Numerical Probabilistic Analysis under Aleatory and Epistemic Uncertainty*

Boris S. Dobronets
Institute of Space and Information Technologies,
Siberian Federal University, Krasnoyarsk, Russia
BDobronets@sfu-kras.ru

Olga A. Popova
Institute of Space and Information Technologies,
Siberian Federal University, Krasnoyarsk, Russia
olgaarc@yandex.ru

Abstract

This paper discusses Numerical Probabilistic Analysis (NPA) for problems under aleatory and epistemic uncertainty. The basis of NPA are numerical operations on probability density functions of the random values and probabilistic extensions. The numerical operations of the histogram arithmetic constitute the major component of NPA. The concepts of natural, probabilistic and histogram extensions of a function are considered. Using NPA approach, we construct numerical methods that enable us to solve systems of linear and nonlinear algebraic equations with stochastic parameters. To facilitate a more detailed description of the epistemic uncertainty, we introduce the concept of second order histograms. Relying on specific practical examples, we show that using second order histograms may prove helpful in decision making. In particular, we consider risk assessment of investment projects, where histograms of factors such as Net Present Value (NPV) and Internal Rate of Return (IRR) are computed.

*Submitted: February 23, 2013; Revised: March 8, 2014; Accepted: April 15, 2014.

Keywords: Numerical probabilistic analysis, epistemic uncertainty, second order histogram, risk assessment.
AMS subject classifications: 65G40

1 Introduction

Many important practical problems involve different uncertainty types. In practice, several sources of uncertainty of the required information impede optimal decision making in the classical sense. When only uncertain information is available (which is most often the case), then decision making requires more complex methods for data representation and their analysis.

We consider Numerical Probabilistic Analysis (NPA) for problems under aleatory and epistemic uncertainty. The general approach is to choose the representation of the uncertainty according to the type and amount of information that can be made available in the process. Aleatory uncertainty characterizes inherent randomness in the behavior of the system under study. Alternative terms possibly used in connection with this kind of uncertainty are 'variability', 'stochastic uncertainty', 'irreducible uncertainty', and 'Type A uncertainty'. On the other hand, epistemic uncertainty characterizes a lack of knowledge about a considered value. Generally, epistemic uncertainty may be inadequate for the frequency interpretation typical for classical probability and hence, for uncertainty description in the traditional probability theory. Other terms for epistemic uncertainty are 'state of knowledge uncertainty', 'subjective uncertainty', and 'irreducible uncertainty' [13].

Once an uncertainty representation has been chosen, the problem is to carry out subsequent calculations in a way that produces correct results with no additional uncertainty being introduced into the answer. Clearly, processing aleatory and epistemic uncertainties may require the use of special methods

The Monte Carlo method [12] is a powerful approach for problems with uncertainties, but it has serious shortcomings. These are difficulties in handling uncertain quantities having unknown dependency relationships or those with imprecise probabilities, that is, with not fully specified probability distributions.

Non-Monte Carlo methods have been developed since the 1960's [2, 8, 15]. A major non-Monte Carlo approach is interval analysis. For the solution of our problems, interval analysis has developed such powerful tools as probabilistic discretizations of random variables [1, 7], p-boxes [6], or clouds [9]. An interesting comparison of interval analysis approaches vs. Monte Carlo techniques is given in [10] and repeated in the Conclusions of the book [11].

Several alternative approaches to the decision making and risk assessments under epistemic uncertainty have been considered in [14, 16, 17].

In our work, we develop a technique that uses Numerical Probabilistic Analysis to solve various problems with stochastic data uncertainty.

NPA is based on numerical operations on probability density functions of the random values approximated by histograms. These are operations "+", "−", "·", "/", "↑", "max", "min", binary relations "≤", "≥", and some others. The numerical operations of histogram arithmetic constitute the major component of NPA. One of the first implementations of the numerical histogram arithmetic is in [2].

Using the arithmetic of probability density functions and probabilistic extensions of usual functions, we can construct numerical methods that enable us to solve systems of linear and nonlinear algebraic equations with stochastic parameters [3]. To facilitate more a detailed description of the epistemic uncertainty, we introduce the concept of second order histograms, which are defined as piecewise histogram functions [4]. Second order histograms can be constructed using experience and/or intuition of experts.

Relying on specific practical examples, we show that using second order histograms may prove very helpful in decision making. In addition, we present methods that are able to cope with uncertainties in strategic decision situations, in particular, involving calculations of the economic uncertainties. We consider risk assessment of investment projects, where histograms of factors such as Net Present Value (NPV) and Internal Rate of Return (IRR) are computed.

