Теория вероятностей, V семестр

Осень 2016, лектор: Виноградов Александр Станиславович

Автор: Всеволод Степанов

Собрано: 18 января 2017 г. 21:37

Оглавление

Глава 1

Теория вероятностей

1.1. Введение, аксиоматика

Принятые обозначения: Ω – множество исходов
$\omega \in \Omega$ – исход (обозначаем маленькой буквой)
$A \subset \Omega$ – событие (обозначаем большими буквами)
$\mathcal{F}$ – множество всех событий
Часто мы будем сокращать и вместо $A \cap B$ будем писать просто AB.

Замечание 1.1.1. Не все множества исходов являются событиями, например, плохо, если нельзя измерить его объем. Может получиться парадокс Банаха-Тарского, когда мы берем, разрезаем один шар на несколько кусков и получаем два точно таких же шара.

$P : \mathcal{F} \to [0, 1]$ – вероятность событий.

Def 1.1.1. Ω – произвольное множество. $\mathcal{F}$ замкнуто относительно $\cap, \cup, \setminus$. Тогда $\mathcal{F}$ – алгебра событий.

Def 1.1.2. $\mathcal{F}$ – σ-алгебра, если $A_k \in \mathcal{F} \Rightarrow \bigcup\limits_{k=1}^{\infty}, \bigcap\limits_{k=1}^{\infty} \in \mathcal{F}$.

Упражнение: привести пример алгебры, не являющейся σ-алгеброй.

Пример 1.1.1. $\{\emptyset, \Omega\}$ – тривиальная σ-алгебра.

Пересечение σ-алгебр тоже σ-алгебра, а вот объединение – нет (например, в их объединении нет множеств, которые частично лежат в одном, а частично в другом, хотя они должны быть из-за замкнутости).

Def 1.1.3. σ-оболочка $\mathcal{A}$ — это пересечение всех σ-алгебр, содержащих $\mathcal{A}$ (тоже σ-алгебра):

$$\sigma(\mathcal{A}) = \bigcap_{\mathcal{F}_x \supseteq \mathcal{A}} \mathcal{F}_x$$

Множество σ-алгебр является решеткой: наибольшей нижней границей A, B будет $A \cap B$, а наименьшей верхней будет $\sigma(A \cap B)$

Def 1.1.4. Тройка $(\Omega, \mathcal{F}, P)$ называется вероятностным пространством (или экспериментом), если $\mathcal{F}$ – σ-алгебра и у P выполняются две аксиомы вероятности:

1. $AB = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$ (аддитивность)

2. $P(\Omega) = 1$ (нормировка)

Так же есть более сильная аксиома 1^* для случая счетных или бесконечных множеств: $A_kA_j = \emptyset \Rightarrow P(\bigcup\limits_k A_k) = \sum\limits_k P(A_k)$ – σ-аддитивность.

Замечание 1.1.2. Функция P, удовлетворяющая аксиомам $1^*, 2$ называется σ-аддитивной.

Свойства:

1. $P(\bar{A}) = 1 - P(A)$. В частности, $P(\emptyset) = 0$
2. Монотонность: $A \subset B \Rightarrow P(A) \leq P(B)$

   ▶

$$B = A \cup B\bar{A}$$
$$P(B) = P(A) + P(B\bar{A}) \geq P(A)$$

◀

3. $P(A \cup B) = P(A) + P(B) - p(AB)$.

   Общий случай: $P(\bigcup\limits_k^n A_k) = \sum P(A_i) - \sum\limits_{i<j} P(A_iA_j) + \sum\limits_{i<j<k} P(A_iA_jA_k) - ...$ – просто формула включений-исключений.

   ▶ Это у нас когда-то там было, когда говорили про множества. ◀

   Следствие 1.1.0.1. $P(\bigcup\limits_1^n A_k) \leq \sum P(A_k)$

4. $P(\lim A_n) = \lim P(A_n)$ (непрерывность P).

   Под пределом в данном случае имеется следующее:

   Если $A_i \subset A_{i+1}$, то $\lim A_n = \bigcup\limits^{\infty} A_k$.

   Если $A_{i+1} \subset A_i$, то $\lim A_n = \bigcap\limits^{\infty} A_k$.

   Доказательство будет следовать из

   Теорема 1.1.1. *P удовлетворяет аксиомам 1 и 2. Тогда равносильны:*

   (a) *Аксиома 1^**
   (b) *Непрерывность P*
   (c) *Непрерывность P в нуле: $B_k \searrow \emptyset \Rightarrow P(B_k) \to 0$*

   ▶ $1 \Rightarrow 2$ Пусть $A_i \subset A_{i+1}$ (иначе перейдем к дополнениям). Рассмотрим следующее объединение: $A_n \cup A_{n+1}\bar{A}_n \cup A_{n+2}\bar{A}_{n+1} \cup \dots = \lim(A_n) =: A$. Это объединение несовместных событий. Тогда $P(A) = P(A_n) + \sum\limits_{k=n}^{\infty} P(A_{k+1}\bar{A}_k)$. Устремим n к бесконечности, второе слагаемое стремится к 0 как хвост сходящегося ряда

   $2 \Rightarrow 3$ – очевидно

   $3 \Rightarrow 1$ Рассмотрим события A_i, $A_iA_j = \emptyset$. Возьмем $B_n = \bigcup\limits_{k=n}^{\infty} A_k$, $B_1 = A = \bigcup\limits^{\infty} A_k$

   B_n убывают, у них есть предел B. Хотим доказать, что $B = \emptyset$. Пусть это не так, тогда в B есть какой-то элемент $b \in B$, тогда $b \in B_1$, значит, $b \in A_l$ для какого-то l. Тогда $b \notin A_k, k \neq l \Rightarrow b \notin \bigcup\limits_{l+1}^{\infty} A_k = B_{l+1} \Rightarrow b \notin \bigcap B_k = B$, противоречие.

   Тогда $P(A) = \sum\limits^{k} P(A_i) + P(B_{k+1})$ – применили 1-ю аксиому. Устремим k к бесконечности, тогда $P(B_{k+1}) \to 0$ по непрерывности в нуле, получим σ-аддитивность.

◀

5. $P(\bigcup\limits^{\infty}) \le \sum\limits^{\infty} P(A_k)$

▶ Знаем для конечного случая, перейдем к пределу.
Получим $\lim P(\bigcup\limits_{i=0}^{n} A_i) \le \lim \sum\limits_{i=0}^{n} A_i = \sum\limits^{\infty} A_i$.
По 4 свойству, левая часть равна $P(\lim \bigcup\limits_{i=0}^{n} A_i) = P(\bigcup\limits^{\infty} A_i)$, что и требовалось. ◀

1.2. Условная вероятность

𝔇𝔢𝔣 1.2.1. $P(A|B) = P_B(A) = \frac{P(A\cap B)}{P(B)}$ – условная вероятность, вероятность события A при условии события B.

Теорема 1.2.1. $P_B(A)$ – вероятность (то есть, выполнены обе аксиомы)

▶ 1. Рассмотрим произвольные несовместные события A_i, проверим аддитивность:
$P_B(\cup A_i) \overset{\text{Def}}{\equiv} \frac{P(B\cap\bigcup A_i)}{P(B)} = \frac{\cup BA_i}{P(B)} = \frac{\sum P(BA_i)}{P(B)} = \sum P_B(A_i)$
2. Нормировка: $P_B(\Omega) = \frac{P(B\Omega)}{P(B)} = 1$

◀

𝔇𝔢𝔣 1.2.2. A, B два события. Если $P(AB) = P(A)P(B)$, то они называются независимыми.

𝔇𝔢𝔣 1.2.3 (Независимость в совокупности)**.** Множество событий $\{A_i\}$ называется независимым, если для любого конечного его подмножества $\{A_{i_j}\}(j \in [1;k])$ выполняется:

$$P(\bigcap_{j=1}^{k} A_{i_j}) = \prod_{j=1}^{k} P(A_{i_j})$$

Замечание 1.2.1. Это определение является более сильным, чем просто попарная независимость любых двух событий из множества.

Пример 1.2.1. Рассмотрим тетраэдр, покрасим три его стороны соответственно в красный, синий и зеленый цвета, а последнюю покрасим во все цвета сразу. Тогда вероятность того, что сторона покрашена и в красный и в синий равна $P(RB) = \frac{1}{4} = P(R)P(B)$, поэтому события “сторона красная” и “сторона синяя” независимы. Но события “сторона красная”, “сторона синяя” и “сторона зеленая” не являются независимыми, так как $P(RBG) = \frac{1}{4} \ne P(R)P(B)P(G) = \frac{1}{8}$. Тем не менее, они являются попарно независимыми.

Замечание 1.2.2. По умолчанию, когда говорится про независимость множества событий, имеется в виду сильная независимость.

