

Построение орбитальных структур статистических взаимодействий случайных событий как метод динамической визуализации вероятностных распределений

Е.Е. Голденюк

Красноярский государственный торгово-экономический институт
ул. Лиды Прушинской 2, Красноярск, 660075
email: golde@rambler.ru

МЕТРИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ КОВАРИАЦИИ СТАТИСТИЧЕСКИХ ЗАВИСИМОСТЕЙ СЛУЧАЙНЫХ СОБЫТИЙ

Традиционной мерой зависимости в теории вероятностей и математической статистике служит ковариация. Для пары случайных величин ξ и η ковариация определяется как центральный смешанный момент второго порядка:

$$\text{Cov}(\xi, \eta) = \mathbf{E}(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta).$$

Недавно [3] была дана метрическая интерпретация *ковариации случайных величин* в виде

$$2\text{Cov}(\xi, \eta) = \mathbf{E}^i(\xi - \eta)^2 - \mathbf{E}(\xi - \eta)^2$$

— разности средних квадратов расстояний между ними при общем $(\mathbf{E}(\xi - \eta)^2)$ и независимом $(\mathbf{E}^i(\xi - \eta)^2)$ совместном распределении. *Ковариация случайных событий*, определяемая как величина

$$\text{Kov}_{xy} = \mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y),$$

и являющаяся мерой зависимости случайных событий (нулевая ковариация эквивалентна независимости случайных событий), также может быть метрически интерпретирована как

$$2\text{Kov}_{xy} = \mathbf{P}^i(x \triangle y) - \mathbf{P}(x \triangle y)$$

— разность средних расстояний между событиями при общем: $\mathbf{P}(x \triangle y)$ и независимом: $\mathbf{P}^i(x \triangle y)$ совместном распределении. Аналогия здесь очевидна, так как

$$\mathbf{P}(x \triangle y) = \mathbf{E}(\mathbf{1}_x - \mathbf{1}_y)^2, \quad \mathbf{P}^i(x \triangle y) = \mathbf{E}^i(\mathbf{1}_x - \mathbf{1}_y)^2$$

— вероятность симметрической разности двух событий x и y — это средний квадрат расстояния между их индикаторами — случайными величинами $\mathbf{1}_x$ и $\mathbf{1}_y$.

Лемма 1 (метрическая интерпретация ковариации статистической зависимости случайных событий). *Ковариацию Kov_{xy} двух случайных событий x и y можно представить как*

$$2\text{Kov}_{xy} = \mathbf{P}^i(x \triangle y) - \mathbf{P}(x \triangle y)$$

— разность средних расстояний между ними при независимом и общем распределении.

Лемма 2 (о связи между ковариацией случайных событий и ковариацией их индикаторов). *Ковариация случайных событий — это ковариация их индикаторов:*

$$\text{Kov}_{xy} = \text{Cov}(\mathbf{1}_x, \mathbf{1}_y).$$

Лемма 3. *Среднее расстояние между случайными событиями — это средний квадрат расстояния между их индикаторами:*

$$\mathbf{P}(x \triangle y) = \mathbf{E}(\mathbf{1}_x - \mathbf{1}_y)^2, \quad \mathbf{P}^i(x \triangle y) = \mathbf{E}^i(\mathbf{1}_x - \mathbf{1}_y)^2.$$

СИЛЫ СТАТИСТИЧЕСКИХ ВЗАИМОДЕЙСТВИЙ СЛУЧАЙНЫХ СОБЫТИЙ

Идея метрической интерпретации ковариации случайных событий возникла из естественного желания визуализировать их расположение друг относительно друга. А поскольку в пространстве событий нет никаких структур, кроме структур, "навязываемых" вероятностью, идея метрической интерпретации ковариации случайных событий оказалась единственно продуктивной. Дело в том, что в соответствии с метрической интерпретацией ковариации случайных событий,

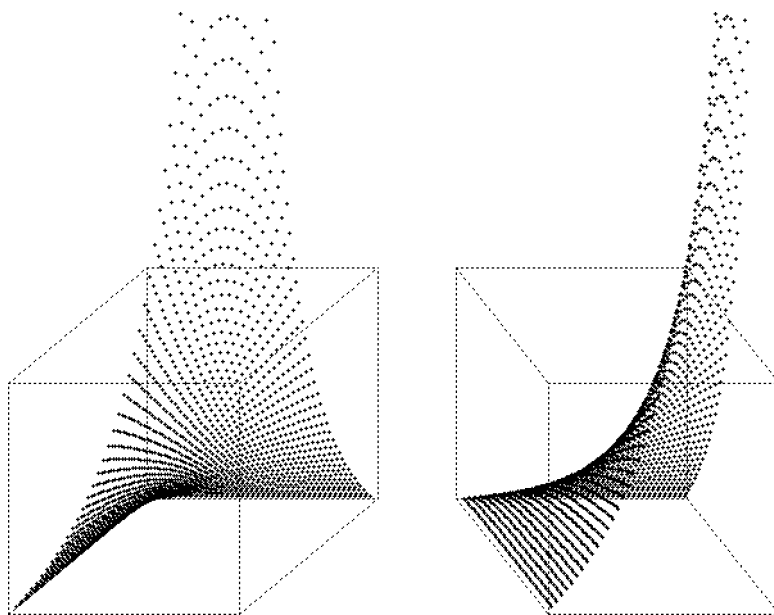


Рис. 1: Независимые события взаимодействуют: график силы взаимодействия двух независимых случайных событий x и y как функции их вероятностей p_x и p_y в двух ракурсах.

если ковариация между ними положительна, то эти события статистически зависимы и притягиваются, а если ковариация отрицательна — события зависимы и отталкиваются.

Нулевая ковариация эквивалентна независимости событий. Изменение расстояния между случайными событиями в общем случае по сравнению с независимым можно объяснить тем, что между событиями происходит *взаимодействие*, напоминающее физическое взаимодействие. По аналогии с физическими законами взаимодействия тел (закон всемирного тяготения Ньютона в гравитационном поле) и взаимодействия зарядов (закон Кулона в электрическом поле) в работе предлагается ввести новое понятие *силы статистического взаимодействия случайных событий*:

$$\text{For}_{xy} = \frac{\mathbf{P}(x)\mathbf{P}(y)}{\mathbf{P}(x\Delta y)}.$$

В качестве аналога масс предлагаются вероятности событий, а в качестве аналога расстояния между событиями — вероятность их симметрической разности. Необходимо отметить одну важную особенность нового понятия. *Сила статистического взаимодействия и ковариация статистической зависимости случайных событий связаны*

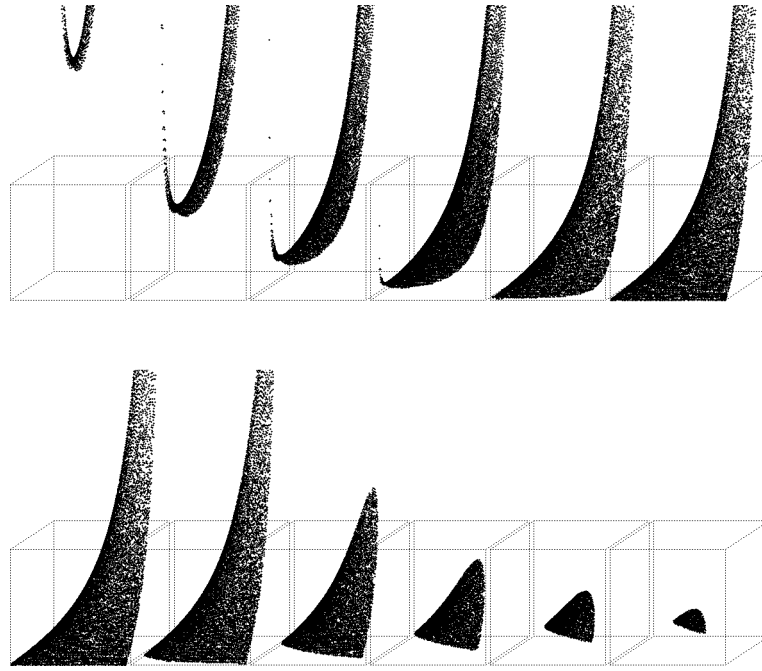


Рис. 2: Сила взаимодействия двух случайных событий как функция их вероятностей p_x и p_y при фиксированной ковариации Kov_{xy} имеет седловую точку $p_x = p_y = 2\text{Kov}_{xy}$. Показаны (слева направо) графики силы взаимодействия двух случайных событий x и y как функции их вероятностей p_x и p_y (в пределах границ Фреше) при фиксированных значениях ковариации: $\text{Kov}_{xy} = 0.20, 0.15, 0.10, 0.05, 0.01, 0.00$ (верхний ряд), $\text{Kov}_{xy} = 0.00, -0.01, -0.05, -0.10, -0.15, -0.20$ (нижний ряд).