2 Types of Probability Density Functions

The basis of NPA are numerical operations on probability density functions of random values and very important operations of probabilistic extension of functions. Before elaborating the details, we describe various types of probability density functions, e.g., a discrete function, a histogram (piecewise constant function), or a piecewise-polynomial function.

Discrete random variables. A discrete random variable ξ may take only a countable number of distinct values x_1, x_2, …, x_n (potentially, an infinite number of them). The probability distribution $p(x)$ of a discrete random variable is a list of probabilities associated with each of its possible values, also called the probability function or the probability mass function.

Histograms. A random variable whose probability density function is represented by a piecewise constant function called a *histogram*. Any histogram P is defined by a grid $\{x_i \mid i = 0, 1, \ldots, n\}$ and a set of values p_i, $i = 1, \ldots, n$, such that the histogram takes the constant value p_i at the interval $[x_{i-1}, x_i]$.

Interval histograms. Random variable is called an *interval histogram* if its probability density function $P(x)$ is a piecewise interval constant function.

Second order histograms. When epistemic uncertainty is present, *second order histograms* may be used along with interval histograms. A probability density function $P(x)$ of a second order histogram is a piecewise-histogram function, i. e., a histogram whose each column is a histogram [5].

Piecewise linear functions. A piecewise linear function is a function composed of straight-line sections. These are the simplest splines. Although such functions are relatively simple, they have good approximating properties, and they can be tools for approximating probability density functions.

Splines. A spline is a piecewise degree n polynomial function with $n-1$ continuous derivatives at its nodes (points where the polynomial pieces connect). We approximate the probability density of the random variables by spline functions.

Analytically defined probability density functions. Random variables and their density distributions can be expressed analytically by explicit formulas.

3 Operations on Probability Densities of Random Variables

In this section, we consider operations on different kinds of probability density functions.

Operations on discrete values. Let $* \in \{+, -, \cdot, /, \uparrow\}$ be an operation between two independent discrete random variables ξ and η. If ξ takes the values x_i with the probability p_i, and η takes the values y_i with the probability q_i, then the result $\xi * \eta$ of any operation between ξ and η is a random variable ψ that takes its possible values $x_i * y_j$ with the probability $p_i q_j$.

Operations on histograms. Let $p(x, y)$ be a joint probability density function of two random variables x and y. Also, let p_z be a histogram approximating the probability density of the operations between two random variables $x * y$, where $* \in \{+, -, \cdot, /, \uparrow\}$. Then the probability to find the value z within the interval $[z_i, z_{i+1}]$ is determined by the formula (see [3])

$$P(z_k \le z \le z_{k+1}) = \int_{\Omega_k} p(x, y)\, dx\, dy\,, \tag{1}$$

where $\Omega_k = \{(x, y) \mid z_k \le x * y \le z_{k+1}\}$.

We outline a numerical implementation of the general approach in the previous paragraph. Let the histogram variables x and y be given on the grids $\{a_i\}$ and $\{b_i\}$, respectively, with the corresponding probabilities $\{p_i\}$ and $\{q_i\}$. Let $[a_0, a_n]$ and $[b_0, b_n]$ be supports of the probability densities of these variables, and let the rectangle $[a_0, a_n] \times [b_0, b_n]$ be the support of the joint probability density $p(x_1, x_2)$. We divide the rectangle $[a_0, a_n] \times [b_0, b_n]$ into n^2 sub-rectangles $[a_i, a_{i+1}] \times [b_j, b_{j+1}]$, in which the probability of getting a constant is equal to $p_i q_j$ for independent random variables and p_{ij} for dependent ones.

To compute the resulting histogram, it is necessary to pass through all the sub-rectangles $[a_i, a_{i+1}] \times [b_j, b_{j+1}]$ calculating for each of them its contribution to every segment $[z_k, z_{k+1}]$ of the resulting histogram. To this end, we define the region Ω'_k, for which the sub-rectangle $[a_i, a_{i+1}] \times [b_j, b_{j+1}]$ intersects Ω_k:

$$\Omega'_k = \Omega_k \cap \left([a_i, a_{i+1}] \times [b_j, b_{j+1}]\right).$$

Then we compute the integral over Ω'_k,

$$p_{zk} = \int_{\Omega'_k} p(x, y)\, dx\, dy\,. \tag{2}$$

Note that, for each $[a_i, a_{i+1}] \times [b_j, b_{j+1}]$, the joint probability density $p(x, y)$ is constant, and so is the ratio of the integration area Ω'_k to the area of the sub-rectangle $[a_i, a_{i+1}] \times [b_j, b_{j+1}]$.

