Теория вероятностей. II семестр 3-й курс, КБ+КН+ФИИТ УрФУ. Конспект лекций

Ю.В. Авербух, Хлопин Д.В.

8 июня 2019 г.

Примечания к тексту

Данный конспект лекций примерно соответствует тому, что было на парах. Упомянутые здесь задачи на баллы не актуальны, для этого есть отдельный задачник. Но зарабатывать баллы можно, вылавливая опечатки и дыры в логике. Ищите их, пересылайте и обрящете баллы (потеря скобки будет оцениваться в $0,2$ балла, двух скобок — в $0,3$ балла). Грамматика и синтаксис тоже оценивается — в $0,04$ —, но отправлять ее минимум десятком с точным указанием где, что и как.

Адрес актуальной версии: https://yadi.sk/i/Ty5oXQhipyLzRQ

Хлопин Д.В.
glukanat@mail.ru

Об экзамене

Оценка за экзамен равна минимуму двух оценок за теорию и практику. За теорию оценка ставится следующим образом: 6 или больше — тройка, 8 или больше — четверка, 9 или больше + решенная задача "мартингал— пятерка.

Существуют три даты для сдачи:

- 10 июня (автоматы + 7 уже избранных для мартингалов),
- 13 июня (автоматы + билет + мартингалы),
- 14 июня (автоматы + билет + мартингалы).

Автомат означает, что Вы согласны с уже имеющейся оценкой. "Мартингал" может быть оказаться задачей совсем не на тему мартингалов и выдается индивидуально каждому.

Билет — это кусок теории, один параграф этого конспекта целиком (примерно с одну лекцию размером), базовая стоимость — 2 балла. Билет из списка ниже Вы выбираете и готовите сами, но в один день все должны сдавать разные билеты.

1. Неравенства концентрации меры
2. Сходимость случайных величин
3. Гарантии почти всюду
4. Слабая сходимость
5. Характеристические функции
6. Центральная предельная теорема
7. Выборочное распределение
8. Точечное оценивание
9. Задача различия двух гипотез
10. Марковские цепи с дикретным временем
11. Марковские цепи с непрерывным временем
12. Мартингалы
13. Теорема о произвольном выборе
14. Сходимость мартингалов
15. Задача об оптимальной остановке

На экзамене можно пользоваться книгами на любых носителях.

Баллы по-прежнему можно заработать, решая задачки и находя опечатки. В 21-00 9 июня их прием будет приостановлен до консультации.

Список литературы

- А.Н. Ширяев. Вероятность-I. М.: МЦНМО, 2011.
- А.Н. Ширяев. Вероятность-II. М.: МЦНМО, 2011.
- Л.Б. Коралов, Я.Г. Синай. Теория вероятностей и случайные процессы. М.: МНЦМО, 2013.
- Б.В. Гнеденко. Курс теории вероятностей. М.: Едиториал УРСС, 2005.
- В.П Чистяков. Курс теории вероятностей. М.: Наука. 1978.
- В.И. Богачев. Введение в теорию вероятностей: конспект лекций. ВШЭ, 2016. https://math.hse.ru/data/2016/11/28/1112782911/TeorVer16.pdf
- G. Casella, R.L. Berger. Statistical Inference. Pacific Grove, 1990.
- Н.И. Чернова Лекции по математической статистике. Новосибирск. НГУ, 2007.
- А.В. Прохоров. Курс лекций по математической статистике. Москва. МГУ. 2006

Список обозначений

$\triangleq$ — равно по определению.
$\mathbf{1}_A(x)$ — индикаторная функция A, определяемая по правилу:

$$\mathbf{1}_A(x) \triangleq \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

$\mathbb{E}\xi$ — математическое ожидание случайной величины ξ.
$D\xi$ — дисперсия случайной величины ξ.
$\xrightarrow{\text{п.в.}}$ — сходимость почти наверное (см. Определение 2).
$\xrightarrow{P}$ — сходимость по вероятности (см. Определение 1).
$\xrightarrow{L^p}$ — сходимость в среднем порядка p (см. Определение 3).
$\xrightarrow{w}$ — слабая сходимость (см. Определение 4).
$\xrightarrow{d}$ — сходимость по распределению (см. Определение 5).
$C_b(\mathbb{R})$ — множество непрерывных ограниченных функций из $\mathbb{R}$ в $\mathbb{R}$.
$C_0^\infty(\mathbb{R})$ — множество бесконечное число раз дифференцируемых финитных функций из $\mathbb{R}$ в $\mathbb{R}$.
$x \vee y = \max\{x, y\}$.
$x \wedge y = \min\{x, y\}$.

Оглавление

Вероятности и случайные величины

В этом параграфе мы кратко напомним основные понятия теории вероятности, введенные в 1-м семестре.

Как известно, в рамках аксиоматики Колмогорова основным объектом теории вероятностей являются вероятностные пространства. Вероятностным пространством называют тройку $(\Omega, \mathcal{F}, \mathbb{P})$, где

- Ω — некоторое множество, называемое множеством событий;
- $\mathcal{F}$ — семейство подмножеств Ω, образующее σ-алгебру; $\mathcal{F}$ называют σ-алгеброй событий.
- $\mathbb{P}$ — функция из $\mathcal{F}$ в $[0,1]$, удовлетворяющая условию σ-аддитивности; саму функцию $\mathbb{P}$ называют вероятностью.

Пару $(\Omega, \mathcal{F})$ часто называют измеримым пространством. На каждом измеримом пространстве в случае $|\Omega| > 1$ вероятность можно определить неединственным способом (докажите это!).

Интерпретация вероятностного пространства следующая. Предположим, что проводится некоторый эксперимент, исход которого не предрешен. Тогда,

- Ω — множество исходов, которые реально могут произойти (полученное число, вектор, последовательность чисел);
- $\mathcal{F}$ — набор событий, про каждое из которых мы можем сказать, случилось оно или нет (был ли результат меньше некоторого числа, лежит ли результат в круге, имеет ли получившаяся последовательность 0 на k-м месте и т.д.);
- $\mathbb{P}$ — оценка, насколько вероятно то или иное событие из $\mathcal{F}$.

Несмотря на универсальность, получившаяся конструкция во многих случаях является достаточно громоздкой. Для того, чтобы уяснить это, читатель может построить вероятностное пространство для опыта, состоящего в подбрасывании монетки до первого выпадения орла. В этом случае Ω является множеством всех последовательностей, состоящих из 0 и 1. А как будут выглядеть $\mathcal{F}$ и $\mathbb{P}$?

Во многих случаях достаточно обойтись более простыми конструкциями. В частности, многие задачи сводятся к случаю, когда $\Omega = \mathbb{R}$ (множеству действительных чисел). В этом случае стандартно в качестве σ-алгебры событий выбирают $\mathcal{B}(\mathbb{R})$ — борелевскую σ-алгебру событий (т.е. минимальную σ-алгебру событий, содержащую все открытые множества). В дальнейшем для краткости будем говорить о вероятности на $\mathbb{R}$, хотя правильно говорить о вероятностях на σ-алгебре $\mathcal{B}(\mathbb{R})$. Вероятности на $\mathcal{B}(\mathbb{R})$ будем обозначать греческими буквами μ, ν и т.д. Заметим, что вероятности на $\mathbb{R}$ намного легче строятся и “обозримы” по сравнению с исходными конструкциями.

Связь между исходным вероятностным пространством и вероятностями на $\mathbb{R}$ задается случайными величинами. Напомним, что случайной величиной называется функция $\xi : \Omega \to \mathbb{R}$, обладающая свойством *измеримости*, т.е. для любого множества $A \subset \mathcal{B}(\mathbb{R})$

$$\xi^{-1}(A) = \{\omega \in \Omega : \xi(\omega) \in A\} \in \mathcal{F}.$$

Свойство измеримости необходимо для того, чтобы мы могли узнать: выполнено или нет событие $\xi(\omega) \in A$.

Каждая случайная величина ξ определяет вероятность на $\mathbb{R}$ следующим образом:

$$\mu_\xi(A) \stackrel{\triangle}{=} \mathbb{P}(\xi^{-1}(A)).$$

Иногда эту вероятность обозначают также $\mathbb{P} \circ \xi^{-1}$ или $\xi\#\mathbb{P}$ и называют *вероятностью, индуцированной случайной величиной* ξ.

Функция распределения F_ξ, заданная правилом

$$F_\xi(x) = \mathbb{P}(\xi \le x),$$

связана с вероятностью μ_ξ следующим образом:

$$\mu_\xi((a,b]) = F_\xi(b) - F_\xi(a).$$

Заметим, что если ϕ борелевская функция из $\mathbb{R}$ в $\mathbb{R}$ (т.е. ϕ — измерима, то есть для всех $A \in \mathcal{B}(\mathbb{R})$ $\phi^{-1}(A) \in \mathcal{B}(\mathbb{R})$), то $\phi \circ \xi$ является также случайной величиной. Поскольку любая непрерывная функция является борелевской (проверьте!), то мы получаем широкий класс преобразований случайных величин. Если $\xi_1, \xi_2, \dots$ являются случайными величинами, то

$$\xi_1 + \xi_2, \quad \xi_1\xi_2, \quad \lim_{n\to\infty} \xi_n, \quad \sup_n \xi_n$$

также являются случайными величинами.

В следующих разделах мы постараемся, используя случайные величины и индуцированные ими вероятности на $\mathbb{R}$, вывести некоторые закономерности, применимые к широкому классу опытов с неясным исходом.

Также для случайных величин вводят числовые характеристики. Важнейшей из них является математическое ожидание. Мы обозначаем математическое ожидание случайной величины ξ через $\mathbb{E}\xi$. По определению

$$\mathbb{E}\xi \stackrel{\triangle}{=} \int_\Omega \xi(\omega) d\mathbb{P}(\omega) = \int_\Omega \xi(\omega)\mathbb{P}(d\omega).$$

Последние два обозначения мы будем считать взаимозаменяемыми.

Напомним, что дисперсия случайной величины ξ определяется по правилу:

$$D\xi \stackrel{\triangle}{=} \mathbb{E}(\xi - \mathbb{E}\xi)^2 = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2.$$

Грубо говоря, дисперсия определяет, насколько случайная величина случайна.

Наконец, отметим полезную формулу замены переменных. Она состоит в том, что для всякой борелевской функции ϕ

$$\mathbb{E}\phi(\xi) = \int_\mathbb{R} \phi(x) d\mu_\xi(x).$$

Глава 1

Предельные теоремы

Здесь и далее мы считаем, что вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ выбрано и зафиксировано.

1 Неравенства концентрации меры

Неравенств концентрации меры достаточно много, их цель — оценить вероятность отклонения, опираясь на какую-либо априорную информацию о случайной величине (величину дисперсии, длину промежутка, содержащего значения этой величины, целочисленность и так далее). Много таких неравенств есть, например, в
Wiki

Сначала выведем простейшие из них.

Предложение 1 (Неравенство Маркова)**.** *Для всякой случайной величины ξ и любого числа $a > 0$ имеет место неравенство*

$$\mathbb{P}(|\xi| \geq a) \leq \frac{\mathbb{E}|\xi|}{a}.$$

Доказательство. Имеем, что

$$\mathbb{E}|\xi| = \mathbb{E}|\xi|\mathbf{1}_{|\xi|<a} + \mathbb{E}|\xi|\mathbf{1}_{|\xi|\geq a}.$$

Поскольку первое слагаемое снизу может быть оценено 0, а второе — величиной $a\mathbb{P}(|\xi| \geq a)$, получаем, что

$$\mathbb{E}|\xi| \geq a\mathbb{P}(|\xi| \geq a).$$

Откуда и следует неравенство Маркова. □

Следствие 1 (Неравенство Маркова, общий случай; иногда называют неравеством Чебышева)**.** *Для всякой случайной величины ξ и любого числа a, для любой монотонно возрастающей положительной функции ϕ имеет место неравенство*

$$\mathbb{P}(\xi \geq a) \leq \frac{\mathbb{E}\phi(\xi)}{\phi(a)}.$$

Доказательство. Действительно,

$$\mathbb{P}(\xi \ge a) = \mathbb{P}(\phi(\xi) \ge \phi(a)) \le \frac{\mathbb{E}\phi(\xi)}{\phi(a)}.$$

$\square$

Подумать: можно ли в последней формулировке брать монотонно неубывающую функцию или функцию, принимающую и отрицательные значения; можно ли написать неравенство для $\mathbb{P}(\xi \le a)$?

Следствие 2. *Для всякой случайной величины ξ имеют место неравенства*

$$\mathbb{P}(\xi \ge a) \le \inf_{t>0} e^{ta}\mathbb{E}e^{-t\xi} \quad \forall a \in \mathbb{R} \text{ (граница Чернова)}$$

$$\mathbb{P}(|\xi| \ge |a|) \le \inf_{t>0} a^{-t}\mathbb{E}|\xi|^t \quad \forall a > 0 \text{ (моментная граница)}.$$

Предложение 2 (Неравенство Чебышева)**.** *Для всякой имеющей конечную дисперсию случайной величины ξ и любого числа $a > 0$ имеет место неравенство*

$$\mathbb{P}(|\xi - \mathbb{E}\xi| \ge a) \le \frac{D\xi}{a^2}.$$

Доказательство. Заметим, что

$$\mathbb{P}(|\xi - \mathbb{E}\xi| \ge a) = \mathbb{P}(|\xi - \mathbb{E}\xi|^2 \ge a^2).$$

Отсюда, применяя неравенство Маркова к случайной величине $|\xi - \mathbb{E}\xi|^2$, математическое ожидание которой равно $D\xi$, мы получаем неравенство Чебышева. $\square$

Следствие 3 (Правило трех сигм)**.** *Для всякой случайной величины ξ с $D\xi = \sigma^2$ имеет место неравенство*

$$\mathbb{P}(|\xi - \mathbb{E}\xi| > 3\sigma) \le \frac{1}{9}.$$

Следствие 4 (Закон больших чисел в форме Чебышева)**.** *Пусть X_i — последовательность попарно независимых, одинаково распределенных случайных величин с конечной дисперсией. Тогда для всех положительных a имеет место неравенство*

$$\mathbb{P}\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}X_1\Big| \ge a\Big) \le \frac{DX_1}{na^2}.$$

Доказательство. Заметим, что $\mathbb{E}\frac{X_1+\cdots+X_n}{n} = \mathbb{E}X_1$, $D\frac{X_1+\cdots+X_n}{n} = \frac{n}{n^2}DX_1 = DX_1/n$. Теперь,

$$\mathbb{P}\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}X_1\Big| \ge a\Big) \le \frac{D\frac{X_1+\cdots+X_n}{n}}{a^2} = \frac{DX_1}{na^2}.$$

$\square$

Подумать: точные ли эти оценки, на каких распределениях достигается точное равенство, можно ли везде в этих утверждениях написать строгие неравенства?

Подумать: ослабить можно как требование существования дисперсии (достаточно потребовать конечности $\mathbb{E}|X_1|^{1+\delta}$), требование независимости (рассматривать слабо коррелированные случайные величины), естественно не требуется и общее распределение у X_i. Как мы позже увидим, можно доказать более продвинутые оценки, но они и будут требовать большего.

Подумать о вечном: что дает закон больших чисел для статистики. Чего бы еще потребовать для нее же...

Выше были полиномиальные оценки. Экспоненциальные оценки имеют принципиально другую скорость сходимости. Доведем до точной формулы одну из них, границу Чернова, для случая ограниченных случайных величин.

Предложение 3 (Chernoff bound)**.** *Пусть X_i — последовательность независимых в совокупности случайных величин, принимающих значения из отрезка $[0,1]$. Тогда для всех натуральных n и положительных δ имеем*

$$\mathbb{P}\Big(X_1 + \dots + X_n \geq (1+\delta)M\Big) \leq \exp(M(\delta - (1+\delta)\ln(1+\delta))),$$

где $M = \mathbb{E}(X_1 + \dots + X_n)$.

Доказательство. Примем $t = \ln(1+\delta) > 0$. Пусть $X = X_1 + \dots + X_n$, $Y_i = e^{tX_i}$, $Y = e^{tX}$.

Заметим, что $e^{tx} \leq 1 - x + e^t x$ для $x \in [0,1]$ в силу

$$e^{xt+(1-x)0} = \exp(tx) \leq x\exp(t) + (1-x)\exp(0),$$

откуда,

$$\mathbb{E}[e^{tX_i}] \leq \mathbb{E}[1 - X_i + e^t X_i] = 1 + (e^t - 1)\mathbb{E}X_i \leq e^{(e^t-1)\mathbb{E}X_i}.$$

Теперь из

$$\mathbb{E}Y = \prod_{i=1}^n \mathbb{E}Y_i \leq \prod_{i=1}^n e^{(e^t-1)\mathbb{E}X_i} \leq e^{(e^t-1)\sum_{i=1}^n \mathbb{E}X_i} \leq e^{(e^t-1)M}$$

имеем

$$\begin{aligned}
\mathbb{P}\Big(X_1 + \dots + X_n \geq (1+\delta)M\Big) &= \mathbb{P}(e^{t(X_1+\dots+X_n)} \geq e^{t(1+\delta)M}) \\
&= \mathbb{P}(Y \geq e^{t(1+\delta)M}) \leq \frac{\mathbb{E}Y}{e^{t(1+\delta)M}} \\
&\leq \frac{e^{(e^t-1)M}}{e^{t(1+\delta)M}} = e^{M(e^t-1-t(1+\delta))}.
\end{aligned}$$

Подставляя $t = \ln(1+\delta)$, получаем

$$\mathbb{P}\Big(X_1 + \dots + X_n \geq (1+\delta)M\Big) \leq e^{M(\delta-(1+\delta)\ln(1+\delta))}.$$

□

Подумать: как изменится оценка, если нас интересуют вероятности малых значений случайной величины. Как изменится неравенство, если значения берутся из другого отрезка, например $[-K, K]$?

Из этой оценки можно получить и более простые для применения. Из

$$\frac{1}{2}\ln(1+\delta) = t/2 \geq \tanh t/2 = \frac{e^t - 1}{e^t + 1} = \frac{\delta}{2+\delta}$$

следует $(1+\delta)\ln(1+\delta) \geq 2(1+\delta)\frac{\delta}{2+\delta} = \delta + \frac{\delta^2}{2+\delta}$, откуда имеем

Замечание 1. В условиях предложения 3 выполнено

$$\begin{aligned}
\mathbb{P}\Big(X_1 + \cdots + X_n \geq (1+\delta)M\Big) &\leq e^{-M\delta^2/(2+\delta)} \qquad \forall\delta > 0,\\
\mathbb{P}\Big(X_1 + \cdots + X_n \geq (1+\delta)M\Big) &\leq e^{-M\delta^2/3} \qquad \forall\delta \in [0,1],\\
\mathbb{P}\Big(X_1 + \cdots + X_n \geq (1+\delta)M\Big) &\leq e^{-M\delta/3} \qquad \forall\delta > 1.
\end{aligned}$$

Аналогично, взяв $t = \ln(1-\delta)$, получаем

Замечание 2. В условиях предложения 3 выполнено

$$\mathbb{P}\Big(X_1 + \cdots + X_n \leq M(1-\delta)\Big) \leq e^{-M(\delta + (1-\delta)\ln(1-\delta))} \leq e^{-M\delta^2/2} \qquad \forall\delta \in (0,1).$$

Сравните закон больших чисел, полученный из неравенства Чебышева, с его аналогом из границы Чернова.

Следствие 5. *Пусть X_i — последовательность независимых в совокупности, одинаково распределенных случайных величин, принимающих значения из отрезка* $[0,1]$. *Тогда для всех натуральных n и положительных δ имеем*

$$\mathbb{P}\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}X_1\Big| \geq \delta\mathbb{E}X_1\Big) \leq 2e^{-n\delta^2\mathbb{E}X_1/2} \qquad \forall\delta \in [0,1].$$

Различие принципиально, хотя и требует много большего — независимости в совокупности. А проверить сие невозможно, в это можно только верить.

Пример 1. В поселке $N = 23025$ жителей, каждый из которых случайно и независимо от остальных ездит в город в среднем два раза в месяц. Поезд из поселка в город идет один раз в сутки. Найдите распределение числа K пассажиров, их среднее число и дисперсию. Какова должна быть вместимость w поезда, чтобы он переполнился с вероятностью, не превышающей $0,01$?

Неравенство Чебышева обещает $w \sim 1914$, граница Чернова — $w \sim 1652$. Поскольку здесь медиана (как и среднее) — $N/15 = 23025/15 = 1535$ пассажиров, откуда $w > 1535$, то оценка 1652 кажется много лучше 1914.

2 Сходимость случайных величин

Мы будем исследовать вопрос о сходимости последовательности случайных величин $\{\xi_n\}_{n=1}^{\infty}$. Если предел существует, то (возможно) мы сможем описать последовательность, исследовав лишь предельную случайную величину.

Определение 1. Будем говорить, что последовательность случайных величин ξ_n сходится к случайной величине ξ по вероятности, и записывать $\xi_n \xrightarrow{P} \xi$, если для каждого $\varepsilon > 0$

$$\mathbb{P}\left\{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\right\} \to 0 \text{ при } n \to \infty.$$

Отметим несколько свойств этой сходимости, их доказательства в целом очевидны:

- если ξ_n сходятся к ξ по вероятности, и ζ_n сходятся к ζ по вероятности, то $\xi_n - \zeta_n$ сходятся к $\xi - \zeta$ по вероятности;
- если ξ_n сходятся к ξ по вероятности, и $c \in \mathbb{R}$, то $c\xi_n$ сходятся к $c\xi$ по вероятности;
- если ξ_n сходятся к ξ по вероятности, и ξ_n сходятся к ζ по вероятности, то $\xi = \zeta$ почти всюду.

Таким образом, наша сходимость совместима с линейной структурой пространства всех случайных величин, но не различает совпадающие почти всюду случайные величины. Тогда она задает топологию на линейном пространстве всех случайных величин после отождествления $\mathbb{P}$-почти всюду совпадающих функций. Более того, эта топология метризуема, метрику (расстояние) можно ввести правилом $d(\xi, \eta) \overset{\triangle}{=} \mathbb{E}(1 - \frac{1}{1+|\xi-\eta|})$ или правилом $d(\xi, \eta) \overset{\triangle}{=} \mathbb{E} \operatorname{arctg} |\xi - \eta|$. Это, в свою очередь, означает, что на этом пространстве можно вводить привычные из матанализа понятия, например ε-окрестности. Впрочем, для выкладок это нам не понадобится.

Определение 2. Будем говорить, что последовательность случайных величин ξ_n сходится к случайной величине ξ $\mathbb{P}$-почти наверное (синонимы: почти наверное, почти наверняка, с вероятностью 1, почти всюду), и записывать $\xi_n \xrightarrow{\text{п.в.}} \xi$, если вероятность тех ω, для которых $\xi_n(\omega) \nrightarrow \xi(\omega)$, равна 0; т.е.

$$\mathbb{P}\{\omega : \xi_n(\omega) \nrightarrow \xi(\omega)\} = 0.$$

Предложение 4. *Сходимость случайных величин ξ_n к ξ почти наверное эквивалентна следующему условию: для всех $\varepsilon > 0$*

$$\mathbb{P}\left\{\omega : \sup_{k \geq n} |\xi_k(\omega) - \xi(\omega)| \geq \varepsilon\right\} \to 0 \text{ \textit{при} } n \to \infty.$$

Доказательство. Пусть

$$A_k^\varepsilon \overset{\triangle}{=} \{\omega : |\xi_k(\omega) - \xi(\omega)| \geq \varepsilon\}, \quad A^\varepsilon \overset{\triangle}{=} \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k^\varepsilon.$$

Заметим, что

$$\{\omega : \xi_n(\omega) \nrightarrow \xi(\omega)\} = \bigcup_{\varepsilon > 0} A^\varepsilon = \bigcup_{m=1}^{\infty} A^{1/m}.$$

Следовательно,

$$\mathbb{P}\{\omega : \xi_n(\omega) \nrightarrow \xi(\omega)\} = 0 \quad \Leftrightarrow \quad \mathbb{P}\left(\bigcup_{m=1}^{\infty} A^{1/m}\right) = 0.$$

Последнее равенство выполнено в том и только в том случае, когда для всех m $\mathbb{P}(A^{1/m}) = 0$, или, что эквивалентно, для всех $\varepsilon > 0$ $\mathbb{P}(A^{\varepsilon}) = 0$. Наконец, равенство $\mathbb{P}(A^{\varepsilon}) = 0$ эквивалентно тому, что

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k^{\varepsilon}\right) = 0.$$

$\square$

Следствие 6. *Сходимость почти наверное влечет сходимость по вероятности.*

Доказательство. Достаточно заметить, что

$$\{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\} \subset \left\{\omega : \sup_{k\geq n} |\xi_k(\omega) - \xi(\omega)| \geq \varepsilon\right\}.$$

$\square$

Утверждение, обратное к следствию 6, неверно.

Пример 2. Пусть $\Omega = [0, 1)$, на нем рассматриваем борелевскую σ-алгебру и равномерную вероятность. Выберем $f_1 = 1$, $f_{2,1}$ положим равной 1 на $[0, 1/2)$ и 0 на $[1/2, 1)$, $f_{2,2}$ положим равной 0 на $[0, 1/2)$ и 1 на $[1/2, 1)$, далее примем $f_{i,j} = \mathbf{1}_{[(j-1)/i, j/i)}$ для всех натуральных i, j. Если мы теперь последовательно занумеруем функции $f_{i,j}$, то полученная последовательность будет сходиться по вероятности к $\xi \equiv 0$, а последовательность $\xi_n(\omega)$ содержит бесконечно много 0 и 1 для всех ω.

Но есть и хорошие новости. Теорема Рисса утверждает, что если ξ_n сходится по вероятности к ξ, то можно выделить подпоследовательность некоторых ξ_{n_k}, сходящуюся почти наверное (тоже к ξ). Более того,

Как достаточно легко выводится из теоремы Рисса, сходимость почти наверное не может быть описана какой-либо метрикой. Это существенно усложняет выкладки, но можно рассмотреть близкую к ней сходимость, допускающую такое представление: сходимость в среднем.

Определение 3. Пусть $p \geq 1$. Будем говорить, что последовательность случайных величин ξ_n сходится к ξ в среднем порядка p, и записывать $\xi_n \xrightarrow{L^p} \xi$, если

$$\mathbb{E}|\xi_n - \xi|^p \to 0.$$

При $p = 2$ говорят также “в среднеквадратичном”.

