

ДИАГРАММЫ РАССЕЯНИЯ В АНАЛИЗЕ СЛУЧАЙНЫХ ПОТОКОВ СОБЫТИЙ

В. И. Хименко^а, доктор техн. наук, профессор
^аОАО «НИО ЦИТ «Петрокомета», Санкт-Петербург, РФ

Цель: вероятностный анализ случайных потоков событий на основе двумерных диаграмм рассеяния и исследование общей структуры диаграмм для нескольких наиболее распространенных моделей случайных потоков. **Результаты:** показаны особенности построения диаграмм рассеяния, которые заключаются в отображении двумерной информации в задачах вероятностного анализа случайных потоков событий. Введены простые числовые характеристики, отражающие информационную структуру случайных данных на фазовой плоскости. Выполнен вероятностный анализ диаграмм рассеяния для класса простых пуассоновских потоков, случайных потоков с неоднородной вероятностной структурой, потоков с двойной стохастичностью и класса случайных альтернирующих потоков событий. **Практическая значимость:** потенциальная полезность полученных результатов и практического использования диаграмм рассеяния состоит в наглядности и высокой информативности представления двумерной информации на фазовой плоскости. Диаграммы рассеяния позволяют накапливать, усреднять и систематизировать выборочные данные, оперативно оценивать однородность потоков событий, рассматривать эффекты коррелированности, группировки, эффекты случайных «загрязнений» и эффекты случайного пропуска данных.

Ключевые слова — обработка информации, случайные процессы, потоки событий, точечные процессы, диаграммы рассеяния, вероятностный анализ, неоднородные данные.

Введение

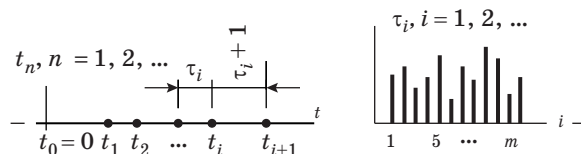
Наиболее часто модели случайных потоков используются при описании различных последовательностей однородных событий. Необходимость исследования таких потоков возникает во многих задачах физики, биологии, техники, теории надежности и теории рисков, теории транспортных потоков, разнообразных задачах логистики и задачах теории массового обслуживания. Типовыми примерами случайных потоков событий могут быть последовательности телефонных вызовов, потоки заявок при многоканальном доступе в спутниковых системах связи, моменты отказов вычислительных систем, потоки импульсов в нейронных сетях, случайные потоки электронов, фотонов или каких-либо элементарных частиц при ядерно-физических экспериментах. Это могут быть и потоки переходов какой-либо сложной системы из одного состояния в другое.

Если рассматривается некоторая последовательность однородных событий, каждое из которых может происходить в случайный момент времени t_i , то поток таких событий условно можно представить (рис. 1) на оси времени $0t$ в виде последовательности точек $\{t_n\}$ — случайного точечного процесса. Каждая точка $t_i, i = 1, 2, \dots$ соответствует при этом моменту появления i -го события, а значения $\tau_i = t_i - t_{i-1}$ — длительностям интервалов между последовательными событиями.

Учитывая, что случайный поток состоит из однородных событий, анализ точечных процессов $\{t_n, n = 1, 2, \dots\}$ во многих практических за-

дачах основан на исследованиях вероятностных характеристик временных интервалов $\{\tau_i\}$. При таком исследовании важное значение имеет не только описание длительностей τ_i , но и рассмотрение совместных свойств случайных величин τ_i и τ_j при $j \neq i$. Одним из возможных (и наиболее простых) подходов к выявлению особенностей совместного поведения длительностей τ_i и τ_j является анализ диаграмм рассеяния для τ_i, τ_j на фазовой плоскости (τ_i, τ_j) . Подобные представления используются в анализе вероятностной структуры временных рядов, например, при корреляционном, регрессионном и факторном анализе [1–4], в задачах распознавания образов [5], задачах обработки данных радиофизических экспериментов [6] и данных аэрокосмического мониторинга [7], задачах анализа информации в эконометрике [8], кардиологии [9] и нейрофизиологии [10].

Характерными свойствами диаграмм рассеяния, с одной стороны, являются простота и наглядность отображения двумерной информации, возможность накопления и усреднения выборочных данных при возрастании общей длительности



■ **Рис. 1.** Отдельная реализация точечного процесса $t_n, n = 1, 2, \dots$ и случайная последовательность интервалов τ_i между отдельными событиями

сти наблюдений. С другой стороны, исследования самих диаграмм рассеяния [1–10], как правило, носят лишь описательный характер, выполняются на уровне визуального анализа и не используют значительную часть информации, содержащейся в вероятностной структуре получаемых данных.

Цель данной работы — анализ особенностей представления случайных величин τ_i и τ_j на фазовой плоскости (τ_i, τ_j) и вероятностное исследование диаграмм рассеяния для нескольких практически важных моделей случайных потоков событий $\{t_n\}$.

Диаграммы рассеяния для процессов Пуассона

Общая структура и основные свойства процессов Пуассона достаточно хорошо изучены; такие процессы рассматриваются во многих работах (например, [1, 11, 12]) и используются в качестве базовой модели при построении большинства других, более сложных моделей случайных потоков событий [13, 14]. Именно поэтому с модели пуассоновского процесса полезно начать вероятностный анализ диаграмм рассеяния.

Предположим, что $\{t_n, n=1, 2, \dots\}$ — простой пуассоновский поток событий, и рассмотрим случайную последовательность $\{\tau_i, i=1, 2, \dots\}$ временных интервалов $\tau_i = t_i - t_{i-1}$ между соседними событиями. Характерной особенностью данной модели является то, что величины $\tau_1, \tau_2, \dots, \tau_n$ относятся здесь к классу независимых одинаково распределенных случайных величин. Они описываются экспоненциальным распределением [11–14], и если через λ обозначить интенсивность потока (т. е. среднее число событий в единицу времени), то плотность вероятностей $p_\tau(\tau)$ и функция распределения $F_\tau(\tau)$ для τ_i будут иметь вид

$$p_\tau(\tau) = \lambda \exp(-\lambda\tau), \quad F_\tau(\tau) = 1 - \exp(-\lambda\tau), \quad \tau \geq 0. \quad (1)$$

