

Стохастический анализ: лекции

Дарья Демидова, Максим Каледин, Денис Ракитин

6 июня 2024 г.

Оглавление

Описание курса	6
0 Обзор теории вероятности	7
0.1 Вероятностное пространство	7
0.1.1 Измеримое пространство	7
0.1.2 Измеримые функции	8
0.1.3 Мера	9
0.2 Независимость	10
0.3 Интеграл Лебега	11
0.4 Условное матожидание	14
1 Случайные процессы: определения	16
1.1 Случайный процесс	16
1.2 Семейство конечномерных распределений и согласованность	18
1.3 Пример: гауссовский процесс (отнесено в семинар)	21
1.4 Числовые характеристики и стационарность	22
2 Винеровский процесс	25
2.1 Броуновское движение	25
2.2 Непрерывность траекторий	30
2.3 Многомерный Винеровский процесс	34
2.4 Формула Иссерлиса: моменты гауссовых векторов	35
3 Броуновское движение: из физики в финансы	36
3.1 Откуда идёт финансовая математика	36
3.2 Геометрическое Броуновское движение	40
3.3 Задача оценки опционов	42
4 Цепи Маркова	47
4.1 Определение цепи Маркова	47
4.1.1 Стохастические матрицы	48
4.1.2 Однородная цепь Маркова	50
4.1.3 Эргодическая теорема	51
4.2 Обобщение цепи Маркова на случай непрерывного времени	54
4.2.1 Инфинитезимальная матрица	56

5 Пуассоновский процесс	57
5.1 Мотивационный пример	57
5.2 Предельная теорема Пуассона	57
5.3 Определение процесса Пуассона	58
5.3.1 Проверка теоремы Колмогорова	58
5.4 Траектории процесса Пуассона	60
5.5 Свойства процесса Пуассона	62
5.6 Процесс Пуассона как марковская цепь с непрерывным временем	63
6 Мартингалы	67
6.1 Фильтрации, моменты остановки и мартингалы	67
6.1.1 Фильтрации	67
6.1.2 Момент остановки	67
6.1.3 σ -алгебра событий, определенных до момента остановки	68
6.1.4 Случайный процесс, согласованный с фильтрацией	69
6.1.5 Мартингалы	69
6.2 Мартингалы с дискретным временем	70
7 Мартингалы: 2	74
7.1 Сходимость мартингалов и непрерывные мартингалы	74
7.1.1 Сходимость мартингалов	74
7.1.2 Мартингалы с непрерывным временем	77
8 Процессы Леви	81
8.1 Хорошо иметь хорошие разрывы	81
8.2 От классики к процессам Леви: Пуассоновский процесс	83
8.3 От классики к процессам Леви: Винеровский процесс	85
8.4 Что-то третье: сложенный процесс Пуассона	85
9 Процессы Леви: 2	88
9.1 Что общего у трёх процессов?	88
9.2 Экспонента процесса Леви	89
9.3 Разложение Леви-Ито и отказ от плотности	91
9.4 Больше примеров процессов Леви	93
9.5 Бонус: ещё один смысл меры Леви	96
10 Стохастический интеграл Ито	98
10.1 Дифференциальные уравнения с "шумом"	98
10.2 Первая попытка: простые функции	101
10.3 Интеграл Ито	102
10.4 Первые расчёты	105

11 Исчисление Ито	107
11.1 Свойства интеграла Ито	107
11.2 Процессы Ито	109
11.3 Формула Ито и интегрирование по частям	110
11.4 Используем формулу Ито	112
12 Стохастические дифференциальные уравнения	115
12.1 Стохастические дифференциальные уравнения	115
12.2 Численные методы решения	118
12.3 Формула Блэка-Шоулза	119
12.4 Уравнение Блэка-Шоулза	121
13 Задача фильтрации	123
13.1 Выделение целевого сигнала из шума	123
13.2 Постановка задачи	124
13.3 Строим оценку сигнала(простой случай)	126
13.4 Фильтр Калмана-Бюси (непрерывное время)	128
13.4.1 Проекция – это условное матожидание	129
13.4.2 Линейные оценки против измеримых	130
13.4.3 Процесс невязок	131
13.4.4 Финальные штрихи	132
14 Диффузия	135
14.1 Процесс диффузии	135
14.2 Марковское свойство	139
14.3 Сильное Марковское свойство	140
14.4 Генератор диффузии	143
15 Уравнения Колмогорова	151
15.1 Обратное уравнение Колмогорова	151
15.2 Новые переходные вероятности	152
15.3 Прямое уравнение Колмогорова	155
15.4 Стационарные распределения и эргодичность	157
16 Диффузионные модели	159
17 Продвинутое семплирование	160
17.1 Зачем нужно семплирование?	160
17.2 Цепь Маркова	162
17.3 Как построить подходящую цепь?	163
17.4 В непрерывном случае	167

17.5 Ещё альтернативы МСМС	169
18 Назад в финансы	172
18.1 Формула Фейнмана-Каца	172
18.2 Ещё о Винеровском	174
18.3 Замена времени	175
18.4 Теоремы Гирсанова	176
18.5 Наличие арбитража на рынке	179

Описание курса

Предисловие

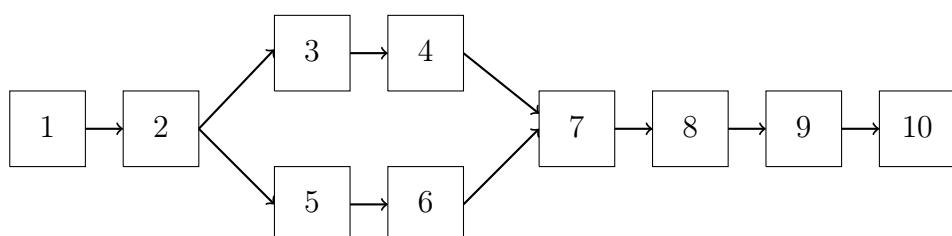
Случайные процессы – широко используемый набор инструментов для моделирования во многих прикладных областях: финансовой математике, макроэкономике, физике. Эволюции цен, диффузия, потоки заявок для обработки, путешествие пользователя по ресурсам сети Интернет – это лишь малая часть явлений, которые можно исследовать с их помощью. В нашем курсе мы хотели бы, с одной стороны, предоставить хороший фундамент для дальнейшего более самостоятельного изучения специальных областей, а с другой – предоставить несколько интересных примеров, как связать это знание с компьютером и научиться не только доказывать интересные факты, но и смело строить решения практических задач в виде программ.

Пререквизиты

1. Математический анализ 1-2;
2. Линейная алгебра;
3. Теория вероятности;
4. Программирование на Python, научные пакеты (numpy, sklearn,..) в плюс;
5. Обыкновенные дифференциальные уравнения (самые основы и общее представление).

Схема курса

В зависимости от вашего интереса курс лекций можно изучать в разном порядке. Основные зависимости обозначены на рисунке ниже и могут помочь сориентироваться и составить свой план. Лекция 1 является вводной и может быть опущена.



Обзор теории вероятности

Главная цель этой лекции – напомнить необходимые базовые понятия из теории вероятности, которые потом мы будем постоянно использовать в течение курса. Мы начинаем с понятия сигма-алгебры, измеримых множеств, измеримых пространств, измеримых функций и меры. Далее мы говорим о понятии независимости и случайных величинах и заканчиваем кратким обзором конструкции интеграла Лебега и определениями матожидания, дисперсии и ковариации случайных величин.

0.1 Вероятностное пространство

0.1.1 Измеримое пространство

Измеримое пространство - это пространство, на котором можно задать меру. Давайте вспомним, как оно строится.

Определение 0.1. Пространство элементарных событий - это некоторое непустое множество, обычно обозначаемое Ω , элементы которого $\omega \in \Omega$ называются элементарными событиями.

Определение 0.2. Совокупность \mathcal{G} подмножеств множества Ω называется алгеброй (множеством), если выполняются следующие три условия:

1. $\Omega \in \mathcal{G}$
2. из $C \in \mathcal{G}$ следует, что $\Omega \setminus C \in \mathcal{G}$
3. из $C_1, \dots, C_n \in \mathcal{G}$ следует, что $\bigcup_{i=1}^n C_i \in \mathcal{G}$

Пример 0.1. Наименьшая алгебра состоит всего из двух элементов: $\{\Omega, \emptyset\}$.

Наибольшая алгебра - это алгебра всех возможных подмножеств множества Ω .

Пример 0.2. Пусть задано конечное или счетное разбиение на Ω (совокупность непересекающихся подмножеств $\{B_i\}_{i \in I}$, такое что $\Omega = \bigcup_{i \in I} B_i$). Тогда это разбиение порождает алгебру, состоящую из элементов вида $C = \bigcup_{i \in J \subset I} B_i$, где J - это любое подмножество индексов.

Упражнение 0.1. Докажите, что любая конечная алгебра порождается каким-то разбиением.

Определение 0.3. Алгебра \mathcal{F} называется σ -алгеброй, если она замкнута относительно счетных объединений, т.е. если $C_i \in \mathcal{F}, i \in \mathbb{N}$, то $\bigcup_{i \in \mathbb{N}} C_i \in \mathcal{F}$.

Пример 0.3. Рассмотрим множество целых чисел \mathbb{Z} . Множество всех его подмножеств будет σ -алгеброй.

Пример 0.4. На вещественной прямой \mathbb{R} чаще всего используют σ -алгебру, порожденную открытыми множествами. Она называется борелевской σ -алгеброй $\mathcal{B}(\mathbb{R})$ и задается как наименьшая σ -алгебра, содержащая все открытые множества. Причина, по которой сигма-алгебра всех подмножеств неудобна состоит в том, что на ней в общем случае нельзя определить вероятностную меру – в этой сигма-алгебре существуют множества, мера которых не определена (вне зависимости от выбранной меры!), так называемые неизмеримые множества.

Пример 0.5. Аналогично можно задать борелевскую σ -алгебру $\mathcal{B}(\mathbb{R}^n)$ на \mathbb{R}^n .

Определение 0.4. Измеримым пространством называется пара (Ω, \mathcal{F}) , где Ω – пространство элементарных событий, а \mathcal{F} – некоторая σ -алгебра подмножеств Ω . Элементы σ -алгебры \mathcal{F} называются измеримыми множествами.

0.1.2 Измеримые функции

Измеримые функции – это базовое понятие для задания случайной величины.

Определение 0.5. Пусть на измеримом пространстве (Ω, \mathcal{F}) задана функция $f : \Omega \rightarrow \Xi$ в другое измеримое пространство (Ξ, \mathcal{G}) . Тогда функция f называется измеримой, если прообраз любого измеримого множества измерим, т.е. $\forall A \in \mathcal{G} f^{-1}(A) \in \mathcal{F}$.

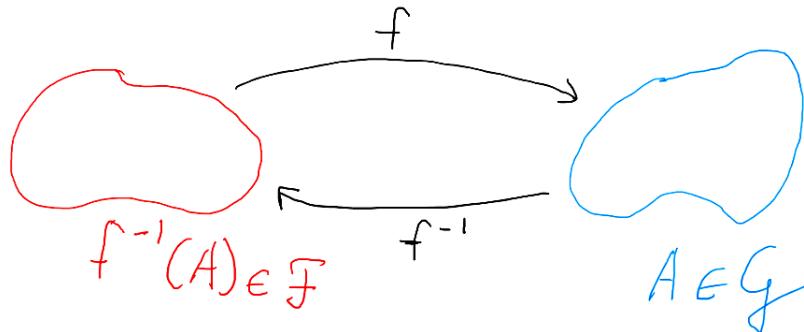


Рис. 1: Измеримая функция $f : \Omega \rightarrow \Xi$ переводит измеримые множества (относительно сигма-алгебры \mathcal{F}) в измеримые множества (относительно сигма-алгебры \mathcal{G})

Иными словами, мы требуем, чтобы измеримые множества отображались такой функцией в измеримые множества (измеримость множеств определяется относительно заданных сигма-алгебр). Такая регулярность позже позволяет корректно определять вероятностные меры, порождённые случайной величиной.

Упражнение 0.2. Если \mathcal{F} порождена разбиением $\{B_i\}_{i \in I}$, и функция f является \mathcal{F} -измеримой, тогда на каждом B_i она принимает постоянное значение.

Пример 0.6. Любая \mathcal{F} -измеримая функция порождает σ -алгебру $\mathcal{F}^f \subset \mathcal{F}$ на Ω :

$$\mathcal{F}^f := \{f^{-1}(A) : A \in \mathcal{G}\}.$$

Можно показать, что \mathcal{F}^f является наименьшей σ -алгеброй, относительно которой функция f измерима.

В качестве более конкретного примера рассмотрим $\Omega = \{1, 2, 3, 4, 5\}$, сигма алгебру

$$\mathcal{F} = \{\emptyset, \Omega, \{1, 2, 3\}, \{4, 5\}\}$$

и измеримую функцию $f : \Omega \rightarrow \{0, 1\}$ такую, что для всех $\omega \in \Omega$ $f(\omega) = 1$. Зададим на $\{0, 1\}$ сигма-алгебру, состоящую из всех подмножеств. Заметим теперь, что

$$\begin{aligned} f^{-1}(\{0\}) &= \emptyset, & f^{-1}(\{1\}) &= \Omega, \\ f^{-1}(\{0, 1\}) &= \Omega, & f^{-1}(\emptyset) &= \emptyset, \end{aligned}$$

следовательно, $\mathcal{F}^f = \{\emptyset, \Omega\} \subset \mathcal{F}$.

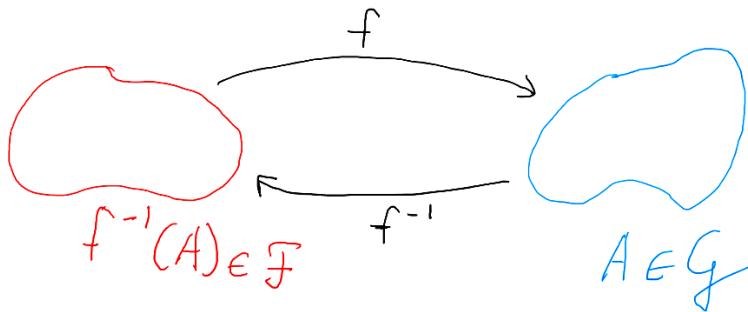


Рис. 2: Сигма-алгебра \mathcal{F}^f состоит из всех прообразов измеримых множеств

Упражнение 0.3. Докажите утверждения:

1. Пусть $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$, $(\Omega_3, \mathcal{F}_3)$ – измеримые пространства. Если $f : \Omega_1 \rightarrow \Omega_2$ и $g : \Omega_2 \rightarrow \Omega_3$ измеримы, то их композиция тоже измерима.
2. Рассмотрим измеримые пространства (Ω, \mathcal{F}) и $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Если $f : \Omega \rightarrow \mathbb{R}^n$ и $g : \Omega \rightarrow \mathbb{R}^n$ измеримы, то их линейная комбинация тоже измерима.
3. Рассмотрим измеримые пространства (Ω, \mathcal{F}) и $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Если $f : \Omega \rightarrow \mathbb{R}$ и $g : \Omega \rightarrow \mathbb{R}$ измеримы, то их произведение тоже измеримо.

0.1.3 Мера

Определение 0.6. Пусть задано измеримое пространство (Ω, \mathcal{F}) . Функция $\mu : \mathcal{F} \rightarrow [0, \infty)$ называется конечной неотрицательной мерой, если для любого счетного семейства измеримых непересекающихся множеств мера объединения равна сумме мер, т.е.

$\mu\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \mu(C_i)$ для любых $\{C_i\}_{i \in \mathbb{N}}$ таких, что $C_i \in \mathcal{F}, C_i \cap C_j = \emptyset (i \neq j)$. Это свойство называется σ -аддитивностью.

Если $\mu(\Omega) < \infty$, то мера называется конечной, а мера, для которой $\mu(\Omega) = 1$, – вероятностной.

Пример 0.7. Рассмотрим $S = [0, 1]$ и измеримое пространство $(S, \mathcal{B}(S))$, где $\mathcal{B}(S)$ – борелевская σ -алгебра, порожденная интервалами в S . На $(S, \mathcal{B}(S))$ можно определить меру Лебега $\lambda(A) = b - a$ для любого интервала $A = (a, b) \in \mathcal{B}(S)$. Так как $\lambda(S) = 1$, это вероятностная мера.

Мера Лебега на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ определяется так же, но она не будет конечной, так как $\mu(\mathbb{R}) = \infty$.

Определение 0.7. Если снабдить измеримое пространство (Ω, \mathcal{F}) вероятностной мерой P , тогда оно будет называться вероятностным пространством и обозначаться (Ω, \mathcal{F}, P) , а измеримые функции $f : \Omega \rightarrow \mathbb{R}$ – случайными величинами.

Определение 0.8. Пусть P – вероятностная мера на вероятностном пространстве (Ω, \mathcal{F}, P) . Если утверждение, зависящее от $\omega \in \Omega$, выполняется для подмножества $A \subset \Omega$, такого, что $P(A) = 1$, то говорят, что утверждение выполняется почти наверное.

Пример 0.8. Случайную величину $X : \Omega \rightarrow \mathbb{R}$ рассматривают как отображение из измеримого пространства (Ω, \mathcal{F}) в измеримое пространство $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, т.к. она переводит \mathcal{F} -измеримые множества в $\mathcal{B}(\mathbb{R})$ -измеримые множества. Если на (Ω, \mathcal{F}) задана вероятностная мера P , то случайная величина X естественным образом индуцирует вероятностную меру на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ как

$$P^X(B) := P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\})$$

для любых $B \in \mathcal{B}(\mathbb{R})$. Чаще всего используют короткую нотацию $P(X \in B) := P^X(B)$, чтобы обозначить вероятность того, что случайная величина принимает значения во множестве B .

Меру P^X в этом случае в англоязычной литературе часто называют *pushforward*-мерой меры P при отображении X .

0.2 Независимость

Независимость – это свойство, описывающее взаимосвязь двух или более событий, σ -алгебр или случайных величин.

Определение 0.9. Пусть задано вероятностное пространство (Ω, \mathcal{F}, P) и $A, B \in \mathcal{F}$.

Если $P(B) \neq 0$, то условной вероятностью события A при условии события B называется число $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

События A и B называются независимыми, если $P(A \cap B) = P(A)P(B)$, что эквивалентно следующему равенству $P(A|B) = P(A)$.

События $A_1, A_2, \dots, A_n \in \mathcal{F}$ называются независимыми в совокупности, если $P(A_{i_1}, A_{i_2}, \dots, A_{i_k}) = P(A_{i_1})P(A_{i_2})\dots P(A_{i_k})$ для любых неповторяющихся индексов $1 \leq i_1 < i_2 < \dots < i_k \leq n$.

Определение 0.10. σ -алгебры $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ называются независимыми, если любой набор событий $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \dots, A_n \in \mathcal{F}_n$ независим.

Мы уже знаем, что случайная величина $X : \Omega \rightarrow \mathbb{R}$ порождает σ -алгебру \mathcal{F}^X , состоящую из множеств вида $X^{-1}(B), B \in \mathcal{B}(\mathbb{R})$.

Определение 0.11. Случайные величины X_1, X_2, \dots, X_n независимы, если они порождают независимые σ -алгебры $\mathcal{F}^{X_1}, \mathcal{F}^{X_2}, \dots, \mathcal{F}^{X_n}$.

Упражнение 0.4. (Эквивалентное определение независимости) Пусть (Ω, \mathcal{F}, P) – вероятностное пространство. Докажите, что случайные величины X_1, \dots, X_n независимы тогда и только тогда, когда для любых $A_1, \dots, A_n \in \mathcal{F}$

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \dots P(X_n \in A_n).$$

Для набора случайных величин можно требовать разной структуры независимости. Самая сильная – независимость в совокупности.

Определение 0.12. Случайные величины X_1, X_2, \dots, X_n независимы в совокупности, если любое подмножество X_{i_1}, \dots, X_{i_k} , $k \leq n$ из этого набора – это независимые случайные величины.

Пример 0.9. В классической статистике выборкой называют набор независимых (в совокупности) и одинаково распределённых (с распределением F) случайных величин. Часто пишут $X_1, \dots, X_n \sim^{i.i.d.} F$, i.i.d означает *independent and identically distributed*.

0.3 Интеграл Лебега

Пусть (Ω, \mathcal{F}, P) – пространство с мерой P (не обязательно вероятностной). Мы построим интеграл Лебега для ограниченных функций $f : \Omega \rightarrow \mathbb{R}$, для неограниченных требуется существенно больше технической работы. Мы приведём общую схему, которая может быть критична для понимания, но за всеми техническими деталями рекомендуем [27, Гл.3.1].

Рассмотрим не более, чем счетное множество $\mathcal{A} \subset \mathcal{F}$, состоящее из непересекающихся измеримых множеств A_i , $i = 1, 2, \dots, N$, где N может равняться бесконечности. Начнём с рассмотрения положительных простых функций $f_{\mathcal{A}} : \Omega \rightarrow \mathbb{R}_+$, то есть, таких, что они равны константе на каждом A_i : $f_{\mathcal{A}}(\omega) = a_i \geq 0$ для всех $\omega \in A_i$. Определим интеграл Лебега для таких функций как

$$\int_{\Omega} f_{\mathcal{A}}(\omega) dP := \sum_{i=1}^N a_i P(A_i), \quad (1)$$

если ряд сходится.

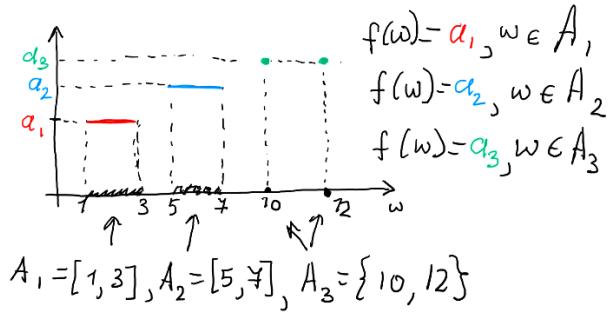


Рис. 3: Пример простой функции, равной нулю везде, кроме $x \in A_1, A_2, A_3$

Пример 0.10. Может показаться, что мы строим интеграл точно так же, как интеграл Римана, но это не так. В отличие от последнего мы теперь перечисляем значения функции и смотрим, на каком измеримом множестве они достигаются, причём последние запросто могут быть не отрезками (см. Рис. 3). Даже на этом этапе мы определили интеграл для одной не интегрируемой по Риману функции, функции Дирихле

$$D(x) = \begin{cases} 1, & x \in \mathbb{Q}, \\ 0, & \text{иначе.} \end{cases}$$

Зададим пространство с мерой Лебега λ как $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. Функция D – простая положительная функция: на измеримом множестве $A_1 = \mathbb{Q}$ она принимает значение 1, а значение 0 она принимает на другом измеримом множестве $A_2 = \mathbb{R} \setminus A_1$. Интеграл Лебега по мере λ (мере Лебега) будет равен нулю: множество \mathbb{Q} имеет меру $\lambda(\mathbb{Q}) = 0$. Интеграл Римана такой функции не существует.

Утверждение 0.1. Для любой положительной измеримой функции f существует последовательность $\{f_n\}$ простых положительных измеримых функций такая, что

$$\max_{\omega \in \Omega} |f(\omega) - f_n(\omega)| \rightarrow 0 \quad \text{при } n \rightarrow 0.$$

Используя идею, что каждая ограниченная положительная измеримая функция f аппроксимируется в смысле бесконечной нормы (иначе, равномерно, как выше), мы можем определить интеграл Лебега как

$$\int_{\Omega} f(\omega) dP = \lim_{n \rightarrow \infty} \int_{\Omega} f_n(\omega) dP. \quad (2)$$

Для аккуратности нужно доказать дополнительно

Утверждение 0.2. *Предел (2) не зависит от выбора аппроксимирующей последовательности.*

Наконец, остаётся обобщить понятие интеграла на любую ограниченную измеримую функцию. Заметим, что множества $[0, \infty)$ и $(-\infty, 0)$ измеримы относительно $\mathcal{B}(\mathbb{R})$, следовательно мы можем представить функцию f как разность двух измеримых функций:

$$f(\omega) = f_+(\omega) - f_-(\omega) := f(\omega)\mathbf{1}_{x:f(x)\geq 0}(\omega) - f(\omega)\mathbf{1}_{x:f(x)<0}(\omega), \quad (3)$$

для каждой из которых мы определили интеграл Лебега.

Интеграл Лебега будет обладать многими свойствами интеграла Римана (в том числе, линейностью) и будет совпадать с ним, если интеграл Римана существует.

Упражнение 0.5. *Пусть $(\Omega, \mathcal{F}, \mu)$ – пространство с мерой. Покажите, что*

$$\int_A d\mu = \mu(A).$$

Матожидание случайной величины мы определяем с помощью интеграла Лебега.

Определение 0.13. *Пусть (Ω, \mathcal{F}, P) – вероятностное пространство, на котором определена случайная величина X . Число $\mathbb{E}[X] := \int_{\Omega} X(\omega)dP$, где интеграл берётся в смысле Лебега, называют матожиданием случайной величины X .*

Число $\mathbb{E}[(X - \mathbb{E}[X])^2]$ называется дисперсией X .

Если X, Y – две случайные величины на заданном вероятностном пространстве, то их ковариация определяется как

$$\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])],$$

а для случайных векторов $X, Y : \Omega \rightarrow \mathbb{R}^n$ определяется ковариационная матрица

$$\text{Cov}(X, Y) := (\text{Cov}(X_i, Y_j))_{i,j=1,\dots,n}.$$

Пример 0.11. *Пусть $X \sim \text{Ber}(p)$, то есть, она равна 1 с вероятностью p и 0 с вероятностью $1 - p$ (распределение Бернулли). Заметим, что мера P^X на $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X)$ определена вполне понятно, хотя мы не определили явно вероятностное пространство (Ω, \mathcal{F}, P) :*

$$P^X((a, b)) := P(X \in (a, b)) = \begin{cases} p, & 1 \in (a, b) \\ 1 - p, & \text{иначе.} \end{cases}$$

Мера P^X – это пример так называемой дискретной меры, которая порождается дискретными случайными величинами. Случайная величина X – это простая функция и (согласно определению через интеграл Лебега) матожидание $\mathbb{E}[X] = 1 \cdot p = p$.

Упражнение 0.6. Как будет устроена мера P^X , если X – это случайная величина из распределения $\text{Binomial}(n, p)$ (биномиальное распределение с параметрами n, p) ?

Упражнение 0.7. Пусть X – это случайная величина из гауссовского распределения $\mathcal{N}(0, 1)$, она абсолютно непрерывная, то есть, имеет плотность. Как будет устроена мера P^X ?

Пример 0.12. Бывает, что существует матожидание, но дисперсия бесконечна (например, распределение Парето). Также бывает, что не существует матожидания и, следовательно, дисперсии (распределение Коши при некоторых параметрах). Наконец, может быть матожидание, но дисперсии не существует (распределение Стьюдента при некоторых параметрах).

0.4 Условное матожидание

Пусть (Ω, \mathcal{F}, P) – вероятностное пространство. На основе определения условной вероятности мы можем определить условное матожидание случайной величины X при условии события $B \in \mathcal{F}$ как

$$\mathbb{E}[X | B] = \frac{\int_B X(\omega) dP}{P(B)},$$

если $P(B) > 0$. Но в случае $P(B) = 0$, очень часто, у нас возникнет проблема, поэтому нужно действовать по-другому.

Определение 0.14. Пусть (Ω, \mathcal{F}, P) – вероятностное пространство, $\mathcal{G} \subset \mathcal{F}$ – σ -алгебра. Условным матожиданием случайной величины X при условии \mathcal{G} , обозначаемое $\mathbb{E}[X | \mathcal{G}]$, называется измеримая функция $f : \Omega \rightarrow \mathbb{R}$ такая, что

$$\int_A X(\omega) dP = \int_A f(\omega) dP.$$

выполнено для всех $A \in \mathcal{G}$.

Иными словами, условное матожидание – снова случайная величина.

Пример 0.13. Проверьте, что данное определение совпадает с более ранним в случае $P(B) > 0$. При условии какой σ -алгебры будет это матожидание?

Пример 0.14. Пусть X, Y – две случайные величины. Условное матожидание $g(c) = \mathbb{E}[X | Y = c]$, обозначаемое часто более кратко как $\mathbb{E}[X | Y]$, есть не что иное, как условное матожидание при условии \mathcal{F}^Y – σ -алгебры, порожденной Y :

$$\mathbb{E}[X | Y] := \mathbb{E}[X | \mathcal{F}^Y].$$

Очень полезный инструмент для вычислений матожиданий от выражений, включающих в себя зависимые случайные величины – это формула полного матожидания.

Утверждение 0.3. Пусть (Ω, \mathcal{F}, P) – вероятностное пространство, на котором определена случайная величина $X : \Omega \rightarrow \mathbb{R}$, и $\mathcal{G} \subset \mathcal{F}$ – сигма-алгебра. Верна формула полного матожидания

$$\mathbb{E}[\mathbb{E}[X | \mathcal{G}]] = \mathbb{E}[X].$$

Эта формула позволяет вычислять достаточно неочевидные на первый взгляд матожидания.

Упражнение 0.8. Саша и Паша очень много играют в шашки и известно, что Саша выигрывает с вероятностью $p > 0$. Предположим, что их партии в шашки независимы. Какое в среднем число игр им нужно сыграть до того, как у Саши наберётся первая серия из n побед?

Ответ: $\sum_{i=1}^n \frac{1}{p^i}$.

Случайные процессы: определения

В этой лекции мы вводим понятия случайного процесса, его основных характеристик (конечномерные распределения, матожидание, ковариационная функция), стационарности в узком и широком смысле, а также обсуждаем первые примеры.

1.1 Случайный процесс

В мире достаточно большое число динамических процессов, которые имеют в себе некоторую неопределённость.

1. Датчики ЭКГ(электрокардиография), как и любые датчики выдают на выходе зашумлённые данные. Замеренное датчиком напряжение в момент времени $t \in \mathbb{R}_+$ – это X_t , случайная величина со значением в \mathbb{R} .
2. Очереди на кассах в супермаркете случайные и меняются со временем. Мы не можем знать, в какой момент к нам придут клиенты. Число клиентов X_t в момент времени $t \in \mathbb{R}_+$ – это случайная величина принимающая целые, но сколь угодно большие значения, ограниченные только размером магазина и прилегающей территории.
3. Сам клиент (с позиции руководителя магазина) – это тоже очень неопределённый субъект. Мы знаем, что он может случайно переходить по отделам, но достоверно не известно, как именно. При этом очевидно, что магазин может влиять на модель клиента путём разных управлеченческих решений (правильного размещения отделов, скидок и пр.). Для выручки магазина может быть не важно, сколько времени клиент провёл на каждом отделе, но важно, как именно он между ними перемещался. Таким образом, положение клиента X_t в момент $t \in \mathbb{Z}_+$ – это случайная величина, принимающая значения в конечном множестве $\{1, \dots, n\}$, если в магазине всего n отделов.
4. Давление в газопроводе зависит от расстояния от точки замера до источника газа очень неопределенным образом в силу засорённости трубы и её геометрии. Давление X_t в точке трубы $t \in D \subset \mathbb{R}^4$ – это случайная величина, принимающая значения в \mathbb{R}_+ .

Очевидно, что одних случайных величин становится недостаточно для описания динамики таких систем – например, положение клиента в магазине зависит от того, где он был раньше и от части может определять, где он будет потом. Поэтому на все X_t нужен

некоторый *общий* взгляд. Для начала заметим, что если мы зафиксируем время (или точку пространства, как в примере 4), то показатели (напряжение, размер очереди на кассе, давление газа, номер отдела, где находится клиент) – это случайные величины, которых очень много и при этом они ещё нетривиально зависят друг от друга.

Определение 1.1. Пусть (Ω, \mathcal{F}, P) – вероятностное пространство. Семейство случайных величин $X_t : \Omega \rightarrow \Xi$, принимающих значения в измеримом пространстве (Ξ, \mathcal{G}) , проиндексированное значениями $t \in T$ из некоторого множества T , называют *случайным процессом* и обозначают $(X_t)_{t \in T}$. Часто для краткости мы будем писать X .

В примерах 1 и 2 индексное пространство $T = \mathbb{R}_+$, в третьем примере $T = \mathbb{Z}_+$; так, индекс – это время. Тем не менее, часто индексом может быть, например, подмножество \mathbb{R}^3 , как в случае с примером 4. Пространство значений $\Xi = \mathbb{R}_+$ в примерах 1 и 4, в примере 2 это $\Xi = \mathbb{Z}$, а в третьем примере $\Xi = \{1, \dots, n\}$.

Случайный процесс обозначают как $X_t(\omega)$ или $X(t, \omega)$, таким образом явно обозначая, что его можно рассматривать как функцию $X : T \times \Omega \rightarrow \Xi$. Если мы фиксируем индекс $t = t_0$, то мы получаем $X_{t_0} : \Omega \rightarrow \Xi$, случайную величину, называемую иногда *сечением процесса в точке t_0* . Если же мы зафиксируем $\omega = \omega_0$, то $X_t(\omega_0) : T \rightarrow \Xi$ – это детерминированная функция индекса t , называемая *реализацией* или *траекторией* случайного процесса X . Иными словами, выпало событие $\omega \in \Omega$ и сразу реализовалась вся траектория процесса $X_t(\omega)$.

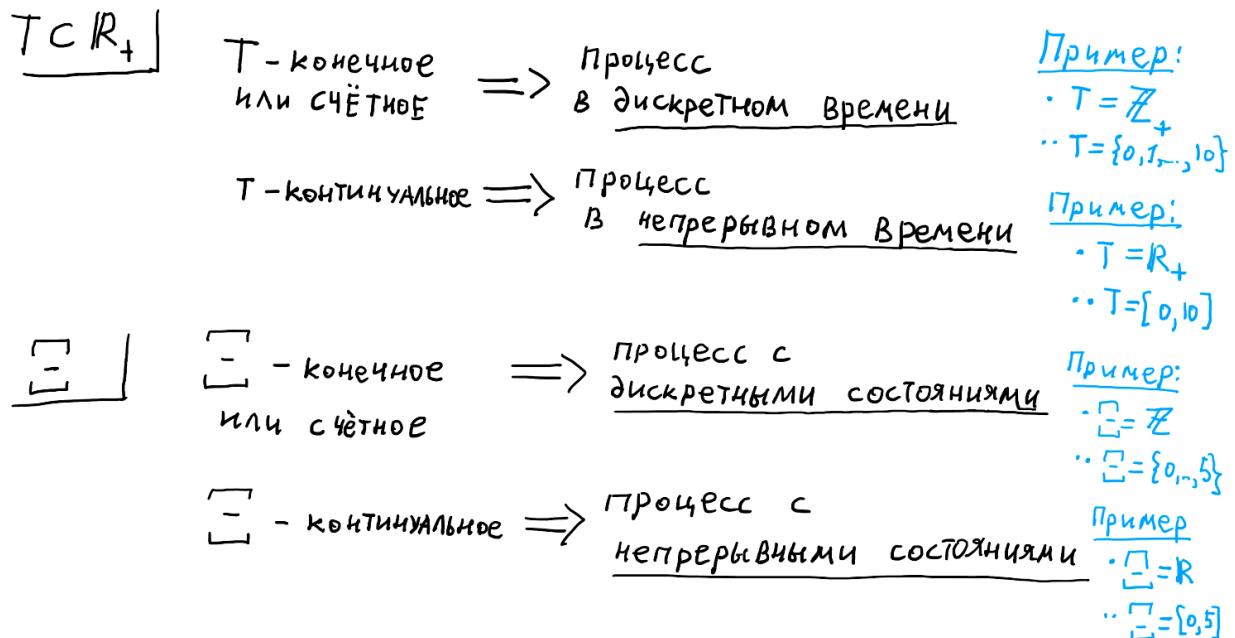


Рис. 1.1: Одна из известных классификаций случайных процессов

Для многих результатов существенно, как именно устроено индексное пространство T и множество значений Ξ . Рассмотрим одну известную классификацию (см. Рис. 1.1).

Пусть $T \subset \mathbb{R}_+$. Если T счётно (например, $T = \mathbb{Z}_+$), то говорят о процессе в *дискретном времени*; если T континуально (к примеру, $T = \mathbb{R}_+$), то X – процесс в *непрерывном времени*. Похожее разделение есть для пространства значений: если Ξ не более, чем счётно, то X – процесс с *дискретными состояниями*; если Ξ континуально, то X – процесс с *непрерывными состояниями*.

Пример 1.1. Пусть $U \in \mathcal{U}[0, 1]$, а случайный процесс задаётся $X_t = U \sin(t)$. Траекториями такого процесса будут синусы со случайными амплитудами U (См. Рис. 1.4), а сечениями будут случайные величины $X_t \sim \mathcal{U}[0, \sin t]$ для $t \in (2\pi k, \pi + 2\pi k)$, $X_t \sim \mathcal{U}(\sin t, 0)$ для $t \in (\pi + 2\pi k, 2\pi + 2\pi k)$ и $X_t = 0$ для $t = 2\pi k$.

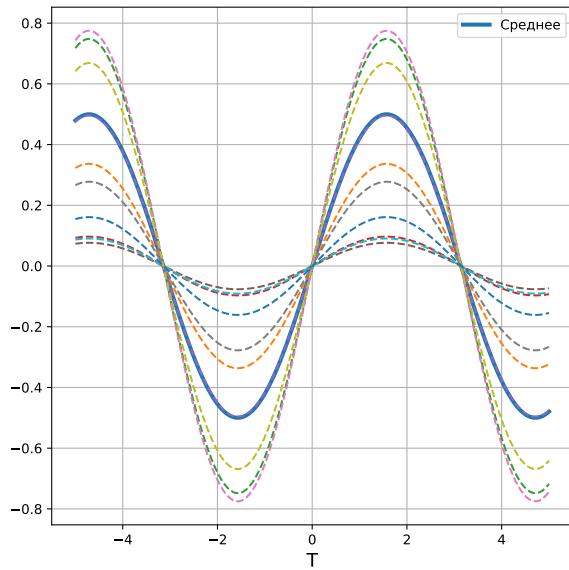


Рис. 1.2: Траектории процесса из Примера 1.1

1.2 Семейство конечномерных распределений и согласованность

Пусть (Ω, \mathcal{F}, P) – вероятностное пространство и $(X_t)_{t \in T}$ – случайный процесс на нём со значениями в измеримом пространстве (Ξ, \mathcal{G}) . Если $\tilde{\Omega}$ – пространство всех функций $\tilde{\omega} : T \rightarrow \Xi$, то мы можем определить *цилиндрическое множество*, как

$$B_{t_1, \dots, t_n}^{A_1, \dots, A_n} := \left\{ \tilde{\omega} \in \tilde{\Omega} : \tilde{\omega}(t_1) \in A_1, \dots, \tilde{\omega}(t_n) \in A_n \right\} \quad (1.1)$$

для $t_1, \dots, t_n \in T$ и $A_1, \dots, A_n \in \mathcal{G}$. Если мы определим *сигма-алгебру цилиндров* \mathcal{B} как порождённую всеми возможными цилиндрическими множествами, то мы получим, что каждый цилиндр – это измеримое множество. Более того, мы можем определить меру

каждого цилиндра как

$$\tilde{P} \left(B_{t_1, \dots, t_n}^{A_1, \dots, A_n} \right) = P \left(\{\omega \in \Omega : X_{t_1}(\omega) \in A_1, \dots, X_{t_n}(\omega) \in A_n\} \right). \quad (1.2)$$

Таким образом, мы получили, что случайный процесс X задаёт также вероятностное пространство $(\tilde{\Omega}, \mathcal{B}, \tilde{P})$ на множестве траекторий.

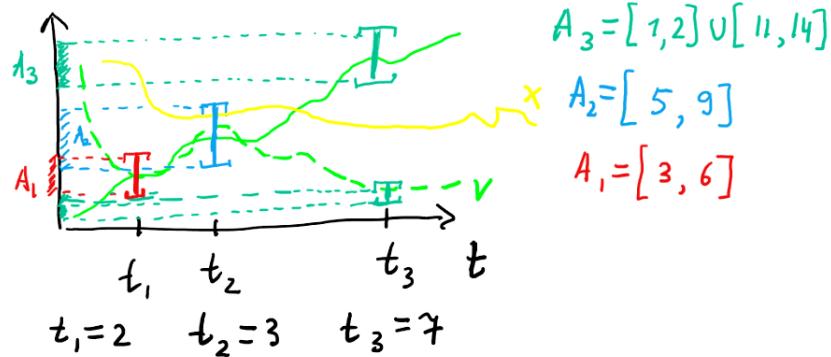


Рис. 1.3: Пример цилиндрического множества: две зелёных функции принадлежат цилинду $B_{t_1, t_2, t_3}^{A_1, A_2, A_3}$, а жёлтая – нет

Поскольку в каждый фиксированный момент X_t – случайная величина, мы можем исследовать процесс и сложные зависимости между разными величинами X_{t_1}, \dots, X_{t_n} изучая совместные распределения, которые имеют самое прямое отношение к пространству $(\tilde{\Omega}, \mathcal{B}, \tilde{P})$.

Определение 1.2. Пусть $(X_t)_{t \in T}$ – случайный процесс, определённый на вероятностном пространстве (Ω, \mathcal{F}, P) и принимающий значения в измеримом пространстве (Ξ, \mathcal{G}) и $t_1, \dots, t_n \in T$. Совместный закон распределения

$$F(t_1, \dots, t_n; A_1, \dots, A_n) = F_{t_1, \dots, t_n}(A_1, \dots, A_n) := P(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n)$$

называют n -мерным распределением в точках $t_1, \dots, t_n \in T$. Часто вводят также функцию распределения

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) := P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n).$$

Одномерных распределений оказывается недостаточно для описания случайного процесса – они ничего не говорят о зависимости случайных величин X_t в разных точках $t \in T$. Но в некоторых простых случаях их может быть достаточно чтобы задать полное семейство конечномерных распределений.

Упражнение 1.1. Выпишите конечномерные распределения для процесса из Примера 1.1.

Пример 1.2. Случайный процесс X_t состоящий из независимых в совокупности (то есть, любые k сечений X_{t_1}, \dots, X_{t_k} независимы в совокупности) и одинаково распределённых случайных величин называется белым шумом. Например, если $X_t \sim \mathcal{N}(0, 1)$ для всех $t \in T$ и любые сечения X_{t_1}, \dots, X_{t_k} независимы в совокупности, получим (стандартный) гауссовский белый шум.

Так мы задали процесс X , задав семейство всех конечномерных распределений. Одномерные распределения $F(t_1; A_1) = f(A_1)$ фиксированы, а любое n -мерное распределение будет иметь вид

$$F(t_1, \dots, t_n; A_1, \dots, A_n) = \prod_{k=1}^n P(X_{t_k} \in A_k) = \prod_{k=1}^n f(A_k)$$

в силу условия независимости.

Пример 1.3. Рассмотрим модель авторегрессии порядка 1 (обозначают $AR(1)$) с параметром α , $x_0 \in \mathbb{R}$. Это процесс в дискретном времени. Пусть $T = \mathbb{Z}_+$, $\Xi = \mathbb{R}$ и ε_t – процесс гауссовского $\mathcal{N}(0, 1)$ белого шума. Авторегрессионный процесс $(X_t)_{t \in T}$ определяется следующим образом: $X_0 = x_0$, а далее X_t подчиняется закону

$$X_{t+1} = \alpha X_t + \varepsilon_{t+1}.$$

Более общие модели авторегрессии задаются путём добавления дополнительных лагов X_{t-1}, X_{t-2}, \dots с другими коэффициентами (весами) $\alpha_{-1}, \alpha_{-2}, \dots$. В таком случае процесс $AR(p)$ задаётся рекуррентным уравнением

$$X_{t+1} = \sum_{i=0}^{p-1} \alpha_{-i} X_{t-i} + \varepsilon_{t+1}.$$

В теории временных рядов есть несколько методов, позволяющих оценивать веса лагов на основе реализации части траектории процесса X .

До сих пор мы рассматривали простые процессы: те, которые можно задать с помощью нескольких случайных величины и процесса белого шума, который в свою очередь, сам задаётся с помощью семейства одномерных распределений и требования независимости сечений. Один из известных способов задания более общего случайного процесса – задание семейства конечномерных распределений всех порядков. Инструмент для этого нам даёт известная теорема Колмогорова о существовании.

Теорема 1.1. (Теорема Колмогорова о существовании) Пусть (Ξ, \mathcal{G}) – измеримое пространство, T – некоторое множество и для всех $k \in \mathbb{Z}_+$ и $t_1, \dots, t_k \in T$ заданы меры ν_{t_1, \dots, t_k} на Ξ^k такие, что выполнены условия согласованности:

1. Для любого набора измеримых множеств $F_1, \dots, F_k \in \mathcal{G}$ и любой перестановки σ на множестве $\{1, \dots, k\}$ выполнено

$$\nu_{t_{\sigma(1)}, \dots, t_{\sigma(k)}}(F_1 \times \dots \times F_k) = \nu_{t_1, \dots, t_k}(F_{\sigma^{-1}(1)} \times \dots \times F_{\sigma^{-1}(k)}).$$

2. Для всех $k, m \in \mathbb{Z}_+$ и $t_1, \dots, t_k, t_{k+1}, \dots, t_{k+m} \in T$ выполнено

$$\nu_{t_1, \dots, t_k}(F_1 \times \dots \times F_k) = \nu_{t_1, \dots, t_{k+m}}(F_1 \times \dots \times F_k \times \Xi \times \dots \times \Xi),$$

где справа число множителей равно $k + m$.

Тогда существует вероятностное пространство (Ω, \mathcal{F}, P) и случайный процесс $(X_t)_{t \in T}$, определённый на нём, такой, что

$$\nu_{t_1, \dots, t_k}(F_1 \times \dots \times F_k) = P(X_{t_1} \in F_1, \dots, X_{t_k} \in F_k)$$

для всех $k \in \mathbb{Z}_+$, индексов $t_1, \dots, t_k \in T$ и измеримых множеств $F_1, \dots, F_k \in \mathcal{G}$.

Так, разумным образом строя семейства конечномерных распределений, мы можем корректно задать случайный процесс.

1.3 Пример: гауссовский процесс (отнесено в семинар)

Возможно, вы уже встречали гауссовые процессы в курсах по машинному обучению или анализу данных. Они являются очень мощным средством для разного рода моделирования и регрессии. Для python имплементацию можно найти, например, в пакетах `scikit-learn` и `gpy`.

Пусть $T = \mathbb{R}$, $\Xi = \mathbb{R}$. Для задания процесса нам понадобится задать функции $m : T \rightarrow \mathbb{R}$ и $K : T \times T \rightarrow \mathbb{R}$, причём последняя должна быть симметричной и удовлетворять следующему условию.

Определение 1.3. Функцию $K : T \times T \rightarrow \mathbb{R}$ назовём положительно полуопределённой, если для любого набора $t_1, \dots, t_k \in T$ и любых ненулевых векторов $x \in \mathbb{R}^k$

$$x^\top B x \geq 0,$$

где матрица $B \in \mathbb{R}^{k \times k}$ имеет элементы $B_{ij} = K(t_i, t_j)$

Иными словами, мы требуем, чтобы любая матрица B , составленная как выше, была ковариационной матрицей случайного вектора для любого набора индексов.

Мы задаём гауссовский процесс $(X_t)_{t \in T}$ с матожиданием $m : T \rightarrow \mathbb{R}$ и ковариационной функцией $K : T \times T \rightarrow \mathbb{R}$ полагая, что для любого набора индексов $t_1, \dots, t_k \in T$ случайный вектор $[X_{t_1}, \dots, X_{t_k}]^\top$ имеет гауссовское распределение с матожиданием $[m(t_1), \dots, m(t_k)]$ и ковариационной матрицей $B = (K(t_i, t_j))_{j=1, \dots, k}$. Набор гауссовых распределений удовлетворяет условиям согласованности (проверьте!) в Теореме 1.1, следовательно, существует и случайный процесс с конечномерными гауссовскими распределениями, заданными выше.

В качестве простейшей ковариационной функции можно использовать, например, линейную или квадратично-экспоненциальную функцию с параметром $l \in \mathbb{R}$ (который можно настраивать методом максимального правдоподобия в задаче регрессии):

$$K(t, s) = ts, \quad K(t, s) = e^{-\frac{(t-s)^2}{2l^2}}. \quad (1.3)$$

Упражнение 1.2. Как выглядят функции $K_{wn}(t, s)$, $m(t)$, которые задают процесс гауссовского белого шума из Примера 1.2?

Упражнение 1.3. Приведите пример функции $K : T \times T \rightarrow \mathbb{R}$, которая является симметричной, но не положительно полуопределённой.

Упражнение 1.4. Приведите пример функции $K : T \times T \rightarrow \mathbb{R}$, которая является несимметричной, но положительно полуопределённой.

Гауссовские процессы – очень гибкий инструмент, потому что можно использовать самые разные ковариационные функции и моделировать очень сложные зависимости.

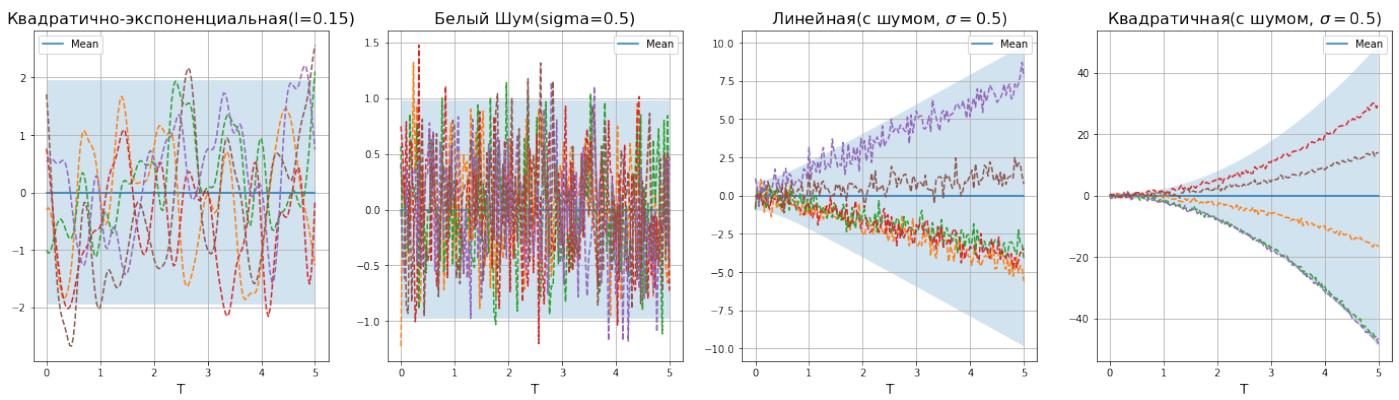


Рис. 1.4: Примеры траекторий гаусsovских процессов с разными ковариационными функциями: квадратично-экспоненциальной, белошумной $K_{wn}(t, s)$, линейной с шумом $K(t, s) = ts + K_{wn}(t, s)$ и квадратичной с шумом $K(t, s) = (ts)^2 + K_{wn}(t, s)$

1.4 Числовые характеристики и стационарность

Пусть процесс X принимает значения в $\Xi \subset \mathbb{R}$ или $\Xi \subset \mathbb{R}^n$. Мы уже познакомились с двумя характеристиками, когда обсуждали гауссовские процессы. Определим их явно и добавим ещё несколько.

Определение 1.4. Пусть $(X_t)_{t \in T}$ – случайный процесс со значениями в \mathbb{R} или \mathbb{R}^n .

Функция $m(t) := \mathbb{E}[X_t]$ называется матожиданием процесса X .

Функция $K(t, s) := \text{Cov}(X_t, X_s) = \mathbb{E}[(X_t - m(t))(X_s - m(s))^\top]$ называется ковариационной функцией.

Функция $v(t) := \text{Tr}(K(t, t))$, где Tr – это след матрицы, называется дисперсией процесса X .

Упражнение 1.5. Чему будут равны матожидания и ковариационные функции процессов из примеров 1.1, 1.3?

В случае гауссовых процессов мы задали целый случайный процесс используя только матожидание и ковариационную функцию. Причём оказывается, что свойство положительной полуопределённости (Опр. 1.3) является существенным.

Утверждение 1.2. (Упражнение) Функция $K(t, s)$ является ковариационной функцией некоторого процесса тогда и только тогда, когда она является симметричной и положительно полуопределённой в смысле Опр. 1.3.

Пусть $(X_t)_{t \in T}$ – случайный процесс, определённый на вероятностном пространстве (Ω, \mathcal{F}, P) и с векторным пространством T в качестве индексного (чтобы мы могли делать сдвиги на элемент $s \in T$).

Определение 1.5. Процесс X называется стационарным в узком смысле (*strongly stationary*), если для любого конечного набора индексов t_1, \dots, t_k конечномерные распределения не меняются со сдвигом всего набора на $s \in T$, то есть,

$$P(X_{t_1} \in A_1, \dots, X_{t_k} \in A_k) = P(X_{t_1+s} \in A_1, \dots, X_{t_k+s} \in A_k)$$

для любых измеримых множеств $A_1, \dots, A_k \in \mathcal{G}$.

Определение 1.6. Процесс X называется стационарным в широком смысле (*weakly stationary*), если

1. Существуют конечные матожидание $m(t)$ и дисперсия $v(t)$ в каждой точке $t \in T$;
2. Матожидание – константа: $m(t) = \text{const}$;
3. Ковариационная функция зависит только от разности аргументов: существует функция $\tilde{K} : T \rightarrow \mathbb{R}$ такая, что $K(t, s) = \tilde{K}(|t - s|) < \infty$ для любых $t, s \in T$.

Пример 1.4. В общем случае из стационарности в узком смысле следует стационарность в широком смысле (докажите, для этого нужно выписать ковариационные функции $K(t, t + s)$ и $K(t - s, t)$ и показать, что они равны для любых $t, s \in T$, используя определение стационарности в узком смысле). В обратную сторону, как оказывается,

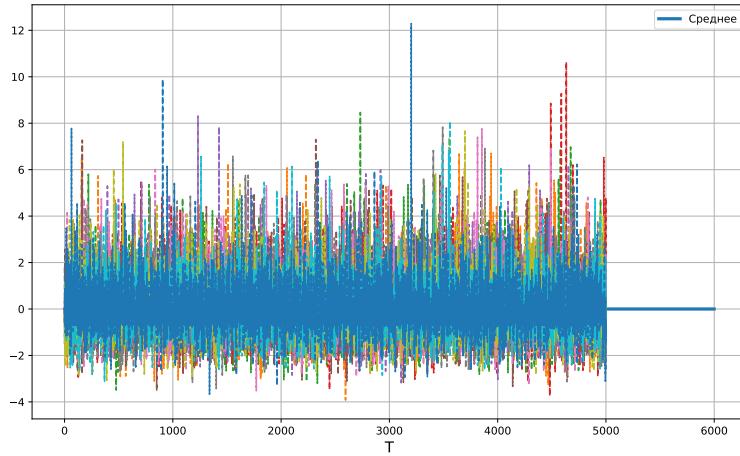


Рис. 1.5: Траектории процесса из Примера 1.4

утверждение неверно – существуют процессы, которые стационарны в широком смысле, но не стационарны в узком. Пусть $T = \mathbb{Z}$, $\Xi = \mathbb{R}$, ε_t – процесс белого шума с распределением $\mathcal{N}(0, 1)$. Процесс X_t , определённый как

$$X_t = \begin{cases} \varepsilon_t, & t \text{ – чётное}, \\ \frac{1}{\sqrt{2}}(\varepsilon_t^2 - 1), & t \text{ – нечётное}, \end{cases}$$

не будет стационарным в узком смысле, но будет стационарным в широком смысле.

