

# Заметки к курсу стохастического анализа\*

Борис Демешев, [boris.demeshev@gmail.com](mailto:boris.demeshev@gmail.com)

16 сентября 2016 г.

## Содержание

Цель: строгий курс для студентов не имеющих подготовки по теории меры, но знающих элементарный курс теории вероятностей. Дойти до простых SDE и модели БШ. Книга должна быть открыто доступна в интернете.

Пригодность для самостоятельной подготовки: после каждой subsection идет небольшое количество (5?) задач, которые нужно решить для себя. На них есть ответы. После каждой section идут дополнительные задачи (уже не обязательно с ответами?)

Идеология курса:

Все определения - строго

Все теоремы - строго формулируются

Несложные теоремы - строго доказываются

Сложные теоремы - по возможности дается набросок доказательства и ссылки для любопытных

Задачи на док-во лучше формулировать «Верно ли что, ...», а не «Докажите...»

Если доказательство больше страницы - в приложение.

Греческие буквы, аддитивный вариант, задачи из Crack

Глава 14 «Практическая». Модель Блэка-Шоулса.

Предпосылка модели.

Решение в явном виде для  $S_t$

Уравнение на  $e^{-rt}S_t$

Уравнение на  $e^{-rt}X_t$

Теорема. Гирсанов.

Способ оценки актива - 1.  $\mathbb{E}(X_T|\mathcal{F}_0)$

Примеры (в т.ч. оценить 75 центов в случае  $> 1$  - см. Crack)

Необходимое условие - дифференциальное уравнение.

Способ оценки актива - 2. Решение ДУ с начальными условиями.

Пример (оценить 75 центов в случае  $> 1$  - см. Crack)

Глава 15. Блэк-Шоулс глубже.

Греческие буквы.

Аддитивный вариант.

Быстрые вычисления

Задачи из Crack

Глава 16. Компьютерная.

Компьютерные симуляции броуновского движения.

Оценки мат. ожидания, дисперсии и функции плотности методом Монте-Карло.

Численное решение ДУ.

---

\*Свежая версия: [https://github.com/bdemeshev/sc401/tree/master/sc\\_book](https://github.com/bdemeshev/sc401/tree/master/sc_book)

Оценка опционов через Монте-Карло и решение ДУ.

Глава 17. Path-Dependent Euro Option.

Knock-out options. Перестает действовать, если. . .

Look-back options. Выигрыш зависит от максимальной цены.

Asian options. Выигрыш зависит от средней цены.

Source: Shreve-II

Аддон: комменты к популярным книжкам

Список литературы обязательно с комментариями! Возможно абзац на книжку!

## 1 Прививки для туристов

Здесь собраны сюжеты, которые нужно знать до изучения стохастического анализа.

### 1.1 Бесконечности бывают разные

Икеевских карандашей никогда не наберешь достаточно!

**Определение 2.1.** Множества  $A$  и  $B$  называются **равномощными**, если между элементами этих множеств существует взаимно-однозначное соответствие.

**Пример 2.2.** Множества  $\{1, 2, 3, 4, 5\}$  и  $\{5, 11, 12, 17, 22\}$  равномощны, так как есть взаимно-однозначное соответствие:

$$1 \leftrightarrow 5$$

$$2 \leftrightarrow 11$$

$$3 \leftrightarrow 12$$

$$4 \leftrightarrow 17$$

$$5 \leftrightarrow 22$$

**Пример 2.3.** Множество  $\mathbb{N}$  натуральных чисел и множество четных натуральных чисел равномощны в силу соответствия:

$$1 \leftrightarrow 2$$

$$2 \leftrightarrow 4$$

$$3 \leftrightarrow 6$$

$$4 \leftrightarrow 8$$

...

Заметим, что в определении требуется только существование взаимно-однозначного соответствия. Это не противоречит тому, что могут существовать другие, не взаимно-однозначные соответствия. В примере ?? можно добиться и того, что среди натуральных чисел останутся «лишние»:

$$1 \leftrightarrow ?$$

$$2 \leftrightarrow 2$$

$$3 \leftrightarrow 4$$

$$4 \leftrightarrow 6$$

...

И того, что среди четных останутся «лишние», а именно:

$$? \leftrightarrow 2$$

$$1 \leftrightarrow 4$$

$$2 \leftrightarrow 6$$

$$3 \leftrightarrow 8$$

...

**Пример 2.4.** Множества  $\{1, 2, 3, 4\}$  и множество  $\mathbb{N}$  неравномощны. При любой попытке построить взаимно-однозначное соответствие среди натуральных чисел останутся «лишние».

Возникает естественный вопрос, все ли бесконечные множества равномощны?

Пусть  $S$  множество всех бесконечных вправо последовательностей из 0 и 1. Например, одним из элементов  $S$  является последовательность 1010101010...

**Теорема 3.1.** *Множество  $S$  бесконечно, но не равномощно множеству  $\mathbb{N}$ .*

*Доказательство.* Допустим противоположное, что  $S$  и  $\mathbb{N}$  равномощны. Тогда существует взаимно-однозначное соответствие между натуральными числами и последовательностями. К примеру оно могло бы выглядеть так:  $1 \leftrightarrow 000000 \dots$   $2 \leftrightarrow 011000 \dots$   $3 \leftrightarrow 101011 \dots$   $4 \leftrightarrow 101001 \dots$

... Оказывается какое бы соответствие ни было создано, всегда существует последовательность, которой не сопоставлено ни одно число!

Создадим последовательность  $a$  по следующему принципу: возьмем первую цифру из первой последовательности, затем вторую из второй, затем третью из третьей и т.д. В нашем примере  $a = 0110 \dots$ :  $1 \leftrightarrow 000000 \dots$   $2 \leftrightarrow 011000 \dots$   $3 \leftrightarrow 101011 \dots$   $4 \leftrightarrow 101001 \dots$

Затем построим последовательность  $b$  заменив единицы на нули, а нули на единицы в последовательности  $a$ . В нашем примере  $b = 1001 \dots$

Вне зависимости от того, какое соответствие мы взяли (как в примере, или любое другое) последовательность  $b$  не может идти в нем ни под каким номером! Она не может идти под номером 1, так как отличается от первой последовательности первой цифрой. Она не может идти под номером 2, так как отличается от второй последовательности второй цифрой и т.д.

Мы пришли к противоречию, в  $S$  есть «лишняя» незанумерованная последовательность  $b$ . Значит  $S$  и  $\mathbb{N}$  неравномощны.  $\square$

**Определение 3.2.** Мы говорим, что множество  $A$  имеет мощность **континуум**, если оно равномощно множеству  $S$  бесконечных вправо последовательностей из 0 и 1.

**Определение 3.3.** Множество  $A$  называется **счетным**, если оно конечно или равномощно множеству  $\mathbb{N}$  натуральных чисел.

Будьте бдительны при чтении других источников: некоторые авторы определяют счетные как равномощные натуральным числам, но таких авторов меньшинство.

Из данного определения следует, что **несчетные** множества - это бесконечные множества не равномощные множеству  $\mathbb{N}$  натуральных чисел.

(...) картинка с классификацией конечные-бесконечные, счетные, несчетные

Оказывается, что любые бесконечные множества можно сравнивать: либо они равномощные, либо одно из них «больше».

**Теорема 3.4.** *Если  $A$  и  $B$ , два произвольных множества, то возможна одна и только одна из трех ситуаций:*

1.  $A$  и  $B$  равномощны.
2.  $A$  и  $B$  неравномощны, но  $A$  равномощно какому-нибудь подмножеству множества  $B$ .
3.  $A$  и  $B$  неравномощны, но  $B$  равномощно какому-нибудь подмножеству множества  $A$ .

*Доказательство.* Доказательство можно найти (...). Может возникнуть вопрос, а что тут собственно доказывать, все же «очевидно»? Доказывать нужно два утверждения. Во-первых, что ситуации 2) и 3) не могут произойти одновременно. Во-вторых, что невозможна гипотетическая ситуация «несравнимости», когда  $A$  и  $B$  неравномощны, в  $A$  нет части равномощной  $B$ , и в  $B$  нет части равномощной  $A$ .  $\square$

Из этой теоремы следует важное следствие:

**Теорема 3.5.** *Если  $A$  равномощно подмножеству  $B$ , то  $A \preceq B$ , т.е. либо мощность  $A$  меньше мощности  $B$ , либо  $A$  и  $B$  равномощны.*

Важные примеры!

- Бесконечные счетные множества:

–  $\mathbb{N}^2$ . Картинка:

Фактически мы доказали, что объединение счетного количества счетных множеств счетно<sup>1</sup>.

–  $\mathbb{Z}$ . Картинка:

–  $\mathbb{Z}^n$  Доказательство — по индукции. Мы уже доказали, что  $\mathbb{Z}^1$  — счетное множество. Теперь предположим, что  $\mathbb{Z}^{n-1}$  — счетное. Получаем цепочку:  $\mathbb{Z}^n \simeq \mathbb{Z}^{n-1} \times \mathbb{Z}^1 \simeq \mathbb{N} \times \mathbb{N} \simeq \mathbb{N}$ .

–  $\mathbb{Q}$ . Пусть  $f(q)$  — это сумма модулей числителя и знаменателя дроби (после сокращения), например,  $f(-2/3) = 5$ . Сначала нумеруем дроби с  $f(q) < 2$ , потом — дроби с  $f(q) < 3$ , потом — дроби с  $f(q) < 4$  и т.д. В результате каждая дробь получает свой номер.

• Множества мощности континуум:

– интервал  $(0; 1)$ .

– прямая  $\mathbb{R}$

– пространство  $\mathbb{R}^n$

– функции непрерывные на отрезке  $[0; 1]$

Для целей стохастического анализа нам хватит этих сведений про мощности множеств, однако для общего развития полезно знать еще пару фактов.

Факт 1. Бесконечные множества не исчерпываются равномощными множеству  $\mathbb{N}$  и множеству последовательностей  $S$ . Бесконечности бывают разные, и их бесконечно много. Если есть одно бесконечное множество  $A$ , то найдется бесконечное множество  $B$ , где элементов еще больше, чем в  $A$ !  $A$  именно:

**Теорема 4.1.** Если  $A$  произвольное множество, то  $B$ , множество всех подмножеств множества  $A$ , обозначаемое  $2^A$ , не равномощно множеству  $A$ .

*Доказательство.* Доказательству этого утверждения посвящено упражнение (...) □

Иными словами, подмножеств действительных чисел больше, чем действительных чисел. И так далее..

Факт 2. Невинный вопрос: а есть ли мощности промежуточные между континуумом и мощностью множества натуральных чисел? оказывается неожиданно сложным! Еще более неожиданно то, что любой ответ на него, и «да», и «нет» оказывается верным. Для ответа на этот вопрос интуитивных представлениях о множествах не хватает и нужно аксиоматизировать теорию множеств. Так вот разные системы аксиом приводят к разным ответам. В данном курсе мы это обсуждать не будем. Но заинтересовавшиеся могут посмотреть книгу Верещагин Шень, Начала теории множеств, <http://www.mccme.ru/free-books/>

## Ссылки

Любопытствующим: Всего 40 страниц заметок «Set theory» на <http://math.uga.edu/~pete/expositions.html>.

## Задачи

**Задача 1.1.** Чудо-чудное, диво-дивное! Существует ли список  $S$  всех возможных мощностей?

---

<sup>1</sup>Для знатоков может быть интересен тот факт, что в этом доказательстве неявно используется аксиома выбора. Без нее можно представить  $\mathbb{R}$  как счетное объединение счетных множеств (ссылка, найти в Williams, Weighting the odds)

## 1.2 Частичный предел. Верхний и нижний.

Если есть последовательность  $a_i$ , то, забыв порядок чисел, можно рассмотреть ее как множество точек на числовой прямой.

**Пример 5.0.** Последовательности  $a_n = (-1)^n + \frac{1}{n}$  на прямой соответствует множество точек: (рисунок)

На прямой могут существовать точки в любой окрестности которых есть бесконечное количество членов последовательности  $a_n$ , в нашем примере это точки  $x = -1$  и  $x = 1$ . Это точки называются точками накопления.

**Определение 5.1.** точка  $x \in \mathbb{R}$  называется точкой **накопления** или **частичным пределом** для последовательности  $a_n$ , если в любой  $\varepsilon$ -окрестности точки  $x$  содержится бесконечное количество членов последовательности  $a_n$ .

Заметим, что если точка  $x$  является точкой накопления, то к ней можно «подобраться» сколь угодно близко выбрав некоторую подпоследовательность из последовательности  $a_n$ . Поэтому можно дать альтернативное определение:

**Определение 5.2.** точка  $x \in \mathbb{R}$  называется точкой **накопления** или **частичным пределом** для последовательности  $a_n$ , если из последовательности  $a_n$  можно выбрать подпоследовательность  $b_k$ , сходящуюся к  $x$ .

Наконец про верхний и нижний частичные пределы. Наибольшая из точек накопления называется верхним частичным пределом, наименьшая - нижним частичным пределом. Обозначаются они  $\limsup$  и  $\liminf$ , соответственно. Чтобы сделать эту мысль совсем точной, остается лишь добавить два крайних случая:

- Верхний частичный предел  $\limsup a_n = +\infty$ , если в последовательности  $a_n$  можно найти сколь угодно большие положительные числа
- Нижний частичный предел  $\liminf a_n = -\infty$ , если в последовательности  $a_n$  можно найти сколь угодно сильно отрицательные числа

Если от последовательности отрезать сколько-то элементов в начале (будь то 10, 100 или  $10^{100}$ ), то на точки накопления это никак не повлияет. Если вокруг некой точки  $x$  было бесконечное количество членов последовательности, то после отрезания начала последовательности вокруг точки  $x$  существенных изменений не произойдет. Именно на этой идее базируется формальное определение. Заодно оно проливает свет на происхождение обозначения:

**Определение 5.3.** **Верхний частичный предел**,  $\limsup_n a_n := \lim_{n \rightarrow \infty} \sup_{k \geq n} a_k$ .

**Нижний частичный предел**,  $\liminf_n a_n := \lim_{n \rightarrow \infty} \inf_{k \geq n} a_k$ .

Несколько примеров.

**Пример 5.4.** Если  $a_n = (-1)^n + \frac{1}{n}$ , то  $\limsup a_n = 1$ ,  $\liminf a_n = -1$

**Пример 5.5.** Если  $a_n = n(-1)^n$ , то  $\limsup a_n = +\infty$ ,  $\liminf a_n = -\infty$

**Пример 5.6.** Если  $a_n = n(1 + (-1)^n)$ , то  $\limsup a_n = +\infty$ ,  $\liminf a_n = 0$

## 1.3 Равномерная непрерывность

Понадобится для доказательств связанных с характеристическими функциями.

Рассмотрим непрерывную функцию  $f : \mathbb{R} \rightarrow \mathbb{R}$  и произвольную точку  $x_0$ . Если мы хотим чтобы  $f(x)$  не сильно отличалось от  $f(x_0)$ , то нам достаточно взять значение  $x$  не сильно отличающееся от  $x_0$ . А именно:

**Определение 5.7.** Функция  $f : \mathbb{R} \rightarrow \mathbb{R}$  называется непрерывной в точке  $x_0$ , если для любого  $\delta$  найдется такая  $\varepsilon$ -окрестность точки  $x_0$ , что для любого  $x \in (x_0 - \varepsilon; x_0 + \varepsilon)$  будет выполнено неравенство  $|f(x) - f(x_0)| < \delta$ .

При этом нужное  $\varepsilon$  в общем случае будет зависеть и от желаемого  $\delta$  и от точки  $x_0$ :

**Пример 5.8.** Функция  $f(t) = t^2$  на  $\mathbb{R}$ . Если мы хотим чтобы  $f(x)$  отличалось от  $f(2) = 4$  не более, чем на 5, то достаточно взять  $\varepsilon = 1$ . Действительно, если  $|x - 2| < 1$ , то  $|f(x) - 4| < 5$ .

Однако, чтобы  $f(x)$  отличалось от  $f(10) = 100$  не более, чем на 5, необходимо подойти к точке  $x_0 = 10$  гораздо ближе чем на  $\varepsilon = 1$ . Достаточным окажется лишь  $\varepsilon = \sqrt{105} - 10 \approx 0.25$ .

Однако для некоторых функций необходимое  $\varepsilon$  можно выбирать даже не зная точки  $x_0$ .

**Пример 6.1.** Функция  $f(t) = 2t + 7$  на  $\mathbb{R}$ . Если мы хотим, чтобы  $f(x)$  отличалось от  $f(x_0)$  не более, чем на 5, то достаточно взять  $\varepsilon = 2.5$ , вне зависимости от  $x_0$ .

Вот такие функции и называются равномерно непрерывными.

**Определение 6.2.** Функция  $f : \mathbb{R} \rightarrow \mathbb{R}$  называется **равномерно непрерывной** на множестве  $A \subset \mathbb{R}$ , если для любого  $\delta$  найдется такое  $\varepsilon$ , что для любых  $x$  и  $y$ , таких что  $|x - y| < \varepsilon$  будет выполнено неравенство  $|f(x) - f(y)| < \delta$ .

**Пример 6.3.** Функция  $f(t) = \cos(t)$  является равномерно непрерывной на  $\mathbb{R}$ .

**Пример 6.4.** Функция  $f(t) = t^2$  не является равномерно непрерывной на  $\mathbb{R}$ .

## 1.4 Еще задачи

**Задача 1.2.** Счетно ли множество  $S_1$  бесконечных вправо последовательностей из 0 и 1, в которых количество 1 конечно? Для ясности,  $1010101010 \dots$  (нули и единицы чередуются)  $\notin S_1$ , но  $010000000000 \dots \in S_1$  (в последовательности только одна единица, стоящая на втором месте).

**Задача 1.3.** Декартово произведение конечного количества счетных множеств является счетным множеством. Да или нет?

**Задача 1.4.** Декартово произведение счетного количества счетных множеств является счетным множеством. Да или нет?

**Задача 1.5.** An enemy submarine is somewhere on the number line (consider only integers for this problem). It is moving at some rate (again, integral units per minute). You know neither its position nor its velocity.

You can launch a torpedo each minute at any integer on the number line. If the submarine is there, you hit it and it sinks. You have all the time and torpedoes you want. You must sink this enemy sub - devise a strategy that is guaranteed to eventually hit the enemy sub.

**Задача 1.6.** Пусть  $A$  произвольное множество. Докажите, что мощность множества  $2^A$  (множество всех подмножеств множества  $A$ ) больше, чем мощность  $A$ .

**Задача 1.7.** Верно ли следующее утверждение: у последовательности  $a_n$  существует предел если и только если  $\limsup a_n = \liminf a_n$

## 2 Списки событий и сигма-алгебры

### 2.1 Наделенность информацией и определение

Рассмотрим простой случайный эксперимент: игральный кубик подбрасывают два раза. Множество  $\Omega$  исходов данного эксперимента содержит 36 элементов. Вася знает результат подбрасываний, и сообщает Пете значение случайной величины  $Z$  - произведения очков на выпавших гранях.

Всегда ли сможет ли Петя владея своей информацией, определить произошли ли события:

$A$  - оба раза выпала единица,

$B$  - хотя бы раз выпала пятерка,

$C$  - хотя бы раз выпала шестерка?

Про события  $A$  и  $B$  Петя всегда сможет сказать произошли ли они: в первом случае достаточно сравнить  $Z$  с единицей, во втором - определить, делится ли  $Z$  на 5. Однако может сложиться такая ситуация, что Петя не будет уверен, произошло ли  $C$ : например, если  $Z$  окажется равным шести, то может быть это была пара (6,1) и тогда  $C$  произошло, а может это была пара (2,3) и тогда  $C$  не произошло.

Можно составить список событий, про которые Петя, зная  $Z$ , всегда сможет сказать, произошли ли они. Обозначим его буквой  $\mathcal{F}$ . В отличие от Пети Вася может уверенно сказать произошло ли любое событие и аналогичный список событий для него - все подмножества множества  $\Omega$ .

(упр) Приведите примеры еще 3 событий, входящих в  $\mathcal{F}$ , и еще 3-х событий не входящих в  $\mathcal{F}$ .

В список  $\mathcal{F}$  входят довольно много событий, а сам список обладает рядом важных свойств, а именно:

- 1)  $\Omega \in \mathcal{F}$  - даже ничего не зная, можно быть уверенным, что  $\Omega$  произошло.
- 2)  $\emptyset \in \mathcal{F}$  - даже ничего не зная, можно быть уверенным, что  $\emptyset$  не произошло.
- 3) Если  $A$  и  $B$  входят в список  $\mathcal{F}$ , то  $A \cup B$  и  $A \cap B$  входят в список  $\mathcal{F}$ . Если Петя способен определить, произошли ли  $A$  и  $B$  по отдельности, то он путем простых логических заключений способен определить, произошли ли  $A \cup B$  и  $A \cap B$ .

Подобные списки событий являются очень важными для нас и называются  $\sigma$ -алгебрами («сигма-алгебрами»).

**Определение 7.1.** Набор подмножеств множества  $\Omega$  называется  $\sigma$ -алгеброй<sup>1</sup>, если обладает следующими тремя свойствами:

SA1.  $\Omega \in \mathcal{F}$

SA2. Если  $A \in \mathcal{F}$ , то  $A^c \in \mathcal{F}$

SA3. Если  $A_1 \in \mathcal{F}$ ,  $A_2 \in \mathcal{F}$ ,  $A_3 \in \mathcal{F}$ , и т.д., то  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ .

**Пример 7.2.**

**Пример 7.3.**

Почему в определении  $\sigma$ -алгебры мы не потребовали, чтобы выполнялись все свойства (...)? Оказывается, неупомянутые свойства следуют из свойств SA1-SA3.

**Пример 7.4.** Применив SA1, а затем SA2 можно понять, что в любую  $\sigma$ -алгебру входит пустое множество.

**Пример 7.5.** Пересечение множеств можно заменить на несколько объединений и дополнений,  $\bigcap_{i=1}^{\infty} A_i = (\bigcup_{i=1}^{\infty} A_i^c)^c$ . А значит из свойств SA2, SA3 следует также, что:

Если  $A_1 \in \mathcal{F}$ ,  $A_2 \in \mathcal{F}$ ,  $A_3 \in \mathcal{F}$ , и т.д., то  $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$ .

Ссылка на упражнение про симметрическую разность (...)

Проще говоря,  $\sigma$ -алгебра - это набор событий замкнутый относительно любых операций с множествами ( $\cup$ ,  $\cap$ ,  $()^c$ ,  $\Delta$ ) взятых в счетном количестве.

Самое время упомянуть два полезных множества, которые точно находятся в  $\sigma$ -алгебре  $\mathcal{F}$  (если известно, что все  $A_i$  лежат в  $\mathcal{F}$ ):

- $\limsup_n A_n$  - те исходы  $w \in \Omega$ , которые входят в бесконечное количество  $A_n$ . Почему это множество обязательно лежит в  $\mathcal{F}$ ? Легко заметить, что  $B_n = \bigcup_{k \geq n} A_k$  - это те исходы  $w$ , которые входят хотя бы в один  $A_k$  при  $k \geq n$ . Чтобы  $w$  входил в бесконечное количество  $A_n$  необходимо и достаточно того, чтобы исход  $w$  входил во все  $B_n$ . Значит,  $\limsup_n A_n = \bigcap_n \bigcup_{k \geq n} A_k$ .
- $\liminf_n A_n$  - те исходы  $w \in \Omega$ , которые входят во все  $A_n$ , начиная с некоторого. Рассуждая аналогично,  $C_n = \bigcap_{k \geq n} A_k$  - это те исходы  $w$ , которые входят во все  $A_k$  при  $k \geq n$ . И, получается, что  $\liminf_n A_n = \bigcup_n \bigcap_{k \geq n} A_k$

**Пример 7.6.** При заданном наборе  $\Omega$  исходов случайного эксперимента самая «подробная»  $\sigma$ -алгебра - это  $2^\Omega$ , множество всех подмножеств  $\Omega$ ; а самая «бедная»  $\sigma$ -алгебра - это набор  $\{\emptyset, \Omega\}$ . Убедитесь, что аксиомы SA1-SA3 выполнены для обоих случаев!

**Теорема 7.7.** Пересечение произвольного количества  $\sigma$ -алгебр -  $\sigma$ -алгебра.

<sup>1</sup>Правильный английский термин  $\sigma$ -algebra. В некоторых текстах встречается устаревшее  $\sigma$ -field. Для знающих определения поля и алгебры отметим, что  $\sigma$ -алгебра действительно будет полем, только если  $\mathcal{F} = \{\emptyset, \Omega\}$ . Вопрос для знатоков: относительно каких операций и над чем  $\sigma$ -алгебра будет алгеброй?

В силу этого простого наблюдения корректно говорить о минимальной  $\sigma$ -алгебре, порождаемой данным набором событий.

**Определение 8.0.** Пусть  $\mathcal{H}$  - произвольный набор событий, не обязательно  $\sigma$ -алгебра.  $\sigma$ -алгебра  $\mathcal{F}$  называется **минимальной  $\sigma$ -алгеброй порожденной** набором  $\mathcal{H}$ , если:  $\mathcal{F}$  содержит набор  $\mathcal{H}$ , и любая другая  $\sigma$ -алгебра  $\mathcal{F}'$ , содержащая набор  $\mathcal{H}$ , содержит в себе  $\mathcal{F}$ . Обозначается порожденная  $\sigma$ -алгебра так:  $\sigma(\mathcal{H})$ .

Другими словами, чтобы найти минимальную  $\sigma$ -алгебру, содержащую заданный набор событий, можно рассмотреть все  $\sigma$ -алгебры, содержащие заданный набор событий, и взять их пересечение.

Заметим, что объединение даже двух  $\sigma$ -алгебр может не быть  $\sigma$ -алгеброй!

**Пример 8.1.** Пусть  $\mathcal{F}_1 = \{\mathbb{R}, \emptyset, (-\infty; 1), [1; +\infty)\}$ ,  $\mathcal{F}_2 = \{\mathbb{R}, \emptyset, (-\infty; 2), [2; +\infty)\}$  - две  $\sigma$ -алгебры. Тогда,  $\mathcal{F}_1 \cup \mathcal{F}_2 = \{\mathbb{R}, \emptyset, (-\infty; 1), [1; +\infty), (-\infty; 2), [2; +\infty)\}$  - не  $\sigma$ -алгебра.

В качестве небольшого итога:  $\sigma$ -алгебра, это список событий различимых рациональным индивидом. То есть список событий, про каждое из которых рациональный индивид (умеющий делать простые логические заключения) может гарантированно сказать, произошло они или нет. Забежим немного вперед и скажем, что именно для событий из  $\sigma$ -алгебры мы будем определять вероятность. У каждого события из данной  $\sigma$ -алгебры  $\mathcal{F}$  будет определена вероятность, число от 0 до 1. А у событий не входящих в  $\sigma$ -алгебру  $\mathcal{F}$  вероятности не будет вообще.

## 2.2 Борелевская сигма-алгебра

Из всех искусств для нас важнейшим является кино<sup>1</sup>. (Ленин)

Из всех  $\sigma$ -алгебр для нас важнейшей является борелевская  $\sigma$ -алгебра!!!

**Определение 8.2.** Борелевская  $\sigma$ -алгебра - это  $\sigma$ -алгебра порожденная открытыми подмножествами множества  $\Omega$ .

Чаще всего нам понадобятся борелевские  $\sigma$ -алгебры:  $\mathcal{B}(\mathbb{R})$  и  $\mathcal{B}(\mathbb{R}^n)$ .

Заметим, что сам набор открытых множеств не является  $\sigma$ -алгеброй. Например, множество  $A = (-\infty; 2)$  является открытым, а множество  $\mathbb{R} \setminus A$  не является открытым.

**Определение 8.3.** Множество называется **борелевским** если оно является элементом борелевской  $\sigma$ -алгебры

**Пример 8.4.** Например на  $\Omega = \mathbb{R}$  борелевскими являются:

- любой интервал, например  $(2; 100)$  - т.к. он является открытым множеством
- любой отрезок, например  $[2; 100]$  - т.к. его можно получить в виде  $\mathbb{R} \setminus ((-\infty; 2) \cup (100; +\infty))$
- любую произвольную точку, например  $\{7\}$  - т.к. ее можно получить в виде  $\mathbb{R} \setminus ((-\infty; 7) \cup (7; +\infty))$
- любой полуинтервал, например  $[2; 100)$  - т.к. его можно представить в виде объединения  $[2; 3] \cup (3; 100)$

Самое время решить упражнение (...), которое показывает, что все «привычные» множества - борелевские.

Существуют ли неборелевские множества? Да, существуют! Но они очень «страшные» и не будут изучаться в этой книжке. Их даже больше, чем борелевских. Ситуация с ними похожа на ситуацию с иррациональными числами: иррациональных больше, чем рациональных, но используют при практических расчетах, как правило, рациональные. А именно:

**Теорема 8.5.** Мощность борелевской  $\sigma$ -алгебры  $\mathcal{B}(\mathbb{R})$  - континуум.

*Доказательство.* ссылка или в апшедникс? (...)

□

Из этой теоремы следует, что неборелевских множеств больше, а именно  $2^{\text{continuum}}$  (...)

Чуть позже (...) мы приведем три примера неборелевских множеств.

---

<sup>1</sup>Об этой цитате можно прочесть на <http://liveuser.livejournal.com/62878.html>



Борелевскую  $\sigma$ -алгебру можно породить также с помощью более простых наборов множеств. Например:

**Теорема 9.0.** Если  $\mathcal{H} = \{(-\infty; t] | t \in \mathbb{R}\}$ , то  $\sigma(\mathcal{H}) = \mathcal{B}$ .

*Доказательство.* Шаг 1. Заметим, что  $\mathcal{H} \subset \mathcal{B}$  и следовательно  $\sigma(\mathcal{H}) \subset \mathcal{B}$ . Для того, чтобы доказать равенство обеих частей, осталось доказать, что любое открытое множество войдет в  $\sigma(\mathcal{H})$ .

Шаг 2. Множества вида  $(a; b]$  входят в  $\sigma(\mathcal{H})$ , т.к. в рамках  $\sigma$ -алгебры можно брать разность множеств, а  $(a; b] = (-\infty; b] \setminus (-\infty; a]$ .

Шаг 3. Пусть  $A$  - произвольное открытое множество. Мы докажем, что можно построить последовательность  $A_1, A_2, \dots$ , такую, что каждое  $A_i$  лежит в  $\sigma(\mathcal{H})$  и  $A = \cup A_i$ . И следовательно,  $A \in \sigma(\mathcal{H})$ :

В качестве  $A_1$  возьмем объединение тех множеств вида  $(n; n+1]$ , где  $n$  - целое, а  $(n; n+1]$  лежит в  $A$ . В качестве  $A_2 \dots (0.5n; 0.5(n+1)] \dots$  В качестве  $A_k \dots (0.5^{k-1}n; 0.5^{k-1}(n+1)]$ .

Иллюстрация для  $A = (0; \sqrt{3}) = (0; 1.73\dots)$ :  $A_1 = (0; 1]$ ,  $A_2 = (0; 1.5]$ ,  $A_3 = (0; 1.5]$ ,  $A_4 = (0; 1.625]$  и т.д.

Почему нас всегда ждет успех? Рассмотрим произвольную точку  $a \in A$ . Вокруг любой точки открытого множества  $A$  можно найти  $\varepsilon$ -окрестность, которая целиком лежит  $A$ . Длина кусочков, из которых состоит каждое  $A_k$  стремится к 0. Значит наступит, такой момент, когда кусочки очередного  $A_k$  «залезут» в произвольную  $\varepsilon$ -окрестность и покроют точку  $a$ . Следовательно, любая точка из  $A$  лежит в некотором  $A_k$ . А, значит,  $A = \cup_i A_i$ .  $\square$

...

Заметим, что в качестве порождающих множеств можно взять, например, такие наборы:

....

## Задачи

**Задача 2.1.** Верно ли, что борелевскими являются следующие подмножества  $\mathbb{R}$ : а) множество целых чисел  $\mathbb{Z}$  б) множество решений уравнения  $\cos(2x) + x = 0$  в) множество решений неравенства  $x^5 - 5x^4 + x^3 - x + 8 \geq 0$

## 2.3 Еще задачи

**Задача 2.2.** За час до Нового года Дед Мороз дарит Вовочке одну конфету. За полчаса он забирает ее, съедает сам и дарит две другие конфеты. За четверть часа, забирает их, съедает и дарит четыре конфеты. И так далее. Пусть  $A_i$  - это номера тех конфет которые находятся в распоряжении Вовочки после  $i$ -го действия Деда Мороза. Например,  $A_1 = \{1\}$ ,  $A_2 = \{2, 3\}$ ,  $A_3 = \{4, 5, 6, 7\}$  и так далее.

Существует ли  $\lim A_i$ ? Если да, то чему он равен?

**Задача 2.3.** Пусть  $N$  - произвольная функция сопоставляющая каждому исходу число или плюс-минус бесконечность,  $N : \Omega \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$ . Можно думать об  $N$  как о случайной величине. Найдите  $\lim_{k \rightarrow +\infty} (N < k)$

**Задача 2.4.** Игральный кубик подбрасывается один раз. В первый момент времени наблюдатель узнает, выпала ли четная грань или нет. Во второй момент времени наблюдатель узнает, выпала ли грань большая двух или нет. В третий момент времени наблюдатель точно узнает, какая грань выпала. Укажите множество элементарных событий и соответствующие три  $\sigma$ -алгебры событий.

## 3 Случайные величины

Сумма чисел на противоположных сторонах игральной кости всегда равна семи.

Обратите внимание, что мы начинаем главу про случайные величины перед главой про вероятность! Без последующего определения вероятности случайные величины окажутся бесполезны, но важно понимать, что понятие случайной величины не требует для себя наличия вероятности. Например, вероятность можно поменять, сохранив случайную величину, при этом, конечно, все интересные количества как то  $\mathbb{P}(X < 0)$ ,  $\mathbb{E}(X)$  могут измениться.

### 3.1 Определение случайной величины

**Определение 10.1.** Случайная величина  $X$  - это функция, сопоставляющая каждому исходу случайного эксперимента некое действительное число,  $X : \Omega \rightarrow \mathbb{R}$ .

Если рациональному наблюдателю известно значение  $X$ , то он может определить, произошли ли например события  $X < 0$ ,  $X \in [2; 3]$ ,  $X = \sqrt{3}$  и др. Таким образом, с каждой случайной величиной  $X$  связана сигма-алгебра  $\sigma(X)$  событий гарантированно различимых наблюдателем, знающим значение  $X$ . Доведем эту идею до формального определения. Что означает, что я знаю  $X$ ? Это означает, что я могу сравнить  $X$  с любым действительным числом, т.е. для любого заданного  $t$  могу сказать произошло ли событие вида  $\{X < t\}$ . Получаем определение:

**Определение 10.2.** Сигма-алгеброй порожденной случайной величиной  $X$ ,  $\sigma(X)$ , называется минимальная сигма-алгебра, содержащая все события вида  $\{X < t\}$ , т.е.  $\sigma(X) := \sigma(\{X < t | t \in \mathbb{R}\})$ .

Как мы видели (...) с помощью подмножеств вида  $(-\infty; t)$  можно породить нашу любимую борелевскую сигма-алгебру, а значит определение можно переформулировать и так:

**Определение 10.3.** Сигма-алгеброй порожденной случайной величиной  $X$ ,  $\sigma(X)$ , называется минимальная сигма-алгебра, содержащая все события вида  $\{X \in A\}$ , где  $A$  — борелевское подмножество  $\mathbb{R}$ , т.е.  $\sigma(X) := \sigma(\{X \in A | A \in \mathcal{B}(\mathbb{R})\})$ .

Тут упражнение ...

Для понимания структуры  $\sigma(X)$  полезным может оказаться понятие прообраза:

**Определение 10.4.** Если задана функция  $X : \Omega \rightarrow \mathbb{R}$ , то **прообразом** (pullback) множества  $A$  называется множество  $\{w \in \Omega | X(w) \in A\}$ , обозначается прообраз  $X^{-1}(A)$ .

**Пример 10.5.** Функция  $f(t) = t^2$ . Примеры прообразов:  $f^{-1}(5) = \{\sqrt{5}, -\sqrt{5}\}$ ,  $f^{-1}(0) = \{0\}$ ,  $f^{-1}(-1) = \emptyset$ .

Используя понятие прообраза можно понять, что  $\sigma(X)$  состоит ровно из прообразов всех борелевских множеств. Действительно, прообразы борелевских множеств обязаны входить в  $\sigma(X)$  исходя из определения (...). А никакое лишнее множество нам не нужно, т.к.

**Теорема 10.6.** Пусть задана случайная величина  $X$ . Набор прообразов всех борелевских множеств  $\{X^{-1}(B) | B \in \mathcal{B}\}$  является  $\sigma$ -алгеброй.

*Доказательство.* Проверяем три свойства  $\sigma$ -алгебры.

- $\Omega = X^{-1}((-\infty; +\infty))$
- Если  $A = X^{-1}(B)$ , то  $A^c = X^{-1}(B^c)$
- 

□

Если рациональный индивид помимо событий из  $\sigma(X)$  различает какие-то еще события, то он конечно же, также может определить значение случайной величины  $X$ . Для описания такой ситуации служит:

**Определение 10.7.** Случайная величина  $X : \Omega \rightarrow \mathbb{R}$  называется **измеримой** относительно  $\sigma$ -алгебры  $\mathcal{F}$ , если  $\sigma(X) \subseteq \mathcal{F}$ .

**Пример 10.8.** Рассмотрим  $\Omega$ , состоящее из 3-х элементов,  $\sigma$ -алгебру  $\mathcal{F} = \{\emptyset, \Omega, a, \{b, c\}\}$  и случайные величины  $X$  и  $Y$ : (...)

В данном случае  $X$  является  $\mathcal{F}$ -измеримой случайной величиной, а  $Y$  - не является  $\mathcal{F}$ -измеримой.

Иногда символом  $\mathcal{F}$  обозначают не только саму  $\sigma$ -алгебру, но и множество всех случайных величин, являющихся  $\mathcal{F}$ -измеримыми. Это дает право использовать короткое обозначение  $X \in \mathcal{F}$ , означающее, что случайная величина  $X$  является  $\mathcal{F}$ -измеримой величиной. Буква  $\mathcal{F}$  оказывается слегка перегруженной, но проблем при этом не возникает. Если  $A$  - событие, то  $A \in \mathcal{F}$  следует понимать буквально: «множество  $A$  входит в список  $\mathcal{F}$ ». Если  $X$  - случайная величина, то  $X \in \mathcal{F}$  следует понимать как « $X$  является  $\mathcal{F}$ -измеримой случайной величиной». В конце концов - букв мало, а событий и случайных величин - много!

**Теорема 11.1.** Любая случайная величина  $X : \Omega \rightarrow \mathbb{R}$  является измеримой относительно самой «подробной»  $\sigma$ -алгебры  $2^\Omega$  (все подмножества множества  $\Omega$ ).

*Доказательство.* Самая «подробная»  $\sigma$ -алгебра содержит все подмножества множества  $\Omega$  и, в частности, содержит сигма-алгебру  $\sigma(X)$ , куда входят все прообразы борелевских множеств.  $\square$

Пусть  $X : \Omega \rightarrow \mathbb{R}$ , на  $\Omega$  задана  $\sigma$ -алгебра  $\mathcal{F}$ , а на  $\mathbb{R}$ - борелевская  $\sigma$ -алгебра  $\mathcal{B}$ . Оказывается, даже если  $X$  не является  $\mathcal{F}$ -измеримой ...

Множества, у которых «хороший» прообраз (прообраз попадает в  $\mathcal{F}$ ) образуют  $\sigma$ -алгебру (не обязательно совпадающую с борелевской).

**Теорема 11.2.** Пусть задана случайная величина  $X$ . Набор числовых подмножеств с прообразом, попадающим в  $\mathcal{F}$ ,  $\mathcal{D} := \{D | X^{-1}(D) \in \mathcal{F}\}$  является  $\sigma$ -алгеброй.

*Доказательство.* Проверяем три свойства  $\sigma$ -алгебры. 1)  $X^{-1}(\mathbb{R}) = \Omega$ , значит  $\mathbb{R} \in \mathcal{D}$

2)  $X^{-1}(D^c) = (X^{-1}(D))^c$ , значит имеем цепочку  $D \in \mathcal{D} \Rightarrow X^{-1}(D) \in \mathcal{F} \Rightarrow (X^{-1}(D))^c \in \mathcal{F} \Rightarrow X^{-1}(D^c) \in \mathcal{F} \Rightarrow D^c \in \mathcal{D}$ .

3) Аналогично б. в силу того, что прообраз объединения равен объединению прообразов.  $\square$

Говоря доступным языком,  $\sigma(X)$  - это список событий, про которые мы сможем гарантированно сказать, произошли они или нет, если мы знаем значение  $X$ . Кроме того, если мы различаем события из  $\sigma(X)$ , то мы можем определить значение  $X$ .

Борелевская  $\sigma$ -алгебра - довольно сложный объект, мы даже не можем явно перечислить все входящие в нее множества. Как же на практике проверить, является ли данная случайная величина  $X$  измеримой относительно данной  $\sigma$ -алгебры  $\mathcal{F}$ ?

Оказывается вместо проверки всех борелевских множеств достаточно проверить только некоторые. А именно, достаточно убедиться, что ...

**Теорема 11.3.** Пусть  $\mathcal{H}$  - произвольный набор подмножеств  $\mathbb{R}$  порождающий борелевскую  $\sigma$ -алгебру  $\mathcal{B}$ ,  $\sigma(\mathcal{H}) = \mathcal{B}$ . Случайная величина  $X$  измерима относительно  $\sigma$ -алгебры  $\mathcal{F}$  если и только если для любого  $H \in \mathcal{H}$  событие  $\{X \in H\} \in \mathcal{F}$ .

*Доказательство.* туда:

Пусть  $X$  измерима относительно  $\mathcal{F}$ . Значит прообразы всех борелевских множеств лежат в  $\mathcal{F}$ . так как  $\mathcal{H} \subset \mathcal{B}$ , то и прообразы всех множеств из  $\mathcal{H}$  лежат в  $\mathcal{F}$ .

обратно:

Мы доказывали, что  $\mathcal{D} := \{D | X^{-1}(D) \in \mathcal{F}\}$  -  $\sigma$ -алгебра.

Мы знаем, что  $\mathcal{H} \subset \mathcal{D}$ ,  $\mathcal{D}$  -  $\sigma$ -алгебра и  $\sigma(\mathcal{H}) = \mathcal{B}$ . Значит  $\mathcal{B} \subset \mathcal{D}$   $\square$

Пример..

Хорошая новость:  $\mathcal{F}$ -измеримые случайные величины можно складывать, вычитать, умножать и делить!

**Теорема 11.4.** Если  $X, Y$  -  $\mathcal{F}$ -измеримые случайные величины, то:  $X + Y, X - Y, X \cdot Y$  -  $\mathcal{F}$ -измеримые случайные величины. Если выполнено дополнительное условие  $Y \neq 0$ , то и  $\frac{X}{Y}$  -  $\mathcal{F}$ -измеримая случайная величина.

*Доказательство.* Мы рассмотрим только случай  $X + Y$ , т.к. при доказательстве остальных трех случаев используется та же идея. (...) □

Кроме того, можно смело брать пределы!

**Теорема 12.1.** Если  $X_i$  - последовательность  $\mathcal{F}$ -измеримых случайных величин, то:

a) Множество  $A = \{w | \exists \lim X_i(w)\} \in \mathcal{F}$ .

b) Если для  $\forall w \in \Omega$  существует предел  $\lim X_n(w)$ , то случайная величина  $\lim X_n$  является  $\mathcal{F}$ -измеримой.

