



ФАКУЛЬТЕТ  
ВЫЧИСЛИТЕЛЬНОЙ  
МАТЕМАТИКИ И  
КИБЕРНЕТИКИ  
МГУ ИМЕНИ  
М.В. ЛОМОНОСОВА



*teach-in*  
ЛЕКЦИИ УЧЕНЫХ МГУ

# ВЕРОЯТНОСТНЫЕ МОДЕЛИ

ШЕСТАКОВ  
ОЛЕГ ВЛАДИМИРОВИЧ

ВМК МГУ

КОНСПЕКТ ПОДГОТОВЛЕН  
СТУДЕНТАМИ, НЕ ПРОХОДИЛ  
ПРОФ. РЕДАКТУРУ И МОЖЕТ  
СОДЕРЖАТЬ ОШИБКИ.  
СЛЕДИТЕ ЗА ОБНОВЛЕНИЯМИ  
НА [VK.COM/TEACHINMSU](https://vk.com/teachinmsu).

ЕСЛИ ВЫ ОБНАРУЖИЛИ  
ОШИБКИ ИЛИ ОПЕЧАТКИ,  
ТО СООБЩИТЕ ОБ ЭТОМ,  
НАПИСАВ СООБЩЕСТВУ  
[VK.COM/TEACHINMSU](https://vk.com/teachinmsu).



**БЛАГОДАРИМ ЗА ОЦИФРОВКУ КОНСПЕКТА  
СТУДЕНТА ВМК МГУ  
БАБКИНА ИЛЬЮ АНДРЕЕВИЧА**

---

## Содержание

<b>Лекция 1</b>	<b>4</b>
Что такое «вероятностная модель»? . . . . .	4
Условия адекватности применимости вероятностной модели . . . . .	5
Парадокс Бертрана . . . . .	6
Вероятностное пространство . . . . .	7
Зачем нужна $\sigma$ -алгебра? . . . . .	7
Случайная величина . . . . .	8
Примеры измеримости и неизмеримости функций . . . . .	8
Распределение случайной величины . . . . .	9
Определение . . . . .	9
Функция распределения случайной величины . . . . .	9
Виды распределений . . . . .	10
Модели центра случайной величины . . . . .	11
Математическое ожидание . . . . .	11
Медиана . . . . .	11
<b>Лекция 2</b>	<b>13</b>
Модели центра случайной величины . . . . .	13
Медиана (продолжение) . . . . .	13
Мода . . . . .	14
Другие модели центра . . . . .	15
Модели разброса случайной величины . . . . .	16
Дисперсия . . . . .	16
Среднее абсолютное отклонение . . . . .	16
Размах. Интерквартильный размах . . . . .	17
Медианные меры разброса . . . . .	17
Независимость событий . . . . .	17
Независимость случайных величин . . . . .	18
Ковариация и коэффициент корреляции . . . . .	19
<b>Лекция 3</b>	<b>21</b>
Ранговые коэффициенты корреляции . . . . .	21
Подходы к построению вероятностных моделей . . . . .	22
Виды сходимости последовательности случайных величин . . . . .	22
Теоремы, основанные на сходимости . . . . .	24
(Усиленный) закон больших чисел . . . . .	24
Центральные предельные теоремы . . . . .	24

<b>Лекция 4</b>	<b>28</b>
Центральные предельные теоремы (продолжение)	28
Репрезентативность выборки	28
Теорема Пуассона	30
Устойчивое распределение	32
Теорема Пуассона (общий случай)	33
<b>Лекция 5</b>	<b>35</b>
Теория информации	36
Определение количества информации	36
Энтропия	37
<b>Лекция 6</b>	<b>40</b>
Дифференциальная энтропия	40
Случайные процессы	42
Определения	42
Пуассоновский процесс	44
<b>Лекция 7</b>	<b>45</b>
Пуассоновский процесс (продолжение)	45
Связь с пуассоновским распределением	45
Первое информационное свойство	46
Второе информационное свойство	47
<b>Лекция 8</b>	<b>50</b>
Связь пуассоновского процесса с нормальным распределением	50
Случайные суммы	51
Определение, свойства	51
Пуассоновские случайные суммы	52
Геометрические случайные суммы	53
Связь между пуассоновскими и геометрическими случайными суммами	54
<b>Лекция 9</b>	<b>56</b>
Случайные суммы	56
Связь между пуассоновскими и геометрическими случайными суммами (продолжение)	56
Теорема переноса	56
<b>Лекция 10</b>	<b>61</b>
Смеси распределений	61
Смешанные пуассоновские распределения	61

Определения . . . . .	62
Идентифицируемость смесей . . . . .	63
<b>Лекция 11</b>	<b>65</b>
Идентифицируемость смесей (продолжение) . . . . .	65
Процессы Кокса . . . . .	66
Обобщение пуассоновского процесса . . . . .	66
Определения . . . . .	66
Обобщённые процессы Кокса . . . . .	67
<b>Лекция 12</b>	<b>69</b>
Предельные теоремы для обобщённых процессов Кокса . . . . .	69
Свойства масштабных смесей нормальных законов . . . . .	71
<b>Лекция 13</b>	<b>73</b>
Устойчивость смесей нормальных законов относительно смешивающего рас- пределения . . . . .	73
Прямая задача . . . . .	73
Обратная задача . . . . .	75
Метрика Леви . . . . .	76
<b>Лекция 14</b>	<b>77</b>
Обратная задача (продолжение) . . . . .	77
Моделирование распределений приращений финансовых индексов смесями нормальных законов . . . . .	79

## Лекция 1

Цель данного курса: получить дополнительные знания по теории вероятностей и о том, как эти знания применять на практике в зависимости от ситуации.

### Рекомендованная литература

- [1] *Королев В. Ю.* Теория вероятностей и математическая статистика. — М.: Проспект, 2006.
- [2] *Феллер В.* Введение в теорию вероятностей и ее приложения. — М.: Мир, 1984.
- [3] *Королев В. Ю., Бенинг В. Е., Шоргин С. Я.* Математические основы теории риска. — М.: Физматлит, 2007.
- [4] *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Издательство Московского Университета, 2011.

### Что такое «вероятностная модель»?

**Вероятностная модель** – это математическая модель реального явления, содержащего элементы принципиально неустранимой неопределённости (случайности).

Но что из себя представляет «принципиально неустраняемая неопределённость»? Принципиальность может заключаться в нежелании её устранять (например, из-за неоправданных затрат). При игре в «орёл или решка» можно принять во внимание все возможные факторы, начиная от силы подбрасывания и заканчивая магнитным полем Земли, и с точностью вычислить сторону, на которую упадёт монета. Но никто так не делает, поэтому возникает неопределённость, хотя теоретически результат предрешён. Или возьмём к примеру социальные опросы: можно опросить всех жителей страны и выяснить точное отношение к вопросу, но так тоже не делают, а берут некоторую выборку, опрашивают её, делают предположение, что остальные ведут себя также, и далее уже из полученных ответов делают некий вывод. Мы сами ограничиваем себя в наборе параметров, которые измеряем.

Другая причина неопределённости заключается в природной случайности. В XIX веке господствовала теория детерминизма: все наши неприятности связаны с тем, что мы не знаем всех начальных данных, но на самом деле ничего случайного нет и всё в мире предрешено. Первым ударом по этой теории стало появление теории относительности: из процесса наблюдения нельзя устранить наблюдателя, сам факт

наличия измерения вносит изменения в процесс и узнать исход эксперимента невозможно. Но теория детерминизма на этом не закончилась: если бы вдруг существовал идеальный наблюдатель, который не вмешивается в процесс, то тогда всё было бы предрешено. Потом появилась квантовая механика и выяснилось, что элементарные частицы с определённой вероятностью могут находиться в нескольких состояниях одновременно. Электрон, в зависимости от того наблюдают за ним или нет, может вести себя и как волна, и как частица.

Самой известной попыткой перенести квантовые эффекты в макромир является мысленный эксперимент Шрёдингера. В непрозрачной коробке находится кот и ампула с ядом. Ампула разбивается, если прибор регистрирует, что некоторая элементарная частица приняла некоторое энергетическое состояние. Поскольку частица с определённой вероятностью может находиться во всех энергетических состояниях, то и кот с определённой вероятностью и жив, и мёртв.

На сегодняшний день считается, что природная случайность на квантовом уровне есть, хотя противником этого в своё время был даже Альберт Эйнштейн. (Одним из последних этому возразил Стивен Хокинг.)

## Условия адекватности применимости вероятностной модели

### 1. Наличие случайности.

2. *Воспроизводимость условий.* Эксперимент должен быть воспроизводим (можно повторить сколько угодно раз). Вообще говоря это своего рода договорённость. Считается, что макроусловия всё время одинаковы и воспроизводимость такая, что эффект случайности не меняется и всегда действует одинаковым образом. Бывают ситуации, когда воспроизвести эксперимент нельзя. Например, несколько лет назад был зарегистрирован сигнал из космоса и до сих пор спорят сигнал ли это разумной цивилизации или просто импульс небесного тела. После этого такой импульс больше не регистрировали.

3. *Устойчивость частот.* Если мы  $n$  раз будем воспроизводить один и тот же эксперимент, и в результате событие  $A$  наступит  $n(A)$  раз, тогда при увеличении  $n$  мы увидим, что отношение  $\frac{n(A)}{n}$  колеблется у какого-то числа  $p$ , которое эмпирически определяет вероятность этого события. То есть мы получим что-то вроде  $\frac{n(A)}{n} \xrightarrow{n \rightarrow \infty} p$ . В дальнейшем мы рассмотрим это подробнее.

В случае соблюдения всех этих условий можно строить вероятностную модель этого явления/процесса. Но следует соблюдать аккуратность, так как словесное описание эксперимента и его математическая модель это не одно и то же.

## Парадокс Бертрана

**Задача.** Определить вероятность того, что наугад выбранная хорда окружности имеет длину больше, чем длина стороны вписанного в эту окружность правильного треугольника.

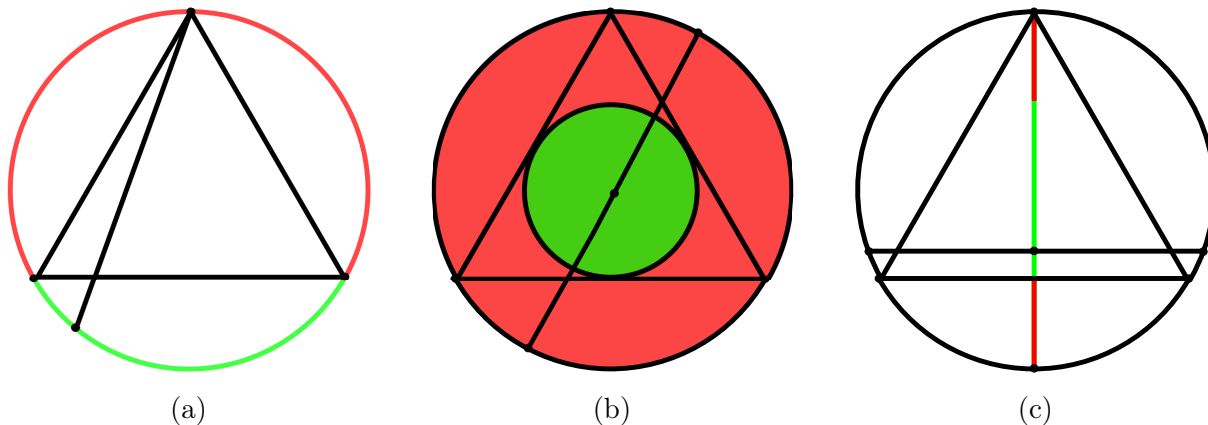


Рис. 1

Оказывается, что слово «наугад» здесь и создаёт проблемы. Рассмотрим три примера, в которых мы по-разному определим его и получим три различные вероятности.

1. Так как правильный треугольник можно вписывать в окружность как угодно, то мы можем зафиксировать первую точку хорды на одной из вершин. Другую же точку мы будем выбирать как угодно на окружности. (Рис. 1а) Очевидно нас будет интересовать дуга, противоположная первой точке, и её отношение ко всей окружности  $p = \frac{1}{3}$ .
2. Для любой точки круга существует ровно одна хорда окружности, для которой эта точка является серединой. Будем выбирать в круге точку и уже по ней восстанавливать хорду. (Рис. 1б) Нам подходит область внутри вписанной окружностями. Получим  $p = \frac{1}{4}$ .
3. Будем рассматривать хорды, перпендикулярные диаметру, а значит будем выбирать точку на диаметре. (Рис. 1с) В таком случае нас устроит отрезок равный половине диаметре, следовательно  $p = \frac{1}{2}$ .

На самом деле в этой задаче слово «наугад» можно определить так, что вероятность получится любым числом от 0 до 1. Обратите внимание, что во всех трёх случаях мы получили однозначный ответ и все наши модели были корректными.



## Вероятностное пространство

Вероятностное пространство состоит из трёх объектов  $(\Omega, \mathcal{A}, P)$ .

- $\Omega$  – непустое множество, пространство элементарных исходов.
- $\mathcal{A}$  – множество подмножеств  $\Omega$ , удовлетворяющее условиям:

1)  $\Omega \in \mathcal{A}$ ;

2)  $A \in \mathcal{A} \implies \bar{A} \in \mathcal{A}$ ;

3)  $A_1, A_2, \dots \in \mathcal{A} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ .

Такое множество называют  $\sigma$ -алгеброй.

- $P : \mathcal{A} \rightarrow [0, 1]$  – вероятностная мера. Она удовлетворяет условиям:

1)  $P(A) \geq 0$  для  $\forall A \in \mathcal{A}$ ;

2)  $P(\Omega) = 1$ ;

3)  $A_1, A_2, \dots : A_i \cap A_j = \emptyset$  для  $\forall i, j : i \neq j \implies P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$  –  
свойство счётной аддитивности.

## Зачем нужна $\sigma$ -алгебра?

Наличие  $\sigma$ -алгебры позволяет сузить класс множеств, на которых мы можем определить вероятность. Те множества, которые входят в  $\sigma$ -алгебру мы называем **событиями**, а остальные множества событиями не являются. И только на событиях можно определить вероятность.

**Пример.**  $\Omega = [0, 1]$ , случайным образом бросаем точку на отрезок. Понятно, что для  $[a, b] \subset [0, 1]$  естественно определить вероятность попадания точки на полуинтервал как  $P([a, b)) = \frac{b-a}{1-0} = b-a$ . Почему именно полуинтервал? Чисто из эстетических соображений, далее мы разберём это подробнее. Таким образом мы можем определить вероятность на всех множествах такого вида.

Далее мы строим  $\sigma$ -алгебру  $\sigma([a, b))$ , которая порождается всеми возможными полуинтервалами вида  $[a, b)$ , то есть это минимальная  $\sigma$ -алгебра, которая содержит все эти полуинтервалы. (По сути мы берём все  $\sigma$ -алгебры, которые содержат эти полуинтервалы, и пересекаем их.)

**Теорема 1** (о продолжении меры). *Существует и единственна такая мера  $P^*$  на  $\sigma([a, b))$ , что  $P^*([a, b)) = P([a, b)) = b-a$ .*

Есть примеры неизмеримых множеств, для которых можно доказать, что разумным способом определить на них меру заданную таким образом нельзя (например, мера Бореля). Следовательно надо сужать класс множеств, чтобы можно было задать вероятность. Для этого и нужна  $\sigma$ -алгебра: она позволяет корректно определить вероятность на всех множествах, которые мы называем событиями.

## Случайная величина

Случайная величина нужна для кодирования исходного вероятностного пространства в то вероятностное пространство, которое оперирует только с числами:

$$\xi : \Omega \rightarrow \mathbb{R}$$

Но этого *недостаточно*, нам необходимо дополнить это определение борелевской  $\sigma$ -алгеброй. **Борелевская  $\sigma$ -алгебра  $\mathcal{B}(\mathbb{R})$**  – это  $\sigma$ -алгебра, порождённая  $[a, b]$  для любых  $a, b \in \mathbb{R}$ . Она содержит в себе все нужные для практики множества: отрезки, полуинтервалы, полупрямые, точки и т. д.

Так почему же всё-таки полуинтервалы? Потому что ими можно непересекающимся образом покрыть всю прямую, но на самом деле никакой разницы нет.

$$B \in \mathcal{B}(\mathbb{R}), \quad P(\xi \in B) = P(\omega : \xi(\omega) \in B);$$

$$\forall B \in \mathcal{B}(\mathbb{R}) : \xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathcal{A}.$$

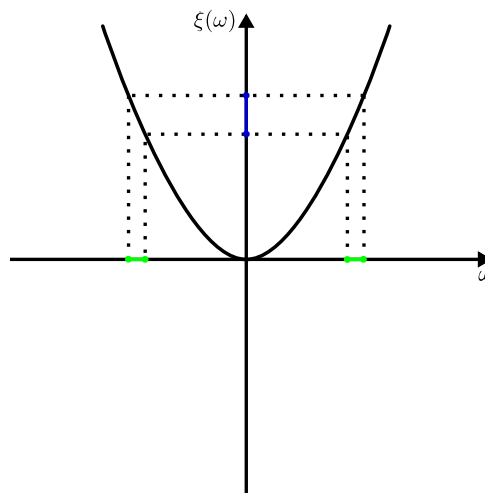


Рис. 2: Пример полного прообраза

И только после этого дополнения (*свойство измеримости*) функция  $\xi$  по праву может зваться **случайной величиной**.

## Примеры измеримости и неизмеримости функций

Измеримость напрямую зависит от того какая задана  $\sigma$ -алгебра на исходном пространстве. Рассмотрим пару примеров на следующих множествах:

$$\Omega = [0, 1], \quad \mathcal{A} = \{\emptyset, \Omega\}.$$

Все требования  $\sigma$ -алгебры здесь выполняются. Такую  $\sigma$ -алгебру называют **тривиальной  $\sigma$ -алгеброй**.

1.  $\xi(\omega) = 7$ . Проверим, является ли эта функция измеримой:

$$\forall B \in \mathcal{B}(\mathbb{R}) \begin{cases} 7 \notin B \rightarrow \xi^{-1}(B) = \emptyset \\ 7 \in B \rightarrow \xi^{-1}(B) = \Omega \end{cases} \implies \xi(\omega) - \text{случайная величина.}$$

2.  $\xi(\omega) = \omega$ . В данном случае функция уже не будет случайно величиной. Покажем это:

$$B = \left[0, \frac{1}{2}\right], \quad \xi^{-1}(B) = \left[0, \frac{1}{2}\right] \notin \mathcal{A} \implies \xi(\omega) - \text{нелучайная величина.}$$

Этот отрезок не входит в  $\mathcal{A}$ , значит значения, которая функция принимает, нельзя померить при помощи множеств из  $\mathcal{A}$  и она является неизмеримой относительно заданной  $\sigma$ -алгебры.

*Вывод:* необходимо корректно определять вероятностную меру. Без приведённых ранее свойств это сделать нельзя.

## Распределение случайной величины

### Определение

Итак, вероятностная модель – это вероятностное пространство, а вероятностное пространство с помощью случайной величины кодируется в новое вероятностное пространство уже на прямой:

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_\xi),$$

где  $P_\xi(B) = P(\omega : \xi(\omega) \in B)$ .

Такое вероятностное пространство называется **индуцированным вероятностным пространством**, а мера  $P_\xi$  называется **распределением случайной величины  $\xi$** .

### Функция распределения случайной величины

Гораздо чаще используется не само распределение, а **функция распределения**:

$$F_\xi(x) = P_\xi((-\infty, x)) = P(\xi < x). \quad (1)$$

При этом неравенство необязательно строгое, это просто дело привычки.

Известно, что распределение и функция распределения взаимно-однозначно определяют друг друга. То, что распределение определяет функцию распределения, понятно, это всего лишь частный случай борелевских множеств. А вот обратное следует из теоремы о продолжении меры (см. Теорема 1).

Давайте вспомним свойства функции распределения:

- 1)  $0 \leq F_{\xi}(x) \leq 1$ ;
- 2)  $F_{\xi}(x)$  не убывает;
- 3)  $\lim_{x \rightarrow +\infty} F_{\xi}(x) = 1, \lim_{x \rightarrow -\infty} F_{\xi}(x) = 0$ ;
- 4)  $F_{\xi}(x)$  непрерывна слева, а если взять в (1) нестрогое неравенство, то непрерывна справа.

## Виды распределений

1. Абсолютно непрерывное распределение (относительно меры Лебега):

$$\exists f_{\xi}(x) \geq 0 : F_{\xi}(x) = \int_{-\infty}^x f_{\xi}(u) du,$$

где  $f_{\xi}(x)$  – это **плотность распределения**  $\xi$ .

2. Дискретное распределение. Нагляднее всего показать это на графике. (Рис. 3)

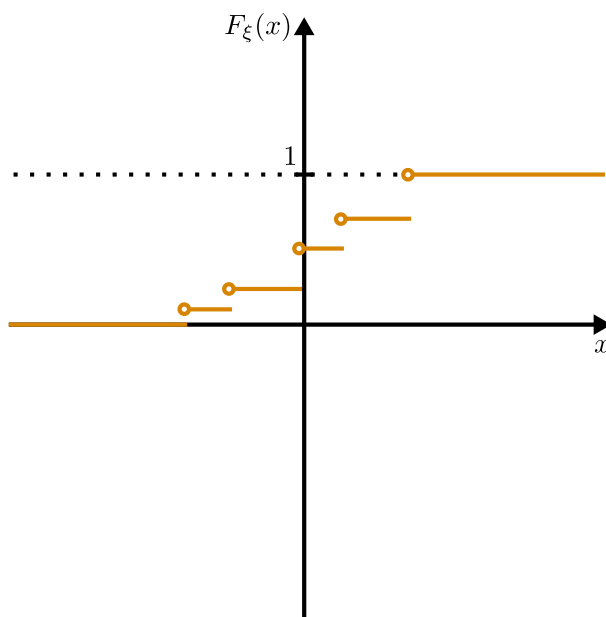


Рис. 3

3. *Сингулярное распределение* – это такое распределение, функция которой непрерывна, но мера (Лебега) точек роста равна нулю. В одномерном случае это выглядит экзотически, но в многомерных случаях эти распределения встречаются часто. И вообще говоря дискретное распределение на самом деле тоже является сингулярным распределением, но их выделили в отдельный класс в виду их важности в теории.

Также верно следующее:

$$\left. \begin{aligned} P(\xi \in [a, b]) &= \int_a^b f_\xi(x) dx \\ P(\xi \in [a, b]) &= \sum_{a \leq x_k \leq b} P(\xi = x_k) \end{aligned} \right\} = \int_a^b dF_\xi(x) - \text{интеграл Лебега-Стилтьеса.}$$

## Модели центра случайной величины

### Математическое ожидание

**Математическое ожидание:**  $E\xi = \int_{-\infty}^{+\infty} x dF_\xi(x)$ . Оно бывает:

- *абсолютно непрерывное:*  $E\xi = \int_{-\infty}^{+\infty} x f_\xi(x) dx$ ;
- *дискретное:*  $E\xi = \sum_{k=1}^{\infty} x_k P(\xi = x_k)$ , где  $x_k$  пробегает все возможные значения, которые принимает  $\xi$ .

**Свойства математического ожидания:**

- 1) линейность:  $E(a\xi + b\eta) = aE\xi + bE\eta$ ;
- 2) не всегда существует, но если существует, то только единственное;
- 3) не устойчиво к выбросам: то есть, если существует значение случайной величины, которое очень большое (или маленькое), то несмотря на малость его вероятности, она повлияет на математическое ожидание, тем самым исказив «картину»; собственно говоря, этот недостаток и привёл к поиску других моделей центра для случайной величины.

### Медиана

Перед определением медианы мы определим, что такое квантиль. Пусть  $q \in (0, 1)$ . Тогда  $x_q$  –  $q$ -квантиль (или **квантиль порядка  $q$** ) распределения  $F_\xi(x)$ , если  $P(\xi <$

$x_q) < q$ ,  $P(\xi \leq x_q) \geq q$ . Чтобы было яснее, давайте рассмотрим следующие равенства:

$$P(\xi < x_q) = F_\xi(x_q) = F_\xi(x_q-), \quad P(\xi \leq x_q) = F_\xi(x_q+).$$

Тогда понятно, что  $F_\xi(x_q-) \leq q$ ,  $F_\xi(x_q+) \geq q$ , а если  $F_\xi(x)$  непрерывна в  $x_q$ , то  $F_\xi(x_q-) = q$ . Наконец, мы можем определить **медиану** как квантиль распределения порядка  $\frac{1}{2}$ . *Обозначение:*  $\text{med } \xi = x_{\frac{1}{2}}$ .

Грубо говоря, медиана – это такая точка, относительно которой вероятность попасть вправо или влево  $\frac{1}{2}$ .