Пример 1.5. Гауссовские процессы интересны ещё и тем, что для них стационарность в широком смысле влечёт стационарность в узком смысле (проверьте, это следует напрямую из того, что совместные распределения всех конечных наборов сечений гауссовые).

Упражнение 1.6. Проверьте, являются ли процесса из Примера 1.1 и гауссовский белый шум из Примера 1.2 стационарными в узком или широком смыслах?

Упражнение 1.7. Пусть $T \in \mathbb{R}$, $(X_t)_{t \in T}$ – гауссовский процесс с $m(t) = 0$. Проверьте, будет ли процесс стационарным (в широком смысле), если ковариационная функция

1. линейная $K(t, s) = ts$;
2. квадратично-экспоненциальная $K(t, s) = e^{-\frac{(t-s)^2}{2l^2}}$ с параметром $l \in \mathbb{R}$.

Винеровский процесс

В этой лекции мы построим и исследуем один из самых важных и самых известных процессов в непрерывном времени в финансовой математике: Винеровский процесс. Винеровский процесс – это базовый блок построения финансовых моделей, с его помощью определяются строятся стохастические интегралы и стохастические дифференциальные уравнения. Мы начинаем с интересного физического примера – Броуновского движения – и затем с помощью теоремы Колмогорова о существовании строим случайный процесс, который помогает описать это явление, и доказываем несколько его полезных свойств. По пути мы рассмотрим вопрос о непрерывности траекторий Винеровского процесса и заново посмотрим на гауссовские процессы, которые мы исследовали в лекции 1.

2.1 Броуновское движение

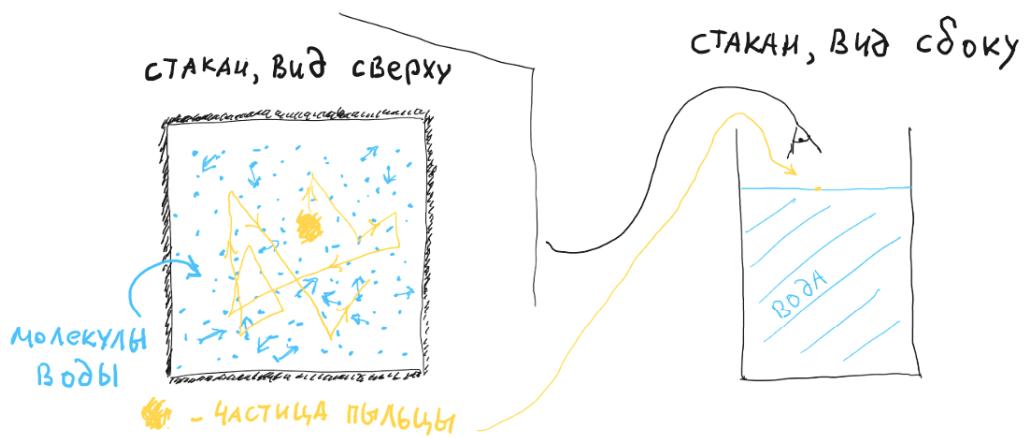


Рис. 2.1: Хаотичный процесс движений частицы пыльцы на поверхности воды

Броуновское движение – известное физическое явление, про которое рассказывают ещё в школьном курсе физики. История его уходит в 19й век, когда в 1828г. Роберт Браун обратил внимание на необычное хаотическое движение частичек пыльцы в жидкости. Интуитивное объяснение данного явления достаточно простое: молекулы жидкости находятся в постоянном движении и случайным образом соударяются, а частица пыльцы значительно больше их и обладает большей инерцией. Молекулы жидкости, таким образом, подталкивают пыльцу случайным образом в разных направлениях. Чем меньше частица, тем быстрее и хаотичнее она движется (Рис. 2.1). Несмотря на то, что про Броуновское движение написано даже в школьных учебниках по физике, его математическую

модель долгое время было очень непросто построить и это потребовало много времени и усилий со стороны как физиков, так и математиков, так как в 19м веке инструментарий был принципиально другой.

Броуновское движение описывается с помощью Винеровского процесса, названного в честь Норберта Винера, который занимался этой задачей почти через 100 лет после открытия Роберта Брауна. Как нам известно из лекции 2, мы можем задавать случайный процесс путём задания измеримого пространства значений (Ξ, \mathcal{G}) , индексного пространства T и семейства мер ν_{t_1, \dots, t_k} на (Ξ^k, \mathcal{G}^k) для всех конечных наборов $t_1, \dots, t_k \in T$. Последние будут конечномерными распределениями итогового процесса.

Шаг 1. Для начала нужно задать пространство состояний, индексов и определить семейство мер, на основе которого мы построим случайный процесс. Рассматриваем одномерную модель, многомерный случай строится аналогично, об этом есть заметка в конце лекции. Фиксируем $\Xi = \mathbb{R}$ и $\mathcal{G} = \mathcal{B}(\mathbb{R})$, борелевскую сигма-алгебру, а также индексное пространство $T = \mathbb{R}_+$, которое будет выполнять роль времени. Зададим также функцию $p : T \times \Xi \times \Xi \rightarrow \mathbb{R}$ для $t > 0$ как

$$p(t, x, y) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}. \quad (2.1)$$

Заметьте, что при фиксированных t, x или при фиксированных t, y функция p является плотностью гауссовского распределения $\mathcal{N}(0, t)$. Фиксируем число $\tilde{x} \in \mathbb{R}$ и теперь зададим для любого конечного отсортированного по возрастанию набора $t_1, \dots, t_k \in T$ меры ν_{t_1, \dots, t_k} как

$$\nu_{t_1, \dots, t_k}(A_1 \times \dots \times A_k) := \int_{A_1 \times \dots \times A_k} p(t_1, \tilde{x}, x_1) p(t_2 - t_1, x_1, x_2) \dots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \dots dx_k. \quad (2.2)$$

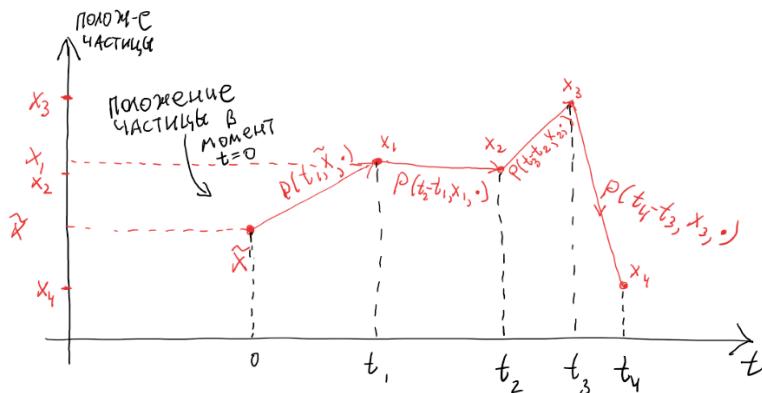


Рис. 2.2: При условии текущего положения x_k следующее x_{k+1} будет иметь нормальное распределение в соответствии с функцией p .

Теперь обобщим на случай неотсортированного набора. Пусть t_1, \dots, t_k – отсортированный набор, как выше, σ – перестановка $\{1, \dots, k\}$, зададим произвольную неотсортированную последовательность как $t_{\sigma(1)}, \dots, t_{\sigma(k)}$ и определим её меру как

$$\nu_{t_{\sigma(1)}, \dots, t_{\sigma(k)}}(A_{\sigma(1)} \times \dots \times A_{\sigma(k)}) := \nu_{t_1, \dots, t_k}(A_1 \times \dots \times A_k), \quad (2.3)$$

то есть, мера, соответствующая неотсортированному набору, – это мера, соответствующая отсортированному набору тех же индексов.

Шаг 2. Положим набор индексов отсортированным; заметим, что случай неотсортированного множества индексов можно обработать с помощью (2.3). Фиксируем произвольный(отсортированный) набор $t_1, \dots, t_k \in T$ и измеримых множеств $A_1, \dots, A_k \in \mathcal{G}$ и положим, что σ – это перестановка на множестве $\{1, \dots, k\}$. Заметим теперь, что

$$\nu_{t_1, \dots, t_k}(A_{\sigma(1)} \times \dots \times A_{\sigma(k)}) := \int_{A_{\sigma(1)} \times \dots \times A_{\sigma(k)}} p(t_1, \tilde{x}, x_1) p(t_2 - t_1, x_1, x_2) \dots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \dots dx_k,$$

где мы поменяли порядок интегрирования. Так, теперь мы берём интеграл по области, где $x_1 \in A_{\sigma(1)}, \dots, x_k \in A_{\sigma(k)}$. Это эквивалентно тому, что мы поменяем местами переменные с помощью σ^{-1} , а множества менять местами не будем. В итоге имеем

$$\nu_{t_{\sigma^{-1}(1)}, \dots, t_{\sigma^{-1}(k)}}(A_1 \times \dots \times A_k) = \nu_{t_1, \dots, t_k}(A_{\sigma(1)} \times \dots \times A_{\sigma(k)}),$$

подтверждая первое условие согласованности.

Шаг 3. Фиксируем произвольный(отсортированный) набор $t_1, \dots, t_k, t_{k+1} \in T$, нужно проверить, что

$$\nu_{t_1, \dots, t_k, t_{k+1}}(A_1 \times \dots \times A_k \times \mathbb{R}) = \nu_{t_1, \dots, t_k}(A_1 \times \dots \times A_k).$$

Используя конкретный вид меры $\nu_{t_1, \dots, t_k, t_{k+1}}$, мы можем вычислить последний интеграл:

$$\nu_{t_1, \dots, t_k, t_{k+1}}(A_1 \times \dots \times A_k \times \mathbb{R}) = \quad (2.4)$$

$$= \int_{A_1} \dots \int_{A_k} \int_{\mathbb{R}} p(t_1, \tilde{x}, x_1) \dots p(t_k - t_{k-1}, x_k, x_{k+1}) dx_1 \dots dx_k dx_{k+1} = \quad (2.5)$$

$$= \int_{A_1} \dots \int_{A_k} p(t_1, \tilde{x}, x_1) \dots p(t_k - t_{k-1}, x_{k-1}, x_k) \left\{ \int_{\mathbb{R}} p(t_k - t_{k-1}, x_k, x_{k+1}) dx_{k+1} \right\} dx_1 \dots dx_k. \quad (2.6)$$

Заметим, что интеграл в скобках – это интеграл по \mathbb{R} от плотности гауссовского распределения и он равен 1 для любых фиксированных $t_k - t_{k-1}, x_k$. Всё, что остаётся, – это и есть желаемый результат:

$$\nu_{t_1, \dots, t_k, t_{k+1}}(A_1 \times \dots \times A_k \times \mathbb{R}) = \quad (2.7)$$

$$= \int_{A_1} \dots \int_{A_k} p(t_1, \tilde{x}, x_1) \dots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \dots dx_k = \quad (2.8)$$

$$= \nu_{t_1, \dots, t_k}(A_1 \times \dots \times A_k). \quad (2.9)$$

Итак, второе условие согласованности тоже выполнено.

Согласно Теореме 1.1, существует вероятностное пространство (Ω, \mathcal{F}, P) и случайный процесс $(W_t)_{t \in T}$ на нём с конечномерными распределениями ν_{t_1, \dots, t_k} , заданными выше. Построенный процесс $(W_t)_{t \in T}$ называется одномерным Винеровским процессом (другое название – Броуновское движение) с началом в точке \tilde{x} .

Такой процесс уже обладает несколькими интересными свойствами.

Утверждение 2.1. Процесс $(W_t)_{t \in T}$ – гауссовский.

▷ Рассмотрим конечный набор сечений W_{t_1}, \dots, W_{t_k} , запишем его в виде вектора $Z = [W_{t_1} \dots W_{t_k}]^\top$. Для доказательства гауссности можно посчитать характеристическую функцию $f(\xi)$ и сравнить её с гауссовой.

$$f(\xi) := \mathbb{E} [e^{i\xi^\top Z}] = \int_{\mathbb{R}^k} e^{i\xi^\top x} p(t_1, \tilde{x}, x_1) \dots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \dots dx_k, \quad x := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}, \xi := \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_k \end{bmatrix}. \quad (2.10)$$

Запишем интеграл в виде повторного и рассмотрим самый последний интеграл:

$$\int_{\mathbb{R}} e^{i\xi_k x_k} p(t_k - t_{k-1}, x_{k-1}, x_k) dx_k;$$

это по определению – характеристическая функция распределения $\mathcal{N}(x_{k-1}, t_k - t_{k-1})$, которая нам известна:

$$\int_{\mathbb{R}} e^{i\xi_k x_k} p(t_k - t_{k-1}, x_{k-1}, x_k) dx_k = e^{ix_{k-1}\xi_k - \frac{1}{2}(t_k - t_{k-1})\xi_k^2}. \quad (2.11)$$

Будем вычислять повторный интеграл далее, получим снова, что интеграл – это характеристическая функция гауссского распределения от аргумента $\xi_{k-1} + \xi_k$ и с другими параметрами:

$$e^{-\frac{1}{2}(t_k - t_{k-1})\xi_k^2} \int_{\mathbb{R}} e^{ix_{k-1}(\xi_{k-1} + \xi_k)} p(t_{k-1} - t_{k-2}, x_{k-2}, x_{k-1}) dx_{k-1} = \quad (2.12)$$

$$= e^{ix_{k-2}(\xi_{k-1} + \xi_k) - \frac{1}{2}(t_k - t_{k-1})\xi_k^2 - \frac{1}{2}(t_{k-1} - t_{k-2})(\xi_{k-1} + \xi_k)^2}. \quad (2.13)$$

По индукции мы получим, что

$$\mathbb{E} [e^{i\xi^\top Z}] = e^{i\tilde{x}(\xi_1 + \dots + \xi_k) - \frac{1}{2}\xi^\top C \xi} = e^{im^\top \xi - \frac{1}{2}\xi^\top C \xi}, \quad (2.14)$$

где (проверьте!)

$$m = [\tilde{x} \dots \tilde{x}]^T \in \mathbb{R}^k, \quad C = \begin{bmatrix} t_1 & t_1 & \dots & \dots & t_1 \\ t_1 & t_2 & t_2 & \dots & t_2 \\ \dots & \dots & \dots & \dots & \dots \\ t_1 & t_2 & t_3 & \dots & t_k \end{bmatrix} \in \mathbb{R}^{k \times k}.$$

Это в точности характеристическая функция многомерного гауссовского распределения с матожиданием t и ковариационной матрицей C . Таким образом, $(W_t)_{t \in T}$ – гауссовский процесс с матожиданием $\mathbb{E}[W_t] = \tilde{x}$ и ковариационной функцией $K(t, s) = t \wedge s := \min(t, s)$, которую мы можем видеть в матрице C . \square

Следствие 2.2. $W_0 = \tilde{x}$ почти наверное.

Упражнение 2.1. Докажите, что матрица C – положительно полуопределённая, следовательно, функция $K(t, s) = t \wedge s$ является положительно полуопределённой в смысле Опр. 1.3. В доказательстве мы этого не проверили, но это требуется для того, чтобы понять, что характеристическая функция действительно соответствует гауссовскому распределению.

Таким образом, ещё одно определение Винеровского процесса может звучать так.

Определение 2.1. Винеровский процесс, начинающийся в точке $\tilde{x} \in \mathbb{R}$ – это гауссовский процесс $(W_t)_{t \in T}$ на индексном пространстве $T = \mathbb{R}_+$ со значениями в $\Xi = \mathbb{R}$, который имеет матожидание $\mathbb{E}[W_t] = \tilde{x}$ и ковариационную функцию $K(t, s) = t \wedge s := \min(t, s)$.

У Винеровского процесса есть ещё одно свойство которое является очень полезным для симуляции траекторий.

Определение 2.2. Случайный процесс $(X_t)_{t \in \mathbb{R}_+}$ называется процессом с независимыми приращениями, если для любого конечного отсортированного по возрастанию набора $t_1, \dots, t_k \in T$ случайные величины

$$X_{t_k} - X_{t_{k-1}}, \dots, X_{t_2} - X_{t_1}$$

являются независимыми в совокупности.

Утверждение 2.3. Винеровский процесс $(W_t)_{t \in T}$ – это процесс с независимыми приращениями.

▷ Поскольку $(W_t)_{t \in T}$ – гауссовский процесс, приращения тоже будут иметь нормальное распределение (в том числе, совместное). Для проверки независимости, таким образом, достаточно показать, что ковариация любых двух приращений равна нулю. Мы знаем точное выражение для ковариации $K(t, s) = \mathbb{E}[W_t W_s] = t \wedge s$, используя его, получим для $t_2 > t_1 > t_0$

$$\begin{aligned} \mathbb{E}[(W_{t_2} - W_{t_1})(W_{t_1} - W_{t_0})] &= \mathbb{E}[W_{t_2} W_{t_1}] + \mathbb{E}[W_{t_1} W_{t_0}] - t_1 - \mathbb{E}[W_{t_2} W_{t_0}] = \\ &= t_1 + t_0 - t_1 - t_0 = 0. \end{aligned}$$

\square

Это утверждение позволяет сконструировать гораздо более экономный метод для симуляции процесса W , чем тот, что мы используем для гауссовых процессов. Положим, что мы хотим сгенерировать (дискретизированную) траекторию процесса в заданные моменты времени t_1, \dots, t_k , тогда наш алгоритм будет выглядеть так:

1. Задать стартовое значение $w_0 = \tilde{x}$;
2. Сгенерировать дискретизированную траекторию процесса гауссовского белого шума $\mathcal{N}(0, 1)$: ξ_1, \dots, ξ_k ;
3. Для $j = 1, \dots, k$ задать $w_{t_j} = w_{t_{j-1}} + \sqrt{t_j - t_{j-1}}\xi_j$;
4. Набор $w_0, w_{t_1}, \dots, w_{t_k}$ – это дискретизированная траектория Винеровского процесса.

В отличие от симуляции гауссовского процесса, нам не требуется работать с ковариационной матрицей и метод требует всего лишь $O(k)$ операций вместо как минимум $O(k^2)$.

2.2 Непрерывность траекторий

До сих пор мы не очень интересовались свойствами траекторий случайного процесса, хотя в лекции 2 мы видели, что некоторые гауссовские процессы чудесным образом обладали непрерывными (и даже гладкими) траекториями. Заметим: теорема Колмогорова (Теорема 1.1) позволяет задать случайный процесс на каком-то вероятностном пространстве, но не утверждает ничего о свойствах траекторий процесса. Дело в том, что в результате теоремы говорится о *существовании* вероятностного пространства и случайного процесса на нём (имеющего заданные конечномерные распределения), но ничего про *единственность*. По этой причине существуют случайные процессы, которые имеют одинаковые конечномерные распределения, но свойства траекторий могут отличаться.

Определение 2.3. Пусть $(X_t)_{t \in T}$ и $(Y_t)_{t \in T}$ – два случайных процесса на вероятностном пространстве (Ω, \mathcal{F}, P) . Если для всех $t \in T$

$$P(\{\omega \in \Omega : X_t(\omega) = Y_t(\omega)\}) = 1,$$

то процесс X называется модификацией (в англ. литературе *version*) процесса Y .

Пример 2.1. Рассмотрим вероятностное пространство $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \mu)$, где μ – вероятностная мера, которая удовлетворяет условию $\mu(\{\omega\}) = 0 \forall \omega \in \mathbb{R}_+$. Зададим индексное пространство $T = \mathbb{R}_+$ и определим два простых случайных процесса :

$$Y_t(\omega) = 0 \quad \forall t \in \mathbb{R}_+, \quad \omega \in \mathbb{R}_+,$$

$$X_t(\omega) = \begin{cases} 1, & \text{если } t = \omega, \\ 0, & \text{иначе.} \end{cases}$$

Заметьте, что X и Y в данном случае являются модификациями друг друга. Действительно, если мы возьмём фиксированный $t \in \mathbb{R}_+$, то

$$P(X_t = Y_t) = 1,$$

так как отличаться они будут лишь в одной точке $\omega = t$. При этом мы видим, что Y имеет непрерывные траектории (константа 0), а X всегда будет иметь разрыв в одной точке.

Утверждение 2.4. *Если процесс X – модификация процесса Y то два процесса имеют одинаковые конечномерные распределения.*

▷ Фиксируем моменты t_1, \dots, t_k и рассмотрим соответствующее конечномерное распределение процесса X для произвольных измеримых A_1, \dots, A_k . Пусть событие

$$G := \{\omega \in \Omega : X_{t_1}(\omega) \in A_1, \dots, X_{t_k}(\omega) \in A_k \text{ и есть } j \text{ такой, что } Y_{t_j}(\omega) \neq X_{t_j}(\omega)\},$$

то есть, X_{t_i} лежат в A_i и при этом не все Y_{t_i} совпадают с X_{t_i} . Заметим, что

$$P(\{\omega \in \Omega : X_{t_1}(\omega) \in A_1, \dots, X_{t_k}(\omega) \in A_k\}) = \quad (2.15)$$

$$= P(\{\omega \in \Omega : Y_{t_1}(\omega) \in A_1, \dots, Y_{t_k}(\omega) \in A_k\}) + P(G). \quad (2.16)$$

Так как

$$P(G) \leq P\left(\bigcup_{j=1}^k \{\omega \in \Omega : Y_{t_j}(\omega) \neq X_{t_j}(\omega)\}\right) = 0$$

в силу того, что X – модификация Y , это означает, что

$$P(\{\omega \in \Omega : X_{t_1}(\omega) \in A_1, \dots, X_{t_k}(\omega) \in A_k\}) = \quad (2.17)$$

$$= P(\{\omega \in \Omega : Y_{t_1}(\omega) \in A_1, \dots, Y_{t_k}(\omega) \in A_k\}), \quad (2.18)$$

то есть, конечномерные распределения процессов будут совпадать. \square

Если мы попробуем просимулировать Винеровский процесс, то мы увидим, что траектории очень похожи на непрерывные. В некотором конкретном смысле это действительно так. Это следует из второй известной теоремы Колмогорова.

Теорема 2.5. (*Теорема Колмогорова о непрерывности*) Пусть $(X_t)_{t \in \mathbb{R}_+}$ – случайный процесс на вероятностном пространстве (Ω, \mathcal{F}, P) со значениями в $\Xi = \mathbb{R}$. Если для всех $T > 0$ существуют такие константы $\alpha, \beta, D > 0$, что

$$\forall t, s \in [0, T] \quad \mathbb{E}[|X_t - X_s|^\alpha] \leq D|t - s|^{1+\beta},$$

то существует модификация $(\tilde{X}_t)_{t \in \mathbb{R}_+}$ с непрерывными траекториями.

Если посмотреть немного по-другому, это означает, что в каждой точке $t \in \mathbb{R}_+$ почти наверное траектория процесса X будет непрерывной. Винеровский процесс, как оказывается, удовлетворяет этой теореме.

Утверждение 2.6. Существует модификация Винеровского процесса $(W_t)_{t \in T}$ с непрерывными траекториями.

▷ Возьмём $\alpha = 4$, положим $t > s$ и рассмотрим

$$\mathbb{E}[(W_t - W_s)^4] = \mathbb{E}[W_t^4 - 4W_t^3W_s + 6W_t^2W_s^2 - 4W_tW_s^3 + W_s^4].$$

Все величины выше – гауссовские и для их моментов есть формула Иссерлиса (см. конец лекции):

$$\begin{aligned}\mathbb{E}[W_t^4] &= 3(\mathbb{E}[W_t^2])^2 = 3t^2, \quad \mathbb{E}[W_t^3W_s] = 3\mathbb{E}[W_t^2]\mathbb{E}[W_tW_s] = 3ts, \\ \mathbb{E}[W_t^2W_s^2] &= \mathbb{E}[W_t^2]\mathbb{E}[W_s^2] + 2(\mathbb{E}[W_tW_s])^2 = ts + 2s^2, \quad \mathbb{E}[W_s^4] = 3s^2, \quad \mathbb{E}[W_tW_s^3] = 3s^2.\end{aligned}$$

В результате имеем

$$\mathbb{E}[(W_t - W_s)^4] = 3t^2 - 12ts + 6(ts + 2s^2) - 12s^2 + 3s^2 = 3(t-s)^2.$$

Так, с константами $\alpha = 4, \beta = 1, D = 3$ Винеровский процесс удовлетворяет Теореме 2.5 и, следовательно, существует его непрерывная модификация. \square

Упражнение 2.2. Проверьте с помощью теоремы Колмогорова о непрерывности, что траектории одномерного гауссовского процесса с линейной ковариационной функцией $K(t, s)$ почти наверное непрерывны.

Пример 2.2. Рассмотрим гауссовский белый шум $(X_t)_{t \in T}$ с распределением $\mathcal{N}(0, \sigma^2)$. Взгляд на симулированные траектории позволяет предположить, что они в каждой точке будут почти наверное разрывными. Это действительно так. Мы можем заметить, что для всех $\delta > 0$ и t, s таких, что $|t - s| < \delta$, разность $|X_t - X_s| > 0$ с вероятностью 1. Формально, однако, это потребует доказательства; приведём идею, опирающуюся на лемму Бореля-Кантелли.

Лемма утверждает следующее. Пусть (Ω, \mathcal{F}, P) – вероятностное пространство, (E_n) – последовательность измеримых множеств $E_n \in \mathcal{F}$. Если ряд $\sum_{k=1}^{\infty} P(E_k)$ сходится, то

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right) = 0,$$

то есть, вероятность того, что бесконечное число событий из последовательности сбываются одновременно, равна нулю, если события E_n достаточно малые (согласно мере P).

Пойдём от противного: предположим, что в фиксированной точке t траектория процесса X с вероятностью больше 0 непрерывна. Это требование эквивалентно тому, что с вероятностью строго больше 0

$$\forall \varepsilon > 0 \exists \delta(\varepsilon) : |t - t'| < \delta \Rightarrow |X_t(\omega) - X_{t'}(\omega)| < \varepsilon.$$

Возьмём последовательность $\varepsilon_n = 1/n^2$ и $\delta_n = \delta(\varepsilon_n)/2$, а также произвольные точки $t_n \in [t - \delta_n, t + \delta_n]$. По нашему предположению, с вероятностью больше 0 условие непрерывности соблюдается и для нашего выбора $\varepsilon_n, \delta_n, t_n$, значит

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{\omega \in \Omega : |X_t - X_{t_k}| < \varepsilon_k\}\right) > 0.$$

Теперь обратимся к лемме Бореля-Кантелли: пусть

$$E_n := \{\omega \in \Omega : |X_t - X_{t_n}| < \varepsilon_n\}.$$

Поскольку разность $X_t - X_{t_n}$ – это гауссовская случайная величина с дисперсией 2, можно оценить (см. Рис. 2.3)

$$P(|X_t - X_{t_n}| < \varepsilon_n) \leq \frac{2\varepsilon_n}{2\sqrt{\pi}} = \frac{\varepsilon_n}{\sqrt{\pi}}.$$

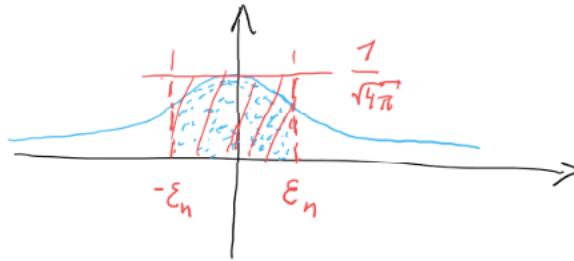


Рис. 2.3: Идея оценки вероятности события E_n : плотность нормального распределения имеет максимум в нуле, равный $1/(\sqrt{4\pi})$

Пяд

$$\sum_{n=1}^{\infty} P(E_n) = \sum_{n=1}^{\infty} \frac{\varepsilon_n}{\sqrt{\pi}} = \sum_{n=1}^{\infty} \frac{1}{n^2 \sqrt{\pi}},$$

как нам известно, будет сходить к 0 по лемме Бореля-Кантелли это означает, что

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{\omega \in \Omega : |X_t - X_{t_k}| < \varepsilon_k\}\right) = 0,$$

что противоречит предположению.

Правда, ничего лучше в плане гладкости про Винеровский процесс доказать не выйдет.

Теорема 2.7. Траектории Винеровского процесса $(W_t)_{t \in \mathbb{R}_+}$ почти наверное нигде не дифференцируемы.

Эту теорему мы приводим без доказательства, так как оно достаточно техническое и опирается на лемму Бореля-Кантелли и похоже на приведённое выше. Рекомендуем, если интересно, посмотреть [25, Гл.3, Теор.1].

2.3 Многомерный Винеровский процесс

Построим теперь многомерный Винеровский процесс как случайный векторный процесс с независимыми координатами, каждая из которых является одномерным Винеровским процессом (в \mathbb{R}^3 он вполне будет описывать физическое Броуновское движение). Все результаты, которые были приведены выше, оказываются верны и в многомерном случае.

Зададим стартовое положение $\tilde{x} \in \mathbb{R}^n$, индексное пространство $T = \mathbb{R}_+$, пространство значений $\Xi = \mathbb{R}^n$ и сигма-алгебру \mathcal{G}^n как наименьшую сигма-алгебру, включающую в себя все декартовые произведения $A_1 \times \dots \times A_n$ для всех $A_1, \dots, A_n \in \mathcal{G} = \mathcal{B}(\mathbb{R})$. Так же, как и в начале, зададим функцию $p : T \times \Xi \times \Xi \rightarrow \mathbb{R}$ как

$$p(t, x, y) := \frac{1}{(\sqrt{2\pi t})^n} e^{-\frac{\|x-y\|_2^2}{2t^n}}. \quad (2.19)$$

При фиксированных t, x и t, y это плотность некоррелированного гауссовского вектора.

Упражнение 2.3. Проверьте выполнение условий согласованности из Теоремы Колмогорова о существовании (Теорема 1.1).

Процесс $(W_t)_{t \in T}$, построенный с помощью теоремы Колмогорова о существовании, называется многомерным Винеровским процессом. Как и одномерный Винеровский, он будет гауссовским, но в отличие от гауссовых процессов из лекции 2, будет принимать значения в $\Xi = \mathbb{R}^n$ и описание его ковариационной функции будет немного более сложным. Многомерный Винеровский процесс $(W_t)_{t \in T}$ – это вектор состоящий из независимых одномерных Винеровских процессов (так как мы строили его именно таким образом), поэтому все теоремы, которые мы доказывали, верны и здесь (как, например, утверждение о независимости приращений), но требуют небольшого уточнения.

Теорема 2.8. (Теорема Колмогорова о непрерывности, векторный вариант). Пусть $(X_t)_{t \in \mathbb{R}_+}$ – это случайный процесс со значениями в $\Xi = \mathbb{R}^n$. Если для любого $T > 0$ существуют такие константы $\alpha, D, \beta > 0$, что

$$\forall t, s < T \quad \mathbb{E} [\|X_t - X_s\|_2^\alpha] \leq D \|t - s\|_2^{1+\beta},$$

то существует непрерывная версия $(\tilde{X}_t)_{t \in T}$ процесса X .

Следствие 2.9. Многомерный Винеровский процесс почти наверное имеет непрерывные траектории.

▷ Для проверки нужно всё так же взять $\alpha = 4$ и проделать те же действия, что в одномерном случае. Получится, что с константами $\alpha = 4, \beta = 1, D = n(n+2)$ многомерный Винеровский процесс W будет удовлетворять Теореме 2.8 \square

2.4 Формула Иссерлиса: моменты гауссовских векторов

Этот результат очень старый, известен также как формула Вика (переоткрытый результат в статистической физике, спустя примерно 40 лет), доказательство можно найти в [11] и в приложении статьи. Формула комбинаторная и не очень распространена в классической вероятности, но иногда очень сильно спасает время.

Теорема 2.10. *Пусть $[X_1, \dots, X_K]^\top$ – гауссовский случайный вектор с нулевым матожиданием и ковариационной матрицей $\Sigma = (\sigma_{ij})_{i,j=1,\dots,K}$ и число K является чётным. Обозначим за \mathcal{P} множество идеальных паросочетаний, то есть, множество вида $\{(i_1, i_2), \dots, (i_{K-1}, i_K)\}$, где все $i_j \in \{1, \dots, K\}$, для всех $j \neq k$ индексы $i_j \neq i_k$ и при этом задействованы все элементы $\{1, \dots, K\}$.*

$$\mathbb{E}[X_1 X_2 \dots X_K] = \sum_{\pi \in \mathcal{P}} \prod_{(i,j) \in \pi} \text{Cov}(X_i, X_j) = \sum_{\pi \in \mathcal{P}} \prod_{(i,j) \in \pi} \sigma_{ij}.$$

В случае, если K нечётное, то

$$\mathbb{E}[X_1 X_2 \dots X_K] = 0.$$

В лекции мы её использовали для того, чтобы посчитать моменты Винеровского процесса, вспомним два для примера. Пусть $t > s$ – два момента времени.

$$\begin{aligned} \mathbb{E}[W_t^4] &= \mathbb{E}[W_t W_t W_t W_t] = |\text{формула Ис.}| = 3\mathbb{E}[W_t W_t] \mathbb{E}[W_t W_t] = 3t^2, \\ \mathbb{E}[W_t^2 W_s^2] &= \mathbb{E}[W_t W_t W_s W_s] = \\ &= |\text{формула Ис.}| = \mathbb{E}[W_t^2] \mathbb{E}[W_s^2] + \mathbb{E}[W_t W_s] \mathbb{E}[W_t W_s] + \mathbb{E}[W_t W_s] \mathbb{E}[W_t W_s] = ts + 2s^2. \end{aligned}$$

Броуновское движение: из физики в финансы

Эта лекция – небольшой экскурс в финансовую математику. Мы начнём с того, что попробуем понять, как Броуновское движение из физики оказалось в финансах, а затем рассмотрим одну базовую модель для финансового моделирования: геометрическое Броуновское движение. Наконец, посмотрим на некоторые финансовые задачи, которые приходится решать с помощью подобных моделей.

3.1 Откуда идёт финансовая математика

Можно сказать, что финансовая математика возникла примерно в 30-х годах XX века, когда уже были биржи и люди как-то умели анализировать временные ряды (согласно нашим определениям, временной ряд – это вещественнозначный случайный процесс в дискретном времени). По утверждению А.Н. Ширяева, до этого по сути вся финансовая математика сводилась в самом сложном случае к подсчёту сложных процентов [29, Гл.1]. В 30-х годах уже сформировался какой-то статистический инструментарий, данные стали привлекать интерес экономистов. В это время выходит несколько работ, посвящённых исследованию финансовых рядов, среди которых стоит упомянуть [1, 21, 2], где анализировались архивные данные цен. Вывод этих работ был весьма интересный: очень похоже, что логарифмы приращений цен $\ln S_t/S_{t-1}$ в этих рядах ведут себя как независимые. Эти работы, к сожалению, не получили в своё время должного внимания, как сейчас можно предполагать, отчасти из-за того, что экономисты тогда верили (некоторые до сих пор верят), что в экономических рядах есть разные ритмы и циклы, а отчасти из-за того, что для экономики проблема динамики цен была сильно второстепенна. В целом, и сейчас макроэкономика, в основном, занимается несколько другими вопросами. Третья причина была в том, что далеко не все экономисты тогда владели статистикой, которая ещё не была сильно распространена.

Поворотным моментом стала в 1953г. работа Кендалла [14], которая вышла в другое время, написана более зрелым статистическим языком и опубликована в журнале Королевского Статистического Общества. В ней автор анализирует различные ценовые временные ряды: месячные средние цены на пшеницу в Чикаго (1883-1934), на хлопок в Нью-Йорке (1816-1951), а также недельные данные цен 19 акций (1928-1938). Кендалл пришёл, более обоснованно, к тем же выводам, что и предыдущие статьи. Изначально его

работа была нацелена на то, чтобы выявить цикличность в поведении цен, но к своему великому удивлению он обнаружил (и эта фраза вошла в историю), что исследуемые ряды выглядят как если бы "...the Demon of Chance drew a random number.. and added it to the current price to determine the next .. price". Иными словами, если положить, что логарифмы процесса цен в дискретном времени $H_t = \ln S_t / S_{t-1}$ независимы, то процесс логарифмов цен – это случайное блуждание, то есть, процесс вида

$$\ln S_t = \ln S_{t-1} + \varepsilon_t,$$

где $(\varepsilon_t)_{t \in \mathbb{Z}_+}$ – процесс (предположительно, гауссовского) белого шума с нулевым матожиданием и начальная цена $S_0 \in \mathbb{R}$ фиксирована. Следовательно, сами цены ведут себя как

$$S_t = S_0 e^{\varepsilon_1 + \dots + \varepsilon_t}.$$

Как это ни удивительно, но подобные идеи возникали ранее: например, в работе Башелье *Théorie de la Speculation* (1900) [3] уже была другая, но похожая гипотеза. Башелье предположил, что цены сами являются случайным блужданием (*модель Башелье*)

$$S_{k\Delta} = S_{(k-1)\Delta} + \varepsilon_{k\Delta}$$

с гауссовским белым шумом $\varepsilon_{k\Delta}$, Δ – шаг дискретизации по времени. Однако, понятно, что никому в то время не нравилась идея со случайными блужданиями. Во-первых, это процесс, который нельзя предсказать (можно, но наилучший прогноз цены в будущем – это цена сегодня). Во-вторых, в такой модели цены могут быть отрицательными, то есть, нельзя сказать, что она адекватна. Заметим, что в пределе (в подходящем смысле) при $k = t/\Delta$ и $\Delta \rightarrow 0$ мы приходим к процессу в непрерывном времени

$$S_t = S_0 + \sigma W_t, \tag{3.1}$$

где $(W_t)_{t \in \mathbb{R}_+}$ – это уже известный нам Винеровский процесс.

Теорема 3.1. (Донскер, вариант) Пусть (U_n) – последовательность независимых и однаково распределённых случайных величин с матожиданием 0 и дисперсией 1. Рассмотрим

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i$$

и для $t \in [0, +\infty)$ линейную интерполяцию S_n , функцию

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} U_k + \frac{1}{\sqrt{n}} (nt - [nt]) U_{[nt]+1}.$$

Тогда при $n \rightarrow \infty$ конечномерные распределения процесса X_n с индексами t сходятся к конечномерным распределениям Винеровского процесса W .

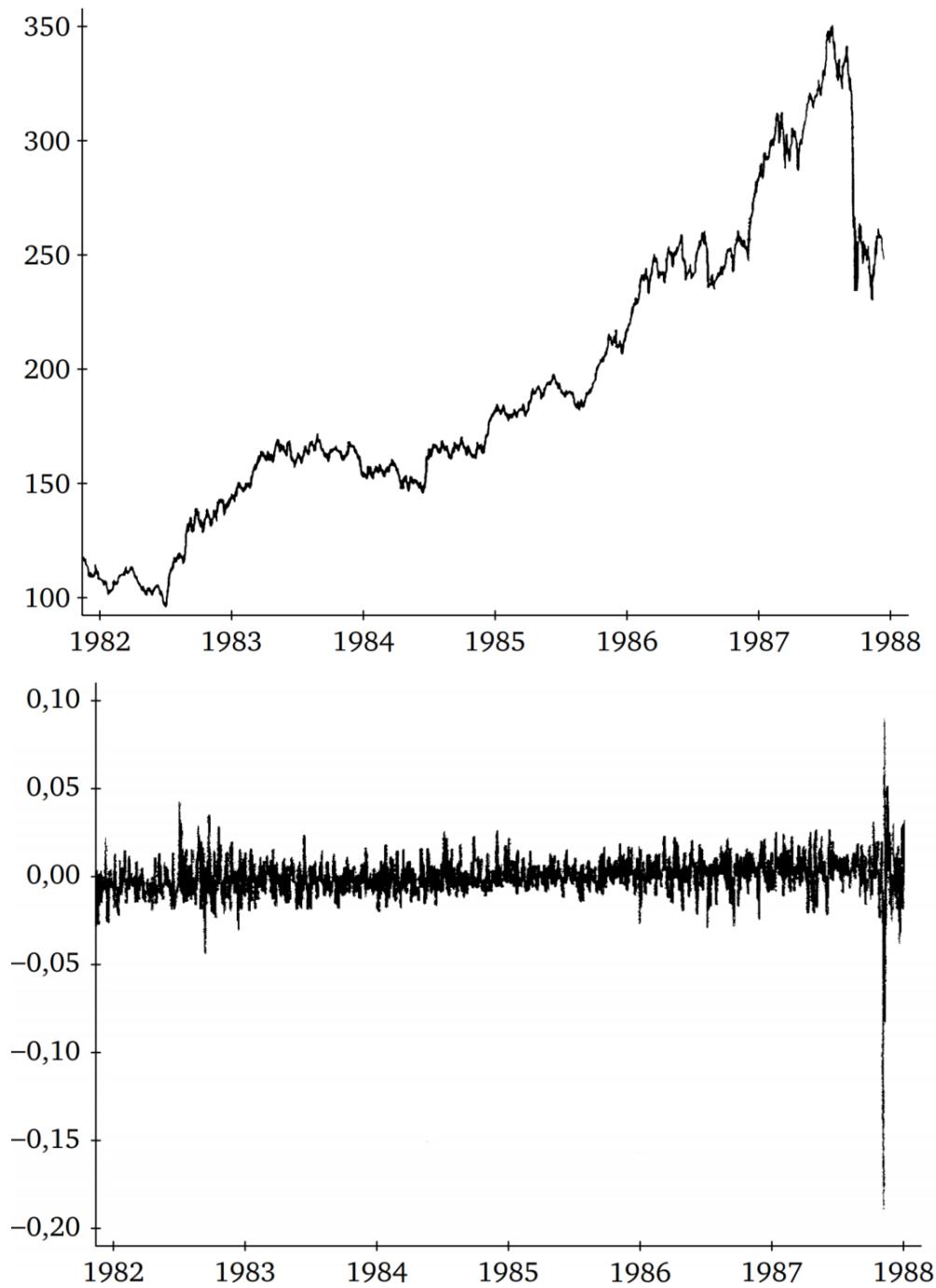


Рис. 3.1: Индекс S&P500 (верх) и его логарифмы приращений (низ), Рис. из [29, Гл. 1]. Несмотря на то, что гипотеза белошумности очень притягательна, в реальности она не до конца адекватна: оказывается, что последовательность шумов устроена более сложно и для её описания используют более сложные модели временных рядов.

▷ Давайте для краткости запишем

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} U_k + \psi_{n,t}, \quad X_n(0) = 0 \text{ п.н. .}$$

Поскольку величины U_k независимы, одинаково распределены, имеют нулевое матожида-

ние и конечную дисперсию, для них работает классическая ЦПТ:

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} U_k \xrightarrow[n \rightarrow \infty]{d} N(0, t).$$

С другой стороны, $\psi_{n,t} \xrightarrow[n \rightarrow \infty]{d} 0$. Рассмотрим теперь вектор зависимых случайных величин для $t_1 < \dots < t_k$.

$$[X_{t_1}, \dots, X_{t_k}]^T.$$

Попробуем сначала перейти к приращениям:

$$\begin{bmatrix} X_n(t_1) \\ X_n(t_2) \\ \dots \\ X_n(t_k) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \dots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} X_n(t_1) - X_n(0) \\ X_n(t_2) - X_n(t_1) \\ \dots \\ X_n(t_k) - X_n(t_{k-1}) \end{bmatrix}.$$

Вектор справа состоит из почти независимых случайных величин

$$X_n(t_p) - X_n(t_{p-1}) = \frac{1}{\sqrt{n}} \sum_{k=[nt_{p-1}]+1}^{[nt_p]} U_k + \psi_{n,t_p} - \psi_{n,t_{p-1}}.$$

По ЦПТ

$$X_n(t_p) - X_n(t_{p-1}) \xrightarrow[n \rightarrow \infty]{d} N(0, t_p - t_{p-1}),$$

а целый вектор сходится к

$$[W_{t_1} - W_0, W_{t_2} - W_{t_1}, \dots, W_{t_k} - W_{t_{k-1}}]^T,$$

так как части $\psi_{n,t_p} - \psi_{n,t_{p-1}}$ в пределе стремятся к 0 (и зависимости полностью исчезают) и разности по ЦПТ имеют гауссовское распределение как у приращения Винеровского процесса. \square

По настоящему значимой точкой, после которой финансовая математика начала бурно развиваться вместе с теорией анализа временных рядов, стала работа Пола Самуэльсона [18], который ввёл в обращение на основе уже имеющихся эмпирических фактов модель цен

$$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}, \quad (3.2)$$

с параметрами μ, σ^2 , которую он назвал экономическим Броуновским движением, сейчас мы её знаем как геометрическое Броуновское движение. Несмотря на то, что эта модель в данный момент признана не вполне соответствующей реальности (см. например, Рис. 3.1, где, очевидно, логарифмы приращений – это не совсем белый шум), она явилась очень мощным первым шагом и на основе неё было сделано много других решений, которые применяются и сейчас. В теории временных рядов есть много моделей, которые появились, можно сказать, хотя бы отчасти благодаря геометрическому Броуновскому движению; например, это условно гауссовские модели ARCH[10] и GARCH[4]. Они были призваны обощить эту модель, конкретнее, они моделируют параметр волатильности σ_t как отдельный случайный процесс.

3.2 Геометрическое Броуновское движение

Модель Самуэльсона использует геометрическое Броуновское движение.

Определение 3.1. Пусть $(W_t)_{t \in \mathbb{R}_+}$ – это Винеровский процесс такой, что $W_0 = 0$, $\mu \in \mathbb{R}, \sigma \in \mathbb{R}, x_0 \in \mathbb{R}_+$ – параметры модели (называемые сносом, волатильностью и начальной ценой). Одномерное геометрическое Броуновское движение (geometric Brownian motion, GBM) – это процесс $(X_t)_{t \in \mathbb{R}_+}$, определяемый как

$$X_t = x_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}. \quad (3.3)$$

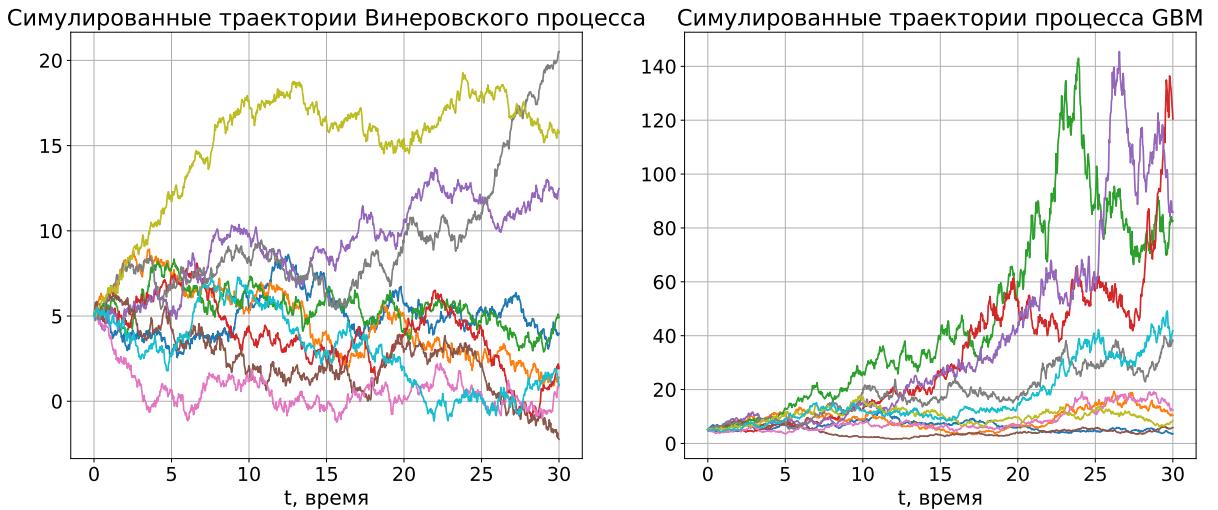


Рис. 3.2: Очевидно, что самого Винеровского процесса мало для моделирования цен (как предлагалось в [3]), поскольку он может уходить ниже нуля. При этом траектории GBM на первый взгляд очень похожи на типичные финансовые ряды.

Перечислим некоторые свойства этого процесса.

1. Легко заметить, что логарифм

$$\ln X_t = \ln X_0 + \left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_t$$

есть не что иное, как Винеровский процесс со сносом (из-за второго слагаемого) и волатильностью (из-за умножения на σ), и имеет нормальное распределение со средним $\ln X_0 + \left(\mu - \frac{\sigma^2}{2} \right) t$ и дисперсией $\sigma^2 t$. По определению, случайная величина X_t – логнормальная, то есть, её логарифм – нормальная величина.

Упражнение 3.1. Докажите, что плотность логнормальной случайной величины $X \sim LogN(\mu, \sigma^2)$ есть

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}},$$

где μ и σ^2 – параметры распределения $\ln X$.

Для того, чтобы вычислить некоторые характеристики такой случайной величины нам понадобится новый объект: производящая функция моментов (Moment Generating Function, MGF).

Определение 3.2. Пусть X – случайная величина. Производящей функцией моментов (MGF) называется функция

$$g_X(\lambda) = \mathbb{E}[e^{\lambda X}],$$

где $\lambda \in \mathbb{R}$.

Производящая функция хранит информацию о моментах случайной величины и обладает многими полезными свойствами. Для нас сейчас важно, что у гауссовой случайной величины с параметрами μ и σ^2 производящая функция моментов равна

$$g_X(\lambda) = e^{\mu\lambda + \sigma^2\lambda^2/2}.$$

Упражнение 3.2. Используя выражение для производящей функции моментов, докажите, что матожидание и ковариационная функция процесса $(X_t)_{t \in \mathbb{R}_+}$ равны

$$\mathbb{E}[X_t] = x_0 e^{\mu t}, \quad K(t, s) = x_0^2 e^{\mu(t+s)} \left(e^{\sigma^2 t \wedge s} - 1 \right). \quad (3.4)$$

Чему равна дисперсия $\mathbb{E}[(X_t - \mathbb{E}[X_t])^2]$?

Подобная ковариационная функция означает что сечения очень сильно зависят, причём зависимость растёт экспоненциально в зависимости от дальности сечений от нулевого, это объясняет взрывной и резкий характер траекторий процесса X_t (см. Рис. 3.2). А дисперсия показывает, что неопределенность относительно будущих значений нарастает очень быстро.

2. По заданию процесса, для любых $t, h \in \mathbb{R}_+$

$$X_{t+h} = X_t e^{\left(\mu - \frac{\sigma^2}{2}\right)h + \sigma W_h}, \quad (3.5)$$

что даёт нам некоторый инструмент для прогноза.

Утверждение 3.2. Для $t, s \in \mathbb{R}_+$ таких, что $t > s$, условное матожидание

$$\mathbb{E}[X_t | X_s] = X_s \mathbb{E}[X_{t-s}] = X_s \mathbb{E}\left[e^{(\mu - \sigma^2/2)(t-s) + \sigma W_{t-s}}\right].$$

▷ Воспользуемся свойством (3.5) и тем, что X_s можно вынести за знак условного матожидания:

$$\mathbb{E}[X_t | X_s] = X_s \mathbb{E}\left[e^{(\mu - \sigma^2/2)(t-s) + \sigma(W_t - W_s)} | X_s\right].$$

Заметим теперь, что по свойству независимости приращений Винеровского процесса

$$\mathbb{E} \left[e^{(\mu - \sigma^2/2)(t-s) + \sigma(W_t - W_s)} \mid X_s \right] = \mathbb{E} \left[e^{(\mu - \sigma^2/2)(t-s) + \sigma(W_t - W_s)} \right]$$

и, наконец из того, что $W_t - W_s$ и W_{t-s} имеют одинаковые распределения, получаем

$$\mathbb{E}[X_t \mid X_s] = X_s \mathbb{E} \left[e^{(\mu - \sigma^2/2)(t-s) + \sigma W_{t-s}} \right].$$

□ Условное матожидание, как мы увидим далее, в данном случае является наилучшим прогнозом, но из-за дисперсии даже этот прогноз будет очень неточным для больших $t-s$.

3. Процесс GBM очень легко симулировать, можно предложить, например, следующий простой алгоритм, чтобы получить дискретизированную траекторию X_{t_1}, \dots, X_{t_k} , состоящую из сечений в моменты $t_1, \dots, t_k \in \mathbb{R}_+$:

- (1) Симулировать траекторию Винеровского процесса (см. лекцию 2) W_{t_1}, \dots, W_{t_k} ;
- (2) Для всех $j = 1, \dots, k$ вычислить

$$X_{t_j} = x_0 e^{\left(\mu - \frac{\sigma^2}{2} \right) t_j + \sigma W_{t_j}}.$$

Кратко скажем про многомерную версию. Если на рынке d товаров (акций), то мы определяем процесс $(X_t)_{t \in \mathbb{R}_+}$ со значениями в \mathbb{R}_+^d аналогично одномерному случаю, но вводим многомерный Винеровский процесс $(W_t)_{t \in \mathbb{R}_+}$ и матрицу $A \in \mathbb{R}^{d \times d}$ для задания коррелированного многомерного Винеровского процесса.

Определение 3.3. Пусть $(W_t^i)_{t \in \mathbb{R}_+}$ – координаты d -мерного Винеровского процесса, а также заданы в качестве параметров $\mu_i \in \mathbb{R}^d$, $x_0 \in \mathbb{R}^d$ и $A \in \mathbb{R}^{d \times d}$. Тогда многомерным геометрическим Броуновским движением называется процесс $(X_t)_{t \in \mathbb{R}_+}$ с координатами

$$X_t^i = x_0^i e^{\left(\mu_i - \sum_{j=1}^d a_{ij}^2 / 2 \right) t + \sum_{j=1}^d a_{ij} W_t^j}.$$

Видно, что в многомерном случае координаты процесса зависимы через коррелированные Винеровские процессы.

3.3 Задача оценки опционов

Упрощённо, биржа существует достаточно давно для того, чтобы осуществлять сделки по продаже и покупке товаров. На бирже в качестве товаров выступают как физические товары (например, драгоценные металлы, валюта, нефть), так и ценные бумаги (акции и облигации). Работает это так: продавец приносит на биржу свой товар и ставит ценник. Далее биржа по всем продавцам вычисляет (по специальной методике, средняя, но есть тонкости) рыночную цену. С другой стороны, покупатели ставят заявки на покупку товара

и тоже ставят свои цены. Как покупатели, так и продавцы могут изменять свои цены. Как только рыночная цена доходит до уровня, когда есть продавец и есть покупатель такие, что цена их обоих устраивает, биржа проводит сделку. Всё совершается через механизм биржи, продавцы и покупатели друг друга даже не знают и все расчёты проводят с биржей, где специальный клиринговый оператор занимается техническими деталями сделки.

