Numerical Probabilistic Analysis for the Digital Economy

Boris Dobronets and Olga Popova

Siberian Federal University
E-mail: BDobronets@yandex.ru, OlgaArc@yandex.ru

Sino-Russian Global Engagement Models in the Context of Digitalization of Social And Economic Processes
2018

Introduction

⋆ The first problem is related to the big data processing and knowledge discovery in data base.

⋆ The second problem concerns the issue of the uncertainty level reducing and data variability study in the big data bases.

⋆ Our approach is based on Big Data technologies including the data aggregation procedures for the input and output parameters and Numerical Probabilistic Analysis (NPA)

Digital Economy

The digital economy is a very broad concept. Here are some definitions:

The digital economy is the worldwide network of economic activities, commercial transactions and professional interactions that are enabled by information and communications technologies (ICT).

A digital economy based on big data is predictive in its type: here the forecast, the plan and the fact tend to equality. Its basic tool is prognostic analytics, the main type of production is personalized to the needs of the client, and the competition goes not so much for the redistribution of existing markets, as for the formation of new ones, where no longer goods and technologies compete, but digital management systems based on digital platforms.

The Dominating Paradigms of Economic Theories

- Nowadays the dominating paradigms of economic theories are based on the classical mathematics and presented in terms of probabilistic and statistical methods.

- It should be emphasized that in applications, the probabilistic and statistical methods are often and successfully used in the synthesis with modern methods of soft computing.

- Now it is understood that in applications we often deal with aleatory and epistemic uncertainty.

Computational economics

Computational economics: a perspective from computational intelligence / Shu-Heng chen and Lakhmi Jain, editors. London. Idea Group Inc. 2006. p. 339.

Valeriu I.

Economic Intelligence // Journal of Knowledge Management, Economics and Information Technology. Special Issue December 2013 p. 182-198

Shu-Heng Chen

Computational intelligence in economics and finance: Carrying on the legacy of Herbert Simon // Information Sciences 170 (2005) 121–131

Big Data

Kim Hua Tan, Guojun Ji, Chee Peng Lim & Ming-Lang Tseng

Using big data to make better decisions in the digital economy // International Journal of Production Research, 2017. 55:17, 4998-5000, DOI: 10.1080/00207543.2017.1331051

Data Aggregation

Although there are many ways of data aggregation, including simple average, we argue that the use of piecewise linear and piecewise polynomial aggregation functions will offer a more informative representation of the variability in the Big Data, than other forms of Data Aggregation.

Developed methods reduce the level of uncertainty in the information flow; significantly reduce the processing time and the implementation of numerical procedures.

Uncertainty Modeling

⋆ There are many modern methods for uncertainty modeling developed in last decades. Generally, they are not in conflict with the traditional probabilistic approach since they deal with another (non-probabilistic) types of uncertainties.

⋆ The treatment of uncertainty in analysis, design, and decision making is going through a paradigm shift from a probabilistic framework to a generalized framework that includes both probabilistic and non probabilistic methods.

Numerical Probabilistic Analysis

The basis of NPA are numerical operations on probability density functions of the random values and probabilistic extensions.
The numerical operations of the probabilistic arithmetic constitute the major component of NPA.
The concepts of probabilistic extensions of a function are considered.

Gerasimov V.A., Dobronets B.S., Shustrov M.Y.

Numerical Operations for Histogram Arithmetic and Their Applications // Automation and Remote Control. 1991. № 2. C. 83-88.

Probabilistic Extensions

One of the most important problems that NPA deals with is to construct probability density functions of random variables. Let us start with the general case when $(x_1, \dots, x_n)$ is a system of continuous random variables with joint probability density function $p(x_1, \dots, x_n)$ and the random variable z is a function $f(x_1, \dots, x_n)$

$$z = f(x_1, \dots, x_n).$$

By probabilistic extension of the function f, we mean a probability density function of the random variable z.