Walking through all the rectangles, we compute the desired histogram p_z. Overall, the number of arithmetic operations required for the construction of the histogram is $O(n^2)$.

Operations between histograms and discrete random variables. We consider the operations of the form $x * c$, where $* \in \{+, -, \cdot, /\}$, c is a constant, and x is a random variable with the probability density f_x. If $* \in \{+, -\}$, the probability density function f_z of the random variable $z = x * c$ can be easily expressed as $f_z(\xi * c) = f_x(\xi)$, $\xi \in \mathbb{R}$. Let "$*$" be multiplication and $c \neq 0$. Then $f_z(\xi) = f_x(\xi/c)/c$. If $c = 0$, the random variable z takes only one value 0 with probability 1. The operation of division by $c \neq 0$ is treated analogously, which results in $f_z(\xi) = f_x(\xi \cdot c) \cdot c$, $\xi \in \mathbb{R}$.

In the case of operations between discrete and histogram random variables, we have one-dimensional segments $A_i \times [b_j, b_{j+1}]$ instead of rectangles $[a_i, a_{i+1}] \times [b_j, b_{j+1}]$. Similar to the previous case, going through all these segments, we calculate the contribution of each one to the final histogram. As compared to the previous case, the difference in the numerical implementation is that we do not have to divide the result by the measure of the region under study (or the length of the subinterval). However, if the number of values that the discrete random variable can take is large, the calculations can be very laborious. In this situation, it makes sense to represent the discrete random variable in the form of a histogram, and then operate as described in the preceding case.

Operations between histograms and an analytically determined density function. In this case, the computation algorithm is similar to the case of two histograms. However, the joint probability density is not constant, which requires the numerical calculation of integrals of the form of Equation (2). The result is again a histogram that approximates the density distribution of the unknown random variable.

Operations on second order histograms. Let X and Y be second order histograms defined by grids $\{ v_i \mid i = 0, 1, \ldots, n \}$ and $\{ w_i \mid i = 0, 1, \ldots n \}$ and sets of histograms $\{Px_i\}$ and $\{Py_i\}$. By $Z = X * Y$, we denote a second order histogram resulting in the operation $* \in \{+, -, \cdot, /, \uparrow\}$ between X and Y. After having defined a grid $\{z_i \mid i = 0, 1, \ldots, n\}$, the histogram Pz_i on the interval $[z_k, z_{k+1}]$ is determined, following Equation (1), by the formula

$$Pz_k = \iint_{\Omega_k} X(\xi)\, Y(\eta)\, d\xi\, d\eta \;/\; (z_{k+1} - z_k)\,,$$

where $\Omega_k = \{ (\xi, \eta) \mid z_k \leq \xi * \eta \leq z_{k+1} \}$.

The function $X(\xi)\, Y(\eta)$ on each rectangle $[v_{i-1}, v_i] \times [w_{j-1}, w_j]$ is a constant histogram $Px_i \cdot Py_j$. In particular, the integral of a histogram over a certain region is equal to the sum of the histogram values multiplied by the areas of the regions.

4 Probabilistic Extensions

One of the most important problems that NPA handles is the computation of probability density functions of random variables expressed as functions of other random variables with given densities. Let us start with the general case when $(x_1, \ldots, x_n)$ is a vector of continuous random variables that have the joint probability density function $p(x_1, \ldots, x_n)$, while the random variable z is a function f of $(x_1, \ldots, x_n)$, i. e.

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

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

We construct a histogram F that approximates the probability density function of the variable $z = f(x_1, \ldots, x_n)$. Suppose the histogram F is defined on a grid $\{ z_i \mid i = 0, 1, \ldots, n \}$. The region is denoted as $\Omega_i = \{(x_1, \ldots, x_n) \mid z_i \leq f(x_1, \ldots, x_n) \leq z_{i+1}\}$. Then the value F_i of the histogram on the interval $[z_i, z_{i+1}]$ is

$$F_i \;=\; \int_{\Omega_i} p(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n \Big/ (z_{i+1} - z_i). \tag{3}$$

By *histogram probabilistic extension* of the function f, we mean a histogram F constructed according to (3).

