SCAN — 2012

Numerical Probabilistic Analysis under Aleatory and Epistemic Uncertainty

Boris S. Dobronets & Olga A. Popova

Sibirian Federal University
E-mail: BDobronets@sfu-kras.ru, olgaarc@yandex.ru

Novosibirsk, 2012

Numerical Probabilistic Analysis (NPA)

NPA is the section of Computing Mathematics.
Subject of NPA is a decision of the problems with stochastic uncertainty in data.
Methods of NPA use numerical operations under probability density functions of random variables and their functions.

Numerical Operations
$\{+, -, \cdot, /, \uparrow, \max, \min\}$

Binary relations
$\{\leq, \geq\}$

Types of Uncertainty
- Aleatory
- Epistemic

NPA

PDF
- discrete
- histogram
- piecewise-polynomial

Probabilistic Extension
- Natural
- Histogram

Second-Order Histogram

Solution
- SLAE
- Nonlinear Equations

Practical Applications
- Decision Making
- Risk assessment
- NPV, IRR

Types to uncertainties

L.P. Swiler, A.A. Giunta. Sandia Technical Report (2007).

"Where it is practical, calculation input characterizations should separate aleatory and epistemic uncertainties."

Aleatory & Epistemic uncertainties

Aleatory uncertainties are characterized by frequency distributions. Alternative terminologies include: variability, stochastic uncertainty, irreducible uncertainty, and Type A uncertainty.

Epistemic uncertainties are characterized degrees of “belief” and should not be given a frequency interpretation.

Numerical Probabilistic Analysis

The basis of NPA is numerical operations on probability density functions of the random values.
Arithmetic on probability density function uses operations as $* \in \{+, -, \cdot, /, \uparrow, \mathsf{max}, \mathsf{min}\}$, and binary relations as $\{\leq, \geq\}$.

Histogram arithmetic

Numerical operations of histogram arithmetic is one of NPA components.
The first idea histogram arithmetic was published in the article

V.A. Gerasimov, B.S. Dobronets, and M.Yu. Shustrov

Numerical operations of histogram arithmetic and their applications. Automation and Remote Control, (Feb 1991), 52(2), pp. 208–212.

Histogram arithmetic

Base idea of the histogram approach is concluded in following:
probability density function of random value can be written in the histogram form (piecewise constant function). For example one — dimensional random value histogram function is defined by mesh $\{x_i | i = 0, ..., n\}$ and means constant P_i on each segment $[x_i, x_{i+1}]$, $h = \max_{i=0}^{n-1}\{x_{i+1} - x_i\}$.

Probabilistic extensions

Definition 1.

Let $(x_1, ..., x_n)$ be a system of continuous random variables with joint probability density function $p(x_1, ..., x_n)$ and random variable z is the function $f(x_1, ..., x_n)$

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

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

Histogram probabilistic extensions

Suppose the histogram F is defined mesh $\{z_i | i = 0, ..., n\}$.
The region is denoted as $\Omega_i = \{(x_1, ..., x_n) | z_i < f(x_1, ..., x_n) < z_{i+1}\}$.
Then the histogram value F_i on the interval $[z_i, z_{i+1}]$ is defined as

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

Definition 2.

By histogram probabilistic extension of the function f we mean an histogram F constructed from (1).

Natural histogram extensions

Let $f(x_1, ..., x_n)$ be rational function.
To construct of histogram of F replaced by the arithmetic operation on the histogram operation, and variables x_1, x_2,..., x_n replaced by histogram of values.

Definition 3.

The resulting histogram of F is called a natural histogram extension.

Histogram probabilistic extensions and arithmetic operations

Let P be a histogram of the probability density function $z = x * y$, and $* \in \{+, -, \cdot, /, \uparrow\}$. Then the value of P_i on the interval $[z_i, z_{i+1}]$ is defined by formula

$$P_i = \int_{\Omega_i} p(x, y) dxdy / (z_{i+1} - z_i), \tag{2}$$

where $\Omega_i = \{(x, y) | z_i \leq x * y \leq z_{i+1}\}$.