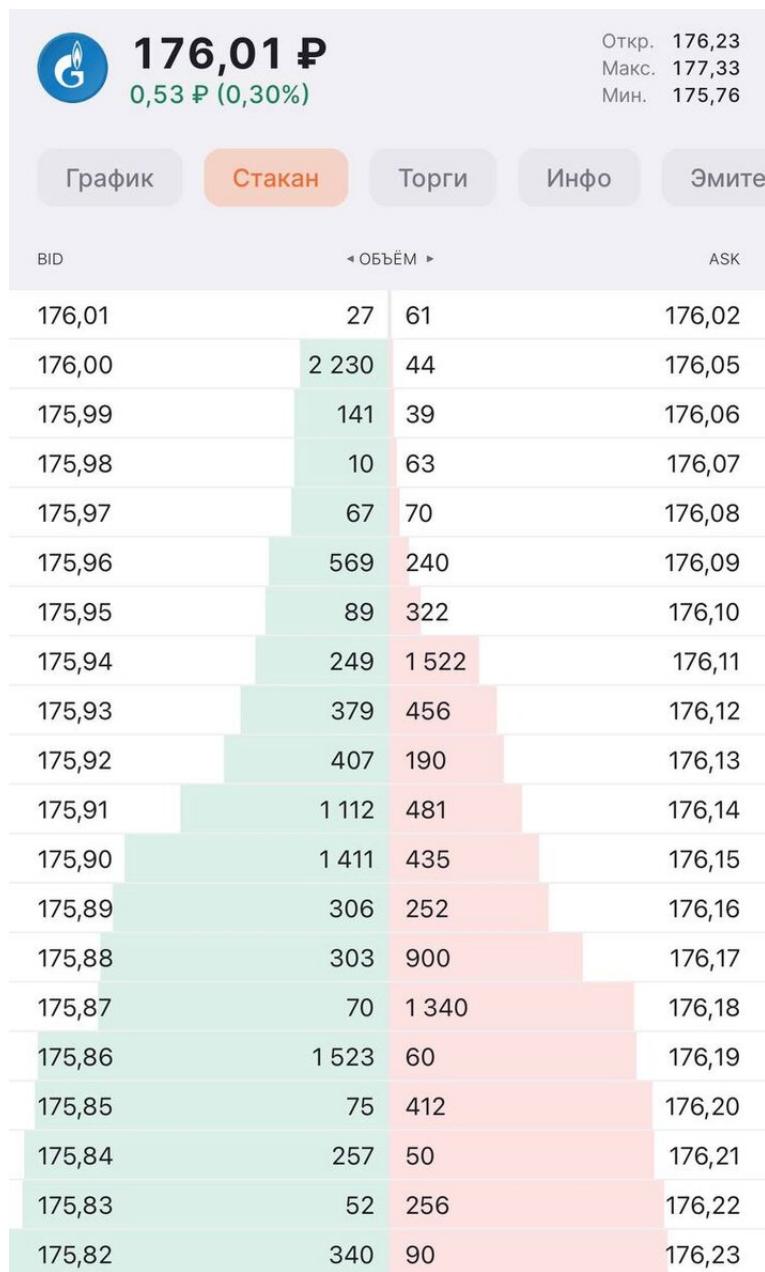


Рис. 3.3: Стакан с заявками ()

Помимо товаров на бирже существуют производные финансовые инструменты (дерибивативы), наиболее популярные – опционы и фьючерсы. Чтобы понять, зачем вообще нужны деривативы, обратимся к простому примеру. Сейчас это уже кажется несколько странным сюжетом, но, скажем, в XIX веке биржи работали именно так. Фермер Егор выращивает картофель, у него есть большое хозяйство и он использует биржу как посредника для

торгов. Есть сеть ресторанов Fried, которая заинтересована в покупке картофеля. Естественно, на рынке бывают урожайные и неурожайные времена, поэтому цена картофеля будет меняться со временем неопределенным образом. Для фермера продажа картофеля – это риск, так как он вложил некоторые средства в его производство и он будет разочарован и может потерять деньги, если цена на картофель будет очень низкая, когда он собирается его продавать. С другой стороны, Fried может потерпеть большие издержки, чем хотелось бы, при покупке картофеля по высокой цене. Для таких случаев биржа предлагает решение в виде деривативов, специальных контрактов на будущие покупки или приобретения, с разными сроками действия.

Егор может приобрести у биржи (или у другого агента на бирже) *фьючерс*, контракт, который обязывает его продать какое-то количество картофеля через год (в строго определенный момент) по заранее(!) заданной цене. С другой стороны, Fried может купить похожий фьючерс, который обязывает его купить какое-то количество картофеля через год по заранее заданной цене. Тем самым, и Егор, и Fried уменьшают риск неудовлетворения от неопределенности цены на картофель через год.

Также Егор или Fried могут приобрести или оформить *опцион*, который отличается тем, что он даёт лишь право на продажу/покупку заданного количества картофеля по заданной цене через год в строго определённое время. То есть, дождавшись момента исполнения, Егор может решить, что текущая цена на бирже выше той, что указаны в опционном контракте и он просто продаст картофель, не исполняя контракта. А Fried, соответственно, может увидеть, что цена на картофель так низка, что опцион исполнять невыгодно и тоже просто купит картофель по текущей цене.

Сумма в контракте на поставку называется *страйком* (strike price). Опционы на покупку имеют приставку *call* (call-опцион), а на продажу – *put* (put-опцион), есть также двусторонние (double), в которых обе стороны осуществляют поставку. Существует много видов опционов, самые популярные – *европейские* и *американские* (на Мосбирже есть и те, и те, очень популярны опционы на индексы и валюту). Первые работают как в приведенном примере: владелец опциона ждёт конца срока действия, а потом решает, исполнять его или нет. Что касается американских опционов, то их можно исполнить в *любой момент* до конца срока действия. В опционах может быть также задана необычная *функция выплаты*: например, если опцион покупается на несколько товаров (корзина, basket-опцион), то выплата будет равняться страйку, если опцион исполняется, и сумме от продажи/покупки заданного количества всех товаров, указанных в контракте иначе.

Для простоты положим, что рассматриваем европейский опцион. Вопрос: если биржа или агент на бирже хочет выпустить опцион, то как назначить его цену?

Определение 3.4. Пусть задана выплата опциона $f : \mathbb{R}^d \rightarrow \mathbb{R}$ и случайный процесс цен $(X_t)_{t \in \mathbb{R}_+}$, принимающий значения в \mathbb{R}^d . Честной ценой европейского опциона со временем исполнения $T > 0$ называется условное матожидание

$$\mathbb{E}[f(X_T) | X_0].$$

Пример 3.1. Рассмотрим европейский *rit-опцион*, который позволяет продать по одной единице каждого из товаров $1, \dots, d$ по цене K (страйк). В случае, если в момент исполнения биржевые цены X_i таковы, что продажа товаров по текущим ценам невыгодна (то есть, мы получаем меньше страйка K), опцион выгоднее исполнить, тогда после продажи мы получим выгоду $K - \sum_{i=1}^d X_i$. Если же страйк ниже суммы текущих цен, то нам выгоднее не исполнять опцион, получая нулевой выигрыш. Описанный опцион имеет выплату

$$f(x_1, \dots, x_d) = \left(K - \sum_{i=1}^d x_i \right)_+,$$

где знак $+$ означает, что всё выражение равно нулю, если в скобках число меньше нуля.

Упражнение 3.3. Как выглядит платёжная функция call-опциона с таким же контрактом?

Если бы биржа продавала опцион по цене ниже честной, то она бы проиграла: в среднем она сама может выиграть с ним больше, чем если продаст этот опцион. Если цена получается выше, то такой опцион, вероятно, никто не купит – в конце концов, не только биржа их умеет оценивать. В реальности же (отчасти из-за неточностей прогнозов покупателя и продавца) есть некоторое окно, окрестность честной цены, в котором покупатель и продавец могут договориться, так что оценка честной цены имеет вполне практическое значение для реализации эффективного рынка. Для случая одного товара есть знаменитая формула Блека-Шоулза (из-за которой модель геометрического Броуновского движения тоже часто называют моделью Блека-Шоулза). Её доказательство сильно техническое и мы его приводить не будем (три разных её доказательства можно найти в [29, Т.2, Гл.8]).

Утверждение 3.3. (Формула Блека-Шоулза) Пусть $(X_t)_{t \in \mathbb{R}_+}$ – одномерное геометрическое Броуновское движение с параметрами x_0, μ, σ . Тогда честная цена опциона call со сроком исполнения $T > 0$ и страйком K равна

$$C_T = x_0 \Phi \left(\frac{\ln \frac{x_0}{K} + T\sigma^2/2}{\sigma\sqrt{T}} \right) - K \Phi \left(\frac{\ln \frac{x_0}{K} - T\sigma^2/2}{\sigma\sqrt{T}} \right),$$

в частности, если $K = x_0$, то

$$C_T = x_0 \left[\Phi \left(\frac{\sigma\sqrt{T}}{2} \right) - \Phi \left(-\frac{\sigma\sqrt{T}}{2} \right) \right],$$

где

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx,$$

функция распределения величины $\mathcal{N}(0, 1)$.

Для случая корзинных опционов (много товаров) формул, как правило, нет, для оценки используется метод Монте-Карло и модель цен. Положим, у биржи есть модель многомерного геометрического Броуновского движения $(X_t)_{t \in \mathbb{R}_+}$ для моделирования цен на d товаров, задействованных в опционе, с параметрами $\mu \in \mathbb{R}^d$ и матрицей $A \in \mathbb{R}^{d \times d}$. Тогда задача оценки опциона может быть решена методом Монте-Карло. В реальности могут использоваться более сложные предсказательные модели, в частности, со случайными сносом и волатильностью.

Заметим одно важное экономическое свойство честной цены: она растёт с ростом срока истечения T . Это естественное следствие того, что долгосрочный опцион страхует большую неопределённость, чем короткосрочный. По этой причине долгосрочные опционы редкие и очень дорогие, как правило, срок истечения у них не больше двух лет. Кроме того, опционы после их выпуска становятся дешевле со временем, так как срок истечения приближается и страхуемая неопределенность тоже становится меньше.

Мы обсудили, как можно оценивать европейские опционы. Американские опционы и разные более экзотические варианты деривативов оценивать сложнее: для этого нужно решать задачу оптимальной остановки (в случае американского опциона) или стохастического оптимального управления (для более сложных деривативов).

Цепи Маркова

Давайте вспомним пример из второй лекции про клиента магазина, перемещающегося по отделам в какой-то последовательности. Это случайный процесс с дискретным временем (т.к. нас интересует только последовательность переходов между отделами, а не время, проведенное в каждом отделе), значения которого лежат в конечном множестве (множество отделов). Этот процесс можно описать с помощью цепи Маркова.

4.1 Определение цепи Маркова

Пусть $(X_t)_{t \in \mathbb{Z}_+}$ - случайный процесс на вероятностном пространстве (Ω, \mathcal{F}, P) , описывающий поведение клиента магазина. В данном случае моменты времени t принимают дискретные значения в $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$. Пусть $\Xi = \{1, 2, \dots r\}$ - пространство состояний (множество значений, которые могут принимать случайные величины X_t). Т.к. Ξ - дискретное пространство, логичнее всего ввести на нем наибольшую σ -алгебру \mathcal{F} , состоящую из всех возможных подмножеств Ξ .

Чтобы определить случайный процесс $(X_t)_{t \in \mathbb{Z}_+}$, нам нужно задать согласованное (т.е. удовлетворяющее теореме Колмогорова) семейство совместных распределений

$$F(t_1, \dots, t_n; A_1, \dots, A_n) = P(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n), \quad A_1, \dots, A_n \in \mathcal{F}.$$

Давайте сделаем это для последовательных моментов времени $t_0 = 0, t_1 = 1, \dots, t_n = n$ и для одноэлементных множеств вида $A_s = \{i\}$. Пусть задано $\mu \in \mathbb{R}^{1 \times r}$ - вектор-строка, описывающая начальное распределение, а также $p_{ij}(t)$ - вероятности перехода из состояния i в состояние j в момент времени t :

$$\mu = (P(X_0 = 1), \dots, P(X_0 = r)) \in \mathbb{R}^{1 \times r},$$

$$p_{ij}(t) = P(X_t = j | X_{t-1} = i).$$

Зададим совместное распределение следующим образом:

$$P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = \mu_{i_0} p_{i_0 i_1}(1) p_{i_1 i_2}(2) \dots p_{i_{n-1} i_n}(n). \quad (4.1)$$

В дальнейшем вы должны будете проверить, что эта формула задает семейство согласованных мер.

Упражнение 4.1. Как в этом случае будут выглядеть совместные распределения для произвольных моментов времени t_1, t_2, \dots, t_k и произвольных множеств $A_1, A_2, \dots, A_n \in \mathcal{F}$?

4.1.1 Стохастические матрицы

Переходные вероятности $p_{ij}(t)$ удобно записать в виде матрицы $P(t) = (p_{ij}(t))_{i,j=1}^r \in \mathbb{R}^{r \times r}$. Однако не любая матрица задает семейство переходных вероятностей. Для этого нужно, чтобы элементы матрицы $p_{ij}(t) = P(X_t = j | X_{t-1} = i)$ в самом деле были условными вероятностями.

Упражнение 4.2. Вспомните, что такое условная вероятность. Докажите, что $f(A) := P(A|B) = \frac{P(A \cap B)}{P(B)}$ является вероятностной мерой относительно A , т.е. $0 \leq f(A) \leq 1$, $f(\Omega) = 1$, $f(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} f(A_i)$.

Таким образом, чтобы матрица $P(t) = (p_{ij}(t))_{i,j=1}^r \in \mathbb{R}^{r \times r}$ задавала семейство переходных вероятностей, должны быть выполнены следующие свойства:

1. Элементы матрицы должны быть неотрицательны:

$$p_{ij}(t) = P(X_t = j | X_{t-1} = i) \geq 0,$$

2. Сумма элементов в любой строке должна быть равна единице:

$$\sum_{j=1}^r p_{ij}(t) = P(X_t \in \Xi | X_{t-1} = i) = 1.$$

Определение 4.1. Матрицы, удовлетворяющие выше перечисленным свойствам, называются стохастическими.

Упражнение 4.3. Докажите, что если $Q \in \mathbb{R}^{r \times r}$ - стохастическая матрица, $\mu = (\mu_1, \dots, \mu_r)$ - распределение вероятностей, то μQ - также распределение вероятностей.

Упражнение 4.4. Докажите, что если $Q', Q'' \in \mathbb{R}^{r \times r}$ - стохастические матрицы, то $Q := Q'Q''$ - также стохастическая матрица.

Определение 4.2. Таким образом, по теореме Колмогорова начальное распределение μ и набор стохастических матриц $(P(t))_{t \in \mathbb{N}}$ задают цепь Маркова, определяемую с помощью согласованного семейства совместных распределений по формуле 4.1.

Упражнение 4.5. Проверьте, что формула 4.1 задает вероятностную меру на пространстве последовательностей $\Omega = \{\omega = (i_0, i_1, i_2, \dots, i_n), i_0, \dots, i_n \in \Xi\}$, снабженном наибольшей σ -алгеброй $\mathcal{G} = 2^\Omega$, т.е.

$$0 \leq P(\omega) \leq 1, P(\Omega) = 1, P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad \forall A_i \in \mathcal{G},$$

где $\bigcup_{i=1}^{\infty} A_i$ обозначает объединение непересекающихся множеств A_i .

Упражнение 4.6. Почему семейство распределений, заданное формулой 4.1, будет согласованым?

Теорема 4.1. Для цепи Маркова, порожденной начальным распределением μ и стохастическими матрицами $P(t)$, выполняются следующие равенства:

$$P(X_n = j | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i) = p_{ij}(n), \quad (4.2)$$

$$P(X_{t+s} = j | X_t = i) = (P(t+1) \dots P(t+s))_{ij}, \quad (4.3)$$

$$P(X_t = j) = (\mu P(1) P(2) \dots P(t))_j. \quad (4.4)$$

▷

Сначала докажем формулу 4.2. Для этого запишем определение условной вероятности $P(A|B) = \frac{P(A \cap B)}{P(B)}$, а после этого используем определение совместного распределения 4.1.

$$\begin{aligned} P(X_k = j | X_0 = i_0, X_1 = i_1, \dots, X_{k-1} = i_{k-1}) &= \frac{P(X_0 = i_0, X_1 = i_1, \dots, X_{k-1} = i_{k-1}, X_k = j)}{P(X_0 = i_0, X_1 = i_1, \dots, X_{k-1} = i_{k-1})} = \\ &= \frac{\mu_{i_0} p_{i_0 i_1}(1) p_{i_1 i_2}(2) \dots p_{i_{n-2} i_{n-1}}(n-1) p_{i_{n-1} i_n}(n)}{\mu_{i_0} p_{i_0 i_1}(1) p_{i_1 i_2}(2) \dots p_{i_{n-2} i_{n-1}}(n-1)} = p_{i_{n-1} i_n}(n). \end{aligned}$$

Для доказательства формулы 4.3 снова используем определение условной вероятности, после этого применим формулу полной вероятности. А в конце используем матричную запись для стохастических матриц $P(t)$, порождающих цепь Маркова.

$$\begin{aligned} P(X_{t+s} = j | X_t = i) &= \frac{P(X_{t+s} = j, X_t = i)}{P(X_t = i)} = \\ &= \frac{\sum_{i_0, i_1, \dots, i_{t-1}, i_{t+1}, \dots, i_{t+s-1} \in \Xi} P(X_0 = i_0, X_1 = i_1, \dots, X_{t-1} = i_{t-1}, X_t = i, \dots, X_{t+s-1} = i_{t+s-1}, X_{t+s} = j)}{\sum_{i_0, i_1, \dots, i_{t-1} \in \Xi} P(X_0 = i_0, X_1 = i_1, \dots, X_{t-1} = i_{t-1}, X_t = i)} = \\ &= \frac{\sum_{i_0, i_1, \dots, i_{t-1} \in \Xi} \mu_{i_0} p_{i_0 i_1}(1) \dots p_{i_{t-1} i}(t) \sum_{i_{t+1}, \dots, i_{t+s-1} \in \Xi} p_{i t+1}(t+1) \dots p_{i_{t+s-1} j}(t+s)}{\sum_{i_0, i_1, \dots, i_{t-1} \in \Xi} \mu_{i_0} p_{i_0 i_1}(1) \dots p_{i_{t-1} i}(t)} = \\ &= \sum_{i_{t+1}, \dots, i_{t+s-1} \in \Xi} p_{i t+1}(t+1) p_{i_{t+1} i_{t+2}}(t+2) \dots p_{i_{t+s-2} i_{t+s-1}}(t+s-1) p_{i_{t+s-1} j}(t+s) = \\ &= (P(t+1) P(t+2) \dots P(t+s))_{ij}. \end{aligned}$$

Для доказательства утверждения 4.4 используем формулу полной вероятности, а после этого используем определение совместного распределения 4.1, а также матричную запись для стохастических матриц $P(t)$, порождающих цепь Маркова.

$$\begin{aligned} P(X_t = j) &= \sum_{i_0, i_1, \dots, i_{t-1} \in \Xi} P(X_0 = i_0, X_1 = i_1, \dots, X_{t-1} = i_{t-1}, X_t = j) = \\ &= \sum_{i_0, i_1, \dots, i_{t-1} \in \Xi} \mu_{i_0} p_{i_0 i_1}(1) \dots p_{i_{t-2} i_{t-1}}(t-1) p_{i_{t-1} j}(t) = (\mu P(1) P(2) \dots P(t))_j. \end{aligned}$$

□

4.1.2 Однородная цепь Маркова

Давайте рассмотрим частный случай цепи Маркова, в котором переходные вероятности не зависят от времени. Такая цепь Маркова называется однородной. Эти модели удобны тем, что их можно легко визуализировать с помощью графов. В однородных цепях Маркова выражения $P(X_{t+s} = j | X_t = i)$ и $P(X_t = j)$ принимают более лаконичную форму.

Определение 4.3. Цепь Маркова называется однородной, если стохастические матрицы, ее определяющие, не зависят от времени, т.е. $P(t) = P, \forall t \in \mathbb{N}$.

Давайте представим однородную цепь Маркова в виде ориентированного взвешенного графа $G = (V, E, W)$. Множество вершин графа V соответствует пространству состояний $\Xi = \{1, 2, \dots, r\}$ цепи Маркова. Ребро графа $e_{ij} \in E$ имеет вес W_{ij} равный вероятности перехода p_{ij} из состояния i в состояние j . Любая реализация случайного процессса, соответствующего цепи Маркова, представляется как путь в графе. Следовательно, однородную цепь Маркова можно представить как распределение вероятностей на пространстве путей в полученном графике.

Пример 4.1. Марковской цепи с переходными вероятностями $P = \begin{pmatrix} 0.1 & 0.9 & 0 \\ 0 & 0.2 & 0.8 \\ 0.7 & 0 & 0.3 \end{pmatrix}$ соответствует следующий график, изображенный на рисунке 4.1.

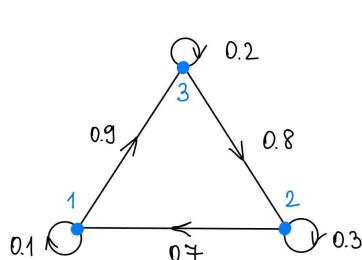


Рис. 4.1: Граф, описывающий цепь Маркова

Упражнение 4.7. Для однородной цепи Маркова, порожденной начальным распределением μ и стохастической матрицей P найдите вероятность перехода за s шагов $P(X_{t+s} = j | X_t = i)$ и распределение вероятностей для состояния в момент времени t , т.е. $P(X_t = j)$.

4.1.3 Эргодическая теорема

Вернемся к нашему примеру с отделами магазина. Предположим, покупатель первый раз пришел в магазин, он еще не привык к расположению отделов, поэтому перемещается по ним случайным образом. Однако с течением времени он изучит расположение отделов и привыкнет; можно посмотреть, как часто он посещает каждый из отделов, т.е. вычислить вероятности $\lim_{s \rightarrow \infty} P(X_s = j)$. Давайте изучим вопрос о существовании этих пределов.

Определение 4.4. Стохастическая матрица P называется эргодической, если существует такое T , что $(P^{(T)})_{ij} = p_{ij}^{(T)} > 0$ для всех i, j .

Заметим, что поскольку множество возможных состояний $\Xi = 1, 2, \dots, r$ конечно, можно найти такое α что $(P^{(T)})_{ij} = p_{ij}^{(T)} \geq \alpha$ для всех i, j .

Упражнение 4.8. Приведите пример неэргодической матрицы и соответствующей ей цепи Маркова.

Теорема 4.2. (Эргодическая теорема) Пусть цепь Маркова порождена начальным распределением μ и эргодической матрицей P . Тогда верны следующие утверждения:

1. переходные вероятности за s шагов сходятся к распределению π , т.е. $\lim_{s \rightarrow \infty} p_{ij}^{(s)} = \pi_j$,
2. распределение вероятностей $P(X_s = j)$ для случайной величины X_s также сходится к π независимо от начального распределения μ , т.е. $\lim_{s \rightarrow \infty} (\mu P^{(s)})_j = \pi_j$
3. такое распределение π единственно и инвариантно относительно стохастической матрицы P , т.е. $\pi P = \pi$.

Определение 4.5. Распределение π из предыдущей теоремы называется инвариантным или стационарным.

Для доказательства теоремы нам понадобится следующая лемма.

Лемма 4.3. Пусть $\mathcal{P} = \{\mu \in \mathbb{R}^{1 \times r} : \sum_{i=1}^r \mu_i = 1, \mu_i \geq 0 \quad \forall i\}$ - пространство вероятностей на Ξ . Тогда $d(\mu', \mu'') := \frac{1}{2} \sum_{i=1}^r |\mu'_i - \mu''_i|$ задает метрику на \mathcal{P} , при этом пространство (\mathcal{P}, d) является полным метрическим пространством, т.е. выполняется критерий Коши. Также выполнены следующие свойства

$$d(\mu', \mu'') \leq 1, \tag{4.5}$$

$$d(\mu'Q, \mu''Q) \leq d(\mu', \mu''), \quad (4.6)$$

где Q - стохастическая матрица.

$$d(\mu'Q, \mu''Q) \leq (1 - \alpha)d(\mu', \mu''), \quad (4.7)$$

где Q - стохастическая матрица, такая что $q_{ij} \geq \alpha \quad \forall i, j$.

Мы будем использовать эту лемму для доказательства теоремы, в частности применим формулу 4.7 для $Q = P^{(T)}$, где T такое, что $(P^{(T)})_{ij} = p_{ij}^{(T)} \geq \alpha$ для всех i, j (см. определение эргодической матрицы).

▷

Пусть $\mu', \mu'' \in \mathcal{P}$. Обозначим через \sum^+ суммирование по тем индексам, для которых слагаемые положительны.

1. Для начала сформулируем несколько технических утверждений:

$$0 = \sum_{i=1}^r \mu'_i - \sum_{i=1}^r \mu''_i = \sum_{i=1}^r {}^+(\mu'_i - \mu''_i) - \sum_{i=1}^r {}^+(\mu''_i - \mu'_i),$$

$$d(\mu', \mu'') = \frac{1}{2} \sum_{i=1}^r {}^+(\mu'_i - \mu''_i) + \frac{1}{2} \sum_{i=1}^r {}^+(\mu''_i - \mu'_i) = \sum_{i=1}^r {}^+(\mu'_i - \mu''_i).$$

2. Ясно, что $d(\mu', \mu'') \leq 1$, т.к.

$$d(\mu', \mu'') := \frac{1}{2} \sum_{i=1}^r |\mu'_i - \mu''_i| \leq \frac{1}{2} \sum_{i=1}^r |\mu'_i| + \frac{1}{2} \sum_{i=1}^r |\mu''_i| \leq 1.$$

3. Пусть Q - стохастическая матрица. Тогда $\mu'Q, \mu''Q$ будут распределениями вероятностей. Используя доказанные выше технические утверждения, запишем оценку на $d(\mu'Q, \mu''Q)$:

$$\begin{aligned} d(\mu'Q, \mu''Q) &= \sum_{j=1}^r {}^+(\mu'Q - \mu''Q)_j = \sum_{j \in J} (\mu'Q - \mu''Q)_j = \\ &= \sum_{j \in J} \sum_{i=1}^r (\mu'_i q_{ij} - \mu''_i q_{ij}) = \sum_{i=1}^r (\mu'_i - \mu''_i) \sum_{j \in J} q_{ij} \leq \sum_{i=1}^r {}^+(\mu'_i - \mu''_i) \sum_{j \in J} q_{ij}. \end{aligned}$$

В общем случае $\sum_{j \in J} q_{ij} \leq \sum_{j=1}^r q_{ij} = 1$. Тогда

$$d(\mu'Q, \mu''Q) \leq \sum_{i=1}^r {}^+(\mu'_i - \mu''_i) = d(\mu', \mu''),$$

В случае, если $q_{ij} \geq \alpha \quad \forall i, j$, мы можем записать следующие неравенства:

$$\sum_{j \in J} q_{ij} = 1 - \sum_{j \notin J} q_{ij} \leq 1 - \alpha,$$

$$d(\mu'Q, \mu''Q) \leq (1 - \alpha) \sum_{i=1}^r (\mu'_i - \mu''_i) = (1 - \alpha)d(\mu', \mu'').$$

4. Утверждение, что $d(\mu', \mu'') := \frac{1}{2} \sum_{i=1}^r |\mu'_i - \mu''_i|$ задает метрику, следует из того, что $d_1(m', m'') = |m' - m''|$ является метрикой в \mathbb{R} . Критерий Коши для d аналогичен критерию Коши для d_1 .

□

▷ (Доказательство теоремы)

1. Для начала докажем, что распределение для X_s сходится при $s \rightarrow \infty$. Пусть μ - распределение X_0 , тогда $\mu^s := \mu P^{(s)}$ - распределение для X_s . Применим формулу 4.7 для $Q = P^{(T)}$.

$$d(\mu^s, \mu^{s+t}) = d(\mu P^{(s)}, \mu P^{(s+t)}) \leq (1 - \alpha)d(\mu P^{(s-T)}, \mu P^{(s+t-T)}).$$

Повторим эту операцию m раз, где m такое, что $0 < n - mT < s$. Получим, что

$$d(\mu^s, \mu^{s+t}) \leq (1 - \alpha)^m d(\mu P^{(s-mT)}, \mu P^{(s+mT)}) \leq (1 - \alpha)^m.$$

Для любого сколь угодно малого $\varepsilon > 0$ можно найти такие s, m , что $d(\mu^s, \mu^{s+t}) \leq (1 - \alpha)^m < \varepsilon$. Таким образом последовательность $\{\mu_s\}$ является последовательностью Коши, а следовательно сходится, т.к. по лемме (\mathcal{P}, d) - полное метрическое пространство.

2. Давайте проверим, что $\pi P = \pi$.

$$\pi P = \lim_{s \rightarrow \infty} \mu P^{(s)} P = \lim_{s \rightarrow \infty} \mu P^{(s+1)} = \pi.$$

3. Теперь давайте докажем, что такое π единственно.

Пусть $\pi_1 P = \pi_1, \pi_2 P = \pi_2$. Тогда

$$d(\pi_1, \pi_2) = d(\pi_1 P^{(T)}, \pi_2 P^{(T)}) \leq (1 - \alpha)d(\pi_1, \pi_2).$$

Т.е. $d(\pi_1, \pi_2) = 0$, что означает, что π_1 и π_2 совпадают и распределения $\mu^s = \mu P^{(s)}$ сходятся к π независимо от начального распределения μ .

4. Теперь проверим, что переходные вероятности за s шагов сходятся к тому же пределу. Для этого возьмем $\mu = (0, 0, \dots, 0, 1, 0, \dots, 0)$, где единица стоит на i -ой позиции. Тогда переходные вероятности можно выразить следующим образом: $p_{ij}^{(s)} = (\mu P^{(s)})_j$. Нам уже известно, что $\mu_0 P^{(s)}$ сходится к π , т.е. $\lim_{s \rightarrow \infty} (\mu_0 P^{(s)})_j = \pi_j$. Тогда

$$\lim_{s \rightarrow \infty} p_{ij}^{(s)} = \lim_{s \rightarrow \infty} (\mu_0 P^{(s)})_j = \pi_j.$$

□

Пример 4.2. Классическими примерами неэргодических цепей Маркова являются несвязанные цепи маркова (число сообщающихся классов больше одного) и цепи с циклами.

Теорема 4.4. Условие детального баланса $\pi_i P_{ij} = \pi_j P_{ji}$ является достаточным условием стационарности π .

Это теорема доказывается прямой проверкой уравнения $\pi P = \pi$. Заметим, что условие детального баланса играет ключевую роль в построении алгоритмов семплирования с помощью цепей Маркова, т.к. для того, чтобы получить семпл с целевым распределением π , нужно сконструировать цепь Маркова с переходными вероятностями P , удовлетворяющими условию баланса, а сходимость обеспечивается эргодической теоремой.

4.2 Обобщение цепи Маркова на случай непрерывного времени

Для практических приложений бывает полезно обобщить цепи Маркова на непрерывное время. Такие процессы могут описывать, например, процесс Пуассона (количество случайных событий, происходящее с заданной интенсивностью). Для того, чтобы определить такой процесс, нам снова понадобится воспользоваться теоремой Колмогорова, которая позволяет задать случайный процесс набором согласованных мер. Для начала рассмотрим, как выглядели согласованные меры в случае обычной цепи Маркова.

$$P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = \mu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n}.$$

$$F(t_1, \dots, t_m; A_{t_1}, \dots, A_{t_m}) = P(X_{t_1} \in A_{t_1}, \dots, X_{t_m} \in A_{t_m}) =$$

$$\begin{aligned} &= \sum_{i_{t_1} \in A_{t_1}, \dots, i_{t_m} \in A_{t_m}, i_0 \in \Xi, \dots, i_{t_{n-1}} \in \Xi} P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) \\ &= \sum_{i_{t_1} \in A_{t_1}, \dots, i_{t_m} \in A_{t_m}, i_0 \in \Xi, \dots, i_{t_n} \in \Xi} \mu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n}. \end{aligned}$$

Заметим, что суммирование ведется по всем индексам, но i_{t_j} , которые параметризуют F , суммируются по A_{t_j} , а остальные - по всему допустимому пространству состояний. Формулы выглядят довольно громоздко, но оказывается, что для случая непрерывного времени они значительно упрощаются. Если мы решаем отказаться от дискретности времени, то вероятности перехода из состояния i_k в состояние i_{k+1} за 1 шаг следует заменить на такую же вероятность, но теперь за время $t_{k+1} - t_k$. В таком случае согласованные меры будут выглядеть следующим образом:

$$F(t_1, \dots, t_m; A_{t_1}, \dots, A_{t_m}) = \sum_{i_{t_1} \in A_{t_1}, \dots, i_{t_m} \in A_{t_m}} \mu_{i_0} p_{i_0 i_{t_1}}(t_1 - 0) p_{i_{t_1} i_{t_2}}(t_2 - t_1) \dots p_{i_{t_{m-1}} i_{t_m}}(t_m - t_{m-1}) \quad (4.8)$$

Заметим, что согласованные меры для цепи Маркова с непрерывным временем сильно напоминают согласованные меры для Виеровского процесса, отличие лишь в том, что вместо интегрирования мы использовали суммирование, т.к. пространство состояний остается дискретным.

$$\nu_{t_1, \dots, t_k}(A_1 \times \dots \times A_k) := \int_{A_1 \times \dots \times A_k} p(t_1, \tilde{x}, x_1) p(t_2 - t_1, x_1, x_2) \dots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \dots dx_k .$$

Осталось проверить, что введенные нами конечномерные распределения удовлетворяют условию согласованности, что по теореме Колмогорова является достаточным условием для существования случайного процесса с заданными конечномерными вероятностями. Делается это по аналогии с согласованностью мер для Винеровского процесса.

Теорема 4.5. *Меры, определенные формулой 4.8, удовлетворяют условию согласованности, если матрицы $P(t)$ являются стохастическими и образуют полугруппу, т.е. $P(s)P(t) = P(s+t)$.*

Полугрупповое свойство интуитивно означает следующее: комбинация перехода за s шагов и перехода за t шагов эквивалентно переходу за $s+t$ шагов. Явное доказательство согласованности оставляет читателю.

Упражнение 4.9. Проверьте условие согласованности мер 4.8.

Упражнение 4.10. Проверьте непрерывность $P(t)$.

Заметим, что конечность дискретного пространства состояний не вносила до сих пор каких-либо значительных ограничений, поэтому цепь Маркова можно определять с для счетного числа состояний. Осмысление этого факта оставляем читателю.

Определение 4.6. *Цепь Маркова в непрерывном времени с стохастическими матрицами перехода за t шагов $P(t)$, удовлетворяющими полугрупповому свойству, и начальным распределением μ - это случайный процесс $(X_t)_{t \in \mathbb{R}_+}$ со значениями в \mathbb{Z}_+ , имеющий конечномерные распределения, удовлетворяющие формуле 4.8.*

Упражнение 4.11. Проверьте, что цепь Маркова в непрерывном времени удовлетворяет следующему свойству:

$$P(X_{t_n} = j | X_{t_0} = i_0, X_{t_1} = i_1, \dots, X_{t_{n-1}} = i) = p_{ij}(t_n - t_{n-1}), \quad (4.9)$$

где $p_{ij}(t)$ - это элементы матрицы $P(t)$.

По аналогии с обычной цепью Маркова можно задать стационарное распределение для цепи Маркова с непрерывным временем и сформулировать эргодическую теорему.

Определение 4.7. Стационарное распределение для цепи Маркова с непрерывным временем, заданное переходными вероятностями $P(t)$ и начальным распределением μ , это распределение π , такое, что $\pi P(t) = \pi, \forall t \geq 0$.

Теорема 4.6. Пусть $P(t) > 0$ для некоторого t . Тогда существует единственное стационарное распределение.

Упражнение 4.12. Докажите эргодическую теорему для цепи Маркова с непрерывным временем.

4.2.1 Инфинитезимальная матрица

Рассмотрим цепи Маркова с дифференцируемыми в нуле переходной матрицей $P(t)$.

Определение 4.8. Инфинитезимальная матрица Q для полугруппы $P(t)$ задается следующим образом $Q = \lim_{t \rightarrow 0^+} \frac{P(t) - I}{t}$.

Теорема 4.7. Если Инфинитезимальная матрица Q существует, то существует $P'(t) = P(t)Q = QP(t)$. Данное дифференциальное уравнение вместе с начальным условием $P(t) = I$ имеет решение $P(t) = \exp(tQ)$.

Эта теорема доказывается прямым применением полугруппового свойства.

Теорема 4.8. Если существует инфинитезимальная матрица, то уравнение $\pi Q = 0$ является необходимым и достаточным условием стационарности распределения π .

Для доказательства достаточности обратитесь к определению матричной экспоненты, необходимости - к определению матрицы Q .

Пример 4.3. Процесс Пуассона (марковская цепь с переходными вероятностями $p_{ij}(t) = Pois(j - i; \lambda t)$) обладает следующей инфинитезимальной матрицей:

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

Пуассоновский процесс

На основе Пуассоновского процесса можно понять связь между более общими случайными процессами и цепями Маркова. Эта связь основана на задании процесса с помощью независимых приращений. Мы уже встречались с независимыми приращениями, когда изучали Винеровский процесс. В цепях Маркова аналогом независимых приращений является свойство независимости перехода из состояния $X_{t-1} = i$ в состояние $X_t = j$ от переходов в предыдущие моменты времени.

5.1 Мотивационный пример

Допустим, мы хотим описать поведение n пользователей приложения Яндекс.Такси. Задаем интервал времени (t, s) , например $(00 : 00 - 01 : 30)$. Пусть Y_i - случайная величина, моделирующая решение пользователя вызвать такси в интервале времени (t, s) . Она подчиняется распределению Бернулли с параметром $p_n(t, s) = P(Y_i = 1)$, тогда число вызовов такси в заданный интервал времени можно описать случайной величиной $\nu^n = \sum_{i=0}^n Y_i$. Если число пользователей n велико, то логично рассмотреть предельное распределение случайной величины ν^n . Введем следующее допущение: пусть среднее число заказов такси в заданный интервал времени (t, s) равно $\lambda(s - t)$. Тогда предельное распределение случайной величины ν^n - распределение Пуассона с параметром $\lambda(s - t)$.

5.2 Предельная теорема Пуассона

Теорема 5.1. Пусть $X_i, i = 1, 2, \dots, n$ - набор независимых Бернуллиевских случайных величин с параметром $p_n = P(X_i = 1)$. Определим ν^n как число успехов в n независимых испытаниях Бернулли, т.е. $\nu^n = \sum_{i=0}^n X_i$. Если $\lim_{n \rightarrow \infty} np_n = \lambda$, то $\lim_{n \rightarrow \infty} P\{\nu^n = k\} = \frac{\lambda^k}{k!} \exp(-\lambda)$.

▷ Для начала запишем распределение случайной величины ν_n при фиксированном n :

$$P(\nu_n = k) = \frac{n!}{k!(n-k)!} p_n^k \cdot (1-p_n)^{n-k}.$$

Теперь рассмотрим предел при $n \rightarrow \infty$ первого множителя:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} p_n^k &= \lim_{n \rightarrow \infty} \frac{1}{k!} \cdot n(n-1)\dots(n-k+1)p_n^k = \\ &= \lim_{n \rightarrow \infty} \frac{1}{k!} 1\left(1 - \frac{1}{n}\right)\dots\left(1 - \frac{k-1}{n}\right)(np_n)^k = \frac{1}{k!} \lambda^k. \end{aligned}$$

Рассмотрим второй множитель:

$$\lim_{n \rightarrow \infty} (1 - p_n)^{n-k} = \lim_{n \rightarrow \infty} \exp[(n-k) \ln(1-p_n)] = \lim_{n \rightarrow \infty} \exp\left[\left(1 - \frac{k}{n}\right) \cdot n(-p_n)\right] = \exp(-\lambda).$$

Следовательно,

$$\lim_{n \rightarrow \infty} P(\nu_n = k) = \frac{1}{k!} \lambda^k \exp(-\lambda).$$

□

Определение 5.1. Распределение вероятностей на дискретном множестве \mathbb{Z}_+ , задаваемое формулой

$$p(k; \lambda) = \frac{\lambda^k}{k!} \exp^{-\lambda} \mathbf{1}\{k \geq 0\}, \quad (5.1)$$

называется распределением Пуассона с параметром λ , обозначаемым как $Pois(\lambda)$.

Упражнение 5.1. Покажите, что формула 5.1 в самом деле задает распределение вероятностей.

5.3 Определение процесса Пуассона

Давайте построим процесс Пуассона $(X_t)_{t \in \mathbb{R}}$, задав его приращения $X_t - X_s \quad \forall s < t \in \mathbb{R}_+$, а потом с помощью теоремы Колмогорова докажем, что такой процесс существует.

Определение 5.2. Пусть $(X_t)_{t \in \mathbb{R}_+}$ - случайный процесс с пространством значение $\Xi = \mathbb{Z}_+$ (σ -алгебра \mathcal{F} состоит из всех подмножеств), обладающий следующими свойствами:

1. $X_0 = 0$ почти наверное,
2. $X_t - X_s \sim Pois(\lambda(t-s))$, где $\lambda > 0$ - параметр, $0 \leq s < t$,
3. $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}}$ - независимые случайные величины $\forall 0 \leq t_1 \leq \dots \leq t_k$.

Такой процесс называется процессом Пуассона.

5.3.1 Проверка теоремы Колмогорова

Для начала давайте найдем выражение для мер $\nu_{t_1, \dots, t_k}(A_1, \dots, A_k) = P(X_{t_1} \in A_1, \dots, X_{t_k} \in A_k)$, задающих конечномерные распределения, для упорядоченных моментов времени $0 \leq t_1 \leq t_2 \leq \dots \leq t_k$ (примем $t_0 = 0$). Пусть $\{\eta_i\}_{i=1}^k$ - независимые случайный величины, каждая из которых имеет распределение Пуассона $Pois(\lambda(t_i - t_{i-1}))$ (см. формулу 5.1). С помощью этих величин мы закодируем приращения процесса Пуассона. Рассмотрим

случайный вектор $\eta = (\eta_1, \eta_1 + \eta_2, \dots, \eta_1 + \dots + \eta_k)$. По определению процесса Пуассона его конечномерное распределение для времен t_1, \dots, t_k совпадают с распределениями вектора η .

$$\begin{aligned} P(X_{t_1} \in A_1, \dots, X_{t_k} \in A_k) &= \\ &= \sum_{z_i \in A_i} P(\eta_1 = z_1, \eta_1 + \eta_2 = z_2, \dots, \eta_1 + \dots + \eta_k = z_k) = \\ &= \sum_{z_i \in A_i} P(\eta_1 = z_1, \eta_2 = z_2 - z_1, \dots, \eta_k = z_k - z_{k-1}) = \\ &= \sum_{z_i \in A_i} p(z_1; \lambda t_1) p(z_2 - z_1; \lambda(t_2 - t_1)) \dots p(z_k - z_{k-1}; \lambda(t_k - t_{k-1})). \end{aligned}$$

Несложно показать, что меры согласованы, а следовательно, процесс Пуассона существует.
Давайте сделаем это:

- Для произвольного набора моментов времени (не обязательного упорядоченного) t_1, \dots, t_k найдем такую (сортирующую) перестановку σ , что $\sigma(t_1) \leq \dots \leq \sigma(t_k)$, и зададим меру следующим образом:

$$\nu_{t_1, \dots, t_k}(A_1, \dots, A_k) = \nu_{\sigma(t_1), \dots, \sigma(t_k)}(\sigma(A_1), \dots, \sigma(A_k)).$$

- Пусть t_1, \dots, t_k, t_{k+1} – упорядоченный набор времён и $\Delta t_k := t_{k+1} - t_k$, рассмотрим

$$\begin{aligned} \nu_{t_1, \dots, t_k, t_{k+1}}(A_1, \dots, A_k, \mathbb{Z}_+) &= \\ &= \sum_{z_1 \in A_1, \dots, z_{k+1} \in \mathbb{Z}_+} p(z_1; \lambda t_1) p(z_2 - z_1; \lambda \Delta t_1) \dots p(z_k - z_{k-1}; \lambda \Delta t_{k-1}) p(z_{k+1} - z_k; \lambda \Delta t_k) = \\ &= \sum_{z_1 \in A_1, \dots, z_k \in A_k} \left[p(z_1; \lambda t_1) p(z_2 - z_1; \lambda \Delta t_1) \dots p(z_k - z_{k-1}; \lambda \Delta t_{k-1}) \sum_{z_{k+1} \in \mathbb{Z}_+} p(z_{k+1} - z_k; \lambda \Delta t_k) \right]. \end{aligned}$$

Рассмотрим последнюю сумму:

$$\begin{aligned} \sum_{z_{k+1} \in \mathbb{Z}_+} p(z_{k+1} - z_k; \lambda(t_{k+1} - t_k)) &= \sum_{z_{k+1} \in \mathbb{Z}_+} \frac{\lambda^{(z_{k+1} - z_k)}}{(z_{k+1} - z_k)!} \exp^{-\lambda} \mathbf{1}\{z_{k+1} - z_k \geq 0\} = \\ &= \sum_{z_{k+1} - z_k \geq 0} \frac{\lambda^{(z_{k+1} - z_k)}}{(z_{k+1} - z_k)!} \exp^{-\lambda} = \sum_{m \in \mathbb{Z}_+} \frac{\lambda^m}{m!} \exp^{-\lambda} = 1. \end{aligned}$$

В итоге

$$\begin{aligned} \nu_{t_1, \dots, t_k, t_{k+1}}(A_1, \dots, A_k, \mathbb{Z}_+) &= \\ &= \sum_{z_1 \in A_1, \dots, z_k \in A_k} p(z_1; \lambda t_1) p(z_2 - z_1; \lambda(t_2 - t_1)) \dots p(z_k - z_{k-1}; \lambda(t_k - t_{k-1})) = \\ &= \nu_{t_1, \dots, t_k}(A_1, \dots, A_k). \end{aligned}$$

5.4 Траектории процесса Пуассона

Траектории процесса Пуассона можно рассматривать как пути на графе $G = (V, E)$ с множеством вершин $V = \mathbb{Z}_+$, при этом переходы между вершинами совершаются не в дискретные моменты времени $t \in \mathbb{Z}_+$, а в непрерывные $t \in \mathbb{R}_+$. Путь на графе, описывающий процесс Пуассона, определяет последовательность вершин $V_0 \neq V_1 \neq V_2 \neq V_3 \dots$. Заметим, что по определению процесса Пуассона пути на графике начинаются в вершине $V_0 = 0$ почти наверное. Рассуждая в терминах марковских цепей, можно сказать, что начальное распределение сосредоточено в нуле.

На самом деле путь на графике, соответствующий почти любой реализации процесса Пуассона, является последовательностью $(0, 1, 2, 3, \dots)$. Это означает, что если за ξ_i мы обозначим время, проведенное в вершине $V_{i-1} = i - 1$ до перехода в вершину $V_i = i$, то процесс Пуассона удовлетворяет следующему свойству:

$$X_t = \sup\{n \geq 0 : \sum_{i \leq n} \xi_i \leq t\}. \quad (5.2)$$

При этом время между скачками ξ_i имеет показательное распределение:

$$P(\xi_i = t) = \lambda \exp(-\lambda t) \mathbf{1}\{t \geq 0\}. \quad (5.3)$$

Заметим, что это распределение не зависит от номера скачка i и от момента предыдущего скачка ξ_{i-1} , т.е. случайные величины ξ_i являются независимыми и одинаково распределенными. Таким образом, траектория - кусочно-постоянная, непрерывная справа, скачки равны 1.

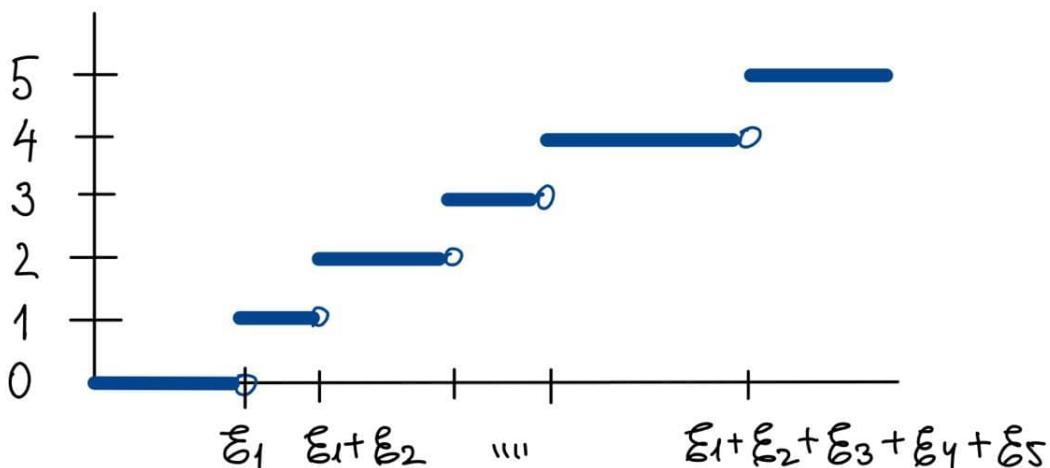


Рис. 5.1: Траектория Пуассоновского процесса

Теорема 5.2. (*Эквивалентное определение процесса Пуассона*)

Процесс Пуассона можно также задать как процесс $(X_t)_{t \in \mathbb{R}_+}$, проводящий случайное время $\xi_i \sim Exp(\lambda)$ в состоянии $i - 1$, а затем попадающий в состояние i , $i = 1, 2, \dots$, т.е.

$$X_t = \sup\{n \geq 0 : \sum_{i \leq n} \xi_i \leq t\},$$

$$P(\xi_i = t) = \lambda \exp(-\lambda t) \mathbf{1}\{t \geq 0\},$$

где ξ_i - независимые случайные величины.

▷

Для начала докажем, что из определения процесса пуассона через моменты скачков следует, что приращения независимые пуассоновские. Запишем распределение для X_t через случайные величины ξ_i :

$$P(X_t = k) = P(\xi_1 + \dots + \xi_k < t < \xi_1 + \dots + \xi_k + \xi_{k+1}).$$

Случайная величина $H_k = \xi_1 + \dots + \xi_k$, определяющая время перехода в состояние k , является суммой независимых экспоненциальных случайных величин с параметром λ , а поэтому имеет гамма-распределение с плотностью $f_k(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$. Также для экспоненциальной случайной величины ξ верно следующее: $P(\xi \geq x) = \exp(-\lambda x)$ при $x > 0$. Из этого следует, что

$$\begin{aligned} P(X_t = k) &= \int_0^t P(\xi_{k+1} > t-x) f_k(x) dx = \int_0^t e^{-\lambda(t-x)} \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} dx = \\ &= \frac{\lambda^k e^{-\lambda t}}{(k-1)!} \int_0^t x^{k-1} dx = \frac{(\lambda t)^k e^{-\lambda t}}{k!}. \end{aligned}$$

Экспоненциальная случайная величина обладает свойством отсутствия памяти, а именно:

$$P(\xi > t+s | \xi > s) = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = P(\xi > t).$$

В нашем случае оно означает, что если скачок не произошел к моменту времени s , то время от момента s до следующего скачка распределено так же, как время от момента $t = 0$ до следующего скачка. То есть если нам дана информация до момента s , то мы можем как бы обнулить наше время ожидания и начать ждать заново.

Из этого свойства для $s = t_{n-1}$, следует независимость приращений:

$$\begin{aligned} P(X_{t_n} - X_{t_{n-1}} = k_n | X_{t_{n-1}} - X_{t_{n-2}} = k_{n-1}, \dots, X_{t_1} - X_0 = k_1) &= \\ &= P(X_{t_n - t_{n-1}} = k_n) = \frac{(\lambda(t_n - t_{n-1}))^{k_n} e^{-\lambda(t_n - t_{n-1})}}{k_n!}. \end{aligned}$$

Теперь докажем утверждение в обратную сторону: если процесс имеет независимые Пуассоновские приращения, то интервалы времени переходов - независимые экспоненциальные

случайные величины. Первым шагом определим распределения для приращений за малый интервал времени.

$$\begin{aligned} P(X_{t+h} - X_t = k) &= \frac{(\lambda h)^k}{k!} \exp(-\lambda h), \\ P(X_{t+h} - X_t = 0) &= \exp(-\lambda h) = 1 - \lambda h + o(h), \\ P(X_{t+h} - X_t = 1) &= (\lambda h) \exp(-\lambda h) = \lambda h + o(h), \\ P(X_{t+h} - X_t \geq 2) &= o(h). \end{aligned}$$

Вторым шагом рассмотрим вероятность того, что на интервале времени $(0, t]$ нет скачков размера больше 1. Для этого представим интервал $(0, t]$ как объединение $m = t/h$ непересекающихся интервалов длины h . На них приращения независимы. Поэтому эта вероятность равна

$$[P(X_{t+h} - X_t = 0) + P(X_{t+h} - X_t = 1)]^m = [1 - \lambda h + \lambda h + o(h)]^m \rightarrow 1, m \rightarrow \infty.$$

Последним шагом найдем совместное распределение интервалов между скачками. Пусть $H_k = \sum_{i=1}^k \xi_i$ - время k -го скачка. Тогда плотность совместного распределения для H_1, H_2, \dots, H_k принимает следующую форму:

$$\begin{aligned} f_{H_1, \dots, H_k}(t_1, \dots, t_k) &= \lim_{h \rightarrow 0} \frac{P(t_1 < H_1 \leq t_1 + h, \dots, t_k < H_k \leq t_k + h)}{h^k} = \\ &= \lim_{h \rightarrow 0} \frac{P(X_{t_1} = 0, X_{t_1+h} - X_{t_1} = 1, \dots, X_{t_k} - X_{t_{k-1}+h} = 0, X_{t_k+h} - X_{t_k} = 1)}{h^k} = \\ &= \lim_{h \rightarrow 0} \frac{\exp(-\lambda t_1) \lambda h \exp(-\lambda h) \dots \exp(-\lambda(t_k - t_{k-1} - h)) \lambda h \exp(-\lambda h)}{h^k} = \\ &= \lambda^k \exp(-\lambda t_k). \end{aligned}$$

Чтобы получить плотность совместного распределения для ξ_1, \dots, ξ_k , достаточно умножить плотность совместного распределения для H_1, H_2, \dots, H_k на якобиан, который равен 1. Поэтому

$$f_{\xi_1, \dots, \xi_k}(s_1, \dots, s_k) = f_{H_1, \dots, H_k}(s_1, \dots, s_1 + \dots + s_k) = \lambda^k \exp(-\lambda(s_1 + \dots + s_k)) = \prod_{i=1}^k [\lambda \exp(-\lambda s_i)].$$

Это означает, что величины ξ_i являются независимыми и экспоненциально распределёнными. \square

5.5 Свойства процесса Пуассона

Упражнение 5.2. Пуассоновский процесс $(X_t^\lambda)_{t \in \mathbb{R}_+}$ с интенсивностью λ можно получить из пуассоновского процесса $(X_t^1)_{t \in \mathbb{R}_+}$ с интенсивностью 1 с помощью замены времени:

$$(X_t^\lambda) \sim (X_{\lambda t}^1).$$

Упражнение 5.3. Пусть $(X_t^{\lambda_1})_{t \in \mathbb{R}_+}$ и $(X_t^{\lambda_2})_{t \in \mathbb{R}_+}$ два независимых Пуассоновских процесса. Тогда их сумма $X_t = X_t^{\lambda_1} + X_t^{\lambda_2}$ является Пуассоновским процессом с параметром $\lambda = \lambda_1 + \lambda_2$.