Let $f(x_1, \ldots, x_n)$ be a rational function. To construct its histogram extension, we replace the variables $x_1, x_2, \ldots, x_n$ by histograms of their possible values, while the arithmetic operations are replaced by the histogram operations. We call the resulting histogram F the *natural histogram extension* of the function f (similar to “natural interval extension”).

Case 1. [4] Let $x_1, \ldots, x_n$ be independent random variables. If $f(x_1, \ldots, x_n)$ is a rational expression in which each variable x_i occurs no more than once, then the natural histogram extension approximates the probabilistic extension.

Case 2. [4] Let us assume that, through the change of variables, the expression of the function $f(x_1, \ldots, x_n)$ can be transformed to a rational expression $f(z_1, \ldots, z_k)$, depending on new variables $z_1, \ldots, z_k$, that satisfies the conditions of Case 1, while the variable z_i is a function of x_i, $i \in \mathrm{Ind}_i \subseteq \{1, 2, \ldots, n\}$, and the index sets Ind_i are mutually disjoint. Suppose that, for each z_i, it is possible to construct a probabilistic extension. Then the natural histogram extension of $f(z_1, \ldots, z_k)$ approximates the probabilistic extension of $f(x_1, \ldots, x_n)$.

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 and z_2, and then compute $f = z_1 \cdot z_2$, which is a rational function satisfying the conditions of Case 1. Hence, the natural histogram extension approximates the probabilistic extension for the function $f(x_1, x_2)$.

Case 3. We have to find a probabilistic extension for the function $f(x_1, x_2, \ldots, x_n)$, but the conditions of Case 1 and Case 2 are not fulfilled. Suppose that only one variable has several occurrences in the expression for f; let it be x_1 for definiteness. If, instead of the random variable x_1, we substitute a determinate value t, then it is possible to construct a natural histogram extension for the function $f(t, x_2, \ldots, x_n)$. We shall assume that t is a discrete random value approximating x_1 in the following sense: t takes values t_i with probability P_i, and it is possible to construct a natural histogram extension φ_i for each one of the functions $f(t_i, x_2, \ldots, x_n)$. Then a probabilistic extension of the function $f(x_1, \ldots, x_n)$ can be approximated by such a probability density function φ (see [4]),

$$\varphi(\xi) = \sum_{i=1}^{n} P_i \varphi_i(\xi)\,.$$

Clearly, this technique can be applied recursively to functions whose expressions have several occurrences of few variables.

Example. Let $f(x, y) = x^2 y + x$, and let x and y be random values uniformly distributed over the interval $[0, 1]$. We change x to a discrete random value t, $\{\, t_i \mid t_i = (i - 0.5)/m,\ i = 1, 2, \ldots, m\}$, $P_i = 1/m$, and then calculate natural histogram extensions φ_i.

More detailed analysis of the calculated results shows that φ approximates f with the order $\alpha = 1.4998$, that is, $\|f - \varphi\|_2 \le O(1/m^\alpha)$.

m	$\|f-\varphi\|_2$
10	1.2888E-03
20	4.5593E-04
40	1.6121E-04
80	5.6996E-05
160	2.0151E-05

Table 1: Approximation error of the probabilistic extensions

Case 4. We consider the problem of construction the probability density function for

$$y = f(x_1, x_2, \dots, x_n)\,,$$

when repeated samples for the vector $(x_1, x_2, \dots, x_n)$ are known.

Suppose that x_1, x_2, $\dots$, x_n are dependent variables, and repeated samples $X_1 = (x_1, x_2, \dots, x_n)_1$, $X_2 = (x_1, x_2, \dots, x_n)_2$, $\dots$, $X_N = (x_1, x_2, \dots, x_n)_N$ are known. We shall construct a histogram approximation P_y of the probability density function of the random value y. Let histogram P_y be defined on a grid $\{z_i \mid i = 0, 1, \dots, m\}$ takes the value p_j over the interval $[z_{i-1}, z_i]$, such that

$$p_j = \frac{m_j}{N(z_j - z_{j-1})}\,,$$

where m_j is the number of points $y_i = f(X_i)$ that falls into the interval $[z_{i-1}, z_i]$.

5 Solving Equations and Systems of Equations

Using NPA techniques, we construct numerical methods that solve linear and nonlinear equations and systems of algebraic equations with stochastic parameters [3].