Условие взаимной независимости интервалов τ_i и τ_j при $i \neq j$ позволяет по выражениям (1) определить совместную плотность вероятностей

$$p_2(\tau_i, \tau_j) = p_\tau(\tau_i)p_\tau(\tau_j) = \lambda^2 \exp[-\lambda(\tau_i + \tau_j)], \quad \tau_{i,j} \geq 0. \quad (2)$$

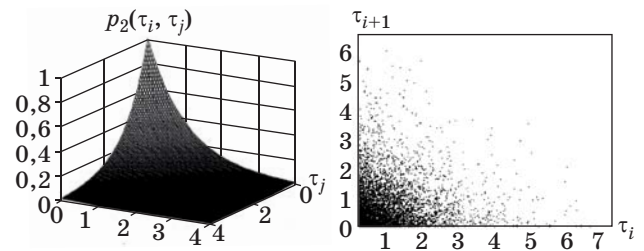
Формально функция (2) содержит всю информацию о совместном поведении временных интервалов τ_i и τ_j , и является основой для исследования двумерных вероятностных характеристик случайных последовательностей $\{\tau_j\}$. Однако в задачах статистического анализа построение по экспериментальным данным двумерных распределений требует больших объемов выборки и существенных вычислительных затрат. Кроме того, подобная информация в функции (2) содер-

жится в неявном виде, и для ее выделения нужны дополнительные операции.

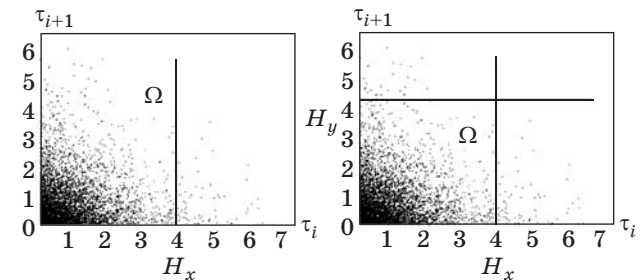
Рассмотрим здесь более простой способ отображения информации о совместном поведении длительностей τ_i и τ_j . Выберем для определенности значение $j=i+1$ и будем интерпретировать случайные величины τ_i и τ_{i+1} как координаты точки на плоскости (τ_i, τ_{i+1}) . Тогда поведение последовательности $\{\tau_i, i=1, 2, \dots\}$ можно представить геометрически в виде некоторой диаграммы рассеяния. В качестве простой иллюстрации на рис. 2 приведена типовая диаграмма для длительностей интервалов τ_i и $\tau_j = \tau_{i+1}$ пуассоновского потока событий. Ее общий вид непосредственно связан с особенностями поведения совместной плотности вероятностей $p_2(\tau_i, \tau_j)$.

Действительно, характер функции (2) подтверждает (см. рис. 2), что для рассматриваемой модели процесса $\{t_n, n=1, 2, \dots\}$ диаграмма рассеяния должна обладать относительно высокой концентрацией (или «плотностью») точек в окрестностях малых значений τ и не должна иметь каких-либо явно выраженных областей группировки событий на плоскости (τ_i, τ_{i+1}) при возрастании длительностей τ . С увеличением значений τ плотность точек (или плотность событий потока) уменьшается по экспоненциальному закону.

Для определения вероятностных характеристик диаграмм рассеяния зададим на фазовой плоскости (τ_i, τ_j) некоторую область «допустимых» значений Ω . Границы этой области (рис. 3)



■ Рис. 2. Характер представления двумерной информации в виде совместной плотности вероятностей и в виде диаграммы рассеяния



■ Рис. 3. Типовая структура диаграммы рассеяния для пуассоновского потока событий и заданные области Ω допустимых значений

определим из условия $\tau_i \in (0, H_x)$, где H_x может рассматриваться как некоторый пороговый уровень в диапазоне изменений длительностей τ_i .

Предположим, что исследуется случайный поток событий $\{t_n, n=1, 2, \dots\}$ и наблюдению доступно m временных интервалов $\tau_1, \tau_2, \dots, \tau_m$. При отображении этих наблюдений на фазовой плоскости (τ_i, τ_{i+1}) диаграмма рассеяния будет содержать $(m-1)$ точек, каждая из которых характеризуется координатами τ_i, τ_j , где $j=i+1$, а $i=1, 2, \dots, m-1$. Некоторое число этих точек $n(\Omega, m)$ может находиться в заданной области Ω допустимых значений $\tau_i \in (0, H_x)$, а часть точек, обозначим их число через $n^+(\Omega, m)$, — за границами выбранной области Ω . Если какая-либо произвольная точка ξ_i описывается координатами τ_i, τ_{i+1} и при этом выполняется условие $\tau_i < H_x$, то данная точка находится в области допустимых значений Ω . Вероятность такого события $P\{\xi \in \Omega\}$ можно выразить через функции (1) как

$$P\{\xi \in \Omega\} = P\{\tau_i \leq H_x\} = \int_0^{H_x} p_\tau(\tau) d\tau = F_\tau(H_x), \quad H_x \geq 0. \quad (3)$$

С другой стороны, если значение $\tau_i > H_x$, то рассматриваемая точка ξ_i выходит за границу заданной области Ω , и вероятность этого события

$$P\{\xi \notin \Omega\} = P\{\tau_i > H_x\} = \int_{H_x}^{\infty} p_\tau(\tau) d\tau = 1 - F_\tau(H_x). \quad (4)$$

Таким образом, если наблюдению доступно m случайных величин $\tau_1, \tau_2, \dots, \tau_m$, то при анализе диаграмм рассеяния, с учетом вероятностей (3) и (4), можно определить математическое ожидание (среднее значение) числа точек $N(\Omega, m) = M\{n(\Omega, m)\}$, попадающих на фазовой плоскости в область Ω , и среднее число точек $N^+(\Omega, m) = M\{n^+(\Omega, m)\}$, выходящих за границы области допустимых значений Ω :

$$\begin{aligned} N(\Omega, m) &= (m-1)P\{\xi \in \Omega\}; \\ N^+(\Omega, m) &= (m-1) - N(\Omega, m) = \\ &= (m-1)[1 - P\{\xi \in \Omega\}]. \end{aligned} \quad (5)$$