## Лекция 2

### Модели центра случайной величины

#### Медиана (продолжение)

**Модельная задача** (о перемещениях). Допустим, мы каждый день должны путешествовать в различные места: работа, магазин, поездка в гости и так далее. И мы перемещаемся в разные места с разной частотой. Пусть  $\xi$  – это место, куда мы должны попасть. (Будем рассматривать на прямой.) Наша задача: узнать в какой точке  $a$  на прямой выгодней всего жить, чтобы затраты и потери при перемещениях были минимальны. Предположим, что затраты по перемещению из  $a$  в  $\xi$  задаются через формулу  $(a - \xi)^2$ . Почему квадрат? Это парабола: у неё единственный глобальный минимум и хорошие аналитические свойства. Почему именно такая функция? Потому что при перемещении на небольшие расстояния затраты небольшие, а на больших расстояниях мы устаём больше (пусть мы ходим пешком) и затраты большие, значит мы затрачиваем ресурсы не линейно, а квадратично.

Естественным образом мы перейдём к следующей функции потерь (затрат):  $E(a - \xi)^2$ . Где тогда будет центр, то есть место, где функция потерь принимает минимальные значения?

$$E(a^2 - 2a\xi + \xi^2) - \text{парабола} \implies \underset{a}{\operatorname{argmin}} (a - \xi)^2 = E\xi.$$

А что если мы пересядем на машину? Тогда у нас будет затрачиваться только топливо и функция потерь будет без квадрата:  $E|\xi - a|$ . В этом случае центром будет не математическое ожидание, а медиана. Убедимся в этом. Пусть  $a > \operatorname{med} \xi$  (аналогично в случае  $<$ ). Мы хотим доказать, что

$$E|\xi - a| - E|\xi - \operatorname{med} \xi| \geq 0.$$

Рассмотрим следующее выражение:

$$|\xi - a| - |\xi - \operatorname{med} \xi| = \begin{cases} a - \operatorname{med} \xi, & \text{если } \xi \in (-\infty, \operatorname{med} \xi]; \\ \operatorname{med} \xi - a, & \text{если } \xi \in [a, +\infty); \\ a - 2\xi + \operatorname{med} \xi, & \text{если } \xi \in (\operatorname{med} \xi, a). \end{cases}$$

Для случая, когда  $\xi \in (\operatorname{med} \xi, a)$ , будет верно, что  $a - 2\xi + \operatorname{med} \xi \geq \operatorname{med} \xi - a$ . Вернёмся к математическим ожиданиям:

$$\begin{aligned} E|\xi - a| - E|\xi - \operatorname{med} \xi| &\geq (a - \operatorname{med} \xi) E I(\xi \in (-\infty, \operatorname{med} \xi]) + \\ &+ (\operatorname{med} \xi - a) E I(\xi \in [a, +\infty)) + (\operatorname{med} \xi - a) E I(\xi \in (\operatorname{med} \xi, a)), \end{aligned}$$

где  $I(\cdot)$  – индикатор события. Что такое  $E I(\cdot)$ ? Это вероятность события. Тогда можем записать выражение справа следующим образом:

$$\begin{aligned} (a - \text{med } \xi) P(\xi \leq \text{med } \xi) - (a - \text{med } \xi) P(\xi > \text{med } \xi) &= \\ &= \underbrace{(a - \text{med } \xi)}_{>0} \underbrace{(P(\xi \leq \text{med } \xi) - P(\xi > \text{med } \xi))}_{\geq 0} \geq 0. \end{aligned}$$

Таким образом мы показали, что в медиане достигается минимум функции потерь.

### Свойства медианы:

- 1) всегда существует;
- 2) устойчива к выбросам: то есть если есть экстремально большие значения, то медиана «не почувствует» их, если они имеют малую вероятность;
- 3) может быть не единственной.

## Мода

*Дискретный случай.* **Мода** ( $\text{mod}$ ) – это такое значение, которое принимается с наибольшей вероятностью:

$$\xi : \begin{matrix} x_1, \dots, x_n, \dots \\ p_1, \dots, p_n, \dots \end{matrix} \implies \text{mod } \xi = x_k, \text{ что } P(\xi = x_k) \geq P(\xi = x_i) \text{ для } \forall i = 1, 2, \dots$$

*Абсолютно непрерывный случай.* **Мода** – это такое значение, на которой достигается максимум плотности распределения:

$$f_\xi(\text{mod } \xi) \geq f_\xi(x) \text{ для } \forall x \in \mathbb{R}.$$

Но нам также надо потребовать, чтобы функция  $f_\xi$  была кусочно-непрерывной, потому что плотность определена почти всюду и, вообще говоря, мы говорим не об одной функции, а о классе функций, поэтому нам необходимо это уточнение, иначе определение будет бессмысленно (можно взять любое число).

### Замечания:

- мода может быть не единственной: распределения с одной модой называют **унимодальными**, а с несколькими – **мультимодальными**;
- определение в абсолютно непрерывном случае необходимо скорее для различных утверждений (иногда удобно использовать понятие унимодальности), а вот практическое применение она находит в дискретном случае;
- если распределение мультимодально, то данные скорее всего разнородные (о смешанных моделях мы будем говорить позже).



**Модельная задача** (о манипуляции общественным мнением). Предположим, что исследуется уровень благосостояния граждан и, исходя из статистических данных, выясняется, что кривая благосостояния подчиняется экспоненциальному закону:

$$P(\xi > x) = e^{-x} \text{ для } x \geq 0.$$

Нужно подготовить отчёт начальству, которое понятие не имеет, что такое экспонента. Значит нужно представить результат в более понятном виде: фактически, нужно выбрать модель для описания среднего дохода. И возникает вопрос: что взять в качестве среднего? Всё зависит от ваших целей. Если вы хотите:

- 1) показать, что все богаты и жизнь налаживается, то нужно брать математическое ожидание:  $E\xi = 1$  (условно);
- 2) исключить влияние очень богатых людей, то нужно брать медиану:  $\text{med } \xi = \ln 2 \approx 0.7$ ;
- 3) призвать к революции или подсадить начальство, то нужно брать моду:  $\text{mod } \xi = 0$ .

*Вывод:* с помощью статистики можно объяснить всё, что угодно, и так можно пудрить мозги несведущим людям; учите вероятность: если Вы будете владеть всеми этими методами, то иногда вы сможете обернуть некоторые выводы в свою сторону.

## Другие модели центра

**Усечённое математическое ожидание.** По сути мы просто убираем экстремальный выброс:

$$E(\xi I(x_{q_1} < \xi < x_{q_2})).$$

Такая модель тоже используется, но конечно реже, чем медиана или обычное математическое ожидание.

**Моральное ожидание.** Возникло это понятие в связи с «петербургским парадоксом»: Вы играете в «орёл или решка» и Вам даётся выигрыш  $2^n$ , если впервые орёл выпал на  $n$ -ом подбрасывании. Легко видеть, что вероятность такого исхода это  $2^{-n}$ , а математическое ожидание бесконечно. Значит справедливая цена за игру – бесконечность. Даниил Бернулли, разбирая этот парадокс, предположил, что выгода прямо пропорциональна приросту и обратно пропорциональна начальному капиталу ( $A$  – начальный капитал,  $Z$  – выгода,  $k$  – некоторый коэффициент):

$$\Delta Z = k \frac{\Delta A}{A}, \quad k > 0.$$

Записав в дифференциалах, мы получим, что  $Z = k \ln A$ . Теперь предположим, что

в результате игры капитал изменяется на случайную величину  $X$ :

$$E k \ln(X + A) = k \ln(Mr + A),$$

где  $Mr$  и есть **моральное ожидание**.

В дискретном случае:  $Mr = \prod_{i=1}^n (x_i + A)^{p_i} - A$ .

Фактически моральное ожидание является первым шагом к «теории полезности».

Что касается «петербургского парадокса»: если игру считать справедливой, то моральное ожидание окажется меньше нуля, и Бернулли назвал это явным указанием природы на то, что надо воздержаться от азартных игр.

## Модели разброса случайной величины

### Дисперсия

**Дисперсия:**  $D\xi = E(\xi - E\xi)^2 = E\xi^2 - (E\xi)^2$ .

**Свойства дисперсии:**

- 1)  $D\xi \geq 0$ , причём  $D\xi = 0 \iff P(\xi = E\xi) = 1$ ;
- 2)  $D(c\xi) = c^2 D\xi$ ;
- 3) обладает хорошими аналитическими свойствами (например, хорошо минимизируется);
- 4) по большому счёту имеет те же достоинства недостатки, что и математическое ожидание: может не существовать и неустойчива к выбросам; причём в случае дисперсии эти недостатки усиливаются (из-за квадрата).

Чаше всего работают не с самой дисперсией, а с её корнем – **среднеквадратичным отклонением** ( $\sqrt{D\xi}$ ), – чтобы случайная величина и её разброс были одной размерности.

### Среднее абсолютное отклонение

**От медианы:**  $E|\xi - \text{med } \xi|$ . Используется реже, так как при вычислении модуль разбивается на разные промежутки и в явном виде это вычислить трудно.

**От математического ожидания:**  $E|\xi - E\xi|$ , также называемая **инженерная метрика**.

## Размах. Интерквартильный размах

**Интерквартильный размах:**  $IQR = x_{\frac{3}{4}} - x_{\frac{1}{4}}$ , то есть из наблюдений отсекается по 25% с каждой стороны и вычисляется размах по оставшимся. В отличие от предыдущих характеристик эта к выбросам устойчива.

Если Вы хотите отбросить не 25%, а какое-то другое количество, то это уже будет **интерквантильный размах**. Возможен также просто **размах** (для случайных величин с конечным носителем):  $\mathcal{R} = \max \xi - \min \xi$ .

## Медианные меры разброса

Такие меры ещё более устойчивы к выбросам. Вот пара примеров:

$$\text{med}|\xi - E\xi|, \quad \mathcal{MAD} = \text{med}|\xi - \text{med} \xi|.$$

Кстати, если выбросов нет, то характеристики интерквартильного размаха и медианы абсолютного отклонения от медианы оказываются асимптотически эквивалентными.

Перед тем, как перейти к следующей теме, давайте рассмотрим как модель разброса может помочь нам понять, как отличаются разные модели центра:

$$|E\xi - \text{med} \xi| \leq E|\xi - \text{med} \xi| \leq E|\xi - E\xi| \leq \sqrt{D\xi} \quad (\text{по неравенству Ляпунова}).$$

## Независимость событий

Имеем вероятностное пространство  $(\Omega, \mathcal{A}, P)$  и  $A, B \in \mathcal{A}$ .  $A$  и  $B$  называются **независимыми событиями**, если  $P(AB) = P(A)P(B)$ .

И вроде всё понятно, но оказывается, что одно и то же словесное описание событий может в различных математических моделях приводить как к их независимости, так и к зависимости. Рассмотрим такой случай.

**Модельная задача** (о джокере). Предположим, что у нас есть колода карт из 52 карт (от 2 до туза, всего 4 масти). Мы хотим исследовать на независимость два события: что карта масти пик и что карта – туз:

$$P(\text{туз пик}) = \frac{1}{52}, \quad P(\text{туз}) = \frac{4}{52}, \quad P(\text{пик}) = \frac{13}{52}.$$

И мы заключаем, что да, они независимы (в стохастическом смысле), так как  $P(\text{туз пик}) = P(\text{туз})P(\text{пик})$ .

Далее мы добавим в колоду ещё одну карту – джокер. Вопрос: останутся ли те же самые события независимыми в этом случае? Проверяем:

$$P(\text{туз пик}) = \frac{1}{53}, \quad P(\text{туз}) = \frac{4}{53}, \quad P(\text{пик}) = \frac{13}{53}.$$

Как мы видим – нет, теперь они зависимы, так как  $P(\text{туз пик}) \neq P(\text{туз})P(\text{пик})$ . Иногда узнать зависимы события или нет можно только посчитав.

Если событий больше, чем два, то для них независимость определяется следующим образом:  $A_1, A_2, \dots, A_n \in \mathcal{A}$  называются **независимыми событиями**, если для  $\forall k = \overline{2, n}$  и  $1 \leq i_1 < \dots < i_k \leq n$  верно, что  $P(A_{i_1} \dots A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$ .

## Независимость случайных величин

Случайные величины  $\xi$  и  $\eta$  называются **независимыми**, если для  $\forall A, B \in \mathcal{B}(\mathbb{R})$  верно, что  $P(\xi \in A, \eta \in B) = P(\xi \in A)P(\eta \in B)$ .

В отличие от событий в случае независимости большего количества случайных величин дополнительные условия не нужны и определение будет выглядеть похожим образом.

Стохастическая независимость не связана с функциональной независимостью. Рассмотрим пару примеров.

**Пример 1.**  $\Omega = [0, 1]$ ,  $\mathcal{A} = \mathcal{B}([0, 1])$ ,  $P = \lambda$  – мера Лебега.

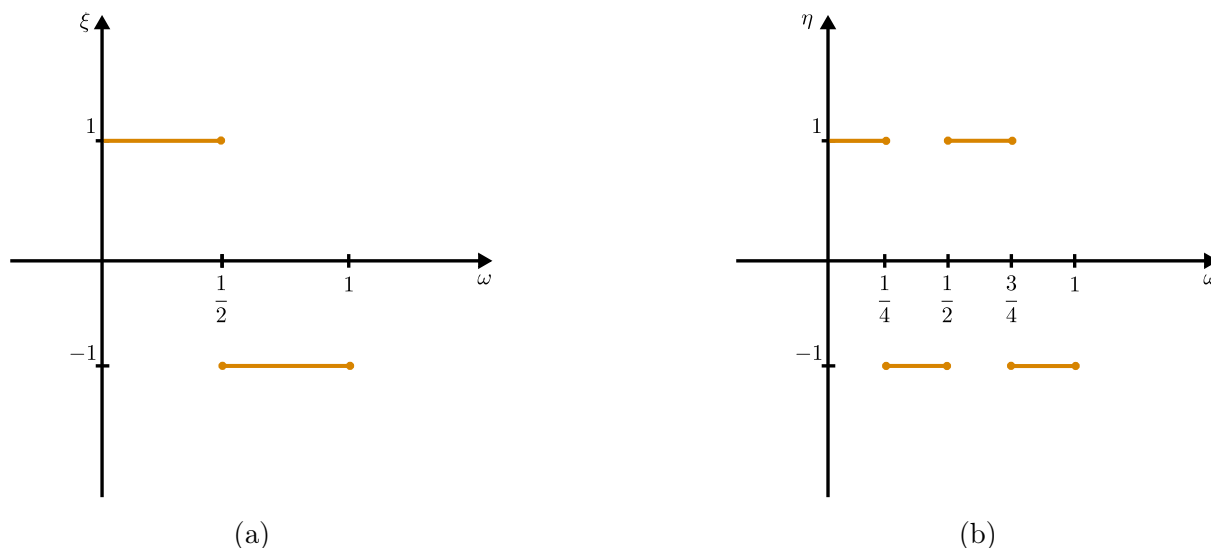


Рис. 4

Очевидно, что  $\xi^2 + \eta^2 = 2$ , но в то же время эти случайные величины стохастически

независимы:

$$P(\xi = 1, \eta = 1) = \frac{1}{4} = \underbrace{P(\xi = 1)}_{\frac{1}{2}} \underbrace{P(\eta = 1)}_{\frac{1}{2}}.$$

**Пример 2.**  $\xi \sim R[0, 2\pi]$ ,  $\eta_1 = \cos \xi$ ,  $\eta_2 = \sin \xi$ .  $\eta_1^2 + \eta_2^2 = 1$ , но в то же время они стохастически зависимы. Проверьте самостоятельно. :)

## Ковариация и коэффициент корреляции

**Ковариация:**  $\text{cov}(\xi, \eta) = E[(\xi - E\xi)(\eta - E\eta)] = E\xi\eta - E\xi E\eta$ .

Если  $\xi$  и  $\eta$  независимы, то  $\text{cov}(\xi, \eta) = 0$ . Обратное не всегда верно. (см. Пример 2) Некоторые свойства ковариации: она линейна по обоим аргументам, инварианта относительно сдвига случайных величин на константу, может не существовать.

Ковариация даёт представление о качественном характере зависимости: если она положительная, то это значит, что при изменении одной случайной величины в среднем в одну сторону (например, увеличения), то другая в среднем тоже будет изменяться в эту сторону; если же она отрицательная, то стороны меняются. Но абсолютная величина ковариации о силе зависимости ничего не говорит. Именно для этого ввели **коэффициент корреляции**:  $\text{corr}(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{\sqrt{D\xi D\eta}}$ . Если какая-то из случайных величин является константой с вероятностью 1, то коэффициент корреляции по определению равняется нулю. Важное свойство:

$$|\text{corr}(\xi, \eta)| \leq 1, \text{ причём } |\text{corr}(\xi, \eta)| = 1 \iff \eta = a\xi + b.$$

То есть, если коэффициент равен единице, то между величинами полная линейная зависимость, если единице, то линейной зависимости нет, но может быть есть какая-то другая.

Пусть  $\xi$  и  $\eta$  – случайные величины и  $E\xi = 0$ ,  $E\eta = 0$ . Возьмём математическое ожидание от квадратичной формы и раскроем его:

$$E(x\xi + y\eta)^2 = Ex^2\xi^2 + 2E\xi\eta xy + Ey^2\eta^2 \geq 0 \text{ для } \forall x, y \in \mathbb{R}.$$

Мы получили некоторый квадратный трёхчлен относительно  $x$  или  $y$ , который всегда больше или равен нулю. Значит дискриминант отрицательный. Отсюда и следует свойство, которое мы привели выше (знаменатель в коэффициенте всегда больше или равен числителя по модулю).

Стоит отметить, что приведённые меры зависимости нельзя использовать для выявления причинно-следственной связи: если у вас много пожарных частей в районе, где много пожаров, то это не значит, что пожарные части вызывают пожары. Скорее наоборот: пожары вызвали появление пожарных частей в этом районе.

Иногда может возникать ложная корреляция из-за разнородности данных. Простой пример: если взять некоторую группу людей и вычислить эмпирическую корреляцию между ростом и длиной волос, то небольшая зависимость всё-таки обнаружится. Это будет связано с третьим фактором: в группе людей есть и мужчины, и женщины, а мужчины в среднем выше женщин и волосы у мужчин в среднем короче, чем у женщин. А вот если разбить группу на две подгруппы по половому признаку и искать корреляцию уже в них, то она не обнаружится.

## Лекция 3

### Ранговые коэффициенты корреляции

Как выглядит коэффициент корреляции в выборочном виде:

$$\begin{matrix} (x_1, x_2, \dots, x_n) \\ (y_1, y_2, \dots, y_n) \end{matrix} \Rightarrow \hat{c} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}},$$

где в числителе находится выборочная ковариация, а внизу корень из произведения выборочных дисперсий. Такую характеристику называют **коэффициентом корреляции Пирсона**.

А если у нас для распределения не существует вторых моментов, или мы хотим найти нелинейную зависимость, то коэффициент корреляции может не сработать. Для таких случаев используют **ранговые коэффициенты корреляции**.

Две пары  $(x_i, y_i)$  и  $(x_j, y_j)$  называются **согласованными**, если  $\text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j) = 1$ .

Запишем через  $T$  превышение согласованности над несогласованностью, а через  $\tau$  количество возможностей выбрать пары:

$$T = \sum_{i < j} \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j), \quad \tau = \frac{T}{T_{\max}} = \frac{2T}{n(n-1)}.$$

Величину  $\tau$  называют **коэффициентом корреляции Кендалла**.

**Коэффициент корреляции Спирмена**  $\rho$  определяется как коэффициент корреляции Пирсона по ранговым выборкам  $x$  и  $y$  (на месте  $x_i$  или  $y_i$  будет стоять их номер в вариационном ряду, построенному по выборке):

$$\begin{aligned} \begin{matrix} (x_1, x_2, \dots, x_n) \\ (y_1, y_2, \dots, y_n) \end{matrix} &\Rightarrow \begin{matrix} (Rx_1, Rx_2, \dots, Rx_n) \\ (Ry_1, Ry_2, \dots, Ry_n) \end{matrix} \Rightarrow \rho = \frac{\sum_{i=1}^n (Rx_i - \overline{Rx})(Ry_i - \overline{Ry})}{\sqrt{\sum_{i=1}^n (Rx_i - \overline{Rx})^2 \sum_{i=1}^n (Ry_i - \overline{Ry})^2}} = \\ &= 1 - \frac{6}{n(n-1)(n+1)} \sum_{i=1}^n (Rx_i - Ry_i)^2. \end{aligned}$$

Ранговые коэффициенты корреляции, также как и коэффициент корреляции Пирсона, удовлетворяют следующим условиям:  $|\tau| \leq 1$ ,  $|\rho| \leq 1$ . Если равно 1, то зависимость *прямая*, а если  $-1$ , то *обратная*.

В отличие от коэффициент Кендалла коэффициент Спирмена более чувствителен к степени согласованности из-за того, что мы рассматриваем на сколько отличаются ранги, а не просто согласованность.

## Подходы к построению вероятностных моделей

Рассмотрим пару подходов:

- **Эмпирический (статистический).** Самый прямой подход. По сути он состоит из построения выборочных оценок тех или иных характеристик (математическое ожидание, дисперсия, функция распределения и так далее), и уже по ним делаются выводы и прогнозы. Такой подход не очень подойдёт для сложных манипуляций с данными, для этого вероятностная модель должна быть достаточно простой, а эмпирическая модель таковой не является:  $n$  может быть настолько велико, что с расчётами не справится ни один компьютер, к тому же эмпирическая функция является ступенчатой и из-за этого с ней неудобно работать.
- **Асимптотический.** Состоит из подмены истинного распределения каким-то предельным в силу некоторой предельной теоремы. Но для того, чтобы «привлечь» распределению к моделированию чего-либо, нужен факт, что оно возникает в какой-нибудь предельной схеме.

## Виды сходимости последовательности случайных величин

**1. Сходимость почти всюду (почти наверное, с вероятностью 1).** При этом «почти всюду» является более широким понятием, так как в отличие от «почти наверное» и «с вероятностью 1» сходимость необязательно по вероятностной мере.

$$\xi_n \xrightarrow{\text{п.в.}} \xi, \text{ если } P(\omega : \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)) = 1.$$

Такая сходимость - это что-то вроде обычной функциональной сходимости, то есть функциональная последовательность  $\xi_n(\omega)$  сходится поточечно, но не во всех точках, а почти во всех. Там, где она не сходится, общая вероятностная мера равна нулю.

### 2. Сходимость по вероятности:

$$\xi_n \xrightarrow{P} \xi, \text{ если для } \forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P(\omega : |\xi_n(\omega) - \xi(\omega)| > \varepsilon) = 0.$$

Это означает, что последовательность становится всё ближе и ближе к предельной функции, но, всегда остаётся вероятность, что произойдёт выброс. (Рис. 5)



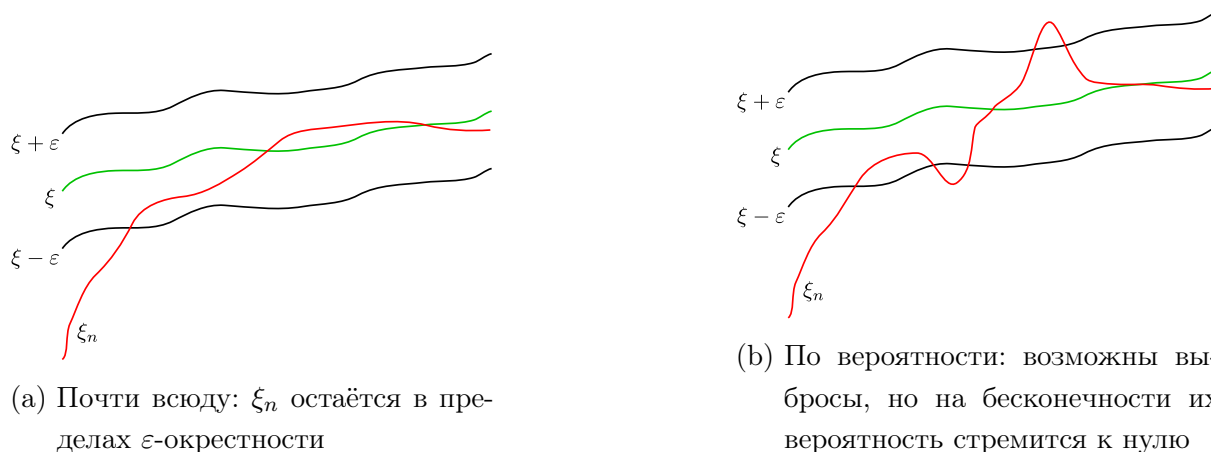


Рис. 5

### 3. Сходимость в среднем порядка $r > 0$ :

$$\xi_n \xrightarrow{r} \xi, \text{ если } E|\xi_n - \xi|^r \xrightarrow{n \rightarrow \infty} 0.$$

Если в сходимости по вероятности к нулю стремится вероятность выброса, то в сходимости в среднем дополнительно ограничивается величина выброса. В практических задачах величина ошибка чаще всего измеряется через  $E|\xi_n - \xi|^r$  (обычно при  $r = 2$ ).