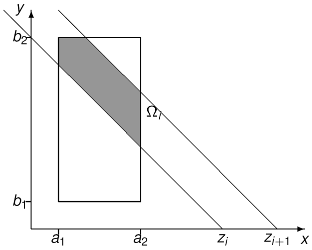
The histogram of the sum for two random variables

$$z = x + y,$$

then P_z is a histogram of the probability density function of z and

$$p_{zi} = (\int_{\Omega_i} p(x, y)dxdy/(z_{i+1} - z_i). \quad (3)$$

Support of $p(x, y)$ is a rectangle $[a_1, a_2] \times [b_1, b_2]$ and
$\Omega_i = \{(x, y)|z_i \leq x + y \leq z_{i+1}\}$.



Theorem 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 histogram extension approximates a probabilistic extension to $O(h^\alpha)$, $\alpha \geq 1$.

Example 1.

$$f(x, y) = xy + x + y + 1$$

Example 1.

$$f(x, y) = xy + x + y + 1 = (x + 1)(y + 1).$$

Theorem 2.

Let the function $f(x_1, ..., x_n)$ 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 Theorem 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 $f(z_1, \ldots, z_k)$ would be approximated by a probabilistic extension $f(x_1, ..., x_n)$.

Example 2.

Let $f(x_1, x_2) = (-x_1^2 + x_1)sin(x_2)$.
Then $z_1 = (-x_1^2 + x_1)$ and $z_2 = sin(x_2)$.
We shell notice that possible to construct a probabilistic extension for functions z_1, z_2 and $f = z_1 * z_2$ be a rational function satisfying the conditions of Theorem 1. So natural extension will approximate probabilistic extension to function $f(x_1, x_2)$.

General case

Consider case when necessary to find probabilistic extension for function $f(x_1, x_2..., x_n)$ but conditions of Theorem 2 are not fulfilled.
Suppose for definiteness that only x_1 occurs a few times.
If instead of random variable x_1 to substitute determinate value t then possible construct natural probabilistic extension to function $f(t, x_2..., x_n)$.

Suppose t is an discrete random value approximating x_1 the following let t takes values t_i with probability P_i and each one function $f(t_i, x_2 ..., x_n)$ possible to construct natural probabilistic extension.
Then a probabilistic extension f of the function $f(x_1, ..., x_n)$ can be approximated by a probability density φ as follows:

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

Example 3.

Let $f(x, y) = x^2 y + x$ and x, y are uniformly distributed on $[0, 1]$ interval random values.
We shell change x to discrete random value t, $\{t_i | t_i = (i - 0.5)/n, i = 1, 2, ..., n\}$, $P_i = 1/n$ and shell calculate natural probabilistic extensions φ_i.

Table 1. Approximating error of the probabilistic extensions

n	$\|\|f - \varphi\|\|_2$
10	1.2887825282E-03
20	4.5592973952E-04
40	1.6120775967E-04
80	5.6996092139E-05
160	2.0151185588E-05

Analysis of calculated results has shown that φ approximates f with $\alpha = 1.4998$, here α is approximation order.

Construction to probabilistic density function for dependent variables under aleatory uncertainty

$$y = f(x_1, x_2, ..., x_m).$$

We shell consider the problem of construction to probabilistic density function when repeated samples for vector $(x_1, x_2, ..., x_m)$ are known.
Suppose $x_1, x_2, \ldots, x_m$ be dependent variables and repeated samples
$X_1 = (x_1, x_2, \ldots, x_m)_1$,
$X_2 = (x_1, x_2, \ldots, x_m)_2, \ldots,$
$X_N = (x_1, x_2, \ldots, x_m)_N$ are known.
We shell constructed to histogram estimation P_y of probabilistic density function for random value y.
Let histogram P_y is defined on mesh $\{z_i, i = 0, 1, \ldots, n\}$.
Then histogram takes value P_j on interval $[z_{i-1}, z_i]$ where

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

n_j — number of points $y_i = f(X_i)$, from interval $[z_{i-1}, z_i]$.

Comparison of NPA and Monte Carlo Methods

Monte Carlo method displays convergence $1/\sqrt{N}$. Monte Carlo Errors reduce by a factor of $1/\sqrt{N}$. Where N is the number of sampled points.
Error of histogram extension is $O(1/n^{\alpha})$, $\alpha \approx 2$.
In practice using the histogram extensions is more efficient than Monte Carlo Methods more than $10^2 - 10^3$ times.