Упражнение 5.4. Пусть $(X_t^\lambda)_{t \in \mathbb{R}_+}$ - Пуассоновский процесс с параметром λ , а $(Y_t)_{t \in \mathbb{R}_+}$ - это прореженный процесс $(X_t^\lambda)_{t \in \mathbb{R}_+}$, такой, что каждый скачок разрешается с вероятностью p , где $0 < p < 1$. Тогда $(Y_t)_{t \in \mathbb{R}_+}$ - это процесс Пуассона с параметром $(p\lambda)$.

5.6 Процесс Пуассона как марковская цепь с непрерывным временем

Определение 5.3. Q -матрицей называется матрица Q , удовлетворяющая следующим условиям:

1. диагональные элементы неположительны $q_{ii} \leq 0, \forall i$,
2. внедиагональные элементы неотрицательны $q_{ij} \geq 0, \forall i \neq j$,
3. выполнено условие баланса $\sum_j q_{ij} = 0$.

Элементы Q -матрицы q_{ij} описывают скорости перехода из состояния i в состояние j .

Теорема 5.3. Пусть Q является Q -матрицей. Тогда матрица $P(t) = \exp(tQ)$ является стохастической и обладает полугрупповым свойством:

$$P(t+s) = P(s)P(t) \quad \forall s, t \geq 0. \quad (5.4)$$

▷ Напомним определение матричной экспоненты:

$$\exp(tQ) = I + \sum_{k \geq 1} \frac{(tQ)^k}{k!} = \sum_{k \geq 0} \frac{(tQ)^k}{k!}.$$

Полугрупповое свойство следует из определения матричной экспоненты, если использовать биноминальное разложение:

$$((t+s)Q)^k = (t+s)^k Q^k = \sum_{l=0}^k C_k^l t^l s^{k-l} Q^k = \sum_{l=0}^k C_k^l (tQ)^l (sQ)^{k-l}.$$

Тогда

$$\exp((t+s)Q) = \sum_{k \geq 0} \frac{((t+s)Q)^k}{k!} = \sum_{k \geq 0} \frac{1}{k!} \sum_{l=0}^k C_k^l (tQ)^l (sQ)^{k-l} = \sum_{k \geq 0} \sum_{l=0}^k \frac{(tQ)^l}{l!} \frac{(sQ)^{k-l}}{(k-l)!}$$

и

$$\exp(tQ) \cdot \exp(sQ) \sum_{l \geq 0} \frac{1}{l!} (tQ)^l \sum_{m \geq 0} \frac{1}{m!} (sQ)^m = \sum_{k \geq 0} \sum_{l=0}^k \frac{(tQ)^l}{l!} \frac{(sQ)^{k-l}}{(k-l)!}.$$

Теперь займемся доказательством стохастичности матрицы $P(t) = \exp(tQ)$.

Для начала заметим, что по определению $\sum_j q_{ij} = 0$, т.е. сумма элементов в строке матрицы Q равна нулю. То же самое верно и для матрицы Q^n :

$$\sum_j q_{ij}^{(n)} = \sum_{j,l} q_{il}^{(n-1)} q_{lj} = \sum_l q_{il}^{(n-1)} \sum_j q_{lj} = \sum_l q_{il}^{(n-1)} = 0.$$

Т.к. по определению

$$P(t) = \exp(tQ) = I + \sum_{k \geq 1} \frac{(tQ)^k}{k!},$$

сумма элементов в строке матрицы $P(t)$ равна 1.

Теперь докажем, что элементы матрицы $P(t)$ неотрицательны для малых $t > 0$:

$$\begin{aligned} P(t) &= I + tQ + o(t), \\ p_{ij}(t) &= \delta_{ij} + tq_{ij} + o(t). \end{aligned}$$

Это означает, что $p_{ii}(t) > 0$ и $p_{ij}(t) > 0$ для $i \neq j$, при условии $q_{ij} > 0$.

В случае, если $q_{ij} = 0$, мы можем рассмотреть разложение Тейлора для более высоких порядков. Рассуждая по индукции, приходим к выводу: если найдется такое $n > 0$, что $q_{ij}^n > 0$, тогда $p_{ij}(t) > 0$. В противном случае $p_{ij} = 0$.

Чтобы перейти от малых $t > 0$ к произвольным $t > 0$, используем полугрупповое свойство:

$$P(t) = P\left(\frac{t}{n}\right) P\left(\frac{t}{n}\right) \dots P\left(\frac{t}{n}\right) = \left[P\left(\frac{t}{n}\right)\right]^n.$$

Таким образом мы доказали, что $P(t)$ является стохастической матрицей. \square

Определение 5.4. Цепь Маркова в непрерывном времени с Q -матрицей Q и начальным распределением μ - это случайный процесс $(X_t)_{t \in \mathbb{R}_+}$ со значениями в \mathbb{Z}_+ , удовлетворяющий следующим свойствам:

1. $P(X_0 = i) = \mu_i,$
2. $P(X_0 = i_0, X_{t_1} = i_1, \dots, X_{t_n} = i_n) = \mu_{i_0} p_{i_0 i_1}(t_1) \dots p_{i_{n-1} i_n}(t_n - t_{n-1}), 0 < t_1 < \dots < t_n,$

где $p_{ij}(t)$ - это элементы матрицы $P(t) = \exp(tQ)$.

Утверждение 5.4. Траектория Марковской цепи с непрерывным временем с Q -матрицей Q проводит в состоянии i случайное время, распределенное экспоненциально с параметром $q_i = -q_{ii}$, а затем переходит в состояние $j \neq i$ с вероятностью $\frac{q_{ij}}{q_i}$.

Доказательство этого утверждения оставляем читателю.

Опишем конструктивный способ задания марковской цепи с непрерывным временем, аналогичный способу, которым мы строили процесс Пуассона. Зададим следующие случайные величины:

1. ξ_i^n - интервал времени, проведенный в состоянии i , между $(n-1)$ -ым переходом и n -ым переходом, распределен экспоненциально с параметром $-q_{ii}$,
2. H^n - время n -го перехода,
3. η_i^n - состояние после n -го перехода из состояния i , имеет распределение $P(\eta_i^n = j) = -\frac{q_{ij}}{q_{ii}}$,
4. $S^n = \eta_{S^{n-1}}^n$ - n -е состояние.

Тогда процесс можно построить следующим образом:

1. $X_0 = S^0 \sim \mu$ - в начальный момент времени нулевое состояние процесса распределено по μ ,
2. $H^0 = 0, H^n = H^{n-1} + \xi_{S^{n-1}}^n$ - генерируем момент времени n -го перехода,
3. $S^n = \eta_{S^{n-1}}^n$ - генерируем n -ое состояние,
4. $X_t = S^n$ при $t \in [H^n, H^{n+1})$ - процесс находится в состоянии S^n до следующего перехода.

Теорема 5.5. Для Марковской цепи в непрерывном времени верны свойства, аналогичные свойствам Марковской цепи в дискретном времени:

1. $P(X_t = j | X_0 = i) = P(X_{t+s} = j | X_s = i) = p_{ij}(t),$
2. $P(X_{t_n} = j | X_{t_{n-1}} = i_{n-1}, \dots, X_0 = i_0) = P(X_{t_n} = j | X_{t_{n-1}} = i_{n-1}) = p_{ij}(t_n - t_{n-1}),$
3. $P(X_t = j) = (\mu P(t))_j.$

▷ Доказательства этих свойств аналогичны доказательствам свойств цепей Маркова в дискретном времени. □

Пример 5.1. Процесс Пуассона можно описать как Марковскую цепь в непрерывном времени с Q -матрицей вида

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

Матрицу $P(t) = \exp(tQ)$ можно записать в следующем виде:

$$P(t) = \begin{pmatrix} e^{-\lambda t} & \frac{\lambda t}{1!} e^{-\lambda t} & \frac{(\lambda t)}{2!} e^{-\lambda t} & \dots \\ 0 & e^{-\lambda t} & \frac{\lambda t}{1!} e^{-\lambda t} & \dots \\ 0 & 0 & e^{-\lambda t} & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}.$$

Можно заметить, что строки матрицы $P(t)$ описывают распределение Пуассона с параметром (λt) . Доказательство оставляем читателю.

Мартингалы

6.1 Фильтрации, моменты остановки и мартингалы

Как много мы можем сказать о процессе в будущем, если мы уже пронаблюдали какую-то часть его траектории? Формальным ответом на то, что такое информация, является введение фильтрации и более детальное исследование процессов в её контексте.

6.1.1 Фильтрации

Пусть (Ω, \mathcal{F}) - измеримое пространство, T - подмножество \mathbb{R} или \mathbb{Z} .

Определение 6.1. Совокупность σ -подалгебр $\mathcal{F}_t \subseteq \mathcal{F}$, $t \in T$, называется фильтрацией, если $\mathcal{F}_s \subseteq \mathcal{F}_t$ при любых $s \leq t$.

Определение фильтрации означает, что если множество $B \in \mathcal{F}$ измеримо относительно \mathcal{F}_s , то оно также измеримо относительно \mathcal{F}_t при $s \leq t$, однако σ -алгебра \mathcal{F}_t может содержать множества, неизмеримые относительно \mathcal{F}_s .

6.1.2 Момент остановки

Определение 6.2. Случайная величина τ на измеримом пространстве (Ω, \mathcal{F}) со значениями в T называется моментом остановки относительно фильтрации $(\mathcal{F}_t)_{t \in T}$, если $\{\tau \leq t\} \in \mathcal{F}_t$ для любого $t \in T$.

Давайте рассмотрим простейший пример момента остановки. Представим себе игру, каждый раунд которой может закончиться либо выигрышем (получаем 1 доллар), либо проигрышем (теряем 1 доллар). Эту игру закодируем бесконечной последовательностью, элементы которой либо 1 (выигрыш), либо -1 (проигрыш). Пусть τ_1 - первый момент времени, когда накопленный выигрыш составит 1 доллар, а τ_2 - минимальный номер раунда, такой, что следующий раунд будет проигран. Давайте выясним, являются случайные величины τ_1 и τ_2 моментами остановки. Для этого формализуем выше сказанное:

$$\Omega = \{\omega = (\omega_1, \omega_2, \dots), \omega_i \in \{-1, +1\}\},$$

$$\mathcal{F}_n = \sigma(\omega : \omega_1 = a_1, \dots, \omega_n = a_n, a_i \in \{-1, +1\}),$$

$$\mathcal{F} = \sigma(\mathcal{F}_n, n \in \mathbb{N}),$$

$$\tau_1(\omega) = \min\{n : \sum_{i=1}^n \omega_i = 3\},$$

$$\tau_2(\omega) = \min\{n : \omega_{n+1} = -1\},$$

Теперь стало очевидно, что множество $\{\tau_1 \leq n\} \in \mathcal{F}_n$, поэтому τ является моментом остановки, а $\{\tau_2 \leq n\} \notin \mathcal{F}_n$, т.е. τ_2 не является моментом остановки.

Можно доказать много разных свойств про моменты остановки. Давайте для примера рассмотрим следующее:

Утверждение 6.1. *Если τ_1, τ_2 - моменты остановки фильтрации $(\mathcal{F}_t)_{t \in T}$, то $\tau_1 \wedge \tau_2 := \min(\tau_1, \tau_2)$ также является моментом остановки.*

▷

$$\{\tau_1 \wedge \tau_2 \leq t\} = \{\tau_1 \leq t\} \cup \{\tau_2 \leq t\} \in \mathcal{F}_t.$$

□

Аналогичные свойства можно доказать для $\tau_1 \vee \tau_2 := \max(\tau_1, \tau_2)$, $\tau_1 + \tau_2$. Оставим это читателю.

6.1.3 σ -алгебра событий, определенных до момента остановки

Определение 6.3. *Пусть τ - момент остановки фильтрации $(\mathcal{F})_{t \in T}$. Совокупность событий $A \in \mathcal{F}$, для которых $A \cap \{\tau \leq t\} \in \mathcal{F}_t$ для любого $t \in T$, называется σ -алгеброй событий, определенных до момента τ , и обозначается \mathcal{F}_τ .*

Утверждение 6.2. *Выполнены следующие свойства:*

1. \mathcal{F}_τ является σ -алгеброй,
2. τ измерима относительно \mathcal{F}_τ ,
3. если $\tau_1 \leq \tau_2$ - моменты остановки, то $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$

▷ 1. Очевидно, первое свойство выполнено, т.к.

$$\Omega \cap \{\tau \leq t\} \in \mathcal{F}_t.$$

Если $A \in \mathcal{F}_\tau$, то

$$(\Omega \setminus A) \cap \{\tau \leq t\} = \{\tau \leq t\} \setminus (A \cap \{\tau \leq t\}) \in \mathcal{F}_t.$$

Если $A_i \in \mathcal{F}_\tau$, то

$$(\bigcup_{i \in \mathbb{N}} A_i) \cap \{\tau \leq t\} = \bigcup_{i \in \mathbb{N}} (A_i \cap \{\tau \leq t\}) \in \mathcal{F}_t.$$

2. Второе свойство выполнено, т.к.

$$\{\tau \leq c\} \cap \{\tau \leq t\} \in \mathcal{F}_{\min(c,t)} \in \mathcal{F}_t.$$

3. Третье свойство выполнено, т.к. если $A \in \mathcal{F}_{\tau_1}$, то

$$A \cap \{\tau_2 \leq t\} = (A \cap \{\tau_1 \leq t\}) \cap \{\tau_2 \leq t\} \in \mathcal{F}_t.$$

□

6.1.4 Случайный процесс, согласованный с фильтрацией

Определение 6.4. Случайный процесс $(X_t)_{t \in T}$ называется согласованным с фильтрацией $(\mathcal{F}_t)_{t \in T}$, если случайная величина X_t измерима относительно \mathcal{F}_t при любом $t \in T$.

Простым примером фильтрации служит фильтрация, порожденная самим процессом: $\mathcal{F}_t^X := \sigma(X_s, s \leq t)$. Очевидно, что процесс $(X_t)_{t \in T}$ согласован с фильтрацией $(\mathcal{F}_t^X)_{t \in T}$.

Утверждение 6.3. Пусть $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ - согласованный процесс, τ - момент остановки. Тогда случайная величина X_τ измерима относительно \mathcal{F}_τ .

▷

Случайная величина τ принимает целые значения $m \in \mathbb{Z}_+$. Для любого $m \leq n$ верно следующее свойство: $\{X_m \in B\} \in \mathcal{F}_m \subseteq \mathcal{F}_n$ для любого $B \in \mathcal{B}(\mathbb{R})$, т.к. $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ - согласованный процесс. Поэтому верно следующее:

$$\{X_\tau \in B\} \cap \{\tau \leq n\} \in \mathcal{F}_n \quad \text{для любого } B \in \mathcal{B}(\mathbb{R})$$

□

6.1.5 Мартингалы

Определение 6.5. Семейство $(X_t, \mathcal{F}_t)_{t \in T}$ называется мартингалом, если выполнены следующие свойства:

1. $(X_t)_{t \in T}$ согласован с фильтрацией $(\mathcal{F}_t)_{t \in T}$,
2. $X_t \in L_1(\Omega, \mathcal{F}, P)$,
3. $X_s = \mathbb{E}[X_t | \mathcal{F}_s]$ для любых $s \leq t$.

Если в третьем свойстве знак равенства заменить на \leq или \geq , то $(X_t, \mathcal{F}_t)_{t \in T}$ будет называться субмартингалом или супермартингалом соответственно.

Утверждение 6.4. Если $(X_t)_{t \in T}$ - процесс с независимыми приращениями, $\mathbb{E}[X_t] = c$ для любых $t \in T$, тогда $(X_t, \mathcal{F}_t^X)_{t \in T}$ - мартингал.

▷ Пусть $s \leq t$. Тогда верно следующее:

$$\mathbb{E}[X_t | \mathcal{F}_s^X] = \mathbb{E}[X_t - X_s + X_s | \mathcal{F}_s^X] = \mathbb{E}[X_t - X_s | \mathcal{F}_s^X] + X_s = X_s \quad \text{п.н.}$$

□

Примерами процессов, для которых выполнено предыдущее утверждение (а, следовательно, мартингалов) служит винеровский процесс и процесс $Y_t = X_t - \mathbb{E}[X_t]$, где X_t - процесс Пуассона. Примером субмартингала служит процесс, полученный применением выпуклой функции к мартингалу.

Утверждение 6.5. Пусть $(X_t, \mathcal{F}_t)_{t \in T}$ - мартингал, $h : \mathbb{R} \rightarrow \mathbb{R}$ - выпуклая функция, $Y_t := h(X_t)$ интегрируема для любого $t \in T$. Тогда $(Y_t, \mathcal{F}_t)_{t \in T}$ - субмартингал.

▷ Для доказательства достаточно применить условное неравенство Йенсена:

$$Y_s = h(X_s) = h(\mathbb{E}[X_t | \mathcal{F}_f]) \leq \mathbb{E}[h(X_t) | \mathcal{F}_f] = \mathbb{E}[Y_t | \mathcal{F}_f]$$

□

6.2 Мартингалы с дискретным временем

Мы докажем важные свойства мартингалов с дискретным временем. Эти свойства можно обобщить на мартингалы и субмартингалы с непрерывным временем, однако мы оставляем читателю возможность разобраться в этом самостоятельно.

Теорема 6.6. (*Разложение Дуба*)

Если $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ - субмартингал, то существует два случайных процесса M_n и A_n со следующими свойствами:

1. $X_n = M_n + A_n$ при $t \geq 1$,
2. (M_n, \mathcal{F}_n) - мартингал,
3. $A_1 = 0$, случайная величина A_n измерима относительно \mathcal{F}_{n-1} при $n \geq 2$,
4. $A_n(\omega) \leq A_{n+1}(\omega)$ п.н. при всех $n \geq 1$, т.е. A_n не убывает

Более того, если другая пара \bar{M}_n, \bar{A}_n обладает теми же свойствами, то $M_n = \bar{M}_n, A_n = \bar{A}_n$ п.н.

▷ 1. Определим процессы M_n, A_n по индукции.

$$M_1 = X_1, \quad A_1 = 0,$$

$$A_n = A_{n-1} + \mathbb{E}[X_n | \mathcal{F}_{n-1}] - X_{n-1}, \quad M_n = X_n - A_n, \quad n \leq 2$$

Очевидно, эти процессы обладают свойствами 1,3,4. Проверим свойство 2:

$$\mathbb{E}[M_n | \mathcal{F}_{n-1}] = \mathbb{E}[X_n - A_n | \mathcal{F}_{n-1}] = \mathbb{E}[X_n | \mathcal{F}_{n-1}] - A_n = X_{n-1} - A_{n-1} = M_{n-1}.$$

2. Теперь докажем единственность. Для этого покажем, что случайные величины M_n, A_n однозначно определяются по X_{n-1}, X_n

$$A_1 = 0, \quad M_1 = X_1 - A_1 = X_1,$$

$$X_n - X_{n-1} = M_n - M_{n-1} + A_n - A_{n-1},$$

$$\mathbb{E}[X_n - X_{n-1} | \mathcal{F}_{n-1}] = \mathbb{E}[M_n - M_{n-1} + A_n - A_{n-1} | \mathcal{F}_{n-1}],$$

$$\mathbb{E}[X_n | \mathcal{F}_{n-1}] - X_{n-1} = A_n - A_{n-1},$$

$$M_n = X_n - A_n.$$

□

Теорема 6.7. (*Теорема о свободном выборе*) Пусть $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ - субмартингал, σ, τ - моменты остановки, $\sigma \leq \tau \leq k$ для некоторого $k \in \mathbb{N}$. Тогда $X_\sigma \leq \mathbb{E}[X_\tau | \mathcal{F}_\sigma]$.

Если $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ - маргингаль или супермаргингаль, то выполнено такое же утверждение, но с заменой \leq на $=$ или \geq .

▷ Докажем для случая, когда $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ - субмартингал.

Пусть $A \in \mathcal{F}_\sigma$. Введем следующие события для $1 \leq m \leq n$:

$$A_m = A \cap \{\sigma = m\} \quad A_{m,n} = A \cap \{\sigma = m\} \cap \{\tau = n\}$$

$$B_{m,n} = A \cap \{\sigma = m\} \cap \{\tau > n\} \quad C_{m,n} = A \cap \{\sigma = m\} \cap \{\tau \geq n\}$$

Очевидно, $C_{m,n} = A_{m,n} \cup B_{m,n}$, $B_{m,n} = C_{m,n+1}$, $A_m = C_{m,m}$, $C_{m,k+1} = \emptyset$, $\bigcup_{m=1}^k A_m = A$.

Заметим, что $\{\tau > n\} = \Omega \setminus \{\tau \leq n\} \in \mathcal{F}_n$, поэтому $B_{m,n} \in \mathcal{F}_n$.

По определению субматрингала $X_n \leq \mathbb{E}[X_{n+1} | \mathcal{F}_n]$, поэтому верно следующее:

$$\int_{B_{m,n}} X_n dP = \int_{B_{m,n}} X_{n+1} dP,$$

Используя свойства $C_{m,n} = A_{m,n} \cup B_{m,n}$, $B_{m,n} = C_{m,n+1}$, получим:

$$\int_{C_{m,n}} X_n dP \leq \int_{A_{m,n}} X_n dP + \int_{B_{m,n}} X_n dP \leq \int_{A_{m,n}} X_n dP + \int_{B_{m,n}} X_{n+1} dP = \int_{A_{m,n}} X_n dP + \int_{C_{m,n+1}} X_{n+1} dP$$

$$\int_{C_{m,n}} X_n dP - \int_{C_{m,n+1}} X_{n+1} dP \leq \int_{A_{m,n}} X_n dP$$

Возьмем сумму по n от m до k от левой части неравенства:

$$\sum_{n=m}^k [\int_{C_{m,n}} X_n dP - \int_{C_{m,n+1}} X_{n+1} dP] = \int_{C_{m,m}} X_m dP - \int_{C_{m,k+1}} X_{k+1} dP = \int_{C_{m,m}} X_m dP = \int_{A_m} X_m dP$$

Возьмем сумму по n от m до k от правой части неравенства:

$$\sum_{n=m}^k \int_{A_{m,n}} X_n dP = \sum_{n=m}^k \int_{A_m \cap \{\tau=n\}} X_n dP = \int_{A_m} X_\tau dP$$

Получим следующее неравенство:

$$\int_{A_m} X_m dP \leq \int_{A_m} X_\tau dP$$

Теперь сделаем суммирование по m от 1 до k :

$$\sum_{m=1}^k \int_{A_m} X_m dP = \sum_{m=1}^k \int_{A \cap \{\sigma=m\}} X_m dP = \int_A X_\sigma dP$$

$$\sum_{m=1}^k \int_{A_m} X_\tau dP = \int_A X_\tau dP$$

Получим следующее неравенство:

$$\int_A X_\sigma dP \leq \int_A X_\tau dP$$

Т.к. множество A было выбрано произвольно, можно сделать вывод, что $X_\sigma \leq \mathbb{E}[X_\tau | \mathcal{F}_\sigma]$.

Для доказательства аналогичного свойства для супермартингала $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ достаточно заметить, что $(-X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ является субмартингалом. Для доказательства аналогичного свойства для мартингала $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ достаточно заметить, что $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ является одновременно субмартингалом и супермартингалом.

□

Теорема о свободном выборе верна для неограниченных σ, τ при условии, что случайные величины X_n , $n \in \mathbb{Z}_+$ равномерно интегрируемы.

Определение 6.6. Множество случайных величин $\{f_s\}_{s \in S}$ называется равномерно интегрируемым, если

$$\lim_{\lambda \rightarrow \infty} \sup_{s \in S} \int_{|f_s| > \lambda} |f_s| dP = 0$$

Введем следующее обозначение: $A(\lambda, n) = \{\omega : \max_{1 \leq i \leq n} X_i(\omega) \geq \lambda\}$.

Теорема 6.8. (Неравенство Дуба) Пусть $(X_n, \mathcal{F}_n)_{n \in \mathbb{Z}_+}$ - субмартингал. Тогда для любого $n \in \mathbb{N}$ и $\lambda > 0$ выполняется неравенство:

$$\lambda P(A(\lambda, n)) \leq \int_{A(\lambda, n)} X_n dP \leq \max(X_n, 0)$$

▷ Введем следующие моменты остановки:

1. σ_0 - первый момент, когда $X_i \geq \lambda$
2. $\sigma = \sigma_0 \wedge n$
3. $\tau = n$

Поскольку $\sigma \leq \tau$, можно применить теорему о свободном выборе:

$$X_\sigma \leq \mathbb{E}[X_n | \mathcal{F}_\sigma]$$

Напомним, что $A(\lambda, n) = \{\omega : \max_{1 \leq i \leq n} X_i(\omega) \geq \lambda\}$. Заметим, что $A(\lambda, n) \in \mathcal{F}_\sigma$, т.к.

$$A(\lambda, n) \cap \{\sigma \leq m\} = \{\omega : \max_{1 \leq i \leq m} X_i(\omega) \geq \lambda\}.$$

Поэтому верно следующее:

$$\int_{A(\lambda, n)} X_\sigma dP \leq \int_{A(\lambda, n)} X_n dP$$

Очевидно, что можно записать следующее неравенство:

$$\int_{A(\lambda, n)} X_n dP \leq \int_{A(\lambda, n)} \max(X_n, 0) dP \leq \mathbb{E}[\max(X_n, 0)]$$

Если $\omega \in A(\lambda, n)$, то $X_{\sigma(\omega)}(\omega) \geq \lambda$. Следовательно,

$$\int_{A(\lambda, n)} X_\sigma dP \geq \int_{A(\lambda, n)} \lambda dP = \lambda P(A(\lambda, n))$$

А теперь соберем все вместе:

$$\lambda P(A(\lambda, n)) \leq \int_{A(\lambda, n)} X_\sigma \leq \int_{A(\lambda, n)} X_n dP \leq \mathbb{E}[\max(X_n, 0)]$$

□

Мартингалы: 2

Оказывается, что мартингалы в силу своего определяющего свойства обладают интересными предельными свойствами. В этой лекции мы немного исследуем сходимость мартингалов, а также посмотрим, как уже известные факты можно перенести на случай непрерывного времени.

7.1 Сходимость мартингалов и непрерывные мартингалы

7.1.1 Сходимость мартингалов

Одним из первых и очевидных примеров мартингалов является процесс $X_n = \mathbb{E}[X|\mathcal{F}_n]$ для интегрируемой случайной величины X . Его смысл заключается в следующем: распределение случайной величины X содержит в себе всю информацию, а распределение X_n - только информацию до момента времени t . Естественно задаться таким вопросом: каким условиям должен удовлетворять мартингал X_t , чтобы существовала случайная величина X со свойством $X_n = \mathbb{E}[X|\mathcal{F}_n]$? Этим вопросом мы и займемся.

Формализацией понятия информации к моменту времени n является сигма-алгебра \mathcal{F}_n (как элемент фильтрации). Тогда всю информацию естественно описывать сигма-алгеброй, порожденной объединением \mathcal{F}_n .

Определение 7.1. Замыканием справа мартингала (X_n, \mathcal{F}_n) называется случайная величина $X'_\infty \in L_1(\Omega, \mathcal{F}, P)$, такая, что $X_n = \mathbb{E}[X'_\infty | \mathcal{F}_n]$, при этом случайная величина $X_\infty = \mathbb{E}[X'_\infty | \mathcal{F}_\infty]$ так же является замыканием.

Если замыкание существует, то принято считать, что оно \mathcal{F}_∞ -измеримо.

Теорема 7.1. Замыкание X_∞ мартингала (X_n, \mathcal{F}_n) существует тогда и только тогда, когда он равномерно интегрируем.

Докажет теорему только в одну сторону (более простую).

▷

Пусть замыкание существует, мы хотим показать равномерную интегрируемость, т.е. $\lim_{\lambda \rightarrow \infty} \sup_{n \in \mathbb{N}} \int_{|X_n| > \lambda} |X_n| dP = 0$. Давайте для начала оценим $|X_n|$, применив определение замыкания $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$, неравенство Йенсена и условное неравенство Йенсена:

$$|X_n| = |\mathbb{E}[X_\infty | \mathcal{F}_n]| \leq \mathbb{E}[|X_\infty| | \mathcal{F}_n]$$

$$\int_{|X_n| > \lambda} |X_n| dP \leq \int_{|X_n| > \lambda} |X_\infty| dP.$$

Т.к. $X_\infty \in L_1(\Omega, \mathcal{F}, P)$, достаточно показать, что $\limsup_{\lambda \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{P}|X_n| > \lambda = 0$. Сделаем это с помощью неравенства Маркова и оценки на $|X_n|$, полученной в начале доказательства:

$$\mathbb{P}\{|X_n| > \lambda\} \leq \frac{\mathbb{E}[|X_n|]}{\lambda} \leq \frac{\mathbb{E}[|X_\infty|]}{\lambda}$$

□

Теперь рассмотрим, предельные теоремы для мартингалов.

Теорема 7.2. Пусть (X_n, \mathcal{F}_n) - замыкаемый мартингал, тогда $X_n \xrightarrow{n.h., L_1} X_\infty$

Теорема 7.3. Пусть (X_n, \mathcal{F}_n) - L_1 -ограниченный мартингал, тогда $X_n \xrightarrow{n.h.} Y$, причем $Y \in L_1$.

Доказательство этих теорем слишком громоздко, однако мы можем поупражняться в применении разложения Дуба и доказать следующие следствия.

Следствие 7.4. Пусть (X_n, \mathcal{F}_n) - L_1 -ограниченный субмартингал, тогда $X_n \xrightarrow{n.h., L_1} Y$

▷ Давайте сведем утверждение к соответствующей теореме для мартингалов, применив разложение Дуба.

$$X_n = M_n + A_n \in L_1$$

Видим, что достаточно рассмотреть по-отдельности сходимость A_n и сходимость M_n . Начнем с A_n . Разложение Дуба нам говорит о том, что A_n не убывает. Поэтому, чтобы доказать, что $A_n \xrightarrow{n.h.} A \in L_1$, достаточно показать L_1 -ограниченность A_n . Сделаем это используя L_1 -ограниченность X_n и свойства мартингала M_n

$$\mathbb{E}[|A_n|] \leq \mathbb{E}[|X_n|] + \mathbb{E}[|M_n|] = \mathbb{E}[|X_n|] + \mathbb{E}[|M_1|] < c$$

Теперь рассмотрим поведение $M_n = X_n - A_n$. Из L_1 -ограниченности X_n и A_n следует L_1 -ограниченность M_n . Поэтому мартингал $M_n \xrightarrow{n.h.} M \in L_1$.

В итоге, $X_n = M_n + A_n \xrightarrow{n.h.} M + A = Y \in L_1$.

□

Следствие 7.5. Пусть (X_n, \mathcal{F}_n) - равномерно интегрируемый субмартингал, тогда $X_n \xrightarrow{n.h., L_1} Y$

▷

Доказательство аналогично доказательству предыдущего следствия. Начнем с разложения Дуба:

$$X_n = M_n + A_n \in L_1$$

В ходе доказательства предыдущего следствия мы уже выяснили, что A_n ограничены в L_1 . Это означает, что A_n равномерно интегрируемы (т.к. они неотрицательны и монотонно не убывают). Поэтому $M_n = X_n - A_n$ является равномерно интегрируемым мартингалом, а следовательно, $M_n \xrightarrow{\text{п.н., } L_1} M_\infty$. Поэтому $X_n = M_n + A_n \xrightarrow{\text{п.н., } L_1} M_\infty + A = Y \in L_1$. □

Давайте рассмотрим пример мартингала с дискретным временем, называемый уроновой схемой Пойа. Пусть у нас есть коробка, в которой в момент времени $n = 0$ лежат 1 черный шар и 1 белый шар. На шаге n мы вынимаем из коробки случайный шар и кладем в коробку вместо него 2 шага того же цвета. Введем следующие обозначения: A_n - число шаров черного цвета в коробке в момент времени n , B_n - число шаров белого цвета в коробке в момент времени n , $X_n = \frac{A_n}{A_n + B_n} = \frac{A_n}{n+2}$ - доля черных шаров в коробке в момент времени n , ξ_n - индикатор события "на шаге n вынули черный шар". Тогда верны следующие рекурентные уравнения:

$$A_0 = B_0 = 1, A_n = A_{n-1} + \xi_n, B_n = B_{n-1} + (1 - \xi_n)$$

Пусть сигма алгебра \mathcal{F}_n содержит всю информацию до шага n , т.е. $\mathcal{F}_n = \sigma\{A_k, B_k, k \leq n\}$. Заметим, что распределение ξ_n зависит только от того, сколько черных и белых шаров было в коробке в момент $n - 1$, а именно:

$$P(\xi_n = 1 | \mathcal{F}_{n-1}) = \frac{A_{n-1}}{A_{n-1} + B_{n-1}}, P(\xi_n = 0 | \mathcal{F}_{n-1}) = \frac{B_{n-1}}{A_{n-1} + B_{n-1}}$$

Теперь мы можем доказать, что X_n является мартингалом.

$$\begin{aligned} \mathbb{E}[X_n - X_{n-1} | \mathcal{F}_{n-1}] &= \mathbb{E}\left[\frac{A_n}{A_n + B_n} - \frac{A_{n-1}}{A_{n-1} + B_{n-1}} | \mathcal{F}_{n-1}\right] = \mathbb{E}\left[\frac{A_{n-1} + \xi_n}{A_{n-1} + B_{n-1} + 1} - \frac{A_{n-1}}{A_{n-1} + B_{n-1}} | \mathcal{F}_{n-1}\right] = \\ &= \mathbb{E}\left[\frac{(A_{n-1} + \xi_n)(A_{n-1} + B_{n-1}) - A_{n-1}(A_{n-1} + B_{n-1} + 1)}{(A_{n-1} + B_{n-1} + 1)(A_{n-1} + B_{n-1})} | \mathcal{F}_{n-1}\right] = \\ &= \mathbb{E}\left[\frac{\xi_n(A_{n-1} + B_{n-1}) - A_{n-1}}{(A_{n-1} + B_{n-1} + 1)(A_{n-1} + B_{n-1})} | \mathcal{F}_{n-1}\right] = \frac{1}{A_{n-1} + B_{n-1} + 1} \mathbb{E}[\xi_n - \frac{A_{n-1}}{A_{n-1} + B_{n-1}} | \mathcal{F}_{n-1}] = 0 \end{aligned}$$

X_n мартингал, равномерно интегрируем (т.к. ограничен единицей), а следовательно, существует его замыкание X_∞ , и $X_n \xrightarrow{\text{п.н., } L_1} X_\infty$.

Можно показать, что $X_\infty \sim U[0, 1]$. Для этого докажем, что $P(A_n = k) = \frac{1}{n+1}$ при $k \in \{1, 2, \dots, n+1\}$, используя индукцию.

$$P(A_0 = 1) = 1$$

$$P(A_n = k) = P(A_{n-1} = k-1, \xi_n = 1) + P(A_{n-1} = k, \xi_n = 0) = \frac{1}{n} \frac{k-1}{n+1} + \frac{1}{n} \frac{n+1-k}{n+1} = \frac{1}{n+1}$$

Тогда $P(X_n = \frac{k}{n+2}) = P(A_n = k) = \frac{1}{n+1}$.

7.1.2 Мартингалы с непрерывным временем

Давайте сформулируем утверждения для мартингалов с непрерывным временем, аналогичные утверждениям для мартингалов с дискретным временем.

Для начала нам потребуется ввести дополнительное (довольно техническое) определение непрерывной справа фильтрации.

Определение 7.2. Непрерывной справа фильтрацией называется фильтрация $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$, такая, что $\mathcal{F}_t = \bigcap_{s < t} \mathcal{F}_s$.

Рассмотрим пример, как можно применять это свойство фильтрации.

Утверждение 7.6. Пусть фильтрация $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$ непрерывна справа. Тогда случайная величина T является моментом остановки тогда и только тогда, когда $\{T < t\} \in \mathcal{F}_t$.

▷ Пусть T является моментом остановки, т.е. $\{T \leq t\} \in \mathcal{F}_t$. Тогда $\{T < t\} = \bigcap_{s < t} \{T \leq s\} \in \mathcal{F}_t$.

Пусть $\{T < t\} \in \mathcal{F}_t$. Тогда $\{T \leq t\} = \bigcap_{s < t} \mathcal{F}_s = \mathcal{F}_t$ (по свойству непрерывности фильтрации). \square

Далее сформулируем обычные условия на фильтрацию, т.е. условия, которые обычно накладываются на фильтрацию в теоремах про непрерывные мартингалы.

Определение 7.3. Фильтрация $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$ удовлетворяет обычным условиям, если она непрерывна справа и \mathcal{F}_0 содержит все P -пренебрежимые множества.

Полезно рассматривать непрерывные права процессы. Рассмотрим следующую теорему без доказательства.

Теорема 7.7. Пусть фильтрация $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$ удовлетворяет обычным условиям. Тогда если (X_t, \mathcal{F}_t) - субмартингал с непрерывным справа матожиданием (т.е. $\mathbb{E}[X_t]$ непрерывно справа как функция от времени), то у него существует непрерывная справа модификация.

Пришло время сформулировать теоремы, аналогичные теоремам для дискретных мартингалов.

Теорема 7.8. (*Разложение Дуба - Мейера*).

Пусть (X_t, \mathcal{F}_t) - непрерывный субмартингал, $X_{\tau \in S_a}$ равномерно интегрируемо для любого a , где $S_a = \{\tau \leq a, \tau \text{ момент остановки}\}$. Тогда существуют непрерывные согласованные процессы M_t, A_t , такие, что

1. $X_t = M_t + A_t$ при $t \geq 0$ п.н.,
2. (M_t, \mathcal{F}_t) - мартингал,
3. $A_0 = 0$,
4. $A_s(\omega) \leq A_t(\omega)$ п.н. при всех $s \leq t$, т.е. A_t не убывает.

Более того, если другая пара \bar{M}_t, \bar{A}_t обладает теми же свойствами, то $M_t = \bar{M}_t, A_t = \bar{A}_t$ п.н.

Если применить разложение Дуба-Мейера к квадрату квадратично интегрируемого непрерывного мартингала Y_t , то можно доказать, что в разложении $Y_t^2 = M_t + A_t$ процесс A_t будет являться квадратичной вариацией процесса Y_t , т.е. $A_t = \lim_{|\pi| \rightarrow 0} \sum_{k=0}^{m-1} |Y_{t_{k+1}} - Y_k|^2$, где $\pi = \{0 = t_0 < t_1 < \dots < t_m = t\}$ - разбиение интервала $[0, t]$. В частности, квадратическая вариация Винеровского процесса это просто t , поэтому $W_t^2 - t$ является мартингалом, т.к. $\mathbb{E}[W^{t_2} - t | \mathcal{F}_{s_1}] = \mathbb{E}[(W_{s_1} + W^{t_2} - W_{s_1})^2 | \mathcal{F}_{s_1}] - t = W_{s_1}^2 + 2W_{s_1}\mathbb{E}[W^{t_2} - W - s_1 | \mathcal{F}_{s_1}] + \mathbb{E}[(W^{t_2} - W_{s_1})^2 | \mathcal{F}_{s_1}] - t = W_{s_1}^2 + t - s_1 - t = W_{s_1}^2 - s_1$. Мы вернемся к этому утверждению во время обсуждения стохастических интегралов.

Теорема 7.9. (*Теорема о свободном выборе*). Пусть (X_t, \mathcal{F}_t) - непрерывный справа субмартингал, $\leq \tau \leq r$ - ограниченные моменты остановки. Тогда $X_{\tau} \leq \mathbb{E}[X_{\tau} | \mathcal{F}_r]$.

Заметим, что так же, как и в случае дискретным мартингалов, это утверждение остается верным для неограниченных моментов остановки, если потребовать равномерную интегрируемость субмартингала.

Теорема 7.10. (*Неравенство Дуба*)

Пусть (X_t, \mathcal{F}_t) - непрерывный справа субмартингал. Тогда для любого $t \geq 0$ и $\lambda > 0$ выполняется неравенство:

$$\lambda P(A(\lambda, t)) \leq \int_{A(\lambda, t)} X_t dP \leq \max(X_t, 0)$$

$$\text{где } A(\lambda, t) = \{\omega : \sup_{0 \leq s \leq t} X_s(\omega) \geq \lambda\}.$$

Теорема 7.11. (*Сходимость равномерно интегрируемого мартингала*)

Пусть (X_t, \mathcal{F}_t) - непрерывный справа замыкаемый мартингал, фильтрация $\mathbb{F} = (\mathcal{F}_t)$ удовлетворяет обычным условиям. Тогда $X_n \xrightarrow{n.h., L_1} X_{\infty}$.

Теорема 7.12. (*Сходимость ограниченного мартинала*)

Пусть (X_t, \mathcal{F}_t) - непрерывный справа L_1 -ограниченный мартинал, фильтрация $\mathbb{F} = (\mathcal{F}_t)$ удовлетворяет обычным условиям. Тогда $X_n \xrightarrow{n.n.} Y$, причем $Y \in L_1$.

Следствие 7.13. Пусть (X_n, \mathcal{F}_n) - непрерывный L_1 -ограниченный субмартинал, фильтрация $\mathbb{F} = (\mathcal{F}_t)$ удовлетворяет обычным условиям, тогда $X_n \xrightarrow{n.n., L_1} Y$.

Следствие 7.14. Пусть (X_n, \mathcal{F}_n) - непрерывный равномерно интегрируемый субмартинал, фильтрация $\mathbb{F} = (\mathcal{F}_t)$ удовлетворяет обычным условиям, тогда $X_n \xrightarrow{n.n., L_1} Y$.

Давайте рассмотрим простой пример непрерывных мартиналов.

Пример 7.1. Докажем, что $X_t = e^{\theta W_t - \theta^2 t/2}$ является мартиналом, используя характеристическую функцию:

$$\begin{aligned}\mathbb{E}[X_t | \mathcal{F}_s] &= \mathbb{E}[e^{\theta W_t - \theta^2 t/2} | \mathcal{F}_s] = \mathbb{E}[e^{\theta(W_s + W_t - W_s) - \theta^2 t/2} | \mathcal{F}_s] = e^{\theta W_s} + \mathbb{E}[e^{\theta(W_t - W_s)} | \mathcal{F}_s] - e^{\theta^2 t/2} = \\ &= e^{\theta W_s} + e^{\theta^2(t-s)/2} - e^{\theta^2 t/2} = e^{\theta W_s - \theta^2 s/2}\end{aligned}$$

Рассмотрим любопытный пример, демонстрирующий другой способ определения Винеровского процесса.

Пример 7.2. Пусть $X_0 = 0$, $Y_t^\lambda = e^{i\langle \lambda, X_t \rangle + \|\lambda\|^2 t/2}$ является мартиналом при любом λ . Докажем, что из этого следует то, что X_t Винеровским процессом.

Для начала проверим нормальность приращений, т.е. $X_t - X_s \sim \mathcal{N}(0, (t-s)I)$. Для этого воспользуемся мартинальным свойством процесса Y_t^λ и формулой полного матожидания:

$$\mathbb{E}[Y_t | \mathcal{F}_s] = Y_s$$

$$\mathbb{E}[e^{i\langle \lambda, X_t \rangle + \|\lambda\|^2 t/2} | \mathcal{F}_s] = e^{i\langle \lambda, X_s \rangle + \|\lambda\|^2 s/2}$$

$$\mathbb{E}[e^{i\langle \lambda, X_t - X_s \rangle} | \mathcal{F}_s] = e^{-\|\lambda\|^2(t-s)/2}$$

$$\phi_{X_t - X_s}(\lambda) = \mathbb{E}[e^{i\langle \lambda, X_t - X_s \rangle}] = e^{-\|\lambda\|^2(t-s)/2}$$

Теперь докажем независимость приращений, т.е. $X_t - X_s \perp \mathcal{F}_s$ при $s \leq t$, что означает $X_t - X_s \perp W$ для любой \mathcal{F}_s -измеримой случайной величины W . Для этого воспользуемся следующим свойством характеристических функций: $X_t - X_s \perp W$ тогда и только тогда, когда характеристическая функция случайного вектора $(X_t - X_s, W)$ разбивается

в произведение характеристической функции $X_t - X_s$ и характеристической функции W . Рассмотрим случайный вектор $(X_t - X_s, W)$. В точке $\lambda = (\lambda_1, \lambda_2)$ характеристическая функция этого вектора принимает следующий вид:

$$\phi_{(X_t - X_s, W)}(\lambda) = E[e^{i\langle \lambda, (X_t - X_s, W) \rangle}] = E[e^{i\langle \lambda_1, X_t - X_s \rangle} e^{i\langle \lambda_2, W \rangle}]$$

Так как W является \mathcal{F}_s -измеримой случайной величиной, полезно воспользоваться формулой полного матожидания, тогда мы сможем применить следующее уже доказанное утверждение:

$$\mathbb{E}[e^{i\langle \lambda, X_t - X_s \rangle} | \mathcal{F}_s] = e^{-\|\lambda\|^2(t-s)/2} = \mathbb{E}[e^{i\langle \lambda, X_t - X_s \rangle}]$$

Применим формулу полного матожидания:

$$E[e^{i\langle \lambda_1, X_t - X_s \rangle} e^{i\langle \lambda_2, W \rangle}] = E[E[e^{i\langle \lambda_1, X_t - X_s \rangle} e^{i\langle \lambda_2, W \rangle} | \mathcal{F}_s]] = E[e^{i\langle \lambda_2, W \rangle} E[e^{i\langle \lambda_1, X_t - X_s \rangle} | \mathcal{F}_s]] =$$

$$= E[e^{i\langle \lambda_2, W \rangle} E[e^{i\langle \lambda_1, X_t - X_s \rangle}]] = E[e^{i\langle \lambda_1, X_t - X_s \rangle}] E[e^{i\langle \lambda_2, W \rangle}] = \phi_{X_t - X_s}(\lambda_1) \phi_W(\lambda_2)$$

В итоге мы доказали, что $\phi_{(X_t - X_s, W)}(\lambda) = \phi_{X_t - X_s}(\lambda_1) \phi_W(\lambda_2)$ для любой \mathcal{F}_s измеримой случайной величины W , откуда следует независимость приращений процесса X_t .

Процессы Леви

Мы уже знакомы с обширным арсеналом нетривиальных случайных процессов, среди которых были Винеровский и Пуассоновский процессы. В этой лекции мы посмотрим на более общий класс процессов: процессы Леви, – и поймём как далеко может завести идея независимых и стационарных приращений.

8.1 Хорошо иметь хорошие разрывы

Мы строили Винеровский и Пуассоновский процессы очень похожим образом, но за счёт разных деталей получали принципиально разные свойства траекторий. У первого непрерывные траектории, у второго – кусочно константные и причём неубывающие. Для моделирования финансовых рядов мы до сих пор использовали Винеровский процесс, но он имеет существенный недостаток: траектории непрерывны; процесс GBM как экспонента непрерывной функции тоже будет иметь непрерывные траектории. При этом давайте взглянем на один типичный финансовый ряд:

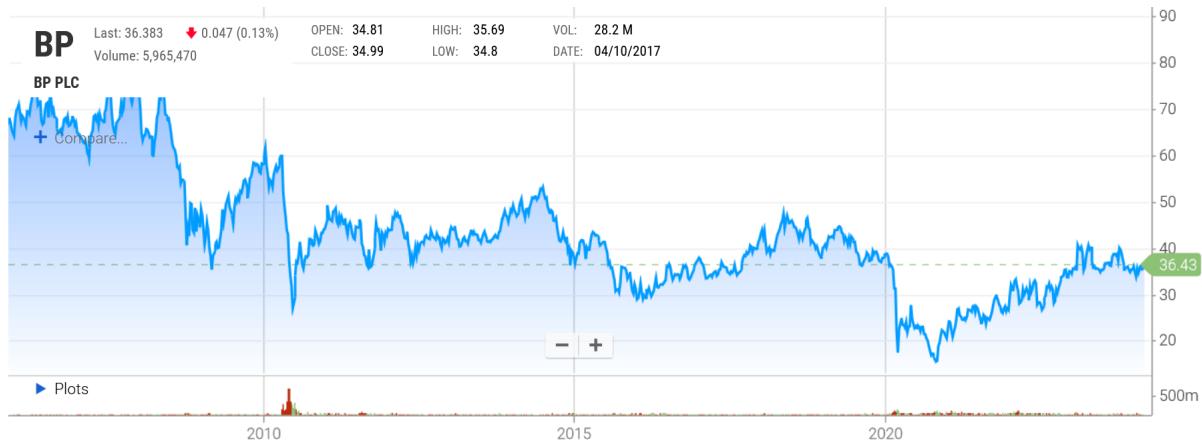


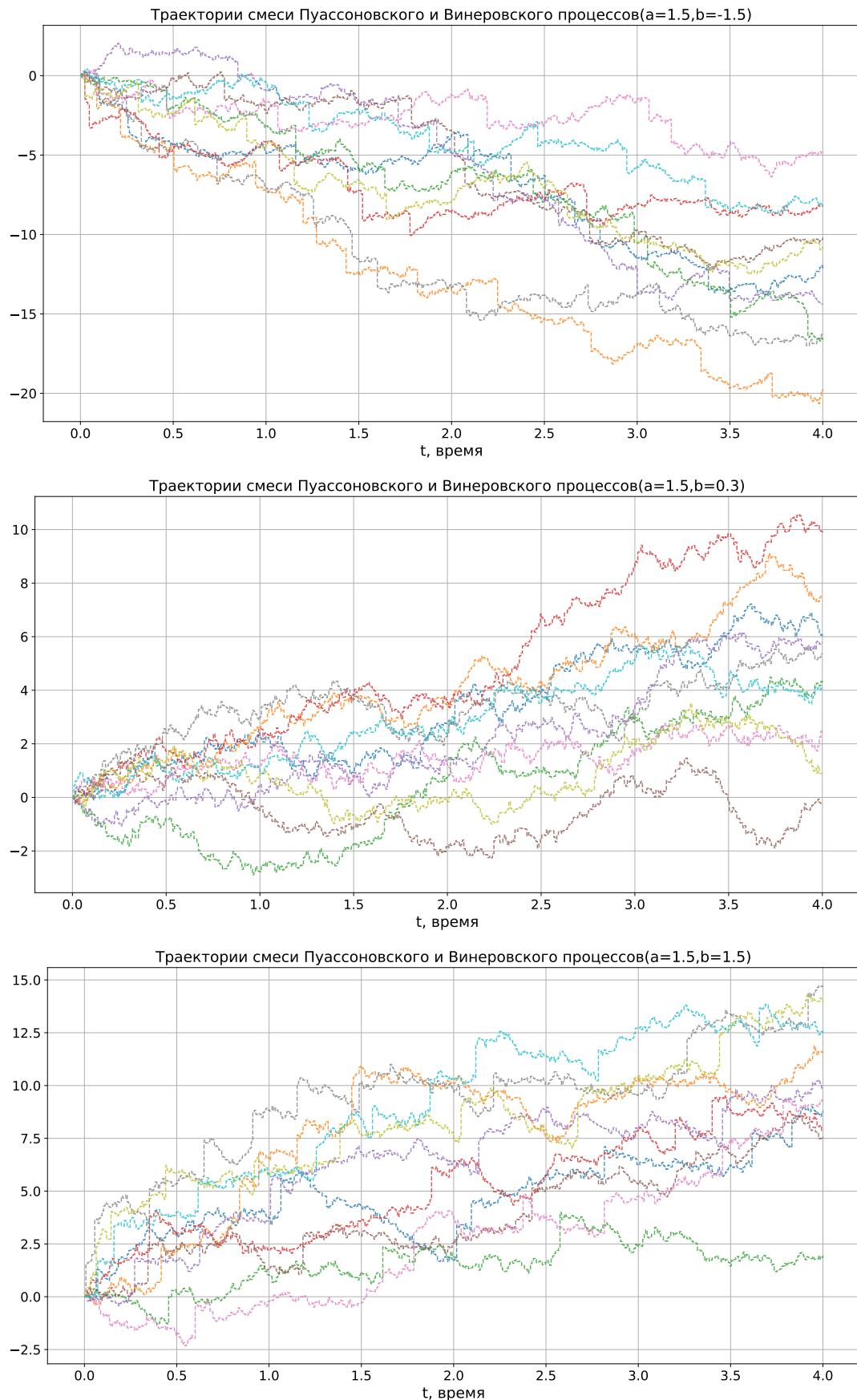
Рис. 8.1: Цены торгов акций British Petroleum, биржа NASDAQ

В данных заметны резкие прыжки цен за счёт каких-то шоковых событий, приводящих инвесторов к резким решениям, из-за которых цена актива может резко падать или возрастать, а потом снова приходить в более спокойный режим. Мы могли бы, например, использовать процесс Пуассона N_t , рассмотрев процесс

$$X_t = aW_t + bN_t, \quad a, b \in \mathbb{R}.$$

Процесс X_t будет иметь прыжки в траекториях, причём в силу свойств траекторий N_t и W_t , траектории X_t будут непрерывны справа и иметь не совпадающий с ним предел слева. Но

для практических целей ни один такой процесс не подойдёт: прыжки происходят только в одну сторону (в зависимости от знака b) и прыжки имеют каждый раз один и тот же размер b (ниже $b = 3$).



Мы хотели бы построить процесс более общего вида, который был бы лишен таких ограничивающих черт и был бы более гибким с точки зрения большого количества возможных различных процессов такого класса.

Процесс Леви мы строим таким же образом как и раньше, но дополнительно вводим требование на свойство траекторий. Мы будем рассматривать везде $T = \mathbb{R}_{\geq 0}$ в качестве индексного пространства.

Определение 8.1. Процесс X_t называется процессом Леви, если

1. $X_0 = 0$ почти наверное;
2. Для $t > s$ приращения $X_t - X_s$ независимы и имеют то же распределение, что X_{t-s} , то есть, стационарны;
3. Траектории процесса X_t в каждой точке непрерывны справа и имеют предел слева (свойство càdlàg, от французского *continu à droite et limite à gauche*).

К этому моменту, нам уже ясно, что такие процессы существуют. Действительно, Винеровский и Пуассоновский процесс – это процессы Леви, потому что мы их строили ровно так: процессы с независимыми и стационарными приращениями, – и получали совсем разные процессы. Траектории càdlàg, поэтому они также обладают свойством непрерывности справа по вероятности.