Более полное вероятностное описание случайных величин $n(\Omega, m)$ и $n^+(\Omega, m)$ в подобных задачах может быть выполнено на основе классической для теории вероятностей схемы Бернулли [14–16]. Так, будем считать, что результаты наблюдений $\tau_1, \tau_2, \dots, \tau_m$ относятся к классу независимых случайных величин и отображаются на фазовой плоскости (τ_i, τ_{i+1}) в виде совокупности случайно расположенных точек $\xi_1, \xi_2, \dots, \xi_{m-1}$. Каждая точка ξ_i с вероятностью $p = P\{\xi \in \Omega\} = F_\tau(H_x)$ попадает в область допусти-

мых значений Ω и с вероятностью $p^+ = 1 - p = 1 - F_\tau(H_x)$ выходит за пределы заданной области Ω . Тогда, по аналогии с результатами работы [15], вероятность выхода за пределы области ровно k точек из рассматриваемой совокупности $\xi_1, \xi_2, \dots, \xi_{m-1}$ определится выражением

$$\begin{aligned} p_k(n^+) &= P\{n^+(\Omega, m) = k\} = \\ &= C_{m-1}^k (p^+)^k (1 - p^+)^{m-1-k}, \end{aligned} \quad (6)$$

где $k=0, 1, 2, \dots, m-1$, а $C_{m-1}^k = (m-1)!/k!(m-1-k)!$ — число сочетаний из $(m-1)$ элементов по k .

По своему виду результат (6) соответствует одному из наиболее распространенных в теории вероятностей [14, 16] биномиальному закону распределения. Форма и параметры этого закона в данном случае зависят лишь от общего числа наблюдений m и вероятности p^+ . В частности, для математического ожидания и дисперсии числа точек, выходящих на фазовой плоскости за границы заданной области Ω , на основе распределения (6) получим

$$\begin{aligned} N^+(\Omega, m) &= M\{n^+(\Omega, m)\} = (m-1)(p^+) = \\ &= (m-1)[1 - F_\tau(H_x)]; \\ D[n^+(\Omega, m)] &= M\left\{\left[n^+(\Omega, m) - N^+(\Omega, m)\right]^2\right\} = \\ &= (m-1)(p^+)(1 - p^+). \end{aligned} \quad (7)$$

Эти формулы просты и удобны для численных расчетов. Полезно, кроме того, заметить, что среднее число точек $N^+(\Omega, m) = M\{n^+(\Omega, m)\}$, попадающих за пределы области Ω , как и следовало ожидать, совпадает с выражением (5) и с учетом функции распределения (1) может быть представлено в форме

$$N^+(\Omega, m) = (m-1)\exp(-\lambda H_x), \quad H_x \geq 0. \quad (8)$$

Если учесть здесь дисперсию σ_τ^2 экспоненциального распределения (1): $\sigma_\tau^2 = 1/\lambda^2$, и для удобства задания области Ω ввести нормированное значение порогового уровня $H_x / \sigma_\tau = \lambda H_x = h_x$, то формулу (8) можно переписать еще в более наглядном виде

$$N^+(\Omega, m) = (m-1)\exp(-h_x), \quad h_x = \lambda H_x \geq 0. \quad (9)$$

Это выражение показывает, что при фиксированном объеме выборки $m = \text{const}$ изменение порогового уровня h_x на фазовой плоскости (τ_i, τ_{i+1}) приводит к экспоненциальному изменению среднего числа отображающих точек (или событий), выходящих за границы заданной области Ω .

Выделим теперь еще одну важную особенность вероятностного анализа. Очевидно, что в зависимости от решаемых задач область допустимых

значений Ω может иметь различную форму. Так, например, в качестве простого обобщения определим область Ω двумя условиями: $\tau_i \in (0, H_x)$, $\tau_{i+1} \in (0, H_y)$, где значения H_x и H_y рассматриваются как некоторые пороговые уровни в диапазоне изменений τ (см. рис. 3).

Вероятность $P\{\xi \in \Omega\}$ нахождения произвольной точки ξ с координатами τ_i, τ_j в заданной области Ω запишется при этом в виде

$$P\{\xi \in \Omega\} = P\{\tau_i < H_x, \tau_j < H_y\} = \int_0^{H_x} \int_0^{H_y} p_2(\tau_i, \tau_j) d\tau_i d\tau_j, \quad (10)$$

а при независимости интервалов τ_i и τ_j значение вероятности (10) будет определяться как

$$P\{\xi \in \Omega\} = P\{\tau_i < H_x\} P\{\tau_j < H_y\} = \int_0^{H_x} p_\tau(\tau_i) d\tau_i \int_0^{H_y} p_\tau(\tau_j) d\tau_j. \quad (11)$$

Таким образом, если наблюдению доступно m временных интервалов $\tau_1, \tau_2, \dots, \tau_m$ между последовательными событиями точечного процесса $\{t_n, n=1, 2, \dots\}$, то с учетом вероятностей (10), (11) для диаграммы рассеяния на плоскости (τ_i, τ_j) при $j=i+1$ можно определить среднее число точек $N(\Omega, m)$, попадающих в область Ω , и среднее число точек $N^+(\Omega, m)$, выходящих за границы заданной области допустимых значений Ω :

$$N(\Omega, m) = (m-1)P\{\xi \in \Omega\};$$

$$N^+(\Omega, m) = (m-1)[1 - P\{\xi \in \Omega\}]. \quad (12)$$

Формально этот результат совпадает с выражением (5), однако значение вероятности $P\{\xi \in \Omega\}$ в формулах (12) определяется выражениями (10) или (11) с учетом иной области допустимых значений Ω .

Так, в частности, применительно к моделям пуассоновских процессов $\{t_n, n=1, 2, \dots\}$ на основе распределений (1) и (2) записанная в формулах (12) вероятность будет определяться значением

$$P\{\xi \in \Omega\} = \left(1 - e^{-\lambda H_x}\right) \left(1 - e^{-\lambda H_y}\right),$$

$$H_x, H_y \in (0, \infty). \quad (13)$$

Для удобства практических расчетов в этом выражении можно выполнить переход от пороговых уровней H_x и H_y к их нормированным значениям:

$$H_x/\sigma_\tau = \lambda H_x = h_x, \quad H_y/\sigma_\tau = \lambda H_y = h_y.$$

В целом полезно здесь подчеркнуть, что рассмотренные результаты (3)–(13) позволяют выполнять вероятностный анализ общей структуры

диаграмм рассеяния, определять вероятностные характеристики и исследовать количественные соотношения в расположении точек (событий) на фазовой плоскости (τ_i, τ_j) при различном задании границ H_x и H_y области допустимых значений Ω .

