

Лекции по случайным процессам

Широбоков М.Г.
ФПМИ МФТИ

21 ноября 2022 г.

Оглавление

Введение	5
Список обозначений и сокращений	6
Программа курса	7
1. Введение в случайные процессы	9
2. Пуассоновский процесс	17
3. Гауссовские процессы и векторы	22
3.1. Свойства гауссовских векторов	23
3.2. Винеровский процесс	26
4. Стохастический анализ	30
4.1. Свойства с.к.-сходимости	31
4.2. С.к.-непрерывность процессов	33
4.3. С.к.-дифференцируемость процессов	34
4.4. С.к.-интегрируемость процессов	36
4.5. Некоторые полезные формулы	38
5. Эргодические процессы	39
6. Стационарные процессы	41
6.1. Введение	41
6.2. Комплексные случайные процессы	45
6.3. Свойства корреляционной функции	46
6.4. Фурье-анализ стационарных процессов	49
6.5. Спектральная функция и плотность	51
6.6. Теорема Крамера	52
6.7. О физическом смысле спектральной плотности	56
6.8. О белом шуме	56
6.9. Линейная теория стационарных процессов	57
7. Марковские случайные процессы	59
8. Дискретные цепи Маркова	61
8.1. Базовые понятия и свойства	61
8.2. Производящие функции	67
8.3. Классификация состояний	68
8.4. Эргодические дискретные цепи Маркова	75

9. Непрерывные цепи Маркова	84
9.1. Эргодические непрерывные цепи Маркова	93
9.2. Процессы гибели и рождения	95
9.3. Взрывные марковские цепи	95
9.4. Потoki событий	97
9.5. Эквивалентные определения пуассоновского процесса	99
10. Непрерывные процессы Маркова	100
10.1. Базовые понятия и свойства	100
10.2. Диффузионные процессы	104
11. Вопросы на понимание	107
12. Задачи по курсу	110
12.1. Первое задание	110
12.2. Второе задание	113
Заключение	115
Литература	116

Введение

Пособие содержит семестровый курс лекций по «Случайным процессам», который автор читает студентам 3-го курса физтех-школы прикладной математики и информатики МФТИ на основе курса лекций, много лет читаемого на кафедре Математических основ управления. Курс охватывает ключевые разделы теории случайных процессов: гауссовские, стационарные, эргодические, марковские процессы, а также стохастический анализ.

В пособии уделяется внимание как формальным, так и прикладной стороне дела, даются многочисленные отсылки к разделам науки и технологиям, в которых активно применяется излагаемый материал, даются ссылки на полезную для более глубокого изучения литературу. Практически все доказательства теорем (или части доказательств теорем), не являющиеся необходимыми для понимания сути излагаемого материала и содержащие, с точки зрения автора, только технические выкладки, вынесены за пределы пособия. При этом даются ссылки на литературу, в которой хорошо излагается вынесенный материал. Теоремы без доказательств снабжаются обозначением * («звездочка»).

По традиции этот курс делится на две части: первые шесть разделов составляют содержание так называемого первого задания, а следующие четыре раздела посвящаются целиком марковским процессам и называются вторым заданием. Пособие самодостаточно и может использоваться обучающимися в течение семестра и при подготовке к экзаменам и сдаче заданий. В конце пособия читатель может найти списки вопросов на понимание, которые автор предлагает студентам во время приема заданий и экзаменов. Там же можно найти списки задач для сдачи заданий в течение семестра, эти задачи касаются всех понятий, вводимых в курсе. Замечания к тексту можно направлять на электронный адрес автора shirobokov@phystech.edu.

Список обозначений и сокращений

$(\Omega, \mathcal{F}, \mathbb{P})$ – вероятностное пространство (Ω – множество исходов, \mathcal{F} – сигма-алгебра, \mathbb{P} – вероятностная мера).

$\mathbb{E}X$ – математическое ожидание случайной величины X .

$\mathbb{D}X$ – дисперсия случайной величины X .

$\text{cov}(X, Y)$ – корреляционный момент (ковариация) случайных величин X и Y .

$\overset{\circ}{X}$ – «центрированная» случайная величина X , то есть $\overset{\circ}{X} = X - \mathbb{E}X$.

$\text{Be}(p)$ – распределение Бернулли.

$\text{Bi}(n, p)$ – биномиальное распределение, $\text{Bi}(1, p) = \text{Be}(p)$.

$\text{Po}(\lambda)$ – распределение Пуассона.

$\text{U}(a, b)$ – равномерное непрерывное распределение на отрезке $[a, b]$.

$\text{N}(\mu, \sigma^2)$ – нормальное (гауссовское) распределение с математическим ожиданием μ и дисперсией σ^2 .

$\text{Exp}(\lambda)$ – показательное распределение с параметром λ , плотность распределения $f(x) = \lambda \exp(-\lambda x)$, $x \geq 0$.

$\xrightarrow{\text{с.к.}}$ – сходимость в среднем квадратичном.

$\xrightarrow{\text{п.н.}}$ – сходимость почти наверное.

$\xrightarrow{\mathbb{P}}$ – сходимость по вероятности.

\xrightarrow{d} – сходимость по распределению.

l.i.m. – предел в среднем квадратичном (limit in mean).

$\stackrel{d}{=}$ – равенство по распределению.

$\stackrel{\text{п.н.}}{=}$ – равенство почти наверное.

\bar{z} – комплексное сопряжение комплексного числа z .

$I(A)$ – индикаторная функция события A : $I(A) = 1$, если A верно, и

$I(A) = 0$ – иначе.

с.к. – в среднем квадратичном.

п.н. – почти наверное.

х.ф. – характеристическая функция.

ЗБЧ – закон больших чисел.

ЦПТ – центральная предельная теорема.

НОД – наибольший общий делитель.

Программа курса

1. Определение понятия *случайный процесс*. Система конечно-мерных распределений случайного процесса, ее свойства. Выборочное пространство случайного процесса. Теорема о существовании непрерывной модификации случайного процесса. Моментные функции случайного процесса. Корреляционная и взаимная корреляционная функции случайных процессов, их свойства.

2. Пуассоновский процесс, его свойства. Теорема о явной конструкции пуассоновского процесса и ее следствия. Процессы восстановления. Сложный пуассоновский процесс, неоднородный пуассоновский процесс.

3. Свойства нормальных (гауссовских) векторов. Нормальный (гауссовский) случайный процесс. Винеровский процесс, его свойства. Теорема Вика. Винеровский процесс как предел случайных блужданий.

4. Непрерывность, дифференцируемость и интегрируемость по Риману случайного процесса в среднем квадратичном, почти наверное, по вероятности и по распределению. Критерии с.к.-непрерывности, с.к.-дифференцируемости и с.к.-интегрируемости по Риману. С.к.-интеграл Римана—Стилтьеса по случайному процессу и критерий его существования.

5. Эргодические по математическому ожиданию (дисперсии, корреляционной функции) в среднем квадратичном случайные процессы. Критерий и достаточное условие эргодичности по математическому ожиданию в среднем квадратичном.

6. Стационарные процессы в узком и широком смыслах. Свойства корреляционной функции стационарных процессов. Теорема Крамера о спектральном представлении стационарного случайного процесса. Теорема Хинчина о спектральном представлении корреляционной функции случайного процесса. Спектральная функция и спектральная плотность случайного процесса, их свойства и приложения. Случайный процесс типа «белый шум».

7. Марковский случайный процесс. Дискретная марковская цепь. Переходные вероятности, уравнения Колмогорова—Чепмена. Однородные дискретные марковские цепи. Классификация состояний дискретной марковской цепи, теорема о «солидарности» их свойств. Эргодические дискретные цепи Маркова, критерий эргодичности, теорема о предельном распределении. Закон больших чисел для марковских цепей.

8. Непрерывная марковская цепь. Переходные вероятности, уравнения Колмогорова—Чепмена. Стандартная непрерывная марковская

цепь. Прямые и обратные уравнения Колмогорова эволюции матрицы перехода цепи и распределения вероятностей состояний цепи. Q-матрица непрерывной марковской цепи. Время пребывания в состоянии и вероятности перехода между состояниями в момент прыжка. Эргодические непрерывные цепи Маркова, критерий эргодичности.

9. Процессы гибели и рождения. Взрывные непрерывные цепи Маркова. Потoki событий, свойства однородности, ординарности и отсутствия последействия. Системы массового обслуживания. Эквивалентные определения пуассоновского процесса.

10. Непрерывный марковский процесс. Переходная функция, уравнение Колмогорова—Чепмена. Необходимое и достаточное условие для того, чтобы нормальный (гауссовский) процесс был марковским процессом. Диффузионные процессы, прямые и обратные уравнения Колмогорова. Винеровский процесс как непрерывный марковский процесс.

1. Введение в случайные процессы

Бывают такие явления, развивающиеся во времени, предсказать поведение которых мы по каким-то причинам не можем. Неформально говоря, они и называются *случайными процессами*. Бывают разные причины, почему мы не можем предсказать поведение случайного процесса. Например, если каждый новый запуск процесса дает какое-то новое, может быть, неожиданное, его поведение. Процесс может быть и детерминированным в бытовом понимании этого слова, но бывает так, что определить будущее таких процессов тоже весьма не просто. Со случайными процессами мы встречаемся повсюду. Например если мы подкидываем монету, мы не знаем, на что она выпадет в очередной раз. Последовательность результатов бросков можно интерпретировать как случайный процесс. Другой пример – мы приходим на автобусную остановку и не знаем, когда придет очередной автобус. В этом случае случайный процесс – это индикатор присутствия автобуса на остановке (сначала он какое-то случайное время равен нулю, потом случайное время равен единице, потом снова нулю и так далее). Курс валюты или акций можно рассматривать как случайные процессы, потому что будущее однозначно не известно. Кривые обучения в машинном обучении каждый раз разные, когда мы перезапускаем обучение модели. Радиосигналы содержат случайные шумы, которые делают и весь сигнал случайным. Если поместить маленький шарик в жидкость, то можно наблюдать, как он «трясется» из-за столкновения с ним молекул жидкости, так что координаты шарика являются случайными процессами. Короче говоря, случайные процессы встречаются в быту, в финансовой математике, машинном обучении, IT-сфере, математической физике, обработке сигналов и много еще где. В общем случае мы не можем однозначно определить, как будет развиваться случайный процесс. Однако мы можем попробовать вычислять или оценивать вероятности того, что процесс поведет себя тем или иным образом, вычислять или оценивать поведение процесса в среднем, разброс его значений. Для этого нужна только подходящая модель случайного явления, чтобы ее можно было содержательно изучать. Эту модель мы построим на базе теории вероятностей.

Определение. *Случайным процессом* называется однопараметрическое семейство $\{\xi(\omega, t), t \in T \subseteq \mathbb{R}\}$ случайных величин, определенных на одном и том же вероятностном пространстве.

Параметр t обычно называют *временем*. Переменная ω называется *исходом*, это элемент пространства исходов. А именно, если процесс определен на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$, где как обычно

Ω – пространство исходов, \mathcal{F} – сигма-алгебра подмножеств множества Ω , \mathbb{P} – вероятностная мера на \mathcal{F} , то ω принимает значения в Ω . Эта переменная часто опускается, и тогда мы пишем, например, что дан процесс $\xi(t)$, но, конечно, при этом подразумевается, что он зависит и от исхода тоже. Когда же нам важно подчеркнуть зависимость от исхода у процесса, мы эту переменную пишем явно.

Для каждого фиксированного t_0 из множества T функция $\xi(\omega, t_0)$ представляет собой случайную величину. Она называется *сечением* процесса в момент времени t_0 . Если зафиксировать исход $\omega_0 \in \Omega$, то функция времени $\xi(\omega_0, t)$ называется *траекторией (реализацией)* процесса, отвечающей исходу ω_0 . Это неслучайная, детерминированная функция. Когда мы наблюдаем в реальности какой-то процесс, мы наблюдаем именно траекторию процесса для какого-то неизвестного исхода ω_0 . Из-за того, что мы не знаем этот исход и процесс как функцию исхода и времени, нам не известно будущее процесса. Кроме того, бывает так, что разным исходам соответствует одно и то же прошлое и настоящее процесса, но разное будущее.

Для того чтобы можно было вычислять или оценивать вероятности каких-то условий, связанных с процессом, можно ввести понятие *о распределении процессов* (аналогично распределениям случайных величин). Для случайных величин главной характеристикой распределения является функция распределения. Для случайных процессов – это *семейство конечномерных распределений*. Перейдем к определению.

Пусть даны $n \geq 1$ сечений: $\xi(\omega, t_1), \dots, \xi(\omega, t_n)$. Так как все сечения принадлежат одному и тому же вероятностному пространству, то и вектор

$$(\xi(\omega, t_1), \dots, \xi(\omega, t_n)),$$

составленный из этих сечений, определен на том же вероятностном пространстве и является случайным вектором. Запишем его функцию распределения:

$$F_\xi(x_1, \dots, x_n; t_1, \dots, t_n) = \mathbb{P}(\omega : \xi(\omega, t_1) < x_1, \dots, \xi(\omega, t_n) < x_n). \quad (1)$$

Здесь следует обратить внимание на то, что эта функция распределения зависит от двух вещей: как обычно для функций распределения – от x_i , но теперь и от моментов времени t_i , потому что при разных наборах t_i сечения могут иметь различные распределения. Функция (1) называется *n-мерной функцией распределения* случайного процесса $\xi(t)$.

Определение. Совокупность функций (1) для всех $n \geq 1$ и $t_1, \dots, t_n \in T$ называется *семейством конечномерных распределений* процесса $\{\xi(t), t \in T\}$.

Нетрудно показать, что семейство конечномерных распределений произвольного случайного процесса $\xi(t)$ обладает следующими свойствами:

1. $0 \leq F_\xi(x_1, \dots, x_n; t_1, \dots, t_n) \leq 1$.
2. Функции $F_\xi(x_1, \dots, x_n; t_1, \dots, t_n)$ непрерывны слева по каждой переменной x_i .
3. Если хотя бы одна из переменных $x_i \rightarrow -\infty$, то

$$F_\xi(x_1, \dots, x_n; t_1, \dots, t_n) \rightarrow 0,$$

если же все переменные $x_i \rightarrow +\infty$, то

$$F_\xi(x_1, \dots, x_n; t_1, \dots, t_n) \rightarrow 1.$$

4. Функции $F_\xi(x_1, \dots, x_n; t_1, \dots, t_n)$ монотонны в следующем смысле:

$$\Delta_1 \dots \Delta_n F_\xi(x_1, \dots, x_n; t_1, \dots, t_n) \geq 0,$$

где Δ_i – линейный оператор конечной разности по переменной x_i :

$$\begin{aligned} \Delta_i F &= F(x_1, \dots, x_{i-1}, x_i + h_i, x_{i+1}, \dots, x_n; t_1, \dots, t_n) - \\ &\quad - F(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n; t_1, \dots, t_n), \end{aligned}$$

а $h_1 \geq 0, \dots, h_n \geq 0$ произвольны.

5. Для любой перестановки $\{k_1, \dots, k_n\}$ индексов $\{1, \dots, n\}$

$$F_\xi(x_1, \dots, x_n; t_1, \dots, t_n) = F_\xi(x_{k_1}, \dots, x_{k_n}; t_{k_1}, \dots, t_{k_n}).$$

6. Для любых $1 \leq k < n$ и $x_1, \dots, x_k \in \mathbb{R}$

$$F_\xi(x_1, \dots, x_k; t_1, \dots, t_k) = F_\xi(x_1, \dots, x_k, +\infty, \dots, +\infty; t_1, \dots, t_n).$$

Поясним пункт 4 в этом списке. Заметим, что

$$\Delta_1 F_\xi(x_1; t_1) = F_\xi(x_1 + h_1; t_1) - F_\xi(x_1; t_1) = \mathbb{P}(\xi(t_1) \in [x_1, x_1 + h_1]),$$

то есть вероятности того, что в момент t_1 процесс принимает значения из интервала $[x_1, x_1 + h_1)$. Также несложно видеть, что

$$\begin{aligned} \Delta_1 \Delta_2 F_\xi(x_1, x_2; t_1; t_2) &= \Delta_1 (F_\xi(x_1, x_2 + h_2; t_1, t_2) - F_\xi(x_1, x_2; t_1, t_2)) = \\ &= F_\xi(x_1 + h_1, x_2 + h_2; t_1, t_2) - F_\xi(x_1, x_2 + h_2; t_1, t_2) - \\ &\quad - F_\xi(x_1 + h_1, x_2; t_1, t_2) + F_\xi(x_1, x_2; t_1, t_2) = \end{aligned}$$

$$= \mathbb{P}(\xi(t_1) \in [x_1, x_1 + h_1), \xi(t_2) \in [x_2, x_2 + h_2)).$$

Последнее равенство вытекает из геометрических соображений и формулы включений–исключений. В общем же случае

$$\begin{aligned} & \Delta_1 \dots \Delta_n F_\xi(x_1, \dots, x_n; t_1, \dots, t_n) = \\ & = \mathbb{P}(\xi(t_1) \in [x_1, x_1 + h_1), \dots, \xi(t_n) \in [x_n, x_n + h_n)). \end{aligned}$$

Оказывается, справедлива следующая теорема.

Теорема Колмогорова о существовании процесса*. Пусть задано семейство функций

$$F = \{F(x_1, \dots, x_n; t_1, \dots, t_n), x_i \in \mathbb{R}, t_i \in T, i = 1, \dots, n, n \geq 1\},$$

удовлетворяющее условиям 1–6. Тогда существуют вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и случайный процесс $\{\xi(t), t \in T\}$, определенный на этом вероятностном пространстве, такие, что семейство конечномерных распределений F_ξ случайного процесса $\xi(t)$ совпадает с семейством функций F .

Смысл этой теоремы состоит в двух вещах. Во-первых, из нее следует, что определять процессы можно не только с помощью функций $\xi(\omega, t)$ двух переменных, но и через задание семейства конечномерных распределений. Вообще случайные процессы удобно задавать именно через семейство конечномерных распределений, а не как функции двух аргументов, точно так же как и случайные величины обычно задают функцией распределения, а не как функции исхода. Существование различных случайных процессов с одними и теми же вероятностными свойствами приводит к желанию (а иногда и необходимости) в некотором смысле отождествлять процессы, у которых конечномерные распределения совпадают.

Определение. Пусть $\{\xi(t), t \in T\}$ и $\{\eta(t), t \in T\}$ – два случайных процесса, определенные на одном и том же вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$. Тогда если для любого $t \in T$

$$\mathbb{P}(\omega : \xi(\omega, t) = \eta(\omega, t)) = 1,$$

то эти процессы называются *стохастически эквивалентными* или просто *эквивалентными*. При этом $\xi(t)$ называется *модификацией* процесса $\eta(t)$, а $\eta(t)$ – *модификацией* процесса $\xi(t)$.

В качестве упражнения предлагается доказать, что если процессы стохастически эквивалентны, то они имеют равные n -мерные функции распределения и одинаковые семейства конечномерных распределений.

Во-вторых, у теоремы Колмогорова есть конструктивное доказательство, в котором явно предлагаются пространство, на котором определен процесс, и процесс как функция времени и исхода. Конечно, и пространство, и соответствующая функция $\xi(\omega, t)$ не обязательно определены однозначно.

Вообще, касаясь темы задания процессов на вероятностных пространствах, удобно ввести понятие *выборочного (вторичного) пространства* для случайных процессов. Это можно сделать, например, таким образом. Рассмотрим пространство \mathcal{X} функций $x(t)$, $t \in T$, в котором лежат траектории процесса $\xi(t)$, $t \in T$. Пусть далее $\mathcal{B}_{\mathcal{X}}^T$ есть σ -алгебра подмножеств из \mathcal{X} , порожденная множествами

$$C = \{x \in \mathcal{X} : x(t_1) \in B_1, \dots, x(t_n) \in B_n\}$$

при варьировании $n \geq 1$, времен $t_1, \dots, t_n \in T$ и борелевских множеств B_1, \dots, B_n . Множества такого вида называются *цилиндрическими*. Отображение $\xi(\omega, t)$ определяет измеримое отображение (Ω, \mathcal{F}) в $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}^T)$, потому что

$$\{\omega : \xi(\omega, \cdot) \in C\} \in \mathcal{F}$$

для любого цилиндрического множества C , а стало быть,

$$\{\omega : \xi(\omega, \cdot) \in B\} \in \mathcal{F}$$

для любого множества $B \in \mathcal{B}_{\mathcal{X}}^T$. Наконец, на пространстве $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}^T)$ вероятностную меру можно определить так:

$$\mathbb{P}_{\xi}(B) = \mathbb{P}(\omega : \xi(\omega, \cdot) \in B), \quad B \in \mathcal{B}_{\mathcal{X}}^T.$$

Вероятностное пространство $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}^T, \mathbb{P}_{\xi})$ называется *выборочным (вторичным) вероятностным пространством*. В этом пространстве исход ω отождествляется с траекторией процесса, а мера \mathbb{P}_{ξ} называется *распределением процесса* ξ . Для простоты мы почти всегда будем писать \mathbb{P} вместо \mathbb{P}_{ξ} .

Множество T может быть разным. Если T состоит из одного элемента, то, по сути, случайный процесс представляет собой одну случайную величину. Если T состоит из $n > 1$ элементов, то случайный процесс можно представить как случайный вектор с n компонентами. Если T состоит из счетного числа элементов, например $T = \mathbb{N}$, то случайный процесс представляет собой случайную последовательность. Если множество T непрерывное, например $T = [a, b]$, где $-\infty < a < b < \infty$, $T = \mathbb{R}$, $T = [0, +\infty)$, то такой случайный процесс обычно называется *случайной функцией*. В этом курсе нас будут в первую очередь интересовать случайные последовательности и случайные функции.

В дальнейшем нам придется вычислять вероятности событий, связанных с процессами. Например, вероятности $\mathbb{P}(\xi(t) \in [a, b])$, $\mathbb{P}(\xi(t_1) \in [a_1, b_1], \xi(t_2) \in [a_2, b_2])$ и вероятности более сложно устроенных событий (например, вероятности того, что процесс возрастает на участке времени, максимум процесса на участке времени не превышает некоторого уровня, на некотором участке времени найдется точка, в которой процесс обращается в ноль). В принципе, события могут касаться процесса в конечном, счетном и даже несчетном числе моментов времени.

И вот тут, когда речь заходит о событиях, связанных с процессом в несчетном числе моментов времени, нужно быть особенно аккуратным. Например, пусть на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$ дан какой-нибудь процесс $\{\xi(t), t \geq 0\}$ с непрерывным временем и непрерывным множеством значений на \mathbb{R} , и, допустим, что нам требуется вычислить следующую вероятность:

$$\mathbb{P} \left(\sup_{t \in [0,1]} \xi(t) \leq x \right).$$

Напомню, что вероятность определена лишь для элементов сигма-алгебры \mathcal{F} , но принадлежит ли событие, записанное под вероятностью, этой сигма-алгебре? Супремум процесса не превосходит x на отрезке $[0, 1]$ тогда и только тогда, когда в каждый момент времени $t \in [0, 1]$ выполнено $\xi(t) \leq x$. И хотя событие $\{\xi(t) \leq x\}$ является элементом сигма-алгебры \mathcal{F} , несчетное пересечение таких событий для разных $t \in [0, 1]$ может не быть элементом сигма-алгебры. Однако если все или почти все траектории процесса $\xi(t)$ непрерывны на $[0, 1]$, то супремум по $[0, 1]$ можно заменить супремумом по рациональным точкам отрезка $[0, 1]$, а это счетное пересечение событий сигма-алгебры. А вообще, траектории этого процесса не обязаны быть непрерывными для того, чтобы существовала эта вероятность. Достаточно лишь, чтобы у процесса существовала его непрерывная модификация (модификация с почти всеми непрерывными траекториями).

Установить существование непрерывной модификации процесса помогает следующая теорема.

Теорема Колмогорова существования непрерывной модификации. Пусть $\{\xi(t), t \in T\}$ – случайный процесс и $T = [a, b]$. Если существуют $\alpha > 0$, $\beta > 0$, $c < \infty$, такие, что при всех $t, t+h \in [a, b]$ выполнено неравенство*

$$\mathbb{E} |\xi(t+h) - \xi(t)|^\alpha \leq c|h|^{1+\beta},$$

то $\xi(t)$ имеет непрерывную модификацию.

Можно доказать, что если непрерывная модификация процесса $\xi(t)$ существует, то всегда можно так определить вероятностное пространство и сигма-алгебру \mathcal{F} (с сохранением семейства конечномерных распределений), на которых задан этот процесс, что условия на процесс в несчетном числе точек можно заменить на условия в счетном числе точек, и множества типа $\{\sup_{t \in [0,1]} \xi(t) \leq x\}$ становятся измеримыми, вероятность их определена.

Определение. *Математическим ожиданием случайного процесса $\{X(t), t \in T\}$* называется функция $m_X : T \rightarrow \mathbb{R}$, значение которой в каждый момент времени равно математическому ожиданию соответствующего сечения, то есть $\forall t \in T \ m_X(t) = \mathbb{E}X(t)$.

В общем случае может оказаться так, что в какие-то моменты времени математическое ожидание процесса существует, а в какие-то моменты оно не существует. Кроме того, заметим, что в терминах одномерного распределения процесса его математическое ожидание равно

$$m_X(t) = \mathbb{E}X(t) = \int x dF_X(x; t),$$

где интеграл понимается в смысле интеграла Лебега–Стилтьеса. Процессы, у которых математическое ожидание всюду существует и равно нулю, называются *центрированными*.

Определение. *Корреляционной функцией случайного процесса $\{X(t), t \in T\}$* называется функция двух переменных $R_X : T \times T \rightarrow \mathbb{R}$, которая каждой паре моментов времени сопоставляет корреляционный момент соответствующих сечений процесса:

$$\begin{aligned} R_X(t_1, t_2) &= \text{cov}(X(t_1), X(t_2)) = \mathbb{E}\overset{\circ}{X}(t_1)\overset{\circ}{X}(t_2) = \\ &= \mathbb{E}X(t_1)X(t_2) - \mathbb{E}X(t_1)\mathbb{E}X(t_2). \end{aligned}$$

Корреляционная функция определяется через корреляционный момент сечений процесса, поэтому корреляционная функция наследует все свойства корреляционных моментов, среди которых, например,

- 1) $R_X(t, t) = \text{cov}(X(t), X(t)) = \mathbb{D}X(t) \geq 0$,
- 2) $R_X(t_1, t_2) = R_X(t_2, t_1)$ (симметричность),
- 3) $R_X^2(t_1, t_2) \leq R_X(t_1, t_1)R_X(t_2, t_2)$.

Неравенство в третьем пункте прямо следует из неравенства Коши–Буняковского для скалярного произведения между случайными величинами $(\xi_1, \xi_2) \doteq \mathbb{E}\xi_1\xi_2$.

Определение. *Ковариационной функцией случайного процесса $\{X(t), t \in T\}$* называется функция двух переменных $K_X : T \times T \rightarrow \mathbb{R}$,

которая каждой паре моментов времени сопоставляет число

$$K_X(t_1, t_2) = \mathbb{E}X(t_1)X(t_2).$$

Заметим по поводу корреляционной и ковариационной функций, что эта терминология соответствует нашему классическому учебнику по случайным процессам¹, но прямо противоположна терминологии, используемой в некоторых других источниках, например в учебнике².

Как и в случае математического ожидания, корреляционная и ковариационная функции в самом общем случае не обязаны быть определены для всех пар (t_1, t_2) , как и не обязаны быть определены хотя бы для одной пары времен. Корреляционную и ковариационную функции можно выразить в терминах двумерного распределения процесса следующим образом:

$$R_X(t_1, t_2) = \int (x - m_X(t_1))(y - m_X(t_2)) dF_X(x, y; t_1, t_2),$$

$$K_X(t_1, t_2) = \int xy dF_X(x, y; t_1, t_2),$$

где интегралы понимаются в смысле Лебега–Стилтьеса. Между собой эти функции связаны формулой

$$R_X(t_1, t_2) = K_X(t_1, t_2) - m_X(t_1)m_X(t_2).$$

Определение. *Дисперсией случайного процесса* $\{X(t), t \in T\}$ называется функция $D_X : T \rightarrow \mathbb{R}$, в каждый момент времени равная дисперсии соответствующего сечения процесса, то есть

$$D_X(t) = \mathbb{E}\overset{\circ}{X}^2(t) = \mathbb{E}X^2(t) - (\mathbb{E}X(t))^2.$$

Легко видеть, что дисперсия процесса связана с корреляционной и ковариационной функциями по формулам

$$D_X(t) = R_X(t, t) = K_X(t, t) - m_X^2(t).$$

Определение. *Взаимной корреляционной функцией* процессов $\{X(t), t \in T\}$ и $\{Y(t), t \in T\}$ называется функция двух переменных $R_{X,Y} : T \times T \rightarrow \mathbb{R}$ такая, что

$$R_{X,Y}(t_1, t_2) = \mathbb{E}\overset{\circ}{X}(t_1)\overset{\circ}{Y}(t_2).$$

¹Натан А.А., Горбачев О.Г., Гуз С.А. Основы теории случайных процессов. Москва : МЗ-Пресс, 2003.

²Миллер Б.М., Панков А.Р. Теория случайных процессов в примерах и задачах. – Москва : ФИЗМАТЛИТ, 2002.

Определение. *Характеристической функцией* случайного процесса $\{X(t), t \in T\}$ называется функция

$$\varphi_{X(t)}(s) = \mathbb{E} \exp(isX(t)), \quad s \in \mathbb{R}.$$

Характеристическая функция определена всюду на \mathbb{R} для любой случайной величины и для любого случайного процесса.

2. Пуассоновский процесс

Введем теперь наш первый содержательный и чрезвычайно важный случайный процесс – пуассоновский процесс. Представим себе последовательность наступающих время от времени каких-то событий. Пусть эти события происходят через случайные интервалы времени, тогда число событий, наступивших в интервале времени $[0, t]$ обозначим за $K(t)$. Очевидно, всякая реализация такого процесса (счетчика событий) есть неубывающая ступенчатая функция.

Что может описывать такой процесс? Число фотонов, испускаемых веществом при радиоактивном распаде, число телефонных вызовов из данного района, число происшествий на данном перекрестке, число ошибок на странице машинописного текста, число выходов из строя некоторого механизма, число поступивших заявок на обслуживание и многое другое.

На практике естественно предполагать, что число событий на одном интервале времени не зависит от числа событий на другом интервале времени. Кроме того, часто естественно предполагать, что случайная величина $K(t+h) - K(t)$ зависит только от h и не зависит от t . Наконец, часто бывает естественным моделировать число событий на интервале как пуассоновскую случайную величину (отсюда и название процесса). Выбор такой модели основывается на законе редких событий (предельной теореме Пуассона). Представим себе большое число испытаний Бернулли с малой вероятностью успеха (наступления события) и постоянным средним числом успеха. Формально говоря, пусть имеются случайные величины $\{\xi_k\}_{k=1}^n$, каждая из которых имеет распределение $\xi_k \in \text{Be}(p_n)$, причем $np_n \rightarrow \lambda > 0$, $n \rightarrow \infty$. Итак, пуассоновские процессы моделируют ситуации, когда имеется большое количество чего-то, что редко наступает, и представляют собой счетчики событий, наступление которых на одних интервалах времени не зависит от наступления их на непересекающихся интервалах времени. Теперь введем формальное определение пуассоновского процесса.

Определение. Случайная функция $\{X(t), t \geq 0\}$ называется

процессом с независимыми приращениями, если для любого $n \geq 1$ и любых моментов времени $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ случайные величины

$$X(0), X(t_1) - X(0), X(t_2) - X(t_1), \dots, X(t_n) - X(t_{n-1})$$

независимы в совокупности.

Определение. Пуассоновским процессом с интенсивностью $\lambda > 0$ называется случайная функция $\{K(t), t \geq 0\}$, удовлетворяющая свойствам:

1. $K(0) = 0$ почти всюду.
2. $K(t)$ – процесс с независимыми приращениями.
3. $\forall t > s \geq 0$ выполнено $K(t) - K(s) \in \text{Po}(\lambda(t - s))$.

Что можно сразу понять про пуассоновский процесс, глядя на это определение? На шкале времени он определен на положительной полуоси и начинается из нуля. Так как приращения процесса $K(t) - K(s)$ для $t > s$ имеют пуассоновское распределение, то они могут принимать только значения, которые принимает пуассоновская случайная величина, то есть 0, 1, 2 и так далее. А это значит, что процесс принимает значения только 0, 1, 2 и так далее, причем траектории этого процесса монотонно не убывают и в какие-то моменты времени испытывают прыжки от меньших натуральных значений к большим натуральным значениям.

