Вычислительный Вероятностный Анализ: модели и методы

Б.С. Добронец

СФУ, Красноярск
E-mail: BDobronets@yandex.ru

апрель 2023

Вычислительный Вероятностный Анализ

Отличительной особенностью ВВА является наличие развитых арифметических операций над функциями плотности вероятности.

⋆ Вычисления функций от случайных аргументов с использованием процедур построения вероятностных расширений.

⋆ В рамках ВВА решаются различные задачи численного анализа, в том числе задачи интерполяции, аппроксимации и оптимизации.

Новая парадигма. Distributions Are the Numbers of the Future

Schweizer B.

Distributions are the numbers of the future. In Proceedings of the mathematics of fuzzy systems meeting (Naples, Italy), (1984) pp. 137-149.

Герасимов В.А., Добронец Б.С., Шустров М.Ю.

Численные операции гистограммной арифметики и их приложения // Автоматика и телемеханика. 1991. № 2. С. 83-88.

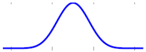
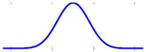
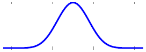
Javier Arroyo.

Distributions are the numbers of today: from histogram data to distributional data (Facultad de Informatica. Universidad Complutense de Madrid), (2013).

Risk analysis

R.E. Moore.

Risk analysis without Monte Carlo methods. Freiburger Intervall-Berichte, №. 84/1, 1984, pp. 1–48

Данные	Модель	Результат
$\boldsymbol{x}_1$  $\rightarrow$ $\vdots$ $\rightarrow$ $\boldsymbol{x}_n$  $\rightarrow$	$\boxed{y = f(x_1, ..., x_n)}$	$\rightarrow$ $\boldsymbol{y}$: 

Кусочно-полиномиальные функции

$$\omega = \{-\infty, x_0, x_1, \ldots, x_n, \infty\}$$

⋆ Кусочно-постоянные функции (гистограммы);
⋆ Кусочно-линейные функции (частотный полигон);
⋆ сплайны;
⋆ обобщенные кусочно-полиномиальные функции.

Вероятностные расширения

Одной из проблем, которая рассматривается ВВА, является задача построения функции плотности вероятности случайных величин рассмотрим общий случай, когда $(x_1, \ldots, x_n)$ есть система непрерывных случайных величин с совместной функцией плотности вероятности /joint probability density function / $p(x_1, \ldots, x_n)$ и случайная величина z есть функция от случайных функций $f(x_1, \ldots, x_n)$

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

Носителем функции плотности вероятности $\boldsymbol{f}$ будем называть множество

$$\text{supp}(\boldsymbol{f}) = \{x | \boldsymbol{f}(x) > 0\}.$$

Вероятностное продолжение

Будем говорить, что случайная функция $\boldsymbol{f} : \boldsymbol{R}^n \to \boldsymbol{R}$ является вероятностным продолжением детерминированной функции $f : R^n \to R$ на множестве $D \subset R^n$, если $\boldsymbol{f}(x) = f(x)$ для всех аргументов $x \in D$.

Вероятностное расширение

Случайная функция $\boldsymbol{f} : \boldsymbol{R}^n \to \boldsymbol{R}$ называется probabilistic extension детерминированной функции $f : R^n \to R$ на множестве $D \subset R^n$, если она

(i) является вероятностным продолжением f на D,

(ii) $\boldsymbol{f}$ совпадает с $\boldsymbol{z}$ плотностью вероятности случайной величины z

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

Таким образом, мы можем записать

$$\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n).$$

В тех случаях, когда надо указать непосредственно значение $\boldsymbol{f}$ в некоторой точке ξ, будем использовать обозначение

$$\boldsymbol{z}(\xi) = \boldsymbol{f}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)(\xi).$$

Теорема 1

Пусть $\boldsymbol{f}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ есть вероятностное расширение функции $f(x_1, x_2, \ldots, x_n)$ и для всех вещественных t функция $\boldsymbol{f}(t, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ есть вероятностное расширение функции $f(t, x_2, \ldots, x_n)$. Тогда

$$\boldsymbol{f}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)(\xi) = \int_{\mathrm{supp}(\boldsymbol{x}_1)} \boldsymbol{x}_1(t)\boldsymbol{f}(t, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)(\xi)dt \tag{1}$$

Замечание

Теорема 1 предполагает возможность рекурсивных вычислений для общего вида вероятностных расширений и сведение вычислительного процесса к вычислению одномерного случая.

Рассмотрим вычисление интеграла (1). Для простоты представим (1) как квадратуру

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

Далее для вычисления $\boldsymbol{f}(t_l,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n)(\xi)$ мы можем также использовать квадратуры and так далее. В общем случае, это NP-сложная проблема, актуально распараллеливание.

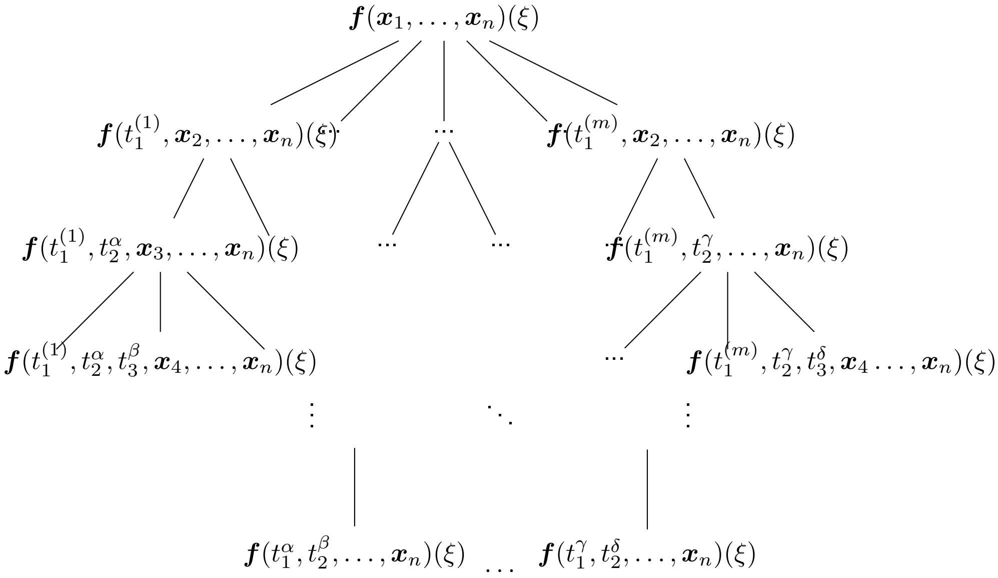


Рис.: Дерево параллельно-рекурсивной организации вычислительного процесса

Число операций

Монте-Карло

$$N \sim \varepsilon^{-2}.$$

Вероятностные расширения

$$N \sim \varepsilon^{-(n-1)/\alpha},$$

n — размерность вектора,
α — порядок сходимости квадратурной формулы.

Естественные Вероятностные расширения

Пусть $f(x_1, \ldots, x_n)$ есть рациональная функция. Для построения вероятностного расширения f заменим арифметические операции вероятностными, переменные x_1, x_2, ..., x_n заменим их плотностями вероятности. В результате вероятностное расширение f является естественным вероятностным расширением.

Случай 1.

Пусть $x_1, \ldots, x_n$ — независимые случайные переменные. Если $f(x_1, \ldots, x_n)$ рациональная функция, x_i встречаются не более одного раза, тогда естественное вероятностное расширение совпадает с вероятностным.

Случай 2.

Пусть функция $f(x_1, \ldots, x_n)$ может быть заменой переменных, поэтому что $f(z_1, \ldots, z_k)$ является рациональной функцией переменных $z_1, \ldots, z_k$, удовлетворяющих условиям случая 1. Переменная z_i является функцией x_i, $i \in Ind_i$. и Ind_i взаимно не пересекаются. Предположим, для каждого z_i можно построить вероятностное расширение. Тогда естественное расширение $f(z_1, \ldots, z_k)$ будет аппроксимировать вероятностное расширение $f(x_1, \ldots, x_n)$.

Одномерный случай

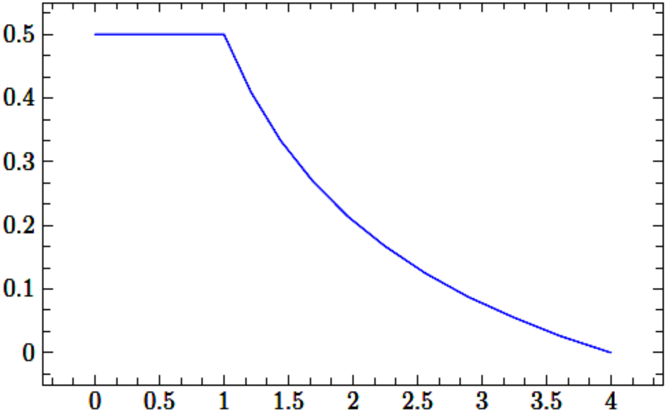
Пусть дана функциональная зависимость

$$z = f(x),$$

где x — случайная величина, $\boldsymbol{x}$ — функция плотности вероятности случайной величины x с носителем $[\underline{x}, \overline{x}]$. Далее $\{x_i(z) \in [\underline{x}, \overline{x}] | i = 1, \ldots, n\}$ — корни уравнения $z = f(x)$.