### 4. Слабая сходимость:

$$\xi_n \xrightarrow{w} \xi, \text{ если для } \forall \varphi - \text{непрерывная и ограниченная функция: } E\varphi(\xi_n) \xrightarrow{n \rightarrow \infty} E\varphi(\xi)$$

или в терминах интегралов

$$\int \varphi(x) dF_n(x) \xrightarrow{n \rightarrow \infty} \int \varphi(x) dF(x).$$

Этот вид сходимости немного отличается от предыдущих. В частности при арифметических операциях могут не получиться те же арифметические операции в пределе. Также в 1) – 3) требуется, чтобы случайные величины были определены на одном вероятностном пространстве, а для слабой сходимости это не нужно, тут мы сразу работаем с индуцированным вероятностным пространством и распределениями случайных величин.

### 5. Сходимость по распределению:

$$\xi_n \Rightarrow \xi, \text{ если для } \forall x - \text{точка непрерывности } F(x) : F_n(x) \xrightarrow{n \rightarrow \infty} F(x).$$

Слабая сходимость эквивалентна сходимости по распределению, причём в одномерном случае лучше всего пользоваться сходимостью по распределению, а в многомерном слабой сходимостью. Большинство предельных теорем связаны с этими двумя сходимостями.



Рис. 6: Взаимосвязи между типами сходимости

## Теоремы, основанные на сходимости

### (Усиленный) закон больших чисел

Это не одна теорема, а большой класс различных теорем. Давайте приведём только одну из них.

**Теорема 2** (Колмогоров). Если случайные величины последовательности  $X_1, X_2, \dots$

независимы и одинаково распределены, то  $\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{n.б.} a$  при  $n \rightarrow \infty \iff \exists EX_i = a$ .

Именно этот факт служит обоснованием для того, чтобы использовать среднее арифметическое наблюдение как оценку математического ожидания. Также это применимо в схеме испытаний Бернулли как обоснование того, что относительная частота наступления события стремится к вероятности этого события.

Стоит отметить, что мы всё-таки делаем некоторое допущение: для того, чтобы написать сумму этих случайных величин, мы предполагаем, что они определены на одном вероятностном пространстве, и что все они зависят от одной и той же  $\omega$ . Поэтому элементарный исход в данном контексте стоит понимать не как то, что наступило в одном испытании, а как «общее состояние природы», в котором вы производите все испытания из выборки.

### Центральные предельные теоремы

**Теорема 3.** Если случайные величины последовательности  $X_1, X_2, \dots$  независимы и одинаково распределены, и существуют  $EX_i = a$  и  $DX_i = \sigma^2$ , то

$$P\left(\frac{\sum_{i=1}^n X_i - na}{\sigma\sqrt{n}} < x\right) \xrightarrow{n \rightarrow \infty} \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du,$$

причём сходимость равномерная.

Эта теорема служат обоснованием для того, чтобы в ошибках измерений считать погрешности нормально распределёнными. Тогда сумму факторов можно будет заменить на один с нормальным распределением.

С помощью неё также можно получить представление о том, с какой скоростью в законе больших чисел среднее арифметическое стремится к математическому ожиданию:

$$z_n = \left| \frac{\sum_{i=1}^n X_i}{n} - a \right|.$$

Будем подбирать такую неслучайную последовательность, зависящую от  $n$ , что при умножении или делении на эту последовательность будет получаться нетривиальный предел (в смысле сходимости по распределению), тогда это будет означать в какой-то степени их стремление к нулю будет совпадать:

$$r_n \sim \frac{1}{\sqrt{n}} : \quad \frac{z_n}{r_n} = \left| \frac{\sum_{i=1}^n X_i - na}{\sqrt{n}} \right|, \text{ тогда для } x > 0:$$

$$\mathbb{P} \left( \left| \frac{\sum_{i=1}^n X_i - na}{\sqrt{n}} \right| < x \right) = \mathbb{P} \left( -x < \frac{\sum_{i=1}^n X_i - na}{\sqrt{n}} < x \right) = \mathbb{P} \left( -\frac{x}{\sigma} < \frac{\sum_{i=1}^n X_i - na}{\sigma \sqrt{n}} < \frac{x}{\sigma} \right).$$

В силу центральной предельной теоремы мы получим следующее:

$$\mathbb{P} \left( -\frac{x}{\sigma} < \frac{\sum_{i=1}^n X_i - na}{\sigma \sqrt{n}} < \frac{x}{\sigma} \right) \xrightarrow{n \rightarrow \infty} \Phi \left( \frac{x}{\sigma} \right) - \Phi \left( \frac{-x}{\sigma} \right) = 2\Phi \left( \frac{x}{\sigma} \right) - 1.$$

Значит мы можем заключить, что скорость сходимости в законе больших чисел имеет порядок  $\mathcal{O} \left( \frac{1}{\sqrt{n}} \right)$ .

Существуют оценки скорости сходимости и в самой центральной предельной теореме. Рассмотрим, например, **неравенство Бэрри-Эссена**. Для него также потребуется наличие третьего абсолютного момента  $\mu_3 = \mathbb{E} |X_i - a|^3$ :

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left( \frac{\sum_{i=1}^n X_i - na}{\sigma \sqrt{n}} < x \right) - \Phi(x) \right| < \frac{c_0 \mu_3}{\sqrt{n} \sigma^3}. \quad (2)$$

Про  $c_0$  заведомо известно, что  $0.4 < c_0 < 0.5$ , но, конечно, существуют более точные оценки.

Это неравенство даёт при замене допредельного распределения предельным гарантированную оценку той ошибки, которую Вы совершаете при замене настоящего распределения суммы асимптотическим нормальным распределением.

Если требовать существования момента чуть больше, чем второй, но меньше, чем третий, то существуют аналоги этого неравенства с более медленной сходимостью по  $n$ . А если требовать момента больше, чем третий, то это не приведёт к увеличению скорости сходимости.

Стоит помнить, что нормальное распределение – это модельное распределение, которое в природе не встречается. Поэтому ошибки порой могут во много раз превосходить искомую вероятность и нужно всегда иметь ввиду ошибку при использовании центральной предельной теоремы.

С нормальным распределением также связано так называемое «правило трёх сигм»:  $P(|X - a| < 3\sigma) \approx 0.997$ . Это не совсем так. Настоящее «правило трёх сигм» должно быть определено так: если  $X$  имеет нормальное распределение и существует её дисперсия, то можно утверждать, что  $P(|X - a| < 3\sigma) \geq \frac{8}{9} \approx 0.89$ . То есть допускается ошибка в целых 10% вместо тысячных долей! Так что нужно быть осторожным.

Иногда бывает сложно рассчитывать на одинаковую распределённость случайных величин. Но есть утверждения, которые показывают, что иногда, когда случайные величины распределены по-разному, асимптотическая нормальность всё равно имеет место.

$X_1, X_2, \dots$  – независимые случайные величины. Предположим, что существуют  $a_i = EX_i$  и  $\sigma_i^2 = DX_i$  (причём  $0 < DX_i < \infty$ ).

$$A_n = \sum_{i=1}^n a_i, \quad B_n^2 = \sum_{i=1}^n \sigma_i^2,$$

$$F_n(x) = P\left(\frac{\sum_{i=1}^n X_i - A_n}{B_n} < x\right),$$

то есть по сути мы центрируем и нормируем функцию распределения. Теперь можем сформулировать обобщение центральной предельной теоремы.

**Теорема 4** (Линдеберг). Если случайные величины последовательности  $X_1, X_2, \dots$  независимы, существуют  $EX_i = a$  и  $DX_i = \sigma^2$ , а также выполнено условие Линдеберга

$$\frac{1}{B_n^2} \sum_{i=1}^n \int_{|x-a_i| > tB_n} |x - a_i|^2 dF_i(x) \xrightarrow{n \rightarrow \infty} 0 \text{ для } \forall t > 0, \quad (3)$$

то

$$\sup_x |F_n(x) - \Phi(x)| \xrightarrow{n \rightarrow \infty} 0,$$

$$\lim_{n \rightarrow \infty} \sup_{1 \leq i \leq n} \mathbf{P}(|X_i - a_i| > \varepsilon B_n) = 0 \text{ для } \forall \varepsilon > 0. \quad (4)$$

Условие 4 называют *условием равномерной предельной малости*.

Доказывать мы её не будем, но обсудим условие Линдеберга. Оно фактически означает, что вклады слагаемых в смысле дисперсии малы по сравнению со всей суммой. Пусть  $\sigma_k^2$  – дисперсия  $X_k$ . Убедимся в том, что величина  $\frac{\sigma_k^2}{B_n^2}$  бесконечно убывает:

$$\sigma_k^2 = \int |x - a_k|^2 dF_k(x) = \int_{|x-a_k| > tB_n} |x - a_k|^2 dF_k(x) + \int_{|x-a_k| \leq tB_n} |x - a_k|^2 dF_k(x).$$

Подставим это в  $\frac{\sigma_k^2}{B_n^2}$ :

$$\begin{aligned} \frac{\int_{|x-a_k| > tB_n} |x - a_k|^2 dF_k(x) + \int_{|x-a_k| \leq tB_n} |x - a_k|^2 dF_k(x)}{B_n^2} &\leq \\ &\leq \frac{\int_{|x-a_k| > tB_n} |x - a_k|^2 dF_k(x) + \int_{|x-a_k| \leq tB_n} t^2 B_n^2 dF_k(x)}{B_n^2} \leq \\ &\leq \frac{\int_{|x-a_k| > tB_n} |x - a_k|^2 dF_k(x) + t^2 B_n^2}{B_n^2}. \end{aligned}$$

Пусть  $t$  такое, что  $t^2 < \frac{\varepsilon}{2}$ :

$$\frac{\int_{|x-a_k| > tB_n} |x - a_k|^2 dF_k(x) + t^2 B_n^2}{B_n^2} \leq \underbrace{\frac{\sum_{k=1}^n \int_{|x-a_k| > tB_n} |x - a_k|^2 dF_k(x)}{B_n^2}}_{< \frac{\varepsilon}{2}} + \frac{\varepsilon}{2} < \varepsilon.$$

Может сложиться впечатление, что условие Линдеберга слишком сильное. На самом деле это не так.

**Теорема 5** (Линдеберг-Феллер). Если случайные величины последовательности  $X_1, X_2, \dots$  независимы, существуют  $\mathbf{E}X_i = a$  и  $\mathbf{D}X_i = \sigma^2$ , а также выполнено условие 4 из теоремы 4 и  $\sup_x |F_n(x) - \Phi(x)| \xrightarrow{n \rightarrow \infty} 0$ , то будет выполнено и условие Линдеберга (3).

## Лекция 4

### Центральные предельные теоремы (продолжение)

Давайте покажем, что из условия Линдеберга следует условие равномерной предельной малости:

$$\begin{aligned} P(|X_i - a_i| > \varepsilon B_n) \varepsilon^2 B_n^2 &= \int_{|x-a_i|>\varepsilon B_n} \varepsilon^2 B_n^2 dF_i(x) \leq \\ &\leq \int_{|x-a_i|>\varepsilon B_n} |x - a_i|^2 dF_i(x) \leq \sum_{i=1}^n \int_{|x-a_i|>\varepsilon B_n} |x - a_i|^2 dF_i(x). \end{aligned}$$

Из этого получаем:

$$P(|X_i - a_i| > \varepsilon B_n) \leq \frac{\sum_{i=1}^n \int_{|x-a_i|>\varepsilon B_n} |x - a_i|^2 dF_i(x)}{\varepsilon^2 B_n^2} \xrightarrow{n \rightarrow \infty} 0,$$

как раз по условию Линдеберга. Значит из (3) следует (4).

**Теорема 6 (Ляпунов).** Если случайные величины последовательности  $X_1, X_2, \dots$  независимы, существуют третьи моменты  $\beta_i^3 = E|X_i - a_i|^3$  и  $M_n^3 = \sum_{i=1}^n \beta_i^3$ , а также

$$\frac{M_n^3}{\beta_n^3} \xrightarrow{n \rightarrow \infty} 0,$$

то выполняются все условия теоремы 3.

То есть это обобщение центральной предельной теоремы для одинаково распределённых случайных величин на различно распределённые. И если выполнены условия теоремы Ляпунова, то выполнено и условие Линдеберга:

$$\int_{|x-a_i|>tB_n} |x - a_i|^2 dF_i(x) \leq \int_{|x-a_i|>tB_n} |x - a_i|^2 \frac{|x - a_i|}{tB_n} dF_i(x) \leq \frac{1}{tB_n} \int_{-\infty}^{+\infty} |x - a_i|^3 dF_i(x).$$

### Репрезентативность выборки

Вернёмся к ситуации, когда случайные величины распределены одинаково. Сколько наблюдений необходимо для репрезентативной выборки?

**Пример.**  $X_1, \dots, X_n$  — независимые и одинаково распределённые случайные величины. Тогда  $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$  является хорошей оценкой математического ожидания.

Что можно сказать о точности и надёжности такой оценки? Применим центральную предельную теорему для следующей вероятности:

$$\begin{aligned} P(|\bar{X} - a| < \varepsilon) &= P(-\varepsilon < \bar{X} - a < \varepsilon) = \\ &= P\left(-\frac{\varepsilon\sqrt{n}}{\sigma} < \frac{\sqrt{n}(\bar{X} - a)}{\sigma} < \frac{\varepsilon\sqrt{n}}{\sigma}\right) \approx \\ &\approx \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1. \end{aligned}$$

Потребуем, чтобы надёжность была не меньше некоего уровня  $\gamma$ :

$$2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1 = \gamma, \quad \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) = \frac{1+\gamma}{2}.$$

Пусть  $u_{\frac{1+\gamma}{2}}$  – квантиль нормального распределения порядка  $\frac{1+\gamma}{2}$ , тогда:

$$\varepsilon = \frac{u_{\frac{1+\gamma}{2}} \sigma}{\sqrt{n}}.$$

Опять же, мы здесь совершаем ошибку, заменяя истинные вероятности предельными, и получаем более оптимистичные точности, чем на самом деле. А какую гарантированную точность можно получить? Предположим, что существует третий момент, тогда возьмём величину из неравенства Бэрри-Эссеена (см. (2)) и оценим её:

$$L_3 = \frac{c_0 \mu^3}{\sqrt{n} \sigma^3} > \frac{0,4}{\sqrt{n}}.$$

Теперь вернёмся к прошлым выкладкам и используем эту величину:

$$P(|\bar{X} - a| < \varepsilon) \geq 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1 - 2L_3.$$

Тогда мы получим  $u_{\frac{1+\gamma}{2}+L_3}$  – квантиль нормального распределения порядка  $\frac{1+\gamma}{2} + L_3$  и более пессимистичную оценку точности:

$$\varepsilon = \frac{u_{\frac{1+\gamma}{2}+L_3} \sigma}{\sqrt{n}}$$

и если

$$\frac{1+\gamma}{2} + L_3 > 1, \text{ то есть } n < \frac{0,64}{(1-\gamma)^2},$$

то вообще нет смысла говорить о точности, так как квантиль будет порядка больше 1, а такого квантиля очевидно не существует (иногда и порядка 1 не существует).

О репрезентативности выборки можно отчасти что-то сказать для непрерывных распределений. Построим **эмпирическую функцию распределения**:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i < x).$$

Несмотря на то, что настоящая функция распределения может быть неизвестна, мы можем рассмотреть следующее:

$$D_n = \sup_x |\hat{F}_n(x) - F(x)|.$$

Это известная *статистика Колмогорова*, которая применяется при проверке гипотезы о равенстве функции распределения какому-то конкретному распределению. По теореме Колмогорова:

$$P(D_n < \varepsilon) = P(\sqrt{n}D_n < \varepsilon\sqrt{n}) \approx K(\varepsilon\sqrt{n}),$$

где  $K(\cdot)$  – *функция распределения Колмогорова*. Мы снова совершаем ошибку при переходе к предельному распределению, но здесь она менее существенна: при  $n > 20$  её можно игнорировать, к тому же при непрерывном  $F(x)$ , то распределение  $D_n$  не зависит от  $X_i$ , оно известно и при малых  $n$  можно по таблице найти функцию распределения для конкретного  $n$ .

Чтобы достичь какой-то точности с заданной надёжностью  $\gamma$ , нужно снова решить уравнение и получить:

$$\varepsilon = \frac{K_\gamma}{\sqrt{n}}, \quad n = \frac{K_\gamma^2}{\varepsilon^2},$$

где  $K_\gamma$  – квантиль.

Получается, что для заданного уровня надёжности  $\gamma$ , достичь репрезентативности можно, произведя минимум  $\frac{K_\gamma^2}{\varepsilon^2}$  наблюдений.

## Теорема Пуассона

Рассмотрим схему Бернулли.  $X_1, \dots, X_n$  – независимые одинаково распределённые случайные величины:

$$X_i \sim \begin{cases} 1, & \text{с вероятностью } p, \\ 0, & \text{с вероятностью } 1 - p; \end{cases} \quad S_n = \sum_{i=1}^n X_i, \quad E S_n = np, \quad D S_n = np(1 - p).$$

Если  $D S_n$  велико, то хорошо применима центральная предельная теорема, но в других случаях она может дать недопустимую погрешность и там может быть более уместно применение теоремы Пуассона. Перед тем, как разбирать её, мы сформулируем другую теорему (не предельную), которая даст гарантированные оценки для пуассоновской аппроксимации.



**Теорема 7.** Пусть  $X_1, \dots, X_n$  – независимые одинаково распределённые случайные величины, которые удовлетворяют схеме Бернулли, а также известно, что  $k-1 \leq \frac{n}{4}$  и  $p \leq \frac{1}{4}$ , тогда

$$P(S_n = k) = \frac{e^{-\lambda + r_n(k)} \lambda^k}{k!},$$

где

$$\lambda = np \quad \text{и} \quad \frac{\lambda k}{n} + 2 \ln \left( \frac{4}{3} \right) \frac{k(1-k)}{n} - \frac{2\lambda^2}{3n} \leq r_n(k) \leq \frac{\lambda k}{n} + \frac{k(1-k)}{2n}.$$

То есть фактически  $r_n(k) = \mathcal{O}\left(\frac{1}{n}\right)$ .

*Доказательство.*

$$\begin{aligned} P(S_n = k) &= C_n^k p^k (1-p)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} p^k (1-p)^{n-k} = \\ &= \frac{(np)^k}{k!} e^{-np} \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) (1-p)^{n-k} e^{np} = \frac{e^{-\lambda + r_n(k)} \lambda^k}{k!}, \end{aligned}$$

$$\text{где } r_n(k) = \ln \left[ \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) (1-p)^{n-k} e^{np} \right].$$

Воспользуемся тем, что при  $0 < x < 1$  верно неравенство  $\ln(1-x) < -x$ :

$$r_n(k) \leq - \sum_{i=1}^{k-1} \frac{1}{n} - (n-k)p + n = \frac{k(1-k)}{2n} + kp = \frac{k(1-k)}{2n} + \frac{k\lambda}{n},$$

где  $\lambda = np$ .

Таким образом мы вывели оценку сверху. Оценка снизу получается применением условий теоремы, что  $k-1 \leq \frac{n}{4}$  и  $p \leq \frac{1}{4}$ , и разложением логарифма в ряд до второй степени. ■

**Схема серий**  $\{X_{n,i}\}$ . Вместо строки случайных величин мы работаем со следующей «матрицей» случайных величин:

$$\begin{array}{c} X_{1,1} \\ X_{2,1}, X_{2,2} \\ \vdots \\ X_{n,1}, \dots, X_{n,n} \end{array} \quad X_{n,i} \sim \begin{cases} 1, & \text{с вероятностью } p_n; \\ 0, & \text{с вероятностью } 1 - p_n. \end{cases}$$

Каждая строка называется **серией**. В каждой серии случайные величины независимы и одинаково распределены, но у разных серий разное распределение.

Схема нарастающих сумм в центральной предельной теореме – частный случай схемы серий:

$$Y_{n,i} = \frac{X_i - a}{\sqrt{n}\sigma}.$$

**Теорема 8** (Пуассон). Пусть  $\{X_{n,i}\}$  – схема серий и  $S_n = \sum_{i=1}^n X_{n,i}$ , а также известно, что  $np_n \xrightarrow{n \rightarrow \infty} \lambda$ , тогда

$$\lim_{n \rightarrow \infty} P(S_n = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{для } k = 0, 1, \dots$$

Эту теорему, которую также называют «законом малых чисел», мы доказывать не будем, но позже докажем более общую теорему для различно распределённых величин.

## Устойчивое распределение

Функция распределения  $G(x)$  и соответствующая её характеристическая функция  $g(t)$  называются **устойчивыми**, если для  $\forall a_1 > 0, a_2 > 0 : \exists a > 0, b \in \mathbb{R}$  такие, что

$$g(a_1 t) g(a_2 t) = g(at) e^{it\sigma},$$

или для  $\forall a_1 > 0, a_2 > 0, b_1 \in \mathbb{R}, b_2 \in \mathbb{R} : \exists a > 0, b \in \mathbb{R}$  такие, что

$$G(a_1 x + b_1) * G(a_2 x + b_2) = G(ax + b),$$

где  $*$  – оператор свёртки:  $F * G(x) = \int_{-\infty}^{+\infty} F(x-y) dG(y)$ .

Можно заметить, что свёртка в определении использована неправильно. Правильная запись будет выглядеть так:

$$\left. \begin{aligned} G_1(x) &= G(a_1 x + b_1) \\ G_2(x) &= G(a_2 x + b_2) \end{aligned} \right\} \implies G_1(x) * G_2(x) = G(ax + b). \quad (5)$$

Так менее наглядно, но *правильно*.

Характеристическую функцию любого устойчивого распределения можно написать в явном виде, и она зависит от четырёх параметров:

$$g_\alpha(t) = \exp \left\{ iat - c|t|^\alpha \left( 1 + ib \frac{t}{|t|} Q(t, \alpha) \right) \right\}, \quad (6)$$

где

$$Q(t, \alpha) = \begin{cases} \operatorname{tg} \left( \frac{\pi \alpha}{2} \right), & \text{если } \alpha = 1, \\ \frac{2}{\pi} \ln |t|, & \text{если } \alpha \neq 1; \end{cases} \quad \alpha \in (0, 2].$$

При  $\alpha = 2$  мы получим нормальное распределение, при  $\alpha = 1$  распределение Коши, при  $\alpha = \frac{1}{2}$  распределение Леви. И только для них известен явный вид плотности.

Устойчивые распределения всегда абсолютно непрерывны (кроме вырожденного случая) и для них верно:

$$G_{\alpha}(-x) + 1 - G_{\alpha}(x) \underset{x \rightarrow +\infty}{\sim} \frac{c}{x^{\alpha}},$$

$$\begin{aligned} \mathbb{E}|X_{\alpha}|^{\gamma} &< \infty, \text{ если } \gamma < \alpha; \\ &= \infty, \text{ если } \gamma \geq \alpha. \end{aligned}$$

Таким образом для всех устойчивых распределений кроме нормального не существует дисперсии (а иногда и математического ожидания).

**Теорема 9 (Леви).** Пусть  $X_1, \dots, X_n$  – независимые одинаково распределённые случайные величины. Функция распределения  $G(x)$  может быть предельной для  $\frac{\sum_{i=1}^n X_i - A_n}{B_n}$  при некотором выборе  $\{A_n\}$  и  $\{B_n\}$  тогда и только тогда, когда  $G(x)$  устойчива.

Центральная предельная теорема – это частный случай теоремы Леви, когда существует дисперсия. Если у слагаемых дисперсия не существует, то получаются устойчивые законы.

## Теорема Пуассона (общий случай)

Характеристическая функция  $f(t)$  называется **безгранично делимой**, если для  $\forall n \in \mathbb{N} : \exists f_n(t)$  – такая характеристическая функция, что  $f(t) = (f_n(t))^n$ . То есть если вероятностное пространство позволяет описать это определение в терминах случайных величин, то безгранично делимая случайная величина для  $\forall n \in \mathbb{N}$  представима в виде суммы  $n$  независимых одинаково распределённых случайных величин.

Если характеристическая функция безгранично делима, то она не обращается в ноль при любом  $t$ . Это свойство позволяет избавиться от деления на ноль и помогает при оценках избавиться от неопределённости.