Theorem 1

Let $\boldsymbol{f}(\cdot, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)$ be Probabilistic Extensions of function $f(x_1, x_2, \dots, x_n)$ and for all real $\boldsymbol{t}$ function $\boldsymbol{f}(\cdot, t, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)$ be probabilistic extensions of the function $f(t, x_2, \dots, x_n)$. Then

$$\boldsymbol{f}(z, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n) = \int \boldsymbol{x}_1(t)\boldsymbol{f}(z, t, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)dt \tag{1}$$

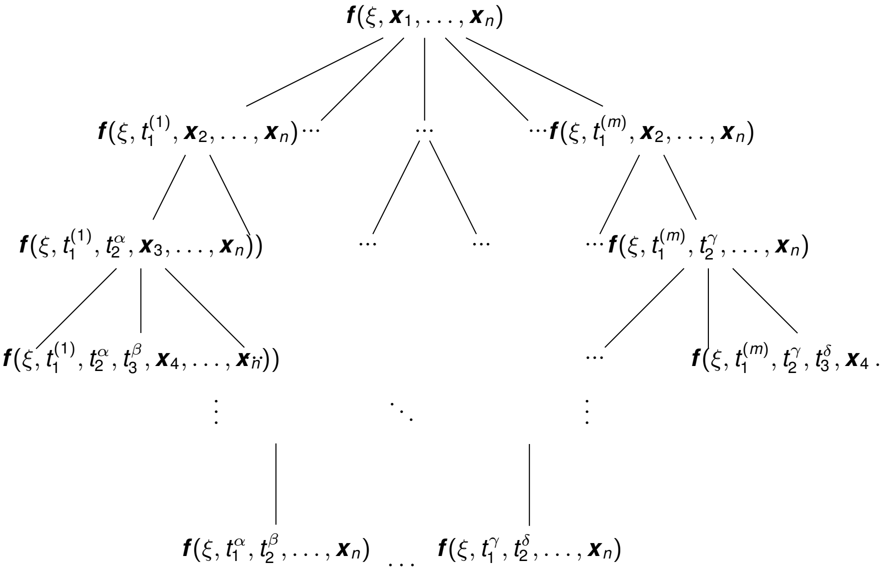
Corollary

Theorem 1 infers the possibility of recursive computations for the general form of probability extensions and reduction to the calculation of the one-dimensional case

Let us consider the computing of the integral (1). For simplicity, we represent (1) as a quadrature

$$\int \boldsymbol{x}_1(t)\boldsymbol{f}(z,t,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n)dt \approx \sum_{l=1}^{m} \gamma_l \boldsymbol{x}_1(t_l)\boldsymbol{f}(z,t_l,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n)$$

Further, for the computing $\boldsymbol{f}(z,t_l,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n)$ we can also use numerical quadratures and so on. In general, it is NP-hard problem with actual parallelization.



In Figure the tree of parallel-recursive organization of the computational

Parallel-Recursive Organization of the Computational Process

Thus, on the lower layer, it is necessary to calculate probabilistic extensions for functions only one variable. All calculations on each layer are independent and can be computed simultaneously.

Natural Probabilistic extension

Let $f(x_1, \ldots, x_n)$ be a rational function. To construct a Probabilistic Extensions of f, we replaced the arithmetic operation by the arithmetic operation, while the variables x_1, x_2, …, x_n are replaced by pdf of their possible values. It makes sense to call the resulting Probabilistic Extensions of f as natural Probabilistic extension.

Case 1.

Let $x_1, \ldots, x_n$ be independent random variables. If $f(x_1, \ldots, x_n)$ is a rational expression where each variable x_i occurs not more than once, then the natural Probabilistic extension approximates a probabilistic extension.

Case 2.

Let the function $f(x_1, \ldots, x_n)$ can be a change of variables, so that $f(z_1, \ldots, z_k)$ is a rational function of the variables $z_1, \ldots, z_k$ satisfying the conditions of Case 1. The variable z_i is a function of x_i, $i \in Ind_i$. and Ind_i be mutually disjoint. Suppose for each z_i is possible to construct a probabilistic extension. Then the natural extension $\boldsymbol{f}(z_1, \ldots, z_k)$ would be approximated by a probabilistic extension $\boldsymbol{f}(x_1, \ldots, x_n)$.

Example. Let $f(x_1, x_2) = (-x_1^2 + x_1) \sin x_2$, and $z_1 = (-x_1^2 + x_1)$, $z_2 = \sin x_2$. Notice that it is possible to construct probabilistic extensions for the functions z_1, z_2, and then compute $f = z_1 * z_2$, which is a rational function satisfying the conditions of Case 1. So, natural extension will approximate probabilistic extension for the function $f(x_1, x_2)$.