Далее, так как $K(t) = K(t) - K(0) \in \text{Po}(\lambda t)$, то сразу получаем

$$\mathbb{E}K(t) = \mathbb{D}K(t) = \lambda t.$$

Таким образом, математическое ожидание и дисперсия пуассоновского процесса определены всюду. Найдем корреляционную функцию

$$R_K(t, s) = \mathbb{E}K(t)K(s) - \mathbb{E}K(t)\mathbb{E}K(s) = \mathbb{E}K(t)K(s) - \lambda^2 ts.$$

Сначала предположим, что $t > s$. В этом случае запишем:

$$\begin{aligned} R_K(t, s) &= \mathbb{E}(K(t) - K(s))K(s) + \mathbb{E}K^2(s) - \lambda^2 ts = \\ &= \mathbb{E}(K(t) - K(s))\mathbb{E}K(s) + \mathbb{D}K(s) + (\mathbb{E}K(s))^2 - \lambda^2 ts = \\ &= \lambda^2(t - s)s + \lambda s + \lambda^2 s^2 - \lambda^2 ts = \lambda s. \end{aligned}$$

Аналогично для $t < s$ мы получим $R_K(t, s) = \lambda t$. Если же $t = s$, то

$$R_K(t, s) = \mathbb{E}K^2(t) - \lambda^2 t^2 = \lambda t + \lambda^2 t^2 - \lambda^2 t^2 = \lambda t.$$

Итак, в самом общем случае мы получаем, что

$$R_K(t, s) = \lambda \min(t, s).$$

Ковариационная функция пуассоновского процесса равна

$$K_K(t, s) = R_K(t, s) + \mathbb{E}K(t)\mathbb{E}K(s) = \lambda \min(t, s) + \lambda^2 ts.$$

Видно, что обе эти функции определены для всех $(t, s) \in [0, +\infty)^2$ и что если $t, s > 0$, то $R_K(t, s) \neq 0$, поэтому сечения пуассоновского процесса являются зависимыми случайными величинами.

Теперь зададимся следующими вопросами: в какие моменты времени осуществляются прыжки пуассоновского процесса, как распределены интервалы между прыжками, зависимы или независимы они, на какие величины прыгает процесс? Ответить разом на все эти вопросы и даже получить конструктивное (а не аксиоматическое) описание процесса нам позволит следующая теорема.

Теорема (явная конструкция пуассоновского процесса). Пусть $\{\xi_n\}$ – последовательность независимых случайных величин с распределением $\text{Exp}(\lambda)$. Обозначим $S_n = \xi_1 + \dots + \xi_n$ ($S_0 = 0$). Введем процесс

$$X(t) = \sup\{n : S_n \leq t\}.$$

Тогда $X(t)$ – пуассоновский процесс интенсивности λ .

Доказательство. Рассмотрим случайный вектор (S_1, \dots, S_n) . Этот вектор связан со случайным вектором (ξ_1, \dots, ξ_n) линейным преобразованием:

$$\begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \vdots \\ \xi_n \end{bmatrix}.$$

Это наблюдение позволяет вычислить плотность вектора (S_1, \dots, S_n) . Действительно, определитель матрицы равен единице, поэтому плотность вектора (S_1, \dots, S_n) равна

$$p_{S_1, \dots, S_n}(x_1, \dots, x_n) = p_{\xi_1, \dots, \xi_n}(x_1, x_2 - x_1, \dots, x_n - x_{n-1}).$$

Остается вспомнить, что случайные величины ξ_1, \dots, ξ_n независимы, значит,

$$\begin{aligned} p_{S_1, \dots, S_n}(x_1, \dots, x_n) &= \prod_{j=1}^n \lambda e^{-\lambda(x_j - x_{j-1})} I(x_j - x_{j-1} > 0) = \\ &= \lambda^n e^{-\lambda x_n} I(x_n > x_{n-1} > \dots > x_1). \end{aligned}$$

Зафиксируем моменты времени $0 = t_0 < t_1 < \dots < t_n$. Пусть $k_1, \dots, k_n \in \mathbb{Z}_+$ и $k_1 \leq k_2 \leq \dots \leq k_n$. Чтобы доказать, что процесс $X(t)$ пуассоновский, нужно доказать, что приращения его независимые и имеют пуассоновское распределение. Для этого мы вычислим вероятность

$$\mathbb{P}(X(t_n) - X(t_{n-1}) = k_n - k_{n-1}, \dots, X(t_2) - X(t_1) = k_2 - k_1, X(t_1) = k_1),$$

покажем, что она распадается на произведение вероятностей и что эти вероятности равны тому, чему нужно, чтобы эти случайные величины имели пуассоновское распределение. Эта вероятность равна

$$\begin{aligned} & \mathbb{P}(\{S_1, \dots, S_{k_1}\} \in (0, t_1), \{S_{k_1+1}, \dots, S_{k_2}\} \in (t_1, t_2), \dots, \\ & \dots \{S_{k_{n-1}+1}, \dots, S_{k_n}\} \in (t_{n-1}, t_n), S_{k_n+1} > t_n) = \\ & = \int \dots \int \lambda^{k_n+1} e^{-\lambda x_{k_n+1}} I(0 < x_1 < \dots < x_{k_n+1}) dx_1 \dots dx_{k_n+1}, \end{aligned}$$

где интеграл берется по области

$$A = \{(x_1, \dots, x_{k_n+1}) \in \mathbb{R}^{k_n+1} : \{x_1, \dots, x_{k_1}\} \in (0, t_1), \dots, x_{k_n+1} > t_n\}.$$

Далее, интеграл выше равен

$$\begin{aligned} & \int_{t_n}^{\infty} \lambda e^{-\lambda x_{k_n+1}} dx_{k_n+1} \lambda^{k_n} \prod_{j=1}^n \int \dots \int I(x_{k_{j-1}+1} < \dots < x_{k_j}) dx_1, \dots, dx_{k_n} = \\ & = e^{-\lambda t_n} \lambda^{k_n} \prod_{j=1}^n \frac{(t_j - t_{j-1})^{k_j - k_{j-1}}}{(k_j - k_{j-1})!} = \prod_{j=1}^n \frac{(\lambda(t_j - t_{j-1}))^{k_j - k_{j-1}}}{(k_j - k_{j-1})!} e^{-\lambda(t_j - t_{j-1})}. \end{aligned}$$

Интеграл в выражении выше без индикатора равен объему многомерного прямоугольника, а с индикатором – объему симплекса, который равен объему прямоугольника поделенному на факториал размерности пространства. По индукции из полученного выражения следует, что все случайные величины $X(t_j) - X(t_{j-1})$ независимы, причем вероятность того, что

$$\mathbb{P}(X(t_j) - X(t_{j-1}) = k_j - k_{j-1}) = \frac{(\lambda(t_j - t_{j-1}))^{k_j - k_{j-1}}}{(k_j - k_{j-1})!} e^{-\lambda(t_j - t_{j-1})},$$

то есть $X(t_j) - X(t_{j-1}) \in \text{Po}(\lambda(t_j - t_{j-1}))$. А это и есть то, что мы хотели доказать: приращения процесса независимы и распределены по Пуассону с нужными параметрами. \square

Замечание. Мы доказали, что процесс $X(t)$ является пуассоновским процессом. Это значит, что у этих процессов семейства конечномерных распределений совпадают, поэтому все вероятностные свойства одного процесса совпадают с вероятностными свойствами другого процесса. Доказывать, что всякий пуассоновский процесс представим в виде процесса $X(t)$, не нужно.

Теперь разберемся с тем, какие следствия мы получаем из явной конструкции пуассоновского процесса.

1) Скачки происходят в моменты времени $\tau_1 = S_1$, $\tau_2 = S_2$ и так далее. Так как S_n представляет собой сумму n независимых случайных величин с распределением $\text{Exp}(\lambda)$, то $\tau_n = S_n$ имеет распределение Эрланга $\text{Erl}(n, \lambda)$. Распределение моментов τ_n можно было бы получить и напрямую, не пользуясь теоремой о явном виде. Например,

$$\mathbb{P}(\tau_1 > t) = \mathbb{P}(K(t) = 0) = \exp(-\lambda t), \quad t \geq 0,$$

что и означает показательное распределение.

2) Между скачками проходит случайное время

$$\tau_n - \tau_{n-1} = S_n - S_{n-1} = \xi_n \in \text{Exp}(\lambda).$$

Времена между соседними последовательными скачками – независимые случайные величины.

3) С вероятностью 1 все скачки пуассоновского процесса являются единичными. Действительно, скачки происходят только в моменты времени $\tau_1 = S_1$, $\tau_2 = S_2$ и так далее, поэтому

$$\mathbb{P}(\exists \text{ скачок размера } \geq 2) = \mathbb{P}(\exists n : S_n = S_{n+1}) = \mathbb{P}(\exists n : \xi_{n+1} = 0) = 0,$$

так как все ξ_j имеют непрерывное распределение.

Процессы вида

$$X(t) = \sup \left\{ n : \sum_{k=1}^n \xi_k \leq t \right\},$$

где ξ_k – независимые случайные величины (не обязательно одинаково или показательно распределенные) называются еще *процессами восстановления*. Теорема, сформулированная и доказанная выше, говорит о том, что пуассоновский процесс – это процесс восстановления, построенный по случайным величинам с показательным распределением.

Термин *процесс восстановления* связан с тем, что $X(t)$ и S_n часто используют для описания работы различных физических устройств. Если, например, ξ_k – время работы какого-то устройства, после которого

он выходит из строя и требует замены или «восстановления», то S_n будет означать n -й момент восстановления прибора. Если считать $S_0 = 0$, то $X(t)$ будет означать число восстановлений, которые произошли до момента t .

Существует множество обобщений пуассоновского процесса, приведем некоторые из них:

1) Пуассоновский процесс можно записать в виде

$$K(t) = \sum_{i=1}^{K(t)} 1,$$

то есть как сумму единиц со случайным числом слагаемых (полагаем сумму равной нулю, если $K(t) = 0$). Процесс

$$K_c(t) = \sum_{i=1}^{K(t)} \xi_i$$

для случайных величин ξ , независимых между собой и от пуассоновского процесса $K(t)$, называется *сложным пуассоновским процессом* (*compound Poisson process*). Если ξ_i распределены как $\text{Be}(p)$, то $K_c(t)$ – это процесс, который получен из пуассоновского после удаления из него каждого скачка с вероятностью $1 - p$. Можно доказать, что этот процесс тоже является пуассоновским, но с интенсивностью λp , если у исходного пуассоновского процесса интенсивность была λ . В общем случае сложный пуассоновский процесс пуассоновским может не являться.

2) Мы выше предполагали, что интенсивность процесса λ постоянна во времени, но можно определить пуассоновский процесс с интенсивностью, зависящей от времени, если в качестве третьего пункта определения взять

$$K(t) - K(s) \in \text{Po} \left(\int_s^t \lambda(\tau) d\tau \right), \quad t > s,$$

где $\lambda(\tau)$ – неотрицательная интегрируемая на любом ограниченном борелевском множестве функция. Такой процесс называется *неоднородным пуассоновским процессом* (*inhomogeneous Poisson process*).

3. Гауссовские процессы и векторы

Определение. Случайный процесс $\{X(t), t \in T\}$ называется *гауссовским* (*нормальным*), если все его конечномерные распределения

являются нормальными, то есть для любых $n \geq 1$ и любых $t_1, \dots, t_n \in T$ случайный вектор $(X(t_1), \dots, X(t_n))$ имеет нормальное распределение.

Изучение нормальных процессов часто сводится к работе над векторами из их сечений, а эти векторы из сечений являются нормальными. Полезно поэтому напомнить определение и свойства нормальных случайных векторов.

Напомним, что по определению вектор $X = (X_1, \dots, X_n)$ имеет нормальное распределение, если его характеристическая функция имеет вид

$$\varphi_X(s) = \exp\left(i\mu^T s - \frac{1}{2}s^T R s\right)$$

для любого $s \in \mathbb{R}^n$. Здесь вектор $\mu = \mathbb{E}X \in \mathbb{R}^n$ – математическое ожидание вектора X , а матрица

$$R = \mathbb{E}\overset{\circ}{X}\overset{\circ}{X}^T = \mathbb{E}(X - \mathbb{E}X)(X - \mathbb{E}X)^T \in \mathbb{R}^{n \times n}$$

– это корреляционная матрица вектора X . Если нормальный вектор имеет математическое ожидание μ и корреляционную функцию R , то пишут, что $X \in N(\mu, R)$.

3.1. Свойства гауссовских векторов

1. Каждая компонента нормального вектора $X \in N(\mu, R)$ имеет нормальное распределение, причем $X_j \in N(\mu_j, R_{jj})$. Обратное неверно, то есть если у некоторого случайного вектора все компоненты по отдельности имеют нормальное распределение, то это не значит, что вектор целиком имеет нормальное распределение³. Дело в том, что совместное распределение компонент вектора в общем случае не определяется их частными распределениями. Однако если у случайного вектора компоненты независимы в совокупности и имеют нормальное распределение, то тогда и вектор имеет нормальное распределение.

2. Если у нормального случайного вектора $X \in N(\mu, R)$ матрица R невырождена, то существует плотность распределения, и она равна

$$f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det R}} \exp\left(-\frac{1}{2}(x - \mu)^T R^{-1}(x - \mu)\right), \quad x \in \mathbb{R}^n.$$

Если же матрица R вырождена, то плотности распределения у вектора X нет. В этом случае найдется вектор $c \neq 0$ такой, что $Rc = 0$, то есть

³Натан А.А., Горбачев О.Г., Гуз С.А. Теория вероятностей. Москва : МЗ Пресс, 2007. С. 179–180.

R имеет линейно зависимые столбцы и строки. Отсюда будет следовать, что и компоненты вектора X линейно связаны. Действительно,

$$\begin{aligned} Rc = 0 &\Rightarrow c^T Rc = 0 \Rightarrow \mathbb{E} c^T \overset{\circ}{X} \overset{\circ}{X}^T c = 0 \Rightarrow \\ &\Rightarrow \mathbb{E} (\overset{\circ}{X}^T c)^T (\overset{\circ}{X}^T c) = \mathbb{E} (\overset{\circ}{X}^T c)^2 = 0, \end{aligned}$$

откуда получаем $\overset{\circ}{X}^T c = 0$ п. н. Геометрически это значит, что вероятностная мера сконцентрирована на подпространстве меньшей размерности, а не на всем пространстве \mathbb{R}^n .

3. Пусть $X \in N(\mu, R)$, $X \in \mathbb{R}^n$. Тогда для любой матрицы $A \in \mathbb{R}^{k \times n}$ (не обязательно невырожденной и даже не обязательно квадратной) вектор AX тоже будет нормальным, причем

$$AX \in N(A\mu, ARA^T).$$

Для себя можно отметить слово *амарат*, или просто *Марат*, которое помогает запомнить место для знака транспонирования в формуле. Утверждение доказывается по определению нормального вектора, просто выписыванием характеристической функции вектора AX .

4. Вектор X является нормальным тогда и только тогда, когда любая линейная комбинация его компонент дает нормальную случайную величину⁴. Доказательство этого утверждение тоже не сложное, через характеристическую функцию, его можно найти в книге⁵.

5. Компоненты нормального случайного вектора некоррелированы тогда и только тогда, когда они независимы. Здесь нужно обратить внимание на то, что речь идет именно о компонентах нормального вектора, а не просто о нормальных случайных величинах! Просто нормальные случайные величины могут быть некоррелированными и зависимыми. Но если случайные величины не только нормальные, но являются компонентами нормального вектора (более сильное свойство), то тогда из их некоррелированности следует их независимость.

6. Условное распределение подвектора нормального вектора является нормальным распределением. А именно, пусть дан нормальный вектор $(\xi, \eta) \in N(\mu, R)$ размерности $n + m$, где подвектор $\xi \in N(\mu_\xi, R_{\xi\xi})$ имеет размерность $n \geq 1$, а подвектор $\eta \in N(\mu_\eta, R_{\eta\eta})$ имеет размерность $m \geq 1$. В этом случае математическое ожидание и корреляционная

⁴Или константу. Чтобы не делать все время эту оговорку, константу можно грубо воспринимать как нормальную случайную величину с нулевой дисперсией.

⁵Ширяев А.Н. Вероятность. В 2-х кн. 6-е изд., испр. Москва : МЦНМО, 2017. Кн. 1. С. 414.

матрица исходного вектора равны

$$\mu = \begin{bmatrix} \mu_\xi \\ \mu_\eta \end{bmatrix}, \quad R = \begin{bmatrix} R_{\xi\xi} & R_{\xi\eta} \\ R_{\eta\xi} & R_{\eta\eta} \end{bmatrix},$$

где матрицы $R_{\xi\eta} = R_{\eta\xi}^T$ состоят из ковариаций между компонентами ξ и η . Пусть матрица $R_{\eta\eta}$ не вырождена. Тогда условное распределение подвектора ξ при условии $\eta = x$ имеет нормальное распределение, причем⁶

$$(\xi | \eta = x) \in N(\mu_\xi + R_{\xi\eta}R_{\eta\eta}^{-1}(x - \mu_\eta), R_{\xi\xi} - R_{\xi\eta}R_{\eta\eta}^{-1}R_{\eta\xi}).$$

Эту формулу легко получить из следующих соображений. Компоненты нормальных векторов независимы тогда и только тогда, когда они некоррелированы. Поэтому для скалярных ξ и η достаточно написать:

$$\mathbb{P}(\xi < y | \eta = x) = \mathbb{P}(\xi - a\eta < y - ax | \eta = x)$$

и найти такую постоянную a , чтобы корреляционный момент между $\xi - a\eta$ и η был равен нулю, тогда это будет означать независимость этих величин и от условия в выражении условной вероятности можно избавиться. Этот прием можно использовать и для векторов ξ и η .

Приведем без доказательства еще один факт из теории вероятностей, связанный с моментами произвольного центрированного нормального случайного вектора.

Теорема Вика*. Пусть дан нормальный вектор $(X_1, \dots, X_n) \in N(0, R)$ с нулевым математическим ожиданием и произвольной корреляционной матрицей $R = \|R_{pq}\|_{p,q=1}^n$. Тогда если n нечетно, то $\mathbb{E}X_1 \dots X_n = 0$. Если же n четно, то

$$\mathbb{E}X_1 \dots X_n = \sum R_{p_1 q_1} \dots R_{p_{n/2} q_{n/2}},$$

где сумма берется по всем неупорядоченным разбиениям множества $\{1, \dots, n\}$ на $n/2$ неупорядоченных пар.

Примеры. Пусть $X = (X_1, X_2, X_3, X_4) \in N(0, R)$. Учтем, что любой подвектор нормального вектора является нормальным вектором, и поэтому для любого подвектора можно применять теорему Вика. Тогда из теоремы следует, что $\mathbb{E}X_1 X_2 X_3 = 0$, так как под математическим ожиданием стоит нечетное число множителей. Далее,

$$\mathbb{E}X_1 X_2 X_3 X_4 = R_{12}R_{34} + R_{13}R_{24} + R_{14}R_{23},$$

⁶Ширяев А.Н. Вероятность. В 2-х кн. 6-е изд., испр. Москва : МЦНМО, 2017. Кн. 1. С. 417.

так как число слагаемых здесь $n = 4$ – четно, а неупорядоченных разбиений множества $\{1, 2, 3, 4\}$ на $n/2 = 2$ неупорядоченные пары здесь ровно три: $12|34$, $13|24$ и $14|23$. Неупорядоченность разбиений означает, что мы не различаем, например, разбиения $12|34$ и $34|12$. Неупорядоченность же пар означает, что мы не различаем, например, разбиение $12|34$ от $12|43$. Если бы мы считали разбиения на упорядоченные пары, то таких разбиений было бы больше. Кроме того, теорема Вика позволяет сказать сразу, что $\mathbb{E}X_1^3 = 0$, так как (X_1, X_1, X_1) – это тогда нормальный вектор (обсуждали это выше), а под математическим ожиданием стоит нечетное число слагаемых. С формальной точки зрения на это можно посмотреть как на введение вектора $Y = (Y_1, Y_2, Y_3)$ такого, что все $Y_i = X_1$. Точно так же $\mathbb{E}X_1X_2^2 = 0$, потому что число слагаемых нечетное.

Несколько слов о том, откуда следует утверждение теоремы. Дело в том, что моменты $\mathbb{E}X_1^{k_1}X_2^{k_2}\dots X_n^{k_n}$ связаны с характеристической функцией $\varphi_X(s)$ вектора $X = (X_1, \dots, X_n)$ и выражаются через смешанные производные этой характеристической функции в точке $s = 0$, причем момент $\mathbb{E}X_1^{k_1}X_2^{k_2}\dots X_n^{k_n}$ выражается через смешанную производную $\partial^n \varphi(0)/\partial s_1 \partial s_2 \dots \partial s_n$, детали см., например, в учебнике⁷. Характеристическая функция нормального вектора с нулевыми математическими ожиданиями зависит от своих аргументов s через квадратичную форму $s^T R s$, поэтому если n нечетно, то соответствующая производная при $s = 0$ равна нулю и только для четных n могут получиться ненулевые производные. А так как характеристическая функция зависит только от ковариационной матрицы R , то и любые производные (и как следствие – моменты) тоже будут зависеть только от R . Точные выражения для моментов получаются при их соответствии коэффициентам при $s_1 \dots s_n$ в разложении в ряд Тейлора характеристической функции нормального вектора.

3.2. Винеровский процесс

Перейдем теперь к важнейшему примеру нормального случайного процесса.

Определение. Случайный процесс $\{W(t), t \geq 0\}$ называется *винеровским процессом* (*процессом броуновского движения*), если

1. $W(0) = 0$ п. н.
2. $W(t)$ – процесс с независимыми приращениями.
3. $\forall t, s \geq 0$ приращение $W(t) - W(s) \in N(0, |t - s|)$.

⁷Боровков А.А. Теория вероятностей. Москва : Эдиториал УРСС, 1999. С. 143.

История возникновения этого процесса началась с наблюдений шотландского ботаника Роберта Броуна в 1827 г., заметившего, что маленькие частицы, помещенные в жидкость, совершают непрерывное беспорядочное движение. В 1905 г. А. Эйнштейн объяснил это тем, что наблюдаемые частицы подвержены непрерывным соударениям с молекулами окружающей среды. Процесс назван в честь американского математика Норберта Винера, глубоко изучавшего этот процесс и его свойства.

Определение винеровского процесса обусловлено следующими соображениями. Пусть $X(t)$ – координата броуновской частицы в момент t . Смещение $X(t) - X(s)$ на интервале времени $[s, t]$ можно рассматривать как сумму большого числа случайных смещений, которые с равной вероятностью могут происходить как в одну, так и другую сторону. В этой ситуации можно воспользоваться центральной предельной теоремой, так что смещение моделируется как нормальная случайная величина с нулевым математическим ожиданием. Кроме того, естественно предположить, что распределения величин $X(t) - X(s)$ и $X(t+h) - X(s+h)$ совпадают при любом h (стационарность) и что смещения на непересекающихся интервалах времени независимы. Наконец, естественно считать, что $X(t) - X(s)$ зависит лишь от разности $t - s$, но не от абсолютных значений моментов времени.

Винеровский процесс и его свойства имеют огромное значение в науке и технологиях. Он используется в прикладной математике и математической физике, экономике, финансовой математике, эволюционной биологии, физике, используется как модель шума при разработке электроники, теории фильтрации, а также как модель возмущений в теории управления.

Отметим наконец, что винеровский и пуассоновский процессы объединяет то, что они являются моделями ситуаций, когда что-то в большом количестве происходит в малом интервале времени, и связаны, тем самым, с суммами независимых случайных величин и предельными теоремами. И хотя свойства и поведение этих процессов во многом отличаются, на самом деле они являются частью чего-то общего – они являются частными случаями процессов Леви – процессов с независимыми и стационарными приращениями. Но и на этом не все. Оказывается, что всякий процесс Леви в некотором смысле является композицией двух типов процессов – нормального и сложного пуассоновского.

Перейдем теперь к изучению свойств винеровского процесса.

Теорема. Винеровский процесс является нормальным процессом.

Доказательство. Возьмем произвольные $n \geq 1$, произвольные

упорядоченные по возрастанию моменты времени $t_1 \leq t_2 \leq \dots \leq t_n$, рассмотрим вектор $X = (W(t_1), \dots, W(t_n))$ и докажем, что он имеет нормальное распределение.

Из определения винеровского процесса видно, что каждое его сечение $W(t) = W(t) - W(0) \in N(0, t)$ имеет нормальное распределение. Вектор, составленный из нормальных случайных величин, не обязательно является нормальным вектором. Однако вектор, составленный из независимых нормальных случайных величин, является нормальным вектором. В нашем случае независимыми и нормальными являются случайные величины $W(t_j) - W(t_{j-1}) \in N(0, t_j - t_{j-1})$. Это значит, что вектор

$$Y = (W(t_1) - W(0), W(t_2) - W(t_1), \dots, W(t_n) - W(t_{n-1}))$$

является нормальным вектором, причем с нулевым математическим ожиданием $m = 0$ и диагональной корреляционной матрицей D с $D_{jj} = t_j - t_{j-1}$. Вектор Y связан с вектором X линейным преобразованием:

$$\begin{bmatrix} W(t_1) \\ W(t_2) \\ W(t_3) \\ \vdots \\ W(t_n) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} W(t_1) - W(0) \\ W(t_2) - W(t_1) \\ W(t_3) - W(t_2) \\ \vdots \\ W(t_n) - W(t_{n-1}) \end{bmatrix}.$$

Так как линейные преобразования переводят нормальные векторы в нормальные векторы, то вектор $(W(t_1), \dots, W(t_n))$ тоже является нормальным.

Вектор из сечений, взятых в неупорядоченные моменты времени, получается как линейное преобразование вектора из сечений, взятых в упорядоченные моменты времени, поэтому такой вектор тоже является нормальным. \square

Действуя по аналогии с пуассоновским процессом, легко доказать, что математическое ожидание и корреляционная функция винеровского процесса равны

$$\mathbb{E}W(t) = 0, \quad R_W(t, s) = \min(t, s).$$

Отсюда следует, что вектор $(W(t_1), \dots, W(t_n))$ имеет нулевое математическое ожидание и корреляционную матрицу с компонентами

$$R_{ij} = \text{cov}(W(t_i), W(t_j)) = R_W(t_i, t_j) = \min(t_i, t_j).$$

Отметим, наконец, что вектор из сечений винеровского процесса является нормальным вектором с нулевым математическим ожиданием,

поэтому, например, $\mathbb{E}W^3(t) = 0$, $\mathbb{E}W(t)W^2(s) = 0$,

$$\mathbb{E}W^4(t) = 3R_W(t, t)R_W(t, t) = 3t^2. \triangle$$

В заключение сформулируем и докажем одну теорему о винеровском процессе как предела (в некотором смысле) случайных блужданий.

Теорема. Пусть случайные величины ξ_1, ξ_2, \dots независимы и одинаково распределены со средним 0 и конечной ненулевой дисперсией σ^2 . Тогда конечномерные распределения процесса

$$X_n(t) = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{[nt]} \xi_k$$

поточечно сходятся к конечномерным распределениям винеровского процесса (считается, что $X_n(t) = 0$, если $[nt] = 0$).

Доказательство. Введем обозначение

$$S_n = \sum_{k=1}^n \xi_k.$$

В первую очередь заметим, что при фиксированном t

$$\frac{1}{\sigma\sqrt{n}} S_{[nt]} \xrightarrow{d} W(t), \quad n \rightarrow \infty.$$

Действительно, сходимость по распределению равносильна поточечной сходимости характеристических функций, а характеристическая функция левой части равна

$$\begin{aligned} \varphi_{\frac{1}{\sigma\sqrt{n}} S_{[nt]}}(s) &= \varphi_{S_{[nt]}}(s/(\sigma\sqrt{n})) = \mathbb{E} \exp \left(\frac{is}{\sigma\sqrt{n}} (\xi_1 + \dots + \xi_{[nt]}) \right) = \\ &= \prod_{k=1}^{[nt]} \varphi_{\xi_k}(s/(\sigma\sqrt{n})) = \left(1 - \frac{s^2\sigma^2}{2\sigma^2n} + o\left(\frac{1}{n}\right) \right)^{[nt]} = \\ &= \left[\left(1 - \frac{s^2\sigma^2}{2\sigma^2n} + o\left(\frac{1}{n}\right) \right)^n \right]^{\frac{[nt]}{n}} \rightarrow e^{-s^2t/2}, \end{aligned}$$

а $e^{-s^2t/2}$ – это характеристическая функция случайной величины с распределением $N(0, t)$ – таким же распределением, как у $W(t)$. Так мы доказали сходимость одномерных распределений.

Рассмотрим теперь два момента времени s и t , $s < t$. Докажем, что

$$(X_n(s), X_n(t)) \xrightarrow{d} (W(s), W(t)).$$

Мы докажем это, если докажем, что

$$(X_n(s), X_n(t) - X_n(s)) \xrightarrow{d} (W(s), W(t) - W(s)),$$

потому что если $(\eta_n, \zeta_n) \xrightarrow{d} (\eta, \zeta)$, то $(\eta_n, \eta_n + \zeta_n) \xrightarrow{d} (\eta, \eta + \zeta)$ ⁸. Но сходимость

$$\left(\frac{1}{\sigma\sqrt{n}} S_{[ns]}, \frac{1}{\sigma\sqrt{n}} S_{[nt]} - \frac{1}{\sigma\sqrt{n}} S_{[ns]} \right) \xrightarrow{d} (W(s), W(t) - W(s))$$

просто следует из сходимости характеристических функций и того, что в векторе слева компоненты независимы (ξ_i входят из несовпадающих множеств). Аналогичным способом рассматриваются случаи трех и более чисел моментов времени. Следовательно, конечномерные распределения процесса $X_n(t)$ сходятся к конечномерным распределениям винеровского процесса. \square

Известны и гораздо более сильные результаты о сходимости случайных блужданий к винеровскому процессу: см., например, Афанасьев В.И. Случайные блуждания и ветвящиеся процессы. Москва : Миан, 2007, лекция 5. Здесь сходимость описывается в пространстве $D[0, 1]$ числовых функций, имеющих пределы слева и непрерывные справа на отрезке $[0, 1]$ с метрикой Скорохода, а сходимость случайных блужданий к винеровскому процессу – это сходимость по распределению в пространстве $D[0, 1]$.

4. Стохастический анализ

В этом разделе вводятся свойства непрерывности, дифференцируемости и интегрируемости случайных процессов. Случайные процессы, как известно, являются функциями двух переменных – исхода и времени, а свойства непрерывности, дифференцируемости и интегрируемости мы введем относительно времени. Непрерывность, дифференцируемость и интегрируемость можно ввести в различных смыслах, это зависит от того, как определяемый предел относится к исходу. Из

⁸Этот факт мы оставляем без доказательства. См. следствие 1 теоремы 5.1 в монографии Биллингсли П. Сходимость вероятностных мер. Москва : Наука, 1977. С. 49.

курса теории вероятностей мы знаем, что есть несколько видов предела, наиболее часто используемые среди них – это предел в среднем квадратичном, предел почти наверное, предел по вероятности и предел по распределению. В этом разделе предел будет пониматься в среднеквадратичном смысле (сходимость в «среднем»), для этого типа предела можно сравнительно просто получить целый ряд полезных результатов. Если же требуется, например, изучать свойства отдельных траекторий, то по этому случаю следует смотреть сходимость почти наверное и при необходимости обращаться к специализированной литературе. Иногда для краткости тот или иной вид сходимости будет обозначаться приставкой п.н.-, с.к.-, \mathbb{P} - или d - для сходимости почти наверное, в среднем квадратичном, по вероятности и по распределению соответственно.

Естественно, если мы говорим о непрерывности, дифференцируемости и интегрируемости, это предполагает, что мы относим эти понятия к случайным функциям, то есть случайным процессам с непрерывным временем. Но прежде всего установим несколько вспомогательных фактов, касающихся вообще сходимости в среднем квадратичном для случайных последовательностей.

4.1. Свойства с.к.-сходимости

Напомним, что случайная последовательность $\{X_n\}_{n=1}^{\infty}$ с $\mathbb{E}X_n^2 < \infty$ для любого n сходится к случайной величине X с $\mathbb{E}X^2 < \infty$ в среднем квадратичном, если

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n - X)^2 = 0.$$

Для этого типа сходимости используются обозначения

$$X_n \xrightarrow{\text{с.к.}} X, \text{ l.i.m. } X_n = X,$$

где l.i.m. означает «limit in mean» (сходимость в среднем). Нам потребуются три вспомогательные леммы.

Лемма 1. Пусть $\{X_m\}_{m=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$, X , Y – случайные величины, заданные на одном вероятностном пространстве и

$$X_m \xrightarrow{\text{с.к.}} X, Y_n \xrightarrow{\text{с.к.}} Y, \mathbb{E}X^2 < \infty, \mathbb{E}Y^2 < \infty.$$

Тогда

$$\lim_{m,n \rightarrow \infty} \mathbb{E}X_m Y_n = \mathbb{E} \text{ l.i.m. } X_m \text{ l.i.m. } Y_n = \mathbb{E}XY.$$

Доказательство. Заметим, что

$$X_m Y_n - XY = (X_m - X)(Y_n - Y) + (X_m - X)Y + (Y_n - Y)X,$$

$$|\mathbb{E}X_m Y_n - \mathbb{E}XY| \leq |\mathbb{E}(X_m - X)(Y_n - Y)| + |\mathbb{E}(X_m - X)Y| + |\mathbb{E}(Y_n - Y)X|.$$

Применяя неравенство Коши–Буняковского и используя условия леммы, при $m, n \rightarrow \infty$ получим

$$|\mathbb{E}(X_m - X)(Y_n - Y)| \leq \sqrt{\mathbb{E}(X_m - X)^2 \mathbb{E}(Y_n - Y)^2} \rightarrow 0,$$

$$|\mathbb{E}(X_m - X)Y| \leq \sqrt{\mathbb{E}(X_m - X)^2 \mathbb{E}Y^2} \rightarrow 0,$$

$$|\mathbb{E}(Y_n - Y)X| \leq \sqrt{\mathbb{E}(Y_n - Y)^2 \mathbb{E}X^2} \rightarrow 0,$$

откуда получаем требуемое $\lim_{m,n \rightarrow \infty} \mathbb{E}X_m Y_n = \mathbb{E}XY$. \square

Из этой леммы следует, в частности, что при $Y_n = Y = 1$

$$\lim_{m \rightarrow \infty} \mathbb{E}X_m = \mathbb{E} \lim_{m \rightarrow \infty} X_m = \mathbb{E}X.$$

Это чрезвычайно важная лемма – она говорит о допустимости перестановки местами предела и математического ожидания, что очень часто используется при доказательстве теорем и на практике. Следующая теорема говорит о свойстве полноты пространства случайных величин со среднеквадратичной нормой, то есть о том, что любая фундаментальная последовательность в этом пространстве сходится (замечу, что п.н.-фундаментальные, \mathbb{P} -фундаментальные и d -фундаментальные последовательности тоже обязательно являются сходящимися⁹).