очевидным соотношением

$$\text{For}_{xy} = \frac{\mathbf{P}(x)\mathbf{P}(y)}{\mathbf{P}^i(x\Delta y) - 2\text{Kov}_{xy}},$$

из которого следует непривычный факт — даже независимые события (с нулевой ковариацией) статистически взаимодействуют между собой с силой

$$\text{For}_{xy}^i = \mathbf{P}(x)\mathbf{P}(y)/\mathbf{P}^i(x\Delta y)$$

(см. рис. 1). Или, иначе говоря, сила статистического взаимодействия независимых событий For_{xy}^i удерживает их на "независимом" расстоянии $\mathbf{P}^i(x\Delta y)$.

Лемма 4 (о седловой точке силы статистического взаимодействия (рис. 2)). При фиксированной положительной ковариации Kov_{xy} сила статистического взаимодействия For_{xy} как функция вероятностей событий $\mathbf{P}(x)$ и $\mathbf{P}(y)$ имеет седловую точку всякий раз, когда

$$\mathbf{P}(x) = \mathbf{P}(y) = 2\text{Kov}_{xy} = \mathbf{P}^i(x\Delta y) - \mathbf{P}(x\Delta y)$$

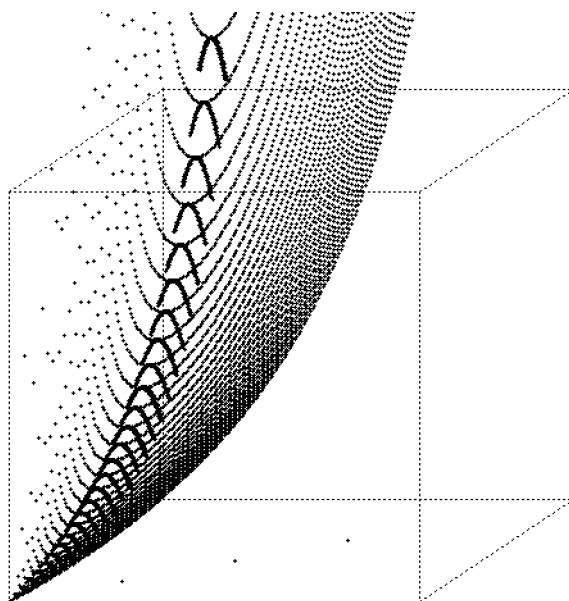


Рис. 3: Седловые точки силы взаимодействия двух случайных событий как функция их вероятностей p_x и p_y при фиксированных неотрицательных ковариациях Kov_{xy} . При возрастании ковариации в пределах границ Фреше сила взаимодействия возрастает, чтобы устремиться к бесконечности в точке $[1/2, 1/2]$, где ковариация достигает максимально разрешенного ей значения: $\text{Kov}_{xy} = 1/4$.

— вероятности событий равны между собой и совпадают с разностью средних расстояний между ними в независимом и общем распределении.

Следствие 1 (о значении силы статистического взаимодействия в седловой точке (рис. 3)). При фиксированной положительной ковариации Kov_{xy} сила статистического взаимодействия For_{xy} как функция вероятностей событий $\mathbf{P}(x)$ и $\mathbf{P}(y)$ принимает в седловой точке значение, равное

$$\text{For}_{xy} = \frac{2\text{Kov}_{xy}}{1 - 4\text{Kov}_{xy}}. \quad (*)$$

Следствие 2.

$$\text{For}_{xy} \rightarrow \infty, \quad \text{при} \quad \text{Kov}_{xy} \rightarrow 1/4. \quad (**)$$

Следствие 3 (метрическая интерпретация силы статистического взаимодействия случайных событий).

$$\text{For}_{xy} = \frac{\mathbf{P}^i(x \triangle y) - \mathbf{P}(x \triangle y)}{1 - 2[\mathbf{P}^i(x \triangle y) - \mathbf{P}(x \triangle y)]}. \quad (***)$$

Сила статистического взаимодействия случайных величин. Новое понятие силы статистического взаимодействия случайных событий позволило, во-первых, обобщить его и ввести аналогичное понятие для случайных величин:

$$\text{For}(\xi, \eta) = \frac{\mathbf{E}(\xi)\mathbf{E}(\eta)}{\mathbf{E}(\xi - \eta)^2},$$

где роль масс случайных величин играют их математические ожидания, и, во-вторых, сформулировать задачу N случайных событий, которой посвящен следующий раздел.

Свойства силы статистического взаимодействия случайных величин. Очевидно, когда случайные величины имеют нулевое математическое ожидание (т.е. когда они центрированы) — сила их статистического взаимодействия по определению равна нулю. Итак, центрирование случайных величин ”зануляет” силу их взаимодействия.

Математическое ожидание $\mathbf{E}\xi$ — это характеристика положения случайной величины, дисперсия $\mathbf{D}\xi = \mathbf{E}(\xi - \mathbf{E}\xi)^2$ — это характеристика её разброса. Сила статистического взаимодействия случайных величин определяется как отношение произведения характеристик положения к характеристике разброса: $\mathbf{E}(\xi - \eta)^2$ — *среднему квадрату расстояния между с.в.*

Заметим, что из метрической интерпретации ковариации случайных величин следует, что

$$\mathbf{E}(\xi - \eta)^2 = \mathbf{E}^i(\xi - \eta)^2 - 2\text{Cov}(\xi, \eta).$$

А поскольку

$$\begin{aligned} \mathbf{E}^i(\xi - \eta)^2 &= \mathbf{E}\xi^2 + \mathbf{E}\eta^2 - 2\mathbf{E}\xi\mathbf{E}\eta = \\ &= \mathbf{E}\xi^2 - (\mathbf{E}\xi)^2 + \mathbf{E}\eta^2 - (\mathbf{E}\eta)^2 - 2\mathbf{E}\xi\mathbf{E}\eta + (\mathbf{E}\xi)^2 + (\mathbf{E}\eta)^2 = \\ &= \mathbf{D}\xi + \mathbf{D}\eta + (\mathbf{E}\xi - \mathbf{E}\eta)^2, \end{aligned}$$

то получаем эквивалентную запись формулы, явно выражающую силу статистического взаимодействия через характеристики положения и характеристики разброса:

$$\text{For}_{xy} = \frac{\mathbf{E}\xi\mathbf{E}\eta}{\mathbf{D}\xi + \mathbf{D}\eta - 2\text{Cov}(\xi, \eta) + (\mathbf{E}\xi - \mathbf{E}\eta)^2}.$$

Из этой записи видно, что в знаменателе находятся не только характеристики разброса (дисперсии: $\mathbf{D}\xi$, $\mathbf{D}\eta$ и ковариация $\text{Cov}(\xi, \eta)$, которые не меняются при центрировании с.в.), но и относительная характеристика их средних положений:

$$(\mathbf{E}\xi - \mathbf{E}\eta)^2$$

— *квадрат расстояния между средними случайных величин* (математическими ожиданиями), который не обязан совпадать со *средним квадратом расстояний между с.в.* $\mathbf{E}(\xi - \eta)^2$.