Этот вид сходимости задает сходимость, метрику и норму в пространстве Лебега $L^p(\Omega, \mathcal{F}, \mathbb{P})$ всех случайных величин с $\mathbb{E}|\xi|^p < \infty$ после отождествления всех $\mathbb{P}$-почти всюду совпадающих случайных величин.

Отметим, что функция $|y|^p$ задает здесь штраф, а именно: разность $\xi_n(\omega) - \xi(\omega)$ наказывается штрафом $|\xi_n(\omega) - \xi(\omega)|^p$. Выбор параметра p дает возможность откалибровать кары за большие/малые отклонения. В принципе можно брать отличные от $|y|^p$ функции, сохраняя неотрицательность, монотонный рост с нуля на положительной полуоси и выпуклость.

Подумать: пусть ξ — какая-то случайная величина. Будут ли сходиться ξ/n к 0 по вероятности, почти всюду, в среднем?

Предложение 5. *Сходимость в среднем порядка p влечет сходимость по вероятности.*

Доказательство. По неравенству Маркова (см. Предложение 1) имеем

$$\mathbb{P}(|\xi_n - \xi| \geq \varepsilon) = \mathbb{P}(|\xi_n - \xi|^p \geq \varepsilon^p) \leq \frac{\mathbb{E}|\xi_n - \xi|^p}{\varepsilon^p} \to 0, \quad n \to \infty.$$

□

Сходимость почти наверное не влечет сходимость в среднем.

Пример 3. $\Omega = [0, 1]$, $\mathbb{P}$ — мера Лебега.

$$\xi_n(\omega) = \begin{cases} 2^n, & \omega \in [0, 1/n], \\ 0, & \omega \notin [0, 1/n]. \end{cases}$$

Хотя $\xi_n \xrightarrow{\text{п.в.}} 0$, ξ_n не сходится в среднем к 0.

В то же время сходимость в среднем получается из сходимости почти наверное при дополнительных предположениях.

Предложение 6. *Пусть ξ_n неотрицательны, $\xi_n \xrightarrow{п.в.} \xi$ и $\mathbb{E}\xi_n \to \mathbb{E}\xi$ (в частности, начиная с некоторого, эти матожидания существуют). Тогда* $\mathbb{E}|\xi_n - \xi| \to 0$.

Доказательство. Без ограничения общности можно считать, что $\mathbb{E}\xi_n < \infty$. Имеем

$$\mathbb{E}|\xi_n - \xi| = \mathbb{E}(\xi - \xi_n)\mathbf{1}_{\xi \geq \xi_n} + \mathbb{E}(\xi_n - \xi)\mathbf{1}_{\xi_n > \xi} = 2\mathbb{E}(\xi - \xi_n)\mathbf{1}_{\xi \geq \xi_n} + \mathbb{E}(\xi_n - \xi).$$

Второе слагаемое сходится к 0 по условию. Также $(\xi - \xi_n)\mathbf{1}_{\xi \geq \xi_n}$ поточечно сходится к 0. Используя неотрицательность ξ_n, получаем, что $0 \leq (\xi - \xi_n)\mathbf{1}_{\xi \geq \xi_n} \leq \xi$. Отсюда, по теореме Лебега о мажорируемой сходимости, получаем, что первое слагаемое тоже сходится к 0. □

Подумать: можно ли сформулировать аналог этого предложения для $p > 1$ (ну и доказать/опровергнуть этот аналог)?

В общем случае сходимость в среднем не влечет сходимости почти наверное.

Пример 4. Пусть ξ_n независимы, $\mathbb{P}(\xi_n = 1) = p_n$, $\mathbb{P}(\xi_n = 0) = 1 - p_n$. Имеем, что

$$\xi_n \xrightarrow{P} 0 \iff p_n \to 0,$$

$$\xi_n \xrightarrow{L^p} 0 \iff p_n \to 0,$$

$$\xi_n \xrightarrow{\text{п.в.}} 0 \iff \sum_{n=1}^{\infty} p_n < \infty.$$

В частности, при $p_n = 1/n$ получаем, что из $\xi_n \xrightarrow{L^p} \xi \not\Rightarrow \xi_n \xrightarrow{\text{п.в.}} \xi$.

Подумать: а откуда здесь взялось условие $\sum_{n=1}^{\infty} p_n < \infty$?

В качестве приложения докажем еще один вариант закона бьльших чисел.

Теорема 1. *Пусть дана последовательность случайных величин ξ_n с математическими ожиданиями $\mathbb{E}\xi_i = m_i$ и равномерно ограниченными дисперсиями. Предположим, что ξ_i попарно некоррелированы, т.е. $\mathbb{E}(\xi_i - m_i)(\xi_j - m_j) = 0$ для $i \neq j$. Тогда*

$$\mathbb{E}\left|\frac{(\xi_1 - m_1) + \ldots + (\xi_n - m_n)}{n}\right|^2 \to 0,$$

т.е. последовательность $\big(\xi_1 - m_1 + \ldots + \xi_n - m_n\big)/n$ сходится к 0 в среднеквадратичном, а значит, и по вероятности.

Доказательство. Можно считать, что $m_i = 0$. Тогда нам требуется доказать, что $D(\xi_1 + \ldots + \xi_n)/n$ сходится к нулю. Имеем, что

$$\mathbb{E}\frac{(\xi_1 + \ldots + \xi_n)^2}{n^2} = \frac{1}{n^2} D(\xi_1 + \ldots + \xi_n) = \frac{D\xi_1 + \ldots + D\xi_n}{n^2} \to 0$$

в силу равномерной ограниченности $D\xi_i$. $\square$

Оценить скорость сходимости можно и на основе неравенства Маркова (см. предыдущую формулировку закона больших чисел про сходимость по вероятности). При этом, и в том, и в другом доказательстве для независимых случайных величин можно потребовать лишь ограниченность $\mathbb{E}|\xi_n - \mathbb{E}\xi_n|^p$ для некоторого $p > 1$, получая соответствующие оценки скорости сходимости. Если конкретные оценки вам не столь интересны, вам нужен лишь факт, то в теореме Хинчина (см. теорему 8) единственное требование к независимым одинаково распределенным случайным величинам — существование матожидания.

3 Гарантии почти всюду

Представленные в этом разделе утверждения позволяют вычислить вероятности предельных событий, равные нулю или единице. Ну и дают представление о “почти всюду”.

3.1 Леммы Бореля-Кантелли

Лемма 1 (первая лемма Бореля-Кантелли)**.** *Пусть* $\{A_n\} \subset \mathcal{F}$ *— последовательность событий такая, что*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty.$$

Положим

$$A \stackrel{\triangle}{=} \{\omega : \text{ существует последовательность } n_k, \text{ что } \omega \in A_{n_k}\}.$$

Тогда $\mathbb{P}(A) = 0$.

Доказательство. Ясно, что

$$A = \bigcap_{l=1}^{\infty} \bigcup_{n=l}^{\infty} A_n.$$

Имеем, что

$$\mathbb{P}(A) \leq \mathbb{P}\left(\bigcup_{n=l}^{\infty} A_n\right) \leq \sum_{n=l}^{\infty} \mathbb{P}(A_n) \to 0 \text{ при } l \to \infty.$$

□

Лемма 2 (вторая лемма Бореля-Кантелли)**.** *Пусть* $\{A_n\} \subset \mathcal{F}$ *— последовательность* ***независимых*** *в совокупности событий такая, что*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty.$$

Положим

$$A \stackrel{\triangle}{=} \{\omega : \text{ существует последовательность } n_k, \text{ что } \omega \in A_{n_k}\}.$$

Тогда $\mathbb{P}(A) = 1$.

Доказательство. Имеем, что

$$\Omega \setminus A = \bigcup_{l=1}^{\infty} \bigcap_{n=l}^{\infty} (\Omega \setminus A_n).$$

$$\mathbb{P}(\Omega \setminus A) = \lim_{l \to \infty} \mathbb{P}\left(\bigcap_{n=l}^{\infty} (\Omega \setminus A_n)\right).$$

Независимость A_n влечет независимость $\Omega \setminus A_n$. Отсюда

$$\mathbb{P}\left(\bigcap_{n=l}^{\infty} (\Omega \setminus A_n)\right) = \prod_{n=l}^{\infty} (1 - \mathbb{P}(A_n)).$$

Так как $\sum_{n=l}^{\infty} \mathbb{P}(A_n) = \infty$, $\prod_{n=l}^{\infty} (1 - \mathbb{P}(A_n)) = 0$ (можно взять логарифм с двух сторон и оценить). □

Отметим, что условие независимости событий A_n во второй лемме Бореля-Кантелли существенно, но достаточно обеспечить их попарную независимость.

3.2 Усиленный закон больших чисел. Закон нуля-единицы

Следующая теорема дает более сильную сходимость нежели сходимость по вероятности.

Теорема 2 (Усиленный закон больших чисел (в форме Колмогорова))**.** *Пусть* $\{\xi_n\}$ *— последовательность независимых одинаково распределенных случайных величин со средним* m*. Потребуем также* $\mathbb{E}|\xi_n|^4 = \mathbb{E}|\xi_1|^4 < \infty$*. Тогда*

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\text{п.в.}} m \text{ при } n \to \infty.$$

Доказательство. Без ограничения общности считаем, что $m = 0$ (в противном случае можно "исправить" случайные величины ξ_n). Пусть B — множество тех точек ω, где

$$\lim_{n\to\infty} \frac{\xi_1 + \ldots + \xi_n}{n} \nrightarrow 0.$$

Тогда $B = \bigcup_{k=1}^{\infty} B_k$, где B_k — множество тех ω, что $|\xi_1 + \ldots + \xi_n|/n \geq 1/k$ для бесконечного количества чисел n. Докажем, что $\mathbb{P}(B_k) = 0$.

Пусть

$$A_n^k \triangleq \{\omega : |\xi_1 + \ldots + \xi_n|/n \geq 1/k\}.$$

Пользуясь независимостью случайных величин, предположением $\mathbb{E}\xi_i = 0$ и комбинаторикой, раскрыв скобки и убирая все произведения с нечетными степенями в силу $\mathbb{E}\xi_1 = 0$ и аккуратно посчитав количество оставшихся членов, получаем, что

$$\mathbb{E}|\xi_1 + \ldots + \xi_n|^4 = n\mathbb{E}|\xi_1|^4 + C_4^2 C_n^2 (\mathbb{E}|\xi_1|^2)^2 \leq K n^2,$$

где K — константа, не зависящая от n.

Тогда, по неравенству Маркова,

$$\mathbb{P}(A_n^k) = \mathbb{P}\{\omega : |\xi_1 + \ldots + \xi_n|^4 \geq n^4/k^4\} \leq \frac{\mathbb{E}|\xi_1 + \ldots + \xi_n|^4}{n^4/k^4} \leq Kk^4/n^2.$$

Поскольку $\sum_{n=1}^{\infty} \mathbb{P}(A_n^k) < \infty$, по первой лемме Бореля-Кантелли заключаем, что $\mathbb{P}(B_k) = 0$. Отсюда $\mathbb{P}(B) = 0$.

□

Замечание 3. На самом деле у Колмогорова все доказано в существенно более слабых предположениях. Можно отказаться от требования общего распределения, а значит, и общего матожидания; уже из существования конечных дисперсий (а не четвертых моментов) и неравенства $\sum_{n=1}^{\infty} D\xi_n/n^2 < \infty$ следует

$$\lim_{n\to\infty} \frac{\xi_1 + \ldots + \xi_n}{n} - \mathbb{E}\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\text{п.в.}} 0.$$

Позже Хинчин смог избавиться и от независимости, предполагая лишь слабую коррелируемость.

Еще одна тема, близкая к усиленному закону больших чисел, интересуется не тем, сойдется ли среднее к матожиданию с вероятностью 1, а тем, насколько оно при этом может "максимум" (с вероятностью 1) отклониться: как растет самое большое отклонение от среднего значения. Такого рода вопросами занимается закон повторного логарифма. Одну из простейших формулировок (не доказательств!) см. ниже:

Теорема 3 (Закон повторного логарифма, без д-ва)**.** *Пусть $\{\xi_n\}$ — последовательность независимых одинаково распределенных случайных величин со средним 0 и дисперсией $\sigma^2 > 0$. Тогда, для всякого положительного ε*

$$\mathbb{P}(\xi_1 + \ldots + \xi_n > (1+\varepsilon)\sigma\sqrt{2n \ln\ln n} \text{ случится бесконечно много раз}) = 0,$$
$$\mathbb{P}(\xi_1 + \ldots + \xi_n < (1-\varepsilon)\sigma\sqrt{2n \ln\ln n} \text{ случится бесконечно много раз}) = 1.$$

Вот некоторый пример задачки, дающей представление, зачем нужны эти формулы (но через цепи Маркова задачка считается точно).

Пример 5. Вы ("орел") играете с соседом ("решка"), бросая правильную монетку, в каждом раунде на кону одна монета. У Вас 10^3 монет, у соседа они не кончаются в принципе. Оцените, при каком числе партий у Вас кончатся монеты? А если у Вас 10^6 монет?

Теорема 4 (Закон нуля-единицы)**.** *Пусть имеется последовательность событий A_n, введем минимальные σ-алгебры $\mathcal{A}_n = \sigma(A_1, \ldots, A_n)$, $\mathcal{A}_\infty = \cap_{k\in\mathbb{N}}\mathcal{A}_k$. Докажите, что для любого $A \in \mathcal{A}_\infty$ или $\mathbb{P}(A) = 1$, или $\mathbb{P}(A) = 0$.*

Доказательство. Отметим, что A не принадлежит σ-алгебре $\sigma(\xi_1, \xi_2, \ldots, \xi_{n-1})$, следовательно не зависит от нее, тогда не зависит от объединяющей их $\sigma(\xi_1, \xi_2, \ldots, \xi_{n-1}, \xi_n \ldots)$, но $A \in \sigma(\xi_1, \xi_2, \ldots, \xi_{n-1}, \xi_n \ldots)$, следовательно A не зависит от самого себя, то есть $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}^2(A)$. Решая квадратное уравнение, получаем требуемое. □

4 Слабая сходимость

В этом параграфе мы будем изучать сходимость вероятностей на $\mathbb{R}$. В некоторых случаях мы будем предполагать, что вероятности индуцированы случайными величинами. Тогда сходимость вероятностей на $\mathbb{R}$ позволяет изучить еще один вид сходимости случайных величин.

Напомним, что для случайных величин мы вводили функцию распределения. Это же понятие можно ввести для вероятностей. Если μ — вероятность на $\mathbb{R}$, то

$$F_\mu(x) \triangleq \mu((-\infty, x]).$$

Заметим, что если $\mu = \xi\#\mathbb{P}$, то

$$F_\mu = F_\xi.$$

Определение 4. Будем говорить, что μ_n слабо сходится к вероятности μ, обозначая эту сходимость $\mu_n \xrightarrow{w} \mu$, если для каждой непрерывной ограниченной на $\mathbb{R}$ функции ϕ со значениями в $\mathbb{R}$

$$\int_{\mathbb{R}} \phi(x)\mu_n(dx) \to \int_{\mathbb{R}} \phi(x)\mu(dx).$$

Подумать: докажите, что если каждая μ_n сосредоточена в некоторой x_n, то есть $\mu_n(A) \stackrel{\triangle}{=} |A \cap \{x_n\}|$, то μ_n сходится тогда и только тогда, когда сходится x_n.

Подумать: стоит ли требовать сходимость интегралов для всех измеримых функций?

Класс пробных функций можно существенно сузить. Напомним, что функция $\phi : \mathbb{R} \to \mathbb{R}$ называется финитной, если она равна нулю вне некоторого отрезка $[a, b]$. Множество непрерывных финитных функций обозначим через $C_0(\mathbb{R})$, множество бесконечное число раз дифференцируемых финитных функций — через $C_0^\infty(\mathbb{R})$.

Предложение 7. *μ_n сходится слабо к μ тогда и только тогда, когда для каждой бесконечное число раз дифференцируемой финитной на $\mathbb{R}$ функции ϕ со значениями в $\mathbb{R}$*

$$\int_{\mathbb{R}} \phi(x)\mu_n(dx) \to \int_{\mathbb{R}} \phi(x)\mu(dx).$$

Доказательство. Очевидно, что надо доказать лишь достаточность. Прежде всего заметим, что для каждого $\varepsilon > 0$ существует отрезок $[a, b]$ такой, что $\mu[a, b] \geq 1 - \varepsilon$.

Пусть теперь ϕ — произвольная непрерывная функция. Будем считать, что $|\phi| \leq c$. Пусть ϕ^ε — финитная бесконечное число раз дифференцируемая функция такая, что $|\phi^\varepsilon| \leq c$, $|\phi^\varepsilon(x) - \phi(x)| \leq \varepsilon$ на $x \in [a, b]$ и $|\phi^\varepsilon| \leq 2c$. Имеем, что для достаточно больших n

$$\begin{aligned} \left| \int_{-\infty}^{+\infty} \phi(x)\mu_n(dx) - \int_{-\infty}^{+\infty} \phi^\varepsilon(x)\mu_n(dx) \right| &\leq \int_a^b |\phi(x) - \phi^\varepsilon(x)|\mu_n(dx) \\ + \left| \int_{\mathbb{R}\setminus[a,b]} \phi(x)\mu_n(dx) \right| &+ \left| \int_{\mathbb{R}\setminus[a,b]} \phi^\varepsilon(x)\mu_n(dx) \right| \leq \varepsilon + 3c\varepsilon. \end{aligned} \tag{1.1}$$

Аналогично,

$$\left| \int_{-\infty}^{+\infty} \phi(x)\mu(dx) - \int_{-\infty}^{+\infty} \phi^\varepsilon(x)\mu(dx) \right| \leq \varepsilon + 3c\varepsilon. \tag{1.2}$$

Поскольку, по предположению, ϕ^ε — финитна и бесконечное число раз дифференцируема, то

$$\int_{-\infty}^{+\infty} \phi^\varepsilon \mu_n(dx) \to \int_{-\infty}^{+\infty} \phi^\varepsilon \mu(dx) \text{ при } n \to \infty.$$

Устремляя ε к нулю и принимая во внимание оценки (1.1), (1.2), мы получаем сходимость $\int \phi(x)\mu_n(dx)$ к $\int \phi(x)\mu(dx)$. □

Слабая сходимость задает топологию на множестве всевозможных вероятностных мер. Эта топология может быть описана метрикой, например введенной ранее метрикой Канторовича. В функане, откуда и пошло название, рассматривают такую сходимость как сходимость на линейных отображениях вида

$$\phi \mapsto \int_{\mathbb{R}} \phi(x)\mu(dx).$$

На любом полном сепарабельном метрическом пространстве Ω имеет место:

Теорема 5 (Теорема Александрова; без д-ва)**.** *Следующие условия эквивалентны:*

- $\mu_n \xrightarrow{w} \mu$;
- $\mu(A) \le \liminf_{n\to\infty} \mu_n(A)$ *для всякого открытого множества* A*;*
- $\mu(A) \ge \limsup_{n\to\infty} \mu_n(A)$ *для всякого замкнутого множества* A*;*
- $\mu(A) = \lim_{n\to\infty} \mu_n(A)$ *для всякого борелевского множества* A*, у границы которого мера* μ *равна нулю.*

На основе слабой сходимости вероятностей можно ввести сходимость случайных величин (напомним обозначение $\mu_\xi \overset{\triangle}{=} \xi\#\mathbb{P}$).

Определение 5. Будем говорить, что ξ_n сходится к ξ по распределению, и писать $\xi_n \xrightarrow{d} \xi$, если μ_{ξ_n} сходится к μ_ξ слабо.

Подумать: Приведите пример $\xi : \Omega \to \{-1, 1\}$ такой, что $\xi_n = \xi \xrightarrow{d} \xi$, $\xi_n = \xi \xrightarrow{d} -\xi$. Можно ли сделать вывод, что $\xi = -\xi$? Еще раз прочитайте предыдущий абзац.

Подумать: проверьте определение, вдруг там что напутано...

Термин "сходимость по распределению" проясняется в следующем утверждении.

Предложение 8. *Последовательность случайных величин* ξ_n *сходится по распределению к случайной величине* ξ *тогда и только тогда, когда для всех точек непрерывности функции распределения* F_ξ

$$F_{\xi_n}(t) \to F_\xi(t), \text{ при } n \to \infty.$$

Доказательство. Мы докажем несколько более общее утверждение: $\mu_n \xrightarrow{w} \mu$ тогда и только тогда, когда

$$F_{\mu_n}(t) \to F_\mu(t)$$

во всех точках непрерывности F_μ.

Пусть вначале $\mu_n \xrightarrow{w} \mu$. Заметим, что поскольку $x \mapsto F_\mu$ не убывает, то количество точек разрыва не более чем счетно. Если t — точка непрерывности F_μ, то для любого $\varepsilon > 0$ можно найти $\delta > 0$ такое, что $F_\mu(t - \delta) \ge F_\mu(t) - \varepsilon$.

Выберем ϕ равным 1 на $(-\infty, t-\delta]$ и 0 на $[t, +\infty)$. На $[t-\delta, t]$ продлим ϕ аффинно. Поскольку $\int_{-\infty}^{+\infty} \phi\mu_n(dx) \to \int_{-\infty}^{+\infty} \phi\mu(dx)$, при достаточно больших n имеем, что

$$
\begin{aligned}
F_{\mu_n}(t) = \int_{-\infty}^{t} \mu_n(dx) \geq \int_{-\infty}^{+\infty} \phi(x)\mu_n(dx) \\
> \int_{-\infty}^{+\infty} \phi(x)\mu(dx) - \varepsilon > \int_{-\infty}^{t-\delta} \phi(x)\mu(dx) - \varepsilon = F_{\mu}(t-\delta) - \varepsilon \geq F_{\mu}(t) - 2\varepsilon.
\end{aligned}
$$

Аналогичным образом, выбирая δ так, чтобы $F_\mu(t+\delta) \leq F_\mu(t) + \varepsilon$, положив кусочно-линейную функцию ϕ равной 1 на $(-\infty, t]$ и 0 на $[t+\delta, +\infty)$, мы получаем, что

$$
F_{\mu_n}(t) \leq F_\mu(t) + 2\varepsilon.
$$

Отсюда следует требуемая сходимость.

Докажем теперь обратное утверждение. Пусть ϕ — финитная, бесконечное число раз дифференцируемая функция. Тогда для не более чем счетного числа точек t не имеет место сходимость $\phi'(t)F_{\mu_n}(t) \to \phi'(t)F_\mu(t)$. По теореме Лебега о мажорируемой сходимости, используя интегрирование по частям, заключаем, что $\int_{-\infty}^{+\infty} \phi(x)\mu_n(dx)$ сходятся к $\int_{-\infty}^{+\infty} \phi(x)\mu(dx)$.

Остается доказать лишь формулу интегрирования по частям: для произвольной $\phi \in C_0^\infty(\mathbb{R})$

$$
\int_{-\infty}^{+\infty} \phi(x)\mu(dx) = -\int_{-\infty}^{+\infty} \phi'(x)F_\mu(x)dx.
$$

Для этого представим $\phi(x)$ в виде

$$
\phi(x) = \int_{-\infty}^{x} \phi'(t)dt = \int_{-\infty}^{+\infty} \mathbf{1}_{(-\infty,x]}(t)\phi'(t)dt.
$$

Тогда, переставляя интегралы и используя равенство $\mathbf{1}_{(-\infty,x]}(t) = \mathbf{1}_{[t,+\infty)}(x)$ для всех x, t и почти всюду выполненное равенство $\int_t^{+\infty} \mu(dx) = 1 - F_\mu(t)$, мы получаем, что

$$
\begin{aligned}
\int_{-\infty}^{+\infty} \phi(x)\mu(dx) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathbf{1}_{(-\infty,x]}(t)\phi'(t)dt\mu(dx) \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathbf{1}_{[t,+\infty)}(x)\mu(dx)\phi'(t)dt = \int_{-\infty}^{+\infty} (1 - F_\mu(t))\phi'(t)dt \\
&= -\int_{-\infty}^{+\infty} F_\mu(t)\phi'(t)dt.
\end{aligned}
$$

Заметим, что в последнем равенстве существенно использовалась финитность ϕ. □

Сравним введенные ранее виды сходимости со сходимостью по распределению (которая выросла из слабой сходимости). Напомню, что сходимость в среднем и сходимость почти всюду влекли сходимость по вероятности.

Предложение 9. *Если* $\xi_n \xrightarrow{P} \xi$, *то* $\xi_n \xrightarrow{d} \xi$.

Доказательство. Чтобы упростить обозначения, положим

$$\mu \triangleq \mu_\xi, \quad \mu_n \triangleq \mu_{\xi_n}.$$

Пусть далее $\phi \in C_0^\infty$. Существует константа c такая, что

$$|\phi(x)| \le c, \quad |\phi'(x)| \le c.$$

Выберем некоторое $\varepsilon > 0$. Обозначим

$$G_n^\varepsilon \triangleq \{\omega : |\xi_n - \xi| < \varepsilon\}.$$

Имеем, в силу $|\phi(y_1) - \phi(y_2)| = |\phi'(y_1 + \theta(y_2 - y_1))| \cdot |y_1 - y_2| \le c|y_1 - y_2|$,

$$\begin{aligned} \left| \int_{-\infty}^{+\infty} \phi(x)\mu(dx) - \int_{-\infty}^{+\infty} \phi(x)\mu_n(dx) \right| &= |\mathbb{E}\phi(\xi_n) - \mathbb{E}\phi(\xi)| \\ &\le |\mathbb{E}\phi(\xi_n) - \mathbb{E}\phi(\xi)|\mathbf{1}_{G_n^\varepsilon} + |\mathbb{E}\phi(\xi_n)|\mathbf{1}_{\Omega\setminus G_n^\varepsilon} + \mathbb{E}|\phi(\xi)|\mathbf{1}_{\Omega\setminus G_n^\varepsilon} \\ &\le c\varepsilon + 2c\mathbb{P}(\Omega \setminus G_n^\varepsilon). \end{aligned}$$

Поскольку $\mathbb{P}(\Omega \setminus G_n^\varepsilon) = \mathbb{P}(|\xi_n - \xi| \ge \varepsilon) \to 0$ при $n \to \infty$, получаем, что

$$\lim_{n\to\infty} \left| \int_{-\infty}^{+\infty} \phi(x)\mu(dx) - \int_{-\infty}^{+\infty} \phi(x)\mu_n(dx) \right| \le c\varepsilon.$$

Устремляя ε к нулю, получаем, что для каждой функции $\phi \in C_0^\infty(\mathbb{R})$ имеет место сходимость

$$\int_{-\infty}^{+\infty} \phi(x)\mu(dx) \to \int_{-\infty}^{+\infty} \phi(x)\mu_n(dx).$$

Это и есть слабая сходимость. □

Обратное верно в том случае, если предел равен константе.