Формула полной вероятности: пусть $\{H_k\}$ – конечное или счетное разбиение Ω. Тогда

$$P(A) = P(A\Omega) = P(A \cap \bigcup_k H_k) = P(\bigcup AH_k) = \sum_k P(AH_k) = \sum_k P(H_k)P(A|H_k)$$

Формула Байеса:

$$P(H_k|A) = \frac{P(H_kA)}{P(A)} = \frac{P(H_kA)}{\sum\limits_i P(H_i)P(A|H_i)} = \frac{P(H_k)P(A|H_k)}{\sum\limits_i P(H_i)P(A|H_i)}$$

Лемма 1.2.1 (Бореля, Кантелли)**.** Есть бесконечное счетное множество событий $\{A_k\}_{k=1}^{\infty}$. Событие A — «при исходе выполнилось бесконечно много событий среди A_k». Тогда:

1. $\sum P(A_k) < \infty \Rightarrow P(A) = 0$
2. $\sum P(A_k) = \infty$ и A_k независимы $\Rightarrow P(A) = 1$

▶ Построим событие A явно: исход $w \in A$ тогда и только тогда, когда он лежит в бесконечном количестве A_k (то есть начиная с любого места в некотором A_k встретится w):

$$A = \overline{\lim} A_k := \bigcap_{n=1}^{\infty} \underbrace{\bigcup_{k=n}^{\infty} A_k}_{B_n}$$

Тут B_n — это те исходы, которые встречаются в каком-нибудь A_k при $k \geqslant n$.

Заметим, что $B_n \searrow$ по включению: каждая точка либо есть в конечном числе A_k, тогда с какого-то момента ее в B_n не будет (а до этого — будет), либо точка есть в бесконечном числе A_k и тогда она есть во всех B_n.

Теперь доказываем пункты леммы:

1. Так как B_n есть объединение A_k с некоторого места, то:

$$P(B_n) \leq \sum_{k=n}^{\infty} P(A_k)$$

А часть справа стремится к нулю, так как является хвостом сходящегося ряда из условия. При этом $P(A) = P\left(\lim\limits_{n\to\infty} B_n\right) = \lim\limits_{n\to\infty} P(B_n)$ (последнее равенство по непрерывности вероятности). То есть $P(A) = 0$, что и требовалось.

2. Перейдем к дополнениям: $\bar{A} = \bigcup\limits_{n=1}^{\infty} \bigcap\limits_{k=n}^{\infty} \bar{A}_k$

$P(\bigcap\limits_{n}^{N} \bar{A}_k) = \prod\limits_{n}^{N} P(\bar{A}_k)$, так как пересечение конечное, события независимы. Устремим N к бесконечности, получим:

$$P(\bigcap_{n}^{\infty}) = \prod_{n}^{\infty} P(\bar{A}_k)$$

$$\ln P(\bigcup_{n}^{\infty} \bar{A}_k) = \sum_{n}^{\infty} \ln(1 - P(A_k)) <$$

Заметим, что $\ln(1-x) < -x$ при $x \in (0;1)$, тогда

$$< -\sum_{n}^{\infty} P(A_k) = -\infty \Rightarrow P(\bigcap_{n}^{\infty} \bar{A}_k) = 0$$

$$P(A) = 1 - P(\bar{A}) = 1 - P(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \bar{A}_k) \geq 1 - \sum_{n=1}^{\infty} P(\bigcap_{k=n}^{\infty} \bar{A}_k) = 1 - 0 = 1$$

◀

1.3. Последовательность независимых событий

Пусть у нас есть последовательность $(\Omega_n, \mathcal{F}_n, P_n)$ – последовательность независимых экспериментов. Хотим взять и построить вероятностное пространство для этой последовательности.

Понятно, что $\Omega = \bigtimes_n \Omega_n$. В качестве события A можно рассматривать $A = A_1 \times A_2 \times A_3 \ldots$. Не хотим возиться с бесконечными произведениями, хотим работать только с конечными, так гораздо проще. Тогда рассмотрим следующее произведение: $A = A_1 \times A_2 \times \ldots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \ldots$. Оно означает, что с какого-то момента мы считаем все события достоверными и нам не важно, что конкретно там просходило.

$\mathfrak{Def}$ 1.3.1. Такое A называется цилиндром. $A_1 \times \ldots \times A_n$ – его n-мерное основание.

Замечание 1.3.1. Почему цилиндр? Цилиндр это когда мы берем фигуру меньшей размерности, начинаем ее двигать в пространстве, очерчивая некоторую фигуру. Так и здесь: есть основание, по остальным координатам условий нет, берем и просто двигаем вдоль них основание.

Цилиндры не образуют σ-алгебру, поэтому нельзя сказать, что $\mathcal{F}$ состоит из цилиндров. Пересечение цилиндров это, конечно, цилиндр, а объединение – не факт. Но можно сказать, что $\mathcal{F} = \sigma(\mathcal{E})$, где $\mathcal{E}$ – набор всех цилиндров.

Для цилиндра A логично определить $P(A) = \prod\limits_{k=1}^{n} P_k(A_k)$. Для остальных событий есть следующая теорема.

$\mathfrak{Def}$ 1.3.2. $\mathcal{E}$ называется полукольцом множеств, если это множество, замкнутое относительно операции пересечения, при этом, для любых его двух элементов A, B должно выполняться $A \setminus B = \bigcup\limits_{j=1}^{m} C_j, C_j \in \mathcal{E}$

Теорема 1.3.1 (Каратеодори)**.** Пусть множество наборов $\mathcal{E}$ – полукольцо, и на нем задана σ-аддитивная функция P (то есть удовлетворяет аксиомам $1^*, 2$).

Тогда существует единственное ее σ-аддитивное продолжение на $\sigma(\mathcal{E})$.

▶ Без доказательства. ◀

Согласно этой теореме, вероятность, определенную на цилиндрах, можно единственным образом продолжить на все $\mathcal{F}$.

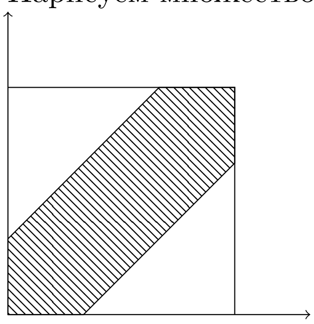
Замечание 1.3.2. Аналогия с линейной алгеброй: достаточно задать линейное отображение на базисе, на остальное пространство продолжается однозначно.

Пример 1.3.1. Задача о встречах. Два человека обговорили часовой интервал, в течение которого они хотят встретиться. Они приходят независимо друг от друга в какой-то случайный момент времени, каждый ждет 20 минут и потом уходит. Надо посчитать вероятность встречи.

$\Omega_k = [0; 1], k = 1..2, \mathcal{F}$ – все измеримые подмножества отрезка. Вероятность попасть – длина.

Смотрим на прямое произведение двух отрезков, это квадрат. Пусть точка (x, y) – моменты времени, когда пришли первый и второй человек. Тогда они встретились, если $|x - y| \leq \frac{1}{3}$.

Нарисуем множество точек, которые нам дают встречу:



Интуитивно понятно, что достаточно просто посчитать площадь очерченной фигуры, чтобы получить вероятность встречи.

Тем не менее, вообще говоря, надо обосновать, почему мы так можем делать. Мы умеем считать вероятность для каждого человека попасть в какой-то временной отрезок: это просто его длина.

Умеем считать для двух человек вероятность того, что каждый из них попадет в свой отрезок: это просто произведение их длин, так как события независимы. Ну а произведение длин это площадь прямоугольника.

Для произвольной фигуры считать не умеем, но можем применить теорему Каратеодори. У нас есть прямоугольники, они являются цилиндрами, которые бразуют полукольцо. Для событий, которые задают прямоугольники, знаем, что их вероятность, это площадь. Возьмем σ-оболочку прямоугольников, получим все измеримые множества (мы это делали и доказывали на матане).

Можем рассмотреть функцию, которая равна площади фигуры, на прямоугольниках она совпадает с вероятностью события, которое эта фигура задает. Тогда, на произвольных фигурах эта функция тоже совпадает с вероятностью (в силу единственности по теореме). Таким образом, для подсчета вероятности события достаточно посчитать площадь фигуры, ему соответствущей.

Замечание 1.3.3. В дальнейшем, конечно, проговоренный формализм будет так или иначе опускаться.

Пример 1.3.2. Теперь экспериментов будет бесконечно, но они будут простыми: каждый раз берем и бросаем монетку, вероятности орла и решки (0 и 1) равны $\frac{1}{2}$.

$\Omega = \{0,1\}^{\mathbb{N}} \leftrightarrow [0,1]$. Цилиндры – последовательности с фиксированным началом: если первое число 0, то это отрезок $[0;0.5]$, иначе $[0.5;1]$. Если мы знаем два числа, то получаем отрезок длины 0.25 и так далее.

Для всех цилиндров их длина совпадает с их вероятностью, тогда вероятность любого события = длина на $[0;1]$.

1.4. Схема Бернулли

$\mathfrak{Def}$ 1.4.1. Схема Бернулли – последовательность независимых испытаний в одинаковых условиях.

Замечание 1.4.1. Иными словами, это последовательность выборов с возвращениями. Если выбор без возвращений, то это схема Лапласа.

Обычно следят за двумя исходами: событие A произошло (успех) или не произошло (неудача) и рассматривают следующую величину: $P_{n,p}(k) = P(\text{ровно } k \text{ успехов в } n \text{ испытаниях})$, p — вероятность успеха, а $q := 1-p$ — вероятность неуспеха.

Теорема 1.4.1. $P_{n,p}(k) = \binom{n}{k}p^k(1-p)^{n-k}$

▶ $\Omega = \{0,1\}^n, |\Omega| = 2^n$ (1 – успех, 0 – неудача)
$P(\{1,1,0,1,0,\dots\}) = p\cdot p\cdot q\cdot p\cdot q\cdot\dots = p^aq^b$, где a,b – число успехов и неудач соответственно.
Наше событие состоит из $\binom{n}{k}$ таких равновероятных точек, по аддитивности получаем нужную вероятность. ◀

$\mathfrak{Def}$ 1.4.2. Соответствующее разбиение называется биноминальным.