*Доказательство.* □

Если преобразовать  $\mathcal{F}$ -измеримую случайную величину с помощью «хорошего» преобразования, то снова получится  $\mathcal{F}$ -измеримая случайная величина. Сначала уточним, что значит «хорошее» преобразование:

**Определение 12.2.** Функция  $f : \mathbb{R} \rightarrow \mathbb{R}$  называется **борелевской** если прообраз любого борелевского множества является борелевским множеством ( $f^{-1}(B) \in \mathcal{B}$  для  $\forall B \in \mathcal{B}$ ).

**Пример 12.3.**  $f(t) = t^2, f(t) = \cos(t), f(t) = |t|$  - борелевские функции.

Более того, все «привычные» функции - борелевские, о чем говорит следующая теорема:

**Теорема 12.4.** Если  $f$  имеет счетное число разрывов, то она - борелевская (... проверить)

(...) Используя пример неборелевского множества, постройте пример неборелевской функции (упр.)

Итак, если использовать «хорошее» преобразование, то  $\mathcal{F}$ -измеримость сохраняется:

**Теорема 12.5.** Если  $X : \Omega \rightarrow \mathbb{R}$  -  $\mathcal{F}$ -измеримая случайная величина, и  $f : \mathbb{R} \rightarrow \mathbb{R}$  - борелевская функция, то  $f(X) : \Omega \rightarrow \mathbb{R}$  -  $\mathcal{F}$ -измеримая случайная величина.

*Доказательство.* □

В случае, когда  $\mathcal{F} = \sigma(X)$  эту теорему можно уточнить:

**Теорема 12.6.** Случайная величина  $Y$  является  $\sigma(X)$ -измеримой если и только если  $Y = f(X)$ , где  $f$  - борелевская функция.

Туда.  $X$  является  $\sigma(X)$ -измеримой. Поэтому согласно (...) если  $f$  - борелевская функция, то  $f(X)$  является  $\sigma(X)$ -измеримой.

Обратно. Интуитивно: если зная  $X$  можно определить, чему равно  $Y$ , то  $Y$  - это функция от  $X$ . Строго:

## 3.2 Сигма-алгебры и случайные величины

На самом деле большинство  $\sigma$ -алгебр связано со случайными величинами. Для иллюстрации приведем три примера:

- Сигма-алгебра, порожденная случайной величиной,  $\sigma(X)$ .
- Сигма-алгебра, связанная с моментом остановки  $T, \mathcal{F}_T$ .
- Остаточная (хвостовая) сигма-алгебра.

...

### 3.3 Обобщение случайных величин

Можно обобщать определение случайной величины по-разному. В любом случае у нас будет следующая картина. Функция  $X : \Omega \rightarrow S$ , где  $S$  - некое (не обязательно числовое) множество. Есть две  $\sigma$ -алгебры:  $\mathcal{F}$ , заданная на  $\Omega$  и  $\mathcal{H}$ , заданная на  $S$ . Чаще всего на  $S$  определено понятие открытого множества и  $\mathcal{H}$  является борелевской  $\sigma$ -алгеброй, т.е. минимальной  $\sigma$ -алгеброй, содержащей все открытые множества.

(...)

Самые популярные обобщения такие:

- Векторные случайные величины.

В  $\mathbb{R}^n$  есть понятие открытого множества. Поэтому есть и понятие борелевской  $\sigma$ -алгебры в  $\mathbb{R}^n$ , т.е. минимальной  $\sigma$ -алгебры, содержащей все открытые множества. Впрочем, принципиально ничего нового не возникает:

**Теорема 13.1.** Вектор  $(X_1, X_2, \dots, X_n)$  измерим относительно  $\sigma$ -алгебры  $\mathcal{F}$  если и только если каждая случайная величина  $X_i$  измерима относительно  $\sigma$ -алгебры  $\mathcal{F}$ .

Хотя «чудеса, леший бродит и русалка на ветвях» сидит даже в  $\mathbb{R}^2$ :

**Теорема 13.2.** Существует измеримое по Лебегу множество в  $\mathbb{R}^2$ , такое, что его проекция на любую координату является неизмеримым в  $\mathbb{R}$ .

*Доказательство.* Пример можно найти... строится он примерно так... □

- Добавление бесконечностей.

Иногда случайная величина может принимать значение плюс или минус бесконечность. Например, время первого выпадения орла при подбрасывании монетки может никогда не наступить и тогда удобно говорить, что оно равно плюс бесконечности. Порою может понадобиться и минус бесконечность.

Как выглядят определения в этом случае? Вместо старого множества значений  $\mathbb{R}$  будет множество  $S = \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$ . В качестве  $\sigma$ -алгебры  $\mathcal{H}$  будем использовать борелевскую  $\sigma$ -алгебру. Она порождается множествами вида  $[-\infty; a]$ , где  $a \in S$ . Например,  $[-\infty; 0)$  открыто,  $(-\infty; 0)$  открыто,  $\{-\infty\}$  замкнуто.

- Комплексные случайные величины,  $X : \Omega \rightarrow \mathbb{C}$ .

В силу того, что комплексная плоскость  $\mathbb{C}$  может быть естественным образом сопоставлена с плоскостью  $\mathbb{R}^2$ , то на  $\mathbb{C}$  есть понятие открытых множеств. Вместо одной комплексной случайной величины можно рассматривать вектор из двух действительных случайных величин. Первая компонента вектора отвечает за действительную часть, вторая - за мнимую. Понятие измеримости для комплексной случайной величины совпадает с измеримостью для вектора.

- Случайные процессы. Об этом позже, в главе ??.

### 3.4 Еще задачи

## 4 Вероятность

### 4.1 Определение вероятности

Вероятность - это размер событий! А именно:

**Определение 13.3.** Пусть  $\mathcal{F}$  -  $\sigma$ -алгебра. Функция  $\mu : \mathcal{F} \rightarrow [0; +\infty]$  называется **мерой**, если выполнены два условия:

M1.  $\mu(\emptyset) = 0$

M2. Если все  $A_i \in \mathcal{F}$  и они попарно не пересекаются ( $A_i \cap A_j = \emptyset$  при  $i \neq j$ ), то  $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$

Обратить внимание следует на то, что мера может принимать бесконечные значения, при этом они складываются по естественным правилам:  $\infty + (\text{любое число}) = \infty$ ,  $\infty + \infty = \infty$ .

**Пример 14.0.** Например,  $\Omega = \mathbb{R}$ ,  $\mathcal{F} = 2^{\mathbb{R}}$ , и  $\mu(A)$  - количество элементов в множестве  $A$ . Легко проверить, что это мера, которая может принимать бесконечные значения.

**Пример 14.1.** Например,  $\Omega = \mathbb{R}$ ,  $\mathcal{F} = 2^{\mathbb{R}}$ , и мера  $\mu(A)$  - равна единице, если множество  $A$  содержит  $\sqrt{3}$ , и равна нулю, если  $A$  не содержит  $\sqrt{3}$ .

**Пример 14.2.**  $\Omega = \mathbb{R}$ ,  $\mathcal{F} = 2^{\mathbb{R}}$ . Функция  $\mu(A) = \sup(A)$  - не является мерой, т.к.  $\mu([0; 1]) \neq \mu([0; 0.5]) + \mu([0.5; 1])$ .

**Пример 14.3.** Длина. Если  $\Omega = \mathbb{R}$ , то на борелевской  $\sigma$ -алгебре  $\mathcal{B}$  можно определить меру  $\lambda$  которая будет соответствовать естественному понятию длины. Например,  $\lambda([0; 100]) = 100$ . К сожалению (?), эту меру нельзя определить на  $\sigma$ -алгебре  $2^{\mathbb{R}}$ . Этот пример содержит в себе сразу два недоказанных утверждения. Во-первых, на  $\mathcal{B}$  можно определить такую меру, там не возникнет проблем с аксиомой M2. Во-вторых, не получится определить на  $2^{\mathbb{R}}$  меру, которая соответствует естественному понятию длины. Доказательство чуть позже.

**Определение 14.4.** Мера  $P$  называется **вероятностью** если  $P(\Omega) = 1$ .

**Пример 14.5.** (простенький пример с конечным  $\Omega$ )

В данном курсе мы будем иметь дело с вероятностью, которая является конечной мерой и с длиной подмножеств  $\mathbb{R}$ , которая является  $\sigma$ -конечной мерой.

**Определение 14.6.** Мера  $\mu$  называется конечной, если  $\mu(\Omega) < \infty$

**Определение 14.7.** Мера  $\mu$  называется  $\sigma$ -конечной, если  $\Omega$  можно разбить на счетное количество непересекающихся  $\Omega_i$ , так что на каждом  $\Omega_i$  мера  $\mu$  будет конечной:  $\Omega = \bigcup \Omega_i$ ,  $\mu(\Omega_i) < \infty$ .

Длина является  $\sigma$ -конечной в силу того, что числовую прямую можно разбить на счетное количество полуинтервалов единичной длины вида  $[n; n + 1)$ .

## Задачи

**Задача 4.1.** Пусть  $\mu_1$  и  $\mu_2$  - конечные меры на  $\Omega$  и  $a > 0$ . Верно ли, что: а)  $\mu := a\mu_1$  - конечная мера? б)  $\mu := \mu_1 + \mu_2$  - конечная мера?

**Задача 4.2.** Пусть  $\mu$  - конечная мера на  $\Omega$ . Верно ли, что  $\nu(A) := \frac{\mu(A)}{\mu(\Omega)}$  - вероятность на  $\Omega$

**Задача 4.3.** Приведите пример  $\sigma$ -алгебры  $\mathcal{F}$  и функции  $R : \mathcal{F} \rightarrow [0; 1]$  такой, что:

1.  $R(\Omega) = 1$ ,  $R(\emptyset) = 0$
2. Если  $A \in \mathcal{F}$ ,  $B \in \mathcal{F}$  и  $A \cap B = \emptyset$ , то  $R(A \cup B) = R(A) + R(B)$
3. Существует возрастающая последовательность  $A_1 \subset A_2 \subset A_3 \subset \dots$  такая, что  $R(\bigcup A_n) \neq \lim R(A_n)$

## 4.2 Свойства вероятности

Нам часто понадобятся следующие свойства:

**Теорема 14.8.** Если  $P$  - вероятность, то:

PP1.  $P(\bigcup A_i) \leq \sum P(A_i)$ .

PP2. Если  $P(A_i) = 0$ , то  $P(\bigcup A_i) = 0$ .

PP3. Если  $P(A_i) = 1$ , то  $P(\bigcap A_i) = 1$ .

PP4. Вероятность выдерживает взятие пределов:

PP4a. Если  $A_1 \subset A_2 \subset A_3 \dots$ , то  $P(\bigcup A_i) = \lim P(A_i)$

PP4b. Если  $A_1 \supset A_2 \supset A_3 \dots$ , то  $P(\bigcap A_i) = \lim P(A_i)$

PP4c. Если  $\lim A_i$  существует, то  $P(\lim A_i) = \lim P(A_i)$

**Доказательство.** PP1: Воспользуемся следующим трюком, от последовательности  $A_n$  перейдем к последовательности «добавок»  $A'_n$ . То есть  $A'_n$  - это те новые элементы, которые приносит с собой  $A_n$ , которых еще нет в  $\bigcup_{i=1}^{n-1} A_i$ . Формально,  $A'_i = A_i \setminus (\bigcup_{i=1}^{n-1} A_i)$ . В силу этого  $A'_n$  попарно не пересекаются.

Остается заметить, что  $\mathbb{P}(A'_n) \leq \mathbb{P}(A_n)$  и  $\mathbb{P}(\cup A_n) = \mathbb{P}(\cup A'_n) = \sum \mathbb{P}(A'_n) \leq \sum \mathbb{P}(A_n)$ .

PP2: Следует из PP1.

PP3: Перейдем к дополнениям,  $B_n = A_n^c$ . Тогда  $\mathbb{P}(B_n) = 0$  и к ним применимо свойство PP2,  $\mathbb{P}(\cup B_n) = 0$  и  $\mathbb{P}((\cup B_n)^c) = 1$ . Но это как раз и есть желаемое:  $(\cup B_n)^c = (\cup A_n^c)^c = \cap A_n$ .  $\square$

Еще несколько «именных» свойств, связанных с множеством  $\limsup A_i$  («произошло бесконечное количество  $A_i$ »):

Лемма Фату для вероятностей: если есть бесконечно много  $A_i$  с вероятностью не ниже  $p$ , то вероятность того, что произойдет бесконечное количество  $A_i$  не ниже  $p$ .

**Теорема 15.1.**  $\mathbb{P}(\limsup A_i) \geq \limsup \mathbb{P}(A_i)$ .

*Доказательство.* Шаг 1. Среди  $A_i$  выбираем подпоследовательность  $B_k$  такую, что  $\lim \mathbb{P}(B_k) = \limsup \mathbb{P}(A_i)$ . Заметим, что  $\limsup B_i \subset \limsup A_i$  (если произошло бесконечное количество  $B_i$ , то произошло бесконечное количество  $A_i$ , но обратное в общем случае неверно),  $\mathbb{P}(\limsup A_i) \geq \mathbb{P}(\limsup B_i)$ .

Шаг 2.  $\mathbb{P}(\limsup B_i) = \mathbb{P}(\cap_{n=1}^{\infty} \cup_{k \geq n} B_k) = \lim_{n \rightarrow \infty} \mathbb{P}(\cup_{k \geq n} B_k) \geq \lim_{n \rightarrow \infty} \mathbb{P}(B_n)$ .  $\square$

Первая лемма Бореля-Кантелли: если сумма вероятностей  $A_i$  не «велика», то вероятность того, что произойдет бесконечное количество  $A_i$  равна нулю. Если быть точным, «не велика» в данном случае - меньше бесконечности:

**Теорема 15.2.** Если  $\sum \mathbb{P}(A_i) < \infty$ , то  $\mathbb{P}(\limsup A_i) = 0$ .

*Доказательство.* Шаг 1.  $\mathbb{P}(\limsup A_i) = \mathbb{P}(\cap_{n=1}^{\infty} \cup_{k \geq n} A_k) = \lim_{n \rightarrow \infty} \mathbb{P}(\cup_{k \geq n} A_k) \leq \lim_{n \rightarrow \infty} \sum_{k \geq n} \mathbb{P}(A_k)$ .

Шаг 2. Частичные суммы  $\sum_{i=1}^n \mathbb{P}(A_i)$  сходятся к сумме  $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$ , а это значит, что непосчитанный «хвост» стремится к нулю,  $\lim_{n \rightarrow \infty} \sum_{k \geq n} \mathbb{P}(A_k) = 0$ .  $\square$

### 4.3 Построение равномерной вероятности

Оказывается, что некоторые интересные вероятности нельзя определить для произвольного подмножества числовой прямой!!!

Очень важный пример!

#### У некоторых множеств на прямой нет длины!

**Пример 15.3.** Допустим мы хотим создать равномерное распределение на интервале  $[0; 1)$ . Казалось бы, чего проще! Берем в качестве  $\mathcal{F}$  все подмножества данного полуинтервала и в качестве вероятности попасть в данное подмножество берем его длину! Однако это не получится! Обязательно будет нарушена аксиома M2 в определении меры.

Давайте разберемся. Чего мы ждем от равномерного распределения? Если взять произвольное подмножество  $A$  и его копию  $A'$  сдвинутую, скажем чуть правее, то мы хотим, чтобы  $\mathbb{P}(A) = \mathbb{P}(A')$ . И все. Больше на ничего от равномерной вероятности ничего не нужно.

Для краткости будем использовать следующее обозначение:

**Определение 15.4.**  $A \oplus r$  - это множество  $A$  сдвинутое вправо на  $r$ . Если при этом результат вылезает за пределы интервала  $[0; 1)$ , то мы выступающую часть «внесем» обратно в  $[0; 1)$  с другой стороны. Например:  $[0; 0, 5] \oplus 0, 1 = [0, 1; 0, 6]$ ,  $(0, 5; 0, 8) \oplus 0, 3 = (0, 8; 1) \cup [0; 0, 1)$ .

А сейчас мы построим довольно «хитрое» множество  $A$ . Для этого мы раскрасим весь интервал  $[0; 1)$  в разные цвета по следующему принципу:

Числа  $x \in [0; 1)$  и  $y \in [0; 1)$  красим в один цвет, если существует рациональное число  $r \in \mathbb{Q}$ , такое что  $x \oplus r = y$ .

Из этого принципа следует, что если взять любое число  $x$ , то есть счетное количество покрашенных в тот же цвет чисел. Так, например, есть счетное количество чисел покрашенных в тот же цвет, что и  $\sqrt{2}/2$ . Заметим, что все рациональные числа покрашены в тот же цвет, что и число 0.

Множество  $A$  создадим так: возьмем по одному числу каждого цвета. Например, от рациональных можно взять ноль, от покрашенных в тот же цвет, что и  $\sqrt{2}/2$  можно взять  $\sqrt{2}/2 + 0, 1$  и т.д. Неважно какого представителя мы берем, главное от каждого цвета ровно по одному.

Что у нас получилось? Если взять рациональное  $r \in [0; 1)$ , то  $A \oplus r$  не совпадает с  $A$ , т.к. каждый представитель своего цвета превратится в другое число своего цвета. Если взять два разных рациональных числа  $r_1 \in [0; 1)$ ,  $r_2 \in [0; 1)$ , то  $A \oplus r_1$  не пересекается с  $A \oplus r_2$ : в каждой группе представитель переходит в разных членов своей группы. Если перебрать все рациональные числа, то  $\bigcup_{r \in \mathbb{Q} \cap [0; 1)} (A \oplus r) = [0; 1)$ , в каждой группе представитель успеет побывать в роли каждого члена своей группы.

Здесь и рождается противоречие! Напомню, что мы хотим, чтобы  $\mathbb{P}(A) = \mathbb{P}(A \oplus r)$ :

С одной стороны,  $\mathbb{P}(\bigcup_{r \in \mathbb{Q} \cap [0; 1)} (A \oplus r)) = \mathbb{P}([0; 1)) = 1$

С другой стороны,  $\mathbb{P}(\bigcup_{r \in \mathbb{Q} \cap [0; 1)} (A \oplus r)) = \sum_{r \in \mathbb{Q} \cap [0; 1)} \mathbb{P}(A \oplus r) = \sum_{r \in \mathbb{Q} \cap [0; 1)} \mathbb{P}(A)$ .

Но сумма бесконечного количества одинаковых слагаемых не может равняться единице! Либо нулю, если складываются нули, либо бесконечности, если складываются ненулевые числа.

Этот пример показал, что равномерную вероятность нельзя построить для произвольных подмножеств отрезка  $[0; 1)$ . Поскольку от понятия длины мы ждем ровно той же неизменности при сдвиге вправо-влево на прямой, становится ясно, что нельзя присвоить понятие длины произвольному подмножеству  $\mathbb{R}$ .

Что же делать? Нужно отказаться от определения вероятности на этом «страшном» множестве  $A$ ! Т.е. для некоторых множеств вероятность определена и является числом от нуля до единицы и подчиняется свойствам M1, M2, а для некоторых подмножеств вероятность просто не существует. В данном случае  $\mathbb{P}(A \oplus r)$  и  $\mathbb{P}(A)$  не будут существовать и противоречия не возникнет.

Возникает естественный вопрос: каким же подмножествам интервала  $[0; 1)$  можно присвоить вероятность, которая не менялась бы при сдвиге? Или каким подмножествам прямой можно присвоить понятие длины?

## Длина есть у всех борелевских множеств!

Нас спасет уже знакомая нам борелевская  $\sigma$ -алгебра  $\mathcal{B}([0; 1])$ ! На ней можно определить равномерную вероятность. А как мы видели эта  $\sigma$ -алгебра достаточно богата и включает все подмножества отрезка  $[0; 1]$  с которыми реально приходится работать.

Борелевских множеств много и они могут иметь достаточно сложную структуру, поэтому попытка явно определить «длину» каждого борелевского множества обречена на провал. Мы поступим следующим путем. Определим понятие «длины» на очень простом наборе множеств. Этот набор множеств будет порождать борелевскую  $\sigma$ -алгебру. И сформулируем теорему, которая будет гарантировать, что у каждого борелевского множества есть длина.

Итак, наш «простой» набор множеств это полуинтервалы вида  $[a; b)$ :

$$\mathcal{H} = \{[a; b) \mid 0 \leq a < b < 1\}.$$

Он, конечно,  $\sigma$ -алгеброй на  $\Omega = [0; 1)$  не является:  $[0.1; 0.2) \in \mathcal{H}$ , но  $[0.1; 0.2)^c \notin \mathcal{H}$ . Зато на нем тривиально определить длину:

$$\lambda([a; b)) := b - a \tag{16.1}$$

Кроме того, этот набор порождает борелевскую  $\sigma$ -алгебру, т.е.  $\sigma(\mathcal{H}) = \mathcal{B}([0; 1))$ .

Формально наш набор  $\mathcal{H}$  является полу-алгеброй:

### Определение 16.2.

Спасительная теорема формулируется так:

**Теорема 16.3** (Каратеодори). *Если функция  $\mu$  задана на полу-алгебре  $\mathcal{H}$  и на ней является  $\sigma$ -аддитивной ( $\sigma$ ), то существует единственная мера  $\mu^*$ , заданная на  $\sigma$ -алгебре  $\sigma(\mathcal{H})$ , совпадающая с  $\mu$  на множестве  $\mathcal{H}$ .*



*Доказательство.* Идея проста. Для каждого подмножества  $\Omega$  определим «внешнюю меру»  $\mu^*$ . Эта «внешняя мера»  $\mu^*$  не всегда будет  $\sigma$ -аддитивна, т.е. она не является мерой! Но если выбрать «хорошие» подмножества  $\Omega$ , а именно такие, на которых  $\mu^*$  будет мерой, то они будут образовывать  $\sigma$ -алгебру! Единственность будет следовать из того, что каждое «хорошее» множество можно «приблизить» с помощью объединения множеств из  $\mathcal{H}$ , а на них любая другая мера должна совпадать с  $\mu^*$ . Полное доказательство в приложении.  $\square$

Заметим, что конкретно в нашем случае  $\sigma$ -алгебра «хороших» окажется строго больше  $\mathcal{B}([0; 1]) = \sigma(\mathcal{H})$ . Длина будет определена не только у борелевских множеств, но и у некоторых других. Это не страшно, и подробнее об этом чуть позже.

**Пример 17.1.** <http://math.stackexchange.com/questions/2949/which-one-result-in-maths-has-2953> Пример, как можно каждое рациональное число накрыть маленьким интервальчиком и в результате длина всего множества будет конечная..

## Как построить любой случайный эксперимент?

Почему так важно было установить существование равномерной вероятности? Оказывается, что имея одну равномерную случайную величину можно получить бесконечное количество независимых равномерных случайных величин! А имея бесконечно много равномерных случайных величин можно такого понастроить...

Пусть  $X$  - равномерно распределенная случайная величина на  $[0; 1)$ . Мы хотим на базе одной  $X$  построить последовательность независимых равномерных  $X_1, X_2, X_3 \dots$

Нам окажется полезной такая неслучайная последовательность натуральных чисел:

$$[a_i] = [1, 1, 2, 1, 2, 3, 1, 2, 3, 4, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 6 \dots] \quad (17.2)$$

Сначала сказали 1, потом посчитали от 1 до 2, потом от 1 до 3, потом от 1 до 4 и т.д. Эта последовательность нам будет говорить какому  $X_i$  «отдать» очередную цифру из десятичной записи  $X$ .

Например, пусть  $X(w) = 0,583601935 \dots$ . Первую цифру после запятой отдаем  $X_1$ , вторую - снова  $X_1$ , третью -  $X_2$ , четвертую - снова  $X_1$ , пятую -  $X_2$ , шестую -  $X_3 \dots$  Получаем:

$$X_1 = 0,5869 \dots$$

$$X_2 = 0,303 \dots$$

$$X_3 = 0,15 \dots$$

Изобразим на картинке:

$$X(w) = 0,583601935 \dots$$

$$a_i = 1121331234 \dots$$

$$X_1 = 0,58 \ 6 \ 9 \dots$$

$$X_2 = 0, \quad 3 \ 0 \ 3 \dots$$

$$X_3 = 0, \quad 1 \ 5 \dots$$

В последовательности  $a_i$  каждое натуральное число  $k$  упомянуто бесконечное количество раз. Значит каждая случайная величина  $X_k$  получит от  $X$  бесконечное количество цифр. Разные цифры в десятичном разложении величины  $X$  независимы, отсюда будет следовать независимость  $X_k$ . Каждая цифра в десятичном разложении  $X$  равновероятно принимает значения от 0 до 9, значит и каждая цифра в десятичном разложении каждого  $X_k$  также равновероятно принимает значения от 0 до 9. Следовательно, все  $X_k$  равномерны на  $[0; 1)$ .

Имея последовательность независимых равномерных  $X_k$  можно получить независимую последовательность  $Y_k$  с произвольным законом распределения. А в главе про броуновское

движение мы узнаем, как из последовательности нормальных случайных величин сварить это самое броуновское движение.

Упр. докажите, что если  $X$  - равномерное, то ... - экспоненциальное, а ... - пуассон.

#### 4.4 Почти на верное и пополнение вероятностного пространства!

неборелевское - не лежит в  $\mathcal{B}$ . неизмеримое (или, точнее, не измеримое по Лебегу, не лебеговское) - не лежит в пополнении  $\mathcal{B}$ .

#### 4.5 Про «хвостовую» сигма-алгебру. Для опоздавших.

Пусть кто-то подбрасывает монетку бесконечное количество раз. Раз за разом, скажем одно подбрасывание в минуту. Введем случайную величину  $X_i$ , которая будет равна 1, если в  $i$ -ый раз выпала решка и 0, если орел. Величины  $X_1, X_2, \dots$  - независимы и одинаково распределены.

Известно, что Вася (любитель подольше поспать) опоздал к началу эксперимента. Опоздал на неизвестное нам время. То есть он наблюдает значения  $X_i$  только начиная с некоторого номера  $T$ .

Есть ли какой-то список событий, которые Вася различает вне зависимости от того, на сколько минут он опоздал?

Казалось бы ответ «нет». Ни про одну из  $X_i$  нельзя быть уверенным в том, что она известна Васе. Он мог опоздать на 10 минут и не знать  $X_1, \dots, X_9$ . А мог опоздать на 1000 минут и не знать значений  $X_1, \dots, X_{999}$ .

Однако это не так! На сколько бы Вася не опоздал, он всегда сможет сказать, произошли ли события «начиная с некоторого момента были только орлы» или «орел сменялся на решку бесконечное количество раз».

Насколько бы Вася не опоздал он станет свидетелем целой кучи событий!

**Определение 18.1.** Пусть  $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \dots)$  -  $\sigma$ -алгебра событий различимых наблюдателем, пришедшем в момент времени  $n$ . **Остаточной**  $\sigma$ -алгеброй называется  $\sigma$ -алгебра  $\bigcap_n \mathcal{F}'_n$ .

(упр) Приведите еще три примера событий, наблюдателем которых станет Вася, вне зависимости от величины опоздания.

Правда все события из этой  $\sigma$ -алгебры в каком-то смысле тривиальны и ради них не стоило и торопиться. Можно было не выходя из дома, сказать произойдут они или нет (почти на верное)!

**Теорема 18.2** (Закон 0-1 Колмогорова). Если  $X_i$  - последовательность независимых одинаково распределенных случайных величин, и  $\mathcal{H}$  - остаточная  $\sigma$ -алгебра, то для любого события  $H \in \mathcal{H}$  вероятность  $\mathbb{P}(H) = 0$  или  $\mathbb{P}(H) = 1$ .

Доказательство. □

**Пример 18.3.** Рассмотрим событие  $A$  — «последовательность «Решка-Решка-Орел» встречается бесконечное количество раз». Это событие входит в остаточную  $\sigma$ -алгебру, поэтому его вероятность либо 0, либо 1. ... В результате она равна 1.

Примеры событий из хвостовой сигма-алгебры: <http://mathoverflow.net/questions/2904/>

#### 4.6 Еще задачи

**Задача 4.4.** Можно ли придумать последовательность  $\sigma$ -алгебр  $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \dots$ , такую, что одновременно выполнены два условия: 1) для любого  $i$ :  $\mathcal{F}_i \subset \mathcal{F}_{i+1}$ ; 2)  $\bigcup \mathcal{F}_i$  - не  $\sigma$ -алгебра.

## 5 Зоопарк!

### 5.1 Канторово множество - дикое но симпатичное

Пример достаточно «сложного» борелевского множества - Канторово множество. Строится оно следующим образом...

Для начала возьмем отрезок  $C_0 = [0; 1]$ :

(...) рисунок

Затем вырежем из него середину, получим  $C_1 = [0; 1/3] \cup [2/3; 1]$ .

(...) рисунок

Затем из каждого отрезка вырежем его середину, получим  $C_2 = [0; 1/9] \cup [2/9; 3/9] \cup [6/9; 7/9] \cup [8/9; 1]$ .

(...) рисунок

Продолжая процесс вырезания середины каждого отрезка бесконечно долго мы получим последовательность вложенных множеств  $C_0 \supset C_1 \supset C_2 \supset C_3 \supset \dots$

**Определение 19.1.** Канторово множество -  $C = \bigcap_{i=1}^{\infty} C_i$ .

Может показаться, что кроме концов отрезков там ничего не остается, однако это не так.

**Пример 19.2.** (проверить) точка  $1/4$  не является концом отрезка, но лежит в каждом  $C_i$ , а следовательно, лежит в  $C$ .

Каждое  $C_i$  лежит в борелевской  $\sigma$ -алгебре, а значит и их счетное пересечение, Канторово множество лежит в борелевской  $\sigma$ -алгебре. Раз Канторово множество является борелевским, значит у него есть длина (согласно (...) на борелевских множествах можно определить классическое понятие длины). Суммарная длина выкинутых отрезков равна единице:  $1/3 + 2/9 + 4/27 + \dots = \frac{1/3}{1-2/3} = 1$ . Значит Канторово множество имеет длину 0.

Несмотря на свою нулевую длину Канторово множество несчетно! точнее говоря, имеет мощность континуум. Существует взаимно однозначное соответствие между множеством бесконечных вправо последовательностей из 0 и 1 и точками Канторова множества. Берем произвольную точку в Канторовом множестве: в  $I_1$  она попадает либо в правый отрезок, либо в левый (0 или 1); в рамках выбранного отрезка, в  $I_2$  она попадает либо в правый отрезок, либо в левый (0 или 1) и т.д. И наоборот, если задана последовательность из 0 и 1, то ее можно переделать в последовательность «вправо»-«влево», а каждой такой последовательности сопоставить единственную точку. На  $\mathbb{R}$  любая последовательность Коши имеет предел, поэтому получается взаимнооднозначное соответствие.

Внутри него можно найти неборелевское! Существование неборелевского множества внутри Канторова можно установить путем разных рассуждений. Например, неконструктивно: подмножеств Канторова множества больше, чем континуум, а борелевских (...) (ссылка еще раз) - континуум. А можно построить в явном виде, что мы сделаем с помощью Канторовой лестницы (...).

Оно совершенно. Каждая точка является пределом последовательности из него.

Оно нигде не плотно.

**Определение 19.3.** Подмножество  $A \subset \mathbb{R}$  называется **нигде не плотным**, если (...)

Почему оно так важно?

Во-первых, оно чудесным образом сочетает характеристики, которые на первый взгляд кажутся противоречивыми (несчетность и нулевую длину, несчетность и нигде не плотность). «Чудо», коротко говоря.

Во-вторых, характеристики Канторова множества типичны для множества нулей броуновского движения. Если взять типичную (выпадающую с вероятностью 1) траекторию броуновского движения, и посмотреть, где она пересекает горизонтальную ось, то это множество нулей будет несчетным, нулевой длины, нигде не плотным и совершенным. Подробнее об этом в главе про броуновское движение.

## 5.2 Канторова лестница

## 5.3 Парадокс Банаха-Тарского

## 5.4 Еще один пример неборелевского множества

Мы приведем еще один пример с целью свыкнуться с мыслью: неборелевских множеств очень много, но на них нелегко наткнуться! Неизмеримые множества - это очень странные математические объекты, способные поражать воображение. Но они лишены физического смысла! Легко отрезать веревку длиной (хотя бы примерно)  $\sqrt{2}$ , но разрезать арбуз так, чтобы из кусков получилось два арбуза (хотя бы примерно) не получится!

Каждое действительное число можно представить в виде непрерывной дроби:

$$x = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\ddots}}}, \text{ где } a_0 - \text{целое, а остальные } a_i - \text{натуральные числа.}$$

(представление иррациональных чисел - единственно, у каждого рационального числа есть два представления, более короткое считается каноническим).

Пусть  $A$  - те числа, для которых непрерывная дробь обладает условием: последовательность целых чисел  $a_i$  содержит в себе подпоследовательность, где каждый последующий член делится на предыдущий.

**Теорема 20.1.**  $A$  - неборелевское, однако  $A$  - измеримо по Лебегу.

*Доказательство.* Lusin (...)

□

## 5.5 Еще задачи

**Задача 5.1.** У нас есть счетное количество пленников. Пленники занумерованы натуральными числами и каждый из них видит цвета колпаков пленников с большим номером. Каждому пленнику надевают колпак, равновероятно белый или черный. У каждого пленника есть одна попытка угадать цвет своего колпака. Пленники заранее могут договориться об общей стратегии. Уточнение: никто из пленников не знает, ни что говорили предыдущие, ни угадали ли они. Сколько пленников можно гарантированно спасти?

**Задача 5.2.** Верно ли, что  $1/4$  лежит в Канторовом множестве? Верно ли, что  $1/4$  является концом одного из выброшенных интервалов?

# 6 Математическое ожидание - интеграл Лебега

## 6.1 Интеграл Лебега

Математическое ожидание - это синоним интеграла Лебега. Единственное отличие в том, что интеграл Лебега иногда считают относительно произвольной меры, а математическое ожидание - только относительно вероятности.

Нам потребуется строгое определение математического ожидания, подходящее для любой случайной величины. Мы будем строить математическое ожидание пошагово. Начнем с простых случайных величин. Затем рассмотрим произвольные неотрицательные, затем действительные. И, когда это требуется, комплексные.

Итак, перед нами четыре объекта:  $\Omega$  - множество исходов  $\mathcal{F}$  -  $\sigma$ -алгебра событий  $P$  - вероятность, определенная для событий из  $\mathcal{F}$   $X$  - случайная величина, измеримая относительно  $\mathcal{F}$  Сразу уточним, что все рассматриваемые случайные величины должны быть измеримы относительно  $\mathcal{F}$ .

**Определение 20.2.** Случайная величина  $X$  называется простой, если принимает конечное число значений.

Если случайная величина простая, то ее можно представить в виде:  $X = \sum_i^n x_i 1_{A_i}$ , где  $A_i$  не пересекаются. Таких представлений может быть несколько (можно множество  $A_1$ , где  $X$  равна, скажем 5, разбить на два подмножества).

Шаг 1. Простая неотрицательная случайная величина.

**Определение 21.1.** Пусть  $X$  - простая неотрицательная случайная величина, т.е.  $X = \sum_{i=1}^n x_i 1_{A_i}$ . Интегралом Лебега по мере  $P$  (или математическим ожиданием) называется число

$$\mathbb{E}(X) := \int X dP := \sum_{i=1}^n x_i \mathbb{P}(A_i) \quad (21.2)$$

В определении неявно задействована измеримость. Вероятность  $P$  определена только для событий из некоей  $\sigma$ -алгебры  $\mathcal{F}$ , поэтому чтобы  $\mathbb{P}(A_i)$  существовали необходимо, чтобы случайная величина  $X$  была измерима относительно этой  $\sigma$ -алгебры  $\mathcal{F}$ . Легко заметить, что определение корректно, т.к. не зависит от выбранного разбиения.

Кроме того, выполняются следующие свойства.

Если  $X, Y$  - простые неотрицательные случайные величины и  $a \geq 0$  - константа, то:

1.  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ .
2.  $\mathbb{E}(aX) = a\mathbb{E}(X)$ .
3. Если  $X \geq Y$ , то  $\mathbb{E}(X) \geq \mathbb{E}(Y)$ .

Шаг 2. Произвольная неотрицательная случайная величина.

**Определение 21.3.** Пусть  $X$  - произвольная неотрицательная случайная величина. Интегралом Лебега по мере  $P$  (или математическим ожиданием) называется величина

$$\mathbb{E}(X) := \int X dP := \sup_{Z \leq X, Z \text{ - simple}} \mathbb{E}(Z) \quad (21.4)$$

Т.е. мы перебираем все простые случайные величины  $Z$ , не превосходящие  $X$  и выбираем наибольшее из получающихся математических ожиданий. При этом нужно отметить, что в результате может получиться уже не число, а плюс бесконечность. Также заметим, что новое определение не противоречит старому, т.к. если изначально  $X$  - простая, то взяв  $Z := X$  мы получим старое значение математического ожидания, а больше получить невозможно в силу неравенства (...).

Все те же свойства:

Шаг 3. Произвольная случайная величина.

Любую случайную величину  $X$  всегда можно представить в виде  $X = X^+ - X^-$ , где  $X^+ = \max\{X, 0\}$  - неотрицательная часть  $X$ , а  $X^- = -\min\{X, 0\}$  - неположительная часть  $X$ , домноженная на (-1). Величины  $X^+$  и  $X^-$  являются неотрицательными, а для неотрицательных у нас уже мат. ожидание придумано.

Картинка [...]

**Определение 21.5.** Пусть  $X$  - произвольная случайная величина. Интегралом Лебега по мере  $P$  (или математическим ожиданием) называется величина

$$\mathbb{E}(X) := \int X dP := \mathbb{E}(X^+) - \mathbb{E}(X^-) \quad (21.6)$$

,

PE1.

PE2.

PE3. Если  $X \geq 0$ , то  $\mathbb{E}(X) \geq 0$ .

PE4. Если  $X \geq 0$  и  $\mathbb{E}(X) = 0$ , то  $\mathbb{P}(X = 0) = 1$ . (часто используется)

PE5. Если  $X \geq 0$  и  $\mathbb{E}(X) < \infty$ , то  $\mathbb{P}(X = \infty) = 0$ .

Обобщение для комплексных случайных величин!

## 6.2 MCT, DCT, Fatou's lemma

**Теорема 21.7.** Если:

1.  $X_n$  - неубывающая последовательность неотрицательных случайных величин измеримых относительно  $\mathcal{F}$ ,  $X_1 \leq X_2 \leq X_3 \leq \dots$

2.  $\lim X_n = X$

То:

1.  $X$  - измерима относительно  $\mathcal{F}$

2.  $\lim \mathbb{E}(X_n) = \mathbb{E}(X)$

Теорема учитывает случай, когда предел с обеих сторон равен плюс бесконечности.

*Доказательство.* Во-первых, заметим, что  $\mathbb{E}(X_n) \leq \mathbb{E}(X_{n+1})$ . Т.е. предел слева обязательно существует, хотя возможно и равен плюс бесконечности.

Случай 1. Среди  $X_n$  есть такая величина  $X_k$ , что  $\mathbb{E}(X_k) = \infty$ .

Если  $\mathbb{E}(X_k) = \infty$ , то для любого числа  $M$  найдется такая простая случайная величина  $S$ , что  $S \leq X_k$ , но  $\mathbb{E}(S) > M$ . Замечаем, что  $X \geq X_k \geq S$ , а значит,  $\mathbb{E}(X) \geq \mathbb{E}(S) > M$ . Но  $M$  произвольное число, значит  $\mathbb{E}(X) = \infty$ .

Случай 2. Все  $\mathbb{E}(X_n)$  конечны, но  $\mathbb{E}(X_n) \rightarrow \infty$ .

Случай 3. Все  $\mathbb{E}(X_n)$  конечны, и  $\mathbb{E}(X_n) \rightarrow \text{const}$ .

□

**Теорема 22.1.** Если:

1. Последовательность  $X_n$  ограничена интегрируемой случайной величиной  $Y$ , т.е.  $|X_n| \leq Y$  и  $\mathbb{E}(Y) < \infty$ .

2.  $\lim X_n = X$

То:

$\lim \mathbb{E}(X_n) = \mathbb{E}(X)$

*Доказательство.*

□

Если применить лемму Фату к индикаторам, то получается лемма Фату для вероятностей.

Добавить Витали (?) Vitali convergence theorem

## 6.3 Неравенства, леммы Бореля Кантелли

Еще раз про леммы Бореля-Кантелли.

Первая лемма Бореля-Кантелли: если сумма вероятностей  $A_i$  конечна, то вероятность того, что произойдет бесконечное количество  $A_i$  равна нулю. теперь можно дать другое прочтение этой формуле. Пусть  $N$  - количество произошедших  $A_i$ , и случайная величина  $N$  может принимать значение  $+\infty$ . Если среднее значение  $N$  конечно, то вероятность того, что  $N = \infty$  равна нулю.

**Теорема 22.2.** Если  $\sum \mathbb{P}(A_i) < \infty$ , то  $\mathbb{P}(\limsup A_i) = 0$ .

*Доказательство.*  $\mathbb{E}(N) = \mathbb{E}(\sum_i 1_{A_i})$ . Применяя MCT получаем, что  $\mathbb{E}(N) = \sum \mathbb{E}(1_{A_i}) = \sum \mathbb{P}(A_i) < \infty$ . А значит (используя PE5) и  $\mathbb{P}(N = \infty) = 0$ .

□

Вторая лемма Бореля-Кантелли: если  $A_i$  независимы (или отрицательно коррелированы) и сумма вероятностей  $A_i$  «велика» (равна бесконечности), то вероятность того, что произойдет бесконечное количество  $A_i$  равна 1. Обобщение на отрицательно зависимые случайные величины заимствовано из [ross:scpr]

(самое время сделать упр. про  $\text{Cov}(1_A, 1_B) = 0 \Leftrightarrow$  независимость)

**Теорема 22.3.** Если события  $A_i$  независимы или неположительно коррелированы ( $\text{Cov}(1_{A_i}, 1_{A_j}) \leq 0$  для  $\forall i, j$ ) и  $\sum \mathbb{P}(A_i) = \infty$ , то  $\mathbb{P}(\limsup A_i) = 1$ .

*Доказательство.* Пусть  $N_k$  - число произошедших событий среди первых  $k$  событий,  $N_k = 1_{A_1} + 1_{A_2} + \dots + 1_{A_k}$ .

Шаг 1.  $Var(N_k) \leq \mathbb{E}(N_k)$ :  $Var(N_k) \leq \sum Var(1_{A_i})$  (т.к. ковариация либо нулевая, либо отрицательная),  $\sum Var(1_{A_i}) = \sum \mathbb{P}(A_i)(1 - \mathbb{P}(A_i)) \leq \sum \mathbb{P}(A_i) = \mathbb{E}(N_k)$ .

Шаг 2. Рассмотрим произвольное число  $x < \mathbb{E}(N_k)$ . Для такого  $x$ :  $\mathbb{P}(N_k < x) = \mathbb{P}(\mathbb{E}(N_k) - N_k > \mathbb{E}(N_k) - x) \leq \mathbb{P}(|\mathbb{E}(N_k) - N_k| > \mathbb{E}(N_k) - x) \leq \frac{Var(N_k)}{(\mathbb{E}(N_k) - x)^2} \leq \frac{\mathbb{E}(N_k)}{(\mathbb{E}(N_k) - x)^2}$ .

Шаг 3. Рассмотрим произвольное число  $x$ . В силу того, что  $\mathbb{E}(N_k) \rightarrow \mathbb{E}(N) = \infty$ , наступит такой момент, что неравенство из шага 2 начнет работать. А значит  $\lim_{k \rightarrow \infty} \mathbb{P}(N_k < x) = 0$ . Но,  $\mathbb{P}(N < x) \leq \mathbb{P}(N_k < x)$ , значит для  $\forall x$ , вероятность  $\mathbb{P}(N < x) = 0$ .

Шаг 4. Для множеств действует предел  $\lim(N < k) = N < \infty$ , вероятность непрерывна, а значит  $\mathbb{P}(N < \infty) = 0$ . □

Неравенство Чебышева (Маркова).  $\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}$ .