Таким образом, проясняется следующее:

- новое понятие
 - сила статистического взаимодействия случайных величин
 - это *мера статистического взаимодействия их положений* (зависит от центрирования), в то время как
- старое понятие
 - ковариация статистической зависимости случайных величин
 - это *мера статистической зависимости их разброса* (не зависит от центрирования).

Задача N случайных событий

Предложенная в работе метрическая интерпретация ковариации статистической зависимости случайных событий как разности средних расстояний между ними в независимом и общем распределении, а также введенное в работе новое понятие силы, как меры статистического попарного взаимодействия случайных событий, привело к идее динамической визуализации случайного множества событий посредством численного решения так называемой задачи N случайных событий.

Действительно, рассмотрим N случайных событий, вероятностное распределение которых определяется случайным множеством наступивших событий

$$K : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (2^{\mathcal{X}}, 2^{2^{\mathcal{X}}})$$

под конечным множеством событий $\mathfrak{X} \subseteq \mathcal{F}$, состоящим из заданных $N = |\mathfrak{X}|$ событий. Для каждой пары событий x и y из \mathfrak{X} сила их статистического парного взаимодействия определена формулой

$$\text{For}_{xy} = \frac{\mathbf{P}(x)\mathbf{P}(y)}{\mathbf{P}(x\Delta y)}.$$

Если эти события поместить (визуализировать) на евклидову плоскость \mathbf{R}^2 , сопоставив каждому событию x и его дополнению x^c некоторые точки плоскости $\mathbf{r}_x \in \mathbf{R}^2$ и $\mathbf{r}_{x^c} \in \mathbf{R}^2$, которые естественно назвать *визуализационными точками*, или *положениями событий на визуализационной плоскости*, то возникает картина, напоминающая известную классическую задачу N тел в физике:

”На плоскости \mathbf{R}^2 визуализированы N случайных событий x , имеющие вероятностные массы $\mathbf{P}(x)$, для каждой пары которых определены величина For_{xy} силы их статистического парного взаимодействия и её направление, определяемое вектором $\mathbf{r}_x - \mathbf{r}_y$. Требуется найти динамику движения положений событий на визуализационной плоскости под действием заданных сил статистического парного взаимодействия”.

Физическая задача N тел приводит к системе дифференциальных уравнений второго порядка:

$$m_i \frac{d^2 x_i}{dt^2} = - \sum_{j \neq i} \frac{m_i m_j (x_i - x_j)}{r_{ij}^3}, \quad i = 1, \dots, N. \quad (1)$$

Еще Пуанкаре показал, что для $N > 2$ задача не имеет общего решения. Давно известны классические частные решения Эйлера (1765) и Лагранжа (1772) для задачи трёх тел (рис. 4 вверху). Недавно [1] американским математиком Ричардом Монтгомери (2001) и его коллегами было получено еще одно частное решение, в котором три тела одинаковой массы, расположенные на одинаковом расстоянии друг от друга, движутся по восьмерке (рис. 4 внизу). Это говорит о том, что поиск частных решений физической задачи N тел чрезвычайно актуален до сих пор.

По аналогии с задачей N тел сформулируем задачу N случайных событий, которая, на наш взгляд, актуальна не менее её физического прототипа.

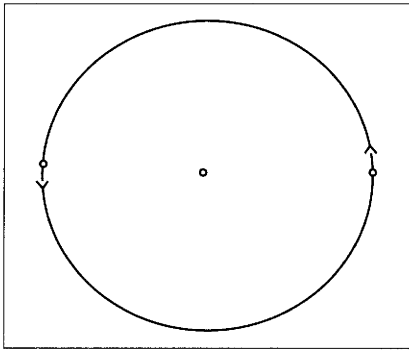


Figure 1. Euler's solution in the equal mass case.

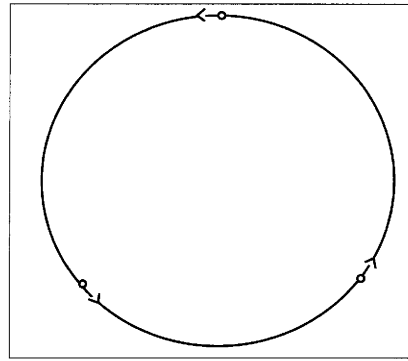


Figure 2. Lagrange's solution in the equal mass case.

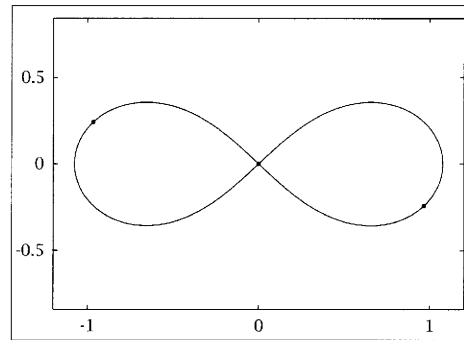


Figure 3. The figure-eight solution.

Рис. 4: Два классических (Эйлер, 1765) и (Лагранж, 1772) и одно новое (Монтгомери и др., 2001) частные решения физической задачи трёх тел.

Формулировка задачи N случайных событий. Пусть $x \in \mathfrak{X}$ — произвольное событие из конечного множества \mathfrak{X} , которое ”находится под управлением” случайного множества наступивших событий K , а x^c — его дополнение. Обозначим

$$\mathbf{r}_x(t), \quad \mathbf{r}_{x^c}(t)$$

— вектора положения этих событий на визуализационной плоскости в момент t . Для сокращения записи введем несколько искусственное, но зато более общее обозначение

$$z = \begin{cases} x \\ x^c \end{cases}$$

— для события $x \in \mathfrak{X}$ и его дополнения $x^c \in \mathfrak{X}^c$, подразумевая, что $z \in Z$, где $Z = \mathfrak{X} \cup \mathfrak{X}^c$ — множество, содержащее все события и их дополнения, соответствующее конечному множеству событий \mathfrak{X} . Тогда визуализация K обеспечивает каждому $z \in Z$ положение на визуализа-

ционной (видимой) плоскости \mathbf{R}^2 , которое будет обозначаться

$$\mathbf{r}_z(t).$$

Все результаты, полученные в работе, могут быть использованы и в случае *трехмерной визуализации*. Только тогда в качестве "экрана" визуализации надо использовать "трёхмерный экран" — *видимое пространство* \mathbf{R}^3 .

Сформулируем задачу N случайных событий, как задачу Коши поиска решения системы дифференциальных уравнений второго порядка:

$$\mathbf{P}(z) \frac{d^2 \mathbf{r}_z}{dt^2} = - \sum_{z'} \text{For}_{zz'} [\mathbf{r}_z - \mathbf{r}_{z'}] \quad (2)$$

при заданных начальных условиях. Задача формулируется для случайных событий $z \in \{x, x^c\}$, $x \in \mathfrak{X}$, где

$$\text{For}_{zz'} = \frac{\mathbf{P}(z)\mathbf{P}(z')}{\mathbf{P}(z \triangle z')}$$

— сила статистического парного взаимодействия случайных событий z и z' , находящихся на среднем расстоянии

$$\mathbf{P}(z \triangle z').$$

Сходство и отличие задачи N тел и задачи N случайных событий. При почти буквальном внешнем сходстве физическая система (1) и система (2) имеют существенно разное содержательное толкование своих составных частей. Физическая система, на самом деле, имеет вид

$$m_i \frac{d^2 x_i}{dt^2} = - \sum_j \frac{m_i m_j}{|x_i - x_j|^3} (x_i - x_j), \quad i = 1, \dots, N,$$

из которого ясно, что расстояния

$$r_{ij} = |x_i - x_j|$$

между телами i и j входят в решение данной системы дифференциальных уравнений и изменяются с течением времени так, как этого

требует решение. Совсем по другому дело обстоит с системой (2), в которой средние расстояния

$$P(z\Delta z')$$

между случайными событиями z и z' , определяющие силу их статистического парного взаимодействия

$$\text{For}_{zz'} = \frac{P(z)P(z')}{P(z\Delta z')}$$

не входят в решение системы, а являются её исходными параметрами, определяемыми распределением случайного множества K , — это исходные константы данной системы дифференциальных уравнений. Поэтому имеются все основания говорить о своеобразной задаче, которую по аналогии с физической *задачей N тел* мы назвали *задачей N случайных событий*, памятуя, что если физика интересуется *взаимодействиями физических тел*, то для теории случайных событий интересны *статистические взаимодействия случайных событий*.