Вернёмся к схеме серий. Только теперь мы откажемся от квадратной матрицы и совпадения распределений внутри серий:

$$\begin{aligned} &X_{1,1} \\ &\vdots \\ &X_{n,1}, \dots, X_{n,m} \end{aligned}$$

Также потребуем, чтобы

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} \sup_{1 \leq i \leq m_n} \mathbb{P}(|X_{n,i}| > \varepsilon) = 0. \quad (7)$$

Это аналог условия равномерной предельной малости (4).

**Теорема 10** (Хинчин). Пусть  $\{X_{n,i}\}$  – схема серий, выполнено условие (7) и  $S_n = \sum_{i=1}^{m_n} X_{n,i}$ . Тогда  $F(x)$  может быть предельной для  $S_n$  тогда и только тогда, когда  $f(t)$  безгранично делима.

Теперь докажем теорему Пуассона для общего случая.

**Теорема 11.** Пусть  $\{X_{n,i}\}$  – схема серий:

$$X_{n,i} \sim \begin{cases} 1, & \text{с вероятностью } p_{n,i}, \\ 0, & \text{с вероятностью } 1 - p_{n,i}; \end{cases} \quad \max_{1 \leq i \leq m_n} p_{n,i} \xrightarrow{n \rightarrow \infty} 0.$$

А также пусть

$$\sum_{i=1}^{m_n} p_{n,i} \xrightarrow{n \rightarrow \infty} \lambda.$$

Тогда

$$\lim_{n \rightarrow \infty} P(S_n = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{для } k = 0, 1, \dots$$

*Доказательство.*

$$\psi_{n,i}(s) = \mathbb{E} s^{X_{n,i}} = 1 - p_{n,i} + s p_{n,i} = 1 + p_{n,i}(s - 1),$$

$$\psi_{S_n}(s) = \prod_{i=1}^{m_n} \psi_{n,i}(s),$$

$$\ln \psi_{S_n}(s) = \sum_{i=1}^{m_n} \ln \psi_{n,i}(s) = \sum_{i=1}^{m_n} \ln (1 + p_{n,i}(s - 1)).$$

Заметим, что

$$\sum_{i=1}^{m_n} p_{n,i}^2 \leq \underbrace{\max_i p_{n,i}}_{\rightarrow 0} \underbrace{\sum_{i=1}^{m_n} p_{n,i}}_{\rightarrow \lambda} \xrightarrow{n \rightarrow \infty} 0.$$

То есть мы получаем, что

$$\sum_{i=1}^{m_n} \ln (1 + p_{n,i}(s - 1)) \underset{n \rightarrow \infty}{\sim} \underbrace{\sum_{i=1}^{m_n} p_{n,i}(s - 1)}_{\rightarrow \lambda} \xrightarrow{n \rightarrow \infty} \lambda(s - 1)$$

И по итогу

$$\psi_{S_n}(s) \xrightarrow{n \rightarrow \infty} e^{\lambda(s-1)} - \text{производящая функция распределения Пуассона.}$$

В силу взаимнооднозначности и взаимонепрерывности производящей функции и неотрицательного целочисленного распределения можем заключить, что распределение является пуассоновским. ■

## Лекция 5

**Пример.** Есть два завода  $A$  и  $B$ , которые производят большое количество одних и тех же электронных приборов. Они расположены в удалённых друг от друга местах и работают независимо. За год на заводе  $A$  было  $n_A$  бракованных приборов, а на заводе  $B$  их было  $n_B$  штук. Известно, что  $n_A > n_B$ . Вопрос: говорит ли это о существовании фактора на заводе  $A$ , которое приводит к большему количеству брака, или это случайность? Ответ зависит от того, что это за числа, насколько одно больше другого и так далее. Но попробуем формализовать задачу и попробовать использовать какие-нибудь теоремы.

Заводы производят много приборов, значит можно применить предельную аппроксимацию. Кроме того, разумно предположить, что количество брака по отношению к общему количеству устройств мало и вероятность появления брака в приборе независима от остальных приборов. С такими сведениями мы можем предположить, что удовлетворяются условия теоремы Пуассона. Тогда обозначим через  $N_A$  и  $N_B$  случайные величины, которые дают за исследуемый период времени количество бракованных устройств на заводах.

$$N_A = X_1 + \dots + X_N, \quad X_i \sim \begin{cases} 1, & \text{с вероятностью } p_A, \\ 0, & \text{с вероятностью } 1 - p_A; \end{cases}$$

$$N_B = Y_1 + \dots + Y_N, \quad Y_i \sim \begin{cases} 1, & \text{с вероятностью } p_B, \\ 0, & \text{с вероятностью } 1 - p_B. \end{cases}$$

Применяем теорему Пуассона:

$$N_A \sim \Pi(\lambda_A), \quad \lambda_A = EN_A,$$

$$N_B \sim \Pi(\lambda_B), \quad \lambda_B = EN_B.$$

Рассмотрим условную вероятность:

$$\begin{aligned} P(N_A = n_A | N_A + N_B = n_A + n_B) &= \frac{P(N_A = n_A, N_A + N_B = n_A + n_B)}{P(N_A + N_B = n_A + n_B)} = \\ &= \frac{P(N_A = n_A, N_B = n_B)}{P(N_A + N_B = n_A + n_B)} = \frac{e^{-\lambda_A} \lambda_A^{n_A}}{n_A!} \frac{e^{-\lambda_B} \lambda_B^{n_B}}{n_B!} = \frac{e^{-(\lambda_A + \lambda_B)} (\lambda_A + \lambda_B)^{n_A + n_B}}{(n_A + n_B)!} = \\ &= \frac{(n_A + n_B)!}{n_A! n_B!} \left( \frac{\lambda_A}{\lambda_A + \lambda_B} \right)^{n_A} \left( \frac{\lambda_B}{\lambda_A + \lambda_B} \right)^{n_B} \sim Bi \left( \frac{\lambda_A}{\lambda_A + \lambda_B}, n_A + n_B \right). \end{aligned}$$

Наша задача – убедиться будет ли  $\lambda_A$  больше, чем  $\lambda_B$ . Нам не важно, что проверять:  $\lambda_A = \lambda_B$  или  $\lambda_A \neq \lambda_B$ . Но с точки зрения выкладок гораздо проще проверять первый случай. То есть мы будем проверять, является ли при заданных  $n_A$  и

$n_B$  условная вероятность биномиальной с параметрами  $\frac{1}{2}$  и  $n_A + n_B$ . Допустим, что  $\lambda_A = \lambda_B$ :

$$\begin{aligned} P(N_A \geq n_A | N_A + N_B = n_A + n_B) &= \\ &= \left(\frac{1}{2}\right)^{n_A+n_B} (n_A + n_B)! \sum_{k=n_A}^{n_A+n_B} \frac{1}{k! (n_A + n_B - k)!} \approx \\ &\approx 1 - \Phi\left(\frac{n_A - \frac{1}{2}(n_A + n_B)}{\sqrt{\frac{1}{4}(n_A + n_B)}}\right). \end{aligned}$$

Последний переход выполнен с помощью предельной теоремы Муавра-Лапласа, поэтому не забываем про возникновение ошибки. Возьмём к примеру  $n_A = 125$  и  $n_B = 100$ :

$$1 - \Phi\left(\frac{n_A - \frac{1}{2}(n_A + n_B)}{\sqrt{\frac{1}{4}(n_A + n_B)}}\right) < 0,05.$$

Значит что-то не так на заводе  $A$ . Но понятно, что ответ зависит не только от того, насколько  $n_A$  превосходит  $n_B$ .

## Теория информации

### Определение количества информации

Информацию очень трудно формализовать, потому что возникает очень много неоднозначных вопросов: что такое информация, как её мерить, в каких единицах? Например, есть лист с текстом. Сколько в нём информации? Это зависит от того, что на нём написано и насколько это влияет на остальные процессы, несёт ли это в себе новое знание, насколько это помогает в нахождении новых знаний и так далее. А если два листа с текстом? Значит ли это, что информации стало в два раза больше? Ну понятно, что нет, так как на втором листе может быть написано то же самое, что и на первом, но другими словами. Поэтому можно сделать вывод: количеством символов информацию мерить нельзя. Но тогда как?

Один из подходов предложил Шеннон в середине XX века. Он связал информацию с вероятностью. Допустим, что у нас есть события  $A$  и  $B$  и одно событие несёт информацию о другом, тогда:

$$I(A|B) = \log_a \frac{P(A|B)}{P(A)}.$$

Так определяется **количество информации, которая содержится в событии  $B$  относительно события  $A$ .**

Но надо обосновать это определение. Если  $A$  и  $B$  независимы, то  $P(A|B) = P(A)$  и  $I(A|B) = 0$ , что вполне разумно. Если  $A = B$ , то  $I(A|A) = I(A) = -\log_a P(A)$  – **количество информации, которая содержится в событии  $A$** . То есть, если вероятность события мала, то информация велика, что тоже вполне разумно. Потому что, если Вы пытаетесь получить информацию о частом событии, то много нового вы не узнаете, а вот редкое событие несёт за собой много новой информации. Также, если  $A$  и  $B$  независимы, то  $I(AB) = -\log_a P(AB) = -\log_a P(A) - \log_a P(B) = I(A) + I(B)$ .

Но что это за основание  $a$ ? Оно большой роли не играет, но требуется, чтобы  $a > 1$ . Как правило применяются  $a = 2$  (то есть единица информации содержится в событии, вероятность которого  $\frac{1}{2}$ ), тогда говорят, что информация измеряется в **битах**. Также используется  $a = e$ , тогда говорят, что информация измеряется в **натах**.

Можно построить пример с такими событиями  $A$  и  $B$ , когда  $a$  будет отрицательным. Как правило это когда наступление события  $B$  снижает вероятность наступления события  $A$ .

## Энтропия

Рассмотрим эксперимент  $E$  с непересекающимся множеством событий  $A_1, \dots, A_n$ :

$$E : \begin{matrix} A_1, \dots, A_n \\ p_1, \dots, p_n \end{matrix} \quad \sum_{i=1}^n p_i = 1, \quad p_i \geq 0,$$

случайную величину  $Q(E)$  и её математическое ожидание  $H(E)$ :

$$Q(E) : \begin{matrix} I(A_1), \dots, I(A_n) \\ p_1, \dots, p_n \end{matrix} \quad H(E) = M(Q(E)) = - \sum_{i=1}^n p_i \log_a p_i.$$

$H(E)$  называется **энтропией** эксперимента  $E$ . Это средняя информация, которую вы получаете, проведя этот эксперимент. Как правило о ней говорится, как о мере неопределённости.

В дальнейших выкладках будет считаться, что  $a = e$  и информация измеряется в натах.

**Теорема 12.** Для любого эксперимента  $E$  с  $n$  исходами верно, что:

- 1)  $H(E) \geq 0$ , причём  $H(E) = 0 \iff \exists i^* \in \{1, \dots, n\} : p_{i^*} = 1, \forall i \neq i^* : p_i = 0$ ;
- 2) пусть  $E_0$  – эксперимент с  $n$  исходами и  $p_i = \frac{1}{n}$ , тогда  $H(E_0) \geq H(E)$ ;

- 3) пусть  $E_1$  – эксперимент, полученный из  $E$  объединением некоторых  $A_i$  и  $A_j$ ,  
а  $E_2$  – эксперимент с двумя исходами  $A_i$  и  $A_j$  с вероятностями  $\frac{p_i}{p_i+p_j}$  и  $\frac{p_j}{p_i+p_j}$   
соответственно, тогда  $H(E) = H(E_1) + (p_i + p_j)H(E_2)$ ;
- 4)  $H(E)$  зависит от  $p_1, \dots, p_n$  (и не зависит от  $A_1, \dots, A_n$ ) и с договорённостью,  
что  $p_i \ln p_i = 0$  при  $p_i = 0$  (так как  $\lim_{p_i \rightarrow 0} p_i \ln p_i = 0$  и мы можем доопределить),  
 $H(p_1, \dots, p_n)$  непрерывна и симметрична.

Доказательство. 1)  $f(p) = -p \ln p \geq 0$  при  $p \in [0, 1]$ , более того  $-p \ln p = 0 \iff p = 0$  или  $p = 1$ , значит  $p_i = 0$  или  $p_i = 1$  при  $\forall i \in \{1, \dots, n\}$ , а так как  $\sum_{i=1}^n p_i = 1$ , то  $\exists i^* : p_{i^*} = 1$ , а  $\forall i \neq i^* : p_i = 0$ .

2)  $g(p) = p \ln p$ ,  $g'(p) = \ln p + 1$ ,  $g''(p) = \frac{1}{p} > 0$  при  $p \in (0, 1]$ , значит  $g(p)$  выпукла  
вниз, следовательно для  $\forall \alpha_1, \dots, \alpha_n : \alpha_i \geq 0$ ,  $\sum_{i=1}^n \alpha_i = 1$  и  $\forall y_1, \dots, y_n$  из той области,  
где  $g(p)$  выпукла вниз:

$$g\left(\sum_{i=1}^n \alpha_i y_i\right) \leq \sum_{i=1}^n \alpha_i g(y_i).$$

Применим это свойство с  $\alpha_i = \frac{1}{n}$  и  $y_i = p_i$ :

$$H(E_0) = -\ln \frac{1}{n} = -\frac{n}{n} \ln \frac{1}{n} = -n \left(\sum_{i=1}^n \frac{p_i}{n}\right) \ln \left(\frac{\sum_{i=1}^n p_i}{n}\right) \geq -\sum_{i=1}^n p_i \ln p_i = H(E).$$

3) Убедимся, что  $H(E) \geq H(E_1)$ :

$$\begin{aligned} H(E_1) &= -\sum_{\substack{k=1 \\ k \neq i,j}}^n p_k \ln p_k - (p_i + p_j) \ln(p_i + p_j) = \\ &= -\sum_{\substack{k=1 \\ k \neq i,j}}^n p_k \ln p_k - p_i \ln(p_i + p_j) - p_j \ln(p_i + p_j) \leq \\ &\leq -\sum_{\substack{k=1 \\ k \neq i,j}}^n p_k \ln p_k - p_i \ln p_i - p_j \ln p_j = H(E). \end{aligned}$$



Рассмотрим следующее:

$$\begin{aligned} H(E) - H(E_1) &= -p_i \ln p_i - p_j \ln p_j + (p_i + p_j) \ln(p_i + p_j) = \\ &= -p_i \ln p_i - p_j \ln p_j + p_i \ln(p_i + p_j) + p_j \ln(p_i + p_j) = \\ &= (p_i + p_j) \left( -\frac{p_i}{p_i + p_j} \ln p_i - \frac{p_j}{p_i + p_j} \ln p_j + \frac{p_i}{p_i + p_j} \ln(p_i + p_j) + \frac{p_j}{p_i + p_j} \ln(p_i + p_j) \right) = \\ &= (p_i + p_j) \left( -\frac{p_i}{p_i + p_j} \ln \frac{p_i}{p_i + p_j} - \frac{p_j}{p_i + p_j} \ln \frac{p_j}{p_i + p_j} \right) = (p_i + p_j) H(E_2). \end{aligned}$$

4) В принципе это очевидно. ■

Таким образом, это обосновывает, что введённая нами энтропия может являться мерой неопределённости эксперимента. Более того, справедливо и обратное утверждение.

**Теорема 13** (Фадеев). Пусть  $H(p_1, \dots, p_n)$  – функционал от  $p_1, \dots, p_n$ , который удовлетворяет условиям теоремы 12. Тогда  $H(p_1, \dots, p_n) = -\sum_{i=1}^n p_i \ln p_i$ .

Естественным образом возникает потребность перенести определение энтропии на случайные величины. Пусть  $\xi$  – случайная величина с конечным числом значений:

$$\begin{aligned} \xi : & x_1, \dots, x_n \\ & p_1, \dots, p_n \end{aligned}$$

**Энтропией случайной величины  $\xi$**  будет являться  $H(\xi) = -\sum_{i=1}^n p_i \ln p_i$ .

Можно немного преобразить выражение, введя функцию  $p(x)$ :

$$p(x) = \begin{cases} p_i, & x = x_i, \quad i = 1, \dots, n; \\ 0, & \text{иначе.} \end{cases}$$

Тогда  $H(\xi) = -E \ln p(\xi)$ . Введённая нами  $p(x)$  – аналог плотности для дискретных случайных величин. Напрашивается следующее определение энтропии для абсолютно непрерывных величин:  $H(\xi) = -E \ln p(\xi)$ , где  $p(\xi)$  уже непосредственно плотность. Но оказывается, что это совсем другой объект и имеет другой смысл на бесконечности.

## Лекция 6

### Дифференциальная энтропия

Пусть  $X$  – случайная величина. Тогда существует  $\{X_n\}$  – такая последовательность простых случайных величин, что  $X_n \xrightarrow[n \rightarrow \infty]{\text{п.в.}} X$ . Идея заключается в том, что для абсолютно непрерывной случайной величины можно построить последовательность простых случайных величин и рассмотреть предел.

Разобьём множество значений  $X$  на отрезки  $\Delta_i$  длиной  $\delta$ . Пусть  $p(x)$  непрерывна,  $X$  имеет конечный носитель,  $X_\delta$  – такая последовательность простых случайных величин, что  $X_\delta \xrightarrow[\delta \rightarrow 0]{\text{п.в.}} X$ . То есть  $X_\delta: x_i \in \Delta_i, P(X_\delta = x_i) = P(X \in \Delta_i)$ . Значит она сходится и по распределению:  $X_\delta \Rightarrow X$ . Проверим верно ли, что  $H(X_\delta) \rightarrow H(X)$ . Воспользуемся теоремой о среднем значении:

$$\begin{aligned} H(X_\delta) &= - \sum_{i=1}^n P(X \in \Delta_i) \ln P(X \in \Delta_i) = - \sum_{i=1}^n p(X_i^*) \delta \ln (p(X_i^*) \delta) = \\ &= - \sum_{i=1}^n p(X_i^*) \delta \ln p(X_i^*) - \sum_{i=1}^n p(X_i^*) \delta \ln \delta = - \sum_{i=1}^n p(X_i^*) \delta \ln p(X_i^*) - \ln \sum_{i=1}^n P(X \in \Delta_i) = \\ &= - \sum_{i=1}^n p(X_i^*) \delta \ln p(X_i^*) - \ln \delta. \end{aligned}$$

Этот возникший логарифм можно трактовать как информацию, которая содержится в событии, вероятность которого  $\delta$ . И рост энтропии при увеличении количества исходов на него обусловлен растущим количеством исходов.

При  $\delta \rightarrow 0$  из-за логарифма мы получим бесконечность. Но благодаря тому, что у нас сходимость почти всюду, мы можем компенсировать логарифм независимо от выбора последовательности:

$$H(X) = \lim_{\delta \rightarrow 0} (H(X_\delta) + \ln \delta) = - \int p(x) \ln p(x) dx.$$

Так мы получаем конечную осмысленную величину, называемую **дифференциальной энтропией**.

К сожалению, так определённая характеристика обладает не всеми свойствами обычной энтропии (например, она может быть отрицательной). Её можно использовать только в относительном смысле как меру неопределённости эксперимента, в основе которого абсолютно непрерывное распределение, и по неопределённости можно сравнивать между собой различные распределения. Более того, сравнивать их между собой можно только в пределах некоторого класса распределения. Иначе надо накладывать дополнительные ограничения.

Мы рассмотрим три класса распределений и найдём для каждого из них максимально неопределённое распределение в смысле дифференциальной энтропии. В этом нам помогут некоторые методы вариационного анализа.

Пусть есть функция  $F(x, p(x))$ , где  $x$  – независимая переменная, а  $p(x)$  – функция от  $x$ . Также заданы  $\varphi_i(x, p(x))$  и числа  $a_i$  при  $i = \overline{1, n}$ . Нужно найти максимум функционала  $J(p) = \int_a^b F(x, p(x)) dx$  на  $[a, b]$  при условиях  $\int_a^b \varphi_i(x, p(x)) dx = a_i$ ,  $i = \overline{1, n}$ .

В целом это похоже на задачу на поиск условного экстремума, которая решается методом Лагранжа. И эта задача решается похожим методом через следующее дифференциальное уравнение:

$$0 = \frac{\partial F}{\partial p} + \lambda_1 \frac{\partial \varphi_1}{\partial p} + \dots + \lambda_n \frac{\partial \varphi_n}{\partial p}, \quad \text{где } \lambda_1, \dots, \lambda_n \text{ – неопределённые коэффициенты.}$$

А  $\lambda_1, \dots, \lambda_n$  мы находим подстановкой решения в заданные условия.

**Теорема 14.** 1. Пусть  $X \sim U([-a, a])$ . Тогда  $H(X) \geq H(Y)$ , где  $Y$  – такая любая абсолютно непрерывная случайная величина, что  $P(|Y| \leq a) = 1$ .

2. Пусть  $X \sim P(\lambda)$ . Тогда  $H(X) \geq H(Y)$ , где  $Y$  – такая любая абсолютно непрерывная случайная величина, что  $P(Y \geq 0) = 1$  и  $EY = \frac{1}{\lambda}$ .

3. Пусть  $X \sim N(a, \sigma^2)$ . Тогда  $H(X) \geq H(Y)$ , где  $Y$  – такая любая абсолютно непрерывная случайная величина, что  $EY = a$  и  $DY = \sigma^2$ .

*Доказательство.* 1.  $F(x, p(x)) = p(x) \ln p(x)$ ,  $\varphi_1(x, p(x)) = p(x)$ . Нужно максимизировать  $J(p) = - \int_{-a}^a p(x) \ln p(x) dx$  при условии  $\int_{-a}^a \varphi_1(x, p(x)) dx = 1$ . Запишем уравнение:

$$-(1 + \ln p) + \lambda_1 = 0 \implies p(x) = e^{\lambda_1 - 1} - \text{константа на } [-a, a].$$

Подставляем в условие и получаем  $p(x) = \frac{1}{2a}$  при  $x \in [-a, a]$ .

2.  $F(x, p(x)) = p(x) \ln p(x)$ ,  $\varphi_1(x, p(x)) = p(x)$ ,  $\varphi_2(x, p(x)) = xp(x)$ . Нужно максимизировать  $J(p) = - \int_0^{+\infty} p(x) \ln p(x) dx$  при условиях  $\int_0^{+\infty} \varphi_1(x, p(x)) dx = 1$  и  $\int_0^{+\infty} \varphi_2(x, p(x)) dx = \frac{1}{\lambda}$ . Запишем уравнение:

$$-(1 + \ln p) + \lambda_1 + \lambda_2 x = 0 \implies p(x) = e^{\lambda_2 x + \lambda_1 - 1}.$$

Подставляем в условия и получаем  $p(x) = \lambda e^{-\lambda x}$  при  $x \in [0, +\infty)$ .

3. Не ограничивая общности, будем считать, что  $a = 0$  (тогда будет два условия).  $F(x, p(x)) = p(x) \ln p(x)$ ,  $\varphi_1(x, p(x)) = p(x)$ ,  $\varphi_2(x, p(x)) = x^2 p(x)$ . Нужно максимизировать  $J(p) = - \int_{-\infty}^{+\infty} p(x) \ln p(x) dx$  при условиях  $\int_{-\infty}^{+\infty} \varphi_1(x, p(x)) dx = 1$  и  $\int_{-\infty}^{+\infty} \varphi_2(x, p(x)) dx = \sigma^2$ . Запишем уравнение:

$$-(1 + \ln p) + \lambda_1 + \lambda_2 x^2 = 0 \implies p(x) = e^{\lambda_2 x^2 + \lambda_1 - 1}.$$

Подставляем в условия и получаем  $p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$  при  $x \in \mathbb{R}$ . ■

## Случайные процессы

### Определения

$(\Omega, \mathcal{A}, P)$  – исходное вероятностное пространство.

**Случайный процесс** – это семейство случайных величин  $X(t, \omega)$ , определённых на  $(\Omega, \mathcal{A}, P)$ , где  $t \in T \subset \mathbb{R}^1$ .

Не стоит забывать, что исход – это «состояние природы», в котором всё происходит, и он не зависит от времени. За время отвечает переменная  $t$ . И если зафиксировать  $\omega_0$ , то  $X(t, \omega_0)$  будет называться **траекторией случайного процесса**.

Если  $T \subset \mathbb{Z}$ , то случайный процесс называют **случайным процессом с дискретным временем**. Если  $T = [a, b]$  или  $T = [a, +\infty)$ , или  $T = (-\infty, b]$ , или  $T = (-\infty, +\infty)$ , то случайный процесс называют **случайным процессом с непрерывным временем**.

Вообще говоря,  $t$  не обязательно должно представлять из себя время, это может быть просто какой-то параметр. К тому же оно может быть многомерным (тогда мы работаем со **случайным полем**) или вообще бесконечномерным (**случайным элементом**).