*Доказательство.* □

Неравенство Йенсена. Если  $f: \mathbb{R} \rightarrow \mathbb{R}$  - выпуклая функция (уточним, выпуклая вниз, как, например,  $f(t) = t^2$ ), то  $\mathbb{E}(f(X)) \geq f(\mathbb{E}(X))$ .

*Доказательство.* □

Мне тут попало малоизвестное забавное неравенство. Будем его популяризировать!

Неравенство «Раз-два-три»! Если  $X$  и  $Y$  независимые случайные величины, то  $\mathbb{E}(|X - Y| < 2) < 3\mathbb{P}(|X - Y| < 1)$ .

Noga Alon, The 123 theorem and its extensions. Или в упражнения (да, оно дальше не используется)?

## 6.4 Производная Радона-Никодима

Оказывается, если из одной вероятности легко получить новую!

**Теорема 23.1.** Если  $X$ -неотрицательная случайная величина и  $\mathbb{E}(X) > 0$ , то функция  $Q: \mathcal{F} \rightarrow [0; 1]$ , определяемая по формуле:  $Q(A) = \frac{\mathbb{E}(1_A X)}{\mathbb{E}(X)}$  является вероятностью.

*Доказательство.* Легко убедиться, что  $Q(\emptyset) = \frac{\mathbb{E}(0)}{\mathbb{E}(X)} = 0$  и  $Q(\Omega) = \frac{\mathbb{E}(X)}{\mathbb{E}(X)} = 1$ . Осталось проверить счетную аддитивность (аксиому M2 в определении меры).

Пусть  $A_i$  - попарно непересекающиеся события. Определим  $Y_n = \sum_{i=1}^n 1_{A_i} X$ . Последовательность  $X_n$  монотонная и поточечно сходится к  $Y = \sum_{i=1}^{\infty} 1_{A_i} X$ . Согласно МСТ (...)  $\mathbb{E}(Y) = \lim \mathbb{E}(X_n) = \lim \sum_{i=1}^n \mathbb{E}(1_{A_i} X) = \sum_{i=1}^{\infty} \mathbb{E}(1_{A_i} X)$ . Разделив обе части на  $\mathbb{E}(X)$  получаем  $Q(\cup A_i) = \sum Q(A_i)$ . □

**Пример 23.2.** Пусть (табличка),  $Q(A) = \frac{\mathbb{E}(1_A X)}{\mathbb{E}(X)}$  ... Найдите  $Q(Y > 0)$ ,  $E_Q(Y)$

**Теорема 23.3.** Для новой вероятности  $Q(A) = \frac{\mathbb{E}(1_A X)}{\mathbb{E}(X)}$  новое математическое ожидание будет считаться по формуле  $E_Q(Y) = \frac{\mathbb{E}(XY)}{\mathbb{E}(X)}$ .

*Доказательство.* Доказательство использует стандартный пошаговый метод. Сначала для простых  $Y$ .

Теперь для неотрицательных  $Y$ .

Теперь для произвольных  $Y$ . □

**Пример 23.4.** Здесь будет дискретный пример.

**Определение 23.5.** Вероятность  $Q$  называется **абсолютно непрерывной** по отношению к вероятности  $P$ , если из условия  $\mathbb{P}(A) = 0$  всегда следует, что  $Q(A) = 0$ . Обозначается абсолютная непрерывность  $PQ$ .

**Пример 23.6.** Пусть  $\Omega = [0; 1]$ ,  $\mathcal{F} = \mathcal{B}[0; 1]$ ,  $\lambda$  - классическая мера Лебега на  $[0; 1]$ . ... Чтобы была абсолютная и второй пример, чтобы не было.

Оказывается иногда верно и обратное, а именно:

**Теорема 24.1.** Если  $P$  и  $Q$  - две вероятности,  $PQ$ , то существует единственная почти-наверное случайная величина  $X$ , такая что  $E_P(X) = 1$  и  $Q(A) = E_P(X1_A)$ .

*Доказательство.* Доказательство существования - в аппендикс (?) Нетрудно доказать, что  $E_P(X) = E_P(X \cdot 1_\Omega) = Q(\Omega) = 1$ . Легко доказать и единственность: Пусть  $X$  и  $Y$  две функции, удовлетворяющие теореме, тогда  $Q(X > Y) = E_P(X \cdot 1_{X>Y}) = E_P(Y \cdot 1_{X>Y})$ . Отсюда,  $E_P((X - Y)1_{X-Y>0}) = 0$ . Но,  $(X - Y)1_{X-Y>0}$  - неотрицательная случайная величина, значит  $\mathbb{P}((X - Y)1_{X-Y>0} > 0) = 0$  и  $\mathbb{P}(X - Y > 0) = 0$ . Из симметрии  $\mathbb{P}(Y - X > 0) = 0$  и, наконец,  $\mathbb{P}(X = Y) = 1$ .  $\square$

**Пример 24.2.** (или упр.) найдите производную радона-никодима - дискретный случай

**Пример 24.3.** Пусть  $\Omega = [0; 1]$ , (упр.?) - производная радона-никодима непрер. случай

## 6.5 Пространство $L^2$ и его геометрия

**Определение 24.4.** Пусть  $X$  - случайная величина и  $p \geq 1$ . Определим  $\|X\|_p := (\mathbb{E}(|X|^p))^{1/p}$ .

Это определение допускает, что  $\|X\|_p$  может равняться  $+\infty$ .

Упражнение. приведите пример  $X$ , такой что  $\|X\|_1 < \infty$ , а  $\|X\|_2 = \infty$ .

**Определение 24.5.** Пространством  $L^p(\Omega, \mathcal{F}, P)$  назовем множество  $\mathcal{F}$ -измеримых случайных величин, таких, что  $\|X\|_p < +\infty$ .

Когда понятно о каких  $\mathcal{F}$  и  $P$  идет речь мы будем сокращать обозначение  $L^p(\Omega, \mathcal{F}, P)$  до  $L^p(\Omega)$  и даже до  $L^p$ .

**Теорема 24.6.** Пространство  $L^p$  является линейным пространством над  $\mathbb{R}$ . Если  $X \in L^p$ ,  $Y \in L^p$  и  $a \in \mathbb{R}$ , то  $X + Y \in L^p$  и  $aX \in L^p$ .

*Доказательство.*  $\square$

**Теорема 24.7.** Если  $1 \leq p \leq q$ , то  $L^q \subset L^p$

*Доказательство.*  $\square$

Holder, Minkowski

Чуть позже (...) мы докажем полноту пространства  $L^p$ .

Среди пространств  $L^p$  наибольший интерес представляет для нас пространство  $L^2$ . Во-первых, потому, что именно в  $L^2$  мы будем по началу определять интеграл Ито. Во-вторых, потому, что в  $L^2$  есть красивая геометрия, даже две!

(вставка про геометрию)

## 6.6 Сеанс связи с Землей

Может показаться, что этот конспект - какой-то другой курс теории вероятностей. Формально определяется то, что раньше без всяких проблем и заморочек считалось. Осталось связать эти «два» курса, указав, что способ подсчета был верный.

Как до теории меры мы считали математической ожидание случайной величины или функции от нее?

Мы использовали функцию плотности  $p(x)$  и брали обычный (Римановский) интеграл:

$$\mathbb{E}(f(X)) = \int_{-\infty}^{+\infty} f(x)p(x)dx$$

Чтобы обосновать этот способ сначала напомним пару определений:

**Определение 24.8.** Функция  $p(x)$  называется **функцией плотности** случайной величины  $X$ , если...



(дать два варианта? с борелевскими и с нормальными?)

**Определение 24.9.** Функция называется интегрируемой по Риману...

Теоремы, связывающие:

## 6.7 Еще задачи

## 7 Сходимости

Случайные величины - более сложные объекты, чем числа, поэтому сходимость случайных величин бывает разная.

### 7.1 Разные виды сходимостей

**Определение 25.1.** Последовательность случайных величин  $X_n$  сходится к случайной величине  $X$  **поточечно** если для любого  $w \in \Omega$  имеет место сходимость  $X_n(w) \rightarrow X(w)$ .

Поскольку ни вероятность, ни математическое ожидание «не чувствуют» изменения происходящие на множестве меры нуль, то это определение имеет разумное обобщение:

**Определение 25.2.** Последовательность случайных величин  $X_n$  сходится к случайной величине  $X$  **почти наверное** если  $\mathbb{P}(w | X_n(w) \rightarrow X(w)) = 1$ .

В силу определений, из поточечной сходимости следует сходимость почти наверное, т.к.  $\mathbb{P}(\Omega) = 1$ .

**Пример 25.3.**

Несколько другим подходом является рассмотрение случайных величин, как векторов пространства  $L^p$ .

**Определение 25.4.** Последовательность случайных величин  $X_n$  сходится к случайной величине  $X$  **в пространстве  $L^p$**  ( $p \geq 1$ ), если  $\|X_n - X\|_p \rightarrow 0$  или, эквивалентно,  $E(|X_n - X|^p) \rightarrow 0$ .

Следующие примеры показывают, что сходимость в  $L^p$  и сходимость почти наверное не всегда одно и то же.

**Пример 25.5.**

**Пример 25.6.**

**Пример 25.7.**

Однако и сходимость в  $L^p$  и сходимость почти наверное приводят к тому, что случайные величины сходятся по вероятности

**Определение 25.8.** Последовательность случайных величин  $X_n$  сходится к случайной величине  $X$  **по вероятности**, если для любого  $\varepsilon$  вероятность  $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$ .

Как и было обещано, две теоремы:

**Теорема 25.9.** Если  $X_n$  сходится к  $X$  почти наверное, то  $X_n$  сходится к  $X$  по вероятности.

*Доказательство.* □

**Теорема 25.10.** Если  $X_n$  сходится к  $X$  в  $L^1$ , то  $X_n$  сходится к  $X$  по вероятности.

*Доказательство.* □

Для сходимостей по вероятности, почти наверное, в  $L^p$  и поточечной требуется, чтобы и все  $X_n$ , и  $X$  были заданы на одном вероятностном пространстве  $(\Omega, \mathcal{F}, P)$ . Ибо только в этом случае будет определена разность  $X_n - X$ .

От этого требования можно отказаться и получить еще более «слабую» концепцию сходимости.

Мы приведем два эквивалентных определения слабой сходимости и (?) докажем их эквивалентность:

**Определение 25.11.** Последовательность случайных величин  $X_n$  сходится **слабо** (или **по распределению**) к случайной величине  $X$ , если для любой непрерывной ограниченной функции  $f: \mathbb{R} \rightarrow \mathbb{R}$  сходятся математические ожидания  $\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X))$ .

**Пример 26.1.**

Почему в определении включены непрерывность и ограниченность показывают два примера:

**Пример 26.2.** Пусть  $X_n$  равномерно на  $[0; \frac{1}{n}]$ . А  $X$  - тождественно равно нулю. Интуитивно понятно, что  $X_n$  становятся похожи на  $X$  при  $n \rightarrow \infty$ . Возьмем разрывную  $f(t) = 1_{t=0}$ . Получается, что  $\mathbb{E}(f(X)) = \mathbb{E}(1) = 1$ , но  $\mathbb{E}(f(X_n)) = 0$ . Если же функция  $f$  непрерывна, то при достаточно большом  $n$  окажется, что  $f(X_n)$  не может отличаться от  $f(0)$  больше чем на  $\varepsilon$ . Следовательно,  $\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X))$ . Это и означает, что  $X_n$  слабо сходится к 0. Заметим, что  $X_n$  могли быть заданы на разных вероятностных пространствах.

**Пример 26.3.**

Второе определение формулируется с помощью функции распределения:

**Определение 26.4.** Последовательность случайных величин  $X_n$  (с функцией распределения  $F_n(t)$ ) сходится **слабо** (или **по распределению**) к случайной величине  $X$  (с функцией распределения  $F(t)$ ), если  $F_n(t) \rightarrow F(t)$  для любой точки  $t$ , в которой функция распределения  $F(t)$  непрерывна.

Почему в определении требуется сходимость только в тех точках, где  $F_X(t)$  непрерывна можно понять из того же примера ??:

**Пример 26.5.** Пусть  $X_n$  равномерно на  $[0; \frac{1}{n}]$ . А  $X$  - тождественно равно нулю. На рисунках видно, что функции распределения  $F_n(t)$  сходятся к  $F(t)$  во всех точках, кроме точки  $t = 0$ . Но для точки 0 сходимость не требуется. Значит  $X_n$  слабо сходится к нулю.

(картинки)

Доказательство эквивалентности:

«Слабая сходимость» соответствует своему названию, и следует из любой другой.

**Теорема 26.6.** Если  $X_n$  сходится к  $X$  по вероятности, то  $X_n$  слабо сходится к  $X$ .

*Доказательство.*

□

В результате общая картина сходимостей выглядит так:

Неединственность предела.

Все виды сходимостей (кроме поточечной) используют либо понятие математического ожидания, либо понятие вероятности. Ни математическое ожидание, ни вероятность не меняются, если значение случайной величины изменить на множестве меры нуль. Значит любой из пределов (кроме поточечного) не единственный! Т.е. из того, что  $\lim X_n = Y$  и  $\lim X_n = Z$  строго говоря не следует  $X = Y$ .

Конечно, не все так плохо и для всех сходимостей (кроме сходимости по распределению) будет верным утверждение: если  $\lim X_n = Y$  и  $\lim X_n = Z$  то  $Y = Z$  почти наверное.

Старые свойства выполнены:

$$\lim X_n + \lim Y_n = \lim(X_n + Y_n)$$

Три критерия сходимости почти наверное: (тттк)

- Для любого  $\varepsilon$  существует такое  $N$ , что для всех  $n > N$  выполнено  $\mathbb{P}(|X_n - X| < \varepsilon) > 1 - \varepsilon$
- $\sum \mathbb{P}(|X_n - X| > \varepsilon) < \infty$  для всех  $\varepsilon$ .
- $\mathbb{P}(\limsup |X_n - X| > \varepsilon) = 0$  для всех  $\varepsilon$ .

Наиболее полезны, пожалуй, первые два, т.к. они «вытаскивают» предел изнутри вероятности наружу.

Упражнение.

Докажите, что обратные утверждения к леммам BC1 и BC2 неверны.

В силу неравенства Йенсена:  $X \in L_p \Rightarrow X \in L_{p-1}$  для  $p > 1$ .

## 7.2 Обратные теоремы? Иногда!

В некоторых важных случаях можно получить «обратные» теоремы.

**Теорема 27.0.** Если  $X_n$  сходится к  $X$  по вероятности, то из последовательности  $X_n$  можно выбрать подпоследовательность  $Y_k$ , сходящуюся к  $X$  почти наверное.

**Теорема 27.1.** (Скоророход)  $X_n$  слабо сходится к  $X$ , если и только если существует последовательность  $X'_n$ , сходящаяся к  $X'$  почти наверное, такая что для любого  $n$  распределение  $X_n$  совпадает с распределением  $X'_n$  и  $X \sim X'$

**Теорема 27.2.** Если  $X_n$  сходится по вероятности к константе...

Докажите такую характеристику сходимости по распределению:

$X_n$  слабо сходится к  $X$ , если и только если существует последовательность  $X'_n$ , сходящаяся к  $X'$  по вероятности, такая что для любого  $n$  распределение  $X_n$  совпадает с распределением  $X'_n$ , а  $X' \sim X$ .

Пусть  $X_i$  - независимы и  $N(0; 1)$ . Рассмотрим последовательность  $Y_n = \sqrt{n}\bar{X}$ . К чему и в каких смыслах она сходится?

По распределению сходится, т.к.  $Y_n \sim N(0; 1)$ . По вероятности не сходится (трудно).

## 7.3 Еще задачи

# 8 Независимость и произведения пространств

Мы уже употребляли независимость, настала пора ее строго определить.

## 8.1 Независимость случайных величин

**Определение 27.3.**  $\sigma$ -алгебры  $\mathcal{F}_1$  и  $\mathcal{F}_2$  называются независимыми, если независимы любые два события  $F_1 \in \mathcal{F}_1$  и  $F_2 \in \mathcal{F}_2$ , т.е.  $\mathbb{P}(F_1 \cap F_2) = \mathbb{P}(F_1)\mathbb{P}(F_2)$ .

Интуитивно независимость  $\sigma$ -алгебр понимается просто: если про каждое событие из  $\mathcal{F}_1$  вам сказали произошло оно или нет, то это никак не меняет вашу оценку вероятности для событий из  $\mathcal{F}_2$ .

С каждой случайной величиной  $X$  связана  $\sigma$ -алгебра  $\sigma(X)$  поэтому, определив независимость  $\sigma$ -алгебр, мы можем дать...

**Определение 27.4.** Случайные величины  $X$  и  $Y$  называются независимыми, если независимы  $\sigma$ -алгебры  $\sigma(X)$  и  $\sigma(Y)$ .

**Определение 27.5.** Случайная величина  $X$  и  $\sigma$ -алгебра  $\mathcal{H}$  называются независимыми, если независимы  $\sigma$ -алгебры  $\sigma(X)$  и  $\mathcal{H}$ .

Опять же, интуитивное понимание независимости случайных величин просто: если вам сказали  $X$ , то это никак не меняет вашу оценку вероятностей связанных с  $Y$ .

Парочка важных теоремок:

**Теорема 27.6.** Если  $X$  и  $Y$  независимы, то и  $f(X)$  и  $g(Y)$  независимы.

*Доказательство.* Заметим, что  $\sigma(f(X)) \subset \sigma(X)$  и  $\sigma(g(Y)) \subset \sigma(Y)$ . По условию  $\sigma$ -алгебры  $\sigma(X)$  и  $\sigma(Y)$  независимы, значит независимы и  $\sigma$ -алгебры  $\sigma(f(X))$  и  $\sigma(g(Y))$ .  $\square$

**Теорема 27.7.** Если  $X$  и  $Y$  независимы, то  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

*Доказательство.* Доказательство построено пошагово почти так же как и определение математического ожидания.  $\square$

Остается добавить про связь между независимостью и некоррелированностью. Из теоремы ?? следует, что из независимости случайных величин следует их некоррелированность. Можно установить и взаимно однозначную связь между некоррелированностью и независимостью:

**Теорема 27.8.** *Случайные величины  $X$  и  $Y$  независимы, если и только если корреляция  $\text{Corr}(f(X), g(Y)) = 0$  для любых борелевских функций  $f$  и  $g$ .*

*Доказательство.* В одну сторону: если независимы  $X$  и  $Y$ , то независимы  $f(X)$  и  $g(Y)$ , значит  $\text{Corr}(f(X), g(Y)) = 0$ . В другую: возьмем в качестве функции  $f$  индикатор произвольного борелевского множества  $A$ , а в качестве функции  $g$  индикатор произвольного борелевского множества  $B$ . Из некоррелированности индикаторов следует независимость событий  $\{X \in A\}$  и  $\{Y \in B\}$ . Все события из  $\sigma(X)$  имеют вид  $\{X \in A\}$ , а все события из  $\sigma(Y)$  имеют вид  $\{Y \in B\}$ .  $\square$

## 8.2 Произведение пространств

Важные следствия:

Когда можно переставлять производную и математическое ожидание?

Когда можно переставлять интеграл и математическое ожидание?

## 8.3 Еще задачи

## 9 Условное ожидание

Понятие «условное ожидание» используется в двух смыслах:

- $\mathbb{E}(Y|A)$ , где  $Y$  - случайная величина и  $A$  - событие. В этом случае  $\mathbb{E}(Y|A)$  - число.
- $\mathbb{E}(Y|\mathcal{H})$ , где  $Y$  - случайная величина и  $\mathcal{H}$  -  $\sigma$ -алгебра. В этом случае  $\mathbb{E}(Y|\mathcal{H})$  - случайная величина.

### 9.1 Условное ожидание как константа

Условное ожидание  $\mathbb{E}(Y|A)$  скорее всего знакомо человеку, освоившему вводный курс теории вероятностей. Условное ожидание  $\mathbb{E}(Y|A)$  считается точно так же, как и обычное  $\mathbb{E}(Y)$ , только вместо обычной вероятности  $\mathbb{P}(\cdot)$  используется условная вероятность  $\mathbb{P}(\cdot|A)$ . Если  $A$  - событие с положительной вероятностью, то мы можем сменить вероятность каждого события  $C$  с  $\mathbb{P}(C)$  на  $\tilde{\mathbb{P}}(C) := \mathbb{P}(C|A)$ . В упр. (...) мы видели, что  $\tilde{\mathbb{P}}$  - это действительно вероятность на  $\sigma$ -алгебре  $\mathcal{F}$ . Значит можно считать математическое ожидание (интегрировать) относительно нее:

**Определение 28.1.** Условное ожидание величины  $Y$  относительно события  $A$ , с  $\mathbb{P}(A) > 0$  определяется по формуле:

$$\mathbb{E}(Y|A) := E_{\tilde{\mathbb{P}}}(Y) = \int_{\Omega} Y d\tilde{\mathbb{P}}, \quad (28.2)$$

где  $\tilde{\mathbb{P}}(\cdot) := \mathbb{P}(\cdot|A)$ .

Можно дать и эквивалентное определение:

**Определение 28.3.** Условное ожидание величины  $Y$  относительно события  $A$ , с  $\mathbb{P}(A) > 0$  определяется по формуле:

$$\mathbb{E}(Y|A) := \frac{\mathbb{E}(1_A Y)}{\mathbb{P}(A)} \quad (28.4)$$

**Теорема 28.5.** *Определения ?? и ?? эквивалентны.*

*Доказательство.* Стандартная пошаговая процедура.

Для простых случайных величин эти определения эквивалентны:

$$\begin{aligned}\int_{\Omega} Y d\tilde{P} &= \sum_{i=1}^n Y_i \mathbb{P}(Y = y_i | A) = \sum_{i=1}^n y_i \frac{\mathbb{P}(Y = y_i \cap A)}{\mathbb{P}(A)} \\ &= \frac{1}{\mathbb{P}(A)} \sum_{i=1}^n y_i \mathbb{E}(1_A 1_{Y=y_i}) = \frac{1}{\mathbb{P}(A)} E \left( \sum y_i 1_{Y=y_i} 1_A \right) = \frac{\mathbb{E}(1_A Y)}{\mathbb{P}(A)}\end{aligned}\quad (28.6)$$

А дальше применяется МСТ...

□

## 9.2 Условное ожидание как случайная величина. Конечный случай.

Гораздо более интересно условное ожидание относительно  $\sigma$ -алгебры!

Для начала рассмотрим конечно-порожденную  $\sigma$ -алгебру.

Пример. (...)

**Определение 29.1.** Если  $A_1, \dots, A_n$  попарно не пересекаются,  $\mathbb{P}(A_i) > 0$  и  $\mathcal{H} = \sigma(\{A_i\})$ , то

$$\mathbb{E}(Y|\mathcal{H}) := \sum_i \mathbb{E}(Y|A_i) 1_{A_i} \quad (29.2)$$

По этому определению, если произойдет  $A_1$ , то величина  $\mathbb{E}(Y|\mathcal{H})$  будет равна  $\mathbb{E}(Y|A_1)$ , если произойдет  $A_2$ , то величина  $\mathbb{E}(Y|\mathcal{H})$  будет равна  $\mathbb{E}(Y|A_2)$  и т.д. А поскольку мы заранее не знаем, какое из  $A_i$  произойдет, то  $\mathbb{E}(Y|\mathcal{H})$  является случайной величиной.

Для краткости обозначим  $\hat{Y} := \mathbb{E}(Y|\mathcal{H})$ .

Как правильно думать об условном математическом ожидании  $\mathbb{E}(Y|\mathcal{H})$ ?

- Во-первых, полезно представлять себе  $\mathbb{E}(Y|\mathcal{H})$  как «грубую», усредненную, версию  $Y$ . На множестве  $A_i$  случайная величина  $Y$  может принимать разные значения. А случайная величина  $\mathbb{E}(Y|\mathcal{H})$  на множестве  $A_i$  принимает фиксированное значение  $\mathbb{E}(Y|A_i)$ , а это не что иное, как среднее значение величины  $Y$  на множестве  $A_i$ .  
(картинка)
- Проекция  $Y$  на множество  $\mathcal{H}$ -измеримых случайных величин. Если рассмотреть все  $\mathcal{H}$ -измеримые случайные величины, то  $\mathbb{E}(Y|\mathcal{H})$  — самая похожая на случайную величину  $Y$ . «Самая похожая» в том смысле, что расстояние  $\|Y - \mathbb{E}(Y|\mathcal{H})\|_{L^2}$  — наименьшее из возможных.  
(рисунок)
- $\hat{Y}$  — как решение системы уравнений. С помощью ковариаций и мат. ожидания нельзя почувствовать разницу между  $Y$  и  $\hat{Y}$ ! А именно, наш прогноз  $\hat{Y}$  — это единственное решение системы уравнений:

$$\begin{cases} \mathbb{E}(\hat{Y}) = \mathbb{E}(Y) \\ Cov(X_1, \hat{Y}) = Cov(X_1, Y) \\ Cov(X_2, \hat{Y}) = Cov(X_2, Y) \\ \dots \\ Cov(X_n, \hat{Y}) = Cov(X_n, Y) \end{cases} \quad (29.3)$$

где  $X_i$  — это всевозможные<sup>1</sup>  $\mathcal{H}$ -измеримые случайные величины.

**Пример 29.4.** здесь дискретный пример с иллюстрацией трех свойств.

Доказательство корректности второй и третьей интерпретации мы проведем для произвольной  $\mathcal{H}$ . Но для начала надо обобщить само определение условного математического ожидания

<sup>1</sup>Здесь есть некоторая вольность: таких случайных величин бесконечное количество, а вовсе не  $n$ . Но в случае конечно порожденной  $\sigma$ -алгебры можно ограничиться индикаторами  $A_i$

на случай произвольной  $\sigma$ -алгебры. Каждую из трех идей можно довести до определения подходящего в общем случае. Кратко о каждом способе:

- Уточнение грубой версии. Нужно переходить к пределу рассматривая все более и более мелкие  $A_i$ . Легко перейти на счетно-порожденные  $\sigma$ -алгебры. Гораздо трудно перейти на  $\sigma$ -алгебры не являющиеся счетно-порожденными (например такой является хвостовая  $\sigma$ -алгебра).
- Обобщение проекции. Проекция определена там, где есть углы и скалярные произведения, т.е. в  $L^2$ . Самая главная трудность состоит в переходе от  $L^2$  определения к  $L^1$  определению.
- Решение системы уравнений. Никаких проблем, годится для общего случая. Но трудно доказать существование условного ожидания, т.к. в системе бесконечное количество уравнений, которое не сводится к конечному количеству. Мы пойдем этим путем. За доказательство существования будет отвечать т. Радона-Никодима.

Чаще всего приходится иметь дело не с абстрактной  $\sigma$ -алгеброй  $\mathcal{H}$ , а с  $\sigma$ -алгеброй, порожденной некоей случайной величиной  $X$ ,  $\mathcal{H} = \sigma(X)$ . В этом случае используют вместо  $\mathbb{E}(Y|\sigma(X))$  более простое обозначение  $\mathbb{E}(Y|X)$ . Если  $X$  принимает конечное количество значений, то  $\sigma(X)$  будет конечнопорожденной  $\sigma$ -алгеброй, и в этом случае можно искать  $\mathbb{E}(Y|X)$  по формуле ??.

Мы интерпретируем  $\mathbb{E}(Y|X)$  как среднее значение  $Y$  при известном  $X$ . Значение  $\mathbb{E}(Y|X)$  является функцией от  $X$ , а поскольку  $X$  является случайной величиной, то и  $\mathbb{E}(Y|X)$  является случайной величиной.

### 9.3 Условное ожидание как случайная величина. Общий случай.

**Определение 30.1.** Пусть  $\mathcal{H}$  - произвольная  $\sigma$ -алгебра и  $Y$  - случайная величина с  $\mathbb{E}(Y) < \infty$ . Условным ожиданием  $\mathbb{E}(Y|\mathcal{H})$  называется любая случайная величина  $\hat{Y}$  удовлетворяющая двум условиям:

- $\hat{Y}$  является  $\mathcal{H}$ -измеримой
  - Для любого события  $A \in \mathcal{H}$  верно равенство:  $\mathbb{E}(\hat{Y}1_A) = \mathbb{E}(Y1_A)$
- Отметим в частности, что в любую  $\mathcal{H}$  входит множество  $\Omega$ , поэтому  $\mathbb{E}(\hat{Y}) = \mathbb{E}(Y)$ .

Зачем в определении требуется, чтобы  $\mathbb{E}(Y) < \infty$ ? Если этого не потребовать, могут возникнуть некоторые парадоксы.

Например: (две шкатулки)

Конечно же,

**Теорема 30.2.** Если  $\mathcal{H}$  - конечно порожденная  $\sigma$ -алгебра, то определения ... и ... совпадают.

*Доказательство.* □

Условные ожидания всегда существуют и почти наверное единственны:

**Теорема 30.3.** Если  $\mathbb{E}(Y) < \infty$  и  $\mathcal{H}$  -  $\sigma$ -алгебра, то существует  $\hat{Y} = \mathbb{E}(Y|\mathcal{H})$ . Если еще и  $\hat{Y}' = \mathbb{E}(Y|\mathcal{H})$ , то  $\mathbb{P}(\hat{Y} = \hat{Y}') = 1$ .

*Доказательство.* • Существование. Если  $Y \geq 0$ , то на  $\sigma$ -алгебре  $\mathcal{H}$  можно определить конечную меру  $\mu(A) := \mathbb{E}(Y1_A)$ . На  $\mathcal{H}$  эта мера абсолютно непрерывна по отношению к мере  $P$ . По теореме Радона-Никодима существует  $\mathcal{H}$  измеримая случайная величина  $\hat{Y}$ , такая что  $\mu(A) = \mathbb{E}(\hat{Y}1_A)$ . Для произвольного  $Y$  существование доказывается путем разложения  $Y = Y^+ - Y^-$ .

- Единственность. Величины  $\hat{Y}$  и  $\hat{Y}'$  являются  $\mathcal{H}$ -измеримыми, поэтому  $A = \{\hat{Y} - \hat{Y}' > \varepsilon\} \in \mathcal{H}$ .  $\mathbb{E}((\hat{Y} - \hat{Y}')1_A) = \mathbb{E}(\hat{Y}1_A) - \mathbb{E}(\hat{Y}'1_A) = \mathbb{E}(Y1_A) - \mathbb{E}(Y1_A) = 0$ . С другой стороны,  $\mathbb{E}((\hat{Y} - \hat{Y}')1_A) \geq \mathbb{E}(\varepsilon1_A) = \varepsilon P(A)$ . Значит  $P(A) = 0$  для всех  $\varepsilon$ . Событие  $(\hat{Y} > \hat{Y}') = \cup_{\varepsilon=1/n} A(\varepsilon)$ . В силу непрерывности вероятности,  $P(\hat{Y} > \hat{Y}') = 0$ . Из соображений симметрии,  $P(\hat{Y}' > \hat{Y}) = 0$ .

□

Условное ожидание обладает рядом хороших свойств:

## Как считать условное ожидание?

- Для конечной  $\mathcal{H}$  ответ дает формула с которой мы начали определение условного ожидания:
- Если нужно посчитать  $\mathbb{E}(Y|X)$ , где  $y$  пары случайных величин  $X$  и  $Y$  есть совместная функция плотности, то можно воспользоваться готовым рецептом:

**Теорема 31.1.** Если у пары  $X, Y$  есть совместная функция плотности  $p(x, y)$ , то условное ожидание  $\mathbb{E}(Y|X)$  можно найти по формуле:

$$\mathbb{E}(Y|X) = \int_{-\infty}^{+\infty} y \cdot p(y|X) dy \quad (31.2)$$

где  $p(y|x)$  - условная функция плотности,  $p(y|x) = \frac{p(x,y)}{p(x)}$ .

*Доказательство.* □

- В остальных случаях остается одно. Думать. Иногда из интуитивных или геометрических соображений можно догадаться до правильного ответа и тогда остается лишь проверить, что догадка  $\hat{Y}$  удовлетворяет определению условного ожидания. При этом может помочь тот факт, что при проверке равенства  $\mathbb{E}(\hat{Y}1_A) = \mathbb{E}(Y1_A)$  не обязательно проверять все  $A \in \mathcal{H}$ :

**Теорема 31.3.** Пусть  $\mathcal{M}$  - набор событий замкнутый относительно пересечений (если  $A, B$  лежат в  $\mathcal{M}$ , то и  $A \cap B$  лежит в  $\mathcal{M}$ ),  $\mathcal{H} = \sigma(\mathcal{M})$ ,  $\mathbb{E}(Y) = \mathbb{E}(\hat{Y}) < \infty$ . Если  $\mathbb{E}(\hat{Y}1_A) = \mathbb{E}(Y1_A)$  для всех  $A \in \mathcal{M}$ , то  $\mathbb{E}(\hat{Y}1_A) = \mathbb{E}(Y1_A)$  для всех  $A \in \mathcal{H}$ .

*Доказательство.* Рассмотрим набор множеств  $\mathcal{N}$ , для которых  $\mathbb{E}(\hat{Y}1_A) = \mathbb{E}(Y1_A)$ . Мы докажем, что  $\mathcal{N}$  -  $\sigma$ -алгебра. Набор множеств  $\mathcal{N}$  содержит  $\mathcal{M}$  по условию. Если  $\mathcal{N}$  - действительно  $\sigma$ -алгебра, то она обязана содержать наименьшую  $\sigma$ -алгебру  $\mathcal{H}$  порождаемую набором  $\mathcal{M}$ . А это нам и требуется доказать.

По условию в  $\mathcal{N}$  входит  $\Omega$ . Если  $A \in \mathcal{N}$ , то и  $A^c \in \mathcal{N}$ , т.к.  $\mathbb{E}(Y1_{A^c}) = \mathbb{E}(Y) - \mathbb{E}(Y1_A) = \mathbb{E}(\hat{Y}) - \mathbb{E}(\hat{Y}1_A) = \mathbb{E}(\hat{Y}1_{A^c})$ .

Если  $A_i \in \mathcal{N}$ , то и  $\cup A_i \in \mathcal{N}$ . Для непересекающихся  $A_i$  это следует в силу DCT. а от пересекающихся можно легко перейти к непересекающимся. Здесь требуется замкнутость набора  $\mathcal{M}$ . □

Например, в самом распространенном случае, при поиске  $\mathbb{E}(Y|X)$ , т.е. когда  $\mathcal{H} = \sigma(X)$ , достаточно проверять не все множества из  $\sigma X$ , а только множества вида  $X \leq t$ . Вспомните теорему ??.

Примеры...

## Условное ожидание как функция и парадокс Колмогорова-Бореля

Образно говоря, «при прогнозировании монеток не подкидывают»:

**Теорема 31.4.** Случайная величина  $\mathbb{E}(Y|X)$  представима в виде  $\mathbb{E}(Y|X) = f(X)$ , где  $f$  - неслучайная борелевская функция.

*Доказательство.* □

Эта теорема говорит, что вся случайность величины  $\mathbb{E}(Y|X)$  связана только со случайной величиной  $X$ . Случайная величина  $\mathbb{E}(Y|X)$  не может использовать другие источники случайности, не связанные с  $X$ !

Можно нарисовать такую схему:

Уточним, что  $f$  не обязательно единственна! Можно лишь утверждать, что если  $f_1$  и  $f_2$  две подобные функции, то  $f_1(X) = f_2(X)$  почти наверное.

Очень часто хочется говорить о величине  $\mathbb{E}(Y|X = x)$  (среднее значение  $Y$ , если  $X = x$ ). Если  $\mathbb{P}(X = x) > 0$ , то тут нет никаких проблем. Если  $\mathbb{P}(X = x) = 0$ , то можно прибегнуть к представлению  $\mathbb{E}(Y|X) = f(X)$ . А именно, сказать что, по определению  $\mathbb{E}(Y|X = x) := f(x)$ . Например, для непрерывных  $X$  и  $Y$  хочется определить:

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{+\infty} y f(y|x) dy \quad (32.1)$$

Это интуитивно понятно, но не стоит этим злоупотреблять! Дело в том, что функция  $f$  определена не однозначно, а только почти наверное. Это означает, что для каждого конкретного  $x$  можно выбрать любое  $f(x)$ ! Именно поэтому предлагаемое «определение» не устойчиво к преобразованиям.

**Пример 32.2.** Парадокс Колмогорова-Бореля. Пусть совместная функция плотности  $X$  и  $Y$  имеет вид:

$$p_{X,Y}(x, y) = \begin{cases} 1, & 0 < y < 1, -y < x < 1 - y \\ 0, & otherwise \end{cases} \quad (32.3)$$

Пусть  $U = \frac{X}{Y} + 1$ . Найдите:

- а)  $p_{Y|X}(y|x)$
- б)  $p_{U,Y}(u, v)$  и  $p_{Y|U}(y, u)$
- в) Верно ли, что  $X = 0 \Leftrightarrow U = 1$ ?
- г) Верно ли, что  $p_{Y|X}(t|0) = p_{Y|U}(t, 1)$ ?
- д) Мораль?

## Условное ожидание и независимость

Как связаны условное ожидание и независимость?

Из независимости  $X$  и  $\sigma$ -алгебры  $\mathcal{H}$  следует, что  $\mathbb{E}(X|\mathcal{H}) = \mathbb{E}(X)$ . Обратное неверно, как показывает:

**Пример 32.4.** Пусть  $X \sim N(0; 1)$ . Зависимы ли  $X$  и  $X^2$ ? Найдите  $\mathbb{E}(X|X^2)$ .

Решение. Очевидно зависимы, но  $\mathbb{E}(X|X^2) = 0 = \mathbb{E}(X)$

Тем не менее, независимость и условные ожидания тесно связаны:

**Теорема 32.5.** *Случайные величины  $X$  и  $Y$  независимы если и только если  $\mathbb{E}(f(X)|Y) = \mathbb{E}(f(X))$  для любой борелевской функции  $f$ .*

*Доказательство.* Если  $X$  и  $Y$  независимы, то независимы  $f(X)$  и  $Y$ . Значит  $\mathbb{E}(f(X)|Y) = \mathbb{E}(f(X))$ .

...

□

А именно, если  $X$  и  $Y$  независимы, можно ли утверждать, что  $\mathbb{E}(X|\mathcal{H})$  и  $\mathbb{E}(Y|\mathcal{H})$  независимы?

Интуитивно на этот вопрос легко ответить используя геометрическую интерпретацию. Пусть есть две некоррелированные случайные величины  $X$  и  $Y$ . На рисунке (???) они изображены как перпендикулярные векторы. Могут ли их проекции на какую-нибудь плоскость не быть перпендикулярными? Да, конечно могут! Скорее наоборот, чтобы между проекциями получился прямой угол нужно выполнения особого условия (???) вектор  $X - Y$  должен лежать в плоскости (??? проверить утверждение)

В качестве конкретного примера подойдет практически первый попавшийся! Например, правильную монетку подбрасывают 20 раз,  $X$  - число орлов в первых 10 подбрасываниях,  $Y$  - число орлов в последних 10 подбрасываниях,  $Z$  - в подбрасываниях с 6-го по 15-ое. Очевидно, что  $X$  и  $Y$  независимы, однако  $\mathbb{E}(X|Z) = \mathbb{E}(Y|Z) = \frac{Z}{2} + \frac{5}{2}$ .

А может быть из независимости  $X$  и  $Y$  будет следовать хотя бы  $\mathbb{E}(XY|\mathcal{H}) = \mathbb{E}(Y|\mathcal{H}) \cdot \mathbb{E}(X|\mathcal{H})$ , «условная некоррелированность»?



Предыдущий пример также показывает, что это не так. Разложив  $X$  и  $Y$  в сумму индикаторов, можно установить, что  $\mathbb{E}(XY|Z) = 25 \left( \frac{Z(Z+8)}{90} + \frac{1}{4} \right) \neq \left( \frac{Z}{2} + \frac{5}{2} \right)^2$ .

## Задачи

**Задача 9.1.** Пусть  $\hat{Y} = \mathbb{E}(Y|\mathcal{H})$ , где  $\mathcal{H}$  - произвольная  $\sigma$ -алгебра. Пусть  $X$  -  $\mathcal{H}$ -измеримая случайная величина, принимающая конечное множество значений. Докажите, что  $\text{Cov}(X, \hat{Y}) = \text{Cov}(X, Y)$

**Задача 9.2.** Приведите пример, таких  $Y_1, Y_2$  и  $\mathcal{H}$ , что  $\mathbb{E}(Y_1|\mathcal{H})$  и  $\mathbb{E}(Y_2|\mathcal{H})$  независимы, а  $Y_1$  и  $Y_2$  зависимы.

## 9.4 Еще задачи

# 10 Производящие функции

[http://www.stats.uwaterloo.ca/~dlmcleis/s901/s901\\_2005.pdf](http://www.stats.uwaterloo.ca/~dlmcleis/s901/s901_2005.pdf) <http://www.mathematik.uni-muenchen.de/~lerdos/WS04/FA/compl.pdf>

Если есть случайная величина  $X$ , то как описать ее закон распределения?

Есть много способов! Если величина непрерывная, то можно описать ее с помощью функции плотности,  $p(t)$ . Если случайная величина дискретная — с помощью функции  $p(t) := \mathbb{P}(X = t)$ . Есть универсальный способ записи, подходящий для любых случайных величин — функция распределения,  $F(t) := \mathbb{P}(X \leq t)$ . В конце концов можно описать просто текстом: например, «правильную» монетку подкидывают  $n$  раз и  $X$  - это количество «орлов».

Сейчас мы познакомимся еще с одной идеей, позволяющей описать закон распределения случайной величины.

Эта идея будет представлена тремя функциями:

- Производящая функция,  $g_X(t) := \mathbb{E}(t^X)$
- Функция производящая моменты,  $m_X(t) := \mathbb{E}(e^{tX})$
- Характеристическая функция,  $\varphi_X(t) := \mathbb{E}(e^{itX})$

## 10.1 Производящая функция

Если  $X$  — целочисленная неотрицательная случайная величина, то для ее описания подходит

**Определение 33.1.** Функция  $g_X(t) := \mathbb{E}(t^X)$  называется производящей функцией случайной величины  $X$ .

Пример с табличкой. (...)

Найдем производящую функцию для пуассоновской случайной величины.

$$g_X(t) = \mathbb{E}(t^X) =$$

Производящая функция может не существовать...

Пример. ....

По производящей функции можно «опознать» случайную величину:

**Теорема 33.2.** Если совпадают производящие функции  $g_X(t)$  и  $g_Y(t)$ , то случайные величины  $X$  и  $Y$  имеют одинаковое распределение.

Легко найти производящую функцию суммы:

**Теорема 33.3.** Если существуют  $g_X(t)$  и  $g_Y(t)$ , то :

$$g_{X+Y}(t) = g_X(t) \cdot g_Y(t)$$

С помощью производящей функции легко считать математическое ожидание и другие моменты:

**Теорема 33.4.** Если существуют  $g_X(t)$  и  $\mathbb{E}(X)$ , то:

$$\mathbb{E}(X) = g'_X(1)$$

$$\mathbb{E}(X(X-1)) = g''_X(1)$$

$$\mathbb{E}(X(X-1)(X-2)) = g'''_X(1)$$

...

Найдем производящую функцию для биномиальной случайной величины.

Упражнение. Пусть  $X \sim \text{Bin}(n, p)$ . Найдите  $g_X(t)$ .

Решение. Биномиальная случайная величина представима в виде суммы независимых  $X = X_1 + X_2 + \dots + X_n$ . Находим производящую функцию отдельного слагаемого, получаем  $g_{X_i}(t) = pt + 1 - p$ . Значит для биномиальной случайной величины  $g_X(t) = (pt + 1 - p)^n$

Теперь попробуем применить производящую функцию на реальных задачах:

Пример.

Сейчас мы познакомимся еще с одним способом описать закон распределения случайной величины, но для этого нам потребуются комплексные случайные величины.

## 10.2 Комплексные случайные величины

Для работы с характеристическими функциями нам потребуются комплексные случайные величины.