Лемма 8.1. Если X_t – Процесс Леви, то

$$\forall t, \varepsilon > 0 \quad \lim_{s \rightarrow 0+} \mathbb{P}(|X_{t+s} - X_t| > \varepsilon) = 0.$$

Помимо этого, независимость и стационарность приращений даёт также многообещающее свойство бесконечной делимости сейчений.

Лемма 8.2. Характеристическая функция X_t для любого n равна

$$\mathbb{E}[e^{i\xi X_t}] = \mathbb{E}[e^{i\xi X_{t/n}}]^n.$$

Это свойство мы подробно будем обсуждать в следующей лекции и оно позволит получить очень удобное представление процесса Леви, позволяющее гибко моделировать любой процесс такого класса.

8.2 От классики к процессам Леви: Пуассоновский процесс

Пуассоновский процесс – это процесс Леви с неубывающими и кусочно константными траекториями, причём значение в точке t если изменяется, то всегда ровно на 1. Это частный случай так называемых *считывающих процессов*.

Определение 8.2. Если задана неубывающая последовательность случайных времён $(T_n)_{n \geq 1}$ и $T_n \rightarrow \inf$ почти наверное, процесс X_t вида

$$X_t = \sum_{i=1}^{\infty} \mathbb{1}(t \geq T_i)$$

называется считающим процессом.

Мы уже встречали считающие процессы, когда обсуждали Пуассоновский процесс. Можно представлять себе, что T_n – время происхождение некоторого события: например, в магазин зашёл посетитель (вообще, это случайно происходит). Тогда X_t – это количество людей, посетивших магазин до момента t .

Пример 8.1. Пуассоновский процесс N_t с интенсивностью λ – это считающий процесс. Действительно, зададим последовательность $\Delta_n \sim^{iid} \text{Exp}(\lambda)$ и времена

$$T_n = \sum_{i=1}^n \Delta_i.$$

Процесс

$$N_t = \sum_{i=1}^{\infty} \mathbb{1}(t \geq T_i)$$

в точности задаёт Пуассоновский процесс. Величины Δ_i – это интервалы между приходом клиентов, при этом мы видели, что вероятность того, что больше одного клиента одновременно войдут в двери, равна нулю.

Считающий процесс вообще не обязан быть процессом Леви, потому что времена T_i могут быть сложно зависимыми. Пуассоновский процесс в каком-то смысле уникальный.

Утверждение 8.3. Если процесс N – это считающий процесс и процесс Леви, то это Пуассоновский процесс.

▷ Мы предположим сейчас, что время до следующего прыжка – абсолютно непрерывная случайная величина, более общий случай обсудим в следующей лекции. Будем пользоваться конструкцией Пуассоновского процесса через интервалы прыжков. Ещё мы воспользуемся одним свойством экспоненциального распределения, которое мы уже встречали ранее: если неотрицательная абсолютно непрерывная случайная величина T для всех $t, s \geq 0$ удовлетворяет

$$\mathbb{P}(T > t + s \mid T > t) = \mathbb{P}(T > s),$$

то она имеет экспоненциальное распределение.

Рассмотрим теперь T_1 – время первого прыжка считающего процесса N_t , тогда

$$\mathbb{P}(T_1 > t + s \mid T_1 > t) = \mathbb{P}(N_{t+s} = 0 \mid N_t = 0).$$

Воспользуемся свойством независимости и стационарности приращений процесса Леви:

$$\mathbb{P}(N_{t+s} = 0 \mid N_t = 0) = \mathbb{P}(N_s = 0) = \mathbb{P}(T_1 > s).$$

Получается, T_1 имеет экспоненциальное распределение с некоторой константной интенсивностью λ . Аналогично мы можем проверить, что $T_k - T_{k-1}$ имеют такое же распределение.

Чтобы завершить, остаётся убедиться в том, что интервалы $T_k - T_{k-1}$ независимы, это можно получить, перейдя к считающему процессу и воспользовавшись свойством независимых приращений процесса Леви:

$$\begin{aligned} \mathbb{P}(T_k - T_{k-1} > \Delta_{k-1}, \dots, T_2 - T_1 > \Delta_1, T_1 > t) &= \\ &= \mathbb{P}(N_{T_{k-1} + \Delta_{k-1}} - N_{T_{k-1}} = 0, \dots, N_{T_1 + \Delta_1} - N_{T_1} = 0, N_t = 0) = \\ &= \mathbb{P}(T_k - T_{k-1} > \Delta_{k-1}) \dots \mathbb{P}(T_2 - T_1 > \Delta_1) \mathbb{P}(T_1 > t). \end{aligned}$$

□

8.3 От классики к процессам Леви: Винеровский процесс

Мы помним, что у Винеровского процесса (если мы возьмём правильную модификацию) траектории непрерывны и изначально из определения процесса Леви можно задаться вопросом: а какие ещё процессы Леви обладают непрерывными траекториями? Ответ на вопрос неочевидный и совершенно неожиданный.

Теорема 8.4. *Если процесс Леви X_t обладает непрерывными траекториями, то он имеет вид*

$$X_t = \gamma t + \sigma W_t,$$

где $\gamma, \sigma \in \mathbb{R}$ и W_t – это Винеровский процесс с $W_0 = 0$.

Таким образом, такой Винеровский процесс со сносом – единственный, обладающий таким свойством. Это свойство мы тоже обсудим в следующий раз и получим его как несложное следствие. Впрочем, его можно доказать, используя предельные теоремы.

8.4 Что-то третье: сложенный процесс Пуассона

Что мы получили? С одной стороны, есть Пуассоновский процесс: у него неотрицательные и неубывающие кусочно-константные траектории, которые имеют разрывы. С другой стороны, есть Винеровский процесс со сносом, который обладает непрерывными траекториями. Это даёт в некотором смысле две крайности, и ни одна из них не приближает

нас к исходной цели: получить процесс с càdlàg-траекториями, с более гибкой моделью разрывов.

Взглянем ещё раз на считающий процесс:

$$N_t = \sum_{i=1}^{\infty} \mathbb{1}(t \geq T_i) = \sum_{i=1}^{\infty} \xi_i, \quad -$$

причём ξ_i независимы. Это наблюдение приводит нас к ещё одной крайности.

Определение 8.3. Пусть μ – вероятностное распределение, N – Пуассоновский процесс с интенсивностью λ . Случайный процесс

$$X_t = \sum_{i=1}^{N_t} Y_i, \quad Y_i \sim^{iid} \mu,$$

называется сложенным Пуассоновским процессом с распределением прыжков μ и интенсивностью прыжков λ .

Это очень гибкая модель: мы теперь контролируем частоту прыжков и их размер.

Утверждение 8.5. Сложенный Пуассоновский процесс X_t с распределением прыжков μ и интенсивностью λ является процессом Леви.

▷ Проверим определение.

Во-первых, $N_0 = 0$ даёт $X_0 = 0$.

Во-вторых, траектории кусочно константные, почти как у Пуассоновского процесса, следовательно càdlàg.

В-третьих, рассмотрим приращения для $t_1 < t_2 < \dots$

$$X_{t_{i+1}} - X_{t_i} = \sum_{j=N_{t_i}+1}^{N_{t_{i+1}}} Y_j.$$

Они не имеют общих слагаемых и все Y_j независимые, значит, приращения тоже независимы и $X_t - X_s \sim X_{t-s}$. □

При этом сложенный процесс Пуассона не является считающим и не имеет непрерывных траекторий, поэтому даёт нам третью крайность в процессах Леви. Как мы увидим чуть позже из формулы Леви-Хинчина, само определение процесса Леви является чётким ограничением класса возможных моделей: все остальные процессы Леви – что-то среднее между этими тремя идеями. При этом гибкость и общность модели прыжков демонстрирует, что мы можем строить достаточно сложные и необычные процессы.

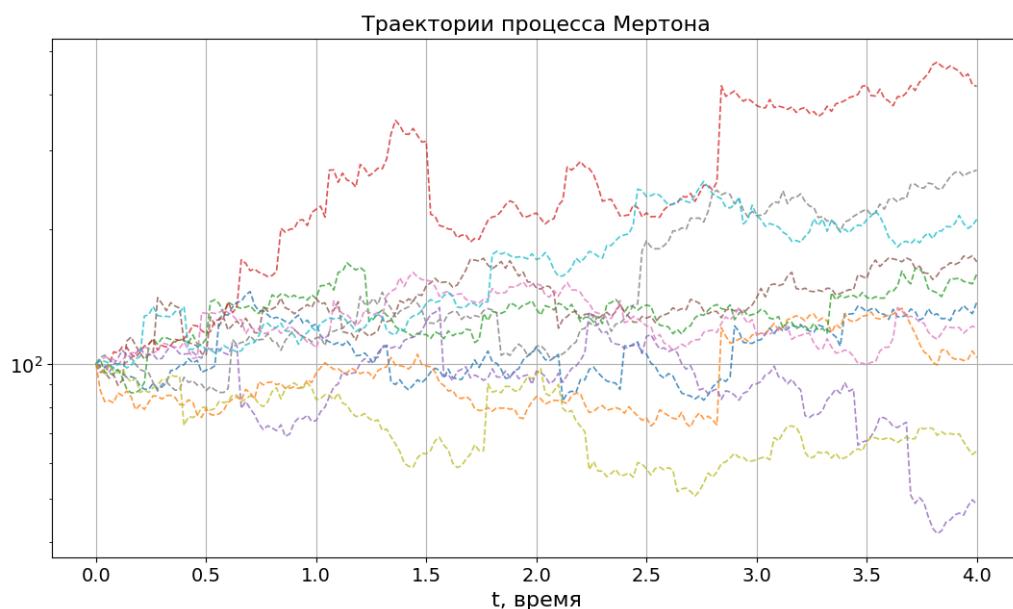
Пример 8.2. (Модель Мертона[15], 1976) Почему вообще цена акций компании может расти? Прежде всего, из-за общей инфляции и реинвестирования прибыли, это нам демонстрирует модель $GBM(\mu, \sigma^2)$, где μ – это безрисковая процентная ставка реинвестирования минус процентные дивиденды (про их размер и время выплат всё известно),

которые надо выплачивать держателям акций. Ещё рост акций может происходить за счёт привлечения внешних инвестиций в виде кредитов и масштабирования компании. Но вместе с платёжными обязательствами приходит и обязательство выплачивать долги и даже риск дефолта. В тяжёлые моменты цена акций может сильно проседать или за счёт дополнительных инвестиций наоборот сильно расти. Мертон в своей статье предложил такую модель цены X_t :

$$X_t = x_0 e^{\gamma t + \sigma W_t + \sum_{i=1}^{N_t} Y_i}, \quad Y_i \sim \text{iid } N(\mu, \sigma^2).$$

Это почти GBM , но в показателе экспоненты сложенный процесс Пуассона с интенсивностью λ и распределением прыжков $N(\mu, \sigma^2)$. Прыжки позволяют моделировать моменты выплат по обязательствам и дополнительных инвестиций, которые не являются открытой информацией, но оказывают резкое влияние на цену.

Гауссовское распределение – скорее технический ход, потому что в этом случае можно вывести формулу цены Европейского опциона, которая будет ощутимо более сложной, чем в случае модели Блэка-Шоулза.



Понятно, что оценка моделей типа модели Мертона по данным – это уже гораздо более сложная задача, которая не поддаётся простому методу максимального правдоподобия прежде всего потому, что мы не знаем, сколько прыжков и были ли они между двумя соседними наблюдениями, а также не знаем их размера. Это задача для EM-алгоритма.

Процессы Леви: 2

Как далеко могут завести ограничения процесса Леви? Насколько сложные процессы можно получать? Ответы на эти вопросы мы будем обсуждать в этой лекции.

9.1 Что общего у трёх процессов?

Что объединяет три рассмотренных процесса: Винеровский, Пуассоновский и сложенный Пуассоновский? Все они процессы Леви, то есть, обладают свойством независимых стационарных приращений.

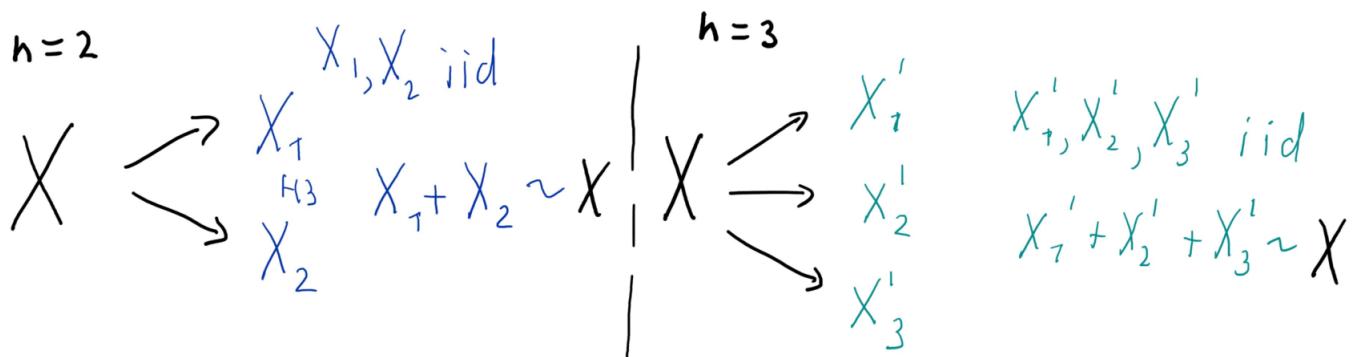
Это очень сильный факт: получается, что если X – процесс Леви, то любая случайная величина X_t представима в виде суммы n независимых и одинаково распределённых случайных величин:

$$X_t = \sum_{i=1}^n (X_{ti/n} - X_{t(i-1)/n}).$$

Каждое из приращений в силу стационарности имеет одно и тоже распределение, совпадающее с распределением $X_{t/n}$. При этом число $n \in \mathbb{Z}_{>0}$ можно брать произвольным.

Определение 9.1. Случайная величина X такая, что для всех n существуют n независимых случайных величин X_i с одинаковым распределением и распределение случайных величин X и $X_1 + \dots + X_n$ равны, называется бесконечно делимой.

Многие являются таковыми, но многие нет.



Упражнение 9.1. Проверьте, является ли бесконечно делимым распределение

1. Пуассона $\text{Poisson}(\lambda)$;
2. Равномерное $\text{Uniform}[a, b]$;

3. Бернуллиевское $Ber(p)$ и биномиальное $Binomial(n, p)$;
4. Гауссовское $N(\mu, \sigma^2)$,

На прошлой неделе мы вспоминали про характеристическую функцию, в случае процесса Леви мы получили, что

$$\mathbb{E}[e^{i\xi X_t}] = \mathbb{E}[e^{i\xi X_{t/n}}]^n,$$

причём для всех t и n , что наводит на мысль, что эти характеристические функции облашают своей структурой.

Пример 9.1. Рассмотрим Пуассоновский процесс X интенсивностью λ , его характеристическую функцию можно вычислить как характеристическую функцию Пуассоновского распределения:

$$\phi_{X_t}(\xi) = \mathbb{E}[e^{i\xi X_t}] = e^{\lambda t(e^{i\xi}-1)}.$$

Пример 9.2. С другой стороны, для Винеровского процесса W характеристическая функция

$$\phi_{W_t}(\xi) = \mathbb{E}[e^{i\xi W_t}] = e^{-\frac{t\xi^2}{2}}.$$

Пример 9.3. Для сложенного Пуассоновского процесса X с интенсивностью λ и распределением размера прыжка $Y_i \sim F$ задача будет немного посложнее, но тут достаточно зафиксировать N_t и использовать формулу полного матожидания, чтобы получить

$$\phi_{X_t}(\xi) = \mathbb{E}[e^{i\xi X_t}] = e^{\lambda t(\mathbb{E}[e^{i\xi Y}] - 1)}$$

или в более популярном виде

$$\phi_{X_t}(\xi) = e^{\lambda t \int_{\mathbb{R}} (e^{i\xi y} - 1) dF(y)}.$$

Если мы будем рассматривать новые и новые процессы Леви, характеристическая функция будет сильно напоминать что-то похожее на эти примеры.

9.2 Экспонента процесса Леви

Оказывается, это часть достаточно общего факта про бесконечно делимые распределения. Для таких распределений есть теорема, фиксирующая вид характеристических функций.

Теорема 9.1. (Формула Леви-Хинчина) Если X – бесконечно делимая случайная величина, то характеристическая функция имеет вид

$$\mathbb{E}[e^{i\xi X}] = \exp\left(a\xi i - \frac{\sigma^2 \xi^2}{2} + \int_{\mathbb{R} \setminus 0} (e^{i\xi y} - 1 - i\xi y \mathbf{1}(|y| < 1)) d\nu(y)\right),$$

где $a \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}$, ν – сигма-конечная (не обязательно конечная!) мера, удовлетворяющая свойству

$$\int_{\mathbb{R} \setminus 0} (1 \wedge y^2) d\nu(y) < \infty.$$

Мы не будем доказывать эту формулу, потому что она потребует очень много времени и приёмов. Но давайте посмотрим, чем это помогает в исследовании характеристических функций процесса Леви.

Теорема 9.2. (*Формула Леви-Хинчина, версия 2*) Пусть X – это процесс Леви, тогда

$$\phi_{X_t}(\xi) = e^{t\Psi(\xi)},$$

где

$$\Psi(\xi) = a\xi i - \frac{\sigma^2 \xi^2}{2} + \int_{\mathbb{R} \setminus 0} (e^{i\xi y} - 1 - i\xi y \mathbf{1}(|y| < 1)) d\nu(y)$$

и a, σ, ν – тройка из формулы Леви-Хинчина для X_1 .

Рассмотрим характеристическую функцию для X_{t+s} , где для краткости зафиксируем ξ и обозначим

$$\phi_{X_{t+s}}(\xi) = \mathbb{E}[e^{i\xi X_{t+s}}] = \zeta_\xi(t+s).$$

По независимости стационарности приращений

$$\zeta_\xi(t+s) = \mathbb{E}[e^{i\xi X_{t+s}}] = \mathbb{E}[e^{i\xi(X_{t+s}-X_t)}] \mathbb{E}[e^{i\xi X_t}] = \zeta_\xi(s)\zeta_\xi(t).$$

Такое свойство мы уже видели ранее, попробуем вычислить производную и составить дифференциальное уравнение:

$$\zeta'_\xi(t) = \lim_{s \rightarrow 0+} \frac{\zeta_\xi(t+s) - \zeta_\xi(t)}{s} = \zeta_\xi(t)\zeta'_\xi(0).$$

Ещё мы знаем, что $\zeta_\xi(0) = 1$. Есть лишь одна функция, удовлетворяющая дифференциальному уравнению с таким условием:

$$\zeta_\xi(t) = e^{\zeta'_\xi(0)t}.$$

Можно считать $\zeta'_\xi(0) = \Psi(\xi)$, некоторой функцией от ξ , таким образом мы получили, что

$$\phi_{X_t}(\xi) = e^{t\Psi(\xi)}.$$

Теперь если мы возьмём $t = 1$ и вспомним о том, что X_1 – бесконечно делимая величина и для неё верна формула Леви-Хинчина, мы получим искомый результат. \square

Функцию $\Psi(\xi)$ называют *характеристической экспонентой* процесса Леви, а тройку (a, σ^2, ν) из формулы Леви-Хинчина – *тройкой Леви* (*Lévy triplet*). Что мы можем узнать из подобного представления? Посмотрим подробнее:

$$\Psi(\xi) = \exp \left(a\xi i - \frac{\sigma^2 \xi^2}{2} + \int_{\mathbb{R}} (e^{i\xi y} - 1 - i\xi y \mathbf{1}(|y| < 1)) d\nu(y) \right).$$

Первые два слагаемых – это части гауссовой характеристической экспоненты. Второе можно переписать в другом виде и тоже увидеть

$$\begin{aligned} \int_{\mathbb{R} \setminus 0} (e^{i\xi y} - 1 - i\xi y \mathbf{1}(|y| < 1)) d\nu(y) = \\ = \nu(\mathbb{R} \setminus (-1, 1)) \int_{\mathbb{R}} (e^{i\xi y} - 1) d\mu(y) + \int_{\mathbb{R}} (e^{i\xi y} - 1 - i\xi y) d\rho(y), \end{aligned}$$

где μ и ρ – это меры

$$\mu = \frac{\nu|_{\mathbb{R} \setminus (-1, 1)}}{\nu(\mathbb{R} \setminus (-1, 1))}, \quad \rho = \nu|_{(-1, 1) \setminus 0}.$$

Первое слагаемое – это часть характеристической экспоненты сложенного процесса Пуассона с интенсивностью $\lambda = \nu(\mathbb{R} \setminus (-1, 1))$ и распределением размера прыжка μ с прыжками больше 1. Что такое второе слагаемое? Это экспонента так называемого *компенсированного сложенного Пуассоновского процесса* с прыжками не более 1. Так в этом контексте называют сложенный Пуассоновский процесс с интенсивностью $\lambda = \rho((-1, 1))$, распределением размера прыжка $\rho/\rho((-1, 1))$, и ещё с вычтенным матожиданием

$$X_t = \sum_{i=1}^{N_t} Y_i - \lambda \mathbb{E}[Y_1], \quad Y_1, \dots \sim_{iid} \rho/\rho((-1, 1)).$$

Очень важно условие интегрируемости для меры ν

$$\int_{\mathbb{R} \setminus 0} (1 \wedge y^2) d\nu(y) < \infty.$$

В силу этого условия первый Пуассоновский процесс имеет прыжки больше 1, но при этом константную интенсивность, а второй имеет (в теории может иметь) бесконечную интенсивность, но прыжки размером не более 1.

9.3 Разложение Леви-Ито и отказ от плотности

По сути, мы только что разложили процесс Леви на части. Если это провести формально, то получается, что и обратно мы можем по тройке Леви восстановить соответствующий процесс Леви.

Теорема 9.3. (*Формула Леви-Хинчина, версия 3*) *Если заданы $a \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}$ и на вероятностном пространстве сигма-конечная мера ν на $\mathbb{R} \setminus 0$ такая, что*

$$\int_{\mathbb{R} \setminus 0} (1 \wedge y^2) d\nu(y) < \infty,$$

то существует вероятностная мера на Ω и процесс Леви (X_t) с характеристикой экспонентой

$$\Psi(\xi) = a\xi i - \frac{\sigma^2 \xi^2}{2} + \int_{\mathbb{R} \setminus 0} (e^{i\xi y} - 1 - i\xi y \mathbf{1}(|y| < 1)) d\nu(y).$$

Таким образом, процесс Леви можно определить, определяя характеристическую экспоненту, причём чаще всего сложные процессы Леви задаются именно таким образом: задаётся тройка Леви (a, σ^2, ν) , где чуть ли не самую важную роль играет мера Леви ν . Структура характеристической экспоненты позволяет более детально разложить процесс на уже известные нам составляющие.

$$X_t = X_t^{(1)} + X_t^{(2)} + X_t^{(3)}$$

Теорема 9.4. (*Разложение Леви-Ито*) Пусть X_t – это процесс Леви, тогда он раскладывается в сумму независимых процессов

$$X_t = X_t^{(1)} + X_t^{(2)} + X_t^{(3)},$$

где

1. Процесс $X_t^{(1)} = at + \sigma W_t$ – Винеровский процесс со сносом a и дисперсией σ^2 ;
2. Процесс $X_t^{(2)}$ – сложенный Пуассоновский процесс с конечной константной интенсивностью и прыжками размера больше 1;
3. Процесс $X_t^{(3)}$ – сложенный скомпенсированный Пуассоновский процесс с (возможно бесконечной) интенсивностью и прыжками размера не более 1.

Давайте ещё раз посмотрим на известные нам примеры и сделаем несколько наблюдений.

Пример 9.4. (*Винеровский процесс*) Это чисто первая компонента $X_t = X_t^{(1)}$. Две другие имеют разрывы, поэтому Винеровский процесс со сносом – это единственный процесс Леви, имеющий модификацию с непрерывными траекториями.

Пример 9.5. (*Пуассоновский процесс*) Это чисто вторая компонента $X_t = X_t^{(2)}$, при этом ν – это дельта-мера $\lambda \delta_1$ в единице (прыжки всегда размера 1, интенсивность λ). И тем самым это единственный считающий процесс Леви, потому что считающий процесс прыгает на 1 и точно с константной интенсивностью, потому что отсечек в любом интервале времени конечное число.

Пример 9.6. (*Сложенный Пуассоновский процесс*) Это константа + вторая и третья компоненты $X_t = a + X_t^{(2)} + X_t^{(3)}$, при этом ν строится так, чтобы меры

$$\mu = \frac{\nu|_{\mathbb{R} \setminus (-1,1)}}{\nu(\mathbb{R} \setminus (-1,1))}, \quad \rho = \nu|_{(-1,1) \setminus 0}$$

удовлетворяли важному условию: $\nu((-1,1)) < \infty$ и $\nu(\mathbb{R}(-1,1)) < \infty$, – в итоге это даёт правильный Пуассоновский процесс с константной интенсивностью в показателе суммы. Распределение размера прыжков тоже зашито в ν .

9.4 Больше примеров процессов Леви

Информация о траекториях может быть извлечена из анализа меры Леви. Например, помимо проверки непрерывности, характеристическая экспонента $\Psi(\xi)$ ограничена тогда и только тогда, когда X_t – это сложенный Пуассоновский процесс, то есть, гауссовской компоненты нет и интенсивность константная. Это отдельная большая тема для обсуждений, которую не покрыть и в нескольких лекциях.

Что даёт нам такая наука? Она позволяет строить более сложные процессы Леви, чем те, которые мы рассматривали.

Пример 9.7. (*Variance-Gamma процесс, VG, Madan, Seneta 1990*) Попробуем ещё раз поменять приращения, рассмотрим гамма-процесс Леви:

$$X_t - X_s \sim \Gamma(\alpha, \beta(t-s)).$$

Характеристическая экспонента такого процесса

$$\Psi_X(\xi) = -\beta \ln(1 - i\xi/\alpha).$$

А мера Леви (после некоторых непростых вычислений) имеет плотность

$$f_\nu(y) = \beta \frac{e^{-\alpha y}}{y} \mathbf{1}(y > 0)$$

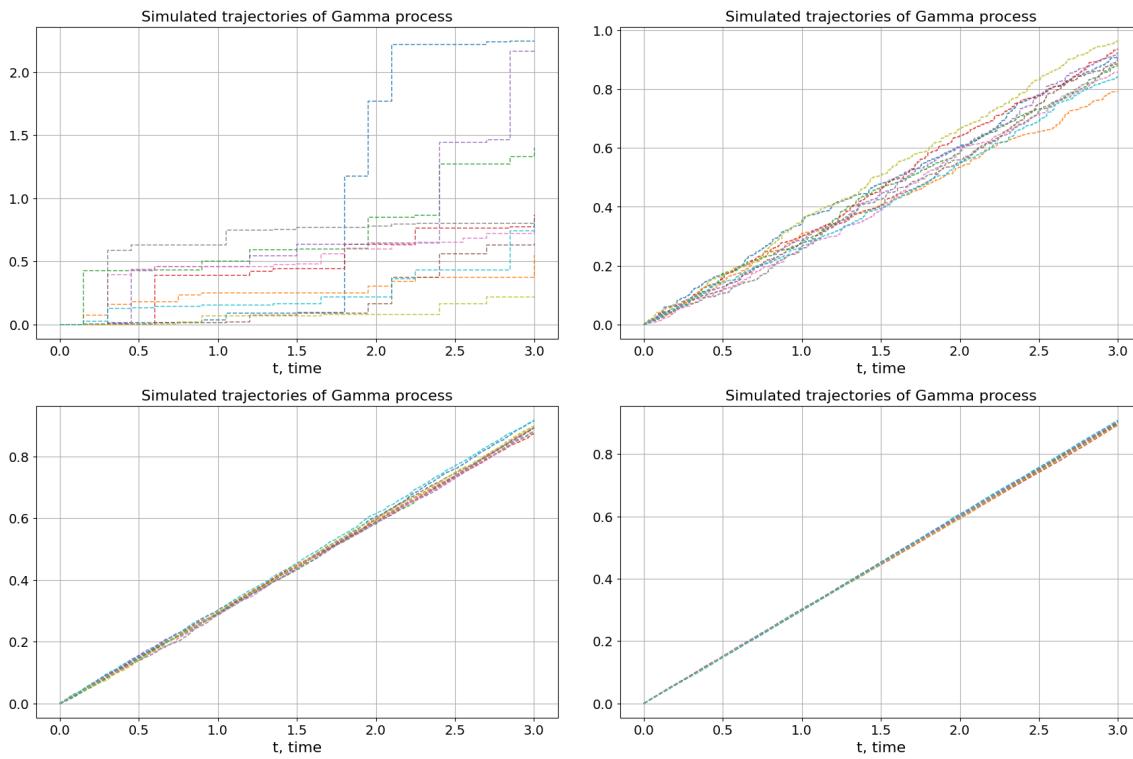
и ещё

$$\sigma = 0, \quad a = - \int_0^1 y d\nu(y).$$

Гамма-процесс, таким образом, – простейший процесс Леви с бесконечной интенсивностью прыжков, то есть, в каждом интервале их счётное количество. Это происходит по двум причинам:

1. Мера $\nu((-1,1)) = \infty$, то есть, скомпенсированный процесс имеет бесконечную интенсивность и бесконечное число прыжков.

2. Прыжков всегда не более, чем счётное число. Это следует из известного факта математического анализа: у любой функции прыжков (разрывов первого рода) всегда не более, чем счётное число. Именно такие разрывы имеют траектории процессов Леви.



Смотря на симуляции очень сложно понять свойства траекторий: при разной дискретизации по времени (на картинках 20, 200, 20000, 200000 точек) мы получаем разную картину и при очень плотной дискретизации траектории практические прямые. Но на самом деле, процесс двигается маленькими прыжками, которых бесконечно много.

На основе гамма-процесса можно построить одну популярную финансовую модель: *Variance-Gamma(VG)* процесс

$$Y_t = \mu X_t + \sigma W_{X_t}.$$

Этот процесс можно использовать, как и модель Мертона, в экспоненте изначальной модели *GBM*.

Это процесс с характеристической экспонентой

$$\Psi(\xi) = -\beta \ln \left(1 - i\xi c/\alpha + \beta^2 \xi^2 / (2\alpha) \right), \quad \alpha, c \in \mathbb{R}, \beta > 0.$$

У этого процесса три параметра, α отвечает за масштаб (связано с дисперсией), c – параметр асимметрии распределения, а β – это эксцесс, тяжесть хвостов распределения.

Симуляция такого процесса уже не такая простая, потому что приращения имеют очень сложное *VG*-распределение, у которого всё ещё можно записать плотность, но сэмплировать из неё?.. Здесь известно два пути:

1. Авторский путь, сгенерировать независимые гамма приращения

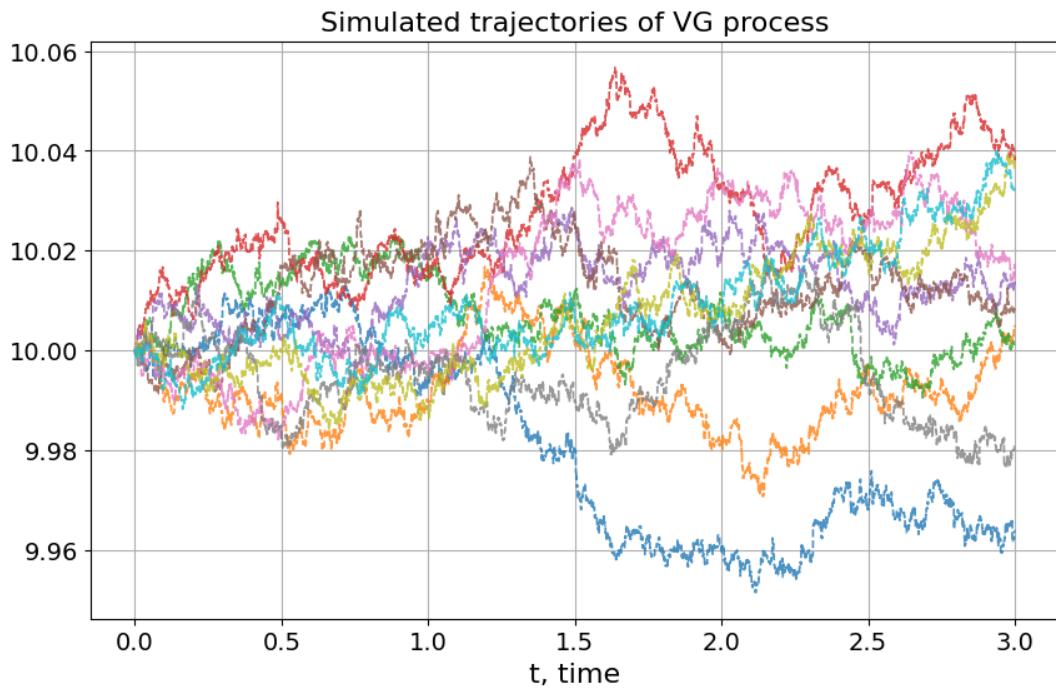
$$\Delta G_{i-1} \sim \Gamma(\alpha, \beta(t_i - t_{i-1}))$$

и далее, задав $Y_{t_0} = 0$,

$$Y_{t_i} = Y_{t_{i-1}} + \mu \Delta G_{i-1} + \sigma \sqrt{\Delta G_{i-1}} Z_i, \quad Z_i \sim_{iid} N(0, 1).$$

2. Ещё другой путь: представить VG -процесс как разность гамма-процессов с пересчитанными параметрами, сэмплировать два гамма-процесса и брать их разность.

По данным такой процесс можно пытаться оценивать EM -алгоритмом, беря за основу пункт 1.



Пример 9.8. (CGMY) Это процесс, который имеет 4 параметра C, G, M, Y , которые совпадают с первыми буквами имён авторов. Они имеют меру Леви с плотностью

$$f_\nu(x) = \begin{cases} C \frac{e^{-G|x|}}{|x|^{1+Y}}, & x < 0, \\ C \frac{e^{-M|x|}}{|x|^{1+Y}}, & x > 0 \end{cases}$$

и характеристическую экспоненту

$$\Psi(\xi) = C\Gamma(Y)(M - i\xi)^Y - M^Y + (G + i\xi)^Y - G^Y.$$

Мера Леви устроена так, что интенсивность прыжков бесконечна, а размер прыжка имеет разную тяжесть хвостов слева и справа. Более того, это процесс, который движется чисто за счёт прыжков, как VG и Пуассоновские процессы. Однако за счёт бесконечной интенсивности обманчиво может показаться, что траектория практически непрерывна, как у Винеровского процесса.

9.5 Бонус: ещё один смысл меры Леви

Почему мера Леви так универсально описывает траектории целого сложного случайного процесса с разрывными траекториями? Оказывается, она имеет чёткую интерпретацию, связанную с измерениями числа прыжков. Поскольку прыжки происходят случайно, здесь мы впервые встречаем *случайные меры*.

Определение 9.2. Пусть (Ω, \mathcal{F}, P) – вероятностное пространство, (E, \mathcal{E}) – измеримое пространство с мерой μ . Тогда $M : \Omega \times \mathcal{E} \rightarrow \mathbb{R}$ называется случайной мерой, если

1. Для всех $\omega \in \Omega$ $M(\omega, \cdot)$ – это мера на \mathcal{E} ;
2. Для всех $A \in \mathcal{E}$ $M(\cdot, A)$ – это измеримая функция.

Иными словами, мера выпадает как результат случайного исхода $\omega \in \Omega$; с другой стороны, мера любого множества A в таком случае – случайная величина, потому что мера множества будет зависеть от того, какая мера попадётся. В частности, нас сейчас будет интересовать конкретный типа таких случайных мер.

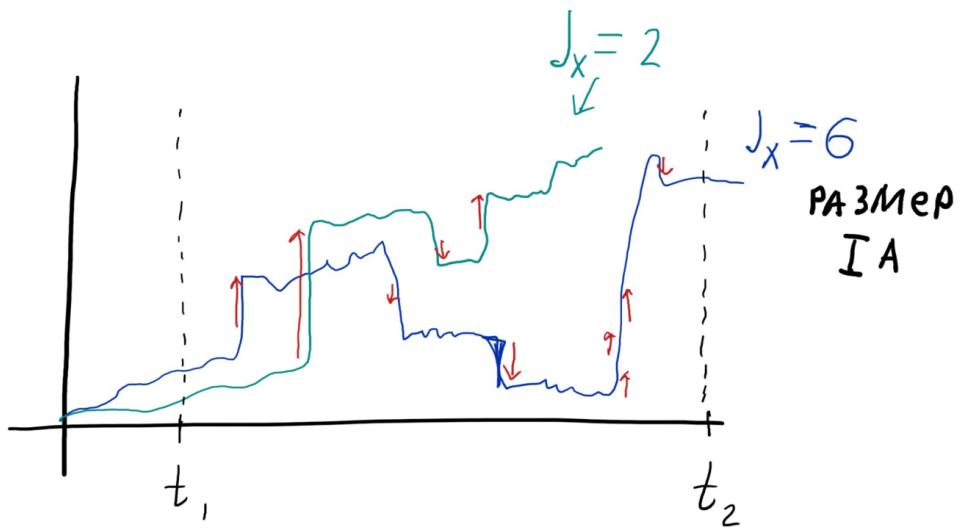
Определение 9.3. Пусть (Ω, \mathcal{F}, P) – вероятностное пространство, (E, \mathcal{E}) – измеримое пространство с мерой μ . Тогда $M : \Omega \times \mathcal{E} \rightarrow \mathbb{R}$ называется случайной Пуассоновской мерой с интенсивностью μ , если

1. Для всех $A \in \mathcal{E}$ при $\mu(A) < \infty$ случайная величина $M(\cdot, A)$ имеет распределение Пуассона с интенсивностью $\mu(A)$;
2. Для любого конечного набора попарно непересекающихся $A_i \in \mathcal{E}$ случайные величины $M(\cdot, A_i)$ независимы в совокупности.

Поскольку мы думаем о процессах Леви, нас интересует Пуассоновская случайная мера, потому что прыжков счётное число и они порождены сложенными Пуассоновскими процессами. Пусть X – это процесс с càdlàg-траекториями, тогда за любой интервал (t_1, t_2) можно измерить число и размер прыжков $\Delta X_t \in A \in \mathcal{B}(\mathbb{R})$ и это количество будет случайно (в теории может быть и бесконечно):

$$J_X(t_1, t_2, A) = \# \{t : (t, \Delta X_t) \in (t_1, t_2) \times A\}.$$

J_X – это случайная мера.



Пример 9.9. Рассмотрим Пуассоновский процесс с интенсивностью λ . Тогда J_X – это случайная Пуассоновская мера с интенсивностью λdt .

В случае Пуассоновского процесса эта мера будет считать количество прыжков за фиксированный интервал времени. Мера Леви агрегирует эту информацию в одну обычную меру, описывающую интенсивность прыжков разного размера.

Определение 9.4. Мерой Леви процесса Леви X_t называется

$$\nu(A) = \mathbb{E} [\# \{t \in [0, 1] : \Delta X_t \neq 0, \Delta X_t \in A\}].$$

Так, меру Леви можно рассматривать, как ожидаемое число прыжков размера из множества A за единицу времени. Если мера Леви $\nu(A)$ бесконечна, значит прыжков такого размера бесконечное количество. Как мы понимаем, сама характеристическая экспонента $\Psi(\xi)$ от времени не зависит, а в силу стационарности приращений в разных отрезках времени среднее число прыжков прямо пропорционально размеру временного отрезка.

Стохастический интеграл Ито

В этой лекции мы определяем известный стохастический интеграл: Интеграл Ито. Как мы увидим, к интегралу Ито естественно приходят в попытке записать более общие дифференциальные уравнения, включающие в себя случайные процессы.

10.1 Дифференциальные уравнения с "шумом"

Рассмотрим классическую задачу Коши из курса обыкновенных дифференциальных уравнений:

$$\frac{dX}{dt} = b(t, X(t)), \quad X : \mathbb{R}_+ \rightarrow \mathbb{R}, \quad X(0) = x_0.$$

Её решение – это некоторая траектория $X(t)$, такая что она начинается в x_0 , так как $X(0) = x_0$. Не секрет, что такие уравнения часто встречаются в физике, а также во многих других областях. Тем не менее, подобные модели обладают естественным ограничением: часто правая часть *идеальна*. Но что делать, если, например, в динамике полёта самолёта нужно учсть ветер? Очевидно, что подобные эффекты не вполне понятно, как описать в силу их хаотичности. Второе важное ограничение – это теорема о существовании решения, которая требует непрерывности правой части, а для единственности требует ещё липшицевости (по X). Таким образом, не всякая функция хорошо подойдёт для модели на основе обыкновенного дифференциального уравнения.

Давайте попробуем посмотреть на эту задачу с несколько эвристической точки зрения, что если мы просто добавим случайный *шум*?

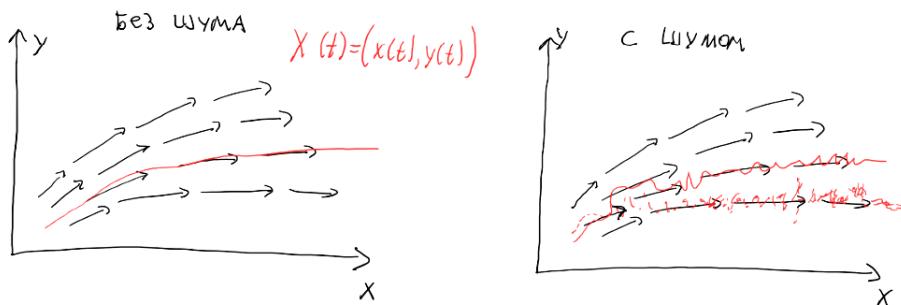


Рис. 10.1: Случайный шум меняет векторное поле и, соответственно, траектории.

$$\frac{dX}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot \xi_t,$$

где $(\xi_t)_{\mathbb{R}_+}$ – это некоторый случайный процесс на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$ (соответственно $(X_t)_{t \in \mathbb{R}_+}$ тоже будет случайным процессом). Мы использовали также функцию $\sigma : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$ для добавления шуму некоторой вариативности. Какие естественные практические предположения мы могли бы сделать насчёт шума? К примеру, такие:

1. ξ_{t_1} и ξ_{t_2} независимы для любых $t_1 \neq t_2$;
2. Процесс ξ стационарен в узком смысле;
3. $\mathbb{E}[\xi_t] = 0$.

Процесс белого гауссовского шума ξ_t подошёл бы, но, к сожалению, он не обладает непрерывными траекториями и, следовательно, у нас проблема с существованием решения. Оказывается, вообще не бывает нетривиального процесса ξ с непрерывными траекториями и свойствами 1-3.

Утверждение 10.1. *Если выполнены свойства (1)-(3), то вероятность того, что траектория будет непрерывна в счётном числе точек равна 0, за исключением случая, когда $\xi_t = 0$ почти наверное.*

▷ Эта задача немного похожа на задачу про разрывность белого шума из первых лекций. Сейчас однако ничего не известно про распределение ξ_t , поэтому мы используем такой подход: если процесс ξ имеет непрерывные траектории, то процесс

$$\xi_t^{(N)} = \begin{cases} N, & \xi_t > N, \\ \xi_t, & \xi_t \in [-N, N], \\ -N, & \xi_t < -N \end{cases}$$

тоже будет иметь непрерывные траектории. Заметим, что этот процесс обладает дисперсией и кроме того

$$\text{Var} [\xi_t^{(N)} - \xi_s^{(N)}] = 2\text{Var} [\xi_t^{(N)}] > 0,$$

если только не $\xi_t = 0$ почти наверное. Покажем, что траектория с ненулевой вероятностью имеет разрыв в точке t_0 , тогда и изначальный процесс тоже будет его иметь. Рассмотрим подпоследовательность $\{t_k\} \rightarrow t_0$, последовательность $\xi_{t_k}^{(N)}$ сходится почти наверное к $\xi_{t_0}^{(N)}$ тогда и только тогда, когда для всех $\varepsilon > 0$

$$\mathbb{P} \left(\limsup_{k \geq n} |\xi_{t_k}^{(N)} - \xi_{t_0}^{(N)}| > \varepsilon \right) = 0.$$

Предел берётся поточечно и всегда есть. Этот критерий удобен тем, что мы можем проанализировать предельную величину под вероятностью с помощью леммы Фату:

$$\mathbb{E} \left[\left(\limsup_{k \geq n} |\xi_{t_k}^{(N)} - \xi_{t_0}^{(N)}| \right)^2 \right] = \mathbb{E} \left[\limsup_{k \geq n} |\xi_{t_k}^{(N)} - \xi_{t_0}^{(N)}|^2 \right] \geq \limsup_{k \geq n} \mathbb{E} [|\xi_{t_k}^{(N)} - \xi_{t_0}^{(N)}|^2] = 2\text{Var} [\xi_{t_0}^{(N)}] > 0.$$

Следовательно, вероятность не может быть нулевой ни для какой подпоследовательности и

$$\mathbb{P} \left(\omega \in \Omega : \lim_{t \rightarrow t_0} \xi_t^{(N)}(\omega) = \xi_{t_0}^{(N)}(\omega) \right) < 1.$$

В силу независимости, вероятность того, что хотя бы в счётом числе точек функция будет непрерывна одновременно равна нулю. \square

Давайте теперь посмотрим на задачу со стороны дискретной модели – попробуем обойти эти тонкие вопросы. Зададим конечный набор времён $t_0 = 0, t_1, \dots, t_{N-1} = T \in \mathbb{R}_+$ и запишем явный метод Эйлера (аппроксимируем производную конечной разностью); добавив в него шумовой член, мы получим вполне понятную модель в дискретном времени:

$$X_{t_{k+1}} = X_{t_k} + b(t_k, X_{t_k})\Delta t_k + \sigma(t_k, X_{t_k})Y_k \Delta t_k, \quad (10.1)$$

где $\Delta t_k := t_{k+1} - t_k$. Рассмотрим поближе $Y_k \Delta t_k$: давайте введём процесс W_{t_k} как

$$W_{t_{k+1}} = W_{t_k} + Y_k \Delta t_k, \quad W_{t_0} = 0.$$

Мы выше предполагали, что шумы $Y_k \Delta t_k$ должны быть независимыми и стационарными в узком смысле. Для процесса W требования для шума означают, что

1. W имеет независимые приращения;
2. Приращения W стационарны в узком смысле;
3. Приращения имеют нулевое матожидание.

Один из процессов в непрерывном времени $(W_t)_{t \in \mathbb{R}_+}$, который удовлетворяет этим свойствам и имеет непрерывные траектории, – это хорошо известный нам Винеровский процесс. Перепишем (10.1) с этой новой информацией и избавимся от процесса ξ :

$$X_{t_{k+1}} = X_{t_k} + b(t_k, X_{t_k})\Delta t_k + \sigma(t_k, X_{t_k})\Delta W_k, \quad \Delta W_k = W_{t_{k+1}} - W_{t_k}. \quad (10.2)$$

Эти уравнения мы можем переписать в виде сумм:

$$X_{t_{k+1}} = X_{t_0} + \sum_{j=0}^k b(t_j, X_{t_j})\Delta t_j + \sum_{j=0}^k \sigma(t_j, X_{t_j})\Delta W_j. \quad (10.3)$$

Вопрос теперь состоит в следующем: если уменьшать шаг дискретизации Δt , то кажется, что мы получим интегральное уравнение

$$X_t = X_{t_0} + \int_0^T b(t, X_t)dt + " \int_0^T \sigma(t, X_t)dW_t ". \quad (10.4)$$

С первым интегралом понятно: он возникал и в обыкновенных дифференциальных уравнениях (интегральное уравнение без второго интеграла называется *уравнением Вольтерра*). Что такое второй интеграл? Поскольку мы говорим о случайных величинах, то в каком смысле можно показать его существование? На этот вопрос есть разные ответы, но мы рассмотрим один очень популярный.

10.2 Первая попытка: простые функции

Итак, фиксируем вероятностное пространство (Ω, \mathcal{F}, P) и непрерывную модификацию Винеровского процесса W , начинающегося в нуле. Мы хотим определить интеграл типа

$$\int_S^T f(t, \omega) dW_t(\omega)$$

для каких-то подходящих (определим чуть позже) функций $f : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$. По некоторой аналогии с интегралами Лебега и Римана, мы можем сначала определить интеграл в кавычках для *простых функций*, которые зададим следующим образом. Фиксируем некоторое разбиение отрезка $[S, T]$ состоящее из точек $t_0 = S, t_1, \dots, t_{N-1} = T$. Простой функцией (см. Рис. 10.2) назовём функцию $\phi : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$ вида

$$\phi(t, \omega) = \sum_{j=0}^{N-1} e_j(\omega) \mathbf{1}_{[t_j, t_{j+1})}(t).$$

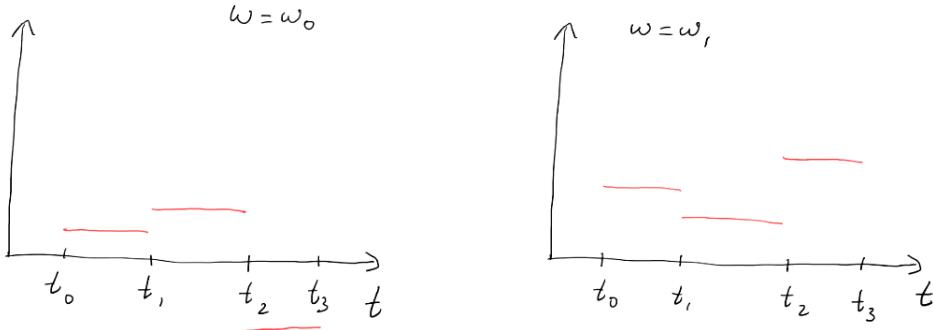


Рис. 10.2: Простые функции при фиксированном $\omega \in \Omega$ кусочно константные по времени

Для простых функций интеграл можно определить, например, так:

$$\int_S^T \phi(t, \omega) dW_t(\omega) := \sum_{j=0}^{N-1} e_j(\omega) (W_{t_{j+1}}(\omega) - W_{t_j}(\omega)). \quad (10.5)$$

Пример 10.1. Определим равномерную сетку $t_j = jT/n$, $j = 0, \dots, n$ на отрезке $[0, T]$ и рассмотрим две простых функции с $e_j(\omega) = W_{t_j}(\omega)$ и $e_j(\omega) = W_{t_{j+1}}(\omega)$:

$$\phi^{(1)}(t, \omega) := \sum_{j=0}^{n-1} W_{t_j}(\omega) \mathbf{1}_{[t_j, t_{j+1})}(t), \quad (10.6)$$

$$\phi^{(2)}(t, \omega) := \sum_{j=0}^{n-1} W_{t_{j+1}}(\omega) \mathbf{1}_{[t_j, t_{j+1})}(t). \quad (10.7)$$

Два интеграла, согласно (10.5), равны

$$F^{(1)}(\omega) = \int_0^T \phi^{(1)}(t, \omega) dW_t(\omega) := \sum_{j=0}^{n-1} W_{t_j}(\omega) (W_{t_{j+1}}(\omega) - W_{t_j}(\omega)), \quad (10.8)$$

$$F^{(2)}(\omega) = \int_0^T \phi^{(2)}(t, \omega) dW_t(\omega) := \sum_{j=0}^{n-1} W_{t_{j+1}}(\omega) (W_{t_{j+1}}(\omega) - W_{t_j}(\omega)). \quad (10.9)$$

Используя свойства независимых приращений Винеровского процесса, можно показать, что $\mathbb{E}[F_1] = 0$, но при этом (независимыми приращениями воспользоваться не получится) в другом случае $\mathbb{E}[F_2] = T$. То есть, два интеграла совершенно разные даже при $n \rightarrow \infty$.

Так, подобный подход работает не для любых функций f .

10.3 Интеграл Ито

Интеграл Ито определяется для специального класса функций, который, тем не менее, не является сильно ограничивающим. Пусть $(\mathcal{F}_t)_{t \in \mathbb{R}}$ – фильтрация, порождённая рассматриваемым Винеровским процессом W .

Определение 10.1. Классом $\mathcal{V}(S, T)$ назовём класс функций $f : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$ таких, что

- Функция f является $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$ -измеримой;
- f является согласованным с фильтрацией $(\mathcal{F}_t)_{t \in \mathbb{R}}$ случайным процессом;
- $\mathbb{E} \left[\int_S^T f(t, \omega)^2 dt \right] \leq \infty$.

Второе требование исключает проблему, которая обозначена в примере выше. Первое и третье требования технические, но оказываются критичными при доказательстве конструкции интеграла Ито. По ходу рассуждений станет ясно, почему именно такие требования нужны.

Начнём с того, что для простых функций $\phi_n \in \mathcal{V}(S, T)$ интеграл Ито определим как в (10.5). Понятно, что в условиях $\mathcal{V}(S, T)$ функции e_j должны быть \mathcal{F}_{t_j} -измеримыми. Прежде всего, в свете пункта 3 отметим один интересный факт.

Утверждение 10.2. (Изометрия Ито) Для любой простой ограниченной функции $\phi_n \in \mathcal{V}(S, T)$ верно тождество

$$\mathbb{E} \left[\left(\int_S^T \phi_n(t, \omega) dW_t(\omega) \right)^2 \right] = \mathbb{E} \left[\int_S^T \phi_n(t, \omega)^2 dt \right].$$

▷ Мы уже определили интеграл для простой функции, раскроем скобки под знаком матожидания:

$$\begin{aligned} \mathbb{E} \left[\left(\int_S^T \phi_n(t, \omega) dW_t(\omega) \right)^2 \right] &= \\ &= \mathbb{E} \left[\sum_{j=0}^{n-1} e_j(\omega)^2 (W_{t_{j+1}} - W_{t_j})^2 + 2 \sum_{i < j} e_i(\omega) e_j(\omega) (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j}) \right]. \end{aligned}$$

Рассмотрим первое слагаемое, для него мы можем использовать формулу полного матожидания с условием \mathcal{F}_{t_j} (это сигма-алгебра из фильтрации Винеровского процесса):

$$\mathbb{E} \left[\sum_{j=0}^{n-1} e_j(\omega)^2 (W_{t_{j+1}} - W_{t_j})^2 \right] = \mathbb{E} \left[\sum_{j=0}^{n-1} e_j(\omega)^2 \mathbb{E} [(W_{t_{j+1}} - W_{t_j})^2 | \mathcal{F}_{t_j}] \right] = \mathbb{E} \left[\sum_{j=0}^{n-1} e_j(\omega)^2 \Delta_j \right].$$

Это интеграл из правой части утверждения. Второе слагаемое с кросс-членами будет равно нулю, так как из-за $i < j$

$$\begin{aligned} \mathbb{E} [e_i(\omega) e_j(\omega) (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j}) | \mathcal{F}_{t_j}] &= \\ &= e_j(\omega) e_i(\omega) (W_{t_{i+1}} - W_{t_i}) \mathbb{E} [(W_{t_{j+1}} - W_{t_j}) | \mathcal{F}_{t_j}] = 0. \end{aligned}$$

□

Шаг 1. Пусть $g \in \mathcal{V}(S, T)$ – непрерывная по t для каждого ω и ограниченная функция.