Лемма 2 (фундаментальные последовательности сходятся)*. Пусть $\{X_k\}_{k=1}^{\infty}$ – последовательность случайных величин, для каждой из которых $\mathbb{E}X_k^2 < \infty$. Необходимым и достаточным условием сходимости в среднем квадратичном последовательности $\{X_k\}_{k=1}^{\infty}$ к некоторой случайной величине X с $\mathbb{E}X^2 < \infty$ является равенство

$$\lim_{m,n \rightarrow \infty} \mathbb{E}(Y_m - Z_n)^2 = 0$$

для любых двух подпоследовательностей $\{Y_m\}_{m=1}^{\infty}$ и $\{Z_n\}_{n=1}^{\infty}$ последовательности $\{X_k\}_{k=1}^{\infty}$.

Лемма 3. Пусть для последовательности $\{X_k\}_{k=1}^{\infty}$, $\mathbb{E}X_k^2 < \infty$ найдется такое $c \in \mathbb{R}$, что для любых подпоследовательностей $\{Y_m\}_{m=1}^{\infty}$ и $\{Z_n\}_{n=1}^{\infty}$ последовательности $\{X_k\}_{k=1}^{\infty}$ при $m, n \rightarrow \infty$ выполнено $\mathbb{E}Y_m Z_n \rightarrow c$. Тогда существует случайная величина X с конечным вторым моментом $\mathbb{E}X^2 < \infty$ и $X_k \xrightarrow{c, \kappa} X$ при $k \rightarrow \infty$.

Доказательство. Утверждение следует из равенств

$$(Y_m - Z_n)^2 = Y_m^2 - 2Y_m Z_n + Z_n^2,$$

⁹Боровков А.А. Теория вероятностей. Москва : Эдиториал УРСС, 1999. С. 116.

$$\lim_{m,n \rightarrow \infty} \mathbb{E}(Y_m - Z_n)^2 = c - 2c + c = 0,$$

с учетом леммы 2. \square

Мы будем называть *процессами второго порядка* такие случайные процессы $\{X(t), t \in T\}$, у которых в любой момент времени $t \in T$ существует конечный второй момент $\mathbb{E}X^2(t) < \infty$. Например, пуассоновский и винеровский процессы являются процессами второго порядка.

4.2. С.к.-непрерывность процессов

Определение. Случайный процесс $\{X(t), t \in T\}$ второго порядка называется *с.к.-непрерывным в точке t* , если

$$X(t + \varepsilon) \xrightarrow{с.к.} X(t), \quad \varepsilon \rightarrow 0.$$

Если это свойство выполнено в любой точке $t \in T$, то такой процесс называется просто *с.к.-непрерывным*.

Теорема (критерий с.к.-непрерывности). *Процесс второго порядка $X(t)$ является с.к.-непрерывным в точке t_0 тогда и только тогда, когда его ковариационная функция $K_X(t, s)$ непрерывна в точке (t_0, t_0) , или, что эквивалентно, тогда и только тогда, когда математическое ожидание $m_X(t)$ непрерывно в точке t_0 и корреляционная функция $R_X(t, s)$ непрерывна в точке (t_0, t_0) .*

Доказательство. Пусть процесс $X(t)$ является с.к.-непрерывным в точке t_0 . Тогда по определению

$$X(t_0 + \varepsilon_1) \xrightarrow{с.к.} X(t_0), \quad X(t_0 + \varepsilon_2) \xrightarrow{с.к.} X(t_0), \quad \varepsilon_1, \varepsilon_2 \rightarrow 0.$$

Из леммы 1 получаем, что

$$\mathbb{E}X(t_0 + \varepsilon_1)X(t_0 + \varepsilon_2) \xrightarrow{\varepsilon_1, \varepsilon_2 \rightarrow 0} \mathbb{E}X^2(t_0),$$

что и означает $K_X(t_0 + \varepsilon_1, t_0 + \varepsilon_2) \rightarrow K_X(t_0, t_0)$ при $\varepsilon_1, \varepsilon_2 \rightarrow 0$.

Предположим теперь, что функция $K_X(t, s)$ непрерывна в точке (t_0, t_0) . Так как

$$\mathbb{E}(X(t_0 + \varepsilon) - X(t_0))^2 = K_X(t_0 + \varepsilon, t_0 + \varepsilon) - 2K_X(t_0 + \varepsilon, t_0) + K_X(t_0, t_0),$$

то, взяв предел при $\varepsilon \rightarrow 0$, мы получаем

$$\mathbb{E}(X(t_0 + \varepsilon) - X(t_0))^2 \rightarrow 0, \quad \varepsilon \rightarrow 0.$$

Если математическое ожидание $m_X(t)$ непрерывно в точке t_0 и корреляционная функция $R_X(t, s)$ непрерывна в точке (t_0, t_0) , то из соотношения

$$K_X(t, s) = R_X(t, s) + m_X(t)m_X(s)$$

сразу следует, что $K_X(t, s)$ непрерывна в точке (t_0, t_0) . Наоборот, если функция $K_X(t, s)$ непрерывна в точке (t_0, t_0) , то $X(t_0 + \varepsilon) \xrightarrow{с.к.} X(t_0)$ при $\varepsilon \rightarrow 0$, и из сходимости

$$\begin{aligned} & \mathbb{E}(X(t_0 + \varepsilon) - X(t_0))^2 = \\ & = \mathbb{E} \left(\overset{\circ}{X}(t_0 + \varepsilon) - \overset{\circ}{X}(t_0) + m_X(t_0 + \varepsilon) - m_X(t_0) \right)^2 = \\ & = \mathbb{E} \left(\overset{\circ}{X}(t_0 + \varepsilon) - \overset{\circ}{X}(t_0) \right)^2 + (m_X(t_0 + \varepsilon) - m_X(t_0))^2 \xrightarrow{\varepsilon \rightarrow 0} 0 \end{aligned}$$

следует непрерывность математического ожидания $m_X(t)$ в точке t_0 и сходимость $\overset{\circ}{X}(t_0 + \varepsilon) \xrightarrow{с.к.} \overset{\circ}{X}(t_0)$, то есть непрерывность $K_{\overset{\circ}{X}}(t, s)$ в точке (t_0, t_0) , то есть непрерывность $R_X(t, s)$ в точке (t_0, t_0) . \square

4.3. С.к.-дифференцируемость процессов

Определение. Случайный процесс $\{X(t), t \in T\}$ второго порядка называется *с.к.-дифференцируемым* в точке $t_0 \in T$, если существует случайная величина η , $\mathbb{E}\eta^2 < \infty$, такая, что

$$\frac{X(t_0 + \varepsilon) - X(t_0)}{\varepsilon} \xrightarrow{с.к.} \eta, \quad \varepsilon \rightarrow 0.$$

Случайная величина η называется *с.к.-производной процесса $X(t)$ в точке t_0* и обозначается $\eta = X'(t_0)$. Если процесс с.к.-дифференцируем в любой точке $t \in T$, то процесс называется просто *с.к.-дифференцируемым*, а соответствующий процесс второго порядка $X'(t)$ называется его *с.к.-производной*.

Теорема (критерий с.к.-дифференцируемости). *Процесс $\{X(t), t \in T\}$ второго порядка является с.к.-дифференцируемым в точке t_0 тогда и только тогда, когда существует конечный предел*

$$\lim_{\varepsilon, \delta \rightarrow 0} \frac{1}{\varepsilon \delta} (K_X(t_0 + \varepsilon, t_0 + \delta) - K_X(t_0 + \varepsilon, t_0) - K_X(t_0, t_0 + \delta) + K_X(t_0, t_0)) < \infty, \quad (2)$$

или, что равносильно, тогда и только тогда, когда математическое ожидание $m_X(t)$ дифференцируемо в точке t_0 и существует конечный указанный выше предел для корреляционной функции $R_X(t, s)$.

Доказательство. Пусть $X(t)$ является с.к.-дифференцируемым в точке t_0 . Введем обозначения

$$Y_\varepsilon = \frac{X(t_0 + \varepsilon) - X(t_0)}{\varepsilon}, \quad Y_\delta = \frac{X(t_0 + \delta) - X(t_0)}{\delta}$$

и заметим, что

$$\mathbb{E}Y_\varepsilon Y_\delta = \frac{1}{\varepsilon\delta}(K_X(t_0+\varepsilon, t_0+\delta) - K_X(t_0+\varepsilon, t_0) - K_X(t_0, t_0+\delta) + K_X(t_0, t_0)).$$

Так как $Y_\varepsilon \xrightarrow{с.к.} X'(t_0)$ и $Y_\delta \xrightarrow{с.к.} X'(t_0)$ при $\varepsilon, \delta \rightarrow 0$, то по лемме 1 получаем, что необходимо $\mathbb{E}Y_\varepsilon Y_\delta \rightarrow \mathbb{E}X'(t_0)^2 < \infty$. Но это и означает, что предел выше существует и конечен. И наоборот, из существования конечного предела $\mathbb{E}Y_\varepsilon Y_\delta \rightarrow c$, $|c| < \infty$ по лемме 3 получается, что существует случайная величина η такая, что $\mathbb{E}\eta^2 < \infty$ и $Y_\varepsilon \xrightarrow{с.к.} \eta$, что по определению означает с.к.-дифференцируемость в точке t_0 .

Если функция $m_X(t)$ дифференцируема в точке t_0 и существует конечный предел

$$\lim_{\varepsilon, \delta \rightarrow 0} \frac{1}{\varepsilon\delta}(R_X(t_0+\varepsilon, t_0+\delta) - R_X(t_0+\varepsilon, t_0) - R_X(t_0, t_0+\delta) + R_X(t_0, t_0)) < \infty,$$

то из формулы $K_X(t, s) = R_X(t, s) + m_X(t)m_X(s)$ следует, что существует конечный предел (2). Наоборот, если существует предел (2), то как, мы выяснили, процесс $X(t)$ с.к.-дифференцируем в точке t_0 , следовательно, выражение

$$\begin{aligned} \mathbb{E}(Y_\varepsilon - X'(t_0))^2 &= \mathbb{E}\left(\frac{X(t_0+\varepsilon) - X(t_0)}{\varepsilon} - X'(t_0)\right)^2 = \\ &= \mathbb{E}\left(\frac{\overset{\circ}{X}(t_0+\varepsilon) - \overset{\circ}{X}(t_0)}{\varepsilon} - \overset{\circ}{X}'(t_0)\right)^2 + \\ &+ \left(\frac{m_X(t_0+\varepsilon) - m_X(t_0)}{\varepsilon} - \mathbb{E}X'(t_0)\right)^2 \end{aligned}$$

стремится к нулю, а значит, каждое из слагаемых стремится к нулю. Сходимость к нулю второго слагаемого означает дифференцируемость $m_X(t)$ в точке t_0 . Сходимость к нулю первого слагаемого означает с.к.-дифференцируемость процесса $\overset{\circ}{X}(t)$ в точке t_0 , а значит (по уже доказанному), существование конечного предела (2) для его ковариационной функции. Но ковариационная функция процесса $\overset{\circ}{X}(t)$ совпадает с корреляционной функцией процесса $X(t)$. \square

Заметим, что предел из утверждения теоремы не является смешанной производной ковариационной функции $K_X(t_1, t_2)$. Дело в том, что по определению смешанная производная второго порядка функции

двух переменных – это производная по одной переменной от производной по другой переменной, то есть повторный предел по каждой переменной. Предел из теоремы – это предел по двум переменным, он отличается от второй смешанной производной так же, как предел по двум переменным отличается от повторного предела. Предел из теоремы обычно называется *обобщенной смешанной производной* (она не имеет никакого отношения к обобщенным функциям!). Из существования обобщенной смешанной производной вытекает существование обычной смешанной производной второго порядка; обратное, вообще говоря, неверно.

4.4. С.к.-интегрируемость процессов

Определение. Пусть процесс $\{X(t), t \in T\}$ определен на отрезке $[a, b] \subseteq T$. На отрезке $[a, b]$ построим некоторое разбиение $a = t_0 < t_1 < \dots < t_{n-1} < t_n = b$, и на каждом из промежутков этого разбиения выберем произвольную точку $\tau_i \in [t_{i-1}, t_i)$, $i = 1, \dots, n$. Если при $n \rightarrow \infty$ и $\max_{i=1, \dots, n} (t_i - t_{i-1}) \rightarrow 0$ существует предел в среднеквадратичном

$$\sum_{i=1}^n X(\tau_i)(t_i - t_{i-1}) \xrightarrow{\text{с.к.}} \eta, \mathbb{E}\eta^2 < \infty,$$

не зависящий от разбиения $\{t_i\}$ и выбора точек $\{\tau_i\}$, то процесс $X(t)$ называется *с.к.-интегрируемым по Риману на $[a, b]$* , а случайная величина η называется его *с.к.-интегралом Римана* на отрезке $[a, b]$ и обозначается

$$\eta = \int_a^b X(t) dt.$$

Теорема (критерий с.к.-интегрируемости по Риману)*. *Процесс второго порядка $\{X(t), t \in T\}$ является с.к.-интегрируемым по Риману на отрезке $[a, b]$ тогда и только тогда, когда существует конечный интеграл Римана:*

$$\int_a^b \int_a^b K_X(t, s) dt ds,$$

или, что равносильно, тогда и только тогда, когда существуют конечные интегралы Римана:

$$\int_a^b m_X(t) dt, \int_a^b \int_a^b R_X(t, s) dt ds.$$

Понятие с.к.-интеграла Римана естественным образом обобщается на случай неограниченных интервалов интегрирования:

$$\int_a^\infty X(t) dt = \text{l.i.m.}_{b \rightarrow \infty} \int_a^b X(t) dt.$$

Аналогичным образом вводятся интегралы от $-\infty$ до числа и от $-\infty$ до $+\infty$.

В заключение отметим, что из критериев с.к.-непрерывности, с.к.-дифференцируемости и с.к.-интегрируемости следует, что если процесс с.к.-дифференцируем в точке, то он и с.к.-непрерывен в этой точке. Если процесс с.к.-непрерывен в каждой точке некоторого отрезка, то он с.к.-интегрируем по Риману на этом отрезке.

Для случайных процессов второго порядка введем понятие *с.к.-интеграла Римана–Стилтьеса*, аналогичное понятию интеграла Римана–Стилтьеса для неслучайных функций.

Определение. Пусть процесс $\{X(t), t \in T\}$ второго порядка определен на отрезке $[a, b] \subseteq T$ и $g(t)$ – непрерывная на отрезке $[a, b]$ функция. На отрезке $[a, b]$ построим разбиение

$$a = t_0 < t_1 < \dots < t_{n-1} < t_n = b,$$

и на каждом из промежутков этого разбиения выберем произвольную точку $\tau_i \in [t_{i-1}, t_i]$, $i = 1, \dots, n$. Если при $n \rightarrow \infty$ и $\max_{i=1, \dots, n} (t_i - t_{i-1}) \rightarrow 0$ существует предел в среднеквадратичном

$$\sum_{i=1}^n g(\tau_i)(X(t_i) - X(t_{i-1})) \xrightarrow{\text{с.к.}} \eta, \quad \mathbb{E}\eta^2 < \infty,$$

не зависящий от разбиения $\{t_i\}$ и выбора точек $\{\tau_i\}$, то случайная величина η называется *с.к.-интегралом Римана–Стилтьеса от функции $g(t)$ по процессу $X(t)$* на отрезке $[a, b]$ и обозначается символом

$$\eta = \int_a^b g(t) dX(t).$$

Теорема (критерий с.к.-интегрируемости по Риману–Стилтьесу)*. *С.к.-интеграл Римана–Стилтьеса*

$$\int_a^b g(t) dX(t)$$

от непрерывной функции $g(t)$ по процессу второго порядка $\{X(t), t \in T\}$ на отрезке $[a, b] \subseteq T$ существует тогда и только тогда, когда существует конечный интеграл Римана–Стилтьеса:

$$\int_a^b \int_a^b g(t)g(s) d^2 K_X(t, s) < \infty,$$

который понимается как предел выражения

$$\sum g(\tau_i)g(\sigma_j)(K_X(t_{i+1}, s_{j+1}) - K_X(t_{i+1}, s_j) - K_X(t_i, s_{j+1}) + K_X(t_i, s_j))$$

при $\max_{i=1, \dots, n} (t_{i+1} - t_i) \rightarrow 0$, $\max_{i=1, \dots, m} (s_{i+1} - s_i) \rightarrow 0$, $n \rightarrow \infty$, $m \rightarrow \infty$ для двух разбиений $a = t_0 < t_1 < \dots < t_n = b$ и $a = s_0 < s_1 < \dots < s_m = b$ отрезка $[a, b]$ и точки $\tau_i \in [t_i, t_{i+1}]$, $\sigma_j \in [s_j, s_{j+1}]$.

Понятие с.к.-интеграла Римана–Стилтьеса естественным образом обобщается на случай неограниченных интервалов интегрирования:

$$\int_a^\infty g(t) dX(t) = \text{l.i.m.}_{b \rightarrow \infty} \int_a^b g(t) dX(t),$$

$$\int_{-\infty}^\infty g(t) dX(t) = \text{l.i.m.}_{\substack{a \rightarrow -\infty \\ b \rightarrow \infty}} \int_a^b g(t) dX(t).$$

4.5. Некоторые полезные формулы

Пусть $X'(t)$ – это с.к.-производная процесса $X(t)$. Воспользовавшись леммой 1 из раздела 4.1, найдем математическое ожидание $X'(t)$:

$$\begin{aligned} \mathbb{E}X'(t) &= \mathbb{E} \text{l.i.m.}_{\varepsilon \rightarrow 0} \left(\frac{X(t+\varepsilon) - X(t)}{\varepsilon} \right) = \lim_{\varepsilon \rightarrow 0} \mathbb{E} \left(\frac{X(t+\varepsilon) - X(t)}{\varepsilon} \right) = \\ &= \lim_{\varepsilon \rightarrow 0} \left(\frac{m_X(t+\varepsilon) - m_X(t)}{\varepsilon} \right) = \frac{d}{dt} m_X(t) = \frac{d}{dt} \mathbb{E}X(t), \end{aligned}$$

то есть математическое ожидание с.к.-производной процесса равно производной математического ожидания процесса. Совершенно аналогично можно найти корреляционную функцию процесса $X'(t)$:

$$\begin{aligned} R_{X'}(t, s) &= \mathbb{E}(X'(t) - \mathbb{E}X'(t))(X'(s) - \mathbb{E}X'(s)) = \\ &= \mathbb{E} \text{l.i.m.}_{\varepsilon, \delta \rightarrow 0} \left(\frac{X(t+\varepsilon) - X(t)}{\varepsilon} - \frac{m_X(t+\varepsilon) - m_X(t)}{\varepsilon} \right) \times \end{aligned}$$

$$\begin{aligned}
& \times \left(\frac{X(s + \delta) - X(s)}{\delta} - \frac{m_X(s + \varepsilon) - m_X(s)}{\varepsilon} \right) = \\
& = \lim_{\varepsilon, \delta \rightarrow 0} \frac{1}{\varepsilon \delta} (R_X(t + \varepsilon, s + \delta) - R_X(t + \varepsilon, s) - R_X(t, s + \varepsilon) + R_X(t, s)) = \\
& = \frac{\partial^2 R_X(t, s)}{\partial t \partial s},
\end{aligned}$$

то есть корреляционная функция с.к.-производной процесса X в точке (t, s) совпадает с обобщенной производной второго порядка процесса X в точке (t, s) . Так как процесс X с.к.-дифференцируем, и, значит, $X'(t)$ является процессом второго порядка, то корреляционная функция $X'(t)$ всюду определена, поэтому обобщенная производная существует в любой точке $(t, s) \in [0, T]^2$. Но обобщенная производная, если существует, совпадает со смешанной производной второго порядка, поэтому вычислять эту производную можно по обычным правилам – сначала вычислив производную по s , а затем по t (или наоборот).

Похожим образом можно вывести формулы математического ожидания и корреляционной функции для с.к.-интегралов процесса $X(t)$. А именно, для процесса

$$Y(t) = \int_a^t X(\tau) d\tau$$

справедливо

$$\mathbb{E} \int_a^t X(\tau) d\tau = \int_a^t \mathbb{E} X(\tau) d\tau, \quad R_Y(t, s) = \int_a^t \int_a^s R_X(\tau, \theta) d\tau d\theta.$$

5. Эргодические процессы

Поднимем один очень практический вопрос: каким образом можно оценить математическое ожидание наблюдаемого процесса? Если у нас есть возможность запускать процесс много раз и регистрировать значения процесса в момент t , то среднее арифметическое этих значений может служить оценкой математического ожидания в момент t . Но что, если нет возможности заново запускать процесс, что если можно все время наблюдать только одну реализацию процесса? Тогда путь через множественные реализации (*ансамбли*) процесса нам недоступен и мы теряем возможность регистрировать возможные значения процесса в один момент времени. Но, оказывается, встречаются такие процессы, которые с течением времени бывают в любой сколь угодно

близкой окрестности любого своего значения! В таких случаях среднее (по исходам) значение процесса совпадает со средним по времени значением процесса.

Определение. Пусть процесс $\{X(t), t \geq 0\}$ второго порядка имеет постоянное математическое ожидание $\mathbb{E}X(t) = m$. Пусть $X(t)$ с.к.-интегрируем на любом отрезке $[0, T]$, $T \geq 0$. Рассмотрим процесс

$$\langle X \rangle_T = \frac{1}{T} \int_0^T X(t) dt, \quad T \geq 0,$$

где интеграл понимается в с.к.-смысле. Тогда если $\langle X \rangle_T \xrightarrow{\text{с.к.}} m$ при $T \rightarrow \infty$, то процесс $X(t)$ называют *эргодическим (в среднеквадратичном смысле) по математическому ожиданию*.

Определение. Процесс $X(t)$ называют *эргодическим по дисперсии*, если процесс $Y(t) = \dot{X}^2 = (X(t) - \mathbb{E}X(t))^2$ эргодичен по математическому ожиданию.

Определение. Процесс $X(t)$ называют *эргодическим по корреляционной функции*, если для каждого $\tau \geq 0$ процесс $Z(t) = \dot{X}(t)\dot{X}(t + \tau)$ эргодичен по математическому ожиданию.

Теорема (критерий эргодичности). *Процесс $\{X(t), t \geq 0\}$ второго порядка с постоянным математическим ожиданием и с.к.-интегрируемый на любом отрезке $[0, T]$ является эргодическим по математическому ожиданию тогда и только тогда, когда*

$$\lim_{T \rightarrow \infty} \frac{1}{T^2} \int_0^T \int_0^T R_X(t, s) dt ds = 0.$$

Доказательство. Достаточно заметить, что

$$\begin{aligned} & \mathbb{E} \left(\frac{1}{T} \int_0^T X(t) dt - m \right)^2 = \\ & = \mathbb{E} \left(\frac{1}{T} \int_0^T X(t) dt - \frac{1}{T} \int_0^T m dt \right)^2 = R_Y(T, T), \end{aligned}$$

где

$$Y(T) = \frac{1}{T} \int_0^T X(t) dt, \quad R_Y(T, T) = \frac{1}{T^2} \int_0^T \int_0^T R_X(t, s) dt ds,$$

и мы воспользовались ранее выписанной формулой для корреляционной функции с.к.-интеграла процесса. \square

Теорема (достаточное условие эргодичности). *Для того, чтобы процесс второго порядка $X(t)$, с.к.-интегрируемый на любом отрезке и с постоянным математическим ожиданием, был эргодичен по математическому ожиданию, достаточно, чтобы*

$$\lim_{|t-s| \rightarrow \infty} R_X(t, s) = 0, \quad \lim_{T \rightarrow \infty} \frac{1}{T} \max_{t \in [0, T]} R_X(t, t) = 0.$$

Доказательство. Пусть дан процесс с нужными свойствами. Тогда для любого $\varepsilon > 0$ найдется T_0 такой, что для любого $T = |t - s| > T_0$ выполнено $|R_X(t, s)| < \varepsilon$. Зафиксируем $\varepsilon > 0$, возьмем подходящий T_0 и разобьем квадрат $[0, T]^2$ на два множества:

$$G_1 = \{(t, s) \in [0, T]^2 : |t - s| > T_0\}, \quad G_2 = \{(t, s) \in [0, T]^2 : |t - s| \leq T_0\}.$$

Обозначим за S_1 и S_2 площади множеств G_1 и G_2 соответственно. Тогда

$$\begin{aligned} & \left| \frac{1}{T^2} \int_0^T \int_0^T R_X(t, s) dt ds \right| = \\ & = \left| \frac{1}{T^2} \iint_{G_1} R_X(t, s) dt ds + \frac{1}{T^2} \iint_{G_2} R_X(t, s) dt ds \right| \leq \\ & \leq \frac{1}{T^2} \left(\iint_{G_1} |R_X(t, s)| dt ds + \iint_{G_2} |R_X(t, s)| dt ds \right) \leq \\ & \leq \frac{\varepsilon S_1 + \max_{G_2} |R_X(t, s)| S_2}{T^2} \leq \varepsilon + 2 \max_{G_2} |R_X(t, s)| \frac{T_0}{T}, \end{aligned}$$

так как $S_1 \leq T^2$ и $S_2 = T^2 - 2(1/2)(T - T_0)^2 = 2TT_0 - T_0^2 \leq 2T_0T$. Из неравенства Коши–Буняковского следует, что

$$|R_X(t, s)|^2 \leq R_X(t, t)R_X(s, s),$$

поэтому $\max_{G_2} |R_X(t, s)| \leq \max_{\tau \in [0, T]} |R_X(\tau, \tau)|$. Теперь достаточно взять $T \geq T_0$ и такое, чтобы $\max_{\tau \in [0, T]} |R_X(\tau, \tau)|/T \leq \varepsilon$. \square

6. Стационарные процессы

6.1. Введение

Приступим к изучению одного из важнейших классов процессов – стационарных процессов. Это процессы, вероятностные свойства

которых не зависят от начала отсчета времени. Распределения их сечений не зависят от времени, их математические ожидания и дисперсии не зависят от времени, двумерные распределения (и любые другие n -мерные распределения) определяются только разницей между моментами времени сечений, но не от их абсолютных значений. Этими процессами описываются характеристики установившихся во времени явлений. Траектории таких процессов могут иметь нерегулярный вид, но при виде на них бросается в глаза общий характер поведения процесса, независимо от рассматриваемого участка времени. Например, пульсации силы тока или напряжения в электрической цепи (шум) моделируются обычно как стационарные процессы. Пульсации скорости в точке турбулентного установившегося потока моделируются как стационарные процессы. Теория и методы обработки сигналов, анализ временных рядов существенно опираются на понятие стационарных процессов и их свойства. Даже если вернуться к задаче оценки математического ожидания процесса, то свойство стационарности позволяет по имеющемуся единственному наблюдению оценивать среднее как среднее арифметическое значения процесса в разные моменты времени (ведь сечения имеют одинаковое распределение). В общем, всем, кто собирается заниматься обработкой данных, следует быть со стационарными процессами «на ты». Внимание к этому классу процессов уделяется неспроста: с одной стороны часто это очень хорошая модель реальности, а с другой – свойство стационарности имеет далеко идущие и очень полезные последствия. Перейдем к формальной части.

Определение. Процесс $\{X(t), t \in T\}$ называется *стационарным в узком смысле*, если все его конечномерные распределения не меняются при эквидистантном изменении моментов времени, то есть если для любых $t_1, \dots, t_n \in T$, и любого τ , при котором $t_1 + \tau, \dots, t_n + \tau \in T$, выполнено

$$F(x_1, \dots, x_n; t_1, \dots, t_n) = F(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau).$$

Другими словами, стационарность в узком смысле значит, что случайные векторы

$$(X(t_1), \dots, X(t_n)), (X(t_1 + \tau), \dots, X(t_n + \tau))$$

имеют одинаковое распределение.

Для $n = 1$ это значит, что $F(x; t)$ не зависит от t , а значит, и все моментные функции, связанные с одномерным распределением, не зависят от t . Например, математическое ожидание $\mathbb{E}X(t) = \text{const}$ (если существует) не зависит от t и дисперсия $\mathbb{D}X(t) = \text{const}$ (если существует) не зависит от t .

При $n = 2$ получаем, что для любых допустимых t_1, t_2, τ функция распределения $F(x_1, x_2; t_1, t_2) = F(x_1, x_2; t_1 + \tau, t_2 + \tau)$, то есть двумерное распределение не меняется, если t_1 и t_2 изменить на одну и ту же величину τ . Этим же свойством будут обладать все моментные функции процесса, основанные на двумерном распределении. Например, для корреляционной функции (если она существует) будет $R_X(t_1, t_2) = R_X(t_1 + \tau, t_2 + \tau) = R_X(0, t_2 - t_1)$. Это значит, что корреляционная функция зависит от t_1 и t_2 только через их разность.

Целесообразно подчеркнуть, что массу полезных свойств стационарных процессов мы получим не обязательно требуя «стационарности всех многомерных распределений», а потребовав лишь стационарность одномерного и двумерного распределений и конечность второго момента процесса.

Определение. Процесс $\{X(t), t \in T\}$ второго порядка называется *стационарным в широком смысле*, если его математическое ожидание не зависит от времени, а корреляционная функция зависит от аргументов только через их разность.

Если процесс является стационарным в узком смысле и *является при этом процессом второго порядка*, то он является стационарным в широком смысле. Если же процесс является стационарным в широком смысле, то это не значит, что он является стационарным в узком смысле, о чем, как нетрудно проверить, говорит пример процесса

$$Z(t) = X \cos t + Y \sin t,$$

где случайные величины X, Y независимы и принимают значения ± 1 с вероятностью $1/2$. Дело в том, что математическое ожидание и корреляционная функция не определяют в общем случае конечномерные распределения для любого набора времени. Но в некоторых случаях задание математического ожидания и корреляционной функции полностью определяет конечномерные распределения процессов.

Теорема. *Гауссовский процесс является стационарным в широком смысле тогда и только тогда, когда он является стационарным в узком смысле.*

Доказательство. Если процесс стационарен в узком смысле, то, так как он гауссовский (и поэтому – второго порядка), он стационарен и в широком смысле. Теперь пусть $\{X(t), t \geq 0\}$ – стационарный в широком смысле гауссовский процесс. Составим вектор из сечений процесса $X = (X(t_1), \dots, X(t_n))$. Тогда, по определению гауссовского процесса, характеристическая функция этого вектора имеет вид

$$\varphi_X(s) = \exp \left(i s^T m - \frac{1}{2} s^T R s \right),$$

где аргумент характеристической функции $s = (s_1, \dots, s_n)$, вектор математических ожиданий $m = (m_1, \dots, m_n)$, $m_i = \mathbb{E}X(t_i)$, корреляционная матрица $R = \|R_{ij}\|_{i,j=1}^n$, $R_{ij} = \mathbb{E}X(t_i)X(t_j) - \mathbb{E}X(t_i)\mathbb{E}X(t_j)$. Так как процесс стационарен в широком смысле, то при эквидистантном изменении моментов времени вектор m и матрица R не изменятся. Значит, не изменится характеристическая функция $\varphi_X(s)$, то есть не изменится распределение вектора X , который был выбран произвольно. По определению, $X(t)$ – стационарный в узком смысле процесс. \square

Замечание. Для того чтобы стационарный в широком смысле процесс второго порядка $X(t)$, с.к.-интегрируемый на любом отрезке и с постоянным математическим ожиданием, был эргодичен по математическому ожиданию, достаточно, чтобы

$$\lim_{t \rightarrow \infty} R_X(0, t) = 0,$$

так как у стационарного в широком смысле процесса дисперсия постоянна и поэтому условие $\lim_{T \rightarrow \infty} \max_{t \in [0, T]} R_X(t, t)/T = 0$ выполнено автоматически.