First, we consider an NPA method for the solution of a system

$$\left\{\begin{array}{ccc} f_1(x,k) = 0, \\ \vdots \quad \ddots \quad \vdots \\ f_n(x,k) = 0\,, \end{array}\right. \tag{4}$$

either linear or nonlinear, where $k \in \mathbb{R}^m$ is a vector of random parameters, its probability density is $p(\xi_1, \xi_2, \dots, \xi_m)$ with the support set K, and $x \in \mathbb{R}^n$ is a random solution vector. The *solution set* of system (4) is the random set

$$\varXi = \bigl\{\, x \in \mathbb{R}^n \mid (\exists k \in K)(\, f_i(x,k) = 0,\, i = 1, 2, \dots, n) \bigr\}\,.$$

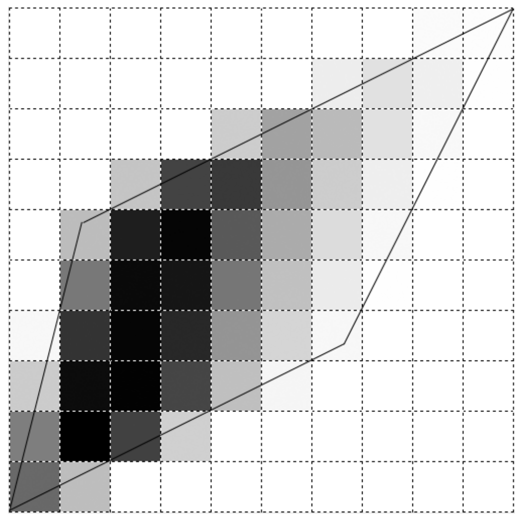

Figure 1: Probability density of the solutions to a linear system.

To every $x \in \Xi$, we can assign a parameter subset $K_x \subseteq K$ such that

$$K_x = \{\, k \in K \mid f_i(x,k) = 0,\ i = 1,2,\dots,n \,\}.$$

Suppose that, for a subset $Y \subseteq \Xi$, we have to find the probability $P(Y)$ of the solution x falling into the subset Y, i. e., of the membership $x \in Y$. If we introduce the parameter subset $K_Y = \{\, K_x \mid x \in Y \,\}$, then the required probability is

$$P(Y) = \int_{K_Y} p(\xi_1, \xi_2, \dots, \xi_m)\, d\xi_1 d\xi_2 \dots d\xi_m\,.$$

Constructing the subsets K_x and K_Y is not an easy task in general. However, it is quite feasible for some particular cases.

Let us consider two illustrative examples.

First, we turn to a system of linear algebraic equations

$$Ax = b\,, \text{ where}$$

$$A = \begin{pmatrix} a_{11} & -1 \\ -1 & 2 \end{pmatrix}, \qquad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

and the variables a_{11}, b_1, and b_2 are independent random variables such that $a_{11} \in [2,4]$, $b_1 \in [0,1]$, and $b_2 \in [0,1]$ with uniform probability distributions on the respective intervals. Discretizing the variables, we can employ the above elaborated technique (see details in [3]).

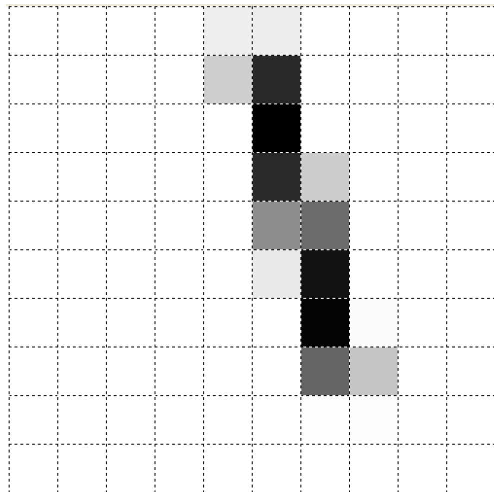

Figure 2: Probability density of the solutions to the system (5).

Figure 1 shows a histogram of the solution density for the step 0.1 along each component. The overall square has the size $[0,1]\times[0,1]$, and the solid line indicates the boundary of the solution set.

As a second example, consider the system of nonlinear equations

$$\begin{cases} ax^2 + by^2 - 4 = 0, \\ xy - c = 0, \end{cases} \tag{5}$$

where a, $\boldsymbol{b}$, and $\boldsymbol{c}$ are uniformly distributed random variables over the intervals $a \in \boldsymbol{a} = [1, 1.1]$, $b \in \boldsymbol{b} = [2, 2.1]$, and $c \in \boldsymbol{c} = [0.505, 0.51]$, respectively. If c is fixed, then for every x and y, the subset $K_{(x,y)}$ of the parameter space of all the (x, y)'s defined by the equation $ax^2 + by^2 - 4 = 0$ is a straight line. Next, it is not hard to find the probability density of the solution vector (x, y) and, as the result, to compute probability of the membership $(x, y) \in [\underline{x}, \overline{x}] \times [\underline{y}, \overline{y}]$, i. e., that the solution falls into a specified rectangle.