Предложение 10. *Если* $\xi_n \xrightarrow{d} m$ *для некоторой константы* m, *то* $\xi_n \xrightarrow{P} m$.

Доказательство. Пусть $\xi = m$. Отметим, что F_ξ непрерывна всюду, кроме точки m, более того $F_\xi(m - \varepsilon) = 0, F_\xi(m + \varepsilon) = 1$ для всех положительных ε. Остальное - голый счет: для всех положительных ε при $n \uparrow \infty$ имеем

$$\begin{aligned} \mathbb{P}(|\xi_n - m| \le \varepsilon) &\ge \mathbb{P}(m - \varepsilon < \xi_n \le m + \varepsilon) \\ &= F_{\xi_n}(m + \varepsilon) - F_{\xi_n}(m - \varepsilon) \to F_\xi(m + \varepsilon) - F_\xi(m - \varepsilon) = 1. \end{aligned}$$

□

Отметим интересную связь между слабой сходимостью и сходимостью почти наверное, установленную А.В. Скороходом.

Упражнение 1. Если $\xi_n \xrightarrow{d} \xi$, то существуют случайные величины η_n и η такие, что $\eta_n \xrightarrow{\text{п.в.}} \eta$ и $F_{\xi_n} = F_{\eta_n}$, $F_\xi = F_\eta$.

В заключение обсудим вопрос о компактности пространства мер. Для этого нам потребуется понятие плотности семейства вероятностей.

Определение 6. Множество вероятностей $\mathcal{K}$ называется *плотным*, если для любого $\varepsilon > 0$ существует отрезок $[a, b]$ такой, что для всех $\nu \in \mathcal{K}$ $\nu[a, b] \geq 1 - \varepsilon$.

Следующая теорема для монотонных ограниченных функций была доказана Хелли, для функции распределения — Прохоровым.

Теорема 6 (Хелли–Прохоров)**.** *Пусть $\{\mu_n\}$ плотна, тогда существуют подпоследовательность μ_{n_k} и вероятность μ такие, что $\mu_{n_k} \xrightarrow{w} \mu$.*

Доказательство. Мы дадим основную схему доказательства. Используя диагональный процесс Кантора, можно построить последовательность $F_{\mu_{n_k}}$ такую, что $F_{\mu_{n_k}}(t)$ сходятся для всех рациональных точек t (напомню, что их счетное число). Обозначим предельную функцию через $F(t)$. Пока она определена лишь для рациональных t. Для простоты считаем, что $F_{\mu_n}(t)$ сходятся к $F(t)$.

Можно продлить функцию F на все $\mathbb{R}$. Заметим, что $F(+\infty) = 1$, $F(-\infty) = 0$ (это следует из плотности μ_n). Также имеем $F_{\mu_n}(t) \to F(t)$ для всех точек непрерывности F. В самом деле, если t — точка непрерывности F, то существуют такие рациональные t_1 и t_2, что $t_1 < t < t_2$ и $F(t_2) - F(t_1) < \varepsilon$. Выберем N так, чтобы для всех $n > N$ $|F_{\mu_n}(t_1) - F(t_1)|, |F_{\mu_n}(t_2) - F(t_2)| < \varepsilon$. Отсюда получаем, что $|F_{\mu_n}(t) - F(t)| < 3\varepsilon$.

Остается изменить функцию F во всех точках разрыва (а их не более чем счетное число) так, чтобы она стала непрерывной справа. Во всех остальных точках сходимость $F_{\mu_n}(t)$ к $F(t)$ сохраняется. Определенная функция F есть функция распределения для некоторой вероятности μ. $\square$

5 Характеристические функции

Пусть μ — вероятность на $\mathbb{R}$. Этой вероятности можно сопоставить функцию по правилу

$$\tilde{\mu}(y) \triangleq \int_{\mathbb{R}} e^{ixy} \mu(dx).$$

Функцию $\tilde{\mu}$ будем называть характеристической функцией вероятности μ.

Если μ — вероятность, порожденная случайной величиной $\mu = \mu_\xi$ (т.е. $\mu = \xi \# \mathbb{P}$), то будем говорить о характеристической функции случайной величины ξ и обозначать ее $\tilde{\xi} = \widetilde{\mu_\xi}$. Заметим, что

$$\tilde{\xi} = \widetilde{\mu_\xi} = \mathbb{E} e^{iy\xi}. \tag{1.3}$$

Если μ имеет плотность $f(\cdot)$, то

$$\tilde{\mu}(y) = \int_{\mathbb{R}} e^{ixy} f(x) dx.$$

Последняя формула сильно напоминает формулу преобразования Фурье (а если быть точным, — обратного преобразования Фурье). Напомню, что преобразование Фурье определяется для одномерного случая по формуле

$$\hat{f}(y) \triangleq \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ixy} f(x)dx,$$

а обратное

$$f(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ixy} \hat{f}(y)dy.$$

0^0 $\tilde{\mu}(y)$ всегда существует.

Подумать: а почему интеграл существует?

1^0 $\widetilde{\mu}(0) = 1$ для любого распределения μ.

2^0 При сложении и умножении на $c \in \mathbb{R}$ случайной величины ее характеристическая функция меняется по правилам:

$$\widetilde{c+\xi}(t) = e^{itc}\widetilde{\xi}(t), \widetilde{c\xi}(t) = \widetilde{\xi}(ct).$$

3^0 Характеристическая функция суммы независимых случайных величин равна произведению характеристических функций: если $\xi_1, \ldots, \xi_n$ — независимы, то

$$\xi_1 \widetilde{+ \ldots +} \xi_n = \widetilde{\xi_1} \cdot \ldots \cdot \widetilde{\xi_n},$$

или в другой записи

$$\widetilde{\mu_{\xi_1+\ldots+\xi_n}} = \widetilde{\mu_{\xi_1}} \cdot \ldots \cdot \widetilde{\mu_{\xi_n}}.$$

Подумать: выразить $a\xi_1 \widetilde{+ \ldots +} a_n\xi_n$.

4^0 [C-но] Если конечен интеграл

$$\int_{\mathbb{R}} |x|^p \mu(dx),$$

т.е. μ имеет k-й момент, то характеристическая функция имеет k-ю производную; в частности $\mathbb{E}\xi = i\frac{d\widetilde{\xi}(0)}{dt}$.

Подумать: обратное верно только для четных k.

Подумать: привести пример, в котором $\mathbb{E}\xi$ нет, а $\frac{d\widetilde{\xi}(0)}{dt}$ есть.

Некоторые характеристические функции можно посчитать на руках.

Вероятность, сосредоточенная в нуле (мера Дирака) δ, имеет характеристическую функцию $\tilde{\delta}(y) = 1$, а случайная величина, почти всюду равная константе m, имеет характеристическую функцию $\tilde{m}(y) = e^{imt}$.

Гауссова вероятность: вероятность $\gamma_{m,\sigma}$ такая, что

$$\gamma[a,b] = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2} dx,$$

имеет характеристическую функцию

$$\widetilde{\gamma_{m,\sigma}} = e^{imy - \frac{\sigma^2 y^2}{2}}.$$

Достаточно показать это для $\widetilde{\gamma_{0,1}}$. Пусть $z(y) = \sqrt{2\pi}\widetilde{\gamma_{0,1}}(y)$. Тогда, дифференцируя под знаком интеграла, а затем интегрируя по частям, имеем

$$\begin{aligned}
\frac{d}{dy} z(y) &= \int_{-\infty}^{+\infty} \frac{d}{dy}(e^{ixy - x^2/2}) dx = \int_{-\infty}^{+\infty} ix e^{ixy - x^2/2} dx \\
&= -i \int_{-\infty}^{+\infty} e^{ixy} d e^{-x^2/2} = -i e^{-x^2/2} e^{ixy} \Big|_{-\infty}^{+\infty} + i \int_{-\infty}^{+\infty} e^{-x^2/2} d(e^{ixy}) \\
&= i \int_{-\infty}^{+\infty} iy e^{-x^2/2} e^{ixy} dx = i^2 y z(y) = -y z(y).
\end{aligned}$$

Решая $z' = -yz$, имеем $z(y) = Ce^{-y^2/2}$ и $\widetilde{\gamma_{0,1}}(0) = De^{-y^2/2}$. Осталось найти D в силу $\widetilde{\gamma_{0,1}}(0) = 1$.

Желающие посчитать характеристические функции известных распределений могут пойти в Wiki

Банальные свойства кончились, дальше только матан.

Получим формулу обращения характеристической функции.

Предложение 11. *Справедливо равенство*

$$\mu((a,b)) + \frac{1}{2}\mu(\{a\}) + \frac{1}{2}\mu(\{b\}) = \frac{1}{2\pi} \lim_{R\to\infty} \int_{-R}^{+R} \frac{e^{-iya} - e^{-iyb}}{iy} \tilde{\mu}(y) dy.$$

Доказательство. Имеем, что

$$\int_{-R}^{+R} \frac{e^{-iya} - e^{-iyb}}{iy} \tilde{\mu}(y) dy = \int_{-R}^{+R} \int_{-\infty}^{+\infty} \frac{e^{-iya} - e^{-iyb}}{iy} e^{iyx} \mu(dx) dy.$$

Переставив интегралы в силу теоремы Фубини, получаем

$$\int_{-R}^{+R} \frac{e^{-iya} - e^{-iyb}}{iy} \tilde{\mu}(y) dy = \int_{-\infty}^{+\infty} \int_{-R}^{+R} \frac{e^{-iya} - e^{-iyb}}{iy} e^{iyx} dy \mu(dx).$$

Пользуясь формулой

$$e^{it} = \cos t + i \sin t,$$

преобразуем внутренний интеграл

$$\begin{aligned}
&\int_{-R}^{+R} \frac{e^{-iya} - e^{-iyb}}{iy} e^{iyx} dy = \int_{-R}^{+R} \frac{e^{iy(x-a)} - e^{iy(x-b)}}{iy} dy \\
&\quad = \int_{-R}^{+R} \frac{\cos y(x-a) - \cos y(x-b)}{iy} dy + \int_{-R}^{+R} \frac{\sin y(x-a) - \sin y(x-b)}{y} dy.
\end{aligned}$$

Первый интеграл равен нулю в силу нечетности подынтегральной функции. Преобразуя второй интеграл, получаем

$$\int_{-R}^{+R} \frac{e^{-iya} - e^{-iyb}}{iy} e^{iyx} dy = 2 \int_0^{+R} \frac{\sin y(x-a)}{y} dy - 2 \int_0^{+R} \frac{\sin y(x-b)}{y} dy.$$

Вводя подстановку $z = y(x-a)$ в первом слагаемом и $z = y(x-b)$ во втором, наконец получаем

$$\int_{-R}^{+R} \frac{e^{-iya} - e^{-iyb}}{iy} e^{iyx} dy = 2 \int_{R(x-b)}^{R(x-a)} \frac{\sin z}{z} dz.$$

Теперь заметим, что

$$\int_0^{+\infty} \frac{\sin z}{z} dz = \frac{\pi}{2}. \tag{1.4}$$

Тогда,

- если $x \notin [a,b]$, то
$$\int_{R(x-b)}^{R(x-a)} \frac{\sin z}{z} dz \to 0, \text{ при } R \to \infty;$$
- если $x \in (a,b)$, то
$$\int_{R(x-b)}^{R(x-a)} \frac{\sin z}{z} dz \to \pi, \text{ при } R \to \infty;$$
- если $x = a$, то
$$\int_{R(x-b)}^{R(x-a)} \frac{\sin z}{z} dz \to \pi/2, \text{ при } R \to \infty;$$
- если $x = b$, то
$$\int_{R(x-b)}^{R(x-a)} \frac{\sin z}{z} dz \to \pi/2, \text{ при } R \to \infty.$$

Подставляя в исходную формулу (и пользуясь теоремой Лебега о мажорируемой сходимости), мы получаем утверждение Предложения.

Для полноты картины докажем формулу (1.4). Имеем, что $1/z = \int_0^{+\infty} e^{-zt} dt$. Отсюда,

$$\int_0^{+\infty} \frac{\sin z}{z} dz = \int_0^{+\infty} \int_0^{+\infty} \sin z e^{-zt} dt dz = \int_0^{+\infty} \int_0^{+\infty} \sin z e^{-zt} dz dt.$$

Дважды взяв внутренний интеграл по частям (используя $e^{-zt}dz$ как dv), мы получаем, что

$$\int_0^{+\infty} \sin z e^{-zt} dz = \frac{1}{1+t^2}.$$

Тогда

$$\int_0^{+\infty} \frac{\sin z}{z} dz = \int_0^{+\infty} \frac{1}{1+t^2} dt = \pi/2.$$

$\square$

Очевидным образом из формулы обращения получаем

Следствие 7. *Вероятности μ_1 и μ_2 равны тогда и только тогда, когда равны их характеристические функции.*

Теперь докажем принципиально важное свойство, связывающее характеристические функции и слабую сходимость (собственно из-за этого свойства нам характеристические функции и нужны).

Теорема 7. *Последовательность вероятностей $\{\mu_n\}$ сходится к вероятности μ слабо тогда и только тогда, когда $\{\tilde{\mu}_n\}$ поточечно сходится к $\tilde{\mu}$.*

Доказательство этой теоремы (точнее достаточности, которая и является нетривиальной) базируется на следующей лемме.

Лемма 3. *Для каждого положительного r имеет место оценка:*

$$\mu([-2/r, 2/r]) \geq \left|\frac{1}{r}\int_{-r}^{r}\tilde{\mu}(y)dy\right| - 1.$$

Доказательство. Вспоминая формулу характеристической функции, переставляя интегралы по теореме Фубини, имеем, что

$$\begin{aligned}\frac{1}{2r}\int_{-r}^{r}\tilde{\mu}(y)dy = \frac{1}{2r}\int_{-r}^{r}\int_{-\infty}^{+\infty} e^{iyx}\mu(dx)dy = \frac{1}{2r}\int_{-\infty}^{+\infty}\int_{-r}^{r} e^{iyx}dy\mu(dx) \\ = \int_{-\infty}^{+\infty}\frac{e^{ixr}-e^{-ixr}}{2ixr}\mu(dx) = \int_{-\infty}^{+\infty}\frac{\sin(rx)}{rx}\mu(dx).\end{aligned}$$

Здесь мы воспользовались формулой

$$\frac{e^{it}+e^{-it}}{2i} = \sin t.$$

Далее,

$$\left|\int_{-\infty}^{+\infty}\frac{\sin(rx)}{rx}\mu(dx)\right| \leq \int_{|x|\leq 2/r}\left|\frac{\sin(rx)}{rx}\right|\mu(dx) + \int_{|x|>2/r}\left|\frac{\sin(rx)}{rx}\right|\mu(dx).$$

Поскольку $|\sin(rx)/rx| \leq 1$ и в области $|x| \geq 2/r$ $|\sin(rx)/rx| \leq 1/2$, получаем, что

$$\begin{aligned}\frac{1}{2r}\int_{-r}^{r}\tilde{\mu}(y)dy \leq \mu([-2/r, 2/r]) + \frac{1}{2}\mu(\mathbb{R}\setminus[-2/r, 2/r]) \\ = \mu([-2/r, 2/r]) + \frac{1}{2}(1-\mu([-2/r, 2/r])) = \frac{1}{2} + \frac{1}{2}\mu([-2/r, 2/r]).\end{aligned}$$

Перенося и умножая на 2, получаем заключение леммы. □

Доказательство теоремы 7. Если $\mu_n \xrightarrow{w} \mu$, то поскольку $e^{iyx} = \cos(yx) + i\sin(yx)$, и обе функции непрерывны и ограничены, характеристические функции $\tilde{\mu}_n(y)$ сходятся к $\tilde{\mu}(y)$ для каждого $y \in \mathbb{R}$.

Теперь предположим обратное, а именно, пусть $\tilde{\mu}_n$ сходится поточечно к $\tilde{\mu}$.

Докажем, что последовательность $\{\mu_n\}$ плотна, что позволит нам применить критерий компактности (теорему 6).

Выберем $\varepsilon > 0$. Поскольку $\tilde{\mu}(0) = 1$ и $\tilde{\mu}$ — непрерывная функция (докажите это!), то существует $r > 0$ такое, что $|\tilde{\mu}(y) - 1| < \varepsilon/4$ для всех $y \in (-r, r)$. Имеем,

$$\left| \int_{-r}^{r} \tilde{\mu}(y)dy \right| = \left| \int_{-r}^{r} (\tilde{\mu}(y) - 1)dy + 2r \right| \geq 2r - \left| \int_{-r}^{r} (\tilde{\mu}(y) - 1)dy \right|$$
$$\geq 2r - \int_{-r}^{r} |\tilde{\mu}(y) - 1|dy \geq 2r - 2r \cdot \frac{\varepsilon}{4} = 2r(1 - \varepsilon/4).$$

Поскольку $\tilde{\mu}_n$ сходится к $\tilde{\mu}$ поточечно и $|\tilde{\mu}_n(y)| \leq 1$, по теореме Лебега о мажорируемой сходимости имеем, что $\int_{-r}^{r} \tilde{\mu}_n(y)dy \to \int_{-r}^{r} \tilde{\mu}(y)dy$. Отсюда следует, что можно подобрать N такое, что для $n \geq N$

$$\left| \int_{-r}^{r} \tilde{\mu}_n(y)dy \right| \geq 2r(1 - \varepsilon/2).$$

Или

$$\left| \frac{1}{r} \int_{-r}^{r} \tilde{\mu}_n(y)dy \right| \geq 2 - \varepsilon.$$

Используя Лемму 3, мы получаем, что для $n \geq N$

$$\mu_n([-R, R]) \geq \mu_n([-2/r, 2/r]) \geq \left| \frac{1}{r} \int_{-r}^{r} \tilde{\mu}(y)dy \right| - 1 \geq 1 - \varepsilon$$

при $R \geq 2/r$. Оценка $\mu_n([-R_n, R_n]) \geq 1 - \varepsilon$ верна и для каждого n при каком-то $R_n > 0$. Принимая $R = \max(2/r, R_1, \ldots, R_N)$, имеем эту оценку с общим R для всех n. Это дает плотность последовательности $\{\mu_n\}$, а следовательно (по Теореме 6) можно выделить сходящуюся (в слабом смысле) подпоследовательность μ_{n_k} к некоторой предельной вероятности ν. Используя уже доказанную выше необходимость, мы получаем, что $\tilde{\mu}_{n_k}$ поточечно сходится к $\tilde{\nu}$. А значит, $\tilde{\nu} = \tilde{\mu}$. Вероятности равны тогда и только тогда, когда равны их характеристические функции (Следствие 7). Следовательно, $\nu = \mu$.

Остается доказать, что сама последовательность μ_n сходится к μ. Для этого предположим обратное, т.е, пусть нашлись число $\varepsilon > 0$, функция $\phi \in C_b(\mathbb{R})$ и подпоследовательность μ_{n_i} такие, что

$$\left| \int_{\mathbb{R}} \phi(x)\mu_{n_i}(dx) - \int_{\mathbb{R}} \phi(x)\mu(dx) \right| \geq \varepsilon.$$

Из μ_{n_i} можно выделить сходящуюся “подподпоследовательность”. Это противоречит предположению. $\square$

Теорема 8 (Закон больших чисел в форме Хинчина)**.** *Пусть дана последовательность независимых одинаково распределенных случайных величин ξ_n со средним m. Тогда имеет место закон больших чисел: $\frac{\xi_1+\cdots+\xi_n}{n}$ сходится по вероятности к m.*

Доказательство. Сделаем лишь набросок. По предыдущей теореме достаточно проверить сходимость характеристических функций. Отметим, что $\tilde{m}$, характеристическая функция случайной величины тождественно равной m, равна e^{itm}. Пусть $\eta_n = \frac{\xi_1+\cdots+\xi_n}{n}$. Тогда для всякого t

$$\begin{aligned}
\ln \tilde{\eta}_n(t) &= \sum_{k=1}^{n} \ln \tilde{\xi}_k(t/n) = n \ln \tilde{\xi}_1(t/n) = t \cdot \frac{\ln \tilde{\xi}_1(t/n) - \ln \tilde{\xi}_1(0)}{t/n}, \\
\lim_{n\to\infty} \ln \tilde{\eta}_n(t) &= t \frac{d \ln \tilde{\xi}_1(s)}{ds}\Big|_{s=0} = t \frac{d\tilde{\xi}_1(s)}{\tilde{\xi}_1(0) ds}\Big|_{s=0} = it\mathbb{E}\xi_1 = itm = \ln \tilde{m}(t),
\end{aligned}$$

то есть $\tilde{\eta}_n(t) \to \tilde{m}(t)$. Теперь по предыдущей теореме, из полученной поточечной сходимости следует, что η_n слабо сходится к m. Осталось воспользоваться Предложением 10. □

Как видно из доказательства, нам не столько нужно само матожидание, как производная у $\tilde{\xi}_1$ в нуле. Это позволяет писать закон больших чисел для распределений, не имеющих матожидания, лишь бы хвосты распределений убывали достаточно быстро. При этом вместо $\mathbb{E}\xi_1$, как интеграла по Лебегу, иногда подходит, например, главное значение несобственного интеграла (ищите слова *“центр распределения”*). Можно здесь требовать еще более слабые условия, нежели существование производной.

6 Центральная предельная теорема

Из показанного закона больших чисел следует, что если у независимых одинаково распределенных случайных величин ξ_n есть матожидание $m = m_n$, то $\frac{\xi_1+\cdots+\xi_n}{n}$ сходится по вероятности к m, то есть

$$\frac{\xi_1 + \cdots + \xi_n - nm}{n} \xrightarrow{P} 0.$$

Как говорит центральная предельная теорема, если у независимых одинаково распределенных случайных величин ξ_n есть дисперсия $\sigma^2 > 0$, то по распределению сходится $\frac{\xi_1+\cdots+\xi_n-nm}{\sqrt{n}}$, точнее

$$\frac{\xi_1 + \cdots + \xi_n - nm}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1).$$

В данном разделе мы будем изучать поведение центрированной и нормированной суммы случайных величин $\{\xi_n\}_{n=1}^{\infty}$. Обозначим $\mu_k \triangleq \xi_k \# \mathbb{P}$, $m_n \triangleq \mathbb{E}\xi_n$, $\sigma_n^2 \triangleq D\xi_n$, $\zeta_n \triangleq \xi_1 + \ldots + \xi_n$, $M_n \triangleq \mathbb{E}\zeta_n = \sum_{k=1}^n m_k$. $\mathcal{D}_n^2 \triangleq D(\zeta_n) = \sum_{k=1}^n \sigma_k^2$. От $\{\xi_n\}$ мы требуем, чтобы

- $\{\xi_n\}$ были независимыми;
- были конечными m_k и σ_k;
- было выполнено **условие Линдеберга**: для каждого $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{1}{\mathcal{D}_n^2} \sum_{k=1}^{n} \int_{\{x:|x-m_k|\geq\varepsilon\mathcal{D}_n\}} (x-m_k)^2 \mu_k(dx) = 0. \tag{1.5}$$

Теорема 9 (Центральная предельная теорема. Условие Линдеберга)**.** *Если $\{\xi_n\}$ независимы, m_n и σ_n конечны и выполнено условие Линдеберга, то распределение случайной величины*

$$\frac{\zeta_n - M_n}{D_n} = \frac{(\xi_1 - m_1) + \ldots + (\xi_n - m_n)}{\sqrt{\sigma_1^2 + \ldots + \sigma_n^2}}$$

слабо сходится к нормальному распределению с нулевым средним и единичной дисперсией.

Доказательство. Без ограничения общности будем считать, что $m_k = 0$ (в противном случае можно “сдвинуть” случайные величины, условие Линдеберга сохранится). Обозначим $\tau_n \triangleq \zeta_n/\mathcal{D}_n$. Необходимо доказать, что распределение τ_n слабо сходится к $N(0,1)$.

Мы воспользуемся методом характеристических функций. По теореме 7 достаточно доказать, что характеристическая функция случайной величины τ_n поточечно сходится к характеристической функции нормального распределения, которая равна $e^{-y^2/2}$. Напомним обозначения: $\tilde{\xi}_k(y)$ — характеристическая функция случайной величины ξ_k, $\tilde{\tau}_n(y)$ — характеристическая функция случайной величины τ_n: $\tilde{\xi}_k(y) = \mathbb{E}e^{i\xi_k y}$, $\tilde{\tau}_k(y) = \mathbb{E}e^{i\tau_k y}$.

Имеем, что

$$\tilde{\tau}_k(y) = \mathbb{E}e^{i\tau_n y} = \mathbb{E}e^{i(\xi_1+\ldots+\xi_n)\frac{y}{\mathcal{D}_n}} = \prod_{k=1}^{n} \tilde{\xi}_k\left(\frac{y}{\mathcal{D}_n}\right). \tag{1.6}$$

Утверждается (и это будет доказано позже), что

$$\tilde{\xi}_k\left(\frac{y}{\mathcal{D}_n}\right) = 1 - \frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + a_k^n, \tag{1.7}$$

где $a_k^n = a_k^n(y)$ и

$$\lim_{n\to\infty} \sum_{k=1}^{n} |a_k^n| = 0. \tag{1.8}$$

Напомним, что по формуле Тейлора для любого комплексного числа z такого, что $|z| < 1/4$, верно равенство $\ln(1+z) = z + \theta(z)|z|^2$ с $|\theta(z)| < 1$. Далее докажем, что

$$\lim_{n\to\infty} \max_{k=1,\ldots,n} \frac{\sigma_k^2}{\mathcal{D}_n^2} = 0. \tag{1.9}$$

В самом деле, для любого $\varepsilon > 0$ справедливо неравенство

$$\max_{k=1,\ldots,n} \frac{\sigma_k^2}{\mathcal{D}_n^2} \le \max_{k=1,\ldots,n} \frac{\int_{\{x:|x|\ge\varepsilon\mathcal{D}_n\}} x^2\mu_k(dx)}{\mathcal{D}_n^2} + \max_{k=1,\ldots,n} \frac{\int_{\{x:|x|<\varepsilon\mathcal{D}_n\}} x^2\mu_k(dx)}{\mathcal{D}_n^2}.$$

Первый член идет к нулю по условию Линдеберга, а второй может быть оценен числом ε на основе неравенства $\int_{\{x:|x|<\varepsilon\mathcal{D}_n\}} x^2\mu_k(dx) \le \varepsilon^2\mathcal{D}_n^2$. В силу того, что ε можно выбрать произвольно малым, (1.9) выполнено.