Example 4. Comparison of approximation errors

Necessary to find p the sum of four standard uniformly distributed random variables.

$$p(x) = \begin{cases} \frac{1}{6}x^3, & \text{in } 0 \le x \le 1; \\ -\frac{1}{2}x^3 + 2x^2 - 2x + \frac{2}{3}, & \text{in } 1 \le x \le 2; \\ \frac{1}{2}x^3 - 4x^2 + 10x - \frac{22}{3}, & \text{in } 2 \le x \le 3. \\ -\frac{1}{6}x^3 + 2x^2 - 8x + \frac{32}{3}, & \text{in } 3 \le x \le 4. \end{cases}$$

Let N be the number of sampled and n be dimension of mesh.
H_n is histogram probabilistic extension of p for n (exact histogram). P_n is natural histogram extension of p for n, $MC_{n,N}$ is histogram approximation of Monte Carlo method of p for n, N

Table 2. Errors of histogram arithmetic and Monte Carlo Methods

n	$N = 10^4$	$N = 10^5$	$N = 10^6$	$\|\|H_n - P_n\|\|_2$.
10	0.0059	0.00168	0.00037	4.16e-3
20	0.0055	0.00198	0.00041	5.39e-4
50	0.0026	0.00103	0.00026	3.47e-5
100	0.0023	0.00062	0.00018	4.35e-6
150	0.0016	0.00055	0.00016	1.28e-6
200	0.0014	0.00044	0.00014	5.44e-7

This table represents the approximation errors $||H_n - P_n||_2$ and $||H_n - MC_{n,N}||_2$. We can see that for a fixed n error of the Monte Carlo method decreases as $\approx 1/\sqrt{N}$, order of convergence natural histogram extension is $\alpha \approx 3.5$.

Second Order Histogram (SOH)

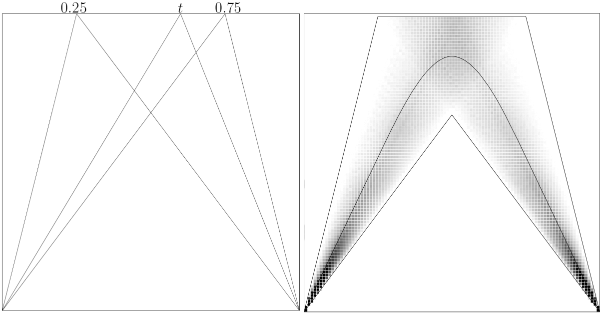
Definition 4.

Second-order histogram is piecewise histogram function.

SOH is determined by the mesh $\{z_i | i = 1, 2, ..., n\}$ and set of histogram $\{P_i | i = 1, 2, ..., n\}$.
On each interval $[z_i, z_{i+1}]$ SOH is a histogram P_i.

Example 5. Second Order Histogram

Let P_t be triangular distributed on $[0, 1]$ random variable with height $h = 2$ and top $(t, 2)$. Let t be triangular distributed on $[0.25, 0.75]$ random variable with top $(0.5, 4)$.



The top and bottom lines corresponds to the interval histogram and the middle line correspond to the mean SOH.
Values probability densities are shades of gray.

Numerical operation under Second Order Histogram

$$Z = X * Y, \quad * \in \{+, -, \cdot, /, \uparrow\},$$

X, Y — Second Order Histograms determined on mesh $\{x_i, i = 0, 1, \ldots n\}$, $\{y_i, i = 0, 1, \ldots n\}$ and set of histograms $\{Px_i\}$ and $\{Py_i\}$.
Present Z as SOH Pz. Pz determined on mesh $\{z_i | i = 0, 1, \ldots, n\}$ and set of histograms $\{Pz_i\}$.
Then Pz_i on $[z_{i-1}, z_i]$ is determined by formula