Проблемы решения задачи N случайных событий. Задача N случайных событий в приведенной постановке имеет отличие от физической задачи N тел, которое заключается в том, что силы статистического взаимодействия не меняются со временем. Это приводит к системе дифференциальных уравнений второго порядка с постоянными коэффициентами, которая имеет известное аналитическое решение задачи Коши. Однако, в работе внимание сосредоточено, главным образом, на методах численного решения задачи N случайных событий, основное преимущество которых заключается в том, что они могут без изменения применяться также и для *переменных сил статистического взаимодействия* (именно этот случай, выходящий за пределы данной работы и являющийся полным аналогом физической задачи N тел, будет интересовать нас в дальнейшем).

Задача N случайных событий в простейших вариантах. Среди случайных событий, фигурирующих в задаче N случайных событий и образующих конечное множество событий \mathcal{X} , можно выделить одно событие, играющее уникальную роль — это достоверное событие Ω . В

работе рассматривается такая задача N случайных событий, для которой исходное конечное множества событий \mathfrak{X} выбирается так, что в него всегда включается достоверное событие Ω . Таким образом, фактически в задаче рассматривается $N + 1$ событие, среди которых N случайных (имеющих вероятность строго между нулем и единицей) и одно — достоверное: Ω , имеющее единичную вероятность: $\mathbf{P}(\Omega) = 1$. Невозможное событие \emptyset , как дополнение достоверного, также формально рассматривается в задаче, но из-за нулевой вероятностной массы: $\mathbf{P}(\emptyset) = 0$, силы его статистического взаимодействия со всеми остальными событиями равны нулю: $\text{For}_{\emptyset x} \equiv 0$. Поэтому фактически невозможное событие не участвует в рассматриваемой системе статистических взаимодействий.

Поскольку роль достоверного события Ω в задаче N случайных событий уникальна, в работе ищется решение этой задачи относительно положения достоверного события на визуализационной плоскости, которое в результате оказывается неподвижно размещенным в начале координат. Относительное положение произвольного события $z \in Z$ на визуализационной плоскости имеет вид

$$\mathbf{r}_z = \bar{\mathbf{r}}_z - \bar{\mathbf{r}}_\Omega,$$

где $\bar{\mathbf{r}}_z$ — его абсолютное положение. Причем $\mathbf{r}_\Omega \equiv \mathbf{0}$ — достоверное событие расположено неподвижно в начале координат.

Таким образом, общую систему дифференциальных уравнений для задачи N случайных событий, в которой выделено достоверное событие, можно переписать в относительной форме. Выделяется уравнение для взаимодействия достоверного события, которое из дифференциального превращается в уравнение связи между векторами положений случайных событий:

$$0 = \sum_z \text{For}_{\Omega z} \cdot \mathbf{r}_z.$$

Остальные дифференциальные уравнения для $z \neq \Omega$ выглядят по-прежнему:

$$\mathbf{P}(z) \frac{d^2 \mathbf{r}_z}{dt^2} = - \sum_{z'} \text{For}_{zz'} [\mathbf{r}_z - \mathbf{r}_{z'}].$$

Задача одного случайного события. Рассмотрим самую простейшую задачу N случайных событий, в которой $N = 1$ и конечное множество событий \mathfrak{X} после добавления достоверного события состоит из двух событий:

$$\mathfrak{X} = \{x, \Omega\}.$$

Этой задаче соответствует следующая система трёх уравнений, из которых два дифференциальных:

$$\begin{aligned} 0 &= \text{For}_{\Omega x} \cdot \mathbf{r}_x + \text{For}_{\Omega x^c} \cdot \mathbf{r}_{x^c}, \\ \mathbf{P}(x) \frac{d^2 \mathbf{r}_x}{dt^2} &= -\text{For}_{x\Omega} \cdot \mathbf{r}_x - \text{For}_{xx^c} [\mathbf{r}_x - \mathbf{r}_{x^c}], \\ \mathbf{P}(x^c) \frac{d^2 \mathbf{r}_{x^c}}{dt^2} &= -\text{For}_{x^c\Omega} \cdot \mathbf{r}_{x^c} - \text{For}_{x^c x} [\mathbf{r}_{x^c} - \mathbf{r}_x]. \end{aligned}$$

Первое уравнение связи позволяет выразить визуализационное положение \mathbf{r}_x события x через визуализационное положение \mathbf{r}_{x^c} события x^c и наоборот. Поэтому данная система сводится к двум системам:

$$\left\{ \begin{aligned} \mathbf{P}(x) \frac{d^2 \mathbf{r}_x}{dt^2} &= -\mathbf{r}_x [\text{For}_{x\Omega} + \text{For}_{xx^c} + \text{For}_{xx^c} \cdot \text{For}_{\Omega x} / \text{For}_{\Omega x^c}], \\ \mathbf{r}_{x^c} &= -\mathbf{r}_x \text{For}_{\Omega x} / \text{For}_{\Omega x^c}. \end{aligned} \right. \quad (*)$$

$$\left\{ \begin{aligned} \mathbf{P}(x^c) \frac{d^2 \mathbf{r}_{x^c}}{dt^2} &= -\mathbf{r}_{x^c} [\text{For}_{x^c\Omega} + \text{For}_{x^c x} + \text{For}_{x^c x} \cdot \text{For}_{\Omega x^c} / \text{For}_{\Omega x}], \\ \mathbf{r}_x &= -\mathbf{r}_{x^c} \text{For}_{\Omega x^c} / \text{For}_{\Omega x}. \end{aligned} \right. \quad (**)$$

Отметим, что по определению силы статистического взаимодействия событий, участвующих в задаче, равны:

$$\text{For}_{\Omega x} = \text{For}_{x\Omega} = \frac{\mathbf{P}(x)}{\mathbf{P}(x^c)}, \quad \text{For}_{\Omega x^c} = \text{For}_{x^c\Omega} = \frac{\mathbf{P}(x^c)}{\mathbf{P}(x)},$$

$$\text{For}_{xx^c} = \text{For}_{x^c x} = \mathbf{P}(x)\mathbf{P}(x^c)$$

и образуют матрицу сил статистического взаимодействия

$$\begin{bmatrix} \text{---} & \text{For}_{\Omega x} & \text{For}_{\Omega x^c} \\ \text{For}_{x\Omega} & \text{---} & \text{For}_{xx^c} \\ \text{For}_{x^c\Omega} & \text{For}_{x^cx} & \text{---} \end{bmatrix} = \begin{bmatrix} \text{---} & \frac{p_x}{p_{x^c}} & \frac{p_{x^c}}{p_x} \\ \frac{p_x}{p_{x^c}} & \text{---} & p_x p_{x^c} \\ \frac{p_{x^c}}{p_x} & p_x p_{x^c} & \text{---} \end{bmatrix},$$

где использованы обычные в работе обозначения для вероятностей событий $p_x = \mathbf{P}(x)$ и $p_{x^c} = \mathbf{P}(x^c)$.