Мы можем представить решение в виде

$$\boldsymbol{f}(\boldsymbol{x})(\xi) = \sum_{i=1}^{n} \frac{\boldsymbol{x}(x_i(\xi))}{|f'(x_i(\xi))|}.$$



Вероятностное расширение функции x^2, x — случайная величина с треугольным распределением.

Например, для нахождения плотности вероятности $p_{x_1+x_2}$ суммы двух случайных величин $\boldsymbol{x}_1 + \boldsymbol{x}_2$ используется соотношение

$$p_{x_1+x_2}(x) = \int_{-\infty}^{\infty} p(x-v, v)dv = \int_{-\infty}^{\infty} p(v, x-v)dv. \tag{2}$$

Плотность вероятности p_{x_1/x_2} частного двух случайных величин $\boldsymbol{x}_1/\boldsymbol{x}_2$ определяется выражением

$$p_{x_1/x_2}(x) = \int_{0}^{\infty} vp(xv, v)dv - \int_{-\infty}^{0} vp(v, xv)dv. \tag{3}$$

Плотность вероятности $p_{x_1x_2}$ произведения двух случайных величин $\boldsymbol{x}_1\boldsymbol{x}_2$ представляется соотношением

$$p_{x_1x_2}(x) = \int_{0}^{\infty} (1/v)p(x/v, v)dv - \int_{-\infty}^{0} (1/v)p(v, x/v)dv. \tag{4}$$

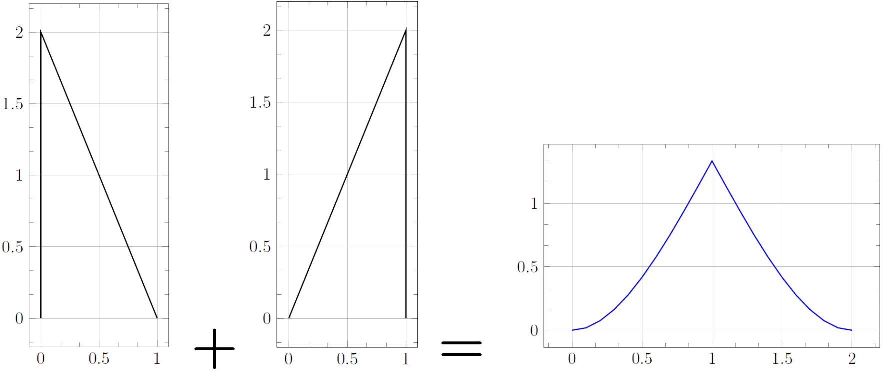


Рис.: Пример сложения двух случайных величин

Рассмотрим сложение четырех независимых равномерных случайных величин $z = x_1 + x_2 + x_3 + x_4$. Поскольку для сложения выполняется ассоциативность, то $z = (x_1 + x_2) + (x_3 + x_4)$. Известно, что z имеет распределение Ирвина-Холла при $n = 4$

$$IH_4(x) = \begin{cases} \frac{1}{6}x^3, & \text{if } 0 \le x \le 1; \\ -\frac{1}{2}x^3 + 2x^2 - 2x + \frac{2}{3}, & \text{if } 1 \le x \le 2; \\ \frac{1}{2}x^3 - 4x^2 + 10x - \frac{22}{3}, & \text{if } 2 \le x \le 3; \\ -\frac{1}{6}x^3 + 2x^2 - 8x + \frac{32}{3}, & \text{if } 3 \le x \le 4. \end{cases}$$

Случайные величины $z_1 = (x_1 + x_2)$ и $z_2 = (x_3 + x_4)$ распределены по треугольному закону с носителем на отрезке $[0, 2]$ с вершиной в точке $(1, 1)$

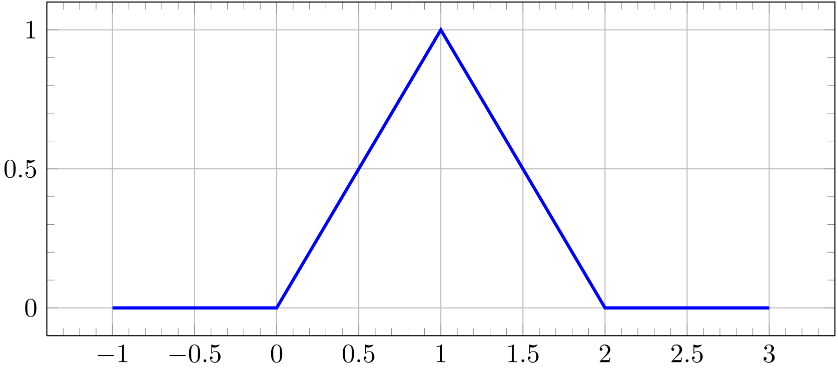
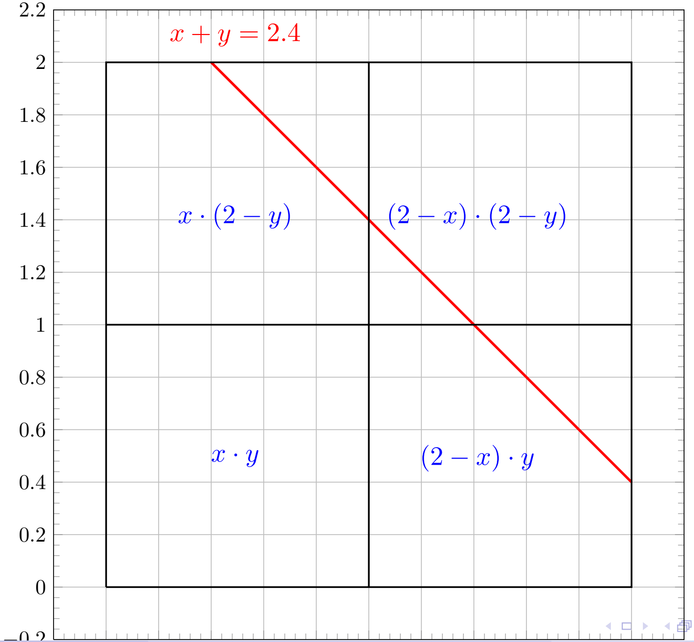


Рис.: Функции плотности вероятности случайной величины распределенной по треугольному закону

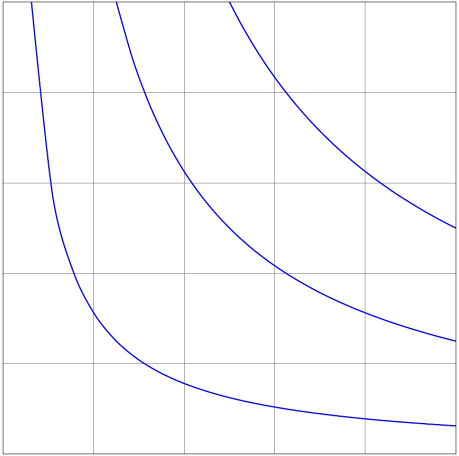

$x + y = 2.4$
$x \cdot (2 - y)$
$(2 - x) \cdot (2 - y)$
$x \cdot y$
$(2 - x) \cdot y$
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
−0.2

Численные операции

Поскольку кубический сплайн на каждом отрезке сетки представляет кубический полином, то $p(x, y)$ в случае вычисления интегралов (2)–(4) будет кусочно полиномиальной функцией шестой степени. Можно найти интегралы точно или использовать квадратуры Гаусса с четырьмя внутренними узлами, которые точны на полиномах седьмой степени.
В качестве примера, рассмотрим построение сплайна, аппроксимирующего $p_{x_1+x_2}$. Для этих целей в области носителя $p_{x_1x_2}$ построим сетку $\omega = \{x_0, x_1, ..., x_n\}$ и вычислим значения $f_i = p_{x_1+x_2}(x_i)$. Используя значения f_i на сетке ω построим кубический сплайн s. В этом случае справедлива оценка

$$||p^{(\nu)}_{x_1+x_2} - s^{(\nu)}|| \leq Kh^{4-\nu}||p^{(4)}_{x_1+x_2}||,\ \nu = 0, 1, 2.$$

В каждой ячейки сетки вычисляем интеграл численно. Число операций для вычисления $f_i \sim Cn$. Для сплайна $\sim Cn^2$.



Далее вычислим

В качестве примера работы рассмотрим вычисление функции плотности вероятности произведения двух независимых случайных величин x_1, x_2. В случае известных функций плотности вероятности $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$ функция плотности вероятности произведения имеет вид

$$(\boldsymbol{x}_1 \cdot \boldsymbol{x}_2)(z) = \int_{-\infty}^{\infty} x_1(x) x_2(z/x) \frac{1}{|x|} dx.$$

Предположим, что x_1, x_2 распределенных по треугольному закону с носителем на отрезке $[-1, 1]$ и вершиной в точке $(0, 1)$ рис. 5.