Теперь выберем $z = -(y^2\sigma_k^2)/(2\mathcal{D}_n^2) + a_k^n$ в формуле $\ln(1+z) = z + \theta(z)|z|^2$ для $\theta \in [-1,1]$ при $|z| \le 1/2$. Из (1.7) и разложения ln получаем, что

$$\sum_{k=1}^n \ln\tilde{\xi}_k\left(\frac{y}{\mathcal{D}_n}\right) = -\sum_{k=1}^n \frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + \sum_{k=1}^n a_k^n + \sum_{k=1}^n \theta\left|-\frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + a_k^n\right|^2, \tag{1.10}$$

где $|\theta_k| \le 1$. Первый член равен $-y^2/2$. Покажем, что остальные стремятся к нулю. Второй член стремится к нулю по выбору (1.8). Остается третий член. Имеем, что

$$\sum_{k=1}^n |\theta|\left|-\frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + a_k^n\right|^2 \le \max_{k=1,\ldots,n}\left[\frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + a_k^n\right]\sum_{k=1}^n\left|-\frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + a_k^n\right|$$
$$\le c(y)\max_{k=1,\ldots,n}\left[\frac{y^2\sigma_k^2}{2\mathcal{D}_n^2} + |a_k^n|\right].$$

Здесь $c(y)$ — некоторая константа (относительно n). Из (1.8) и (1.9) мы получаем, что правая часть неравенства стремится к нулю. Но тогда и третий член в (1.10) стремится к нулю. Отсюда (по модулю (1.7)) мы получаем центральную предельную теорему.

Остается доказать (1.7) и (1.8).

Используя ряд Тейлора (с остаточным членом в форме Лагранжа) получаем, что

$$e^{it} = 1 + it + \frac{\theta_1(t)t^2}{2}, \quad e^{it} = 1 + it - \frac{t^2}{2} + \frac{\theta_2(t)t^3}{6},$$

где $|\theta_1(t)|, |\theta_2(t)| \le 1$. Отсюда

$$\tilde{\xi}_k\left(\frac{y}{\mathcal{D}_n}\right) = \int_{-\infty}^{+\infty} e^{ix\frac{y}{\mathcal{D}_n}}\mu_k(dx)$$
$$= \int_{\{x:|x|\ge\varepsilon\mathcal{D}_n\}} e^{ix\frac{y}{\mathcal{D}_n}}\mu_k(dx) + \int_{\{x:|x|<\varepsilon\mathcal{D}_n\}} e^{ix\frac{y}{\mathcal{D}_n}}\mu_k(dx)$$
$$= \int_{\{x:|x|\ge\varepsilon\mathcal{D}_n\}}\left(1 + i\frac{y}{\mathcal{D}_n}x + \frac{\theta_1(x)(yx)^2}{2\mathcal{D}_n^2}\right)\mu_k(dx)$$
$$+ \int_{\{x:|x|<\varepsilon\mathcal{D}_n\}}\left(1 + i\frac{y}{\mathcal{D}_n}x - \frac{(yx)^2}{2\mathcal{D}_n^2} + \frac{\theta_2(x)(yx)^3}{6\mathcal{D}_n^3}\right)\mu(dx) = 1 - \frac{y^2\sigma_k^2}{\mathcal{D}_n^2} + a_k^n,$$

где

$$a_k^n = \frac{y^2}{2\mathcal{D}_n^2}\int_{\{x:|x|\ge\varepsilon\mathcal{D}_n\}}(1+\theta_1(x))x^2\mu_k(dx) + \int_{\{x:|x|<\varepsilon\mathcal{D}_n\}}\frac{\theta_2(x)(yx)^3}{6\mathcal{D}_n^3}\mu_k(dx).$$

Чтобы доказать (1.8), заметим, что

$$\frac{1}{\mathcal{D}_n^2}\sum_{k=1}^{n}\int_{\{x:|x|\geq\varepsilon\mathcal{D}_n\}}(1+\theta_1(x))x^2\mu_k(dx)$$

сходится к нулю по условию Линдеберга, а

$$\left|\sum_{k=1}^{n}\frac{y^3}{6\mathcal{D}_n^3}\int_{\{x:|x|<\varepsilon\mathcal{D}_n\}}\theta_2(x)x^3\mu_k(dx)\right| \leq \left|\sum_{k=1}^{n}\frac{y^3\varepsilon}{6\mathcal{D}_n^3}\int_{\{x:|x|<\varepsilon\mathcal{D}_n\}}\theta_2(x)x^2\mathcal{D}_n\mu_k(dx)\right| \leq \left|\sum_{k=1}^{n}\frac{y^3\varepsilon\sigma_k^2}{6\mathcal{D}_n^2}\right| = \frac{\varepsilon|y|^3}{6}.$$

Величина, стоящая в правой части, может быть сделана сколь угодно малой за счет выбора ε. □

Следствия из ЦПТ

Прежде всего мы дадим другую (более слабую) формулировку ЦПТ. Для этого мы заменим условие Линдеберга на **условие Ляпунова**, которое состоит в том, что для некоторого положительного $\delta > 0$

$$\lim_{n\to\infty}\frac{1}{\mathcal{D}^{2+\delta}}\sum_{k=1}^{n}\mathbb{E}(\xi_k-m_k)^{2+\delta}=0.$$

Теорема 10 (Центральная предельная теорема. Условие Ляпунова)**.** *Если $\{\xi_k\}$ независимы, m_k и σ_k конечны и выполнено условие Ляпунова, то распределение случайной величины*

$$\frac{\zeta_n-M_n}{D_n}=\frac{(\xi_1-m_1)+\ldots+(\xi_n-m_n)}{\sqrt{\sigma_1^2+\ldots+\sigma_n^2}}$$

слабо сходится к нормальному распределению с нулевым средним и единичной дисперсией.

Доказательство. Мы выведем из условия Ляпунова условие Линдеберга. Пусть $\varepsilon > 0$, имеем, что

$$\lim_{n\to\infty}\frac{1}{\mathcal{D}_n^2}\sum_{k=1}^{n}\int_{\{x:|x-m_k|\geq\varepsilon\mathcal{D}_n\}}(x-m_k)^2\mu_k(dx) \leq \lim_{n\to\infty}\frac{1}{\mathcal{D}_n^2(\varepsilon\mathcal{D}_n)^\delta}\sum_{k=1}^{n}\int_{\{x:|x-m_k|\geq\varepsilon\mathcal{D}_n\}}|x-m_k|^{2+\delta}\mu_k(dx) \leq \varepsilon^{-\delta}\frac{\sum_{k=1}^{n}\mathbb{E}|\xi_k-m_k|^{2+\delta}}{\mathcal{D}_n^{2+\delta}}=0.$$

□

В случае одинаково распределенных случайных величин, имеющих дисперсию, условие Линдеберга эквивалентно

$$\lim_{n\to\infty} \frac{1}{n\sigma_1^2} \int_{\{x:|x-m_k|\ge\varepsilon\sigma_n\sqrt{n}\}} (x-m_k)^2 \mu_1(dx) = 0,$$

оно всегда выполнено, потому имеет место:

Следствие 8. *Пусть $\{\xi_n\}$ независимы и одинаково распределены. Пусть $\mathbb{E}\xi_1 = m$, $D\xi_1 = \sigma^2 > 0$. Тогда распределения случайных величин $\frac{\xi_1+\ldots+\xi_n-nm}{\sigma\sqrt{n}}$ слабо сходятся к нормальному распределению с нулевым средним и единичной дисперсией, в частности, для всех $a, b \in \mathbb{R}\cup\{-\infty,+\infty\}, (a<b)$*

$$\mathbb{P}\Big(a < \frac{\xi_1+\ldots+\xi_n-nm}{\sigma\sqrt{n}} < b\Big) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Для схемы Бернулли, в которой все величины распределены одинаково и принимают значение 1 (успех) с вероятностью p и 0 (неудача) с вероятностью $1-p$, имеем, что $\mathbb{E}\xi_1 = p$, $D\xi_1 = p(1-p)$. Тогда

Следствие 9 (Интегральная предельная теорема Муавра-Лапласа)**.** *Пусть $\xi_1,\ldots,\xi_n$ — испытания в схеме Бернулли. Тогда*

$$\mathbb{P}\left(a < \frac{(\xi_1+\ldots+\xi_n)-np}{\sqrt{np(1-p)}} \le b\right) \to \frac{1}{2\pi}\int_a^b e^{-\frac{x^2}{2}dx} \text{ при } n\to\infty.$$

В условиях Следствия 8 можно оценить погрешность при известном третьем центральном моменте: имеет место неравенство Берри-Эссена

$$\left|F_{\frac{\xi_1+\ldots+\xi_n-nm}{\sigma\sqrt{n}}}(b) - \int_{-\infty}^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx\right| < \frac{1}{2}\cdot\frac{\mathbb{E}|\xi_1-m|^3}{\sigma^3\sqrt{n}} \qquad \forall b\in\mathbb{R}.$$

Соответствующая оценка равномерна по всем b, следовательно крайне грубо оценивает, например, хвосты (в приложениях обычно именно это важно). По этой причине для больших $|b|$ более разумно применять неравенства концентрации меры типа границы Чернова.

В заключение без доказательства сформулируем локальную предельную теорему. Для этого напомним понятие носителя меры (случайной величины). *Носителем меры* μ называют наименьшее замкнутое множество, вероятность которого равна 1; обозначают это множество через $\operatorname{supp}\mu = 1$.

<u>Подумать:</u> почему такое наименьшее замкнутое множество существует.

Теорема 11 (Локальная предельная теорема (без д-ва))**.** *Пусть $\{\xi_n\}$ независимы и одинаково распределены, $\mathbb{E}\xi_1 = m$, $D\xi_1 = \sigma^2$, а ξ_i принимают целочисленные значения, причем НОД$(\operatorname{supp}\mu_{\xi_1}) = 1$. Тогда*

$$\sigma\sqrt{n}\mathbb{P}\{\xi_1+\ldots+\xi_n-nm=k\} - \frac{1}{\sqrt{2\pi}}e^{-(k-nm)^2/2\sigma\sqrt{n}}$$

стремится при $n\uparrow\infty$ к нулю равномерно по всем целым k.

Подумать: приведите простейший пример, показывающий, что без условия НОД$(\operatorname{supp}\mu_{\xi_1}) = 1$ теорема была бы неверна.

Следствие 10 (Локальная предельная теорема Муавра-Лапласа)**.** *Пусть* $\xi_1, \ldots, \xi_n$ *— испытания в схеме Бернулли. Тогда*

$$\mathbb{P}(\xi_1 + \ldots + \xi_n = k) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{np(1-p)}}\right)^2} (1 + \delta_n(k)),$$

где $\delta_n(k)$ *стремятся к нулю равномерно по* k *при* $n \to \infty$.

Глава 2

Статистика

Цель статистики прямо противоположна целям теории вероятностей. Теория вероятностей ставит целью изучить, что мы будем наблюдать в результате тех или иных случайных событий на основании общих представлений о случайных величинах. В свою очередь, статистика по полученным результатам случайных событий пытается сделать выводы о случайной величине.

Задача статистики может быть сформулирована следующим образом. Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство, ξ — случайная величина, функция распределения которой F, вообще говоря, неизвестна. Проводится n испытаний, в результате мы получаем n чисел (каждое из них — случайная величина) $X = (X_1, \ldots, X_n)$, каждая X_i имеет распределение F. На основании полученных данных надо сделать вывод о распределении F или о таких характеристиках ξ, как математическое ожидание и дисперсия.

Ниже мы очень часто будем работать с нормальным распределением. Напомним, что нормальное распределение с параметрами μ и σ обозначается $N_{\mu,\sigma}$ и задается плотностью

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Несложно проверить, что если случайная величина распределена по нормальному закону, то ее математическое ожидание и дисперсия равны μ и σ^2 соответственно.

7 Выборочное распределение

КТО ВЫБЕРЕТ ЭТОТ БИЛЕТ ДОЛЖЕН ТАКЖЕ УМЕТЬ ДОКАЗЫВАТЬ ЗБЧ В ФОРМЕ ХИНЧИНА

Прежде всего предположим, что F нам неизвестна вообще. Тогда логично поставить вопрос о восстановлении распределения. Определим выборочную функцию распределения $F_n(x)$ по правилу:

$$F_n(x) \triangleq \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty,x]}(X_i).$$

Справедливо следующее утверждение.

Предложение 12. *Для любого $x \in \mathbb{R}$ $F_n(x) \xrightarrow{P} F(x)$.*

Доказательство. Заметим, что для каждого x $F_n(x)$ — случайная величина, получаем
$$F_n(x) = \frac{\sum_{i=1}^n \mathbf{1}_{(-\infty,x]}(X_i)}{n}.$$

Имеем, что $\mathbf{1}_{(-\infty,x]}(X_i)$ — одинаково распределены,
$$\mathbb{E}\mathbf{1}_{(-\infty,x]}(X_i) = 1 \cdot \mathbb{P}(X_i \le x) = \mathbb{P}(\xi \le x) = F(x) < \infty.$$

Также легко заметить, что у $\mathbf{1}_{(-\infty,x]}(X_i)$ есть дисперсии. Тогда по закону больших чисел
$$F_n(x) = \frac{\sum_{i=1}^n \mathbf{1}_{(-\infty,x]}(X_i)}{n} \xrightarrow{P} \mathbb{E}\mathbf{1}_{(-\infty,x]}(\xi) = F(x).$$

□

Также математическое ожидание F_n может быть использовано для оценки $F(x)$.

Предложение 13. *Имеют место следующие свойства:*

1. *$\mathbb{E}F_n(x) = F(x)$, т.е. $F_n(x)$ — несмещенная оценка $F(x)$;*
2. *$DF_n(x) = F(x)(1 - F(x))/n$;*
3. *если $0 < F(x) < 1$, то $\sqrt{n}(F_n(x) - F(x)) \xrightarrow{d} N_{0,F(x)(1-F(x))}$;*
4. *$nF_n(x)$ имеет биномиальное распределение с параметром $p = F(x)$.*

Доказательство. Заметим, что $\mathbf{1}_{X_1 \le x}$ распределена по закону Бернулли с вероятностью успеха $F(x)$. Следовательно, $\mathbb{E}\mathbf{1}_{X_1 \le x} = F(x)$, $D\mathbf{1}_{X_1 \le x} = F(x)(1-F(x))$. Поскольку случайные величины $\mathbf{1}_{X_1 \le x}, \dots, \mathbf{1}_{X_n \le x}$ независимы и одинаково распределены, заключаем, что
$$\mathbb{E}F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\mathbf{1}_{X_i \le x} = F(x),$$
$$DF_n(x) = \frac{1}{n^2}\sum_{i=1}^n D\mathbf{1}_{X_i \le x} = \frac{F(x)(1 - F(x))}{n}.$$

Таким образом, первые два утверждения доказаны.

Для того, чтобы доказать третье утверждение, заметим, что для случайных величин $\mathbf{1}_{X_i \le x}$ выполнены условия центральной предельной теоремы, включая условие Ляпунова. Следовательно,
$$\sqrt{n}(F_n(x) - F(x)) = \sqrt{n}\left(\frac{\sum_{i=1}^n \mathbf{1}_{X_i \le x}}{n} - F(x)\right) = \frac{\sum_{i=1}^n (\mathbf{1}_{X_i \le x} - F(x))}{\sqrt{n}}$$
$$\xrightarrow{d} N_{0,F(x)(1-F(x))} \text{ при } n \to \infty.$$

Это доказывает третье утверждение.

Наконец, последнее утверждение следует из того, что

$$nF_n(x) = \sum_{i=1}^{n} \mathbf{1}_{X_i \le x}$$

и $\mathbf{1}_{X_i \le x}$ распределены по закону Бернулли с вероятностью успеха $F(x)$. $\square$

Кроме оценки распределения $F(x)$ целиком очень важно узнать такие его параметры, как математическое ожидание и дисперсия. В качестве оценки математического ожидания используется выборочное среднее

$$\overline{X}_n \triangleq \frac{X_1 + \ldots X_n}{n}.$$

Предложение 14. *Справедливы следующие утверждения:*

1. *если* $\mathbb{E}|X_1| < \infty$, *то* $\mathbb{E}\overline{X}_n = \mathbb{E}X_1$;
2. *если* $\mathbb{E}|X_1| < \infty$, *то* $\overline{X}_n \xrightarrow{P} \mathbb{E}X_1$;
3. *если* $0 < DX_1 < \infty$, *то* $\sqrt{n}(\overline{X}_n - \mathbb{E}X_1) \xrightarrow{d} N_{0,DX_1}$.

Доказательство. Имеем, что

$$\mathbb{E}\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}X_i = \mathbb{E}X_1.$$

Это доказывает первое утверждение.

Чтобы доказать второе утверждение, воспользуемся законом больших чисел. Имеем, что

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n} \xrightarrow{P} \mathbb{E}X_1.$$

Наконец, третье утверждение следует из центральной предельной теоремы (в этом случае выполнено условие Ляпунова):

$$\sqrt{n}(\overline{X}_n - \mathbb{E}X_1) = \frac{\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)}{\sqrt{n}} \xrightarrow{d} N_{0,DX_1}.$$

$\square$

Аналогичное утверждение справедливо и для k-х моментов.

Предложение 15. *Справедливы следующие утверждения:*

1. *если* $\mathbb{E}|X_1|^k < \infty$, *то* $\mathbb{E}(\overline{X^k}_n) = \mathbb{E}X_1^k$;
2. *если* $\mathbb{E}|X_1|^k < \infty$, *то* $\overline{X^k}_n \xrightarrow{P} \mathbb{E}X_1^k$;
3. *если* $0 < D(X_1^k) < \infty$, *то* $\sqrt{n}((\overline{X^k}_n) - \mathbb{E}X_1^k) \xrightarrow{d} N_{0,DX_1^k}$.

Упражнение 2 (0.75 балла). Докажите предложение 15.

Для оценки дисперсии используются две величины

$$S_n^2 \triangleq \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2, \quad S_{0,n}^2 \triangleq \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$

которые называются соответственно смещенной и несмещенной выборочными дисперсиями.

Предложение 16. *Справедливы утверждения:*

1. $S_n^2 \xrightarrow{P} DX_1$, $S_{0,n}^2 \xrightarrow{P} DX_1$;
2. S_n^2 *— смещенная, а* $S_{0,n}^2$ *— несмещенная оценка дисперсии, т.е.*
$$\mathbb{E}S_n^2 = \frac{n-1}{n}DX_1, \quad \mathbb{E}S_{0,n}^2 = DX_1;$$
3. *выборочные дисперсии* S_n^2, $S_{0,n}^2$ *— асимптотически нормальны, т.е.*
$$\sqrt{n}(S_n^2 - DX_1) \xrightarrow{d} N_{0,\mathbb{E}(X_1-\mathbb{E}X_1)^2},$$
$$\sqrt{n}(S_{0,n}^2 - DX_1) \xrightarrow{d} N_{0,\mathbb{E}(X_1-\mathbb{E}X_1)^2}.$$

Доказательство. Имеем, что

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - (\overline{X}_n)^2) = \overline{(X^2)}_n - (\overline{X}_n)^2.$$

Отсюда по закону больших чисел следует, что $S_n^2 \xrightarrow{P} \mathbb{E}(X_1^2) - (\mathbb{E}X_1)^2 = DX_1$. Также, поскольку $\frac{n}{n-1} \to 1$, то мы получаем и сходимость $S_{0,n}^2$ к DX_1, что доказывает первое утверждение.

Чтобы доказать второе утверждение, заметим, что

$$\begin{aligned}\mathbb{E}S_n^2 &= \mathbb{E}(\overline{(X^2)}_n - (\overline{X}_n)^2) = \mathbb{E}(X_1)^2 - \left((\mathbb{E}\overline{X}_n)^2 + D\overline{X}_n\right) \\ &= \mathbb{E}(X_1)^2 - (\mathbb{E}\overline{X}_n)^2 - D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = DX_1 - \frac{1}{n^2}nDX_1 = \frac{n-1}{n}DX_1.\end{aligned}$$

Отсюда же мы получаем, что $\mathbb{E}S_{0,n}^2 = DX_1$.

Далее, чтобы упростить обозначения, положим $a \triangleq \mathbb{E}X_1$, $\sigma^2 \triangleq DX_1$.

Для того, чтобы доказать третье утверждение, заметим, что

$$\begin{aligned}S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 &= \frac{1}{n}\sum_{i=1}^{n}((X_i - a) - (\overline{X}_n - a))^2 \\ &= \overline{(X-a)^2}_n - (\overline{X}_n - a)^2.\end{aligned}$$

Тогда

$$\begin{aligned}\sqrt{n}(S^2-\sigma^2) &= \sqrt{n}(\overline{(X-a)^2}_n - (\overline{X}_n - a)^2 - \sigma^2) \\ &= \sqrt{n}(\overline{(X-a)^2}_n - \mathbb{E}(X_1-a)^2) - \sqrt{n}(\overline{X}_n - a)^2 \\ &= \frac{\sum_{i=1}^n[(X_i-a)^2 - \mathbb{E}(X_1-a)^2]}{\sqrt{n}} - (\overline{X}_n - a)\sqrt{n}(\overline{X}_n - a) \\ &\xrightarrow{d} N_{0,D(X_1-a)^2}.\end{aligned}$$

В самом деле, первое слагаемое сходится к $N_{0,D(X_1-a)^2}$ по распределению в силу центральной предельной теоремы, а второе сходится к нулю как произведение сходящихся к нулю по вероятности случайных величин $((\overline{X}_n - a))$ и сходящихся к $N_{0,D(X_1-a)}$ по распределению величин $\sqrt{n}(\overline{X}_n - a)$.

Случай $S^2_{0,n}$ рассматривается аналогично. □

8 Точечное оценивание

Мы предполагаем, что результаты испытаний $X_1, \dots, X_n$ распределены в соответствии с некоторой функцией распределения F_θ, от этой функции нам известно все, кроме θ. Про θ нам известно, что $\theta \in \Theta$. Наша задача найти θ. Для этого мы построим функцию $\theta^*(X) = \theta^*(X_1, \dots, X_n)$, которая называется *статистикой*, она и будет давать оценку θ. В дальнейшем, если $g : \mathbb{R}^k \to \mathbb{R}$ — некоторая функция, то $\mathbb{E}_\theta g(X_{i_1}, \dots, X_{i_k})$ — математическое ожидание в силу предположения, что θ известно.

Прежде всего выпишем требования, обычно предъявляемые к оценкам.

Определение 7. Статистика $\theta^*(X) = \theta^*(X_1, \dots, X_n)$ называется *несмещенной*, если $\mathbb{E}_\theta \theta^* = \theta$ для любого $\theta \in \Theta$.

Определение 8. Статистика называется состоятельной, если $\theta^* \xrightarrow{P} \theta$ для любого θ.

8.1 Сравнение оценок. Эффективность оценок

Определение 9. Пусть $\hat\theta_1$, $\hat\theta_2$ — две оценки (статистики) параметра θ. Говорят, что $\hat\theta_1$ предпочтительнее $\hat\theta_2$, если для всех $\theta \in \Theta$

$$\mathbb{E}_\theta(\hat\theta_1 - \theta)^2 \le \mathbb{E}_\theta(\hat\theta_2 - \theta)^2,$$

а для некоторого $\theta_0 \in \Theta$

$$\mathbb{E}_\theta(\hat\theta_1 - \theta_0)^2 < \mathbb{E}_\theta(\hat\theta_2 - \theta_0)^2.$$

Определение 10. Оценка θ^* называется эффективной, если она несмещенная и

$$\mathbb{E}_\theta(\theta^* - \theta)^2 = \min_{\hat\theta : \mathbb{E}_\theta \hat\theta = \theta} \mathbb{E}_\theta(\hat\theta - \theta)^2.$$

Определение 11. Говорят, что оценка θ^* — асимптотически нормальна, если для некоторой последовательности $\{c_n\}_{n=1}^{\infty}$, сходящейся к нулю, выполнено

$$\frac{\theta^*(X_1,\ldots,X_n)-\theta}{c_n} \xrightarrow{d} N_{0,1}.$$

Пусть $X=(X_1,\ldots,X_n)$. В дальнейшем обозначим через $f(\theta,x)$ плотность F_θ, если F_θ соответствует абсолютно непрерывной случайной величине, и $\mathbb{P}_\theta(\xi = x)$, если случайная величина дискретная.

Функцией максимального правдоподобия назовем величину

$$L(\theta,X) \triangleq \prod_{i=1}^{n} f(\theta,X_i).$$

Определение 12. Статистическая модель называется регулярной (по Рао-Крамеру), если

1. $L(\theta,X)>0$ и L дифференцируема по θ;

2. функция
$$U(\theta,X)=\frac{\partial \ln L(\theta,X)}{\partial\theta}=\sum_{i=1}^{n}\frac{\partial \ln f(\theta,X_i)}{\partial\theta}$$
имеет ограниченную дисперсию, т.е. $0<D_\theta U(\theta,X)<\infty$;

3. для любой статистики $\hat\theta$ имеет место равенство
$$\frac{\partial}{\partial\theta}\int_{\mathbb{R}^n}\hat\theta(X)L(\theta,X)dX=\int_{\mathbb{R}^n}\hat\theta(X)\frac{\partial}{\partial\theta}L(\theta,X)dX.$$

Заметим, что всегда (поскольку для каждого θ $L(\theta,X)$ — плотность) выполняется равенство
$$\int_{\mathbb{R}^n}L(\theta,X)dX\equiv 1.$$
Дифференцируя это равенство, имеем, что

$$\frac{\partial}{\partial\theta}\int_{\mathbb{R}^n}L(\theta,X)dX=\int_{\mathbb{R}^n}\frac{\partial \ln L(\theta,X)}{\partial\theta}L(\theta,X)dX=0.$$

Следовательно,
$$\mathbb{E}_\theta U(\theta,X)=\mathbb{E}_\theta\frac{\partial \ln L(\theta,X)}{\partial\theta}=0.$$

Определение 13. Информацией Фишера назовем дисперсию U

$$I_n(\theta)\triangleq D_\theta U(\theta,X)=\mathbb{E}_\theta U^2(\theta,X)=\int_{\mathbb{R}^n}\left(\frac{\partial \ln L(\theta,X)}{\partial\theta}\right)^2 L(\theta,X)dX.$$

Теорема 12 (Неравенство Рао-Крамера)**.** *Рассмотрим статистику* $\hat{\theta}(X)$. *Пусть*

$$b(\theta) \triangleq \mathbb{E}_\theta \hat{\theta} - \theta$$

— смещение оценки. Если $b(\theta)$ *дифференцируемо, то*

$$D_\theta(\hat{\theta} - \theta - b(\theta)) \geq \frac{(1 + b'(\theta))^2}{I_n(\theta)}.$$

При этом равенство имеет место лишь тогда, когда $\hat{\theta} - \theta - b(\theta) = a(\theta) U(\theta, X)$, *для некоторой функции* $a(\theta)$.