Поэтому системы (*) и (**) принимают вид

$$\begin{cases} \frac{d^2 \mathbf{r}_x}{dt^2} = -\mathbf{r}_x \frac{1 + p_x^2 + p_{x^c}^2}{p_{x^c}^2}, \\ \mathbf{r}_{x^c} = -\mathbf{r}_x \left(\frac{p_x}{p_{x^c}} \right)^2, \end{cases} \quad \begin{cases} \frac{d^2 \mathbf{r}_{x^c}}{dt^2} = -\mathbf{r}_{x^c} \frac{1 + p_x^2 + p_{x^c}^2}{p_x^2}, \\ \mathbf{r}_x = -\mathbf{r}_{x^c} \left(\frac{p_{x^c}}{p_x} \right)^2. \end{cases}$$

Вводя сокращающие обозначения для неотрицательных коэффициентов:

$$a_x^2 = \frac{1 + p_x^2 + p_{x^c}^2}{p_{x^c}^2}, \quad b_x^2 = \left(\frac{p_x}{p_{x^c}} \right)^2,$$

получим краткую запись этих же систем в виде

$$\begin{cases} \frac{d^2 \mathbf{r}_x}{dt^2} = -\mathbf{r}_x a_x^2, \\ \mathbf{r}_{x^c} = -\mathbf{r}_x b_x^2, \end{cases} \quad \begin{cases} \frac{d^2 \mathbf{r}_{x^c}}{dt^2} = -\mathbf{r}_{x^c} a_{x^c}^2, \\ \mathbf{r}_x = -\mathbf{r}_{x^c} b_{x^c}^2. \end{cases}$$

Известно, что дифференциальные уравнения такого вида описывают незатухающие колебания с собственными частотами $a_x/2\pi$ и $a_{x^c}/2\pi$, а решения соответствующей задачи Коши имеют вид (в не очень корректной векторной записи, которую надо понимать "покомпонентно"):

$$\begin{cases} \mathbf{r}_x = \boldsymbol{\rho}_x \sin(a_x t + \boldsymbol{\theta}_x), \\ \mathbf{r}_{x^c} = -\mathbf{r}_x b_x^2, \end{cases} \quad \begin{cases} \mathbf{r}_{x^c} = \boldsymbol{\rho}_{x^c} \sin(a_{x^c} t + \boldsymbol{\theta}_{x^c}), \\ \mathbf{r}_x = -\mathbf{r}_{x^c} b_{x^c}^2, \end{cases}$$

где $\boldsymbol{\rho}_x$, $\boldsymbol{\theta}_x$ и $\boldsymbol{\rho}_{x^c}$, $\boldsymbol{\theta}_{x^c}$ — пары постоянных векторов, определяемые из начальных условий соответствующих задач Коши, иначе говоря, удов-

летворяющие уравнениям, запись которых также надо понимать "покомпонентно":

$$\begin{cases} \rho_x \sin \theta_x = r_x(0), \\ a_x \rho_x \cos \theta_x = v_x(0), \end{cases} \begin{cases} \rho_{x^c} \sin \theta_{x^c} = r_{x^c}(0), \\ a_x \rho_{x^c} \cos \theta_{x^c} = v_{x^c}(0). \end{cases}$$

Эти покомпонентные уравнения разрешаются в виде

$$\begin{cases} \rho_x^2 = r_x^2(0)[1 + \mathbf{A}_x], \\ \sin^2 \theta_x = [1 + \mathbf{A}_x]^{-1}, \end{cases} \begin{cases} \rho_{x^c}^2 = r_{x^c}^2(0)[1 + \mathbf{A}_{x^c}], \\ \sin^2 \theta_{x^c} = [1 + \mathbf{A}_{x^c}]^{-1}, \end{cases}$$

где ("покомпонентно"):

$$\mathbf{A}_x = \frac{1}{a_x^2} \cdot \frac{v_x^2(0)}{r_x^2(0)}, \quad \mathbf{A}_{x^c} = \frac{1}{a_{x^c}^2} \cdot \frac{v_{x^c}^2(0)}{r_{x^c}^2(0)}.$$

Иллюстрации решения задачи одного случайного событий показаны на рис. 5.

АЛГОРИТМ ЧИСЛЕННОГО РЕШЕНИЯ ЗАДАЧИ N СЛУЧАЙНЫХ СОБЫТИЙ

Алгоритм численного решения задачи N случайных событий является, по-существу, имитацией движения на видимой плоскости случайных событий и их дополнений в результате воздействия на них сил парного статистического взаимодействия. Речь идет о численном алгоритме решения соответствующей задачи Коши с начальными условиями, заданными в момент времени t_0 . В начальный момент времени t_0 на видимой плоскости задаются $2N$ плоских векторов положения и $2N$ плоских векторов нулевых начальных скоростей

$$\mathbf{r}_z^0 = \mathbf{r}_z(t_0), \quad \mathbf{v}_z(t_0) = \mathbf{0}, \quad z \in Z.$$

На каждом шаге алгоритма на основе исходного разностного уравнения

$$\mathbf{P}_{(z)} \frac{\Delta \mathbf{v}_z(t)}{\Delta t} \approx - \sum_{z'} \text{For}_{zz'}(\mathbf{r}_{z'}(t) - \mathbf{r}_z(t)),$$

получают смещенные расположения событий и измененные скорости в следующий момент времени:

$$\mathbf{v}_z(t + \Delta t) = \mathbf{v}_z(t) + \Delta \mathbf{v}_z(t),$$

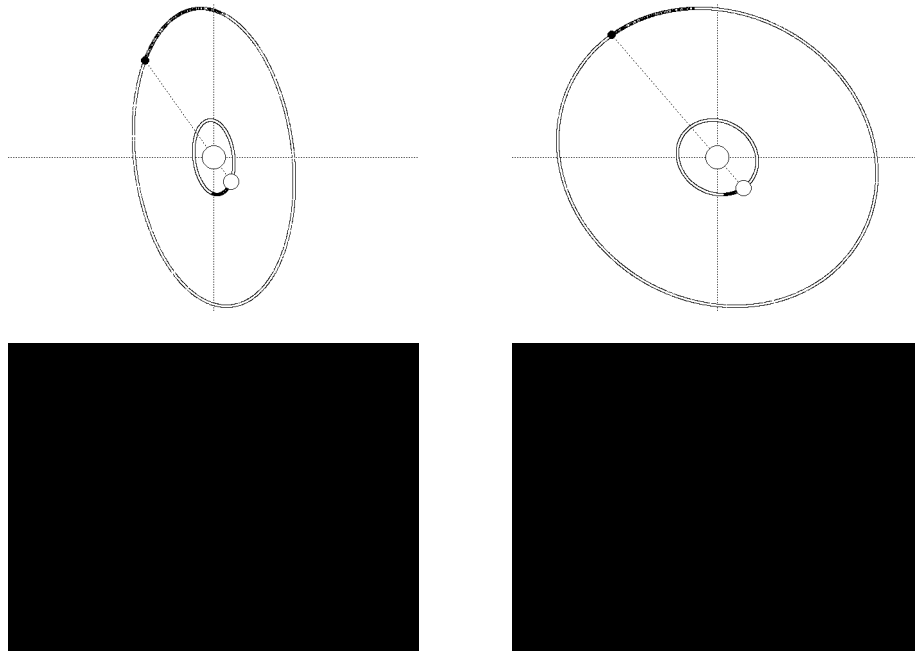


Рис. 5: Иллюстрации аналитического решения задачи одного случайного события. Первая (слева) и вторая (справа) задача Коши с одними и теми же начальными условиями. Достоверное событие Ω расположено в начале координат (большой белый кружок). Верхний вариант: событие x имеет вероятность $p_x = 2/3$ (белый кружок), событие x^c имеет вероятность $p_{x^c} = 1/3$ (черный кружок). Частота колебаний в первой задаче равна $a_x/2\pi \approx 0.34$, во второй задаче — $a_{x^c}/2\pi \approx 0.24$. Нижний вариант: $p_x = 0.4$, $p_{x^c} = 0.6$, $a_x/2\pi \approx 0.25$, $a_{x^c}/2\pi \approx 0.31$.

$$\mathbf{r}_z(t + \Delta t) = \mathbf{r}_z(t) + \mathbf{v}_z(t + \Delta t)\Delta t.$$

Существенной особенностью предложенного в работе алгоритма численного решения является обязательное включение в рассматриваемое конечное множество событий \mathfrak{X} достоверного события $\Omega \in \mathfrak{X}$, вероятностная масса которого равна единице.