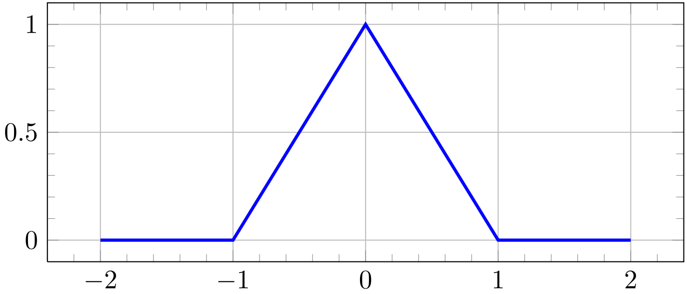


Рис.: Функции плотности вероятности $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$

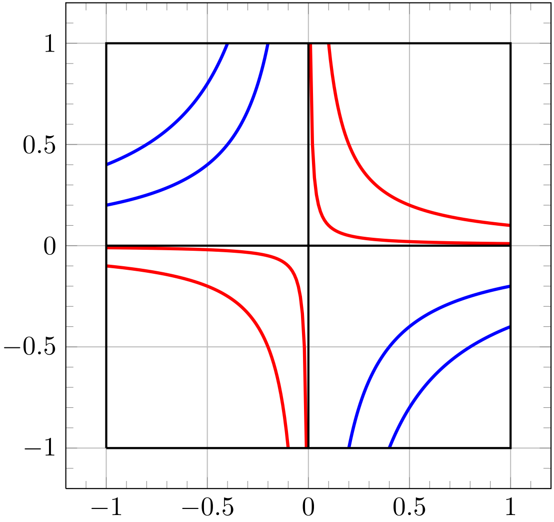
В этом случае $(\boldsymbol{x}_1 \cdot \boldsymbol{x}_2)$ имеет вид

$$(\boldsymbol{x}_1 \cdot \boldsymbol{x}_2)(z) = \int_z^1 (1-x)(1-z/x)\frac{1}{|x|}dx + \int_{-1}^{-z} (1+x)(1+z/x)\frac{1}{|x|}dx. \tag{5}$$

В силу того, что исходные функции плотности вероятности представлены кусочно-полиномиальными функциями интеграл (5) может быть вычислен в явном виде

$$\boldsymbol{x}_3(z) = (\boldsymbol{x}_1 \cdot \boldsymbol{x}_2)(z) = -4 - 4|z| - 2\ln(|z|)(|z|+1).$$

На рисунке 6 показана совместная функция плотности вероятности вектора (x_1, x_2), красные линии — линии интегрирования для вычисления $\boldsymbol{x}_3(\xi)$ при $\xi > 0$, синие линии — линии интегрирования для вычисления $\boldsymbol{x}_3(\xi)$ при $\xi < 0$.



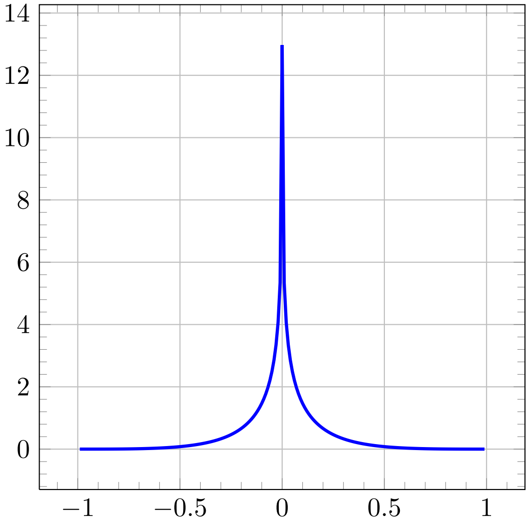


Рис.: Плотность вероятности x_3 произведения случайных величин $(x_1 \cdot x_2)$

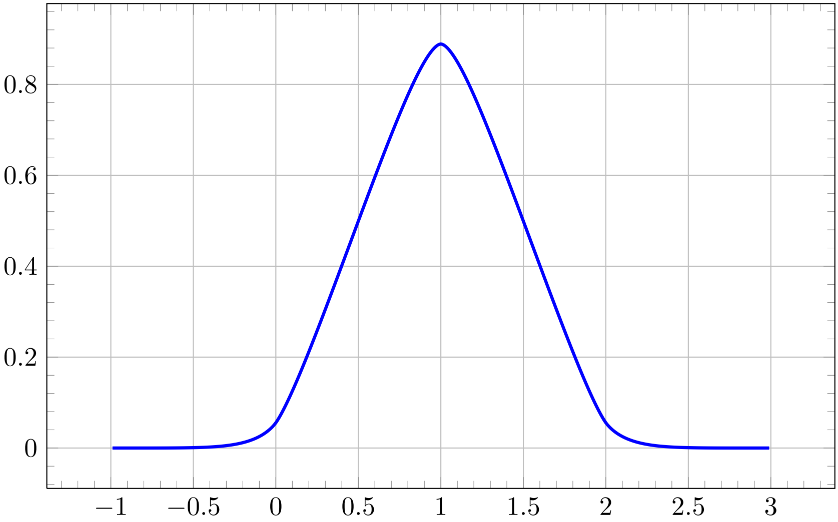


Рис.: Сумма x_3 и случайной величины распределенной по треугольному закону

Заметим, что особенности при вычислении произведений случайных величин возникают только в случае когда носитель совместной функции плотности вероятности содержит точку (0,0).

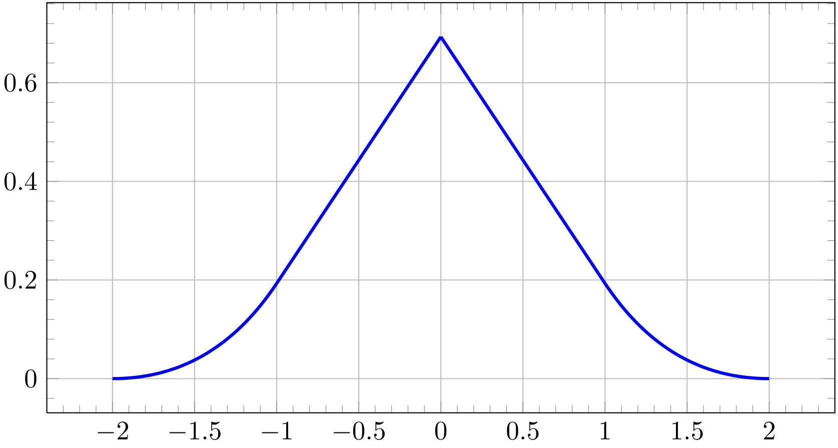


Рис.: Плотность вероятности произведения случайных величин распределенной по треугольному закону с носителем $[-1, 1]$ и равномерной на $[1, 2]$

Системы линейных алгебраических уравнений

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

где $A = (a_{ij})$ — случайная матрица и $b = (b_i)$ — случайный вектор соответственно. Предположим, что случайная матрица A и вектор b имеют независимые компоненты с плотностями вероятности $\boldsymbol{A} = (\boldsymbol{a}_{ij})$, $\boldsymbol{b} = (\boldsymbol{b}_i)$ соответственно

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

Носитель множества решений может быть представлен в виде [?]

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

Построим вероятностное расширение $\boldsymbol{x}_1(\boldsymbol{A}, \boldsymbol{b})$

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

или

$$\boldsymbol{x}_1(\boldsymbol{A}, \boldsymbol{b})(\xi) = \int_\Omega \boldsymbol{a}_{12}(t_{12}) \dots \boldsymbol{a}_{nn}(t_{nn}) \frac{\sum \boldsymbol{b}_i \Delta_i(t_{12}, \dots, t_{nn})}{\sum \boldsymbol{a}_{1i} \Delta_i(t_{12}, \dots, t_{nn})}(\xi) dt_{12} \dots dt_{nn}, \tag{7}$$

где $\Delta_i(t_{12}, \dots, t_{nn}) \in R$ миноры из метода Крамера для решения СЛАУ, $t_{ij} \in$ supp$(\boldsymbol{a}_{ij})$, $\Omega =$ supp$(\boldsymbol{a}_{12})\times$ ... $\times$supp$(\boldsymbol{a}_{nn})$. Выражение

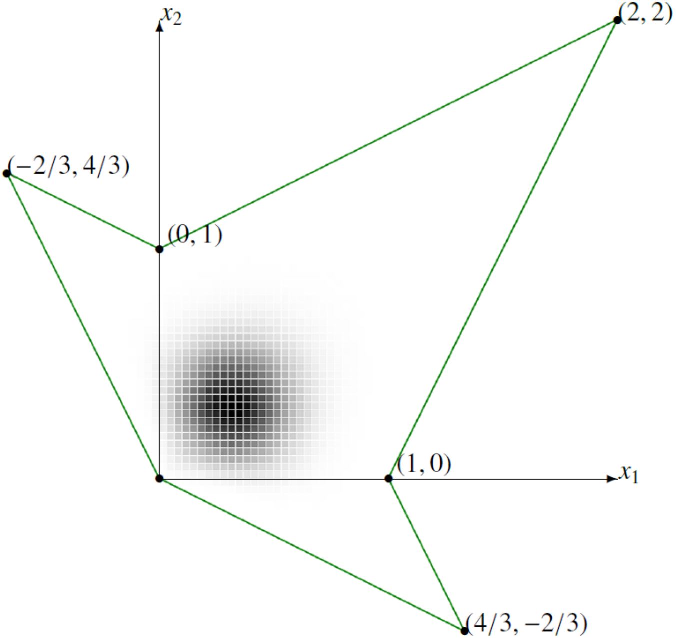
$$\left(\frac{\sum \boldsymbol{b}_i \Delta_i(t_{12}, \dots, t_{nn})}{\sum \boldsymbol{a}_{1i} \Delta_i(t_{12}, \dots, t_{nn})} \right)(\xi)$$

вычисляется используя вероятностные арифметики.