Risk assessment of Investment Projects

we consider risk assessment of investment projects, where pdf of factors such as Net Present Value (NPV) and Internal Rate of Return (IRR) are computed.
Net Present Value (NPV) is a formula used to determine the present value of an investment by the discounted sum of all cash flows received from the project. The formula for the discounted sum of all cash flows can be rewritten as

$$NPV(r) = Cz_1 s_1 \sum_{i=1}^{T} \frac{C_i}{(1+r)^i} - C_0, \tag{2}$$

where $-C_0$ is initial investment, C_i is cash flow, T time, r is the discount rate, s_1 is cost, z_1 is expenditures.
IRR determines the maximum acceptable discount rate in which you can invest without any loss to the owner: IRR $= r$, in which the

$$NPV(r) = 0. \tag{3}$$

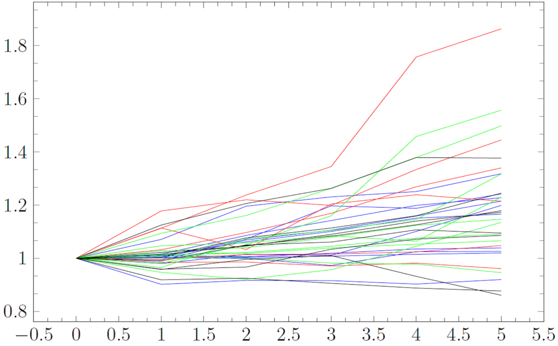


Рис.: The resulted volumes of sales similar firms

Using Big Date, we can build the resulted volumes of sales similar firms. Figure 2 shows the resulted volumes of sales similar firms. Further using the procedure aggregation, constructed cubic splines approximating the probability density of c_i, x_i, s_i, z_i.

NPV

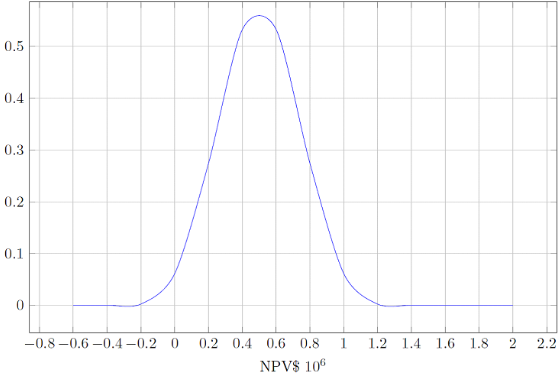


Рис.: The spline approximating the probability density of NPV

The figure 2 shows the splines approximating the probability density of NPV. Support of the NPV is [$ -0.704 106, $ 2.07 106].

IRR

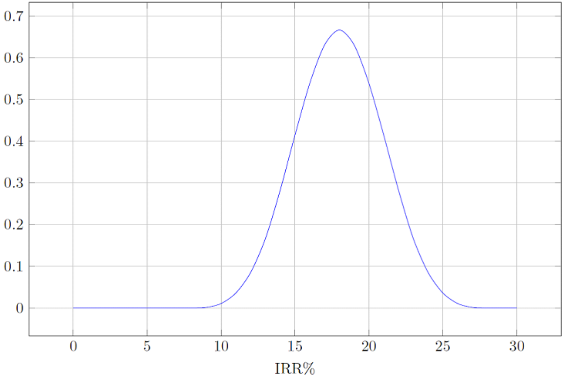


Рис.: The spline approximating the probability density of IRR

The figure 3 shows the splines approximating the probability density of IRR. Support of the IRR is [5%, 30%].

Systems of linear algebraic equations

As one example of numerical simulation, let us consider solution of a system of linear algebraic equations

$$Ax = b, \tag{4}$$

where $A = (a_{ij})$ a random matrix and $b = (\boldsymbol{b}_i)$ a random right-hand side vector. respectively. Suppose that the random matrix A and the vector b have independent components with probability densities $\boldsymbol{A} = (\boldsymbol{a}_{ij})$, $\boldsymbol{b} = (\boldsymbol{b}_i)$ respectively and

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_{11} & \boldsymbol{a}_{12} & \dots & \boldsymbol{a}_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_{n1} & \boldsymbol{a}_{n2} & \dots & \boldsymbol{a}_{nn} \end{pmatrix}.$$

The support of the solution set can be represented as follows

$$\mathcal{X} = \{x | Ax = b, A \in \mathrm{supp}(\boldsymbol{A}), b \in \mathrm{supp}(\boldsymbol{b})\}.$$