Считается, что достоверное событие расположено в начале координат и всегда остается неподвижным. Таким образом, фактически ищется численное решение задачи *относительно* неподвижно расположенного в начале координат *достоверного события*.

С помощью предложенного алгоритма численного решения задачи N случайных событий удалось построить целый ряд частных решений в виде динамических орбитальных структур в различных прикладных областях.

Пример численного решения задачи N случайных событий. Параметры распределения случайного множества событий, необходимые для чис-

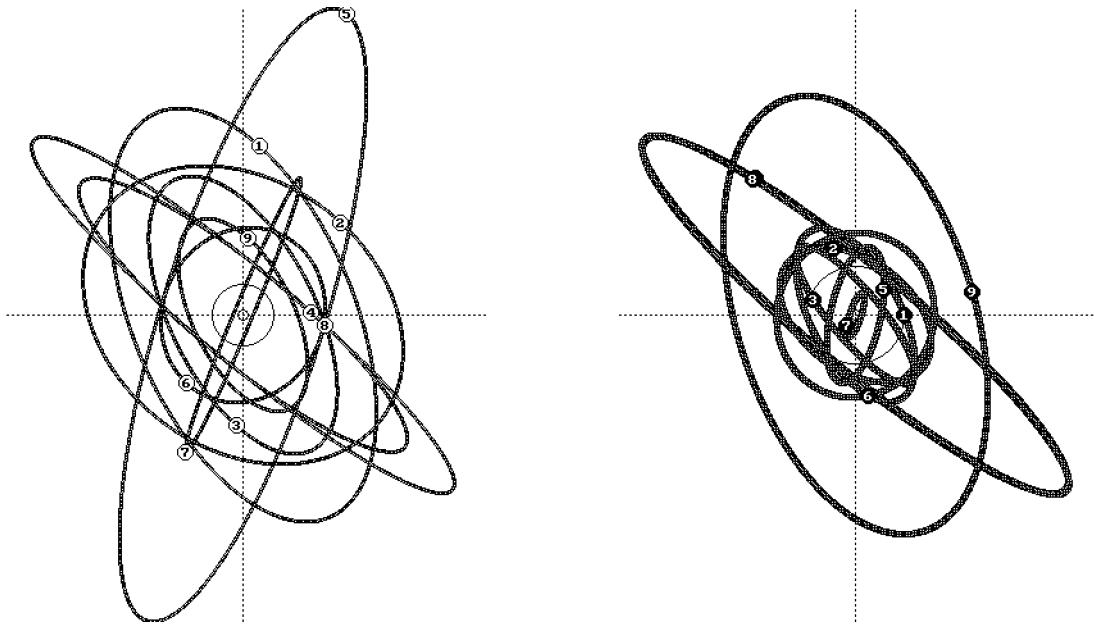


Рис. 6: Орбиты случайных событий, порождённые случайным множеством событий — динамическая визуализация событий из случайного множества под конечным множеством из 9-ти событий. Роль масс взаимодействия случайных событий играют их вероятности. События случайного множества движутся каждый по своей орбите, совершая полный оборот за одно и то же время. Слева орбиты 9-ти событий. Справа в более подробном (3:1) масштабе — 9-ти дополнений событий.

ленного решения задачи N случайных событий, включают вероятности этих событий $\mathbf{P}(x)$, $x \in \mathfrak{X}$, и матрицу сил парного статистического взаимодействия между ними $[\text{For}_{xy}]$, задание которой эквивалентно заданию матрицы ковариаций. Рассмотрим пример численного решения задачи 9-ти случайных событий. Параметры распределения случайного множества событий в этом примере имеют следующий вид.

$$[\mathbf{P}(x), x \in \mathfrak{X}] = [0.39, 0.415, 0.40, 0.37, 0.26, 0.43, 0.30, 0.35, 0.35]$$

— вектор вероятностей событий из \mathfrak{X} ;

$$\begin{bmatrix} \text{---} & 0.122 & -0.054 & -0.127 & 0.117 & 0.097 & 0.022 & -0.047 & 0.014 \\ 0.122 & \text{---} & 0.116 & 0.112 & 0.120 & -0.143 & 0.089 & 0.025 & -0.099 \\ -0.054 & 0.116 & \text{---} & -0.054 & 0.063 & 0.153 & 0.056 & 0.100 & -0.139 \\ -0.127 & 0.112 & -0.054 & \text{---} & -0.083 & -0.152 & 0.033 & 0.164 & -0.050 \\ 0.117 & 0.120 & 0.063 & -0.083 & \text{---} & -0.019 & -0.015 & -0.082 & -0.032 \\ 0.097 & -0.143 & 0.153 & -0.152 & -0.019 & \text{---} & 0.150 & -0.101 & -0.115 \\ 0.022 & 0.089 & 0.056 & 0.033 & -0.015 & 0.150 & \text{---} & -0.053 & -0.021 \\ -0.047 & 0.025 & 0.100 & 0.164 & -0.082 & -0.101 & -0.053 & \text{---} & -0.021 \\ 0.014 & -0.099 & -0.139 & -0.050 & -0.032 & -0.115 & -0.021 & -0.021 & \text{---} \end{bmatrix}$$

— матрица $[Kov_{xy}]$ ковариаций парных зависимостей событий;

$$\begin{bmatrix} \text{---} & 0.680 & 0.266 & 0.199 & 0.475 & 0.577 & 0.283 & 0.243 & 0.311 \\ 0.680 & \text{---} & 0.660 & 0.606 & 0.494 & 0.230 & 0.432 & 0.343 & 0.216 \\ 0.266 & 0.660 & \text{---} & 0.254 & 0.319 & 0.951 & 0.346 & 0.520 & 0.187 \\ 0.199 & 0.606 & 0.254 & \text{---} & 0.159 & 0.202 & 0.291 & 0.971 & 0.231 \\ 0.475 & 0.494 & 0.319 & 0.159 & \text{---} & 0.222 & 0.180 & 0.154 & 0.185 \\ 0.577 & 0.230 & 0.951 & 0.202 & 0.222 & \text{---} & 0.749 & 0.221 & 0.212 \\ 0.283 & 0.432 & 0.346 & 0.291 & 0.180 & 0.749 & \text{---} & 0.193 & 0.218 \\ 0.243 & 0.343 & 0.520 & 0.971 & 0.154 & 0.221 & 0.193 & \text{---} & 0.246 \\ 0.311 & 0.216 & 0.187 & 0.231 & 0.185 & 0.212 & 0.218 & 0.246 & \text{---} \end{bmatrix}$$

— матрица $[For_{xy}]$ сил парных статистических взаимодействий событий;

$$[For_{x\Omega}, x \in \mathfrak{X}] = \left[\frac{P(x)}{P(x^c)}, x \in \mathfrak{X} \right] =$$

$$= [0.64, 0.71, 0.67, 0.51, 0.35, 0.75, 0.43, 0.54, 0.54]$$

— силы статистического взаимодействия случайных событий $x \in \mathfrak{X}$ с достоверным событием.

Результаты численного решения данной задачи 9-ти случайных событий показаны на рис. 6 и имеют вид своеобразной орбитальной структуры.