Случайный процесс с непрерывным временем куда более сложный объект для исследования. Рассмотрим, например, процесс изменения силы тока в цепи. При нагрузках более пиковой величины с большой вероятностью произойдёт отказ прибора. (Рис. 7) Хотим определить вероятность того, что на некотором промежутке времени не произойдёт отказа:

$$P(A) = P\left(\sup_{t \in [a, b]} X(t) < c\right) - ?$$

С точки зрения того, что мы определили, этот вопрос может вообще быть бессмысленным. С дискретным временем всё просто: для любой последовательности случайных величин верно, что супремум этой последовательности тоже является случайной

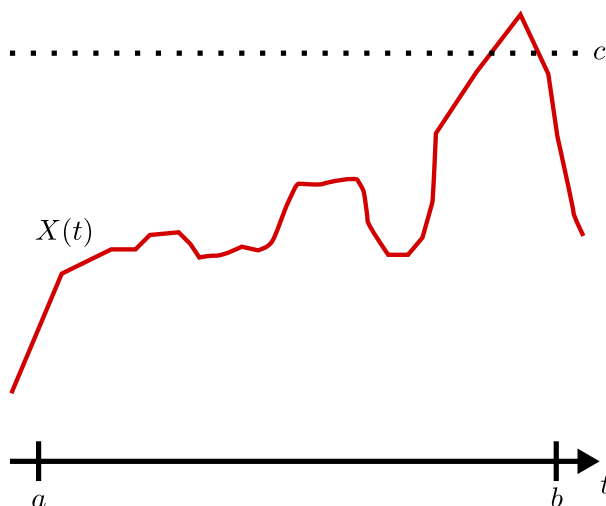


Рис. 7

величиной (это обусловлено определением  $\sigma$ -алгебры). А вот в непрерывном времени нет гарантии, что супремум по несчётному множеству будет случайно величиной.

Но оказывается, что для одномерного случайного процесса существует эквивалентный (с вероятностью 1 совпадает для любого  $t$ ) ему так называемый *сепарабельный* процесс, то есть процесс, в котором те самые проблемные операции можно заменить на те же операции, но по счётному множеству.

А вот в случае с функциональными случайными процессами проблему так не решить. Допустим у нас есть выборка случайных величин  $X_1, \dots, X_n$  и функция  $g(X_i)$ . Рассмотрим типичный *эмпирический случайный процесс*  $\sum_{i=1}^n g(X_i)$ , где  $g \in G$ .

Задача: найти характеристики  $\sup_{g \in G} \sum_{i=1}^n g(X_i)$ . Здесь сепарабельность не поможет, так как супремум берётся по классу функций, то есть как минимум не понятно что делать с размерностью. В этом может помочь *энтропия пространства*, но если она велика, то никакой сепарабельности не будет и вероятность  $P\left(\sup_{g \in G} \sum_{i=1}^n g(X_i) < c\right)$ , вообще говоря, найти не получится. Можно обратиться к *теории внешней меры*, но вероятностный смысл оценок множества не получится напрямую передать на прямую.

Мы определили случайный процесс как семейство случайных величин, но на самом деле его можно понимать как случайный элемент, отображающий элементарный исход в какую-то функцию:

$$X : \Omega \longrightarrow S - \text{множество функций.}$$

Пусть  $\mathcal{B}(S)$  – борелевская  $\sigma$ -алгебра на  $S$ . Тогда  $X$  – это **случайный элемент**, ес-

ли  $\forall B \in \mathcal{B}(S) : \{\omega : X \in B\} \in \mathcal{A}$ . Если  $S$  – вещественные функции, то  $X(t)$  – **случайный процесс**. Такое определение поможет описать бесконечномерное распределение, связанное с этим процессом:

$$\forall B \in \mathcal{B}(S) : P_X(B) = P(\omega : X \in B).$$

То есть определяется точно так же, как и индуцированная мера для случайной величины, и отличие только в множестве  $S$  вместо  $\mathbb{R}$ .

На практике конечно работают не с бесконечномерными  $X$ , а с их конечномерными проекциями: распределения вектора  $X(t_1), \dots, X(t_n)$  называются конечномерными распределениями  $X(t)$ .

### Пуассоновский процесс

Процесс  $X(t)$  называется **процессом с независимыми приращениями**, если для  $\forall n \in \mathbb{N}$  и  $\forall t_0 < t_1 < \dots < t_n, t_i \in T: X(t_n) - X(t_{n-1}), \dots, X(t_1) - X(t_0), X(t_0)$  независимы.

Процесс  $X(t)$  называется **однородным**, если для  $\forall t, s, t+h, s+h \in T: [t, t+h]$  и  $[s, s+h]$  не пересекаются и  $X(t+h) - X(t) \stackrel{d}{=} X(s+h) - X(s)$ .

Процесс  $X(t)$  называется **пуассоновским**, если:

1.  $X(0) \stackrel{п.н.}{=} 0$ .
2.  $X(t)$  – процесс с независимыми приращениями.
3.  $X(t)$  – однородный процесс.
4.  $\exists \lambda > 0$ : при  $h \downarrow 0$   $P(X(h) = 0) = 1 - \lambda h + o(h)$ ,  $P(X(h) = 1) = \lambda h + o(h)$ ,  $P(X(h) \geq 2) = o(h)$ .

## Лекция 7

### Пуассоновский процесс (продолжение)

#### Связь с пуассоновским распределением

Вообще говоря, то, что мы описали в прошлой лекции, не является определением пуассоновского процесса, это его свойства. Построить его довольно сложно и долго. Мы же ограничимся описанием его свойств и будем считать, что такой объект существует и принимает только целые неотрицательные значения. Почему этот процесс называется «пуассоновским»? Придумал его не Пуассон, но зато есть прямая связь с пуассоновским распределением.

Рассмотрим  $X(t)$  при фиксированном  $t$  (это называется **проекцией пуассоновского процесса** на время  $t$ ). Методом производящих функций найдём какое распределение будет иметь эта случайная величина:

$$\varphi_{X(t)}(s) = \varphi_t(s) = \mathbb{E}s^{X(t)}, \quad |s| \leq 1.$$

Воспользуемся тем, что процесс является процессом с независимыми приращениями и однородным:

$$\begin{aligned} \varphi_{t+h}(s) &= \mathbb{E}s^{X(t+h)} = \mathbb{E}s^{X(t+h)-X(t)+X(t)} = \mathbb{E}s^{X(t+h)-X(t)}s^{X(t)} = \\ &= \mathbb{E}s^{X(t+h)-X(t)}\mathbb{E}s^{X(t)} = \mathbb{E}s^{X(h)-X(0)}\varphi_t(s) = \mathbb{E}s^{X(h)}\varphi_t(s). \end{aligned}$$

Воспользуемся свойством:

$$\mathbb{E}s^{X(h)} = 1 - \lambda h + o(h) + s\lambda h + o(h) + o(h) = 1 + \lambda h(s - 1) + o(h).$$

Подставим это в прошлые выкладки:

$$\begin{aligned} \varphi_{t+h}(s) &= \varphi_t(s)(1 + \lambda h(s - 1) + o(h)), \\ \frac{\varphi_{t+h}(s) - \varphi_t(s)}{h} &= \varphi_t(s)(\lambda(s - 1) + o(h)). \end{aligned}$$

При  $h \downarrow 0$  получим:

$$\frac{\partial \varphi_t(s)}{\partial t} = \varphi_t(s)\lambda(s - 1).$$

Наконец воспользуемся начальным условием:

$$\varphi_0(s) = 1 \implies \varphi_t(s) = e^{\lambda t(s-1)},$$

что является производящей функцией пуассоновского распределения. Значит:

$$P(X(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \quad EX(t) = DX(t) = \lambda t.$$

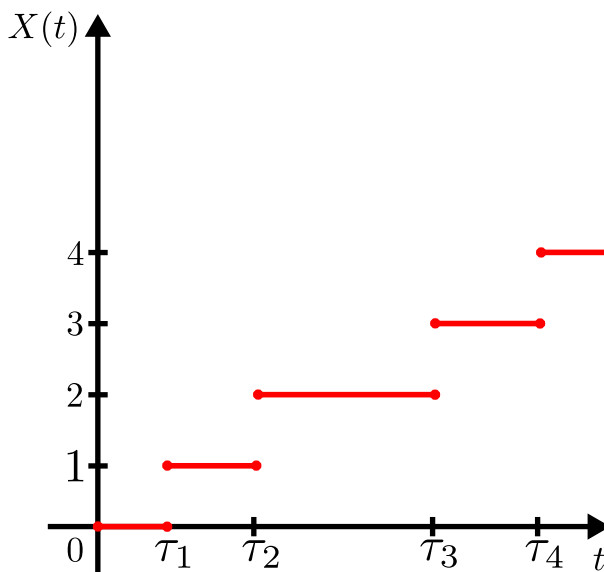


Рис. 8: Типичная траектория пуассоновского процесса

Договоримся, что процессы всегда будут иметь непрерывную траекторию справа. Так как  $\lambda t$  показывает сколько у нас будет в среднем «ступенек» до момента времени  $t$ , то  $\lambda$  называют **интенсивностью пуассоновского процесса**.

### Первое информационное свойство

Будем считать, что времена ожидания скачков независимы и имеют одинаковое распределение (далее мы докажем такое утверждение). Значит достаточно посчитать распределение только первого времени ожидания:

$$P(\tau_1 \leq t) = 1 - P(\tau_1 > t).$$

А событие  $\{\tau_1 > t\}$  эквивалентно событию  $\{X(t) = 0\}$ , значит:

$$P(\tau_1 \leq t) = 1 - e^{-\lambda t}, \quad t \geq 0.$$

$\tau_1$  имеет показательное распределение с параметром  $\lambda$ . Показательное распределение среди всех распределений на положительной полуоси с фиксированным математическим ожиданием является наиболее неопределённым.

Также показательное распределение не имеет «последствия». Что это математически значит? Пусть  $T$  – время жизни, а  $U(t) = P(T > t)$  – функция выживания. Также введём  $U(t|s) = P(T > t + s | T > s)$ ,  $s$  – возраст процесса, а  $t$  – время жизни, которое ему осталось. Тогда  $U(t|s)$  описывает как влияет возраст процесса на оставшееся время жизни.

$$U(t|s) = \frac{P(T > t + s, T > s)}{P(T > s)} = \frac{P(T > t + s)}{P(T > s)} = \frac{U(t + s)}{U(s)}.$$



Отсутствие «последствия» («памяти») означает, что  $U(t|s) = U(t)$ . Это свойство выполнено только, если  $U(t+s) = U(t)U(s)$ . Это функциональное уравнение в классе вероятностных распределений имеет единственное решение  $e^{at}$ . Если выйти за класс вероятностных распределений, то могут найтись и другие решения.

Можем сделать *вывод*: распределение показательно  $\iff$  оно обладает свойством отсутствия «памяти» (но только именно в приведённом смысле, дискретный случай отличается).

## Второе информационное свойство

**Теорема 15.** Пусть на отрезке времени  $[a, b]$  произошло  $n$  скачков пуассоновского процесса, то есть  $X(b) - X(a) = n$ .  $\{\tau_1, \dots, \tau_n\}$  – моменты скачков. Тогда условное распределение  $(\tau_1, \dots, \tau_n)$  при условии  $X(b) - X(a) = n$  совпадает с распределением вариационного ряда, построенного по выборке объёма  $n$  из  $U([a, b])$ .

*Доказательство.* Пусть  $(\eta_1, \dots, \eta_n)$  – случайный вектор с абсолютно непрерывным распределением, а  $f(t_1, \dots, t_n)$  – его плотность. Тогда  $P(\eta_i \in [t_i, t_i + h] \text{ при } i = \overline{1, n}) = f(t_1, \dots, t_n)h^n + o(h^n)$ .

Обозначим  $a = t_0 + h < t_1 < t_1 + h < \dots < t_n < t_n + h < t_{n+1} = b$ ,  $h$  возьмём подходящее этому условию. Тогда при  $i = \overline{1, n}$ :

$$P(\tau_i \in [t_i, t_i + h] | X(b) - X(a) = n) = \frac{P(\tau_i \in [t_i, t_i + h], X(b) - X(a) = n)}{P(X(b) - X(a) = n)},$$

где

$$\begin{aligned} P(X(b) - X(a) = n) &= \frac{e^{-\lambda(b-a)}(\lambda(b-a))^n}{n!}, \\ P(\tau_i \in [t_i, t_i + h], X(b) - X(a) = n) &= \\ &= P(\text{на } [t_i, t_i + h] \text{ ровно по одной точке скачка, вне этих отрезков точек скачков нет}) = \\ &= P(X(t_i + h) - X(t_i) = 1, X(t_i) - X(t_{i-1} + h) = 0 \text{ для } i = \overline{1, n+1}) = \\ &= \prod_{i=1}^n P(X(t_i + h) - X(t_i) = 1) \prod_{i=1}^{n+1} P(X(t_i) - X(t_{i-1} + h) = 0) = \\ &= (\lambda h)^n e^{-\lambda h} e^{-\lambda(t_1 - a + t_2 - t_1 - h + \dots + b - t_n - h)} = (\lambda h)^n e^{-\lambda h} e^{-\lambda(b-a)} e^{\lambda h} = (\lambda h)^n e^{-\lambda(b-a)}. \end{aligned}$$

Наконец можем написать:

$$P(\tau_i \in [t_i, t_i + h] \text{ при } i = \overline{1, n} | X(b) - X(a) = n) = \frac{(\lambda h)^n e^{-\lambda(b-a)} n!}{e^{-\lambda(b-a)} \lambda^n (b-a)^n} = \frac{h^n n!}{(b-a)^n}.$$

Следовательно:

$$p(t_1, \dots, t_n | X(b) - X(a) = n) = \frac{n!}{(b-a)^n},$$

где  $p(\cdot)$  – условная плотность и  $t_1 < \dots < t_n$ . А это и есть плотность вариационного ряда, построенного по выборке объёма  $n$  из  $U([a, b])$ . ■

То есть второе информационное свойство заключается в том, что при фиксированном количестве точек, эти точки распределены на отрезке в соответствии с равномерным законом. А равномерное распределение наиболее неопределённое среди распределений на конечном отрезке.

Условие  $X(b) - X(a) = n$  потребовалось в том смысле, чтобы на последнем отрезке  $[t_n + h, b]$  не было ничего. Если бы этого условия не было, то мы бы ничего не могли сказать о последнем отрезке. Но если взять сразу  $a = h$  и  $b = t_n + h$ , то у нас будет безусловная вероятность и  $p_{\tau_1, \dots, \tau_n}(t_1, \dots, t_n) = \lambda^n e^{-\lambda t_n}$ , где  $t_1 < \dots < t_n$ . Исходя из этого, мы докажем, что времена ожидания независимые и одинаковые показательные распределения. Введём обозначения:

$$\xi_1 = \tau_1 - \tau_0, \quad \xi_2 = \tau_2 - \tau_1, \quad \dots, \quad \xi_j = \tau_j - \tau_{j-1}$$

и сделаем обратное преобразование:

$$\tau_1 = \xi_1, \quad \tau_2 = \xi_1 + \xi_2, \quad \dots, \quad \tau_j = \xi_1 + \dots + \xi_j.$$

Матрица этого перехода выглядит как:

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}$$

Её определить равен 1 и он является якобианом перехода от  $\xi$  к  $\tau$ . Выпишем вероятность:

$$P(\tau_1 < t_1, \dots, \tau_n < t_n) = \int_{u_1 < t_1} \dots \int_{u_n < t_n} p_{\tau_1, \dots, \tau_n}(u_1, \dots, u_n) du_1 \dots du_n.$$

С другой стороны:

$$P(\xi_1 < t_1, \dots, \xi_1 + \dots + \xi_n < t_n) = \int_{v_1 < t_1} \dots \int_{v_1 + \dots + v_n < t_n} p_{\xi_1, \dots, \xi_n}(v_1, \dots, v_n) dv_1 \dots dv_n.$$

Делаем замену

$$u_1 = v_1, \quad u_2 = v_1 + v_2, \quad \dots, \quad u_n = v_1 + \dots + v_n,$$

чтобы в первом интеграле область оказалась такой же, как и во втором. Получаем:

$$\int_{v_1 < t_1} \dots \int_{v_1 + \dots + v_n < t_n} p_{\tau_1, \dots, \tau_n}(v_1, \dots, v_1 + \dots + v_n) dv_1 \dots dv_n,$$

из чего мы заключаем, что:

$$p_{\xi_1, \dots, \xi_n}(v_1, \dots, v_n) = p_{\tau_1, \dots, \tau_n}(v_1, \dots, v_1 + \dots + v_n) = \lambda^n e^{-\lambda \sum_{i=1}^n v_i} = \prod_{i=1}^n \lambda e^{-\lambda v_i}.$$

То есть получилось, что совместная плотность времён ожидания превратилась в произведение показательных плотностей (отдельно каждого из них). Значит они независимы и имеют показательное распределение с параметром  $\lambda$ .

## Лекция 8

### Связь пуассоновского процесса с нормальным распределением

**Теорема 16.** Пусть  $X(t)$  – пуассоновский процесс с интенсивностью  $\lambda > 0$ . Тогда:

$$\mathbb{P}\left(\frac{X(t) - \lambda t}{\sqrt{\lambda t}} < x\right) \xrightarrow{\lambda t \rightarrow \infty} \Phi(x), \quad \text{причём} \sup_x \left| \mathbb{P}\left(\frac{X(t) - \lambda t}{\sqrt{\lambda t}} < x\right) - \Phi(x) \right| < \frac{c_0}{\sqrt{\lambda t}},$$

где  $c_0$  – константа из неравенства Берри-Эссеена (см. (2)).

*Доказательство.* Напишем характеристическую функцию пуассоновского процесса:

$$f_{X(t)}(s) = \mathbb{E}e^{isX(t)} = e^{\lambda t(e^{is}-1)} = \underbrace{\left(e^{\frac{\lambda t}{n}(e^{is}-1)}\right)^n}_{g_n(s)}.$$

То есть для  $\forall n \in \mathbb{N}$ :

$$X_n(t) \stackrel{d}{=} X_{n,1} + \dots + X_{n,n}, \quad X_{n,i} \sim \Pi\left(\frac{\lambda t}{n}\right).$$

Запишем правую часть из неравенства Берри-Эссеена:

$$\frac{c_0 \mathbb{E}|X_{n,1} - \mathbb{E}X_{n,1}|^3}{\sqrt{n}(\mathbb{D} X_{n,1})^{\frac{3}{2}}}, \quad \mathbb{E}X_{n,1} = \mathbb{D} X_{n,1} = \frac{\lambda t}{n},$$

$$\mathbb{E}\left|X_{n,1} - \frac{\lambda t}{n}\right|^3 = \sum_{k=0}^{\infty} \mathbb{P}(X_{n,1} = k) \left|k - \frac{\lambda t}{n}\right|^3.$$

Выберем  $n$  так, чтобы  $\frac{\lambda t}{n} < 1$ :

$$\begin{aligned} \mathbb{E}\left|X_{n,1} - \frac{\lambda t}{n}\right|^3 &= \sum_{k=0}^{\infty} \mathbb{P}(X_{n,1} = k) \left(k - \frac{\lambda t}{n}\right)^3 + 2 \mathbb{P}(X_{n,1} = 0) \left(\frac{\lambda t}{n}\right)^3 = \\ &= \mathbb{E}\left(X_{n,1} - \frac{\lambda t}{n}\right)^3 + 2e^{-\frac{\lambda t}{n}} \left(\frac{\lambda t}{n}\right)^3 = \frac{\lambda t}{n} \left(1 + 2e^{-\frac{\lambda t}{n}} \left(\frac{\lambda t}{n}\right)^2\right) < \frac{\lambda t}{n} \left(1 + 2 \left(\frac{\lambda t}{n}\right)^2\right). \end{aligned}$$

Подставляя полученные результаты, мы и получим искомое неравенство. ■

Некоторые называют это *третьим информационным свойством* пуассоновского процесса, которое связывает его с нормальным распределением. На самом деле, это не совсем информационное свойство. Если у процесса приращение обладает конечной дисперсией и выполнены некоторые другие требования, то тоже будет справедливо утверждение об асимптотической нормальности, так что это не особенность пуассоновского процесса. Но это позволяет при большой интенсивности аппроксимировать настоящее распределение нормальным.

## Случайные суммы

### Определение, свойства

В некоторых статистических задачах информация собирается не для фиксированного количества экспериментов/наблюдений, а в течение фиксированного времени. Если информация поступает в виде случайного процесса, то естественно считать, что количество наблюдений тоже является случайной величиной. В этом случае вводят новый объект.

Имеем вероятностное пространство  $(\Omega, \mathcal{A}, P)$  и последовательность случайных величин  $N, X_1, X_2, \dots$ , причём  $N$  – целочисленная и неотрицательная случайная величина,  $X_1, X_2, \dots$  независимы и одинаково распределены, а  $N, X_1, X_2, \dots$  только независимы. Тогда **случайной суммой** называется  $S_N = \sum_{k=1}^N X_k$ , причём при  $S_0 = 0$ .

Введём обозначения:

$$p_k = P(N = k),$$

$\psi(s)$  – производящая функция  $N$ ,

$f(t)$  – характеристическая функция  $X_i$ ,

$F(x)$  – функция распределения  $X_i$ ,

$p(x)$  – плотность распределения  $X_i$ .

**Теорема 17.** Пусть  $S_N$  – случайная сумма. Тогда:

1.  $F_{S_N}(x) = \sum_{k=0}^{\infty} p_k F^{*k}(x)$ , где  $F^{*k}(x)$  –  $k$ -кратная свёртка  $F(x)$ , причём  $F^{*0}(x)$  – функция распределения с единичным скачком в нуле.
2. Если  $p_0 \neq 0$ , то  $S_N$  не является абсолютно непрерывной, даже если  $X_i$  абсолютно непрерывны. Если  $p_0 = 0$  и  $\exists p(x)$  – плотность  $X_i$ , то  $\exists p_{S_N}(x) = \sum_{k=0}^{\infty} p_k p^{*k}(x)$ .
3.  $f_{S_N}(t) = \psi(f(t))$ .
4.  $ES_N = EX_1 EN$ ,  $DS_N = EN DX_1 + DN(EX_1)^2$ .

*Доказательство. 1.*

$$\begin{aligned} P(S_N < x) &= \sum_{k=0}^{\infty} P(S_N < x | N = k) P(N = k) = \sum_{k=0}^{\infty} p_k \frac{P(S_N < x, N = k)}{P(N = k)} = \\ &= \sum_{k=0}^{\infty} p_k \frac{P(S_k < x, N = k)}{P(N = k)} = \sum_{k=0}^{\infty} p_k P(S_k < x) = \sum_{k=0}^{\infty} p_k F^{*k}(x). \end{aligned}$$

2. Если  $p_0 \neq 0$ , то тогда в сумме присутствует дискретная компонента, абсолютной непрерывности нет. Если  $p_0 = 0$ , то ряд сходится равномерно, и его можно почленно дифференцировать, так мы и получим искомую формулу.

3. Здесь мы воспользуемся тем, что ряд сходится равномерно, и будем почленно интегрировать:

$$\begin{aligned} f_{S_N}(t) &= \mathbb{E} e^{itS_N} = \int_{-\infty}^{+\infty} e^{itx} dF_{S_N}(x) = \int_{-\infty}^{+\infty} e^{itx} d \sum_{k=0}^{\infty} p_k F^{*k}(x) = \\ &= \sum_{k=0}^{\infty} p_k \int_{-\infty}^{+\infty} e^{itx} dF^{*k}(x) = \sum_{k=0}^{\infty} p_k f^k(t) = \psi(f(t)), \end{aligned}$$

так как  $\psi(s) = \sum_{k=0}^{\infty} p_k s^k$ .

4.

$$\mathbb{E} S_N = \frac{1}{i} f'_{S_N}(t) \Big|_{t=0}, \quad f'_{S_N}(t) = \psi'(f(t)) = \frac{\partial \psi}{\partial f} \frac{\partial f}{\partial t} \implies \mathbb{E} S_N = \frac{1}{i} \mathbb{E} N i \mathbb{E} X_1 = \mathbb{E} N \mathbb{E} X_1.$$

$$\begin{aligned} \mathbb{E} S_N^2 &= - \frac{\partial^2 f_{S_N}}{\partial t^2} \Big|_{t=0}, \quad \frac{\partial^2 f_{S_N}}{\partial t^2} = \frac{\partial}{\partial t} \left( \frac{\partial \psi}{\partial f} \frac{\partial f}{\partial t} \right) = \frac{\partial^2 \psi}{\partial f^2} \left( \frac{\partial f}{\partial t} \right)^2 + \frac{\partial \psi}{\partial f} \frac{\partial^2 f}{\partial t^2} \implies \\ \implies \mathbb{E} S_N^2 &= \mathbb{E} N(N-1)(\mathbb{E} X_1)^2 + \mathbb{E} N \mathbb{E} X_1^2, \quad (\mathbb{E} S_N)^2 = (\mathbb{E} N)^2 (\mathbb{E} X_1)^2. \end{aligned}$$

$$\mathbb{D} S_N = \mathbb{E} S_N^2 - (\mathbb{E} S_N)^2 = \mathbb{E} N \mathbb{D} X_1 + \mathbb{D} N (\mathbb{E} X_1)^2.$$

■

## Пуассоновские случайные суммы

Пусть  $N \sim \Pi(\lambda)$ . Тогда  $S_N$  называется **пуассоновской случайной суммой**.