Рассмотрим систему линейных алгебраических уравнений

$$Ax = b, \tag{8}$$

Пусть $A = (a_{ij})$ — случайная матрица $n = 2$. Элементы матрицы $\boldsymbol{A}$ независимы и распределены по треугольному закону, $\boldsymbol{a}_{11}, \boldsymbol{a}_{22}$ распределены на интервале $[2, 4]$, a_{21}, a_{12} распределены на отрезке $[-1, 1]$. Вектор $\boldsymbol{b}$ состоит из независимых компонент $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, распределенных по треугольному закону на отрезке $[0, 2]$.


x_2
$(2, 2)$
$(-2/3, 4/3)$
$(0, 1)$
$(1, 0)$
x_1
$(4/3, -2/3)$

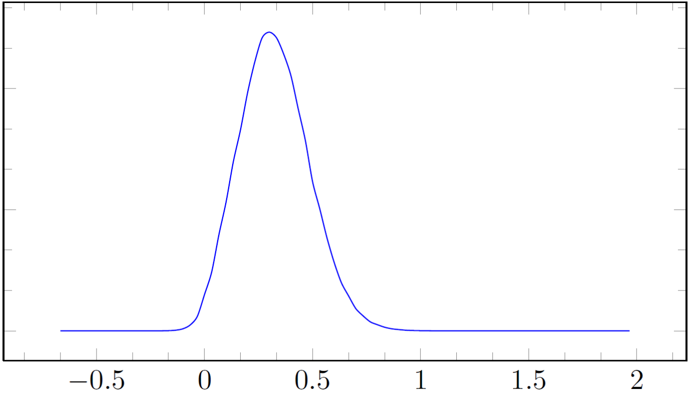


Рис.: Функция плотности вероятности x_1

На рис. 11 показана функция плотности вероятности первой компоненты вектора решений x_1. Плотность вероятности случайной величины x_1 имеет носитель $[-2/3, 2]$. Тем не менее, вне интервала $[-0.1, 1]$ плотность вероятности пренебрежимо мала.

Совместное использование Монте-Карло и ВВА

Для этих целей, например, при вычислении $\boldsymbol{x}_1(\xi, \boldsymbol{A}, \boldsymbol{b})$ (7) реализуем N выборочных значений $\zeta_i = (t^i_{12}, \ldots, t^i_{nn}), i = 1, \ldots, N$ случайного вектора $\zeta = (a_{12}, \ldots, a_{nn})$ согласно плотностей $\boldsymbol{a}_{12}, \ldots, \boldsymbol{a}_{nn}$. Для каждого вектора ζ_i вычислим

$$I_i = t^i_{12} \cdot \ldots \cdot t^i_{nn} \frac{\sum \boldsymbol{b}_i \Delta_i(t_{12}, \ldots, t_{nn})}{\sum \boldsymbol{a}_{1i} \Delta_i(t_{12}, \ldots, t_{nn})}(\xi), i = 1, \ldots, N. \tag{9}$$

Следовательно, согласно методу Монте-Карло

$$\boldsymbol{x}_1(\boldsymbol{A}, \boldsymbol{b})(\xi) \approx \bar{I} = \frac{1}{N} \sum_i^N I_i.$$

Краевые задачи со случайными коэффициентами

$$\begin{aligned} Lu &= f, \qquad x \in D, && (10)\\ u(x) &= 0, \qquad x \in \partial\overline{D}, && (11) \end{aligned}$$

$$Lu = -\sum_{i=1}^{2} \frac{\partial}{\partial x_i}(a_i \frac{\partial}{\partial x_i} u).$$

$a_i(x, \omega_a)$ with $x \in D$ and $\omega \in \Omega_a$, where $(\Omega_a, F_a, \mathbb{P}_a)$ denotes a complete probability space and $(\Omega_f, F_f, \mathbb{P}_f)$, $f(x, \omega_f)$ respectively.

Risk assessment

Gordon A. Fenton, D.V. Griffiths

Risk assessment in geotechnical engineering. John Wiley & Sons 2008

Stochastic Finite element methods

⋆ Monte Carlo method.

Gunzburger M D, Webster C G and Zhang G 2014

Stochastic Finite element methods for partial differential equations with random input data *Acta Numerica* **23** pp 521 – 650

Stochastic Collocation method and Polynomial chaos expansions

Ghanem R G and Spanos P 1991

Stochastic Finite Element: A Spectral Approach (New York: Springer)

D. Xiu 2010

Numerical Methods for Stochastic Computations: A Spectral Method Approach Princeton University Press

Babuška I, Nobile F and Tempone R (2007a)

A stochastic collocation method for elliptic partial differential equations with random input data *SIAM J. Numer. Anal* **45** pp 1005–1034

Shalimova I A and Sabelfeld K K 2017

Solution to a stochastic Darcy equation by the polynomial chaos expansion *Siberian J. Num. Math. / Sib. Branch of Russ. Acad. of Sci. Novosibirsk*, **20**(3) 313–327

Random Boundary value Problem

$$Lu \equiv -(pu')' = f(x), x \in (0,1), \tag{12}$$

с граничными условиями

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

где $p(x) = a_1(1-x) + a_2 x$, a_1, a_2 — случайные константы с совместной функцией плотности вероятности $p_a(a_1, a_2)$, с носителем $\omega = [0.5, 1] \times [0.5, 1]$.

Random Boundary value Problem

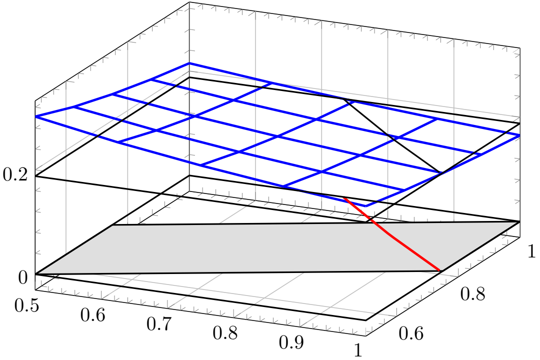
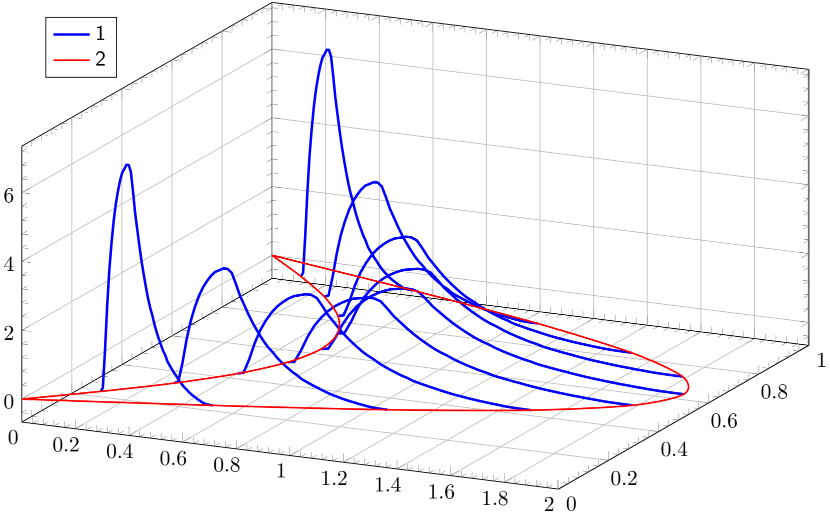


Рис.: Построение функции плотности вероятности решения

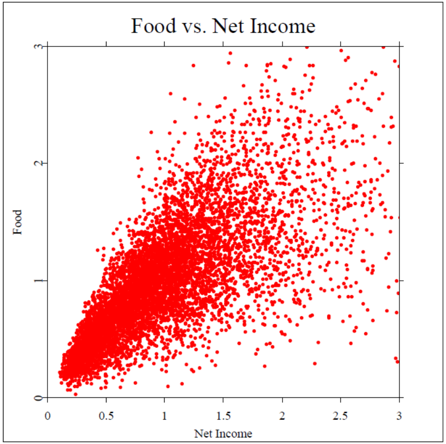
Краевые задачи



Замечание

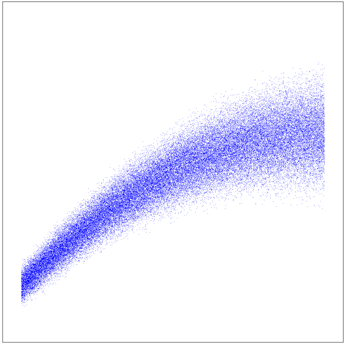
The main computational costs are spent on building the set $u_{ikl} \sim O(KLN)$. Computational costs building of the probability extension $\boldsymbol{u}_i \sim O(m)$. Therefore, once you have u_{ikl}, you can compute relatively quickly $\boldsymbol{u}_i$ for different $\boldsymbol{p}, \boldsymbol{q}$.