Итак, мы ввели понятие *стационарного процесса*. Эти процессы в науке и технологиях встречаются повсеместно, являются базовой моделью в теории обработки сигналов и временных рядов и моделируют установившиеся явления, вероятностные свойства которых не зависят от выбора начала отсчета времени. Значения таких процессов не могут все время расти или все время уменьшаться, они, как правило, нерегулярно колеблются вокруг математического ожидания. Такие процессы не могут быть всякими. Например, процесс вида $X(t) = A \cos \nu t$ со случайной величиной A и неслучайной ν не может быть стационарным, так как он равен нулю в определенные и неслучайные моменты времени, и не равен нулю в другие моменты времени. Интуитивно ясно что процесс $X(t) = A \cos(\nu t + \varphi)$ со случайными A, φ тоже не является стационарным, если φ имеет неравномерное распределение. А если φ имеет равномерное распределение, то такой процесс, как легко показать, является стационарным.

Интересно заметить, что в начале XX века Н. Винер попытался исследовать свойства таких нерегулярных колебаний, развил даже целую теорию на этот счет (обобщенный гармонический анализ), но полученные результаты имели непривычный и непонятный вид. Прогресс с описанием этих кривых появился именно только тогда, когда А. Н. Колмогоров, А. Я. Хинчин и Г. Крамер стали трактовать эти кривые именно как реализации случайных (стационарных) процессов, то есть вместо того, чтобы изучать кривые с определенными свойствами, они стали изучать случайные процессы с определенными свойствами, и

вся наука, которую мы сейчас изучаем, выросла из этих исследований. Эффективным оказался Фурье-анализ стационарных процессов, к которому мы сейчас и приступаем. Для этого нам потребуется обобщить понятие случайного процесса на комплексные величины, привычные и удобные для Фурье-анализа.

6.2. Комплексные случайные процессы

Определение. *Комплексным случайным процессом* $\{Z(t), t \in T\}$, определенном на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$, называется функция

$$Z(\omega, t) = X(\omega, t) + iY(\omega, t),$$

где $i^2 = -1$, $\omega \in \Omega$, а $X(t)$ и $Y(t)$ – два вещественных случайных процесса, определенных при $t \in T$ и принадлежащих тому же пространству $(\Omega, \mathcal{F}, \mathbb{P})$.

Определим математическое ожидание такого процесса $Z(t)$ по формуле

$$\mathbb{E}Z(t) = \mathbb{E}X(t) + i\mathbb{E}Y(t),$$

где $X(t) = \operatorname{Re} Z(t)$ – вещественная часть процесса $Z(t)$, а $Y(t) = \operatorname{Im} Z(t)$ – мнимая часть процесса $Z(t)$. Дисперсией комплексного процесса $Z(t)$ будем называть

$$\mathbb{D}Z(t) = \mathbb{E}\overset{\circ}{Z}(t)\overline{\overset{\circ}{Z}(t)} = \mathbb{E}|\overset{\circ}{Z}(t)|^2,$$

а корреляционной функцией

$$R_Z(t, s) = \mathbb{E}\overset{\circ}{Z}(t)\overline{\overset{\circ}{Z}(s)},$$

где черта означает комплексное сопряжение. Сразу отметим, что

$$R_Z(t, s) = \overline{R_Z(s, t)}.$$

Математическое ожидание и корреляционная функция комплексного процесса могут принимать комплексные значения. Дисперсия, как и прежде, может принимать только вещественные значения.

Определение. *Комплексным процессом второго порядка* будем называть комплексный случайный процесс $\{Z(t), t \in T\}$ со всюду конечным вторым моментом $\mathbb{E}|Z(t)|^2 < \infty$, $t \in T$.

Далее будем иметь дело с процессами, определенными на неотрицательной оси, то есть $T = [0, \infty)$.

6.3. Свойства корреляционной функции

У стационарных процессов корреляционная функция является функцией только разности компонент. Удобно поэтому ввести функцию

$$R_X(t) = R_X(t, 0) = \mathbb{E} \overline{\dot{X}(t) \dot{X}(0)},$$

которую мы так же будем называть корреляционной функцией стационарного процесса X . Эта функция будет определена для всех $t \in \mathbb{R}$, а не только для $t \in T = [0, \infty)$, если мы доопределим функцию $R_X(t, s)$ с помощью равенства $R_X(t, s) = R_X(t + h, s + h)$ для всех $h \in \mathbb{R}$. Займемся теперь выводом свойств корреляционных функций произвольных комплексных стационарных случайных процессов.

1) Эрмитовость.

Из выражений

$$R_X(t, s) = R_X(t - s, 0) = R_X(t - s),$$

$$R_X(t, s) = R_X(0, s - t) = \overline{R_X(s - t, 0)} = \overline{R_X(s - t)}$$

следует, что

$$R_X(t) = \overline{R_X(-t)}, \quad t \in \mathbb{R}.$$

2) Неотрицательность в нуле.

$$\mathbb{D}X(t) = \mathbb{D}X(0) = R_X(0) = \text{const} \geq 0.$$

3) Ограниченность.

Из неравенства Коши–Буняковского следует, что

$$|R_X(t)| \leq R_X(0), \quad t \in \mathbb{R}.$$

4) Непрерывность корреляционной функции и с.к.-непрерывность процесса.

Заметим, что

$$\mathbb{E}|X(t+h) - X(t)|^2 = \mathbb{E}|\dot{X}(t+h) - \dot{X}(t)|^2 = 2(R_X(0) - \text{Re } R_X(h)).$$

Если процесс $X(t)$ с.к.-непрерывен хоть в какой-нибудь точке t , то $\text{Re } R_X(h) \rightarrow R_X(0)$ при $h \rightarrow 0$. Но так как $|R_X(h)| \leq R_X(0)$, то получается, что $\text{Im } R_X(h) \rightarrow 0$, $h \rightarrow 0$. Значит, из с.к.-непрерывности процесса следует непрерывность функции $R_X(t)$ в $t = 0$. Наоборот, если $R_X(t)$

непрерывна в $t = 0$, то процесс $X(t)$ с.к.-непрерывен в любой точке t . Резюмируем: 1) стационарный процесс с.к.-непрерывен в какой-либо точке тогда и только тогда, когда его корреляционная функция $R_X(t)$ непрерывна в нуле, и 2) если стационарный процесс с.к.-непрерывен в одной точке, он с.к.-непрерывен в любой другой точке.

5) Непрерывность корр. функции в нуле и всюду.

Из неравенства Коши–Буняковского следует, что

$$\left| \mathbb{E} \left(\dot{X}(t+h) - \dot{X}(t) \right) \overline{\dot{X}(t-s)} \right|^2 \leq \mathbb{E} |\dot{X}(t+h) - \dot{X}(t)|^2 \cdot \mathbb{E} |\dot{X}(t-s)|^2,$$

что равносильно

$$|R_X(s+h) - R_X(s)|^2 \leq 2(R_X(0) - \operatorname{Re} R_X(h))R_X(0).$$

Отсюда следует, что если функция $R_X(t)$ непрерывна в нуле, то она непрерывна в любой точке. Из непрерывности в любой точке следует, в частности, непрерывность в нуле. В итоге получается, что непрерывность $R_X(t)$ в нуле равносильна непрерывности во всех точках.

6) Неотрицательная определенность.

Заметим, что для любых моментов времени t_1, \dots, t_n и комплексных чисел z_1, \dots, z_n выполняется

$$\sum_{i=1}^n \sum_{j=1}^n z_i \bar{z}_j R_X(t_i, t_j) = \mathbb{E} \left| \sum_{i=1}^n z_i \dot{X}(t_i) \right|^2 \geq 0.$$

Другими словами, это значит, что все матрицы вида $\|R_X(t_i, t_j)\|_{i,j=1}^n$ являются неотрицательно определенными над полем комплексных чисел. Это верно для любых процессов, не только стационарных. Для стационарных процессов можно подставить $R_X(t_i, t_j) = R_X(t_i - t_j)$.

Определение. Вещественная или комплексная функция *двух* переменных $f(t, s)$ называется *неотрицательно определенной*, если для любого $n \geq 1$, любых чисел $z_1, \dots, z_n \in \mathbb{C}$ и любых t_1, \dots, t_n

$$\sum_{i=1}^n \sum_{j=1}^n z_i \bar{z}_j f(t_i, t_j) \geq 0.$$

Другими словами, функция $f(t, s)$ называется *неотрицательно определенной*, если для любого $n \geq 1$ и любых t_1, \dots, t_n матрица $\|f(t_i, t_j)\|_{i,j=1}^n$ является неотрицательно определенной над полем \mathbb{C} .

Определение. Вещественная или комплексная функция *одной* переменной $f(t)$ называется *неотрицательно определенной*, если для любого $n \geq 1$, любых чисел $z_1, \dots, z_n \in \mathbb{C}$ и любых t_1, \dots, t_n

$$\sum_{i=1}^n \sum_{j=1}^n z_i \bar{z}_j f(t_i - t_j) \geq 0.$$

Другими словами, функция $f(t)$ называется *неотрицательно определенной*, если для любого $n \geq 1$ и любых t_1, \dots, t_n матрица $\|f(t_i - t_j)\|_{i,j=1}^n$ является неотрицательно определенной над полем \mathbb{C} .

В определении неотрицательной определенности есть много тонкостей, которые, на первый взгляд, не очевидны. Во-первых, в этих определениях z_i – комплексные числа, даже если рассматриваются вещественные функции f . Во-вторых, эти неравенства означают, что суммы являются *вещественными* и *неотрицательными*. Условие определенности сразу же приводит к тому, что неотрицательно определенная функция является эрмитовой: $f(t, s) = \overline{f(s, t)}$ для функций двух переменных и $f(-t) = \overline{f(t)}$ для функций одной переменной. Для вещественных функций двух переменных $f(t, s)$ эрмитовость означает симметричность $f(t, s) = f(s, t)$, а для вещественных функций одной переменной $f(t)$ эрмитовость означает четность $f(t) = f(-t)$. В-третьих, эрмитова матрица является неотрицательно определенной тогда и только тогда, когда *все* ее главные миноры неотрицательные (все, не только угловые). *Главным минором* называется определитель подматрицы, симметричной относительно главной диагонали, то есть подматрицы, у которой множества задающих её номеров столбцов и строк одинаковые. И, наконец, в-четвертых, из неотрицательной определенности функции $f(t)$ следует сразу, что $f(0) \geq 0$ и $|f(t)| \leq f(0)$ для любого t . Оказывается, что свойство неотрицательной определенности у корреляционной функции процесса является самым главным, из него следуют все остальные приведенные выше свойства. А именно, справедливы следующие теоремы¹⁰.

Теорема*. *Функция $R(t)$ является корреляционной функцией некоторого стационарного в широком смысле случайного процесса тогда и только тогда, когда она является неотрицательно определенной. Более того, если функция $R(t)$ неотрицательно определена, всегда найдется стационарный гауссовский процесс с корреляционной функцией $R(t)$.*

Теорема*. *Функция $R(t, s)$ является корреляционной функцией*

¹⁰См. Булинский А.В., Ширяев А.Н. Теория случайных процессов. Москва : ФИЗМАТЛИТ, 2005. С. 52.

ей некоторого случайного процесса второго порядка тогда и только тогда, когда она является неотрицательно определенной. Более того, если функция $R(t, s)$ неотрицательно определена, всегда найдется гауссовский процесс с корреляционной функцией $R(t, s)$.

6.4. Фурье-анализ стационарных процессов

Рассмотрим процесс

$$X(t) = \xi e^{i\lambda t},$$

где ξ – это случайная величина с конечным вторым моментом и нулевым математическим ожиданием, а λ – это неслучайная величина. Тогда $\mathbb{E}X(t) = 0$, а корреляционная функция

$$R_X(t, s) = \mathbb{E}X(t)\overline{X(s)} - \mathbb{E}X(t)\overline{\mathbb{E}X(s)} = \mathbb{E}|\xi|^2 e^{i\lambda(t-s)}$$

зависит от t и s только через их разность. По определению, этот процесс является стационарным в широком смысле.

Теперь рассмотрим процесс вида $X(t) = \xi_1 e^{i\lambda_1 t} + \xi_2 e^{i\lambda_2 t}$ с $\mathbb{E}\xi_1 = 0$ и $\mathbb{E}\xi_2 = 0$ и частотами $\lambda_1 \neq \lambda_2$ и выясним, при каких условиях этот процесс будет стационарным в широком смысле. Математическое ожидание $\mathbb{E}X(t) = 0$ не зависит от времени. Корреляционная функция

$$\begin{aligned} R_X(t, s) &= \mathbb{E}|\xi_1|^2 e^{i\lambda_1 \tau} + \mathbb{E}(\xi_1 \overline{\xi_2}) e^{i(\lambda_1 - \lambda_2)t + i\lambda_1 \tau} + \\ &+ \mathbb{E}(\xi_2 \overline{\xi_1}) e^{-i(\lambda_1 - \lambda_2)t + i\lambda_2 \tau} + \mathbb{E}|\xi_2|^2 e^{i\lambda_2 \tau}, \end{aligned}$$

где $\tau = t - s$, является функцией лишь τ тогда и только тогда, когда $\mathbb{E}(\xi_1 \overline{\xi_2}) = \mathbb{E}(\xi_2 \overline{\xi_1}) = 0$, что следует из линейной независимости функций перед этими коэффициентами. Получается, что процесс $X(t) = \xi_1 e^{i\lambda_1 t} + \xi_2 e^{i\lambda_2 t}$ для неравных частот λ_1 и λ_2 является стационарным в широком смысле тогда и только тогда, когда случайные величины ξ_1 и ξ_2 некоррелированы. Корреляционная функция в этом случае равна

$$R_X(t, s) = \mathbb{E}|\xi_1|^2 e^{i\lambda_1(t-s)} + \mathbb{E}|\xi_2|^2 e^{i\lambda_2(t-s)}.$$

Рассуждая аналогичным образом, можно доказать следующую теорему.

Теорема*. *Случайный процесс*

$$X(t) = \sum_{k=1}^n \xi_k e^{i\lambda_k t}, \quad \lambda_k \neq \lambda_m, \quad \mathbb{E}\xi_k = 0, \quad (3)$$

является стационарным в широком смысле тогда и только тогда, когда случайные величины ξ_k попарно некоррелированы. В этом случае корреляционная функция равна

$$R_X(t, s) = R(\tau) = \sum_{k=1}^n \mathbb{E}|\xi_k|^2 e^{i\lambda_k \tau}, \quad \tau = t - s. \quad (4)$$

Отметим, что для стационарности в широком смысле необходима и достаточна некоррелированность комплексных гармоник (экспонент) с не совпадающими частотами. Кроме того, в то время как процесс представляется в виде суммы некоррелированных гармоник с некоторыми случайными коэффициентами, корреляционная функция также представляет собой сумму гармоник с амплитудами, равными вторым моментам этих коэффициентов.

Один из важнейших результатов теории стационарных процессов состоит в том, что в некотором смысле все стационарные процессы имеют вид (3) и все корреляционные функции имеют вид (4). Только суммы эти могут быть в общем случае бесконечными или «непрерывными» (интегралы). И чтобы аккуратно определить эти интегралы и сформулировать нужное утверждение (о разложении процессов и корреляционных функций в ряды или интегралы Фурье), нам потребуется одно вспомогательное определение.

Определение. Центрированный комплексный процесс второго порядка $\{V(\lambda), \lambda \in \mathbb{R}\}$ называется процессом с ортогональными приращениями, если для любых $\lambda_1 < \lambda_2 < \lambda_3 < \lambda_4$

$$\mathbb{E}(V(\lambda_2) - V(\lambda_1))\overline{(V(\lambda_4) - V(\lambda_3))} = 0.$$

Этот процесс будет выступать в роли аналога последовательности некоррелированных случайных величин с нулевым математическим ожиданием. Без доказательства сформулируем две важнейшие теоремы.

Теорема Крамера (док-во см. далее). Любому стационарному с.к.-непрерывному процессу $X(t)$ с математическим ожиданием m_X соответствует случайный процесс с ортогональными приращениями $V(\lambda)$, такой, что с вероятностью единица выполняется равенство

$$X(t) = m_X + \int_{-\infty}^{+\infty} e^{i\lambda t} dV(\lambda).$$

Теорема Хинчина*. Для того чтобы непрерывная функция $R(t)$ представляла собой корреляционную функцию некоторого стационарного с.к.-непрерывного процесса $X(t)$, необходимо и достаточно,

чтобы она была представима в виде

$$R(t) = \int_{-\infty}^{+\infty} e^{i\lambda t} dS(\lambda),$$

где $S(\lambda)$ – функция, которая с точностью до неотрицательного множителя и произвольной аддитивной постоянной совпадает с функцией распределения некоторой случайной величины.

Доказательство теоремы Хинчина мы опускаем лишь потому, что оно не более чем техническое. Теорему Крамера докажем из-за идейной красоты ее доказательства.

Выражение из теоремы Крамера представляет $X(t)$ как непрерывную сумму элементарных гармонических составляющих $dV(\lambda)e^{i\lambda t}$, где $dV(\lambda)$ – амплитуда элементарной гармонической составляющей $e^{i\lambda t}$ с угловой частотой λ . Интеграл понимается как несобственный с.к.-интеграл Римана–Стилтьеса, то есть как предел при $a \rightarrow -\infty$ и $b \rightarrow +\infty$ предела в среднем квадратичном интегральной суммы

$$S_{\{\lambda_j\}}^{(a,b)} = \sum_{\lambda_j=a}^b e^{i\lambda_j t} (V(\lambda_j) - V(\lambda_{j-1})).$$

6.5. Спектральная функция и плотность

Функция $S(\lambda)$ в теореме Хинчина называется *спектральной функцией*. Если существует неотрицательная функция $\rho(\lambda) \geq 0$ такая, что

$$S(\lambda) = \int_{-\infty}^{\lambda} \rho(\tau) d\tau,$$

то эта функция $\rho(\lambda)$ называется *спектральной плотностью*. Спектральная плотность не обязана существовать. Но если она существует, то корреляционная функция запишется в виде

$$R_X(t) = \int_{-\infty}^{+\infty} e^{i\lambda t} \rho(\lambda) d\lambda.$$

Полезно отметить связь между неотрицательной определенностью функции и неотрицательностью ее Фурье-образа.

Теорема. Пусть функция $R(t)$ всюду непрерывна, абсолютно интегрируема на числовой прямой и пусть для нее выполнены условия разложения в интеграл Фурье. Тогда свойство неотрицательной определенности функции $R(t)$ равносильно свойству неотрицательности ее Фурье-преобразования.

Замечание к теореме. Условия разложения в интеграл Фурье могут быть разнообразными. Например, это условие выполнено, если функция $R(t)$ является всюду непрерывно дифференцируемой. Условие разложение в интеграл Фурье также выполнено, если Фурье-преобразование функции $R(t)$ является абсолютно интегрируемой на числовой оси функцией.

Доказательство. Из выполнения условий разложения функции в интеграл Фурье следует, что

$$R(t) = \int_{-\infty}^{+\infty} e^{i\lambda t} \rho(\lambda) d\lambda, \quad \rho(\lambda) = \int_{-\infty}^{+\infty} e^{-i\lambda t} R(t) dt.$$

Если функция $\rho(\lambda)$ неотрицательная, то справедливы условия теоремы Хинчина и функция $R(t)$ является корреляционной функцией некоторого стационарного процесса, а значит, является неотрицательно определенной функцией. Наоборот, если функция $R(t)$ является неотрицательно определенной, то она является корреляционной функцией некоторого стационарного процесса, но из теоремы Хинчина и единственности разложения в интеграл Фурье следует, что $\rho(\lambda)$ – это спектральная плотность этого процесса, то есть неотрицательная функция. \square

6.6. Теорема Крамера

Теорема Крамера. Любому стационарному с.к.-непрерывному процессу $X(t)$ с математическим ожиданием m_X соответствует случайный процесс с ортогональными приращениями $V(\lambda)$, такой, что с вероятностью единица выполняется равенство

$$X(t) = m_X + \int_{-\infty}^{+\infty} e^{i\lambda t} dV(\lambda).$$

Доказательство. Рассмотрим множество случайных величин $H(\overset{\circ}{X})$, образованное из сечений случайного процесса $\overset{\circ}{X}(t) = X(t) - m_X$, их линейных комбинаций

$$\alpha_1 \overset{\circ}{X}(t_1) + \dots + \alpha_n \overset{\circ}{X}(t_n)$$

и с.к.-пределов последовательностей таких комбинаций. На этом множестве случайных величин введем скалярное произведение

$$(\eta_1, \eta_2) = \mathbb{E}\eta_1\bar{\eta}_2$$

для любых элементов η_1, η_2 из $H(\overset{\circ}{X})$. Будем считать два элемента тождественными, если расстояние между ними, определяемое этим скалярным произведением, равно нулю. Множество $H(\overset{\circ}{X})$ с введенным таким образом скалярным произведением является гильбертовым пространством. Обратим внимание на то, что все элементы этого пространства имеют нулевые математические ожидания.

Обозначим через $L_2(S)$, где S – спектральная функция процесса $X(t)$, множество всех неслучайных комплексных функций $g(\lambda)$, для которых интеграл Лебега–Стилтьеса

$$\int_{-\infty}^{+\infty} |g(\lambda)|^2 dS(\lambda)$$

существует и конечен. Множество $L_2(S)$ станет гильбертовым пространством, обозначаемым $H(S)$, если помимо сложения и умножения на скаляры, определяемых обычным образом, ввести в нем операцию скалярного произведения g_1 и g_2 , положив

$$(g_1, g_2) = \int_{-\infty}^{+\infty} g_1(\lambda) \overline{g_2(\lambda)} dS(\lambda).$$

Элементы g_1 и g_2 будем рассматривать как тождественные, если расстояние между ними, определяемое скалярным произведением, равно нулю, то есть если

$$\int_{-\infty}^{+\infty} |g_1(\lambda) - g_2(\lambda)|^2 dS(\lambda) = 0.$$

Установим соответствие между $H(\overset{\circ}{X})$ и $H(S)$ следующим образом. Для каждого вещественного t элементу $\overset{\circ}{X}(t) \in H(\overset{\circ}{X})$ сопоставим элемент $e^{it\lambda} \in H(S)$. По теореме Хинчина

$$\mathbb{E} \overset{\circ}{X}(t) \overline{\overset{\circ}{X}(s)} = \int_{-\infty}^{+\infty} e^{it\lambda} e^{-is\lambda} dS(\lambda),$$

так что скалярное произведение при этом соответствии остается тем же самым.

Теперь распространим это соответствие на конечные линейные комбинации элементов $\overset{\circ}{X}(t_k)$ и $e^{it_k\lambda}$, а именно элементу

$$\eta = \alpha_1 \overset{\circ}{X}(t_1) + \dots + \alpha_n \overset{\circ}{X}(t_n) \tag{5}$$

сопоставим элемент

$$g(\lambda) = \alpha_1 e^{it_1\lambda} + \dots + \alpha_n e^{it_n\lambda}. \tag{6}$$

В силу теоремы Хинчина и линейных свойств скалярного произведения для соответствующих пар η_1, η_2 и g_1, g_2

$$(\eta_1, \eta_2) = (g_1, g_2) \quad (7)$$

и для соответствующих квадратов расстояний

$$\mathbb{E}|\eta_1 - \eta_2|^2 = \int_{-\infty}^{+\infty} |g_1(\lambda) - g_2(\lambda)|^2 dS(\lambda). \quad (8)$$

Отсюда следует, что установленное соответствие взаимно однозначно, поскольку η_1 и η_2 совпадают тогда и только тогда, когда совпадают g_1 и g_2 .

Теперь пусть η_1, η_2, \dots – последовательность случайных величин вида (5), сходящаяся в среднем квадратичном к некоторой случайной величине η . Тогда из равенств

$$\mathbb{E}|\eta_m - \eta_n|^2 = \int_{-\infty}^{+\infty} |g_m(\lambda) - g_n(\lambda)|^2 dS(\lambda)$$

следует, что неслучайные функции $g_n(\lambda)$ сходятся в среднем квадратичном относительно спектральной функции $S(\lambda)$. Если $g(\lambda)$ – их предел, то элементу η мы поставим в соответствие элемент $g(\lambda)$ и обратно, если отправляться от сходящейся последовательности $g_n(\lambda)$.

Далее, любой элемент из $H(\overset{\circ}{X})$ по определению является с.к.-пределом некоторой последовательности (5). Аналогично любой элемент из $H(S)$ является с.к.-пределом некоторой последовательности элементов вида (6). Таким образом, мы распространили соответствие на все элементы из $H(\overset{\circ}{X})$ и $H(S)$. Из свойств с.к.-сходимости следует, что соотношения (7) и (8) выполняются для всех $\eta_1, \eta_2 \in H(\overset{\circ}{X})$ и $g_1, g_2 \in H(S)$, так что скалярные произведения и расстояния сохраняются при рассматриваемом соответствии и, следовательно, последнее взаимно однозначно.

Обозначим $g(t) = I(t \leq 0)$. Тогда для любого λ_0 функция

$$g(\lambda - \lambda_0) = I(\lambda \leq \lambda_0)$$

принадлежит $H(S)$. Если обозначить за $V(\lambda_0)$ соответствующий элемент из $H(\overset{\circ}{X})$, то $\mathbb{E}V(\lambda_0) = 0$, так как $V(\lambda_0) \in H(\overset{\circ}{X})$, а приращение $V(\lambda_1) - V(\lambda_0)$ будет отвечать разности

$$g(\lambda - \lambda_1) - g(\lambda - \lambda_0) = I(\lambda \leq \lambda_1) - I(\lambda \leq \lambda_0).$$

При $\lambda_0 < \lambda_1$ эта разность равна единице на интервале $(\lambda_0, \lambda_1]$ и нулю вне его. Если интервалы (λ_0, λ_1) и (λ_2, λ_3) не пересекаются, то в силу инвариантности скалярного произведения

$$\begin{aligned} & \mathbb{E}(V(\lambda_3) - V(\lambda_2))\overline{(V(\lambda_1) - V(\lambda_0))} = \\ & = \int_{-\infty}^{+\infty} (g(\lambda - \lambda_3) - g(\lambda - \lambda_2))\overline{(g(\lambda - \lambda_1) - g(\lambda - \lambda_0))} dS(\lambda) = 0, \end{aligned}$$

так что процесс $V(\lambda)$ – это процесс с ортогональными приращениями. Полагая $\lambda_3 = \lambda_1$ и $\lambda_2 = \lambda_0$, получаем

$$\mathbb{E}|V(\lambda_1) - V(\lambda_0)|^2 = S(\lambda_1) - S(\lambda_0). \quad (9)$$

Пусть $-A = \lambda_1 < \lambda_2 < \dots < \lambda_{n+1} = A$ – некоторое разбиение интервала $(-A, A)$. Тогда случайная величина

$$\eta = \sum_{j=1}^n e^{it\lambda_j} (V(\lambda_{j+1}) - V(\lambda_j))$$

соответствует функции

$$g(\lambda) = \begin{cases} e^{it\lambda_j}, & \lambda \in (\lambda_j, \lambda_{j+1}], \\ 0, & \text{иначе.} \end{cases}$$

Когда A и n стремятся к бесконечности и одновременно максимальное расстояние между соседними точками λ_j стремится к нулю, $g(\lambda)$ сходится в среднем квадратичном к $e^{it\lambda} \in H(S)$, а η сходится к с.к.-интегралу

$$\int_{-\infty}^{+\infty} e^{it\lambda} dV(\lambda) \in H(\overset{\circ}{X}).$$

Пределы являются соответствующими друг другу элементами. Но мы знаем, что $e^{it\lambda}$ соответствует $\overset{\circ}{X}(t)$, поэтому

$$\overset{\circ}{X}(t) = \int_{-\infty}^{+\infty} e^{it\lambda} dV(\lambda).$$

Теорема доказана. \square

Заметим, что для любого η вида (5) и соответствующей функции $g(\lambda)$ вида (6) имеет место соотношение

$$\eta = \int_{-\infty}^{+\infty} g(\lambda) dV(\lambda). \quad (10)$$

Оказывается, что это соотношение выполняется вообще для любой пары $\eta \in H(\overset{\circ}{X})$ и $g(\lambda) \in H(S)$. Это утверждение оставляем без доказательства.

6.7. О физическом смысле спектральной плотности

Пусть дан центрированный процесс (сигнал) $X(t)$. Спектральное представление $X(t) = \int_{-\infty}^{+\infty} e^{it\lambda} dV(\lambda)$ дает разложение этого сигнала на элементарные гармонические составляющие. Мощность сигнала – это временное среднее квадрата амплитуды сигнала, что для гармонических сигналов означает квадрат амплитуды множителя перед гармоникой, то есть $|dV(\lambda)|^2$ в нашем случае. Из (9) следует, что $\mathbb{E}|dV(\lambda)|^2 = dS(\lambda) = \rho(\lambda)d\lambda$, что означает, что $dS(\lambda) = \rho(\lambda)d\lambda$ – это средняя мощность, отвечающая гармонической компоненте с частотой, лежащей в интервале $(\lambda, \lambda + d\lambda)$. Таким образом, спектральная плотность – это плотность распределения средней мощности сигнала $X(t)$ в частотной области. Средняя мощность, соответствующая интервалу частот $\lambda_1 < \lambda \leq \lambda_2$, равна $S(\lambda_2) - S(\lambda_1)$. Для всей области частот $(-\infty, +\infty)$ соответствующая средняя мощность равна

$$\mathbb{E}|X(0)|^2 = S(+\infty) - S(-\infty).$$

6.8. О белом шуме

В приложениях встречаются стационарные процессы со спектральной плотностью

$$\rho_{\lambda_0}(\lambda) = \begin{cases} C, & |\lambda| < \lambda_0 \\ 0 & |\lambda| \geq \lambda_0 \end{cases},$$

где C – постоянная. Такие процессы являются моделями сигналов, излучающих на всех частотах диапазона $|\lambda| < \lambda_0$ с одинаковой мощностью, и называются поэтому *белыми шумами*. Корреляционная функция подобных процессов имеет вид

$$R_{\lambda_0}(t) = 2C \cdot \frac{\sin \lambda_0 t}{t},$$

где считается $R_{\lambda_0}(0) = 2C\lambda_0$. Часто модель таких сигналов идеализируют и считают, что $\lambda_0 = \infty$, то есть что сигнал излучает во всем диапазоне частот с постоянной мощностью. И хотя, строго говоря, таких процессов не существует (ведь у непрерывных стационарных процессов спектральная плотность абсолютно интегрируема), тем не менее удобно считать, что процесс все же существует, а его спектральная плотность есть константа. Такие процессы тоже называют *белыми шумами*. Их корреляционная функция рассматривается как слабый предел функций $R_{\lambda_0}(t)$ при $\lambda_0 \rightarrow \infty$, который, как известно, выражается через

дельта-функцию Дирака:

$$R(t) = 2C\pi\delta(t).$$

За всем этим стоит развитая теория обобщенных случайных процессов¹¹.

Сечения белого шума независимы и в принципе могут иметь всякое распределение, но распространена модель гауссовского белого шума. Важно заметить, что гауссовский белый шум получается, если (в определенном смысле) продифференцировать винеровский процесс. Тут полезна следующая аналогия. Мы знаем, что винеровский процесс является пределом случайных блужданий, которые представляют собой сумму последовательности независимых одинаково распределенных случайных величин (*дискретный белый шум*). При переходе к пределу случайное блуждание переходит в непрерывный по времени винеровский процесс, а последовательность этих величин – в непрерывный по времени белый шум.

6.9. Линейная теория стационарных процессов

Пусть $X(t)$ – стационарный с.к.-непрерывный процесс с нулевым средним, тогда он допускает представление

$$X(t) = \int_{-\infty}^{+\infty} e^{it\lambda} dV_X(\lambda),$$

где $V_X(\lambda)$ – процесс с ортогональными приращениями, причем по формуле (9) $\mathbb{E}|dV_X(\lambda)|^2 = dS_X(\lambda)$. Пусть \mathcal{L} обозначает преобразование, переводящее процесс $X(t)$ в новый случайный процесс

$$Y(t) = \mathcal{L}X(t) = \sum_{n=1}^N c_n X(t_n + t).$$

Используя спектральное представление для $X(t)$, получаем

$$Y(t) = \int_{-\infty}^{+\infty} \Phi(\lambda) e^{it\lambda} dV_X(\lambda), \quad \Phi(\lambda) = \sum_{n=1}^N c_n e^{it_n\lambda}.$$

Тем же способом, как и (10), это соответствие можно распространить на все функции $\Phi(\lambda)$ из $H(S)$, а всякий процесс $Y(t)$, полученный

¹¹Коралов Л.Б., Синай Я.Г. Теория вероятностей и случайные процессы. Москва : МЦНМО, 2014.

таким образом, можно рассматривать как результат применения к $X(t)$ некоторой линейной операции¹².

Заметим, что процесс $Y(t)$ является стационарным – его математическое ожидание равно нулю, так как $Y(t) \in H(\overset{\circ}{X})$, а корреляционная функция равна

$$R_Y(t, s) = \mathbb{E}Y(t)\overline{Y(s)} = \int_{-\infty}^{+\infty} e^{i(t-s)} |\Phi(\lambda)|^2 dS_X(\lambda),$$

как следует из соотношения между расстояниями в пространствах $H(\overset{\circ}{X})$ и $H(S)$. Более того, отсюда сразу следует, что

$$dS_Y(\lambda) = |\Phi(\lambda)|^2 dS_X(\lambda).$$

Если $X(t)$ имеет спектральную плотность $\rho_X(\lambda)$, то и $Y(t)$ имеет спектральную плотность $\rho_Y(\lambda)$, причем

$$\rho_Y(\lambda) = |\Phi(\lambda)|^2 \rho_X(\lambda).$$

Важно отметить, что при линейных преобразованиях сами процессы могут изменяться весьма сложным образом, который трудно описать или записать, в то время как в частотной области преобразование записывается особенно легко: надо просто умножить спектральную плотность на некоторую функцию. Если над процессом совершается ряд линейных операций, то это приводит к существенному усложнению процесса во временной области, но при этом лишь к простому наращиванию множителей перед спектральной плотностью в частотной области. Функция $\Phi(\lambda)$ называется *частотной характеристикой* линейного преобразования \mathcal{L} и характеризует отклик системы на гармонические колебания.