Вариант для схемы с m исходами: $P_{n,\bar{p}}(\bar{k}) = \binom{n}{k_1,\dots,k_m}\cdot p_1^{k_1}\cdot\ldots\cdot p_m^{k_m}$

Можно посмотреть как зависит $P(k)$ от k. Можно понять, что $P(k) < P(k+1)$ если $k < np+q$ (можно например подставить в формулу и проверить).

Говорят, что наиболее вероятное число успехов равно np. Это число, понятное дело, не является целым, но ближайшее целое число к нему будет наиболее вероятным.

Иногда хочется как-то приближенно оценить $P(k)$, для этого есть следующие две теоремы:

Теорема 1.4.2 (Муавра-Лапласа, локальная)**.** Рассмотрим $a_k = \dfrac{e^{-\frac{(k-np)^2}{2}}}{\sqrt{2\pi npq}}$. Тогда

$$\sup_{k\,:\,|k-np|=o(n^{\frac{2}{3}})} |\frac{P(k)}{a_k} - 1| \to 0$$

▶ Без доказательства ◀

Теорема 1.4.3 (Муавра-Лапласа, интегральная)**.** $p \in (0,1)$. Тогда равномерно по $-\infty \le a < b \le \infty$:

$$\sum P_{n,p}(k) \xrightarrow[n\to\infty]{} \int\limits_a^b \underbrace{\frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}}}_{\varphi(t)} \, \mathrm{d}t$$

Погрешность при этом порядка $\frac{1}{\sqrt{npq}}$.

Теорема 1.4.4 (Пуассон)**.** $\pi_k = e^{-\lambda}\frac{\lambda^k}{k!}, \sum\limits_0^\infty \pi_k = 1, \lambda > 0$

Если $p \xrightarrow[n\to\infty]{}\to 0$, $\lambda = np$, то

$$P_{n,p}(k) - \pi_k \xrightarrow[n\to\infty]{} 0$$

Погрешность: $|\sum\limits_{k\in A}(P_{n,p}(k) - \pi_k)| \le np^2 \forall A \subset \{0, 1, \dots, n\}$

Замечание 1.4.2. Все без доказательства. Теорему Муавра докажем в будущем для всех распределений, а теорему Пуассона в чуть более слабой форме.

Замечание 1.4.3. Можно либо честно считать погрешности, либо пользоваться эвристическим правилом: если np меньше нескольких десятков, то используем Пуассона, иначе Лапласа.

Глава 2

Случайные величины, их распределения

$\mathfrak{Def}$ **2.0.3.** Возьмем все открытые лучи в $\mathbb{R}$, которые идут из $-\infty$.

Тогда множество $\mathcal{B} = \sigma(\{(-\infty, x), x \in \mathbb{R}\})$ называется Борелевской σ-алгеброй.

Замечание 2.0.4. Чтобы привести пример не Борелевской алгебры, нужно сильно постараться. В частности, для этого надо как-то использовать аксиому выбора. На практике считаем, что нам такие множества не встретятся.

Упражнение: доказать, что $\mathcal{B}$ – σ-алгебра, $\mathcal{B} = \sigma(\{(x, y)\}) = \sigma(\{[x, y]\}) = ...$

$\mathfrak{Def}$ **2.0.4.** Случайная величина (с.в.) ξ – функция $\Omega \to \mathbb{R}$, если она $\mathcal{F}$-$\mathcal{B}$-измерима, то есть:

$$\forall B \in \mathcal{B} \colon \xi^{-1}(B) \in \mathcal{F}$$

Другими словами: если есть элемент Борелевской алгебры, то его ξ-прообраз есть элемент σ-алгебры вероятностного пространства.

Замечание 2.0.5. Смысл $\mathcal{F} - \mathcal{B}$-измеримости: хотим, чтобы высказывания вида "ξ принимает значение от 3 до 5" и ему подобные были событиями.

Пример 2.0.1. Пусть $\Omega \subset \mathbb{R}$

1. $\mathcal{F} = \{\emptyset, \Omega\}$, тогда ξ может быть только константой.
2. $\mathcal{F} = \sigma(\{[k, k+1), k \in \mathbb{Z}\})$, тогда ξ – любая ступенчатая функция (длина ступеньки 1).
3. $\mathcal{F} = \{A \colon A = -A = \{-x, x \in A\}\}$, тогда ξ – любая четная функция.

$\mathfrak{Def}$ **2.0.5.** $P_\xi \colon \mathcal{B} \to [0, 1]$ – распределение случайной величины, $P_\xi(B) \stackrel{\text{Def}}{=} P(\xi^{-1}(B))$.

От тройки $(\Omega, \mathcal{F}, P)$ мы сейчас перешли перешли к тройке $(\mathbb{R}, \mathcal{B}, P_\xi)$, это тоже вероятностное пространство.

$\mathfrak{Def}$ **2.0.6.** $\mathcal{F}_\xi = \{\xi^{-1}(B), B \in \mathcal{B}\} \subset \mathcal{F}$ – σ-алгебра, порожденная ξ.

Упражнение: проверить, что это действительно σ-алгебра (просто по определению).

Замечание 2.0.6. В нее входят те и только те события, про которые можно сказать, зная только значение ξ.

Пример 2.0.2. Пусть у нас есть колода карт, из которой мы вытягиваем одну карту, ξ – достоинство вытянутой карты.

Тогда $\mathcal{F}_\xi$ будет состоять только из событий вида (вытянута двойка / вытянута либо тройка, либо четверка, и так далее). При этом, событий вида (вытянута пиковая дама) в этом множестве лежать не будут, так как они не следуют только из значения ξ.

$\mathfrak{Def}$ **2.0.7.** Функция распределения (ф.р.) случайной величины ξ: $F_\xi(x) = P((-\infty, x)) = P(\xi < x)$

Замечание 2.0.7. Часто индекс ξ у F опускают.

Несложно заметить, что F однозначно опреляет P_ξ по теореме Каратеодори для набора полуоткрытых интервалов.

Теорема 2.0.5 (Свойства функции распределения)**.** Пусть F – функция распределения. Тогда:

1. $F(x) \nearrow$
2. $F(x) \xrightarrow[x\to-\infty]{} 0$, $F(x) \xrightarrow[x\to\infty]{} 1$
3. F непрерывна слева

▶ 1 – из монотонности P.
2, 3 – из непрерывности P: рассмотрим $x_n = n$, последовательность лучей $(-\infty, n)$ возрастает и стремится к $\mathbb{R}$, тогда $P((-\infty, x_n) \xrightarrow[n\to\infty]{} P(\mathbb{R}) = 1$ ◀

Если x – точка разрыва, то $P(\xi = x) = F(x+0) - F(x)$.

$$P_\xi([x,y)) = F(y) - F(x)$$
$$P_\xi([x,y]) = F(y+0) - F(x)$$
$$P_\xi((x,y)) = F(y) - F(x+0)$$
$$P_\xi((x,y]) = F(y+0) - F(x+0)$$

2.1. Классификация и примеры распределений

I. ξ и ее распределение называются дискретным, если $\exists B\colon P_\xi(B) = 1$, B не более чем счетное.

1. $\xi = C$ – константа, тогда $F_\xi(x) = \begin{cases} 0, x \le c \\ 1, x > c \end{cases}$

   Замечание 2.1.1. Удобно записывать $F_\xi(x) = \mathbb{1}_{(x\ge c)}$, $\mathbb{1}_A(x)$ выдает 1, если $x \in A$, 0 иначе

2. $\xi = \begin{cases} 1, \text{с вероятностью } p \\ 0, \text{с вероятностью } q = 1-p \end{cases}$.

   $F(x) = q\mathbb{1}_{(x\ge 0)} + p\mathbb{1}_{(x\ge 1)}$
   В общем случае: $\{x_k\}, \{p_k\}$ – ряд распределения, $p_k = P(\xi = x_k)$, $F_\xi(x) = \sum\limits_{k\colon x_k<x} p_k$.

3. $\xi \sim Bin(n,p)$, если $x_k = 0 \dots n, p_k = P_{n,p}(k)$
4. $\xi \sim Poiss(x), x_k = 0, 1, 2, \dots, p_k = e^{-\lambda}\frac{\lambda^k}{k!}$

II.

$\mathfrak{Def}$ **2.1.1.** ξ и ее распределене называют непрерывным, если $F_\xi(x) \in C(\mathbb{R})$;

$\mathfrak{Def}$ **2.1.2.** Называют абсолютно непрерывной, если есть $p_\xi(x)$ (*плотность распределения*) такая, что:

$$F_\xi(x) = \int\limits_{-\infty}^{x} p_\xi(x)\mathrm{d}x$$

1. $\xi \sim U[a,b]$, если $F_\xi = \begin{cases} 0, x \leq a; \\ \frac{x-a}{b-a}, a < x \leq b \\ 1, x > b \end{cases}$. Это абсолютно непрерывное распределение: $p_\xi(x) = \frac{\mathbb{1}_{[a,b]}(x)}{b-a}$
2. $\xi \sim N(a,b)$ - нормальное (Гауссово), если $p_\xi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{-(x-a)^2}{2\sigma^2}}$. График распределения — «колокол», σ отвечает за его ширину, a – за местоположение его пика.

III.

$\mathfrak{Def}$ 2.1.3. ξ и ее распределение называют сингулярным, если ξ непрерывна и $\exists B \in \mathcal{B}\colon P_\xi(B) = 1$, при этом $\mu(B) = 0$. Другими словами, случайная величина попадает в множество меры ноль с вероятностью один.