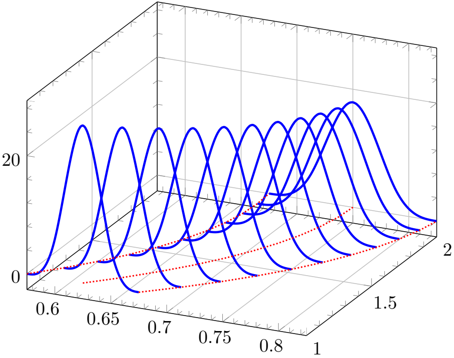
Зависимость расходов на питание от чистого дохода



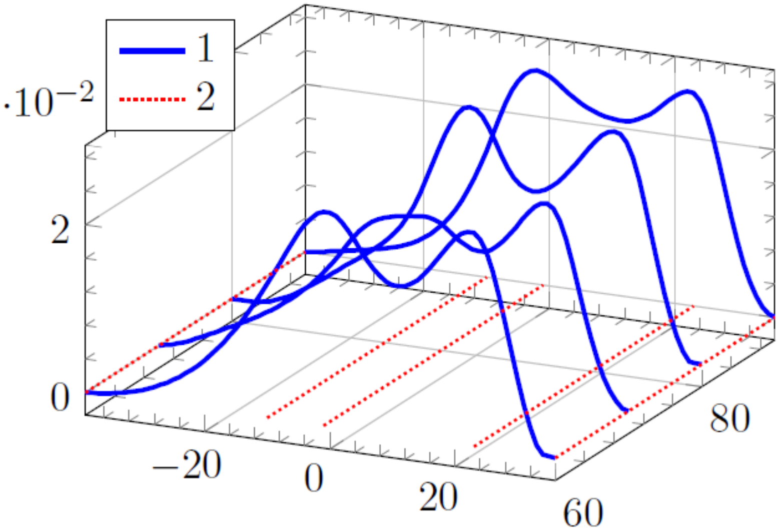
⋆ Хардле В. Прикладная непараметрическая регрессия: Пер. с англ. — М., Мир, 1993. — 349

Модельный пример

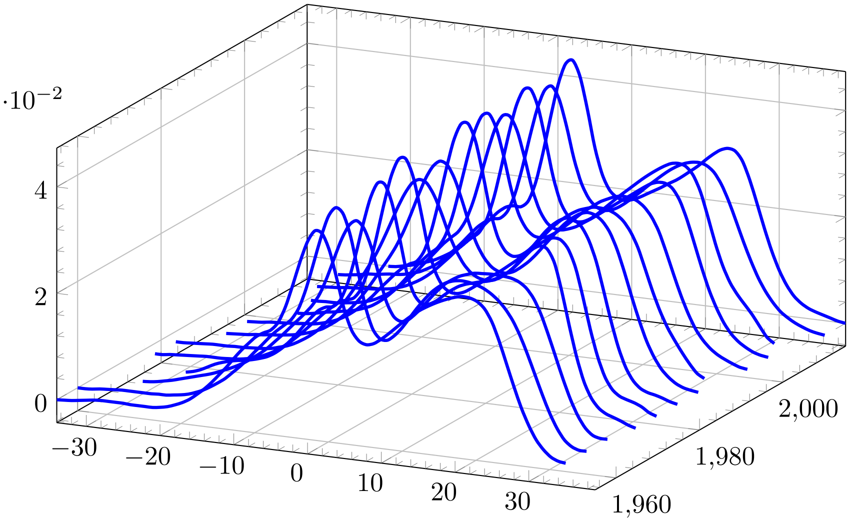




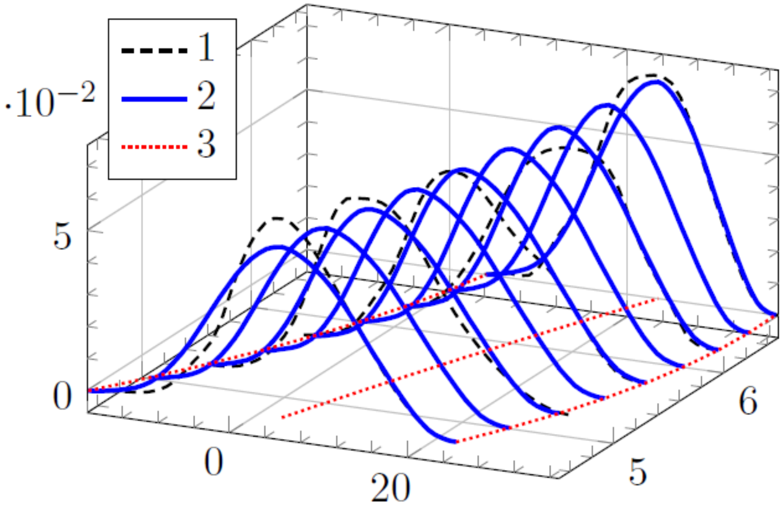
Красноярск, температура по годам



Москва, температура по годам



Красноярск, температура по дням



Достоверные оценки

Под достоверностью оценок статистических показателей следует понимать степень их соответствия отображаемой ими действительности.

Достоверными результатами считаются те, которые не искажают и правильно отражают объективную реальность.

Оценить достоверность результатов исследования означает определить, с какой вероятностью возможно перенести результаты, полученные на основе выборочной совокупности, на всю генеральную совокупность.

Оценка достоверности необходима для того, чтобы по части явления можно было бы судить о явлении в целом, изучая его закономерности

Надежные оценки Эмпирической Функции Распределения

Пусть $x_1, \ldots, x_n$ есть вещественные случайные величины с функцией распределения $F(x)$. тогда эмпирическая функция распределения F_n определяется как

$$F_n(x) = \frac{m_x}{n}. \tag{13}$$

где m_x есть число $x_i < x$.

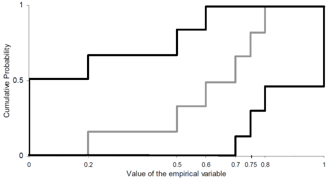
Теорема Колмогорова

$$\sqrt{n} \sup_{x \in R} |F - F_n| \to K, \text{ при } n \to \infty,$$

где K — случайная величина имеющая распределение Колмогорова. На основе этой теоремы построится интервальная функция распределения (P-box), содержащая функцию распределения F с вероятностью γ для $n \to \infty$:

$$F(x) \in F_n(x) + [-\Delta, \Delta],$$

где $\Delta = k_\gamma / \sqrt{n}$ и k_γ определяется как решение уравнения $K(k_\gamma) = \gamma$.



Пусть

$$z_i = F(x_i), i = 1, \ldots, n.$$

Заметим, что $z_i, i = 1, \ldots, n$ есть равномерно распределенные величины на отрезке [0,1].

Если $z_1 \leq z_2 \leq \ldots \leq z_n$, то z_k is k порядковая статистика с

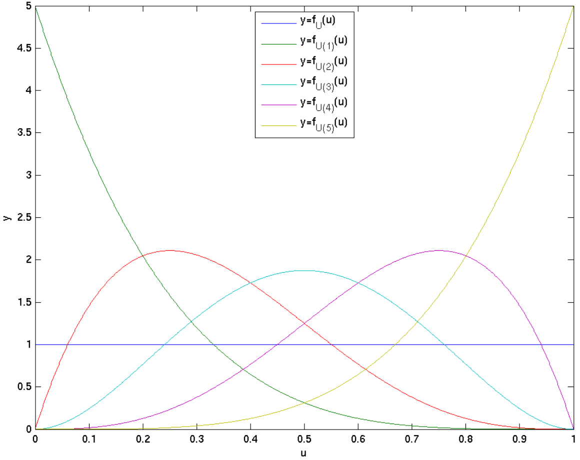
$$\mathsf{M}[z_k] = k/(n+1).$$

Если использовать точные значения z_i, тогда погрешность кусочно-линейной функции s(x) на сетке $\{x_i\}$ с $h = \max(x_{i+1} - x_i), i = 0, \ldots, n$ удовлетворяет оценке

$$||F - l_1|| \leq Kh^2||F^{(2)}||.$$

Таким образом даже при относительно небольших n, построенные оценки достаточно хорошо аппроксимируют функцию распределения F. Относительно z_i известно, что они образуют порядковые статистики.