Construct the probabilistic extension of the solution vector $\boldsymbol{x}(\cdot, \boldsymbol{A}, \boldsymbol{b})$

$$\boldsymbol{x}_1(\cdot, \boldsymbol{A}, \boldsymbol{b}) = \frac{\begin{vmatrix} \boldsymbol{b}_1 & \boldsymbol{a}_{12} & \dots & \boldsymbol{a}_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{b}_n & \boldsymbol{a}_{n2} & \dots & \boldsymbol{a}_{nn} \end{vmatrix}}{\begin{vmatrix} \boldsymbol{a}_{11} & \boldsymbol{a}_{12} & \dots & \boldsymbol{a}_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_{n1} & \boldsymbol{a}_{n2} & \dots & \boldsymbol{a}_{nn} \end{vmatrix}}$$

or

$$\boldsymbol{x}_1(\xi, \boldsymbol{A}, \boldsymbol{b}) = \int \dots \int \boldsymbol{a}_{12}(t_{12}) \dots \boldsymbol{a}_{nn}(t_{nn}) \left(\frac{\sum \boldsymbol{b}_i \Delta_i(t_{12}, \dots, t_{nn})}{\sum \boldsymbol{a}_{1i} \Delta_i(t_{12}, \dots, t_{nn})} \right) (\xi) dt_{12} \dots d \tag{5}$$

where $\Delta_i(t_{12}, \dots, t_{nn}) \in R$ are minors from the Cramer method for solving SLAE, $t_{ij} \in \mathrm{supp}(\boldsymbol{a}_{ij})$. Expression

$$\left(\frac{\sum \boldsymbol{b}_i \Delta_i(t_{12}, \dots, t_{nn})}{\sum \boldsymbol{a}_{1i} \Delta_i(t_{12}, \dots, t_{nn})} \right) (\xi)$$

is computed using probabilistic arithmetic.

Example

$$Ax = b, \tag{6}$$

Let $A = (a_{ij})$ be random matrix $n = 2$. Elements of the matrix $\boldsymbol{A}$ are independent and distributed on the triangular law, $\boldsymbol{a}_{11}, \boldsymbol{a}_{22}$ distributed on the interval $[2, 4]$, a_{21}, a_{12} distributed on the interval $[-1, 1]$. The vector $\boldsymbol{b}$ consists of independent components $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, are distributed according to a triangular law on the interval $[0, 2]$.

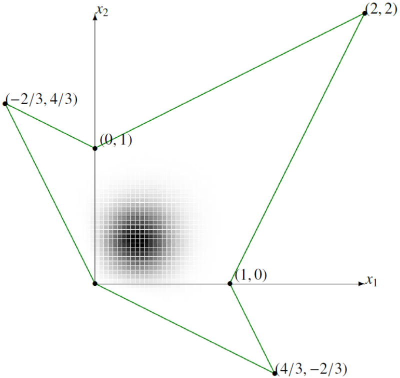


Рис.: The boundary of the solution set and the joint probability density function of (x_1, x_2)

In Figure the joint probability density of the vectors (x_1, x_2) are shown. The value of the probability is represented by shades of gray. Solid line is the boundary of the solution set.

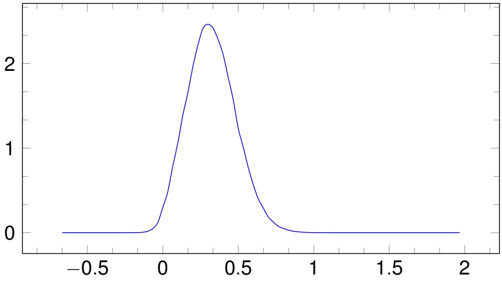


Рис.: Probability density function of random variable x_1

$$\boldsymbol{x}_1(\xi) = \int\int \boldsymbol{a}_{22}(t_{22})\boldsymbol{a}_{12}(t_{12}) \left(\frac{t_{22}\boldsymbol{b}_1 - t_{12}\boldsymbol{b}_2}{\boldsymbol{a}_{11}t_{22} - t_{12}\boldsymbol{a}_{21}} \right)(\xi) dt_{22} dt_{12}.$$

Random Boundary value Problem

$$Lu \equiv -\boldsymbol{p}u'' + \boldsymbol{q}u = \boldsymbol{r}f(x), x \in (0,1), \quad (7)$$

$$u(0) = 0, \ u(1) = 0.$$

$\boldsymbol{p} > 0$, $\boldsymbol{q} \leq 0$, $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}$ — independent random variables.
$\omega_h = \{x_i = ih, i = 1, 2, ..., N-1, h = 1/N\}$

$$L^h u_i^h = -p\frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} + qu_i = rf(x_i), i = 1, 2, ..., N-1.$$