Заметим, что если оценки несмещенные, то $b(\theta) = 0$, $I_n(\theta) = nI_1(\theta)$ и неравенство Рао-Крамера принимает вид

$$D_\theta(X) \geq \frac{1}{nI_1(\theta)}.$$

Доказательство теоремы 12. Имеем, что $\theta + b(\theta) = \mathbb{E}_\theta \hat{\theta}(x)$. Дифференцируя это равенство, мы получаем, что

$$\begin{aligned} 1 + b'(\theta) = \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} \hat{\theta}(x) L(\theta, X) dX = \int_{\mathbb{R}^n} \hat{\theta}(x) \frac{\partial \ln L(\theta, X)}{\partial \theta} L(\theta, X) dX \\ = \int_{\mathbb{R}^n} \hat{\theta}(x) U(\theta, X) L(\theta, X) dX = \mathbb{E}_\theta(\hat{\theta}(x) U(\theta, X)). \end{aligned}$$

Поскольку $\mathbb{E}_\theta U(\theta, X) = 0$, заключаем, что

$$1 + b'(\theta) = \mathbb{E}_\theta \hat{\theta}(x) U(\theta, X) = \mathbb{E}_\theta[(\hat{\theta}(X) - \mathbb{E}_\theta \hat{\theta}(X))(U(\theta, X) - \mathbb{E}_\theta(\theta, X))].$$

Чтобы убедиться в справедливости этого равенства, достаточно просто раскрыть скобки в правой части.

Воспользовавшись неравенством Коши-Буняковского-Шварца, мы получаем, что

$$\begin{aligned} (1 + b'(\theta))^2 = \Big[\mathbb{E}_\theta[(\hat{\theta}(X) - \mathbb{E}_\theta \hat{\theta}(X))(U(\theta, X) - \mathbb{E}_\theta(\theta, X))]\Big]^2 \\ \leq \mathbb{E}_\theta[(\hat{\theta}(X) - \mathbb{E}_\theta \hat{\theta}(X))^2] \mathbb{E}_\theta[(U(\theta, X) - \mathbb{E}_\theta(\theta, X))^2] \\ = \mathbb{E}_\theta(\hat{\theta} - \theta + b(\theta))^2 \cdot I_n(X). \end{aligned}$$

Это и доказывает теорему. Заметим, что неравенство Коши-Буняковского-Шварца заменяется на равенство в случае линейной зависимости. □

Из теоремы Рао-Крамера следует, что дисперсия каждой оценки не превосходит величины $1/I_n$. Введем отношение верхней границы и дисперсии

$$e_n(\hat{\theta}) \triangleq \frac{1/I_n(\theta)}{D_\theta(\hat{\theta})}.$$

Заметим, что $e_n \in (0, 1]$.

Определение 14. Если $e_n(\hat{\theta}) \to 1$, то оценка эффективна.

8.2 Метод максимального правдоподобия

Идея метода максимального правдоподобия очень проста — максимизировать значение $L(\theta, X)$ по θ, т.е. положим

$$\theta^*(X) \stackrel{\triangle}{=} \operatorname{argmax}\{L(\theta, X) : \theta \in \Theta\}.$$

Пример 6. Пусть F_θ равномерные распределения на $[0, \theta]$, плотность каждого распределения равна

$$f_\theta(x) = \frac{1}{\theta}\mathbf{1}_{[0,\theta]}(x).$$

Отсюда

$$L(\theta, x_1, \ldots, x_n) = \left\{ \begin{array}{r} 1/\theta^n, \quad x_i \in [0, \theta], \quad i = 1, \ldots, n \\ 0, \quad \text{иначе.} \end{array} \right\}$$

Максимум L достигается при $\theta^* = \max\{X_i\}$.

Теорема 13 (Дюге). *Предположим, что выполнены условия регулярности:*

1. *Θ — невырожденный замкнутый интервал $\mathbb{R}$, существуют $\partial \ln f_\theta / \partial\theta$, $\partial^2 \ln f_\theta / \partial\theta^2$, $\partial^3 \ln f_\theta / \partial\theta^3$;*

2. *для всех θ*

$$\left|\frac{\partial \ln f_\theta(\theta, x)}{\partial\theta}\right| \le g_1(x), \quad \left|\frac{\partial^2 \ln f_\theta(\theta, x)}{\partial\theta^2}\right| \le g_2(x), \quad \left|\frac{\partial^3 \ln f_\theta(\theta, x)}{\partial\theta^3}\right| \le H(x),$$

где g_1, g_2 интегрируемы, а H удовлетворяет условию

$$\mathbb{E}_\theta H = \int_{\mathbb{R}} H(x) f_\theta(x) dx = \text{const};$$

3.

$$0 < \int_{\mathbb{R}} \left|\frac{\partial \ln f_\theta(\theta, x)}{\partial\theta}\right|^2 f_\theta(x) dx < \infty.$$

Тогда оценка, построенная по методу максимального правдоподобия, состоятельна, асимптотически эффективна и асимптотически нормальна.

8.3 Метод моментов

Метод моментов предлагает оценить θ через некоторые вспомогательные функции. А именно, пусть g и h таковы, что

$$\mathbb{E}g(X_1) = h(\theta).$$

Тогда в качестве θ^* возьмем решение уравнения

$$\overline{g(X)}_n = h(\theta),$$

или что тоже самое

$$\theta^* = h^{-1}\left(\frac{1}{n}\sum_{i=1}^n g(X_i)\right).$$

Пример 7. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на $[0, \theta]$. Выберем $g(y) = y$. Получаем, что $\mathbb{E}X_1 = \theta/2$. Отсюда, $h(\theta) = 2\theta$. Тогда $\theta^* = 2\overline{X}_n$.

Пример 8. Пусть теперь $g(y) = y^k$. Тогда

$$\theta^* = \sqrt[k]{(k+1)\overline{X}_n}.$$

Пример 9. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Выберем $g(y) = \mathbf{1}_{y=1}$. Имеем, что

$$h(\lambda) = \mathbb{P}_\lambda(X_1 = 1) = \lambda e^{-\lambda}.$$

Тогда

$$\lambda^* e^{-\lambda^*} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1} X_i = 1.$$

Пример 10. Пусть $X_1, \ldots, X_n$ выборка, распределенная согласно распределению $N_{a,1}$, где $a \geq 0$. Тогда $a^* = \overline{X}_n$, а т.к. $a \geq 0$, то эту оценку надо немного исправить и выбрать $a^* = \max\{0, \overline{X}_n\}$.

Главное свойство метода моментов — состоятельность.

Предложение 17. *Пусть h^{-1} существует и непрерывна. Тогда оценка $\theta^* = h^{-1}(\overline{g(X)}_n)$ состоятельна.*

Доказательство. По закону больших чисел имеем, что

$$\overline{g(X)}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{P} \mathbb{E}_\theta g(X_1) = h(\theta).$$

Воспользовавшись непрерывностью h^{-1}, мы получаем утверждение предложения. □

Заметим, что несмещенность оценки, построенной по методу моментов, — это скорее исключение. С другой стороны, обычно эта оценка асимптотически нормальна.

9 Задача различия двух гипотез

Предположим, что мы получили выборку и на основе той или иной априорной информации смогли сузить множество возможных решений до двух гипотез. Самыми яркими примерами таких решения являются вопросы о том, совпадают или нет у двух рядов выборочных данных среднее значение или дисперсия. Другой вопрос, который также часто встает в приложениях: есть ли корреляция между двумя рядами данных.

Формализуем данную задачу. Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство, $X = (X_1, \ldots, X_n)$ — некоторая выборка (т.е. X_i результаты испытаний некоторой случайной величины ξ). Пусть нам известно, что функция распределения случайной величины F лежит в некотором семействе Ξ. На основании этой информации мы хотим принять гипотезу H_0, которая состоит в том, что $F \in \widehat{\Xi} \subset \Xi$, или отвергнуть ее и тем самым принять гипотезу $H_1 = \{F \in \Xi \setminus \widehat{\Xi}\}$. Для того, чтобы принять/отвергнуть гипотезу H_0, мы будем использовать критерий, основанный на некотором борелевском множестве (S-критерий). В случае, если $X \notin S$ — гипотеза H_0 принимается, в противном случае — отвергается. Таким образом, множество S — это критическое множество.

Принять решение по имеющимся данным со стопроцентной гарантией невозможно. И ошибки здесь неизбежны. Наша задача их минимизировать.

Ошибка первого рода возникает, когда мы отвергли H_0, в то время как она верна; *ошибка второго рода* — это ситуация, когда мы приняли H_0, в то время когда верна гипотеза H_1.

Если мы считаем, что $\Xi = \{F_\theta : \theta \in \Theta\}$, то задача упрощается. Мы можем считать, что гипотеза H_0 состоит в том, что $\theta \in \Theta_0$, а гипотеза H_1 в том, что $\theta \in \Theta_1$. Ясно, что $\Theta = \Theta_0 \cup \Theta_1$, $\Theta_0 \cap \Theta_1 = \varnothing$. Если $\Theta_0 = \{\theta_0\}$, то, что H_0 — простая гипотеза, в противном случае речь идет о сложной гипотезе.

Напомним, что для заданного θ F_θ определяет вероятность на $\mathbb{R}^n$ по следующему правилу:

$$\mathbb{P}_\theta((a_1, b_1] \times \ldots \times (a_n, b_n]) \stackrel{\triangle}{=} (F_\theta(b_1) - F(a_1)) \cdot \ldots \cdot (F_\theta(b_n) - F(a_n)).$$

Также при фиксированном θ введем функцию мощности S-критерия (определена на Θ):

$$g_S(\theta) \stackrel{\triangle}{=} \mathbb{P}_\theta(S).$$

Заметим, что для $\theta \in \Theta_0$ число $g_S(\theta)$ равно вероятности ошибки первого рода. При $\theta \in \Theta_1$ число $1 - g_S(\theta)$ равно вероятности ошибки второго рода.

- Размером критерия называют

  $$\sup_{\theta \in \Theta_0} g_S(\theta).$$

  Размер критерия — это максимальная вероятность совершить ошибку первого рода. Задав число α, мы можем потребовать, чтобы

  $$\sup_{\theta \in \Theta_0} g_S(\theta) \leq \alpha.$$

  Число α называется уровнем значимости.

- Мощностью критерия называют величину

  $$\inf_{\theta \in \Theta_1} g_S(\theta).$$

  Разница между 1 и мощностью критерия — максимальная вероятность совершить ошибку второго рода.

К сожалению, критерии, основанные на множестве S, не всегда дают хороший результат. Чтобы его улучшить, переходят к рандомизации. Пусть ϕ функция из $\mathbb{R}^n$ в $[0, 1]$. На основе ϕ будем принимать решение следующим образом:

- если $\phi(X) = 1$, то H_0 *отвергается*;
- если $\phi(X) = 0$, то H_0 *принимается*;
- если $\phi(x) \in (0, 1)$, то H_0 отклоняется с вероятностью $\phi(X)$ (для этого необходимо провести дополнительный эксперимент, к примеру, подбросив “монетку”).

Заметим, что если ϕ принимает всего два значения 0 и 1, то $S = \{X : \phi(X) = 1\}$. Также предыдущий случай можно свести к данному, положив $\phi \triangleq \mathbf{1}_S$. Также функция мощности рандомизированного критерия есть $g_\phi(\theta) = \mathbb{E}_\theta \phi$. Размер критерия при этом равен

$$A_1(\phi) = \sup_{\theta \in \Theta_0} g_\phi(\theta) = \sup_{\theta \in \Theta_0} \mathbb{E}_\theta \phi,$$

а мощность

$$\inf_{\theta \in \Theta_1} g_\phi(\theta) = \inf_{\theta \in \Theta_1} \mathbb{E}_\theta \phi.$$

Обозначим

$$A_2(\phi) \triangleq 1 - \inf_{\theta \in \Theta_1} g_\phi(\theta) = 1 - \inf_{\theta \in \Theta_1} \mathbb{E}_\theta \phi.$$

Сравнивать критерии можно различными способами.

Определение 15. Критерий ϕ^1 не хуже ϕ^2 в минимаксном смысле, если

$$\max\{A_1(\phi^1), A_2(\phi^1)\} \le \max\{A_1(\phi^2), A_2(\phi^2)\}.$$

Определение 16. Критерий ϕ^1 не хуже ϕ^2 в байесовском смысле для констант r, s $(r, s > 0,\ r + s = 1)$, если

$$rA_1(\phi^1) + sA_2(\phi^1) \le rA_1(\phi^2) + sA_2(\phi^2).$$

Заметим, что оба определения фактически задают порядок на $\mathbb{R}^2$.

Также сравнивать элементы $\mathbb{R}^2$ можно при фиксировании одной координаты.

Определение 17. Говорят, что критерий ϕ^1 мощнее критерия ϕ^2 при заданном размере α, если $A_1(\phi^1) = A_1(\phi^2) = \alpha$ и $A_2(\phi^1) \le A_2(\phi^2)$.

Далее рассмотрим случай двух простых гипотез, т.е. $\Theta = \{\theta_0, \theta_0\}$. Гипотеза H_0 состоит в том, что $\theta = \theta_0$, гипотеза H_1 в том, что $\theta = \theta_1$.

Теорема 14 (Лемма Неймана-Пирсона)**.** *Пусть*

- $\theta_0, \theta_1 \in \mathbb{R}$*,* $\theta_0 < \theta_1$*;*

- *функции распределения F_{θ_i} абсолютно непрерывны; плотность F_{θ_i} — f_i;*
- $f_i > 0$;
- $L_i(X) = f_1(X_1) \cdot \ldots \cdot f_n(X_n)$;
- $\alpha \in (0,1)$.

Определим стратегию

$$\phi_{c,\varepsilon} \triangleq \begin{cases} 0, & L_0(X) > cL_1(X), \\ \varepsilon, & L_0(X) = cL_1(X), \\ 1, & L_0(X) < cL_1(X). \end{cases}$$

Тогда

1. *если c и ε выбраны так, чтобы $A_1(\phi_{c,\varepsilon}) = A_2(\phi(\varepsilon))$, то $\phi_{c,\varepsilon}$ — наилучший критерий в минимаксном смысле;*
2. *если $c = s/r$, то при любом ε $\phi_{c,\varepsilon}$ — наилучший критерий в байесовском смысле;*
3. *если c и ε таковы, что $\mathbb{E}_0\phi_{c,\varepsilon} = \alpha$, то $\phi_{c,\varepsilon}$ — наиболее мощный критерий.*

Также требуемые константы существуют.

Доказательство. Покажем существование требуемых констант. В случае два это очевидно.

Для случая три определим $h(c) \triangleq \mathbb{P}_0(L_0(Y) < cL_1(Y))$. Заметим, что $h(c)$ не убывает, $h(0) = 0$, $h(+\infty) = 1$. Тогда положим

$$c_\alpha \triangleq \operatorname{argmax}\{c : h(c) \geq \alpha > h(c-0)\}.$$

$$\varepsilon_\alpha \triangleq \begin{cases} 0, & c_\alpha \text{ точка непрерывности } h; \\ \frac{h(c_\alpha)-\alpha}{h(c_\alpha)-h(c_\alpha-0)}, & \text{иначе.} \end{cases}$$

Легко заметить, что при таком выборе c_α и ε_α $A_1(\phi_{c_\alpha,\varepsilon_\alpha}) = \alpha$.

Для первого случая введем также функцию $h_1(c) \triangleq 1 - \mathbb{P}_1(L_0(Y) < cL_1(Y))$. Заметим, что эта функция не убывает и $h_1(0) = 1$, $h_1(+\infty) = 0$. Выберем c_* равным

$$c_* = \operatorname{argmax}\{c : h(c) \leq h_1(c)\}.$$

Далее выберем

$$\varepsilon_* \triangleq \frac{1 - \mathbb{P}_0(L_0 < cL_1) - \mathbb{P}_1(L_0 < cL_1)}{\mathbb{P}_0(L_1 = cL_0) + \mathbb{P}_1(L_0 = cL_1)}.$$

Заметим, что $A_1(\phi_{c_*,\varepsilon_*}) = A_2(\phi_{c_*,\varepsilon_*})$.

Далее, если мы введем функцию

$$\psi_c(Y) \triangleq \min\{L_0(Y), cL_1(Y)\},$$

то

$$A_1(\phi_{c,\varepsilon}) + cA_2(\phi_{c,\varepsilon}) = \int_{\mathbb{R}^n} \psi_c(Y)dY. \tag{2.1}$$

В самом деле,

$$\begin{aligned}
&A_1(\phi_{c,\varepsilon}) + cA_2(\phi_{c,\varepsilon}) \\
&\quad = \mathbb{P}_0(L_0 < cL_1) + \varepsilon\mathbb{P}_0(L_0 = cL_1) + c(1 - \mathbb{P}_1(L_0 < cL_1) - \varepsilon\mathbb{P}_1(L_0 = cL_1)) \\
&\quad = \mathbb{P}_0(L_0 < cL_1) + \varepsilon\mathbb{P}_0(L_0 = cL_1) + c\mathbb{P}_1(L_0 > cL_1) + (1-\varepsilon)c\mathbb{P}_1(L_0 = cL_1) \\
&\quad = \int_{L_0 < cL_1} L_0(y)dy + \varepsilon \int_{L_0 = cL_1} L_0(y)dy \\
&\quad + \int_{L_0 > cL_1} (cL_1(y))dy + (1-\varepsilon) \int_{L_0 = cL_1} (cL_1(y))dy \\
&\quad = \int_{\mathbb{R}^n} \psi_c(y)dy.
\end{aligned}$$

Далее, пусть ϕ — произвольный критерий. Используя (2.1) имеем, что

$$\begin{aligned}
A_1(\phi) + cA_2(\phi) &= \mathbb{E}_0\phi + c(1 - \mathbb{E}_1\phi) \\
&= \int_{\mathbb{R}^n} \phi(y)L_0(y)dy + \int_{\mathbb{R}^n} (1-\phi(y))cL_1(y)dy \\
&\geq \int_{\mathbb{R}^n} \phi(y)\psi_c(y)dy + \int_{\mathbb{R}^n} (1-\phi(y))\psi_c(y)dy \\
&= \int_{\mathbb{R}^n} \psi_c(y)dy = A_1(\phi_{c,\varepsilon}) + cA_2(\phi_{c,\varepsilon}).
\end{aligned} \tag{2.2}$$

Теперь покажем выполнение всех пунктов теоремы.

1. Покажем, что в этом случае критерий — минимаксный. В самом деле, так как $A_1(\phi_{c_*,\varepsilon_*}) = A_2(\phi_{c_*,\varepsilon_*})$, то $A_1(\phi_{c_*,\varepsilon_*}) + c_*A_2(\phi_{c_*,\varepsilon_*}) = (1 + c_*)\max\{A_1(\phi_{c_*,\varepsilon_*}), A_2(\phi_{c_*,\varepsilon_*})\}$. Отсюда с использованием (2.2) для любого критерия ϕ получаем

   $$\begin{aligned}
   (1 + c_*)\max\{A_1(\phi_{c_*,\varepsilon_*}), A_2(\phi_{c_*,\varepsilon_*})\} &= A_1(\phi_{c_*,\varepsilon_*}) + c_*A_2(\phi) \\
   &\leq A_1(\phi_{c_*,\varepsilon_*}) + c_*A_2(\phi) \leq (1 + c_*)\max\{A_1(\phi), A_2(\phi)\}.
   \end{aligned}$$

   Таким образом, критерий ϕ_{c_*,ε_*} — минимаксный.

2. При $c = r/s$ и произвольном ε имеем по (2.2), что

   $$A_1(\phi) + \frac{s}{r}A_2(\phi_{s/r,\varepsilon}) \leq A_1(\phi) + \frac{s}{r}A_2(\phi).$$

   Это и есть определение байесовского наилучшего критерия.

3. Наконец, пусть ϕ таков, что $A_1(\phi) = \alpha$ имеем, что

   $$\begin{aligned}
   \alpha + c_\alpha A_2(\phi_{c_\alpha,\varepsilon_\alpha}) &= A_1(\phi_{c_\alpha,\varepsilon_\alpha}) + c_\alpha A_2(\phi_{c_\alpha,\varepsilon_\alpha}) \\
   &\leq A_1(\phi) + c_\alpha A_2(\phi) = \alpha + c_\alpha A_2(\phi).
   \end{aligned}$$

   Отсюда заключаем, что $\phi_{c_\alpha,\varepsilon_\alpha}$ — наиболее мощный критерий.

□

Глава 3

Случайные процессы

Случайным процессом называют семейство случайных величин X_t ($t \in T$), определенных на одном и том же вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$. Индекс t будем называть временем. В качестве множества T обычно выбирают либо $\mathbb{N}$ (в этом случае будем считать, что $0 \in \mathbb{N}$), либо $\mathbb{Z}$, либо $\mathbb{R}^+ = [0, +\infty)$, либо $\mathbb{R}$.

Для того, чтобы описать случайный процесс $X_\bullet$, изучают совместные распределения случайных векторов $(X_{t_1}, \ldots, X_{t_N})$. Самый простой способ сделать это — ввести совместную функцию распределения $F_{t_1,\ldots,t_N}(x_1, \ldots, x_N) = \mathbb{P}(X_{t_1} \leq x_1, \ldots, X_{t_N} \leq x_N)$. Оказывается, что

1. $F_{t_1,\ldots,t_N}$ — функция распределения, т.е. монотонна по каждой переменной, непрерывна справа, принимает значения из $[0, 1]$, а кроме того $F_{t_1,\ldots,t_{i-1},t_i,t_{i+1},\ldots,t_N}(x_1, \ldots, x_{i-1}, -\infty, x_{i+1}, \ldots, x_N) = 0$, $F_{t_1,\ldots,t_N}(+\infty, \ldots, +\infty) = 1$;

2. выполнено условие согласованности:

$$F_{t_1,\ldots,t_{i-1},t_i,t_{i+1},\ldots,t_N}(x_1, \ldots, x_{i-1}, +\infty, x_{i+1}, \ldots, x_N) \\ = F_{t_1,\ldots,t_{i-1},t_{i+1},\ldots,t_N}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N).$$

Оказывается (и об этом говорит теорема Колмогорова о построении случайных процессов), что по любому семейству функций распределения $F_{t_1,\ldots,t_N}$, удовлетворяющему условию согласованности, можно построить как минимум один случайный процесс. Но изучать случайные процессы "в целом" несколько неудобно, и поэтому логично предположить те или иные свойства, чтобы получить конкретные результаты.

10 Марковские цепи с дикретным временем

Прежде всего мы предположим, что случайный процесс X_t таков, что

- t принимает только целые неотрицательные значения, т.е. время дискретно;

- множество значений X_t не более чем счетно;
- (марковость) вероятность состояния X_{t+1} зависит лишь от состояния X_t, т.е.
$$\mathbb{P}(X_{t+1}|X_t, X_{t-i_1} \ldots, X_{t-i_n}) = \mathbb{P}(X_{t+1}|X_t).$$

В этом случае говорят о марковских цепях (с дискретным временем). Марковское свойство и марковские цепи (как объект) названы в честь А.А. Маркова-старшего (1856-1922).

Особенности марковских цепей позволяют, во-первых, считать, что X_t принимают значения в множестве $\mathbb{N}$. Мы можем задаться вопросом о том, какова вероятность того, что $X_t = i$, где $i = 1, 2, \ldots$. Обозначим $p_i(t) = \mathbb{P}(X_t = i)$, $p(t) = (p_1(t), p_2(t), \ldots)$. Таким образом, $p(t)$ — это вектор-строка. Как мы помним, вероятность состояния i в момент t определяется лишь состояниями в предыдущий момент времени $t-1$. Обозначим $q_{ij}(t) \overset{\triangle}{=} \mathbb{P}(X_t = j|X_{t-1} = i)$. Мы получаем матрицу (возможно бесконечную) $Q(t) = (q_{ij}(t))_{i,j\in\mathbb{N}}$.

Отметим, что правило условных вероятностей приводит нас к формуле:
$$p_j(t) = \mathbb{P}(X_t = j) = \sum_i \mathbb{P}(X_t = j|X_{t-1} = i)\mathbb{P}(X_{t-1} = i) = \sum_i q_{ij}(t)p_i(t).$$

Если записать это в матричной форме, то мы получим уравнение Колмогорова:
$$p(t) = p(t-1)Q(t).$$

Естественно, для того, чтобы решить уравнение Колмогорова, нам необходимо знание о начальном распределении состояний $p(0)$.

Отметим свойства матрицы переходных вероятностей $Q(t)$.

Определение 18. Будем говорить, что (возможно бесконечная) матрица $Q = (q_{ij})_{i,j=1,2,\ldots}$ является стохастической, если

- $q_{ij}(t) \geq 0$;
- $\sum_j q_{ij}(t) = 1$.

Легко видеть, что матрица переходных вероятностей стохастическая.

Предложение 18. *Если Q', Q'' — стохастические матрицы, то $Q = Q'Q''$ тоже стохастическая.*

Доказательство. Доказательство следует из определения стохастической матрицы и произведения матриц. □

В качестве следствия заметим, что если X_t — марковская цепь, то X_{nt} — тоже марковская цепь.

В дальнейшем мы будем рассматривать лишь стационарные (однородные по времени) марковские цепи. В этом случае матрица переходных вероятностей не зависит от момента времени t, т.е. $Q = Q(t)$. Интересным вопросом является вопрос о возможности построить стационарную марковскую цепь по заданному начальному распределению и матрице переходных вероятностей.