Пример: Канторова лестница

Теорема 2.1.1 (О классификации)**.** Пусть $F_\xi(x)$ – функция распределения произвольной случайной величины ξ. Она единственным образом представляется в виде комбинации

$$F_\xi(x) = \lambda_1 F_1(x) + \lambda_2 F_2(x) + \lambda_3 F_3(x),$$
$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

где F_1, F_2, F_3 – функции распределения каких-то дискретной, непрерывной и сингулярной случайной величины соответственно, при этом такое разложение единственно.

Теорема 2.1.2 (Теорема об обратной функции)**.** Пусть какая-то функция F удовлетворяет свойствам функции распределения, тогда F – функция распределения некоторой случайной величины ξ.

▶ Пусть $F^*(y) = \sup\{x\colon F(x) \leq y\}$. Несложно заметить, что $F^* = F^{-1}$, в тех точках, где F^{-1} существует. Еще есть точки разрыва и участки, где F константа, там мы F^* не определили, давайте сделаем это.

Пусть x – точка разрыва, $F(x-0) = a, F(x+0) = b$, тогда скажем, что $F^*(y) = x, y \in [a;b]$. Если $F(x) = y$ на отрезке $[a;b]$, то $F^*(y) = a$.

Отлично, доопределили обратную функцию, F^* называется обобщенной обратной функией. Еще верен факт, что $F^{*^*}(x) = F(x)$, разрывы переводит в участки постоянства и наоборот.

Возьмем случайную величину $\eta \sim U[0,1]$, хотим доказать, что $\xi = F^*(\eta)$.

$$F_\xi(z) = P(F^*(\eta) < z) = P(\eta < F(z)) = F_\eta(F(z)) =$$

Так как η – равномерное распределение

$$= F(z)$$

◀

Замечание 2.1.2. Важна даже не сколько теорема, сколько ее доказательство. Оно полезно, если хотим сгенерировать какую-то случайную величину, зная ее распределение и имея только генератор равномерно распределенной случайной величины (например, функция `rand()` в каком-нибудь языке программирования).

Например, хотим сгенерировать монетку, с вероятностью орла p. Мы знаем ее функцию распределения $F(x) = \begin{cases} p, x \in [0, 0.5] \\ 1, x \in (0.5, 1] \end{cases}$. Строим $F^*(y) = \begin{cases} 1, y \in [p, 1] \\ 0, y \in [0, p) \end{cases}$

А теперь берем генератор равномерно распределенных случайных чисел с отрезка $[0;1]$, подставим число, которое оно выдал, в обратную функцию, получаем единичку или нолик с нужными вероятностями.

2.2. Случайные векторы и их распределения

$\mathfrak{Def}$ 2.2.1. Вектор $\vec{\xi} = (\xi_1, \dots, \xi_n)\colon \Omega \to \mathbb{R}^n$ называется случайным вектором, если все его координаты ξ_k – случайные величины.

Замечание 2.2.1. Значок вектора будет часто опускаться, если из контекста понятно.

Борелевские множества определаются аналогично одномерному случаю: определяем лучи как $(-\infty, \vec{x}) = (-\infty, x_1) \times \dots \times (-\infty, x_n)$, а потом берем их σ-замыкание.

Аналогично определяется и функция распределения: $F_{\vec{\xi}}(\vec{x}) = P_{\vec{\xi}}((-\infty, \vec{x}))$.

Теорема 2.2.1 (Свойства функции распределения)**.** Пусть F – функция распределения. Тогда:

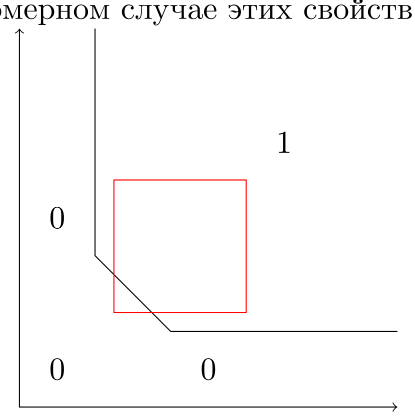
1. $F(x) \nearrow$ по каждой переменной.

2. $\exists i\colon x_i \to -\infty \Rightarrow F(x) \to 0$, а так же $F(x) \to 1$, если $\forall i\colon x_i \to \infty$.

3. F непрерывна слева по каждой переменной.

Если мы живем в двумерном случае и есть квадратик $[a,b] \times [c,d]$, то $P(\xi \in \square) = F(b,d) - F(a,d) - F(b,c) + F(a,c)$. Соответственно, в общем случае тут просто формула включений-исключений.

Проблема: в одномерном случае есть теорема об обратной функции, с помощью которой для любой функции с нужными свойствами можно построить соответствующее распределение. В многомерном случае этих свойств не достаточно. Нарисуем график следующей функции $\mathbb{R}^2 \to \mathbb{R}$.



Несложно заметить, что график этой функции удовлетворяет всем трем свойствам функции распределения. Пусть этому графику соответствует какая-то случайная величина. Тогда можно взять красный квадратик на картинке и посчитать вероятность того, что ξ находится в этом квадрате. Мы получим: $P(\xi \in \square) = 1 - 1 - 1 + 0 = -1$. Такого быть не может, и значит этот график не задает никакую случайную величину и этих трех свойств недостаточно.

Утверждается, что достаточно добавить еще четвертое свойство: для любого квадрата $[a,b] \times [c,d]$ верно: $F(b,d) - F(a,d) - F(b,c) + F(a,c) \geq 0$ (опять же, в случае большей размерности, это свойство обобщится до формулы включений-исключений), и тогда теорема об обратной функции в многомерном случае тоже будет верна.

Аналогично одномерному случаю, ξ – абсолютно непрерывая случайная величина, если она равна

$$\int\limits_{-\infty}^{x} p_\xi(t)\mathrm{d}t = \int\limits_{-\infty}^{x_1} \mathrm{d}t_1 \dots \int\limits_{-\infty}^{x^n} p(t_1, \dots, t_n)\mathrm{d}t_n$$

Равномерным распределением называют распределение с плотностью $p_\xi = \frac{1}{V(D)}\mathbb{1}_D(x)$, где $V(D)$ – мера Лебега. Определяют именно через плотность, так как описывать функцию распределения в многомерном случае сложно из-за кучи случаев.

В отличие от одномерного случая, сингулярных распределений будет гораздо больше. Например, можно взять просто отрезок, вложить его в пространство и получить интеграл, равный нулю.

2.2.1. Независимые случайные величины

Задача: есть распределение, хотим по его проекциям восстановить исходное. К сожалению, это сделать не всегда можно однозначно: вот есть три распределения, проекции на оси одинаковые.

Утверждается, что если координаты независимы, то это все-таки можно сделать.

𝔇𝔢𝔣 2.2.2. $\xi_1, \dots, \xi_n$ называют независимыми, если $\forall B_1, \dots, B_n \in \mathbb{B}$: $P(\xi_1 \in B_1, \dots, \xi_n \in B_n) = \prod P(\xi_i \in B_i)$.

Или же (равносильно), если $\forall x_1, \dots, x_n$: $F_\xi(x) = \prod F_{\xi_i}(x_i)$.

Теорема 2.2.2. Если ξ дискретное, то все ξ_k дискретны и независимость $\{\xi_k\}$ равносильна тому, что $\forall x \in \mathbb{R}^n P(\xi = x) = \prod(\xi_k = x_k)$.

▶ Проекции дискретны потому что различных проекций не более чем количество различных точек, а их счетно. Распосильность просто по определению. ◀

Теорема 2.2.3. ξ абсолютно непрерывная, тогда ξ_k – абсолютно непрерывны и их независимость равносильна тому, что $p_\xi(x) = \prod p_{\xi_k}(x_k)$ в точках непрерывности p_ξ.

Замечание 2.2.2. Условие про точки непрерывности важное: можно взять, как угодно поменять значение p в одной точке, это будет все еще плотность, а вот теорема сломается.

2.3. Условная вероятность

До этого отдельно говорили про дискретный случай, отдельно про абсолютно непрерывный. Хочется их как-то объединить вместе. Для этого используется интеграл Стилтьеса.

Если интеграл Римана определяется как $\int\limits_a^b f(x)\mathrm{d}x = \lim\limits_{\sup x_{k+1}-x_k \to 0} \sum f(x_k)(x_{k+1} - x_k)$, то

интеграл Стилтьеса: $\int\limits_a^b f(x)\mathrm{d}F(x) = \lim\limits_{\sup x_{k+1}-x_k \to 0} \sum f(x_k)(F(x_{k+1}) - F(x_k))$. В каком-то смысле, чем больше плотность в точке, тем она важнее при подсчете интеграла.

В частности, если $F(x) = x$, то это просто интеграл Римана.

Замечание 2.3.1. На самом деле, этот интеграл у нас появлялся тогда, когда мы хотели сделать какую-нибудь замену переменной и что-то заносили под дифференциал.

Интеграл Стилтьеса обладает всеми свойствами интеграла Римана (линейность и прочие).

Можно в качестве $F(x)$ взять ступеньку. Тогда, если раскроем интеграл по определению, то все слагаемые слева от ступени будут равны нулю, справа от ступеньки будут равны нулю, а отрезочек, содержащий ступеньку, будет единственный, который не занулится. То есть, $\int f(x)dF(x) = f(c)$, если у нас ступенька высоты 1 в точке c.