For example, we can find that the probability of being the solution within the rectangle $[0.37, 0.275] \times [1.36, 1.365]$ is equal to 0.1465. The comparable accuracy of the result is achieved by Monte Carlo modeling after $\approx 10^6$ trials.

Figure 2 shows a piecewise constant approximation of the joint probability density for the solution (x, y) to the system (2) (details can be found in [3]).

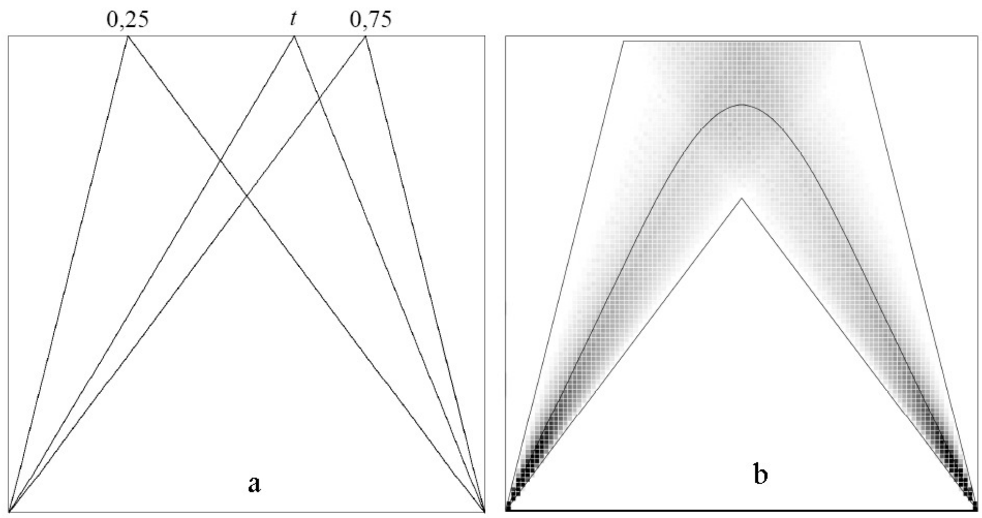

Figure 3: Second order histogram.

6 Second Order Histogram

In this section, we briefly consider the idea of second order histograms and their construction in the case of epistemic uncertainty.

Suppose that we have a series of histograms $\{Y_i, i = 1, 2, \ldots, N\}$. Let each Y_i be assigned a probability p_i such that $\sum_i^N p_i = 1$. For simplicity, we assume that all the histograms Y_i, $i = 1, 2, \ldots, N$, are defined on a common grid $\{z_j \mid j = 0, 1, \ldots, n\}$, and the histogram Y_i takes the value Y_{ik} at the interval $[z_{k-1}, z_k]$. We have thus a random variable denoted as $Y^{(k)}$, taking the value Y_{ik} with the probability p_i on the interval $[z_{k-1}, z_k]$. Using these values, we can restore the histogram Pz_k on each interval $[z_{k-1}, z_k]$.

Example. Let P_t be a random variable triangularly distributed on $[0, 1]$, with the height $h = 2$ and its top at the point $(t, 2)$. Then t is a random variable, triangularly distributed on $[0.25, 0.75]$, with the top $(0.5, 4)$ (see Figure 3a).

Figure 3b shows the corresponding second order histogram where the values of probability densities are shades in gray. Interval distributions (maximum and minimum P_t for all t) are shown by the boundary lines, and the middle line indicates an "efficient" probability density of the second order histogram, that is, the mean of the probability density at a specific point.

7 Risk Assessment

In this section, we consider application of Numerical Probabilistic Analysis to risk assessment procedures for investment projects. We use a priori information about the probability densities of sales and product price and calculate Net Present Value and Internal Rate of Return, important financial variables.

Net Present Value (NPV) of a time series of cash flows is the sum of the present values (PVs) of the individual cash flows. When studying specific investment projects, NPV is used to determine the present value of an investment by the discounted sum of all cash flows anticipated to be received from the project. The formula for the discounted sum of all cash flows is

$$NPV(r) = Cz_1 s_1 \sum_{i=1}^{T} \frac{C_i}{(1+r)^i} - C_0\,, \tag{6}$$

where C_0 is the initial investment, C_i is the i-th cash flow, T is time (assumed to be discrete), r is the discount rate, s_1 is the cost, and z_1 represents the expenditure.