Случайные потоки с неоднородной структурой

В задачах обработки информации и задачах исследования случайных потоков событий достаточно часто возникают ситуации, когда интервалы τ_i между отдельными событиями $\{t_n, n=1, 2, \dots\}$ имеют неоднородную вероятностную структуру. Подобные неоднородности могут быть связаны с физической природой изучаемых процессов, могут появляться при анализе наблюдений со случайными пропусками данных, при анализе случайных потоков с «загрязнениями» или появлением аномальных значений.

Рассмотрим простую и распространенную для подобных задач вероятностную модель

$$p_\tau(\tau) = \alpha_1 p_1(\tau|\lambda_1) + \alpha_2 p_2(\tau|\lambda_2),$$

$$\alpha_1 + \alpha_2 = 1, \quad \alpha_{1,2} \in [0, 1], \quad (14)$$

в которой плотность вероятностей $p_\tau(\tau)$ для временных интервалов $\{\tau_j\}$ представляется в виде смеси двух различных распределений $p_1(\tau|\lambda_1)$ и $p_2(\tau|\lambda_2)$ с параметрами λ_1 и λ_2 . Весовые коэффициенты α_1 и α_2 играют в этой модели роль вероятностей, а их соотношение определяет степень неоднородности случайной последовательности $\{\tau_j\}$.

Для анализа совместного поведения случайных величин τ_i и τ_j рассмотрим применительно к модели (14) особенности поведения диаграмм рассеяния для длительностей интервалов τ_i и τ_j при $j=i+1, i=1, 2, \dots, m-1$ на фазовой плоскости (τ_i, τ_{i+1}) .

Зададим на плоскости (τ_i, τ_{i+1}) в области изменения значений $\tau_i \in (0, \infty)$ некоторый пороговый уровень $H_x > 0$ и будем считать область $\Omega: \tau_i \in (0, H_x)$ областью допустимых значений. Произвольная точка $\xi_i = \xi_i(\tau_i, \tau_{i+1})$ с координатами (τ_i, τ_{i+1}) на фазовой плоскости может находиться в заданной области $\xi_i \in \Omega$ и может выходить за границу H_x этой области, т. е. $\xi_i \notin \Omega$. Вероятности этих событий в соответствии с формулами (3) и (4) определяются выражениями

$$P\{\xi \in \Omega\} = \int_0^{H_x} p_\tau(\tau) d\tau =$$

$$= \alpha_1 \int_0^{H_x} p_1(\tau|\lambda_1) d\tau + \alpha_2 \int_0^{H_x} p_2(\tau|\lambda_2) d\tau;$$

$$P\{\xi \notin \Omega\} = \alpha_1 \int_{H_x}^{\infty} p_1(\tau|\lambda_1) d\tau + \alpha_2 \int_{H_x}^{\infty} p_2(\tau|\lambda_2) d\tau. \quad (15)$$

Следовательно, если наблюдению доступно m случайных величин $\tau_1, \tau_2, \dots, \tau_m$, то на основе общих результатов (5) и вероятностей (15) можно определить для диаграммы рассеяния на плоскости (τ_i, τ_{i+1}) среднее число точек $N(\Omega, m)$, находящихся в области допустимых значений Ω , и среднее число точек $N^+(\Omega, m)$, выходящих за границу H_x заданной области Ω :

$$N(\Omega, m) = (m - 1) - N^+(\Omega, m);$$

$$N^+(\Omega, m) = \alpha_1 N_1^+(H_x, m) + \alpha_2 N_2^+(H_x, m), \quad (16)$$

где

$$N_k^+(H_x, m) = (m - 1) \int_{H_x}^{\infty} p_k(\tau | \lambda_k) d\tau, \quad k = 1, 2. \quad (17)$$

Такой результат показывает, что представление неоднородных данных $\{\tau_i\}$ моделью распределения $p_\tau(\tau)$ в виде вероятностной смеси (14) позволяет представить и характеристики расположения точек ξ_i на диаграмме рассеяния в виде взвешенной суммы слагаемых $N_k^+(H_x, m)$, соответствующих отдельным компонентам $k = 1, 2$ случайной последовательности $\{\tau_i\}$.

Рассмотрим теперь на фазовой плоскости (τ_i, τ_{i+1}) два пороговых уровня H_x и H_y и определим область допустимых значений условием

$$\Omega: \tau_i \in (0, H_x), \tau_{i+1} \in (0, H_y). \quad (18)$$

Для упрощения анализа будем считать, что случайная последовательность $\{\tau_i\}$ является последовательностью независимых и одинаково распределенных случайных величин с общей плотностью вероятностей (14). Допущение о независимости случайных интервалов $\{\tau_i\}$ характерно для большинства моделей точечных процессов $\{t_n\}$. При этом без потери общности можно также в определении пороговых уровней (18) положить $H_x = H_y = H$.

Перечисленные условия позволяют воспользоваться формулой (11) и для диаграммы рассеяния процессов $\{t_n\}$ с неоднородной структурой (14) определить вероятность нахождения произвольной точки $\xi_i = \xi_i(\tau_i, \tau_{i+1})$ с координатами (τ_i, τ_{i+1}) в заданной области Ω :

$$\begin{aligned} P\{\xi \in \Omega\} &= P\{\tau_i < H, \tau_j < H\} = \left(\int_0^H p_\tau(\tau) d\tau \right)^2 = \\ &= \left(\alpha_1 \int_0^H p_1(\tau | \lambda_1) d\tau + \alpha_2 \int_0^H p_2(\tau | \lambda_2) d\tau \right)^2. \end{aligned} \quad (19)$$

В свою очередь, этот результат позволяет на основе общих выражений (12) непосредственно определить среднее число точек $N(\Omega, m)$ и $N^+(\Omega, m)$, находящихся в области Ω допустимых значений (18) и выходящих за границы $H_x = H_y = H$ заданной области Ω .