Предложение 19. *Пусть* $p^0 = (p^0_1, \dots, p^0_r)$ *— некоторый* r*-мерный вектор,* $Q = (q_{ij})_{i,j=1}^r$ *—* $r \times r$*-матрица. Тогда существует марковская цепь (с дискретным временем и состояниями* $1, \dots, r$*)* X_t *такая, что* p^0 *— начальное распределение, а* Q *— матрица переходных вероятностей.*

Доказательство. В качестве Ω выберем множество последовательностей $\bar{\omega} = (\omega_0, \omega_1, \omega_2, \dots, \omega_k, \dots)$, где $\omega_k \in \{1, \dots, r\}$, σ-алгебру измеримых множеств мы определим как минимальную σ-алгебру, которая содержит множества $A_{k_1,\dots,k_n;i_1,\dots,i_n} = \{\bar{\omega} = (\omega_0, \omega_1, \dots, \omega_k, \dots) : \omega_{k_l} = i_l\}$. (Здесь мы считаем, что $k_1 < \dots < k_n$). Далее, вероятность $\mathbb{P}$ определим на множествах $A_{0,1,\dots,n;i_1,\dots,i_n}$ по правилу

- $\mathbb{P}(A_{0,i}) = p^0_i$;
- $\mathbb{P}(A_{0,1,\dots,n;i_1,\dots,i_n}) = p^0_{i_1} Q_{i_0,i_1} Q_{i_1,i_2} \dots Q_{i_{n-1},i_n}$.

После этого, пользуясь известными правилами, распространим $\mathbb{P}$ на всю σ-алгебру. Наконец, положим $X_t(\bar{\omega}) \triangleq \omega_t$. Заметим, что по построению $X_\bullet$ — марковская цепь, и ее переходная матрица равна Q. □

Упражнение 3 (0.75 балла). Можно ли обобщить Предложение 10 на случай, когда матрица Q — бесконечная матрица?

10.1 Стационарное распределение

Определение 19. Стохастическая $r \times r$ матрица $Q = (q_{ij})_{i,j=1}^r$ называется эргодической, если все $q_{ij} > 0$.

Теорема 15. *Пусть матрица переходов* Q *— эргодична. Тогда существует вектор* $\pi = (\pi_1, \dots, \pi_r)$ *такой, что*

$$\pi = \pi Q,$$

и для любого начального вектора вероятностей состояний марковской цепи p^0 *последовательность векторов вероятностей* $p^t \triangleq \pi^0 Q^t$ *сходится к* π.

Доказательство. Нам достаточно доказать, что в некоторой метрике оператор $p \mapsto pQ$ сжимающий. Отсюда (по известной теореме Банаха о сжимающих отображениях) заключение теоремы последует. Напомним, что метрикой на множестве $\mathbb{X}$ называется отображение $d : \mathbb{X} \times \mathbb{X} \to [0, +\infty)$ такое, что

1. $d(x, y) = 0$ тогда и только тогда, когда $x = y$;
2. $d(x, y) = d(y, x)$;
3. $d(x, z) \le d(x, y) + d(y, z)$.

Заметим, что вероятности состояний лежат во множестве $S^r \triangleq \{(p_1, \ldots, p_r) : p_1, \ldots p_r \geq 0, p_1 + \ldots + p_r = 1\}$ (это множество является r-мерным симплексом). На S^r введем метрику d по правилу: если $p' = (p'_1, \ldots, p'_r)$, $p'' = (p''_1, \ldots, p''_r)$, то

$$d(p', p'') \triangleq \frac{1}{2}(|p'_1 - p''_1| + \ldots + |p'_r - p''_r|).$$

Найдем эквивалентную форму метрики d. Для этого обозначим через $a^+ \triangleq \max\{a, 0\}$, $a^- \triangleq \min\{a, 0\}$. Заметим, что $a = a^+ + a^-$, $|a| = a^+ - a^-$. Имеем, что

$$0 = \sum_i (p'_i - p''_i) = \sum_i (p'_i - p''_i)^+ + \sum_i (p'_i - p''_i)^-.$$

Также,

$$d(p', p'') = \frac{1}{2}\sum_i |p'_i - p''_i| = \frac{1}{2}\sum_i (p'_i - p''_i)^+ - \frac{1}{2}\sum_i (p'_i - p''_i)^- = \sum_i (p'_i - p''_i)^+.$$

Напомним, что $q_{ij} > 0$, поскольку Q $r \times r$-матрица, существует $\alpha > 0$ такое, что $q_{ij} \geq \alpha$ для всех i и j.

Пусть J — множество тех индексов, для которых $(p'Q - p''Q)_j > 0$. Имеем, что

$$d(p'Q, p''Q) = \sum_{j \in J} \sum_i q_{ij}(p'_i - p''_i) = \sum_i (p'_i - p''_i) \sum_{j \in J} q_{ij} \leq \sum_i (p'_i - p''_i)^+ \sum_{j \in J} q_{ij}.$$

Заметим, что J не может содержать все индексы, одновременно для каждого i $\sum_j q_{ij} = 1$. Следовательно,

$$\sum_{j \in J} q_{ij} \leq 1 - \alpha.$$

Напомним, что $\alpha = \min_{ij} q_{ij}$. Отсюда,

$$d(p'Q, p''Q) \leq (1 - \alpha) \sum_i (p'_i - p''_i)^+ = (1 - \alpha) d(p', p'').$$

Напомним, что теорема Банаха о сжимающих отображениях говорит, что если $A : \mathbb{X} \to \mathbb{X}$ таково, что для некоторого $\beta \in (0, 1)$ $d(Ax_1, Ax_2) \leq \beta d(x_1, x_2)$, то существует единственный элемент x_* такой, что $d(A^k x, x_*) \to 0$ при $k \to \infty$ (здесь A^k обозначает k-ю степень оператора k). В частности, $Ax_* = x_*$.

Мы доказали, что оператор $p \mapsto pQ$ — сжимающий в S^r. Следовательно, теорема доказана.

□

Замечание 4. Заметим, что если π — стационарное состояние, то $\pi_i > 0$.

Упражнение 4 (0.5 балла). Докажите утверждение замечания 4.

Замечание 5. Имеет место оценка

$$d(\pi, p_0 Q^n) \leq (1 - \alpha)^n.$$

10.2 Закон больших чисел для марковской цепи

Введем случайные величины: ν_i^n, равная числу моментов $t \in \{1, \ldots, n\}$ таких, что $X_t = i$, и ν_{ij}^n, равная числу моментов $t \in \{1, \ldots, n\}$ таких, что $X_{t-1} = i$, $X_t = j$.

Теорема 16. *Пусть переходная матрица марковской цепи Q эргодическая, и пусть π — стационарное распределение, тогда для всех $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\nu_i^n}{n} - \pi_i\right| \geq \varepsilon\right) = 0,$$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\nu_{ij}^n}{n} - \pi_i q_{ij}\right| \geq \varepsilon\right) = 0.$$

Доказательство. Введем вспомогательные случайные величины

$$\chi_i^t \triangleq \begin{cases} 1, & X_t = i, \\ 0, & X_t \neq i, \end{cases}$$

$$\chi_{ij}^t \triangleq \begin{cases} 1, & X_{t-1} = i, X_t = j, \\ 0, & \text{иначе.} \end{cases}$$

Имеем, что $\nu_i^n = \sum_{t=1}^n \chi_i^t$, $\nu_{ij}^n = \sum_{t=1}^n \chi_{ij}^t$.

Для любого начального распределения p

$$\mathbb{E}\chi_i^t = \sum_{l=1}^r p_l q_{li}^{(t)}, \quad \mathbb{E}\chi_{ij}^t = \sum_{l=1}^r p_l q_{li}^{(t-1)} q_{ij}.$$

Здесь, $q_{li}^{(t)}$ — вероятность того, что $X_t = i$ при условии $X_0 = l$.

Согласно замечанию 4 имеем, что $q_{mi}^{(t)}$ сходится к π_i с экспоненциальной скоростью. Отсюда,

$$\mathbb{E}\chi_i^t \to \pi_i, \quad \mathbb{E}\chi_{ij}^t \to \pi_i q_{ij}$$

с экспоненциальной скоростью. В силу определений ν_i^n и ν_{ij}^n получаем, что

$$\mathbb{E}\frac{\nu_i^n}{n} \to \pi_i, \quad \mathbb{E}\frac{\nu_{ij}^n}{n} \to \pi_i q_{ij}.$$

Теперь оценим скорость сходимости, рассмотрим множества событий

$$\left\{\omega : \left|\frac{\nu_i^n(\omega)}{n} - \pi_i\right| \geq \varepsilon\right\}, \quad \left\{\omega : \left|\frac{\nu_{ij}^n(\omega)}{n} - \pi_i q_{ij}\right| \geq \varepsilon\right\}.$$

При достаточно большом n имеем, что

$$\left\{\omega : \left|\frac{\nu_i^n(\omega)}{n} - \pi_i\right| \geq \varepsilon\right\} \subset \left\{\omega : \left|\frac{\nu_i^n(\omega)}{n} - \frac{\mathbb{E}\nu_i^n}{n}\right| \geq \frac{\varepsilon}{2}\right\},$$

$$\left\{\omega : \left|\frac{\nu_{ij}^n(\omega)}{n} - \pi_i q_{ij}\right| \geq \varepsilon\right\} \subset \left\{\omega : \left|\frac{\nu_{ij}^n(\omega)}{n} - \frac{\mathbb{E}\nu_{ij}^n}{n}\right| \geq \frac{\varepsilon}{2}\right\}.$$

Мы можем оценить вероятности с использованием неравенства Чебышева. Имеем, что

$$\mathbb{P}\left(\left|\frac{\nu_i^n}{n} - \pi_i\right| \geq \varepsilon\right) \leq \mathbb{P}\left(|\nu_i^n - \mathbb{E}\nu_i^n| \geq \frac{\varepsilon n}{2}\right) \leq \frac{4D\nu_i^n}{\varepsilon^2 n^2},$$

$$\mathbb{P}\left(\left|\frac{\nu_{ij}^n}{n} - \pi_i q_{ij}\right| \geq \varepsilon\right) \leq \mathbb{P}\left(\left|\nu_{ij}^n - \mathbb{E}\nu_{ij}^n\right| \geq \frac{\varepsilon n}{2}\right) \leq \frac{4D\nu_{ij}^n}{\varepsilon^2 n^2}.$$

Таким образом, нам остается оценить $D\nu_i^n$ и $D\nu_{ij}^n$. Прежде всего напомним, что $\nu_i^n = \sum_{t=1}^n \chi_i^t$. Далее обозначим $M_i^t \triangleq \mathbb{E}\chi_i^t = \sum_{l=1}^r p_l q_{li}^{(t)}$. Имеем,

$$D\nu_i^n = \mathbb{E}\left(\sum_{t=1}^n (\chi_i^t - M_i^t)\right)^2 = \sum_{t=1}^n \mathbb{E}(\chi_i^t - M_i^t)^2 + 2\sum_{t_1<t_2} \mathbb{E}(\chi_i^{t_1} - M_i^{t_1})(\chi_i^{t_2} - M_i^{t_2}).$$

Поскольку χ_i^t принимает значения из множества $\{0,1\}$, $\chi_i^t - M_i^t \in [-1,1]$, а $\mathbb{E}(\chi_i^t - M_i^t)^2 \leq 1$, то

$$\sum_{t=1}^n \mathbb{E}(\chi_i^t - M_i^t)^2 \leq n.$$

Далее имеем, что

$$\mathbb{E}(\chi_i^{t_1} - M_i^{t_1})(\chi_i^{t_2} - M_i^{t_2}) = \mathbb{E}\chi_i^{t_1}\chi_i^{t_2} - M_i^{t_1}M_i^{t_2} = \sum_{l=1}^r p_l q_{li}^{(t_1)} q_{ii}^{(t_2-t_1)} - M_i^{t_1}M_i^{t_2}.$$

Обозначим правую часть этого равенства через R_{t_1,t_2}.

Из замечания 4 следует, что $M_i^t = \pi_i + d_i^t$, $q_{li}^{(t)} = \pi_i + \beta_{li}^t$, где $|d_i^t|, |\beta_{li}^t| \leq c_1\lambda^t$, c_1 — некоторая константа, $\lambda \in (0,1)$.

Имеем, что

$$|R_{t_1,t_2}| = \left|\sum_{l=1}^r p_l q_{li}^{(t_1)} q_{ii}^{(t_2-t_1)} - M_i^{t_1}M_i^{t_2}\right| \leq c_2(\lambda^{t_1} + \lambda^{t_2} + \lambda^{t_2-t_1}).$$

Следовательно,

$$\sum_{t_1<t_2} R_{t_1,t_2} \leq c_3 n.$$

Это, и оценка на $\sum_{t=1}^n \mathbb{E}(\chi_i^t - M_i^t)^2$, влечет первое утверждение теоремы.

Второе утверждение может быть получено аналогично. □

Упражнение 5 (0.5 балла). Получите требуемый результат для ν_{ij}^n.

11 Марковские цепи с непрерывным временем

Пусть $t \in [0, +\infty)$ — непрерывное время. Предположим, что X_t принимает значения в $\{1, \ldots, r\}$.

Определение 20. Случайный процесс X_t называется марковской цепью с непрерывным временем (марковской очередью), если для всех $t_1 < t_2 < \ldots < t_n < \tau$ и всех $i_1, \ldots, i_n, j \in \{1, \ldots, r\}$

$$\mathbb{P}(X_\tau = j | X_{t_n} = i_n, X_{t_{n-1}} = i_{n-1}, \ldots, X_{t_1} = i_1) = \mathbb{P}(X_\tau = j | X_{t_n} = i_n).$$

Обозначим,

$$p_{ij}(t_1, t_2) = \mathbb{P}(X_{t_2} = j | X_{t_1} = i).$$

Также пусть $P(t_2, t_1) = p_{ij}(t_2, t_1)$. Имеем, что

$$\begin{aligned} p_{ij}(t_1, t_3) = \mathbb{P}(X_{t_3} = j | X_{t_1} = i) = \sum_{l=1}^{r} \mathbb{P}(X_{t_3} = j | X_{t_2} = l) \mathbb{P}(X_{t_2} = l | X_{t_1} = i) \\ = \sum_{l=1}^{r} p_{il}(t_1, t_2) p_{lj}(t_2, t_3). \end{aligned}$$

Это равенство можно записать в матричной форме:

$$P(t_1, t_3) = P(t_1, t_2) P(t_2, t_3). \tag{3.1}$$

В дальнейшем мы сосредоточим свое внимание на однородных марковских цепях.

Определение 21. Марковская цепь называется однородной, если $P(t_1, t_2)$ зависит лишь от разности $t_2 - t_1$; в этом случае можно переобозначить P, всюду вместо $P(t_1, t_2)$ мы будем писать $P(t_2 - t_1)$.

Заметим, что для однородных цепей равенство (3.1) принимает вид

$$P(\tau_1 + \tau_2) = P(\tau_1) P(\tau_2). \tag{3.2}$$

Оно говорит о том, что $P(\cdot)$ удовлетворяет полугрупповому свойству. В дальнейшем, отображение $t \mapsto P(t)$, удовлетворяющее свойству (3.2) и равенству $P(0) = I$, где I — единичная матрица, будем называть *полугруппой*.

Так же как и в случае дискретного времени, можно ввести понятие стационарного распределения.

Определение 22. Распределение вероятностей $\pi = (\pi_1, \ldots, \pi_r)$ называется стационарным, если $\pi P(t) = \pi$ для всех t.

Теорема 17. *Пусть для некоторого $t_* > 0$ $P(t_*)$ — эргодична, т.е. $p_{ij}(t_*) > 0$. Тогда существует стационарное распределение марковской цепи π и $|\pi_i - p_{ki}(t)|$ сходится к нулю экспоненциально быстро.*

Упражнение 6 (1 балл). Докажите теорему 17.

11.1 Матрица Колмогорова

Самое естественное действие, которое можно предпринять с функцией времени — это её продифференцировать. В силу полугруппового свойства логично дифференцировать в нуле. Заметим, что $P(0) = I$, где I — единичная матрица, $I = (I_{ij})_{i,j=1}^r$, где $I_{i,j} = 1$ при $i = j$ и нулю в противном случае.

Положим

$$q_{ij} = \lim_{t\downarrow 0} \frac{p_{ij}(t) - I_{ij}}{t}, \tag{3.3}$$

$Q = (q_{ij})_{i,j=1}^r$. Матрицу Q называют матрицей Колмогорова, или инфинитезимальной матрицей переходных вероятностей.

Предложение 20. *Если предел в (3.3) существует, то $P(t)$ дифференцируема и выполнены равенства*

$$\frac{dP(t)}{dt} = P(t)Q \text{ (прямое уравнение Колмогорова)},$$

$$\frac{dP(t)}{dt} = QP(t) \text{ (обратное уравнение Колмогорова)}.$$

Доказательство. Вначале докажем дифференцируемость $P(t)$ (в 0 рассматривается только дифференцируемость справа). Из полугруппового свойства заключаем, что при $h > 0$ $P(t+h) - P(t) = P(t)(P(h) - I)$. Отсюда

$$\lim_{h\downarrow 0} \frac{P(t+h) - P(t)}{h} = P(t) \lim_{h\downarrow 0} \frac{P(h) - I}{h} = P(t)Q \tag{3.4}$$

.

Таким образом, нами показано, что $P(t)$ дифференцируема справа. Докажем теперь, что $P(t)$ непрерывна и дифференцируема слева. Заметим, что для $h > 0$

$$P(t-h) - P(t) = P(t-h)(P(h) - I).$$

Поскольку элементы $P(t-h)$ ограничены, и элементы $P(h)-I$ сходятся к нулю, имеем, что $P(t-h)$ сходится к $P(t)$. Теперь найдем

$$\lim_{h\downarrow 0} \frac{P(t-h) - P(t)}{-h} = \lim_{h\downarrow 0} P(t-h) \lim_{h\downarrow 0} \frac{P(h) - I}{h} = P(t)Q.$$

Отсюда и из (3.4) получаем, что прямое уравнение Колмогорова выполнено.

Аналогично, используя полугрупповое свойство, получаем, что

$$\lim_{h\downarrow 0} \frac{P(t+h) - P(t)}{h} = \lim_{h\downarrow 0} \frac{P(h) - I}{h} \cdot P(t) = QP(t),$$

$$\lim_{h\downarrow 0} \frac{P(t-h) - P(t)}{-h} = \lim_{h\downarrow 0} \frac{P(h) - I}{h} \lim_{h\downarrow 0} P(t-h) = QP(t).$$

Таким образом, мы получили, что выполнено и обратное уравнение Колмогорова. □

Отметим свойства матрицы Колмогорова.

Предложение 21. *Пусть матрица Q определена формулой (3.3). Тогда*

1. $q_{ij} \geq 0$ *для* $i \neq j$;
2. $\sum_j q_{ij} = 0$.

Доказательство. Первое свойство следует из определения q_{ij} для $i \neq j$:

$$q_{ij} = \lim_{t \downarrow 0} \frac{p_{ij}(t)}{t}$$

и неравенства $p_{ij}(t) \geq 0$.

Второе свойство следует из тождества $p_{i1}(t) + \ldots + p_{ir}(t) = 1$ и равенства

$$q_{ij}(t) = \frac{d}{dt} p_{ij}(0).$$

$\square$

Отметим, что решение уравнения Колмогорова может быть записано в виде

$$P(t) = \exp(tQ), \tag{3.5}$$

где $\exp(A)$ — матричная экспонента:

$$\exp(A) = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \ldots + \frac{1}{n!}A^n + \ldots.$$

Отсюда, в частности следует, что π — стационарное распределение тогда и только тогда, когда

$$\pi Q = 0.$$

Наконец, зададимся вопросом, а определяют ли уравнения Колмогорова стационарную марковскую цепь? А именно, пусть $P(t) = \exp(tQ)$, p^0 — некоторое начальное распределение. Мы построим марковскую цепь с начальным распределением p^0 и матрицей переходных вероятностей Q. Для этого определим случайные величины ξ, τ_i^n и η_i^n, $i \in \{1, \ldots, r\}$, $n \in \mathbb{N}$ по следующим правилам.

1. ξ принимает значения из $\{1, \ldots, r\}$ в соответствии с распределением p^0.
2. τ_i^n — случайная величина на $[0, \infty)$, распределенная экспоненциально с параметром $\lambda_i = -q_{ii}$.
3. η_i^n принимает значения из множества $\{1, \ldots, r\} \setminus \{i\}$ с вероятностями $\mathbb{P}(\eta_i^n = j) = -q_{ij}/q_{ii}$.
4. Все случайные величины ξ, τ_i^n и η_i^n независимы.

Фактически ξ — определяет начальное состояние, τ_i^n — время от $(n-1)$-го до n-го прыжка при условии, что после $(n-1)$-го прыжка марковская цепь оказалась в состоянии i, η_i^n — состояние после n-го прыжка при условии того, что мы прыгаем из состояния i. Положим $\xi^0 \triangleq \xi$, $\theta^0 \triangleq 0$. Если ξ^{n-1} и θ^{n-1} уже построены, положим:

$$\theta^n \triangleq \theta^{n-1} + \tau^n_{\xi^{n-1}},$$
$$\xi^n = \eta^n_{\xi^{n-1}}.$$

Наконец, для $t \in [\theta^{n-1}, \theta^n)$ определим

$$X_t \triangleq \xi^{n-1}. \tag{3.6}$$

Предложение 22. *Случайный процесс $(X_t)_{t\geq 0}$, определенный (3.6), является однородной марковской цепью с начальным распределением p^0 и матрицей переходов $P(t) = \exp(tQ)$.*

Упражнение 7 (1 балл). Докажите предложение 22.

11.2 Пример. Модель системы массового обслуживания

Предположим, что у нас есть r устройств (серверов), каждый из которых может обработать не более одного запроса. Если все серверы загружены, запрос не обрабатывается (никогда), запросы поступают независимо. Вероятность поступления каждого запроса определяется экспоненциальным распределением с параметром λ, время обработки запроса также случайно и распределено экспоненциально с параметром μ. Серверы также независимы.

В качестве моделирующей марковской цепи выберем цепь, в которой X_t равно количеству занятых серверов, т.е. $X_t \in \{0, 1, \dots, r\}$. Заметим, что если все серверы свободны, то вероятность того, что придет запрос, распределена экспоненциально с параметром λ, если i серверов заняты, то вероятность того, что хоть один сервер освободится, распределена также экспоненциально с параметром $i\mu$.

Отсюда следует, что время, за которое система не сменит свое состояние при условии того, что она находится в состоянии i, распределено экспоненциально с параметром

$$\gamma(i) \triangleq \begin{cases} \lambda, & i = 0 \\ \lambda + i\mu, & i = 1, \dots, r-1 \\ r\mu, & i = r. \end{cases}$$

Используя эти соображения, можно посчитать матрицу Колмогорова. В этом случае это $(r+1) \times (r+1)$-матрица

$$Q = \begin{pmatrix} -\gamma(0) & \lambda & 0 & 0 & 0 & \dots & 0 & 0 \\ \mu & -\gamma(1) & \lambda & 0 & 0 & \dots & 0 & 0 \\ 0 & 2\mu & -\gamma(2) & \lambda & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & r\mu & -\gamma(r) \end{pmatrix}$$

Также в этом случае $P(t) = \exp(tQ)$ эргодична и

$$\pi_i = \frac{(\lambda/\mu)^i/i!}{\sum_{j=0}^r (\lambda/\mu)^j/j!}.$$

12 Мартингалы

12.1 Фильтрация и моменты остановки

Пусть задано вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$.

Если $t \in T$ (дискретное или непрерывное время), то логично рассмотреть события, произошедшие лишь до момента t. Эти события образуют набор σ-алгебр, который принято называть фильтрацией.

Определение 23. Набор σ-алгебр $\{\mathcal{F}_t\}_{t\in T}$ называют фильтрацией, если для всех $s \leq t$ $\mathcal{F}_s \subset \mathcal{F}_t$.

Определение 24. Пусть $\{\mathcal{F}_t\}_{t\in T}$ — фильтрация, X_t — случайный процесс. Будем говорить, что X_t согласовано с $\{\mathcal{F}_t\}_{t\in T}$, если для каждого t X_t измеримо относительно $\mathcal{F}_t$.

Говорят, что процесс $(X_t)_{t\in T}$ прогрессивно измерим, если для всех $t \in T$ $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-измеримы отображения $[0,t] \times \Omega \ni (s,\omega) \mapsto X_s(\omega)$.

Говорят, что $\{\mathcal{F}_t\}$ — естественная фильтрация для процесса X_t, если $\mathcal{F}_t$ — наименьшая σ-алгебра, относительно которой для всех $s \leq t$ измеримы случайные величины X_s.

Определение 25. Случайную величину τ, принимающую значения в T, называют моментом остановки (относительно фильтрации $\{\mathcal{F}_t\}_{t\in T}$), если для всех t событие $\{\tau \leq t\}$ лежит в $\mathcal{F}_t$.

Подумать: докажите, что все константы — моменты остановки.

Подумать: можно ли было определить момент остановки правилом "$\{\tau < t\}$ лежит в $\mathcal{F}_t$" ?

Пример 11. Пусть $T = \mathbb{N}$, Ω — множество функций $\omega : \mathbb{N} \to \{-1, 1\}$, $\mathcal{F}_n$ — наименьшая σ-алгебра, содержащая множества

$$\{\omega = (\omega_1, \omega_2, \ldots, \omega_n, \ldots) : \omega_1 = a_1, \ldots, \omega_n = a_n\}.$$

Тогда моментами остановки являются

$$\tau(\omega) \stackrel{\triangle}{=} \min\left\{n : \sum_{i=1}^n \omega_i = 3\right\}, \quad \sigma(\omega) = \min\{n : \omega_n = -1\}.$$

Интерпретация этих моментов следующая. Представим игру, в которой игрок выигрывает (проигрывает) рубль, если выпал орел (выпала решка). Игра до τ означает игру до выигрыша в три рубля, игра до σ означает игру до первого проигрыша.

Одновременно легко заметить, что функция

$$\sigma'(\omega) = \min\{n : \omega_{n+1} = -1\}$$

моментом остановки не является.

Напомним обозначения: $x \vee y = \max\{x, y\}$, $x \wedge y = \min\{x, y\}$.

Лемма 4. *Если σ и τ — моменты остановки, то $\sigma \wedge \tau$ тоже момент остановки.*

Доказательство. Нам необходимо доказать, что $\{\sigma \wedge \tau \le t\} \in \mathcal{F}_t$ для всех t. В самом деле имеем, что

$$\{\sigma \wedge \tau \le t\} = \{\sigma \le t\} \cup \{\tau \le t\} \in \mathcal{F}_t.$$

□

Подумать: будет ли верен аналогичный факт для $\sigma \vee \tau$?

Интересный пример момента остановки может быть получен для процессов, принимающих не более чем счетное число значений.