Internal Rate of Return (IRR) determines the maximum acceptable discount rate at which one can invest without any loss to the owner. In other words, $IRR = r$ providing that $NPV(r) = 0$.

To provide an example of computing Net Present Value, we consider a company that decides whether it should invest in a new project. Specifically, let the company be expected to invest $C_0 = \$3.4 \cdot 10^6$ for the development of the new product. At the same time, the company estimates that the cash flow is going to be $C_i = c_i \cdot x_i$ during the i-th year, where c_i is the price, and x_i is sales volume. The expected return of 10% is used as the discount rate.

Under the circumstances, there is very high market uncertainty about the specific values of the main parameters of the model, so that the standard ('point') financial model cannot produce an adequate recommendation on how to make this or that decision. To take into account simultaneously the uncertainty in prices, sales, costs, and expenses, it makes sense to use Numerical Probabilistic Analysis. We represent the main parameters of the financial model, i. e., prices and sales, as random variables that have numerical probability distributions. The technique based on NPA enables one to understand which factors exhibit the greatest influence on the financial results of the project.

First, using expert estimates, we can construct histograms that approximate the probability densities of c_i, x_i, s_1, and z_1.

To model the selling price, we use a triangular distribution, determined by three parameters: the minimum value, the maximum value, and the most probable value ("top value"). Let the sale price in the first year have the minimum value \$5.90, the maximum value \$6.10, while the top value is a random variable with uniform

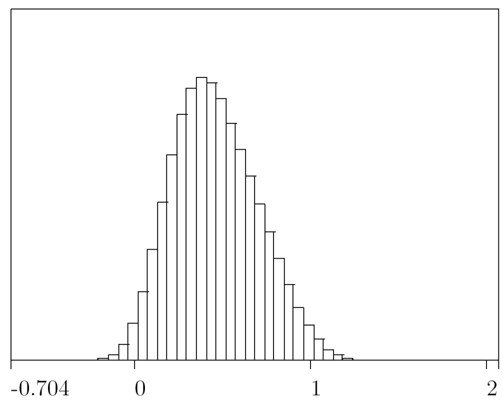

Figure 4: The mean of the second order histogram for NPV.

distribution on [\$5.95, \$6.05]. Similarly, the sale price for the second year has the triangular distribution with the parameters \$5.95, \$6.15, and [\$6.0, \$6.1]. Finally, the third year price has the triangular distribution with the parameters \$6.0, \$6.20, and [\$6.05, \$6.15].

Sales are approximated by random variables with normal Gaussian distributions. In the first year, the mathematical expectation μ_1 of the sale is a uniform random variable on [\$800,000, \$805,000], and the standard deviation σ_1 is a uniform random variable on [\$20,000, \$30,000]. In the second year, the expectation μ_2 is a uniform random variable on [\$950,000, \$1,000,000], and the standard deviation σ_1 is a uniform random variable on [\$20,000, \$30,000]. In the third year, μ_3 is a uniform random variable on [\$1,100,000, \$1,150,000], and the standard deviation σ_3 is a uniform random variable on [\$20,000, \$30,000].

Cost s_1 (as percentage of sales) is assumed to have a triangular distribution with the minimum of 50%, maximum of 65%, and the most likely value of 55%.

The expenditure z_1 (percentage of sales) is modeled as the normal Gaussian distribution with expectation of 15% and standard deviation of 2%.

The presence of various expert assessments leads to the necessity to build a second order histogram. Probability density functions for the histogram variables c_i, x_i, s_1, and z_1 are presented for $n = 50$.

Figures 4 and 5 show the mean of the second order histograms of NPV and IRR. Support of the NPV is $[-\$0.704 \cdot 10^6, \$2.07 \cdot 10^6]$, and support of the IRR is $[5\%, 30\%]$. Considering the histograms of NPV and IRR, one can see the possibility of both negative outcomes and considerable profits, contrary to the results of the standard 'point' analysis that does not capture the variability of the outcome.