В качестве полезного примера линейной операции рассмотрим операцию с.к.-дифференцирования. Предположим, что с.к.-непрерывный стационарный процесс $X(t)$ имеет конечный *второй спектральный момент*

$$\int_{-\infty}^{+\infty} \lambda^2 dS_X(\lambda) < \infty.$$

Из разложения корреляционной функции по теореме Хинчина отсюда следует, что корреляционная функция дважды непрерывно дифференцируема, а значит процесс является с.к.-дифференцируемым. Тогда по

¹²Дуб Дж.Л. Вероятностные процессы. Москва : Издательство иностранной литературы, 1956.

критерию с.к.-интегрируемости по Риману–Стилтьесу с.к.-интеграл

$$\int_{-\infty}^{+\infty} \lambda e^{i\lambda t} dV_X(\lambda)$$

существует и

$$\begin{aligned} \frac{X(t+h) - X(t)}{h} - \int_{-\infty}^{+\infty} i\lambda e^{i\lambda t} dV_X(\lambda) &= \\ &= \int_{-\infty}^{+\infty} \left(\frac{e^{i\lambda h} - 1}{h} - i\lambda \right) e^{it\lambda} dV_X(\lambda). \end{aligned}$$

Следовательно,

$$\begin{aligned} \mathbb{E} \left| \frac{X(t+h) - X(t)}{h} - \int_{-\infty}^{+\infty} i\lambda e^{i\lambda t} dV_X(\lambda) \right|^2 &= \\ &= \int_{-\infty}^{+\infty} \left| \frac{e^{i\lambda h} - 1}{h} - i\lambda \right|^2 dS_X(\lambda). \end{aligned}$$

Используя возможность перехода к пределу под знаком интеграла для мажорируемой последовательности подынтегральных функций, легко показать, что это выражение стремится к нулю при $h \rightarrow 0$. По определению с.к.-производной это означает равенство почти наверное

$$X'(t) = \int_{-\infty}^{+\infty} i\lambda e^{i\lambda t} dV_X(\lambda).$$

Таким образом, с.к.-дифференцирование можно рассматривать как линейную операцию с частотной характеристикой $\Phi(\lambda) = i\lambda$.

7. Марковские случайные процессы

Мы уже познакомились с некоторыми важнейшими классами случайных процессов: стационарные, эргодические, нормальные процессы – и изучили их свойства. Каждый из этих классов процессов определяется связями между сечениями – это то, что делает теорию вероятностей и случайные процессы содержательными, богатыми на результаты и самостоятельными разделами математики, а не просто частью теории меры или функционального анализа.

Мы приступаем к изучению еще одного важнейшего класса процессов, который является обобщением динамических систем, то есть

систем, будущие состояния которых определяются лишь текущим состоянием. Это марковские процессы, и им будет посвящена вся оставшаяся часть учебного пособия.

Определение. Случайный процесс $\{X(t), t \in T\}$ называется *марковским*, если

$$\begin{aligned} \mathbb{P}(X(t_{n+1}) \in B | X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1) = \\ = \mathbb{P}(X(t_{n+1}) \in B | X(t_n) = x_n) \end{aligned} \quad (11)$$

для любого $n \geq 2$, любых $t_1 < t_2 < \dots < t_n < t_{n+1}$ из T , любых чисел x_1, \dots, x_n и любого борелевского множества B , для которых условные вероятности выше определены.

Грубо говоря, случайный процесс называется *марковским*, если вероятностные характеристики его «будущего» (в момент t_{n+1}) зависят лишь от значения процесса, которое он принял в «настоящем» (в момент t_n), и не зависят от значений, которые процесс принимал в «прошлом» (в моменты t_{n-1}, \dots, t_1). Этим свойством обладают обычные, неслучайные детерминированные процессы, когда текущее состояние системы однозначно определяет будущие состояния. В марковских же процессах однозначно определены вероятностные характеристики процесса в будущем, если известно состояние системы в настоящем. Здесь следует сказать о встречающейся чрезмерно вольной трактовке понятий «будущее», «настоящее» и «прошлое» и связи между ними. Обратите внимание, что в определении выше важно, что в условии стоят события вида $X(t_k) = x_k$, а не, например, $X(t_k) \in B_k$ для борелевских множеств B_k , так как иначе определение было бы неэквивалентным и неверным. Под *марковостью* процесса следует понимать именно то, что написано в определении.

Область применения марковских процессов необъятна, они широко используются для описания явлений в физике (особенно в термодинамике и процессах диффузии), химии, экономике, финансовой математике, теории обработки сигналов, навигации, теории информации, теории распознавания речи, в IT-технологиях, являются базовой моделью в теории машинного обучения с подкреплением.

Без доказательства примем следующее утверждение.

Теорема*. *Всякий процесс с независимыми приращениями является марковским процессом.*

Из этой теоремы следует, например, что винеровский процесс, пуассоновский процесс, а также процессы случайного блуждания являются марковскими процессами. Если процесс является марковским, то отсюда в общем случае не следует, что он является процессом с независимыми приращениями.

Марковские процессы классифицируются по множеству значений, которые они могут принимать, и множеству времен T . Множество значений процесса $X(t)$ обозначим буквой S и будем называть его *множеством состояний*, а его элементы – *состояниями* процесса $X(t)$. Бывают марковские процессы, для которых $S = \mathbb{R}$, $S = [a, b]$, $S = \mathbb{Z}$, $S = \{0, 1, 2\}$ и др.

Бывают марковские процессы, в которых

- 1) множество состояний S дискретно и время T дискретно,
- 2) множество состояний S дискретно, время T непрерывно,
- 3) множество состояний S непрерывно, время T дискретно,
- 4) множество состояний S непрерывно и время T непрерывно.

Марковские процессы из первых двух пунктов называются *цепями Маркова*, или *марковскими цепями*. Различают *дискретные цепи Маркова* (первый пункт) и *непрерывные цепи Маркова* (второй пункт). Винеровский процесс относится к пункту 4, пуассоновский – к пункту 2. К пункту 3 относятся, например, случайные блуждания, в которых каждый шаг – это непрерывно распределенная случайная величина.

Приступим к изучению дискретных цепей Маркова, эта модель близка к понятию *конечного автомата* и широко применяется в теории игр, процессах принятия решений, лингвистике, обработке сигналов. На основе марковских цепей строится теория машинного обучения с подкреплением, динамическое программирование, работают системы ранжирования страниц в поисковых системах (например, PageRank компании Google).

8. Дискретные цепи Маркова

8.1. Базовые понятия и свойства

Начнем с изучения дискретных цепей Маркова. Как было сказано ранее, это – марковские процессы с дискретным множеством состояний и дискретным временем. Без ограничения общности будем считать, что множество состояний S процесса есть подмножество целых чисел, возможно, совпадающее с множеством целых чисел. Кроме того, так как нас будут интересовать главным образом асимптотические свойства таких процессов, мы рассмотрим случай со счетным T и для определенности примем $T = \{0, 1, 2, \dots\}$. Тогда дискретная цепь Маркова представляет собой случайную последовательность, обладающую свойством (11). Марковское свойство (11) для дискретных цепей Маркова можно переписать в терминах случайных последовательностей.

Определение. Случайная последовательность $\{X_k\}$, каждая компонента которой принимает значения из некоторого множества

$$S \subseteq \mathbb{Z}, |S| \leq \infty$$

и обладающая свойством

$$\begin{aligned} \mathbb{P}(X_{m_n} = x_n \mid X_{m_{n-1}} = x_{n-1}, \dots, X_{m_0} = x_0) = \\ = \mathbb{P}(X_{m_n} = x_n \mid X_{m_{n-1}} = x_{n-1}) \end{aligned} \quad (12)$$

для каждого $n \geq 1$, $m_0 < m_1 < \dots < m_n$ и $x_0, \dots, x_n \in S$, для которых указанные условные вероятности определены, называется *дискретной цепью Маркова*. Если множество состояний конечно, то есть $|S| < \infty$, то цепь называют *конечной*. Если же множество состояний счетно, то и цепь в таком случае называют *счетной*.

Свойство (12) – это ровно то же свойство (11), записанное в более привычных для последовательностей обозначениях. Но на самом деле свойство (12) равносильно другому, на первый взгляд, более простому свойству:

$$\begin{aligned} \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \\ = \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}) \end{aligned} \quad (13)$$

для каждого $n \geq 1$ и x_0, \dots, x_n , для которых условные вероятности определены. Эту эквивалентность предлагается доказать в одной из задач второго задания. Свойства (11), (12) или (13) называют также *марковским свойством* процесса.

Случайная величина X_n – это состояние цепи в момент времени $n \geq 0$, поэтому событие $\{X_n = j\}$ читается как «на шаге n цепь находится в состоянии $j \in S$ ». Траектория дискретной цепи Маркова – это просто последовательность состояний. Удобно представлять себе цепь в виде некоторого графа, вершинами которого являются состояния, а ориентируемые ребра между ними обозначают возможные переходы между состояниями.

Изучение цепей Маркова начнем с записи их конечномерных распределений. А именно, зададимся произвольными числами $m_0 < m_2 < \dots < m_n$, состояниями x_0, \dots, x_n и, воспользовавшись марковским свойством (12), запишем функцию вероятности случайного вектора $(X_{m_0}, \dots, X_{m_n})$:

$$\mathbb{P}(X_{m_0} = x_0, \dots, X_{m_n} = x_n) = \mathbb{P}\left(\bigcap_{k=m_0}^{m_n} \{X_k = x_k\}\right) =$$

$$\begin{aligned}
&= \mathbb{P} \left(X_{m_n} = x_n \mid \bigcap_{k=m_0}^{m_{n-1}} \{X_k = x_k\} \right) \mathbb{P} \left(\bigcap_{k=m_0}^{m_{n-1}} \{X_k = x_k\} \right) = \\
&= \mathbb{P}(X_{m_n} = x_n \mid X_{m_{n-1}} = x_{n-1}) \mathbb{P} \left(\bigcap_{k=m_0}^{m_{n-1}} \{X_k = x_k\} \right) = \dots = \\
&= \mathbb{P}(X_{m_0} = x_0) \prod_{k=1}^n \mathbb{P}(X_{m_k} = x_k \mid X_{m_{k-1}} = x_{k-1}).
\end{aligned}$$

Это значит, что конечномерные распределения дискретных цепей Маркова определяются двумя вещами: одномерными распределениями (см. первый множитель полученного выражения) и условными вероятностями, характеризующие вероятность перейти из одного состояния в один момент времени в другое состояние в другой момент времени. Введем поэтому следующее определение.

Определение. Число $p_{ij}(m, n) = \mathbb{P}(X_n = j \mid X_m = i)$ называется *вероятностью перехода* из состояния $i \in S$ в момент m в состояние $j \in S$ в момент $n \geq m$. Для простоты мы будем далее предполагать, что эти числа существуют для всех $i, j, m, n \geq m$, иначе придется делать множество оговорок, не привносящих содержание в материал, но затуманивающих дело. Из чисел $p_{ij}(m, n)$ составим матрицу (возможно бесконечную в счетных цепях) $P(m, n) = \|p_{ij}(m, n)\|$ и назовем ее *матрицей перехода* от момента m к моменту n .

Например, в случае $S = \{1, 2, \dots, N\}$, $N < \infty$ матрица перехода имеет вид:

$$P(m, n) = \begin{bmatrix} p_{11}(m, n) & p_{12}(m, n) & \cdots & p_{1N}(m, n) \\ p_{21}(m, n) & & & \\ \vdots & & \ddots & \\ p_{N1}(m, n) & & & p_{NN}(m, n) \end{bmatrix}.$$

Обратим внимание на то, что в любой цепи Маркова сумма

$$\sum_{j \in S} \mathbb{P}(X_n = j \mid X_m = i) = 1$$

для каждого $i \in S$, так как множество событий $\{X_n = j\}$ для $j \in S$ образует полную систему непересекающихся событий. Это значит, что сумма компонент переходной матрицы в любой строке равна единице, то есть

$$\forall i \in S \quad \sum_{j \in S} p_{ij}(m, n) = 1.$$

Кроме того, очевидно, что все компоненты переходной матрицы неотрицательные и не превышают единицу. Вообще матрица с неотрицательными и ограниченными сверху единицей компонентами, сумма компонент в каждой строке которой равна единице, называется *стохастической матрицей*.

Теорема. Для любых $n \geq k \geq m \geq 0$ выполнено равенство

$$P(m, n) = P(m, k)P(k, n), \quad (14)$$

что равносильно

$$p_{ij}(m, n) = \sum_{l \in S} p_{il}(m, k)p_{lj}(k, n). \quad (15)$$

Доказательство. По формуле полной вероятности¹³

$$\begin{aligned} p_{ij}(m, n) &= \mathbb{P}(X_n = j \mid X_m = i) = \\ &= \sum_{l \in S} \mathbb{P}(X_n = j \mid X_m = i, X_k = l) \mathbb{P}(X_k = l \mid X_m = i). \end{aligned}$$

Из марковского свойства (12) следует, что

$$\mathbb{P}(X_n = j \mid X_m = i, X_k = l) = \mathbb{P}(X_n = j \mid X_k = l),$$

откуда сразу получаем $p_{ij}(m, n) = \sum_{l \in S} p_{lj}(k, n)p_{il}(m, k)$. \square

Равенство (14) (или (15)) называют *уравнением Колмогорова–Чепмена*. Уравнением оно называется потому, что, по сути, является функциональным уравнением на функции $p_{ij}(\cdot, \cdot)$, хотя мы будем пользоваться этой формулой не для того, чтобы решать функциональное уравнение, а чтобы вычислять вероятности перехода для разных моментов m, n на основе известных вероятностей перехода для других моментов.

Из упомянутой формулы, в частности, следует, что

$$P(0, n) = P(0, 1)P(1, 2) \cdots P(n-1, n). \quad (16)$$

Определение. Вероятность $\pi_k(n) = \mathbb{P}(X_n = k)$, $k \in S$, называется *вероятностью состояния k* в момент $n \geq 0$. Вектор вероятностей

¹³Здесь формула полной вероятности применяется к уже условной вероятности $\mathbb{P}(X_n = j \mid X_m = i)$. Вывод этой формулы точно такой же, что и для безусловной вероятности. А вообще условная вероятность – это тоже вероятность, просто на специально выбранном вероятностном пространстве, на котором событие из условия происходит с вероятностью 1.

$\pi(n) = [\pi_0(n), \pi_1(n), \dots]^T$ называют *распределением вероятностей состояний* в момент $n \geq 0$.

Теорема. *Распределение вероятностей состояний $\pi(n)$ связано с распределением вероятностей состояний $\pi(n-1)$ соотношением*

$$\pi(n) = P^T(n-1, n)\pi(n-1). \quad (17)$$

Доказательство. По формуле полной вероятности:

$$\begin{aligned} \mathbb{P}(X_n = k) &= \sum_{j \in S} \mathbb{P}(X_n = k | X_{n-1} = j) \mathbb{P}(X_{n-1} = j) = \\ &= \sum_{j \in S} p_{jk}(n-1, n) \pi_j(n-1). \square \end{aligned}$$

В дальнейшем нас будут интересовать только *однородные цепи Маркова*.

Определение. Если $P(m, n) = P(0, n-m)$ для любых m и $n \geq m$, то соответствующую цепь Маркова называют *однородной*.

Так как для однородной цепи Маркова всегда выполнено $P(m, n) = P(0, n-m)$, то введем специальное обозначение $P(n) = P(0, n)$ для переходной матрицы от момента 0 к моменту n . Уравнение Колмогорова–Чепмена (14) для однородной цепи можно переписать как

$$P(n) = P(n-m)P(m),$$

или, что эквивалентно,

$$p_{ij}(n) = \sum_{k \in S} p_{ik}(n-m) p_{kj}(m).$$

Из первой формулы сразу следует

$$P(n_1 + \dots + n_m) = P(n_1) \dots P(n_m),$$

что равносильно

$$p_{ij}(n_1 + \dots + n_m) = \sum_{k_1 \in S} \dots \sum_{k_{m-1} \in S} p_{ik_1}(n_1) \dots p_{k_{m-1}j}(n_m).$$

Отметим важнейшее следствие этой формулы:

$$p_{ij}(n_1 + \dots + n_m) \geq p_{ik_1}(n_1) \dots p_{k_{m-1}j}(n_m)$$

для любых $i, j, k_1, \dots, k_{m-1}, n_1, \dots, n_m$. Этим неравенством мы будем пользоваться очень часто.

Кроме того, будем обозначать буквой P без скобок матрицу $P(1)$, то есть матрицу перехода за один шаг. Заметим, что для переходной матрицы $P(n)$ свойство (16) означает

$$P(n) = P(1)P(1) \cdots P(1) = P^n,$$

а свойство (17) записывается тогда как

$$\pi(n) = P^T \pi(n-1),$$

откуда сразу следует, что

$$\pi(n) = (P^T)^n \pi(0). \quad (18)$$

Таким образом, мы получаем несколько новых свойств. Во-первых, так как матрица P^n равна $P(n) = P(0, n)$, то получается, что P^n – это стохастическая матрица, то есть все ее компоненты неотрицательные, ограничены сверху единицей, а сумма компонент в каждой строке равна 1. Далее, формула (18) позволяет вычислить распределение состояний однородной цепи Маркова в произвольный момент времени n , если известны начальное распределение $\pi(0)$ и матрица перехода за один шаг P . Более того, как мы видели выше, конечномерные распределения цепей Маркова определяются лишь одномерными распределениями и вероятностями перехода. Для однородных цепей это значит, что их конечномерные распределения определяются лишь начальным вектором распределения $\pi(0)$ и матрицей перехода P за один шаг. Оказывается справедлива следующая теорема, доказательство которой можно найти в известной монографии¹⁴.

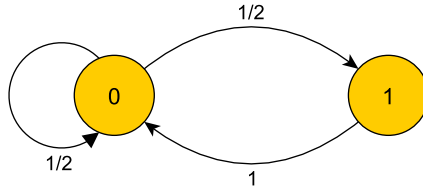
Теорема существования. Пусть заданы конечное или счетное множество S , последовательность π_k и стохастическая матрица P . Тогда существует вероятностное пространство и определенная на нем цепь Маркова X_n с множеством состояний S , начальным распределением $\pi_k(0) = \pi_k$ и матрицей перехода P .*

Дискретные цепи Маркова удобно представлять в виде *стохастического графа*. Это взвешенный ориентированный граф, в вершинах которого расположены состояния цепи, а веса ребер между состояниями равны вероятностям перехода за один шаг между этими состояниями. Например, однородной дискретной цепи Маркова с матрицей переходов

$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix}$$

соответствует стохастический граф

¹⁴Чжун К.-л. Однородные цепи Маркова. Москва : Мир, 1964. С. 21.



8.2. Производящие функции

Приведем некоторые сведения о *производящих* функциях последовательностей. Оказывается, это очень полезный инструмент для исследования цепей Маркова и доказательства некоторых теорем.

Определение. *Производящей функцией* последовательности $\{a_n\}$ (необязательно числовой последовательности) называют формальный ряд

$$\varphi(z) = \sum_{n=0}^{\infty} a_n z^n, \quad z \in \mathbb{C}.$$

Функция

$$\varphi(z) = \frac{1}{1-z} = 1 + z + z^2 + \dots$$

является производящей для последовательности $a_n = \{1, 1, \dots\}$.

Пусть $\varphi(z)$ является производящей функцией последовательности a_n . Тогда производящей функцией последовательности $\{a_{n+1}\}$ является функция

$$\psi(z) = \sum_{n=0}^{\infty} a_{n+1} z^n = \sum_{n=1}^{\infty} a_n z^{n-1} = z^{-1} \left(\sum_{n=0}^{\infty} a_n z^n - a_0 \right) = z^{-1}(\varphi(z) - a_0).$$

Техника производящих функций помогает сформировать один из способов вычисления степеней матрицы перехода P^n и распределений $\pi(n)$ на произвольном шаге. Действительно, пусть $\varphi(z)$ – производящая функция для последовательности $\{P^n\}$. Тогда производящая функция последовательности $\{P^{n+1}\}$ равна

$$\psi(z) = z^{-1}(\varphi(z) - I),$$

но так как $P^{n+1} = P^n P$, то

$$\psi(z) = P\varphi(z) = z^{-1}(\varphi(z) - I),$$

откуда сразу следует

$$\varphi(z) = (I - zP)^{-1}.$$

Так что если удалось функцию справа выразить как ряд $\sum A_n z^n$, то $P^n = A_n$. Отметим, что эта матричнозначная функция $\varphi(z)$ называется *резольвентой* матрицы P . Резольвенты линейных операторов изучаются в функциональном анализе.

Аналогично доказывается, что функция

$$\varphi(z) = (I - zP^T)^{-1}\pi(0)$$

является производящей функцией последовательности

$$\pi(n) = (P^T)^n \pi(0).$$

Поэтому если удастся разложить в ряд

$$(I - zP^T)^{-1} = \sum_{n=0}^{\infty} A_n z^n,$$

то $(P^T)^n = A_n$ и $\pi(n) = A_n \pi(0)$.

8.3. Классификация состояний

Мы ввели понятие *марковского процесса*, это обобщение на стохастический случай динамических систем, то есть систем, в которых будущее процесса (в данном случае распределение состояний в будущем) полностью определяется состоянием процесса в настоящем. Изучение марковских процессов мы начали с частного случая дискретных цепей Маркова – это марковские процессы, у которых время и множество состояний дискретны. Цепи Маркова удобно представлять себе в виде графов, вершинами которых являются состояния, а стрелки обозначают переходы между состояниями. Мы выяснили, что такая система инициализируется с некоторым начальным распределением над состояниями и далее эволюционирует в соответствии с вероятностями перехода из состояния в состояние на каждом шаге. Цепи, у которых вероятности перехода не зависят от номера шага, называются однородными. Мы выяснили, что в таких цепях эволюция матриц перехода и распределения подчиняется следующим линейным рекуррентным уравнениям:

$$P(n) = P \cdot P(n-1), \quad \pi(n) = P^T \pi(n-1).$$

Заметим, что решение этих линейных рекуррентных уравнений (как и решение линейных дифференциальных уравнений) можно описывать в терминах собственных чисел и собственных векторов матрицы перехода за один шаг P . В дальнейшем нас будет в первую очередь интересовать асимптотическое поведение цепей, то есть что происходит с распределением состояний при $n \rightarrow \infty$. Для этого можно построить теорию решения указанных уравнений в терминах собственных чисел и векторов матрицы P , однако этот путь страдает от вычислительной сложности задачи на собственные числа для больших и запутанных цепей, для больших матриц P . Но оказывается, что можно не вычислять никакие собственные числа и векторы, а при одном только взгляде на граф цепи (сколько бы сложным и большим он ни был) сразу сказать, что будет происходить с цепью! Идея состоит в том, чтобы классифицировать должным образом состояния цепи и изучить свойства этих классов.

Всюду далее будем считать цепи Маркова однородными, даже если это не оговорено специально. Вероятность перехода за n шагов из состояния $i \in S$ в состояние $j \in S$ обозначим символом

$$p_{ij}(n) = \mathbb{P}(X_n = j \mid X_0 = i), \quad n \geq 0,$$

где S – множество состояний. Так как цепь однородная, нам не важен момент времени, с которого цепь стартует, поэтому в качестве стартового будем всегда брать момент времени $n = 0$.

Если для состояния i найдется шаг $n \geq 1$ такой, что $p_{ij}(n) > 0$, то будем говорить, что за состоянием i *следует* состояние j . Этот факт мы будем обозначать так: $i \rightarrow j$. Отношение \rightarrow не является рефлексивным (не обязательно $i \rightarrow i$), не является симметричным (не обязательно $i \rightarrow j$ влечет $j \rightarrow i$), но является транзитивным – если $i \rightarrow j$ и $j \rightarrow k$, то $i \rightarrow k$. Если одновременно $i \rightarrow j$ и $j \rightarrow i$, то состояния называются *сообщающимися*. Для сообщающихся состояний i и j мы будем писать: $i \leftrightarrow j$. В развернутом виде это значит, что

$$(\exists m \geq 1 : p_{ij}(m) > 0) \wedge (\exists n \geq 1 : p_{ji}(n) > 0).$$

В противном случае состояния i и j называются *несообщающимися*. Обратим внимание на то, что состояние не обязано сообщаться само с собой. Поэтому отношение \leftrightarrow не является рефлексивным. Но отношение \leftrightarrow является симметричным и транзитивным, то есть если $i \leftrightarrow j$, то $j \leftrightarrow i$, и если $i \leftrightarrow j$, $j \leftrightarrow k$, то $i \leftrightarrow k$. Следовательно, с помощью этого отношения все множество состояний S может быть разбито на непересекающиеся множества, называемые *классами*, так что два состояния

принадлежат одному и тому же классу тогда и только тогда, когда они сообщаются. По определению, состояние, которое не сообщается ни с каким другим состоянием, образует отдельный класс. Таким образом, класс – это либо множество из двух и более сообщающихся между собой состояний, либо одно состояние.

Определение. Состояние i называется *существенным*, если оно сообщается с каждым из следующих за ним состоянием. В противном случае оно называется *несущественным*.

Другими словами, состояние i называется существенным, если для каждого состояния j такого, что $i \rightarrow j$, выполнено $j \rightarrow i$. В развернутом виде это можно записать следующим образом:

$$\forall j \in S ((\forall m \geq 1 p_{ij}(m) = 0) \vee (\exists n \geq 1 p_{ji}(n) > 0)).$$

Для несущественных состояний i найдется состояние j такое, что $i \rightarrow j$, но из j не следует i . В развернутом виде это значит, что

$$\exists j \in S : (\exists m \geq 1 p_{ij}(m) > 0) \wedge (\forall n \geq 1 p_{ji}(n) = 0).$$

Теорема. *За существенным состоянием может следовать только существенное состояние.*

Доказательство. Пусть i существенно и $i \rightarrow j$. Если для какого-то k выполнено $j \rightarrow k$, то и $i \rightarrow k$. Но так как i существенно, то обязательно $k \rightarrow i$ и, следовательно, $k \rightarrow j$. По определению это значит, что состояние j тоже существенно. \square

Из этой теоремы сразу следует важный результат – свойство существенности есть *свойство класса*, то есть если хотя бы одно состояние в классе сообщающихся состояний является существенным, то и все остальные состояния в этом классе являются существенными. Если же есть хотя бы одно несущественное состояние, то и все остальные состояния в данном классе являются несущественными.

Отсюда следует, что все множество состояний S можно разбить на объединение непересекающихся классов, полностью состоящих из существенных или несущественных состояний. Все множество несущественных состояний (которое может быть объединением классов несущественных состояний) мы обозначим символом S_0 . Классы существенных состояний можно пронумеровать и обозначить S_1, S_2 и так далее. Классов несущественных и существенных состояний конечное число в конечных цепях Маркова. В счетных цепях Маркова и классов несущественных состояний, и классов существенных состояний может быть бесконечное число.

Определение. Если в цепи Маркова все состояния сообщаются, то такая цепь называется *неразложимой* или *неприводимой*.

В неразложимой цепи отсутствует множество несущественных состояний и имеется единственный класс существенных состояний. Очень важно заметить, что произвольную цепь Маркова можно представить как состоящую из неразложимых цепей Маркова (эти цепи отвечают классам существенных состояний) и части цепи, отвечающей множеству несущественных состояний. Мы поймем, как эволюционирует со временем исходная цепь, если поймем поведение с течением времени неразложимых цепей Маркова и поведение цепи в множестве несущественных состояний.

Для того чтобы углубить наше понимание происходящего в цепях Маркова, мы введем еще несколько важных понятий. Введем вероятность первого попадания за n шагов из состояния i в состояние j при условии, что цепь вышла из состояния i :

$$f_{ij}(n) = \mathbb{P}(X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j \mid X_0 = i), \quad n \geq 1,$$

а также вероятность $f_i(n) = f_{ii}(n)$ первого возвращения из i в i и вероятность возвращения в i за конечное число шагов:

$$F_i = \sum_{n=1}^{\infty} f_i(n).$$

Заметим, что вероятности $p_{ij}(n)$ и $f_{ij}(n)$ не обязаны совпадать. Вероятность $f_{ij}(n)$ – это вероятность *первого* попадания из i в j при условии, что цепь вышла из состояния i . Вероятность $p_{ij}(n)$ подсчитывается с учетом того, что на промежуточных шагах цепь может бывать в состоянии j , поэтому очевидно, что $p_{ij}(n) \geq f_{ij}(n)$. Можно получить и более точную связь между этими вероятностями:

$$p_{ij}(n) = \sum_{k=1}^n f_{ij}(k) p_{jj}(n-k), \quad n \geq 1. \quad (19)$$

Действительно, сначала заметим, что

$$\begin{aligned} p_{ij}(n) &= \mathbb{P}(X_n = j \mid X_0 = i) = \\ &= \mathbb{P}(X_n = j, X_1 = j \mid X_0 = i) + \\ &+ \mathbb{P}(X_n = j, X_2 = j, X_1 \neq j \mid X_0 = i) + \\ &+ \mathbb{P}(X_n = j, X_3 = j, X_2 \neq j, X_1 \neq j \mid X_0 = i) + \dots + \\ &+ \mathbb{P}(X_n = j, X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j \mid X_0 = i). \end{aligned}$$

Далее,

$$\begin{aligned}
 p_{ij}(n) &= \mathbb{P}(X_n = j \mid X_1 = j, X_0 = i) \mathbb{P}(X_1 = j \mid X_0 = i) + \\
 &+ \mathbb{P}(X_n = j \mid X_2 = j, X_1 \neq j, X_0 = i) \mathbb{P}(X_2 = j, X_1 \neq j \mid X_0 = i) + \\
 &+ \dots = \mathbb{P}(X_n = j \mid X_1 = j) \mathbb{P}(X_1 = j \mid X_0 = i) + \\
 &+ \mathbb{P}(X_n = j \mid X_2 = j) \mathbb{P}(X_2 = j, X_1 \neq j \mid X_0 = i) + \dots + = \\
 &= p_{jj}(n-1) f_{ij}(1) + p_{jj}(n-2) f_{ij}(2) + \dots + p_{jj}(0) f_{ij}(n).
 \end{aligned}$$

Определение. Состояние i называется *возвратным*, если $F_i = 1$, то есть если вероятность возвращения цепи в это состояние за *конечное* число шагов равна единице. В противном случае, то есть если $F_i < 1$, состояние называется *невозвратным*.

Стоит обратить внимание, что «возвратность» состояния означает не то, что цепь в него возвращается с какой-то ненулевой вероятностью, то есть $p_{ii}(n) > 0$ для какого-то n , а то, что цепь возвращается в это состояние *с вероятностью 1 за конечное число шагов*.

Теорема. *Состояние i является возвратным тогда и только тогда, когда*

$$R_i = \sum_{n=1}^{\infty} p_{ii}(n) = \infty.$$

Доказательство. Перепишем формулу (19) для $j = i$ в виде

$$v_n = \sum_{k=1}^n u_k v_{n-k}, \quad n \geq 1, \quad (20)$$

где $v_k = p_{ii}(k)$, $u_k = f_i(k)$, и будем считать, что $v_0 = 1$, $u_0 = 0$. Введем функции

$$V(z) = \sum_{k=0}^{\infty} v_k z^k, \quad U(z) = \sum_{k=0}^{\infty} u_k z^k$$

и заметим, что ряды сходятся при $|z| < 1$, а ряд для $U(z)$ сходится даже для $|z| = 1$. Получаем, что

$$\begin{aligned}
 U(z)V(z) &= \sum_{n=0}^{\infty} \left(\sum_{k=0}^n u_k v_{n-k} \right) z^n = \\
 &= u_0 v_0 z^0 + \sum_{n=1}^{\infty} \left(u_0 v_n + \sum_{k=1}^n u_k v_{n-k} \right) z^n =
 \end{aligned}$$

$$= \sum_{n=1}^{\infty} \left(\sum_{k=1}^n u_k v_{n-k} \right) z^n = \sum_{n=1}^{\infty} v_n z^n = V(z) - 1,$$

откуда

$$V(z) = \frac{1}{1 - U(z)}.$$

Из теории степенных рядов (теорема Абеля) следует, что ряд $\sum_{k=0}^{\infty} a_k$ неотрицательных чисел a_k сходится к $a \leq \infty$ тогда и только тогда, когда существует предел

$$\lim_{z \rightarrow 1-0} \sum_{k=0}^{\infty} a_k z^k = a.$$

Следовательно, если $U(1) = \sum_{k=1}^{\infty} u_k = F_i = 1$, то $\lim_{z \rightarrow 1-0} V(z) = \infty$, поэтому $R_i = \sum_{n=1}^{\infty} p_{ii}(n) = \infty$. С другой стороны,

$$U(z) = \frac{V(z) - 1}{V(z)}.$$

Если $R_i = \infty$, то $V(1-0) = \infty$ и $U(1-0) = 1$, то есть $\sum_{k=1}^{\infty} u_k = F_i = 1$, что по определению означает возвратность состояния i . \square

Определение. Состояние i называется *нулевым*, если $p_{ii}(n) \rightarrow 0$ при $n \rightarrow \infty$. Иначе, то есть если предел последовательности $p_{ii}(n)$ при $n \rightarrow \infty$ не существует или существует, но не равен нулю, состояние i называется *ненулевым*.