Лемма 10.3. Для любой такой функции $g \in \mathcal{V}(S, T)$ существует последовательность простых функций ϕ_n , что

$$\mathbb{E} \left[\int_S^T (g(t, \omega) - \phi_n(t, \omega))^2 dt \right] \rightarrow 0, \quad \text{при } n \rightarrow \infty.$$

▷ Проверим, что последовательность из функций

$$\phi_n(t, \omega) = \sum_{j=0}^n g(t_j, \omega) \mathbf{1}_{[t_j, t_{j+1}]}(t)$$

с равномерным разбиением точками t_j отрезка $[S, T]$ на n отрезков подходит. Для этого запишем

$$\int_S^T (g(t, \omega) - \phi_n(t, \omega))^2 dt = \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} (g(t, \omega) - g(t_j, \omega))^2 dt.$$

в свою очередь в силу непрерывности и ограниченности g

$$\sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} (g(t, \omega) - g(t_j, \omega))^2 dt \leq \max_{j=0, \dots, n-1} \max_{t \in [t_j, t_{j+1}]} (g(t, \omega) - g(t_j, \omega))^2 \sum_{j=0}^{n-1} \Delta_j = M_n T,$$

при этом $M_n \rightarrow 0$. □

Шаг 2. Рассмотрим теперь ограниченную (но не обязательно непрерывную) функцию $h \in \mathcal{V}$.

Лемма 10.4. Для любой такой функции $h \in \mathcal{V}(S, T)$ существует последовательность непрерывных ограниченных функций $g_n \in \mathcal{V}(S, T)$ таких, что

$$\mathbb{E} \left[\int_S^T (h(t, \omega) - g_n(t, \omega))^2 dt \right] \rightarrow 0, \quad \text{при } n \rightarrow \infty.$$

▷ Как получить непрерывную функцию из ограниченной? Один из удобных вариантов – сделать свёртку (по t) с подходящей непрерывной функцией:

$$g_n(t, \omega) = h(\cdot, \omega) * \zeta_n(\cdot),$$

где $\zeta_n(t)$ – любая непрерывная неотрицательная функция, для которой

$$\forall t \notin [-1/n, 0] \quad \zeta(t) = 0, \quad \int_{-1/n}^0 \zeta_n(t) dt = 1.$$

В силу того, что $\zeta_n(t)$ равна нулю при $t \geq 0$, после такой свёртки получится всё ещё измеримая функция и согласованная с фильтрацией (хотя это не так очевидно и требует доказательства), а ещё g_n будет непрерывной и ограниченной. Что особенно важно:

$$g_n(t, \omega) - h(t, \omega) \rightarrow 0 \quad n \rightarrow \infty.$$

□

Шаг 3. Пусть $f \in \mathcal{V}(S, T)$.

Лемма 10.5. Для любой функции $f \in \mathcal{V}(S, T)$ существует последовательность ограниченных функций $h_n \in \mathcal{V}(S, T)$ таких, что

$$\mathbb{E} \left[\int_S^T (f(t, \omega) - h_n(t, \omega))^2 dt \right] \rightarrow 0, \quad \text{при } n \rightarrow \infty.$$

▷ Проверьте, что функции

$$h_n(t, \omega) := \begin{cases} -n, & \text{если } f(t, \omega) < -n, \\ f(t, \omega), & \text{если } f(t, \omega) \in [-n, n], \\ n, & \text{если } f(t, \omega) > n, \end{cases}$$

подходят. □

Используя эти леммы мы можем определить интеграл Ито.

Определение 10.2. Интегралом Ито функции $f \in \mathcal{V}$ называется

$$\int_S^T f(t, \omega) dW_t(\omega) := \lim_{n \rightarrow \infty} \int_S^T \phi_n(t, \omega) dW_t(\omega),$$

где предел понимается в смысле $L^2(\Omega)$, а $\phi_n \in \mathcal{V}(S, T)$ – это простые функции, такие, что

$$\mathbb{E} \left[\int_S^T (f(t, \omega) - \phi_n(t, \omega))^2 dt \right] \rightarrow 0 \quad \text{при } n \rightarrow \infty. \quad (10.10)$$

Давайте внимательно всмотримся в предел в смысле $L^2(\Omega)$, по определению это означает, что

$$\mathbb{E} \left[\left(\int_S^T f(t, \omega) dW_t - \int_S^T \phi_n(t, \omega) dW_t(\omega) \right)^2 \right] \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

Предел, таким образом существует (по набору лемм выше) и не зависит от конкретного выбора последовательности ϕ_n (достаточно иметь (10.10)), так как есть факт более общей изометрии Ито.

Упражнение 10.1. (*Изометрия Ито*) Докажите, что для всех $f \in \mathcal{V}(S, T)$ верна изометрия Ито:

$$\mathbb{E} \left[\left(\int_S^T f(t, \omega) dW_t \right)^2 \right] = \mathbb{E} \left[\int_S^T f(t, \omega)^2 dt \right].$$

По сути нужно доказать, что для последовательности простых функций из определения

$$\int_S^T \phi_n^2(t, \omega) \rightarrow^{L^2} \int_S^T f^2(t, \omega).$$

10.4 Первые расчёты

Рассмотрим несколько примеров. Первый иллюстрирует одно из естественных свойств, а второй показывает, что интеграл Ито ведёт себя не совсем как привычный нам интеграл Римана или Лебега.

Упражнение 10.2. Пусть $f(t, \omega) \equiv 1$. Покажите используя определение интеграла Ито, что

$$\int_0^T 1 dW_t = W_T.$$

Пример 10.2. Покажем, что

$$\int_0^T W_t dW_t = \frac{W_T^2}{2} - \frac{T}{2}.$$

Винеровский процесс как функция $W : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$ конечно же лежит в $\mathcal{V}(0, T)$, так как $(W_t)_{t \in \mathbb{R}_+}$ согласован со своей же фильтрацией и из непрерывности траекторий следует пункт про измеримость и

$$\mathbb{E} \left[\int_0^T W_t(\omega)^2 dt \right] = \int_0^T \mathbb{E}[W_t^2] dt = \frac{T^2}{2}.$$

Выберем аппроксимирующую последовательность с помощью метода левых прямоугольников на равномерной сетке из $n + 1$ точки t_j :

$$\int_0^T W_t(\omega) dW_t(\omega) = \lim_{n \rightarrow \infty} \sum_{j=0}^n W_{t_j}(\omega) (W_{t_{j+1}}(\omega) - W_{t_j}(\omega)).$$

Под суммой добавим и вычтем $W_{t_{j+1}}(\omega)^2$, это позволит выделить полный квадрат:

$$\int_0^T W_t(\omega) dW_t(\omega) = \frac{1}{2} \lim_{n \rightarrow \infty} \sum_{j=0}^n \left[(W_{t_{j+1}}(\omega)^2 - W_{t_j}(\omega)^2) - (W_{t_{j+1}}(\omega) - W_{t_j}(\omega))^2 \right].$$

С левой стороны стоит телескопическая сумма, которая точно равна $W_{t_n}(\omega) = W_T(\omega)^2$. Что касается правого слагаемого, то мы можем показать, что в L^2 (а именно в этом смысле понимается предел в определении интеграла Ито)

$$\lim_{n \rightarrow \infty} \sum_{j=0}^n (W_{t_{j+1}}(\omega) - W_{t_j}(\omega))^2 = T.$$

Упражнение 10.3. Докажите, используя свойства Винеровского процесса, последнее утверждение, которое говорит, что

$$\mathbb{E} \left[\left(\sum_{j=0}^n (W_{t_{j+1}}(\omega) - W_{t_j}(\omega))^2 - T \right)^2 \right] \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Естественно, по определению такие интегралы нелегко вычислять, в следующей лекции мы рассмотрим формулу Ито, которая позволяет это делать существенно проще.

Исчисление Ито

В этой лекции мы продолжаем исследовать интеграл Ито. В частности, мы изучим один из главных инструментов для стохастических интегралов: формулу Ито.

11.1 Свойства интеграла Ито

Интеграл Ито, несмотря на его специфику, не только обладает простейшими свойствами известных нам интегралов, но и приобретает несколько новых. Мы уже упоминали изометрию Ито, добавим ещё несколько фактов. Далее для краткости мы будем по возможности опускать зависимость от ω там, где это не вызывает неопределённости.

Теорема 11.1. *Верны следующие свойства для любых $f, g \in \mathcal{V}(S, T)$:*

1. (Аддитивность) для всех $U \in [S, T]$ выполняется $\int_S^T f dW_t = \int_S^U f dW_t + \int_U^T f dW_t$;
2. (Линейность) Для любых $\lambda \in \mathbb{R}$ верно $\int_S^T (f + \lambda g) dW_t = \int_S^T f dW_t + \lambda \int_S^T g dW_t$;
3. $\mathbb{E} \left[\int_S^T f dW_t \right] = 0$.

▷ Не будем останавливаться на первых двух, они возникают из свойств самого предела. Интересно третье; напомним, что по условию, если $f \in \mathcal{V}(S, T)$, то она является случайным процессом, согласованным с фильтрацией $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$, порождённой Винеровским процессом W . Это означает, что $f(t, \cdot)$ является \mathcal{F}_t -измеримой случайной величиной и её можно вынести за знак условного матожидания при условии \mathcal{F}_t . С помощью формулы полного матожидания получим для простой функции

$$\begin{aligned} \mathbb{E} \left[\int_S^T \phi_n dW_t \right] &= \sum_j \mathbb{E} [\phi_n(t_j) (W_{t_{j+1}} - W_{t_j})] = \\ &= \sum_j \mathbb{E} [\phi_n(t_j) \mathbb{E} [(W_{t_{j+1}} - W_{t_j}) \mid \mathcal{F}_{t_j}]] = 0, \end{aligned}$$

так как условное матожидание равно $\mathbb{E} [W_{t_{j+1}-t_j}] = 0$. Это верно для любого n , поэтому будет верно и для предела, то есть,

$$\mathbb{E} \left[\int_S^T f dW_t \right] = 0$$

□

Добавим ещё одно не очень ожидаемое с первого взгляда свойство.

Теорема 11.2. Пусть $f \in \mathcal{V}(0, T)$. Случайный процесс $M_t(\omega) = \int_0^t f(s, \omega) dW_s(\omega)$, проиндексированный $t \in [0, T]$, является маргином относительно фильтрации, порождённой Винеровским процессом W .

▷ Проверим прямо, пусть $t' < t$, тогда по свойству аддитивности интеграл можно разбить на две части:

$$\mathbb{E}[M_t | \mathcal{F}_{t'}] = \mathbb{E} \left[\int_0^{t'} f(s, \omega) dW_s(\omega) + \int_{t'}^t f(s, \omega) dW_s | \mathcal{F}_{t'} \right].$$

Первая часть – это в точности $M_{t'}$, так как интеграл $\mathcal{F}_{t'}$ -измерим и его можно вытащить из-под условного матожидания, а вторая равна нулю по только что доказанному свойству 3. \square

Другое не очень ожидаемое свойство – непрерывность траекторий кумулятивного интеграла.

Теорема 11.3. Пусть $f \in \mathcal{V}(0, T)$. Случайный процесс $M_t(\omega) = \int_0^t f(s, \omega) dW_s(\omega)$, проиндексированный $t \in [0, T]$, имеет непрерывную модификацию.

Так, если даже функция $f \in \mathcal{V}(0, T)$ была разрывной, то в силу непрерывности траекторий Винеровского процесса интеграл будет непрерывно зависеть от верхнего предела. Наконец, не забываем про изометрию, которую в этот раз мы докажем.

Утверждение 11.4. (Изометрия Ито) Для всех $f \in \mathcal{V}(S, T)$ верна изометрия Ито:

$$\mathbb{E} \left[\left(\int_S^T f(t, \omega) dW_t \right)^2 \right] = \mathbb{E} \left[\int_S^T f(t, \omega)^2 dt \right].$$

▷ Добавим и вычтем под квадратом:

$$\mathbb{E} \left[\left(\int_S^T f(t, \omega) dW_t \right)^2 \right] = \mathbb{E} \left[\left(\int_S^T f dW_t - \int_S^T \phi_n dW_t + \int_S^T \phi_n dW_t \right)^2 \right].$$

Если раскрыть скобки, то получим

$$\begin{aligned} \mathbb{E} \left[\left(\int_S^T f(t, \omega) dW_t \right)^2 \right] &= \mathbb{E} \left[\left(\int_S^T f dW_t - \int_S^T \phi_n dW_t \right)^2 \right] + \\ &+ 2\mathbb{E} \left[\int_S^T (f - \phi_n) dW_t \int_S^T \phi_n dW_t \right] + \mathbb{E} \left[\left(\int_S^T \phi_n dW_t \right)^2 \right], \end{aligned}$$

где первое слагаемое по определению интеграла Ито стремится к нулю. Если последнее сходится, то оно ограничено: в этом случае с помощью неравенства Коши-Буняковского можно показать, что кросс-член сойдётся к нулю. Докажем, что последнее слагаемое стремится к искомому интегралу, для этого рассмотрим, вспомнив про изометрию Ито для простых функций,

$$\mathbb{E} \left[\left(\int_S^T \phi_n dW_t \right)^2 \right] - \mathbb{E} \left[\int_S^T f^2 dt \right] = \mathbb{E} \left[\int_S^T (\phi_n^2 - f^2) dt \right].$$

Если сведём выражение под интегралом к полному квадрату, то по свойству аппроксимирующей функции он будет стремиться к нулю, останется только посмотреть на остаток:

$$\mathbb{E} \left[\int_S^T (\phi_n^2 - f^2) dt \right] = \mathbb{E} \left[\int_S^T (\phi_n - f)^2 dt \right] - 2\mathbb{E} \left[\int_S^T (f - \phi_n) f dt \right].$$

Остаток в свою очередь тоже стремится к нулю, так как по неравенству Коши-Буняковского для подходящего скалярного произведения

$$\mathbb{E} \left[\int_S^T (f - \phi_n) f dt \right] \leq \sqrt{\mathbb{E} \left[\int_S^T (\phi_n - f)^2 dt \right] \mathbb{E} \left[\int_S^T f^2 dt \right]} \xrightarrow{n \rightarrow \infty} 0.$$

□

11.2 Процессы Ито

В обычном математическом анализе мы на самом деле почти никогда не вычисляем интегралы по определению, поэтому не удивительно, что в безусловно более сложном случае с интегралом Ито такой путь является ещё более громоздким. В анализе обычно мы используем набор формул и правил, частично вытекающих из дифференциального исчисления. В случае случайных процессов, как мы видели, с такой стороны зайти не получится – у нас есть только теория интеграла Ито.

Рассмотрим два примера с прошлой лекции. Как и раньше, обозначим за $(W_t)_{t \in \mathbb{R}_+}$ модификацию одномерного Винеровского процесса с непрерывными траекториями и потребуем $W_0 = 0$ почти наверное.

Пример 11.1. Мы показали, что $\int_0^t 1 dW_s = W_t$; на это можно посмотреть под другим углом: Винеровский процесс W представим как интеграл Ито, конкретнее,

$$W_t = \int_0^t dW_s.$$

Пример 11.2. Мы также в прошлой лекции по определению посчитали

$$\int_0^t W_s dW_s = \frac{W_t^2}{2} - \frac{t}{2}.$$

Заметим теперь, что процесс $(W_t^2/2)_{t \in \mathbb{R}_+}$ представим как сумма привычного нам интеграла и интеграла Ито:

$$\frac{W_t^2}{2} = \int_0^t \frac{1}{2} dt + \int_0^t W_s dW_s.$$

Эта идея представления случайного процесса в виде суммы привычного интеграла и интеграла Ито на самом деле оказывается очень общей и очень часто применимой в практических расчётах.

Определение 11.1. Пусть (Ω, \mathcal{F}, P) – вероятностное пространство, $(W_t)_{t \in \mathbb{R}_+}$ – одномерный Винеровский процесс. Процессом Ито называют процесс вида

$$X_t(\omega) = X_0(\omega) + \int_0^t u(s, \omega)ds + \int_0^t v(s, \omega)dW_s, \quad (11.1)$$

где $u, v : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$ функции для которых для любого конечного $T \in \mathbb{R}_+$ верно

1. $v \in \mathcal{V}(0, T)$, при этом согласована с фильтрацией Винеровского процесса W ;
2. интеграл $\int_0^t |u(s, \omega)|ds < \infty$ почти наверное для любого $t \in \mathbb{R}_+$.

Зависимость от ω мы далее будем по возможности опускать для упрощения формул. Равенство (11.1) принято для краткости записывать в *дифференциальной форме*

$$dX_t = u(t, \omega)dt + v(t, \omega)dW_t, \quad X_0 \text{ задано.}$$

Так получается короче, лучше видно связь с дифференциальными уравнениями, но это не более, чем мнемоническое правило, которое, тем не менее, оказывается очень полезным и d частично наследует часть свойств привычного дифференциала, но дифференциалом строго не является.

11.3 Формула Ито и интегрирование по частям

Оказывается, достаточно регулярное преобразование процесса Ито тоже является процессом Ито.

Теорема 11.5. (*Формула Ито*) Пусть X – процесс Ито записанный как

$$dX_t = udt + vdW_t,$$

где $u, v : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$ – непрерывные функции из определения процесса Ито. Если $g : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$ лежит в $C^2(\mathbb{R}_+ \times \mathbb{R})$ (дважды непрерывно дифференцируема), то процесс

$$Y_t = g(t, X_t)$$

тоже является процессом Ито с представлением

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t)(dX_t)^2,$$

причём $(dX_t)^2 = dX_t \cdot dX_t$ вычисляется согласно правилам

$$dt \cdot dt = dt \cdot dW_t = dW_t \cdot dt = 0, \quad dW_t \cdot dW_t = dt.$$

▷ Доказательство этой формулы достаточно объёмное, но с интересными идеями, поэтому мы попробуем набросать примерную схему.

Введём дискретизацию времени с равным шагом Δt и положив $t_0 = 0, t_n = T$. Заметим, что можно записать

$$g(T, X_T) - g(0, X_0) = \sum_{j=0}^{n-1} \Delta g(t_j, X_j), \quad \Delta g(t_j, X_j) = g(t_{j+1}, X_{t_{j+1}}) - g(t_j, X_{t_j}),$$

используя несложную идею телескопической суммы. Разложим каждое слагаемое в ряд Тейлора, используя то, что функция g дважды непрерывно дифференцируема, получим

$$\begin{aligned} g(T, X_T) &= g(0, X_0) + \sum_{j=0}^{n-1} \frac{\partial g}{\partial t}(t_j, X_{t_j}) \Delta t + \sum_{j=0}^{n-1} \frac{\partial g}{\partial x}(t_j, X_{t_j}) \Delta X_j + \\ &\quad \frac{1}{2} \sum_{j=0}^{n-1} \frac{\partial^2 g}{\partial t^2}(t_j, X_{t_j})(\Delta t)^2 + \sum_{j=0}^{n-1} \frac{\partial^2 g}{\partial t \partial x}(t_j, X_{t_j})(\Delta t)(\Delta X_j) + \frac{1}{2} \sum_{j=0}^{n-1} \frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})(\Delta X_j)^2 + R, \end{aligned}$$

где мы обозначили $\Delta X_j = X_{t_{j+1}} - X_{t_j}$.

Попробуем понять, что происходит с этим выражением при $\Delta t \rightarrow 0$. Прежде всего, заметим, что остаток $R = o((\Delta t)^2 + \max_j(\Delta X_j)^2)$ и в силу непрерывности траекторий интеграла Ито и непрерывной зависимости от параметра обычного интеграла Римана, обе компоненты убывают к нулю. Далее, вместо

$$\sum_{j=0}^{n-1} \frac{\partial g}{\partial t}(t_j, X_{t_j}) \Delta t \rightarrow \int_0^T \frac{\partial g}{\partial t}(s, X_s) ds,$$

получится интеграл Римана.

Теперь, вспомнив дискретизацию, рассмотрим подробнее

$$\Delta X_j = u(t_j, \omega) \Delta t + v(t_j, \omega) \Delta W_{t_j}.$$

Вернувшись к вышезаписанным суммам, можем заметить

$$\sum_{j=0}^{n-1} \frac{\partial g}{\partial x} \Delta X_j \rightarrow \int_0^T \frac{\partial g}{\partial x}(s, X_s) u(s, \omega) ds + \int_0^T \frac{\partial g}{\partial x}(s, X_s) v(s, \omega) dW_s,$$

если первый интеграл ожидаемый, то предел в L^2 для второго следует из того, что подынтегральная функция всё ещё из класса $\mathcal{V}(0, T)$.

Перейдём далее к членам второго порядка. В силу непрерывной дифференцируемости g , мы заключаем, что

$$\left| \sum_{j=0}^{n-1} \frac{\partial^2 g}{\partial t^2}(t_j, X_{t_j})(\Delta t)^2 \right| \rightarrow 0$$

поточечно и по теореме Лебега о мажорируемой сходимости в L^2 . Похожая история происходит в кросс-члене, но больше внимания нужно уделить второму слагаемому с Винеровским процессом, которое в пределе в смысле L^2 даст ноль:

$$\left| \sum_{j=0}^{n-1} \frac{\partial^2 g}{\partial t \partial x}(t_j, X_{t_j})(\Delta t)(\Delta X_j) \right| = \left| \sum_{j=0}^{n-1} \frac{\partial^2 g}{\partial t \partial x}(t_j, X_{t_j}) \Delta t(u(t_j, \omega)\Delta t + v(t_j, \omega)\Delta W_{t_j}) \right| \rightarrow 0.$$

В члене с Δt^2 тоже используется теорема Лебега о мажорируемой сходимости. Интересно, что происходит в последнем члене, ради которого нам и пришлось разложить до второго порядка.

$$\sum_{j=0}^{n-1} \frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})(\Delta X_j)^2 = \sum_{j=0}^{n-1} \frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})(A \cdot (\Delta t)^2 + B \cdot (\Delta t)(\Delta W_{t_j}) + C \cdot (\Delta W_{t_j})^2),$$

с какими-то конкретными выражениями A, B, C . Первые два слагаемых мы уже наблюдали выше, они в пределе дают ноль. А третье путём некоторой техники в пределе в смысле L^2 приведёт к ещё одному интегралу Римана. В итоге получим исходную формулу Ито, положив $Y_t = g(t, X_t)$. \square

Из этих формул видно, откуда возникли правила перемножения dt и dW_t : члены с ними в L^2 при уплотнении дискретизации сходятся в L^2 к нулю, что мы проанализировали в процессе доказательства.

11.4 Используем формулу Ито

Применений у формулы Ито огромное количество; она позволяет, к примеру, проще вычислять интегралы.

Пример 11.3. Винеровский процесс W , очевидно сам является процессом Ито, так как его можно записать как

$$W_t = W_0 + \int_0^t dW_s.$$

Предположим, мы хотим вычислить известный нам $\int_0^t W_s dW_s$. Член с dW_t возникает только из второго слагаемого формулы Ито, поэтому мы можем рассмотреть преобразование $Y_t = g(t, W_t) = W_t^2/2$ и согласно формуле Ито получим

$$d\left(\frac{W_t^2}{2}\right) = W_t dW_t + \frac{1}{2}(dW_t)^2 = W_t dW_t + \frac{1}{2}dt$$

или, в интегральной форме,

$$\frac{W_t^2}{2} = \frac{1}{2} \int_0^t ds + \int_0^t W_s dW_s,$$

откуда мы можем выразить

$$\int_0^t W_s dW_s = \frac{W_t^2}{2} - \frac{t}{2}.$$

Упражнение 11.1. Вычислите с помощью формулы Ито $\int_0^t s dW_s$, используя преобразование $g(t, x) = tx$.

Помимо формулы Ито верен также и аналог интегрирования по частям для детерминированных функций.

Теорема 11.6. (*Формула интегрирования по частям*) Пусть $f(s, \omega) = f(s)$ зависит только от $s \in \mathbb{R}_+$, непрерывна и имеет ограниченную вариацию на $[0, t]$. Тогда

$$\int_0^t f(s) dW_s = f(t) W_t - \int_0^t W_s df(s).$$

▷ Мы докажем утверждение с помощью формулы Ито, усилив предположение для f , чтобы не углубляться снова в громоздкие выражения с вариацией. Если f удовлетворяет условиям для формулы Ито, то тогда

$$g(t, x) = f(t)x$$

и $X_t = W_t$ в качестве стартового процесса даёт нужный результат. \square

Пример 11.4. Ответ в Упражнении 11.1 следует мгновенно, но мы можем так же просто, например, записать

$$\int_0^t s^2 dW_s = t^2 W_t - 2 \int_0^t s W_s ds.$$

Последний интеграл упростить не выйдет, но это интеграл Римана и его уже проще исследовать, в частности проще использовать методы Монте-Карло.

Существует и многомерное обобщение формулы Ито, но мы не будем особенно заострять на нём внимание – оно очень естественно, но выглядит более громоздко; подробнее можно найти в [17, Гл. 4.2].

Многомерный процесс Ито задаётся как

$$dX_t = u(t, X_t) dt + v(t, X_t) dW_t,$$

где

$$X_t = \begin{bmatrix} X_t^1 \\ \dots \\ X_t^d \end{bmatrix}, \quad u : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad v : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m},$$

а процесс W_t – это m -мерный Винеровский процесс.

Теорема 11.7. Если $g : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^p$ является дважды непрерывно дифференцируемой, то процесс

$$Y_t = g(t, X_t)$$

является процессом Ито с представлением

$$dY_t^k = \partial_t g^k(t, X_t) dt + \nabla_x g^k(t, X_t) dX_t + \frac{1}{2} dX_t^T \nabla_x^2 g^k(t, X_t) dX_t, \quad k = 1, \dots, p.$$

где произведения раскрываются используя правила

$$dW_t^i dW_t^j = \begin{cases} 0, & i \neq j, \\ dt, & i = j, \end{cases} \quad dtdW_t^i = 0, \quad dt dt = 0.$$

Интересно, что формула Ито позволяет нам построить почти правила из дифференциального исчисления, используя только конструкцию интеграла Ито.

Пример 11.5. Возьмём в качестве базового процесса $X_t = W_t$, это процесс Ито с представлением $dX_t = dW_t$ и $X_0 = 0$ почти наверное. если $f(t)$ – дифференцируемая функция, то взяв $Y_t = g(t, X_t) = f(t)$, по формуле Ито получим

$$df(t) = f'(t)dt,$$

то есть, детерминированные функции дифференцируются так же, как раньше. Вместе с этим d всё ещё линейное отображение, как обычный дифференциал.

Пример 11.6. С другой стороны, используя многомерную формулу Ито для двух процессов Ито X_t и Y_t при помощи трансформации $Z_t = g(t, X_t, Y_t) = X_t Y_t$ мы получим с формулой Ито

$$dZ_t = d(X_t Y_t) = Y_t dX_t + X_t dY_t + dX_t dY_t.$$

По этой причине d , будучи в области определения продолженным на процессы Ито, уже не является дифференциалом, так как не выполнена формула Лейбница. Эта формула после перехода к интегральной нотации даёт ещё одно интересное наблюдение в виде более общей формулы интегрирования по частям:

$$\int_S^T X_t dY_t = X_T Y_T - X_S Y_S - \int_S^T X_t dY_t - \int_S^T dX_t dY_t.$$

Теперь, когда мы, как раньше на матанализе, обзавелись более простыми инструментами для исследования стохастических интегралов и процессов Ито, мы готовы с помощью этого исследовать и дифференциальные уравнения.

Стохастические дифференциальные уравнения

Мы давно собирались, но теперь мы хорошо подготовлены к тому, чтобы разобраться, как думать про стохастические дифференциальные уравнения. В этой лекции мы посмотрим, что ещё интересного можно вытащить из формулы Ито.

12.1 Стохастические дифференциальные уравнения

Одно из, наверное, самых красивых применений формулы Ито – это решение стохастических дифференциальных уравнений (в смысле Ито). Такие уравнения уже отдалённо возникали в начале предыдущей лекции, теперь мы готовы определить их формально. Для простоты будем рассматривать одномерный случай, но на многомерный случай все результаты несложно обобщаются.

Определение 12.1. Пусть (Ω, \mathcal{F}, P) – вероятностное пространство, W – одномерный Винеровский процесс на нём, $b, \sigma : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$. Интегральное уравнение на случайный процесс $(X_t)_{t \in \mathbb{R}_+}$

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s, \quad (12.1)$$

где второй интеграл понимается в смысле Ито, называют стохастическим дифференциальным уравнением Ито.

Уравнение 12.1, как правило записывают в дифференциальной форме, как в предыдущей главе:

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t.$$

Отсюда и идёт название *дифференциальное уравнение*, хотя на самом деле это более короткая запись интегрального уравнения. Как ни странно, несмотря на кажущуюся дальность от настоящих дифференциальных уравнений, для подобного уравнения есть аналогичная теорема существования и единственности, которую, правда, технически существенно сложнее доказать.

Теорема 12.1. Пусть $T > 0$, функции $b, \sigma : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ и W – одномерный Винеровский процесс на вероятностном пространстве (Ω, \mathcal{F}, P) . Пусть выполнены

1. (Условие роста) Существует константа C такая, что для всех t и x верно

$$|b(t, x)| + |\sigma(t, x)| \leq C(1 + |x|);$$

2. (*Условие Липшица*) Существует константа $D > 0$ такая, что для всех $x, y \in \mathbb{R}$

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D|x - y|;$$

3. (*Нач. условие*) Z – случайная величина, независимая от Винеровского процесса и такая, что $\mathbb{E}[Z^2] < \infty$.

Тогда существует и единственное решение $(X_t)_{t \in [0, T]}$ дифференциального уравнения

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = Z \text{ н.н.}$$

Стохастические дифференциальные уравнения возникают ровно в тех ситуациях, которые мы обсуждали на предыдущей лекции: когда есть детерминированная модель, но хочется добавить случайный шум.

Пример 12.1. Рост суммы банковского вклада с капитализацией и непрерывным начислением процентов со ставкой $r\%$ годовых описывается обыкновенным дифференциальным уравнением

$$\frac{dX}{dt} = rX, \quad X(0) = X_0.$$

Известно решение этой задачи Коши:

$$X(t) = X_0 e^{rt}.$$

Как быть, если процентная ставка подвержена некоторым случайным возмущениям, например, от биржи? Можем записать модель в виде стохастического дифференциального уравнения Ито:

$$dX_t = (rdt + \sigma dW_t)X_t.$$

Из решения уравнения без Винеровского процесса можно предложить преобразование $g(t, x) = \ln x$ (попробуйте мысленно поделить обе части на X_t и обратить внимание на то, что слева). Используя формулу Ито, для процесса $Y_t = \ln X_t$ получим

$$\begin{aligned} dY_t &= \frac{1}{X_t}dt - \frac{1}{2X_t^2}(dX_t)^2, \\ dY_t &= (r - \sigma^2/2)dt + \sigma dW_t. \end{aligned}$$

Перейдя к интегральной форме, получим решение:

$$\ln X_t = \ln X_0 + (r - \sigma^2/2)t + \sigma W_t.$$

Так X – это в частности процесс геометрического Броуновского движения с параметрами $r, \sigma, X_0 \in \mathbb{R}$, которое мы рассматривали в Лекции 4.

Пример 12.2. Рассмотрим уравнение

$$dX_t = -\theta X_t dt + \sigma dW_t$$

с ненулевыми $\theta, \sigma \in \mathbb{R}$ и некоторым начальным условием X_0 . Без члена с dW_t получим обыкновенное уравнение с известным решением $X_0 e^{-\theta t}$, однако попробовав формулу Ито, поймём, что нужно немного внимательнее быть со знаком в экспоненте. Рассмотрим

$$Y_t = X_t e^{\theta t}$$

и применим формулу Ито, получим, что

$$dY_t = \sigma e^{\theta t} dW_t$$

или, в интегральной форме и с обратной заменой $X_t = e^{-\theta t} Y_t$,

$$X_t = X_0 e^{\theta t} + \sigma \int_0^t e^{\theta(s-t)} dW_s.$$

Можно показать, что X – гауссовский процесс и, дополнительно, по формуле интегрирования по частям

$$\int_0^t e^{\theta(s-t)} dW_s = W_t - \int_0^t e^{\theta(s-t)} W_s ds.$$

Посчитав ковариационную функцию и матожидание, мы увидим, что X – это в частности процесс Орнштейна-Уленбека из Лекции 4.

Упражнение 12.1. Процесс Орнштейна-Уленбека, возвращающийся к заданному среднему $\mu \in \mathbb{R}$ будет иметь уравнение

$$dX_t = (\mu - X_t) dt + \sigma dW_t.$$

Используя формулу Ито и предыдущий пример, запишите решение этого уравнения (если будет оставаться неупрощаемый интеграл по dt в правой части, то это нормально).

Вполне естественно, что далеко не все стохастические дифференциальные уравнения решаются аналитически. Для целей финансового моделирования используются часто модели, для которых можно что-то доказать, но, как правило, решаются уравнения численными методами (как это делать, см. последнюю главу лекции). Например, мы уже упоминали, что геометрическое Броуновское движение (процесс GBM) можно использовать как базовый строительный блок, в котором можно заменять процентную ставку и волатильность различными случайными процессами и моделями временных рядов, получая при этом более сложную динамику, моделируемую с помощью более сложного дифференциального уравнения (в смысле Ито). Рассмотрим в этом свете две новых модели, которые ранее не встречались.

Пример 12.3. Модель процентных ставок, похожая на Упражнение 12.1,

$$dR_t = \theta(\mu - R_t)dt + \sigma dW_t, \quad \theta > 0, \quad \mu > 0$$

в финансовой литературе называется моделью *Вашичека* (*Vašiček*) и известна с 1977го года. Обобщением модели *Вашичека* является модель *Хола-Уайта* (*Hull-White*), предложенная в 1990 году:

$$dR_t = (\theta(t) - \alpha R(t))dt + \sigma(t)dW_t.$$

Интуитивно, $\theta(t)$ позволяет описывать некоторую детерминированную динамику (например, тренд или сезонность), а член $-\alpha R(t)$ даёт компонент от процесса Орнштейна-Уленбека, заставляя решение двигаться в сторону значения $\theta(t)$. Наконец, $\sigma(t)$ позволяет отказаться от постоянной волатильности и описывать её некоторой детерминированной функцией. Обобщение этой модели на случай переменного $\alpha(t)$ тоже называется моделью *Хола-Уайта*.

Пример 12.4. Другим обобщением модели *Вашичека* явилась модель *Кокса-Ингерсолла-Росса* (1985), которая связала волатильность с самой процентной ставкой.

$$dR_t = \theta(\mu - R_t)dt + \sigma\sqrt{R_t}dW_t, \quad \theta > 0.$$

В этой модели помимо уже известных эффектов стремления к среднему, описывается также и феномен большей волатильности при большей процентной ставке. Интересно, что при любых (!) положительных θ, μ, σ значение R_t остаётся неотрицательным, более того, ноль не достигается, если волатильность достаточно мала по сравнению с $\theta\mu$, то есть, при $2\theta\mu \geq \sigma^2$.

Все эти модели можно оценивать из данных реализаций временного ряда, однако построение таких процедур – это вопрос исследовательского уровня.

12.2 Численные методы решения

Наконец, давайте попробуем привести несколько методов для численного решения стохастических дифференциальных уравнений. Последнее пригождается, например, для вычисления квантилей статистики ADF-теста (расширенный критерий Дики-Фулера).

Рассмотрим стохастическое дифференциальное уравнение в смысле Ито

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \tag{12.2}$$

с начальным условием $X_0 = x_0 \in \mathbb{R}$. Положим, что решение существует и единствено. Общая идея состоит в том, чтобы ввести дискретизацию по времени и затем построить

итеративный алгоритм.

Первый простейший метод, как и в обыкновенных дифференциальных уравнениях, – это *явный метод Эйлера* (для стохастических уравнений такой метод называют также *методом Эйлера-Маруямы*):

$$\begin{aligned} X_{t_{k+1}} &= X_{t_k} + b(t_k, X_{t_k})\Delta t_k + \sigma(t_k, X_{t_k})\Delta W_k, \quad X_0 = x_0 \\ \Delta t_k &= t_{k+1} - t_k, \quad \Delta W_k = W_{t_{k+1}} - W_{t_k}. \end{aligned}$$

Можно показать, что средняя ошибка метода имеет порядок $O(\sqrt{\Delta t})$, на полпорядка меньше, чем для обыкновенного дифференциального уравнения. Тем не менее, в силу вычислительной эффективности часто именно он используется для многих задач.

Если требуется делать более точные вычисления, то для автономных уравнений (b и σ не зависят от t) есть следующий метод, называемый *методом Мильштейна*:

$$\begin{aligned} X_{t_{k+1}} &= X_{t_k} + b(X_{t_k})\Delta t_k + \sigma(X_{t_k})\Delta W_k + \frac{1}{2}b(X_{t_k})b'(X_{t_k})((\Delta W_k)^2 - \Delta t_k), \quad X_0 = x_0 \\ \Delta t_k &= t_{k+1} - t_k, \quad \Delta W_k = W_{t_{k+1}} - W_{t_k}. \end{aligned}$$

Такой метод имеет среднюю ошибку порядка $O(\Delta t)$ и почти так же эффективен, как и метод Эйлера. Требуется, правда, автономность (что не всегда выполняется) и дифференцируемость b .

Наконец, есть аналоги методов Рунге-Кутты для стохастических дифференциальных уравнений, но записываются они более сложно.

12.3 Формула Блэка-Шоулза

Формулу Блэка-Шоулза можно вывести, если знать модель позади процесса цен. Эта формула получается для случая Европейского опциона на $d = 1$ товар. В модели Блэка-Шоулза рассматривается рынок состоящий из безрискового актива B_t (банковский счёт), растущего за счёт постоянной процентной ставки r , и рискового актива на бирже с ценами X_t . Математически они изменяются как

$$\begin{aligned} dX_t &= \mu X_t dt + \sigma X_t dW_t, \quad X_0 = x_0, \\ dB_t &= r B_t dt, \quad B_0 = 1. \end{aligned}$$

Инвестор может распределять свой капитал между двумя активами. Поскольку второй актив безрисковый, удобнее рассматривать его как бэйзлайн и смотреть, сколько можно получить по сравнению с ним. Второе уравнение решается легко, это просто $B_t = e^{rt}$. Так, можно рассматривать преобразованный ряд цен

$$\tilde{X}_t = X_0 e^{(\mu - r - \sigma^2/2)t + \sigma W_t}.$$

В любом случае, если мы оцениваем опцион, то правильнее его оценивать исходя из преобразованного ряда, потому что нет смысла работать с рисковым активом, если он растёт медленнее безрискового. Итак, мы имеем модель цены

$$X_t = X_0 e^{(\mu - \sigma^2/2)t + \sigma W_t},$$

где мы опускаем тильду И напишем μ вместо $\mu - r$ для красоты, держа в уме, что ряд преобразованный.

Вспомним евро-опцион, его цена – это максимально возможный выигрыш при самом выгодном использовании (при условии, что опцион уже на руках). Европейский опцион можно использовать (или не использовать), в момент истечения T , то есть, нам нужно посчитать

$$P_t = e^{-\mu(T-t)} \mathbb{E} [f(X_T) \mid X_t = x].$$

Функция выплат f в случае евро-опциона выглядит как

$$f(x) = (K - x)_+ \quad (\text{Put})$$

или

$$f(x) = (x - K)_+ \quad (\text{Call}).$$

Рассмотрим формулу для Call, для Put она выводится точно так же. Здесь помогает точное вычисление и знание распределения X_T :

$$\begin{aligned} \mathbb{E} [f(X_T) \mid X_t = x] &= \mathbb{E} [(X_T - K) \mathbb{1}(X_T > K) \mid X_t = x] = \\ &= \mathbb{E} [X_T \mathbb{1}(X_T > K) \mid X_t = x] - K \mathbb{P}(X_T > K \mid X_t = x). \end{aligned}$$

При условии $X_t = x$ распределение цены актива

$$X_T \sim \log N \left((\mu - \sigma^2/2)(T-t) + \ln x, \sigma^2(T-t) \right).$$

По этой причине вероятность

$$\mathbb{P}(X_T > K \mid X_t = x) = \Phi \left(\frac{\ln(x/K) + (\mu - \sigma^2/2)(T-t)}{\sigma \sqrt{T-t}} \right),$$

Чтобы это получить, достаточно заметить, что $\ln X_T$ имеет нормальное распределение, у которого $\Phi(x)$ – это функция распределения.

Аналогично можем рассмотреть первое:

$$\mathbb{E} [X_T \mathbb{1}(X_T > K) \mid X_t = x] = \int_K^\infty z \frac{1}{\sqrt{2\pi\sigma^2(T-t)}} e^{-\left(\ln z - ((\mu - \sigma^2/2)(T-t) + \ln x)\right)^2/(2\sigma^2(T-t))} dz.$$

делаем замену $y = \ln z$, тогда

$$\begin{aligned} & \int_K^\infty \frac{1}{\sqrt{2\pi\sigma^2(T-t)}} e^{-\left(\ln z - ((\mu - \sigma^2/2)(T-t) + \ln x)\right)^2/(2\sigma^2(T-t))} dz = \\ &= \int_{\ln K}^\infty \frac{e^y}{\sqrt{2\pi\sigma^2(T-t)}} e^{-\left(y - ((\mu - \sigma^2/2)(T-t) + \ln x)\right)^2/(2\sigma^2(T-t))} dy. \end{aligned}$$

После выделения полного квадрата останется

$$\begin{aligned} & \int_{\ln K}^\infty \frac{e^y}{\sqrt{2\pi\sigma^2(T-t)}} e^{-\left(y - ((\mu - \sigma^2/2)(T-t) + \ln x)\right)^2/(2\sigma^2(T-t))} dy = \\ &= e^{\ln x + \mu(T-t)} \Phi\left(\frac{\ln(x/K) + (\mu + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right) \end{aligned}$$

и в результате получим формулу

$$P_t = x\Phi\left(\frac{\ln(x/K) + (\mu + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right) + Ke^{-\mu(T-t)}\Phi\left(\frac{\ln(x/K) + (\mu - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right).$$

12.4 Уравнение Блэка-Шоулза

Вообще премию Блэку и Шоулзу, конечно, дали не за этот вывод, а за дифференциальное уравнение, которое не только позволяет вычислить цену (с позиции вычислений оно даже не очень удобно), но что важнее, открывает глаза на то, какие факторы оказывают влияние на цену. Такое уравнение было бы, вероятно, невозможно получить без развития инструментов исчисления Ито.

Всё начинается с модели цен; геометрическое Броуновское движение во многом продиктовано эмпирическими наблюдениями.

$$dX_t = \mu X_t dt + \sigma X_t dW_t.$$

Пусть цена опциона (или вообще некоторого дериватива) $C(x, t)$ в момент t при цене x – это функция достаточно гладкая, чтобы можно было применить формулу Ито. Конечно, это скорее всего не так, но это позволит нам получить уравнение и в перспективе доказать, что у него есть и более реалистичные решения в подходящем смысле. Такая ситуация – не редкость. Если можно применить формулу Ито, то тогда

$$dC = \partial_t C(X_t, t) dt + \partial_x C(X_t, t) dX_t + \frac{1}{2} \partial_x^2 C(X_t, t) (dX_t)^2,$$

а после упрощения и подстановки

$$dC = \left(\partial_t C(X_t, t) + \partial_x C(X_t, t) \mu X_t + \frac{1}{2} \partial_x^2 C(X_t, t) \sigma^2 X_t^2 \right) dt + \partial_x C(X_t, t) \sigma X_t dW_t.$$

Далее следует очень интересная идея. Согласно этому уравнению, весь риск и вся случайность динамики даётся последним слагаемым. Мы могли бы докупить в портфель

ещё немного (Δ) актива X_t , чтобы исключить этот риск полностью, тогда общий капитал портфеля будет $C + \Delta X_t$, а в уравнении для портфеля

$$d(C + \Delta X_t) = \left(\partial_t C(X_t, t) + \partial_x C(X_t, t) \mu X_t + \frac{1}{2} \partial_x^2 C(X_t, t) \sigma^2 X_t^2 + \Delta \mu X_t \right) dt + (\partial_x C(X_t, t) + \Delta) \sigma X_t dW_t.$$

Если мы хотим исключить случайность в динамике цены портфеля, то выбор очевиден:

$$\Delta = -\partial_x C(X_t, t).$$

(Хотя мы будто забыли, что у Δ тоже есть производная, это связано с конкретным финансовым смыслом) Такой приём называется *дельта-хеджированием*, а величина $-\partial_x C(X_t, t)$ оценивается для деривативов и называется *дельтой*. По фундаментальной теореме оценивания (Fundamental Theorem of Asset Pricing), на полном эффективном рынке (это отдельная тема для разговоров на уровне построения специальных мер) безрисковый портфель должен расти с процентной ставкой r :

$$d(C + \Delta X_t) = r(C + \Delta X_t) dt.$$

Если приравняем теперь правую часть с тем, что было выше, и уберём с обеих сторон dt , то получим *уравнение Блэка-Шоулза* для цены опциона:

$$\partial_t C + rx \partial_x C + \frac{\sigma^2 x^2}{2} \partial_x^2 C = rC.$$

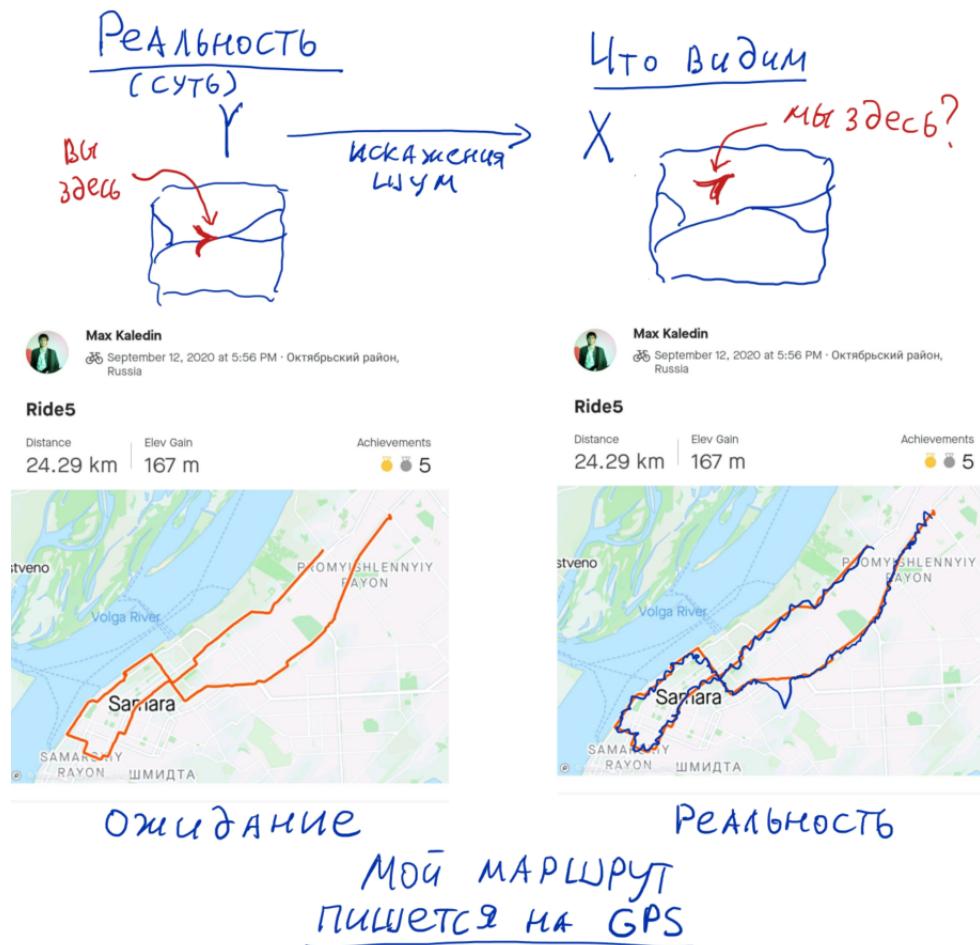
Это уравнение в частных производных на функцию $C(x, t)$, которое можно рассматривать с краевым условием $C(x, T) = c_T(x)$, то есть, известной ценой в момент истечения.

Мы получили интересный факт: некоторая функция от случайного процесса удовлетворяет вполне детерминированному дифференциальному уравнению в частных производных. И это не случайное наблюдение, далее в диффузиях мы увидим ещё больше таких примеров.

Задача фильтрации

В этой лекции мы рассмотрим одну популярную инженерную задачу, актуальную для робототехники, интернета вещей и вообще работы с любыми датчиками. Мы поймём, как можно из наблюдаемых зашумлённых данных эффективно оценить скрытую истину. Мы не в силах покрыть все детали этой огромной области, но попробуем понять, чем нам в этой задаче может помочь знание стохастических дифференциальных уравнений.

13.1 Выделение целевого сигнала из шума



Мы сейчас живём в мире, наполненном огромным количеством разных устройств. В том числе есть огромное количество маленьких простых устройств, которые обслуживают самые разные ежедневные потребности. Огромная развивающаяся быстрыми темпами область – интернет вещей (Internet of Things, IoT), где конструируются различные технические решения, основанные на локальном использовании небольших вычислительных

устройств, состоящих из вычислительной единицы (микроконтроллер или микропроцессор) и системы датчиков: для замера температуры, освещённости, датчики геопозиционирования (GPS или ГЛОНАСС), акселерометры и многое другое.

Проблема в том, что сами по себе датчики несовершены, даже на мощных смартфонах вы столкнётесь с тем, что в подземном переходе или тоннеле геопозиционирование может быть сильно сложнее. Но и в обычных обстоятельствах в поле рядом со Сколтехом точность определения геопозиции по GPS может составлять 5-10 метров в зависимости от используемых датчиков. По этой же причине приложения замера пройденного расстояния, работающие от GPS выдают всегда не совсем точные данные.

Но ситуация бы была гораздо хуже, если бы не использовались дополнительно алгоритмы фильтрации сигнала, которые в предположении некоторой динамики системы помогают из нескольких шумных наблюдений уточнить настоящий сигнал. Знаковое первенство (хотя, говорят, что это было сделано на плечах гигантов) в изобретении таких решений для технических приложений принадлежит Рудольфу Калману и Ричарду Бюси, которые решили задачу при условии известной динамики в дискретном времени в 1960 [13] и немного позже в непрерывном в 1963. Одновременно этой задачей также занимались датский математик и астроном Николай Тиле, американский инженер Петер Сверлинг и советский математик Руслан Стратонович. В 1961 году Калман встречался со Стратоновичем в Москве[19], когда Стратонович успел опубликовать некоторое количество работ по этой теме (из них, в частности, [20, 28]). Стратонович и Калман дальше много переписывались, а сам Калман приезжал в СССР и Россию. Фильтр Калмана является частным случаем фильтра Стратоновича, который строится для нелинейных систем.

13.2 Постановка задачи

Мы предполагаем, что система работает по некоторым своим законам, которые моделируются как дифференциальные уравнения типа Стратоновича

$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t) \circ dU_t,$$

где U – многомерный Винеровский процесс. Сами Y_t ненаблюдаемые, вместо них наблюдается некоторая зашумлённая величина

$$H_t = c(t, Y_t) + \gamma(t, Y_t)\widetilde{W}_t,$$

где \widetilde{W} – процесс гауссовского белого шума. Мы ранее видели проблемы, возникающие с такой постановкой, поэтому чтобы их обойти, вместо H_t будем рассматривать процесс аккумулированных наблюдений

$$X_t = \int_0^t H_s ds.$$

Это не искажает задачу, потому что эти процессы и фильтрация устроены так, что H_s измерима относительно \mathcal{F}_s тогда и только тогда, когда X_s тоже измерима. Согласившись на такую постановку, мы получим уравнение для наблюдений

$$dX_t = c(t, Y_t)dt + \gamma(t, Y_t) \circ dV_t.$$

К счастью для нас, уравнения Ито и Стратоновича взаимозаменяемы, нужно только скорректировать функцию перед dt , поэтому дальше мы будем рассматривать только постановку в смысле Ито, про которую мы уже много знаем.

Итак, мы руководствуемся следующей моделью системы и наблюдений:

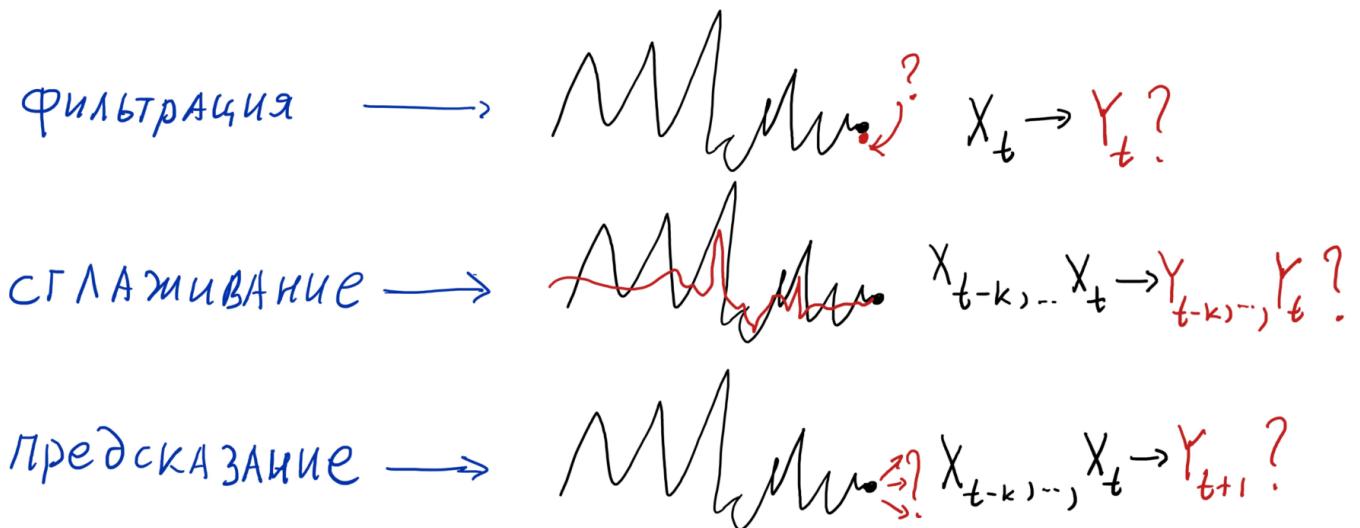
$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t)dU_t, \text{ (модель системы)}$$

$$dX_t = c(t, Y_t)dt + \gamma(t, Y_t)dV_t, \text{ (модель наблюдений),}$$

где U_t и V_t независимые многомерные Винеровские процессы. Мы хотели бы решать фундаментально три задачи, используя только наблюдения

1. *Задача фильтрации.* При известном X_t на ходу оценивать Y_t .
2. *Задача сглаживания.* При известной истории X до t оценить значения Y_s при $s \leq t$.
3. *Задача предсказания.* При известной истории X до t оценить будущий момент $Y_{t+\Delta}$.

