbf{A} are sets of points

$$x_{(0)} = \inf(\mathbf{X}) < x_{(1)} < \dots < x_{(k-1)} < x_{(k)} = \sup(\mathbf{X}),$$

those define the partition of set \mathbf{X} into grouping intervals. We will have to designate these partitions by one way from following ones

$$\alpha = \alpha^k = \{x_{(0)}, \dots, x_{(k)}\} = \{x_{(0)}^{\alpha}, \dots, x_{(k)}^{\alpha}\}.$$

Let us name the quantity $d(\alpha) = \max_{i=2, \dots, k-1} (x_{(i)} - x_{(i-1)})$ as diameter of partition α . To every α

we set in conformity the density $\pi_{\alpha}^{gr}(\theta|\bar{n}) = \pi_{\alpha}^{gr}(\theta)$ depending on partition α . Thus we obtained a sequence of functions $\{\pi_{\alpha}^{gr}(\theta)\}_{\alpha \in \mathbf{A}}$ where indexes α are ordered as follows:

$$\alpha_2 > \alpha_1 \Leftrightarrow d(\alpha_2) < d(\alpha_1) \wedge x_{(1)}^{\alpha_2} < x_{(1)}^{\alpha_1} \wedge x_{(k-1)}^{\alpha_2} > x_{(k-1)}^{\alpha_1}.$$

Then the definition of uniform convergence $\pi_{\alpha}^{\text{gr}}(\theta)$ to $\pi^{\text{ng}}(\theta|\bar{x}) = \pi^{\text{ng}}(\theta)$ can be formulated as follows.

Definition 2. The sequence of functions $\{\pi_{\alpha}^{\text{gr}}(\theta)\}_{\alpha \in \mathbf{A}}$ converges uniformly with respect to θ on the set Θ towards the function $\pi^{\text{ng}}(\theta)$ as $d(\alpha) \rightarrow 0, x_{(1)} \rightarrow x_{(0)}, x_{(k-1)} \rightarrow x_{(k)}$ simultaneously if

$$\forall \varepsilon > 0 \quad \exists \delta_{\varepsilon} > 0 \quad \exists r_{\varepsilon} > x_{(0)}, s_{\varepsilon} < x_{(k)} \quad \forall \alpha \in \mathbf{A}: \quad d(\alpha) < \delta_{\varepsilon}, x_{(1)}^{\alpha} < r_{\varepsilon}, x_{(k-1)}^{\alpha} > s_{\varepsilon} \quad \forall \theta \in \Theta$$

$$|\pi_{\alpha}^{\text{gr}}(\theta) - \pi^{\text{ng}}(\theta)| < \varepsilon.$$

In practice the problem with $\mathbf{X} = (-\infty; +\infty)$ is often substituted by the problem in which $\inf(\mathbf{X}) < \infty$ and $\sup(\mathbf{X}) < \infty$. For this particular case the diameter of partition $d(\alpha)$ can be defined as $d(\alpha) = \max_{i=1, \dots, k} (x_{(i)} - x_{(i-1)})$. Indexes of a considered sequence of functions

$\{\pi_{\alpha}^{\text{gr}}(\theta)\}_{\alpha \in \mathbf{A}}$ will be ordered in accordance with the following rule: $\alpha_2 > \alpha_1 \Leftrightarrow d(\alpha_2) < d(\alpha_1)$.

We will formulate the following theorem about uniform convergence of the a posteriori density functions.

Theorem 1. Let $\pi(\theta)$ be a priori density function of θ ; $f(x; \theta)$ be density function of distribution of $x \in \mathbf{X}$; θ be unknown parameter; $\bar{x} = (x_1, \dots, x_N)$ be a some fixed sample from distribution $f(x; \theta)$; $\pi_{\alpha}^{\text{gr}}(\theta)$ be a posteriori density function constructed according to the grouped sample connected with partition $\alpha \in \mathbf{A}$, where \mathbf{A} is set of all possible partitions of the domain \mathbf{X} into $k \geq 2$ non-intersecting intervals; $\pi^{\text{ng}}(\theta)$ be a posteriori density function constructed according to the non-grouped sample \bar{x} . Let the following conditions are performed:

- 1) $\exists \mathbf{M} < \infty \quad \forall \theta \in \Theta \quad \pi(\theta) < \mathbf{M}$;
- 2) $\exists \mathbf{K} < \infty \quad \forall x \in \mathbf{X} \quad \forall \theta \in \Theta \quad f(x; \theta) < \mathbf{K}$;
- 3) function $f(x; \theta)$ is continuous uniformly on the $\mathbf{X} \times \Theta$.

Then the sequence $\{\pi_{\alpha}^{\text{gr}}(\theta)\}_{\alpha \in \mathbf{A}}$ converges uniformly with respect to $\theta \in \Theta$ towards the function $\pi^{\text{ng}}(\theta)$ as $d(\alpha) \rightarrow 0$.