**Определение 34.1.** Математическим ожиданием комплексной случайной величины  $Z$  называется число

$$z := \mathbb{E}(\text{Re}(Z)) + i\mathbb{E}(\text{Im}(Z)) \quad (34.2)$$

при условии, что  $\mathbb{E}(|Z|) < \infty$ .

У математического ожидания по-прежнему сохраняются все полезные свойства:

- $\mathbb{E}(cZ) = c\mathbb{E}(Z)$ , если  $c \in \mathbb{C}$
- $\mathbb{E}(Z_1 + Z_2) = \mathbb{E}(Z_1) + \mathbb{E}(Z_2)$ , если мат. ожидания существуют

## 10.3 Определение и примеры

Однако есть принципиально другой подход к описанию закона распределения случайной величины, а именно - производящие/характеристические функции. Их много видов:

- Производящая функция,  $g(t) = \mathbb{E}(t^X)$
- Функция производящая моменты,  $m(t) = \mathbb{E}(e^{tX})$
- Характеристическая функция,  $\varphi(t) = \mathbb{E}(e^{itX})$ .

Чем они хороши?

- Любой случайной величине (дискретной, непрерывной...) сопоставляется непрерывная функция!
- Во-первых, они подходят для описания закона распределения случайной величины. Если у величин  $X$  и  $Y$  совпадают производящие функции, то у них совпадает закон распределения.
- Если  $X$  и  $Y$  - независимые случайные величины, то производящая функция суммы легко находится:  $g_{X+Y}(t) = g_X(t)g_Y(t)$ . Для сравнения: при использовании функций плотности придется считать интеграл!
- С помощью них легко считать моменты случайной величины  $\mathbb{E}(X^k)$ , при этом нужно будет только находить производную.
- Некоторые из них, а именно преобразование Фурье и характеристическая функция существуют для любых случайных величин.

Чем они плохи?

- Их труднее интерпретировать. Мы постараемся...
- Вероятности с их помощью считать труднее.
- Многие студенты в 21 веке до сих пор боятся комплексных чисел<sup>1</sup>.

Давайте попробуем посчитать и убедимся, что все не так трудно!

**Пример 35.1.** Пусть  $X$  принимает значения  $-1$ ,  $2$  и  $4$  с вероятностями  $0.1$ ,  $0.4$  и  $0.5$  соответственно. Найдите

1. Производящую функцию
2. Функцию производящую моменты
3. Характеристическую функцию

Выписав значения соответствующих величин получаем:

1.  $g_X(t) = 0.1t^{-1} + 0.4t^2 + 0.5t^4$ ,
2.  $m_X(t) = 0.1e^{-t} + 0.4e^{2t} + 0.5e^{4t}$ ,
3.  $\varphi_X(t) = 0.1e^{-it} + 0.4e^{2it} + 0.5e^{4it}$ . У характеристической функции можно выделить мнимую и действительную части:  $\varphi_X(t) = 0.1 \cos(t) + 0.4 \cos(2t) + 0.5 \cos(4t) + i(-0.1 \sin(t) + 0.4 \sin(2t) + 0.5 \sin(4t))$ .

Уже можно заметить, что производящие функции оказались неограниченными, а характеристическая - ограничена! Еще один пример:

**Пример 35.2.** Какие значения с какими вероятностями принимает случайная величина  $X$  если:

1. Производящая функция равна  $g_X(t) = 0.6t^{-5} + 0.4t^7$
2. Функция производящая моменты равна  $m_X(t) = 0.7e^{3t} + 0.3e^{27t}$
3. Характеристическая функция равна  $\varphi(t) = 0.5(1 + \cos(t)) + i \cdot 0.5 \sin(t)$ .

Ответы, конечно, читаются в условии:

1.  $X$  принимает значения  $-5$  и  $7$  с вероятностями  $0.6$  и  $0.4$  соответственно.
2.  $X$  принимает значения  $3$  и  $27$  с вероятностями  $0.7$  и  $0.3$  соответственно.
3. Заметим, что  $1 = \cos(0 \cdot t)$  и  $0 = \sin(0 \cdot t)$ , поэтому наша функция равна  $\varphi_X(t) = 0.5e^{i \cdot 0 \cdot t} + 0.5e^{i \cdot 1 \cdot t}$ . Значит  $X$  равновероятно принимает значения  $0$  и  $1$ .

В этом примере есть маленькая нестрогость: мы нашли случайную величину с данной производящей функцией, но не доказали, что не существует другой.

**Пример 35.3.** Пример, когда  $g(t)$  не существует

Мы остановимся более подробно на характеристической функции, так как она определена для любой случайной величины.

Для начала маленькое напоминание про математическое ожидание комплексных случайных величин. Если  $Z = X + iY$  - комплексная случайная величина, то (по определению),  $\mathbb{E}(Z) = \mathbb{E}(X) + i\mathbb{E}(Y)$ . Заметим также, что  $e^{itX} = \cos(tX) + i \sin(tX)$ . Поэтому, под выражением  $\mathbb{E}(e^{itX})$  нужно понимать не что иное как  $\mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX))$ . Из этого сразу следует, что характеристическая функция всегда существует. Действительно, косинус и синус ограничены, значит оба математических ожидания всегда существуют.

**Пример 35.4.** Пусть  $X$  равномерна на  $[0; 1]$ . Найдём характеристическую функцию:  $\varphi(t) = \mathbb{E}(e^{itX}) = \mathbb{E}(\cos(tX) + i \sin(tX)) = \int_0^1 \cos(tx) dx + i \int_0^1 \sin(tx) dx = \frac{\sin(t)}{t} + i \frac{\cos(t) - 1}{t}$ . При  $t = 0$ , особый, но простой случай,  $\varphi(0) = 1$ .

Все, конечно, хорошо представляют как выглядит график функции плотности равномерной случайной величины: (...)

А как представить графически характеристическую функцию? Можно поступить несколькими способами:

- Нарисовать кривую в 3-х мерном пространстве (рисунок)
- Нарисовать два графика на плоскости:  $Re(\varphi(t))$  и  $Im(\varphi(t))$ :

---

<sup>1</sup>Впрочем, это скорее недостаток студента, а не характеристической функции

- Нарисовать два графика на плоскости:  $|\varphi(t)|$  и  $\arg(\varphi(t))$  (в качестве аргумента комплексного числа можно выбрать любую версию, например попадающую в интервал  $[0; 2\pi)$  или непрерывную)

Второй способ, по всей видимости, предпочтительнее так как в случае, когда закон распределения случайной величины  $X$  симметричен относительно нуля (т.е.  $\mathbb{P}(X \leq t) = \mathbb{P}(X \geq -t)$  для  $\forall t$ ), оказывается, что  $\operatorname{Im}(\varphi(t)) = 0$  и достаточно одного графика. Чуть позже мы докажем это, а пока еще один пример.

**Пример 36.1.** Случайная величина  $X$  принимает равновероятно значения 1 и  $(-1)$ . Как выглядит ее характеристическая функция?  $\varphi(t) = \mathbb{E}(e^{itX}) = 0.5e^{-it} + 0.5e^{it} = 0.5(\cos(-t) - i\sin(t)) + 0.5(\cos(t) + i\sin(t)) = \cos(t)$ . Как и было обещано, мнимая часть отсутствует.

## 10.4 Свойства характеристической функции

- $\varphi(0) = 1$   
По определению,  $\varphi(0) = \mathbb{E}(\cos(0 \cdot X)) + i\mathbb{E}(\sin(0 \cdot X)) = 1$
- $|\varphi(t)| \leq 1$  для любого  $t$   
 $|\varphi(t)| = |\mathbb{E}(\cos(t \cdot X) + i\sin(t \cdot X))| \leq \mathbb{E}(|\cos(t \cdot X) + i\sin(t \cdot X)|) = 1$
- $\varphi_{-X}(t) = (\varphi_X(t))^*$  - сопряжение поставить!
- $\varphi_{aX+b}(t) = e^{iat}\varphi(bt)$
- Если  $X$  и  $Y$  независимы, то  $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$

$$\begin{aligned}\varphi_{X+Y}(t) &= E(e^{it(X+Y)}) = E(e^{itX}e^{itY}) \\ &= E(e^{itX})E(e^{itY}) = \varphi_X(t)\varphi_Y(t), \quad (36.2)\end{aligned}$$

в силу независимости  $e^{itX}$  и  $e^{itY}$

- Если закон распределения  $X$  симметричен, то  $\operatorname{Im}(\varphi_X(t)) = 0$
- Функция  $\varphi$  является равномерно непрерывной

## 10.5 Доказательство ЦПТ

## 10.6 Поиск вероятностей и среднего с помощью характеристической функции

Допустим нам нужно найти вероятность связанную с суммой случайных величин. В силу того, что для суммы легко ищется характеристическая функция, бывает что проще искать вероятность именно с ее помощью.

Пример. Вероятность вымирания.

Еще Пример. Поиск вероятности с использованием производящей функции.

Мы рассмотрим пример с векторными случайными величинами.

Пример. Мы подбрасываем монетку до тех пор, пока количество орлов не привысит количества решек.  $T$  - количество подбрасываний. Найдите  $\mathbb{E}(1/T)$ .

Заметим, что  $1/T = \int_0^1 x^{T-1} dx$ .

Ссылаясь на теорему ...  $\mathbb{E}(1/T) = \int_0^1 \mathbb{E}(x^{T-1}) dx$ .

А  $\mathbb{E}(x^T)$  - это не что иное, как производящая функция!

Из этого примера следует мораль: производящие/характеристические функции дают возможность заменить дискретную сумму на интеграл непрерывной функции!

Хотите поговорить об этом?

## 10.7 Суммирование рядов с помощью производящих функций. . .

основное из статьи A unified method to sum a variety of infinite series with applications

## 10.8 Еще задачи

# 11 Мартингалы в дискретном времени

## 11.1 О случайном блуждании

Случайное блуждание - пожалуй, самый важный из дискретных случайных процессов. Броуновскому движению в непрерывном времени многие свойства случайного блуждания «перейдут по наследству». Поэтому мы достаточно подробно изучим симметричное случайное блуждание.

**Определение 37.1.** Пусть  $X_i$  - последовательность независимых случайных величин, равновероятно принимающих значение  $+1$  или  $(-1)$ . Процесс  $S_n = \sum_{i=1}^n X_i$  называется **симметричным случайным блужданием**.

Графически  $S_n$  - это координаты частицы, которая начав в нуле равновероятно смещается на один шаг влево или вправо.

Чтобы «почувствовать» некоторые свойства случайного блуждания мы ответим на несколько вопросов:

1. Пусть  $a$  и  $b$  - два натуральных числа. Какова вероятность того, что случайное блуждание никогда не достигнет ни точки  $a$ , ни точки  $-b$ ? Какова вероятность того, что  $S_n$  «коснется» точки  $a$  раньше, чем точки  $-b$ ?
2. Сколько шагов в среднем потребуется, чтобы достигнуть точки  $1$ ?
3. Сколько шагов в среднем потребуется, чтобы достигнуть уровня  $a$  или  $-b$ ?
4. Какова вероятность того, что случайное блуждание посетит точку  $a$ ?
5. Какова вероятность того, что случайное блуждание вернется в точку  $0$ ?
6. Сколько раз в среднем случайное блуждание посетит точку  $a$ ?
7. Сколько раз в среднем случайное блуждание посетит точку  $a$  прежде чем впервые вернется в точку  $0$ ?

Попробуйте для начала ответить на эти вопросы устно, без вычислений, интуитивно. Запишите свои ответы и потом сравните с правильными. В этой главе мы ответим на эти вопросы не произнося слова «мартингал». Поехали!

1. Пусть  $a$  и  $b$  - два натуральных числа. Какова вероятность того, что случайное блуждание никогда не достигнет ни точки  $a$ , ни точки  $-b$ ? Какова вероятность того, что  $S_n$  «коснется» точки  $a$  раньше, чем точки  $-b$ ?

Предположим, что мы стартуем не в нуле, а в произвольной точке  $x \in \{-b, -b+1, \dots, a\}$ . Обозначим интересующую нас вероятность как  $p_x$ . По определению,  $p_a = 1$  и  $p_{-b} = 0$ . Для остальных  $x$  выполнено рекуррентное соотношение:

$$p_x = 0.5p_{x-1} + 0.5p_{x+1}$$

Если разбить  $p_x$  на два одинаковых слагаемых  $p_x = 0.5p_x + 0.5p_x$ , то мы получим

$$p_x - p_{x-1} = p_{x+1} - p_x$$

Это означает, что  $p_x$  растет линейно с постоянной скоростью от  $p_{-b} = 0$  до  $p_a = 1$ . В частности,  $p_0 = \frac{b}{a+b}$ .

2. Сколько шагов в среднем потребуется, чтобы достигнуть точки  $1$ ?

Пусть  $m$  - это интересующее нас мат. ожидание. Заметим, что оно очевидно больше  $1$  и, а-приори, нельзя исключать, что оно бесконечно.

После первого шага мы либо достигли точки 1 (и нам потребовался один шаг), либо отошли еще дальше от точки 1 и нам в среднем потребуется  $m$  шагов, чтобы вернуться в 0 и еще в среднем  $m$  шагов, чтобы попасть в 1. Итого:

$$m = 0.5 \cdot 1 + 0.5(1 + 2m) \quad (38.1)$$

Хм, решений у этого уравнения нет. Точнее есть, а именно,  $m = +\infty$ .

3. Сколько шагов в среднем потребуется, чтобы достигнуть уровня  $a$  или  $-b$ ?

Пусть  $e_k$  - искомое количество ходов, при условии, что сейчас мы находимся в точке  $k$ . Краевые условия:  $e_a = 0$  и  $e_b = 0$ . Разностное уравнение:

$$e_k = 1 + 0.5e_{k-1} + 0.5e_{k+1} \quad (38.2)$$

Чтобы не мучиться с методом неопределенных коэффициентов, который, конечно, сработает, мы поступим так. Заметим, что решение - квадратный трехчлен, обнуляющийся в  $a$  и  $b$ . Значит, это что-то типа  $e_k = -(k-a)(k+b)$ . Проверяем, подходит. Находим, что  $e_0 = ab$ .

4. Какова вероятность того, что случайное блуждание посетит точку  $a$ ?

Для начала посчитаем вероятность когда-либо посетить точку 1. Обозначим ее  $p$ .

После первого шага мы либо уже в точке 1 с вероятностью 0.5, либо отошли влево. Если мы отошли влево, то чтобы посетить когда-нибудь точку 1 придется сначала когда-нибудь посетить точку 0, а затем когда-нибудь точку 1. Итого:

$$p = 0.5 \cdot 1 + 0.5p^2 \quad (38.3)$$

Единственное решение этого уравнения  $p = 1$ . По-честному, нужно еще доказать, что событие «процесс когда-нибудь достигнет точки 1» - это событие, чтобы не случилось так, что вероятность не определена.

Теперь рассмотрим произвольную точку  $a$ . Мы с вероятностью 1 когда-нибудь попадем из точки 0 в точку 1. Когда-нибудь с вероятностью 1 попадем из 1 в 2 и т.д. Значит с вероятностью 1 когда-нибудь попадем в  $a$ .

5. Какова вероятность того, что случайное блуждание вернется в точку 0?

После первого шага мы отойдем на единичку вправо (или влево). А потом с вероятностью 1 вернемся в 0.

6. Сколько раз в среднем случайное блуждание посетит точку  $a$ ?

С вероятностью 1 мы попадем в  $a$  хотя бы раз. Затем мы отойдем на единичку вправо или влево. И снова с вероятностью 1 попадем в  $a$ . Значит в среднем мы посетим  $a$  бесконечное количество раз.

7. Сколько раз в среднем случайное блуждание посетит точку  $a$  прежде чем впервые вернется в точку 0?

Задачи...

Решите все указанные задачи для несимметричного случайного блуждания.

## 11.2 Определение мартингала

**Определение 38.4.** Пусть  $(\Omega, \mathcal{F}, P)$  - вероятностное пространство. Последовательность  $\sigma$ -алгебр  $\{\mathcal{F}_t\}$  называется **фильтрацией**, если  $\mathcal{F}_n \subset \mathcal{F}_{n+1}$  и  $\mathcal{F}_n \subset \mathcal{F}$  для всех  $n$ .

Говоря простым языком,  $\mathcal{F}_n$  - это список событий, которые рациональный наблюдатель способен различить в момент времени  $n$ . С течением времени наблюдатель накапливает информацию (ничего не забывает), поэтому этот список становится все больше и больше.

**Определение 38.5.** Процесс  $X_n$  называется адаптированным к фильтрации  $\mathcal{F}_n$  если для каждого  $n$  случайная величина  $X_n$  является  $\mathcal{F}_n$ -измеримой.

Другими словами, в любой момент времени  $n$ , информации, содержащейся в  $\sigma$ -алгебре  $\mathcal{F}_n$  достаточно, чтобы сказать, чему равна случайная величина  $X_n$ . С каждым процессом всегда связана «естественная» фильтрация, а именно:  $\mathcal{F}_1 := \sigma(X_1)$ ,  $\mathcal{F}_2 := \sigma(X_1, X_2)$ ,  $\mathcal{F}_3 := \sigma(X_1, X_2, X_3)$  и т.д.

**Определение 39.1.** Случайный процесс  $X_n$  называется мартингалом по отношению к фильтрации  $\{\mathcal{F}_n\}$ , если:

- M1. Для всех  $n$  математическое ожидание  $\mathbb{E}(X_n)$  конечно.
- M2. Процесс  $X_n$  адаптирован к фильтрации  $\mathcal{F}_n$ .
- M3.  $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$  a.s.

Самое существенное условие M3 говорит: наилучший прогноз будущего значения мартингала - это его текущее значение. Поэтому бывает удобно представлять мартингал как благосостояние азартного игрока, играющего в «справедливую» игру в казино.

Иногда дают определение мартингала без явного упоминания фильтрации:

**Определение 39.2.** Случайный процесс  $X_n$  называется мартингалом, если:

- M1. Для всех  $n$  математическое ожидание  $\mathbb{E}(X_n)$  конечно.
- M3'.  $\mathbb{E}(X_{n+1}|X_n, X_{n-1}, \dots, X_1) = X_n$  a.s.

Это определение равносильно первому, если в качестве фильтрации взять естественную.

Примеры.

**Пример 39.3.** . Случайное блуждание - мартингал относительно естественной фильтрации:

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}(X_n + I_{n+1}|\mathcal{F}_n) = X_n + \mathbb{E}(I_{n+1}|\mathcal{F}_n) = X_n + \mathbb{E}(I_n) = X_n, \quad (39.4)$$

мы воспользовались тем, что  $\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n) = \sigma(I_1, I_2, \dots, I_n)$ . Поэтому  $I_{n+1}$  не зависит от  $\mathcal{F}_n$ , а  $X_n$  измерима относительно  $\mathcal{F}_n$ .

Можно обобщить идею случайного блуждания. Не обязательно прибавлять плюс или минус единицу равновероятно. Главное, чтобы добавка имела нулевое условное ожидание,  $\mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n) = 0$ . Например, если  $I_n$  - независимые случайные величины, и  $\mathbb{E}(I_n) = 0$ , то  $X_n = \sum_{i=1}^n I_i$  - мартингал.

**Пример 39.5.** . Мультипликативный мартингал.  $X_1 = 1$ , случайные величины  $I_n$  - независимы и  $\mathbb{E}(I_n) = 1$ . Рассмотрим  $X_n = \prod_{i=1}^n I_i$ . Он является мартингалом относительно естественной фильтрации:

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}(X_n \cdot I_{n+1}|\mathcal{F}_n) = X_n \cdot \mathbb{E}(I_{n+1}|\mathcal{F}_n) = X_n \mathbb{E}(I_n) = X_n, \quad (39.6)$$

Хозяйке на заметку! Если у вас в холодильнике есть случайный процесс, не являющийся мартингалом, а до прихода гостей осталось несколько минут, то из него можно сварить превосходный мартингал!

Может эти сюжету включить после? Т.к. не совсем понятно, зачем варить мартингал?

Если в холодильнике завалилась последовательность iid. . .

Пусть  $I_i$  - независимые одинаково распределенные случайные величины с  $\mathbb{E}(I_i) \neq 0$ .

Рецепт 1. Отрежем лишнее! Определим  $X_n := \sum_{i=1}^n I_i - n\mathbb{E}(I_i)$ .

Рецепт 2. Создание «мультипликативного мартингала». Найдем такое число  $a > 0$ , что  $\mathbb{E}(a^{I_i}) = 1$ .

Определим  $X_n := a^{\sum_{i=1}^n I_i}$ .

Если в холодильнике только марковская цепь. . . [Нужна ли глава про марковские цепи?]

Рецепт 3. Если  $X_n$  - цепь маркова, то можно поискать такую функцию  $f$ , что  $M_n = f(X_n)$  - мартингал.

**Пример 39.7.** Пусть  $X_1 = 1$ , а далее в каждый момент времени мы делаем шаг либо вправо, либо влево. Вероятности зависят от  $n$  и равны  $p_{n \rightarrow n+1} = \frac{n+1}{2n}$  и  $p_{n \rightarrow n-1} = \frac{n-1}{2n}$ . Найдите такое преобразование  $f$ , что  $f(X_n)$  - мартингал.

Решение. Если  $f(X_n)$  - мартингал, то и  $af(X_n) + b$  - тоже мартингал

Т.к.  $f(X_n)$  - это мартингал, то получаем следующее рекуррентное соотношение:

$$f(n) = p_{n \rightarrow n-1}f(n-1) + p_{n \rightarrow n+1}f(n+1) \quad (39.8)$$

В нашем случае:

$$f(n) = \frac{n-1}{2n}f(n-1) + \frac{n+1}{2n}f(n+1) \quad (40.1)$$

Как решить такое разностное уравнение? Можно домножить на  $n$  и сделать замену  $g(n) = nf(n)$ . После этого получим линейное уравнение и находим, что  $g(n) = c_1 + c_2n$ . Т.е.  $f(n) = \frac{c_1}{n} + c_2$ .

Мы вольны выбрать любое решение, поэтому выбираем самое простое,  $f(n) = \frac{1}{n}$ . Т.е.  $M_n = 1/X_n$  - это мартингал. Конечно,  $M_n := 0$  тоже мартингал, но он не будет полезен при нахождении вероятностей связанных с процессом  $X_n$ .

Высший пилотаж! В холодильнике всего одна случайная величина!!!

**Пример 40.2.** Рецепт 4. Мартингал Леви. Берем произвольную случайную величину  $X$  с  $\mathbb{E}(X) < \infty$  и произвольную фильтрацию  $\mathcal{F}_n$ . И строим последовательность  $X_n$  по принципу  $X_n := \mathbb{E}(X|\mathcal{F}_n)$ .

Неожиданно, но многие мартингалы устроены по принципу мартингала Леви. Даже если ...

У мартингалов есть два очень хороших свойства:

- Во-первых, многие преобразования оставляют мартингал мартингалом.
- Во-вторых, многие мартингалы сходятся. (...) сюда про Леви

Поэтому очень часто при изучении случайных процессов не являющихся мартингалами полезно увидеть мартингал.

## 11.3 Момент остановки

**Определение 40.3.** Случайная величина  $\tau$ , принимающая значения из множества  $\{1, 2, \dots\} \cup \{+\infty\}$  называется моментом остановки, если  $\{\tau = n\} \in \mathcal{F}_n$  для любого  $n$ .

Другими словами, вывод о том, что надо остановиться в момент времени  $n$  можно сделать в момент времени  $n$ . Возможно, что вывод может быть сделан и раньше (так как вполне возможно, что событие  $\{\tau = n\}$  содержится не только в  $\sigma$ -алгебре  $\mathcal{F}_n$ , но и в более ранней).

**Пример 40.4.** Момент времени, когда индекс Доу-Джонса, достигает своего ежегодного максимума, не является моментом остановки. Этот момент времени не обнаруживается сразу, чтобы понять, что он был в момент времени  $n$ , нужно ждать конца года. Мы не можем использовать стратегию типа «когда индекс достигнет своего годового максимума я сделаю то-то и то-то». Хотя, безусловно, этот момент времени - случайная величина.

**Пример 40.5.** Момент первого касания (first hitting time). Момент  $\tau$  - это тот день, когда индекс Доу-Джонса впервые превысит 1000,  $\tau = \min\{n | X_n \geq 1000\}$ . Если траектория такова, что индекс никогда не превышает 1000, то  $\tau = +\infty$ .

Дадим эквивалентное определение:

Упражнение. Убедитесь, что определение не изменится, если  $\{\tau = n\}$  заменить на  $\{\tau \leq n\}$  или на  $\{\tau > n\}$

Упражнение. Если  $\tau$  - момент остановки, то множества  $\{\tau < n\}$ ,  $\{\tau \leq n\}$ ,  $\{\tau = n\}$ ,  $\{\tau \geq n\}$  и  $\{\tau > n\}$  лежат в  $\mathcal{F}_n$ .

Упражнение.

Верно ли, что  $\tau = a$  - момент остановки?

Пусть  $\tau_1$  и  $\tau_2$  - моменты остановки, верно ли что  $\tau_1 \wedge \tau_2$ ,  $\tau_1 \vee \tau_2$ ,  $\tau_1 + a$  и  $\tau_1 + \tau_2$  - моменты остановки?



### 11.3.1 Информация к моменту остановки.

Каждому моменту остановки  $\tau$  можно сопоставить  $\sigma$ -алгебру, отражающую всю информацию, известную к моменту времени  $\tau$ .

**Определение 41.0.** Пусть  $\tau$  - момент остановки.

$$\mathcal{F}_\tau := \{A \in \mathcal{F} | A \cap \{\tau = n\} \in \mathcal{F}_n, \forall n\} \quad (41.1)$$

**Определение 41.2.** Если  $\tau$  - время первой аварии на недавно построенном шоссе, то событие  $A = \{\text{за день до аварии был туман}\}$  лежит в  $\mathcal{F}_\tau$ . Это событие может не содержаться ни в одной  $\mathcal{F}_n$ .

Событие  $B = \{\text{на следующий день после аварии был туман}\}$  в  $\mathcal{F}_\tau$  может не содержаться.

В  $\mathcal{F}_\tau$  входят те события, про которые в момент времени  $\tau$  точно можно будет сказать, произошли они или нет.

Если возможно несколько сценариев развития события, при одном из которых мы в момент времени  $N$  будем знать, произошло ли  $A$  или нет, а при другом - не будем знать, то  $A$  в  $\mathcal{F}_\tau$  не входит.

$\mathcal{F}_\tau$  - это не случайный объект а 'нормальная'  $\sigma$ -алгебра.

Полезно представлять себе  $\mathcal{F}_n$  как газету, выходящую в день  $n$ , а  $\sigma$ -алгебру  $\mathcal{F}_\tau$  как спецвыпуск газеты, посвященный произошедшей аварии (или другому наступившему моменту остановки).

Упражнение.

Определение не изменится, если  $\tau = n$  заменить на  $\tau \leq n$ .

Упражнение.

Если  $\tau$  - момент остановки, то  $\mathcal{F}_\tau$  - действительно  $\sigma$ -алгебра и  $\tau \in \mathcal{F}_\tau$ .

Упражнение.

$N \equiv n \Rightarrow \mathcal{F}_N = \mathcal{F}_n$  и  $N \equiv +\infty \Rightarrow \mathcal{F}_N = \mathcal{F}$

Упражнение.

Если  $\tau < +\infty$  - момент остановки и  $Y_n \in \mathcal{F}_n$ , то  $Y_\tau \in \mathcal{F}_\tau$ .

В частности, из этого упражнения следует, что  $\sum_{i=1}^{\tau} Y_i$  и  $\max_{n \leq \tau} \sum_{i=1}^n Y_i$  являются  $\mathcal{F}_\tau$ -измеримыми.

Упражнение.

Если  $M$  и  $N$  - моменты остановки, то  $\mathcal{F}_M \cap \mathcal{F}_N = \mathcal{F}_{M \wedge N}$

"Доказательство": Вынесем  $\mathcal{F}$  за скобки и заострим  $\cap$  :)

Упражнение. Если  $N$  - момент остановки, то  $\mathcal{F}_{N \wedge n} \subseteq \mathcal{F}_n$

Упражнение.

Пусть  $M \leq N$  - моменты остановки. Как связаны между собой  $\sigma$ -алгебры  $\mathcal{F}_M$  и  $\mathcal{F}_N$ ?

Упражнение.

Пусть  $T$  - момент остановки,  $A \in \mathcal{F}_T$  и  $m \leq n$ . Тогда  $A \cap \{T = m\} \in \mathcal{F}_{T \wedge n}$

Упражнение.

Пусть  $T_1, T_2, T_3, \dots$  - моменты остановки и существует  $T = \lim_{n \rightarrow +\infty} T_n$ . Тогда  $T$  - момент остановки.

Упражнение.

Пусть  $T_1 \geq T_2 \geq T_3 \geq \dots$  - последовательность моментов остановки с пределом  $T$ . Докажите, что  $\mathcal{F}_T = \bigcap_n \mathcal{F}_{T_n}$ .

Пусть  $T_1 \leq T_2 \leq T_3 \leq \dots$  - последовательность моментов остановки с конечным пределом  $T$ . Докажите, что  $\mathcal{F}_T = \sigma(\bigcup_n \mathcal{F}_{T_n})$ .

Приведите пример к предыдущему утверждению, доказывающий необходимость конечности  $T$ .

Подсказка:

$$\Omega = \bigcup_n \bigcup_{L=1}^{\infty} \bigcap_{l=L}^{\infty} (T_l = T = n)$$

## 11.4 Остановленный мартингал

На стиль изложения во многом повлияли источники [stirzaker:prp], [chang:sp], [ross:scp]

Пусть  $X_t$  - наше благосостояние в справедливой игре в момент времени  $t$  или мартингал. Наша стратегия заключается в том, чтобы в нужный момент завершить игру, скажем, после крупного выигрыша. Другими словами, наша стратегия определяется моментом остановки  $T$ . Этот момент остановки - случайная величина, так как может зависеть от хода игры. Случайной величиной является также и  $X_T$  - выигрыш на момент выхода из игры. Вопрос в том, чему равен средний выигрыш на момент прекращения игры,  $\mathbb{E}(X_T)$ ?

Ответ дает теорема Дуба: Если не ждать «слишком» долго, то каким бы хитрым не был момент остановки ожидаемый выигрыш будет равен начальной сумме,  $\mathbb{E}(X_T) = \mathbb{E}(X_1)$ .

Сразу приведем пример «слишком» долгого ожидания: ждать до выигрыша в один рубль в классическом случайном блуждании. Условие  $\mathbb{E}(X_T) = \mathbb{E}(X_1)$  здесь нарушено: выигрыш на момент выхода из игры равен одному рублю (по построению), а стартовая сумма равна нулю. Почему это слишком долгое ожидание? Потому, что в этом случае можно доказать, что  $\mathbb{E}(T) = +\infty$ .

Заметим, что для того, чтобы случайная величина  $X_T$  была определена (хотя бы почти наверное) необходимо, чтобы  $\mathbb{P}(T < \infty) = 1$ .

Точный смысл понятия «слишком» долго можно увидеть в теореме:

Если  $X_t$  - мартингал,  $T$  - момент остановки,  $\mathbb{P}(T < \infty) = 1$ , и выполнено хотя бы одно из пяти условий:

TD1. Момент  $T$  ограничен, то есть существует число  $M$ , такое что  $T < M$ .

TD2. Процесс  $X_{t \wedge T}$  ограничен, то есть существует число  $M$ , такое что для любого  $t$  верно неравенство  $|X_{t \wedge T}| < M$ .

TD3.  $\mathbb{E}(T) < +\infty$  и существует число  $M$ , такое что для любого  $t$  верно неравенство  $\mathbb{E}(|X_{t+1} - X_t| | \mathcal{F}_t) < M$ .

TD4.  $\mathbb{E}(|X_T|) < \infty$  и  $\lim_{t \rightarrow \infty} \mathbb{E}(X_t 1_{T > t}) = 0$ .

TD5. Мартингал  $X_t$  является равномерно интегрируемым.

То:  $\mathbb{E}(X_T) = \mathbb{E}(X_1)$ .

В большинстве случаев первых трех критериев достаточно для практического применения. Четвертый критерий является следствием любого из первых трех (?).

В пятом критерии используется определение равномерной интегрируемости. . . Набор случайных величин является равномерно интегрируемым, если. . .

Из этих критериев все кроме третьего (может есть какой-то «предельный» аналог и третьего? - я не знаю) работают в непрерывном времени.

Из теоремы Дуба легко вывести тождество Вольда. Wald's identity.

Если складывать случайное количество случайных величин (опять же, не «слишком» много), то среднее значение суммы равно произведению среднего размера слагаемых на среднее количество слагаемых. Это кажется очевидным, но на самом деле не все так просто, так теорема допускает, что количество слагаемых может зависеть от размера слагаемых.

Если  $X_i$  - независимые одинаково распределенные случайные величины и  $T$  - случайная величина принимающая целые неотрицательные значения и  $\mathbb{E}(T) < \infty$ , то  $\mathbb{E}(\sum_{i=1}^T X_i) = \mathbb{E}(X_i) \mathbb{E}(T)$ .

Доказательство:

Случайный процесс  $M_t = \sum_{i=1}^t X_i - t \cdot \mathbb{E}(X_i)$  - мартингал. Применяя к нему теорему Дуба получаем, что  $\mathbb{E}(M_T) = \mathbb{E}(M_1)$ . Здесь  $\mathbb{E}(M_1) = 0$ ,  $M_T = \sum_{i=1}^T X_i - T \cdot \mathbb{E}(X_i)$ , поэтому  $\mathbb{E}(\sum_{i=1}^T X_i) = \mathbb{E}(X_i) \mathbb{E}(T)$ .

Между примерами и задачами нет существенного различия. Если легко, то можно взять условие примера и попробовать его решить как задачу. А если тяжело, то можно посмотреть решение задачи. Пожалуй, тексты примеров в каком-то смысле «классические». Иногда используемые для решения мартингалы настолько изящны, что вспоминаются субтитры при показе сложных трюков по телевизору: «Трюки выполнены профессиональными каскадерами. Не пробуйте повторить их самостоятельно». Только здесь нужно пробовать!

**Задача. Симметричное случайное блуждание.**

Улитка начинает свой путь в точке 0 и за каждую минуту равновероятно смещается влево или вправо на один сантиметр<sup>1</sup>. Справа от улитки на расстоянии  $a$  находится виноградное дерево, слева на расстоянии  $b$  - шелковица.

Какова вероятность того, что улитка доползет до виноградного дерева раньше? Сколько в среднем времени ей потребуется чтобы доползти до любого из деревьев?

Решение с мартингалами:  $X_t$  - координата улитки в момент времени  $T$  - мартингал. Пусть  $T$  - момент достижения одной из границ ( $a$  или  $-b$ ), тогда  $|X_{t \wedge T}| \leq \max\{a, b\}$ . Момент времени  $T$  можно смажорировать геометрической случайной величиной. Разобьем время на интервалы по  $(a + b)$  шагов. Будем засчитывать выход за пределы множества  $(-b; a)$  только на границах этих интервалов. Очевидно, что требуемое для выхода время от этого только возрастет.

За каждый интервал вероятность выхода случайного блуждания за пределы множества  $(-b; a)$  выше чем  $(1/2)^{a+b}$  (это вероятность того, что случайное блуждание сделает  $(a + b)$  шагов вправо). Значит требуемое для выхода количество временных интервалов можно ограничить сверху случайной величиной  $M$  с геометрическим распределением и вероятностью успеха равной  $p = (1/2)^{a+b}$ . Получаем, что  $\mathbb{P}(T = \infty) \leq \mathbb{P}(M = \infty) = 0$  и  $\mathbb{E}(T) \leq (a + b)\mathbb{E}(M) = (a + b)2^{a+b}$ .

Применяя пункт (ii) теоремы Дуба получаем, что  $\mathbb{E}(X_T) = \mathbb{E}(X_1)$ . По условию  $\mathbb{E}(X_1) = 0$ . Величина  $X_T$  может принимать только два значения  $a$  или  $-b$ . Пусть значение  $a$  принимается с вероятностью  $p$ . Тогда  $pa + (1 - p)(-b) = 0$  и  $p = \frac{b}{a+b}$ .

Заметим также, что  $M_t = X_t^2 - t$  - также мартингал:  $\mathbb{E}(M_{t+1} - M_t | \mathcal{F}_t) = \mathbb{E}(2X_t \Delta X_{t+1} + (\Delta X_{t+1})^2 - 1 | \mathcal{F}_t) = 2X_t \mathbb{E}(\Delta X_{t+1} | \mathcal{F}_t) + \mathbb{E}((\Delta X_{t+1})^2) - 1 = 0 + 1 - 1 = 0$ .

Применяя ??? пункт (iii) теоремы Дуба получаем, что  $\mathbb{E}(M_T) = \mathbb{E}(M_1)$ . При этом  $\mathbb{E}(M_1) = 0$ , а  $M_T = X_T^2 - T$ , значит  $\mathbb{E}(X_T^2) = \mathbb{E}(T)$ . Величина  $X_T^2$  принимает только два значения:  $a^2$  и  $b^2$ , значит  $\mathbb{E}(T) = a^2 \frac{b}{a+b} + b^2 \frac{a}{a+b} = ab$ .

**Задача. Несимметричное случайное блуждание**

Улитка начинает свой путь в точке 0 и за каждую минуту смещается влево или вправо на один сантиметр. Однако влево идти улитке не очень хочется, поэтому вправо она идет с вероятностью  $p > 1/2$ , влево - с вероятностью  $(1 - p)$ . Справа от улитки на расстоянии  $a$  находится виноградное дерево, слева на расстоянии  $b$  - шелковица.

Какова вероятность того, что улитка доползет до виноградного дерева раньше? Сколько в среднем времени ей потребуется чтобы доползти до любого из деревьев?

Решение с мартингалами:  $X_t$  - координата улитки в момент времени  $T$  - не мартингал, т.к. в улитка склонна смещаться вправо.

Как сварить из немартингала мартингал? Один из стандартных приемов следующий: вместо процесса  $X_t$  рассмотреть процесс  $M_t = u^{X_t}$ , где число  $u$  подбирается так, чтобы  $M_t$  стал мартингалом. Из нескольких возможных  $u$  выбирают  $u \neq 1$ , т.к. при  $u = 1$  процесс  $M_t$  всегда будет равен 1.

Ищем  $u$  в нашем случае. Требуем, чтобы  $M_t$  был мартингалом:  $\mathbb{E}(M_{t+1} | \mathcal{F}_t) = M_t$ . Это условие упрощается до  $\mathbb{E}(u^{\Delta X_{t+1}}) = 1$  или  $p \cdot u^1 + (1 - p) \cdot u^{-1} = 1$ . Квадратное уравнение имеет два корня,  $u = 1$  и  $u = \frac{1-p}{p}$ , выбираем второй.

Как и в случае симметричного случайного блуждания улитки  $\mathbb{P}(T = \infty) = 0$ ,  $\mathbb{E}(T) < \infty$  так как время  $T$  снова мажорируется геометрической случайной величиной. Также и процесс  $M_{t \wedge T}$  ограничен сверху.

<sup>1</sup>Скорость виноградной улитки около 4 см в минуту, но у нас улитка попала ленивая.

Применяя пункт (ii) теоремы Дуба получаем, что  $\mathbb{E}(M_T) = \mathbb{E}(M_1)$ . Величина  $M_T$  принимает всего два значения:  $u^a$  и  $u^{-b}$ , а  $\mathbb{E}(M_1) = 1$ . Пусть  $p$  - вероятность допозлти сначала до точки  $a$  (виноградного дерева). Из уравнения  $pu^a + (1-p)u^{-b} = 1$  находим  $p = \dots$ .

Чтобы найти  $\mathbb{E}(T)$  можно использовать мартингал  $Y_t = X_t - t \cdot \mathbb{E}(\Delta X_t)$  или, что в принципе то же самое, тождество Вольда. По тождеству Вольда среднее количество слагаемых равно среднему значению суммы делить на средний размер слагаемого:  $\mathbb{E}(T) = \frac{\mathbb{E}(Y_T)}{\mathbb{E}(\Delta X_t)} = \dots$

Задача. ABRACADABRA [ross:scp][williams:pwm]

Если одна мартышка печатает наугад буквы на клавиатуре, то она рано или поздно с единичной вероятностью напечатает роман Льва Толстого «Война и мир». Мы поставим вопрос по-другому: сколько нажатий на клавиши в среднем потребуется чтобы напечатать слово «абракадабра»? Какова дисперсия требуемого количества нажатий?

Решение через мартингалы. Организуем казино! Перед каждым нажатием в казино приходит новый игрок с начальным капиталом в 1 рубль. Каждый входящий игрок действует по одной и той же схеме: ставит все имеющиеся деньги на очередную букву слова АБРАКАДАБРА (войдя в казино игрок ставит на А, потом на Б, потом на Р и т.д. Если слово кончилось, то игрок покидает казино с выигрышем). Если обезьяна напечатала нужную букву, то игрок получает свою ставку, увеличенную в 33 раза, если нет - то игрок покидает казино без денег.

Пусть  $X_t$  - суммарное благосостояние всех игроков, начавших игру до нажатия номер  $t$ . Величина  $X_t$  естественно раскладывается на благосостояния отдельных игроков пришедших в казино  $X_t = Y_{1t} + \dots + Y_{tt}$ . Здесь  $Y_{it}$  - это благосостояние  $i$ -го игрока после нажатия номер  $t$ . Заметим, что для каждого игрока  $Y_{it}$  - это мартингал, т.к. игра справедливая - на каждом шаге игрок либо теряет деньги, либо с вероятностью  $1/33$  увеличивает свой выигрыш в 33 раза. Так как игроки начинают с одного рубля, то  $\mathbb{E}(Y_{it}) = 1$ . Чистый выигрыш игроков к моменту времени  $t$ ,  $M_t = X_t - t$  оказывается мартингалом:  $\mathbb{E}(M_{t+1} - M_t | \mathcal{F}_t) = \mathbb{E}(Y_{1,t+1} - Y_{1,t} | \mathcal{F}_t) + \dots + \mathbb{E}(Y_{t,t+1} - Y_{t,t} | \mathcal{F}_t) + \mathbb{E}(Y_{t+1,t+1} - 1 | \mathcal{F}_t) = 0$ .

Заметим, что когда слово «абракадабра» напечатано впервые, все игроки кроме трех проиграли по одному рублю. Угадавший «абракадабра» имеет  $33^{11}$ , угадавший «абра» имеет  $33^4$  и угадавший «а» имеет 33 рубля. Значит  $M_T = 33^{11} + 33^4 + 33 - T$ . Также получаем, что  $M_t 1_{T>t} \leq (33^{11} + 33^4 + 33) 1_{T>t}$ .

Время  $T$  можно смажорировать с помощью геометрической случайной величины. Можно давать обезьяне чистый лист после каждый 11 нажатий и засчитывать «абракадабру», только если слово напечатано на отдельном листе. Очевидно, что ожидаемое время при этом возрастет. Количество листов  $L$  будет иметь геометрическое распределение с  $p = \frac{1}{33^{11}}$  и  $\mathbb{E}(L) = 33^{11}$ . Значит  $T < 11L$  и  $\mathbb{E}(T) < \mathbb{E}(11 \cdot L) = 11 \cdot 33^{11}$ .

Для геометрического распределения  $\mathbb{P}(L > t)$  стремится к нулю при  $t \rightarrow \infty$ . Значит  $\mathbb{E}(M_t 1_{T>t}) \leq (33^{11} + 33^4 + 33) \mathbb{P}(L > t/11) \leq (33^{11} + 33^4 + 33) \mathbb{P}(L < t/11)$  стремится к нулю.

Применяя пункт (iv) теоремы Дуба получаем, что  $\mathbb{E}(M_T) = \mathbb{E}(M_1)$ . При этом  $\mathbb{E}(M_T) = (33^{11} + 33^4 + 33) - \mathbb{E}(T)$ , а  $\mathbb{E}(M_1) = 0$ , значит  $\mathbb{E}(T) = 33^{11} + 33^4 + 33$ .