Предложение 23. *Пусть t — непрерывное время, $\{\mathcal{F}_t\}$ — фильтрация, X_t — случайный процесс, согласованный с $\{\mathcal{F}_t\}$. Предположим, что $X_t \in \mathbb{N}$. Пусть также для почти всех ω функция $t \mapsto X_t(\omega)$ имеет предел слева и непрерывна справа. Тогда для $s \in \mathbb{R}$, $i \in \mathbb{N}$*

$$\tau_i^s(\omega) \stackrel{\triangle}{=} \inf\{t \ge s : X_t(\omega) = i\}$$

(момент первого попадания в i) есть момент остановки.

Доказательство. Заметим, что множество

$$\{\tau_i^s < t\} = \bigcup_{u \in \{t\} \cup (\mathbb{Q} \cap (s,t))} \{X_u = i\} \in \mathcal{F}_t.$$

□

Подумать: будет ли верно доказательство для момента первого попадания в некоторое множество A без предположения дискретности ($X_t \in \mathbb{N}$)? А если это множество открыто, а если — замкнуто?

Легко видеть, что любой момент остановки можно представить в виде

$$\bar{\tau}_B(\omega) \stackrel{\triangle}{=} \inf\{t \,|\, \exists t \in S\, (t, \omega) \in B\}$$

подходящим выбором $B \subset S \times \Omega$.

Предложение 23'. [без д-ва] Пусть выполнены обычные условия на $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in T}, \mathbb{P})$, а некоторое множество $B \subset T \times \Omega$ прогрессивно измеримо (то есть для всех $t \in T$ $\mathcal{F}_t$-измеримы функции $\omega \to 1_B(t, \omega)$). Тогда $\bar{\tau}_B$ является моментом остановки.

Подумать: для любых прогрессивно измеримого процесса $(X_t)_{t\in S}$ и измеримого множества $A \in \mathcal{F}_s$, τ_A^s является моментом остановки в силу Предложения 23'.

Будем говорить, что функция $f : \mathbb{R} \to \mathbb{R}$ *cádlág* ("*continue á droite, limite á gauche*"), если в каждой точке она имеет предел слева и непрерывна справа.

Подумать: всякий имеющий лишь cádlág траектории процесс прогрессивно измерим.

Пусть τ — момент остановки. Введем σ-алгебру событий, произошедших до τ:

$$\mathcal{F}_\tau \stackrel{\triangle}{=} \{A \in \mathcal{F} : A \cap \{\tau \le t\} \in \mathcal{F}_t \text{ для всех } t\}.$$

Отметим свойства $\mathcal{F}_\tau$.

- $\mathcal{F}_\tau$ — σ-алгебра.
- τ измеримо относительно $\mathcal{F}_\tau$. В самом деле, достаточно показать, что для любого $c \in \mathbb{R}$ $\{\tau \le c\} \in \mathcal{F}_\tau$. Имеем, что для всех t

  $$\{\tau \le c\} \cap \{\tau \le t\} = \{\tau \le c \wedge t\} \in \mathcal{F}_t.$$

  Отсюда $\{\tau \le c\} \in \mathcal{F}_\tau$.
- Если σ, τ — моменты остановки и $\sigma \le \tau$ (почти наверное), то $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$. Действительно, для всех $t \in T$, $A \in \mathcal{F}_\sigma$ достаточно убедиться, что $A \cap \{\tau \le t\} = B \cap \{\tau \le t\} \in \mathcal{F}_\tau$ для $B = (A \cap \{\sigma \le t\}) \cap \{\sigma \wedge t \le \tau \wedge t\} \in \mathcal{F}_t$.

12.2 Мартингалы. Примеры и не только.

Прежде всего напомним, что такое условное математическое ожидание. Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство, $\mathcal{G} \subset \mathcal{F}$ — под-σ-алгебра (или σ-подалгебра), ξ — случайная величина, определенная на $(\Omega, \mathcal{F}, \mathbb{P})$ (т.е. для любого борелевского множества $B \subset \mathbb{R}$ $\xi^{-1}(B) \in \mathcal{F}$). Условное математическое ожидание ξ при заданной σ-подалгебре $\mathcal{G}$ — это случайная величина $\eta : \Omega \to \mathbb{R}$, измеримая относительно $\mathcal{G}$ (т.е. для любого борелевского множества $B \subset \mathbb{R}$ $\eta^{-1}(B) \in \mathcal{G}$), определяемая по правилу: для всех $A \in \mathcal{G}$

$$\mathbb{E}(\eta \mathbf{1}_A) = \mathbb{E}(\xi \mathbf{1}_A).$$

Напомним также свойства условных математических ожиданий.

1. Если $\mathcal{G}_2 \subset \mathcal{G}_1 \subset \mathcal{F}$ σ-алгебры, то

   $$\mathbb{E}(\xi|\mathcal{G}_2) = \mathbb{E}(\mathbb{E}(\xi|\mathcal{G}_1)|\mathcal{G}_2).$$

2. Если $\xi_1 \le \xi_2$, то $\mathbb{E}(\xi_1|\mathcal{G}) \le \mathbb{E}(\xi_2|\mathcal{G})$.
3. $\mathbb{E}(\mathbb{E}(\xi|\mathcal{G})) = \mathbb{E}\xi$.
4. Если $|\xi_n| \le \psi$, существует $\mathbb{E}\psi$ и $\xi_n \xrightarrow{\text{п.в.}} \xi$, то $\mathbb{E}(\xi_n|\mathcal{G}) \xrightarrow{\text{п.в.}} \mathbb{E}(\xi|\mathcal{G})$.

5. (неравенство Иенсена) Если $g : \mathbb{R} \to \mathbb{R}$ выпукла, $\mathbb{E}|\xi|$, $\mathbb{E}|g(\xi)| < \infty$, то почти наверное выполнено неравенство

$$g(\mathbb{E}(\xi|\mathcal{A})) \leq \mathbb{E}(g(\xi)|\mathcal{G}).$$

Определение 26. Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство, $\{\mathcal{F}_t\}_{t\in T}$ — фильтрация, говорят, что

- $(X_t, \{\mathcal{F}_t\}_{t\in T})$ — *мартингал*, если для $s \leq t$ $X_s = \mathbb{E}(X_t|\mathcal{F}_s)$;
- $(X_t, \{\mathcal{F}_t\}_{t\in T})$ — *субмартингал*, если для $s \leq t$ $X_s \leq \mathbb{E}(X_t|\mathcal{F}_s)$;
- $(X_t, \{\mathcal{F}_t\}_{t\in T})$ — *супермартингал*, если для $s \leq t$ $X_s \geq \mathbb{E}(X_t|\mathcal{F}_s)$.

Пример 12. В случае дискретного времени, для процесса X_n иногда берут по умолчанию его естественную фильтрацию $(\sigma(X_0, X_1, \ldots, X_n))_{n\in\mathbb{N}}$. В этом случае X_n — мартингал тогда и только тогда, когда

$$X_{n-1} = \mathbb{E}(X_n|X_0, X_0, \ldots, X_{n-1}) \qquad \forall n \in \mathbb{N}.$$

Пример 13. Пусть Y — $\mathcal{F}$-интегрируемая случайная величина, тогда $X_s \stackrel{\triangle}{=} \mathbb{E}(Y|\mathcal{F}_s)$ является мартингалом в силу $\mathbb{E}(Y|\mathcal{F}_s) = \mathbb{E}(\mathbb{E}(Y|\mathcal{F}_t)|\mathcal{F}_s)$ при $s \leq t$.

Пример 14. Пусть у Вас есть две возможные плотности f и g, и выборка $x_1, \ldots, x_k, \ldots$. Функция правдоподобия

$$X_n = \frac{f(x_1)}{g(x_1)} \cdot \ldots \cdot \frac{f(x_n)}{g(x_n)}$$

является мартингалом.

Пример 15. Пусть вероятностное пространство и фильтрация определены как в примере 11. Тогда в случае равновероятности 1 и -1 (справедливая игра) $X_n = \omega_1 + \ldots + \omega_n$ (выигрыш игрока за n раундов) — мартингал. В то же время,

$$Y_n = \frac{\omega_1 + \ldots + \omega_n}{n} \text{ (средний выигрыш игрока за } n \text{ раундов)}$$

не является ни субмартингалом, ни супермартингалом (треуется знакоопределенность выигрыша).

Для того, чтобы это проверить, заметим, что если $a_1, \ldots, a_n, \ldots$ — какие-то коэффициенты, то для всех $m \leq n$

$$\mathbb{E}(a_1\omega_1 + \ldots + a_n\omega_n|\mathcal{F}_m) = a_1\omega_1 + \ldots + a_m\omega_m.$$

Подумать: в случае если шансы 1 и -1 не равны, но не зависят от времени, когда именно X_n — субмартингал, супермартингал(?); подберите константу R так, чтобы R^{X_n} стал мартингалом.

Предложение 24. *Если X_t, Y_t — мартингалы относительно одной и той же фильтрации, то $aX_t + bY_t$ тоже мартингал для любых $a, b \in \mathbb{R}$.*

Подумать: верно ли это предложение для супермартингалов.

Подумать: как соотносится это предложение с лишь субмартингальностью среднего выигрыша в предыдущем примере.

Предложение 25. *Если* $(X_t, \{\mathcal{F}_t\}_{t\in T})$ *— мартингал,* g *— выпуклая функция,* $\mathbb{E}|g(X_t)| < \infty$ *для всех* t*, то* $(g(X_t), \{\mathcal{F}_t\}_{t\in T})$ *— субмартингал.*

[С-но; 0,5 баллов] Докажите, что супермартингал X_t является мартингалом тогда и только тогда, когда $\mathbb{E}X_t$ не зависит от времени.

[С-но; 0,5 баллов] Докажите, что если X_t, Y_t — супермартингалы относительно одной и той же фильтрации, то $X_t \wedge Y_t$ — тоже супермартингал.

Всюду далее в основном пойдет речь про случай $T = \mathbb{N}$, для непрерывного случая будут приводиться в лучшем случае формулировки аналогичных результатов.

Теорема 18 (разложение Дуба)**.** *Если* $(X_n, \{\mathcal{F}_n\}_{n\in\mathbb{N}})$ *— субмартингал, то существуют два случайных процесса* M_n *и* A_n *такие, что*

1. $X_n = M_n + A_n$, $n \geq 1$;
2. $(M_n, \{\mathcal{F}_n\}_{n\in\mathbb{N}})$ *— мартингал;*
3. $A_1 = 0$, A_n *—* $\mathcal{F}_{n-1}$ *измеримо,* $n \geq 2$;
4. A_n *не убывает, т.е.* $A_n(\omega) \leq A_{n+1}(\omega)$ *почти наверное.*

Если есть другое такое разложение $\overline{M}_n$, $\overline{A}_n$*, то* $\overline{M}_n = M_n$, $\overline{A}_n = A_n$.

Доказательство. Прежде всего докажем единственность. Пусть M_n, A_n уже построены. Для $n \geq 2$ имеем, что

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

Взяв условное математическое ожидание, получаем, что

$$\mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} = A_n - A_{n-1}.$$

Вместе с условием $A_1 = 0$ это показывает, что X_n определяют A_n однозначно. Аналогично, $M_n = X_n - A_n$, что также говорит об однозначности определения M_n.

Теперь докажем существование. Положим $M_1 \triangleq X_1$, $A_1 \triangleq 0$, $A_n \triangleq \mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} + A_{n-1}$, $M_n \triangleq X_n - A_n$.

Заметим, что свойства 1, 3 выполняются по определению. Свойство 4 следует из того, что X_n — субмартингал. Докажем теперь свойство 2, а именно то, что $(M_n, \{\mathcal{F}_n\}_{n\in\mathbb{N}})$ — мартингал. Имеем, что

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

$\square$

В непрерывном времени также есть похожий результат:

Теорема 19 (разложение Дуба–Мейера; без д-ва)**.** *Если* $(X_t, \{\mathcal{F}_t\}_{t\geq 0})$ *— субмартингал с фильтрацией, удовлетворяющей обычным условиям, для каждой ограниченной константы* $a > 0$*, для множества всех ограниченных этой константой моментов остановки* τ *выполнено*

$$\lim_{\lambda\to\infty} \sup_{\tau} \mathbb{E}(|\xi_\tau| \mathbf{1}_{\{|\xi_\tau|>\lambda\}}) = 0.$$

Тогда найдутся случайные процессы M_t *и* A_t *с непрерывными траекториями такие, что*

1. $X_t = M_t + A_t$*,* $t \geq 0$*;*

2. $(M_t, \{\mathcal{F}_t\}_{t\geq 0})$ *— мартингал;*

3. $A_0 = 0$*,* A_t *согласован с* $\mathcal{F}_t$*;*

4. A_t *не убывает, т.е.* $A_s(\omega) \leq A_t(\omega)$ *почти наверное для всех* $s \leq t$*.*

Если есть другая пара процессов $\overline{M}_t$*,* $\overline{A}_t$*, то* $\overline{M}_t = M_t$*,* $\overline{A}_n = A_t$ *почти всюду.*

В некотором смысле такой же результат, но с другого конца (и дело не в замене субмартингал/супермартингал).

Теорема 20 (разложение Рисса; без д-ва)**.** *Пусть* $(X_n, \{\mathcal{F}_n\})$ *— супермартингал. Следующие условия эквивалентны:*

1. *для некоторого субмартингала* $(W_n, \{\mathcal{F}_n\})$ *почти всюду выполнено* $W_n \leq X_n$*;*

2. *для некоторого мартингала* $(M_n, \{\mathcal{F}_n\})$ *и неотрицательного супермартингала* $(A_n, \{\mathcal{F}_n\})$ *выполнено* $X_n = M_n + A_n$ *и* $\mathbb{E}A_n \to 0$ *при* $n \to \infty$*.*

Если есть другая пара процессов $\overline{M}_t$*,* $\overline{A}_t$*, то* $\overline{M}_n = M_n$*,* $\overline{A}_n = A_n$ *п.в.*

В непрерывном случае в этой теореме надо потребовать от фильтрации обычных условий и ограничиться процессами, имеющими лишь cádlág траектории.

13 Теорема о произвольном выборе

Если X_n согласовано с фильтрацией $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, τ — момент остановки, то случайная величина $X_{\tau(\omega)}(\omega)$ измерима относительно σ-алгебры $\mathcal{F}_\tau$. В самом деле, для любого борелевского множества $B \subset \mathbb{R}$ имеем, что $\{X_\tau \in B\} \in \mathcal{F}$ и

$$\{X_\tau \in B\} \cap \{\tau \leq n\} = \bigcup_{l\leq n} \{\omega : X_l \in B, \tau = l\} \in \mathcal{F}_n.$$

Теорема 21 (о произвольном выборе)**.** *Пусть τ, σ — некоторые моменты остановки, $\sigma \le \tau \le k$, где k — некоторое число.*

- *Если $(X_n, \{\mathcal{F}_n\}_{n\in\mathbb{N}})$ — субмартингал, то $X_\sigma \le \mathbb{E}(X_\tau|\mathcal{F}_\sigma)$;*
- *если $(X_n, \{\mathcal{F}_n\}_{n\in\mathbb{N}})$ — мартингал, то $X_\sigma = \mathbb{E}(X_\tau|\mathcal{F}_\sigma)$;*
- *если $(X_n, \{\mathcal{F}_n\}_{n\in\mathbb{N}})$ — супермартингал, то $X_\sigma \ge \mathbb{E}(X_\tau|\mathcal{F}_\sigma)$.*

В непрерывном случае теорема также имеет место, при этом вместо ограниченности можно потребовать равномерно ограниченность.

Для того, чтобы осознать теорему о произвольном выборе, рассмотрим ситуацию игры в орлянку. В этом случае игрок не может изменить своего выигрыша (в среднем), если он будет начинать игру в момент σ, а заканчивать в момент τ, если игра идет случайно, а решения о входе/выходе принимаются в зависимости от текущей ситуации.

Доказательство. Мы докажем теорему лишь для случая, когда $(X_n, \{\mathcal{F}_n\}_{n\in\mathbb{N}})$ — субмартингал. Пусть $A \in \mathcal{F}_\sigma$, $1 \le m \le n$. Положим

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

$$B_{m,n} \triangleq A_m \cap \{\tau > n\}, \quad C_{m,n} \triangleq A_m \cap \{\tau \ge n\}.$$

Имеем, что $B_{m,n} \in \mathcal{F}_n$ т.к. $\{\tau > n\} = \Omega \setminus \{\tau \le n\} \in \mathcal{F}_n$. По определению субмартингала имеем

$$\mathbb{E}(X_n \mathbf{1}_{B_{m,n}}) \le \mathbb{E}(X_{n+1} \mathbf{1}_{B_{m,n}}).$$

Поскольку $C_{m,n} = A_{m,n} \cup B_{m,n}$ имеем, что

$$\mathbb{E}(X_n \mathbf{1}_{C_{m,n}}) \le \mathbb{E}(X_n \mathbf{1}_{A_{m,n}}) + \mathbb{E}(X_{n+1} \mathbf{1}_{B_{m,n}}).$$

А так как $B_{m,n} = C_{m,n+1}$, верно неравенство

$$\mathbb{E}(X_n \mathbf{1}_{C_{m,n}}) - \mathbb{E}(X_{n+1} \mathbf{1}_{C_{m,n+1}}) \le \mathbb{E}(X_n \mathbf{1}_{A_{m,n}}).$$

Просуммировав последнее неравенство по n от m до k и воспользовавшись равенством $A_m = C_{m,m}$, получаем, что

$$\mathbb{E}(X_m \mathbf{1}_{A_m}) \le \mathbb{E}(X_\tau \mathbf{1}_{A_m}).$$

Далее последнее неравенство суммируем по m от 1 до k. Получаем, что

$$\mathbb{E}(X_\sigma \mathbf{1}_A) \le \mathbb{E}(X_\tau \mathbf{1}_A).$$

Это и есть заключение теоремы (для первого случая). Остальные случае доказываются аналогично. □

13.1 Пример. Задача о разорении страховой компании

Пусть u — начальный капитал страховой компании, пусть страховые поступления идут постоянно, со скоростью $c > 0$, а в случайные моменты времени $T_i \in \mathbb{R}_{\geq 0}$ (T_i возрастают) происходят выплаты страховки ξ_i. Тогда $(X_t)_{t\geq 0}$ — эволюция капитала страховой компании — случайный процесс:

$$X_t = u + ct - S_t, \qquad S_t = \sum_{i\in\mathbb{N}} \xi_i 1_{T_i \leq t}.$$

Как обычно, $\mathcal{F}_s$ — все события, что произошли к моменту s.

Нас интересуют $T \triangleq \inf\{t \geq 0 \,|\, X_t \leq 0\} \cup \{+\infty\}$ и $\mathbb{P}(T < +\infty)$

Следуя модели Крамера– Лундеберга предположим, что

1. $T_i - T_{i-1}$ — независимые случайные величины, распределенные по закону $Exp(\lambda)$;
2. ξ_i — независимые неотрицательные случайные величины с общей функцией распределения F_ξ и $\mathbb{E}\xi_i = \mu$;
3. последовательности $(T_i)_{i\in\mathbb{N}}$, $(\xi_i)_{i\in\mathbb{N}}$ независимы.

В рамках такой модели, например через характеристические функции, легко показывается, что для всякого $t \geq 0$, $k \in \mathbb{N} \cup \{0\}$

$$\mathbb{P}(T_k < t, T_{k+1} > t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

В силу

$$\mathbb{E}(X_t - X_0) = ct - \sum_{i\in\mathbb{N}} \mathbb{E}(\xi_i 1_{T_i\leq t}) = ct - \mu \sum_{i\in\mathbb{N}} \mathbb{P}(T_i \leq t) = t(c - \lambda\mu),$$

далее считаем, что $c > \lambda\mu$ (средняя прибыль компании положительна).

Введем $h(z) \triangleq \int_0^\infty (e^{zx} - 1)dF_\xi(x)$ и $g(z) = \lambda h(z) - cz$ для всех неотрицательных z.

Теорема 22. *В модели Крамера–Лундберга при $\lambda\mu < c$ вероятность разорения не превосходит e^{-Ru}, где R — единственный корень уравнения $\lambda g(R) = 0$.*

Доказательство. Для $h(z) \triangleq \int_0^\infty (e^{zx} - 1)dF_\xi(x)$ и $g(z) = \lambda h(z) - cz$ при $z \geq 0$, в силу $\mathbb{E}e^{r\xi_1} = 1 + h(r)$, имеем

$$\begin{aligned}
\mathbb{E}(e^{-r(X_t - X_0)}) &= e^{-rct}\mathbb{E}e^{r\sum_{T_i\leq t}\xi_i} \\
&= e^{-rct}\sum_{k\in\mathbb{N}} \mathbb{E}e^{r\sum_{i=1}^k \xi_i}\mathbb{P}(T_k < t, T_{k+1} > t) \\
&= e^{-rct}\sum_{k\in\mathbb{N}} \mathbb{E}(1 + h(r))^k \frac{e^{-\lambda t}(\lambda t)^k}{k!} \\
&= e^{-rct}e^{-\lambda t}e^{\lambda t(1+h(r))} = e^{tg(r)}.
\end{aligned}$$

Ну тогда $\mathbb{E}(e^{-r(X_t-X_s)}) = \mathbb{E}(e^{-r(X_t-X_s)}|\mathcal{F}_s) = e^{(t-s)g(r)}$.

Поскольку при $s<t$

$$\begin{aligned}\mathbb{E}(e^{-rX_t-tg(r)}|\mathcal{F}_s) &= e^{-tg(r)}\mathbb{E}(e^{-r(X_t-X_s)}e^{rX_s}|\mathcal{F}_s)\\ &= e^{-tg(r)}e^{(t-s)g(r)}e^{rX_s} = e^{-rX_t-sg(r)},\end{aligned}$$

то $e^{-rX_t-tg(r)}$ — мартингал, следовательно для момента остановки $\tau=\min(t,T)$

$$\begin{aligned}e^{-ru} &= \mathbb{E}e^{-rX_t-tg(r)} = \mathbb{E}e^{-rX_\tau-\tau g(r)} \geq \mathbb{E}(e^{-rX_\tau-\tau g(r)}|T\leq t)\mathbb{P}(T\leq t)\\ &= \mathbb{E}(e^{-rX_T-Tg(r)}|T\leq t)\mathbb{P}(T\leq t)\\ &\geq \mathbb{E}(e^{-Tg(r)}|T\leq t)\mathbb{P}(T\leq t)\\ &\geq \min_{s\in[0,T]}\mathbb{E}(e^{-sg(r)}|s\leq t)\mathbb{P}(T\leq t)\\ \mathbb{P}(T\leq t) &\leq \max_{s\in[0,T]} e^{sg(r)-ru}.\end{aligned}$$

Если взять $r=R$ такое, что $g(R)=0$, то $\mathbb{P}(T\leq t)\leq e^{-Ru}$. □

14 Сходимость мартингалов

Начнем с неравенств.

Пусть X_n — случайный процесс. Рассмотрим событие

$$A_{\lambda,n} \triangleq \left\{\omega : \max_{i=\overline{1,n}} X_i(\omega)\geq\lambda\right\}.$$

Неравенство $\lambda\mathbb{P}\{|X|\geq\lambda\}\leq\mathbb{E}|X|$. (Чебышева оно же в других книжках — неравенство Маркова) говорит о том, что

$$\lambda\mathbb{P}\{X_n\geq\lambda\}\leq\mathbb{E}\max\{X_n,0\}.$$

Однако оно ничего не говорит о том, что было на предыдущих шагах. Если же X_n — субмартингал, то мы можем оценить $\mathbb{P}(A_{\lambda_n})$.

Теорема 23 (Неравенство Дуба)**.** *Если $(X_n,\mathcal{F}_n)$ — субмартингал, то для любых n и λ*

$$\lambda\mathbb{P}(A_{\lambda,n})\leq\mathbb{E}(X_n\mathbf{1}_{A_{\lambda,n}})\leq\mathbb{E}\max\{X_n,0\}.$$

Подумать: как доказать то же неравенство в непрерывном случае для имеющих лишь cádlág траектории мартингалов.

Доказательство. Определим момент остановки σ по правилу

$$\sigma(\omega)\triangleq\begin{cases} i, & X_i\geq\lambda,\\ n, & \max_{i=1,\dots,n}X_i<\lambda.\end{cases}$$

Также $\tau\equiv n$. Имеем, что $A_{\lambda,n}\in\mathcal{F}_\sigma$, так как $A_{\lambda,n}\cap\{\sigma\leq m\}=\{\max_{i=1,\dots,m}X_i\geq\lambda\}\in\mathcal{F}_m$.

Поскольку на $A_{\lambda,n}$ $X_\sigma\geq\lambda$, имеем из теоремы 21, что

$$\lambda\mathbb{P}(A_{\lambda,n})\leq\mathbb{E}(X_\sigma\mathbf{1}_{A_{\lambda,n}})\leq\mathbb{E}(X_n\mathbf{1}_{A_{\lambda,n}})\leq\mathbb{E}\max\{X_n,0\}.$$

□

Результаты, полученные при анализе субмартингалов, можно применить к исследованию сумм независимых случайных величин.

Следствие 11 (Неравенство Колмогорова)**.** *Пусть* $\xi_1, \xi_2, \ldots$ *— независимые случайные величины,* $m_i = \mathbb{E}\xi_i$*,* $\mathcal{D}_i = D\xi_i$*, тогда*

$$\mathbb{P}\left(\max_{i=1,\ldots,n} |\xi_1 + \ldots + \xi_i - m_1 - \ldots - m_i| \geq \lambda\right) \leq \frac{1}{\lambda^2}\sum_{i=1}^{n} \mathcal{D}_i.$$

Доказательство. Для того, чтобы доказать неравенство Колмогорова, введем случайный процесс $X_n \triangleq \xi_1 + \ldots + \xi_n$. Заметим, что X_n — мартингал относительно естественной фильтрации, тогда как X_n^2 — субмартингал относительно естественной фильтрации. Остается применить неравенство Дуба. □

Теперь займемся уже сходимостью собственно мартингалов.