Из предыдущей теоремы следует, что если состояние i является невозвратным, то ряд из $p_{ii}(n)$ сходится, и поэтому необходимо $p_{ii}(n) \rightarrow 0$ при $n \rightarrow \infty$. Значит, невозвратные состояния являются нулевыми, а ненулевые состояния – возвратными. Нулевые состояния могут быть как возвратными, так и невозвратными.

Определение. Если $i \rightarrow i$, то наибольший общий делитель множества положительных n , для которых $p_{ii}(n) > 0$, называется *периодом* состояния i и обозначается d_i . Если $d_i = 1$, то такое состояние называют *непериодическим*.

Перейдем к одному из важнейших результатов в теории цепей Маркова.

Теорема (о солидарности).

Пусть состояния i и j сообщаются. Тогда

- 1) *Если i – нулевое, то и j – нулевое;*
- 2) *Если i – возвратное, то и j – возвратное;*
- 3) *Если i имеет период $d \geq 1$, то и j имеет тот же период d .*

Доказательство. Так как эти состояния сообщаются, то

$$\exists M > 0, N > 0 : \alpha = p_{ij}(M) > 0, \beta = p_{ji}(N) > 0.$$

Ранее мы выяснили, что

$$\begin{aligned} p_{ii}(M + n + N) &= \sum_{k \in S} \sum_{l \in S} p_{ik}(M) p_{kl}(n) p_{li}(N) \geq \\ &\geq p_{ij}(M) p_{jj}(n) p_{ji}(N) = \alpha \beta p_{jj}(n). \end{aligned}$$

Аналогично получим

$$p_{jj}(M + n + N) \geq \alpha \beta p_{ii}(n).$$

Теперь пройдемся по пунктам теоремы.

1. Если i – нулевое состояние, то из первого неравенства тривиально следует, что и j – нулевое состояние.

2. Если i – возвратное состояние, то $\sum_{n=1}^{\infty} p_{ii}(n) = \infty$, и тогда

$$\sum_{n=1}^{\infty} p_{jj}(M + N + n) \geq \alpha \beta \sum_{n=1}^{\infty} p_{ii}(n) = \infty,$$

что означает, что и j – возвратное состояние.

3. Введем в рассмотрение множества чисел

$$W_i = \{n > 0 : p_{ii}(n) > 0\}, \quad W_j = \{n > 0 : p_{jj}(n) > 0\}$$

и их наибольшие общие делители d_i и d_j соответственно. Пусть $n \in W_i$, тогда

$$p_{jj}(M + N + n) \geq \alpha \beta p_{ii}(n) > 0,$$

значит, $M + N + n \in W_j$ и делится на d_j . Но для такого n будет $p_{ii}(2n) > 0$ и

$$p_{jj}(M + N + 2n) > 0,$$

откуда $M + N + 2n \in W_j$, и тоже делится на d_j . Следовательно, n делится на d_j , причем $n \in W_i$ выбран произвольно. Получается, что $d_j \leq d_i$. Поменяв i и j местами и записав аналогичные неравенства, получаем в итоге, что $d_i = d_j$. \square

Следствие. В неразложимых цепях Маркова

- 1) если одно из состояний нулевое, то все состояния нулевые,
- 2) если одно из состояний возвратное, то все состояния возвратные,

- 3) если одно из состояний имеет период $d \geq 1$, то все состояния имеют тот же самый период d .

Кроме того, без доказательства приведем следующие теоремы. Доказательство первой теоремы можно найти в классической монографии¹⁵. Доказать вторую теорему предлагается читателю.

Теорема*. *Несущественное состояние всегда является невозвратным и, следовательно, нулевым. Возвратное состояние – всегда существенное.*

Теорема*. *В конечных неразложимых цепях Маркова все состояния являются ненулевыми и, следовательно, возвратными.*

Из этих теорем следует, что в конечных цепях Маркова состояние является существенным тогда и только тогда, когда оно является возвратным и тогда и только тогда, когда оно является ненулевым. Действительно, у любой цепи множество состояний делится на конечные множества несущественных состояний и классы существенных состояний. Несущественные состояния невозвратны и нулевые. Существенные состояния принадлежат классам существенных состояний и в конечных цепях представляют собой неразложимые цепи Маркова, все состояния которых ненулевые и поэтому возвратные. Легко также доказать, что в конечных цепях всегда найдется существенное состояние.

Итак, резюмируем связи между различными свойствами состояний.

В конечных и счетных цепях:

1. Несущественные состояния являются невозвратными.
2. Невозвратные состояния являются нулевыми.
3. Ненулевые состояния являются возвратными.

В конечных цепях:

1. Существенность равносильна возвратности.
2. Возвратность равносильна «ненулевости».
3. Всегда существует существенное состояние.

При этом известны примеры счетных цепей, в которых нет существенных состояний, а также примеры счетных цепей, в которых все состояния одновременно возвратные и нулевые.

8.4. Эргодические дискретные цепи Маркова

Всюду далее речь идет про однородные цепи Маркова.

Определение. Распределение состояний π^0 называется *стационарным*.

¹⁵Чжун К.-л. Однородные цепи Маркова. Москва : Мир, 1964. С. 39.

онарным, если

$$P^T \pi^0 = \pi^0,$$

где P – матрица перехода за один шаг.

В этом определении важно, что π^0 – это не просто какой-то вектор (для конечных цепей) или последовательность (для счетных цепей), удовлетворяющие уравнению $P^T \pi = \pi$, а именно распределение, то есть $\pi_i^0 \geq 0$, $i \in S$, и $\sum \pi_i^0 = 1$. Из этого определения следует, что $(P^T)^n \pi^0 = \pi^0$ для любого $n \geq 0$. Это значит, что если в какой-то момент времени распределение состояний является стационарным распределением π^0 , то оно остается таковым и равным π^0 и во все последующие моменты времени.

Как было сказано в разделе 8.1, вероятностные характеристики однородных цепей Маркова однозначно определяются двумя вещами: начальным распределением состояний и матрицей перехода за один шаг. Цепи Маркова с одинаковыми матрицами перехода, но различными начальными распределениями – это разные цепи. Но иногда нас интересует поведение цепей с одними и теми же матрицами перехода, но отличающиеся начальным распределением. Когда возникают такие вопросы, мы условимся не различать цепи Маркова с различными начальными распределениями, но одинаковыми вероятностями перехода. На практике, когда говорят о том, что дана некоторая цепь Маркова, часто подразумевают, что дана ее матрица переходов, а начальное распределение или неизвестно исследователю, или оно определено неоднозначно (лежит в каком-то множестве распределений), или может свободно выбираться исследователем.

Мы приступаем к изучению асимптотических свойств цепей Маркова, а именно – к изучению явления эргодичности цепей Маркова.

Определение. Дискретная цепь Маркова называется *эргодической*, если для любых $i, j \in S$ существуют положительные пределы

$$\lim_{n \rightarrow \infty} p_{ij}(n) = p_j > 0,$$

не зависящие от i .

Эргодическая теорема для конечных цепей Маркова. *Для того чтобы конечная дискретная цепь Маркова была эргодической, необходимо и достаточно, чтобы она была неразложимой и неперiodической.*

Доказательство.

Необходимость. Пусть конечная цепь Маркова эргодична, то есть для всех $i, j \in S$ существуют положительные пределы $p_{ij}(n) \rightarrow \rightarrow p_j > 0$, не зависящие от i . Это значит, что для каждой пары $i, j \in S$ и

любого $\varepsilon > 0$ найдется шаг $n_0(i, j)$ такой, что для всех $n \geq n_0(i, j)$ получается $p_{ij}(n) > p_j - \varepsilon$. Значит, для $\varepsilon < \min_j p_j$ и $n > n_0 = \max_{i,j} n_0(i, j)$ мы получим, что для всех $i, j \in S$ вероятности $p_{ij}(n) > 0$, $p_{ji}(n) > 0$, $p_{ii}(n) > 0$, то есть все состояния сообщаются, непериодичны и образуют один класс сообщающихся состояний. Другими словами, так как пределы $p_{ij}(n)$ положительны, а число состояний конечно, то, начиная с некоторого номера и далее, все эти числа будут отделены от нуля положительным числом, это влечет свойства сообщаемости и непериодичности. Необходимое условие в теореме доказано.

Достаточность. Теперь предположим, что дана неразложимая непериодическая цепь Маркова, и докажем, что она является эргодической. Сначала докажем, что в этом случае начиная с некоторого номера $p_{ij}(n)$ положительны.

Так как все состояния сообщаются и являются непериодическими, то по определению для каждого состояния $i \in S$

$$\text{НОД}\{n > 0 : p_{ii}(n) > 0\} = 1.$$

Отсюда следует, что найдется некоторое число $r \geq 2$ взаимно простых чисел n_1, \dots, n_r таких, что $p_{ii}(n_k) > 0$, $k = 1, \dots, r$. Перенести результат о положительности вероятностей перехода $p_{ii}(n)$ на случай произвольных достаточно больших n поможет теорема из теории чисел, которая приводится здесь без доказательства.

Лемма*. Для любого конечного набора взаимно простых чисел $\{n_i\}_{i=1}^r$, $r \geq 2$, существует такое n_0 , что для любого $n \geq n_0$ существует совокупность r неотрицательных целых чисел $\{k_i\}_{i=1}^r$, для которых

$$n = \sum_{i=1}^r k_i n_i.$$

Согласно лемме существует $n_0(i)$ такое, что для любого $n \geq n_0(i)$ существует набор r неотрицательных чисел $\{k_j\}_{j=1}^r$ такой, что

$$n = \sum_{j=1}^r k_j n_j.$$

Отсюда, для $n \geq n_0(i)$

$$p_{ii}(n) \geq \prod_{j=1}^r p_{ii}(k_j n_j) \geq \prod_{j=1}^r (p_{ii}(n_j))^{k_j} > 0.$$

Таким образом, вероятности $p_{ii}(n)$ положительны не только для взаимно простых чисел n_1, \dots, n_r , но и вообще, начиная с некоторого номера,

который зависит от состояния i . Так как состояния сообщаются, то для любых $i, j \in S$ найдется шаг $l = l(i, j)$ такой, что $p_{ij}(l) > 0$. Тогда для $n \geq n_0(i) + l(i, j)$ получается

$$p_{ij}(n) \geq p_{ii}(n-l)p_{ij}(l) > 0.$$

Таким образом, начиная с номера шага $N_0 = \max_{i,j}(n_0(i) + l(i, j))$ для всех состояний $i, j \in S$ будет $p_{ij}(n) > 0$.

Перейдем теперь к доказательству того, что $p_{ij}(n)$ сходятся к положительным числам, не зависящим от i , при $n \rightarrow \infty$. Обозначим

$$m_j(n) = \min_i p_{ij}(n), \quad M_j(n) = \max_i p_{ij}(n).$$

Поскольку

$$p_{ij}(n+1) = \sum_{l \in S} p_{il} p_{lj}(n),$$

то

$$\begin{aligned} m_j(n+1) &= \min_i p_{ij}(n+1) = \min_i \sum_{l \in S} p_{il} p_{lj}(n) \geq \\ &\geq \min_i \sum_{l \in S} p_{il} \min_l p_{lj}(n) = m_j(n), \end{aligned}$$

откуда $m_j(n) \leq m_j(n+1)$. Аналогично получаем $M_j(n) \geq M_j(n+1)$. Это значит, что последовательность $m_j(n)$ монотонно не убывает, а последовательность $M_j(n)$ монотонно не возрастает. Поэтому для доказательства сходимости $p_{ij}(n) \rightarrow p_j$ достаточно доказать, что

$$M_j(n) - m_j(n) \rightarrow 0, \quad n \rightarrow \infty, \quad j \in S.$$

Пусть $\delta = \min_{i,j} p_{ij}(N_0)$, где N_0 — это номер шага, начиная с которого все $p_{ij}(n)$ оказываются положительными. Тогда

$$\begin{aligned} p_{ij}(N_0 + n) &= \sum_{l \in S} p_{il}(N_0) p_{lj}(n) = \\ &= \sum_{l \in S} (p_{il}(N_0) - \delta p_{jl}(n)) p_{lj}(n) + \delta \sum_{l \in S} p_{jl}(n) p_{lj}(n), \end{aligned}$$

то есть

$$p_{ij}(N_0 + n) = \sum_{l \in S} (p_{il}(N_0) - \delta p_{jl}(n)) p_{lj}(n) + \delta p_{jj}(2n).$$

Заметим, что $p_{il}(N_0) - \delta p_{jl}(n) \geq \delta - \delta p_{jl}(n) \geq 0$, поэтому

$$\begin{aligned} p_{ij}(N_0 + n) &\geq m_j(n) \sum_{l \in S} (p_{il}(N_0) - \delta p_{jl}(n)) + \delta p_{jj}(2n) = \\ &= m_j(n)(1 - \delta) + \delta p_{jj}(2n). \end{aligned}$$

Отсюда следует, что

$$m_j(N_0 + n) \geq m_j(n)(1 - \delta) + \delta p_{jj}(2n).$$

Аналогично получаем, что

$$M_j(N_0 + n) \leq M_j(n)(1 - \delta) + \delta p_{jj}(2n).$$

Объединяя эти неравенства, получаем

$$M_j(N_0 + n) - m_j(N_0 + n) \leq (M_j(n) - m_j(n))(1 - \delta)$$

и, следовательно,

$$M_j(kN_0 + n) - m_j(kN_0 + n) \leq (M_j(n) - m_j(n))(1 - \delta)^k \rightarrow 0, \quad k \rightarrow \infty.$$

Итак, по некоторой подпоследовательности $\{n_k\}$ $M_j(n_k) - m_j(n_k) \rightarrow 0$ при $n_k \rightarrow \infty$. Но разность $M_j(n) - m_j(n)$ монотонна по n , а значит, $M_j(n) - m_j(n) \rightarrow 0$ при $n \rightarrow \infty$. Отсюда следует, что $p_{ij}(n) \rightarrow p_j$, причем $p_j > 0$, так как все $p_{ij}(n)$ начиная с некоторого шага отделены от нуля и, стало быть, отделены от нуля $m_j(n)$. Достаточное условие теоремы доказано. \square

Теорема о предельном распределении. В конечных эргодических цепях

1) существуют постоянные $C > 0$ и $\rho \in (0, 1)$, такие, что

$$|p_{ij}(n) - p_j| \leq C\rho^n,$$

- 2) числа p_j задают распределение состояний,
- 3) это распределение является стационарным,
- 4) это распределение является единственным стационарным распределением для данной цепи Маркова,
- 5) распределение состояний $\pi_j(n)$ сходится в пределе к p_j .

Доказательство.

Пункт 1. Из полученных оценок вытекает, что

$$|p_{ij}(n) - p_j| \leq M_j(n) - m_j(n) \leq (1 - \delta)^{\lfloor n/N_0 \rfloor - 1},$$

откуда, как легко показать, следует существование $C > 0$ и $\rho \in (0, 1)$ таких, что

$$|p_{ij}(n) - p_j| \leq C\rho^n, \quad n \geq 1.$$

Это значит, что сходимость $p_{ij}(n)$ к предельным значениям p_j происходит с геометрической скоростью.

Пункт 2. Докажем, что p_j образуют распределение на множестве состояний. Для этого достаточно выписать очевидное равенство

$$\sum_{j \in S} p_{ij}(n) = 1$$

и перейти к пределу при $n \rightarrow \infty$. В силу конечности суммы этот предел можно внести под знак суммы, и тогда мы получим, что $\sum_{j \in S} p_j = 1$, причем, как мы выяснили, все $p_j > 0$.

Пункт 3. То, что это распределение является стационарным, прямо следует из равенства

$$p_{ij}(n+1) = \sum_{l \in S} p_{il}(n)p_{lj}.$$

Если перейти в нем к пределу при $n \rightarrow \infty$ и внести предел под сумму, то получится равенство

$$p_j = \sum_{l \in S} p_l p_{lj},$$

то есть $p = P^T p$, где вектор p , как мы выяснили, задает распределение состояний. По определению p – стационарное распределение состояний.

Пункт 4. Докажем, что других стационарных распределений не существует. Пусть q – стационарное распределение, тогда

$$q = P^T q = \dots = (P^T)^n q.$$

Но $(P^T)^n = P^T(n)$, и, значит,

$$q_j = \sum_{l \in S} p_{lj}(n)q_l \rightarrow \sum_{l \in S} p_j q_l = p_j,$$

так что любое стационарное распределение обязано совпадать с p .

Пункт 5. Наконец, пусть дано какое-либо распределение состояний на начальный момент времени $\pi(0)$. Тогда

$$\pi_j(n) = \sum_{l \in S} p_{lj}(n)\pi_l(0) \rightarrow \sum_{l \in S} p_j \pi_l(0) = p_j, \quad n \rightarrow \infty$$

независимо от конкретных значений $\pi(0)$. \square

Из эргодической теоремы следует, что как бы ни была инициализирована эргодическая цепь в начальный момент времени, то есть какое бы начальное распределение она ни имела, распределение состояний устремится к стационарному распределению p , которое единственно и находится из системы уравнений $P^T p = p$.

Что касается счетных цепей, без доказательства приведем следующую теорему.

Эргодическая теорема для произвольных цепей Маркова*. Для того чтобы цепь Маркова была эргодической, необходимо и достаточно, чтобы она была неразложимой, непериодической и ненулевой. Свойства 2–5, указанные в предыдущей теореме свойства предельного распределения $p_j = \lim_{n \rightarrow \infty} p_{ij}(n)$ в эргодических цепях, справедливы и в счетных цепях (в свойстве 5 под сходимостью понимается поточечная сходимость).

Отметим, что в этом случае свойство 5 следует из теоремы Лебега о мажорируемой последовательности (Lebesgue's Dominated Convergence Theorem), а остальные свойства следуют из леммы Фату¹⁶.

Введенное понятие эргодичности цепи естественно связано с ранее введенным понятием эргодичности как связи между временным средним и средним по пространству. А именно, справедлива следующая теорема, которую мы сформулируем и докажем только для конечных цепей Маркова.

Закон больших чисел. В конечных эргодических цепях Маркова

$$\frac{1}{n+1} \sum_{k=0}^n I(X_k = j) \xrightarrow{\mathbb{P}} p_j, \quad n \rightarrow \infty,$$

где $p_j = \lim_{n \rightarrow \infty} p_{ij}(n)$, где $I(A)$ – индикаторная функция события A . Доказательство. Введем обозначение

$$\nu_j(n) = \frac{1}{n+1} \sum_{k=0}^n I(X_k = j)$$

для частот попаданий в состояние j . Рассмотрим состояния i и j (которые могут и совпадать) и покажем сначала, что для любого фиксированного $\varepsilon > 0$

$$\mathbb{P}(|\nu_j(n) - p_j| > \varepsilon \mid X_0 = i) \rightarrow 0, \quad n \rightarrow \infty.$$

¹⁶Ширяев А.Н. Вероятность. В 2-х кн. 6-е изд., испр. Москва : МЦНМО, 2017. Кн. 2. С. 871–872.

Из неравенства Чебышева следует, что

$$\mathbb{P}(|\nu_j(n) - p_j| > \varepsilon | X_0 = i) \leq \frac{\mathbb{E}(|\nu_j(n) - p_j|^2 | X_0 = i)}{\varepsilon^2},$$

поэтому достаточно доказать, что $\mathbb{E}(|\nu_j(n) - p_j|^2 | X_0 = i) \rightarrow 0$ при $n \rightarrow \infty$. Итак,

$$\begin{aligned} \mathbb{E}(|\nu_j(n) - p_j|^2 | X_0 = i) &= \frac{1}{(n+1)^2} \mathbb{E} \left(\left[\sum_{k=0}^n (I(X_k = j) - p_j) \right]^2 | X_0 = i \right) = \\ &= \frac{1}{(n+1)^2} \sum_{k=0}^n \sum_{l=0}^n m_{ij}^{(k,l)}, \end{aligned}$$

где

$$\begin{aligned} m_{ij}^{(k,l)} &= \mathbb{E}((I(X_k = j) - p_j)(I(X_l = j) - p_j) | X_0 = i) = \\ &= \mathbb{E}(I(X_k = j)I(X_l = j) | X_0 = i) - p_j \mathbb{E}(I(X_k = j) | X_0 = i) - \\ &\quad - p_j \mathbb{E}(I(X_l = j) | X_0 = i) + p_j^2 = \\ &= p_{ij}(\min(k,l))p_{jj}(|k-l|) - p_j p_{ij}(k) - p_j p_{ij}(l) + p_j^2. \end{aligned}$$

Для эргодической цепи существуют $C > 0$ и $\rho \in (0, 1)$ такие, что для любых i, j, n справедливо неравенство

$$|p_{ij}(n) - p_j| \leq C\rho^n,$$

поэтому можно записать, что $p_{ij}(n) = p_j + \varepsilon_{ij}(n)$ и $|\varepsilon_{ij}(n)| \leq C\rho^n$. Тогда

$$\begin{aligned} m_{ij}^{(k,l)} &= \\ &= (p_j + \varepsilon_{ij}(\min(k,l)))(p_j + \varepsilon_{ij}(|k-l|)) - p_j(p_j + \varepsilon_{ij}(k)) - p_j(p_j + \varepsilon_{ij}(l)) + p_j^2 = \\ &= p_j \varepsilon_{ij}(\min(k,l)) + p_j \varepsilon_{ij}(|k-l|) - p_j \varepsilon_{ij}(k) - p_j \varepsilon_{ij}(l) + \varepsilon_{ij}(\min(k,l))\varepsilon_{ij}(|k-l|), \end{aligned}$$

откуда следует

$$|m_{ij}^{(k,l)}| \leq C(\rho^{\min(k,l)} + \rho^{|k-l|} + \rho^k + \rho^l + \rho^{\max(k,l)}).$$

Так как степень $\max(k, l)$ – максимальная из всех участвующих в сумме степеней, а $\rho < 1$, то найдется постоянная $C_1 > 0$ такая, что

$$|m_{ij}^{(k,l)}| \leq C_1(\rho^{\min(k,l)} + \rho^{|k-l|} + \rho^k + \rho^l).$$

Отсюда следует, что

$$\frac{1}{(n+1)^2} \sum_{k=0}^n \sum_{l=0}^n |m_{ij}^{(k,l)}| \leq \frac{1}{(n+1)^2} \sum_{k=0}^n \sum_{l=0}^n \left(\rho^{\min(k,l)} + \rho^{|k-l|} + \rho^k + \rho^l \right).$$

Далее воспользуемся очевидным неравенством

$$\sum_{i=0}^n \rho^i \leq \frac{1}{1-\rho}.$$

Во-первых, заметим, что

$$\sum_{k=0}^n \sum_{l=0}^n \rho^k \leq \frac{n+1}{1-\rho}, \quad \sum_{k=0}^n \sum_{l=0}^n \rho^l \leq \frac{n+1}{1-\rho}.$$

Во-вторых,

$$\sum_{k=0}^n \sum_{l=0}^n \rho^{\min(k,l)} = \sum_{k=0}^n \sum_{l=0}^k \rho^l + \sum_{k=0}^n \sum_{l=k+1}^n \rho^k.$$

Первое слагаемое оценивается сверху как $(n+1)/(1-\rho)$. Второе слагаемое равно

$$\sum_{k=0}^n (n-k)\rho^k = n \sum_{k=0}^n \rho^k - \sum_{k=0}^n k\rho^k \leq \frac{n}{1-\rho} \leq \frac{n+1}{1-\rho},$$

поэтому

$$\sum_{k=0}^n \sum_{l=0}^n \rho^{\min(k,l)} \leq \frac{2(n+1)}{1-\rho}.$$

Наконец, в-третьих,

$$\sum_{k=0}^n \sum_{l=0}^n \rho^{|k-l|} = \sum_{k=0}^n \sum_{l=0}^k \rho^{k-l} + \sum_{k=0}^n \sum_{l=k+1}^n \rho^{l-k}.$$

Первое слагаемое равно

$$\sum_{k=0}^n (\rho^k + \rho^{k-1} + \dots + \rho^0) \leq \frac{n+1}{1-\rho}.$$

Второе слагаемое равно

$$\sum_{k=0}^n \sum_{l=1}^{n-k} \rho^l \leq \sum_{k=0}^n \left(\frac{1}{1-\rho} - 1 \right) = \frac{n+1}{1-\rho}.$$

Итак, получается, что

$$\sum_{k=0}^n \sum_{l=0}^n \rho^{|k-l|} \leq \frac{2(n+1)}{1-\rho}.$$

Следовательно,

$$\frac{1}{(n+1)^2} \sum_{k=0}^n \sum_{l=0}^n |m_{ij}^{(k,l)}| \leq \frac{C_1}{(n+1)^2} \frac{6(n+1)}{1-\rho} \rightarrow 0, \quad n \rightarrow \infty.$$

Это значит, что $\mathbb{E}(|\nu_j(n) - p_j|^2 | X_0 = i) \rightarrow 0$ при $n \rightarrow \infty$ и, стало быть,

$$\mathbb{P}(|\nu_j(n) - p_j| > \varepsilon | X_0 = i) \rightarrow 0, \quad n \rightarrow \infty.$$

Из формулы полной вероятности следует, что тогда и

$$\mathbb{P}(|\nu_j(n) - p_j| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty,$$

что и требовалось доказать. \square

9. Непрерывные цепи Маркова

Мы продолжаем изучать марковские процессы. До этого речь шла о марковских процессах с дискретным множеством состояний и дискретным временем (дискретными цепями Маркова). Сегодня начнем изучать марковские процессы с дискретным множеством состояний, но непрерывным временем (непрерывные цепи Маркова). Мы уже сталкивались с примером непрерывной цепи Маркова – это пуассоновский процесс. Непрерывные цепи Маркова широко применяются для моделирования так называемых *систем массового обслуживания*, то есть систем, на вход которых поступают разного рода заявки, и которые эти заявки обрабатывают. Наглядным примером могут служить службы поддержки какой-то фирмы (тогда под заявкой понимается звонок в службу поддержки, который через определенное время обрабатывается). Системы массового обслуживания не обязательно связаны с людьми и службами сервиса, это могут быть сложные электрические цепи с множеством компонентов, которые обрабатывают поступающие на них сигналы. Непрерывные марковские процессы также находят себя в эволюционной биологии (процессом может быть численность какого-то вида при определенных условиях) и кинетике химических реакций (процессом может быть число образующихся в результате реакций молекул вещества).

Как и в случае дискретных цепей Маркова, дадим эквивалентное, но упрощенное определение для этого типа процессов. В качестве множества, на котором определено время, примем $T = [0, \infty)$.

Определение. Случайная функция $X(t)$, $t \geq 0$, в каждый момент времени принимающая значения из множества

$$S \subseteq \mathbb{Z}, \quad |S| \leq \infty,$$

и обладающая свойством

$$\begin{aligned} \mathbb{P}(X(t_n) = x_n \mid X(t_{n-1}) = x_{n-1}, \dots, X(t_0) = x_0) = \\ = \mathbb{P}(X(t_n) = x_n \mid X(t_{n-1}) = x_{n-1}) \end{aligned}$$

для каждого $n \geq 1$, $t_0 < t_1 < \dots < t_n$ и x_0, \dots, x_n , для которых указанные условные вероятности определены, называется *непрерывной цепью Маркова*.

Условные вероятности вида

$$p_{ij}(t_1, t_2) = \mathbb{P}(X(t_2) = j \mid X(t_1) = i)$$

для $t_2 \geq t_1$ называются *переходными вероятностями*. Числа $\|p_{ij}(t_1, t_2)\|$ формируют *переходную матрицу* $P(t_1, t_2)$, которая зависит от непрерывных моментов времени t_1 и t_2 .

Матрица перехода непрерывной цепи, как и матрица перехода дискретной цепи, обладает следующими очевидными свойствами:

- 1) $\sum_{j \in S} p_{ij}(s, t) = 1$ для всех $i \in S$ и $0 \leq s \leq t$,
- 2) $p_{ii}(t, t) = 1$ для всех $i \in S$ и $t \geq 0$,
- 3) $p_{ij}(t, t) = 0$ для всех $i \neq j$ и $t \geq 0$.

Пусть $t_1 < t < t_2$, тогда, используя формулу полной вероятности, в силу марковости процесса $X(t)$ для любых $i, j \in S$ получим

$$\begin{aligned} p_{ij}(t_1, t_2) &= \mathbb{P}(X(t_2) = j \mid X(t_1) = i) = \\ &= \sum_{k \in S} \mathbb{P}(X(t_2) = j \mid X(t) = k, X(t_1) = i) \mathbb{P}(X(t) = k \mid X(t_1) = i) = \\ &= \sum_{k \in S} p_{ik}(t_1, t) p_{kj}(t, t_2). \end{aligned}$$

Это значит, что для любых $t_1 < t < t_2$ справедливо равенство

$$P(t_1, t_2) = P(t_1, t)P(t, t_2),$$

которое называется *уравнением Колмогорова–Чепмена* для непрерывной цепи Маркова.

Вероятностью i -го состояния цепи $X(t)$ в момент $t \geq 0$ называется величина

$$\pi_i(t) = \mathbb{P}(X(t) = i).$$

Очевидно, что

$$\pi_i(t) \geq 0, \quad \sum_{i \in S} \pi_i(t) = 1$$

для любого момента $t \geq 0$. Набор вероятностей $\pi(t) = \{\pi_i(t)\}$ называется *распределением вероятностей состояний*.

Распределение $\pi(t)$ удовлетворяет уравнению

$$\pi(t) = P^T(s, t)\pi(s), \quad 0 \leq s \leq t,$$

что, как и в случае дискретных цепей Маркова, является прямым следствием применения формулы полной вероятности и марковского свойства.

Определение. Непрерывная цепь Маркова с переходной матрицей P называется *однородной*, если для любых $s, t \geq 0$

$$P(s, s + t) = P(0, t).$$

Для однородных цепей мы введем также обозначение $P(t) = P(0, t)$, а элементы этой матрицы будем обозначать просто $p_{ij}(t)$.

Зададимся теперь вопросом: как развиваются во времени непрерывные цепи Маркова? Эволюция дискретных цепей Маркова представляется на каждом шаге как прыжок из состояния i в состояние j с вероятностью p_{ij} , содержащейся в матрице перехода. В непрерывных цепях Маркова время непрерывно, а пространство состояний дискретно. Значит, ее эволюцию тоже можно представить как прыжки из одних состояний в другие, при этом, попав в некоторое состояние, цепь может пребывать в нем некоторое время. Чтобы конкретнее описать поведение непрерывных цепей Маркова, приходится вводить дополнительные предположения о функциях $p_{ij}(t, s)$ и сосредотачиваться таким образом на каких-то подклассах непрерывных цепей Маркова. Обычно эти предположения касаются свойств непрерывности или дифференцируемости этих функций. Мы же введем ограничение не на функции $p_{ij}(t, s)$, а на траектории непрерывных цепей Маркова.

Определение. Случайный процесс $\{X(t), t \geq 0\}$, определенный на вероятностном пространстве $\{\Omega, \mathcal{F}, \mathbb{P}\}$, называется *непрерывным справа*, если его траектории непрерывны справа для \mathbb{P} -п.н. $\omega \in \Omega$.

Это условие для непрерывной цепи Маркова будет значить, что после того, как цепь совершила прыжок в какое-то состояние, она

находится в этом состоянии еще какое-то время. Иначе может быть так, что в любой сколь угодно малой окрестности времени перехода совершаются прыжки между состояниями, мы будем избегать рассмотрения таких абстрактных, нефизических процессов.

Теорема. Пусть $\{X(t), t \geq 0\}$ – непрерывная справа непрерывная цепь Маркова. Тогда для любого $s \geq 0$ покомпонентно

$$P(s, t) \rightarrow I, \quad t \rightarrow s + 0,$$

где I – единичная матрица. В частности, для однородной цепи

$$P(t) \rightarrow I, \quad t \rightarrow 0 + 0.$$

Доказательство. Пусть некоторая последовательность $\{t_n\}$ сходится справа к точке s . Так как состояния цепи – целые числа, то для любого $\omega \in \Omega$ отображение $t \mapsto X(\omega, t)$ непрерывно в точке s справа тогда и только тогда, когда существует такой n_0 , что любого $n \geq n_0$ выполнено $X_{t_n} = i$ для некоторого состояния i . Обозначим за E множество ω , для которых отображение $t \mapsto X(\omega, t)$ непрерывно в точке s справа.