Аналогично, если взять F, у которой много ступенек, то интеграл будет равен сумме значений f в точках.

Таким образом, обычная сумма, или даже сумма ряда, тоже интеграл Стилтьеса, это очень удобное и полезное свойство.

𝔇𝔢𝔣 2.3.1.

$$P(a \mid \xi = X) = \frac{P(A \cap \{\xi = x\})}{P(\xi = x)} = \begin{cases} \lim\limits_{\varepsilon \to 0} \frac{P(A \cap |\xi - x| < \varepsilon)}{P(|\xi - x| < \varepsilon)}, \text{если он существует} \\ 0, \text{если } \exists \varepsilon \colon P(|\xi - x| < \varepsilon) = 0 \end{cases}$$

𝔇𝔢𝔣 2.3.2. $F_\eta(y \mid \xi = x) = P(\eta < y \mid \xi = x)$ – условная функция распределения.

Непрерывная формула полной вероятности:

$$P(A) = \sum_{k=-\infty}^{\infty} P(\underbrace{\xi \in [x_k - \varepsilon, x_k + \varepsilon)}_{H_k})P(A \mid H_k) \xrightarrow[\varepsilon \to 0]{} \int\limits_{-\infty}^{\infty} P(A \mid \xi = x)\mathrm{d}F_\xi(x)$$

Пример 2.3.1. Пусть есть равномерное распределение $\theta \sim U[0,1]$ и распределение ξ, про которое мы знаем, что $(\xi \mid \theta = p) \sim Bin(n, p)$. Хочется как-то найти ξ, это можно сделать из формулы выше:

$$P(\xi = k) = \int\limits_0^1 P_{n,p}(k)\mathrm{d}p = \int\limits_0^1 \binom{n}{k} p^k (1-p)^{n-k}\mathrm{d}p = \frac{1}{n+1}$$

Пример 2.3.2. Закон композиции распределений:

Пусть есть ξ, η – независимые случайные величины, хочется что-нибудь узнать про величину $\xi + \eta$:

$$F_{\xi+\eta}(x) = P(\xi + \eta < x) = \int\limits_{-\infty}^{\infty} P(\xi + \eta < x \mid \xi = t)\mathrm{d}F_\xi(x) = \int\limits_{-\infty}^{\infty} F_\eta(x - t)\mathrm{d}F_\xi(t) = F_\eta * F_\xi$$

Операция “*” называется сверткой.

Для плотностей:

$$p_{\xi+\eta}(t) = \int\limits_{\mathbb{R}} p_\eta(x - t)p_\xi(t)\mathrm{d}t = p_\eta * p_\xi$$

Посмотрим теперь, как условная вероятность выглядит в многомерном случае. Пусть (ξ, η) – абсолютно непрерывный случайный вектор. Тогда

$$F_\eta(y \mid \xi = x) = \lim_{\varepsilon \to 0} \frac{P(\eta < y, |\xi - x| < \varepsilon)}{P(|\xi - x| < \varepsilon)} =$$

Раскроем условную вероятность по определению

$$= \lim_{\varepsilon \to 0} \frac{\int\limits_{x-\varepsilon}^{x+\varepsilon} \mathrm{d}t \int\limits_{-\infty}^{y} p_{\xi,\eta}(t, s)\mathrm{d}s}{\int\limits_{x-\varepsilon}^{x+\varepsilon} \int\limits_{-\infty}^{\infty} p_\xi(t, s)\mathrm{d}t\mathrm{d}s}$$

Применим правило Лопиталя, дифференцируем по ε, получим

$$\lim_{\varepsilon \to 0} = \frac{\int\limits_{-\infty}^{y} (p_{\xi,\eta}(x + \varepsilon, s) + p_{\xi,\eta}(x - \varepsilon, s))\mathrm{d}s}{\int\limits_{-\infty}^{\infty} p_\xi(x + \varepsilon, s) + p_\xi(x - \varepsilon, s)\mathrm{d}s} = \frac{\int\limits_{-\infty}^{y} p_{\xi,\eta}(x, s)\mathrm{d}s}{\int\limits_{-\infty}^{\infty} p_\xi \mathrm{d}s}$$

Продифференцируем по y, получим

$$p_\eta(y \mid \xi = x) = \frac{p_{\xi,\eta}(x,y)}{p_\xi(x)}$$

Интуитивно эта формула угадывается и описывается так: есть двумерный график плотности, поверхность над плоскостью (x,y), высота поверхности в точке это и есть плотность.

Если зафиксировали x, то мы сечем поверхность относительно прямой, получаем какой-то график функции, он нам подходит с точностью до нормировки, надо ее соблюсти, поэтому вылезает знаменатель.

2.4. Моменты случайных величин

$\mathfrak{Def}$ 2.4.1. Математическое ожидание (мат. ожидание, среднее значение) $E\xi = M\xi = \int\limits_{-\infty}^{\infty} x\mathrm{d}F_\xi(x)$

В дискретном случае этот интеграл превращается в $\sum\limits_k x_k p_k$.

В абсолютно непрерывном случае это превращается в Римановский интеграл $\int\limits_{-\infty}^{\infty} xp_\xi(x)\mathrm{d}x$

Упражнение: $E\xi = \int\limits_0^\infty (1 - F(x))\mathrm{d}x - \int\limits_{-\infty}^{0} F_\xi(x)\mathrm{d}x$

Пример 2.4.1. Матожидание существует не всегда. Рассмотрим распределение Коши: $p_\xi(x) = \frac{1}{\pi}\frac{1}{1+x^2}$. Домножим на x, получим величину порядка $\frac{1}{x}$, интеграл которой расходится. И матожидание для этого распределения не определено.

Свойства:

1. Линейность: $E(a\xi + b\eta) = aE\xi + bE\eta$.

   ▶ Пусть $a > 0$. Тогда $Ea\xi = \ \mathrm{d}F_{a\xi}(x) = \int x\mathrm{d}P(a\xi < x) = a\int \frac{d}{a}\mathrm{d}P(\xi < \frac{x}{a}) = a\int y\mathrm{d}F_\xi(y)$.
   Во вторую часть придется поверить, так как доказывается она для интеграла Стилтьеса не очень легко, но там просто надо применить какой-то матанализ. В дискретном/абсолютно непрерывном случае там просто. ◀

2. $Eg(\xi) = \int g(x)\mathrm{d}F_\xi(x)$.

   Замечание 2.4.1. Формула верна и для случайного вектора $\vec{\xi}$. Только интеграл будет многомерным.

3. Монотонность: пусть $P(\xi \geq 0) = 1$ (обычно сокращают до "$\xi \geq 0$ почти всегда (п. в)"). Тогда $E\xi \geq 0$.

   ▶ $0 = P(\xi < 0) = F_\xi(0) \Rightarrow E\xi = \int\limits_0^\infty x\mathrm{d}F_\xi \geq 0$. ◀

4. $\xi \geq 0$ почти всюду и $E\xi = 0$, тогда $\xi = 0$ почти всюду.

   ▶ Пусть не так, тогда $F_\xi(x_0) < 1$ для некоторого $x_0 > 0$. Ну а тогда

$$E\xi = \int\limits_0^\infty \geq \int\limits_{x_0}^\infty \geq \int_{x_0}^\infty x_0\mathrm{d}F = x_0(F(\infty) - F(x_0)) > 0$$

◀

5. ξ, η независимы, тогда $E\xi\eta = E\xi E\eta$ (в записи считаем, что умножение имеет больший приоритет, чем взятие матожидания, поэтому скобки опускаем).

   ▶ В дискретном случае:

   $$E\xi\eta = \sum x_k y_k P(\xi = x_k, \eta = y_k) = \sum x_k y_k P(\xi = x_k)P(\eta = y_k) = \sum x_i P(\xi = x_i) \sum y_j P(\eta = y_j)$$

   В непрерывном случае тоже легко, в общем – сложно. ◀

6. Матожидание смеси: $F_\xi(x) = \sum\limits_k \lambda_k F_k(x)$, тогда

   $$E\xi = \int x\mathrm{d}F = \sum_k \lambda_k \int x\mathrm{d}F_k(x) = \sum \lambda_k E_k$$

Пример 2.4.2. Играем против игроков A и B. Вероятнотсть выиграть против A равна 0.6, против B – 0.55. С игроком A каждый раз мы играем на 5 долларов, с игроком B – на один.

Изначально у нас есть 10 долларов. Вопрос: как играть, чтобы получить побольше выгоды?

Можно посчитать матожидания выигрыша за одну игру, и играть с тем, для кого матожидание больше. $E_a = 5 \cdot 0.6 - 5 \cdot 0.4 = 1$, $E_B = 1 \cdot 0.55 - 1 \cdot 0.45 = 0.1$, поэтому, казалось бы, лучше играть с A: в среднем за игру мы будем у него выигрывать больше.

Но есть еще следующая проблема: при игре с A есть больший шанс взять и проиграть все деньги в самом начале. Оказывается, что если это учесть, то вероятность выиграть у A равна 0.555, у B – 0.865.

Пример 2.4.3. Есть автобусы. Средний интервал между ними равен 10 минутам. Вопрос: сколько в среднем придется ждать автобуса, если придем в случайное время на остановку?

Если автобусы идут равномерно, то ответ – 5 минут. И это минимально возможный ответ. Но можно взять и рассмотреть следующий крайний случай: куча автобусов приходят почти в один момент времени, а следующий – через очень большой интервал. Тогда средний интервал будет все еще 10 минут, а среднее время ожидания – очень большим.