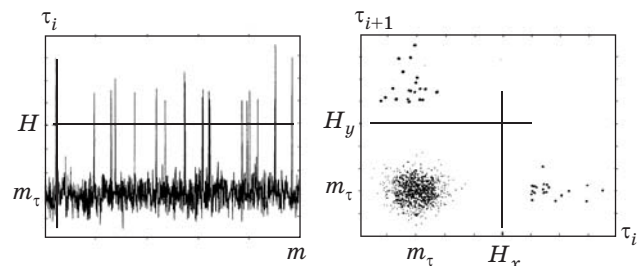
Рассмотренная вероятностная модель (14) наиболее часто используется при описании «загрязненных» наблюдений, анализе экспериментальных данных с «грубыми» ошибками или исследовании потоков событий $\{t_n\}$ с аномальными отклонениями параметров. Функции $p_1(\tau | \lambda_1)$ и $p_2(\tau | \lambda_2)$ или параметры интенсивности λ_1 и λ_2 в подобных ситуациях существенно различаются между собой.

Для наглядности на рис. 4 показан характерный вид отдельной реализации $\tau_1, \tau_2, \dots, \tau_m$ неоднородной случайной последовательности $\{\tau_i\}$ и приведена построенная для этой реализации диаграмма рассеяния на фазовой плоскости (τ_i, τ_{i+1}) . В качестве примера за основу здесь был выбран случайный точечный процесс $\{t_n\}$, порожденный регулярным потоком однородных событий со случайными возмущениями длительностей интервалов τ_i между событиями. По структуре эти возмущения представляют собой последовательность независимых случайных величин, описание которых в модели (14) выполняется на основе плотности вероятностей $p_1(\tau | \lambda_1)$. Вторая компонента $p_2(\tau | \lambda_2)$ в вероятностной модели (14) описывает эффекты резких отклонений длительностей τ_i от их средних значений и, по существу, отражает структуру случайных «загрязнений» последовательности $\{\tau_i\}$.

Неоднородность общей вероятностной структуры точечного процесса $\{t_n\}$ достаточно наглядно проявляется на фазовой плоскости (τ_i, τ_{i+1}) . При весовых коэффициентах $\alpha_1 \gg \alpha_2$ на диаграмме рассеяния (см. рис. 4) отчетливо формируются область основных значений τ , определяемых функцией $p_1(\tau | \lambda_1)$, и области «загрязнений» — скопления редких, но выделяющихся по своей величине значений τ , описываемых плотностью вероятностей $p_2(\tau | \lambda_2)$.

При обработке и анализе подобных наблюдений за счет соответствующего выбора пороговых уровней H_x и H_y на фазовой плоскости (τ_i, τ_{i+1}) могут быть разделены основная и «загрязняющая» компоненты.

Рассмотренный здесь подход к анализу и полученные результаты (14)–(19) позволяют во многих



■ Рис. 4. Реализация случайной последовательности $\{\tau_i\}$ с неоднородной вероятностной структурой и ее отображение на фазовой плоскости

практических задачах выполнять совместное и раздельное исследование вероятностных свойств отдельных составляющих в неоднородной структуре случайных потоков событий.

Случайные потоки с двойной стохастичностью

Подавляющее большинство моделей случайных потоков событий $\{t_n\}$ строится в предположении постоянства вероятностных характеристик потока на интервале наблюдения. Это относится, в частности, и к моделям процессов с неоднородной структурой вида (14). Реальные ситуации нередко оказываются более сложными, исследуемые потоки меняют свои параметры случайным образом, и для их описания необходимы более универсальные вероятностные модели.

Рассмотрим особенности поведения диаграмм рассеяния в задачах исследования случайных потоков событий с двойной стохастичностью. Для этого предварительно отметим, что при фиксированных границах области допустимых значений Ω и постоянных на интервале наблюдения вероятностных характеристиках случайного потока событий $\{t_n\}$ для произвольной точки $\xi_i = \xi_i(\tau_i, \tau_{i+1})$ с координатами (τ_i, τ_{i+1}) на диаграмме рассеяния может быть определена вероятность $p = P\{\xi \in \Omega\}$ события $\xi \in \Omega$ и вероятность $p^+ = P\{\xi \notin \Omega\} = 1 - p$ того, что точка ξ_i выходит за границы заданной области Ω .

Условие постоянства вероятностных характеристик случайного потока $\{t_n\}$ позволяет в данной ситуации считать постоянными и значения вероятностей $p = \text{const}$ и $p^+ = 1 - p = \text{const}$. Эти особенности, в свою очередь, приводят к возможности использования модели биномиального распределения (6) для описания числа точек $n^+(\Omega, m)$, выходящих на фазовой плоскости (τ_i, τ_j) за пределы заданной области Ω .

Перейдем теперь к более сложной ситуации и рассмотрим вероятностную модель (6) при условии случайных изменений значения вероятности $p^+ = P\{\xi \notin \Omega\}$. Такие изменения могут происходить из-за меняющейся помеховой обстановки при обработке информационных процессов, за счет появления нестационарностей, эффектов случайной группировки событий в потоке, флуктуаций границ заданной области, случайных изменений интенсивности потоков и многих других причин.

Будем считать, что в распределении (6) вероятность p^+ является непрерывной случайной величиной и описывается некоторой плотностью вероятностей $w(p^+)$, $p^+ \in [0, 1]$. В этих условиях для описания числа точек $n^+(\Omega, m)$ необходимо выполнить вероятностное усреднение функ-

ции (6) по возможным значениям вероятности $p^+ \in [0, 1]$:

$$p_k(n^+) = P\{n^+(\Omega, m) = k\} = C_{m-1}^k \int_0^1 (p^+)^k (1-p^+)^{m-1-k} w(p^+) dp^+. \quad (20)$$

Выражение (20) по своей сути является обобщением биномиального распределения (6) и может рассматриваться как вероятностная модель случайного потока с двойной стохастичностью. Функция $w(p^+)$ играет при этом роль управляющего закона для случайных изменений вероятности p^+ . Модель (20) справедлива при различных распределениях $w(p^+)$ и независимо от конкретного вида функции $w(p^+)$ дает возможность найти два полезных общих соотношения для основных параметров:

$$N^+(\Omega, m) = M\{n^+(\Omega, m)\} = (m-1)M\{p^+\} = (m-1)\bar{p}^+;$$

$$D[n^+(\Omega, m)] = (m-1)\bar{p}^+(1-\bar{p}^+) + (m-1)(m-2)D[p^+]. \quad (21)$$

Эти соотношения показывают связь математического ожидания $N^+(\Omega, m)$ и дисперсии $D[n^+(\Omega, m)]$ случайной величины $n^+(\Omega, m)$ с соответствующими параметрами $\bar{p}^+ = M\{p^+\}$ и $D[p^+]$ управляющего закона распределения $w(p^+)$.