Предположим, что $\mathbb{P}(X(s) = i) > 0$. Тогда

$$\begin{aligned} 1 &= p_{ii}(s, s) = \mathbb{P}(X(s) = i | X(s) = i) = \\ &= \mathbb{P}(\omega \in \Omega : X(\omega, s) = i | X(\omega, s) = i) = \\ &= \mathbb{P}(\omega \in E : X(\omega, s) = i | X(\omega, s) = i), \end{aligned}$$

так как дополнение множества E до множества Ω имеет нулевую меру. Далее,

$$\begin{aligned} &\mathbb{P}(\omega \in E : X(\omega, s) = i | X(\omega, s) = i) = \\ &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \{\omega \in E : X(\omega, t_k) = i\} | X(\omega, s) = i\right) = \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{k \geq n} \{\omega \in E : X(\omega, t_k) = i\} | X(\omega, s) = i\right). \end{aligned}$$

С другой стороны,

$$\mathbb{P}\left(\bigcap_{k \geq n} \{\omega \in E : X(\omega, t_k) = i\} | X(\omega, s) = i\right) \leq$$

$$\begin{aligned} &\leq \mathbb{P}(\omega \in E : X(\omega, t_n) = i \mid X(\omega, s) = i) = \\ &= \mathbb{P}(\omega \in \Omega : X(\omega, t_n) = i \mid X(\omega, s) = i) \leq 1. \end{aligned}$$

Отсюда следует, что существует предел

$$\lim_{n \rightarrow \infty} \mathbb{P}(X(t_n) = i \mid X(s) = i) = 1,$$

или, другими словами, $p_{ii}(s, t_n) \rightarrow p_{ii}(s, s)$. Так как последовательность t_n произвольная, то $p_{ii}(s, t) \rightarrow 1$, $t \rightarrow s + 0$. Далее, так как для любых i и $j \neq i$ получается

$$p_{ij}(s, t) = \mathbb{P}(X(t) = j \mid X(s) = i) \leq \mathbb{P}(X(t) \neq i \mid X(s) = i) = 1 - p_{ii}(s, t),$$

то получаем $p_{ij}(s, t) \rightarrow 0$ для $i \neq j$ при $t \rightarrow s + 0$. \square

Итак, для непрерывных справа однородных цепей Маркова получаем

$$P(0) = I, \quad P(t + s) = P(t)P(s), \quad P(t) \rightarrow I, \quad t \rightarrow 0.$$

Такое семейство матричнозначных функций называется *стандартной стохастической полугруппой*. Цепи с такими свойствами называют еще *стандартными* непрерывными цепями Маркова.

Теорема. Матрица вероятностей перехода $P(t)$ конечной стандартной цепи Маркова непрерывна в любой точке t .

Доказательство. Действительно, так как $P(h) \rightarrow I$, $h \rightarrow 0 + 0$, то существует окрестность $h = 0$, в которой определитель $P(h)$ отличен от нуля, и поэтому матрица $P(h)$ обратима для каждого h из этой окрестности. Кроме того¹⁷, $P^{-1}(h) \rightarrow I$ при $h \rightarrow 0 + 0$, откуда

$$P(t + h) = P(t)P(h) \rightarrow P(t), \quad h \rightarrow 0 + 0,$$

$$P(t - h) = P(t)P^{-1}(h) \rightarrow P(t), \quad h \rightarrow 0 + 0,$$

таким образом $P(t)$ непрерывна в каждой точке t . \square

Более того, функция $P(t)$ в конечных стандартных цепях Маркова является дифференцируемой в каждой точке $t \geq 0$. Мы оставляем

¹⁷Факт о сходимости обратной матрицы к единичной можно доказать различными способами. Например, обратную матрицу можно представить через миноры и определитель исходной матрицы. Это значит, что обратная матрица непрерывным образом выражается через элементы исходной матрицы. С другой, фундаментальной, точки зрения, в произвольных банаховых пространствах для линейных ограниченных операторов A можно записать оценку $\|(A + \Delta A)^{-1} - A^{-1}\| \leq (\dots) \|\Delta A\|$, если норма матрицы ΔA достаточно мала, см., например, Треногин В.А. Функциональный анализ. Москва : Наука, 1980. С. 141.

этот факт без доказательства, которое можно найти в монографии¹⁸, теоремы 9 и 10, или в монографии¹⁹, теорема С.12, доказательство основывается на свойствах непрерывных полугрупп. Кстати из теоремы 9 первой монографии следует и то, что функция $P(t)$ дифференцируема в нуле даже в счетных стандартных цепях Маркова, если допустить равенство производных $p_{ii}(t)$ равным минус бесконечности. А именно, справедлива следующая теорема.

Теорема. Пусть $P(t)$ – матрица вероятностей перехода стандартной цепи Маркова. Тогда существует матрица*

$$Q = \lim_{h \rightarrow 0+0} \frac{P(h) - I}{h}$$

с компонентами q_{ij} , причем $0 \leq q_{ij} < \infty$, $q_i = -q_{ii} \in [0, \infty]$.

Из этой теоремы следует, что в конечных цепях обязательно $q_i < \infty$, так как $\sum_j p_{ij} = 1$, производная этого выражения равна

$$\sum_j q_{ij} = 0,$$

а q_{ij} для $i \neq j$ конечны. В счетных цепях может быть $q_i = \infty$. Если же в цепи $q_i < \infty$ и $\sum_j q_{ij} = 0$, то ее называют *консервативной*. Получается, что конечные цепи являются консервативными.

Теорема. В конечных стандартных цепях Маркова функция $P(t)$ удовлетворяет системам дифференциальных уравнений

$$\dot{P}(t) = P(t)Q, \quad P(0) = I, \quad (21)$$

$$\dot{P}(t) = QP(t), \quad P(0) = I, \quad (22)$$

которые имеют единственное решение $P(t) = \exp(Qt)$, выражаемое в виде матричной экспоненты, где Q – производная в нуле функции $P(t)$.

Доказательство. Рассмотрим конечную стандартную цепь Маркова. Пусть

$$Q = \lim_{h \rightarrow 0+0} \frac{P(h) - I}{h}.$$

Заметим, что

$$P(t+h) = P(t)P(h) = P(h)P(t).$$

¹⁸Булинский А.В., Ширяев А.Н. Теория случайных процессов. Москва : ФИЗМАТЛИТ, 2005. С. 201–206.

¹⁹Modica G., Poggiolini L. A First Course In Probability and Markov Chains. Wiley, 2013.

Вычитая слева и справа из этих равенств $P(t)$, получим

$$\frac{P(t+h) - P(t)}{h} = P(t) \frac{P(h) - I}{h} = \frac{P(h) - I}{h} P(t),$$

и, устремляя h к нулю, приходим к уравнениям

$$\dot{P}(t) = P(t)Q \text{ и } \dot{P}(t) = QP(t).$$

Вкупе с начальным условием $P(0) = I$ получаются две однородные системы линейных дифференциальных уравнений с постоянными коэффициентами. Из теории дифференциальных уравнений известно, что обе эти системы имеют единственное решение

$$P(t) = \exp(Qt),$$

где под экспонентой понимается матричная экспонента. \square

С некоторыми оговорками эта теорема справедлива и в счетных цепях Маркова²⁰. Что касается распределения состояний, то $\pi(t) = P^T(t)\pi(0)$, откуда после дифференцирования по t получим

$$\dot{\pi}(t) = Q^T \pi(t).$$

Это однородная система линейных дифференциальных уравнений, которая для начального условия $\pi(0)$ имеет решение $P^T(t)\pi(0)$.

Из теоремы выше следует, что переходные вероятности однозначно определяются по независящей от времени матрицы Q , называемой *Q-матрицей*, *матрицей интенсивностей* или *инфинитезимальной матрицей*, а в теории полугрупп операторов ее называют *генератором* полугруппы. По определению, ее элементы удовлетворяют следующим равенствам:

$$q_{ii} = \lim_{h \rightarrow 0+0} \frac{p_{ii}(h) - 1}{h}, \quad q_{ij} = \lim_{h \rightarrow 0+0} \frac{p_{ij}(h)}{h}, \quad j \neq i.$$

Эти равенства равносильны следующим:

$$p_{ii}(h) = 1 + q_{ii}h + o(h), \quad p_{ij}(h) = q_{ij}h + o(h), \quad i \neq j, \quad h \rightarrow 0 + 0.$$

Важно отметить, что хотя эти формулы – разложения вероятностей перехода с точностью всего лишь до первого порядка малости по h , они однозначно определяют вероятности перехода как функции времени,

²⁰Булинский А.В., Ширяев А.Н. Теория случайных процессов. Москва : ФИЗМАТЛИТ, 2005. С. 205–207.

так как вероятности перехода однозначно определяются элементами Q -матрицы. Очень часто эти разложения помогают в решении задач.

Для $i \neq j$ числа q_{ij} называются *интенсивностью перехода* из состояния i в состояние j . Так как $\sum_j p_{ij}(h) = 1$ для любого h , то в *конечной* цепи

$$1 + q_{ii}h + \sum_{j \neq i} q_{ij}h + o(h) = 1,$$

откуда следует

$$\sum_{j \neq i} q_{ij} = -q_{ii} = q_i.$$

Это значит, что сумма элементов каждой строки матрицы Q равна нулю. В счетных же цепях это может быть не так.

Выясним теперь смысл чисел q_{ij} . Пусть $\{X(t), t \geq 0\}$ – консервативная цепь Маркова с непрерывными справа траекториями и конечным или счетным числом состояний. Для каждого состояния $i \in S$ определим функцию исхода

$$\tau_i(\omega) = \inf\{t : X(\omega, t) \neq i\}.$$

Заметим, что если $X(0) = i$, то эта случайная величина есть время пребывания цепи в состоянии i . В силу непрерывности справа траекторий почти наверное

$$\tau_i(\omega) = \inf\{t : X(\omega, t) \neq i\} = \inf\{t \in D : X(\omega, t) \neq i\} \doteq \tau_i^D(\omega),$$

где D – произвольное счетное всюду плотное множество на $[0, \infty)$, например множество рациональных чисел, а равенство посередине понимается в смысле почти наверное. Действительно, так как инфимум для левой части берется по более широкому множеству, то $\tau_i \leq \tau_i^D$. Отсюда следует, что если $\tau_i = +\infty$, то и $\tau_i^D = +\infty$. Если же $\tau_i < +\infty$, то множество D плотно окружает точку τ_i , поэтому в нем найдется последовательность $\{t_n\}$, сходящаяся справа к $\tau_i(\omega) \in [0, +\infty)$, а в силу непрерывности справа процесса начиная с некоторого номера $X(\omega, t_n) \neq i$. Значит, уж точно $\tau_i^D \leq \tau_i$. Значит, $\tau_i(\omega) = \tau_i^D(\omega)$ почти наверное (потому что непрерывность справа – почти наверное). Теперь вместо $\tau_i(\omega)$ будем рассматривать $\tau_i^D(\omega)$, которая является случайной величиной в строгом смысле слова, так как

$$\{\omega : \tau_i^D(\omega) > t\} = \bigcap_{d \in D \cap [0, t]} \{\omega : X(\omega, d) = i\}$$

является измеримым множеством.

Теперь рассмотрим множество $D(t)$:

$$D(t) = \bigcup_{n=0}^{\infty} D_n = \bigcup_{n=0}^{\infty} \left\{ \frac{jt}{2^n} \right\}_{j=0}^{2^n},$$

где $D_n \subset D_{n+1}$. Множество $D(t)$ – это счетное всюду плотное множество на отрезке $[0, t]$. Объединение множеств $D(t2^N)$ при $N \in \mathbb{N} \cup \{0\}$ является счетным всюду плотным множеством на $[0, +\infty)$ и может играть роль множества D в рассуждениях выше. Тогда

$$\begin{aligned} \mathbb{P}(\tau_i^D > t) &= \mathbb{P} \left(\bigcap_{n=0}^{\infty} \bigcap_{j=0}^{2^n} \left\{ X \left(\frac{jt}{2^n} \right) = i \right\} \right) = \\ &= \lim_{n \rightarrow \infty} \mathbb{P} \left(\bigcap_{j=0}^{2^n} \left\{ X \left(\frac{jt}{2^n} \right) = i \right\} \right) = \lim_{n \rightarrow \infty} p_{ii}(t/2^n)^{2^n} \pi_i(0) = \\ &= \lim_{n \rightarrow \infty} (1 + q_{ii}t/2^n + o(1/2^n))^{2^n} \pi_i(0) = e^{q_{ii}t} \pi_i(0). \end{aligned}$$

Если цепь стартует из состояния i , то $\mathbb{P}(X(0) = i) = 1$ и время пребывания в состоянии i имеет показательное распределение интенсивности $-q_{ii} = q_i$, то есть $\tau_i^D \in \text{Exp}(q_i)$.

Введем функцию

$$F_{ij}(h) = \mathbb{P}(X(t+h) = j | X(t) = i, X(t+h) \neq i)$$

для $i \neq j$. Это вероятность попадания цепи в состояние j в момент $t+h$ при условии, что цепь *находится* в состоянии i в момент t и *не находится* в состоянии i в момент $t+h$. По определению условной вероятности при $h \rightarrow 0$

$$\begin{aligned} F_{ij}(h) &= \frac{\mathbb{P}(X(t+h) = j | X(t) = i)}{\mathbb{P}(X(t+h) \neq i | X(t) = i)} = \\ &= \frac{\mathbb{P}(X(t+h) = j | X(t) = i)}{1 - \mathbb{P}(X(t+h) = i | X(t) = i)} \rightarrow \frac{q_{ij}}{q_i}. \end{aligned}$$

Отсюда следует, что числа q_{ij} определяют приоритет одних состояний перед другими в момент прыжка. Можно доказать строго²¹, что q_{ij}/q_i – это вероятность перейти в состояние j из состояния i в момент прыжка. Можно также доказать (см. ту же монографию), что времена пребывания в каждом состоянии независимы в совокупности.

²¹Чжун К.-л. Однородные цепи Маркова. Москва : Мир, 1964.

Итак, поведение цепи можно представлять себе следующим образом. Пусть сначала цепь находится в некотором состоянии $i \in S$. В этом состоянии она находится некоторое время, имеющее показательное распределение $\text{Exp}(q_i)$. По прошествии этого времени наступает момент скачка цепи в другое состояние. В этот момент скачка цепь может попасть в произвольное состояние $j \neq i$ с вероятностью q_{ij}/q_i . После этого перехода цепь находится в состоянии j . В этом состоянии цепь пробудет случайное время, имеющее показательность распределения $\text{Exp}(q_j)$, после чего наступит момент скачка. В этот момент цепь может попасть в произвольное состояние $k \neq j$ с вероятностью q_{jk} и так далее.

9.1. Эргодические непрерывные цепи Маркова

Всюду далее речь идет про однородные цепи Маркова.

Определение. Распределение состояний π^0 называется *стационарным*, если

$$P^T(t)\pi^0 = \pi^0, \quad \forall t \geq 0,$$

где $P(t)$ – матрица перехода.

Так как распределение состояний подчиняется уравнению

$$\dot{\pi}(t) = Q^T \pi,$$

то для стационарного распределения π^0 обязательно

$$Q^T \pi^0 = 0.$$

Это значит, что стационарное распределение непрерывной цепи Маркова является собственным вектором матрицы Q^T для нулевого собственного значения. Уравнение $Q^T \pi^0 = 0$ называется *стационарным уравнением Колмогорова*. В принципе, как видно из определения, стационарное распределение является собственным вектором матриц $P(t)$ для всех $t \geq 0$ для собственного значения 1.

Перенесем на случай непрерывных цепей классификацию состояний, а именно понятия следования одного состояния за другим, понятия сообщаемости, существенности и понятие нулевого состояния. Эти понятия переносятся на случай непрерывных цепей без каких-либо оговорок, нужно только заменить дискретный шаг n на непрерывное время t . Неразложимыми цепями будем так же называть цепи, в которых все состояния сообщаются. Понятие о возвратности нам не потребуется, а понятие периодического состояния для непрерывных цепей не вводится (в дискретном случае периодом был наибольший общий делитель некоторого множества, которое получалось дискретным, потому что

время было дискретным, а наибольший общий делитель для непрерывного множества не определен). Теорема о солидарности переносится и на случай непрерывных цепей, за исключением пункта про периодичность состояний. Действительно, нигде (кроме периодичности) не использовалось то, что время дискретное.

Перенесем и определение эргодической цепи.

Определение. Непрерывная цепь Маркова называется *эргодической*, если для любых $i, j \in S$ существуют положительные пределы

$$\lim_{t \rightarrow \infty} p_{ij}(t) = p_j > 0,$$

не зависящие от i .

Эргодическая теорема для конечных цепей Маркова. *Для того чтобы конечная непрерывная цепь Маркова была эргодической, необходимо и достаточно, чтобы она была неразложимой.*

Доказательство. Необходимость условия показывается точно так же, как для дискретной цепи Маркова, за исключением слов о непериодических состояниях, потому как не опирается на дискретность времени.

Достаточность условия следует из наблюдения, что в неразложимых непрерывных конечных цепях Маркова $p_{ij}(t) > 0$ для любых $i, j \in S$ и $t > 0$. Действительно, так как время жизни в любом состоянии является показательно распределенной случайной величиной и все состояния сообщаются, то из любого состояния можно попасть в любое другое с ненулевой вероятностью, причем за сколь угодно малое время. А раз так, то остается воспользоваться доказательством теоремы для дискретной цепи, заменив всюду дискретный шаг на непрерывное время. \square

Теорема о предельном распределении $\{p_j\}$ тоже легко переносится на случай непрерывной цепи. Опять же, всюду дискретный шаг следует заменить на непрерывное время. То есть предельные вероятности p_j образуют единственное стационарное распределение в конечной непрерывной цепи Маркова.

Наконец, сформулируем без доказательства теорему о связи между введенным нами понятием эргодичности цепи и эргодичностью как связи между временным средним и средним по пространству.

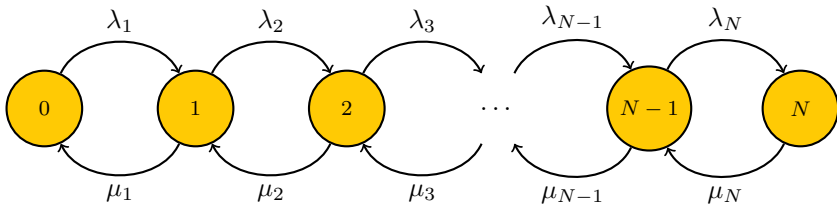
Теорема*. *В конечной эргодической непрерывной цепи Маркова*

$$\frac{1}{T} \int_0^T I(X(t) = j) dt \xrightarrow{\text{п.н.}} p_j, \quad T \rightarrow \infty,$$

где $p_j = \lim_{t \rightarrow \infty} p_{ij}(t)$.

9.2. Процессы гибели и рождения

Определение. Однородная непрерывная цепь Маркова называется *процессом гибели и рождения*, если ее стохастический граф имеет следующий вид:



причем число состояний N может быть как конечным, так и бесконечным.

Можно сказать, что процессы гибели и рождения – это аналоги случайных блужданий, просто переходы между состояниями осуществляются не в дискретном времени (на каждом шаге), а через случайные непрерывные моменты времени (с показательным распределением).

Теорема*. Пусть $X(t)$ – процесс гибели и рождения с конечным числом состояний. Тогда стационарное распределение π^0 существует, единственно и определяется соотношениями

$$\pi_0^0 = \left(1 + \frac{\lambda_1}{\mu_1} + \frac{\lambda_1 \lambda_2}{\mu_1 \mu_2} + \dots + \frac{\lambda_1 \lambda_2 \dots \lambda_N}{\mu_1 \mu_2 \dots \mu_N} \right)^{-1}, \quad \pi_j^0 = \frac{\lambda_j}{\mu_j} \pi_{j-1}^0, \quad j \geq 1.$$

Эти формулы дают точное решение соответствующей системы стационарных уравнений Колмогорова.

Величину $X(t)$ при любом $t \geq 0$ можно интерпретировать как число членов некоторой популяции, где N – максимально возможное число ее членов. Параметры λ_j называют *интенсивностями рождения*, а параметры μ_j – *интенсивностями гибели*. И те и другие могут быть равны нулю. Если все $\mu_j = 0$, то соответствующий процесс называется *процессом рождения* (или чистого рождения). Если все $\lambda_j = 0$, то такой процесс называют *процессом гибели* (или чистой гибели).

9.3. Взрывные марковские цепи

Пусть $T_n, n \geq 1$ – моменты времени, в которые осуществляются переходы между состояниями в непрерывной цепи Маркова. Рассмотрим для каждого исхода равенство

$$T_\infty = \lim_{n \rightarrow \infty} T_n.$$

Определение. Если $\mathbb{P}(T_\infty = \infty) = 1$, то непрерывная цепь Маркова называется *невзрывной*. Иначе она называется *взрывной*.

Проще говоря, во взрывных цепях с ненулевой вероятностью возможны ситуации, когда в цепи осуществляется бесконечное число переходов за конечное время. Такие цепи могут рассматриваться, например, как модель в химической кинетике, ядерных исследованиях, в биологии.

Теорема. Пусть $\{\xi_n\}_{n=1}^N$, $N \leq \infty$ – последовательность независимых случайных величин с показательным распределением, $\xi_n \in \text{Exp}(\lambda_n)$. Тогда

$$\sum_{n=1}^N \frac{1}{\lambda_n} < \infty \Rightarrow \mathbb{P}\left(\sum_{n=1}^N \xi_n < \infty\right) = 1.$$

$$\sum_{n=1}^N \frac{1}{\lambda_n} = \infty \Rightarrow \mathbb{P}\left(\sum_{n=1}^N \xi_n = \infty\right) = 1.$$

Доказательство. Пусть указанная сумма сходится, тогда

$$\mathbb{E} \sum_{n=1}^N \xi_n = \sum_{n=1}^N \frac{1}{\lambda_n} < \infty,$$

и поэтому $\sum_n \xi_n < \infty$ п. н. Теперь пусть указанная сумма расходится (это возможно только, если $N = \infty$). Тогда

$$\mathbb{E} \exp\left(-\sum_{n=1}^{\infty} \xi_n\right) = \prod_{n=1}^{\infty} \mathbb{E} \exp(-\xi_n) = \prod_{n=1}^{\infty} \left(1 + \frac{1}{\lambda_n}\right)^{-1} = 0.$$

Последнее равенство следует из расходимости ряда $\sum_n \ln(1 + 1/\lambda_n)$, чья расходимость равносильна расходимости ряда $\sum_n 1/\lambda_n$. Отсюда следует, что $\sum_n \xi_n = \infty$ п. н. \square

Следствие. Процесс рождения с интенсивностями рождения λ_j является взрывным тогда и только тогда, когда

$$\sum_{j=1}^N \frac{1}{\lambda_j} < \infty.$$

Заметим, что пуассоновский процесс $K(t)$ интенсивности $\lambda > 0$ является однородной непрерывной цепью Маркова, причем он является процессом рождения с одинаковыми интенсивностями рождения $\lambda_n = \lambda$. В этом случае ряд из теоремы расходится, поэтому процесс Пуассона невзрывной, то есть он не может устремиться к бесконечности за конечное время.

9.4. Потоки событий

Пуассоновский процесс и вообще непрерывные цепи Маркова являются ключевыми математическими моделями случайных последовательностей (или *потоков*) событий.

Определение. *Потоком событий* называется последовательность одинаковых событий, происходящих одно за другим через промежутки времени случайной длины.

Пусть здесь и далее $t \geq 0$, $h > 0$. Обозначим через $n(t_1, t_2)$ случайную величину, равную числу событий из потока, происходящих на промежутке $[t_1, t_2)$. Введем некоторые естественные предположения о потоке событий и изучим его свойства.

Определение. Поток называется *однородным* или *стационарным*, если законы распределения случайных величин $n(t_1, t_2)$ и $n(t_1 + s, t_2 + s)$ совпадают при всех $s \geq 0$. Для таких потоков введем обозначение $n(t) = n(0, t)$.

Пусть выбраны произвольные точки $0 = t_0 \leq t_1 \leq \dots \leq t_{m+1}$. Рассмотрим совокупность случайных величин

$$\xi_k = n(t_k, t_{k+1}), \quad k = 0, 1, \dots, m.$$

Определение. Поток событий называется *потоком без последствия*, если случайные величины $\{\xi_0, \dots, \xi_m\}$ независимы в совокупности.

Предположим, наконец, что между событиями в потоке обязательно проходит какое-то конечное время, то есть что после того, как какое-то событие произойдет, проходит какое-то время, когда не происходит никаких событий, а потом, и притом обязательно, происходит новое событие.

Зададимся теперь вопросом о том, что же за процесс $n(t)$ мы только что получили. Так как поток является потоком без последствия, то $n(t)$ – это процесс с независимыми приращениями, а значит, он является марковским процессом. Он определен на непрерывном множестве времени, а его состояния – целые числа, поэтому $n(t)$ – непрерывная цепь Маркова. Из однородности потока следует однородность цепи. Так как между событиями проходит какое-то конечное время, траектории этой цепи непрерывны справа, а значит, применима теория непрерывных справа цепей Маркова, разработанная нами ранее. В частности это значит, что эта цепь Маркова является стандартной, а поэтому вероятности перехода всюду непрерывны. Траектории $n(t)$ не убывают, значит, $p_{ij}(t) \neq 0$ может быть только для $j \geq i$. Из того, что цепь стандартна, следует, что существуют конечные производные

функций $p_{ij}(t)$ в нуле для $j > i$, называемые интенсивностями перехода $0 \leq \lambda_{ij} = \dot{p}_{ij}(0) < \infty$. Более того, существуют и (возможно бесконечные или равные нулю) производные $p_{ii}(t)$ в нуле, то есть λ_{ii} . На самом же деле λ_{ii} в нашем случае не могут быть равны нулю, так как тогда для любого $t \geq 0$

$$p_{ii}(t) = p_{ii}(t/n)^n = (1 + o(t/n))^n \rightarrow 1, \quad n \rightarrow \infty,$$

что будет означать $p_{ii}(t) = 1$. Но мы предположили, что после одних событий обязательно рано или поздно следуют другие события, поэтому $\lambda_{ii} = 0$ не возможно. Предположим дополнительно, что эта производная не равна и бесконечности. В таком случае цепь находится в состоянии i случайное время с показательным распределением $\text{Exp}(\lambda_i)$, где $\lambda_i = -\lambda_{ii}$, причем для произвольного $h > 0$

$$p_{ii}(t) = \mathbb{P}(n(t+h) = i | n(h) = i) = \mathbb{P}(n(h, t+h) = 0)$$

от i не зависит, поэтому и λ_{ii} не зависит от i . Обозначим за $\lambda = \lambda_{i,i+1}$. В момент скачка из состояния i по построению потока с вероятностью 1 цепь переходит в состояние $i+1$. Поэтому $\lambda_{i,i+1} = \lambda$ и $\lambda_{ij} = 0$ для $j > i+1$.

Итак, мы получили марковский процесс, исходящий из нуля, с независимыми приращениями, у которого время между скачками имеет показательное распределение с параметром $\lambda > 0$, а скачки между состояниями происходят по схеме $0 \rightarrow 1 \rightarrow 2 \rightarrow \dots$. Получается, что это процесс восстановления, в котором случайные величины независимы и показательно распределены (утверждение о независимости времен пребывания в каждом состоянии следует из марковского свойства процесса, доказательство этого факта мы опускаем, но его можно найти в монографии²²). Такой процесс, как мы выяснили в разделе 2, является пуассоновским процессом.

Заметим, что в выкладках выше нам понадобилось, чтобы между событиями проходило какое-то время и чтобы $0 < \lambda_i < \infty$. Из этого следуют стандартность цепи Маркова и факты о временах пребывания в состояниях и вероятностях скачков в другие состояния. Если в потоке происходит конечное, пусть и большое и сколь угодно большое, число событий, то цепь конечна и тогда требование $\lambda_i < \infty$ больше не нужно, оно выполнится автоматически. Достаточно будет того, что цепь непрерывна справа, или, что эквивалентно, того, что между событиями проходит всегда конечное время. Единственная оговорка –

²²Чжун К.-л. Однородные цепи Маркова. Москва : Мир, 1964.

время пребывания в последнем состоянии, когда произойдет последнее событие, будет бесконечным, интенсивность будет равна нулю.

Требования на то, что между событиями проходит конечное время и что интенсивность – конечное ненулевое число, можно было бы заменить на требования

$$\mathbb{P}(n(t, t+h) = 1) = \lambda h + o(h), \quad h \rightarrow 0,$$

$$\mathbb{P}(n(t, t+h) = 0) = 1 - \lambda h + o(h), \quad h \rightarrow 0,$$

где $0 < \lambda < \infty$. Потоки с этими двумя свойствами называются *ординарными* («обыкновенными»). Получается, что $n(0, t)$ в однородном ординарном потоке без последействия есть пуассоновский случайный процесс. И вообще, однородный ординарный поток без последействия обычно называют *простейшим пуассоновским потоком событий*.

Конечно, бывают и другие потоки, которые необязательно приводят к пуассоновскому процессу. Например, потоки необязательно бывают однородными. В этом случае интенсивность определяют как предел

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(n(t, t+h) > 0)}{h},$$

и тогда ординарность определяют в виде свойств

$$\mathbb{P}(n(t, t+h) = 1) = \lambda(t)h + o(h), \quad h \rightarrow 0,$$

$$\mathbb{P}(n(t, t+h) = 0) = 1 - \lambda(t)h + o(h), \quad h \rightarrow 0.$$

9.5. Эквивалентные определения пуассоновского процесса

Резюмируем все эквивалентные подходы к определению или заданию процесса Пуассона. Вспомним наше основное определение процесса Пуассона.

Определение 1. *Пуассоновским процессом с интенсивностью $\lambda > 0$ называется случайная функция $\{K(t), t \geq 0\}$, удовлетворяющая свойствам:*

- 1) $K(0) = 0$ почти всюду;
- 2) $K(t)$ – процесс с независимыми приращениями;
- 3) $\forall t > s \geq 0$ выполнено $K(t) - K(s) \in \text{Po}(\lambda(t-s))$.

В разделе 2 мы выяснили, что процесс восстановления с показательными случайными величинами является пуассоновским процессом. Поэтому вместо теоремы о явной конструкции процесса Пуассона можно было бы принять следующее определение.

Определение 2. Пусть $\{\xi_n\}$ – последовательность независимых случайных величин с распределением $\text{Exp}(\lambda)$. Обозначим $S_n = \xi_1 + \dots + \xi_n$ ($S_0 = 0$). Введем процесс

$$X(t) = \sup\{n : S_n \leq t\}.$$

Тогда $X(t)$ называется *пуассоновским процессом* интенсивности λ .

Так как процесс Пуассона имеет независимые приращения, то он является марковским процессом, причем у него непрерывное время и дискретное множество состояний. Следовательно, пуассоновский процесс – это непрерывная цепь Маркова, причем это процесс рождения. Его можно было бы определить следующим образом.

Определение 3. Пусть $K(t)$, $t \geq 0$ – непрерывная цепь Маркова с множеством состояний $S = \{0, 1, 2, \dots\}$ и непрерывным временем с непрерывными справа траекториями и являющийся процессом рождения с одинаковыми интенсивностями рождения λ . Тогда процесс $K(t)$ называется *пуассоновским процессом* интенсивности λ .

Наконец, пуассоновский процесс связан с простейшим пуассоновским потоком.

Определение 4. Пусть дан простейший пуассоновский поток событий интенсивности λ . Тогда процесс $K(t) = n(0, t)$, равный числу событий на интервале времени $[0, t)$, называется *пуассоновским процессом* интенсивности λ .

10. Непрерывные процессы Маркова

10.1. Базовые понятия и свойства

В этом разделе мы рассмотрим марковские процессы с непрерывным временем и непрерывным множеством состояний. Вспомним общее определение марковского процесса, данное еще в разделе 7.

Определение. Случайный процесс $\{X(t), t \in T\}$ называется *марковским*, если

$$\begin{aligned} \mathbb{P}(X(t_{n+1}) \in B \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1) = \\ = \mathbb{P}(X(t_{n+1}) \in B \mid X(t_n) = x_n) \end{aligned}$$

для любого $n \geq 2$, любых $t_1 < t_2 < \dots < t_n < t_{n+1}$ из T , любых x_1, \dots, x_n и любого борелевского множества B , для которых условные вероятности определены выше.

Для дискретных и непрерывных марковских цепей мы вводили очень полезную конструкцию – матрицу переходных вероятностей. Для непрерывных марковских процессов введем более общую конструкцию.

Определение. *Переходной функцией* марковского процесса называется функция четырех переменных

$$P(t_0, x_0, t, B) = \mathbb{P}(X(t) \in B \mid X(t_0) = x_0),$$

где $t_0 \leq t$ – произвольные моменты времени из T , $x_0 \in S$ – произвольное состояние, B – произвольное борелевское множество, для которых условная вероятность определена.

Переходная функция – это функция четырех переменных (трех чисел и множества). Она равна вероятности того, что процесс в будущий момент t попадет в множество B при условии, что в текущий момент t_0 он принимает состояние x_0 , при этом значение этой функции определяется всеми четырьмя аргументами. Воспользовавшись формулой полной вероятности и выбирая произвольный $s \in (t_0, t)$, получаем

$$\begin{aligned} P(t_0, x_0, t, B) &= \mathbb{P}(X(t) \in B \mid X(t_0) = x_0) = \\ &= \int_{-\infty}^{+\infty} \mathbb{P}(X(t) \in B \mid X(s) = y, X(t_0) = x_0) d\mathbb{P}(X(s) < y \mid X(t_0) = x_0) = \\ &= \int_{-\infty}^{+\infty} P(s, y, t, B) P(t_0, x_0, s, dy), \end{aligned}$$

где принято обозначение $P(t_0, x_0, s, dy) = \mathbb{P}(t_0, x_0, s, [y, y + dy))$. Уравнение

$$P(t_0, x_0, t, B) = \int_{-\infty}^{+\infty} P(s, y, t, B) P(t_0, x_0, s, dy)$$

называется *уравнением Колмогорова–Чепмена*, и оно является обобщением уравнений Колмогорова–Чепмена для дискретных и непрерывных марковских цепей.