Вообще говоря, ответ может быть любым числом из луча $[5, \infty)$.

Эти примеры показывают, что просто матожидания не достаточно, чтобы оценить характер распределения, для этого вводят еще одну его характеристику.

$\mathfrak{Def}$ 2.4.2. Дисперсия случайной величины ξ: $D\xi = E(\xi - E\xi)^2$.
$\sigma_\xi = \sqrt{D\xi}$ – среднеквадратическое отклонение.

Свойства дисперсии:

1. $D\xi \geq 0$, $D\xi = 0 \Leftrightarrow \xi = C$ почти всегда.

2. $D\xi = E(\xi^2 - 2\xi E\xi + (E\xi)^2) = E\xi^2 - 2E\xi E\xi + (E\xi)^2 = E\xi^2 - (E\xi)^2$

3. $D(c\xi) = c^2 D\xi$

4. ξ, η – независимые, тогда $D(\xi + \eta) = D\xi + D\eta$

   ▶

   $$D(\xi+\eta) = E\xi^2+E\eta^2+2E\xi\eta-((E\xi)^2+(E\eta)^2+2E\xi E\eta) = D\xi+D\eta+2(E\xi\eta-E\xi E\eta) = D\xi+D\eta$$

   ◀

   Следствие 2.4.0.1. $D(\xi + C) = D\xi + DC = D\xi$

Пример 2.4.4. 1. $EC = C, DC = 0$

2. $\xi = \mathbb{1}_A(\omega) = \begin{cases} 1, w \in A \\ 0, w \notin A \end{cases}$.

$$E\xi = 1P(A) + 0P(\bar{A}) = P(A)$$
$$D\xi = P(A) - P^2(A) = P(A)P(\bar{A})$$

3. $\xi \sim Bin(n, p)$.

$$E\xi = \sum_{k=0}^{n} kP_{n,p}(k)$$

. Можно записать по-другому: $\xi = \sum\limits_{1}^{n} \mathbb{1}_{A_k}$, где A_k – успех на k-м шаге. Тогда

$$E\xi = \sum_{1}^{n} E\mathbb{1}_{A_k} = np$$

.

Заметим, что все броски независимы, тогда

$$D\xi = np(1-p)$$

4. $\xi \sim U[a, b]$.

$$E\xi = \int xp_\xi(x)\mathrm{d}x = \int_a^b \frac{x}{b-a}\mathrm{d}x = \frac{a+b}{2}$$
$$D\xi = E\xi^2 - (E\xi)^2 = \int \frac{x^2}{b-a}\mathrm{d}x - (\frac{a+b}{2})^2 = \frac{(b-a)^2}{12}$$

5. $\xi \sim N(a, \sigma^2)$.

$$E\xi = \int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-a)^2}{2\sigma^2}}\mathrm{d}x =$$

Сделаем замену $t = x - a$

$$\int_{-\infty}^{\infty} (t+a)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{t^2}{2\sigma^2}}\mathrm{d}t = 0 + a = a$$

$D\xi = \sigma^2$ (упражнение)

Введем две операции: $\overset{0}{\xi} = \xi - E\xi$ – центрирование и $\widetilde{\xi} = \frac{\xi}{\sigma_\xi} = \frac{\xi}{\sqrt{D\xi}}$ – нормировка.

Эти операции часто используются на практике. Можно, например, посмотреть в теорему Муавра-Лапласа и их там увидеть. Операции полезны, например, тем, что после их применения гораздо легче проводить какие-либо сравнения распределений.

𝔇𝔢𝔣 2.4.3. $E\xi^k$ – момент k-го порядка (начальный)
$E|\xi|^k$ – абсолютный момент k-го порядка
$E(\xi - E\xi)^k$ – центральный момент k-го порядка
$E|\xi - E\xi|^k$ – абсолютный центральный момент k-го порядка

Замечание 2.4.2. k, вообще говоря, не обязательно целое число, хотя на практике такое редко встречается.

Замечание 2.4.3. Дисперсия – центральный момент 2-го порядка.

Замечание 2.4.4. $\frac{E(\xi-E\xi)^3}{\sigma^3}$ – коэффициент асимметрии.

Замечание 2.4.5. $\frac{E(\xi-E\xi)^4}{\sigma^4}$ – эксцесс (степень остроты пика)

Обозначим $a_k = E|\xi|^k$. Через a_k выражаются все моменты. Например, $D\xi = E(\xi-E\xi)^2 = a_2-a_1^2$. Поэтому, верно следующее утверждение:

Утверждение 2.4.1. Если $\exists E|\xi|^n < +\infty$, то конечны все моменты порядка не больше чем n.

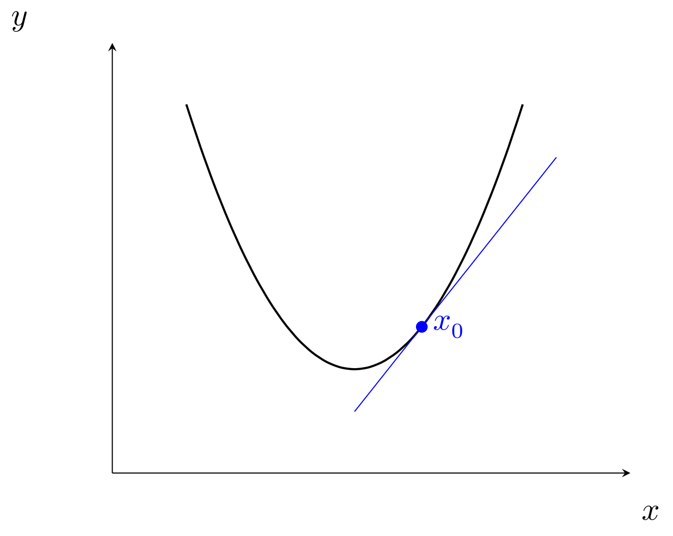
▶ $E|\xi|^n = \int\limits_{-\infty}^{+\infty} |x|^n \mathrm{d}F(x)$, он сходится. Если уменьшим n, то получим интеграл, который будет мажорироваться исходным везде, кроме отрезка $[-1;1]$, значит интеграл на $\mathbb{R}\setminus[-1;1]$ сходится. Ну а на этом отрезке он конечен: $|x| \leq 1, \mathrm{d}F = p\mathrm{d}x$, интеграл плотности конечен, а мы его умножаем на что-то, не большее единицы. ◀

Как мы уже выяснили, просто матожидания недостаточно для того, чтобы однозначно определить случайную величину. Вопрос: а достаточно ли такого множества значений $\{a_k\}_{k=1}^{\infty}$? В общем случае ответ – нет, но:

Теорема 2.4.1. Если $\overline{\lim} \frac{(E|\xi|^n)^{\frac{1}{n}}}{n} < +\infty$, то моменты определяют распределение.

Теорема 2.4.2 (Неравенство Йенсена)**.** Пусть $g(x)$ выпукла вниз, тогда $g(E\xi) \leq Eg(\xi)$

▶ Через каждую точку графика g можно провести прямую, которая ограничивает график снизу



(это будет просто касательная)

Возьмем точку $x_0 = E\xi$, через нее проходит прямая $ax+b$, ограничивающая график снизу: $ax+b \leq g(x), ax_0+b = g(x_0)$.

Тогда

$$g(\xi) \geq g(E\xi) + a(\xi - E\xi)$$

. А теперь возьмем матожидание от обеих частей:

$$E(g(E\xi)) = g(E\xi)\text{как матожидание константы,}$$
$$Ea(\xi - E\xi) = Ea\xi - EaE\xi = Ea\xi - Ea\xi = 0$$

Откуда получаем

$$g(E\xi) \leq Eg(\xi)$$

. ◀

Следствие 2.4.2.1 (Неравенство Лягунова). $0 < s < t$. Тогда

$$(E|\xi|^s)^{\frac{1}{s}} \le (E|\xi|^t)^{\frac{1}{t}}$$

.

▶ Применим Йенсена к $g(x) = |x|^{\frac{t}{s}}$ ◀

Следствие 2.4.2.2. $|E\xi| \le E|\xi|$.

Теорема 2.4.3 (Неравенство Гёльдера). $p > 1, q > 1, \frac{1}{p} + \frac{1}{q} = 1$, $E|\xi|^p, E|\eta|^q < \infty$. Тогда

$$E|\xi\eta| \le (E|\xi|^p)^{\frac{1}{p}}(E|\eta|^q)^{\frac{1}{q}}$$

▶ Если $\xi = 0$ или $\eta = 0$, то $0 \le 0$. Иначе $E > 0$.

Заведем обозначения

$$\tilde{\xi} = \frac{|\xi|}{(E|\xi|^p)^{\frac{1}{p}}}$$

$$\tilde{\eta} = \frac{|\eta|}{(E|\eta|^q)^{\frac{1}{q}}}$$

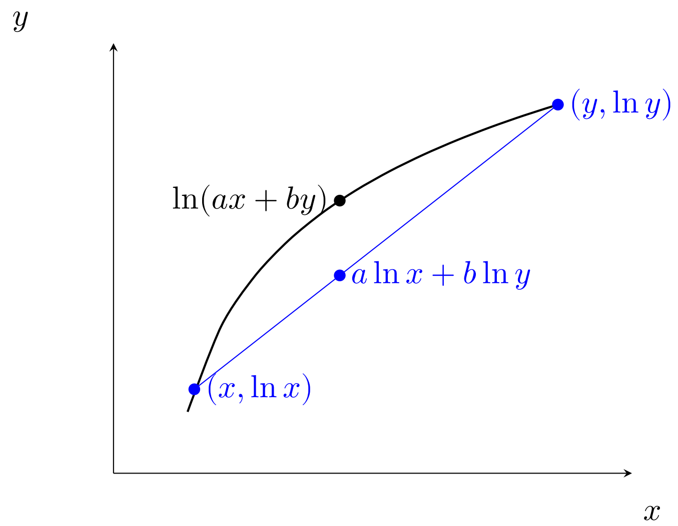
Пусть $x, y, a, b > 0, a + b = 1$. Тогда верно

$$x^a y^b \le ax + by$$

Прологарифмируем это, получим

$$a \ln x + b \ln y \le \ln(ax + by)$$

.