Лучше всего их иллюстрирует задача в дискретном времени.



Формально оценить с использованием истории означает, что мы ищем оценку \hat{Y}_t как согласованную с фильтрацией (\mathcal{F}_t) функцию (мы используем фильтрацию, порождённую Винеровскими процессами). Более того, мы хотели бы оценить *наилучшим образом*, то есть, полученная оценка должна минимизировать

$$\mathbb{E} \left[\left| \hat{Y}_t - Y_t \right|^2 \right],$$

иными словами, быть оптимальным L^2 -прогнозом среди

$$\mathcal{K}_t = \left\{ Y : \Omega \rightarrow \mathbb{R}^n; Y \in L^2(P), Y \text{ измерима относительно } \mathcal{F}_t \right\}.$$

Оказывается, что в данных условиях универсальным решением является алгоритм, который носит название *фильтра Калмана* или *фильтра Калмана-Бюси*. Мы будем рассматривать задачу, где параметры модели и наблюдений известны, но вообще полезно знать, что существуют процедуры их оценки на основе ЕМ-алгоритма, которые заслуживают отдельного курса.

Мы упростим себе задачу, рассматривая только одномерную постановку и линейную, то есть,

$$\begin{aligned} dY_t &= F(t)Y_t dt + C(t)dU_t, \quad (\text{модель системы}) \\ dX_t &= G(t)Y_t dt + D(t)dV_t, \quad (\text{модель наблюдений}), \end{aligned}$$

где $F, C, G, D : \mathbb{R} \rightarrow \mathbb{R}$ предполагаются ограниченными на ограниченных интервалах. Уже здесь нужно будет сделать очень много усилий, чтобы получить результат. Детальный вывод с подробностями (они вполне нам по силам, но их очень много) можно найти в [17, Гл.6]. Технические подробности мы будем по возможности опускать, но попробуем хотя бы поверхностно понять идею позади конструкции фильтра Калмана.

13.3 Строим оценку сигнала(простой случай)

Чтобы понять лучше как можно подойти к решению задачи фильтрации, полезно рассмотреть очень простой случай. Пусть X, W_1, W_2, \dots независимые, с нулевым матожиданием и дисперсиями $\mathbb{E}[X^2] = a^2$, $\mathbb{E}[W_j^2] = m^2$. Пусть

$$Z_j = X + W_j,$$

как получить лучшую L^2 -оценку X среди всех измеримых относительно $\mathcal{F}_k = \sigma(X_j, j \leq k)$? Для начала попробуем сконструировать линейную оценку, то есть, оценку из класса

$$\mathcal{L}_k = \left\{ \sum_{i=1}^k c_i Z_i : c_i \in \mathbb{R} \right\}.$$

Это линейное подпространство $L^2(P)$, а само пространство $L^2(P)$ является Гильбертовым: там есть скалярное произведение через интеграл и относительно метрики, порождённой скалярным произведением, оно полное метрическое пространство. В этом смысле поиск наилучшей в L^2 оценки – это на самом деле задача поиска ортогональной проекции $\mathcal{P}_{\mathcal{L}_k}$ на конечномерное подпространство \mathcal{L}_k .

Вообще Z_j зависимы через X , поэтому в качестве базиса для ортогональной проекции не очень удобны, хотя они действительно линейно независимы, в противном случае $Z_j(\omega)$ выражалось бы линейно через остальные для всех $\omega \in \Omega$ и какого ненулевого α , что противоречит независимости W и X . Чтобы упростить задачу поиска проекции, можно применить процедуру Грама-Шмидта. Так мы конструируем

$$A_1 = Z_1$$

и далее

$$A_j = Z_j - \sum_{i=1}^{j-1} \frac{\langle Z_j, A_i \rangle}{\langle A_i, A_i \rangle} A_i = Z_j - \sum_{i=1}^{j-1} \frac{\mathbb{E}[Z_j A_i]}{\mathbb{E}[A_i^2]} A_i.$$

В новом ортогональном базисе проекции считать сильно проще, оптимальной оценкой в $L^2(P)$ будет

$$\hat{X}_k = \sum_{i=1}^k \frac{\mathbb{E}[X A_i]}{\mathbb{E}[A_i^2]} A_i,$$

где мы посчитали

$$c_i = \frac{\mathbb{E}[X A_i]}{\mathbb{E}[A_i^2]}.$$

Если вернуться к процессу ортогонализации, один шаг – это вычисление

$$A_j = Z_j - \mathcal{P}_{\mathcal{L}_{j-1}}(Z_j) = Z_j - \mathcal{P}_{\mathcal{L}_{j-1}}(X),$$

последнее происходит из-за того, что X и W_j независимы и поэтому $\mathbb{E}[X W_j] = 0$. Из этого также следует интересное наблюдение:

$$\mathbb{E}[X A_j] = \mathbb{E}[(X - \hat{X}_{j-1})^2].$$

То есть, новую оценку можно получить путём некоторой линейной коррекции старой с помощью уже известной в момент k невязки $Z_k - \hat{X}_{k-1}$:

$$\hat{X}_k = \hat{X}_{k-1} + \frac{\mathbb{E}[(X - \hat{X}_{k-1})^2]}{\mathbb{E}[(X - \hat{X}_{k-1})^2] + m^2} (Z_k - \hat{X}_{k-1}).$$

Проанализируем это матожидание, добавив $\pm \hat{X}_{j-1}$:

$$S_j = \mathbb{E}[(X - \hat{X}_j \pm \hat{X}_{j-1})^2] = S_{j-1} + \mathbb{E}[(\hat{X}_j - \hat{X}_{j-1})^2] - 2\mathbb{E}[(X - \hat{X}_{j-1})(\hat{X}_j - \hat{X}_{j-1})].$$

Почти всё мы уже видели, например,

$$\mathbb{E}[(\hat{X}_j - \hat{X}_{j-1})^2] = \frac{S_{j-1}^2}{(S_{j-1} + m^2)^2} \mathbb{E}[(Z_j - \hat{X}_{j-1})^2]$$

и

$$\mathbb{E}[(X - \hat{X}_{j-1})(\hat{X}_j - \hat{X}_{j-1})] = \mathbb{E}[X(\hat{X}_j - \hat{X}_{j-1})] = \frac{S_{j-1}}{S_{j-1} + m^2} \mathbb{E}[X(Z_j - \hat{X}_{j-1})] = \dots = \frac{S_{j-1}^2}{S_{j-1} + m^2},$$

и ещё

$$\mathbb{E} \left[(Z_j - \hat{X}_{j-1})^2 \right] = \mathbb{E} \left[(X + W_j - \hat{X}_{j-1})^2 \right] = S_{j-1} + m^2.$$

Если собрать это вместе, то мы получим уравнение на S_k

$$S_k = S_{k-1} - \frac{S_{k-1}^2}{S_{k-1} + m^2}.$$

На каждом шаге мы можем пересчитывать новое значение S_k основываясь на предыдущем. Таким образом, мы получили совсем простую процедуру вычисления нового значения, уточнённого с помощью невязки предыдущего прогноза

$$\hat{X}_k = \hat{X}_{k-1} + \frac{S_{k-1}}{S_{k-1} + m^2} (Z_k - \hat{X}_{k-1}).$$

Оказывается, можно даже получить чёткую формулу \hat{X}_k , без рекуррентного отношения. Можно подумать, что стоит попробовать искать решение в виде $\hat{X}_k = \alpha_k \sum_{j=1}^k Z_j$. Это ортогональная проекция тогда и только тогда, когда $X - \hat{X}_k$ ортогонально Z_i для всех $i \leq k$. В смысле L^2 это даёт

$$\mathbb{E} \left[(X - \hat{X}_k) Z_i \right] = \mathbb{E} \left[(X - \alpha_k \sum_{j=1}^k Z_j) Z_i \right] = a^2 - \alpha_k (ka^2 + m^2) = 0.$$

Это должно равняться нулю, поэтому

$$\alpha_k = \frac{a^2}{ka^2 + m^2},$$

что приводит к

$$\hat{X}_k = \frac{a^2}{a^2 + m^2/k} \frac{1}{k} \sum_{j=1}^k Z_j.$$

Это не совсем обычное предсказание выборочным средним. Если m^2 небольшое относительно a^2 , то в какой-то момент мы получим, что прогноз примерно равен выборочному среднему наблюдений, а если наоборот большое, то на начальных этапах наблюдения вообще игнорируются и прогноз близок к нулю. И в любом случае в силу закона больших чисел при $k \rightarrow \infty$ прогноз почти наверное сойдётся к среднему, то есть, к нулю.

13.4 Фильтр Калмана-Бюси (непрерывное время)

Что мы узнали из этого простого случая?

1. Мы умеем строить *линейные* оценки в дискретном случае;
2. Мы получаем прогноз, обновляя старый прогноз и *корректируя* его с помощью невязки $Z_k - \hat{X}_{k-1}$;
3. Z_j – процесс наблюдений, но он с ортогональными приращениями, что позволило так много упростить по пути;

4. Мы знали решение модели состояний, это был просто X .

Для того, чтобы перейти к дискретному случаю, нужно суметь обобщить эти наблюдения. Конкретнее:

1. Как соотнести линейную оценку с проекцией на пространство \mathcal{K}_t , которое больше не конечномерное и даже необязательно там есть хотя бы счётный базис.
2. Как сделать замену процесса наблюдений на удобный процесс с ортогональными приращениями.
3. Как получить итоговое соотношение для \hat{X}_t , конкретнее в непрерывном времени нас интересует дифференциальное уравнение, а не рекуррентное отношение.

13.4.1 Проекция – это условное матожидание

И начинаем мы с пункта 1, вспоминая по пути, что такое условное матожидание.

Лемма 13.1. Пусть $\mathcal{H} \subset \mathcal{F}$ – это сигма-алгебра и $X \in L^2(P)$ (то есть, есть второй момент) измерима относительно \mathcal{F} (то есть, является корректной случайной величиной). Обозначим за $\mathcal{P}_{\mathcal{H}}$ ортогональную проекцию на $\mathcal{K}_{\mathcal{H}}$, пространство измеримых относительно \mathcal{H} функций. Тогда

$$\mathcal{P}_{\mathcal{H}}(X) = \mathbb{E}[X \mid \mathcal{H}].$$

▷ Мы неоднократно вспоминаем это свойство, но полезно вернуться к нему в новом контексте. Вспомним, что условное матожидание

1. существует;
2. единственno (если есть другое, то разница почти наверное 0);
3. как функция от ω , измеримо относительно \mathcal{H} ;
4. по определению это такая функция, что для всех $A \in \mathcal{H}$

$$\int_A \mathbb{E}[X \mid \mathcal{H}] dP = \int_A X dP.$$

С другой стороны, $\mathcal{P}_{\mathcal{H}}(X)$ – ортогональная проекция, это верно тогда и только тогда, когда для всех $Y \in \mathcal{K}_{\mathcal{H}}$

$$\int_{\Omega} (X - \mathcal{P}_{\mathcal{H}}(X)) Y dP = 0.$$

Индикаторная функция $\mathbb{1}_A(\omega)$ при $A \in \mathcal{H}$ будет измеримой относительно \mathcal{H} , поэтому для всех A

$$\int_{\Omega} (X - \mathcal{P}_{\mathcal{H}}(X)) \mathbb{1}_A dP = \int_A (X - \mathcal{P}_{\mathcal{H}}(X)) dP = 0.$$

Мы получили в точности определение условного матожидания, которое единственno. \square

Этот факт позволяет нам уточнить, как считать проекцию, то есть, в качестве наилучшего L^2 -прогноза мы должны искать

$$\hat{Y}_t = \mathcal{P}_{\mathcal{K}_t}(Y_t) = \mathbb{E}[Y_t | \mathcal{F}_t],$$

то есть, условное матожидание при условии сигма-алгебры из фильтрации.

13.4.2 Линейные оценки против измеримых

Далее, заметим, что процесс (Y_t, X_t) – гауссовский, мы явно выводили решение дифференциального уравнения с помощью формулы Ито и интегрирующего множителя в одномерном случае, к которому сводится данный путём замены базиса в \mathbb{R}^2 . Это важный момент, потому что теперь мы можем сопоставить наилучшие в L^2 линейные и измеримые оценки, используя гауссовость.

Лемма 13.2. Рассмотрим случайные величины Y и $X_s, s \leq t$ из $L^2(P)$ и такие, что

$$Y, X_{s_1}, \dots, X_{s_n}$$

имеют совместное гауссовское распределение для любого конечного набора индексов $s_i \leq t$. Тогда

$$\mathcal{P}_{\mathcal{L}_t}(Y) = \mathcal{P}_{\mathcal{K}_t}(Y) = \mathbb{E}[Y | \mathcal{F}_t],$$

где (\mathcal{F}_t) – фильтрация, с которой согласован процесс X , \mathcal{K}_t – множество функций, измеримых относительно \mathcal{F}_t , а множество \mathcal{L}_t – это замыкание множества

$$\left\{ \sum_{i=1}^k c_i X_{t_i} : k \geq 1, c_i \in \mathbb{R}, t_i \leq t \right\}.$$

▷ Рассмотрим невязку с линейной проекцией $\tilde{Y} = Y - \mathcal{P}_{\mathcal{L}}(Y)$ и покажем, что она независима от сигма-алгебры \mathcal{F}_t . Предел в L^2 нормальных величин тоже нормальная величина (включая константу), если он существует, поэтому проекция будет нормальной случайной величиной, как и невязка. Поэтому

$$\tilde{Y}, X_{s_1}, \dots, X_{s_n}$$

тоже имеют совместное гауссовское распределение. Так как мы считали ортогональную проекцию,

$$\mathbb{E}[\tilde{Y} X_{s_i}] = 0$$

и в силу гауссности \tilde{Y} не зависит от любого набора $X_s, s \leq t$, которыми порождена \mathcal{F}_t , следовательно, невязка не зависит от \mathcal{F}_t . В этом случае для любого $A \in \mathcal{F}_t$

$$\mathbb{E}[\mathbf{1}_A \tilde{Y}] = \mathbb{E}[\mathbf{1}_A] \mathbb{E}[\tilde{Y}] = 0,$$

а значит, второе слагаемое равно нулю, и это в результате даёт определение условного матожидания. \square

Таким образом, мы соединили результат проекции с конструкцией линейных оценок. Вообще, пространство \mathcal{L} несмотря на свою общность кажется относительно представимым: это все конечные комбинации случайных величин до момента t и все L^2 -пределы таких сумм. Немного знания про L^2 -пространства (доказывать не будем) и можно получить его точное описание, которое не очень удивляет. Вернёмся в рамки изначальной модели с процессами Y, X .

Теорема 13.3. *Пространство \mathcal{L}_T , определённое как замыкание множества*

$$\left\{ \sum_{i=1}^k c_i X_{t_i} : k \geq 1, c_i \in \mathbb{R}, t_i \leq t \right\},$$

в частности

$$\mathcal{L}_T = \left\{ c_0 + \int_0^T f(t) dX_t; c_0 \in \mathbb{R}, f \in L^2[0, T] \right\}.$$

13.4.3 Процесс невязок

Как и в дискретном случае, мы строим коррекцию, рассматривая невязки на предыдущих шагах, аналог процесса $Z_k - \hat{X}_{k-1}$ из начала – это процесс

$$N_t = X_t - \int_0^t \mathcal{P}_{\mathcal{K}_s}(G(s)Y_s) ds = X_t - \int_0^t G(s)\hat{Y}_s ds,$$

что в дифференциальной нотации эквивалентно записывается как

$$dN_t = G(t)(Y_t - \hat{Y}_t) dt + D(t)dV_t.$$

Оказывается, что этот процесс не сильно отличается от Винеровского.

Лемма 13.4. *Процесс N_t*

- гауссовский;
- имеет ортогональные приращения;
- имеет те же пространства \mathcal{L}_t , что и процесс Y .

От Винеровского процесса его отличает только ковариационная функция и в частности дисперсия, потому что используя формулу Ито и взяв матожидание можно получить

$$\mathbb{E}[N_t^2] = \int_0^t D(s)^2 ds.$$

Если развить эту мысль, мы получим, что

$$dR_t = \frac{1}{D(t)} dN_t, R_0 = 0$$

будет настоящим Винеровским процессом с непрерывными траекториями. Более того, поскольку мы только поделили на детерминированную функцию пространства \mathcal{L}_t останутся теми же, что и у N_t . В свою очередь строить проекцию на основе процесса R уже оказывается существенно проще и можно даже выписать точную формулу.

Теорема 13.5. *Проекцию можно вычислить как*

$$\hat{Y}_t = \mathbb{E}[Y_t] + \int_0^t \partial_s \mathbb{E}[Y_t R_s](s) dR_s.$$

▷ По описанию пространства, куда мы проектируем, проекция на пространство наблюдений процесса R имеет вид

$$\hat{Y}_t = c(t) + \int_0^t f(s) dR_s$$

для какой-нибудь $f \in L^2[0, t]$. Поскольку R – Винеровский процесс, интеграл справа – это интеграл Ито и его матожидание равно 0, поэтому

$$\mathbb{E}[\hat{Y}_t] = c(t)$$

и при этом $\mathbb{E}[Y_t] = \mathbb{E}[\hat{Y}_t]$ по свойству нашей проекции. Кроме того из ортогональности проекции следует, что

$$(Y_t - \hat{Y}_t) \text{ и } \int_0^t p(s) dR_s$$

ортогональны для любой $p \in L^2[0, t]$. Воспользуемся:

$$\mathbb{E}\left[Y_t \int_0^t p(s) dR_s\right] = \mathbb{E}\left[\hat{Y}_t \int_0^t p(s) dR_s\right] = \mathbb{E}\left[\int_0^t f(s) dR_s \int_0^t p(s) dR_s\right],$$

откуда получается с помощью изометрии Ито для двойных интегралов (см. ДЗ-3)

$$\mathbb{E}\left[\int_0^t f(s) dR_s \int_0^t p(s) dR_s\right] = \int_0^t f(s) p(s) ds.$$

Поскольку это верно для любых $p \in L^2[0, t]$, выберем $p(t) = \mathbf{1}(t \in [0, r])$ с $r \leq t$. Тогда получится

$$\mathbb{E}[Y_t R_r] = \int_0^r f(s) ds,$$

что изначально мы искали.

□

13.4.4 Финальные штрихи

На семинаре мы как-то выводили точную формулу для решения уравнения для Y_t , используя его и все теоремы до этого мы, наконец, можем подойти к финальному результату, для которого нужно проделать ещё достаточно дополнительных вычислений (в том числе, получить уравнение Рикатти). Финальный результат выглядит так.

Теорема 13.6. (*Фильтр Калмана-Бюси, 1d*) Решение \hat{X}_t задачи фильтрации для уравнений

$$\begin{aligned} dY_t &= F(t)Y_t dt + C(t)dU_t, \quad (\text{модель системы}) \\ dX_t &= G(t)Y_t dt + D(t)dV_t, \quad (\text{модель наблюдений}), \end{aligned}$$

где F, C, G, D ограничены на ограниченных интервалах, задаётся уравнением

$$d\hat{Y}_t = F(t)\hat{Y}_t dt + \frac{G(t)S(t)}{D(t)^2} (dX_t - G(t)\hat{Y}_t dt),$$

где функция $S(t) = \mathbb{E}[(Y_t - \hat{Y}_t)^2]$ берётся из уравнения Рикатти

$$\frac{dS}{dt} = 2F(t)S(t) - \frac{G(t)^2}{D(t)^2} S^2(t) + C(t)^2, \quad S(0) = \mathbb{E}[(Y_0 - \mathbb{E}[Y_0])^2].$$

Величину

$$K(t) = \frac{G(t)S(t)}{D(t)^2}$$

часто называют *Kalman gain*, а множитель справа – невязкой предсказания $G(t)\hat{Y}_t dt$.

Наконец, для полноты сформулируем многомерную версию.

Теорема 13.7. (*Фильтр Калмана-Бюси, многомерная версия*) Решение \hat{X}_t задачи фильтрации для уравнений

$$\begin{aligned} dY_t &= F(t)Y_t dt + C(t)dU_t, \quad (\text{модель системы}) \\ dX_t &= G(t)Y_t dt + D(t)dV_t, \quad (\text{модель наблюдений}), \end{aligned}$$

где F, C, G, D ограничены на ограниченных интервалах, задаётся уравнением

$$d\hat{Y}_t = F(t)\hat{Y}_t dt + SG^T(DD^T)^{-1}(dX_t - G(t)\hat{Y}_t dt),$$

где матричная функция $S(t) = \mathbb{E}[(Y_t - \hat{Y}_t)(Y_t - \hat{Y}_t)^T]$ берётся из уравнения Рикатти

$$\frac{dS}{dt} = FS + SF^T - SG^T(DD^T)^{-1}GS + CC^T, \quad S(0) = \mathbb{E}[(Y_0 - \mathbb{E}[Y_0])(Y_0 - \mathbb{E}[Y_0])^T].$$

Фильтр Калмана по своей конструкции при известных параметрах системы способен решать все три обозначенные в начале задачи сразу.

1. **Задача фильтрации.** Решение уравнения сразу даёт функцию от наблюдения X_t .
2. **Задача сглаживания.** Сглаживание обновляет оценки фильтрации. В момент фильтрации мы строили проекцию только основываясь на \mathcal{F}_t , то есть на истории. Теперь мы можем пройти назад и использовать всё накопившееся знание, чтобы обновить более ранние оценки. Это делается с помощью специального сглаживающего уравнения, которое получил Закриссон[22]. Обозначим сглажённую оценку $\tilde{Y}_s = \mathbb{E}[Y_s | \mathcal{F}_t]$, для неё можно положить $\tilde{Y}_t = \hat{Y}_t$, а для $s < t$ получить уравнение

$$\frac{d}{ds}\tilde{Y}_s = F(s)\tilde{Y}_s + C(s)C^T(s)S^{-1}(s)(\tilde{Y}_s - \hat{Y}_s).$$

3. Задача предсказания Это делается достаточно просто, как раньше мы решали уравнения с помощью интегрирующих множителей, будто нет случайности:

$$\mathbb{E}[Y_{t+h} \mid \mathcal{F}_t] = e^{\int_t^T F(s)ds} \hat{Y}_t.$$

Задача фильтрации изначально была решена в постановке Стратоновича и в теории фильтрации, а также в оптимальном управлении можно встретить много отсылок к его интегралу.



Диффузия

На этой неделе мы займёмся более детальным изучением решений дифференциальных уравнений Ито. Связующим мостом между стохастическим и детерминированным миром являются диффузионные процессы, которые с точки зрения физики можно рассматривать через призму макромасштаба (через плотности, описательные характеристики) и микромасштаба (через описание стохастической динамики отдельных частиц).

14.1 Процесс диффузии

Диффузия – это чисто физическое название. Наверняка все мы помним что-то из курса школьной физики, где диффузией называется взаимное проникновение частиц разных веществ. Также диффузией называют рассеивание энергии, например, тепла. В 3д-рендеринге можно встретить название материала *diffuse*, поверхность из такого материала рассеивает поток света, очень слабо его отражая.

Всё это очень похоже с точки зрения математики модели, при этом они находятся в области молекулярно-кинетической теории, изучающую системы очень большого количества непредсказуемо двигающихся молекул или частиц.

Пример 14.1. Известное уравнение теплопроводности, называемое также уравнением диффузии, выглядит так:

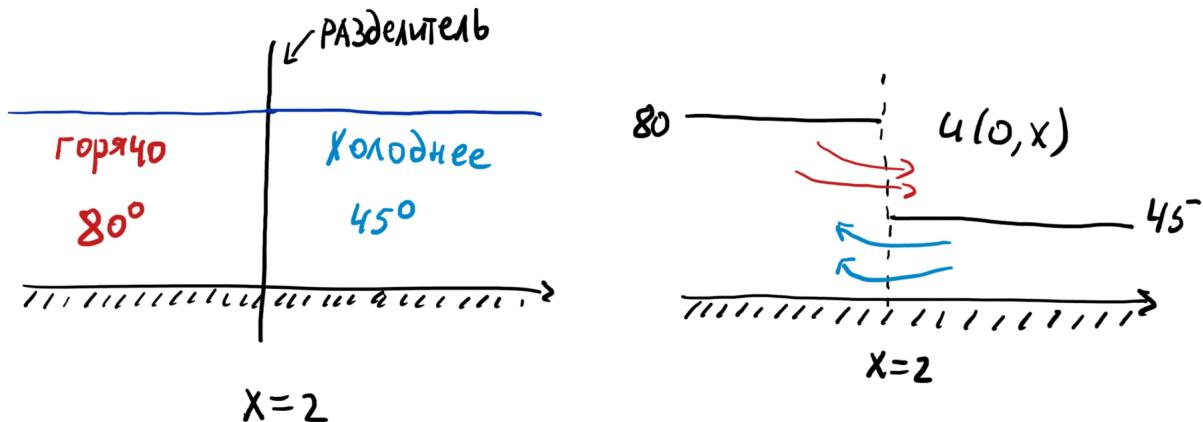
$$\partial_t u = \Delta u.$$

Здесь $u : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$, записываемое как $u(t, x)$ задаёт уровень тепла (температуру, если записать с правильными константами) в момент $t \in \mathbb{R}_+$ в точке пространства $x \in \mathbb{R}^d$.



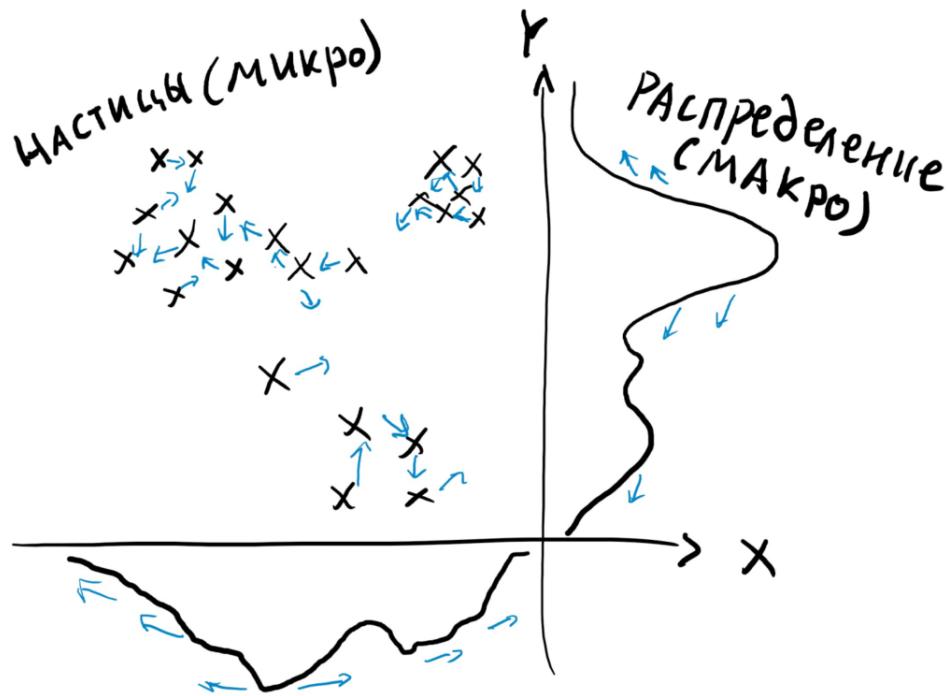
Такое уравнение выводится во многом интуитивными аргументами и с использованием предположения достаточной гладкости функции $u(t, x)$, но оказывается, что у него

есть и не везде гладкие решения, которые изначально такой подход не позволяет получить. Например, в среде может быть разделитель, который отделяет область с двумя разными температурами.

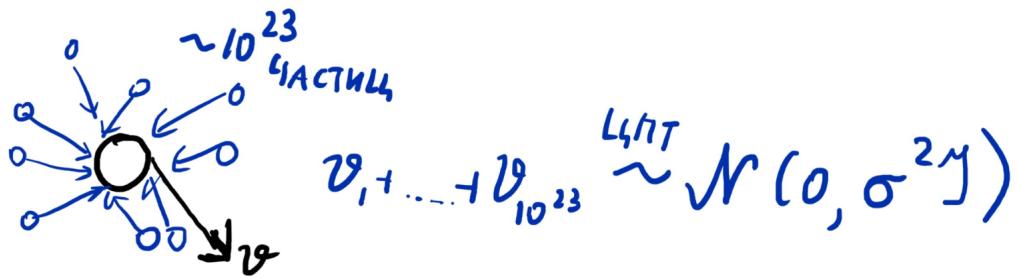


В силу физики после того, как разделитель убрали, температура должна уравниваться, плавно меняясь, потому что диффузия – это перетекание частиц вещества.

Явления, похожие на диффузию, всегда связаны с большим количеством частиц, поэтому уравнения в частных производных описывают в некотором смысле изменение их плотности, но не говорят, где будет каждая частица в какой момент.



Как можно описать одну конкретную частицу во время диффузии? Здесь на помощь приходит модель Броуновского движения, где есть одна инородная частица, всё происходит в стакане, и эта частица претерпевает соударения от соседних частиц жидкости. В силу того, что соударений очень много и они (можно предположить) независимы и маловероятно, что очень большие, то работает центральная предельная теорема: при условии, что частица была в точке $X_s = x_s$, её положение в момент t будет $X_t \sim N(x_s, \sigma^2(t-s))$.



Это обычная модель Броуновского движения, которую можно записать в виде стохастического дифференциального уравнения

$$dX_t = \sigma dW_t, \quad X_0 = x_0.$$

Это сильное упрощение, потому что в исследовании явлений диффузии есть ещё различные внешние факторы и неоднородность среды. Поэтому для нас моделью диффузии будет являться более общее дифференциальное уравнение в смысле Ито (можно рассматривать также уравнения Стратоновича)

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t,$$

где

$$b : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad \sigma : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$$

и W_t – m -мерный Винеровский процесс. Мы до самого конца (если не оговорено иного) будем исследовать однородные по времени диффузии, чтобы упростить выкладки. Неоднородный случай можно несложно обобщить, например, рассматривая время как отдельный процесс.

Определение 14.1. Решение стохастического дифференциального уравнения в смысле Ито

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0$$

при условии, что решение существует и слабо единствено (если есть \mathcal{L} , то конечно-мерные распределения одинаковы), назовём (однородной по времени) диффузией Ито.

Что может быть однородной по времени диффузией?

Пример 14.2. Уравнение для процесса Орнштейна-Уленбека

$$dX_t = -X_t dt + dW_t$$

является простым примером очень интересного класса диффузий. Пусть

$$p(x) = e^{-x^2/2},$$

тогда это уравнение можно записать как

$$dX_t = \nabla_x \ln p(X_t) dt + dW_t,$$

уравнения такого типа задают диффузию Ланжевена (*Langevin diffusion*), которая обладает очень большим количеством интересных свойств, мы уже видели её в действии в домашних заданиях.

Оказывается, что такой процесс действительно будет однородным по времени, достаточно только убрать зависимость от t из коэффициентов. Обозначим $X_t^{s,x}, t \geq s$ решение уравнения процесса диффузии с начальным условием $X_s = x$. Тогда можно заметить из интегральной формы записи, что

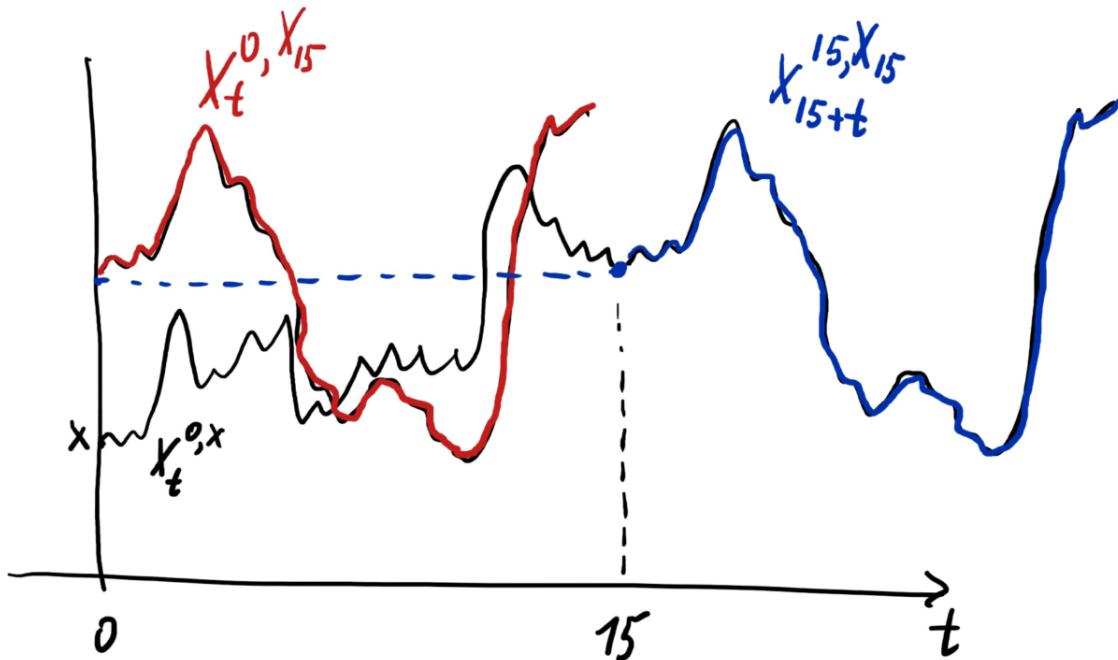
$$X_{s+h}^{s,x} = x + \int_s^{s+h} b(X_u)du + \int_s^{s+h} \sigma(X_u)dW_u$$

и, с другой стороны,

$$X_h^{0,x} = x + \int_0^h b(X_u)du + \int_0^h \sigma(X_u)dW_u.$$

Можно обратить внимание, что под интегралом используются только приращения Винеровского процесса, поэтому в первом выражении можно сделать замену на $\tilde{W}_t = W_t - W_s$ (тоже Винеровский процесс) и увидеть

$$X_{s+h}^{s,x} = x + \int_0^h b(X_u)du + \int_0^h \sigma(X_u)d\tilde{W}_u.$$



В силу слабой единственности решения дифференциального уравнения, решения

$$X_t^{s,x}, t \geq s$$

и

$$X_t^{0,x}, t \geq 0$$

имеют одинаковые конечномерные распределения. Иными словами, X – однородный по времени процесс.

14.2 Марковское свойство

Мы уже много раз почти приближались к этому вопросу, но теперь готовы обсудить его более формально.

Определение 14.2. Процесс (X_t) обладает Марковским свойством относительно фильтрации (\mathcal{F}_t) , если для $t, h > 0$ и любой измеримой ограниченной функции f

$$\mathbb{E}[f(X_{t+h}) \mid \mathcal{F}_t] = \mathbb{E}[f(X_{t+h}) \mid X_t].$$

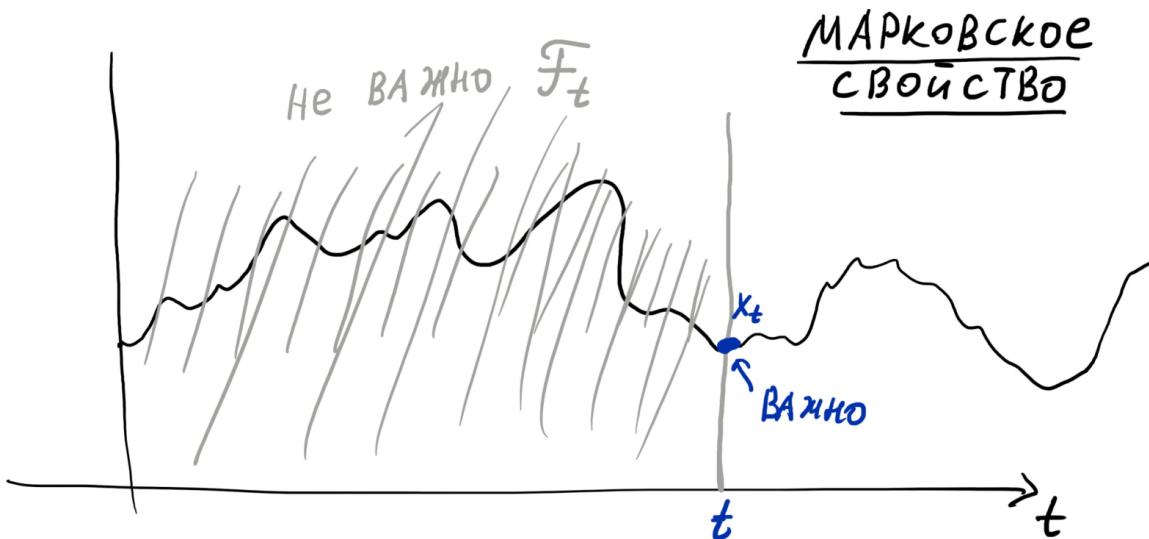
Это свойство можно увидеть также и в другом виде:

$$\mathbb{P}(X_t \in E_t \mid X_{t_1} \in E_{t_1}, \dots, X_{t_k} \in E_{t_k}) = \mathbb{P}(X_t \in E_t \mid X_{t_k} \in E_{t_k})$$

для всех $t > t_k > \dots > t_1$. Это мы уже видели ранее в Марковских цепях.

Пример 14.3. Винеровский процесс, как и любой процесс со стационарными и независимыми приращениями (в том числе, любой процесс Леви), обладает Марковским свойством относительно своей фильтрации. Здесь помогает независимость приращения $X_t - X_s$ относительно сигма алгебры \mathcal{F}_s .

Пример 14.4. Марковские цепи, которые мы рассматривали ранее, по своей конструкции удовлетворяют Марковскому свойству.



Учитывая однородность по времени, неудивительно, что диффузии тоже Марковские относительно правильной фильтрации.

Утверждение 14.1. Диффузия

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t$$

обладает Марковским свойством относительно фильтрации Винеровского процесса (\mathcal{F}_t) .

▷ Используем однородность по времени (так можно, потому что до времени t траектория при условии \mathcal{F}_t уже реализовалась и X_t измерим):

$$\mathbb{E} [f(X_{t+h}^{0,x}) \mid \mathcal{F}_t] = \mathbb{E} [f(X_h^{t,X_t}) \mid \mathcal{F}_t].$$

На самом деле значение под матожиданием зависит только от X_t , потому что интеграл Ито будет пределом интегральной суммы с приращениями Винеровского процесса, которые не зависят от \mathcal{F}_t . При этом X_t измерим относительно \mathcal{F}_t , поэтому полностью от условного матожидания избавиться нельзя. По этой причине

$$\mathbb{E} [f(X_{t+h}^{0,x}) \mid \mathcal{F}_t] = \mathbb{E} [f(X_{t+h}^{0,x}) \mid X_t].$$

□

Заметим, что X_t измерим относительно \mathcal{F}_t , поэтому $(\sigma(X_s), s \leq t)$, сигма-алгебра, порождённая всеми сечениями до момента t , будет вложена в фильтрацию Винеровского процесса. Поэтому X_t обладает Марковским свойством также относительно своей фильтрации.

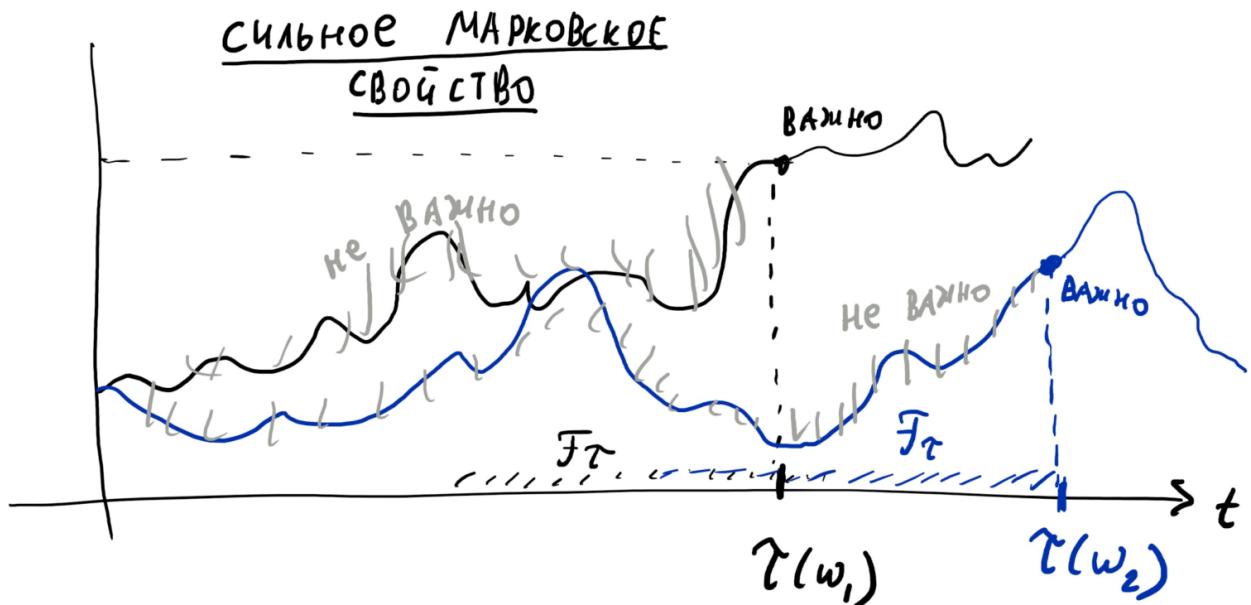
14.3 Сильное Марковское свойство

Само Марковское свойство открывает интересные горизонты: получается, что всё будущее процесса после момента t определяется только значением X_t . Поэтому это свойство часто ассоциируют с отсутствием памяти.

Один из интересных приёмов исследования процессов в финансовой математике и в семплировании состоит в том, чтобы исследовать процесс после наступления момента остановки. Скажем, если цена акции – это диффузия, и она достигла 123 рубля, то как понять, сколько ещё она продержится выше этого уровня? Если бы было что-то похожее на Марковское свойство, можно было бы сказать, что история не важна и нужно просто начать тот же процесс, но с другим начальным условием. Аналогом и обобщением Марковского свойства до уровня моментов остановки является *сильное Марковское свойство*.

Определение 14.3. Процесс X обладает сильным Марковским свойством относительно фильтрации (\mathcal{F}_t) , если для любого времени остановки τ (относительно этой же фильтрации и почти наверное $\tau < \infty$), любой ограниченной измеримой функции f и любого $h > 0$

$$\mathbb{E} [f(X_{\tau+h}) \mid \mathcal{F}_\tau] = \mathbb{E} [f(X_{\tau+h}) \mid X_\tau].$$



Многие примеры Марковских процессов транслируются и на этот случай, но не все.

Пример 14.5. Винеровский процесс обладает сильным Марковским свойством. Это оказывается очень непросто доказать и потребует от нас хорошего уровня владения остановленными сигма-алгебрами. Тем не менее, это вполне посильный факт, хотя и занимает достаточно места.

Пример 14.6. Любой процесс Леви обладает сильным Марковским свойством относительно своей фильтрации.

Пример 14.7. Можно построить процесс, который Марковский, но не сильно Марковский. К примеру, пусть $U \sim Ber(p)$. Если $U(\omega) = 1$, то процесс $X_t(\omega)$ – траектория Винеровского процесса, стартующего из 1, в целой точке t , а иначе тождественный 0. Это Марковский процесс относительно своей фильтрации. С другой стороны, если мы рассмотрим время остановки $\tau - k$ -й момент t , когда $X_t = 0$, – то случайная величина X_τ на самом деле константа 0 и порождает тривиальную сигма-алгебру, в то время как \mathcal{F}_τ включает в себя события $\{\omega : X_0(\omega) = 0\}$. Иначе говоря, значение $f(X_{\tau+h})$ будет зависеть от того, был ли 0 изначально, при этом значение 0 в момент достижения не даёт никакой информации о том, вся траектория ноль или нулевое значение только сейчас достигнуто.

Оказывается, что любая диффузия является сильно Марковской и это следствие того, что мы использовали Винеровский процесс, который тоже сильно Марковский.

Утверждение 14.2. Диффузия

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t$$

обладает сильно Марковским свойством.

Пример 14.8. Сильное Марковское свойство позволяет с помощью времён остановки и остановленных сигма-алгебр получать изначально не очень очевидные формулы для вероятности. Одним из таких примеров является закон отражения.

Рассмотрим W , Винеровский процесс и попробуем вычислить вероятность

$$\mathbb{P} \left(\sup_{s \leq t} W_s \geq a \right).$$

Добавим с помощью формулы полной вероятности в рассмотрение сечение W_t :

$$\mathbb{P} \left(\sup_{s \leq t} W_s \geq a \right) = \mathbb{P} \left(\sup_{s \leq t} W_s \geq a, W_t \geq a \right) + \mathbb{P} \left(\sup_{s \leq t} W_s \geq a, W_t < a \right).$$

Прежде всего, заметим, что

$$\mathbb{P} \left(\sup_{s \leq t} W_s \geq a, W_t \geq a \right) = \mathbb{P} (W_t \geq a),$$

потому что первое событие больше и содержит второе. Рассматриваем фильтрацию Винеровского процесса, задаваемую

$$\mathcal{F}_t = \sigma(W_s, s \leq t),$$

и время остановки

$$\tau = \inf\{s : W_s = a\} \wedge t.$$

Воспользуемся формулой полного матожидания, переходя к индикаторам:

$$\mathbb{P} \left(\sup_{s \leq t} W_s \geq a, W_t < a \right) = \mathbb{E} \left[\mathbb{E} \left[\mathbf{1} \left(\sup_{s \leq t} W_s \geq a \right) \mathbf{1} (W_t < a) \mid \mathcal{F}_\tau \right] \right].$$

Первый индикатор измерим относительно остановленной сигма-алгебры; он имеет два значения: 1 и 0, посмотрим на прообраз каждого.

1. Индикатор равен 1. В этом случае прообраз – это событие

$$\{\omega : \sup_{s \leq t'} W_s \geq a\} = \{\omega : \tau(\omega) \leq t'\} \cap \{\omega : \sup_{s \leq t'} W_s \geq a\} \in \mathcal{F}_{t'},$$

так что оно лежит в \mathcal{F}_τ .

2. Индикатор равен 0. Это происходит только тогда, когда $\tau = t$, а это событие из \mathcal{F}_τ .

Внутри остаётся ещё

$$\mathbb{E} [\mathbf{1}(W_t < a) \mid \mathcal{F}_\tau] = \mathbb{P} (W_t - W_\tau < 0 \mid \mathcal{F}_\tau) = \mathbb{P} (W_t - W_\tau < 0 \mid W_\tau) = 1/2,$$

где нам помогло сильное Марковское свойство. Таким образом, после подстановки и небольших манипуляций получаем

$$\mathbb{P} \left(\sup_{s \leq t} W_s \geq a \right) = 2\mathbb{P} (W_t \geq a).$$

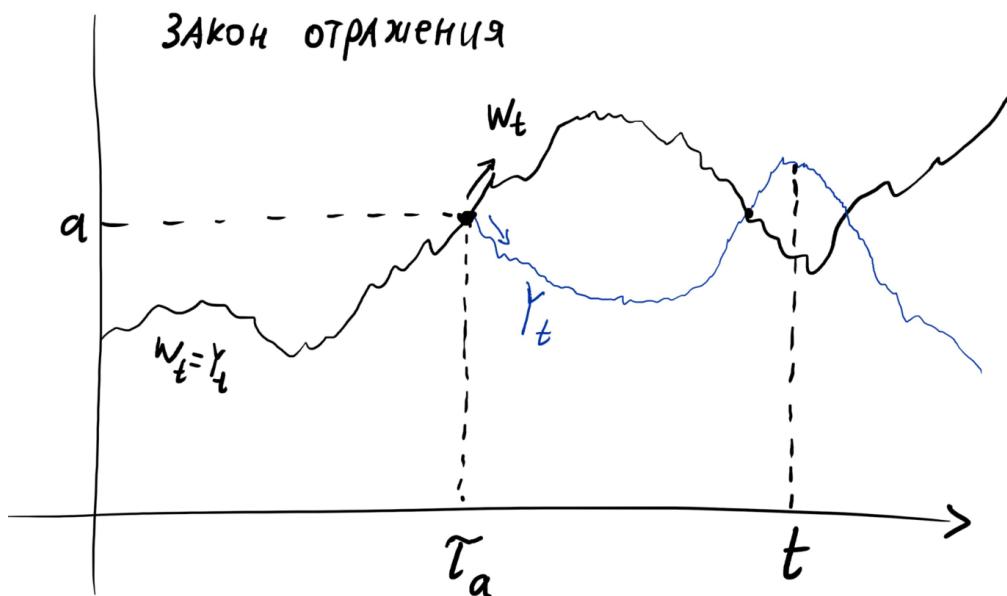
Почему это закон отражения? Есть ещё одна интерпретация. Если мы введём время остановки τ (любое строгое, то есть $\{\omega : \tau \leq t\}$ лежит в \mathcal{F}_t), то процесс

$$Y_t = \begin{cases} 2W_\tau - W_t, & t > \tau \\ W_t, & t \leq \tau \end{cases}$$

тоже будет Винеровским процессом с помощью сильного Марковского свойства. То есть, мы как бы отразили траекторию вертикально относительно времени остановки τ . После момента τ оригиналный W_t идёт дальше, а новый Y_t отражается. В этом контексте два изначальных слагаемых (при τ определённом как изначально) можно рассмотреть также как

$$\mathbb{P} \left(\sup_{s \leq t} W_s \geq a \right) = \mathbb{P} (\tau \leq t, W_t \geq a) + \mathbb{P} (\tau \leq t, W_t < a).$$

При этом после наступления момента остановки согласно сильному Марковскому свойству стартует новый Винеровский процесс $W_{\tau+h} - W_{\tau}$, который с вероятностью 1/2 выше или ниже 0.



14.4 Генератор диффузии

Центральный вопрос для финансовой математики – уметь посчитать или получить удобное представление для матожидания типа

$$\mathbb{E}[f(X_\tau)],$$

то есть, матожидания функции от остановленного процесса. Это важно в частности для того, чтобы получать формулы для цен опционов и для задач хеджирования. С другой стороны, есть физические задачи: как мы можем использовать информацию о вероятностной модели частицы для получения статистического описания явления диффузии?

Как и в случае Марковских процессов, мы приходим к понятию чего-то похожего на Q-матрицу (такой объект называется *переходным ядром* и занимает центральное место в теории Марковских цепей). Для нас эта теория вне горизонта, но мы её ещё немножко увидим в следующий раз. В контексте поставленной задачи нас будет интересовать немножко другой объект, но очень похожий.

Определение 14.4. Генератором однородной по времени диффузии X называется оператор A , действующий на функциях f как

$$A[f](x) = Af(x) = \lim_{t \rightarrow 0^+} \frac{\mathbb{E}[f(X_t) \mid X_0 = x] - f(x)}{t}.$$

Множество функций, для которых предел существует во всех x , обозначается \mathcal{D}_A и называется областью определения оператора A , а множество функций таких, что предел существует в точке x , обозначается $\mathcal{D}_A(x)$.

Иногда оператор A называют также *инфирнитезимальным генератором*, чтобы подчеркнуть предел из определения. Оператор A является по смыслу производной условного матожидания $f(X_t)$ по t , то есть, он может показать, как оно меняется во времени при развитии процесса. Конкретнее и более формально, как мы увидим, условное матожидание

$$u(t, x) = \mathbb{E}[f(X_t) \mid X_0 = x]$$

удовлетворяет обратному уравнению Колмогорова

$$\partial_t u = Au$$

с краевым условием типа

$$u(0, x) = f(x), \quad x \in \mathbb{R}^n$$

или

$$u(T, x) = f(x), \quad x \in \mathbb{R}^n.$$

В частности, уравнение Блэка-Шоулза – частный случай обратного уравнения Колмогорова для диффузии GBM. Мы более подробно вернёмся к этому вопросу в следующей главе.

Оказывается, что знание генератора позволяет нам получить очень полезное представление для матожидания функции от процесса.

Теорема 14.3. Пусть Y – это процесс Ито с представлением

$$dY_t = u(t, \omega)dt + v(t, \omega)dW_t, \quad Y_0 = x.$$

Тогда для $f \in C^2$, которая отображает $\mathbb{R}^d \rightarrow \mathbb{R}$, и времени остановки τ с конечным матожиданием верно

$$\mathbb{E}[f(Y_\tau) \mid Y_0 = x] = f(x) + \mathbb{E}\left[\int_0^\tau \nabla f(Y_s)u(s, \omega) + \frac{1}{2} \sum_{i,j} (vv^T)_{ij}(s, \omega)(\nabla^2 f(Y_s))_{ij} ds \mid Y_0 = x\right]$$

▷ Для начала заметим, что процесс $f(Y_t)$ – это тоже процесс Ито и по многомерной формуле Ито после некоторой арифметики получим

$$\begin{aligned} df(Y_t) &= \nabla f(Y_t)dY_t + \frac{1}{2}dY_t^T \nabla^2 f(Y_t)dY_t = \\ &= \nabla f(Y_s)u(s, \omega)dt + \frac{1}{2} \sum_{i,j} (vv^T)_{ij}(t, \omega)(\nabla^2 f(Y_t))_{ij}dt + \sum_{i,j} \partial_{x_i} f(Y_t)v(t, \omega)_{ij}dW_t^j. \end{aligned}$$

Если мы сейчас возьмём матожидание от интегрального представления, то получим нужную формулу, так как последняя сумма по свойству интеграла Ито даст 0. Но чтобы получить результат не для t , а для времени остановки τ , потребуется ещё немного усилий.

Зададим время остановки τ . Первые два слагаемых уже есть в формуле, нам надо доказать, что матожидание интеграла последнего равно 0. Здесь помогает приём аппроксимации времени остановки. Обозначим

$$\partial_{x_i} f(Y_s)v(s, \omega)_{ik} = g(Y_s, \omega)$$

и тогда

$$\mathbb{E} \left[\int_0^{\tau \wedge m} g(Y_s)dW_s^j \mid Y_0 = x \right] = \mathbb{E} \left[\int_0^m \mathbf{1}(s \leq \tau)g(Y_s, \omega)dW_s^j \mid Y_0 = x \right].$$

Подынтегральная функция из класса интегрируемых по Ито, поэтому по свойству интеграла Ито, матожидание равно нулю. С другой стороны, если мы сравним с интегралом до τ ,

$$\mathbb{E} \left[\left(\int_0^\tau g(Y_s, \omega)dW_s - \int_0^{\tau \wedge m} g(Y_s, \omega)dW_s \right)^2 \mid Y_0 = x \right] = \mathbb{E} \left[\left(\int_{\tau \wedge m}^\tau g(Y_s, \omega)dW_s \right)^2 \mid Y_0 = x \right].$$

Если вместо g рассмотреть $g\mathbf{1}(\tau \wedge m \leq s \leq \tau)$ и перейти к интегралу до ∞ (интеграл на самом деле всегда по конечному множеству, поэтому так можно), получим обычную изометрию Ито и

$$\mathbb{E} \left[\left(\int_{\tau \wedge m}^\tau g(Y_s, \omega)dW_s \right)^2 \mid Y_0 = x \right] = \mathbb{E} \left[\int_{\tau \wedge m}^\tau g(Y_s, \omega)^2 ds \mid Y_0 = x \right] \leq M^2 \mathbb{E} [\tau - \tau \wedge m \mid Y_0 = x] \rightarrow 0$$

при $m \rightarrow \infty$. Из сходимости в L^2 следует сходимость матожиданий, поэтому последняя сумма в случае времени остановки тоже занулится. \square

Из этого утверждения в одно действие можно получить выражение для генератора диффузии.