Результаты (20) и (21) для многих практических задач дают исчерпывающее описание вероятностных характеристик числа точек $n^+(\Omega, m)$, выходящих на фазовой плоскости за границы заданной области Ω . Сравнивая формулы (20), (21) и (6), (7), необходимо здесь заметить, что случайные изменения вероятности p^+ и переход к исследованиям случайных потоков с двойной стохастичностью (20) всегда приводят к возрастанию дисперсии $D[n^+(\Omega, m)]$ случайной величины $n^+(\Omega, m)$. Как и следовало ожидать, это возрастание непосредственно связано с дисперсией $D[p^+]$ управляющего закона распределения $w(p^+)$. Если случайные изменения параметра p^+ уменьшаются, то $D[p^+] \rightarrow 0$, и формулы (20), (21) переходят в соответствующие выражения (6), (7) для типовых задач с фиксированным значением вероятности $p^+ = 1 - p = \text{const}$.

Очевидно, что вероятностная модель (20) с двойной стохастичностью является более общей для исследования случайных потоков событий. Полученные при этом результаты (20), (21) позволяют выполнять анализ вероятностной структуры диаграмм рассеяния в условиях случайных изменений заданной области Ω и случайных изменений вероятностных характеристик рассматриваемых потоков событий.

**Структура
альтернирующих потоков событий**

Остановимся теперь кратко на рассмотрении вероятностной структуры диаграмм рассеяния для класса альтернирующих процессов. Характерной особенностью таких процессов является существование двух типов интервалов $\{\tau_i\}$ и $\{\tau_i^*\}$ между последовательными событиями. При этом сама структура случайного потока $\{t_n\}$ формируется таким образом, что временные интервалы τ_i и τ_i^* , $i = 1, 2, \dots$ следуют один за другим, поочередно заменяя друг друга и формируя двухкомпонентную случайную последовательность вида $\tau_1, \tau_1^*, \tau_2, \tau_2^*, \dots, \tau_n, \tau_n^*$.

На практике обычно предполагается, что все случайные величины τ_i и τ_i^* являются взаимно независимыми, и их вероятностные характеристики описываются некоторыми плотностями вероятностей $p(\tau)$ и $p(\tau^*)$. Такие особенности характерны для многих задач теории массового обслуживания, теории надежности и теории восстановления, задач логистики и задач теории запасов.

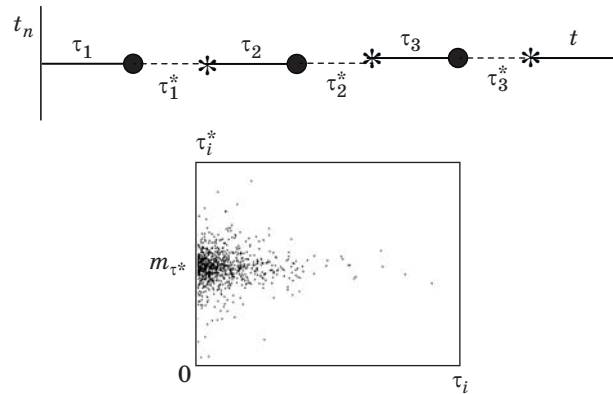
Для анализа вероятностной структуры альтернирующего потока событий $\{t_n\}$ предположим, что наблюдению доступна случайная последовательность

$$\tau_1, \tau_1^*, \tau_2, \tau_2^*, \dots, \tau_m, \tau_m^*, \quad (22)$$

состоящая из m пар временных интервалов τ_i и τ_i^* , $i = 1, 2, \dots$. Такая последовательность (22) может рассматриваться как отдельная реализация τ_i, τ_i^* , $i = 1, \dots, m$ двумерной векторной непрерывной случайной последовательности $\{\tau_i, \tau_i^*\}$ с независимыми компонентами $\{\tau_i\}$ и $\{\tau_i^*\}$. Если при этом значения τ_i и τ_i^* интерпретировать как координаты точки $\xi_i = \xi_i(\tau_i, \tau_i^*)$ на фазовой плоскости (τ_i, τ_i^*) , то реализация (22) может быть представлена на этой плоскости в виде некоторой диаграммы рассеяния. По своей структуре такая диаграмма отличается от рассмотренных ранее диаграмм (см. рис. 3 и 4). Она формируется на фазовой плоскости (τ_i, τ_i^*) из двух классов случайных величин $\{\tau_i\}$ и $\{\tau_i^*\}$ и, по существу, может рассматриваться как диаграмма рассеяния для двумерной векторной случайной последовательности $\{\tau_i, \tau_i^*; i = 1, 2, \dots\}$.

Характерный вид такой диаграммы показан на рис. 5. Для наглядности здесь приведен типовой пример, когда длительности интервалов $\{\tau_i\}$ описываются экспоненциальным распределением $p(\tau)$, а интервалы $\{\tau_i^*\}$ характеризуются гауссовой моделью $p(\tau^*)$.

Различия этих распределений наглядно проявляются в вероятностной структуре расположения точек $\xi_i = \xi_i(\tau_i, \tau_i^*)$. Так, в частности, концентрация отображающих точек на фазовой



■ **Рис. 5.** Особенности общей структуры альтернирующих потоков событий

плоскости (τ_i, τ_i^*) вдоль горизонтальной оси отражает экспоненциальный характер убывания их количества по мере возрастания длительности τ_i . Расположение и характер рассеяния точек ξ_i вдоль вертикальной оси $0\tau_i^*$ отражает особенности гауссова распределения $p(\tau^*)$ интервалов τ_i^* . Здесь наглядно проявляются эффекты группировки точек в окрестностях среднего значения $m_{\tau^*} = M\{\tau_i^*\}$ и эффекты характерного для гауссовой модели убывания количества отображающих точек $\xi_i = \xi_i(\tau_i, \tau_i^*)$ при возрастании отклонений $|\tau_i^* - m_{\tau^*}|$ переменной τ_i^* от значения $M\{\tau_i^*\}$.