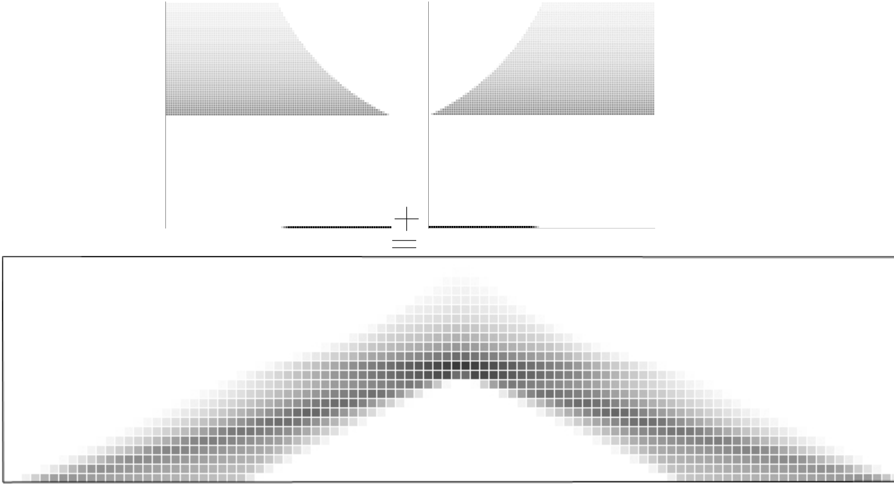
$$Pz_i = \int_{\Omega_i} X(\xi) Y(\eta) d\xi d\eta / (z_i - z_{i-1}), \tag{4}$$

where $\Omega_i = \{(\xi, \eta) | z_i \le \xi * \eta \le z_{i+1}\}$.

Example 6. Addition of SOH

Suppose we want to add two second order histograms X and Y.
SOH X and Y are generated by uniform random variables defined respectively on $[0, t_1]$ and $[t_2, 2]$, where t_1 is uniform random variable defined on the interval [1,2], t_2 is uniform random variable defined on the interval [0,1].

The sum of two second order histograms



The result of the addition of two SOH X and Y constructed in the form SOH Z.
Support of Z is the interval [0,4], the height of 1.
Values of probability densities are shades of gray.

The nonlinear equations

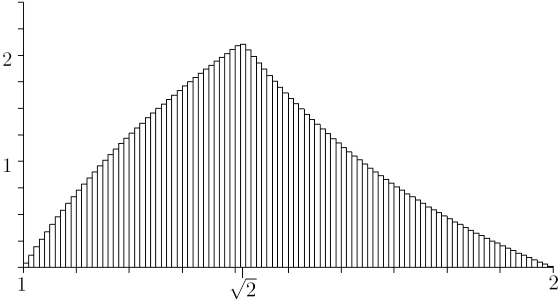
$$f(x,k) = 0,$$

where k — vector of random parameters, $x \in [a, b]$.
Let ϕ_z be probabilistic extensions of $f(z, \boldsymbol{k})$ and $z \in [a, b]$.
Then $P(z)$ is a probability that the root x is to the left (right) point z:

$$P(z) = \int_{-\infty}^{0} \phi_z(\xi) d\xi.$$

The nonlinear equations

$\boldsymbol{a}x^2 - \boldsymbol{b} = 0$, where $\boldsymbol{a}$, $\boldsymbol{b}$ — random variable with uniform distribution on $[1, 2]$, $[2, 4]$.



Histogram of the root of the square equation

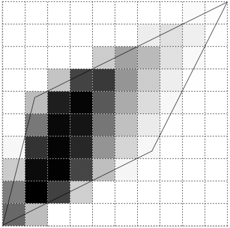
Solution of linear system equations

$$Ax = b.$$

Let $\boldsymbol{b}$ be random vector, and $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ be independing uniformly distributed components on $[0, 1]$ interval.
Suppose that matrix A is

$$A = \begin{pmatrix} a_{11} & -1 \\ -1 & 2 \end{pmatrix}.$$

and component $\boldsymbol{a}_{11}$ is independent random value uniformly distributed on $[2, 4]$ interval.



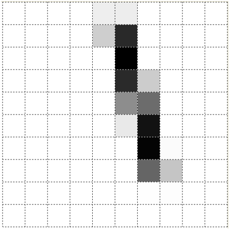
Piecewise constant with step 0.1 approaching the joint density probability $\boldsymbol{x}$. The solid line shows the boundary the set of solutions of the original system.

Solution of nonlinear system equations

$$\boldsymbol{a}x^2 + \boldsymbol{b}y^2 - 4 = 0,$$

$$xy - \boldsymbol{c} = 0,$$

where $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ — independent uniformly distributed components on $[1, 1.1]$, $[2, 2.1]$, $[0.505, 0.51]$.



Piecewise constant with step 0.1 approaching the joint density probability $(\boldsymbol{x}, \boldsymbol{y})$.