Про дисперсию. Теперь игроки приходят в наше казино с возрастающими суммами денег: первый с одним рублем, второй - с двумя и т.д. Снова рассмотрим  $X_t$ , суммарное благосостояние игроков вступивших в игру к моменту времени  $t$  и чистый выигрыш этих игроков,  $M_t = X_t - (1 + 2 + \dots + t) = X_t - \frac{t(t+1)}{2}$ .

При данном изменении выигравших будет также трое и на момент появления «абракадабры» общий чистый выигрыш составит  $M_T = (T - 10) \cdot 33^{11} + (T - 3) \cdot 33^4 + T \cdot 33 - \frac{T(T+1)}{2}$ .

Как и раньше  $\mathbb{E}(T) < \infty$  и  $\mathbb{E}(M_t 1_{T>t}) \rightarrow 0 \dots$

Применяя пункт (iv) теоремы Дуба получаем, что  $\mathbb{E}(M_T) = \mathbb{E}(M_1)$ . Значит  $\mathbb{E}(M_T) = 0$ . Из этого уравнения выражается  $\mathbb{E}(T^2)$ , а затем и по формуле  $Var(T) = \mathbb{E}(T^2) - \mathbb{E}(T)^2$  находим дисперсию  $Var(T) = (33^{11} + 33^4 + 33)^2 - (11 \cdot 33^{11} + 4 \cdot 33^4 + 33)$ .

Задача. «Следующая карта - дама» [morters:m], [winkler:gdpd]

Перед гадалкой колода из 36 карт, хорошо перемешанная. Гадалка должна предсказать появление дамы. Гадалка открывает одну за другой карты из колоды и в любой момент может остановиться и сказать «Следующая карта будет дамой». Если это окажется правдой, то гадалка выиграла.

Какую вероятность выигрыша дает следующая стратегия: дожидаться появления первой дамы и после этого рискнуть и заявить, что следующая карта будет дамой? Какова оптимальная стратегия? Каковы при шансы выиграть при оптимальной стратегии?

Решение через мартингалы. Рассмотрим процесс  $X_t$  - долю дам в еще неоткрытой части колоды после  $t$  открытых карт. После  $t$  открытых карт в колоде остается  $(36 - t)$  карт, из которых  $(36 - t)X_t$  дам. С вероятностью  $X_t$  (доля дам совпадает с вероятностью открыть даму) количество дам уменьшится на одну.

$$\text{Считаем } \mathbb{E}(X_{t+1}|\mathcal{F}_t) = \mathbb{E}(X_{t+1}|X_t) = \frac{(36-t)X_t + X_t(-1)}{36-t-1} = X_t.$$

Значит,  $X_t$  - мартингал. Момент остановки  $T$  в любом случае ограничен 36 картами, следовательно, требования пункта (i) теоремы Дуба выполнены и  $\mathbb{E}(X_T) = \mathbb{E}(X_0) = \frac{4}{36}$  для любой стратегии.

Решение без мартингалов. Шансы гадалки не меняются, если она будет угадывать последнюю карту, а не следующую: информации о последней карте ровно столько, сколько о следующей. А шансы того, что последняя карта будет дамой равны  $\frac{4}{36}$ . Значит все стратегии оптимальны и дают выигрыш в  $4/36$

Задача. «И в воздух чепчики бросали...»

Приезжающих из армии или от двора встречают  $n$  женщин. Они одновременно подбрасывают вверх  $n$  чепчиков. Ловят чепчики наугад, каждая женщина ловит один чепчик. Женщины, поймавшие свой чепчик уходят. А женщины, поймавшие чужой чепчик, снова подбрасывают его вверх. Подбрасывание чепчиков продолжается до тех пор, пока каждая не поймает свой чепчик.

Найдите:

- а) среднее количество женщин, поймавших свой чепчик при одном подбрасывании
- б) среднее количество подбрасываний

Решение с мартингалами. Пусть  $N_1$  - количество женщин поймавших чепчики при первом подбрасывании. Величина  $N$  легко раскладывается в сумму индикаторов:  $N_1 = X_1 + \dots + X_n$ , где  $X_i$  - принимает значение 0 или 1 в зависимости от того, поймала ли  $i$ -ая женщина свой чепчик. Находим  $\mathbb{E}(X_i) = \mathbb{P}(X_i = 1) = 1/n$  и  $\mathbb{E}(N_1) = n \cdot \frac{1}{n} = 1$ . Аналогично и для остальных раундов, где участвует меньшее количество женщин.

Значит в каждом раунде в среднем одна женщина ловит свой чепчик. Пусть  $Y_t = N_1 + \dots + N_t$  - суммарное количество женщин поймавших свой чепчик после раунда  $t$ . Получаем, что  $\mathbb{E}(Y_t) = t$ . Случайный процесс  $M_t = Y_t - t$  - это мартингал, так как  $\mathbb{E}(M_{t+1}|\mathcal{F}_t) = \mathbb{E}(M_{t+1}|M_t) = \mathbb{E}(M_t + N_{t+1} - 1|M_t) = M_t$ . Разность  $M_{t+1} - M_t = N_{t+1} - 1$  ограничена количеством женщин  $n$ .

Смажорируем  $T$ , чтобы увидеть, что  $\mathbb{E}(T) < \infty$ . Вероятность того, что конкретная женщина поймает свой чепчик равна единице делить на количество женщин, участвующих в раунде, а следовательно не меньше  $1/n$ . Если вместо подбрасывания чепчиков в каждом раунде одна (любая) женщина будет уходить с вероятностью  $1/n$ , а с вероятностью  $1 - 1/n$  не будет уходить никто, то ожидаемое количество раундов возрастет (занижена вероятность ухода, возможность ухода нескольких женщин за один раунд исключена). Количество раундов для ухода одной женщины в этом случае будет иметь геометрическое распределение со средним значением  $n$  раундов, а ожидаемое общее количество раундов равно  $n^2$ . Следовательно,  $\mathbb{E}(T) < n^2$ .

Применяя пункт (iii) теоремы Дуба получаем, что  $\mathbb{E}(M_T) = \mathbb{E}(M_1)$ . Так как  $M_T = Y_T - T = n - T$  и  $\mathbb{E}(M_1) = \mathbb{E}(N_1 - 1) = 0$ , получаем, что  $\mathbb{E}(T) = n$ .

Ключи и сейфы [aops:keys]

Ballot problem [ross:scp]

Задача. РРО против ОРО[li:ma].

Правильную монетку подбрасывают до появления последовательности РРО или ОРО. Сколько подбрасываний в среднем нужно? Какова вероятность того, что подбрасывания закончатся

последовательностью РРО?

Решение с мартингалами. ???

Решение без мартингалов. ???

Задача. «Вампиры-гладиаторы» [winkler:gdpd]

Две команды вампиров-гладиаторов борются за победу в турнире. В вашей команде 100 вампиров с силами от  $1, 2, \dots, 100$ . В команде противника 59 вампиров с силами  $72, 73, \dots, 130$ . Турнир состоит из последовательных раундов в каждом из которых участвует по одному гладиатору с каждой стороны. Если встретились гладиаторы с силами  $a$  и  $b$ , то первый побеждает с вероятностью  $\frac{a}{a+b}$ , а второй - с вероятностью  $\frac{b}{a+b}$ . Победитель добавляет к своей силе силу побежденного (получает силу  $a + b$ ), а побежденный выбывает из турнира (получает силу 0). Турнир продолжается до полного выбывания одной из команд.

Вы знаете, что команда противника будет выставлять гладиаторов по следующему принципу: на арену всегда выходит самый слабый из команды.

Какова ваша оптимальная стратегия? Какова вероятность выигрыша при этой стратегии?

Решение с мартингалами. Пусть  $X_t$  - суммарная сила гладиаторов нашей команды после  $t$  боев. Процесс  $X_t$  - мартингал:  $\mathbb{E}(X_{t+1} - X_t | \mathcal{F}_t) = \frac{a}{a+b}b + \frac{b}{a+b}(-a) = 0$ .

Момент окончания турнира  $T$  ограничен, т.к. за каждый раунд выбывает ровно один гладиатор, то есть  $T < 100 + 59$ .

Применяя пункт (i) теоремы Дуба получаем, что  $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ . Начальная суммарная сила нашей команды,  $X_0 = 50 \cdot 101$ . В конце турнира суммарная сила  $X_T$  может принимать два значения: либо 0, если мы проиграли, либо  $50 \cdot 101 + 59 \cdot 101$ . Если  $p$  - вероятность нашей победы, то  $(1 - p) \cdot 0 + p \cdot (109 \cdot 101) = 50 \cdot 101$ . Значит  $p = \frac{50}{109}$ . Причем результат не зависит от используемой стратегии, т.к. она нигде не использовалась при подсчете!

Ряд задач взят из [stirzaker:prp], [stirzaker:otep], [zastawniak:bsp], [blom:pspt]

Задача.

Пусть  $X_t$  - симметричное случайное блуждание.

Найдите  $\mathbb{E}(TS_T)$ ,

Задача. [wilmott:chap]

Бабушка изготовила кисель. В банке киселя плавают 10 вишенек. За один день Вовочка выпивает случайное количество киселя равномерно распределенное от нуля до всей банки. Вовочка пьет кисель прямо с вишенками, если они ему попадают. Чтобы бабушка ничего не заметила каждый день Вовочка доливают в банку воды до полного объема. Вовочка пьет до тех пор пока в банке не останется 5 вишенек. Пусть  $T$  - количество дней, которые Вовочка будет пить кисель, а  $X_t$  концентрация киселя в день  $t$ .

Докажите, что  $\mathbb{E}(T) = -\mathbb{E}(\ln X_T)$

Решение. Заметим, что  $X_{t+1} = X_t \cdot U_{t+1}$ , где  $U_t$  — доля невыпитого Вовочкой в день  $t$  равномерно распределена на  $[0; 1]$ . Применяем народную примету: если есть умножение, значит логарифм где-то рядом. Поигравшись получаем мартингал:  $M_t = \ln(X_t) + t$ . Применяем к нему теорему Дуба:  $\mathbb{E}(\ln X_T) + \mathbb{E}(T) = 0$ .

Задача. Вариация на тему дней рождения.

Мы набираем людей по одному в группу до тех пор, пока в группе не будет хотя бы одного совпадающего дня рождения. Пусть  $T$  количество людей, которое потребуется набрать. Найдите<sup>1</sup>  $\mathbb{E}(T^2) - \mathbb{E}(T)$ .

Решение через мартингалы. Каждый вступающий человек приходит в группу с количеством денег равным количеству людей уже вступивших в группу. Обяжем каждого вступающего человека сыграть с каждым уже вступившим в группу в такую лотерею: входящий ставит на кон 1 рубль, если их дни рождения совпадают, то входящий получает 365 рублей, если нет, то входящий теряет рубль (деньги платит и получает устроитель, существующие члены группы

<sup>1</sup>величина  $\mathbb{E}(T)$  является «некрасивой» в том смысле, что не целая и точное значение имеет громоздкую запись. Оказывается, что  $\mathbb{E}(T) \approx 23$ , но для решения данной задачи это не используется

ничего не платят и не получают). Пусть  $X_t$  - чистый выигрыш всех участников после вступления в группу  $t$  человек. Поскольку каждая лотерея по отдельности справедлива, то  $X_t$  - мартингал.

Момент остановки  $T$  ограничен сверху,  $T \leq 365$ . Применяя пункт (i) теоремы Дуба получаем, что  $\mathbb{E}(X_T) = \mathbb{E}(X_1)$ . При этом  $\mathbb{E}(X_1) = 0$ , а  $X_T = 365 - (1 + 2 + \dots + (T - 1)) = 365 - \frac{T(T-1)}{2}$ . И  $\mathbb{E}(T^2 - T) = 2 \cdot 365$ .

Задача. Улитка отдыхает (симметричное блуждание), [blom:pspt]

Улитка начинает свой путь в точке 0 и за каждую минуту равновероятно смещается влево или вправо на один сантиметр или отдыхает никуда не перемещаясь. Влево (и вправо) улитка ползет с одинаковой вероятностью  $p$ , отдыхает - с вероятностью  $(1 - 2p)$ .

Справа от улитки на расстоянии  $a$  находится виноградное дерево, слева на расстоянии  $b$  - шелковица.

Какова вероятность того, что улитка доползет до виноградного дерева раньше? Сколько в среднем времени ей потребуется чтобы доползти до любого из деревьев?

Задача. Улитка отдыхает (несимметричное блуждание), [blom:pspt]

Улитка начинает свой путь в точке 0 и за каждую минуту смещается влево или вправо на один сантиметр или отдыхает никуда не перемещаясь. Влево улитка ползет с вероятностью  $p_l$ , вправо - с вероятностью  $p_r$ , отдыхает - с вероятностью  $(1 - p_l - p_r)$ .

Справа от улитки на расстоянии  $a$  находится виноградное дерево, слева на расстоянии  $b$  - шелковица.

Какова вероятность того, что улитка доползет до виноградного дерева раньше? Сколько в среднем времени ей потребуется чтобы доползти до любого из деревьев?

Задача. Ждем 1 рубль.

Найдите  $\mathbb{E}(T)$

Задача. Войны Добра и Зла [stirzaker:otep], 12-13

Изначально воюют один воин Добра и один воин Зла. Судьба выбирает одного из воюющих наугад и добавляет еще одного воина той же стороны. Пусть  $T$  - время, когда Судьба впервые добавит война Добра. Найдите  $\mathbb{E}(\frac{T}{T+2})$

Войны Добра и Зла. (Взято из Kingman, Martingales in the OK Corral)

На одном холме выстроились  $d$  воинов Добра, а на другом - столько же воинов Зла. Обычные стрелы не долетают, слишком велико расстояние. Но каждый день Судьба выбирает одного из всех воинов наугад и вкладывает в его колчан Золотую Стрелу. Золотая стрела бьет без промаха. Великая Битва длится пока одна из сторон не будет полностью побеждена. Пусть случайная величина  $S$  - количество выживших в этой Битве. Обозначим,  $X_t$  и  $Y_t$  - количество воинов Добра и Зла в конце дня  $t$ . Разберитесь с асимптотикой  $\mathbb{E}(S^4)$  при  $d \rightarrow \infty$ . Подсказка: может быть найдется функция  $f$  такая, что  $f(X_t, Y_t)$  - мартингал?

Решение.

Из того, что  $f(X_t, Y_t)$  - мартингал получаем:

$$f(m, n) = \frac{n}{n+m} f(m-1, n) + \frac{m}{n+m} f(m, n-1) \quad (47.1)$$

У этого уравнения куча решений, поищем чего-нибудь простое... Попробуем многочлены невысокой степени... Находим,  $f(m, n) = (m + 0.5)^2 - (n + 0.5)^2$ ,  $f(m, n) = (m + n + 1)(m + n + 2)(m + n + 3(m - n)^2)$ , ...

Если применить теорему Дуба к первому получает  $0 = 0$ , вот ко второму уже интереснее... Итак, рассматриваем мартингал

$$M_t = (X_t + Y_t + 1)(X_t + Y_t + 2)(X_t + Y_t + 3(X_t - Y_t)^2) = (2d - t + 1)(2d - t + 2)(2d - t + 3(X_t - Y_t)^2) \quad (47.2)$$

Применяем к нему теорему Дуба:

$$(2d + 1)(2d + 2)2d = M_0 = \mathbb{E}(M_\tau) = \mathbb{E}((S + 1)(S + 2)(S + 3S^2)) \quad (47.3)$$

И мы получаем, что  $\mathbb{E}(S^4) \sim \frac{8}{3}d^3$ .

Задача. Банковский счет улитки. [stirzaker:prp], 12.5

Перед отправкой в свой долгий путь улитка положила все свои сбережения (один рубль) в банк. За каждую минуту банк начисляет небольшой процент  $r > 0$ , такой что  $0 < r < 1/\cos(\frac{\pi}{a+b}) - 1$ . Какая сумма в среднем будет находится на счету улитки к моменту достижения ей любого из деревьев?

Задача. Сумма координат улитки. [stirzaker:prp], 12.7

Каждый раз проходя через точку с координатой  $x$  улитка получает  $x$  рублей. Каков суммарный заработок улитки?

Задача. Улитка на плоскости. [stirzaker:otep], 12-17

**Пример 48.1.** Пусть  $X_1 = 1$ , а далее в каждый момент времени мы делаем шаг либо вправо, либо влево. Вероятности зависят от  $n$  и равны  $p_{n \rightarrow n+1} = \frac{(n+1)^2}{n^2+(n+1)^2}$  и  $p_{n \rightarrow n-1} = \frac{n^2}{n^2+(n+1)^2}$ . Если процесс попадает в точку ноль, то происходит Большой Взрыв. Найдите какую-нибудь  $f$ , чтобы  $f(X_n)$  было мартингалом. А затем посчитайте вероятность того, что когда-нибудь будет Большой Взрыв.

Решение. Для начала уточним мы будем работать с «остановленным» процессом. Т.е. будем считать, что если  $X_n$  достиг точки 0, то он остается равным нулю навсегда. Поэтому величина  $f(0)$  не имеет никакого значения - можем выбрать как хотим. Величина  $f(1)$  влияет только на математическое ожидаемое мартингала. Давайте возьмем самые простые значения  $f(0) = 0$  и  $f(1) = 1$ . Хотя можно было взять и другие. Если взять совсем тривиальные  $f(0) = f(1) = 0$ , то тоже получится мартингал, но он будет скучным,  $M_n := 0$ .

Т.к.  $f(X_n)$  - это мартингал, то получаем следующее рекуррентное соотношение:

$$f(n) = p_{n \rightarrow n-1}f(n-1) + p_{n \rightarrow n+1}f(n+1) \quad (48.2)$$

Из этого соотношения и из начальных условий получаем,  $f(k) = \sum_{i=1}^k \frac{1}{i^2}$ .

Чему равна вероятность когда-нибудь вернуться в 0? Давайте найдем вероятность достигнуть точки 0 раньше, чем точки  $n$ ,  $\mathbb{P}(\tau_0 < \tau_n)$ . Здесь  $\tau_i$  - момент времени, когда впервые достигнута точка  $i$ , если точка 0 оказалась достигнута раньше точки  $i$  и процесс «залип» в нуле, то считаем, что  $\tau_i = \infty$ .

Для нахождения этой вероятности применим теорему Дуба. Получаем, что для любого  $i$ :

$$f(1) = M_1 = \mathbb{E}(M_1) = \mathbb{E}(M_{\tau_i}) = \mathbb{P}(\tau_i < \tau_0)f(n) + \mathbb{P}(\tau_i > \tau_0)f(0) \quad (48.3)$$

Проверка, что теорема Дуба работает...

Из этого уравнения находим, что  $\mathbb{P}(\tau_n < \tau_0) = \frac{f(1)-f(0)}{f(n)-f(0)}$

В связи с нашим выбором  $f(0) = 0$  и  $f(1) = 1$  получаем:  $\mathbb{P}(\tau_n < \tau_0) = \frac{1}{\sum_{i=1}^n \frac{1}{i^2}}$

Если большой взрыв происходит за конечное время, то какие-то точки на прямой остаются непосещенными. Значит событие  $B = \{\text{Большой Взрыв когда-нибудь произойдет}\}$  можно представить в виде  $B = \bigcap_{n=1}^{\infty} \{\tau_n < \tau_0\}$ . Поскольку эта последовательность событий монотонна, то в силу непрерывности вероятности (...) получаем:  $\mathbb{P}(B) = \lim_{n \rightarrow \infty} \mathbb{P}(\tau_n < \tau_0) = \frac{1}{\sum_{i=1}^{\infty} \frac{1}{i^2}}$ .

Как известно знатокам,  $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$ , а кто не в курсе - читает вики [http://en.wikipedia.org/wiki/Basel\\_problem](http://en.wikipedia.org/wiki/Basel_problem). Следовательно,  $\mathbb{P}(B) = \frac{6}{\pi^2}$ .

**Пример 48.4.**



## 11.5 Мартингалы сходятся

## 11.6 Еще задачи

# 12 Многомерное нормальное распределение

### Векторные случайные величины

Векторная случайная величина - это просто вектор-столбец из случайных величин:

$$\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{pmatrix} \quad (49.1)$$

В принципе можно говорить о векторе-строке или о случайной матрице, но нам понадобятся только вектор-столбцы. Две самых распространенных характеристики одномерной случайной величины  $X$  - это среднее  $\mathbb{E}(X)$  и дисперсия  $Var(X)$ . У векторных случайных величин также есть среднее:

$$\mathbb{E}(\vec{X}) = \begin{pmatrix} \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \\ \dots \\ \mathbb{E}(X_n) \end{pmatrix} \quad (49.2)$$

И ковариационная матрица:

$$Var(\vec{X}) = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Var(X_2) & \dots & Cov(X_2, X_n) \\ \dots & \dots & \dots & \dots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & Var(X_n) \end{pmatrix} \quad (49.3)$$

В ковариационной матрице в  $i$ -ой строке в  $j$ -ом столбце находится  $Cov(X_i, X_j)$ . На диагонали  $i$ -ый элемент, это  $Cov(X_i, X_i) = Var(X_i)$ . Из-за этого сразу следует, что ковариационная матрица симметрична.

Для простоты мы будем обозначать транспонирование с помощью штриха, вот так:  $A'$ . Путаницы с производной не возникнет.

С помощью транспонирования легко сказать, что ковариационная матрица симметрична:  $Var(X)' = Var(X)$ .

Конечно же, дисперсию можно записать через математическое ожидание:  $Var(X) = E[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

Для ковариационной матрицы это выглядит так:

$$Var(\vec{X}) = E[(X - \mathbb{E}(X)) \cdot (X - \mathbb{E}(X))'] = \mathbb{E}(XX') - \mathbb{E}(X) \cdot \mathbb{E}(X)' \quad (49.4)$$

Упр. Убедитесь в этом.

Мы быстренько пройдемся по свойствам этих вещей:

Если  $A$  - неслучайная матрица и  $b$  - неслучайный вектор подходящих размеров, то:

1.  $\mathbb{E}(A\vec{X} + b) = A\mathbb{E}(\vec{X}) + b$
2.  $\mathbb{E}(\vec{X}A + b) = \mathbb{E}(\vec{X})A + b$
3.  $Var(AX + b) = AVar(X)A'$ .
4.  $Var(X)$  - неотрицательно определена, имеет  $n$  неотрицательных собственных чисел (с учетом кратности), где  $n$  размерность вектора  $X$ .

Проверка этих свойств - еще одно упражнение по линейной алгебре.

## Краткая схема разных определений

Есть три с половиной подхода определить многомерное нормальное распределение.

Вектор  $\vec{X}$  имеет многомерное нормальное распределение, если выполнено одно из трех равносильных условий:

1. Вектор  $\vec{X}$  представим в виде  $\vec{X} = A\vec{Z} + \mu$ , где  $\vec{Z}$  — это вектор независимых стандартных нормальных случайных величин
2. Любая линейная комбинация  $c_1X_1 + c_2X_2 + \dots + c_nX_n$  имеет одномерное нормальное распределение
3. Характеристическая функция вектора  $\vec{X}$  имеет вид:

$$\varphi(\vec{u}) = \exp\left(i \cdot \mu' \cdot u - \frac{1}{2} u' V u\right) \quad (50.1)$$

В случае  $\det(V) \neq 0$  первые три формулировки становятся эквивалентны четвертой:

Вектор  $\vec{X}$  имеет невырожденное многомерное нормальное распределение, если

1. Функция плотности вектора  $\vec{X}$  имеет вид:

$$p(\vec{x}) = (2\pi)^{-n/2} (\det(V))^{-1/2} \cdot \exp\left(-\frac{1}{2} (\vec{x} - \mu) V^{-1} (\vec{x} - \mu)'\right) \quad (50.2)$$

Мы выбираем первое свойство как определение, а остальные будут доказывать как теоремы.

## Определение и функция плотности

Очень часто бывает удобно говорить о константе как о нормально распределенной случайной величине с нулевой дисперсией. Это просто соглашение. Никакой функции плотности у константы конечно же нет! Просто когда мы говорим  $N(\mu, \sigma^2)$ ,  $\sigma > 0$ , мы имеем ввиду «честное» нормальное распределение с функцией плотности

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (50.3)$$

, а когда говорим  $N(\mu, 0)$  имеем ввиду просто константу  $\mu$ .

Т.е. говоря «нормальное распределение» мы теперь можем иметь ввиду и константу. Если же нам надо подчеркнуть, что речь идет о «честном» нормальном распределении с функцией плотности, то мы будем говорить «невырожденное нормальное распределение».

Теперь мы можем дать несколько необычное определение одномерной нормальной величины... Пусть  $a$  и  $b$  — это константы.

**Определение 50.4.** Если случайная величина  $X$  представима в виде  $X = aZ + b$ , где  $Z \sim N(0, 1)$ , то мы говорим, что  $X$  имеет **нормальное распределение**  $N(b, a^2)$ . Если, кроме того,  $a \neq 0$ , то мы говорим, что  $X$  имеет **невырожденное нормальное распределение**.

Теперь мы готовы сказать, что такое многомерное нормальное распределение. Пусть матрица  $A$  — неслучайная, вектор  $b$  — неслучайный. Размерности подходящие.

**Определение 50.5.** Если  $X = AZ + b$ , где вектор  $Z$  состоит из независимых  $Z_i \sim N(0; 1)$ , то мы говорим, что вектор  $X$  имеет **нормальное распределение**  $N(b, AA')$ . Если, кроме того,  $A$  обратимая матрица,  $\det(A) \neq 0$ , то мы говорим, что  $X$  имеет **невырожденное нормальное распределение**.

В этой главе и в книге в целом мы **постараемся** уточнять, но обычно из контекста понятно, включает ли автор в понятие нормального распределения вырожденные случаи. Как правило, да, включает.

**Пример 50.6.** Вектор  $X$  имеет вырожденное нормальное распределение, у него нет двумерной функции плотности:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_1 \end{pmatrix} \quad (50.7)$$

Невырожденный нормальный вектор имеет функцию плотности...

**Теорема 51.1.** Вектор  $X$  имеет невырожденное нормальное распределение  $N(b, AA')$ , если и только если его функция плотности имеет вид

$$p(\vec{x}) = (2\pi)^{-n/2} (\det(AA'))^{-1/2} \cdot \exp\left(-\frac{1}{2}(\vec{x} - b)(AA')^{-1}(\vec{x} - b)'\right) \quad (51.2)$$

Если не обращать внимания на константу, которая нужна чтобы интеграл под функцией плотности равнялся единице, то:

$$p(\vec{x}) \sim \exp\left(-\frac{1}{2}(\vec{x} - b)(AA')^{-1}(\vec{x} - b)'\right) \quad (51.3)$$

Доказательство. □

## Напоминалка про характеристические функции

Опытный охотник увидев уши, торчащие из кустов, скажет: «Ага, это заяц!» Для этих же целей нужна и характеристическая функция. Она позволяет опознать закон распределения случайной величины.

Узнав например, что характеристическая функция случайной величины  $W$  равна  $\varphi(u) = \frac{\exp(iu)-1}{iu}$ , опытный охотник скажет: «Ага, это равномерная на  $[0; 1]$ !». Увидев характеристическую функцию  $\varphi(u) = (p \cos(u) + ip \sin(u) + 1 - p)^n$ , опытный охотник узнает биномиальное распределение  $Bin(n, p)$ .

Настала пора для формального определения:

**Определение 51.4.** Характеристической функцией случайной величины  $W$  называется функция  $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ :

$$\varphi_W(u) := \mathbb{E}(\cos(uW)) + i\mathbb{E}(\sin(uW)) \quad (51.5)$$

Или более кратко:

$$\varphi_W(u) := \mathbb{E}(\exp(iuW)) \quad (51.6)$$

Косинус и синус - ограниченные функции, поэтому для любой случайной величины  $W$  существуют  $\mathbb{E}(\sin(uW))$  и  $\mathbb{E}(\cos(uW))$ . А значит и характеристическая функция всегда существует. Тот факт, что по характеристической функции можно узнать закон распределения строго формулируется в виде теоремы:

**Теорема 51.7.** Пусть  $X_1$  и  $X_2$  — две случайной величины. Функции распределения  $F_1(t)$  и  $F_2(t)$  совпадают если и только если совпадают характеристические функции  $\varphi_1(u)$  и  $\varphi_2(u)$

Для описания многомерных нормальных величин характеристическая функция оказывается более удобна, чем функция плотности. Дело в том, что у вырожденных нормальных векторов функции плотности нет, а характеристическая функция существует всегда.

## Еще два эквивалентных определения

**Теорема 51.8.** Вектор  $X$  имеет нормальное распределение  $N(b, AA')$ , возможно вырожденное, если и только если его характеристическая функция имеет вид

$$\varphi(\vec{u}) = \exp\left(i \cdot \mu' \cdot u - \frac{1}{2} u' V u\right) \quad (51.9)$$

Важная особенность многомерного нормального распределения.

**Теорема 51.10.** Вектор  $X = (X_1, \dots, X_n)$  имеет многомерное нормальное распределение (возможно вырожденное) если и только если любая линейная комбинация  $Y = c_1X_1 + c_2X_2 + \dots + c_nX_n$  имеет одномерное нормальное распределение (возможно вырожденное).

*Доказательство.* Если  $X$  - многомерное нормальное, то  $X = AZ + b$ . Следовательно,  $Y = c_1X_1 + c_2X_2 + \dots + c_nX_n = \vec{c}X = \vec{c}AZ + \vec{c}b$ . Значит  $Y$  по определению ?? имеет нормальное распределение.

Если любая линейная комбинация нормальна...

□

## Стандартизация

В одномерном случае, если  $Var(X) \neq 0$ , то случайную величину  $X$  можно «стандартизировать», т.е. превратить в случайную величину с нулевым средним и единичной дисперсией:

$$Z := \frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}} \quad (52.1)$$

У случайной величины  $Z$ :  $\mathbb{E}(Z) = 0$ ,  $Var(Z) = 1$ .

В многомерном случае, если матрица  $Var(X)$  обратима, то случайный вектор  $\vec{X}$  можно «стандартизовать», т.е. превратить в вектор некоррелированных случайных величин с нулевым средним и единичной дисперсией.

$$\vec{Z} := (Var(X))^{-1/2}(\vec{X} - \mathbb{E}(\vec{X})) \quad (52.2)$$

Для этой операции нужно понимать, что такое  $A^{\frac{1}{2}}$ . Для положительно определенной матрицы эта операция корректна. Если  $A$  положительно определена, то у нее есть разложение  $A = PDP^{-1}$  причем  $D$  - диагональная матрица, где на главной диагонали стоят положительные собственные значения  $A$ . И, стало быть,  $A^{\frac{1}{2}} = PD^{\frac{1}{2}}P^{-1}$ .

Если нужно решить какую-то задачу, связанную с многомерным нормальным распределением, то стандартизация может здорово облегчить вычисления. Даже в вырожденном случае имеет смысл перейти к рассмотрению независимых нормальных  $N(0; 1)$  случайных величин.

**Пример 52.3.** Пусть рост любой женщины имеет распределение  $N(165, 25)$ . А корреляция между ростом матери и ростом дочери равна  $\rho$ . Какова вероятность того, что дочь высокая (выше среднего роста), если мама - высокая?

Начинаем решать:

Обозначим  $X_1$  - рост матери,  $X_2$  - рост дочери.  $\mathbb{P}(X_2 > 165 | X_1 > 165) = \frac{\mathbb{P}(X_1 > 165 \cap X_2 > 165)}{\mathbb{P}(X_1 > 165)} = 2\mathbb{P}(X_1 > 165 \cap X_2 > 165)$ .

Если решать без стандартизации, «в лоб». Вылезает интеграл, который берется усердным трудом...

Если перейти к стандартным  $Z_1$  и  $Z_2$ ...

## Многомерное — это больше чем несколько одномерных

Из теоремы ?? следует в частности, что:

**Теорема 52.4.** Если вектор  $\vec{X}$  имеет многомерное нормальное распределение, то и любая его компонента  $X_i$  имеет нормальное распределение. Кроме того, если распределение  $\vec{X}$  — невырожденное, то и распределение каждой компоненты  $X_i$  невырожденное.

Обратное утверждение неверно. Мы приведем пример, в котором  $X_1$  и  $X_2$  имеют нормальное распределение по отдельности, но не имеют совместного нормального распределения.

**Пример 52.5.** Пусть  $X_1 \sim N(0; 1)$ , а  $K$  — это случайная величина равновероятно принимающая значения 1 и  $-1$ , причем  $K$  и  $X_1$  независимы. Определим

$$X_2 = K \cdot |X_1| \quad (52.6)$$

Проверим, что  $X_2$  имеет нормальное распределение. Для положительных  $x$ :

$$\begin{aligned} \mathbb{P}(X_2 \leq x) &= \mathbb{P}(K = -1) + \mathbb{P}(K = 1) \cdot \mathbb{P}(|X_1| \leq x) = \\ &= 0.5 + 0.5\mathbb{P}(|X_1| \leq x) = 0.5 + \mathbb{P}(X_1 \in [0; x]) = \mathbb{P}(X_1 \leq x) = F(x) \end{aligned} \quad (53.1)$$

Для отрицательных  $x$  проверьте сами!

Можно заметить, что  $X_1 + X_2$  не является нормально распределенной случайной величиной. Более того, распределение  $X_1 + X_2$  не является непрерывным. Действительно,

$$\mathbb{P}(X_1 + X_2 = 0) = \mathbb{P}(X_1 + K|X_1| = 0) = \mathbb{P}(K \neq \text{sign}(X_1)) = 0.5 \quad (53.2)$$

Если бы вектор  $(X_1, X_2)$  имел совместное нормальное распределение, то сумма  $X_1 + X_2$  была бы нормально распределенной. Кстати, найдем ковариацию между  $X_1$  и  $X_2$ :

$$\begin{aligned} \text{Cov}(X_1, X_2) &= \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = \mathbb{E}(X_1 X_2) = \mathbb{E}(X_1 \cdot K \cdot |X_1|) = \\ &= \mathbb{E}(K)\mathbb{E}(X_1 \cdot |X_1|) = 0 \end{aligned} \quad (53.3)$$

Значит наши  $X_1$  и  $X_2$  были некоррелированы, нормальны по отдельности. Вместе с тем они не были нормальны в совокупности. И конечно, они зависимы, т.к.  $|X_1| = |X_2|$ .

## Некоррелированность и независимость

Для многомерного нормального распределения некоррелированность равносильна независимости.

Доказательство для двумерного...

## Условное распределение

Предположим, что вектор  $(X, Y)$  имеет совместное нормальное распределение. Часто возникает задача прогнозирования случайной величины  $Y$ , если значение случайной величины  $X$  известно.

Для удобства вычислений всегда используем стандартизацию! Вместо исходных  $X$  и  $Y$  рассмотрим:

$$Z_1 = \frac{X - \mathbb{E}(X)}{\sigma_X}, \quad Z_2 = \frac{Y - \mathbb{E}(Y)}{\sigma_Y} \quad (53.4)$$

Естественно,  $Z_1 \sim N(0; 1)$  и  $Z_2 \sim N(0; 1)$  и  $\text{Corr}(Z_1, Z_2) = \text{Corr}(X, Y)$ .

Найдем условную функцию плотности  $Z_2$  при известном  $Z_1$ :

$$p(z_2|z_1) = \frac{p(z_1, z_2)}{p(z_1)} \quad (53.5)$$

Поскольку нас интересует только зависимость от  $z_2$ , то на  $p(z_1)$  можно не обращать внимания.

Значком  $\sim$  мы будем обозначать равенство с точностью до константы не зависящей от  $z_2$ :

$$\begin{aligned} p(z_2|z_1) &\sim p(z_1, z_2) \sim \exp \left( -\frac{1}{2}(z_1, z_2) \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right) = \\ &= \exp \left( -\frac{1}{2(1-\rho^2)}(z_1, z_2) \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right) = \\ &= \exp \left( -\frac{1}{2(1-\rho^2)}(z_2^2 - 2\rho z_1 z_2 + z_1^2) \right) \sim \exp \left( -\frac{1}{2(1-\rho^2)}(z_2 - \rho z_1)^2 \right) \end{aligned} \quad (53.6)$$

Сравним полученный результат с функцией плотности одномерного нормального распределения:

$$p(x) \sim \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right) \quad (54.1)$$

Замечаем, что  $\sigma^2 = 1 - \rho^2$  и  $\mu = \rho z_2$ . И оформляем результат вычислений в виде теоремы:

**Теорема 54.2.** Если  $Z_1$  и  $Z_2$  имеют совместное нормальное распределение,  $\mathbb{E}(Z_i) = 0$ ,  $Var(Z_i) = 1$  и  $Corr(Z_1, Z_2) = \rho$ , то условное распределение  $Z_1$  при известном  $Z_2$  является нормальным и:

$$\mathbb{E}(Z_2|Z_1 = z) = \rho \cdot z \quad (54.3)$$

$$Var(Z_2|Z_1 = z) = 1 - \rho^2 \quad (54.4)$$

Формулы дают нам смысл:

1. Корреляция стандартизированных нормальных величин показывает на сколько в среднем растёт одна случайная величина при росте другой на единицу.
2. Условная дисперсия  $Z_2$  не зависит от конкретного значения  $Z_1$ . Это несколько неожиданно, но интуитивного объяснения я не знаю.

Для возврата к исходным  $X$  и  $Y$  замечаем, что условие  $X = x$  равносильно тому, что  $Z_1 = \frac{x - \mu_x}{\sigma_x}$  и, кроме того,  $Y = \mu_y + \sigma_y Z_2$ .

Сделав обратную замену, получаем:

**Теорема 54.5.** Если  $X$  и  $Y$  имеют совместное нормальное распределение,  $X \sim N(\mu_x, \sigma_x^2)$ ,  $Y \sim N(\mu_y, \sigma_y^2)$  и  $Corr(X, Y) = \rho$ , то условное распределение  $Y$  при известном  $X$  является нормальным и:

$$\mathbb{E}(Y|X = x) = \mu_y + \rho \sigma_y \frac{x - \mu_x}{\sigma_x} \quad (54.6)$$

$$Var(Y|X = x) = (1 - \rho^2) \sigma_y^2 \quad (54.7)$$

Получаем трактовку корреляции: Если  $X$  и  $Y$  имеют совместное нормальное распределение, то корреляция показывает на сколько своих стандартных отклонений в среднем растёт  $Y$  при росте  $X$  на одно свое стандартное отклонение.

## Геометрический смысл ковариационной матрицы

**Определение 54.8.** Матрица  $R$  называется **матрицей поворота** если одновременно выполнены два условия:  $R'R = I$  и  $\det(R) = 1$

Почему определение поворота именно такое?

Условие  $RR' = I$  означает две вещи:

1. Вектор  $R\vec{x}$  имеет такую же длину, как и вектор  $\vec{x}$ :

$$|\vec{x}|^2 = \vec{x}'\vec{x} = \vec{x}'R'R\vec{x} = (R\vec{x})'(R\vec{x}) = |R\vec{x}|^2 \quad (54.9)$$

2. Угол между  $\vec{x}$  и  $\vec{y}$  равен углу между  $R\vec{x}$  и  $R\vec{y}$

$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x}'\vec{y}}{|\vec{x}||\vec{y}|} = \frac{\vec{x}'R'R\vec{y}}{|R\vec{x}||R\vec{y}|} = \frac{(R\vec{x})'(R\vec{y})}{|R\vec{x}||R\vec{y}|} = \cos(R\vec{x}, R\vec{y}) \quad (54.10)$$

Условию  $R'R = I$  подходят матрицы с определителем  $\det(A) = \pm 1$ . Дополнительное условие  $\det(R) = 1$  исключает «отражения».

Упражнение. Мы доказали, что матрицы вида  $R'R = I$  сохраняют углы и длины. Докажите, что никакие другие матрицы не сохраняют одновременно углы и длины.

Решение. Рассмотрим вектор  $e_k = (0, 0, \dots, 0, 1, 0, \dots, 0)'$ .

Из линейной алгебры:

**Теорема 55.1.** Если  $A_{n \times n}$  действительная симметричная положительно полу-определенная матрица, то

1. У  $A$  имеется ровно  $n$  действительных собственных чисел
2.  $A$  представима в виде

$$A = RDR^{-1} = RDR' \quad (55.2)$$

, где  $D$  — диагональная матрица из собственных чисел матрицы  $A$ , а  $R$  — матрица поворота из собственных векторов матрицы  $A$ .

Для начала сформулируем геометрические факты:

**Определение 55.3.** Множество точек называется **эллипсоидом**, если оно задается системой уравнений

$$(\vec{x} - \vec{x}_0)'A(\vec{x} - \vec{x}_0) = 1 \quad (55.4)$$

, где  $A$  — положительно определенная матрица. Точка  $\vec{x}_0$  — центр эллипсоида. В двумерном случае эллипсоид называют **эллипсом**.

Почему определение эллипса именно такое?

Матрицу  $A$  можно представить в виде  $A = R'DR$ , где  $R$  — матрица поворота, а  $D$  — диагональная матрица собственных чисел матрицы  $A$ . Отсюда получаем, что уравнение эллипса можно записать в виде:

$$(R(\vec{x} - \vec{x}_0))'D(R(\vec{x} - \vec{x}_0)) = 1 \quad (55.5)$$

Если ввести обозначения  $\vec{z} = R(\vec{x} - \vec{x}_0)$ , то уравнение примет вид:

$$\sum_{i=1}^n d_{ii} z_i^2 = 1 \quad (55.6)$$

Именно в силу этого представления:

**Определение 55.7.** Для эллипса  $(\vec{x} - \vec{x}_0)'A(\vec{x} - \vec{x}_0) = 1$  собственные векторы матрицы  $A$  называют **направлениями полуосей**. Если  $\lambda_i$  — это собственное число матрицы  $A$ , то величины  $1/\sqrt{\lambda_i}$  называют **длинами полуосей**.

Теперь мы готовы нарисовать линии уровня многомерного нормального распределения!

**Теорема 55.8.** Для невырожденного нормального распределения линии уровня функции плотности являются эллипсоидами. Направления главных осей задаются собственными векторами ковариационной матрицы. Соотношение длин полуосей равно соотношению корней из собственных чисел ковариационной матрицы.

*Доказательство.* Для доказательства нам потребуется технический факт из линейной алгебры:

**Теорема 55.9.** Если  $\vec{a}$  собственный вектор для матрицы  $V$  с собственным числом  $\lambda$ , то  $\vec{a}$  собственный вектор для матрицы  $V^{-1}$  с собственным числом  $1/\lambda$ .

*Доказательство.* Если  $\vec{a}$  собственный вектор матрицы  $V$ , то с одной стороны:

$$V^{-1} \cdot (V \cdot \vec{a}) = V^{-1}(\lambda \vec{a}) = \lambda V^{-1} \vec{a} \quad (55.10)$$

С другой стороны:

$$(V^{-1} \cdot V) \cdot \vec{a} = \vec{a} \quad (55.11)$$

Т.е.  $\lambda V^{-1} \vec{a} = \vec{a}$ . Или:

$$V^{-1} \vec{a} = \frac{1}{\lambda} \vec{a} \quad (56.1)$$

□

Итак, пусть  $X \sim N(\vec{\mu}; V)$ .

Тогда условие

$$p(\vec{x}) = \text{const} \quad (56.2)$$

после преобразований равносильно тому, что:

$$(\vec{x} - \vec{\mu}) \cdot V^{-1} \cdot (\vec{x} - \vec{\mu})' = \text{const} \quad (56.3)$$

Т.е. мы получили наше определение эллипсоида с  $A = \frac{1}{\text{const}} V^{-1}$ .

Направления полуосей задаются собственными векторами  $A$ , значит они совпадают с собственными векторами  $V$ .

Длины полуосей обратно пропорциональны корням из собственных чисел  $A$ , значит они прямо пропорциональны корням из собственных чисел  $V$ . □

**Пример 56.4.** Пусть  $X \sim N(0; V)$  и

$$V = \begin{pmatrix} 5 & 1 \\ 1 & 9 \end{pmatrix} \quad (56.5)$$

Нарисуйте линии уровня функции плотности  $p(x_1, x_2)$

Решение.