Приведем пример для винеровского процесса. Как мы знаем, винеровский процесс имеет независимые приращения, поэтому он является марковским процессом. Найдем его переходную функцию. Пусть $t > t_0 > 0$, тогда

$$\begin{aligned} P(t_0, x_0, t, B) &= \mathbb{P}(W(t) \in B \mid W(t_0) = x_0) = \\ &= \mathbb{P}(W(t) - W(t_0) + W(t_0) \in B \mid W(t_0) = x_0) = \\ &= \mathbb{P}(W(t) - W(t_0) \in B - x_0 \mid W(t_0) = x_0) = \end{aligned}$$

$$= \mathbb{P}(W(t) - W(t_0) \in B - x_0) = \frac{1}{\sqrt{2\pi(t-t_0)}} \int_B \exp\left(-\frac{(y-x_0)^2}{2(t-t_0)}\right) dy,$$

где $B - x_0$ означает множество $\{x : x = z - x_0, z \in B\}$ (каждую точку из B «сдвинули» на x_0). Эти выкладки справедливы и для $t_0 = 0$, если считать $x_0 = 0$, так как по определению винеровского процесса $W(0) = 0$ п. н. Если $t = t_0 \geq 0$, то

$$P(t_0, x_0, t_0, B) = \mathbb{P}(W(t_0) \in B | W(t_0) = x_0) = I(x_0 \in B),$$

то есть переходная функция равна индикатору того, что x_0 находится в B .

Для гауссовских процессов можно доказать очень удобный критерий «марковости». Приведем его без доказательства.

Теорема. Центрированный гауссовский процесс $X(t)$ со всюду положительной дисперсией является марковским процессом тогда и только тогда, когда его корреляционная функция $R_X(t, s)$ удовлетворяет условию*

$$R_X(t_1, t_3) = \frac{R_X(t_1, t_2)R_X(t_2, t_3)}{R_X(t_2, t_2)}$$

для любых $t_1 \leq t_2 \leq t_3$.

Легко видеть, что корреляционная функция винеровского процесса удовлетворяет этому свойству.

Для определения конечномерных распределений марковских процессов достаточно знать его функцию перехода и распределение на множестве состояний в начальный момент времени, поскольку можно доказать, что

$$\begin{aligned} & \mathbb{P}(X(t_1) \in B_1, \dots, X(t_n) \in B_n) = \\ & = \int_{-\infty}^{+\infty} \pi(dx_0; t_0) \int_{B_1} P(0, x_0, t_1, dx_1) \cdots \int_{B_n} P(t_{n-1}, x_{n-1}, t_n, dx_n), \end{aligned}$$

где $t_0 < t_1 < \dots < t_n$, B_1, \dots, B_n – борелевские множества на \mathbb{R} , а $\pi(dx; t_0)$ – вероятностная мера на \mathbb{R} , определяющая распределение на множестве состояний в момент $t = t_0$, то есть $\pi(B; t_0) = \mathbb{P}(X(t_0) \in B)$ для любого борелевского множества B . Если функция перехода P и распределение π обладают плотностями распределения вероятностей, то есть существуют неотрицательные функции $p(t_0, x_0, t, y)$ и $f_X(x_0; t_0)$ такие, что для любого борелевского множества B

$$P(t_0, x_0, t, B) = \int_B p(t_0, x_0, t, y) dy, \quad \pi(B; t_0) = \int_B f_X(x_0; t_0) dx_0,$$

то и любое n -мерное распределение также имеет плотность

$$f_X(x_1, \dots, x_n; t_1, \dots, t_n) = \int_{-\infty}^{+\infty} \prod_{i=1}^n p(t_{i-1}, x_{i-1}, t_i, x_i) f_X(x_0; t_0) dx_0.$$

Вспомним в связи с этим, что в терминах наших обозначений еще из первого раздела n -мерная функция распределения процесса в этом случае равна

$$F_X(x_1, \dots, x_n; t_1, \dots, t_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(y_1, \dots, y_n; t_1, \dots, t_n) dy_1 \dots dy_n.$$

Функция $p(t_0, x_0, t, y)$ называется *переходной плотностью* распределения марковского процесса. Из выкладок выше следует, что у винеровского процесса существует переходная плотность и она равна

$$p(t_0, x_0, t, y) = \frac{1}{\sqrt{2\pi(t-t_0)}} \exp\left(-\frac{(y-x_0)^2}{2(t-t_0)}\right), \quad t > t_0.$$

При $t = t_0$ переходной плотности у винеровского процесса не существует, хотя в этом случае обычно считают, что она все же есть и выражается с помощью обобщенной дельта-функции

$$p(t_0, x_0, t_0, y) = \delta(y - x_0).$$

Такой подход оправдан, как минимум, по двум причинам: во-первых, окажется, что у широкого и практического класса марковских процессов переходные функции и переходные плотности являются решениями некоторых уравнений в частных производных, а для постановки начальных и краевых задач с такими уравнениями и для записи их решений использование обобщенных функций является обычным делом. Во-вторых, легко видеть, что $p(t_0, x_0, t, y)$ в действительности является слабым пределом выражения для $p(t_0, x_0, t, y)$ при $t \rightarrow t_0$. Что касается начального распределения, то при $t > 0$ оно имеет плотность

$$f_W(x_0; t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x_0^2}{2t}\right),$$

так как $W(t) \in N(0, t)$. Опять же, при $t = 0$ плотности у начального распределения нет, но можно считать, что $f_W(x_0; 0) = \delta(x_0)$.

Теперь обратим внимание на то, что у винеровского процесса и переходная функция, и переходная плотность зависят от t и t_0 только через их разность.

Определение. Марковский процесс называется *однородным*, если его переходная функция $P(t_0, x_0, t, B)$ зависит лишь от разности $t - t_0$ для любых t, t_0, B и любых x_0 , для которых условные вероятности существуют.

Для однородных процессов удобно ввести функцию

$$P(x_0, t, B) = P(\tau, x_0, t + \tau, B)$$

для всех τ и допустимых значений x_0 . В этом случае уравнение Колмогорова–Чепмена запишется в виде

$$P(x_0, s + t, B) = \int_{-\infty}^{+\infty} P(y, t, B)P(x_0, s, dy).$$

Как видно из выкладок выше, винеровский процесс является однородным марковским процессом. Для него

$$P(x_0, t, B) = \frac{1}{\sqrt{2\pi t}} \int_B \exp\left(-\frac{(y - x_0)^2}{2t}\right) dy, \quad t > 0.$$

Если $t = 0$, то допустимо только значение $x_0 = 0$ и $P(0, 0, B) = I(0 \in B)$.

10.2. Диффузионные процессы

Определение. Непрерывный марковский процесс называется *диффузионным*, если его переходная функция $P(t_0, x_0, t, B)$ удовлетворяет следующим трем условиям:

- 1) $\lim_{t \rightarrow t_0+} \frac{1}{t - t_0} \int_{|x - x_0| > \delta} P(t_0, x_0, t, dx) = 0$,
- 2) $\lim_{t \rightarrow t_0+} \frac{1}{t - t_0} \int_{|x - x_0| < \delta} (x - x_0) P(t_0, x_0, t, dx) = a(t_0, x_0)$,
- 3) $\lim_{t \rightarrow t_0+} \frac{1}{t - t_0} \int_{|x - x_0| < \delta} (x - x_0)^2 P(t_0, x_0, t, dx) = b(t_0, x_0)$,

для любых $t_0, x_0, t \geq t_0$ и $\delta > 0$.

Обратим внимание на то, что здесь t_0 и x_0 – произвольные переменные, а не какие-то фиксированные числа, так что $a(t_0, x_0)$ и $b(t_0, x_0)$ – это функции двух переменных. Кроме того, обратим внимание, что в пункте 1 стоит $|x - x_0| > \delta$, а в пунктах 2 и 3 стоит $|x - x_0| < \delta$.

Выражения в пунктах 1–3 характеризуют локальные по времени и пространству свойства переходной функции. Первое выражение означает, что если процесс в момент времени t_0 находится в состоянии x_0 , то за малое время от момента t_0 процесс уйдет на конечную величину от точки x_0 лишь с малой вероятностью, которая стремится к нулю со стремлением t к t_0 . Второе выражение характеризует скорость изменения процесса вблизи точки x_0 , а точнее – совпадает с условным

математическим ожиданием выражения

$$(X(t) - X(t_0))I(|X(t) - X(t_0)| < \delta)$$

при $t \rightarrow t_0$. Функция $a(t_0, x_0)$ называется *функцией дрейфа*. Третье выражение характеризует дисперсионные свойства процесса, а точнее – совпадает с условным математическим ожиданием выражения

$$(X(t) - X(t_0))^2 I(|X(t) - X(t_0)| < \delta)$$

при $t \rightarrow t_0$. Функция $b(t_0, x_0)$ называется *коэффициентом диффузии*.

Можно доказать²³, что винеровский процесс является диффузионным процессом с функцией дрейфа $a(t_0, x_0) \equiv 0$ и коэффициентом диффузии $b(t_0, x_0) \equiv 1$.

Каков же смысл введения этого класса процессов? Как мы уже поняли, переходная функция и начальное распределение однозначно определяют семейство конечномерных распределений. Это было видно по отношению как к дискретным, так и непрерывным цепям Маркова. Мы видели, что эволюция переходных вероятностей дискретных цепей Маркова однозначно определялась лишь матрицей перехода *за один шаг*, при этом матрица переходных вероятностей за произвольное число шагов определяется некоторым рекуррентным уравнением. Эволюция переходных вероятностей непрерывных цепей Маркова определяется *производной матрицы перехода* в начальный момент времени (Q -матрицей), а матрица перехода на произвольном интервале времени определяется из системы обыкновенных дифференциальных уравнений. Матрица перехода за один шаг и Q -матрица – это локальные характеристики матриц перехода. Оказывается, что точно так же и эволюция диффузионных процессов определяется функцией дрейфа и коэффициентом диффузии – локальными характеристиками функции перехода. Другими словами, функция перехода диффузионных процессов однозначно определяется функцией дрейфа и коэффициентом диффузии, а уравнение, которому эта функция удовлетворяет, является некоторым уравнением в частных производных (уравнением диффузии, если точнее).

Введем вспомогательные обозначения. Пусть

$$F(x, t | x_0, t_0) = \mathbb{P}(X(t) < x | X(t_0) = x_0) = P(t_0, x_0, t, (-\infty, x)),$$

а $f(x, t | x_0, t_0)$ – ее функция плотности, существование которой мы

²³Миллер Б.М., Панков А.Р. Теория случайных процессов в примерах и задачах. Москва : ФИЗМАТЛИТ, 2002. С. 47.

предполагаем, то есть неотрицательная функция, для которой

$$F(x, t | x_0, t_0) = \int_{-\infty}^x f(y, t | x_0, t_0) dy.$$

Теорема Колмогорова*²⁴. При определенных условиях на функции $f(x, t | x_0, t_0)$, $a(t_0, x_0)$ и $b(t_0, x_0)$ имеют место следующие уравнения:

$$\frac{\partial f(x, t | x_0, t_0)}{\partial t} = -\frac{\partial}{\partial x} [a(t, x)f(x, t | x_0, t_0)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [b(t, x)f(x, t | x_0, t_0)],$$

называемое прямым уравнением Колмогорова, и

$$\frac{\partial f(x, t | x_0, t_0)}{\partial t_0} = -a(t_0, x_0) \frac{\partial f(x, t | x_0, t_0)}{\partial x_0} - \frac{b(t_0, x_0)}{2} \frac{\partial^2 f(x, t | x_0, t_0)}{\partial x_0^2},$$

называемое обратным уравнением Колмогорова.

Оба уравнения были строго обоснованы А. Н. Колмогоровым, хотя прямые уравнения были также независимо получены до этого физиками А. Фоккером и М. Планком, поэтому их часто называют также уравнениями Фоккера–Планка или уравнениями Колмогорова–Фоккера–Планка. В физике в применении к распределению положений частиц прямые уравнения также известны как уравнения Смолуховского и, в этом контексте, эквивалентны уравнениям диффузии или теплопроводности. В статистической механике в случае с нулевой диффузией прямые уравнения называются также уравнениями Лиувилля.

Почему одни уравнения называются прямыми, а другие обратными? В прямых уравнениях время t_0 и состояние x_0 являются параметрами, а функция $f(x, t | x_0, t_0)$ рассматривается как функция только x и t . Получается, что для фиксированных значений x_0 и t_0 ищется функция переменных x и t . В обратных же уравнениях параметрами являются x и t , а x_0 и t_0 – независимые переменные, и функция $f(x, t | x_0, t_0)$ рассматривается как функция переменных x_0 и t_0 .

Если решать эти уравнения аналитически, то неважно, что является параметрами, а что независимыми переменными, так как в любом случае получается некоторое выражение, куда входят все эти переменные. Разница проявляется при численном решении уравнений. В этом случае получается функция не всех четырех переменных, а только двух переменных. Если известны начальные момент времени t_0 и состояние x_0 и требуется найти при этих параметрах распределение

²⁴Гнеденко Б.В. Курс теории вероятностей : учебник. Изд. 8-е, испр. и доп. Москва : Едиториал УРСС, 2005. С. 298.

будущих состояний, то это *прямая* задача, в этом случае следует пользоваться прямыми уравнениями. Наоборот, если начальные момент времени и состояние неизвестны, то с учетом некоторых значений x и t (настоящих) можно решать *обратную* задачу – восстанавливать условные вероятности на множестве условий. Прямые и обратные уравнения дают как бы различные «срезы» четырехмерного пространства переменных (x, t, x_0, t_0) для функции $f(x, t | x_0, t_0)$.

Что касается начальных условий, то тут все однозначно, например, для прямых уравнений следует принять

$$f(x, t_0 | x_0, t_0) = \delta(x - x_0),$$

это условие выхода из точки x_0 в момент t_0 , и уравнения решаются при $t \geq t_0$. Для обратных уравнений аналогичным условием является

$$f(x, t | x_0, t) = \delta(x - x_0),$$

это условие входа в точку x в момент t , а решаются уравнения при $t_0 \leq t$.

У винеровского процесса прямые уравнения Колмогорова запишутся в виде

$$\frac{\partial f(x, t | x_0, t_0)}{\partial t} = \frac{1}{2} \frac{\partial^2 f(x, t | x_0, t_0)}{\partial x^2},$$

а обратные – в виде

$$\frac{\partial f(x, t | x_0, t_0)}{\partial t_0} = -\frac{1}{2} \frac{\partial^2 f(x, t | x_0, t_0)}{\partial x_0^2}.$$

Легко убедиться, что функция

$$f(x, t | x_0, t_0) = \frac{1}{\sqrt{2\pi(t - t_0)}} \exp\left(-\frac{(x - x_0)^2}{2(t - t_0)}\right)$$

удовлетворяет обоим уравнениям.

11. Вопросы на понимание

Вопросы по первому заданию

1. Что такое семейство конечномерных распределений и теорема Колмогорова?
2. Что такое пуассоновский процесс $K(t)$ и какова его связь с процессами восстановления?

3. Как связан пуассоновский процесс с пуассоновским потоком событий?
4. Назовите математическое ожидание и корреляционную функцию процессов $K(t)$ и $W(t)$.
5. Зависимы ли сечения процесса $K(t)$? А процесса $W(t)$?
6. Каково распределение промежутков между последовательными скачками процесса $K(t)$?
7. Что такое винеровский процесс $W(t)$ и какие есть способы доказать его существование?
8. Как связаны процессы $K(t)$ и $W(t)$ с процессами Леви?
9. Как восстановить семейство конечномерных распределений процесса Леви по распределению только одного его сечения?
10. Является ли процесс $K(t)$ дифференцируемым хоть в каком-нибудь смысле?
11. Является ли процесс $W(t)$ дифференцируемым хоть в каком-нибудь смысле?
12. Чему равна с.к.-производная процесса $X(t) = A \exp(Bt)$, где A и B – независимые случайные величины с конечным вторым моментом?
13. Какая общая идея лежит в доказательствах критериев с.к.-интегрируемости, с.к.-непрерывности и с.к.-дифференцируемости?
14. Откуда следует, что корреляционная функция процесса неотрицательно определена?
15. Верно ли, что если функция является неотрицательно определенной, то она является эрмитовой?
16. Верно ли, что если функция одной переменной является неотрицательно определенной, то она является характеристической функцией какой-то случайной величины?
17. Верно ли, что если функция является неотрицательно определенной, то она является корреляционной функцией какого-то случайного процесса?
18. Привести пример стационарного процесса, имеющего корреляционную функцию $R(t) = 1$.
19. Привести пример спектральной функции процесса с $R(t) = 1$.
20. Является ли $R(t) = \cos(t)$ корреляционной функцией какого-либо стационарного процесса?
21. Бывают ли нестационарные эргодические процессы?
22. Что и как позволяет оценить свойство эргодичности процесса?
23. Что такое временное среднее и среднее по пространству?
24. Являются ли $W(t)$ и $K(t)$ эргодическими? А $W(t)/t$ и $K(t)/t$?
25. Что такое спектральное разложение стационарного процесса?

26. Что такое процесс с ортогональными приращениями?
27. Является ли процесс с независимыми приращениями процессом с ортогональными приращениями?
28. Каков физический смысл спектральной плотности стационарного процесса?
29. Как связана спектральная плотность стационарного процесса со спектральной плотностью его с.к.-производной?
30. Какова характеристическая функция произвольного случайного нормального вектора?
31. У всякого ли нормального вектора существует плотность распределения?
32. Как показать линейную зависимость компонент нормального вектора в случае когда корреляционная матрица вырождена?
33. Верно ли, что вектор из нормальных случайных величин является нормальным вектором?
34. Как преобразуются распределения нормальных векторов при линейных преобразованиях?
35. Как вычислить любой момент для нормального вектора?
36. Как найти условное мат. ожидание для нормального вектора?
37. Почему векторы из сечений винеровского процесса являются нормальными?
38. Пусть процесс стационарен в узком смысле. Верно ли, что пятый момент процесса (если он существует) не зависит от времени?
39. Всегда ли стационарный в узком смысле процесс является стационарным в широком смысле?

Вопросы по второму заданию

1. Являются ли $K(t)$ и $W(t)$ марковскими процессами? А процессы вообще Леви?
2. Что такое однородная дискретная цепь Маркова?
3. Пусть X имеет распределение Бернулли. Является ли цепью Маркова последовательность (X, X, X, \dots) ?
4. Может ли состояние быть существенным и невозвратным?
5. Могут ли все состояния цепи быть нулевыми?
6. Могут ли состояния сообщаться и иметь разные периоды?
7. Как доказывается критерий возвратности состояния?
8. Может ли цепь быть неразложимой, непериодической и при этом неэргодической?
9. Всегда ли у конечной цепи существует ненулевое состояние?
10. Всегда ли у конечной цепи существует стационарное распределение? А у счетной?

11. Может ли у цепи быть бесконечное множество стационарных распределений?
12. Всегда ли сумма марковских цепей является марковской цепью?
13. Как исследовать цепь на эргодичность с помощью собственных чисел матрицы перехода?
14. Является ли пуассоновский процесс непрерывной цепью Маркова?
15. Что такое прямое и обратное уравнения Колмогорова? Каково их решение?
16. Каково время пребывания в одном состоянии в непрерывной цепи Маркова?
17. Что стоит на диагонали Q -матрицы и как это связано со временем пребывания?
18. Что такое уравнения Колмогорова–Фоккера–Планка?
19. Как связаны марковские процессы с диффузионными процессами?
20. Как искать аналитическое решение стохастических уравнений?
21. Как искать численное решение стохастических уравнений?
22. Являются ли процессы с независимыми приращениями марковскими? А наоборот?

12. Задачи по курсу

12.1. Первое задание

Задача 1. Пусть случайный процесс $X(\omega, t) = \omega t$, $t \in [0, 1]$, определен на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$, где $\Omega = \{1, 2, 3\}$, \mathcal{F} – множество всех подмножеств множества Ω , а мера \mathbb{P} такова, что $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = 1/3$. Построить вторичное (выборочное) вероятностное пространство процесса.

Задача 2. Поток сделок в фирме моделируется с помощью пуассоновского процесса $K(t)$ с интенсивностью $\lambda = 100$ сделок/час. Каждая сделка приносит доход $V_n \in R(a, b)$, $a = 10$, $b = 100$ условных единиц денег (R означает непрерывное равномерное распределение). Считая, что $K(t)$, $\{V_i\}$ – независимые в совокупности случайные величины, найдите математическое ожидание, дисперсию и характеристическую функцию выручки за время t . Докажите, что она имеет асимптотически нормальное распределение.

Задача 3. Случайный процесс $X(t)$, $t \geq 0$, представляет собой сумму n независимых пуассоновских процессов с интенсивностями λ_i , $i = 1, \dots, n$. Определить тип и параметры процесса $X(t)$.

Задача 4. Пусть $K(t)$, $t \geq 0$ – пуассоновский случайный процесс

интенсивности λ , а $X(t)$ – случайный процесс, полученный в результате удаления из $K(t)$ всех событий, очередной номер которых не кратен s . Определить тип и параметры распределения интервала между соседними событиями в случайном процессе $X(t)$.

Задача 5. Пусть случайный вектор $X \in N(0, R)$ с матрицей ковариаций

$$C = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Вычислить $\mathbb{E}(X_1^2 X_2^2)$, $\mathbb{E}(X_1 X_2^3 X_3)$, $\mathbb{E}(X_1 X_2 X_3^2)$.

Задача 6. Пусть $X(t)$ – нормальный (гауссовский) случайный процесс с математическим ожиданием $m_X(t) = m = \text{const}$ и корреляционной функцией

$$R_X(t, s) = be^{-a|t-s|}, \quad a > 0, \quad b > 0.$$

Найти вероятность $\mathbb{P}(X(t) > c | X(s) = x)$ для $t > s$.

Задача 7. Исследовать винеровский процесс $W(t)$, $t \geq 0$, на непрерывность, дифференцируемость и интегрируемость во всех смыслах (в среднем квадратичном, почти наверное, по вероятности, по распределению).

Задача 8. Исследовать пуассоновский процесс $K(t)$, $t \geq 0$, на непрерывность, дифференцируемость и интегрируемость во всех смыслах (в среднем квадратичном, почти наверное, по вероятности, по распределению).

Задача 9. Пусть A, X, Y – случайные величины такие, что V не зависит от A и X , причем V равномерно распределена на отрезке $[0, 2\pi]$, A и X имеют совместное распределение с функцией плотности распределения $f(a, x)$. Исследовать процесс $Z(t) = A \cos(Xt + V)$, $t \geq 0$, на стационарность в широком и узком смыслах.

Задача 10. Пусть $K(t)$, $t \geq 0$ – пуассоновский случайный процесс интенсивности λ . Исследовать процесс

$$X(t) = K(t+1) - K(t), \quad t \geq 1$$

на стационарность в широком смысле.

Задача 11. Доказать, что сумма независимых стационарных случайных процессов является стационарным случайным процессом. (Доказать это утверждение как для стационарности в широком смысле, так и узком смысле).

Задача 12. Пусть $K(t)$, $t \geq 0$ – пуассоновский случайный процесс интенсивности λ . Исследовать процесс $X(t) = K(t)/t$, $t \geq 1$, на эргодичность по математическому ожиданию.

Задача 13. В модели Блэка–Шоулса–Мертона эволюция цены акции описывается геометрическим броуновским движением

$$S(t) = S(0) \exp(at + \sigma W(t)), \quad t \geq 1,$$

где $S(0)$, a и $\sigma > 0$ – неслучайные постоянные величины, $W(t)$ – винеровский процесс. Воспользовавшись понятием эргодичности случайных процессов, оценить неизвестную величину a .

Задача 14. 1) Может ли функция

$$R(t) = \begin{cases} 1, & t \in [-T, T], \\ 0, & t \notin [-T, T] \end{cases}$$

быть характеристической функцией некоторой случайной величины?

2) А корреляционной функцией некоторого стационарного в широком смысле случайного процесса?

3) Изменится ли ответ, если сгладить разрывы функции $R(t)$ в точках $t = \pm T$?

4) Является ли функция

$$R(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0 \end{cases}$$

неотрицательно определенной?

5) Верно ли, что если функция является неотрицательно определенной, то она является характеристической функцией некоторой случайной величины? А корреляционной функцией случайного процесса?

6) Как связана неотрицательная определенность функции и неотрицательность ее Фурье-образа?

7) Какой физический смысл имеет спектральная плотность стационарного в широком смысле случайного процесса?

Задача 15. Построить пример стационарного процесса $X(t)$ с корреляционной функцией $R_X(t) = \cos t$. Найти спектральную функцию этого процесса. Существует ли спектральная плотность этого процесса?

Задача 16. Рассмотрим колебательный контур, состоящий из последовательно соединенных катушки индуктивности, конденсатора, сопротивления и источника сторонних ЭДС. ЭДС $\mathcal{E}(t)$, заряд $q(t)$ на обкладках конденсатора и производные $\dot{q}(t)$, $\ddot{q}(t)$ считаются достаточно с.к.-гладкими стационарными в широком смысле случайными процессами с нулевым математическим ожиданием. Считая известной спектральную плотность $\rho_{\mathcal{E}}(\lambda)$ процесса $\mathcal{E}(t)$, вычислить спектральную плотность $\rho_q(\lambda)$ процесса $q(t)$.

12.2. Второе задание

Задача 17. Показать, что для дискретной марковской цепи при $m_1 < m_2 < m_3$ выполнено равенство

$$\begin{aligned} \mathbb{P}(X_{m_1} = x_1, X_{m_3} = x_3 \mid X_{m_2} = x_2) &= \\ &= \mathbb{P}(X_{m_1} = x_1 \mid X_{m_2} = x_2) \mathbb{P}(X_{m_3} = x_3 \mid X_{m_2} = x_2). \end{aligned}$$

Задача 18. Доказать, что условие

$$\mathbb{P}(X_{m_n} = x_n \mid X_{m_{n-1}} = x_{n-1}, \dots, X_{m_0} = x_0) = \mathbb{P}(X_{m_n} = x_n \mid X_{m_{n-1}} = x_{n-1})$$

в определении марковской цепи равносильно условию

$$\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}).$$

Здесь $n \geq 1, m_0 < m_1 < \dots < m_n$.

Задача 19. Пусть X_0, X_1, \dots, X_n – дискретная марковская цепь. Является ли марковской последовательность X_n, X_{n-1}, \dots, X_0 ?

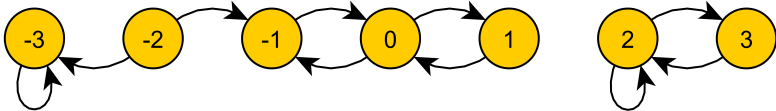
Задача 20. Пусть $\{X_n\}$ – последовательность независимых одинаково распределенных случайных величин, принимающих значения -1 и 1 с вероятностями p и $q = 1 - p$ соответственно. Выяснить, будет ли последовательность $\{Y_n\}$ марковской цепью, если

$$1) Y_n = X_n X_{n+1}, \quad 2) Y_n = \max_{0 \leq i \leq n} X_i, \quad 3) Y_n = \prod_{i=0}^n X_i.$$

Задача 21. Пусть $\{X_n\}$ и $\{Y_n\}$ – две марковские цепи. Будет ли марковской последовательность $\{X_n + Y_n\}$?

Задача 22. Пусть $\{X_n\}$ – марковская цепь, а $\psi(x)$ – некоторая функция. Будет ли последовательность $\{\psi(X_n)\}$ марковской цепью?

Задача 23. Классифицировать состояния однородной дискретной цепи Маркова с множеством состояний $S = \{-3, -2, -1, 0, 1, 2\}$ (стрелками изображены переходы, имеющие ненулевые вероятности).



Задача 24. Доказать, что в конечной неразложимой однородной дискретной цепи Маркова все состояния ненулевые.

Задача 25. Доказать, что если j -е состояние невозвратно, то для всех i сходится ряд $\sum_{n=1}^{\infty} p_{ij}(n)$.

Задача 26. Доказать, что неразложимая дискретная марковская цепь, у матрицы переходных одношаговых вероятностей которой

хотя бы один диагональный элемент положителен, не может быть периодической. Может ли неразложимая дискретная цепь Маркова, у матрицы одношаговых переходных вероятностей которой все диагональные элементы равны нулю, быть непериодической?

Задача 27. Проведение некоторого эксперимента состоит в осуществлении большого числа шагов. На каждом шаге может быть выбрано одно из двух возможных действий. Каждое действие может привести как к успеху, так и к неудаче данного шага. Существуют вероятности успеха p_1 и p_2 первого и второго действий соответственно и вероятности их неудач $q_1 = 1 - p_1$, $q_2 = 1 - p_2$, которые экспериментатору неизвестны. Цель экспериментатора состоит в максимизации математического ожидания числа успехов в эксперименте в целом. Сравнить две стратегии проведения эксперимента: а) равновероятный выбор на каждом шаге каждого действия, б) повторение на следующем шаге действия, приведшего к успеху на предшествующем шаге, и смена действия, приведшего к неудаче.

Задача 28. Поток событий является однородным и не имеет последействия. Величина длины интервала времени между двумя последовательными событиями имеет показательное распределение $\text{Exp}(\lambda)$, $\lambda > 0$. Длины интервалов времени между событиями независимы в совокупности. Доказать, что поток – простейший.

Задача 29. Поток отказов прибора – простейший с интенсивностью λ . Если прибор отказал, то отказ обнаруживается в течение случайного времени, имеющего распределение $\text{Exp}(\nu)$. Ремонт осуществляется после обнаружения отказа и продолжается случайное время с распределением $\text{Exp}(\mu)$. Найти стационарную вероятность того, что прибор исправен.

Задача 30. Пусть $X(t)$ – центрированный нормальный процесс. Доказать, что для того, чтобы этот процесс был марковским, необходимо, чтобы его корреляционная функция $R_X(t, s)$ удовлетворяла при $t_1 \leq t_2 \leq t_3$ равенству

$$R_X(t_1, t_3) = \frac{R_X(t_1, t_2)R_X(t_2, t_3)}{R_X(t_2, t_2)}.$$

(Это свойство является и достаточным для марковости нормального процесса, но доказать это сложнее.)

Задача 31. Пусть $\{X(t), t \in [0, 1]\}$ – броуновский мост, т.е.

$$X(t) = W(t) - tW(1),$$

где $W(t)$ – винеровский процесс. Показать, что процесс $X(t)$ является

марковским. Показать, что он не является процессом с независимыми приращениями.

Задача 32. Доказать, что процесс Орнштейна–Уленбека (Ornstein–Uhlenbeck)

$$X(t) = \frac{\sigma}{2\theta} e^{-\theta t} W(e^{2\theta t}), \quad \sigma, \theta, t > 0,$$

где $W(t)$ – винеровский процесс, является нормальным стационарным марковским процессом.

Заключение

Учебное пособие содержит семестровый лекций по случайным процессам и отражает ключевые разделы теории случайных процессов: гауссовские, стационарные, эргодические, марковские процессы, а также стохастический анализ. Содержащиеся в пособии вопросы на понимание и задачи помогают студентам осваивать курс и рекомендуются автором для приема экзаменов, контрольных работ и заданий по предмету. Литература в конце пособия рекомендуется как для углубления и расширения знаний в вышеуказанных разделах теории случайных процессов, так и для знакомства с приложениями.

Литература

1. Боровков А.А. Теория вероятностей. Москва : Эдиториал УРСС, 1999.
2. Булинский А.В. Случайные процессы. Примеры, задачи и упражнения. Москва : МФТИ, 2010.
3. Лекции по случайным процессам : учебное пособие / под ред. А. В. Гасникова. Москва : ЛЕНАНД, 2022.
4. Булинский А.В., Ширяев А.Н. Теория случайных процессов. Москва : ФИЗМАТЛИТ, 2005.
5. Кельберт М.Я., Сухов Ю.М. Вероятность и статистика в примерах и задачах. Т. 2. Марковские цепи как отправная точка теории случайных процессов и их приложения. Москва : МЦНМО, 2010.
6. Крамер Г., Лидбеттер М. Стационарные случайные процессы. Москва : Мир, 1969.
7. Миллер Б.М., Панков А.Р. Теория случайных процессов в примерах и задачах. Москва : Физматлит, 2002.
8. Натан А.А., Горбачев О.Г., Гуз С.А. Основы теории случайных процессов : учеб. пособие. Москва : МЗ Пресс; МФТИ, 2003.
9. Ширяев А.Н. Вероятность. В 2-х кн. 6-е изд., испр. Москва : МЦНМО, 2017.
10. Чжун К.-л. Однородные цепи Маркова. Москва : Мир, 1964.
11. Яглом А.М. Корреляционная теория стационарных случайных функций. Ленинград : Гидрометеиздат, 1981.