**Утверждение 1.** Пусть  $S_N$  – пуассоновская случайная сумма. Тогда:

$$1. \mathbb{E} S_N = \lambda \mathbb{E} X_1, \quad \mathbb{D} S_N = \lambda \mathbb{E} X_1^2.$$

2.  $S_N$  безгранично делима.

*Доказательство. 1.*

$$\mathbb{E} S_N = \mathbb{E} X_1 \mathbb{E} N = \lambda \mathbb{E} X_1, \quad \mathbb{D} S_N = \mathbb{E} N \mathbb{D} X_1 + \mathbb{D} N (\mathbb{E} X_1)^2 = \lambda (\mathbb{D} X_1 + (\mathbb{E} X_1)^2) = \lambda \mathbb{E} X_1^2$$

2.

$$\psi_N(s) = e^{\lambda(s-1)}, \quad f_{S_N}(t) = e^{\lambda(f(t)-1)} = \left( e^{\frac{\lambda}{n}(f(t)-1)} \right)^n,$$

значит мы всегда можем представить пуассоновскую случайную сумму в виде суммы независимых одинаково распределённых случайных величин, каждая из которых тоже является пуассоновской случайной суммой с параметром  $\frac{\lambda}{n}$ . ■

Рассмотрим аналог центральной предельной теоремы для пуассоновских случайных сумм.

**Утверждение 2.** Пусть  $S_\lambda$  – пуассоновская случайная сумма и  $EX_1^2 < \infty$ ,  $EX_1 = a$ ,  $DX_1 = \sigma^2$ . Тогда:

$$P\left(\frac{S_\lambda - \lambda a}{\sqrt{\lambda(a^2 + \sigma^2)}} < x\right) \xrightarrow{\lambda \rightarrow \infty} \Phi(x), \quad \text{причём} \sup_x \left| P\left(\frac{S_\lambda - \lambda a}{\sqrt{\lambda(a^2 + \sigma^2)}} < x\right) - \Phi(x) \right| < \frac{c_0 L_1}{\sqrt{\lambda}},$$

где  $c_0$  – константа из неравенства Берри-Эссеена (см. (2)), а  $L_1 = \frac{E|X_1|^3}{(\lambda(a^2 + \sigma^2))^{\frac{3}{2}}}$ .

Пуассоновскую случайную сумму используют в качестве счётчика неопределённого (хаотического) потока событий. Её широко применяют, начиная от моделирования броуновского движения и заканчивая работой финансовых рынков и страховых компаний.

## Геометрические случайные суммы

Пусть  $M \sim \text{Geom}(p)$ , то есть  $P(M = k) = p(1-p)^k$ ,  $k = 0, 1, \dots$ . Тогда  $S_M$  называется геометрической случайной суммой.

$$EM = \frac{1-p}{p}, \quad DM = \frac{1-p}{p^2}, \quad \psi_M(s) = \frac{p}{1 - (1-p)s}.$$

**Теорема 18** (Реньи). Пусть  $X_i \stackrel{n.n.}{\geq} 0$ ,  $\exists EX_i = a > 0$  и  $a < \infty$ . Тогда:

$$P\left(\frac{p}{a} S_M < x\right) \xrightarrow{p \rightarrow 0} G(x),$$

где

$$G(x) = \begin{cases} 1 - e^{-x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Если также  $\exists X_i^2 = b^2$ , то

$$\sup_x \left| P\left(\frac{p}{a} S_M < x\right) - G(x) \right| \leq \frac{pb^2}{(1-p)a^2}.$$

Геометрическую случайную сумму используют для подсчёта случайных величин до наступления какого-то события (успеха или неуспеха).

## Связь между пуассоновскими и геометрическими случайными суммами

Если случайная сумма  $S_N$  пуассоновская ( $N \sim \Pi(\lambda)$ ), то её распределение называется **обобщённым пуассоновским распределением**. То есть если обобщённая пуассоновская случайная величина представляет из себя распределение суммы целочисленных слагаемых, то она тоже будет целочисленной и её можно подставить в качестве верхнего индекса в сумму.

Пусть  $M$  – целочисленная обобщённая пуассоновская случайная величина. Если  $M$  неотрицательна, то  $\exists \psi_M(s) = e^{\lambda(\varphi(s)-1)} = \psi_\lambda(\varphi(s))$ . Подставим  $M$  в случайную сумму:

$$S_M = X_1 + \dots + X_M, \quad f_{S_M}(t) = \psi_M(f(t)) = e^{\lambda\varphi(f(t))}.$$

Значит, если у нас есть  $Y_i \sim f_{Y_i}(t) = \varphi(f(t))$ , то  $Y_i \stackrel{d}{=} X_1 + \dots + X_L$ , а  $L \sim \varphi(s)$ .

**Лемма 1.** Пусть  $M$  – целочисленная обобщённая пуассоновская случайная величина,  $S_M$  – случайная сумма. Тогда  $S_M \stackrel{d}{=} S_N$ , где  $S_N \stackrel{d}{=} Y_1 + \dots + Y_N$ ,  $Y_i \stackrel{d}{=} X_1 + \dots + X_L$ .

То есть любая случайная сумма, у которой верхний индекс является обобщённой пуассоновской случайной величиной, есть пуассоновская случайная сумма.

**Лемма 2.** Пусть  $M \sim \text{Geom}(p)$ . Тогда  $M$  – целочисленная обобщённая пуассоновская случайная величина.

*Доказательство.*

$$\lambda = \ln \frac{1}{p}, \quad \psi(s) = \frac{1}{\lambda} \ln \frac{1}{1 - (1-p)s}.$$

Если мы убедимся, что

$$\psi_M(s) = \frac{p}{1 - (1-p)s},$$

то всё будет доказано. Сделать это легко:

$$\psi_M(s) = e^{\lambda(\varphi(s)-1)} = e^{\ln \frac{1}{1-(1-p)s} - \lambda} = \frac{p}{1 - (1-p)s}.$$

Осталось убедиться, что  $\psi(s)$  – производящая функция, иначе это не имеет смысла:

$$\ln(1-x) = -\sum_{k=1}^{\infty} \frac{x^k}{k} \implies \frac{1}{\lambda} \ln \frac{1}{1 - (1-p)s} = \frac{1}{\lambda} \sum_{k=1}^{\infty} \frac{(s(1-p))^k}{k},$$

$$\frac{1}{\lambda} \sum_{k=1}^{\infty} \frac{(1-p)^k}{k} = 1.$$

То есть мы убедились, что данной функции соответствует некоторое вероятностное распределение (такое распределение называют **логарифмическим**), значит это производящая функция. ■



**Теорема 19.** Пусть  $M \sim \text{Geom}(p)$ ,  $S_M \stackrel{d}{=} X_1 + \dots + X_M$  – случайная сумма. Тогда  $S_M$  – пуассоновская случайная сумма и  $S_M \stackrel{d}{=} S_N$ , где  $S_N \stackrel{d}{=} Y_1 + \dots + Y_N$ ,  $Y_1, \dots, Y_N$  независимы,  $Y_i \stackrel{d}{=} X_1 + \dots + X_L$ , а  $L$  – логарифмическое распределение,  $N \sim \Pi\left(\ln \frac{1}{p}\right)$ .

## Лекция 9

### Случайные суммы

#### Связь между пуассоновскими и геометрическими случайными суммами (продолжение)

**Следствие 1** (теоремы 19). *Распределение любой геометрической случайной суммы безгранично делимо.*

Утверждение, обратное теореме 19, верно только при определённых условиях.

**Утверждение 3.** Пусть задана функция  $g(t) = \frac{1 - e^{-\lambda f(t)}}{1 - e^{-\lambda}}$ ,  $S_N = X_1 + \dots + X_N$  ( $N \sim \Pi(\lambda)$ ) и  $f(t)$  – характеристическая функция  $X_i$ . Если  $g(t)$  является характеристической функцией некоторого распределения, то

$$S_N \stackrel{d}{=} S_M, \quad M \sim \text{Geom}(e^{-\lambda}), \quad S_M = Y_1 + \dots + Y_M$$

и характеристическая функция  $Y_i$  равна  $g(t)$ .

*Доказательство.* Мы знаем, что

$$\psi_M(s) = \frac{p}{1 - (1-p)s},$$

и нам надо убедиться, что

$$\psi_M(g(t)) = e^{\lambda(f(t)-1)} = \psi_N(f(t)).$$

Пусть  $p = e^{-\lambda}$ :

$$\psi_M(g(t)) = \frac{e^{-\lambda}}{1 - (1 - e^{-\lambda}) \left( \frac{1 - e^{-\lambda f(t)}}{1 - e^{-\lambda}} \right)} = e^{\lambda(f(t)-1)}$$

■

#### Теорема переноса

Рассмотрим схему серий  $\{X_{n,j}\}$ :

$$\begin{array}{l} X_{1,1}, \dots \\ X_{2,1}, X_{2,2}, \dots \\ \vdots \\ X_{n,1}, X_{n,2}, \dots \end{array}$$

$X_{n,1}, X_{n,2}, \dots$  – независимые одинаково распределённые случайные величины. Также обозначим  $S_{n,k} = \sum_{j=1}^k X_{n,j}$ .

**Теорема 20** (переноса). Пусть  $\{X_{n,j}\}$  – схема серий,  $N_n$  – такая последовательность положительных целочисленных случайных величин, что для  $\forall n: N_n, X_{n,1}, \dots$  – независимые случайные величины,  $S_{n,N_n}$  – случайная сумма. Также существует  $\{m_n\}$  – такая последовательность неслучайных положительных целых чисел, что

$$P(S_{n,m_n} < x) \xrightarrow{n \rightarrow \infty} H(x) \text{ и } P\left(\frac{N_n}{m_n} < u\right) \xrightarrow{n \rightarrow \infty} A(u).$$

Тогда

$$P(S_{n,N_n} < x) \xrightarrow{n \rightarrow \infty} F(x),$$

где  $F(x)$  соответствует характеристической функции  $f(t)$ , причём

$$f(t) = \int_0^{\infty} h^u(t) dA(u),$$

где  $h(t)$  – характеристическая функция  $H(x)$ .

В терминах случайных величин это означает, что если  $S_{n,m_n} \Rightarrow Y$  (с характеристической функцией  $h(t)$ ) и  $\frac{N_n}{m_n} \Rightarrow U$  (с функцией распределения  $A(u)$ ), то  $S_{n,N_n} \Rightarrow Z$  (с характеристической функцией  $f(t) = Eh^u(t)$ ).

*Доказательство.* Пусть  $g_n(t)$  – характеристическая функция  $X_{n,1}$ , а  $h_n(t) = g_n^{m_n}(t)$  – характеристическая функция суммы  $m_n$  слагаемых. Тогда:

$$\begin{aligned} f_n(t) &= Ee^{itS_{n,N_n}} = Ee^{it \sum_{j=1}^{N_n} X_{n,j}} = \sum_{k=1}^{\infty} P(N_n = k) Ee^{it \sum_{j=1}^k X_{n,j}} = \sum_{k=1}^{\infty} P(N_n = k) g_n^k(t) = \\ &= \int_0^{\infty} g_n^u(t) dP(N_n < u) = \{u = u'm_n\} = \int_0^{\infty} g_n^{u'm_n}(t) dP\left(\frac{N_n}{m_n} < u'\right). \end{aligned}$$

Обозначим  $A_n(u) = P\left(\frac{N_n}{m_n} < u\right)$  и  $u = u'$ :

$$f_n(t) = \int_0^{\infty} g_n^{um_n}(t) dA_n(u) = \int_0^{\infty} h_n^u(t) dA_n(u).$$

Нам нужно доказать, что для  $\forall \varepsilon > 0$  при  $\forall$  фиксированном  $t \in \mathbb{R}$ :  $\exists n_0 : \forall n > n_0 : |f_n(t) - f(t)| < \varepsilon$ . Фактически нам надо узнать следующее:

$$\begin{aligned} & \left| \int_0^\infty h_n^u(t) dA_n(u) - \int_0^\infty h_n^u(t) dA(u) \right| = \\ & = \left| \int_0^\infty h_n^u(t) dA_n(u) - \int_0^\infty h^u(t) dA_n(u) + \int_0^\infty h^u(t) dA_n(u) - \int_0^\infty h_n^u(t) dA(u) \right| \leq \\ & \leq \underbrace{\left| \int_0^\infty h_n^u(t) dA_n(u) - \int_0^\infty h^u(t) dA_n(u) \right|}_{I_1(n)} + \underbrace{\left| \int_0^\infty h^u(t) dA_n(u) - \int_0^\infty h_n^u(t) dA(u) \right|}_{I_2(n)} < \varepsilon? \end{aligned}$$

Для начала разберёмся с  $I_2(n)$ . Пусть  $N$  – такая случайная величина, что  $\frac{N_n}{m_n} \Rightarrow N$ . То есть по определению для  $\forall \varphi$  – непрерывна и ограниченная функция:  $E\varphi\left(\frac{N_n}{m_n}\right) \xrightarrow{n \rightarrow \infty} E\varphi(N)$ .

Возьмём  $\varphi_z(u) = z^u$ . При  $|z| \leq 1$  она непрерывная и ограничена. Тогда  $\varphi_{h(t)}(u) = h^u(t)$ . Из того, что  $A_n \xrightarrow{n \rightarrow \infty} A$  и  $|h(t)| \leq 1$  (свойство характеристической функции), следует, что, начиная с некоторого  $n$ ,  $I_2(n) < \varepsilon'$  ( $\varepsilon'$  зависит от  $\varepsilon$ , его можно сделать равным, например,  $\frac{\varepsilon}{2}$ ).

Теперь перейдём к  $I_1(n)$ . Возьмём достаточно большое число  $M$ :

$$\begin{aligned} I_1(n) &= \left| \int_0^\infty h_n^u(t) dA_n(u) - \int_0^\infty h^u(t) dA_n(u) \right| \leq \\ & \leq \underbrace{\left| \int_0^M h_n^u(t) dA_n(u) - \int_0^M h^u(t) dA_n(u) \right|}_{I_{11}(n)} + \underbrace{\left| \int_M^\infty h_n^u(t) dA_n(u) - \int_M^\infty h^u(t) dA_n(u) \right|}_{I_{12}(n)}. \end{aligned}$$

Снова для начала разберёмся со вторым слагаемым  $I_{12}(n)$ :

$$I_{12}(n) = \left| \int_M^\infty (h_n^u(t) - h^u(t)) dA_n(u) \right| \leq \int_M^\infty (|h_n^u(t)| + |h^u(t)|) dA_n(u) \leq 2 \int_M^\infty dA_n(u).$$

Здесь нам понадобится следующее определение: семейство случайных величин  $\{\xi_n\}$  (семейство соответствующих функций распределений  $\{F_n\}$ ) называется **слабо компактным**, если из любой последовательности этого семейства можно выбрать слабо сходящуюся подпоследовательность. Более того,  $\{\xi_n\}$  слабо компактно  $\iff$

$\lim_{M \rightarrow \infty} \sup_n \mathbb{P}(|\xi_n| > M) = 0$  (говорят, что  $\{\xi_n\}$  плотно). Вернёмся к выкладкам:

$$I_{12}(n) \leq 2 \int_M^\infty dA_n(u) \leq 2 \sup_n \int_M^\infty dA_n(u) = 2 \sup_n \mathbb{P}\left(\frac{N_n}{m_n} > M\right) \leq \varepsilon''.$$

Ну и осталось разобраться с  $I_{11}(n)$ :

$$I_{11}(n) = \left| \int_0^M h_n^u(t) dA_n(u) - \int_0^M h^u(t) dA_n(u) \right| \leq \int_0^M |h_n^u(t) - h^u(t)| dA_n(u).$$

Воспользуемся следующим утверждением (по сути формулой Лагранжа): пусть  $\psi(z)$  непрерывно дифференцируема в области  $U$  и отрезок с концами  $a$  и  $b$  целиком лежит в этой области, тогда:

$$|\psi(b) - \psi(a)| \leq |b - a| \sup_{0 \leq \alpha \leq 1} |\psi'(\alpha a + (1 - \alpha)b)|.$$

Возьмём  $\psi(z) = z^u$ ,  $\psi'(z) = uz^{u-1}$ :

$$\begin{aligned} |h_n^u(t) - h^u(t)| &\leq u |h_n(t) - h(t)| \sup_{0 \leq \alpha \leq 1} |\alpha h_n(t) + (1 - \alpha)h(t)|^{u-1} \leq \\ &\leq u |h_n(t) - h(t)| \sup_{0 \leq \alpha \leq 1} |\alpha h_n(t) + (1 - \alpha)h(t)|^{-1}. \end{aligned}$$

Так как  $h(t)$  безгранично делима (по теореме Хинчина), то  $h(t) \neq 0$  при  $\forall t \in \mathbb{R}$ . Это значит, что, начиная с некоторого  $n_0$ :  $h_n(t) \neq 0$ . Более того, для некоторого  $\delta$  и  $\forall n > n_0$ :  $|h_n(t)| > \delta > 0$ . То есть, зная, что  $\min(h_n(t), h(t)) > \delta > 0$ , можем заключить:

$$|h_n^u(t) - h^u(t)| \leq \frac{u |h_n(t) - h(t)|}{\min(h_n(t), h(t))} < \frac{u}{\delta} |h_n(t) - h(t)|.$$

Ну и вернёмся к  $I_{11}(n)$ :

$$I_{11}(n) \leq \int_0^M \underbrace{\frac{u}{\delta} |h_n(t) - h(t)|}_{< \varepsilon'''} dA_n(u) \leq \frac{\varepsilon'''}{\delta} \int_0^M u dA_n(u) \leq \frac{M \varepsilon'''}{\delta}.$$

Подобрав подходящие  $\varepsilon'$ ,  $\varepsilon''$  и  $\varepsilon'''$ , мы и получим, что:

$$\left| \int_0^\infty h_n^u(t) dA_n(u) - \int_0^\infty h^u(t) dA(u) \right| < \varepsilon.$$

■

Заметим, что в отличие от центральной предельной теоремы, здесь мы сумму не центрируем, то есть подразумевается, что слагаемые либо уже с нулевым математическим ожиданием, либо их математические ожидания от серии к серии стремятся к нулю. Если сумма не случайна, то нет разницы что центрировать: каждое слагаемое или всю сумму. А вот в случае со случайной суммой, если центрировать каждое слагаемое константой, то в результате вся сумма центрируется случайной величиной. Тогда мы придём к доказанной нами теореме переноса. Но, вообще говоря, это не очень разумно. Желательно центрировать сумму чем-то неслучайным. В таком случае теорема поменяется.

**Теорема 21** (переноса для центрированных сумм в терминах случайных величин). Пусть  $\{X_{n,j}\}$  – схема серий и существуют  $\{a_n\}$ ,  $\{c_n\}$ ,  $\{m_n\}$  – последовательности неслучайных чисел (причём  $\{m_n\}$  состоит только из положительных целых чисел). Также известно, что

$$S_{n,m_n} - a_n \Rightarrow Y, \quad \frac{N_n}{m_n} \Rightarrow U, \quad a_n \frac{N_n}{m_n} - c_n \Rightarrow V.$$

Тогда

$$S_{n,m_n} - c_n \Rightarrow Z$$

с характеристической функцией  $f(t) = E h^U(t) e^{itV}$ .

Если  $U \stackrel{\text{п.н.}}{=} 1$ , то  $f(t) = h(t) E e^{itV} \sim Y + V$ , причём  $Y$  и  $V$  независимы. Это называется **сдвиговой смесью**.

## Лекция 10

### Смеси распределений

#### Смешанные пуассоновские распределения

Рассмотрим семейство случайных величин  $\{X_{j,p}\}$  ( $0 < p < 1$ ):

$$X_{j,p} \sim \begin{cases} 1, & \text{с вероятностью } p, \\ 0, & \text{с вероятностью } 1 - p; \end{cases}$$

и семейство целочисленных отрицательных величин  $\{N_p\}$  и потребуем, чтобы  $N_p, X_{1,p}, X_{2,p}, \dots$  были независимы. Обозначим  $S_p = \sum_{j=1}^{N_p} X_{j,p}$  (такие суммы описывают редкий поток событий при  $p \rightarrow 0$ ; например, в случае со страховой компанией,  $X_{j,p}$  – индикатор неблагоприятного события,  $N_p$  – объём портфеля за фиксированный интервал времени, а  $S_p$  – число страховых выплат по портфелю).

**Теорема 22** (аналог теоремы Пуассона для случайных сумм). Пусть  $pN_p \xrightarrow[p \rightarrow 0]{} N$ . Тогда  $S_p \xrightarrow[p \rightarrow 0]{} S$ , причём

$$P(S = k) = \frac{1}{k!} \int_0^\infty e^{-z} z^k dP(N < z).$$

*Доказательство.* Возьмём бесконечно малую последовательность  $\{p_n\}$ ,  $m_n = \left\lfloor \frac{1}{p_n} \right\rfloor$ .

Тогда для  $S_{m_n} = \sum_{j=1}^{m_n} X_{j,p_n}$  по теореме Пуассона верно, что  $S_{m_n} \xrightarrow[n \rightarrow \infty]{} Y \sim \Pi(1)$  (поскольку  $p_n m_n \xrightarrow[n \rightarrow \infty]{} 1$ ) и  $f_Y(t) = e^{it-1}$ .

$$p_n N_{p_n} = p_n m_n \frac{N_{p_n}}{m_n} \xrightarrow[n \rightarrow \infty]{} N$$

и по теореме переноса

$$S_{p_n} \xrightarrow[n \rightarrow \infty]{} S, \quad f_S(t) = \int_0^\infty f_Y^z(t) dA(z) = \int_0^\infty e^{z(e^{it}-1)} dA(z),$$

где  $A(z)$  – функция распределения случайной величины  $N$ .

Далее можно воспользоваться разложением в ряд и получить искомую формулу. ■

Давайте взглянем на эти вероятности. Пуассоновские вероятности представляются из себя  $\pi_k(\lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$ , а вероятность из теоремы  $P(S = k) = E\pi_k(N)$ . Общий вид выглядит так:  $F(x, y)$  – функция распределения ( $x$  – переменная,  $y$  – параметр),  $G(y)$  – функция распределения  $N$ , тогда функция распределения для вероятности из теоремы будет выглядеть как  $H(x) = \int F(x, y) dG(y)$ . Такая процедура называется **рандомизацией (смешиванием)**, а результат **смесью** (более строгое определение дадим далее). Распределение, которое мы получим в пределе, называется **смешанным пуассоновским распределением**.

Вернёмся к теореме переноса. Пусть в ней  $H(x) = \Phi(x)$ . Тогда

$$F(x) = \int_0^{\infty} \Phi\left(\frac{x}{\sqrt{u}}\right) dA(u)$$

называется **масштабной смесью нормальных законов**. В случае теоремы переноса для центрированных случайных сумм мы получим уже **сдвиг-масштабную смесь нормальных законов**. Эти смеси представляют из себя очень разные объекты. Теперь рассмотрим характеристическую функцию:

$$\begin{aligned} f(t) &= \int_{-\infty}^{+\infty} e^{itx} dF(x) = \int_{-\infty}^{+\infty} e^{itx} d \int_0^{\infty} \Phi\left(\frac{x}{\sqrt{u}}\right) dA(u) = \int_{-\infty}^{+\infty} e^{ity\sqrt{u}} d \int_0^{\infty} \Phi(y) dA(u) = \\ &= \int_0^{\infty} \int_{-\infty}^{+\infty} e^{ity\sqrt{u}} d\Phi(y) dA(u) = \int_0^{\infty} \varphi(t\sqrt{u}) dA(u) = \int_0^{\infty} \varphi^u(t) dA(u), \end{aligned}$$

где  $\varphi(t)$  – характеристическая функция распределения  $N(0, 1)$ . То есть, если для неслучайных сумм справедлива центральная предельная теорема, то в теореме переноса получаются масштабные смеси нормальных законов. Этот класс достаточно богат: в него входит само нормальное распределение  $\left(\sim \frac{e^{-\gamma x^2}}{|x|} \text{ при } |x| \rightarrow \infty\right)$ ,

распределение Лапласа  $f(x) = \frac{1}{2} \mu e^{-\mu|x|}$  ( $\sim e^{-\mu x}$  при  $x \rightarrow \infty$ ), распределение Стюдента  $\left(\sim \frac{1}{|x|} \gamma \text{ при } x \rightarrow \infty \text{ и } \gamma > 0\right)$  и распределение Коши  $\left(\sim \frac{1}{|x|} \text{ при } x \rightarrow \infty\right)$ . То, какое распределение получится в хвосте, зависит от поведения верхнего индекса случайной суммы.