Using estimates of NPV and IRR densities in the form of histograms and second

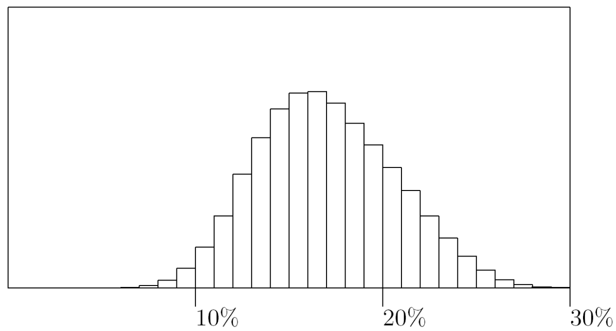

Figure 5: The mean of the second order histograms for IRR.

order histograms, we can assess the risk that the investment project is loss-making. If P_{NPV} is the probability density function of NPV, the probability P that the investment project is loss-making is

$$P = \int_{-\infty}^{0} P_{NPV}(\xi)\, d\xi\,.$$

8 Conclusions

We have considered the representation of uncertainty information and computational aspects of its processing with the use of Numerical Probabilistic Analysis (NPA).

Relying on practical examples, we demonstrate that using the histograms and second order histograms may prove very helpful in decision making. In particular, we applied the new technique in a computation of NPV and IRR estimates for risk assessment of investment projects.

Comparison of NPA and Monte Carlo method showed good agreement of their results. For instance, for the number of samples $N = 10^6$ and histogram dimension $n = 30$, four significant digits matched. At the same time, the numerical experiments demonstrate that the histogram arithmetic is more than 100 times faster than the Monte Carlo method. As a result, the approach based on NPA can be applied successfully to the solution of certain economic problems for which Monte Carlo simulation is infeasible.

References

[1] D. Berleant. Automatically verified reasoning with both intervals and probability density functions. *Interval Computations,* 1993, No. 2, pp. 48–70.

[2] V.A. Gerasimov, B.S. Dobronets, and M.Yu. Shustrov. Numerical operations of histogram arithmetic and their applications. *Automation and Remote Control,* vol. 52, 1991, No. 2, pp. 208–212.

[3] B.S. Dobronets and O.A. Popova. Numerical operations on random variables and their application. *Journal of Siberian Federal University. Mathematics & Physics,* vol. 4, 2011, No. 2, pp. 229–239. (in Russian)

[4] B.S. Dobronets and O.A. Popova. Elements of numerical probability analysis. *SibSAU Vestnik,* vol. 42, 2012, No. 2, pp. 19–23. (in Russian)

[5] B.S. Dobronets and O.A. Popova. Numerical probabilistic analysis for the study of systems with uncertainty. *Journal of Control and Computer Science,* vol. 21, 2012, No. 4, pp. 39–46.

[6] S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, and K. Sentz. Constructing probability boxes and Dempster-Shafer structures. *Technical Report SAND2002–4015,* Sandia National Laboratories, 2003. `http://www.sandia.gov/epistemic/Reports/SAND2002-4015.pdf`

[7] W. Li, J. Hym. Computer arithmetic for probability distribution variables. *Reliability Engineering and System Safety,* vol. 85, 2004, pp. 191–209.

[8] R.E. Moore. Risk analysis without Monte Carlo methods. *Freiburger Intervall-Berichte,* No. 84/1, 1984, pp. 1–48.

[9] A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing,* vol. 10, 2004, pp. 249–272.

[10] S.P. Shary. Interval analysis or Monte-Carlo methods? *Computational Technologies,* vol. 12, 2007, No. 1, pp. 103–112. (in Russian)

[11] S.P. Shary. *Finite-Dimensional Interval Analysis.* XYZ, Novosibirsk, 2013. Electronic book (in Russian) available at `http://www.nsc.ru/interval/Library/InteBooks/SharyBook.pdf`

[12] I.M. Sobol. A Primer for the Monte Carlo Method. CRC Press, Boca Raton, 1994.

[13] L.P. Swiler and A.A. Giunta. Aleatory and epistemic uncertainty quantification for engineering applications. *Technical Report SAND2007–2670C*, Sandia National Laboratories, 2007.

[14] W.T. Tucker and S. Ferson. Probability bounds analysis in environmental risk assessments. *Technical report,* Applied Biomathematics, 2003.

[15] R. Williamson and T. Downs. Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning,* vol. 4, 1990, pp. 89–158.

[16] R. Yager. Uncertainty modeling and decision support. *Reliability Engineering & System Safety,* vol. 85, 2004, No. 1–3, pp. 341–354.

[17] R. Yager and V. Kreinovich. Decision making under interval probabilities. *International Journal of Approximate Reasoning,* vol. 22, 1999, pp. 195–215.