$[\underline{p}, \overline{p}]$, $[\underline{q}, \overline{q}]$ $[\underline{r}, \overline{r}]$ — supports pdf.
$\omega_p = \{p_0 = \underline{p} < p_1 < \ldots < p_K = \overline{p}\}$,
$\omega_q = \{q_0 = \underline{q} < q_1 < \ldots < q_L = \overline{q}\}$
положим $r = 1$ и решим численно kl задач

$$-p_k \frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} + q_l u_i = f(x_i), i = 1, 2, ..., N-1.$$

$u_{ikl} = u_i(p_k, q_l)$

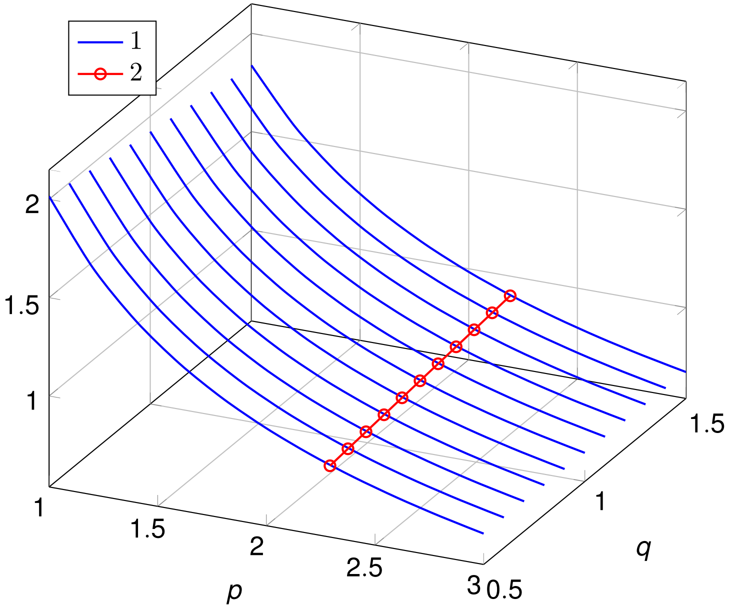


Рис. :

References

⋆ Dobronets B., Popova O. Numerical Probabilistic Approach for Optimization Problems // Scientific Computing, Computer Arithmetic, and Validated Numerics. Lecture Notes in Computer Science 9553. Springer International Publishing, Cham, 43–53 (2016)

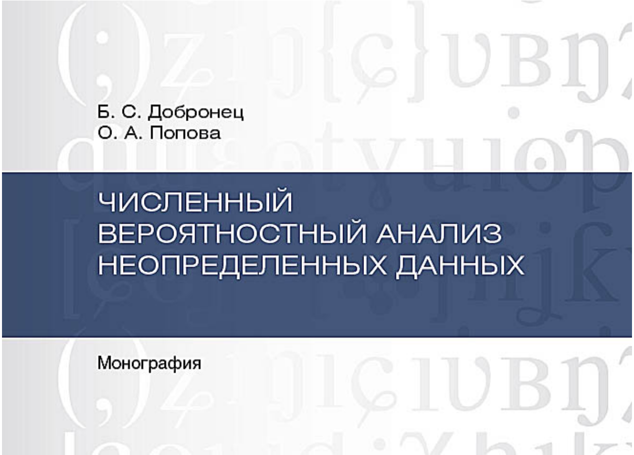
⋆ Dobronets B.S., Popova O.A. Numerical probabilistic analysis under aleatory and epistemic uncertainty // Reliable Computing. 19, (2014), 274–289.

⋆ Dobronets B.S., Popova O.A. Improving the accuracy of the probability density function estimation, Journal of Siberian Federal University, Mathematics and Physics, 10 (1), 16–21 (2017).

⋆ Dobronets B.S., Popova O.A. Piecewise Polynomial Aggregation as Preprocessing for Data Numerical Modeling // IOP Conf. Series: Journal of Physics: Conf. Series 1015 (2018) 032028 doi :10.1088/1742-6596/1015/3/032028

⋆ Dobronets B.S., Popova O.A. Improving reliability of aggregation, numerical simulation and analysis of complex systems by empirical data // IOP Conf. Series: Materials Science and Engineering 354 (2018) 012006 doi:10.1088/1757-899X/354/1/012006

Conclusion



- Добронец Б.С., Попова О.А. Численный вероятностный анализ неопределенных данных Сибирский федеральный университет, Институт космический и информационных технологий. Красноярск, 2014.