Порядковые статистики



Вероятность k-ой порядковой статистики

Случайная кусочно-линейная интерполяция

Let $a = x_0 < x_1 < x_2 < \ldots < x_n < b = x_{n+1}$ be mesh,
$h = \max(x_{i+1} - x_i), i = 0, \ldots, n$ and $h = O(1/n)$.
В этом случае на каждом отрезке x_i, x_{i+1} реализуется полином Лагранжа первой степени.

$$\boldsymbol{l}_1(x) = \boldsymbol{f}_i \frac{x_{i+1} - x}{x_{i+1} - x_i} + \boldsymbol{f}_{i+1} \frac{x - x_i}{x_{i+1} - x_i}.$$

Случайное поле

В вероятностном пространстве $(\Omega, \mathcal{F}, P)$ случайный процесс представляет собой набор случайных величин

$$\{a(x,\omega), x \in D, \omega \in \Omega\}. \tag{14}$$

Термин «случайное поле» обычно относится к случайному процессу, принимающему значения в евклидовом пространстве R^d $d = 1, 2, 3$. Случайное поле можно посмотреть двумя способами:

– для фиксированного $x \in D$, $a(x, \cdot)$ является случайной величиной в Ω;

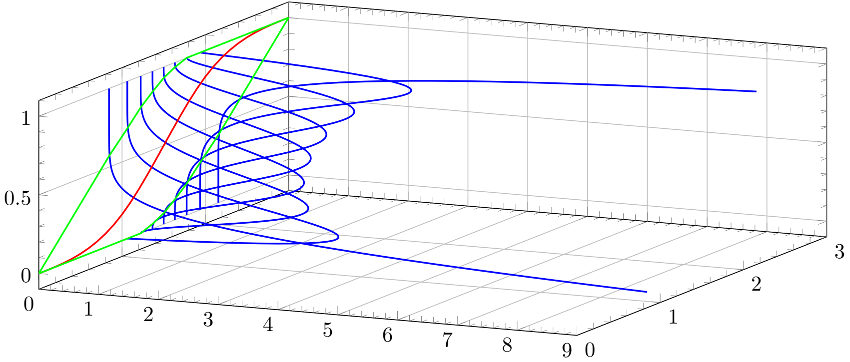
– для фиксированного $\omega \in \Omega$, $a(\cdot, \omega)$ является реализацией случайного поля в D.

Распределение второго порядка

Распределение второго порядка

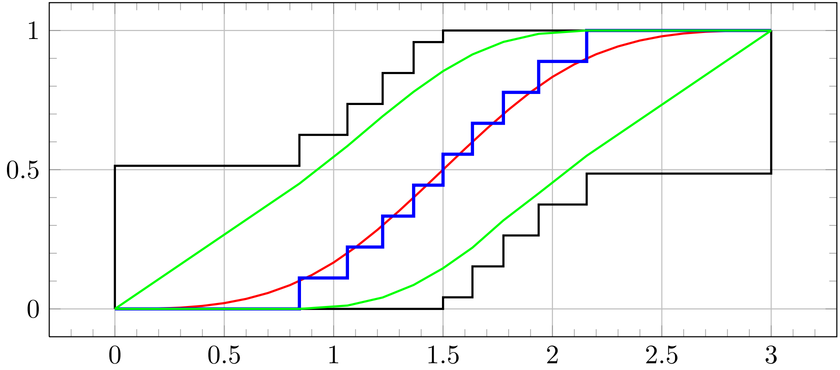
$\boldsymbol{f}^{(2)}$ — случайное поле $f(x, \omega)$, $x \in D, \omega \in \Omega$ заданное на $D \subset \mathbb{R}$. где $(\Omega, \mathcal{F}, P)$ — вероятностное пространство. Обладает следующими свойствами: для фиксированного $\omega \in \Omega$, $f(\cdot, \omega)$ является функцией распределения.

Метод построения функции распределения второго порядка для надежной оценки на малых выборках



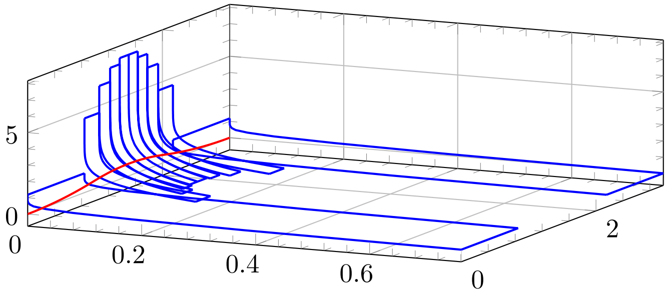
Представлена функция распределения второго порядка аппроксимирующая распределения кусочно-линейных интерполяций распределений Ирвина-Холла $n = 3$, построенная на выборке случайной величины размерности 9. Голубые линии — плотности вероятности случайной кусочно-линейной функции. Красная линия — точная функция распределения. Зеленые линии — границы 95% доверительной области.

Сравнение границ Колмогорова-Смирнова и надежных оценок



Красная линия — точная функция распределения. Черные линии — границы Колмогорова-Смирнова. Зеленые линии — границы 95% доверительной области ВВА. Голубая линия — эмпирическая функция распределения.

Пример применения метода построения производной от функции распределения второго порядка



Производная от кусочно-линейной функции — кусочно-постоянная функция. Приведен пример кусочно-постоянной функция — функции плотности вероятности второго порядка (надежной оценке ф.п.в.)
Красная линия — точная функция плотности вероятности. Синии линии — функции плотности вероятности надежной оценки, построенной на выборке случайной величины размерности 9.

Операции над распределениями второго порядка

Пусть $\boldsymbol{f}^{(2)}, \boldsymbol{g}^{(2)}$ — распределения второго порядка, $(\Omega_f, \mathcal{F}_f, P_f)$, $(\Omega_g, \mathcal{F}_g, P_g)$ — соответственно их вероятностные пространства. Тогда результат операции $\boldsymbol{f}^{(2)} \star \boldsymbol{g}^{(2)}$, распределение второго порядка $\boldsymbol{F}^{(2)}$

$$F(\cdot, \omega_\star) = \{f(\cdot, \omega_f) \star g(\cdot, \omega_g) | (\omega_f, \omega_g) \in \Omega_\star\}$$

где $(\Omega_\star, \mathcal{F}_\star, P_\star)$ — вероятностное пространство, $\Omega_\star = \Omega_f \times \Omega_g$.

Сглаживающие сплайны

$$S(x) = fv(x-1) + mw(x-1) + v(x-2).$$

$v(x) = (|x|-1)^2(2|x|+1), w(x) = x(|x|-1)^2$

$$\sum_{i=1}^{n}(s(\xi_i) - z_i)^2 + \alpha\|s''\|_2^2 \to \min,$$

$$Ay = b$$

$y = (f, m)$

α выбиралось таким образом, (f, m) определяли монотонный сплайн. (f, m) должны лежать внутри области

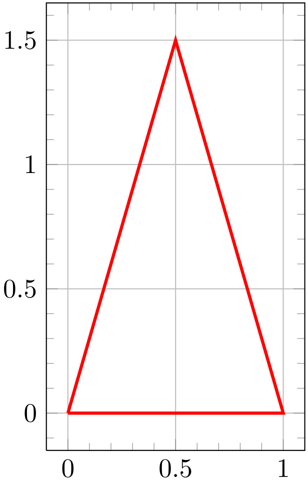


Рис.: Область монотонности для (f, m)

Таким образом, распределение второго порядка $\boldsymbol{S}$ можно представить в параметризованном виде

$$\boldsymbol{S}(x, \boldsymbol{f}, \boldsymbol{m}) = \boldsymbol{f} v(x-1) + \boldsymbol{m} w(x-1) + v(x-2).$$