Вот эта картинка поясняет, почему неравенство действительно верно. Возьмем

$$a = \frac{1}{p}, b = \frac{1}{q}$$

$$x = \tilde{\xi}^p, y = \tilde{\eta}^q$$

Подставляем в неравенство, берем матожидание

$$Eax = E\frac{1}{p}\frac{|\xi|^p}{E|\xi|^p} = \frac{1}{p}\frac{1}{E|\xi|^p}E|\xi|^p = \frac{1}{p}$$

$$E\tilde{\xi}\tilde{\eta} \le \frac{1}{p} \cdot 1 + \frac{1}{q} \cdot 1 = 1$$

$$E|\xi||\eta| \le (E|\xi|^p)^{\frac{1}{p}}(E|\eta|^q)^{\frac{1}{q}}$$

◀

Другие характеристики распределения:

1. Носитель – множество, куда мы попадаем с вероятностью 1.
2. Определим $z_p = \sup\{x \mid F_\xi(x) \le p\}$. Тогда $z_{\frac{1}{2}}$ – медиана, $z_{\frac{1}{4}}, z_{\frac{3}{4}}$ – квартили.
3. Мода $= \arg\max p_\xi(x)$.

Def 2.4.4. $E\vec{\xi} = (E_{\xi_1}, \dots, E_{\xi_n})$.

Def 2.4.5. Ковариацией пары случайных величин ξ и η называется

$$cov(\xi, \eta) = E((\xi - E\xi)(\eta - E\eta) = E\xi\eta - E\xi E\eta$$

В частости, при $\xi = \eta$ ковариация это дисперсия.

Теорема 2.4.4 (Свойства ковариации).

1. $cov(\xi, \eta)$ – билинейная по каждому аргументу, симметричная форма.
2. Если ξ, η независимы, то $cov(\xi, \eta) = 0$.

Def 2.4.6. Рассмотрим случайный вектор $\xi = (\xi_1, \dots, \xi_n$, введем обозначение $\sigma_{ij} = cov(\xi_i, \xi_j)$

$\sigma^n_{i,j=1}$ – матрица ковариаций – аналог дисперсии. Эта матрица симметрична и неотрицательно определена.

Доказательство неотрицательной определенности:

▶ Напоминание: A неотрицательно определена, если $x^T Ax \ge 0$ для любого вектора x.

Возьмем случайную величину ξ из векторного пространства, она представляется в виде линейной комбинации ξ_i. Тогда

$$0 \le D(\sum_{i=1}^n c_i\xi_i) = E(\sum c_i(\xi_i - E\xi_i))^2 = \sum_{i,j=1}^n c_ic_j\sigma_{ij}$$

Что и требовалось. ◀

Def 2.4.7. Коэффициентом корреляции называют $\rho(\xi, \eta) = \frac{cov(\xi,\eta)}{\sqrt{D_\xi D_\eta}}$

Если $\rho(\xi, \eta) = 0$, то ξ, η называют некоррелироваными.

Замечание 2.4.6. $|\rho| \le 1$: применим неравенство Гёльдера с $p = q = 2$ к $\xi - E\xi$ и $\eta - E\eta$.

Замечание 2.4.7. Из независимости следует некоррелируемость, а обратно нет. Возьмем любое распределение ξ, симметричное относительно начала координат, $\rho(\xi, \xi) = 0$, так как **TODO** на лекции так сказали, но это неверно, ниже правильный пример из гугла **TODO**

Возьмем $\xi \sim U[-\pi, \pi]$, $\eta_1 = \sin\xi, \eta_2 = \cos\xi$. η_1, η_2 зависят, но корреляция равна нулю.

Замечание 2.4.8. ρ, по сути, является мерой линейной зависимости между случайных величин.

При этом, его модуль – степень зависимости, а знак – направление: если $\rho > 0$, то обе величины вместе растут/падают, иначе наоборот.

Рассмотрим задачу линейного прогноза: есть случайная величина η, мы ее не знаем, но хотим ее как-то угадать. Для этого мы ее будем приближать какой-нибудь линейной функцией $g(\xi) = a\xi + b$.

Надо взять и подобрать в каком-то смысле наилучшую функцию, то есть, оптимальные параметры a и b. Будем минимизировать среднеквадратичное отклонение: $E(\eta - g(\xi))^2$

Замечание 2.4.9. А еще это иногда называют задачей линейной регрессии, например, встречается в машинном обучении.

Упражнение: если $E(\eta - a^*\xi - b^*)^2$ – минимум (a^*, b^*) – оптимальные параметры), то $a^* = \frac{cov(\xi,\eta)}{D\xi}, b^* = E\eta - a^*E\xi$.

Можно заметить, что $E(\eta - a^*\xi - b^*)^2 = (1-\rho^2)D\eta$ (просто раскроем скобки). Тогда $\eta = a\xi + b$ тогда и только тогда когда $|\rho| = 1$

Замечание 2.4.10. Наличие корелляции вовсе не гарантирует какую-либо причинно-следственную связь.

Например, есть известная шутка про то, что численность пиратов уменьшается из-за глобального потепления (корелляция то действительно есть между этими двумя величинами, а вот связь – вряд ли)

$\mathfrak{Def}$ 2.4.8. Пространство L^2 случайных величин – пространство величин с конечным вторым моментом: $E\xi^2 < \infty$

Упражнение: в нем $cov(\xi, \eta)$ – скалярное произведение (можно просто руками проверить все свойства скалярного произведения).

"Длина" ξ это $\sqrt{(\xi,\xi)} = \sqrt{D\xi} = \sigma\xi$ – среднеквадратичное отклонение.

"Косинус угла" это $\rho(\xi, \eta)$, некореллируемые величины – перпендикулярные.

Лучшее линейное приближение η – проекция на плоскость $a\xi + b = 0$: расстояние будет минимальным, когда получится перпендикуляр. А он вылезает как разность вектора и проекции. **TODO** картинку бы **TODO**

2.5. Общий (функциональный) прогноз

Более общая задача: теперь $g(\xi)$ – произвольная борелевская фукнция, не обязательно линейная. И хотим приблизить η этой функцией.

Оказывается, что решением этой задачи является условное матожидание.

$\mathfrak{Def}$ 2.5.1. Пусть $\forall x\colon g(x) = E(\eta \mid \xi = x)$. Тогда $E(\eta \mid \xi) = g(\xi)$. Это и называется условным матожиданием.

$\mathfrak{Def}$ 2.5.2. Пусть $\mathbb{A} \subset \mathbb{F}$ – σ-алгебра.

$\eta_{\mathbb{A}} = E(\eta \mid \mathbb{A})$ – $\mathbb{A}$-измеримая случайная величина (то есть $\forall B \in \mathbb{B}\colon E(\eta \mid \mathbb{A})^{-1}(B) \in \mathbb{A}$), такая что $\forall A \in \mathbb{A}\colon E\eta\mathbb{1}_A = E\eta_{\mathbb{A}}\mathbb{1}_A$

Пример 2.5.1. Пусть $\Omega = \mathbb{R}$

1. $\mathbb{A} = \{\emptyset, \mathbb{R}\}$, тогда $\eta_{\mathbb{A}} = E\eta$
2. $\mathbb{A} = \sigma(\{[k, k+1)\})$. Новая случайная величина должна быть постоянна на этих интервалах (потому что она $\mathbb{A}$-измерима), и при этом интегралы должны совпадать с интегралами исходной величины, откуда получаем значение на каждом интервале.
3. $\mathbb{A} = \{A\colon A = -A\}$, $\eta_{\mathbb{A}}(w) = \frac{\eta(w)+\eta(-w)}{2}$

$\mathfrak{Def}$ 2.5.3. Ξ – набор множеств, тогда $E(\eta \mid \Xi) = E(\eta \mid \sigma(\Xi))$

ξ – случайная величина, тогда $E(\eta \mid \xi) = E(\eta \mid \mathbb{F}_\xi)$

Замечание 2.5.1. Это определение и первое равносильны, доказывать это мы, конечно же, не будем.

Теорема 2.5.1. Свойства условного матожидания:

1. $E(\xi \mid \mathbb{F}_\xi) = \xi$

2. Пусть ξ не зависит от σ-алгебры $\mathbb{A}$ (то есть $\xi^{-1}(B \in \mathbb{B})$ и $A \in \mathbb{A}$ независимы). Тогда $E(\xi \mid \mathbb{A}) = E\xi$ (потому что в силу независимости ξ от $\mathbb{A}$ мы не получаем никакой информации от того, произошло ли какое-то событие A, или нет).
3. Формула полного матожидания: $E(E(\xi \mid \mathbb{A})) = E\xi$
4. $E(E(\xi \mid \mathbb{A}_1) \mid \mathbb{A}_2) = E(\xi \mid \mathbb{A}_1 \cap \mathbb{A}_2)$
5. η – $\mathbb{A}$-измерима ($\mathbb{F}_\eta \subset \mathbb{A}$). Тогда $E(\xi\eta \mid \mathbb{A}) = \eta E(\xi \mid \mathbb{A})$.