Тут обязательно картинка. Эллипс. Главные оси. С кодом R и Sage!

В случае независимых нормальных случайных величин с одинаковой дисперсией линиями уровня будут окружности (сферы при более высоких размерностях). Поскольку поворот никак не влияет на линию уровня мы бесплатно получаем следующую теорему:

**Теорема 56.6.** Если матрица  $R$  — это матрица поворота и вектор  $\vec{Z} \sim N(\vec{0}, I)$  то вектор  $R\vec{Z} \sim N(\vec{0}, I)$

## Хи-квадрат распределение

Как известно,

**Определение 56.7.** Если случайная величина  $W$  представима в виде  $W = \sum_{i=1}^k Z_i^2$ , где  $Z_i$  — независимые стандартные нормальные случайные величины, то говорят, что  $W$  имеет **хи-квадрат распределение с  $k$  степенями свободы**.

К сожалению, проверять, что что-то имеет хи-квадрат распределение напрямую очень неудобно.

**Пример 56.8.** Давайте попробуем по определению показать, что если  $X_i \sim N(\mu; \sigma^2)$  и независимы, то

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \sim \chi_{n-1}^2 \quad (56.9)$$

Как всегда, сначала стандартизируем наши  $X_i$ .

...

Гораздо более удобным оказывается следующий способ:



**Теорема 56.10.** Если вектор  $\vec{Z}$  состоит из независимых стандартных нормальных случайных величин и все собственные числа симметричной матрицы  $A$  равны либо нулю, либо единице, то

$$\vec{Z}' A \vec{Z} \sim \chi_r^2 \quad (56.11)$$

где  $r$  — количество собственных чисел матрицы  $A$  равных единице.

*Доказательство.* Матрица  $A$  представима в виде  $A = R' D R$ . Поэтому:

$$\vec{Z}' A \vec{Z} = \vec{Z}' R' D R \vec{Z} = (R \vec{Z})' D (R \vec{Z}) \quad (57.1)$$

Остается заметить, что  $R \vec{Z}$  — это вектор независимых стандартных нормальных случайных величин в силу теоремы ???. Если обозначить этот вектор буквой  $\vec{W} = R \vec{Z}$ , то

$$\vec{Z}' A \vec{Z} = \vec{W}' D \vec{W} = \sum_{i=1}^n d_{ii} W_i^2 \quad (57.2)$$

Матрица  $D$  — это матрица собственных чисел матрицы  $A$ , т.е.  $d_{ii}$  равны либо нулю, либо единице. Получается, что  $\vec{Z}' A \vec{Z}$  — это сумма  $r$  стандартных независимых нормальных случайных величин.  $\square$

**Теорема 57.3.** Собственные числа симметричной матрицы  $A$  равны либо нулю, либо единице, если и только если  $A^2 = A$

*Доказательство.* С одной стороны:

$$A^2 = (R' D R)(R' D R) = R' D (R R') D R = R' D D R = R' D^2 R \quad (57.4)$$

С другой стороны  $A = R' D R$ . Значит  $A^2 = A$  если и только если  $D^2 = D$ . Но  $D$  — диагональная матрица, поэтому условие  $D^2 = D$  равносильно тому, что на диагонали стоят либо нули, либо единицы.  $\square$

**Пример 57.5.** Тот же пример только быстрее... ..

## Границы на хвостовые вероятности

При изучении броуновского движения полезны два неравенства.

**Теорема 57.6.** Если  $Z \sim N(0; 1)$ , то:

$$\frac{1}{\sqrt{2\pi}} \frac{\exp(-x^2/2)}{x + x^{-1}} \leq \mathbb{P}(Z \geq x) \leq \frac{1}{\sqrt{2\pi}} \frac{\exp(-x^2/2)}{x}, \quad x > 0 \quad (57.7)$$

*Доказательство.* Заметим, что  $\exp(-t^2/2)' = -t \exp(-t^2/2)$ . Получаем верхнюю границу:

$$\begin{aligned} \mathbb{P}(Z \geq x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt \leq \\ &\leq \int_x^\infty \left( \frac{t}{x} \right) \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt = \frac{1}{\sqrt{2\pi}} \frac{\exp(-x^2/2)}{x} \end{aligned} \quad (57.8)$$

Заметим, что  $(t^{-1} \exp(-t^2/2))' = (1 + t^{-2}) \exp(-t^2/2)$ . Получаем нижнюю границу:

$$\begin{aligned} \mathbb{P}(Z \geq x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt \geq \int_x^\infty \left( \frac{1+t^{-2}}{1+x^{-2}} \right) \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt = \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{1+x^{-2}} \frac{\exp(-x^2/2)}{x} = \frac{1}{\sqrt{2\pi}} \frac{\exp(-x^2/2)}{x+x^{-1}} \end{aligned} \quad (57.9)$$

□

**Теорема 58.1.** Если  $Z \sim N(0; 1)$ , то:

$$\frac{2x}{\sqrt{2\pi}e} \leq \mathbb{P}(|Z| \leq x) \leq \frac{2x}{\sqrt{2\pi}}, \quad 0 < x \leq 1 \quad (58.2)$$

*Доказательство.* Если  $|t| \leq 1$ , то

$$\frac{1}{\sqrt{2\pi}e} \leq \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) \leq \frac{1}{\sqrt{2\pi}} \quad (58.3)$$

Интегрируя это неравенство от  $-x$  до  $x$  получаем требуемое. □

Еще кусок из блога... (ссылка?) ...

## Откуда взялось $\pi$ в формуле?

Есть много способов объяснить, откуда берется  $\pi$  в формуле функции плотности. Вот то, которое нравится мне.

Рассмотрим пару независимых стандартных нормальных случайных величин  $X$  и  $Y$ . Их функция плотности имеет вид:

$$p(x, y) = c \cdot \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \quad (58.4)$$

Рассмотрим более подробно функцию

$$f(x, y) = \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \quad (58.5)$$

Значение  $f(x, y)$  зависит только от расстояния от точки  $(x, y)$  до начала координат. Значит объем под «шляпой» является фигурой вращения.

Картинка (слева и справа повернутая):

...

Уже понятно, что  $\pi$  — в деле. Остается вспомнить, что объем фигуры вращения определяется по формуле  $Vol = \int_a^b \pi r^2(t) dt$ .

В нашем случае:  $a = 0$ ,  $b = f(0, 0) = 1$ , а  $r(t)$  находится из условия:

$$\exp\left(-\frac{1}{2}r^2(t)\right) = t \quad (58.6)$$

Находим  $r^2(t)$  и получаем:

$$r^2(t) = -2 \ln(t) \quad (58.7)$$

Находим объем фигуры вращения:

$$Vol = \int_0^1 \pi r^2(t) dt = \int_0^1 \pi(-2 \ln(t)) dt = -2\pi \int_0^1 \ln(t) dt = 2\pi \quad (58.8)$$

Интеграл от  $\ln(t)$  можно взять либо по частям, либо заметив, что:

$$\int_0^1 \ln(t) dt = - \int_0^\infty \exp(-t) dt \quad (59.1)$$

Картинка:

Настоящие знатоки берут интеграл  $\int_0^\infty \exp(-t) dt$  по методу Мамикона Мнацаканяна [apostol:visual\_calculus] в уме глядя на картинку:

Пояснение к картинке:

Производная функции  $\exp(-x)$  равна ей самой умноженной на минус один. Поэтому тень от касательной всегда имеет длину один. Интересующая нас площадь равна площади треугольника плюс оставшаяся площадь. Касательные заметающие оставшуюся площадь можно перенести так, чтобы они замели треугольник. Значит интеграл равен удвоенной площади треугольника.

Мы доказали, что объем под функцией  $f(x, y)$  равен  $2\pi$ . Следовательно, функция плотности  $p(x, y)$  должна иметь вид:

$$p(x, y) = \frac{1}{2\pi} \cdot \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \quad (59.2)$$

Для независимых величин  $p(x, y) = p(x)p(y)$ , следовательно:

$$p(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}x^2\right) \quad (59.3)$$

## ЦПТ — доказательство через характеристические функции

**Теорема 59.4.** Пусть  $X_n$  — последовательность случайных величин с характеристическими функциями  $\varphi_n(u)$ . Пусть кроме того,  $X$  — случайная величина с характеристической функцией  $\varphi(u)$ . Последовательность  $X_n$  сходится по распределению к случайной величине  $X$  если и только если последовательность функций  $\varphi_n(u)$  сходится поточечно к функции  $\varphi(u)$ .

## 13 Процессы в непрерывном времени

Начиная с этой главы читателю предстоит бороться с дополнительной трудностью - с меняющимися обозначениями. Почему? Потому, что писать  $X_t$  (значение процесса в момент времени  $t$ ) проще, чем писать  $X(t)$ , но иногда приходится иметь дело с последовательностями процессов, и тогда удобнее будет обозначение  $X_n(t)$ .

### 13.1 Версии, непрерывность справа

### 13.2 Гауссовские процессы

Гауссовский случайный процесс.

**Определение 59.5.** Случайный процесс  $X_t$  называется гауссовским, если любой вектор  $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$  имеет многомерное нормальное распределение (возможно вырожденное).

Из определения ясно, что разные гауссовские процессы могут различаться математическим ожиданием, функцией  $\mu(t) = \mathbb{E}(X_t)$ , и ковариационной функцией,  $\gamma(s, t) = Cov(X_t, X_s)$ . Важнее, конечно, ковариационная функция, т.к. простым преобразованием  $Y_t = X_t - \mathbb{E}(X_t)$  всегда можно добиться того, что математическое ожидание равно нулю.

## Задачи

**Задача 13.1.** Пусть величины  $X$  и  $Y$  имеют совместное нормальное распределение с нулевым средним, единичной дисперсией и корреляцией  $\rho$ .

а) Представьте  $Y$  в виде  $Y = aX + bZ$ , так, чтобы  $Z \sim N(0; 1)$ , а  $Z$  и  $X$  не были бы коррелированы.

б) Что представляет собой множество  $X > 0 \cap Y > 0$  в осях  $(X, Z)$ ?

в) Чему равна вероятность  $\mathbb{P}(XY > 0)$ ?

г) Постройте график  $\mathbb{P}(XY > 0)$  как функции от  $\rho$

**Задача 13.2.** Пусть величины  $X, Y, Z$  - имеют совместное нормальное распределение, с математическим ожиданием 0 и некоей ковариационной матрицей  $B$ .

Как зависит от  $B$  вероятность  $\mathbb{P}(XYZ > 0)$ ?

## 13.3 Случайный процесс - это одна случайная величина

Буквально только что мы сказали, что случайный процесс - это набор случайных величин. Оказывается, что можно и иногда полезно рассматривать целый случайный процесс как **одну** случайную величину. Вот как это можно сделать:

Для примера возьмем случайный процесс для  $t \in [0; 1]$ . Случайным образом выберем траекторию, затем случайно (равномерно) выберем момент времени  $t$ . И вуаля, вместо целого процесса  $X_t$  мы получили одну случайную величину  $X$ ! При этом мы изменили пространство  $\Omega$  на котором был задан случайный процесс  $X_t$ . А именно, новая случайная величина  $X$  задана на множестве  $\Omega' = \Omega \times [0; 1]$ . А вероятности считаются с помощью новой меры  $P' = P \times \lambda$ , где  $\lambda$  - классическая мера Лебега, задающая равномерное распределение.

При такой интерпретации корректно говорить, например, о  $P'(X < 0.2)$  и о  $\int_{\Omega'} X dP' = E'(X)$ . Давайте попробуем на паре конкретных задач:

- Какой случайный процесс соответствует случайной величине  $1_{\Omega'}$ ?

Очень просто! Для любого  $w' = (w, t)$  значение  $1_{\Omega'}(w') = 1$ , значит наш случайный процесс  $X_t$  всегда равен 1 (на любой траектории и в любой момент времени).

- Пусть  $Y$  равномерная случайная величина на  $[0; 1]$ , и  $X_t = Y - t$ , процесс определенный для  $t \in [0; 1]$ . Нарисуйте несколько разных траекторий процесса  $X_t$ . Рассмотрев процесс как одну случайную величину, найдите  $P'(X < 0, 2)$  и  $E'(X)$ .

Любая траектория начинается в точке  $y \in [0; 1]$  и с единичной скоростью убывает на отрезке  $[0; 1]$ . Какова при заданном  $y$  вероятность того, что выбрав случайную точку на отрезке  $[0; 1]$  мы получим ординату меньше 0,2? Если  $y \leq 0.2$ , то вероятность равна 1; если  $y > 0.2$ , то вероятностью равна  $(y - 0, 2)$ . Находим искомую вероятность:

$$P'(X < 0, 2) = \mathbb{P}(Y < 0.2) + \mathbb{P}(Y > 0.2)\mathbb{E}(Y - 0.2 | Y > 0.2) = 0.2 + 0.8 \cdot 0.4 = 0.52 \quad (60.1)$$

Переходим к  $E'(X)$ . При фиксированном  $y \in [0; 1]$  траектория с постоянной скоростью опускается с отметки  $y$  до отметки  $y - 1$ , поэтому при фиксированном  $y$  средняя ордината равна  $\frac{y+y-1}{2} = y - 0.5$ . Усредняя по  $y$  получаем, что  $E'(X) = \mathbb{E}(Y - 0.5) = 0$ . Попутно отметим, что  $E'(X)$  можно интерпретировать как среднюю площадь под траекторией процесса.

Если процесс задан на  $[0; T]$  (или на  $[0; \infty)$ ), то новая мера  $P' := P \times \lambda$  не будет вероятностью (т.к. длина  $\lambda([0; T]) = T$ ). Но принципиально это ничего не меняет!

По прежнему можно будет считать выражения типа  $P'(X < 0.2)$ . только это будет не «вероятность», а «средняя длина» множества  $X < 0.2$ . Для каждой отдельно взятой траектории (для каждого  $w$ ) можно посчитать длину множества  $\{t | X_t(w) < 0.2\}$ , а поскольку на  $\Omega$  задана вероятность  $P$ , то интеграл означает среднее значение этой длины:

$$P'(X < 0.2) = \int_{\Omega'} 1_{X < 0.2} dP' = \int_{\Omega} \int_{[0; T]} 1_{X < 0.2} d\lambda dP = \mathbb{E}(\lambda(X < 0.2)), \quad (60.2)$$

при раскрытии интеграла  $\int_{\Omega'} (\cdot) dP'$  мы воспользовались теоремой Фубини.

Обозначение  $E'(X)$  не совсем корректно (оно использовалось бы, если бы  $P'$  была вероятностью). Осталось обозначение с интегралом  $\int_{\Omega'} X dP'$ . Сохранилась и интерпретация:  $\int_{\Omega'} X dP'$  - средняя площадь под траекторией процесса:

$$\int_{\Omega'} X dP' = \int_{\Omega} \int_{[0;T]} X d\lambda dP = \mathbb{E} \left( \int_{[0;T]} X d\lambda \right) \quad (61.1)$$

Для усвоения еще пара примеров: упр.

Единственное, что мы не проверили - это измеримость. За кадром должна быть некая  $\sigma$ -алгебра  $\mathcal{F}'$ , относительно которой функция  $X$  должна быть измеримой. Какая же?

Процесс  $X_t$  определен на пространстве  $(\Omega, \mathcal{F}, P)$  и все случайные величины  $X_t$  являются  $\mathcal{F}$  измеримыми. Может еще присутствовать фильтрация  $\mathcal{F}_t$  и процесс  $X_t$  может быть адаптирован к ней, но сейчас это не обязательно. Новое множество исходов,  $\Omega' := \Omega \times [0; T]$  - это обычное декартово произведение, т.е. множество всех пар вида  $(w, x)$ , где  $w \in \Omega$ ,  $x \in [0; T]$ . Естественная  $\sigma$ -алгебра событий на нем - это  $\mathcal{F} \times \lambda$ , минимальная  $\sigma$ -алгебра порожденная декартовым произведением  $\sigma$ -алгебр  $\mathcal{F}$  и  $\mathcal{B}$ ,  $\mathcal{F} \times \mathcal{B} := \sigma((F, B) | F \in \mathcal{F}, B \in \mathcal{B})$ . Эта же  $\sigma$ -алгебра  $\mathcal{F} \times [0; T]$  является областью определения меры  $P' = P \times \lambda$ .

**Определение 61.2.** Процесс  $X_t$  на пространстве  $(\Omega, \mathcal{F}, P)$  для  $t \in [0; T]$  (также и для  $t \in [0; \infty)$ ) называется **измеримым**, если функция  $X$  измерима относительно  $\sigma$ -алгебры  $\mathcal{F} \times \mathcal{B}$ .

Зачем нам важна измеримость функции  $X$  относительно  $\mathcal{F} \times \mathcal{B}$ ? Именно это условие является позволяющим интегрировать и получать «хороший» результат:

**Теорема 61.3.** Если функция  $X$  измерима относительно  $\mathcal{F} \times \lambda$ , то для любого  $t \in [0; T]$  интеграл

$$Y_t = \int_0^t X_t d\lambda$$

корректно определен и является  $\mathcal{F}$ -измеримой случайной величиной.

*Доказательство.* □

Сформулируйте и докажите теорему (...) для случая  $t \in [0; \infty)$  (упр.)

Иногда оказывается желательным, чтобы  $Y_t = \int_0^t X_t d\lambda$  был не только  $\mathcal{F}$ -измерим, но и  $\mathcal{F}_t$ -измерим. Это требует более сильных условий на  $X_t$ :

**Определение 61.4.** Процесс  $X_t$  на пространстве  $(\Omega, \{\mathcal{F}_t\}, \mathcal{F}, P)$  для  $t \in [0; T]$  (также и для  $t \in [0; \infty)$ ) называется **прогрессивно измеримым**, если для любого  $t \in [0; T]$  (или, соответственно,  $t \in [0; \infty)$ ) случайная величина  $X$  измерима относительно  $\sigma$ -алгебры  $\mathcal{F}_t \times \mathcal{B}$ .

**Теорема 61.5.** Если процесс  $X_t$  прогрессивно измерим на  $(\Omega, \{\mathcal{F}_t\}, \mathcal{F}, P)$ , то интеграл

$$Y_t = \int_0^t X_t d\lambda$$

корректно определен и является  $\mathcal{F}_t$ -измеримой случайной величиной.

*Доказательство.* Зафиксируем момент времени  $t$ , в качестве  $\mathcal{F}$  берем  $\mathcal{F}_t$  и получаем желаемое в силу теоремы (...) □

Оказывается, что связь между адаптированностью  $X_t$  к фильтрации  $\mathcal{F}_t$  и измеримостью  $X$  относительно  $\mathcal{F} \times \lambda$  нетривиальная. Этому и посвящены два следующих примера:

**Пример 61.6.** Процесс  $X_t$  адаптирован к  $\mathcal{F}_t$ , но случайная величина  $X$  не измерима относительно  $\mathcal{F} \times \mathcal{B}$ .

**Пример 61.7.** Процесс  $X_t$  не адаптирован к  $\mathcal{F}_t$ , но случайная величина  $X$  измерима относительно  $\mathcal{F} \times \mathcal{B}$ .

Общая картина выглядит так:

Если процесс  $X_t$  «хороший», то  $X$  измерим даже если он не адаптирован:

**Теорема 61.8.** Если каждая траектория процесса  $X_t$  является непрерывной слева (или непрерывной справа), то процесс  $X_t$  измерим.

Доказательство. □

Для любопытных приведем пример измеримого, но не прогрессивно измеримого процесса:

## 13.4 Еще задачи

# 14 Броуновское движение

## 14.1 Определение

Wiener process - Винеровский процесс

Wiener dog - такса

Дадим два определения броуновского движения.

С  $\sigma$ -алгебрами.

**Определение 62.1.** Случайный процесс  $W_t$  - броуновское движение по отношению к фильтрации  $\mathcal{F}_t$  если:

1.  $W_0 = 0$
2.  $W_t - W_s \sim N(0; t - s)$  при  $t > s$
3.  $W_t - W_s$  не зависит от  $\mathcal{F}_s$  при  $t > s$
4. Траектория  $W_t$  непрерывна с вероятностью 1.
5.  $W_t$  адаптирован к  $\mathcal{F}_t$

**Определение 62.2.** Случайный процесс  $W_t$  - броуновское движение если:

1.  $W_0 = 0$
2.  $W_t - W_s \sim N(0; t - s)$  при  $t > s$
3.  $W_{t_1} - W_{t_2}, W_{t_2} - W_{t_3}, \dots, W_{t_{n-1}} - W_{t_n}$  не зависимы
4. Траектория  $W_t$  непрерывна с вероятностью 1.

Второе определение получается из первого если «забыть» про фильтрацию.

Первое определение получается из второго если «добавить» подходящую фильтрацию. Например, подходящей фильтрацией будет естественная  $\mathcal{F}_t = \sigma(W_s, s \in [0; t])$

## 14.2 Три преобразования

## 14.3 Множество нулей

## 14.4 Непрерывное и недифференцируемое

Непрерывным броуновское движение является по определению. Оказывается, с вероятностью 1 траектория броуновского движения является нигде ни дифференцируемой. Интуиция довольно тонкая штука. С одной стороны, кажется что это невозможно: попробуйте нарисовать непрерывную нигде ни дифференцируемую линию! С другой стороны, с чего бы броуновскому движению быть дифференцируемым: берем произвольный момент времени  $t_0$  и отступаем от него на  $\Delta t$  вправо и влево. Два получаемых приращения,  $(W(t_0) - W(t_0 - \Delta t))$  и  $(W(t_0 + \Delta t) - W(t_0))$ , независимые случайные величины. А ведь для дифференцируемости они должны быть примерно равны при малом  $\Delta t$ !

## 14.5 Построения

## 14.6 Вариации

Скорости у него нет, но бегают быстро...

## 14.7 Внимательно посмотрим на приращение

Возьмем небольшой промежуток времени  $\Delta t > 0$ .

И посмотрим на случайные величины  $\Delta W$ ,  $(\Delta W)^2$ ,  $(\Delta W)^3$ ... в масштабе  $\Delta t$ .

$$\mathbb{E}(\Delta W) = 0, \text{Var}(\Delta W) = \Delta t$$

$$\mathbb{E}((\Delta W)^2) = \Delta t, \text{Var}((\Delta W)^2) = o(\Delta t)$$

Для  $k > 2$ :

$$\mathbb{E}((\Delta W)^k) = o(\Delta t), \text{Var}((\Delta W)^k) = o(\Delta t)$$

Надо продумать тут интуицию

Почему нам важна малость дисперсии а не стандартного отклонения? Ведь (упрощенно говоря), множество значений случайной величины - это  $[\mu - 2\sigma; \mu + 2\sigma]$ ? И согласно этой логике...

Дело в том, что мы будем работать только с суммами...

Мораль: если работать в масштабе  $\Delta t$  при  $\Delta t \approx 0$ , то  $\Delta W$  - это случайная величина с нулевым матожиданием, а  $(\Delta W)^2$  - это константа равная  $\Delta t$ .

Хм... А как ответить на этот аргумент?: На всякий случай напомним, что диапазон возможных значений примерно определяется с помощью стандартного отклонения и математического ожидания.

Пусть  $\sigma$  и  $\mu$  - это стандартное отклонение и мат. ожидание для  $X$ .

диапазон возможных значений не может быть уже, чем (...)

Неравенство Чебышева задает максимальную ширину диапазона «частых» значений  $X$ .

Например, если  $d = 3\sigma$ , то по неравенству Чебышева  $\mathbb{P}(X \in [\mu - 3\sigma; \mu + 3\sigma]) \geq 8/9$ .

Для наглядности мы на прямой покажем среднее значение плюс-минус три стандартных отклонения.

$$\mathbb{E}(\Delta W) = 0, \sqrt{\text{Var}(\Delta W)} = \sqrt{\Delta t} \dots$$

## 14.8 Пространство Омега для броуновского движения

Многих интересует вопрос, можно ли как-то более наглядно предъявить какое-нибудь пространство  $(\Omega, \mathcal{F}, \mathbb{P})$ , так чтобы на нем было броуновское движение  $W_t$ ?

Приведем два в каком-то смысле «полярных» примера, которые как-то проясняют ответ...

- Простые  $\Omega$ ,  $\mathcal{F}$ ,  $\mathbb{P}$ , но не очень явный  $W_t$ .

Берем в качестве  $\Omega$  отрезок  $[0; 1]$ , борелевскую  $\sigma$ -алгебру  $\mathcal{B}[0; 1]$  на этом отрезке и равномерную вероятность, «длину»,  $\lambda$  в качестве меры  $P$ . Само броуновское движение строится не очень явно. Сначала строим последовательность  $X_n$  независимых равномерно распределенных случайных величин. Ее мы строили (...). Затем преобразуем последовательность равномерных  $X_n$  в последовательность нормальных и по-прежнему независимых  $Z_n$ . А на базе последовательности  $Z_n$  строим броуновское движение. Способ указан!

- Простое  $\Omega$ ,  $\mathcal{F}$ , простой процесс  $W_t$ , но не очень явная  $\mathbb{P}$ .

Мы знаем, что траектории броуновского движения непрерывны, поэтому в качестве  $\Omega$  возьмем множество непрерывных на  $[0; \infty)$  функций. Обычно оно обозначается  $C[0; \infty)$ . Рассмотрим какую-нибудь метрику на этом пространстве. Например, в качестве расстояния между двумя непрерывными функциями можно взять  $d(f, g) = \sup_{x \geq 0} |f(x) - g(x)|$ . Эта метрика определяет открытые множества. А именно,  $\varepsilon$ -окрестность функции  $f$  это все функции  $g$ , такие, что  $d(f, g) < \varepsilon$ . С помощью понятия открытого множества зададим борелевскую  $\sigma$ -алгебру - минимальную  $\sigma$ -алгебру порожденную открытыми множествами.

Как определить  $X_t(w)$ ? Поскольку  $w$  - это готовая траектория, то  $X_t(w) := w(t)$ . А вот вероятность  $P$  задается неявно. От этой вероятности нужно многого потребовать. Например, того, что  $\mathbb{P}(w(10) < 0) = 0.5$ . Ее полное описание совпадет с определением броуновского движения. Утешает лишь теорема (...), доказывающая, что такая вероятность существует.

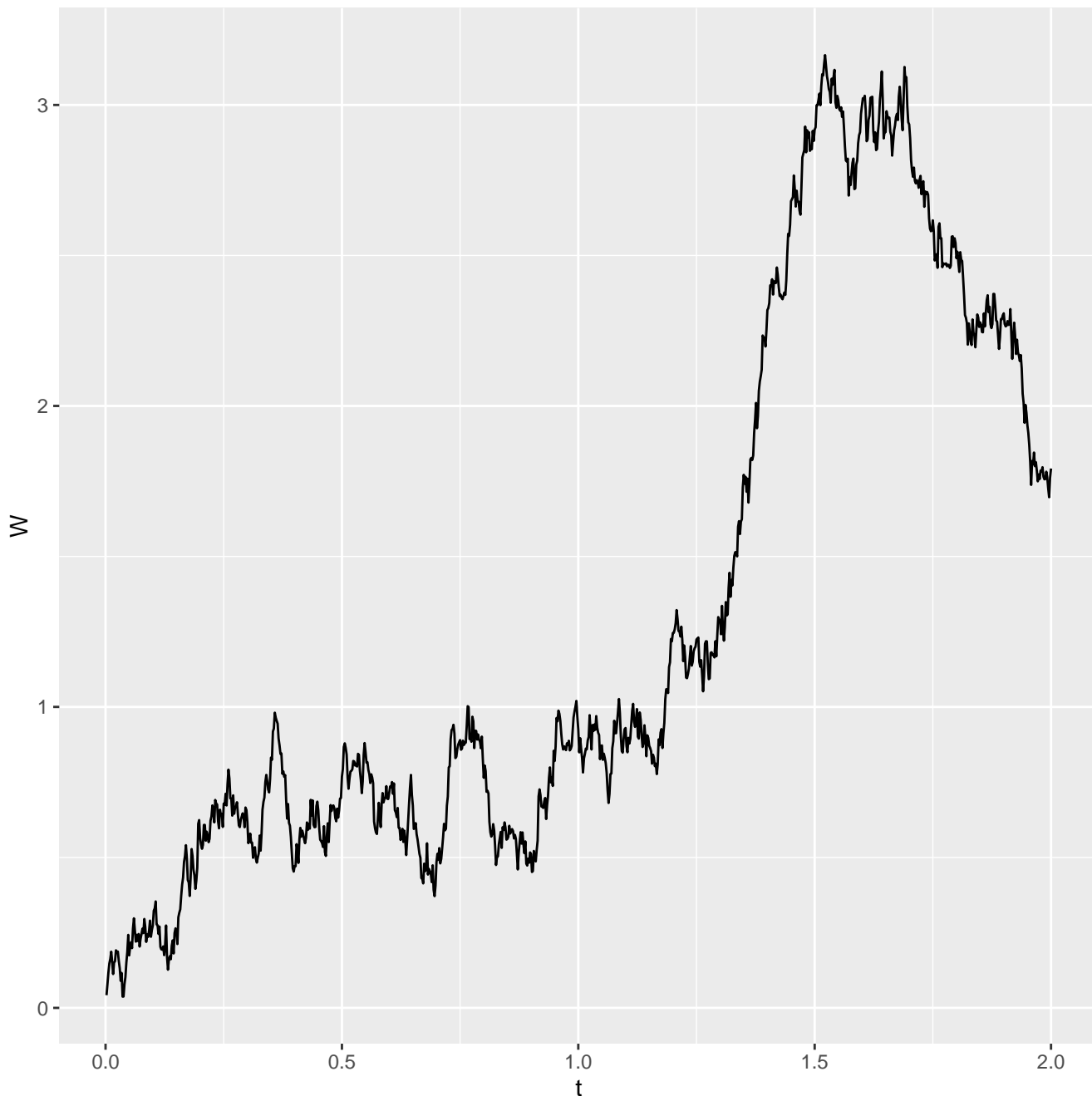
## 14.9 На компьютере

```
library("ggplot2")
## Warning: package 'ggplot2' was built under R version 3.2.4
library("dplyr")
## Warning: package 'dplyr' was built under R version 3.2.4
##
## Attaching package: 'dplyr'
## Следующие объекты скрыты от 'package:stats':
##
##   filter, lag
## Следующие объекты скрыты от 'package:base':
##
##   intersect, setdiff, setequal, union
n <- 1000
T <- 2
dt <- T/n

obs <- data_frame(dW = rnorm(n, mean=0, sd=sqrt(dt)),
                  t = seq(T/n, T, by=T/n),
                  W = cumsum(dW))

ggplot(data = obs, aes(x = t, y = W)) + geom_line()
```





## 14.10 Модель для игры в нарды

### 14.11 Еще задачи

**Задача 14.1.** Пусть  $W_t$  - винеровский процесс для  $t \in [0; 1]$ . Определим новый процесс для  $t \in [0; \infty)$  по формуле:

$$X_t = (1 + t) \left( W_{\left(\frac{t}{1+t}\right)} - \frac{tW_1}{1+t} \right) \quad (65.1)$$

Верно ли, что  $X_t$  - винеровский процесс?

**Задача 14.2.** Пусть  $W_t$  - стандартное броуновское движение и момент остановки  $\tau = \min\{t | W_t = 1\}$ . Найдите  $\mathbb{E}(\exp())$

**Задача 14.3.** Two questions on Brownian Motion:

1) A standard Brownian motion  $Z_t$  starting at position  $a > 0$ . There is an absorbing boundary at 0 - i.e. after  $Z_t$  first hits zero at time  $T_0$ ,  $Z_t$  stays at zero for  $t > T_0$ . What is the probability that  $Z_t$  is at zero at any arbitrary time  $t$ ?

2) Two independent standard Brownian motions  $Z_t$  and  $Z'_t$ .  $Z_t$  starts at position  $a > 0$ ,  $z'_t$  starts at position 0.  $Z_t$  and  $Z'_t$  are «mutually absorbing» - i.e. when  $Z_t = Z'_t$  for the first time at  $T_0$ , they both stay at  $Z_{T_0} = Z'_{T_0}$  for  $t > T_0$ . What is the probability that  $Z_t = Z'_t$  at any arbitrary time  $t$ ?

## 15 Интеграл Ито и лемма Ито

### 15.1 Интуиция и интеграл Ито для простых функций

Многие знают геометрический смысл обычного интеграла  $\int_0^t f(x)dx$  - это площадь под кривой  $f(x)$  на участке от  $[0; t]$ . Отдельные лица даже знают физический смысл обычного интеграла,  $\int_0^t f(x)dx$  - это расстояние пройденной машиной за промежуток времени  $[0; t]$ , если она будет двигаться со скоростью  $f(x)$ .

А какой смысл можно придать  $\int_0^t q(x)dp(x)$ ? Многие студенты прекрасно считают интегралы такого типа. Вряд ли кто затруднится посчитать что-то типа  $\int_0^1 \cos(x^2)d(x^2)$ . Настала пора раскрыть экономический смысл интеграла! А именно:  $\int_0^T q(t)dp(t)$  - это прибыль инвестора, если:

Цена акции в момент времени  $t$  задается функцией  $p(t)$

Количество акций в момент времени  $t$  на руках у инвестора задается функцией  $q(t)$

Ровно точно такой же смысл у интеграла-Ито. Интеграл  $\int_0^T s(t)dW(t)$  - это чистая прибыль инвестора за период  $[0; T]$ , если в момент времени  $t$  у него в наличии было  $s(t)$  акций, которые стоили  $W_t$ .

Для возрастающих функций можно предложить еще одну геометрическую интерпретацию. . . Площадь под линией  $(p(t), q(t))$  - тут есть некоторая сложность  $W(t)$  запросто может убывать, но эта мысль не повредит

Даже не зная формального определения, уже можно считать простейшие интегралы Ито. Единственное предостережение: функция  $s$  предполагается непрерывной слева. Представить себе это можно так: в момент времени  $t$ , когда у меня на руках  $s(t)$  акций я принимаю решение о покупке (продаже) и в любой момент времени  $(t + \varepsilon)$  у меня уже новое количество акций,  $s(t + \varepsilon)$ . Новые акции я покупал (продавал) по цене  $W_t$ . При интегрировании по броуновскому движению можно было бы взять и  $s(t)$  непрерывную справа - это ничего бы не изменило, т.к. броуновское движение непрерывно, однако при обобщении интеграла Ито в (...) это окажется важно.

**Пример 66.1.** Например, Вася руководствуется следующей стратегией: купить в момент времени  $t = 0$  одну акцию, в момент времени  $t = 1$  докупить еще одну акцию, если они будут стоить меньше 5 рублей. Найдите его чистую прибыль к моменту времени  $T = 3$ . В данном случае  $s(t, W_t) = \{$ . Вася получит прибыль  $W_3 - W_0$  с первой акции и  $W_3 - W_1$  со второй акции (если он ее купит, т.е. если  $W_1 < 5$ ), а значит  $\int_0^3 s(t, W_t)dW(t) = (W_3 - W_0) + 1_{W_1 < 5}(W_3 - W_1)$ .

Если покупка-продажа акций идет непрерывно, то необходимо заменить ее на частую покупку-продажу и переходить к пределам. Для интеграла Ито используется предел в смысле  $L^2$ .

**Пример 66.2.** Попробуем посчитать  $\int_0^T W dW$ , то есть в момент  $t$  мы держим на руках столько акций, сколько они сейчас стоят.

Разобьем интервал  $[0; T]$  на  $n$  равных частей, посчитаем интеграл Ито, а затем устремим  $n$  к бесконечности. Обозначаем для краткости:  $W_i = W(i \cdot \frac{T}{n})$  и  $\Delta W_i = W_i - W_{i-1}$ .

Сразу после момента времени  $(i - 1) \cdot \frac{T}{n}$  у нас на руках  $W_{i-1}$  акций, а сразу после момента времени  $i \cdot \frac{T}{n}$  должно быть  $W_i$  акций. Значит в момент времени  $t = i \cdot \frac{T}{n}$ , когда акции стоят  $W_i$  рублей, мы будем докупать  $W_i - W_{i-1}$  акций. Таким образом после докупки у нас будет как раз  $W_i$  акций.

В конце у нас будет  $W_n$  акций и мы их продадим по  $W_n$  рублей.

Наша прибыль составит выручка от продаж в конце минус расходы на все закупки:  $W_n^2 - \sum_{i=1}^n W_i \Delta W_i$

Заметим, что  $\sum_{i=1}^n (\Delta W_i)^2 = \sum_{i=1}^n (W_i^2 + W_{i-1}^2 - W_i W_{i-1} - W_i W_{i-1}) = \sum_{i=1}^n (W_i^2 + W_{i-1}^2 - W_i W_{i-1} - W_i W_{i-1}) - W_n^2 = 2 \sum_{i=1}^n W_i \Delta W_i - W_n^2$

Т.е. интересующую нас прибыль можно представить как  $\frac{W_n^2}{2} - \frac{1}{2} \sum_{i=1}^n (\Delta W_i)^2$

В (...) мы доказали, что  $\sum_{i=1}^n (\Delta W_i)^2 \rightarrow t$  в  $L^2$  при  $n \rightarrow \infty$

А это значит, что  $\int_0^T W dW = \frac{W(T)^2}{2} - \frac{t}{2}$

Вот стохастический анализ - это о том, как брать такие интегралы. И какими свойствами они обладают. Можно и такую аналогию провести. Обычный матанализ изучает функции, у которых существует производная, т.е. функции, локально похожие на прямую. Роль самой простой функции («прямой») в стохастическом анализе выполняет броуновское движение. Стохан изучает диффузии - случайные процессы локально похожие на броуновское движение.

Настала пора для формального определения. Мы построим интеграл Ито пошагово, как строили интеграл Лебега.

Наша цель - определить интеграл для достаточно широкого класса случайных процессов:

**Определение 67.1.** Пусть задано броуновское движение  $W_t$  и  $\mathcal{F}_t$  - его естественная фильтрация. Пусть  $T$  - фиксированный момент времени. Обозначим с помощью  $\mathcal{H}^2[0; T]$  множество процессов, удовлетворяющих трем условиям:

Н1. Процесс  $X_t$  адаптирован к фильтрации  $\mathcal{F}_t$

Н2. Процесс  $X_t$  измерим

Н3.  $\mathbb{E}(\int_0^T X_t^2 dt) < \infty$

Но начнем мы с более узкого класса, с **простых** случайных процессов:

**Определение 67.2.** Пусть задано броуновское движение  $W_t$  и  $\mathcal{F}_t$  - его естественная фильтрация. Пусть  $T$  - фиксированный момент времени. Назовем **простыми** и обозначим с помощью  $\mathcal{H}_0^2[0; T]$  множество процессов, удовлетворяющих условиям:

Н01. Процесс  $X_t \in \mathcal{H}^2[0; T]$

Н02. Процесс  $X_t$  представим в виде  $X_t = \sum_{i=0}^{n-1} A_i 1_{t \in (t_i; t_{i+1}]}$ , где  $t_i$  - константы, удовлетворяющие условию  $0 = t_0 < t_1 < \dots < t_n = T$ , а  $A_i$  -  $\mathcal{F}_{t_i}$ -измеримые случайные величины.

**Пример 67.3.**

Комментарий: из Н02 всегда следует Н1, Н2(?).

**Определение 67.4.** Для **простых** процессов назовем **интегралом Ито** случайную величину, определяемую по формуле:

$$\int_0^T X_t dW_t := \sum_{i=0}^{n-1} A_i (W(t_{i+1}) - W(t_i)).$$

Строго говоря, нужно убедиться, что определение корректно, а именно, что если один и тот же простой процесс имеет два разных представления, то использование любого из них приводит к одному и тому же результату. Это небольшое упражнение (...)

**Теорема 67.5.** Для **простых** случайных процессов интеграл Ито обладает свойствами:

РН01.  $\int_0^T a X_t dW_t = a \int_0^T X_t dW_t.$

РН02.  $\int_0^T (X_t + Y_t) dW_t = \int_0^T X_t dW_t + \int_0^T Y_t dW_t.$

РН03. Изометрия Ито.  $E \left( \left( \int_0^T X_t dW_t \right)^2 \right) = E \int_0^T X_t^2 dt$

В доказательствах мы для краткости будем использовать обозначение  $I(X(t))$  вместо  $\int_0^T X(t) dW(t).$

*Доказательство.* РН01. Сравнив обе части выражения видим, что они равны:

Левая:  $\int_0^T a X_t dW_t = \sum_{i=0}^{n-1} a A_i (W(t_{i+1}) - W(t_i)).$

Правая:  $a \int_0^T X_t dW_t = a \sum_{i=0}^{n-1} A_i (W(t_{i+1}) - W(t_i)).$

РН02. Как-то тавтологично даже. Возьмем  $0 = t_0 < t_1 < \dots < t_n = T$ , с помощью которого можно представить оба процесса. Заметим, что если  $X_t$  представим в виде  $X_t = \sum_{i=0}^{n-1} A_i 1_{t \in (t_i; t_{i+1}]}$ , а  $Y_t$  представим в виде  $Y_t = \sum_{i=0}^{n-1} B_i 1_{t \in (t_i; t_{i+1}]}$ , то процесс  $X_t + Y_t$  представим в виде:  $\sum_{i=0}^{n-1} (A_i + B_i) 1_{t \in (t_i; t_{i+1}]}$ . А теперь сравним:

Левая:  $\sum_{i=0}^{n-1} (A_i + B_i) (W(t_{i+1}) - W(t_i))$

Правая:  $\sum_{i=0}^{n-1} A_i(W(t_{i+1}) - W(t_i)) + \sum_{i=0}^{n-1} B_i(W(t_{i+1}) - W(t_i))$

РНОЗ. Снова сравним обе части:

Левая. Шаг 1.  $(\int_0^T X_t dW_t)^2 = \sum_{i,j} A_i A_j (W(t_{i+1}) - W(t_i))(W(t_{j+1}) - W(t_j))$ . Если взять  $i \neq j$ , например,  $i < j$ , то сомножитель  $(W(t_{j+1}) - W(t_j))$  не зависит от остальных сомножителей. Остальные сомножители являются  $\mathcal{F}_{t_j}$ -измеримыми, а  $(W(t_{j+1}) - W(t_j))$  не зависит  $\sigma$ -алгебры  $\mathcal{F}_{t_j}$ . Значит при  $i \neq j$  ожидание  $\mathbb{E}(A_i A_j (W(t_{i+1}) - W(t_i))(W(t_{j+1}) - W(t_j))) = \mathbb{E}(\dots)\mathbb{E}(W(t_{j+1}) - W(t_j)) = (\dots) \cdot 0 = 0$ .

Шаг 2. После шага 1 от всей суммы осталось только:  $\sum_i A_i^2 (W(t_{i+1}) - W(t_i))^2$ . Замечаем, что  $A_i^2$  не зависит от  $(W(t_{i+1}) - W(t_i))^2$ , т.к.  $A_i^2$  - это  $\mathcal{F}_{t_i}$ -измеримая случайная величина. И, наконец,  $\mathbb{E}((\int_0^T X_t dW_t)^2) = \mathbb{E}(\sum_i A_i^2 (W(t_{i+1}) - W(t_i))^2) = \sum_i \mathbb{E}(A_i^2) \mathbb{E}(W(t_{i+1}) - W(t_i))^2 = \sum_i \mathbb{E}(A_i^2)(t_{i+1} - t_i)$ .

Правая. С правой проще. Заметив, что  $X_t^2 = \sum_{i=0}^{n-1} A_i^2 1_{t \in (t_i; t_{i+1}]}$  (при  $i \neq j$  произведение индикаторов обнуляется), получаем, что  $\mathbb{E}(\int_0^T X_t^2 dt) = \sum_{i=0}^{n-1} \mathbb{E}(A_i^2)(t_{i+1} - t_i)$ .  $\square$

## 15.2 Интеграл Ито для аш2

Итак, интеграл Ито определен для процессов из  $\mathcal{H}_0[0; T]$ . Как расширить его на процессы из  $\mathcal{H}[0; T]$ ?