Определение 27. Будем говорить, что $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ — *замкнутый справа мартингал*, если для некоторой суммируемой случайной величины X_∞ выполнено при всех $n \in \mathbb{N}$

$$X_n = \mathbb{E}(X_\infty | \mathcal{F}_n).$$

Теорема 24 (Теорема Дуба о замкнутых мартингалах)**.** *. Пусть* $(X_n, \mathcal{F}_n)$ *— замкнутый справа мартингал, тогда*

$$X_n \xrightarrow{п.в.} X_\infty, \qquad X_n \xrightarrow{L^1} X_\infty.$$

Подумать: а что такое тогда $\mathcal{F}_\infty$?
Подумать: чем такая теорема может быть полезна для $X_n = Y_{1-1/n}$,или в случае $\Omega = \mathbb{R}$ для

$$X_n(\omega) = \frac{f\big(2^{-n}\lceil\omega 2^n\rceil + 2^{-n}\big) - f\big(2^{-n}\lceil\omega 2^n\rceil\big)}{2^{-n}}.$$

Прежде чем доказывать теорему, докажем факт попроще.

Предложение 26. *Если мартингал замыкаем, то последовательность случайных величин* X_n *равномерно интегрируема.*

Замечание 6. Обратный факт тоже верен, но доказывается сложно.

Доказательство. Нам нужно лишь доказать, что

$$\lim_{\lambda\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}1_{\{|X_n|>\lambda\}}|X_n| = 0.$$

По неравенству Йенсена $|X_n| = |\mathbb{E}(X_\infty \,|\, \mathcal{F}_n)| \leq \mathbb{E}(|X_\infty| \,|\, \mathcal{F}_n)$, в частности $\mathbb{E}1_{\{|X_n|>\lambda\}}|X_n| \leq \mathbb{E}1_{\{|X_n|>\lambda\}}|X_\infty|$, и осталось заметить

$$\lim_{\lambda\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}\{|X_n| > \lambda\} \leq \lim_{\lambda\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}|X_n|/\lambda \leq \lim_{\lambda\to\infty} \mathbb{E}|X_\infty|/\lambda = 0.$$

□

Вернемся к доказательству теоремы.

Доказательство. Зафиксируем $\varepsilon > 0$. Рассмотрим $\mathcal{F}_\infty = \sigma(\cup_{i=1}^\infty \mathcal{F}_i)$. Для каждого $A \in \mathcal{F}_\infty$ найдутся $N \in \mathbb{N}$ и $B \in \mathcal{F}_N$, для которых $\mathbb{E}(|1_A - 1_B|) < \varepsilon$.

Всевозможные индикаторы элементов $\mathcal{F}_n$ и их конечные суммы всюду плотны в множестве всех суммируемых $\mathcal{F}_n$-измеримых функций. Значит индикаторы элементов $\cup_{i=1}^\infty \mathcal{F}_i$ и их конечные суммы всюду плотны в множестве всех суммируемых $\mathcal{F}_\infty$-измеримых функций. Тогда объединение конечных сумм индикаторов элементов $\mathcal{F}_n$ (по всем n) всюду плотно в множестве всех суммируемых $\mathcal{F}_\infty$-измеримых функций.

В частности, $\mathbb{E}(|X_\infty - Y_\infty|) < \varepsilon^2$ для некоторой $\mathcal{F}_N$-измеримой функции Y_∞.

Примем $Y_n = \mathbb{E}(Y_\infty|\mathcal{F}_n)$, тогда по неравенству Чебышева $\mathbb{P}(|X_\infty - Y_\infty| > \varepsilon) \leq \varepsilon$.

Теперь $Y_n = \mathbb{E}(Y_\infty|\mathcal{F}_n)$ — мартингал, $X_n - Y_n = \mathbb{E}(X_\infty - Y_\infty|\mathcal{F}_n)$ — мартингал, $(X_n - Y_n)^2 = \mathbb{E}((X_\infty - Y_\infty)^2|\mathcal{F}_n)$ — субмартингал.
По неравенству Дуба, для $n > N$,

$$\begin{aligned}\mathbb{P}(\limsup_{n\in\mathbb{N}} |X_n - Y_\infty| > \varepsilon) \leq \mathbb{P}(\sup_{n\in\mathbb{N}} |X_n - Y_n| > \varepsilon) \quad &\leq \quad \sup_{n\in\mathbb{N}} \mathbb{E}|X_n - Y_n|/\varepsilon \\ &\leq \quad \mathbb{E}|X_\infty - Y_\infty|/\varepsilon \leq \varepsilon.\end{aligned}$$

Таким образом, $\mathbb{P}(\limsup_{n\in\mathbb{N}} |X_n - X_\infty| > 2\varepsilon) \leq 2\varepsilon$, и сходимость почти всюду показана. Для сходимости в L_1 достаточно заметить сходимость матожиданий в силу $\mathbb{E}|X_n - Y_n| \leq \mathbb{E}|X_\infty - Y_\infty| \leq \varepsilon^2$. □

Подумать: а зачем нам предыдущее предложение, вроде обошлись без него?

Теорема 25 (без д-ва)**.** *Пусть $(X_n, \mathcal{F}_n)$ — супермартингал, причем $\mathbb{E}|X_n|$ ограничены, тогда*

$$X_n \xrightarrow{п.в.} Y$$

для некоторой суммируемой случайной величины Y. Если X_n равномерно интегрируемы, то имеет место сходимость и в L_1.

Замечание 7. В принципе можно обойтись ограниченностью $\mathbb{E}\max(X_n, 0)$, аккуратно поправляя все формулировки о суммируемости. Тогда достаточно потребовать неположительность $\mathbb{E}X_n$.

Замечание 8. В непрерывном случае все вышесказанное работает, если предположить для фильтрации обычные условия и ограничиться имеющими лишь cádlág траектории мартингалами.

Для мартингалов есть и аналог центральной предельной теоремы, с почти таким же условием, и почти тем же доказательством, те же характеристические функции, подробности смотрите в двухтомнике Ширяева.

15 Задача об оптимальной остановке

15.1 Общий случай. Мартингальный подход

Задача об оптимальной остановке формулируется в случае дискретного времени следующим образом. Пусть набор моментов времени конечен $n = 0, 1, \dots, N$, фиксировано вероятностное пространство с фильтрацией $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\in N}, \mathbb{P})$, и для каждого момента времени задана случайная величина, описывающая качество остановки $f_n : \omega \to [0, +\infty)$. Естественно предполагать, что f_n $\mathcal{F}_n$-измеримо. Требуется найти

$$V_0 = \sup_{\tau\in\mathfrak{M}_0^N} \mathbb{E} f_\tau.$$

Здесь и далее $\mathfrak{M}_n^N$ обозначает семейство моментов остановки, принимающих значения во множестве $\{n, \dots, N\}$.

Подумать: если f_n — супермартингал, тогда оптимален $\tau = 0$; если f_n — субмартингал, тогда оптимален $\tau = N$; если f_n — мартингал, тогда оптимален любой момент остановки.

Метод решения очень напоминает метод динамического программирования и предполагает попятную процедуру. Для этого мы определим набор функций $\{v_n\}$ по правилу:

$$v_N \overset{\triangle}{=} f_N, \quad v_n \overset{\triangle}{=} \max\{f_n, \mathbb{E}(v_{n+1}|\mathcal{F}_n)\}.$$

Положим

$$\tau_n \overset{\triangle}{=} \min\{k \in \overline{n, N} \,|\, v_k = f_k\}.$$

Теорема 26. *Выполнено*

1. *моменты остановки τ_n оптимальны в классе $\mathfrak{M}_n$:*

$$\mathbb{E} f_{\tau_n} = V_n \overset{\triangle}{=} \sup_{\tau\in\mathfrak{M}_n} \mathbb{E} f_\tau;$$

2. *"стохастические цены" совпадают с v_n, т.е.*

$$\underset{\tau\in\mathfrak{M}_n}{\text{ess-sup}} \mathbb{E}(f_\tau|\mathcal{F}_n) = v_n.$$

Подумать: свяжите числа V_n и случайные величины v_n; можно ли утверждать, что $V_0 = v_0$ и $V_N = \mathbb{E} f_N$?

Прежде, чем мы перейдем к доказательству теоремы, необходимо объяснить, что такое операция ess-sup и почему мы не использовали sup. Операция ess-sup обозначает существенный супремум. Необходимость в ее использовании обусловлена неизмеримостью обычного супремума. А именно. Пусть $\mathfrak{A}$ — некоторое множество индексов, и для каждого $\alpha \in \mathfrak{A}$ существует случайная величина ξ_α. Вопрос: является ли функция $\omega \mapsto \xi(\omega) \overset{\triangle}{=} \sup_{\alpha\in\mathfrak{A}} \xi_\alpha(\omega)$ случайной величиной? Оказывается, что в случае несчетности $\mathfrak{A}$ не всегда. Действительно, условие измеримости ξ означает, что для всех $c \in \mathbb{R}$

$$\left\{\omega : \sup_{\alpha\in\mathfrak{A}} \xi_\alpha\right\} = \bigcap_{\alpha\in\mathfrak{A}} \{\omega : \xi_\alpha(\omega) \le c\}$$

лежит в $\mathcal{F}$. Если $\mathfrak{A}$ не более чем счетно, то это гарантируется свойствами σ-алгебры. А если нет, то может оказаться, что это множество не лежит в $\mathcal{F}$ (приведите пример!).

Определение 28. Будем говорить, что расширенная (т.е. принимающая значения в $\mathbb{R} \cup \{+\infty\}$) случайная величина ξ есть существенный супремум ξ_α и писать

$$\xi = \operatorname*{ess\text{-}sup}_{\alpha\in\mathfrak{A}} \xi_\alpha,$$

если

1. для всех $\alpha \in \mathcal{A}$ $\xi \ge \xi_\alpha$ $\mathbb{P}$-п.н.
2. если расширенная случайная величина η такова, что для всех $\alpha \in \mathcal{A}$ $\eta \ge \xi_\alpha$ $\mathbb{P}$-п.н., то
$$\eta \ge \xi \ \ \mathbb{P}\text{-п.н.}$$

Предложение 27. $\xi = \operatorname{ess\text{-}sup}_{\alpha\in\mathfrak{A}} \xi_\alpha$ *существует. Более того, найдется счетное множество* $\mathfrak{A}_0 \subset \mathfrak{A}$ *такое, что*

$$\xi(\omega) = \sup_{\alpha\in\mathfrak{A}_0} \xi_\alpha(\omega).$$

Доказательство. Предположим вначале, что ξ_α равномерно ограничены константой c, т.е. $|\xi_\alpha| \le c$. Пусть A — конечное множество индексов $\alpha \in \mathfrak{A}$. Положим

$$S(A) \triangleq \mathbb{E} \max_{\alpha\in A} \xi_\alpha.$$

Пусть S есть супремум значений $S(A)$ по всем конечным множествам $A \subset \mathfrak{A}$.

Обозначим через A_n конечное множество A такое, что

$$\mathbb{E} \max_{\alpha\in A_n} \xi_\alpha \ge S - \frac{1}{n}.$$

Пусть

$$\mathfrak{A}_0 \triangleq \bigcup_{n=1}^{\infty} A_n.$$

Положим

$$\xi(\omega) \triangleq \sup_{\alpha\in\mathfrak{A}_0} \xi_\alpha(\omega).$$

По построению ξ — случайная величина и $\xi = \operatorname{ess\text{-}sup}_{\alpha\in\mathfrak{A}} \xi_\alpha$.

В общем случае надо рассмотреть $\tilde{\xi}_\alpha(\omega) \triangleq \operatorname{arctg}(\xi_\alpha(\omega))$. Заметим, что $\tilde{\xi}_\alpha$ ограничены $\pi/2$. Тогда мы можем построить $\tilde{\xi} = \operatorname{ess\text{-}sup}_{\alpha\in\mathfrak{A}} \tilde{\xi}_\alpha$. Остается получить ξ по формуле $\xi(\omega) = \operatorname{tg}(\tilde{\xi}(\omega))$. $\square$

Доказательство теоремы 21. Пусть N зафиксировано. Мы будем использовать обратную индукцию по n.

Если $n = N$, то $v_N = f_N$ и все доказано. Пусть теперь теорема доказана для $n = N, N-1, \dots, k$. Докажем ее для $n = k-1$. Пусть $\tau \in \mathfrak{W}_{k-1}^N$ и $A \in \mathcal{F}_{k-1}$. Положим $\bar{\tau} \stackrel{\triangle}{=} \max\{\tau, k\}$. Заметим, что $\bar{\tau} \in \mathfrak{W}_k^N$. Также отметим, что событие $\{\tau \geq k\}$ лежит в $\mathcal{F}_{k-1}$.

Имеем, что

$$\begin{aligned}
\mathbb{E}(\mathbf{1}_A f_\tau) &= \mathbb{E}[\mathbf{1}_{A\cap\{\tau=k-1\}} f_\tau] + \mathbb{E}[\mathbf{1}_{A\cap\{\tau\geq k\}} f_\tau] \\
&= \mathbb{E}[\mathbf{1}_{A\cap\{\tau=k-1\}} f_\tau] + \mathbb{E}[\mathbf{1}_{A\cap\{\tau\geq k\}} \mathbb{E}(f_\tau|\mathcal{F}_{k-1})] \\
&= \mathbb{E}[\mathbf{1}_{A\cap\{\tau=k-1\}} f_\tau] + \mathbb{E}[\mathbf{1}_{A\cap\{\tau\geq k\}} \mathbb{E}(\mathbb{E}(f_{\bar{\tau}}|\mathcal{F}_k)|\mathcal{F}_{k-1})] \\
&\leq \mathbb{E}[\mathbf{1}_{A\cap\{\tau=k-1\}} f_\tau] + \mathbb{E}[\mathbf{1}_{A\cap\{\tau\geq k\}} \mathbb{E}(v_k|\mathcal{F}_{k-1})] \\
&\leq \mathbb{E}(\mathbf{1}_A v_{k-1}).
\end{aligned} \tag{3.7}$$

Это означает, что для всех $\tau \in \mathfrak{W}_n^N$

$$\mathbb{E}(f_\tau|\mathcal{F}_{k-1}) \leq v_{k-1}. \tag{3.8}$$

Теперь покажем, что

$$\mathbb{E}(f_{\tau_{k-1}}|\mathcal{F}_{k-1}) = v_{k-1} \quad \mathbb{P}\text{-п.н.}$$

Заметим, что на множестве $\{\tau_{k-1} \geq k\}$ и по предположению индукции $\tau_{k-1} = \tau_k$ и $\mathbb{E}(f_{\tau_k}|\mathcal{F}_k) = v_k$ $\mathbb{P}$-п.н. Аналогично (3.7) получаем:

$$\begin{aligned}
\mathbb{E}(\mathbf{1}_A f_{\tau_{k-1}}) &= \mathbb{E}[\mathbf{1}_{A\cap\{\tau_{k-1}=k-1\}} f_{k-1}] + \mathbb{E}[\mathbf{1}_{A\cap\{\tau_{k-1}\geq k\}} f_{\tau_{k-1}}] \\
&= \mathbb{E}[\mathbf{1}_{A\cap\{\tau_{k-1}=k-1\}} f_{k-1}] + \mathbb{E}[\mathbf{1}_{A\cap\{\tau_{k-1}\geq k\}} \mathbb{E}(f_{\tau_k}|\mathcal{F}_{k-1})] \\
&= \mathbb{E}[\mathbf{1}_{A\cap\{\tau_{k-1}=k-1\}} f_{k-1}] + \mathbb{E}[\mathbf{1}_{A\cap\{\tau_{k-1}\geq k\}} \mathbb{E}(v_k|\mathcal{F}_{k-1})] \\
&= \mathbb{E}(\mathbf{1}_A v_{k-1}).
\end{aligned}$$

Последнее равенство выполнено в силу определения Поскольку $v_{k-1} = \max\{f_{k-1}, \mathbb{E}(v_k|\mathcal{F}_{k-1})\}$ и свойств $v_{k-1} = f_{k-1}$ при $\tau_{k-1} = k-1$ и $v_{k-1} = \mathbb{E}(v_k|\mathcal{F}_{k-1})$ в случае $\tau_{k-1} \geq k$. Тем самым нами доказано, что

$$\mathbb{E}(f_{\tau_{k-1}}|\mathcal{F}_{k-1}) = v_{k-1} \quad \mathbb{P}\text{-п.н.}$$

Вместе с (3.8) это и дает заключение теоремы. □

При этом фактически показано чуть больше, не только, что v — мажоранта для f, то есть $v_n \geq f_n$; или, что v_n — супермартингал, то есть что :

$$v_n \geq \mathbb{E}(v_{n+1}|\mathcal{F}_n);$$

показано, что v — наименьшая супермартингальная мажоранта для f (наименьшая последовательность, обладающая предыдущими двумя свойствами), тогда она может быть найдена как наименьшее решение вариационного неравенства:

$$\gamma_n \geq \max\{f_n, \mathbb{E}(\gamma_{n+1}|\mathcal{F}_n)\}, \quad \gamma_n = f_n.$$

15.2 Случай марковской цепи

Применим эту теорему для стационарной марковской цепи с дискретным временем.

Пусть Ω состоит из всевозможных последовательностей $\omega = (x_0, x_1, \ldots, x_n, \ldots)$, где x_n — произвольные элементы некоторого измеримого пространства $(E, \mathcal{E})$,
$X_n(\omega)$ — n-я координата у ω, а $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. Тогда $(X_n)_{n\in\mathbb{N}}$ — стационарная марковская цепь, если задана переходная вероятность $P(dy; X_n = x) = P(dy; X_n)$.

Подумать: хотя для хороших $(E, \mathcal{E})$ разницы нет, переходная вероятность требует чуть больше, чем условная вероятность (при любом фиксированном $x \in E$, отображение $\mathcal{E} \ni \Gamma \mapsto P(\Gamma; X_n = x)$ — вероятность на $(E, \mathcal{E})$, при любом фиксированном $\Gamma \in \mathcal{E}$ отображение $E \ni x \mapsto P(\Gamma; X_n = x)$ — $\mathcal{E}$-измеримая функция).

Теперь для $\mathcal{E}$-измеримой функции $g : E \to \mathbb{R}$ можно рассмотреть при каждом $x \in E$ свое матожидание:

$$Tg(x) = \mathbb{E}_x g(X_1) = \int_E g(y) P(dy; x = X_1);$$

для простоты будем считать, что все наши функции суммируемы для всех $x \in E$.

Пусть $\mathfrak{W}^n$ — моменты остановки, принимающие значения из $\{0, \ldots, n\}$, а качество остановки задается функцией $g : E \to \mathbb{R}$, в частности оно не зависит от времени. Требуется найти оптимальное значение

$$s_n(x) \overset{\triangle}{=} \sup_{\tau\in\mathfrak{W}^n} \mathbb{E}_x g(X_\tau) = \operatorname*{ess\text{-}sup}_{\tau\in\mathfrak{W}^n} \mathbb{E}(g(X_\tau) | X_0 = x),$$

и тот момент остановки, на котором это значение достигается.

Имея ввиду $f_k = g(X_k), v_k^N = s_{N-k}$, напрямую из теоремы имеем для момента остановки

$$\tau^n \overset{\triangle}{=} \min\{k \in \overline{0, n} \,|\, s_{n-k}(X_k(\omega)) = g(X_k(\omega))\}$$

Следствие 12. *1. моменты остановки τ^n оптимальны в $\mathfrak{W}^n$: $\mathbb{E}_x g(X_{\tau^n}) = s_n(x)$;*

2. “цены” s_n могут быть найдены по формуле

$$s_n(x) = \max\{g(x), \mathbb{E}_x s_{n-1}(x)\} = \max\{g(x), Ts_{n-1}(x)\};$$

3. для $Q(x) \overset{\triangle}{=} \max\{g(x), \mathbb{E}_x g(X)\} = \max\{g(x), Tg(x)\}$ имеем

$$s_n(x) = Q(s_{n-1}(x)) = Q^{(n)}(g(x)).$$

Удивительно, но если не убрать n, как ограничение сверху на момент остановки, но при этом по-прежнему требовать его конечность, то найти оптимальные цены будет даже проще.

Пусть $\mathfrak{W}^\infty$ — моменты остановки, принимающие значения из $\mathbb{N} \cup \{0\}$.

Следствие 13. *Для* $s(x) \triangleq \sup\limits_{\tau \in \mathfrak{W}^{\infty}} \mathbb{E}_x g(X_\tau) = \operatorname*{ess\text{-}sup}\limits_{\tau \in \mathfrak{W}^{\infty}} \mathbb{E}(g(X_\tau)|X_0 = x)$, *имеет место*

1. $$s(x) = \lim_{n\to\infty} s_n(x) = \sup_{n\to\infty} Q^n g(x);$$

2. *выполнено уравнение Вальда–Беллмана* $s(x) = \max\{g(x), Ts(x)\}$*;*

3. $s(X_n)$ *— наименьший супермартингал среди не меньших* g*;*

4. $\tau_0^{\infty} \triangleq \min\{k \in \overline{n, N} \,|\, s_\infty(X_k(\omega)) = g(X_k(\omega))\}$ *оптимален в* $\mathfrak{W}^{\infty}$ *при конечном* E*;*

5. *момент остановки* $\tau_\varepsilon^{\infty} \triangleq \min\{k \in \overline{n, N} \,|\, s_\infty(X_k(\omega)) \le g(X_k(\omega) + \varepsilon)\}$ ε-*оптимален в классе* $\mathfrak{W}^{\infty}$*, то есть* $\mathbb{E}_x g(X_{\tau_\varepsilon^{\infty}}) + \varepsilon \ge s_\infty(x)$.

15.3 Пример. Задача о разборчивой невесте

В качестве применения последнего следствия решим задачу о разборчивой невесте (другие варианты названия: о выборе наилучшего объекта, о выборе секретаря).

Имеется *a priori* известное число N кандидатов *a priori* неизвестного качества; предполагается, что приходят на смотрины кандидаты в случайном порядке, не зависящем от свойств каждого кандидата. Нужно найти наилучшего из них, хотя бы одного кандидата выбрать необходимо. Качество каждого кандидата измеряется точно, но, поскольку кандидаты уходят сразу после замера, решение брать/не брать надо принять до прихода следующего.

Итак, нужно подобрать марковскую цепь так, чтобы для каждого состояния мы знали вероятность каждого другого состояния быть следующим. При продумывании такой цепи можно попытаться ответить на такие вопросы. Какая информация наблюдаема? Когда приходит новая информация? Какую, пусть и наблюдаемую, информацию можно заведомо не рассматривать, и напротив, какой наблюдаемой информации достаточно для принятия решений? В момент прихода новой информации какую часть старой информации можно забыть? Что назвать переходом, из каких состояний в какие?

Рассуждая, заметим для начала, что любой кандидат, не ставший в момент прихода "пока лучшим" , нам неинтересен: он заведомо не может быть выбран, а его приход хоть и уточняет информацию о вероятности быть самым лучшим последнего на этот момент "пока лучшего" , эта информация уже бесполезна, "пока лучший" уже ушел. С другой стороны, поскольку все порядки среди претендентов равноценны, каждый "пока лучший" может оказаться "самым лучшим" с вероятностью, зависящей лишь от числа ещё непроверенных претендентов (то есть от номера его прихода). Следовательно номер "пока лучшего" кандидата — та информация, которую следует помнить до прихода следующего "пока лучшего"; свойство марковости при этом уже выполнено!

Итак, в качестве базовой информации возьмем *какой по счету кандидат сейчас имеет титул "пока лучший"* . Мы получили состояния $1, 2, \ldots, N$. Если мы не взяли очередного "пока лучшего" , то можно считать (переход!), что нам сразу предъявили номер в очереди следующего "пока лучшего" (а значит, скольким он уже не проиграл), или же сказали, что таких нет, переход в состояние 0^*. По той причине, что фактически мы принимаем решение в момент прихода очередного "пока лучшего" , а значит, и знаем его номер-время, наша марковская цепь автоматически стационарна. Осталось найти матрицу переходов.

Итак, берем марковскую цепь, в которой состояний $N+1$, от 0 до N; где состояние k (от 1 до N) — «k-й был "пока лучшим"», состояние 0^* означает, что самый лучший уже ушел.

Теперь $p_{0^*,0^*} = 1$, ($p_{0^*,0^*} = 1$ означает, что состояние 0^* поглощающее, символ $*$ — издавна устоявшийся маркер для таких состояний), $p_{i,0^*} = i/N$ — это вероятность того, что оказавшийся не хуже уже i претендентов, будет не хуже N претендентов. Аналогично, вероятность того, что оказавшийся уже не хуже i претендентов, будет не хуже $j-1$ (соответственно j) претендентов равна $\frac{i}{j-1}$ (соответственно $\frac{i}{j}$). Теперь, вероятность того, что оказавшийся не хуже уже i претендентов, оказался хуже лишь j-го претендента равна $\frac{i}{j-1} - \frac{i}{j} = \frac{i}{j(j-1)}$. Таким образом, мы показали $p_{i,j} = \frac{i}{j(j-1)}$ при $0 < i < j$. Легко видеть, что $p_{i,j} = 0$ во всех остальных случаях.

Как фактически показано выше, вероятность быть наилучшим у пришедшего i-м и ставшего "пока лучшим" равна i/N. Примем $g(i) = i/N$. Тогда

$$\sup_\tau \mathbb{P}(\text{"выбран наилучший''}) = \sup_\tau \frac{\mathbb{E}X_\tau}{N} = \sup_\tau \mathbb{E}g(X_\tau).$$

Теперь формализация закончена, и можно применить теорию; при этом, поскольку τ ничем неограниченно, удобно воспользоваться последним следствием. Итак, нам достаточно решить уравнение Вальда-Беллмана:

$$s(x) = \max(g(x), Tv(x)) = \max\Big(\frac{x}{N}, \sum_{k=x+1}^{N} \frac{x}{k(k-1)} s(k)\Big),$$

или

$$Ns(x)/x = \max\Big(1, \sum_{k=x+1}^{N} \frac{Ns(k)/k}{k-1}\Big), \quad s(N) = 1 \qquad \forall x \in \{1, \ldots, N-1\},$$

или

$$\bar{s}(x) = \max\Big(1, \sum_{k=x+1}^{N} \frac{\bar{s}(k)}{k-1}\Big), \quad \bar{s}(N) = 1 \qquad \forall x \in \{1, \ldots, N-1\}.$$

Поскольку $\bar{s}$ не возрастает, то τ_0^∞ — последний момент, когда $\bar{s} = 1$:

$$\tau_0^\infty \triangleq \max\Big\{k \,\Big|\, \frac{1}{X_k - 1} + \cdots + \frac{1}{N-1} \leq 1\Big\}.$$

При этом вероятность удачи, выбора действительно лучшего, равна

$$P_{best} = \frac{\tau_0^\infty}{N},$$

1288 и при больших n стремится (простейший матан!) к $e^{-1} \approx 0,368\ldots.$