## Определения

Пусть функция  $F(x, y)$  задано на  $\mathbb{R} \times Y$ , где  $Y \subset \mathbb{R}^m$ .  $\Sigma$  –  $\sigma$ -алгебра на  $Y$ . При  $\forall$  фиксированном  $x \in \mathbb{R}$ :  $F(x, y)$  измерима относительно  $\Sigma$ . При  $\forall$  фиксированном



$y \in Y$ :  $F(x, y)$  является функцией распределения.  $G$  – вероятностная мера на  $(Y, \Sigma)$ . Тогда  $H(x) = \int F(x, y)G(dy)$  – смесь  $F$  относительно  $G$ ,  $F$  – смешиваемое распределение, а  $G$  – смешивающее распределение.

Пусть на  $(Y, \Sigma, G)$  задана тождественная случайная величина  $Z$ . Тогда смесь можно записать как  $H(x) = EF(x, Z)$ . Если существует плотность  $f(x, y)$ , то и у смеси тоже существует плотность  $h(x) = Ef(x, Z)$ . Если  $Z$  дискретна, то  $H(x) = \sum_{j=1}^n p_j F(x, y_j)$  и  $h(x) = \sum_{j=1}^n p_j f(x, y_j)$ , причём  $F(x, y_j)$  называют **компонентами смеси**, а  $p_j$  **весами смеси**.

Теперь рассмотрим сдвиг-масштабные смеси. В этом случае  $y = (u, v)$  ( $Y = (U, V)$ ), где  $u > 0, v \in \mathbb{R}$ . Потребуем, чтобы для  $\forall u, v$  из  $y$ :  $F\left(\frac{x-v}{u}\right)$  измерима по  $u$  и  $v$  и является функцией распределения по  $x$ .  $G(u, v)$  – смешивающее распределение, тогда сдвиг-масштабная смесь записывается как

$$H(x) = EF\left(\frac{x-V}{U}\right) = \int_0^{\infty} \int_{-\infty}^{+\infty} F\left(\frac{x-v}{u}\right) G(du, dv).$$

Если существует плотность, соответствующая  $F\left(\frac{x-v}{u}\right)$ , равная  $\frac{1}{u}f\left(\frac{x-v}{u}\right)$ , то

$$h(x) = E \frac{1}{U} f\left(\frac{x-V}{U}\right).$$

Если вектор  $(U, V)$  имеет дискретное распределение с парами  $(a_j, \sigma_j)$ , вероятность которых  $p_j$ :

$$H(x) = \sum_{j=1}^n p_j F\left(\frac{x-a_j}{\sigma_j}\right), \quad h(x) = \sum_{j=1}^n \frac{p_j}{\sigma_j} f\left(\frac{x-a_j}{\sigma_j}\right).$$

**Пример.** Предположим, что рассматривается популяция, состоящая из  $k$  субпопуляций, обладающих разными признаками.  $F_j(x)$  – функция распределения признака у  $j$ -ой субпопуляции, а  $p_j$  – доля  $j$ -ой субпопуляции. Тогда дискретная смесь вида  $H(x) = \sum_{j=1}^k p_j F_j(x)$  позволяет рассматривать неоднородную группу как однородную, то есть работать со всей популяцией, несмотря на различные группы. То есть происходит «смешивание всего в однородную кучу».

## Идентифицируемость смесей

Более интересной является обратная задача (разложение/идентификация смеси): имеется смесь и нужно установить из каких групп она состоит и какими свойствами

обладают их функции распределения. Причём нам нужно определить ведь не только веса и их вероятности, но и их количество; делается это в основном итерационными методами (например, ЕМ-алгоритмами и его модификациями). Но вообще говоря, не каждую смесь можно разделить.

Пусть функция  $F(x, y)$  удовлетворяет определению смеси,  $\mathcal{Q}$  – семейство случайных величин на  $(Y, \Sigma, G)$ ,  $\mathcal{H} = \{H(x) = EF(x, Q), Q \in \mathcal{Q}\}$  – семейство смесей. Тогда  $\mathcal{H}$  называется **идентифицируемым семейством смесей**, если из равенства  $EF(x, Q_1) = EF(x, Q_2)$  при  $\forall x \in \mathbb{R}$  следует, что  $Q_1 \stackrel{d}{=} Q_2$ .

Конечно же, не каждая смесь идентифицируема. Запишем очевидное равенство:

$$\frac{2}{3} 3 \mathbb{I} \left( x \in \left[ 0, \frac{1}{3} \right) \right) + \frac{2}{3} \frac{3}{2} \mathbb{I} \left( x \in \left[ \frac{1}{3}, 1 \right) \right) = \frac{1}{2} 2 \mathbb{I} \left( x \in \left[ 0, \frac{1}{2} \right) \right) + \frac{1}{2} 2 \mathbb{I} \left( x \in \left[ \frac{1}{2}, 1 \right) \right).$$

Это равенство показывает, что смеси равномерных законов неидентифицируемы. И такая ситуация не редкая, но очень часто всё равно пытаются разделять, но никакого математического обоснования этому нет.

Вернёмся к сдвиг-масштабным смесям. То есть  $H(x) = EF \left( \frac{x - V}{U} \right)$ . Пусть  $(U, V)$  – случайный вектор (причём  $U > 0$ ),  $X \sim F(x)$ . Тогда  $H(x)$  – функция распределения случайной величины  $UX + V$ , где  $X$  и  $(U, V)$  независимы. Пусть  $\mathcal{Q} = \{(U, V)\}$ ,  $\mathcal{H} = \left\{ H(x) = EF \left( \frac{x - V}{U} \right), (U, V) \in \mathcal{Q} \right\}$ . Тогда  $\mathcal{H}$  называется **идентифицируемым** если из равенства  $X_1 U_1 + V_1 \stackrel{d}{=} X_2 U_2 + V_2$  (где  $X_1, X_2 \sim F(x)$  и  $(U_i, V_i) \in \mathcal{Q}$ ) следует, что  $(U_1, V_1) \stackrel{d}{=} (U_2, V_2)$ .

Ну и рассмотрим дискретный случай. То есть  $H(x) = \sum_{j=1}^n p_j F \left( \frac{x - a_j}{\sigma_j} \right)$ . Тогда  $\mathcal{H}$  называется **идентифицируемым** если из равенства

$$\sum_{j=1}^n p_j F \left( \frac{x - a_j}{\sigma_j} \right) = \sum_{i=1}^k q_i F \left( \frac{x - b_i}{d_i} \right) \text{ для } \forall x \in \mathbb{R}$$

следует, что  $n = k$  и для  $\forall j \in \{1, \dots, n\}$ :  $\exists i \in \{1, \dots, n\}$ :  $p_j = q_i$ ,  $a_j = b_i$ ,  $\sigma_j = d_i$ .

**Пример.** Пусть  $F(x, y) = F(x - y)$  и характеристическая функция  $f(t) \neq 0$  для  $\forall t \in \mathbb{R}$ . Также  $\mathcal{Q}$  – семейство всех возможных случайных величин, а  $\mathcal{H} = \{EF(x - Y), Y \in \mathcal{Q}\}$ . Тогда  $\mathcal{H}$  идентифицируема. То есть семейство сдвиговых смесей любых законов, у которых характеристическая функция нигде не обращается в ноль, идентифицируема (в частности семейство сдвиговых смесей безгранично делимых законов). Как это доказывается?  $EF(x - Y) \sim X + Y$ , где  $X$  и  $Y$  независимы и  $X \sim F(x)$ . Тогда  $f_{X+Y}(t) = f_X(t)f_Y(t)$  и если есть равенство  $f_{X_1}(t)f_{Y_1}(t) = f_{X_2}(t)f_{Y_2}(t)$  (где  $X_1, X_2 \sim F(x)$ ), то можем просто получить  $f_{Y_1}(t) = f_{Y_2}(t)$  для  $\forall t \in \mathbb{R}$ .

## Лекция 11

### Идентифицируемость смесей (продолжение)

Семейство функций  $F(x, y)$  ( $y > 0$ ) называется **аддитивно замкнутым**, если  $F(x, y_1 + y_2) = F(x, y_1) * F(x, y_2)$ , где  $*$  – оператор свёртки (причём вспомните, что данная запись некорректна, но мы её используем для наглядности, см. (5)).

**Утверждение 4.**  $\mathcal{H} = \{EF(x, Q), Q : P(Q > 0) = 1\}$ ,  $F(x, y)$  аддитивно замкнуто. Тогда  $\mathcal{H}$  идентифицируема.

Например, рассмотрим  $\Phi\left(\frac{x}{\sqrt{s}}\right)$  ( $s > 0$ ) – семейство нормальных законов, у которых разная дисперсия. Это семейство в силу свойств нормального распределения будет аддитивно замкнутым. Значит семейство масштабных смесей нормальных законов с нулевым математическим ожиданием идентифицируемо. Также это значит, что в теореме переноса для нецентрированных случайных сумм мы получаем идентифицируемые смеси.

**Утверждение 5.**  $F(x, y) = F(xy)$  ( $y > 0, F(0) = 0$ ),  $\mathcal{H} = \{EF(x, Q), Q : P(Q > 0) = 1\}$ . Тогда  $\mathcal{H}$  идентифицируема.

**Утверждение 6.** Семейство конечных сдвиг-масштабных смесей идентифицируемо.

Например, из смеси вида  $H(x) = \sum_{k=1}^n p_k \Phi\left(\frac{x - a_k}{\sigma_k}\right)$  можно однозначно определить  $p_k, a_k$  и  $\sigma_k$ .

**Пример.** Пусть  $X_1, X_2 \sim N(0, 1)$  и они независимы,  $U_1 : P(U_1 = 1) = 1$ ,  $U_2 : P(U_2 > 0) = 1$ ,  $U_2$  не вырождена и не зависит от  $X_1$  и  $X_2$ . Рассмотрим случайную величину  $Y = U_2 X_1 + X_2$ . Тогда

$$H_Y(x) = \int_0^{\infty} \int_{-\infty}^{+\infty} \Phi\left(\frac{x - v}{u}\right) dP(U_2 < u) dP(X_2 < v),$$

$$H_Y(x) = \int_0^{\infty} \int_{-\infty}^{+\infty} \Phi\left(\frac{x - v}{u}\right) dP(U_1 < u) dP(U_2 X_1 < v).$$

То есть мы рассматривали некоторую величину вида  $UX + V$ , в первом случае получили  $(U, V) \stackrel{d}{=} (U_2, X_2)$ , а во втором  $(U, V) \stackrel{d}{=} (U_1, U_2 X_1)$ . Понятно, что  $(U_2, X_2) \neq$

$(U_1, U_2 X_1)$ . То есть одну и ту же смесь можно представить двумя разными смешивающими законами, следовательно сдвиг-масштабная смесь не является идентифицируемой. В то же время конечная сдвиг-масштабная смесь является идентифицируемой. Можно представить распределение в виде предела простых распределений, для которых есть идентифицируемость, но данное свойство не сохраняется при переходе к пределу.

## Процессы Кокса

### Обобщение пуассоновского процесса

Пусть  $N_1(t)$  – пуассоновский процесс с интенсивностью  $\lambda = 1$ , а  $N_\lambda(t)$  – пуассоновский процесс с некоторой  $\lambda$ . Тогда  $P(N_\lambda(t) = k) = P(N_1(\lambda t) = k)$ . То есть интенсивность можно деформировать нелинейным образом для неоднородности.

Давайте откажемся от того, что  $\lambda$  – константа. Пусть  $\lambda(t)$  – некоторая положительная интегрируемая функция, а  $\Lambda(t) = \int_0^t \lambda(\tau) d\tau$ . Тогда  $N^*(t)$  называется **неоднородным пуассоновским процессом**, если

1.  $N^*(0) \stackrel{\text{п.н.}}{=} 0$ .

2.  $P(N^*(t) = k) = e^{-\Lambda(t)} \frac{\Lambda^k(t)}{k!}$  или же  $N^*(t) = N_1(\Lambda(t))$ .

$\lambda(t)$  называется **мгновенной интенсивностью**  $N^*(t)$ :

$$EN^*(t) = \Lambda(t), \quad \frac{EN^*(t+h) - EN^*(t)}{h} = \frac{\Lambda(t+h) - \Lambda(t)}{h} \xrightarrow{h \rightarrow 0} \lambda(t).$$

$\Lambda(t)$  называется **накопленной интенсивностью**  $N^*(t)$ .

Траектория неоднородного пуассоновского процесса (Рис. 9) отличается от траектории обычного пуассоновского процесса (Рис. 8) тем, что в зависимости от  $\lambda(t)$  скачки могут сгущаться или, наоборот, редеть.

### Определения

Пусть  $\Lambda(t)$  удовлетворяет следующим условиям:

1.  $\Lambda(0) \stackrel{\text{п.н.}}{=} 0$ .
2. Траектории  $\Lambda(t)$  не убывают и непрерывны справа (с какой-то одной стороны).
3. Для  $\forall t > 0: P(\Lambda(t) < \infty) = 1$ .
4.  $N_1(t)$  и  $\Lambda(t)$  независимы.

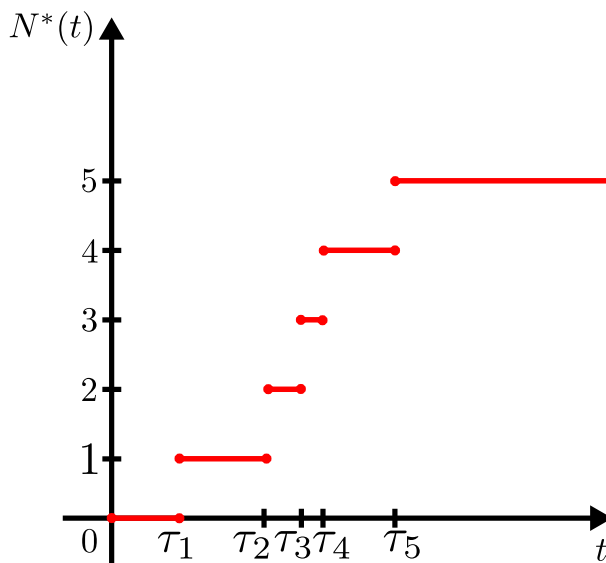


Рис. 9: Типичная траектория неоднородного пуассоновского процесса

Тогда  $N(t) = N_1(\Lambda(t))$  называется **дважды стохастическим пуассоновским процессом**, или **процессом Кокса**. Причём говорят, что  $\Lambda(t)$  управляет процессом Кокса и называется **управляющим процессом**.

Пусть  $\Lambda(t) = \int_0^t \lambda(\tau) d\tau$ , где  $\lambda(\tau)$  – процесс с неотрицательными интегрируемыми траекториями. Тогда  $\lambda(\tau)$  называется **мгновенной стохастической интенсивностью процесса**  $N(t)$ .

$$\begin{aligned} \mathbb{E} N^2(t) &= \mathbb{E} \mathbb{E}(N(t) | \Lambda(t)) = \mathbb{E} \Lambda(t), \\ \mathbb{E} N^2(t) &= \mathbb{E} \mathbb{E}(N^2(t) | \Lambda(t)) = \mathbb{E} \Lambda^2(t) + \mathbb{E} \Lambda(t), \\ D N(t) &= \mathbb{E} N^2(t) - (\mathbb{E} N(t))^2 = \mathbb{E} \Lambda(t) + D \Lambda(t). \end{aligned}$$

Отличия траектории такого процесса от траектории неоднородного пуассоновского процесса (Рис. 9) по одному рисунку не передать. Главное различие заключается в том, что в силу случайности интенсивности области сгущения могут смещаться, оставаться на месте долгое время. То есть если меняется интенсивность, то меняется и траектория, а для неоднородного пуассоновского процесса она остаётся в среднем одного вида.

### Обобщённые процессы Кокса

Пусть  $X_1, X_2, \dots$  – последовательность независимых одинаково распределённых случайных величин,  $N(t)$  – процесс Кокса, причём  $N(t), X_1, X_2, \dots$  независимы при  $\forall t > 0$ . Тогда  $S(t) = \sum_{j=1}^{N(t)} X_j$  называется **обобщённым процессом Кокса** (если

$N(t)$  – пуассоновский процесс, то  $S(t)$  – **обобщённый пуассоновский процесс**), причём при  $N(t) = 0$ :  $S(t) = 0$ .

**Пример.** Возможно вы слышали, что броуновское движение подчиняется нормальному закону. На самом деле исходно эта модель есть предельная аппроксимация... Пусть  $X_j$  – изменение координаты частицы в результате  $j$ -го столкновения,  $S(t) = \sum_{j=1}^{N(t)} X_j$  – координата частицы в момент времени  $t$ . Если среда однородна, то это обобщённый пуассоновский процесс. Если среда очень неоднородна, то, вообще говоря,  $N(t)$  будет обобщённым процессом Кокса и предельное распределение не обязательно будет нормальным.

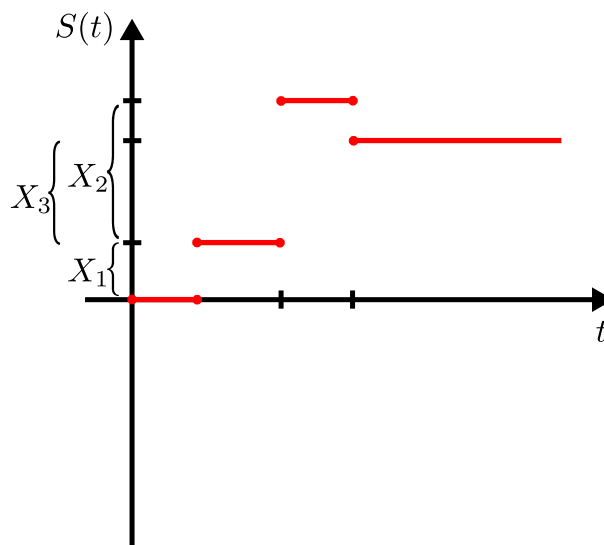


Рис. 10: Типичная траектория обобщённого процесса Кокса

## Лекция 12

### Предельные теоремы для обобщённых процессов Кокса

С этого момента будем предполагать, что  $EX_j = a$ ,  $D X_j = \sigma^2$ . Тогда

$$ES(t) = aE\Lambda(t), \quad D S(t) = E\Lambda(t)(a^2 + \sigma^2) + a^2 D \Lambda(t).$$

Далее будем предполагать, что  $a = 0$ .

**Теорема 23** (Ц.П.Т. для обобщённых процессов Кокса). Пусть функция  $d(t)$  неограниченно возрастает при  $t \rightarrow \infty$ ,  $\Lambda(t) \xrightarrow[t \rightarrow \infty]{P} \infty$ . Тогда

$$\frac{S(t)}{\sigma \sqrt{d(t)}} \Rightarrow Z$$

тогда и только тогда, когда:

- 1) существует такая неотрицательная случайная величина  $U$ , что  $P(Z < x) = \int_0^\infty \Phi\left(\frac{x}{\sqrt{y}}\right) dP(U < y)$ ;
- 2)  $\frac{\Lambda(t)}{d(t)} \Rightarrow U$ .

**Следствие 2.** В условиях Ц.П.Т. для обобщённых процессов Кокса  $Z \sim N(0, 1)$  тогда и только тогда, когда  $\frac{\Lambda(t)}{d(t)} \Rightarrow 1$ . Это следует из идентифицируемости масштабных смесей нормальных законов.

Функция распределения  $G_{\alpha, \theta}(x)$  называется **строго устойчивой**, если соответствующая характеристическая функция

$$g_{\alpha, \theta}(t) = \exp \left\{ -|t|^\alpha \exp \left\{ -i \frac{\pi \theta \alpha}{2} \operatorname{sgn} t \right\} \right\},$$

где  $|\theta| \leq \tilde{\theta} = \min \left( 1, \frac{2}{\alpha} - 1 \right)$ . Можете вернуться к определению устойчивого распределения (см. (6)) и понять, что строгая устойчивость – это частный случай устойчивости.

Как частный случай Ц.П.Т. сформулируем критерий сходимости к строго устойчивым законам.

**Теорема 24** (критерий сходимости одномерных распределений обобщённых процессов Кокса к строго устойчивым законам). Пусть функция  $d(t)$  неограниченно возрастает при  $t \rightarrow \infty$ ,  $\Lambda(t) \xrightarrow[t \rightarrow \infty]{P} \infty$ . Тогда

$$P\left(\frac{S(t)}{\sigma\sqrt{d(t)}} < x\right) \xrightarrow[t \rightarrow \infty]{} G_{\alpha,0}(x)$$

тогда и только тогда, когда

$$P\left(\frac{\Lambda(t)}{d(t)} < x\right) \xrightarrow[t \rightarrow \infty]{} G_{\frac{\alpha}{2},1}(x).$$

Оказывается, что между разными устойчивыми законами есть связь. Например,  $G_{\alpha,0}(x) = \int_a^\infty \Phi\left(\frac{x}{\sqrt{y}}\right) dG_{\frac{\alpha}{2},1}(y)$ . (Этот факт доказан в книге Золотарёва за 1983 год.)

Из этого факта и идентифицируемости масштабных смесей нормальных законов сразу же следует вышеприведённый критерий.

Теперь рассмотрим дискретный случай этого процесса. Пусть есть моменты времени  $1, 2, \dots, n$  и управляющий процесс зависит не от  $t$ , а от  $n$ :  $\Lambda(n) = \sum_{i=1}^n Z_i$ , где  $Z_i$  – независимые одинаково распределённые случайные величины (приращения процесса  $\Lambda(n)$ ). Тогда  $S(n) = \sum_{j=1}^{N_1(\Lambda(n))} X_j$ .

**Теорема 25.**

$$P\left(\frac{S(n)}{\delta_n} < x\right) \xrightarrow[n \rightarrow \infty]{} G_{\alpha,0}(x)$$

при некотором выборе нормирующей последовательности  $\{\delta_n\}$  тогда и только тогда, когда

$$\forall k > 0 : \lim_{x \rightarrow \infty} \frac{P(Z_i > x)}{P(Z_i > kx)} = k^{\frac{\alpha}{2}}.$$

Всё это время мы рассматривали нулевое математическое ожидание. Пусть теперь  $EX_j = a \neq 0$ . Сформулируем теперь закон больших чисел (он верен и при нулевом  $a$ , но ненулевой случай интереснее).

**Теорема 26** (З.Б.Ч. для обобщённых процессов Кокса). Пусть  $\Lambda(t) \xrightarrow[t \rightarrow \infty]{P} \infty$ . Тогда

$$\frac{S(t)}{t} \Rightarrow Z$$

тогда и только тогда, когда

$$\frac{\Lambda(t)}{t} \Rightarrow U$$

и при этом  $Z \stackrel{d}{=} aU$ .



## Свойства масштабных смесей нормальных законов

Вспомним функцию распределения и его плотность:

$$H(x) = E\Phi\left(\frac{x}{Y}\right) = \int_0^\infty \Phi\left(\frac{x}{y}\right) dP(Y < y), \quad h(x) = E\frac{1}{Y}\varphi\left(\frac{x}{Y}\right) = \int_0^\infty \frac{1}{y}\varphi\left(\frac{x}{y}\right) dP(Y < y),$$

где  $Y = \sqrt{U}$ . Ну и дискретный случай:

$$H(x) = \sum_{i=1}^n p_i \Phi\left(\frac{x}{y_i}\right), \quad h(x) = \sum_{i=1}^n \frac{p_i}{y_i} \varphi\left(\frac{x}{y_i}\right)$$

Для плотности распределения существует понятие «островершинности»: если у графика плотности острая вершина, то большая часть массы уходит на хвосты, и они становятся тяжёлыми; если у графика пологая вершина, то большая часть массы сосредоточена в самой вершине.

Один из показателей измерения «островершинности» – это **коэффициент эксцесса**. Мы предполагаем, что существует четвёртый момент  $EZ^4 < \infty$ . Тогда коэффициент эксцесса  $\kappa(Z) = E\left(\frac{Z - EZ}{\sqrt{DZ}}\right)^4$ . Чем он больше, тем острее вершина.

Обычно четырёх моментов достаточно для того, чтобы получить представление о модельном распределении, поэтому у них есть названия (математическое ожидание, дисперсия, коэффициент асимметрии, коэффициент эксцесса).