For the sequence of density functions $\{\pi_{\alpha}^{\text{gr}}(\theta)\}_{\alpha \in \mathbf{A}}$ there exists a sequence of independent random values $\{\theta_{\alpha}\}_{\alpha \in \mathbf{A}}$. Hence we can speak about a sequence of Bayesian estimations $\{\hat{\theta}_{\alpha}^{\text{gr}}\}_{\alpha \in \mathbf{A}}$ obtained as solution of the problem of minimization for corresponding a-posteriori risk.

Theorem 2. Let the conditions of Theorem 1 be performed. For squared-error loss function, if set θ is compact, the convergence of Bayesian estimates $\hat{\theta}_\alpha^{BR}$ in the grouped data structure towards Bayesian estimate $\hat{\theta}^{NB}$ in the non-grouped data structure takes place as $d(\alpha) \rightarrow 0$.

6 COMPUTATIONAL DIFFICULTIES OF BAYESIAN ESTIMATION ACCORDING TO GROUPED DATA

As was mentioned above, the ideas of Bayesian approach are equally acceptable both for grouped and non-grouped data. Nevertheless, the Bayesian approach is mostly applied for constructing estimates from non-grouped and censored samples. For computing parameter estimates from completely grouped data, the Bayesian method practically is not applied. It may be due to the difficulties of computational character, the researcher faces in constructing Bayesian estimates with grouped data structure.

It is rather rare that one can succeed in obtaining an analytical expression for the Bayesian estimate from grouped data. That is why in the process of computing the estimate there is constantly a necessity to carry out numerical integration, and the integrals can have large dimension. The other practical difficulty consists in the fact that in the work with the grouped data it is necessary to compute the probabilities of the random value x falling in the i th

grouping interval: $p_i(\theta) = \int_{R_i} f(x; \theta) dx$, $i = 1, \dots, k$. In the case when the dimension of the

domain X is larger than 1, in computing $p_i(\theta)$, there appear some technical difficulties owing to the dimensions of the integrals and to the fact that the region of grouping R_i in the general case can have an arbitrary form.

One more problem, specific for the Bayesian estimation from the grouped data, consists in violating the property of closure of the conjugate distribution families. As is known, the conjugate distribution families which are widely applied to the Bayesian estimation, are characterized by the closure property in relation to the process of experimental data choice. In other words, if the a priori distribution $\pi(\theta)$ belongs to some conjugate family, then, with any sample \bar{x} , the a posteriori density function $\pi(\theta|\bar{x})$ will also belong to this family. Due to the violation of the closure property, the analytical or numerical computation of the Bayesian estimates from grouped data becomes even more complicated. In this paper, the following scheme of the Bayesian inference from grouped data, which allows to get round the problem of violating the property described.

Let us apply the binomial formula to the likelihood function from grouped data

$l(\theta|\bar{n}) = \gamma \prod_{i=1}^k p_i^{n_i}(\theta) = \gamma \prod_{i=1}^k (F(x_{(i)}; \theta) - F(x_{(i-1)}; \theta))^{n_i}$. Then the a posteriori density function

from the grouped data can be written down in the following way

$$\pi^{gr}(\theta|\bar{n}) = \sum_{i_k=0}^{n_k} \dots \sum_{i_2=0}^{n_2} \mathbf{K}_{i_2 \dots i_k}(\bar{n}) \pi_{i_2 \dots i_k}(\theta|\bar{n}),$$

where

$$\mathbf{K}_{i_2 \dots i_k}(\bar{n}) = \frac{\mathbf{C}_{n_2}^{i_2} \dots \mathbf{C}_{n_k}^{i_k} (-1)^{i_2 + \dots + i_k} \int_{\Theta} \mathbf{G}_{i_2 \dots i_k}(\theta|\bar{n}) \pi(\theta) d(\theta)}{\sum_{i_k=0}^{n_k} \dots \sum_{i_2=0}^{n_2} \mathbf{C}_{n_2}^{i_2} \dots \mathbf{C}_{n_k}^{i_k} (-1)^{i_2 + \dots + i_k} \int_{\Theta} \mathbf{G}_{i_2 \dots i_k}(\theta|\bar{n}) \pi(\theta) d(\theta)},$$

$$\pi_{i_2 \dots i_k}(\theta|\bar{n}) = \frac{\mathbf{G}_{i_2 \dots i_k}(\theta|\bar{n}) \pi(\theta)}{\int_{\Theta} \mathbf{G}_{i_2 \dots i_k}(\theta|\bar{n}) \pi(\theta) d\theta},$$

$$\mathbf{G}_{i_2 \dots i_k}(\theta|\bar{n}) = F^{n_1+i_2}(x_{(1)}) F^{n_2-i_2+i_3}(x_{(2)}) \dots F^{n_{k-1}-i_{k-1}+i_k}(x_{(k-1)}), \quad \mathbf{C}_n^i = \frac{n!}{i!(n-i)!}.$$

With this approach we can construct for the density functions $\pi(\theta)$ and $\pi_{i_2 \dots i_k}(\theta|\bar{n})$ the conjugate density families corresponding to the grouped data structure. Note that in some cases the conjugate families for grouped and non-grouped data will coincide. The scheme suggested allows, in a number of cases, to obtain an analytical form of Bayesian estimates from grouped data and avoid numerical integration. In the next section, an example of constructing a Bayesian estimate according to the scheme described is given.