Теорема 14.4. *Генератор диффузии*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t$$

задаётся как

$$Af(x) = \nabla f(x)b(x) + \frac{1}{2} \sum_{i,j} (\sigma\sigma^T)_{ij}(x)(\nabla^2 f)_{ij}(x)$$

и если $f \in C^2$, то $f \in \mathcal{D}_A$.

Наконец, первый мостик в статистическую физику.

Пример 14.9. Броуновское движение – это диффузия

$$dX_t = dW_t,$$

а её генератор – это почти оператор Лапласа

$$A = \Delta = \frac{1}{2} \sum_{i=1}^d \partial_{x_i}^2.$$

По этой причине уравнение теплопроводности (оно же уравнение диффузии) связаны с Броуновским движением. Так мы ввели вероятностную модель частицы на микроравнении и поняли, что её также можно описать в макромасштабе с помощью уравнения в частных производных. Обратное уравнение Колмогорова говорит, какое будет ожидаемое значение $f(X_t)$ при условии, что частица стартовала с $X_0 = x_0$. Но есть нюанс...

Пример 14.10. Уравнение переноса

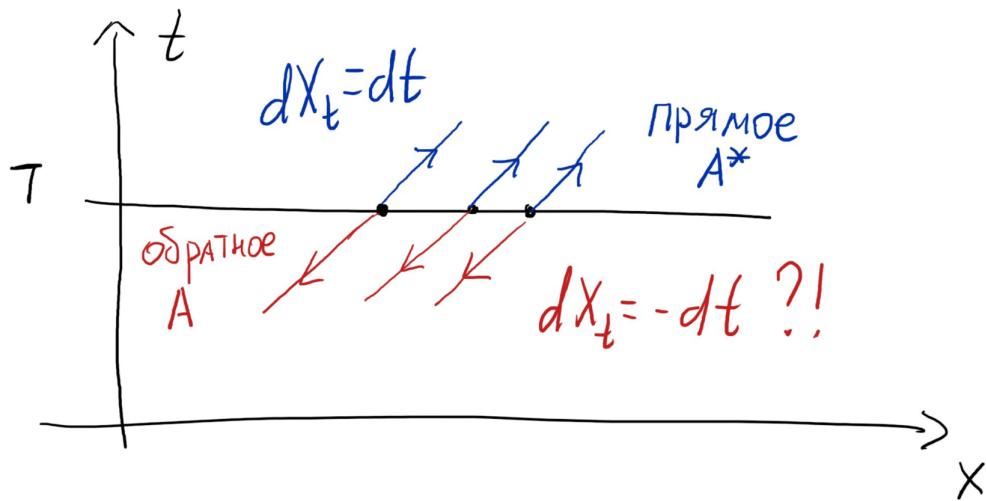
$$\partial_t u + \partial_x u = 0$$

с начальным условием $u(0, x) = u_0(x)$ имеет решение $u(t, x) = u_0(x - t)$. Если рассмотреть это уравнение как обратное уравнение Колмогорова, то окажется, что такому генератору соответствует диффузия

$$dX_t = -dt.$$

Кажется странным, почему частицы движутся влево, если решение $u(t, x)$ наоборот сдвигается вправо. Обратное уравнение Колмогорова оценивает матожидание функции в координате частицы в момент t . Уравнение переноса говорит, что в момент t функция должна сдвинуться правее. То есть, в момент t в стартовой точке x должно быть то же значение, которое было в той точке, откуда в x за время t пришла другая частица. На формальном уровне это означает, что построенная диффузия на самом деле как бы идёт назад во времени. Поэтому это уравнение называется обратным.

Есть ещё прямое уравнение Колмогорова, которое в физике называют уравнением Фоккера-Планка, которое записывается с оператором A^* . Оператор Лапласа самосопряжённый, поэтому $A = A^*$ и Винеровский процесс действительно описывает продвижение частиц вперёд во времени, но в этом случае сопряжённым для $A = -\partial_x$ будет оператор $A^* = \partial_x$. Об этом на следующей неделе.



Пока же подытожим наш основной результат, который носит название формулы Дынкина, на основе сбора наших последних результатов.

Теорема 14.5. (Формула Дынкина) Для диффузии

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

времени остановки τ (относительно фильтрации Винеровского процесса или процесса X , с конечным матожиданием) и $f \in C^2$ верно

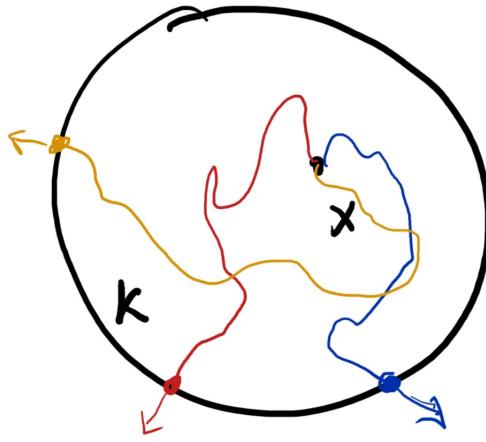
$$\mathbb{E}[f(X_\tau) \mid X_0 = x] = f(x) + \mathbb{E}\left[\int_0^\tau Af(X_s)ds \mid X_0 = x\right].$$

Пример 14.11. Чем ещё полезно знание генератора и формулы Дынкина? Например, можно исследовать разные глобальные свойства процессов с помощью времён остановки. Одним из интересных примеров является Винеровский процесс. Для $d = 1$ вообще сработает похожий аргумент, но там можно сильно проще, поэтому рассмотрим Винеровский процесс W в d -мерном пространстве, при этом $d \geq 2$. Какое среднее время выхода из некоторой окрестности нуля, например, множества

$$K = \{x \in \mathbb{R}^d : \|x\| \leq R\}?$$

Норма – это евклидова 2-норма, мы опускаем нижний индекс 2 для краткости. Это время случайное и более того, является моментом остановки

$$\tau = \inf \{t : W_t \notin K\} = \inf \{t : \|W_t\| = R\}.$$



Последнее верно в силу непрерывности траекторий и того, что мы рассматриваем инфимум, то есть, точную нижнюю грань. Генератор Винеровского процесса – это оператор Лапласа $A = \frac{1}{2}\Delta$. Присмотримся к формуле Дынкина и положим $x \in K$:

$$\mathbb{E}[f(W_\tau) \mid W_0 = x] = f(x) + \mathbb{E} \left[\int_0^\tau Af(W_s)ds \mid W_0 = x \right].$$

Если бы мы подобрали такую дважды дифференцируемую функцию, которая (1) во время остановки была бы константой или очень простой случайной величиной и (2) $Af = \text{const}$, то мы получим выражение для матожидания времени остановки. Такой функцией является, например,

$$f(x) = \|x\|^2.$$

Применив формулу Дынкина, увидим

$$R^2 = \|x\|^2 + d\mathbb{E}[\tau \mid W_0 = x],$$

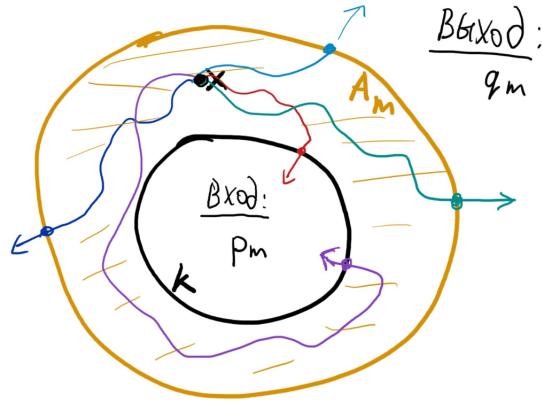
то есть,

$$\mathbb{E}[\tau \mid W_0 = x] = \frac{R^2 - \|x\|^2}{d}.$$

(На самом деле, тут нужно проверить конечность матожидания времени остановки, в это можно поверить (с), но это можно сделать трюком с взятием $\tau_k = k \wedge \tau$, которое точно обладает таким свойством и устремить k к бесконечности). Отсюда видно ещё один момент: это время уменьшается с ростом размерности, то есть в больших размерностях процесс быстрее покидает любую окрестность нуля.

Оказывается, мы можем понять, вернётся ли процесс обратно в K , если покинет её. Здесь тоже помогает формула Дынкина. давайте предположим, что процесс startует вне K , но при этом всё ещё в ограниченном множестве, например, в

$$A_m = \{x \in \mathbb{R}^d : R \leq \|x\| \leq 2^m R\}.$$



В силу динамики Винеровского процесса, он точно за конечное время выйдет из A_m , вопрос только в том, где. Есть два варианта:

1. Через внутреннюю границу, обратно в K ;
2. Через внешнюю границу, вовне.

Мы обозначим вероятность сценария (1) за p_m , а вероятность сценария (2) за q_m , при этом $p_m + q_m = 1$, так как других вариантов выхода нет. Положим $W_0 = x \in A_m$ и определим момент остановки

$$\tau_m = \tau = \inf \{t : W_t \notin A_m\}.$$

Здесь помогает найти какую-нибудь функцию f , которая является гармонической, то есть $\Delta f = 0$, и при этом связана с нормой. Хорошим выбором будет

$$f(x) = \begin{cases} \ln \|x\|, & d = 2, \\ \|x\|^{2-d}, & d > 2. \end{cases}$$

Такой выбор атмосферой немного навеян из физики и уравнений в частных производных, конкретнее, из уравнений диффузии, уравнений Лапласа и Гельмгольца. Что это даёт?

Случай $d = 2$. В этом случае формула Дынкина показывает

$$p_m \ln R + q_m \ln(R2^m) = \ln \|x\|.$$

Справа конечное число, p_m, q_m находятся между 0 и 1, при этом с ростом t слагаемое с q_m растёт до бесконечности. Поэтому неизбежно должно быть $q_m \rightarrow 0$ при $t \rightarrow \infty$. Таким образом, Винеровский процесс с вероятностью 1 вернётся в любую окрестность старта, то есть в $d = 2$ он обладает свойством возвращенности.

Случай $d > 2$. Здесь такой аргумент не работает.

$$p_m R^{2-d} + q_m (R2^m)^{2-d} = \|x\|^{2-d}.$$

Если присмотреться, то

$$p_m = \frac{\|x\|^{2-d} - q_m(R2^m)^{2-d}}{R^{2-d}},$$

при этом $(R2^m)^{2-d} \rightarrow 0$ при $m \rightarrow \infty$ и q_m ограничена. Значит,

$$p_m \rightarrow \frac{\|x\|^{2-d}}{R^{2-d}},$$

это число < 1 . Следовательно, с ненулевой вероятностью Винеровский процесс может и не вернуться в окрестность старта.

На следующей неделе мы больше уйдём в сторону связи с уравнениями в частных производных, которые нам откроют способы изучения распределений сечений диффузии. Теория вероятности неожиданно успешно описывает физические феномены.

Уравнения Колмогорова

В этой лекции мы посмотрим детальнее, как физика связана с диффузиями и уравнениями Колмогорова. Идеи из марковских цепей в силу марковских свойств диффузии пригодятся и здесь.

15.1 Обратное уравнение Колмогорова

Формально, до сих пор мы не выводили обратного уравнения Колмогорова, пользуясь вместо этого только некоторой интуицией о том, как связан макро- и микро-масштаб процесса диффузии. Кроме того, попытка сопоставить уравнение переноса и диффузию обернулась неожиданным результатом: на макроуровне плотность двигалась в одну сторону, а на микроуровне частицы двигались в другую, будто двигаясь обратно во времени.

Время привести физику в порядок.

Утверждение 15.1. *Пусть дана диффузия X_t с генератором A . Тогда для $f \in \mathcal{D}_A$ условное матожидание*

$$u(t, x) = \mathbb{E}[f(X_t) \mid X_0 = x]$$

удовлетворяет обратному уравнению Колмогорова

$$\partial_t u = Au$$

с краевым условием $u(0, x) = f(x)$. Более того, если $w(t, x)$ – некоторая $C^{1,2}$ функция, которая удовлетворяет уравнению с этим краевым условием, то она совпадает с

$$u(t, x) = \mathbb{E}[f(X_t) \mid X_0 = x].$$

▷ Мы выведем только само уравнение, но ещё немного технической работы с формулой Дынкина с подходящим временем остановки и можно получить утверждение о единственности.

Обозначим

$$u(t, x) = \mathbb{E}[f(X_t) \mid X_0 = x].$$

По формуле Дынкина

$$\mathbb{E}[f(X_{t+h}) \mid X_0 = x] = f(x) + \mathbb{E}\left[\int_0^{t+h} Af(X_s)ds \mid X_0 = x\right].$$

Поэтому $u(t, x)$ дифференцируема по t и

$$\partial_t u(t, x) = \mathbb{E}[Af(X_t) \mid X_0 = x]$$

Но это не всё, оператор A до сих пор внутри. С одной стороны, так как мы знаем, что u дифференцируема по времени, при $h \rightarrow 0$ получим, что

$$\frac{\mathbb{E}[f(X_{t+h}) \mid X_0 = x] - \mathbb{E}[f(X_t) \mid X_0 = x]}{h} = \frac{u(t+h, x) - u(t, x)}{h} \rightarrow \partial_t u(t, x).$$

С другой стороны, по определению генератора, если мы его применим по x игнорируя t , то получим

$$\begin{aligned} Au &= \lim_{h \rightarrow 0+} \frac{\mathbb{E}[u(t, X_h) \mid X_0 = x] - u(t, x)}{h} = \lim_{h \rightarrow 0+} \frac{\mathbb{E}[\mathbb{E}[f(X_{t+h}) \mid X_h] \mid X_0 = x] - u(t, x)}{h} = \\ &= \lim_{h \rightarrow 0+} \frac{\mathbb{E}[f(X_{t+h}) \mid X_0 = x] - \mathbb{E}[f(X_t) \mid X_0 = x]}{h}. \end{aligned}$$

Это совпадает с тем, что выше; таким образом, как мы и хотели,

$$\partial_t u = Au, \quad u(0, x) = f(x).$$

□

Можно заметить, что из-за однородности по времени если взять от функции

$$q(t, x) = \mathbb{E}[f(X_T) \mid X_t = x] = \mathbb{E}[f(X_{T-t}) \mid X_0 = x] = u(T-t, x)$$

производную по времени, по получится

$$\partial_t q = -\partial_t u(T-t, x) = -Au(T-t, x)$$

или по-другому опять уравнение Колмогорова

$$\partial_t u(T-t, x) = Au(T-t, x),$$

но с другим краевым условием $u(T, x) = q(0, x) = f(x)$. То есть, его можно рассматривать обратно во времени, отсюда название *обратного уравнения*. Но в матожиданиях это не так заметно, чтобы понять реальный мотив, нам нужно задуматься о том, как меняется распределение X_t .

15.2 Новые переходные вероятности

Мы обошли вопрос о том, как меняются сами распределения X_t . Чтобы это лучше понять, нам пригодится знание об условных распределениях. Поскольку мы сейчас находимся в случае общих Марковских процессов, а не Марковских цепей, техника усложняется, но аналогия сохраняется.

Определение 15.1. Пусть (Ω, \mathcal{F}) и (Ξ, \mathcal{G}) – измеримые пространства. Переходным Марковским ядром называется отображение

$$P : \Xi \times \mathcal{F} \rightarrow [0, 1],$$

которое

1. является вероятностной мерой при фиксированном $x \in \Xi$, то есть, когда считаем $P(x, \cdot)$;
2. является измеримой функцией (случайной величиной) при фиксированном $A \in \mathcal{F}$, то есть, когда считаем $P(\cdot, A)$.

Мы уже встречали случайные меры в процессах Леви, это за исключением мелких технических моментов практически то же самое. Переходное ядро в самом общем виде описывает для цепи Маркова вероятности получить в следующем моменте времени значения процесса $X_{t+1} \in A$ при условии, что $X_0 = x$. По этой причине и записаны 2 пункта выше. Поскольку, в непрерывном времени не очень понятно, что значит в *следующий момент*, одного ядра не всегда достаточно.

Пример 15.1. Помните однородные по времени цепи Маркова с конечным числом состояний? Их можно задать, задав одно переходное ядро с помощью переходной матрицы как

$$P(x, X_{t+1}^{-1}(\{y\})) = \mathbb{P}(X_{t+1} = y \mid X_t = x).$$

Интеграл по такому ядру – это интеграл по дискретной мере, то есть, просто сумма. Все формулы для вероятностей в цепи Маркова, которые мы видели до этого остаются верны.

Пример 15.2. Пуассоновский процесс X_t с интенсивностью λ – это общая цепь Маркова, в которой состояния дискретны, но время уже непрерывно. Мы можем вспомнить, что количество прыжков в среднем растёт с ростом временного горизонта, поэтому одного ядра нам будет недостаточно – нужно свою для каждого интервала $h > 0$.

$$\mathbb{P}(X_{t+h} = y \mid X_t = x) = P_h(x, X_{t+h}^{-1}(\{y\})) = \begin{cases} \frac{(\lambda h)^{y-x}}{(y-x)!} e^{-\lambda h}, & y \geq x, \\ 0, & y < x. \end{cases}$$

Как мера, ядро формально записывается так:

$$P_h(x, A) = \left(\sum_{y \geq x} \frac{(\lambda h)^{y-x}}{(y-x)!} e^{-\lambda h} \delta_{y-x} \right) (A),$$

где δ_p понимается как дельта-мера, то есть, определяемая как

$$\delta_p(A) = \begin{cases} 1, & p \in A, \\ 0, & p \notin A. \end{cases}$$

Пример 15.3. Для однородных диффузий получается сложнее, но с той же мыслью. Для Винеровского процесса, например,

$$\mathbb{P}(W_{t+h} \in A \mid W_t = x) = P_h(x, W_{t+h}^{-1}(A)) = \int_A \frac{1}{\sqrt{2\pi h}} e^{-(y-x)^2/(2h)} dy.$$

В этом случае в литературе часто можно встретить более краткую запись сразу для ядра

$$P_h(x, dy) = \frac{1}{\sqrt{2\pi t}} e^{-(y-x)^2/(2t)} dy$$

и

$$\int_A P_h(x, dy),$$

чтобы подчеркнуть, что интегрирование происходит по y .

Для однородных диффузий в силу Марковского свойства тоже определены вероятности

$$\mathbb{P}(X_t \in A \mid X_0 = x),$$

которые можно также записать с помощью переходного ядра – оно всегда существует, так как эта условная вероятность по своей сути мера по A . И как условная вероятность, она также является измеримой функцией условия, то есть, x . Так, мы можем также записать в другом виде:

$$\mathbb{P}(X_{t+h} \in A \mid X_t = x) = \int_A dP_h(y, x)$$

или более кратко, как часто пишут,

$$\mathbb{P}(X_{t+h} \in A \mid X_t = x) = \int_A P_h(dy, x).$$

где для промежутка времени h мы определили ядро P_h .

Для случая $h = 0$ говорят, что $P_0(x, \cdot)$ – это дельта-мера δ_x .

Почему же уравнение Колмогорова, которое мы раньше записывали называется обратным? Рассмотрим случай, когда ядро задаётся переходной плотностью, которая дважды непрерывно дифференцируема по условию x и один раз дифференцируема по времени s :

$$P_h(x, dy) = p(y, t|x, s)dy, \quad t > s.$$

Далее если записывается интеграл без пределов, подразумевается, что он берётся по всему пространству.

Утверждение 15.2. Переходная плотность $p(y, T|x, T-s) = u(T-s, x)$ удовлетворяет по переменным x, s уравнению

$$\partial_s u = A_x u$$

с генератором A , применённым по x , и краевым условием $u(T, y) = \delta_x$

▷ Если записать матожидание как интеграл по плотности, то получим для $f \in C^2$, что из-за однородности по времени

$$\mathbb{E}[f(X_T) \mid X_{T-s} = x] = \mathbb{E}[f(X_s) \mid X_0 = x] = \int f(y)p(y, s|x, 0)dy.$$

Обратное уравнения Колмогорова для

$$\partial_s \mathbb{E}[f(X_s) \mid X_0 = x] = \int f(y)\partial_s p(y, s|x, 0)dy,$$

матожидания посередине, говорит, что

$$\partial_s \mathbb{E}[f(X_s) \mid X_0 = x] = A_x \mathbb{E}[f(X_s) \mid X_0 = x] = \int f(y)A_x p(y, s|x, 0)dy.$$

Мы пользуемся наличием плотности и требованиями гладкости, чтобы занести дифференцирования под интеграл. Таким образом,

$$\int_A f(y) (\partial_s p(y, s|x, 0) - A_x p(y, s|x, 0)) dy = 0.$$

Поскольку мы взяли произвольную f ,

$$\partial_s p(y, s|x, 0) - A_x p(y, s|x, 0) = 0$$

Или в другой форме с использованием однородности по времени

$$\partial_s p(y, T|x, T-s) - A_x p(y, T|x, T-s) = 0.$$

При $s = T$ по нашему определению переходного ядра

$$P_0(x, \cdot) = \delta_x,$$

поэтому краевое условие гарантировано. \square

Видно, что на самом деле, в силу граничного условия обратное уравнение Колмогорова нужно понимать как по-настоящему *обратное во времени*, поэтому в уравнении переноса ранее мы получили, что частицы летят в обратном направлении.

15.3 Прямое уравнение Колмогорова

Теперь мы знаем, как меняется плотность обратно во времени, но оказывается, мы можем получить и прямое уравнение. В физике его принято называть уравнением Фоккера-Планка, а в математике прямым уравнением Колмогорова или уравнением Фоккера-Планка-Колмогорова.

Теорема 15.3. (*Уравнение Фоккера-Планка*) Для плотности $p(y, t)$ сечения X_t , дважды дифференцируемой по аргументу y и дифференцируемой по t верно уравнение

$$\partial_t p = A_y^* p$$

с генератором, сопряжённым A и применённым к y , а также с краевым условием $p(y, 0) = p_0(y)$.

▷

Рассмотрим при $T > t$ матожидание

$$u(t, x) = \mathbb{E}[f(X_t) \mid X_0 = x]$$

и матожидание

$$q(t, x) = \mathbb{E}[f(X_T) \mid X_t = x] = \int f(y)p(y, T|x, t)dy.$$

Заметим, что

$$\mathbb{E}[q(t, X_t)] = \int q(t, y)p(y, t)dy = \mathbb{E}[f(X_T)]$$

не зависит от t , то есть, его производная по t равна нулю. Пользуясь тем, что можно заносить производную под знак интеграла, получаем, что

$$0 = \int \partial_t(qp)dy = \int (\partial_tqp + q\partial_tp)dy = \int (-A_y[q]p + q\partial_tp)dy,$$

где в конце мы вспомнили про обратное уравнение Колмогорова и

$$\partial_tq(t, y) = \partial_t\mathbb{E}[f(X_T) \mid X_t = y] = \partial_t\mathbb{E}[f(X_{T-t}) \mid X_0 = y] = -A_xu(T-t, y) = -A_yq(t, y).$$

Сопряжение получится, когда мы сделаем интегрирование по частям, предположив, что $u(q, x)$ по x убывает к нулю на бесконечности, это зависит от функции f . Конкретнее,

$$\int A_y[u]p(t, y)dx = \int uA_x^*[p]dy,$$

откуда в силу произвольности f (и соответственно u) получаем

$$\partial_tp - A_y^*p = 0.$$

□

Из этого уравнения также следует прямое уравнение Колмогорова, которое немного отличается краевым условием.

Следствие 15.4. Для переходной плотности $p(y, t|x, 0)$, дважды дифференцируемой по аргументу y и дифференцируемой по t верно уравнение

$$\partial_tp = A_y^*p$$

с генератором, сопряжённым A и применённым к y , а также с краевым условием $p(y, 0|x, 0) = \delta_x$.

На самом деле, уравнения верны не только для таких гладких функций, но и даже для некоторых обобщённых функций, а значит, для более общих вероятностных распределений. Поэтому условие про C^2 не является ограничивающим плюс отдельная большая теория уравнений частных производных и уравнений Фоккера-Планка в частности позволяют судить о существовании и единственности решений в конкретных случаях [24].

Пример 15.4. Уравнение диффузии

$$\partial_t u = \frac{1}{2} \partial_{xx}^2 u, \quad u : \mathbb{R}_{\geq 0} \times \mathbb{R}$$

является и прямым, и обратным уравнением Колмогорова для Винеровского процесса W_t , так как генератор

$$A = \frac{1}{2} \partial_{xx}^2$$

самосопряжённый.

Пример 15.5. Уравнение переноса

$$\partial_t u + \partial_x u = 0, \quad u : \mathbb{R}_{\geq 0} \times \mathbb{R}$$

является прямым для диффузии

$$dX_t = -dt$$

и обратным для

$$dY_t = dt,$$

так как для $A = \partial_x$ сопряжённым оператором будет $A^* = -\partial_x$.

Есть формула для сопряжённого оператора генератора, которую можно получить, сделав несколько раз интегрирование по частям.

Утверждение 15.5. Для диффузии

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t$$

генератор определяется как

$$Af = \nabla f b + \frac{1}{2} \sum_{i,j} \sigma_{ij} (\nabla^2 f)_{ij},$$

а сопряжённый ему как

$$A^*f = -\operatorname{div}(b(x)f(x)) + \frac{1}{2} \sum_{i,j} \partial_{x_i x_j}^2 ((\sigma \sigma^T)_{ij} f).$$

15.4 Стационарные распределения и эргодичность

Уравнение Фоккера-Планка задаёт в каком-то смысле поток материи или вероятности. Оно показывает, куда и как должна смещаться плотность. С помощью уравнения Фоккера-Планка можно пытаться искать стационарные распределения Марковских процессов, а также доказывать сходимость к ним.

Определение 15.2. Распределение (плотность) μ называется стационарной при заданных граничных условиях, если

$$\int_A P_t(x, A) d\mu(x) = \mu(A).$$

Уравнение Фоккера-Планка позволяет посмотреть на стационарность с другой стороны: стационарная плотность не должна меняться со временем, значит, удовлетворяет

$$A^* \mu = 0,$$

причём на месте μ необязательно может быть именно плотность, возможно искать более общие решения.

Пример 15.6. В динамике Ланжевена

$$dX_t = (\nabla \ln p)(X_t)dt + dW_t$$

стационарная плотность определяется с помощью уравнения Фоккера-Планка. Для этого нужно записать условие стационарности

$$A_x^* v(t, x) = 0.$$

Зная формулу для сопряжённого

$$A^* f = -\operatorname{div}((\nabla \ln p)f(x)) + \frac{\sigma^2}{2} \Delta f,$$

можно получить с точностью до констант C

$$C - \sum_{i=1}^d \partial_{x_i} \ln p v(t, x) + \frac{\sigma^2}{2} \sum_{i=1}^d \partial_{x_i} v(t, x) = 0.$$

Тут можно заметить закономерность, если посмотреть на отдельные слагаемые:

$$\partial_{x_i} v(t, x) = v(t, x) \partial_{x_i} \ln p.$$

Это наводит на мысль, что решение нужно искать в форме

$$v(t, x) = C_0 e^{(2/\sigma^2) \ln p} + \sum_{i=1}^d c_i x_i,$$

но константы c_i обнуляются (иначе не получится условие нормировки), а константа C_0 – это нормировочная константа. Если дополнительного сделать $\sigma = \sqrt{2}$, то получится вообще

$$v(t, x) = C_0 p(x).$$

Более того, можно показать, что динамика Ланжевена действительно сходится (в разных метриках) к этому инвариантному распределению при соблюдении некоторых технических условий, среди которых, например, лог-вогнутость функции плотности или похоже условия. В дискретном времени (при использовании метода Эйлера) сходимость медленнее. На этой идеи основаны методы семплирования *ULA*, *MALA* и их различные модификации из области *MCMC*.

Диффузионные модели

придумать, как вставить PDF

Продвинутое семплирование

На этой неделе мы попробуем посмотреть на цепи Маркова и диффузию Ланжевена немного в другом контексте. Мы посмотрим, как можно использовать известное нам для семплирования значений случайных величин.

17.1 Зачем нужно семплирование?

Семплирование – это симуляция значений случайных величин. Прежде всего, мы знаем огромное количество алгоритмов, где в каком-то месте нужно накинуть случайность: от симуляции моделей до обучения и инференса диффузионных моделей типа StableDiffusion. С другой стороны, нам часто интересно получать Монте-Карло семплы из распределений для того, чтобы оценивать матожидания и соответственно интегралы в высокой размерности, где квадратурные формулы не работают.

Оказывается, в каких-то простых и даже не очень простых случаях мы уже умеем генерировать семплы.

Пример 17.1. Равномерное распределение на отрезке (или на конечном множестве). На самом деле очень неочевидный и непростой концептуально метод с точки зрения того, чтобы поверить, что работает.

Есть алгоритмы (они называются линейные конгруэнтные генераторы, LCG), которые упрощённо основаны на следующей идее. Рассмотрим единичную окружность: отрезок $[0, 2\pi)$, на котором по любому вещественному значению из \mathbb{R} берётся остаток $x \text{ mod } 2\pi$. Если взять произвольную стартовую точку x_0 и добавлять сдвиги

$$x_i = x_{i-1} + \Delta,$$

то получится детерминированная последовательность точек. Оказывается, что для любого рационального Δ множество различных x_i в последовательности конечно. Но если задать Δ иррациональным, то это множество будет бесконечным и плотным (в любом открытом множестве найдётся как минимум одна точка из траектории). Такая система будет эргодической [5]: вне зависимости от того, откуда мы начали, после некоторого момента семплы будут выглядеть очень похоже на то, как будто они пришли из распределения $U[0, 2\pi]$. Более того, корреляция между типо-семплами будет убывать со временем, то есть, можно в том числе получать и наборы близкие к независимым.

Компьютер пока не умеет численно задавать иррациональные числа, но можно задавать достаточно сложные рациональные типа

$$\Delta = 10^8 \frac{27474919237474883664711}{237474800984748447474787348748}$$

так, что в результате различных точек будет очень много и их обход будет очень сложным и непредсказуемым после взятия остатка. Для численного моделирования этого достаточно. При этом эта система тоже эргодическая: при больших i значения x_i будто приходят из дискретного равномерного распределения. Стартовое значение x_0 называется обычно *seed* (наверняка вы его видели в питоне) и математически это прямой аналог элементарного исхода $\omega \in \Omega$.

Пример 17.2. (*Обратная функция распределения*) В одномерном случае очень часто есть универсальный рецепт семплирования из произвольной плотности $f(x)$. Для этого задают

$$F^{-1}(t) = \inf \{x : F(x) \geq t\}.$$

Такое определение подходит даже в случае, когда строго обратной функции нет. Тогда если взять $U \sim U[0, 1]$, то $F^{-1}(U) \sim f$. Таким образом можно семплировать, например, экспоненциальные и гамма-величины.

Пример 17.3. (*Гауссовское распределение*) Равномерное распределение также открывает дорогу для семплирования гауссовского. Оказывается (после полярной замены и некоторых манипуляций), что если $U \sim [0, 1]$, то

$$X = \sqrt{-2 \ln U} \sin(2\pi U), \quad Y = \sqrt{-2 \ln U} \cos(2\pi U)$$

являются двумя независимыми гауссовскими величинами $\mathcal{N}(0, 1)$. Отсюда можно просто получить $Z = \mu + \sigma X$, который будет иметь распределение $\mathcal{N}(\mu, \sigma^2)$.

Пример 17.4. (*Гауссовский вектор*) Если мы умеем генерировать значения независимых $\mathcal{N}(0, 1)$, то мы можем получать семплы векторов из

$$X \sim \mathcal{N}(0, I).$$

Чтобы получить $Y \sim \mathcal{N}(\mu, \Sigma)$, достаточно найти любую матрицу A такую, чтобы получилось $AA^T = \Sigma$ и тогда

$$AX + \mu \sim \mathcal{N}(\mu, \Sigma).$$

Один из самых популярных методов для поиска A – это разложение Холецкого $\Sigma = AA^T$, где A – нижнетреугольная.

Но это примеры, не выходящие за классическую теорию вероятности. На самом деле мы знаем более сложные примеры вероятностных моделей.

Пример 17.5. (*Гауссовская смесь*) Гауссовская смесь это двухшаговая вероятностная модель, в которой задаётся набор средних μ_1, \dots, μ_K , набор ковариационных матриц $\Sigma_1, \dots, \Sigma_K$ и набор весов $\alpha_1, \dots, \alpha_K$, которые суммируются в 1 и неотрицательные. С точки зрения плотности это сумма гауссовых плотностей $\mathcal{N}(\mu_k, \Sigma_k)$ с весами α_k . Рецепт семплирования здесь очень простой:

1. Семплируется k из распределения на конечном множестве (вероятность – это вес компонента смеси α_k);
2. Семплируется $X \sim \mathcal{N}(\mu_k, \Sigma_k)$.

Пример 17.6. (*Гауссовский процесс*) Гауссовский процесс нам хорошо знаком, мы научились его семплировать на самой первой неделе.

Можно ли взять из головы произвольную плотность $f(x)$ и научиться получать семплы? Что особенно актуально для апостериорных распределений Байесовских методов: что если мы знаем форму плотности, но она не интегрируется в 1 ? Есть более-менее универсальный рецепт: Монте-Карло на Марковских цепях (Markov Chain Monte Carlo, MCMC).

17.2 Цепь Маркова

Давайте немного вспомним классические цепи Маркова с конечным числом состояний. Это процесс в дискретном времени с конечным числом N возможных значений. Такой процесс задаётся стартовым распределением μ_0 и переходной матрицей $P = (p(i, j))$. Можно себе представлять некоторого агента, который с вероятностью $\mu_0(i)$ оказывается в начальный момент в состоянии i и на каждом шаге по времени совершает независимые переходы $i \rightarrow j$ с вероятностью $p(i, j)$.

Мы помним, что есть эргодические цепи, где есть единственное *инвариантное распределение*

$$\pi : \pi P = \pi,$$

к которому сходится при $t \rightarrow \infty$ распределение $\mu_0 P^t$, вероятности нахождения в момент t в состоянии i . Важный момент состоит в том, что со временем цепь вообще будто забывает всю историю и в реальности коррелированные семплы последовательных состояний X_t можно до какого-то предела рассматривать как почти независимые и ковариационная функция быстро убывает с ростом расстояния между семплами. К примеру, для них есть закон больших чисел (для него достаточно эргодичности) и даже разные центральные предельные теоремы [12].

Основная идея методов МСМС – построить цепь Маркова, которая

- Будет эргодической;
- В качестве стационарного распределения будет иметь то, из которого нужно семплировать.

В этом случае наш алгоритм будет следующим:

1. Стартовать процесс из какого-то стартового распределения $X_0 \sim \mu_0$;
2. подождать до момента $T....$
3. Собирать семплы $X_{T+1}, \dots, X_{T+k}, \dots$

Или же таким:

1. Стартовать N процессов из какого-то стартового распределения $X_0 \sim \mu_0$;
2. подождать до момента $T....$
3. Собрать выборку из последних значений N цепей.

Единственный вопрос состоит в том, как именно построить по заданному распределению π такую цепь Маркова (переходные вероятности P). А в более общем (не конечном и не дискретном) случае нужно построить правильную последовательность переходных ядер P_h переходов за время h , в этом случае условие стационарности меры π перепишется так: для любого $A \in \mathcal{F}$ при вычислении распределения после времени h

$$(\pi P_h)(A) = \int_A P_h(x, A) d\pi(x)$$

распределение меняться не должно:

$$\pi P_h = \pi.$$

17.3 Как построить подходящую цепь?

Пока попробуем поработать в конечном случае (с дискретным временем и конечным числом состояний). На самом деле, даже здесь с практической точки зрения МСМС часто будет полезен. Пример ниже приводит в своём тексте известный учёный в области теории вероятности Перси Диаконис [7].

Предположим, что нам приходит зашифрованное сообщение типа

рассеаэдтынпрзңцыпкнЁарезва

и нам известно, что оно получено путём какой-то замены символов кириллицы и знака пробела. Это называется подстановочным кодом (substitution code), частный случай – это

код Цезаря, который получается просто сдвигом алфавита на t символов. Если мы хотим сломать код по-грубому (брутфорсом), то нам придётся перебрать порядка $N!$ различных расшифровок, что даже при небольшом алфавите типа 35 букв составляет

$$35! = 1033314796638614492966651337523200000000 \approx 10^{40}.$$

Допустим, мы знаем язык сообщения (русский). Тогда для русского можно найти в сети Интернет таблицу частот би-грамм (последовательностей двух символов) либо большой русский текст и посчитать её вручную. Помимо этого можно пытаться дополнительно использовать любую другую языковую статистику и даже скоры языковой модели. Исходя из неё можно предположить модель текста: каждый следующий символ текста генерируется независимо в соответствии с вероятностями би-грамма kv . Тогда для любого сообщения за линейное время можно вычислить его вероятность:

$$\mathbb{P}(ab\bar{b}) = \mathbb{P}(b | \bar{b}) \mathbb{P}(b | a) \mathbb{P}(a),$$

где переходные вероятности можно взять из таблицы би-грамм и поделить на частоту символа, а частоту символа – из статистики встречаемости букв русского языка.

Естественная идея тогда следующая. Будем начинать с произвольной расшифровки (таблицы, показывающей какую букву в какую нужно перевести при расшифровке) и цепь Маркова будет перебирать расшифровки, переходя в случайную равномерно с вероятностью $1/N!$. Такая цепь эргодическая, но предельное распределение у неё равномерное. То есть, такой перебор – это брутфорс, причём даже хуже изначального.

Чтобы исправить ситуацию, мы можем использовать уже выясненную информацию о языке. Мы не можем перебрать все расшифровки, но мы можем вычислить вероятность текста при условии, что расшифровка дана. Поэтому задача сводится по сути к поиску моды распределения с $N!$ значениями и вероятностями

$$\pi_i = \mathbb{P}(\sigma_i | text) = \frac{\mathbb{P}(text | \sigma_i)}{\sum_{j=1}^{N!} \mathbb{P}(text | \sigma_j)}.$$

Нормировочная константа нам недоступна, но мы попробуем построить цепь Маркова, перебирающую расшифровки (перестановки символов) $\sigma \in S^N$; если эта цепь будет иметь заданное инвариантное распределение и будет эргодической, но после некоторого прогрева мы будем генерировать семплы из распределения π . Далее вычислить моды легко: нужно всего лишь выбрать самые частые и посмотреть на текст. Либо выбрать топ-10 расшифровок, чтобы защититься от риска плохой оценки статистики языка. По сути задача у нас такая: есть K состояний (у нас $K = N!$), есть распределение π на K состояниях и нам нужно построить цепь Маркова (то есть, переходную матрицу P), которая была бы эргодической и качестве инвариантного распределения имела бы π .

Начнём с того, что нам нужен более простой критерий инвариантности π для матрицы P . По определению инвариантное распределение – это такое, что

$$\pi P = \pi$$

или

$$\sum_{i=1}^K \pi_i p_{ij} = \pi_j.$$

Это также иногда называют условием *баланса*. Его можно упростить, получив несложное достаточное условие.

Утверждение 17.1. *Если выполнено условие детального баланса*

$$\pi_i p_{ij} = \pi_j p_{ji},$$

то выполнено условие баланса (и, следовательно, π – инвариантное распределение).

▷ Просуммируем обе части по j :

$$\pi_i = \sum_{j=1}^K \pi_j p_{ji}, \quad -$$

это уравнение баланса. \square

Условие детального баланса называют также *обратимостью по времени* (time reversibility).

Это имеет вполне чёткий физический смысл: цепь, запущенная по шагам t_1, \dots, t_n будет иметь те же распределения, что цепь, запущенная с развернутым ядром в обратную сторону по времени (КАРТИНКА).

Теперь нам нужна эргодичность. Вспомним, что для эргодичности цепи тоже есть достаточно простые и проверяемые условия.

Утверждение 17.2. *Цепь Маркова эргодическая, если она*

- Несократимая (из каждого состояния в каждое можно дойти);
- Апериодическая (НОД времён всех возможных возвратов для каждого состояния равен 1).

Теперь остаётся явно предложить цепь, удостовериться, что она эргодическая и имеет нужное инвариантное распределение. Один из вариантов – это процедура Метрополиса–Гастингса (1953[16]? говорят, было известно раньше). Мы строим двухэтапную процедуру для одного шага по времени. На каждом шаге t

1. Генерируем предложение(proposal) $X' \sim q(\cdot | X_t)$;
2. С вероятностью $a(X' | X_t)$ принимается предложение и $X_{t+1} = X'$, иначе отвергается и $X_{t+1} = X_t$.

Заметим, что мы получили всё ещё цепь Маркова, причём такая цепь уже точно апериодическая, так как мы закладываем возможность оставаться на месте. Ещё мы можем потребовать несократимость, если цепь с ненулевой вероятностью сможет дойти до любого состояния. Чтобы завершить алгоритм, нужно подобрать распределение пропоузала и вероятность принятия. В качестве пропоузала мы, как изначально хотели, можем взять, к примеру, просто случайное состояние из равномерного распределения на K состояниях. Вероятность принятия поможет получить уравнения детального баланса.

Итак, мы строим цепь с переходными вероятностями

$$p(j \mid i) = q(j \mid i)a(j \mid i),$$

так как подброшенная монета на шаге принятия/отверждения не зависит от сгенерированного пропоузала. Уравнение детального баланса требует, чтобы

$$\pi_i q(j \mid i)a(j \mid i) = \pi_j q(i \mid j)a(i \mid j)$$

или по-другому

$$\frac{\pi_i}{\pi_j} = \frac{q(i \mid j)a(i \mid j)}{q(j \mid i)a(j \mid i)}.$$

Метрополис предложил такой вариант:

$$a(i \mid j) = 1 \wedge \frac{\pi_i q(j \mid i)}{\pi_j q(i \mid j)}.$$

Заметим, что если $a(i \mid j) < 1$ то $a(j \mid i) = 1$ и наоборот, поэтому уравнение детального баланса будет выполнено.

Подытожим алгоритм **RWMH**, Random Walk Metropolis-Hastings.

1. $X' \sim Q(\cdot | X_t)$, в качестве Q берём равномерное распределение на всех состояниях.
2. С вероятностью

$$a(X' \mid X_t) = 1 \wedge \frac{\pi_{X'} q(X_t \mid X')}{\pi_{X_t} q(X' \mid X_t)}$$

задаётся $X_{t+1} = X'$, иначе $X_{t+1} = X_t$.

Этот алгоритм несовершенен и можно его в разные стороны улучшать*. Например, мы не использовали знание текущего состояния в пропоузале (можно, к примеру, предлагать состояния из окрестности текущего, чтобы обход был более локальным). Ещё пропоузал можно делать более сложным, но всё это по большей части более специфичные детали.

*У меня такой брутфорс с русским не заработал за 30 минут времени, но по интернету у кого-то завелось. Думаю, проблема в очень долгом прогреве и секрет в правильной априорной информации для пропоузала.

17.4 В непрерывном случае

Как мы можем обобщить подход на случай, когда нам нужно сэмплировать из данной плотности $f(x)$? Оказывается, что как и раньше нам нужна цепь Маркова в дискретном времени, которая будет несократимой (в смысле, что из любого состояния траектория из точки x проходит с ненулевой вероятностью любую окрестность любой точки y за конечное время), апериодической (цепь вернётся в любую окрестность старта за конечное время, но без периода). Здесь нам, как и раньше, помогает уравнение детального баланса, которое для случая, когда переходное ядро имеет плотность $p(y|x)$ и желаемое распределение имеет плотность $f(x)$, можно записать как

$$f(x)p(y|x) = f(y)p(x|y).$$

Пример 17.7. Поэтому если мы находимся в \mathbb{R}^d и плотность $f(x)$ ненулевая на всём пространстве, то мы можем построить непрерывную версию **RWMH(cont)**, к примеру, так:

1. $X' \sim Q(\cdot|X_t)$, в качестве Q берём гауссовское распределение $\mathcal{N}(X_t, \sigma^2)$.

2. С вероятностью

$$a(X' | X_t) = 1 \wedge \frac{f(X')q(X_t | X')}{f(X_t)q(X' | X_t)}$$

задаётся $X_{t+1} = X'$, иначе $X_{t+1} = X_t$.

Мы могли бы взять пропузал $\mathcal{N}(0, \sigma^2)$, но проблема такого подхода была бы в том, что если большая часть вероятностной массы $f(x)$ находится далеко от 0, то все пропузалы бы с очень высокой вероятностью отвергались и мы бы ничего не добились. С другой стороны такая идея случайного блуждания позволяет уверенно обойти всю плоскость в надежде, что когда-то мы точно зайдём в область, где $f(x)$ ощутимо больше нуля (КАРТИНКА). Вообще алгоритмы типа RWMH дают несколько узкий взгляд на задачу. Мы сумели построить цепь с нужными свойствами, но для пропузала никак не используется знание плотности $f(x)$. Это приводит к тому, что имеющаяся информация используется неэффективно. Проблема алгоритмов МН в том, что они в теории дают эргодическую цепь, но почти нет гарантий, что она обязательно сойдётся за разумное время.

В ядре алгоритма МСМС находится эргодическая цепь. Но не так давно мы с помощью уравнения Фоккера-Планка узнали, что инвариантное распределение есть у модели диффузии Ланжевена. Например, у диффузии

$$dX_t = \nabla \ln f(X_t)dt + \sqrt{2}dW_t$$

инвариантным распределением будет плотность $C_0f(x)$ (даже в случае дискретизации по методу Эйлера), с точностью до нормировочной константы то, что нужно.

Пример 17.8. Первый простой подход – просто запустить диффузию с помощью метода Эйлера, подождать и потом собирать семплы. Это приводит к алгоритму **ULA** (*Unadjusted Langevin Algorithm*).

Но есть нюанс. Как показывают исследования, дискретизация не проходит бесследно[9]: с постоянным шагом по времени алгоритм ULA асимптотически смещён относительно настоящей стационарной плотности и смещение зависит от шага. Впрочем, если брать переменный шаг, этот эффект можно попробовать починить, но это непросто с точки зрения подбора параметров [8]. В общем, это недостаток схемы, который вместе со временем прогрева делает её сложной в использовании. С другой стороны, что мешает использовать диффузию в качестве пропуска в схеме Метрополиса-Гастингса? После дискретизации у нас останется Марковская цепь, пропуск больше не будет симметричным, хотя условие детального баланса будет выполняться. При этом в силу самой природы метода Метрополиса-Гастингса, инвариантное распределение будет таким, как надо.

Пример 17.9. Второй способ приводит к алгоритму **MALA** (*Metropolis-Adjusted Langevin Algorithm*).

1. Сгенерировать пропуск из ULA, формально

$$X' \sim \mathcal{N}(X_{t_k} + \nabla \ln f(X_{t_k}), 2\Delta t_k).$$

2. С вероятностью

$$a(X' \mid X_t) = 1 \wedge \frac{f(X')q(X_t \mid X')}{f(X_t)q(X' \mid X_t)}$$

задаётся $X_{t_{k+1}} = X'$, иначе $X_{t_{k+1}} = X_{t_k}$.

Преимущества MALA в том, что мы ликвидировали асимптотическое смещение ULA, а кроме того улучшили эффективность: цепи (во всяком случае теоретически) нужно меньше времени на прогрев, чем в случае ULA. Но ULA отмечать полностью не стоит. Что будет, если мы стартанём и

Наконец, в непрерывном случае есть ещё один способ, который относится к МСМС, но формально не совсем такой. Это Гамильтонов Монте-Карло (**HMC**). Методы HMC полностью детерминированные и используют правильно собранную систему обыкновенных дифференциальных уравнений типа

$$\dot{x} = p, \quad \dot{p} = g(x)$$

Если система гамильтонова, то есть, существует функция $H(x, p)$ такая, что эта система имеет вид

$$\dot{x} = \nabla_p H, \quad \dot{p} = -\nabla_x H.$$

Это история из самой классической физики и очень хорошо изученные системы с позиции математической физики в недавнее время [26]. Про такие системы известно, что они эргодические и имеют инвариантную меру. Например, если задать гамильтониан

$$H(x, p) = U(x) + \frac{1}{2}p^T V^{-1} p,$$

то инвариантная мера будет $e^{-U(x)}$. С точки зрения физики, первый член – это потенциальная энергия системы, а второй – кинетическая.

Пример 17.10. (*дописать про техническую сторону*) В теории и на практике это достаточно неплохой подход, но он страдает от некоторых недостатков:

1. Чувствительность к стартовому условию: может не повезти и ждать сходимости нужно долго.
2. Схема дискретизации важна. Популярный стандартный выбор – схема leapfrog, но есть более сложные варианты (например, No U-turn).
3. Ещё динамика рекуррентна в том смысле, что если ждать достаточно долго, то система постарается вернуться в окрестность старта. Чтобы получить достаточно хаотическое поведение, нужно подбирать параметры и добавлять случайность в разных местах, например задавать поле скоростей с использованием плотности $f(x)$ и дополнительной случайности.

Всё это чинится разными техническими улучшениями НМС и сам базовый алгоритм НМС сильно отличается от того, чтобы просто как есть использовать обыкновенное дифференциальное уравнение. Отдельного внимания заслуживает то, что можно применить шаг Метрополиса-Гастингса, и как именно, это отдельная история [23, 6].

17.5 Ещё альтернативы МСМС

МСМС – неплохой универсальный рецепт семплирования в высокой размерности в случае, когда дана известная плотность $f(x)$. Для эргодических средних формально зависимых наблюдений есть законы больших чисел, то есть, МСМС можно использовать для вычисления матожиданий. Более того, других способов в общий ситуации почти нет. Всё же есть достаточно много недостатков:

- (-) Нужно прогревать цепь, ждать сходимости к инвариантному распределению. Сложно понять, когда именно это происходит. И можно ждать очень долго.
- (-) Очень много дополнительной инженерии для того, чтобы всё заработало, как хотелось бы и добиваться быстрой сходимости.

Всё это приводит к вычислительной неэффективности: большое количество арифметики проходит впустую во время прогрева. По этой причине стал развиваться генеративный подход: вместо того, чтобы (к примеру, с помощью ММП) оценить вероятностную модель и думать, как из неё семплировать при известном распределении, пытаются изначально строить такую модель, которая бы явно включала в себя явно блок генерации, который при обученной модели мог бы быстро генерировать семплы.

Это немного другая постановка: теперь у нас нет явно плотности $f(x)$, но есть данные, семплы из неё. Генеративные модели пытаются что-то узнать из данных, чтобы сразу генерировать семплы из распределения, близкого к $f(x)$. Среди главных моделей здесь выступают

1. **Гауссовские смеси** (да, это простейшая генеративная модель). Эта модель оценивается ЕМ-алгоритмом и семплирование происходит по очень простой схеме.
2. **Гауссовские процессы** (да, это генеративная модель, но немного сложнее). Она оценивается с помощью метода максимального правдоподобия и потом можно просто генерировать траектории как гауссовые векторы.
3. **Фильтр Калмана.** В известных пределах отлично себя показывает на симуляции финансовых данных для задач, связанных с биржей. Его можно обучать в том числе в онлайн-режиме с помощью ЕМ и генерировать итеративно с известными функциями перехода и ковариационными матрицами.
4. **Вариационный автознекодер (VAE).** Это модель, которая имеет функцию вложения в скрытое пространство (энкодер) и функцию реконструкции из скрытого представления в пространство данных (декодер). Две эти функции правильным образом параметризуются и обучаются на задачу хорошей реконструкции и хорошей аппроксимации распределения. После обучения для генерации нужно сгенерировать вектор стандартного гауссова шума и прогнать через декодер.
5. **Нормализующие потоки (NF).** Это модель, немного похожая на VAE, но энкодер $f(x)$ явно обратим, поэтому декодер задаётся как $f^{-1}(y)$.
6. **Генеративно-состязательные сети (GAN).** В этой модели генератор – это функция, которая гауссовский стандартный вектор должна перевести в экземпляр, похожий на данные. Ещё есть блок дискриминатора, который по данному экземпляру классифицирует фейк (сгенерировано генератором) и не фейк (оальный семпл данных). По итогам обучения имеется модель генератора, которая тоже способна быстро генерировать семплы.

7. **Диффузионные модели.** Здесь после приглашённой лекции добавить нечего :).

Тоже модель, которая предназначена отойти от модели семплирования из готовой плотности.

Разумеется, у всех этих подходов есть свои проблемы и преимущества, но отчасти их развитие мотивировано в том числе недостатками МСМС для задач с данной выборкой наблюдений и явно неизвестной плотностью.

- [14] M. G. Kendall and A. Bradford Hill. The analysis of economic time-series-part i: Prices. *Journal of the Royal Statistical Society. Series A (General)*, 116(1):11–34, 1953.
- [15] Robert C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1):125–144, 1976.
- [16] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 06 1953.
- [17] Bernt Oksendal. *Stochastic Differential Equations (3rd Ed.): An Introduction with Applications*. Springer-Verlag, Berlin, Heidelberg, 1992.
- [18] Paul A. Samuelson. Proof that properly discounted present values of assets vibrate randomly. *The Bell Journal of Economics and Management Science*, 4(2):369–374, 1973.
- [19] O. A. Stepanov. Kalman filtering: Past and present. an outlook from russia. (on the occasion of the 80th birthday of rudolf emil kalman). *Gyroscopy and Navigation*, 2(2):99–110, Apr 2011.
- [20] R. L. Stratonovich. Conditional markov processes. *Theory of Probability & Its Applications*, 5(2):156–178, 1960.
- [21] Holbrook Working. A random-difference series for use in the analysis of time series. *Journal of the American Statistical Association*, 29(185):11–24, 1934.
- [22] L.E. Zachrisson. On optimal smoothing of continuous time kalman processes. *Information Sciences*, 1(2):143–172, 1969.
- [23] Guangyao Zhou. Metropolis augmented hamiltonian monte carlo. In *Fourth Symposium on Advances in Approximate Bayesian Inference*, 2022.
- [24] Н. В. Рекнер М. Шапошников С. В. Богачев, В. И. Крылов. *Уравнения Фоккера – Планка – Колмогорова*. Институт компьютерных исследований, 2013.
- [25] Ширяев А.Н. Булинский А.В. *Теория случайных процессов*. М:ФИЗМАТЛИТ, 2005.
- [26] Тиморин В.А. *Геометрия гамильтоновых систем и уравнений с частными производными*. Москва : ВШЭ, 2017.
- [27] Синай Я.Г. Коралов Л.Б. *Теория вероятностей и случайные процессы*. МЦНМО, 2013.
- [28] Р.Л. Стратонович. *Условные марковские процессы и их применение к теории оптимального управления*. Московский государственный университет, 1966.
- [29] А.Н. Ширяев. *Основы стохастической финансовой математики*. МЦНМО, 2016.