Если  $Z \sim N(0, 1)$ , то  $\kappa(Z) = 3$ . Если у распределения  $\kappa(Z) > 3$ , то вершина более острая, хвосты более тяжёлая. Если  $\kappa(Z) < 3$ , то более пологая. Так уж принято, что коэффициент эксцесса сравнивают с нормальным, поэтому обычно из него вычитают 3. То есть получается некоторый  $\gamma(Z) = \kappa(Z) - 3$ .

Вернёмся к масштабным смесям. Функция распределения соответствует случайно величине  $Z = XY$ , где  $X \sim N(0, 1)$ , а  $Y$  имеет некоторое заданное распределение. Такое утверждение мы конечно можем сделать только при условии, что вероятностное пространство достаточно богато.

**Следствие 3.** Пусть  $X$  и  $Y$  независимы,  $P(Y > 0) = 1$ ,  $EX = 0$ . Тогда  $\kappa(XY) \geq \kappa(X)$  и более того  $\kappa(XY) = \kappa(X) \iff P(Y = c) = 1$ .

*Доказательство.*

$$\begin{aligned} \kappa(XY) &= E\left(\frac{XY - EXY}{\sqrt{D(XY)}}\right)^4 = E\left(\frac{XY - EXEY}{\sqrt{E(XY - EXEY)^2}}\right)^4 = \\ &= \frac{E(XY)^4}{(E(XY)^2)^2} = \frac{EX^4EY^4}{(EX^2)^2(EY^2)^2} = \kappa(X) \frac{EY^4}{(EY^2)^2}, \end{aligned}$$

ну и так как  $EY^4 \geq (EY^2)^2$  (и  $EY^4 = (EY^2)^2 \iff P(Y = c) = 1$ ), то мы и получаем данное неравенство. ■

Небольшое замечание: если говорить о процессах Кокса, то не стоит забывать, что  $Y = \sqrt{U}$  и для  $U$  должен существовать второй момент, а не четвёртый.

Итак, как же будет выглядеть хвост случайной величины  $Z = X\sqrt{U}$ :  $P(Z > x) = 1 - E\Phi\left(\frac{X}{\sqrt{U}}\right)$ ? Пусть  $p_Z(x)$  – плотность  $Z$ .

**Следствие 4.**

$$P(Z > x) \geq 1 - \Phi\left(\sqrt{2\pi}p_Z(0)x\right).$$

А если  $EU^{-\frac{1}{2}} = 1$ , то  $p_Z(0) = \int_0^\infty \frac{1}{\sqrt{y}} \varphi\left(\frac{0}{\sqrt{y}}\right) dP(U < y) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{y}} dP(U < y) =$

$$\frac{1}{\sqrt{2\pi}} EU^{-\frac{1}{2}} = \frac{1}{\sqrt{2\pi}} \text{ и тогда}$$

$$P(Z > x) \geq 1 - \Phi(x).$$

Это свойство указывает нам на то, что всегда получаются более плохие распределение в классе масштабных смесей, чем чисто нормальные.

## Лекция 13

### Устойчивость смесей нормальных законов относительно смешивающего распределения

#### Прямая задача

Введём понятие **равномерного расстояния между случайными величинами**:  
 $\rho(X, Y) = \rho(F_X, F_Y) = \sup_x |F_X(x) - F_Y(x)|$ .

У нас есть  $X_i \sim F_i(x)$  – смешиваемое распределение,  $Y_i \sim G_i(x)$  – смешивающее распределение,  $X_i$  и  $Y_i$  независимы,  $P(Y_i \geq 0) = 1$ . Тогда прямой задачей является получить оценку для  $\rho(Z_1, Z_2)$  в терминах  $\rho(X_1, X_2)$  и  $\rho(Y_1, Y_2)$ .

Рассмотрим такую задачу: выразить  $\rho\left(\frac{X_1}{Y_1}, \frac{X_2}{Y_2}\right)$  через  $\rho(X_1, X_2)$  и  $\rho(Y_1, Y_2)$ .

$$\begin{aligned} P\left(\frac{X}{Y} < x\right) &= P(X < xY) = P(X < 0, Y = 0) + P(X < x, Y > 0) = \\ &= P(X < 0) P(Y = 0) + \int_{0+}^{\infty} F_X(xy) dF_Y(y) = \\ &= F_X(0)(F_Y(0+) - F_Y(0)) + \int_{0+}^{\infty} F_X(xy) dF_Y(y). \end{aligned}$$

Пусть //починить что внизу//

$$\begin{aligned} \Delta x &= \left| P\left(\frac{X_1}{Y_1} < x\right) - P\left(\frac{X_2}{Y_2} < x\right) \right| = \\ &= \left| F_1(0)(G_1(0+) - G_1(0)) + \int_{0+}^{\infty} F_1(xy) dG_1(y) - F_2(0)(G_2(0+) - G_2(0)) - \int_{0+}^{\infty} F_2(xy) dG_2(y) \right| \leq \\ &\leq \underbrace{P(X_1 < 0) P(Y_1 = 0) + P(X_2 < 0) P(Y_2 = 0)}_{Q_1} + \underbrace{\left| \int_{0+}^{\infty} F_1(xy) dG_1(y) - \int_{0+}^{\infty} F_2(xy) dG_2(y) \right|}_{Q_2}. \end{aligned}$$

Рассмотрим  $Q_2$ :

$$Q_2 \leq \underbrace{\left| \int_{0+}^{\infty} (F_1(xy) - F_2(xy)) dG_1(y) \right|}_{Q_{21}} + \underbrace{\left| \int_{0+}^{\infty} F_2(xy) d(G_1(y) - G_2(y)) \right|}_{Q_{22}},$$

$$Q_{21} \leq \int_{0+}^{\infty} \rho(X_1, X_2) dG_1(y) = \rho(X_1, X_2) P(Y_1 > 0),$$

Переходя к  $Q_{22}$ , мы будем рассматривать две ситуации:

$$\begin{aligned} \text{а) } x = 0 : Q_{22} &= \int_{0+}^{\infty} F_2(0) d(G_1(y) - G_2(y)) = \\ &= |P(X_2 < 0)(P(Y_1 > 0) - P(Y_2 > 0))| \leq P(X_2 < 0) |P(Y_1 = 0) - P(Y_2 = 0)|; \end{aligned}$$

$$\begin{aligned} \text{б) } x \neq 0 : Q_{22} &= F_2(xy)(G_1(y) - G_2(y)) \Big|_{0+}^{\infty} - \int_{0+}^{\infty} (G_1(y) - G_2(y)) d_y F_2(xy) \leq \\ &\leq \underbrace{P(X_2 \leq 0) |P(Y_1 = 0) - P(Y_2 = 0)|}_{Q_{221}} + \underbrace{\rho(Y_1, Y_2) \int_{0+}^{\infty} d_y F_2(xy)}_{Q_{222}}, \end{aligned}$$

$$Q_{222} = \rho(Y_1, Y_2) \underbrace{\int_{0+}^{\infty} d_y F_2(xy)}_{I(x)}, \quad I(x) = \int_{0+}^{\infty} d_y F_2(xy) = \begin{cases} P(X_2 > 0), & x > 0, \\ P(X_2 < 0), & x < 0. \end{cases}$$

Используя все полученные выкладки, сформулируем утверждение.

**Утверждение 7.** Пусть  $X_i \sim F_i(x)$ ,  $Y_i \sim G_i(x)$ ,  $P(Y_i \geq 0) = 1$ ,  $X_i$  и  $Y_i$  независимы. Тогда

$$\begin{aligned} \rho\left(\frac{X_1}{Y_1}, \frac{X_1}{Y_1}\right) &\leq P(X_1 < 0) P(Y_1 = 0) + P(X_2 < 0) P(Y_2 = 0) + \rho(X_1, X_2) P(Y_1 > 0) + \\ &+ P(X_2 \leq 0) |P(Y_1 = 0) - P(Y_2 = 0)| + \rho(Y_1, Y_2) I(x). \end{aligned}$$

**Следствие 5.** Пусть  $P(Y_1 = 0) = P(Y_2 = 0) = 0$ . Тогда

$$\rho\left(\frac{X_1}{Y_1}, \frac{X_1}{Y_1}\right) \leq \rho(X_1, X_2) + \rho(Y_1, Y_2) \min_{i=1,2} \max(P(X_i > 0), P(X_i < 0)).$$

Более того:

$$\rho\left(\frac{X_1}{Y_1}, \frac{X_1}{Y_1}\right) \leq \rho(X_1, X_2) + \rho(Y_1, Y_2).$$

**Следствие 6.** Пусть  $P(Y_1 = 0) = P(Y_2 = 0) = 0$  и хотя бы для одного  $X_i$  выполнено  $P(X_i > 0) = P(X_i < 0)$ . Тогда

$$\rho\left(\frac{X_1}{Y_1}, \frac{X_1}{Y_1}\right) \leq \rho(X_1, X_2) + \frac{1}{2} \rho(Y_1, Y_2).$$

Пусть  $P(Y_1 = 0) = P(Y_2 = 0) = 0$ . Рассмотрим следующее свойство:

$$\begin{aligned}\rho(Y_1^{-1}, Y_2^{-1}) &= \sup_{x>0} \left| P\left(\frac{1}{Y_1} < x\right) - P\left(\frac{1}{Y_2} < x\right) \right| = \sup_{x>0} \left| P\left(Y_1 > \frac{1}{x}\right) - P\left(Y_2 > \frac{1}{x}\right) \right| = \\ &= \sup_{x>0} \left| P\left(Y_1 \leq \frac{1}{x}\right) - P\left(Y_2 \leq \frac{1}{x}\right) \right| = \sup_{z>0} |P(Y_1 < z) - P(Y_2 < z)| = \rho(Y_1, Y_2).\end{aligned}$$

А это означает, что в вышеприведённых следствиях нам ничего не мешает рассматривать не дроби, а произведения.

## Обратная задача

Сразу будем рассматривать масштабные смеси нормальных законов:  $Z_i = X_i\sqrt{U_i}$ ,  $X_i$  и  $U_i$  независимы,  $X_i \sim N(0, 1)$ ,  $P(U_i > 0) = 1$ . Имеем характеристическую функцию  $Z_i$ :

$$f_i(t) = Ee^{itX_i\sqrt{U_i}} = E\left(Ee^{itX_i\sqrt{U_i}} | U_i\right) = Ee^{-\frac{t^2 U_i}{2}}.$$

С другой стороны:

$$\psi_{U_i}(s) = Ee^{sU_i} \text{ — преобразование Лапласа-Стилтьеса.}$$

Соответственно

$$f_i(t) = \psi_{U_i}\left(\frac{t^2}{2}\right).$$

Далее рассмотрим формулу обращения

$$\begin{aligned}P(Y_i < x) &= \lim_{t \rightarrow \infty} \sum_{n < \sqrt{2tx}} \frac{(-1)^n (2t)^{\frac{n}{2}}}{n!} f_i^{(n)}(\sqrt{2t}), \\ |P(Y_1 < x) - P(Y_2 < x)| &\leq \lim_{t \rightarrow \infty} \sum_{n < \sqrt{2tx}} \frac{(-1)^n (2t)^{\frac{n}{2}}}{n!} \left| f_1^{(n)}(\sqrt{2t}) - f_2^{(n)}(\sqrt{2t}) \right|.\end{aligned}$$

Это значит, что для оценивания разности функциями смешивающих распределений недостаточно информации о близости только характеристических функций, также необходимо знать о близости всех производных этих функций. Это наводит на мысль, что обратная задача не будет корректно поставлена.

**Пример.** Пусть  $Y_1$  вырождена в 1, а  $Y_2$  вырождена в  $1 + \varepsilon'$  (где  $\varepsilon' = \varepsilon\sqrt{2\pi e}$ , а  $\varepsilon$  — некоторое малое число). Рассмотрим разность смесей:

$$\begin{aligned}\left| E\Phi\left(\frac{x}{Y_1}\right) - E\Phi\left(\frac{x}{Y_2}\right) \right| &= \left| \Phi(x) - \Phi\left(\frac{x}{1 + \varepsilon'}\right) \right| \leq \left| \frac{\varepsilon' x}{1 + \varepsilon'} \right| \varphi\left(x \left( \delta + (1 - \delta) \frac{1}{1 + \varepsilon'} \right)\right) \leq \\ &\leq \left| \frac{\varepsilon' x}{1 + \varepsilon'} \varphi\left(\frac{x}{1 + \varepsilon'}\right) \right| \leq \sup_{x' > 0} (\varepsilon' x' \varphi(x')) \leq \frac{\varepsilon'}{\sqrt{2\pi e}} = \varepsilon\end{aligned}$$

где  $\delta \in (0, 1)$ , а  $x' = \frac{x}{1 + \varepsilon'}$ .

Из данных выкладок мы можем сделать вывод, что смеси отличаются не больше чем на  $\varepsilon$  по равномерной метрике и обратная задача состоит в том, чтобы понять как отличаются смешивающие распределения. И оказывается, что они отличаются очень сильно:  $\rho(Y_1, Y_2) = 1$ . Значит в данном примере обратная задача поставлена некорректна.

Это пример также наводит на мысль, что равномерную метрику здесь применять неправильно. Тогда можно, например, рассматривать сглаженные распределения или другую метрику.

## Метрика Леви

**Метрикой Леви** называется

$$L(F, G) = L(G, F) = \left\{ \inf h : h > 0, F(x - h) - h \leq G(x) \leq F(x + h) + h \right\}.$$

По сути она представляет из себя максимальную длину стороны квадрата, чьи стороны параллельны осям координат и который можно вписать между двумя кривыми распределения.

Будем рассматривать двухкомпонентную смесь. Введём функцию  $F_{\sigma,p}(x) = p\Phi(x) + (1 - p)\Phi(\sigma x)$ , где  $\sigma \in (0, 1)$ . Такая модель была введена в 60-ых годах американским математиком Тьюки для описания выборки с «загрязнением». Считается, что «чистые» наблюдения берутся из нормального закона с вероятностью  $p$ , а  $(1 - p)$  – вероятность появления выброса, который моделируется нормальным распределением с большей дисперсией.

Такое распределение возникает, например, при моделировании принятия некоторого сигнала. Также такую модель используют для построения ритмограммы из кардиограммы.

## Лекция 14

### Обратная задача (продолжение)

Нетрудно убедиться, что функция распределения  $F_{\sigma,p}(x) = p\Phi(x) + (1-p)\Phi(\sigma x)$ , где  $\sigma \in (0, 1)$ , соответствует случайной величине  $Z = XW_{\sigma,p}$ , где  $X$  и  $W_{\sigma,p}$  независимы, а

$$W_{\sigma,p} = \begin{cases} 1, & \text{с вероятностью } p, \\ \sigma^{-1}, & \text{с вероятностью } 1-p. \end{cases}$$

Пусть  $L(\Phi, F_{\sigma,p}) < \varepsilon$  и

$$U_{\sigma,p} = \begin{cases} 1, & \text{с вероятностью } p, \\ \sigma, & \text{с вероятностью } 1-p; \end{cases}$$

то есть по сути обратная  $W_{\sigma,p}$  величина. При таком условии известно, что

$$\rho(\Phi, F_{\sigma,p}) \leq \left(1 + \frac{1}{\sqrt{2\pi}}\right) L(\Phi, F_{\sigma,p}) < \varepsilon \left(1 + \frac{1}{\sqrt{2\pi}}\right).$$

Рассмотрим это расстояние:

$$\begin{aligned} \rho(\Phi, F_{\sigma,p}) &= (1-p) \sup_x |\Phi(x) - \Phi(\sigma x)| \geq (1-p) |\Phi(1) - \Phi(\sigma)| = \\ &= (1-p)(1-\sigma)\varphi(\sigma') \geq (1-p)(1-\sigma)\varphi(1) = \frac{(1-p)(1-\sigma)}{\sqrt{2\pi e}}, \end{aligned}$$

где  $\sigma' \in (0, 1)$ . Из этого следует, что

$$(1-p)(1-\sigma) \leq \sqrt{e}(1 + \sqrt{2\pi})\varepsilon.$$

Но давайте перейдём к метрике Леви:

$$L(U_{\sigma,p}, 1) = \min(1-p, 1-\sigma),$$

$$L^2(U_{\sigma,p}, 1) = (\min(1-p, 1-\sigma))^2 \leq (1-p)(1-\sigma).$$

**Утверждение 8.** Пусть  $L(\Phi, F_{\sigma,p}) \leq \varepsilon$ . Тогда

$$L(U_{\sigma,p}, 1) \leq c\sqrt{\varepsilon},$$

где  $c = (\sqrt{e}(1 + \sqrt{2\pi}))^{\frac{1}{2}} < 2.5$ .

Аналогично можно рассмотреть сдвиг-масштабные смеси. Введём функцию распределения  $G_{a,p}(x) = p\Phi(x) + (1-p)\Phi(x-a)$ , которая соответствует случайной величине  $Z = X + V_{a,p}$ , где  $X$  и  $V_{a,p}$  независимы, а

$$V_{a,p} = \begin{cases} 0, & \text{с вероятностью } p; \\ a, & \text{с вероятностью } 1-p. \end{cases}$$

Предположим, что  $|a| < 1$ .

**Утверждение 9.** Пусть  $L(\Phi, G_{a,p}) \leq \varepsilon$ . Тогда

$$L(V_{a,p}, 0) \leq c\sqrt{\varepsilon},$$

где  $c = (\sqrt{e}(1 + \sqrt{2\pi}))^{\frac{1}{2}} < 2.5$ .

Теперь вернёмся к масштабным смесям, только заменим  $\Phi$  на  $F_{\sigma,q}$ . Тогда оценка снизу останется той же, а само равномерное расстояние изменится:

$$\rho(F_{\sigma,q}, F_{\sigma,p}) = |p - q| \sup_x |\Phi(x) - \Phi(\sigma x)| \geq \dots$$

И мы придём к похожему неравенству:

$$|p - q|(1 - \sigma) \leq \sqrt{e}(1 + \sqrt{2\pi})\varepsilon.$$

К тому же расстояние Леви будет тоже похожим:

$$\begin{aligned} L(U_{\sigma,p}, 1) &= \min(|p - q|, 1 - \sigma), \\ L^2(U_{\sigma,p}, 1) &= (\min(|p - q|, 1 - \sigma))^2 \leq |p - q|(1 - \sigma). \end{aligned}$$

Далее приведём ряд утверждений, в которых  $\sigma, \sigma_i \in (0, 1)$  и  $|a|, |a_i| < 1$ .

**Утверждение 10.** Пусть  $L(F_{\sigma,q}, F_{\sigma,p}) \leq \varepsilon$ . Тогда

$$L(U_{\sigma,q}, U_{\sigma,p}) \leq c\sqrt{\varepsilon},$$

где  $c = (\sqrt{e}(1 + \sqrt{2\pi}))^{\frac{1}{2}} < 2.5$ .

**Утверждение 11.** Пусть  $L(G_{a,q}, G_{a,p}) \leq \varepsilon$ . Тогда

$$L(V_{a,q}, V_{a,p}) \leq c\sqrt{\varepsilon},$$

где  $c = (\sqrt{e}(1 + \sqrt{2\pi}))^{\frac{1}{2}} < 2.5$ .

**Утверждение 12.** Пусть  $L(F_{\sigma_1,p}, F_{\sigma_2,p}) \leq \varepsilon$ . Тогда

$$L(U_{\sigma_1,p}, U_{\sigma_2,p}) \leq c\sqrt{\varepsilon},$$

где  $c = (\sqrt{e}(1 + \sqrt{2\pi}))^{\frac{1}{2}} < 2.5$ .

**Утверждение 13.** Пусть  $L(G_{a_1,p}, G_{a_2,p}) \leq \varepsilon$ . Тогда

$$L(V_{a_1,p}, V_{a_2,p}) \leq c\sqrt{\varepsilon},$$

где  $c = (\sqrt{e}(1 + \sqrt{2\pi}))^{\frac{1}{2}} < 2.5$ .



## Моделирование распределений приращений финансовых индексов смесями нормальных законов

Считается, что существует процесс, согласно которому меняется  $S$  (цена базового актива).  $\Delta S$  – относительное приращение цены актива, которое пропорционально самой цене.

$$\frac{\Delta S}{S} = r\Delta t + \sigma\Delta W_t,$$

где  $r$  – безрисковая ставка,  $\sigma$  – волатильность,  $W_t$  – винеровский процесс. Так задаётся **модель Блэка-Шоулза**. Далее происходит предельный переход при  $\Delta t \rightarrow 0$ , после которого получается стохастическое дифференциальное уравнение, на основе которого выводятся цены для финансовых производных. Основной недостаток такой модели – это независимость  $\sigma$  от времени. Но существуют различные модификации, благодаря которым данную модель используют до сих пор.

Другой подход состоит в том, что не существует какого-то базового процесса, и цена формируется в каждом контракте, и на неё действуют различные игроки в своих интересах.  $P_j$  – цена  $j$ -го контракта,  $\tau_j$  – случайный момент заключения  $j$ -го контракта,  $P_0$  – начальная цена. То есть таким образом задаётся точечный процесс  $(P_j, \tau_j)$ . Такой процесс можно описать и в терминах непрерывного времени. Пусть  $N(\tau)$  – число контрактов, заключённых к моменту  $\tau$ , и эта величина постоянно равна на отрезках  $[\tau_{j-1}, \tau_j)$ , а также на этих отрезках постоянна цена  $P(\tau) = P_j$ .

На самом деле формирует цену величина  $\frac{P_j}{P_{j-1}}$ . Тогда

$$P(\tau) = \prod_{j=1}^{N(\tau)} \frac{P_j}{P_{j-1}}, \quad S(\tau) = \ln P(\tau) = \sum_{j=1}^{N(\tau)} \underbrace{(\ln P_j - \ln P_{j-1})}_{X_j}.$$

Сделаем предположение, что  $EX_j = 0$ ,  $D X_j = \sigma^2$ ,  $X_j$  независимы,  $N(\tau) = N_1(\Lambda(\tau))$ , где  $\Lambda(\tau)$  – управляющий процесс и  $E\Lambda(\tau) = \lambda\tau$ . Тогда вводится зависимость цены от  $\lambda$ , а также минимальная и максимальная цена на отрезке времени:

$$\underline{P}(\lambda, t) = \min_{0 < \tau < t} P(\lambda, t), \quad \bar{P}(\lambda, t) = \max_{0 < \tau < t} P(\lambda, t).$$

**Теорема 27.** Пусть  $X_j = \ln P_j - \ln P_{j-1}$  – независимые одинаково распределённые случайные величины,  $EX_j = 0$ ,  $D X_j = \sigma^2$ ,  $E\Lambda(t) = \lambda t$ . Тогда эквиваленты следующие утверждения:

- 1)  $P \left( \frac{\ln P(\lambda, t) - \ln P(0)}{\sigma\sqrt{\lambda t}} < x \right) \xrightarrow{\lambda \rightarrow \infty} F(x);$
- 2)  $P \left( \frac{\ln \bar{P}(\lambda, t) - \ln P(0)}{\sigma\sqrt{\lambda t}} < x \right) \xrightarrow{\lambda \rightarrow \infty} \bar{F}(x);$

$$3) \mathbb{P} \left( \frac{\ln \underline{P}(\lambda, t) - \ln P(0)}{\sigma \sqrt{\lambda t}} < x \right) \xrightarrow{\lambda \rightarrow \infty} \underline{F}(x);$$

$$4) \frac{\Lambda(t)}{\lambda t} \xRightarrow{\lambda \rightarrow \infty} U.$$

$$\text{Причём } F(x) = \mathbb{E} \Phi \left( \frac{x}{\sqrt{U}} \right), \bar{F}(x) = 2 \mathbb{E} \Phi \left( \frac{\max(0, x)}{\sqrt{U}} \right) - 1, \underline{F}(x) = 2 \left( 1 - \mathbb{E} \Phi \left( \frac{-\min(0, x)}{\sqrt{U}} \right) \right).$$

А если устремить к бесконечности  $t$ , а не  $\lambda$ , то разумно предположить, что

$$\Lambda(t) = \sum_{i=1}^{n(t)} \Lambda_i, \text{ где } \Lambda_i - \text{независимые одинаково распределённые случайные величины.}$$

Тогда  $\frac{\Lambda(t)}{\lambda t} \xRightarrow{t \rightarrow \infty} 1$  и мы придём к модели Блэка-Шоулза.



ФАКУЛЬТЕТ  
ВЫЧИСЛИТЕЛЬНОЙ  
МАТЕМАТИКИ И  
КИБЕРНЕТИКИ  
МГУ ИМЕНИ  
М.В. ЛОМОНОСОВА



*teach-in*  
ЛЕКЦИИ УЧЕНЫХ МГУ