Идея нового, более широкого, определения проста:

Каждый процесс  $X(t)$  из  $\mathcal{H}[0; T]$  является пределом некоей последовательности простых процессов из  $\mathcal{H}_0[0; T]$ . Для каждого простого процесса  $X_n(t)$  интеграл Ито определен и является случайной величиной, обозначим ее  $I_n$ . Оказывается, что последовательность интегралов Ито  $I_n$  всегда сходится в пространстве  $L^2(\Omega)$  к некоторой случайной величине  $I$ . И этот предел  $I$  не зависит от выбора последовательности простых процессов  $X_n(t)$ , сходящейся к  $X(t)$ . Этот предел  $I$  мы и назовем интегралом Ито для процесса  $X_t$ .

Сразу следует заметить, что интеграл Ито определен как предел в смысле  $L^2(\Omega)$ , поэтому единственности в поточечном смысле мы гарантировать не сможем. Интеграл Ито в  $\mathcal{H}^2[0; T]$  определен с точностью до множества меры нуль. Несколько случайных величины равных почти наверное могут представлять один и тот же интеграл Ито. теперь все утверждения касающиеся Интеграла Ито будут формулироваться с припиской «почти наверное».

Формальное определение выглядит так:

**Определение 68.1.** Пусть  $X(t) \in \mathcal{H}^2[0; T]$  и  $X_n(t)$  произвольная последовательность простых процессов, сходящаяся к процессу  $X(t)$  в  $L^2(\Omega \times \mathbb{R})$ . Пусть последовательность случайных величин  $I(X_n(t))$  сходится в  $L^2(\Omega)$  к случайной величине  $I$ . Интегралом Ито случайного процесса  $X(t)$  называется случайная величина  $I$ , обозначается она как  $\int_0^T X(t) dW(t)$ .

Корректность определения подкрепляют 3 «кита»:

**Теорема 68.2.** Множество  $\mathcal{H}_0^2[0; T]$  плотно в  $\mathcal{H}^2[0; T]$ . Для любого процесса  $X(t) \in \mathcal{H}^2$  существует последовательность  $X_n(t) \in \mathcal{H}_0^2[0; T]$ , такая что  $\|X_n(t) - X(t)\|_{L^2(\Omega \times \mathbb{R})} \rightarrow 0$ .

*Доказательство.* Доказательство кита отложим до (...)  $\square$

**Теорема 68.3.** Если есть последовательность простых процессов  $X_n(t) \in \mathcal{H}_0^2[0; T]$  сходится к процессу  $X_t$ , т.е.  $\|X_n(t) - X(t)\|_{L^2(\Omega \times \mathbb{R})} \rightarrow 0$ , то последовательность  $I_n = \int_0^T X_t^{(n)} dW_t$  сходится к некоторой случайной величине  $I$  в  $L^2(\Omega)$ .

*Доказательство.* Шаг 1. Последовательность  $(X_n(t) - X(t))$  сходится к нулю в  $L^2(\Omega \times \mathbb{R})$ , значит  $X_n(t)$  - это последовательность Коши. Действительно,  $\|X_i(t) - X_j(t)\| = \|X_i(t) - X(t) + X(t) - X_j(t)\| \leq \|X_i(t) - X(t)\| + \|X(t) - X_j(t)\|$ . Для любого  $\varepsilon > 0$ , можно выбрать такое  $N$ , начиная с которого, т.е. при  $i > N$  и  $j > N$  оба слагаемых будут меньше  $\varepsilon/2$ .

Шаг 2. Последовательность  $X_n(t)$  - это последовательность Коши в  $L^2(\Omega \times \mathbb{R})$ . В силу изометрии Ито последовательность  $I_n = \int_0^T X_n(t) dW(t)$  также будет последовательностью Коши

в  $L^2(\Omega)$ . Подробнее. Процесс  $Z(t) = X_i(t) - X_j(t)$  является простым, поэтому  $\|Z(t)\|_{L^2(\Omega \times \mathbb{R})} = \|\int_0^T Z(t)dW(t)\|_{L^2(\Omega)}$ . Кроме того,  $\int_0^T Z(t)dW(t) = \int_0^T X_i(t)dW(t) - \int_0^T X_j(t)dW(t) = I_i - I_j$ . Значит  $\|X_i(t) - X_j(t)\|_{L^2(\Omega \times \mathbb{R})} = \|I_i - I_j\|_{L^2(\Omega)}$ .

Шаг 3. Пространство  $L^2(\Omega)$  полное, поэтому любая последовательность Коши  $I_n$  сходится к некоторой случайной величине  $I$ .  $\square$

**Теорема 69.1.** *Если две различные последовательности процессов  $X_n(t)$  и  $Y_n(t)$  сходятся в  $L^2(\Omega \times \mathbb{R})$  к одному пределу  $X(t)$ , то и последовательность интегралов Ито,  $I(X_n(t))$  и  $I(Y_n(t))$  сходятся к одному пределу в  $L^2(\Omega)$ .*

*Доказательство.* Шаг 1. Заметим, что разница  $\|X_n(t) - Y_n(t)\|$  сходится к нулю в силу неравенства треугольника:  $\|X_n(t) - Y_n(t)\| = \|X_n(t) - X(t) + X(t) - Y_n(t)\| \leq \|X_n(t) - X(t)\| + \|X(t) - Y_n(t)\|$ . В этой сумме каждое слагаемое стремится к нулю с ростом  $n$ .

Шаг 2. В силу изометрии Ито,  $\|X_n(t) - Y_n(t)\| = \|I(X_n(t)) - I(Y_n(t))\|_{L^2(\Omega)}$ . Значит последовательность  $\|I(X_n) - I(Y_n)\|$  стремится к нулю.

Шаг 3. В силу теоремы (...) последовательность  $I(X_n(t))$  сходится к некоей случайной величине  $I$ . Используем неравенство треугольника еще раз,  $\|I(Y_n(t)) - I\| = \|I(Y_n(t)) - I(X_n(t)) + I(X_n(t)) - I\| \leq \|I(Y_n(t)) - I(X_n(t))\| + \|I(X_n(t)) - I\|$ . Оба слагаемых стремятся к нулю, первое в силу шага 2, второе в силу сходимости  $I(X_n(t))$ . Значит  $I(Y_n(t))$  стремится к той же случайной величине  $I$ .  $\square$

теперь интеграл Ито корректно определен для случайных величин из  $\mathcal{H}^2[0; T]$ ! Попутно заметим, что для простых процессов новое определение совпадает со старым, т.к. в качестве последовательности сходящейся к простому процессу  $X(t)$  можно взять последовательность  $X(t), X(t), X(t), \dots$ . Поэтому можно использовать одно и то же обозначение для интеграла Ито на  $\mathcal{H}^2[0; T]$  и на  $\mathcal{H}_0^2[0; T]$

Интеграл Ито снова обладает хорошими свойствами:

**Теорема 69.2.** *Для случайных процессов из  $\mathcal{H}^2[0; T]$  интеграл Ито обладает свойствами:*

РН1.  $\int_0^T aX(t)dW(t) = a \int_0^T X(t)dW(t)$  (почти наверное).

РН2.  $\int_0^T (X_t + Y_t)dW_t = \int_0^T X_t dW_t + \int_0^T Y_t dW_t$  (почти наверное).

РН3. Изометрия Ито.  $E \left( \left( \int_0^T X_t dW_t \right)^2 \right) = E \int_0^T X_t^2 dt$

РН4. Условная изометрия Ито.

*Доказательство.* РН1. Берем последовательность  $X_n(t)$ , сходящуюся к  $X(t)$ , автоматически оказывается, что  $aX_n(t)$  сходится к  $aX(t)$ , обе сходимости, конечно, в  $L^2(\Omega \times \mathbb{R})$ . Но для простых процессов свойство РН01 доказано, значит  $I(aX_n(t)) = aI(X_n(t))$ . Последовательность  $I(aX_n(t))$  сходится к  $I(aX(t))$ , а последовательность  $aI(X_n(t))$  сходится к  $aI(X(t))$ , на этот раз обе сходимости в  $L^2(\Omega)$ . Значит  $I(aX(t))$  и  $aI(X(t))$  равны почти наверное.

РН2. Берем последовательность  $X_n(t)$ , сходящуюся в  $L^2(\Omega \times \mathbb{R})$  к процессу  $X(t)$  и последовательность  $Y_n(t)$ , сходящуюся в  $L^2(\Omega \times \mathbb{R})$  к процессу  $Y(t)$ . Последовательность  $X_n(t) + Y_n(t)$  будет сходиться к  $X(t) + Y(t)$ . Для простых процессов свойство РН02 доказано, значит  $I(X_n(t) + Y_n(t)) = I(X_n(t)) + I(Y_n(t))$ . По определению, предел последовательности  $I(X_n(t))$  равен  $I(X(t))$ , предел последовательности  $I(Y_n(t))$  равен  $I(Y(t))$ , а предел последовательности  $I(X_n(t) + Y_n(t))$  равен  $I(X(t) + Y(t))$ . Значит  $I(X(t) + Y(t)) = I(X(t)) + I(Y(t))$ .

РН3. Существует последовательность простых процессов  $X_n(t)$ , сходящаяся к  $X(t)$ . Применяем к каждому ее члену изометрию Ито (доказанную для простых процессов:  $\|X_n(t)\|_{L^2(\Omega \times \mathbb{R})} = \|I(X_n(t))\|_{L^2(\Omega)}$ ). Последовательность процессов  $X_n(t)$  сходится в  $L^2(\Omega \times \mathbb{R})$  к  $X(t)$ , а значит  $\|X_n(t)\| \rightarrow \|X(t)\|$ . Последовательность случайных величин  $I(X_n(t))$  сходится в  $L^2(\Omega)$  к  $I(X(t))$ , а значит  $\|I(X_n(t))\| \rightarrow \|I(X(t))\|$ . И, следовательно,  $\|I(X(t))\| = \|X(t)\|$ .  $\square$

## 15.3 Интеграл Ито как процесс

## 15.4 Ito's lemma

Почему когда я происходил от обезьяны меня не предупредили, что придется учить доказательство леммы Ито?

Когда надо посчитать обычный интеграл типа  $\int_0^T \cos(3x + 7)dx$ , никто не будет считать его как предел сумм. Обычно пользуются теоремой Ньютона-Лейбница, знанием производных основных функций и свойствами производной. Аналогично и в стохастическом анализе. В тех редких случаях когда стохастический интеграл можно посчитать в явном виде, для подсчета обычно хватает леммы Ито и понимание смысла краткой записи.

Лемма Ито в краткой записи:

**Теорема 70.1.**

Лемма Ито в полной записи:

**Теорема 70.2.**

## 15.5 Использование процесса Ито при моделировании

Данный раздел списан с источника [lones:dte]

Допустим у нас есть некий процесс  $Y_t$ , поведение которого без случайных шоков описывается обыкновенным дифференциальным уравнением  $dY = f(t, Y)dt$ .

Для описания шоков предположим, что время поделено на маленькие отрезочки длины  $1/n$ . И на каждом отрезке может с вероятностью  $O(1/n)$  произойти случайный шок, который превращает  $Y_t$  в  $(1 + O(1/n))Y_t$ .

Оказывается, что в этом случае процесс  $Y_t$  хорошо описывается с помощью процесса Ито.

Рассмотрим пару конкретных примеров.

(lones)

(steele, exam)

<http://math.stackexchange.com/questions/284874/why-do-people-simulate-with-brownian-mo>

**Определение 70.3.** Процессом Ито называется случайный процесс представимый в виде:

Функция

## 15.6 Еще задачи

**Задача 15.1.** Пусть  $dX = udt + v dW$ , где  $u$  и  $v$  - не обязательно константы. Пусть  $Y_t = \int_0^t X_a da$ . Найдите  $dY_t$ .

# 16 Основные теоремы стохастического анализа по броуновскому движению

## 16.1 Медвежья услуга

Инопланетяне давно уже прибыли на Землю. Они пытаются общаться с нами с помощью дорожных знаков, но инопланетян никто не понимает. . .

Условные обозначения, которые были призваны упростить запись, оказались дополнительным барьером для понимания стохастического анализа. Например, часто используется обозначение типа  $\mathbb{E}(dW) = 0$ . Проблема в том, что  $dW$  - не существует! Это сбивает с толку тех, кто только начинает осваивать стохастический анализ. Несмотря на неудачные условные обозначения, книга может быть очень удачной, поэтому приведем полезный «словарь»:

$$dX = a(t, X)dt + b(t, X)dW \quad X_T = X_0 + \int_0^T a(t, X)dt + \int_0^T b(t, X)dW$$

$$\mathbb{E}(b(t, X)dW) = 0 \quad \mathbb{E}(\int_0^T b(t, X)dW) = 0, \text{ если } \dots$$

$$\mathbb{E}(dW) = 0 \quad \mathbb{E}(\Delta W) = 0$$

$$\text{Var}(b(t, X)dW) = b^2(t, X)dt \quad \text{Var}(\int_0^T b(t, X)dW) = \int_0^T \mathbb{E}(b^2(t, X))dt$$

$$\text{Var}(dW) = dt \quad \text{Var}(\Delta W) = \Delta t$$

$$dW \cdot dW = dt, \quad dt \cdot d(\text{неважно что}) = 0 \quad \text{способ запомнить формулу Ито}$$

Мы говорим Ленин, подразумеваем - партия, мы говорим партия, подразумеваем - Ленин.  
В.В. Маяковский. Владимир Ильич Ленин.

Вот так мы уже пятьдесят лет говорим одно, а подразумеваем совсем другое. (анекдот)

Вот потому, что вы говорите то, что не думаете и думаете то, что не думаете, вот в клетках и сидите. И вообще, весь этот горький катаклизм, который я тут наблюдаю... и Владимир Николаич тоже... (Скрипач, Кин-дза-дза)

## 16.2 теорема Гирсанова

Простейшая версия теоремы Гирсанова формулируется

**Теорема 71.1.** Если  $W_t$  — винеровский процесс относительно вероятности  $\mathbb{P}$  и  $\tilde{W}_t = \alpha t + W_t$ , то существует вероятность  $\tilde{\mathbb{P}}$ , относительно которой процесс  $\tilde{W}_t$  будет винеровским.

Более сложные версии теоремы Гирсанова

## 16.3 Martingale representation theorem

## 16.4 Feynman-Kac

Иногда теория вероятностей помогает решать задачи из других областей математики. Например, позволяет находить решения детерминистических дифференциальных уравнений.

**Теорема 71.2.** Пусть  $q : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f : \mathbb{R} \rightarrow \mathbb{R}$  две ограниченные функции. Решение уравнения  $\frac{\partial u(t, x)}{\partial t} = \frac{1}{2} \frac{\partial^2 u(t, x)}{\partial x^2} + q(x)u(t, x)$  с начальным условием  $u(0, x) = f(x)$  является единственным и его можно представить в виде  $u(t, x) = \mathbb{E}(f(x + W_t) \exp(\int_0^t q(x + W_s)ds))$ .

До доказательства рассмотрим пару примеров. (...)

*Доказательство.* Шаг 1. Применим лемму Ито к функции  $f(s, y) := u(t - s, x + y)$ :

$$df(s, W_s) = (-u_1(t - s, x + W_s) + \frac{1}{2}u_{22}(t - s, x + W_s))ds + u_2(t - s, x + W_s)dW_s.$$

Шаг 2. Функция  $u$  удовлетворяет исходному детерминистическому дифференциальному уравнению, поэтому:

$$df(s, W_s) = -q(x + W_s)u(t - s, x + W_s)ds + u_2(t - s, x + W_s)dW_s.$$

Шаг 3. Применим лемму Ито к случайному процессу  $M_s := u(t - s, x + W_s) \exp(\int_0^s q(x + W_v)dv)$ :

Процесс  $M_s$  является произведением, поэтому полезно вспомнить упражнение (...), в котором мы доказали, что  $d(XY) = (dX)Y + (dY)X + dX dY$ .

$$\begin{aligned} dM_s &= df \cdot \exp(\int_0^s q(x + W_v)dv) + f \cdot d(\exp(\int_0^s q(x + W_v)dv)) + df \cdot d(\exp(\int_0^s q(x + W_v)dv)) = \\ &= -q(x + W_s)u(t - s, x + W_s) \exp(\int_0^s q(x + W_v)dv)ds + u_2(t - s, x + W_s) \exp(\int_0^s q(x + W_v)dv)dW_s + f q(x + W_s) \exp(\int_0^s q(x + W_v)dv)ds \\ &= u_2(t - s, x + W_s) \exp(\int_0^s q(x + W_v)dv)dW_s \quad (71.3) \end{aligned}$$

Шаг 4. Значит  $M_v = M_0 + \int_0^v u_2(t - s, x + W_s) \exp(\int_0^s q(x + W_v)dv)dW_s$ , и  $M_v$  является мартингалом.

Шаг 5. По построению,  $M_0 = u(t, x)$  - неслучайная функция,  $M_0 = \mathbb{E}(M_0)$ . В силу того, что  $M_v$  - мартингал,  $\mathbb{E}(M_0) = \mathbb{E}(M_t)$ . Откуда и следует утверждение теоремы:

$$u(t, x) = \mathbb{E}(f(x + W_t) \exp(\int_0^t q(x + W_s) ds)).$$

□

(откуда следует единственность?)

Изложить по Steele!!!!

Взять примеры из Bjork!

<http://galton.uchicago.edu/~lalley/Courses/391/Lecture12.pdf>

## 16.5 Еще задачи

## 17 Стохастические уравнения

Нам важно четко ответить на несколько вопросов:

Что такое стохастическое дифференциальное уравнение?

Что мы подразумеваем под решением?

Какие методы решения есть? Какие есть самые простые виды уравнений?

Существует ли решение и единственно ли оно?

Итак, по порядку.

### 17.1 Определение и виды решений

**Определение 72.1.** Уравнение вида  $dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$  называется **стохастическим дифференциальным уравнением**.

Естественно, это уравнение не более чем сокращенная запись и полное определение выглядело бы так:

**Определение 72.2.** Уравнение вида  $X_t = x_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s$  называется **стохастическим дифференциальным уравнением**.

Решения бывают двух типов, сильные и слабые. Если помимо уравнения (...) также заданы и броуновское движение, и фильтрация и нужно найти процесс  $X_t$ , который был бы функцией заданного броуновского движения и адаптирован (прогрессивно измерим и т.д.) к данной фильтрации, то речь идет о поиске сильного решения. Если задано только уравнение (...) и в качестве решения можно предъявить любое броуновское движение и любой  $X_t$  связанный с ним уравнением (...), то речь идет о слабом решении.

**Определение 72.3.** Процесс  $X_t$  называется ...

**Определение 72.4.** Процесс  $X_t$  называется ...

Из этих двух определений непосредственно следует теорема

**Теорема 72.5.** Если существует сильное решение стохастического дифференциального уравнения, то существует и слабое решение.

Можно продемонстрировать ситуации, когда слабое решение существует, а сильное - нет.

**Пример 72.6.** ...

### 17.2 Простые СДУ и аналитические методы решения

Несколько идей, которые могут помочь решить СДУ в явном виде.

- Домножение на некоторую функцию.  $Y = f(t, W) \cdot X$ . Например, если в уравнении есть кусок вида  $dX = (aX + \dots)dt + (\dots)dW$ , то может помочь замена  $Y = e^{-at}X$ .
- Поиск решения простого вида.  $Y = f(t)g(W)$ .
- Замена  $Y' = f(Y)$ .

**Пример 72.7.** Иногда  $f$  можно просто угадать...

**Пример 72.8.** Функцию  $f$  можно искать специально, чтобы сделать постоянным коэффициент при  $dW$ .



- Сначала решить детерминистическое уравнение.  $dY = (\dots)dt$ . Это может привести к хорошей замене.

### 17.3 Численные методы решения СДУ

### 17.4 Существование и единственность

## 18 Простые модели опционов

Для всех рассматриваемых моделей мы будем предполагать постоянную процентную ставку  $r$ , наличие двух оригинальных ценных бумаг: акций с ценой  $S_t$  в момент времени  $t$  и облигаций с ценой  $B_t$  в момент времени  $t$ .

### 18.1 Binomial model

Время дискретно,  $t \in \{0, 1, 2, \dots\}$ .

...

Если «правильным» образом взять предел дискретной модели, то можно получить те же результаты, что и в непрерывном случае.

В дискретной модели время исполнения опциона и количество шагов - это одно и то же. При переходе к пределу нужно разделить эти два понятия. Время исполнения опциона нужно приравнять к константе  $T$ , а число шагов  $n$  устремить к бесконечности.

### 18.2 Портфель в непрерывном времени

Для того, чтобы описать портфель ценных бумаг, необходимо указать в каком количестве входит в портфель каждая ценная бумага. Портфель  $V$  будет состоять из двух ценных бумаг, акций в количестве  $X$  и облигаций в количестве  $Y$ . т.е. стоимость портфеля будет равна  $V_t = X_t S_t + Y_t B_t$ . При этом количество ценных бумаг в портфеле, конечно, может зависеть от их прошлых цен, но не может «заглядывать» вперед, т.е. есть процессы  $X_t$  и  $Y_t$  должны быть предсказуемыми. Формально,

**Определение 73.1.** Портфель  $(X, Y)$  - это набор из двух предсказуемых процессов,  $X_t$  и  $Y_t$ .

Более того: дядюшки, который нам оставил бы наследство у нас нет, поэтому кроме как из самого портфеля денег на покупку ценных бумаг нам взять неоткуда. Портфель должен финансировать сам себя!

**Определение 73.2.** Портфель  $(X, Y)$  называется **self-financing**, если  $dV = X_t dS_t + Y_t dB_t$

### 18.3 Модель Блэка-Шоулса

Предпосылки:

1. Время непрерывно
2. Есть два актива, акции и облигации
3. Активы можно покупать в дробном количестве
4. Стоимость акций,  $S_t = S_0 \exp((\mu - \frac{\sigma^2}{2})t + \sigma W_t)$
5. Стоимость облигаций,  $B_t = \exp(rt)$
6. Транзакционные издержки отсутствуют

Факт 1.

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

Факт 2. Если  $A_t$  — количество акций на руках, а  $X_t$  — стоимость портфеля из акций и облигаций, то  $(X_t - A_t S_t)/B_t$  — количество облигаций.

$$dX_t = A_t dS_t + (X_t - A_t S_t)/B_t dB_t$$

Факт 3. Существует портфель, реплицирующий опцион.

$$d(e^{-rt} X_t) = \dots = \left( \frac{\mu - r}{\sigma} t + W_t \right)$$

Из предположения о невозможности арбитража следует, что текущая цена актива должна равняться текущей цене реплицирующего портфеля.

Сделаем замену  $\tilde{W}_t = \frac{\mu - r}{\sigma} t + W_t$ . При этом  $d\tilde{W}_t = \frac{\mu - r}{\sigma} dt + dW_t$ .

Значит

$$d(e^{-rt} X_t) = (\dots) d\tilde{W}_t$$

Теорема Гирсанова.

По теореме Гирсанова существует вероятность  $\tilde{\mathbb{P}}$ , относительно которой процесс  $\tilde{W}_t = at + W_t$  будет винеровским процессом.

Здесь  $\tilde{W}_t = \frac{\mu - r}{\sigma} t + W_t$

Относительно вероятности  $\tilde{\mathbb{P}}$  процесс  $e^{-rt} X_t$  будет мартингалом, и, следовательно,

$$X_0 = \mathbb{E}_{\tilde{\mathbb{P}}}(e^{-rT} X_T) = e^{-rT} \mathbb{E}_{\tilde{\mathbb{P}}}(X_T)$$

Примеры подсчета стоимости активов.

Как реплицировать? pde...

## 18.4 Модель Луи Башелье

Время непрерывно. Есть два актива, которые можно покупать в любом (даже в отрицательном количестве):

Наличные деньги (на них процент не начисляется), стоимость  $B_t = 1$ .

Акции, стоимость  $S_t = S_0 + \sigma W_t$ .

## 18.5 Еще задачи

## 19 Приложения

### 19.1 Напоминалка по мат. анализу

А может мы ее вперед вынесем вместе с бесконечностями...

### 19.2 Доказательство теоремы Каратеодори

**Теорема 74.1.** Вероятность  $\mathbf{P}$ , заданная на алгебре может быть продолжена до вероятности  $\mathbf{P}^*$ , определенной на  $\sigma$ -алгебре  $\mathcal{U}$ , содержащей алгебру  $\mathcal{A}$ . Это продолжение единственно.

Заметим, что указанная  $\sigma$ -алгебра  $\mathcal{U}$  окажется пополнением  $\sigma(\mathcal{A})$ .

Доказательство мы проведем в три этапа и, как оказалось, в две лекции.

1. Для произвольного подмножества  $A \subset \Omega$  определим внешнюю меру  $\mathbf{P}^*(A)$ . Эта внешняя мера будет обладать рядом неплохих свойств.

2. Набор таких подмножеств, где  $\mathbf{P}^*$  обладает очень хорошими свойствами (а именно, всеми свойствами вероятности) окажется  $\sigma$ -алгеброй  $\mathcal{U}$ .

3. Любая другая вероятность на  $\mathcal{U}$  совпадет с  $\mathbf{P}^*$ .

Для определения внешней меры нам потребуется понятие инфинума. Инфинум следует понимать как минимум в тех ситуациях, когда сам минимум не достижим. Пример. Пусть  $x \in (0; 1]$ . Чему равен минимальный  $x$ ? Хотелось бы сказать нулю, но нельзя, потому, что ноль - выколота точка. Так вот в этом примере инфинум равен нулю. Ниже следует формальное определение.

Пусть  $A \subset \mathbb{R}$ . Число  $x \in \mathbb{R}$  называется *инфинумом* множества  $A$ , если  $\forall a \in A \Rightarrow x \leq a$  и в любой  $\varepsilon$ -окрестности точки  $x$  найдется хотя бы одна точка из множества  $A$ .

Если множество  $A$  не ограничено снизу, то инфинум (равно как и минимума) не будет.

Упражнение. Дайте формальное определение супремума - аналога максимума.

Вернемся к определению внешней меры  $\mathbf{P}^*$ .

Рассмотрим произвольное множество  $A \subset \Omega$ . Накроем множество  $A$  набором попарно непересекающихся множеств из алгебры.

Все наборы множеств  $\{A_i\}$ , накрывающие  $A$  и принадлежащие алгебре, будем обозначать символом  $\gamma(A)$ .

Для произвольного  $A$  определим  $\mathbf{P}^*(A) = \inf_{\{A_i\} \in \gamma(A)} (\sum_i \mathbf{P}(A_i))$ .

Внешняя мера (external measure)  $\mathbf{P}^*$  обладает свойствами:

ЕМ1. Если  $A \subset B$ , то  $\mathbf{P}^*(A) \leq \mathbf{P}^*(B) \leq 1$ .

ЕМ2. Если  $A_i \in \mathcal{A}$  и попарно не пересекаются, то  $\mathbf{P}^*(\bigcup_i A_i) = \sum_i \mathbf{P}(A_i)$ .

ЕМ3.  $\mathbf{P}^*(\bigcup_i A_i) \leq \sum_i \mathbf{P}^*(A_i)$

Доказательство ЕМ1 остается в качестве упражнения.

Перейдем к доказательству ЕМ2. Итак,  $A = \bigcup_i A_i$ , где  $A_i$  попарно не пересекаются и лежат в алгебре.

Во-первых, докажем, что  $\mathbf{P}^*(A) \geq \sum_i \mathbf{P}(A_i)$ .

Для этого рассмотрим произвольную последовательность  $\{C_i\}$ , накрывающую  $A$ .

Из того, что последовательность  $\{C_i\}$  накрывает  $A$ , следует, что  $\bigcup_m (A_i \cap C_m) = A_i$ .

Далее Суворов переходит через Альпы с комментариями:

$$\sum_i \mathbf{P}(A_i) = \sum_i \mathbf{P}\left(\bigcup_m (A_i \cap C_m)\right)$$

В силу того, что  $C_m$  не пересекаются:

$$\sum_i \mathbf{P}\left(\bigcup_m (A_i \cap C_m)\right) = \sum_i \sum_m \mathbf{P}(A_i \cap C_m)$$

В силу первого свойства внешней меры  $\mathbf{P}^*$  этот ряд сходится и значки сумм можно переставлять:

$$\sum_i \sum_m \mathbf{P}(A_i \cap C_m) = \sum_m \sum_i \mathbf{P}(A_i \cap C_m)$$

Поскольку для любого  $n$  выполнено  $\bigcup_{i=1}^n (A_i \cap C_m) \subset C_m$ , и объединяемые множества попарно не пересекаются, то для любого  $n$  будет верно неравенство  $\sum_{i=1}^n \mathbf{P}(A_i \cap C_m) \leq \mathbf{P}(C_m)$ . Если неравенство верно для произвольного  $n$ , то оно верно и для соответствующего предела  $\sum_{i=1}^\infty \mathbf{P}(A_i \cap C_m) \leq \mathbf{P}(C_m)$ . Получаем решающий переход:

$$\sum_m \sum_i \mathbf{P}(A_i \cap C_m) \leq \sum_m \mathbf{P}(C_m)$$

Выходит, что каким бы ни было покрытие  $\{C_m\}$  множества  $A$  будет выполнено неравенство:

$$\sum_i \mathbf{P}(A_i) \leq \sum_m \mathbf{P}(C_m)$$

Следовательно и

$$\mathbf{P}^*(A) = \inf_{\{C_i\} \in \gamma(A)} \left( \sum_m \mathbf{P}(C_m) \right) \geq \sum_i \mathbf{P}(A_i)$$

Во-вторых, докажем, что  $\mathbf{P}^*(A) \geq \sum_i \mathbf{P}(A_i)$ . Это очень легко, т.к.  $\{A_i\}$  - сами представляют собой последовательность накрывающую  $A$ , то  $\mathbf{P}^*(A) = \inf_{\{C_i\} \in \gamma(A)} (\sum_i \mathbf{P}(C_i)) \leq \sum_i \mathbf{P}(A_i)$ .

Один из тайных смыслов свойства ЕМ2 заключается в том, что на алгебре  $\mathcal{A}$  внешняя мера  $\mathbf{P}^*$  совпадает с уже имеющейся вероятностью  $\mathbf{P}$ .

Доказательство ЕМ3:

Сначала докажем свойство ЕМ3 для множеств  $A_i$ , лежащих в  $\mathcal{A}$ :

Старый трюк! От последовательности  $A_i$ , где возможны попарные пересечения, перейдем к последовательности  $\widetilde{A}_i$ , где попарных пересечений нет.

$$\mathbf{P}^*\left(\bigcup_i A_i\right) = \mathbf{P}^*\left(\bigcup_i \widetilde{A}_i\right)$$

По только что доказанному второму свойству:

$$\mathbf{P}^*\left(\bigcup_i \widetilde{A}_i\right) = \sum_i \mathbf{P}(\widetilde{A}_i)$$

По свойству старого трюка:

$$\sum_i \mathbf{P}(\widetilde{A}_i) \leq \sum_i \mathbf{P}(A_i)$$

В силу ЕМ2:

$$\sum_i \mathbf{P}(A_i) = \sum_i \mathbf{P}^*(A_i)$$

Итого, для  $A_i \in \mathcal{A}$  мы получили:

$$\mathbf{P}^*\left(\bigcup_i A_i\right) \leq \sum_i \mathbf{P}^*(A_i)$$

Перейдем к доказательству общего случая (general case). Особо внимательный слушатель или читатель может заметить, что для этого доказательства необходимо существование упомянутого general case. Недавно математикам стало известно, что General Case существует!

... General Case, a graduate of West Point. The General gave long ... service to the U.S. Army

Смысл числа  $\mathbf{P}^*(D)$  состоит в том, что найдется такое покрытие  $\{D_i\}$ , которое будет накрывать  $D$  с произвольной точностью. Т.е. сумма  $\sum_i \mathbf{P}(D_i)$  будет не меньше чем внешняя мера  $\mathbf{P}^*(D)$ , но при этом сколь угодно близка к ней.

Итак, имеется последовательность произвольных  $A_i$ . Зафиксируем произвольное  $\varepsilon$ .

Подберем последовательность  $A_1^k$  накрывающую  $A_1$  с точностью до  $\frac{\varepsilon}{2}$ , т.е.  $\sum_k \mathbf{P}(A_1^k) \leq \mathbf{P}^*(A_1) + \frac{\varepsilon}{2}$ .

Подберем последовательность  $A_2^k$  накрывающую  $A_2$  с точностью до  $\frac{\varepsilon}{4}$ .

Подберем последовательность  $A_3^k$  накрывающую  $A_3$  с точностью до  $\frac{\varepsilon}{8}$ .

И т.д.

Заметим, что получившаяся последовательность  $\{A_i^k\}$  будет накрывать  $A$ .

По свойству ЕМ1 внешней меры

$$\mathbf{P}^*\left(\bigcup_i A_i\right) \leq \mathbf{P}^*\left(\bigcup_{i,k} A_i^k\right)$$

Т.к. для множеств из  $\mathcal{A}$  свойство ЕМЗ уже доказано, то

$$\mathbf{P}^*(\bigcup_{i,k} A_i^k) \leq \sum_{i,k} \mathbf{P}^*(A_i^k)$$

На алгебре вероятность и внешняя мера совпадают, поэтому

$$\sum_{i,k} \mathbf{P}^*(A_i^k) = \sum_i \sum_k \mathbf{P}(A_i^k)$$

В силу выбора накрытий  $A_i^k$  получаем переход:

$$\sum_i \sum_k \mathbf{P}(A_i^k) \leq \sum_i (\mathbf{P}^*(A_i) + \frac{\varepsilon}{2^i})$$

Вспомнив девятый класс и формулу суммы геометрической прогрессии, получаем:

$$\sum_i (\mathbf{P}^*(A_i) + \frac{\varepsilon}{2^i}) \leq \sum_i \mathbf{P}^*(A_i) + \varepsilon$$

### 19.3 Расстояние. Метрика.

Слова расстояние и метрика - это синонимы.

Если есть произвольный набор объектов, то между ними можно придумать расстояние. Расстояние - это не обязательно то, что можно померить линейкой. Например, расстояние между городами можно мерить стоимостью телефонного звонка.

Мера - это размер подмножества, а метрика - это расстояние между двумя элементами множества.

**Определение 77.1.** Расстояние  $\rho(x, y)$  - любая функция удовлетворяющая трем аксиомам:

$$R1: \rho(x, y) = \rho(y, x) \geq 0$$

$$R2: \rho(x, y) = 0 \Leftrightarrow x = y$$

$$R3: \text{Неравенство треугольника. } \rho(x, y) \leq \rho(x, z) + \rho(z, y)$$

Классический пример метрики - привычное расстояние на плоскости:

$$\rho((x_1, x_2), (y_1, y_2)) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

Можно измерять расстояние другим способом, например:  $\rho((x_1, x_2), (y_1, y_2)) = |x_1 - y_1| + |x_2 - y_2|$

В качестве упражнения можно убедиться, что эти две функции удовлетворяют аксиомам R1-R3.

Мы будем измерять расстояние между множествами:  $\rho(A, B) = \mathbf{P}^*(A \Delta B)$

Тем кто не знал и забыл, что такое симметрическая разность множеств,  $A \Delta B$ , напомним: под  $A \Delta B$  понимается то, чем  $A$  отличается от  $B$ . Формально  $A \Delta B = (A \setminus B) \cup (B \setminus A)$ .

Симметрическая разность обладает свойствами:

$$S1. A \Delta B = B \Delta A = \overline{A \Delta B} \subseteq A \cup B$$

$$S2. A \Delta A = \emptyset$$

$$S3. A \Delta \emptyset = A$$

$$S4. (A \Delta B) \Delta C = A \Delta (B \Delta C)$$

$$S5. (\bigcup A_i) \Delta (\bigcup B_i) \subseteq \bigcup (A_i \Delta B_i)$$

Доказательство свойств S1-S4 - нетрудное упражнение.

Доказательство S5:

Пусть некий  $x \in (\bigcup A_i) \Delta (\bigcup B_i)$ .

Если  $x \in (\bigcup A_i)$ , то  $x \notin (\bigcup B_i)$ , а значит найдется такой номер  $i$ , что  $x \in A_i$  и  $x \notin B_i$ . Следовательно,  $x \in A_i \Delta B_i$ , а значит такой  $x$  принадлежит правой части.

Случай  $x \in (\bigcup B_i)$  рассматривается точно так же.

Полезно понимать, что в множество  $A \Delta B \Delta C$  входят только те элементы, которые встречаются нечетное число раз в упомянутых  $A$ ,  $B$  и  $C$ .

В силу свойств симметрической разности, придуманное нами расстояние между множествами обладает свойствами:

$$D1. \rho(A, B) = \rho(B, A) = \rho(\overline{A}, \overline{B}) \geq 0$$

$$D2. \rho(A, A) = 0$$

Тут можно заметить, что введенное нами  $\rho(A, B)$  не совсем расстояние, потому что из  $\rho(A, B) = 0$  не следует, что  $A = B$ . Ну и ладно.

$$D3. \rho(A, B) \leq \rho(A, C) + \rho(C, B)$$

$$D4. \rho(\bigcup A_i, \bigcup B_i) \leq \sum \rho(A_i, B_i)$$

$$D5. |\mathbf{P}^*(A) - \mathbf{P}^*(B)| \leq \rho(A, B)$$

Доказательство свойств D1-D4 - самостоятельное упражнение.

Доказательство D5:

$$A \subseteq B \bigcup (A \Delta B)$$

$$\mathbf{P}^*(A) \leq \mathbf{P}^*(B \bigcup (A \Delta B)) \leq \mathbf{P}^*(B) + \mathbf{P}^*(A \Delta B)$$

$$\mathbf{P}^*(A) - \mathbf{P}^*(B) \leq \mathbf{P}^*(A \Delta B) = \rho(A, B)$$

По аналогии можно получить, что

$$\mathbf{P}^*(B) - \mathbf{P}^*(A) \leq \rho(A, B)$$

На практике D5 означает, что из  $\rho(A_i, A) \rightarrow 0$  следует, что  $\mathbf{P}^*(A_i) \rightarrow \mathbf{P}^*(A)$ .

**Определение 78.1.** Назовем множество  $A$  аппроксимируемым («хорошим»), если найдется последовательность  $A_i \in \mathcal{A}$ , такая что  $\rho(A, A_i) \rightarrow 0$ .

Другими словами множество называется хорошим, если найдется сколь угодно похожее на него множество в исходной алгебре  $\mathcal{A}$ .

**Теорема 78.2.** Аппроксимируемые множества образуют  $\sigma$ -алгебру.

Доказательство.

Обозначим набор всех аппроксимируемых множеств буквой  $\mathcal{U}$ .

Докажем, что  $\mathcal{U}$  удовлетворяет аксиомам  $\sigma$ -алгебры:

$$A1. \Omega \in \mathcal{U}$$

$$A2. A \in \mathcal{U} \Rightarrow \overline{A} \in \mathcal{U}$$

$$A3. A_i \in \mathcal{U} \Rightarrow \bigcup A_i \in \mathcal{U}$$

Доказательство A1 и A2 не вызывает особых трудностей и оставлено в качестве упражнения.

Доказательство A3:

Для начала докажем более слабое утверждение:  $A_i \in \mathcal{A} \Rightarrow \bigcup A_i \in \mathcal{U}$

Старый трюк снова подвернулся под руку: от возможно пересекающихся  $A_i$  перейдем к непересекающимся  $\widetilde{A}_i$ .

Итак,  $A = \bigcup \widetilde{A}_i$ .

$$\rho(A, \bigcup_{i=1}^k \widetilde{A}_i) = \mathbf{P}^*(\bigcup_{i=k+1}^{+\infty} \widetilde{A}_i) = \sum_{i=k+1}^{+\infty} \mathbb{P}(\widetilde{A}_i)$$

В силу того, что ряд  $\sum_{i=1}^{+\infty} \mathbb{P}(\widetilde{A}_i) = \mathbf{P}^*(A) \leq 1$  сходится предел остаточных сумм  $\lim_{k \rightarrow +\infty} \sum_{i=k+1}^{+\infty} \mathbb{P}(\widetilde{A}_i) = 0$ .

Таким образом, мы доказали, что  $A$  можно сколь угодно приблизить множествами из  $\mathcal{U}$ .

Перейдем к более общему случаю  $A_i \in \mathcal{U}$ .

Выберем произвольное  $\varepsilon > 0$ . Т.к.  $A_1$  "хорошее" множество, то найдется  $C_1 \in \mathcal{A}$  такое, что  $\rho(A_1, C_1) < \frac{\varepsilon}{2^1}$ . Аналогично, каждое  $A_i$  мы заменим на множество  $C_i$  из исходной алгебры  $\mathcal{A}$  так чтобы  $\rho(A_i, C_i) < \frac{\varepsilon}{2^i}$ .

В силу свойства расстояния D4 получаем:  $\rho(\bigcup A_i, \bigcup C_i) \leq \sum \rho(A_i, C_i) < \sum \frac{\varepsilon}{2^i} = \varepsilon$ .

Другими словами, выбрав подходящее  $\varepsilon$  можно добиться того, что расстояние между  $\bigcup A_i$  и  $\bigcup C_i$  будет сколь угодно мало.

Но само множество  $\bigcup C_i$  является аппроксимируемым, т.к. все  $C_i \in \mathcal{A}$ . Следовательно, найдется  $C \in \mathcal{A}$  такое, что  $\rho(\bigcup C_i, C) < \varepsilon$ .

Мы получили, что  $\bigcup A_i$  похоже на  $\bigcup C_i$ , а  $\bigcup C_i$  похоже на  $C$ . Используя неравенство треугольника,  $\rho(\bigcup A_i, C) \leq \rho(\bigcup A_i, \bigcup C_i) + \rho(\bigcup C_i, C) < 2\varepsilon$ .

Т.е. множество  $\bigcup A_i$  можно сколь угодно точно приблизить с помощью множества  $C$ . Следовательно,  $\bigcup A_i \in \mathcal{U}$ .

**Теорема 79.1.** На  $\sigma$ -алгебре  $\mathcal{U}$  внешняя мера  $\mathbf{P}^*$  обладает всеми свойствами вероятности.

Требуется доказать три свойства вероятности:

P1.  $\forall A \in \mathcal{U} \Rightarrow \mathbf{P}^*(A) \geq 0$

P2. Если  $A_i$  попарно непересекаются, то  $\mathbf{P}^*(\bigcup A_i) = \sum \mathbf{P}^*(A_i)$

P3.  $\mathbf{P}^*(\Omega) = 1$

Доказательство свойств P1 и P3 - упражнение.

Докажем свойство P2:

Пусть  $A$  и  $B$  лежат в  $\mathcal{U}$  и  $A \cap B = \emptyset$ .

Рассмотрим последовательности множеств  $A_i$  и  $B_i$ , приближающиеся к множествам  $A$  и  $B$ .

$$\mathbf{P}^*(A_i \bigcup B_i) = \mathbf{P}(A_i \bigcup B_i) = \mathbf{P}(A_i) + \mathbf{P}(B_i) - \mathbf{P}(A_i \cap B_i) = \mathbf{P}^*(A_i) + \mathbf{P}^*(B_i) - \mathbf{P}^*(A_i \cap B_i)$$

Или  $\mathbf{P}^*(A_i) + \mathbf{P}^*(B_i) - \mathbf{P}^*(A_i \bigcup B_i) = \mathbf{P}^*(A_i \cap B_i)$

Возьмем предел  $i \rightarrow +\infty$  от обеих частей.

В силу выбора  $A_i$  и  $B_i$  получаем  $\mathbf{P}^*(A_i) \rightarrow \mathbf{P}^*(A)$  и  $\mathbf{P}^*(B_i) \rightarrow \mathbf{P}^*(B)$ .

В силу D5:  $|\mathbf{P}^*(A \bigcup B) - \mathbf{P}^*(A_i \bigcup B_i)| \leq \rho(A \bigcup B, A_i \bigcup B_i)$

В силу D4:  $\rho(A \bigcup B, A_i \bigcup B_i) \leq \rho(A_i, A) + \rho(B_i, B)$ .