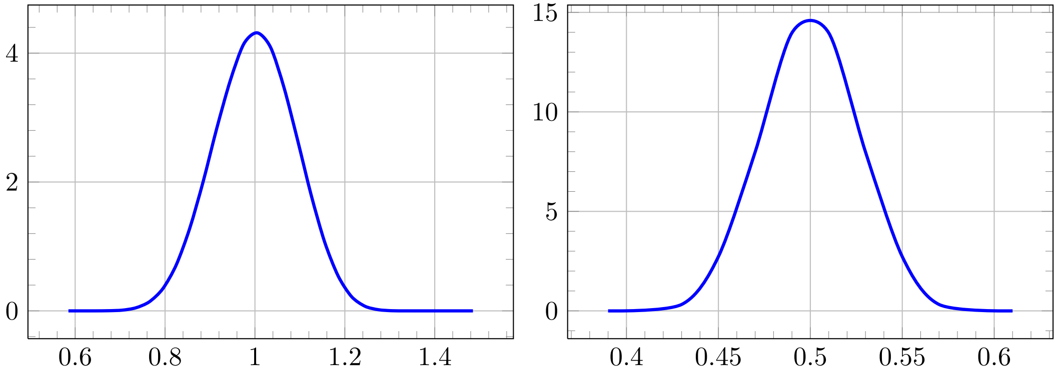


Рис.: Плотность вероятности константы $\boldsymbol{f}$, $\boldsymbol{m}$

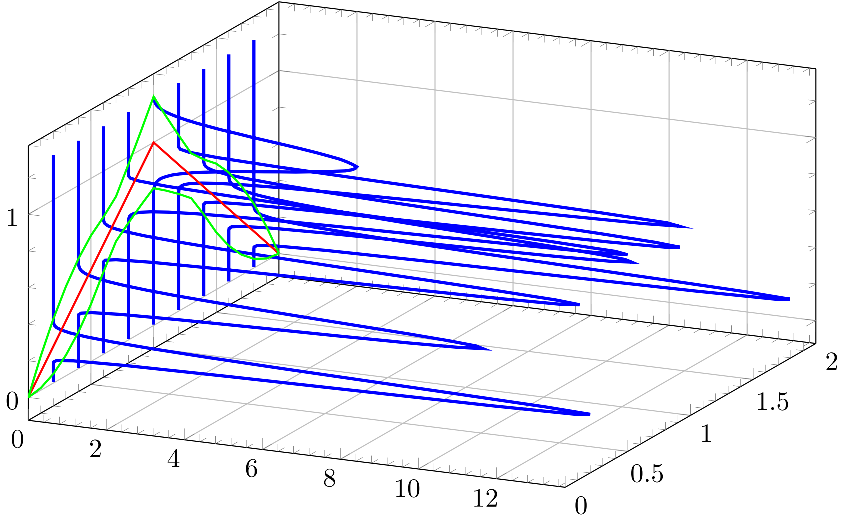


Рис.: Функция плотности вероятности распределения второго порядка

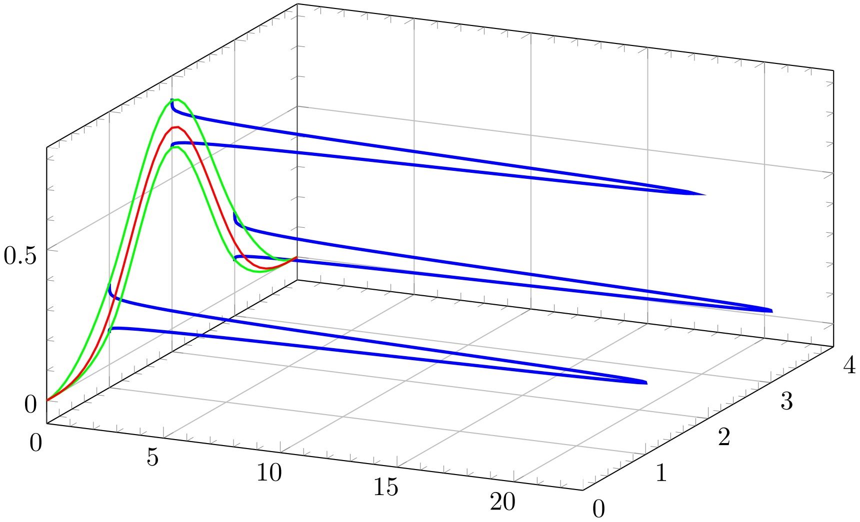


Рис.: Сумма распределений второго порядка

Applications

- Risk assessment of investment projects.
- Optimization of hydroelectric power generation.

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

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

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

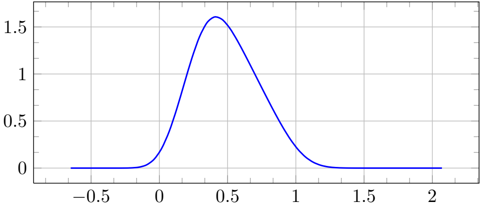
$$NPV(r) = 0.$$

NPV & IRR

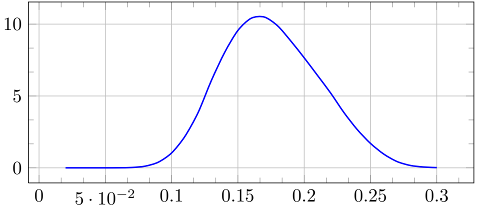
- Probability density function for the variables c_i, x_i, s_1, z_1 histograms were presented with $n = 50$.
- Comparison of NPV calculations and Monte Carlo simulation showed that when the number of experiments $N = 1000000$ coincides with the results of the Histogram calculation of up to three or four 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.

NPV & IRR

NPV



IRR



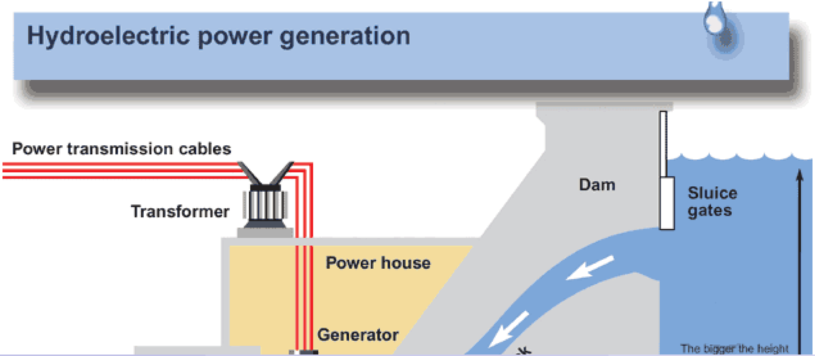
Analysis of NPV and IRR can see that as very likely negative outcomes, and the possibility of considerable profit compared with the standard analysis.

Optimization of hydroelectric power generation

Power generating electricity p can be represented

$$p = Chu,$$

where
C — constant;
h — height of the water level, $h \in [h_{min}, h_{max}]$,
u — water passing through the turbine, $u \in [u_{min}, u_{max}]$.

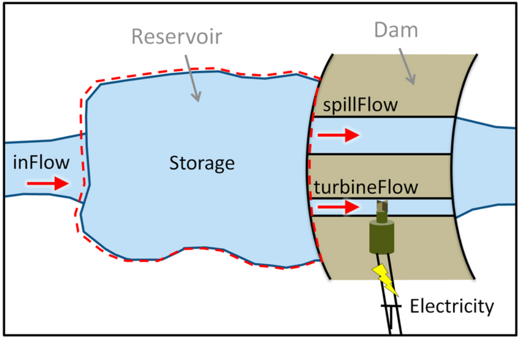


Statement of the problem

Height h depends on the amount of water in the reservoir of V:

$$h = h(V).$$

$$V(t) = V_0 + \int_0^t q(\xi) - u(\xi) - u_x(\xi)d\xi.$$



$q(t)$ — inflow;

$u_x(t)$ — water passing through the spillway;

Statement of the problem

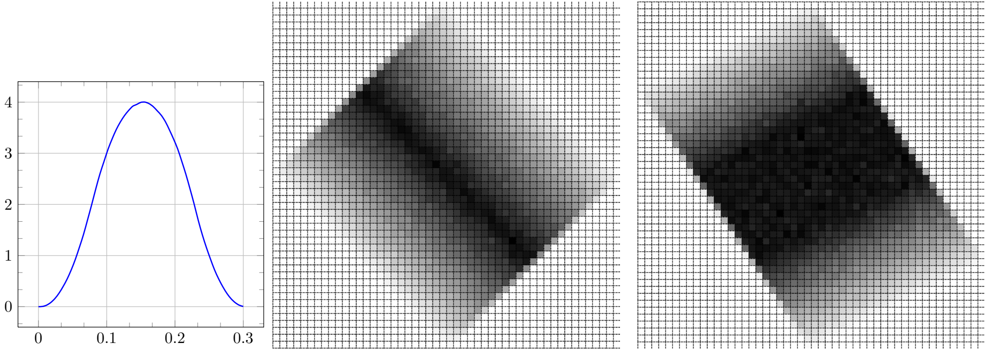
Suppose we want to maximize the generation of electricity in the time interval $[0, T]$. The task of optimal control

$$P(u) = \int_0^T C\ h\left(V_0 + \int_0^T q(t) - u(t) - u_x(t)dt\right)\ u(t)dt \to \max,$$

where u — control.

Numerical example

Let $\boldsymbol{q}_i \in [\underline{q}_i, \overline{q}_i]$ be uniform random variables, $n = 3$, $S = 1$, supports $q_1 = [0.1, 0.2]$, $q_2 = [0.2, 0.3]$, $q_3 = [0.3, 0.4]$ and $h_0 = 0.9$.



Probability density function $\boldsymbol{u}_1$ and joint probability density (u_1, u_2), (u_2, u_3).
Supports $u_1 = [0.0, 0.3]$, $u_2 = [0.25, 0.35]$, $u_3 = [0.575, 0.725]$.

Оценка интенсивности отказов

Интенсивность отказов

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \le \xi < t + \Delta t | t \le \xi)}{\Delta t} = \frac{f(t)}{P(t)} = \frac{f(t)}{1 - F(t)}.$$

$$\lambda(t) = -\frac{P'(t)}{P(t)}. \tag{15}$$

где $P(t)$ — вероятность безотказной работы.

Численное моделирование интенсивности отказов

Пусть $(\xi_1, \xi_2, \ldots, \xi_n)$ статистика отказов полученная опытным путем.
Тогда

$$-\ln(z_i) = \int_0^{\xi_i} \lambda(\xi) d\xi,$$

где $z_i = P(\xi_i)$.