Если применительно к подобным диаграммам рассеяния (см. рис. 5) воспользоваться результатами (3)–(9) и (11), (12), то появляется возможность исследовать вероятностную структуру альтернирующих потоков событий. Так, в частности, задавая различные пороговые уровни для значений $\{\tau_i\}$, можно исследовать вероятностные характеристики распределений $p(\tau)$. Если рассматривать длительности интервалов $\{\tau_i^*\}$, то при различных пороговых уровнях для τ_i^* можно исследовать вероятностные характеристики распределений $p(\tau^*)$.

Заключение

Результаты данной работы показывают возможность вероятностного анализа диаграмм рассеяния и особенности их использования в задачах исследования общей структуры случайных потоков событий. Диаграммы рассеяния обладают достаточно высокой информативностью и позволяют отображать, накапливать и систематизировать на фазовой плоскости двумерную информацию о совместном поведении случайных величин. На их основе удастся оценивать однородность выборки, рассматривать эффекты коррелированности, группировки, эффекты «загрязнения» и случайного пропуска данных, выполнять приближенную оценку простых вероятностных

характеристик — математического ожидания, степени разброса выборочных данных, вероятностей нахождения случайных величин в заданной области допустимых значений. Подобные операции представляют особый интерес в задачах оперативной обработки данных и экспресс-анализе вероятностной структуры различных по своей физической природе случайных потоков событий.

Следует также отметить, что полученные в данной работе результаты допускают различные обобщения. В частности, это может быть связано с выбором вероятностных моделей рассматриваемых случайных потоков, принципами выбора областей допустимых значений, исследованиями нестационарных и неоднородных вероятностных структур.

Литература

1. Вентцель Е. С. Теория вероятностей. — М.: Академия, 2005. — 573 с.
2. Forster E., Ronz B. Methoden der Korrelations und Regressionsanalyse. — Berlin: Verlag die Wirtschaft, 1979.
3. Боровиков В. П. STATISTICA: искусство анализа данных на компьютере. — СПб.: Питер, 2001. — 656 с.
4. Uberla K. Faktorenanalyse. — Berlin: Springer-Verlag, 1971.
5. Tou J. T., Gonzalez R. C. Pattern Recognition Principles. — London: Addison — Wesley Publishing Company, 1974.
6. Яковлев О. И. Космическая радиофизика. — М.: Научная книга, 1998. — 432 с.
7. Schowengerdt R. A. Remote Sensing: Models and Methods for Image Processing. — Amsterdam: Elsevier, 2007.
8. Носко В. П. Эконометрика. — М.: Дело, 2011. — 672 с.
9. Новые методы электрокардиографии / под ред. С. В. Грачева, Г. Г. Иванова, А. Л. Сыркина. — М.: Техносфера, 2007. — 552 с.
10. Naatanen R. Attention and Brain Function. — Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers, 1998. — 560 p.
11. Snyder D. L. Random Point Processes. — N. Y.: John Wiley, 1975. — 486 p.
12. Kingman J. Poisson Processes. — Oxford: Clarendon Press, 2007. — 136 p.
13. Коваленко И. Н., Кузнецов Н. Ю., Шуренков В. М. Случайные процессы: справочник. — Киев: Наук. думка, 1983. — 368 с.
14. Вероятность и математическая статистика: энциклопедия/ под ред. Ю. В. Прохорова. — М.: БРЭ, 2003. — 910 с.
15. Хименко В. И. Характеристики типа превышений уровней для случайных точечных процессов // Радиотехника и электроника. 2000. Т. 45. № 4. С. 436–443.
16. Гнеденко Б. В. Курс теории вероятностей. — М.: УРСС, 2007. — 400 с.

UDC 519.21:537.86

doi:10.15217/issn1684-8853.2016.4.85

Scatterplots to Analysis of Random Streams of Events

Khimenko V. I.^a, Dr. Sc., Tech., Professor, vih.13@yandex.ru

^aJSC “The Scientific Research and Experimental Centre of Intelligent Technologies “Petrokometa”, 67, B. Morskaiia St., 190000, Saint-Petersburg, Russian Federation

Purpose: Our goals are probabilistic analysis of random streams of events on the base of two-dimensional scatterplots, and studying a common structure of charts for several most common models of random streams. **Results:** We have shown the specific features of building scatterplots which are displaying two-dimensional information in the problems of probability analysis of random streams of events. We introduced simple numerical characteristics reflecting the information structure of random data on the phase plane. We performed probabilistic analysis of scatterplots for a class of simple Poisson streams, for random streams with inhomogeneous probabilistic structure, for streams with double stochasticity and for a class of random alternating streams of events. **Practical relevance:** Presenting two-dimensional information on a phase plane is highly informative, therefore scatterplots are potentially useful. They allow you to store, average out and systematize sampled data, quickly assess the homogeneity of the event streams, consider correlation effects, grouping, random “contamination” effects and randomly missed data.

Keywords — Information Processing, Random Processes, Event Streams, Point Processes, Scatterplots, Probabilistic Analysis, Inhomogeneous Data.

References

1. Vencel E. S. *Teoriia veroiatnostei* [Probability Theory]. Moscow, Akademiia Publ., 2005. 576 p. (In Russian).
2. Forster E., Ronz B. *Methoden der Korrelations und Regressionsanalyse*. Berlin, Verlag die Wirtschaft, 1979. (In German).
3. Borovikov V. P. *STATISTICA: iskusstvo analiza dannykh na komp'iutere* [STATISTICA: Data Analysis a Computer Oriented Approach]. Saint-Petersburg, Piter Publ., 2001. 656 p. (In Russian).
4. Uberla K. *Faktorenanalyse*. Berlin, Springer-Verlag, 1971.