ОРБИТАЛЬНЫЕ СТРУКТУРЫ СТАТИСТИЧЕСКИХ ВЗАИМОДЕЙСТВИЙ СЛУЧАЙНЫХ СОБЫТИЙ

Предлагаемую *орбитальную структуру* — динамическую визуализацию — статистических взаимодействий систем случайных событий, которая возникает в результате решения задачи N случайных событий, можно по своему значению сравнить с динамической визуализацией ежедневного прогноза погоды в хороших ТВ программах, когда зрителю показывают карту движущихся фронтов циклонов и антициклонов *вместо того, чтобы перекармливать его множеством цифр* температур, давлений, скоростей ветра и другими излишними метеорологическими деталями.

Глядя на движение погодных фронтов зритель сам может составить представление не только о том какая будет погода, и о том, когда и как быстро она будет меняться, но и о том насколько точна вся эта метеоинформация и насколько ей можно доверять.

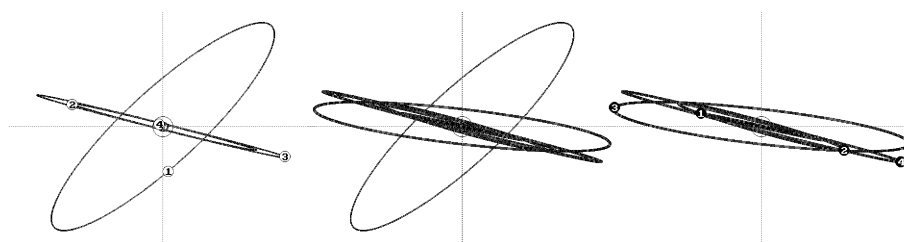


Рис. 7: Численное решение задачи 4-х случайных событий на примере четырёх-товарного рынка напитков — кофе (4), коньяк (3), чай (2) и пиво (1); Показаны орбиты: (слева) — покупки напитков, (в центре) — покупки и "не-покупки", (справа) — "не-покупки".

Ровно таким же образом, глядя на динамику орбитальной структуры взаимодействий N случайных событий, любой пользователь сможет составить общее представление о множестве параметров (их число равно $N(N + 1)/2$), характеризующих эту систему, чтобы с помощью такого визуального образа иметь возможность, например, хотя бы сравнивать общий характер статистического взаимодействия в различных системах событий.

Предлагается использовать численное решение задачи N случайных событий как метод динамической визуализации их взаимодействий в статистических системах, который каждой такой системе сопоставляет её динамический образ в виде *орбитальной структуры* случайных событий на визуализационной плоскости.

Исходными параметрами задачи N случайных событий служат вектор вероятностей событий и матрица сил их парных статистических взаимодействий. Кроме того, необходимо задать начальные положения точек на плоскости, визуализирующих случайные события и их дополнения. Затем применение алгоритма численного решения задачи N случайных событий, позволяет вычислять положения визуализационных точек в следующие моменты времени. В результате на визуализационной плоскости образуется *орбитальная структура взаимодействий*, по орбитам которой вращаются точки — визуализации случайных событий. Таким образом, можно дать следующее

Определение. *Орбитальной структурой статистических взаимодействий N случайных событий на визуализационной плоскости будем*

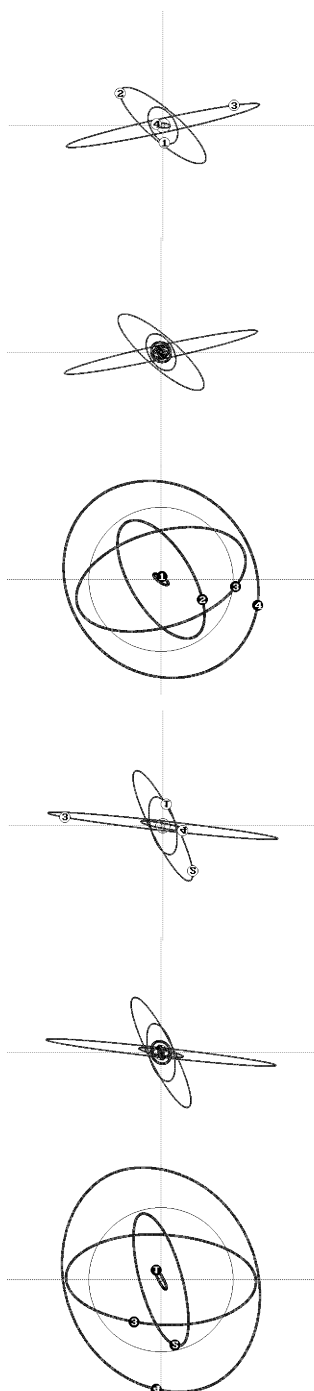
называть численное решение задачи N случайных событий.

В основе орбитальной визуализации случайного множества событий лежит принцип метрической интерпретации ковариации и новое понятие взаимодействия случайных событий. Принцип метрической интерпретации ковариации толкует ковариацию между случайными событиями как меру их статистической близости. При этом положительная ковариация трактуется как сближение случайных событий в случайном множестве, а отрицательная — как их отдаление друг от друга относительно их положения при независимом совместном распределении.

Орбитальная визуализация случайных множеств событий показывает, что каждому случайному множеству событий под конечным множеством \mathcal{X} соответствует некоторая равновесная динамика его случайных событий, которая весьма напоминает динамику электронов в атоме или звезд в галактике. События из случайного множества визуализируются "движущимися" вокруг общего центра по центрально симметричным орбитам (рис. 6). Это, пожалуй, совершенно новый подход в проблеме визуализации, когда в качестве результата визуализации предлагается не неподвижная, а "движущаяся картинка" на плоскости или в пространстве.

Применение понятия *орбитальной структуры* взаимодействий случайных событий открывает совершенно новый принцип визуализации, в соответствии с которым каждое случайное множество порождает своеобразную "планетную систему" взаимодействующих случайных событий.

Здесь важно понять, что случайное множество событий — распределение вероятности между всеми подмножествами конечного множества событий — это самый общий способ описания всех статистических взаимозависимостей, равно как и всех статистических взаимодействий, которые только могут существовать между случайными событиями и подмножествами событий конечного множества \mathcal{X} . Можно сказать, что случайное множество событий под \mathcal{X} способно *управлять* любым типом статистически взаимозависимого и взаимодействующе-



34

my soul is dark - oh! quickly string
the harp i yet can brook to hear;
and let thy gentle fingers fling
its melting murmurs o'er mine ear.
if in this heart a hope be dear,
that sound shall charm it forth again:
if in these eyes there lurk a tear,
'twill flow, and cease to burn my brain.
but bid the strain be wild and deep,
nor let thy notes of joy be first:
i tell thee, minstrel, i must weep,
or else this heavy heart will burst;
for it hath been by sorrow nursed,
and ached in sleepless silence long;
and now 'tis doomed to know the worst,
and break at once - or yield to song.

George Byron

35

душа моя мрачна. скорей, певец, скорей!
вот арфа золотая;
пускай персты твои, промчавшись по ней,
пробудят в струнах звуки рая.
и если не навек надежды рок унёс, -
они в груди моей проснутся,
и если есть в очах застывших капля слёз, -
они растают и прольются.
пусть будет песнь твоя дика, как мой венец,
мне тягостны веселья звуки!
я говорю тебе: я слёз хочу, певец,
иль разорвётся грудь от муки.
страданиями была упитана она,
томилась долго и безмолвно;
и грозный час настал - теперь она полна,
как кубок смерти яда полный.

Михаил Лермонтов

Рис. 8: Численное решение задачи 4-х случайных событий — динамическая визуализация взаимодействий четырёх звуковых категорий в стихотворении Байрона и его переводе Лермонтовым — прообраз визуализации статистических взаимодействий категорий в контент – анализе текстов.

го поведения его событий и подмножеств событий.