Для нахождения $\lambda(t)$ будем использовать метод наименьших квадратов. Пусть $\varphi_1, \varphi_2, \dots, \varphi_m$ — линейно независимые функции и $\lambda(t)$ будем искать в виде

$$\int_0^t \lambda(\xi) d\xi \approx \sum_{i=1}^{m} a_i \varphi_i(t).$$

Для нахождения $a_1, a_2, \dots, a_m$ рассмотрим функционал

$$\Phi(a_1, \dots, a_m) = \sum_{i=1}^{n} (-\ln(z_i) - \sum_{j=1}^{m} a_j \varphi_j(\xi_i))^2 \to \min .$$

Задача сводится к решению СЛАУ

$$A\vec{a} = b,$$

где $G = (g_{ij})$ — матрица Грама, $\vec{a} = (a_1, a_2, \dots, a_m)$, $b = (b_i)$ $g_{ij} = (\varphi_i, \varphi_j)$, $b_i = (\vec{z}, \varphi_i)$ и

$$(\vec{z})_i = \ln(z_i),$$

$$(x, y) = \sum_{i=1}^{n} x(\xi_i) y(\xi_i).$$

Используя вместо $z_1, z_2, \dots, z_n$ совместную функцию плотности $p(z_1, z_2, \dots, z_n)$, можем построить вероятностное расширение $\lambda(t)$.

Модельный пример

Имеем $(\xi_1, \xi_2, \ldots, \xi_n)$, $n = 29$. Предположим $\lambda(t)$ имеет вид

$$\lambda(t) = \begin{cases} 0.2, & t \in [0, 0.7] \\ 0.2 + 12(t - 0.7) & t > 0.7 \end{cases}$$

Представим $\lambda(t)$ случайную кусочно-линейную функцию

$$\boldsymbol{l}(t) = \sum \boldsymbol{a}_i \psi_i(t),$$

$\{\psi_i\}$ — базис в пространстве кусочно-линейных функций.

$$\int_0^t \lambda(\xi) d\xi \approx \sum_{i=1}^{m} a_i \varphi_i(t),$$

где

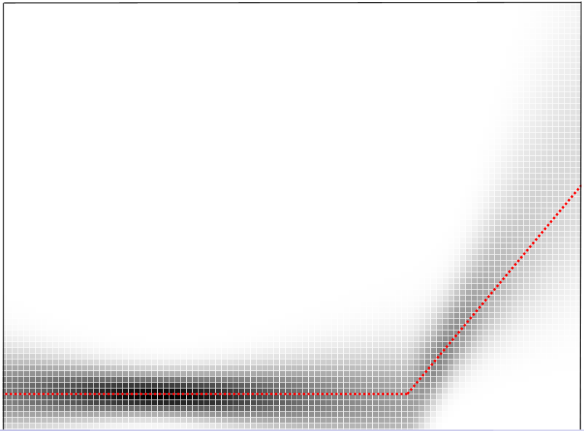
$$\varphi_i(t) = \int_0^t \psi_i(t) dt.$$

В силу детерминированности матрицы МНК G и того факта, что компоненты вектора $\boldsymbol{b}$ есть линейные комбинации $\ln z_i$, вектор $\vec{\boldsymbol{a}} = G^{-1}\boldsymbol{b}$ можно выразить как линейную комбинацию $\ln z_i$.

Достоверные оценки

$$\lambda(t) \approx \sum_{i=1}^{n} \gamma_i(t) \ln z_i,$$

где $\gamma_i(t)$ — некоторые вещественные функции. Используя вероятностные расширения, получаем



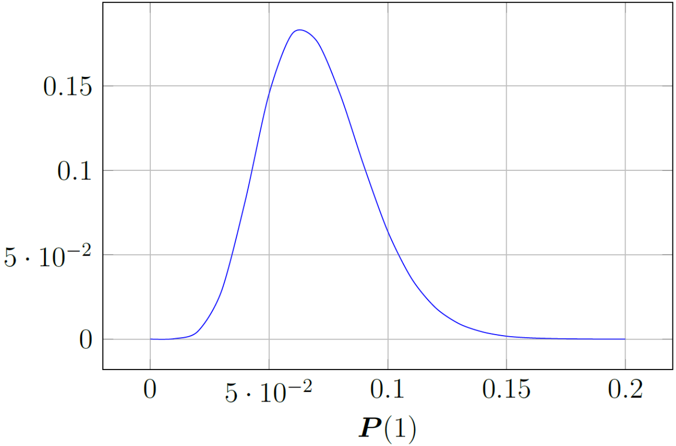


Рис. 5

Using probabilistic extension of λ can calculate the probability density function evaluation values $\boldsymbol{P}(t)$ at any time. Figure shows the spline evaluation of probability density function $\boldsymbol{P}(t)$ at the time $t = 1$.

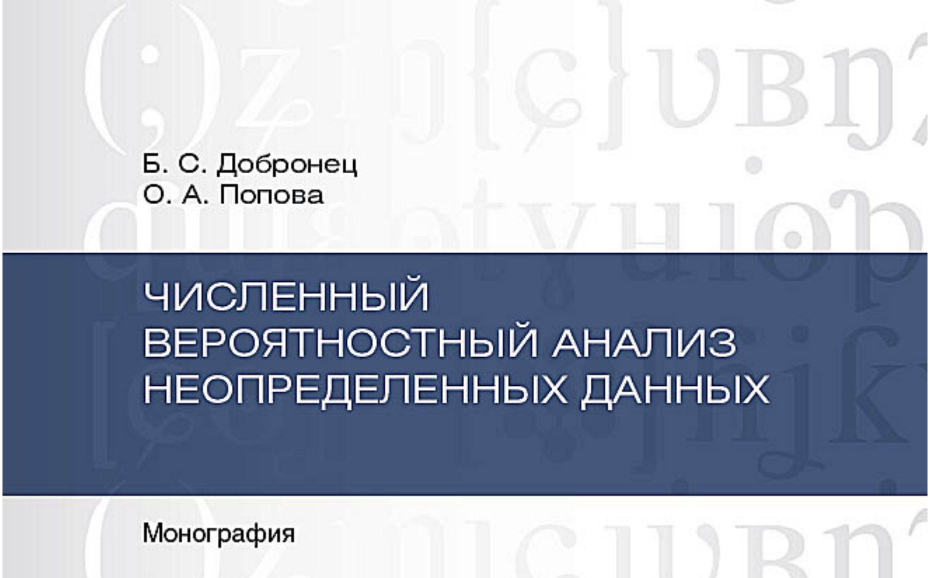
Литература

⋆ Dobronets B.S.& Popova O.A. (2019) Computational Aspects of Probabilistic Extensions. // Tomsk State University Journal of Control and Computer Science. 47. pp. 41-48 DOI: 10.17223/19988605/47/5
Dobronets, B. & Popova, O. (2016) Numerical Probabilistic Approach for Optimization Problems. Scientific Computing, Computer Arithmetic, and Validated Numerics. Lecture Notes in Computer Science. Vol. 9553. Springer International Publishing, Cham. pp. 43-53.
Dobronets, B.S. & Popova, O.A. (2017) Improving the accuracy of the probability density function estimation. Journal of Siberian Federal University, Mathematics and Physics. 10(1). pp. 16-21. DOI: 10.17516/1997-1397-2017-10-1-16-21
Dobronets, B.S. & Popova, O.A. (2018) Piecewise Polynomial Aggregation as Preprocessing for Data Numerical Modeling. IOP Conf. Series: Journal of Physics: Conf. Series. 1015. DOI:10.1088/1742-6596/1015/3/032028
Dobronets, B.S. & Popova, O.A. (2018) Improving reliability of aggregation, numerical simulation and analysis of complex systems by empirical data. IOP Conf. Series: Materials Science and Engineering. 354. DOI:10.1088/1757-899X/354/1/012006

Литература


Б. С. Добронец, О. А. Попова
СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
SIBERIAN FEDERAL UNIVERSITY
Изложен подход к использованию вычислительного вероятностного анализа для решения задач с неопределенными входными данными. Основное внимание уделено процессу обработки, представления, моделирования и анализа информации для разных типов неопределенности. Рассмотрены различные математические модели и численные методы их обработки, вопросы надежности результатов численного моделирования для разнообразных задач в условиях ограниченного и большого объемов информации. Даны примеры применения рассматриваемого подхода для практических задач цифровой экономики, надежности технических систем и оборудования. Разработанные алгоритмы могут быть использованы для исследования сложных систем с входными данными, обусловленными различными типами неопределенности.
ISBN 978-5-7638-4232-6
ВЫЧИСЛИТЕЛЬНЫЙ ВЕРОЯТНОСТНЫЙ АНАЛИЗ: МОДЕЛИ И
Б. С. Добронец
О. А. Попова
ВЫЧИСЛИТЕЛЬНЫЙ
ВЕРОЯТНОСТНЫЙ АНАЛИЗ:
МОДЕЛИ И МЕТОДЫ

References


Б. С. Добронец
О. А. Попова
ЧИСЛЕННЫЙ ВЕРОЯТНОСТНЫЙ АНАЛИЗ НЕОПРЕДЕЛЕННЫХ ДАННЫХ
Монография