7 EXAMPLE

Let the random value x belongs to the exponential distribution with distribution function $F(x) = F(x; \theta) = 1 - e^{-\theta x}$ and density function $f(x; \theta) = \theta e^{-\theta x}$, $x \geq 0$, $\theta \geq 0$, θ is unknown parameter. We will construct the Bayesian estimate of parameter θ from grouped data. As a priori distribution we take the gamma distribution with parameters α , β and with density function $\pi(\theta) = \alpha^\beta \theta^{\beta-1} e^{-\alpha\theta} / \Gamma(\beta)$. Consider the grouped sample $\bar{n} = (n_1, \dots, n_k)$. Likelihood function has the following form

$$l^{gr}(\theta|\bar{n}) = \gamma \prod_{i=1}^k (F(x_{(i)}) - F(x_{(i-1)}))^{n_i} = \gamma \prod_{i=1}^k (e^{-\theta x_{(i-1)}} - e^{-\theta x_{(i)}})^{n_i}.$$

In this example it is more convenient to apply the scheme from the section 6 to the likelihood

function written as $l^{gr}(\theta|\bar{n}) = \gamma \prod_{i=1}^k (\bar{F}(x_{(i-l)}) - \bar{F}(x_{(i)}))^{n_i}$, where $\bar{F}(x) = 1 - F(x)$. Then we obtain

$$\pi^{gr}(\theta|\bar{n}) = \sum_{i_{k-1}=0}^{n_{k-1}} \dots \sum_{i_1=0}^{n_1} \mathbf{K}_{i_1 \dots i_{k-1}}(\bar{n}) \pi_{i_1 \dots i_{k-1}}(\theta|\bar{n}),$$

$$\mathbf{K}_{i_1 \dots i_{k-1}}(\bar{n}) = \frac{\mathbf{C}_{n_1}^{i_1} \dots \mathbf{C}_{n_{k-1}}^{i_{k-1}} (-1)^{i_1 + \dots + i_{k-1}} (\varphi_{i_1 \dots i_{k-1}}(\bar{n}))^{-\beta}}{\sum_{i_{k-1}=0}^{n_{k-1}} \dots \sum_{i_1=0}^{n_1} \mathbf{C}_{n_1}^{i_1} \dots \mathbf{C}_{n_{k-1}}^{i_{k-1}} (-1)^{i_1 + \dots + i_{k-1}} (\varphi_{i_1 \dots i_{k-1}}(\bar{n}))^{-\beta}},$$

$$\pi_{i_1 \dots i_{k-1}}(\theta|\bar{n}) = (\varphi_{i_1 \dots i_{k-1}}(\bar{n}))^\beta \theta^{\beta-1} \exp\{-\theta \varphi_{i_1 \dots i_{k-1}}(\bar{n})\} / \Gamma(\beta),$$

where $\varphi_{i_1 \dots i_{k-1}}(\bar{n}) = \sum_{l=1}^{k-2} x_{(l)}(i_l - i_{l+1} + n_{l+1}) + x_{(k-1)}(i_{k-1} + n_k) + \alpha$. We see that $\pi_{i_1 \dots i_{k-1}}(\theta|\bar{n})$

is density function of the gamma distribution with parameters $\varphi_{i_1 \dots i_{k-1}}(\bar{n})$, β . Thus we can write the analytical expression for the Bayesian estimate from grouped data with the squared-error loss function:

$$\hat{\theta}^{gr}(\bar{n}) = \sum_{i_{k-1}=0}^{n_{k-1}} \dots \sum_{i_1=0}^{n_1} \mathbf{K}_{i_1 \dots i_{k-1}}(\bar{n}) \frac{\beta}{\varphi_{i_1 \dots i_{k-1}}(\bar{n})}.$$

REFERENCES

Denisov, V.I., Lemesenko, B.Yu., and Tsoi, Ye.B.(1993) Optimal grouping, parameter estimation, and design of regression experiments, (in Russian). Novosibirsk.
 Kulldorf, G. (1961) Estimation from grouped data and partially grouped samples. John Wiley, New York.
 Zacks, S. (1971) The Theory of Statistical Inference. John Wiley and Sons Inc., New York.