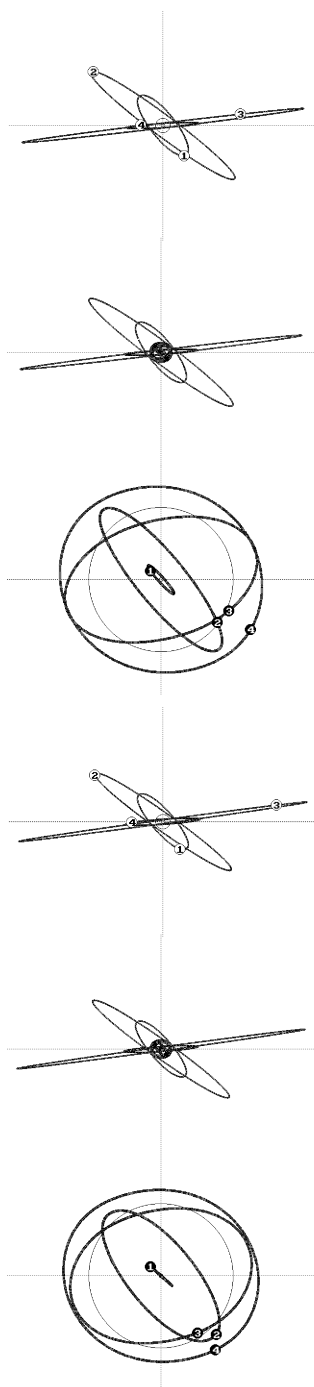
На рис. 7 приведены результаты численного решения задачи 4 случайных событий — пример динамической визуализации статистического поведения случайного потребителя 4-товарного рынка напитков. На рис. 6 приведены результаты численного решения задачи 9 случайных событий — пример динамической визуализации статистического поведения случайного множества событий, который показывает насколько сложной может быть орбитальная структура случайных событий, визуализирующая соответствующий тип их зависимостей и взаимодействий.

Итак, результаты численного решения данной задачи N случайных событий можно рассматривать как динамическую визуализацию структуры их статистических взаимодействий. Такие структуры, названные в работе *орбитальными структурами взаимодействий случайных событий*, сопоставляют совокупности параметров (вектор вероятностей и матрица сил статистического взаимодействия), характеризующих распределение случайного множества наступивших событий, видимый динамический образ, помогающий представить общую картину взаимодействий и упрощающий статистический анализ систем взаимодействующих случайных событий. Так в примере численного решения задачи 9-ти событий (рис. 6) анализ 54 параметров (9 вероятностей, 36 сил парного взаимодействия и 9 сил взаимодействия с достоверным событием) можно заменить анализом одной орбитальной структуры взаимодействий системы из 9-ти случайных событий.

ПРИМЕРЫ ОРБИТАЛЬНЫХ СТРУКТУР СТАТИСТИЧЕСКИХ ВЗАИМОДЕЙСТВИЙ ЗВУКОВЫХ КАТЕГОРИЙ В ПОЭТИЧЕСКИХ ТЕКСТАХ

Задача N случайных событий может быть решена для визуализации взаимодействий звуковых категорий в поэтических текстах.

Текст — это упорядоченное мультимножество слов, а слово — упорядоченное мультимножество букв. Текст можно единственным образом с помощью специальной процедуры — сет-нарезки — разбить



36

душа моя объята тьмою...
 певец, скорее арфы звон
 ты пробуди своей игрою;
 пускай мой слух лелеет он.
 и, если сердца святы грёзы,
 откликнусь я; коль может течь
 слеза из глаз, то хлынут слёзы
 и перестанут мозг мне жечь.
 пусть в песне той молчит отрада;
 в напеве будь уныл и дик.
 поверь, певец, – мне плакать надо.
 иль разорвётся сердце вмиг.
 оно томимо вечной скукой,
 тоской взлелеяно оно –
 и вот покончить разом с мукой
 иль песни власть признать должно.

Павел Козлов

37

душе моей грустно. спой песню, певец!
 любезен глас арфы душе и унылой...
 мой слух очаруй ты волшебством сердец,
 гармонии сладкой всемогущей силой.
 коль искра надежды есть в сердце моем,
 её вдохновенная арфа пробудит;
 когда хоть слеза сохранилась в нём –
 прольётся – и сердца сжигать мне не будет.
 но песни печали, певец, мне воспой;
 для радости сердце моё уж не бьётся;
 заставь меня плакать – иль долгой тоской
 гнетомое сердце моё разорвётся.
 довольно страдал я, довольно терпел –
 устал я... пусть сердце или сокрушится
 и кончит земной мой несносный удел,
 иль с жизнью арфой златой примирится.

Николай Гнедич

Рис. 9: Орбитальные структуры взаимодействий четырёх звуковых категорий в переводах Байрона, выполненных Козловым и Гнедичем, позволяют, сравнивая динамические образы, сравнить качество передачи "звучания" одного и того же стихотворения у различных переводчиков.

на пра-слова, каждое из которых — обычное множество букв. Пра-слова содержат звуковые категории — гласные, согласные, пробелы и знаки препинания, присутствия или отсутствия которых в пра-слове рассматриваются в качестве случайных событий.

На рис. 8 показано численное решение задачи 4 случайных событий — динамическая визуализация орбитальной структуры взаимодействий четырёх звуковых категорий: согласных (2) и гласных (1) букв, ”пробела” (3) и ”знаков препинания” (4) в пра-словах известного стихотворения Байрона и его не менее известного перевода Лермонтовым. Показаны: вверху — орбиты присутствий категорий, в центре — орбиты присутствия и отсутствия, внизу в другом масштабе — орбиты отсутствия категорий. Построенные орбитальные структуры статистических взаимодействий звуковых категорий можно рассматривать как прообраз визуализации статистических взаимодействий категорий в контент – анализе текстов.

Показанные рядом на рис. 9 орбитальные структуры взаимодействий четырёх звуковых категорий в переводах того же стихотворения Байрона, выполненных Козловым и Гнедичем, позволяют, сравнивая динамические образы, сравнить качество передачи ”звучания” одного и того же стихотворения у различных переводчиков.

Построение орбитальных структур взаимодействий звуковых категорий позволяет визуально сравнивать между собой различные системы статистически взаимодействующих категорий, сравнивая видимые образы их орбитальных структур. Подобные задачи динамической визуализации взаимодействия текстовых категорий — можно аналогичным образом формулировать и решать, например, в актуальном и модном сегодня *контент-анализе* текстов, цель которого так переработать огромный поток текстовой информации, чтобы облегчить её усвоение человеком.

Л И Т Е Р А Т У Р А

- [1] MONTGOMERY RICHARD (2001) A New Solution to the Three-Body Problem. *Notices of the AMS*, May, Volume 48, Number 5, 471–481.
- [2] ВОРОБЬЕВ О.Ю. (2001) *Предметный указатель шести томов персональных Записок ФАМ семинара (1–6, 1997–2002)*. Красноярск: ИВМ СО РАН, 11 с.
- [3] ВОРОБЬЕВ О.Ю. (2002, в печати) *Теория случайных событий и её применение*. Красноярск: ИВМ СО РАН, 340 с.
- [4] ГОЛДЕНОК Е.Е., и О.Ю. ВОРОБЬЕВ (2001) Захват вероятности событиями и визуализация элементов случайного множества. *Труды V ФАМ Конференции*. Красноярск: ИВМ СО РАН, 36–50.
- [5] ГОЛДЕНОК Е.Е., и О.Ю. ВОРОБЬЕВ (2001) Зависимости и взаимодействия в теории вероятностей. *Труды V ФАМ конференции*. Красноярск: ИВМ СО РАН, 231–238.
- [6] ГОЛДЕНОК Е.Е., АНОХИНА А.В., и О.Ю. ВОРОБЬЕВ (2001) О динамической де-визуализации случайных множеств. *Труды V ФАМ Конференции*. Красноярск: ИВМ СО РАН, 239–250.
- [7] ВОРОБЬЕВ О.Ю., и Е.Е. ГОЛДЕНОК (2001) Ковариация между случайными множествами и ковариация одного случайного множества. *Труды V ФАМ Конференции*. Красноярск: ИВМ СО РАН, 183–204.