5. Tou J. T., Gonzalez R. C. *Pattern Recognition Principles*. London, Addison — Wesley Publishing Company, 1974.
6. Iakovlev O. I. *Kosmicheskaja radiofizika* [Space Radio-Physics]. Moscow, Nauchnaia kniga Publ., 1998. 432 p. (In Russian).
7. Schowengerdt R. A. *Remote Sensing: Models and Methods for Image Processing*. Amsterdam, Elsevier, 2007.
8. Nosko V. P. *Ekonometrika* [Econometrics]. Moscow, Delo Publ., 2011. 672 p. (In Russian).
9. *Novye metody elektrokardiografii* [Modern Methods in Electrocardiography]. Ed. by S. V. Grachev, G. G. Ivanov, A. L. Syrkin. Moscow, Tekhnosfera Publ., 2007. 552 p. (In Russian).
10. Naatanen R. *Attention and Brain Function*. Hillsdale, New Jersey, Lawrence Erlbaum Associates Publishers, 1998. 560 p.
11. Snyder D. L. *Random Point Processes*. N. Y., John Wiley, 1975. 486 p.
12. Kingman J. *Poisson Processes*. Oxford, Clarendon Press, 2007. 136 p.
13. Kovalenko I. N., Kuznetsov N. Iu., Shurenkov V. M. *Sluchainye protsessy* [Random Processes]. Kiev, Naukova dumka Publ., 1983. 368 p. (In Russian).
14. *Veroiatnost' i matematicheskaja statistika* [Probability and Mathematical Statistics]. Ed. by Iu. V. Prokhorov. Moscow, BRE Publ., 2003. 910 p. (In Russian).
15. Khimenko V. I. Level Crossing Type Characteristics for Random Point Processes. *Radiotekhnika i elektronika* [Journal of Communications Technology and Electronics], 2000, vol. 45, no. 5, pp. 436–443 (In Russian).
16. Gnedenko B. V. *Kurs teorii veroiatnoستي* [Probability Theory]. Moscow, URSS Publ., 2007. 400 p. (In Russian).

Уважаемые авторы!

При подготовке рукописей статей необходимо руководствоваться следующими рекомендациями.

Статьи должны содержать изложение новых научных результатов. Название статьи должно быть кратким, но информативным. В названии недопустимо использование сокращений, кроме самых общепринятых (РАН, РФ, САПР и т. п.).

Объем статьи (текст, таблицы, иллюстрации и библиография) не должен превышать эквивалента в 20 страниц, напечатанных на бумаге формата А4 на одной стороне через 1,5 интервала Word шрифтом Times New Roman размером 13, поля не менее двух сантиметров.

Обязательными элементами оформления статьи являются: индекс УДК, заглавие, инициалы и фамилия автора (авторов), ученая степень, звание (при отсутствии — должность), полное название организации, аннотация и ключевые слова на русском и английском языках, электронные адреса авторов, которые по требованию ВАК должны быть опубликованы на страницах журнала. При написании аннотации не используйте аббревиатур и не делайте ссылок на источники в списке литературы.

Статьи авторов, не имеющих ученой степени, рекомендуется публиковать в соавторстве с научным руководителем, наличие подписи научного руководителя на рукописи обязательно; в случае самостоятельной публикации обязательно предоставляйте заверенную по месту работы рекомендацию научного руководителя с указанием его фамилии, имени, отчества, места работы, должности, ученого звания, ученой степени — эта информация будет опубликована в ссылке на первой странице.

Формулы набирайте в Word, не используя формульный редактор (MathType или Equation), при необходимости можно использовать формульный редактор; для набора одной формулы не используйте два редактора; при наборе формул в формульном редакторе знаки препинания, ограничивающие формулу, набирайте вместе с формулой; для установки размера шрифта никогда не пользуйтесь вкладкой Other..., используйте заводские установки редактора, не подгоняйте размер символов в формулах под размер шрифта в тексте статьи, не растягивайте и не сжимайте мышью формулы, вставленные в текст; в формулах не отделяйте пробелами знаки: + = -.

Для набора формул в Word никогда не используйте Конструктор (на верхней панели: «Работа с формулами» — «Конструктор»), так как этот ресурс предназначен только для внутреннего использования в Word и не поддерживается программами, предназначенными для изготовления оригинал-макета журнала.

При наборе символов в тексте помните, что символы, обозначаемые латинскими буквами, набираются светлым курсивом, русскими и греческими — светлым прямым, векторы и матрицы — прямым полужирным шрифтом.

Иллюстрации в текст не заверстаются и предоставляются отдельными исходными файлами, поддающимися редактированию:

— рисунки, графики, диаграммы, блок-схемы предоставляются в виде отдельных исходных файлов, поддающихся редактированию, используя векторные программы: Visio 4, 5, 2002-2003 (*.vsd); Coreldraw (*.cdr); Excel (*.xls); Word (*.doc); AdobeIllustrator (*.ai); AutoCad (*.dxf); Matlab (*.ps, *.pdf или экспорт в формат *.ai);

— если редактор, в котором Вы изготавливаете рисунок, не позволяет сохранить в векторном формате, используйте функцию экспорта (только по отношению к исходному рисунку), например, в формат *.ai, *.esp, *.wmf, *.emf, *.svg;

— фото и растровые — в формате *.tif, *.png с максимальным разрешением (не менее 300 pixels/inch).

Наличие подписанных подписей обязательно (желательно не повторяющих дословно комментарии к рисункам в тексте статьи).

В редакцию предоставляются:

— сведения об авторе (фамилия, имя, отчество, место работы, должность, ученое звание, учебное заведение и год его окончания, ученая степень и год защиты диссертации, область научных интересов, количество научных публикаций, домашний и служебный адреса и телефоны, e-mail), фото авторов: анфас, в темной одежде на белом фоне, должны быть видны плечи и грудь, высокая степень четкости изображения без теней и отблесков на лице, фото можно представить в электронном виде в формате *.tif, *.png с максимальным разрешением — не менее 300 pixels/inch при минимальном размере фото 40×55 мм;

— экспертное заключение.

Список литературы составляется по порядку ссылок в тексте и оформляется следующим образом:

— для книг и сборников — фамилия и инициалы авторов, полное название книги (сборника), город, издательство, год, общее количество страниц;

— для журнальных статей — фамилия и инициалы авторов, полное название статьи, название журнала, год издания, номер журнала, номера страниц;

— ссылки на иностранную литературу следует давать на языке оригинала без сокращений;

— при использовании web-материалов указывайте адрес сайта и дату обращения.

Список литературы оформляйте двумя отдельными блоками по образцам lit.dot на сайте журнала (<http://i-us.ru/paperrules>) по разным стандартам: Литература — СИБИБД РФ, References — один из мировых стандартов.

Более подробно правила подготовки текста с образцами изложены на нашем сайте в разделе «Оформление статей».

Контакты

Куда: 190000, Санкт-Петербург,
Б. Морская ул., д. 67, ГУАП, РИЦ
Кому: Редакция журнала «Информационно-управляющие системы»
Тел.: (812) 494-70-02
Эл. почта: ius.spb@gmail.com
Сайт: www.i-us.ru