Так как оба последних расстояния стремятся к 0, то и  $\mathbf{P}^*(A_i \bigcup B_i) \rightarrow \mathbf{P}^*(A \bigcup B)$ .

Осталось доказать, что  $\mathbf{P}^*(A_i \cap B_i) \rightarrow 0$ :

Заметим, что  $A_i \cap B_i \subseteq (A \Delta A_i) \cup (B \Delta B_i)$ . Пусть  $x$  лежит в левой части. Значит  $x \in A_i$  и  $x \in B_i$ . Раз  $A$  и  $B$  не пересекаются, то  $x$  не попадет хотя бы в одно из этих множеств, но тогда он попадет в соответствующую симметрическую разность.

По свойству ЕМЗ,  $\mathbf{P}^*(A_i \cap B_i) \leq \mathbf{P}^*(A_i \Delta A) + \mathbf{P}^*(B_i \Delta B)$ . Здесь видно, что правая часть стремится к нулю.

Итак, мы доказали, что  $\mathbf{P}^*(A) + \mathbf{P}^*(B) = \mathbf{P}^*(A \bigcup B)$  для непересекающихся  $A$  и  $B$ . Из этого автоматически следует аддитивность внешней меры для любых конечных объединений. Если  $A_i$  попарно не пересекаются, то  $\mathbf{P}^*(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbf{P}^*(A_i)$ .

Рассмотрим теперь счетное объединение.

$$\mathbf{P}^*(\bigcup_{i=1}^{+\infty} A_i) \geq \mathbf{P}^*(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbf{P}^*(A_i).$$

Взяв предел  $n \rightarrow +\infty$  слева и справа, получаем  $\mathbf{P}^*(\bigcup_{i=1}^{+\infty} A_i) \geq \sum_{i=1}^{+\infty} \mathbf{P}^*(A_i)$ .

Обратное неравенство  $\mathbf{P}^*(\bigcup_{i=1}^{+\infty} A_i) \leq \sum_{i=1}^{+\infty} \mathbf{P}^*(A_i)$  - это доказанное свойство ЕМЗ.

**Теорема 79.2.** Не существует вероятности, отличной от  $\mathbf{P}^*$  на наборе  $\mathcal{U}$  и являющейся продолжением вероятности  $\mathbf{P}$

Доказательство.

Допустим, что существует  $\mathbf{P}' \neq \mathbf{P}^*$ .

Во-первых, возможно, что  $\exists A \in \mathcal{U}$ , такое что  $\mathbf{P}'(A) > \mathbf{P}^*(A)$  или  $\mathbf{P}'(A) - \mathbf{P}^*(A) = d > 0$ .

Докажем, что таких  $A$  не существует. По определению  $\mathbf{P}^*(A)$  существует такое покрытие множества  $A$  с помощью набора непересекающихся множеств  $B_i$ , что  $A \subseteq \bigcup B_i$  и  $\mathbf{P}^*(A) < \sum \mathbf{P}(B_i) + \frac{d}{2}$ . Такое покрытие существует, так как  $\mathbf{P}^*$  определялась как инфимум покрытий.

Итак,  $\mathbf{P}^*(A) \leq \sum \mathbf{P}(B_i) < \mathbf{P}^*(A) + \frac{d}{2}$ .

$$\mathbf{P}'(A) = \mathbf{P}^*(A) + d > \sum \mathbf{P}(B_i) + \frac{d}{2} = \sum \mathbf{P}'(B_i) + \frac{d}{2} = \mathbf{P}'(\bigcup B_i) + \frac{d}{2}$$

В результате мы получили, что  $\mathbf{P}'(A) > \mathbf{P}'(\bigcup B_i) + \frac{d}{2}$ , что невозможно в силу того, что  $\mathbf{P}'$  - это вероятность, а  $A \subseteq \bigcup B_i$ .

Во-вторых, возможно, что  $\exists A \in \mathcal{U}$ , такое что  $\mathbf{P}'(A) < \mathbf{P}^*(A)$ .

Рассмотрим произвольное  $A$ . Используя только что опровергнутое «во-первых» опровергаем «во-вторых»:  $\mathbf{P}^*(\bar{A}) \geq \mathbf{P}'(\bar{A}) \Rightarrow 1 - \mathbf{P}^*(A) \geq 1 - \mathbf{P}'(A) \Rightarrow \mathbf{P}^*(A) \leq \mathbf{P}'(A)$

## 19.4 Еще задачи

## 20 Неразобрано!

Глава 2. Сигма-алгебры, измеримость.

Знаком  $\sigma(\mathcal{E})$  мы будем обозначать минимальную  $\sigma$ -алгебру, порожденную множеством  $\mathcal{E}$

**Определение 80.1.** Пусть  $\Omega = \mathbb{R}$ , и  $\mathcal{E} = \{\text{Открытые множества}\}$ . Борелевской  $\sigma$ -алгеброй  $\mathcal{B}$  называется  $\mathcal{B} = \sigma(\mathcal{E})$

Элементы этой  $\sigma$ -алгебры называются борелевскими множествами.

**Теорема 80.2.** Пусть  $\mathcal{A} = \{\text{Множества вида } (-\infty; t)\}$ , и  $\mathcal{E} = \{\text{Открытые множества}\}$ . Тогда  $\sigma(\mathcal{A}) = \sigma(\mathcal{E})$ .

Из  $\mathcal{A} \subseteq \mathcal{E}$  следует, что  $\sigma(\mathcal{A}) \subseteq \sigma(\mathcal{E})$

Выберем произвольное открытое  $A$ . Доказательство того, что  $A \in \sigma(\mathcal{A})$  разобьем на три этапа:

Во-первых, множества вида  $[a; b) \in \sigma(\mathcal{A})$  в силу того, что  $[a; b) = (-\infty; b) \cap (\Omega \setminus (-\infty; a))$

Во-вторых, построим последовательность множеств  $A_i$ .

$A_1$  будет состоять из полуинтервалов длины 1 вида  $[n; n+1)$ , целиком содержащихся в  $A$ .

$A_2$  будет состоять из полуинтервалов длины  $1/2$  вида  $[\frac{n}{2}; \frac{n+1}{2})$ , целиком содержащихся в  $A$ .

Каждое  $A_i$  лежит в  $\sigma(\mathcal{A})$

В-третьих,  $A = \bigcup_{i=1}^{+\infty} A_i$

**Теорема 80.3.** Если  $\sigma(\mathcal{E}) = \mathcal{B}$ , то для того, чтобы  $f$  была измеримой, необходимо и достаточно, чтобы  $\forall A \in \mathcal{E} \Rightarrow f^{-1}(A) \in \mathcal{F}$

**Теорема 80.4.** Если  $X$  - случайная величина, а  $f$  - борелевская функция, то  $f \circ X() = f(X(\cdot))$  - случайная величина.

Пример. Если  $X$  - случайная величина, то и  $X^2$  - случайная величина.

Корректно говорить, что функций  $f$  является  $\mathcal{F}$ -измеримой. Указывать о какой  $\sigma$ -алгебре идет речь необходимо потому, что одна и та же функция может быть измеримой относительно одной  $\sigma$ -алгебры и быть неизмеримой относительно более бедной  $\sigma$ -алгебры. Однако, если из контекста понятно, о какой  $\sigma$ -алгебре идет речь, мы будем пропускать явное указание.

Идея replicating portfolio.

Пример (оценить 75 центов в случае  $> 1$  - см. Crack)

Теорема. Формула Блэка-Шоулса в непрерывном времени как предел дискретной формулы.

Независимость сигма-алгебр. Независимость случайных величин. Из независимости  $X$  и  $Y$  следует независимость  $f(X)$  и  $g(Y)$ .

Для более удобной работы используются такие обозначения:

$A_i \uparrow A$  -  $A_i$  "стремится снизу" к  $A$ . Это означает, что  $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$  и  $\bigcup A_i = A$ .

$A_i \downarrow A$  -  $A_i$  "стремится сверху" к  $A$ . Это означает, что  $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$  и  $\bigcap A_i = A$ .

Исходя из свойств P1-P3 можно вывести, что у вероятности также имеется ряд других удобных свойств.

Если  $A \in \mathcal{F}$ ,  $B \in \mathcal{F}$ ,  $\forall A_i \in \mathcal{F}$ , то выполнены, например, следующие свойства:

1.  $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$



2. Если  $B \subseteq A$ , то  $\mathbf{P}(B) \leq \mathbf{P}(A)$
3.  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$
4.  $\mathbf{P}(\bigcup_{i=1}^{+\infty} A_i) = \lim_{n \rightarrow +\infty} \mathbf{P}(\bigcup_{i=1}^n A_i)$
5.  $A_i \uparrow A \Rightarrow \mathbf{P}(A) = \lim_{i \rightarrow +\infty} \mathbf{P}(A_i)$
6.  $A_i \downarrow A \Rightarrow \mathbf{P}(A) = \lim_{i \rightarrow +\infty} \mathbf{P}(A_i)$

**Определение 81.1.** Измеримое пространство - это набор  $(\Omega, \mathcal{F})$

**Определение 81.2.** Вероятностное пространство - это набор  $(\Omega, \mathcal{F}, \mathbf{P})$

Пополнение.

Пусть  $(\Omega, \mathcal{F}, \mathbf{P})$  - вероятностное пространство, и  $\mathbf{P}(A) = 0$ . Иногда может найтись такое множество  $C$ , что  $C$  неизмеримо, но  $C \subseteq A$ . Таких ситуаций хотелось бы избежать, для этого существует процедура пополнения (completion) вероятностного пространства.

Для начала определим набор множеств  $\mathcal{N} = \{C \notin \mathcal{F} | \exists A \supseteq C, \mathbf{P}(A) = 0\}$ .

$\mathcal{N}$  не является алгеброй. Например,  $\Omega \notin \mathcal{N}$ .

Определим набор множеств  $\overline{\mathcal{F}} = \{F \cup N | F \in \mathcal{F}, N \in \mathcal{N}\}$ .

Для  $\overline{F} = F \cup N$  определим  $\overline{\mathbf{P}}(\overline{A}) = \mathbf{P}(F)$ .

Докажем, что  $\overline{\mathcal{F}}$  -  $\sigma$ -алгебра:

Во-первых,  $\emptyset \in \overline{\mathcal{F}}$  в силу того, что  $\emptyset = \emptyset \cup \emptyset$ , где  $\emptyset \in \mathcal{F}$  и  $\emptyset \in \mathcal{N}$ .

Во-вторых,

В-третьих,

$$\bigcup_i (F_i \cup N_i) = \left( \bigcup_i F_i \right) \cup \left( \bigcup_i N_i \right)$$

При этом  $\bigcup_i F_i$  измеримо и  $\bigcup_i N_i \in \mathcal{N}$ .

Докажем, что  $\overline{\mathbf{P}}$  - вероятность.

Сначала установим, что определение внутренне непротиворечиво.

Допустим в  $\overline{\mathcal{F}}$  появится некоторое множество допускающее два разложения  $C = F_1 \cup N_1 = F_2 \cup N_2$ , где  $N_1, N_2 \in \mathcal{N}$ , а  $F_1, F_2 \in \mathcal{F}$ , причем  $\mathbf{P}(F_1) \neq \mathbf{P}(F_2)$ .

Заметим, что  $F_1 \subset F_2 \cup N_1 \cup N_2$ , следовательно,  $F_1 \setminus F_2 \subset (F_2 \cup N_1 \cup N_2) \setminus F_2$ .

Но  $(F_2 \cup N_1 \cup N_2) \setminus F_2 \in \mathcal{N}$ , следовательно,  $\mathbf{P}(F_1 \setminus F_2) = 0$ .

Далее воспользовавшись тем, что  $F_1 = (F_1 \cap F_2) \cup (F_1 \setminus F_2)$  получаем, что  $\mathbf{P}(F_1) = \mathbf{P}(F_1 \cap F_2)$ .

По аналогичному рассуждению получаем, что  $\mathbf{P}(F_2) = \mathbf{P}(F_1 \cap F_2)$ , что доказывает непротиворечивость определения новой вероятности.

Остается доказать, что  $\overline{\mathbf{P}}$  является  $\sigma$ -аддитивной.  $\overline{\mathbf{P}}(\bigcup_i (F_i \cup N_i)) = \overline{\mathbf{P}}((\bigcup_i F_i) \cup (\bigcup_i N_i)) = \mathbf{P}(\bigcup_i F_i)$

Маленькая сенсация!

Оказывается, «почти» - это строгий математический термин.

Говорят, что некоторое свойство выполнено почти наверное (almost surely, a.s.), если существует событие с вероятностью единица, на котором свойство выполнено.

Например,  $X = Y$  a.s. означает, что  $\mathbf{P}(X \neq Y) = 0$ .

Понятие «почти наверное» играет важную роль в силу того, что случайные величины равные почти наверное не отличимы по большинству характеристик. Многие интуитивные выводы становятся абсолютно точными, если к ним добавлять «почти наверное». Например, (скорее всего, это будет упражнением через лекцию или две) из того, что  $Var(X) = 0$  следует, что  $X = const$  почти наверное.

Итак **Каратеодори** - человек и теорема!

Пусть задан набор  $(\Omega, \mathcal{A}, \mathbf{P})$ , где  $\mathcal{A}$  - алгебра, а  $\mathbf{P}$  - вероятность, заданная на алгебре.

Вспомним, что мы определяли  $\mathbf{P}$  на  $\sigma$ -алгебре. Там требовалось выполнение свойств:

P1.  $\mathbf{P}(A) \geq 0$

P2. Если  $A_i$  попарно не пересекаются, то  $\mathbf{P}(\bigcup A_i) = \sum \mathbf{P}(A_i)$

P3.  $\mathbf{P}(\Omega) = 1$

Т.к.  $\mathcal{A}$  - всего лишь алгебра, то в пункте P2 нужно добавить условие  $\bigcup_i A_i \in \mathcal{A}$ . В  $\sigma$ -алгебре оно было бы выполнено само собой.

P2': Если  $A_i$  попарно не пересекаются и  $\bigcup_i A_i \in \mathcal{A}$ , то  $\mathbf{P}(\bigcup A_i) = \sum \mathbf{P}(A_i)$

Глава 6. Интеграл Лебега.

«Четыре шага»

Для индикатора.

Для простой.

Для неотрицательной.

Для любой действительной.

FL. Fatou's Lemma.

Если  $X_n \geq 0$ , то

$$E\left(\liminf_{n \rightarrow +\infty} X_n\right) \leq \liminf_{n \rightarrow +\infty} \mathbb{E}(X_n)$$

MCT. Monotone Convergence Theorem.

Если  $X_{n+1} \geq X_n \geq 0$ , то

$$\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = E\left(\lim_{n \rightarrow +\infty} X_n\right)$$

DCT. Dominated Convergence Theorem.

Если  $X_n \rightarrow X$  п.н. и  $|X_n| < Y$ , где  $\mathbb{E}(Y) < +\infty$ , то  $\mathbb{E}(|X|) < +\infty$  и

$$\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$$

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|) = 0$$

В MCT и FL пределы могут принимать значение  $+\infty$ .

BCT. Bounded Convergence Theorem.

Если  $X_n \rightarrow X$  (in P) и  $|X_n| < M$ , то  $\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$ .

Примеры:

Мат. ожидание  $= \int dP$

Площадь под кривой  $= \int d\lambda$

Теорема:

Интеграл Римана = интеграл Лебега

Теорема:

Пусть с.в.  $X$  индуцирует меру  $Pr$  на прямой. Тогда  $\int f(X)dP = \int f dPr$

Док-во: «Четыре шага»

Определение. Функция плотности.

Теорема:

Если у с.в.  $X$  есть функция плотности, то  $\int f dPr = \int f p d\lambda$

Док-во: «Четыре шага»

Задачи:

доказать MCT, DCT для почти наверное

доказать FL для вероятности

Теорема: Если  $X$  и  $Y$  независимы, то  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

Док-во: «Четыре шага»

Неравенства: Jensen, Cauchy-Schwarz, Holder.

Jensen - да, остальные - ?

Теорема Фубини (4 шага, возможно шаг 1 неполностью)

Взятие производной под интегралом.

Глава 8. Условное ожидание.

Условное математическое ожидание. Мыслим как о проекции. Формальное определение.  
Свойства

Теорема Пифагора!

Теорема Радона-Никодима (без док-ва?).

Свойства в куче:

1. Единственность почти наверное. Если  $Y_1 = \mathbb{E}(X|\mathcal{H})$  и  $Y_2 = \mathbb{E}(X|\mathcal{H})$  то  $Y_1 = Y_2$  ае
- Пусть  $X, X_n$  и  $Y$  - интегрируемые (из  $L^1$ ) случайные величины.
2.  $\mathbb{E}(\mathbb{E}(X|\mathcal{H})) = \mathbb{E}(X)$
3.  $\mathcal{H} = \{\emptyset, \Omega\} \Rightarrow \mathbb{E}(X|\mathcal{H}) = \mathbb{E}(X)$
4.  $X \in \mathcal{H} \Rightarrow \mathbb{E}(X|\mathcal{H}) = X$  (ас)
5.  $\mathbb{E}(aX + bY|\mathcal{H}) = a\mathbb{E}(X|\mathcal{H}) + b\mathbb{E}(Y|\mathcal{H})$
6.  $X \geq 0$  (ас)  $\Rightarrow \mathbb{E}(X|\mathcal{H}) \geq 0$  (ас)
7.  $X \geq Y$  (ас)  $\Rightarrow \mathbb{E}(X|\mathcal{H}) \geq \mathbb{E}(Y|\mathcal{H})$  (ас)
8.  $|\mathbb{E}(X|\mathcal{H})| \leq \mathbb{E}(|X||\mathcal{H})$  (ас)
9.  $X \in \mathcal{H}, XY \in L_1 \Rightarrow \mathbb{E}(XY|\mathcal{H}) = X\mathbb{E}(Y|\mathcal{H})$  (ас)
10.  $X_n \uparrow X \Rightarrow \mathbb{E}(X_n|\mathcal{H}) \rightarrow \mathbb{E}(X|\mathcal{H})$  (ас) (МСТ)
11. Если  $\mathcal{H} \subseteq \mathcal{F}$ , то  $\mathbb{E}(X|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{F})|\mathcal{H})$  (ас) (Tower property, теорема о трех перпендикулярах)
12. Неравенство Йенсена
13. Если  $X$  и  $\mathcal{H}$  независимы, то  $\mathbb{E}(X|\mathcal{H}) = \mathbb{E}(X)$  ае
14. Геометрический смысл - проекция
- $\mathbb{E}((X - Y)^2)$  достигает своего минимума для  $Y \in \mathcal{H}$  при  $Y = \mathbb{E}(X|\mathcal{H})$
15. Т. Пифагора  $\mathbb{E}(X^2) = \mathbb{E}(\mathbb{E}(X|\mathcal{H})^2) + \mathbb{E}((X - \mathbb{E}(X|\mathcal{H}))^2)$

Док-во 9.

1. Пусть  $X$  - неотрицательная.  $Y$  - индикатор,  $Y$  - простая,  $Y$  - неотрицательная.
2.  $X$  и  $Y$  - любые.

Упражнения:

Глава 9. «Практическая».

Метод первого шага. Задача про разборчивую невесту.

Первым шагом:

1. Вероятность достичь уровня А раньше, чем (-В)
2. Ожидаемое время достижения уровня А или уровня (-В)
3. Вероятность рано или поздно выиграть 1 рубль.
4. Ожидаемое время до выигрыша в 1 рубль.
5. Ожидаемое время обнаружения пчел Винни-Пухом (время возвращения пьяницы домой из кабака)
6. Вероятность возвращения когда-либо в точку старта
7. Задача про хрюшек-копилок
8. вероятность для асимметричного блуждания,

Марковские цепи.

Марковская цепь неформально (график состояний со стрелочками)

Что лучше ООР или РОО?

Глава 10. Мартингалы в дискретном времени.

Тема. Определение мартингала и момента остановки. Мартингальное преобразование.  
Upcrossing inequality.

**Определение мартингалов**

Пусть задано вероятностное пространство  $(\Omega, \mathcal{F}, \mathbf{P})$ .

Определение. Фильтрацией называется последовательность вложенных  $\sigma$ -алгебр:

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \dots \subseteq \mathcal{F}$$

$\mathcal{F}_n$  - те события, которые мы отличаем в момент времени  $n$ .

$\mathcal{F}$  - самая полная (подробная)  $\sigma$ -алгебра. Все рассматриваемые величины являются  $\mathcal{F}$ -измеримыми.

$\mathcal{F}_0$  как правило состоит только из  $\emptyset$  и  $\Omega$ .

Определение. Случайный процесс - это набор случайных величин, заданных на одном вероятностном пространстве.

Определение. Случайный процесс  $\{X_t\}$  адаптирован к фильтрации  $\{\mathcal{F}_t\}$  если  $X_t \in \mathcal{F}_t$  (условная запись для фразы " $X_t$  является  $\mathcal{F}_t$ -измеримой случайной величиной").

Определение. Естественная фильтрация (natural filtration).  $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$ .

Определение. Интегрируемый адаптированный случайный процесс  $X$  называется [по отношению к  $\mathcal{F}_t$ ]:

Мартингалом (martingale):  $\mathbb{E}(X_{t+1}|\mathcal{F}_t) = X_t$  п.н.

Субмартингалом:  $\mathbb{E}(X_{t+1}|\mathcal{F}_t) \geq X_t$  п.н.

Супермартингалом:  $\mathbb{E}(X_{t+1}|\mathcal{F}_t) \leq X_t$  п.н.

Мартингал полезно понимать, как денежное благосостояние игрока после  $n$  раундов честной игры.

Субмартингал соответствует игре в пользу игрока.

С точки зрения игрока, в супермартингале нет ничего "суперского":)

Обозначения. Вместо формального  $\{X_n\}_{n=0,1,2,\dots}$  мы будем писать просто  $X_n$  или даже  $X$ .

Упражнение 1.

Докажите, что если  $X$  - мартингал, то

1.  $\mathbb{E}(X_n|\mathcal{F}_m) = X_m$  п.н. для  $m \leq n$

2.  $\mathbb{E}(X_{n+1}) = \mathbb{E}(X_n)$

Докажите аналогичные свойства для суб- и супермартингалов

Упражнение 2.

Докажите, что если процесс  $X$  - супермартингал по отношению к некоей фильтрации  $\{\mathcal{F}_n\}$ , то он также супермартингал по отношению к естественной фильтрации  $\sigma(X_0, X_1, \dots, X_n)$ .

Докажите, что  $X$  - супермартингал  $\Leftrightarrow (-X)$  - субмартингал.

Докажите, что

$$\begin{cases} X\text{- субмартингал} \\ X\text{- супермартингал} \end{cases} \Leftrightarrow X\text{- мартингал}$$

Упражнение 3.

3.1. Докажите, что если  $X$  - мартингал,  $\varphi$  - выпуклая вниз и  $\varphi(X_n)$  - интегрируемая величина то  $\varphi(X)$  - субмартингал.

3.2. Докажите аналогичный результат для случая, когда  $X$  - субмартингал, а  $\varphi$  - выпуклая вниз и неубывающая

Пример 1.

$S_0 = 0$ ,  $\mathbb{E}(X_i) = 0$  /кстати, из этого, в частности, следует, что  $\mathbb{E}(|X_i|) < +\infty$  - интегрируемость/,  $X_i$  - независимы,  $S_n = \sum_{i=1}^n X_i$

$S$  - мартингал по отношению к фильтрации, порождаемой последовательностью  $X_i$ .

Пример 2.

$X_i$  - независимы,  $\mathbb{E}(X_i) = 0$ ,  $Var(X_i) = \sigma^2$ ,  $M_n = S_n^2 - n\sigma^2$

$M_n$  - мартингал по отношению к  $X_i$

Пример 3. Мартингал Леви.

$Y$  - произвольная интегрируемая случайная величина,  $\mathcal{F}_n$  - произвольная фильтрация.

$X_n = \mathbb{E}(Y|\mathcal{F}_n)$  - мартингал.

Пример 4.

Возрастающая последовательность действительных чисел - субмартингал.

Пример 5.

$X_n \geq 0$ ,  $X_i$  - независимы,  $\mathbb{E}(X_n) = 1$ ,  $M_0 = 1$ ,  $M_n = X_1 \cdot X_2 \cdot \dots \cdot X_n$

$M_n$  - мартингал по отношению к  $X_n$

Пример 6.

$Y_i$  - iid,  $\varphi(\lambda) = \mathbb{E}(\exp(\lambda Y_i)) < +\infty$ , тогда  $X_i = \exp(\lambda Y_i) / \varphi(\lambda)$  подходят в условия предыдущего примера, и имеется целый набор мартингалов (для каждого  $\lambda$ ):

$$M_n = \exp\left(\lambda \sum_{i=1}^n Y_i\right) / \varphi(\lambda)^n$$

Среди этого набора мартингалов может быть особо полезен мартингал, получающийся в случае такого  $\lambda_0$ , что  $\varphi(\lambda_0) = 1$ .

Пример 7.

Пуассоновский процесс.

P1. При  $0 = t_0 < t_1 < t_2 < \dots < t_n$  величины  $N(t_k) - N(t_{k-1})$  являются независимыми.

P2. Величина  $N(t) - N(s)$  имеет распределение Пуассона с параметром  $\lambda \cdot (t - s)$

Распределение Пуассона с параметром  $\lambda$ :  $\mathbb{P}(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}$

Пусть  $t_i$  - фиксированные моменты времени и  $X_i = N(t_i) - \lambda \cdot t_i$ .

Тогда  $(X_i, \mathcal{F}_i = \sigma(N(t), t \leq t_i))$  - мартингал.

Упражнение.

Руководствуясь примером 6 постройте мартингал из iid  $Y_i = \begin{cases} 1, p \\ -1, 1-p \end{cases}$

При подборе  $\lambda$  заметьте, что уравнение  $\varphi(\lambda) = 1$  будет иметь два корня!

Упражнение.

Постройте «мультипликативный» мартингал из  $Y_i = \begin{cases} 1, p \\ 0, 1-p \end{cases}$

Упражнение.

Докажите, что максимум двух субмартингалов (по отношению к одной фильтрации) - также субмартингал

Игрок может не просто играть в честную игру, но и делать ставки. Естественно ставка должна определяться только из прошлой информации.

Определение. Процесс  $C$  на  $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbf{P})$  называется предсказуемым (predictable, nonanticipating) если  $C_n \in \mathcal{F}_{n-1}$

$C_n$  - это размер ставки.

Игра со ставками будет формально выглядеть как:

Определение. Мартингальным преобразованием (martingale transform) процесса  $X$  называется процесс:

$$(C \cdot X)_n = \sum_{i=1}^n C_i (X_i - X_{i-1}),$$

$$(C \cdot X)_0 = 0$$

Нетрудно заметить, что  $X_i - X_{i-1}$  - это выигрыш в  $i$ -ой игре, а  $C_i$  - ставка, т.е. множитель для выигрыша.

Мартингальное преобразование - это дискретная версия интеграла.

Теорема. Если  $X$  - супермартингал,  $C_n \geq 0$  - предсказуемый процесс и каждая  $C_i$  ограничена (т.е.  $\forall i \exists a : |C_i| < a$ ), то  $C \cdot X$  - также супермартингал.

Доказательство.

В силу ограниченности всех  $C_i$ , следует интегрируемость  $(C \cdot X)_n$ .

$$\begin{aligned} \mathbb{E}((C \cdot X)_{n+1} | \mathcal{F}_n) &= (C \cdot X)_n + \mathbb{E}(C_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n) = \\ &= (C \cdot X)_n + C_{n+1} \mathbb{E}(X_{n+1} - X_n | \mathcal{F}_n) \leq (C \cdot X)_n \end{aligned}$$

Упражнение.

Убедитесь в том, что эта теорема верна для субмартингалов без изменений, а для мартингалов верна даже без условия  $C_n \geq 0$ .

Мораль. В честной игре бесполезно изменять ставку в зависимости от текущего везения-невезения: это не изменит ожидаемого выигрыша. Нельзя победить систему. "System theorem".

С другой стороны, честную или благоприятную игру нельзя испортить меняя ставки даже если очень стараться :)

Нетрудно догадаться, что и стратегия выхода из игры после крупного выигрыша не повлияет на ожидаемый выигрыш. Выход из игры после крупного выигрыша равносильно нулевой ставке.

Хотя с помощью игровых стратегий невозможно сделать деньги, они пригодятся для доказательства теорем :)

А можно ли сделать деньги с помощью теорем?

Определим остановленный процесс. Остановленный процесс  $X_{N \wedge n}$  совпадает с исходным процессом  $X_n$  до момента  $N$  и равен  $X_N$  после.

Теорема.

Пусть  $N$  - момент остановки и  $X$  - супермартингал. Тогда остановленный процесс  $X_{N \wedge n}$  является супермартингалом.

Доказательство

Определим  $C_n = 1_{N \geq n}$ . Стратегия «ставим по рублю вплоть до остановки». Заметим, что процесс  $C$  является предсказуемым, т.к.  $\{N \geq n\} = \{N \leq n-1\}^c \in \mathcal{F}_{n-1}$ . Из теоремы о мартингальном преобразовании следует, что  $(C \cdot X)_n = X_{N \wedge n} - X_0$  - также является супермартингалом.

Упражнение. Сформулируйте аналогичную теорему для субмартингалов и мартингалов.

Пусть  $X_n$  - субмартингал. Например, можно считать, что это цена какой-нибудь акции. Построим график одной из реализаций  $X_n$ .

Медитируем, смотрим на отсутствующий график.

Отметим два уровня цен  $a < b$ . Вопрос, на который мы сейчас в некотором смысле ответим, звучит так: Сколько раз к моменту времени  $n$  мартингал пересечет диапазон  $(a; b)$  снизу-вверх?

На графике таких пересечений снизу-вверх (upcrossing) ... штуки.

Введем последовательность случайных величин  $N_i$ . Так, чтобы  $N$  с нечетными номерами означали время начала пересечения, а  $N$  с четными номерами - время окончания пересечения.

Формально:

$$N_{2k-1} = \inf\{m > N_{2k-2} | X_m \leq a\}$$

$$N_{2k} = \inf\{m > N_{2k-1} | X_m \geq b\}$$

$$N_0 = -1$$

Заметим, что все  $N_i$  являются моментами остановки (о том, что произошло событие  $N_i = k$ , можно сделать зная  $X_1, \dots, X_k$ ).

Рассмотрим событие {момент времени  $m$  приходится на  $k$ -ое пересечение}

$$\{N_{2k-1} < m \leq N_{2k}\} = \{N_{2k-1} \leq m-1\} \cap \{N_{2k} \leq m-1\}^c \in \mathcal{F}_{m-1}$$

$$\text{Рассмотрим } H_m = \begin{cases} 1, \text{ если } N_{2k-1} < m \leq N_{2k} \text{ для какого-нибудь } k \\ 0, \text{ иначе} \end{cases}$$

$H_m$  - предсказуемый процесс.

$U_n = \sup\{k | N_{2k} \leq n\}$  - число законченных пересечений снизу-вверх к моменту  $n$ .

Теорема. The upcrossing inequality (Doob)

$$(b-a)\mathbb{E}(U_n) \leq \mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+$$

Доказательство.

Вместо процесса  $X_n$  рассмотрим процесс  $Y_n$ , который совпадает с  $X_n$ , когда  $X_n > a$ , и равен  $a$ , если  $X_n \leq a$ .

Заметим, что  $Y_n = a + (X_n - a)^+$ , поэтому  $Y_n$  - субмартингал. Количество пересечений снизу вверх у  $X_n$  и  $Y_n$  одинаково.

Будем делать ставки на  $Y$  согласно игровой стратегии  $H$ . Т.е. будем каждый раз во время подъема от  $a$  к  $b$  делать ставку в 1 рубль.

Каждый полный подъем от  $a$  к  $b$  дает выигрыш  $\geq (b - a)$ , поэтому  $(b - a)U_n \leq (H \cdot Y)_n$ .

В силу того, что мы делаем ставки на процесс  $Y$ , тот факт, что последний подъем может быть неполным, а процесс  $X$  может уйти далеко вниз, нас нисколько не смущает.

Обозначим  $K_n = 1 - H_n$ . В силу теоремы о мартингальном преобразовании,  $(K \cdot Y)$  - субмартингал, поэтому,  $\mathbb{E}(K \cdot Y)_n \geq \mathbb{E}(K \cdot Y)_0 = 0$ .

Заметим, что  $Y_n - Y_0 = (H \cdot Y)_n + (K \cdot Y)_n$  (на сколько раз рублей изменится благосостояние игрока, если каждый раз делать ставку в 1 рубль?).

Итого:

$$\begin{aligned}(b - a)\mathbb{E}(U_n) &\leq \mathbb{E}(H \cdot Y)_n = \mathbb{E}(Y_n - Y_0) - \mathbb{E}(K \cdot Y)_n \leq \mathbb{E}(Y_n) - \mathbb{E}(Y_0) = \\ &= \mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+\end{aligned}$$

В доказательстве использовался тот факт, что самая «глупая» игровая стратегия  $K$  приносит прибыль. :)

Подглава. Нужна ли? Равномерная интегрируемость.

Определение. Набор случайных величин  $\{X_\alpha | \alpha \in A\}$  (необязательно счетный) называется равномерно интегрируемым, если:

$$\lim_{M \rightarrow +\infty} \sup_{\alpha \in A} \mathbb{E}(|X_\alpha| 1_{|X_\alpha| > M}) = 0$$

Пример 1. Конечный набор интегрируемых случайных величин равномерно интегрируем. По DCT.

Пример 2. Если  $|X_\alpha| < Y$  и  $\mathbb{E}(Y) < +\infty$ , то набор  $\{X_\alpha\}$  равномерно интегрируем.

Пример 3. Если набор случайных величин ограничен в смысле  $L_p$ ,  $p > 1$ , то он равномерно интегрируем.

Ограничен в смысле  $L_p$ :  $\sup_{\alpha} \mathbb{E}(|X_\alpha|^p) < +\infty$ .

Упражнение 1.

Равномерно интегрируемый набор ограничен в  $L_1$ .

Упражнение 2.

Сходящаяся в  $L_1$  последовательность является равномерно интегрируемой.

Упражнение 3.

Пусть  $\{X_\alpha\}$  и  $\{Y_\beta\}$  равномерно интегрируемы. Тогда равномерно интегрируемо их объединение.

Определение. Случайная величина  $X$  называется равномерно интегрируемой, если для любой последовательности событий  $A_i$ , такой что  $\mathbb{P}(A_i) \rightarrow 0$ , выполнено условие  $\mathbb{E}(X \cdot 1_{A_i}) \rightarrow 0$ .

Утверждение. Из равномерной интегрируемости набора  $\{X_\alpha\}$  следует:

$$\text{UI1. } \sup_{\alpha} \mathbb{E}(|X_\alpha|) < +\infty$$

$$\text{UI2. } \lim_{M \rightarrow +\infty} \sup_{\alpha} \mathbb{P}(|X_\alpha| > M) = 0$$

$$\text{UI3. Для любой } A_i \text{ такой что } \mathbb{P}(A_i) \rightarrow 0 \text{ следует } \lim_{i \rightarrow +\infty} \sup_{\alpha} \mathbb{E}(|X_\alpha| \cdot 1_{A_i}) = 0$$

В частности из UI3 следует равномерная интегрируемость каждой  $X_\alpha$

Утверждение.

Если выполнены свойства UI1 и UI3, то набор  $\{X_\alpha\}$  равномерно интегрируем.

Если выполнены свойства UI2 и UI3, то набор  $\{X_\alpha\}$  равномерно интегрируем.

Теорема.

Если  $X_n \rightarrow X$  по вероятности, то утверждения UE1, UE2 и UE3 эквивалентны

Утверждение.

Если  $X \in L_1$ , то набор случайных величин  $\{\mathbb{E}(X|\mathcal{Y})|\mathcal{Y} \subseteq \mathcal{F}\}$  является равномерно интегрируемым.

Если не ждать «слишком» долго, то ожидаемый выигрыш будет равен начальной сумме,  $\mathbb{E}(X_T) = \mathbb{E}(X_1)$ . Вот пример «слишком» долгого ожидания: ждать до выигрыша в один рубль в классическом случайном блуждании. Если  $T$  - первый момент достижения суммы в один рубль, то  $\mathbb{E}(X_T) = X_T = 1 \neq X_0 = 0$  по построению, но ждать счастливого момента действительно придется долго:  $\mathbb{E}(T) = +\infty$ .

Точный смысл понятия «слишком» долго можно увидеть в теореме:

Если  $X_t$  - мартингал,  $T$  - момент остановки и выполнено хотя бы одно из пяти условий:

(i) Момент  $T$  ограничен, то есть существует число  $M$ , такое что  $T < M$ .

(ii)  $\mathbb{P}(T < +\infty) = 1$  и процесс  $X_{t \wedge T}$  ограничен, то есть существует число  $M$ , такое что для любого  $t$  верно неравенство  $|X_{t \wedge T}| < M$ .

(iii)  $\mathbb{E}(T) < +\infty$  и существует число  $M$ , такое что для любого  $t$  верно неравенство  $\mathbb{E}(|X_{t+1} - X_t| | \mathcal{F}_t) < M$ .

(iv)  $\mathbb{P}(T < +\infty) = 1$ ,  $\mathbb{E}(|X_T|) < \infty$  и  $\lim_{t \rightarrow \infty} \mathbb{E}(X_t 1_{T > t}) = 0$ .

(v)  $\mathbb{P}(T < +\infty) = 1$  и мартингал  $X_t$  является равномерно интегрируемым.

В большинстве случаев первых трех критериев достаточно для практического применения. Четвертый критерий является следствием любого из первых трех (?).

В пятом критерии используется определение равномерной интегрируемости. . . Набор случайных величин является равномерно интегрируемым, если. . .

Из этих критериев все кроме третьего (может быть какой-то «предельный» аналог и третьего? - я не знаю) работают в непрерывном времени.

Примеры:

Тождество Вольда. Wald's identity.

Симметричное случайное блуждание,  $T = \min\{t | S_t = a \cup S_t = -b\}$  - подсчет  $\mathbb{P}(S_T = a)$  (через  $S_t$ ) и  $\mathbb{E}(T)$  (через  $S_t^2 - t$ )

Несимметричное случайное блуждание - подсчет  $\mathbb{P}(S_T = a)$  (через  $(p/q)^{S_t}$ )

ABRACADABRA -  $\mathbb{E}(T)$ ,  $Var(T)$  - см. Ross, Second course

Следующая карта - красная (через долю красных карт)

Раунды при подбрасывании шляп (через  $\sum X_i - n$ )

Ключи и сейфы

Ballot problem

Second heart problem

Задачи:

$\mathbb{E}(TS_T)$  для классического случайного блуждания - ?

Задача.

Пусть  $Y_k$  - количество посещений точки 0 симметричным случайным блужданием до посещения точки  $k$ . Заметим, что случайное блуждание может никогда не посетить точку  $k$ , но это событие происходит с вероятностью 0, поэтому величину  $Y_k$  можно считать корректно определенной. Верно ли, что  $Y_k - 2k$  - мартингал?

Решение:

## 21 Решения

**1.1** Нет! От противного, допустим  $S$  существует. Для каждого  $s \in S$  существует множество  $A_s$  соответствующей мощности. Построим множество  $A = \cup_s A_s$ . Здесь используется аксиома выбора. И теперь построим множество  $B = 2^A$ . Оно больше любого из  $A_s$ ! Значит оно не было упомянуто в списке  $S$ .

### 1.2



**1.3** Да. Считаем «змейкой»

**1.4** Нет. Бесконечные вправо последовательности из 0 и 1 - это декартово произведение счетного количества  $A_i$ , где каждое  $A_i = \{0, 1\}$

**1.5** Объединение счетного количества счетных множеств, считаем змейкой

**1.6** Мощность  $2^A$  не меньше мощности  $A$ , т.к. есть одноточечные подмножества, которые можно сопоставить с элементами  $A$ . Допустим все же, что  $2^A$  и  $A$  равномощны. Значит есть взаимно однозначное соответствие  $b \longleftrightarrow B$ , где  $b \in A$  и  $B \in 2^A$ . Построим  $C \subset A$  по принципу: будем включать туда только такие  $b$ , которые не входят в соответствующее  $B$ . Этому множеству  $C$  должен соответствовать некий элемент  $c$ . С одной стороны  $c$  не может входить в  $C$ , с другой стороны - обязан. Противоречие.

**1.7** да

**2.1** да, так как все эти множества представимы в виде счетного объединения точек, отрезков или интервалов.

**2.2**  $\lim A_i = \emptyset$

**2.3**  $N < +\infty$

**2.4**  $\Omega = \{1, 2, 3, 4, 5, 6\} \dots$

**4.1** а) да, б) да

**4.2** да, если  $\mu(\Omega) > 0$

**4.3** Пусть  $\Omega = [0; 1] \cap \mathbb{Q}$ ,  $\mathcal{F}$  — это все конечные множества и множества с конечными дополнениями. Функция  $R$  равна 0 для конечных множеств и единице для множеств с конечным дополнением. Занумеруем все рациональные числа на  $[0; 1]$ :  $r_i$  —  $i$ -ое по счету рациональное число. Возьмем  $A_n = \{r_1, r_2, \dots, r_n\}$ . Тогда  $R(A_n) = 0$  т.к.  $A_n$  конечно, но  $\cup A_n = \Omega$  и  $R(\Omega) = 1$ .

**4.4** да

**5.1** Гарантированно спасаем всех кроме конечного количества! Решение основано на неизмеримых множествах. В парадоксе Банаха-тарского - разрежали и меняем после перестановки объем. Здесь разрежали и мат. ожидание доли перестало равняться вероятности

**5.2** да, лежит; нет, не конец интервала

**9.1**  $X$  допускает разложение  $X = \sum x_i 1_{A_i}$ . В силу линейности ковариаций достаточно проверить равенство для индикаторов. Проверяем:  $Cov(1_{A_i}, \hat{Y}) = \mathbb{E}(\hat{Y} 1_{A_i}) - \mathbb{E}(\hat{Y})\mathbb{E}(1_{A_i}) = \mathbb{E}(Y 1_{A_i}) - \mathbb{E}(Y)\mathbb{E}(1_{A_i}) = Cov(1_{A_i}, Y)$ .

**9.2** Например, пусть  $Y_1$  - распределено геометрически и  $Z$ , независимо от  $Y_1$ , принимает равновероятно значения  $+1$  или  $-1$ . Определим  $Y_2 := ZY_1$  и  $\mathcal{H} := \sigma(Y_1)$ . Получаем, что  $Y_1 = |Y_2|$ , но  $\mathbb{E}(Y_2|\mathcal{H}) = 0$ , поэтому  $\mathbb{E}(Y_1|\mathcal{H})$  и  $\mathbb{E}(Y_2|\mathcal{H})$  независимы.

**13.1** а)  $Y = \rho X + \sqrt{1 - \rho^2} Z$

б) Угол с градусной мерой  $\theta = \pi/2 + \arcsin(\rho)$

в)  $\mathbb{P}(XY > 0) = \theta/\pi$

**13.2** Никак. Если рассмотреть величины  $-X$ ,  $-Y$ ,  $-Z$ , то у них такое же математическое ожидание и такая же ковариационная матрица. Значит  $\mathbb{P}(XYZ > 0) = \mathbb{P}((-X)(-Y)(-Z) > 0)$ . Но эти вероятности в сумме дают 1, значит они равны по 0.5.

**14.1** да

**14.2**

**14.3** (check!)

Let  $T_a = \min t > 0 : W_t = a$ . Using simple properties of B.M. you can easily show that  $\mathbb{P}(T_a < T) = 2\mathbb{P}(W_T > a)$ . That's all you need to solve both problems.

1.  $2N(-a/\sqrt{T})$

2.  $2N(-a/\sqrt{2T})$

**15.1**  $X_t dt$