Risk Assessment

Consider risk assessment of investment projects. We use a priori information about the probability densities of sales and product price and calculate NPV and IRR.

NPV & IRR

Net Present Value (NPV) and Internal Rate of Return (IRR)

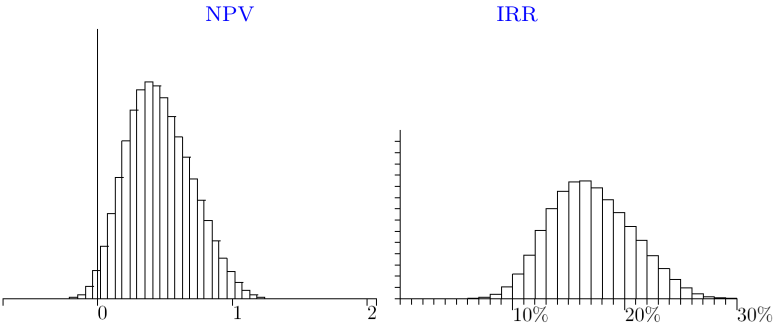
$$NPV(r) = 0.8181818 \cdot 0.68 z_1 s_1 \sum_{i=1}^{3} \frac{c_i x_i}{(1+r)^i} - 3400000, \qquad (5)$$

r — the discount rate,
c_i — price,
x_i — volume of sales,
s_1 — cost,
z_1 — expenditures.
IRR determines the maximum acceptable discount rate in which you can invest without any loss to the owner: $IRR = r$, in which the

$$NPV(r) = 0. \qquad (6)$$

Using expert estimates were constructed histogram approximating the probability density c_i, x_i, s_1, z_1. Presence of various expert assessments can build SOH.

NPV & IRR



The figure shows the mean of the second order histograms of NPV and IRR. Histogram analysis of NPV and IRR can see that as very likely negative outcomes, and the possibility of considerable profit compared with the standard analysis. Using estimates of density NPV and IRR in the form of histograms and second order histograms, we can assess the risk that the investment project will be loss-making. So if P_{NPV} is histogram probability density NPV, then the probability that the investment project will be loss-making can be calculated by the formula

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

NPV & IRR

- Probability density function for the variables c_i, x_i, s_1, z_1 histograms were presented with $n = 50$.
- Comparison of NPV calculations and Monte Carlo simulation showed that when the number of experiments $N = 1000000$ coincides with the results of the Histogram calculation of up to three or four decimal places.
- Numerical experiments have shown that this histogram arithmetic more than three hundred times faster.
- To calculate the IRR to solve nonlinear equations. In the case of a numerical probability analysis, the computation of the histogram of the root of a nonlinear equation is reduced to the computation of the integrals of the corresponding histogram extensions.

References:

[1] L.P. Swiler, A.A. Giunta, Aleatory and Epistemic Uncertainty Quantification for Engineering Applications, Sandia Technical Report, SAND2007-2670C

[2] V.A. Gerasimov, B.S. Dobronets, and M.Yu. Shustrov, Numerical operations of histogram arithmetic and their applications. Automation and Remote Control, (Feb 1991), 52(2), pp. 208–212.

[3] B.S. Dobronets, O.A. Popova, Numerical probabilistic analysis and probabilistic extension, Proceedings of the XV International EM'2011 Conference. Oleg Vorobyev, ed. — Krasnoyarsk: SFU, RIFS (2011), pp. 67–69.

[4] B.S. Dobronets, O.A. Popova, Numerical Operations on Random Variables and their Application, Journal of Siberian Federal University. Mathematics & Physics (2011), 4(2), pp. 229–239.

[5] B.S. Dobronets, O.A. Popova, Histogram time series, Proceedings of the X International FAMES'2011 Conference. Oleg Vorobyev, ed. — Krasnoyarsk: RIFS, SFU, KSTEI, (2011), pp. 127–130.

[5] Skyrms B. Higher Order Degrees of Belief. // Prospects for Pragmatism: Essays in Memory of F.P. Ramsey, D.H. Mellor, ed. Cambridge; New York: Cambridge University Press, 1980., pp. 109–137.

[6] B.S. Dobronets, O.A. Popova, Elements of a numerical probabilistic analysis, SibSAU Vestnik (2012), 2(42), pp. 19–23.