▶ 1. Просто по определению.
2. На пальцах: можно рассмотреть $E(\xi \mid A_1)$ и $E(\xi \mid A_2)$. Если предположить, что они разные, то A_1 и A_2 оказывают влияние на ξ. Это же противоречит независимости.
Эти рассуждения при желании, видимо, можно формализовать.
3. Частный случай 4 свойства.
Остальные два – без доказательства.

◀

Теперь докажем, что g^* – условное матожидание ξ будет оптимальной функцией. Для этого надо доказать, что расстояние до произвольной функции хотя бы расстояние до g^*:

$$E(\eta - g(\xi))^2 = E(\eta - g^* + g^* - g)^2 = E(\eta - g^*)^2 + \underbrace{E(g - g^*)^2}_{\geq 0} + 2E(\eta - g^*)(g^* - g) \overset{?}{\geq} E(\eta - g^*)^2$$

По 3 свойству:

$$E(\eta - g^*)(g^* - g) = E(E(\eta - g^*)(g^* - g) \mid \mathbb{F}_\xi) =$$

По свойству 5 (потому что вторая скобка $\mathbb{F}_\xi$-измерима, так как обе функции зависят от ξ)

$$= E(g^* - g)E(\eta - g^*(\xi) \mid \mathbb{F}_\xi)$$

Правый множитель равен (по 1 свойству)

$$E(\eta \mid \mathbb{F}_\xi) - E(g^*(\xi) \mid \mathbb{F}_\xi) = g^*(\xi) - g^*(\xi) = 0$$

2.6. Характеристические производящие функции

𝔇𝔢𝔣 2.6.1. Характеристическая функция $f_\xi(t) = Ee^{it\xi} = \int\limits_{-\infty}^{\infty} e^{itx}\mathrm{d}F_\xi(x)$.

Для дискретных: $= \sum\limits_k e^{itx_k}p_k$

Для абсолютно непрерывных: $\int\limits_{-\infty}^{\infty} e^{itx}p_\xi(x)\mathrm{d}x$. (А еще это называется преобразованием Фурье для функции p_ξ)

$f_\xi(t) \colon \mathbb{R} \to \mathbb{C}$.

Замечание 2.6.1. Взяли плотность, растянули ее в t раз, после чего свернули в единичную окружность. Потом усреднили = взяли центр масс

𝔇𝔢𝔣 2.6.2. Если ξ – неотрицательная дискретная целочисленная случайная величина, то у нее есть производящая функция $\varphi_\xi(t) = Et^\xi = \sum p_k t^k$

Замечание 2.6.2. $f_\xi(t) = \varphi_\xi(e^{it})$. Свойства почти одинаковые, производящая функция удобна только тем, что она вещественная. В общем случае же пользуемся характеристической.

Основное свойство: если ξ, η независимы, то $f_{\xi+\eta}(t) = Ee^{it(\xi+\eta)} = f_\xi(t) f_\eta(t)$. Аналогично для производящих.

$f_{a\xi+b}(t) = Ee^{it(a\xi+b)} = e^{itb} Ee^{ita\xi} = e^{itb} f_\xi(ta)$

Пример 2.6.1. $\xi = C$, $f_\xi(t) = e^{itc}$

Пример 2.6.2. $\xi = \begin{cases} 1, p \\ 0, q \end{cases}$, $f_\xi(t) = pe^{it1} + q^{eit0} = p^{it} + q$; $\varphi(t) = pt + q$

Пример 2.6.3. $\xi \sim Bin(n, p)$. $\xi = \sum\limits_{k=1}^{n} \mathbb{1}_{A_k}$, $f_\xi(t) = (pe^{it} + q)^n$

$\xi \sim N(0, 1)$ – нормальное распределение. Тогда (упражнение): $f_\xi(t) = e^{-\frac{t^2}{2}}$.

$\xi \sim N(a, \sigma^2)$, $\xi = \sigma\tilde{\xi} + a$, где $\tilde{\xi} \sim N(0, 1)$. Откуда $f_\xi(t) = e^{ita - \frac{t^2\sigma^2}{2}}$.

Теорема 2.6.1. (Формулы обращения)

1. $a < b, F_\xi$ непрерывно в точках a, b. Тогда $F(b) - F(a) = \frac{1}{2\pi} v.p. \int\limits_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} f_\xi(t) \mathrm{d}t$, ($v.p.$ – главное значение интеграла)

2. Если $\int\limits_{\mathbb{R}} |f_\xi(t)| \mathrm{d}t < +\infty$, то ξ – абсолютно непрерывна и $p_\xi(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx} f_\xi(t) \mathrm{d}t$

▶ Без доказательства. Вообще – должно доказываться в матанализе как существование обратного преобразования Фурье ◀

Поэтому характеристической функции достаточно, чтобы описать распределение.

Теорема 2.6.2 (Свойства характеристической функции)**.** Везде ниже $f = f_\xi$.

1. $|f(t)| \le f(0) = 1$

2. f равномерно непрерывна на $\mathbb{R}$

3. $f(-t) = \overline{f(t)}$

4. $\forall t f(t) \in \mathbb{R} \Leftrightarrow \forall B \in \mathbb{B} P_\xi(B) = P_{-\xi}(B)$

5. Если $E|\xi|^n < \infty$, то $\forall k \le n \exists f^{(k)}(t) = \int (ix)^k e^{itx} dF_\xi(x)$. $f^{(k)}(0) = i^k E\xi^k$

6. Если $\exists f^{(2n)}(0)$, то $E|\xi|^{2n} < \infty$

7. $f(t) = \sum\limits_{0}^{n} \frac{i^k E\xi^k}{k!} t^k + o(t^n)$. Если радиус сходимости положителен ($\overline{\lim} \frac{(E|\xi|^n)^{\frac{1}{n}}}{n} = \frac{1}{R} < \infty$, то $f(t) = \sum\limits_{k=0}^{\infty} ...$ при $|t| < R$.

   Отсюда, в частности, следует теорема про моменты, когда они определяют распределение.

▶ 1 и 3 свойства очевидно следуют из определения.

2:

$$|f(t+h)-f(t)| = |\int (e^{i(t+h)}x - e^{itx})\mathrm{d}F| \le \int \underbrace{|e^{itx}|}_{=1} |e^{ixh}-1|\mathrm{d}F =$$

$$\int\limits_{-T}^{T} + \int\limits_{\mathbb{R}\setminus[-T,T]} |e^{ixh}-1|\mathrm{d}F(x)$$

Возьмем T такое, что $P(R\setminus[-T,T])<\varepsilon$, и $h\colon \max(|e^{ihT}-1|, |e^{ih-T}-1|)<\varepsilon$

$$|e^{ihx}-1|\le 2 \Rightarrow \int\limits_{\mathbb{R}\setminus[-T,T]} + \int\limits_{-T}^{T} \le 2\varepsilon+\varepsilon = 3\varepsilon$$

Мы взяли настолько маленькое h, что получили какую-то маленькую окрестность 1 на окружности, на обоих концах у нас модуль разности мелкий, тогда интеграл тоже будет мелким

4: $f(t)=\int \cos tx \mathrm{d}F(t) + i\int \sin tx \mathrm{d}F(t)$. Второе слагаемое равно 0 для симметричных распределений (синус нечетный).

В обратную сторону: $f_\xi(-t)=\overline{f(t)}=f(t)$. С другой стороны, минус можно от t перекинуть к ξ и получить $f_\xi(-t)=f_{-\xi}(t)$ ◀

Теорема 2.6.3 (Бохнер, Хинчин)**.** $f\in C(\mathbb{R}), f(0)=1$

Тогда f – характеристическая функция $\Leftrightarrow$ f неотрицательно определена, то есть, для любых $t_1,\dots,t_n\in\mathbb{R}$ и $c_1,\dots,c_n\in R$: $\sum\limits_{j,l=1}^{n} c_j\bar{c}_l f(t_j-t_l)\ge 0$

Теорема 2.6.4 (Пойа)**.** Если $f\in\mathbb{R}$ четная, $f(0)=1$, непрерывная, $f\xrightarrow[t\to\infty]{}0$, f выпуклая вниз на $[0,+\infty)$, то f – характеристическая функция.

Задача: (сумма случайного числа независимых случайных величин):

Пусть θ – случайная величина с носителем $\mathbb{Z}_+$, ξ_k – iid (сокращение для независимо и одинаково распределенных величин).

$\eta=\xi_1+\xi_2+\dots+\xi_\theta=?$

$$f_\eta(t)=\int e^{itx}\mathrm{d}P(\eta<x)=\int e^{itx}\mathrm{d}(\sum_k P(\theta=k)P(\eta<x\mid\theta=k))=$$

$$=\sum_k P(\theta=k)\int e^{itx}\mathrm{d}\underbrace{F_{\xi_1+\dots+\xi_k}(x)}_{f_{\xi+1+\dots+\xi+k}(t)=f_x i_1^k(t)} =$$

$$=\sum P(\theta=k)f_{\xi_1}^k(t)=\varphi_\theta(f_{\xi_1}(t))$$

Пример 2.6.4. Двое кидают монетку: первый человек кидает ее n раз, второй – столько раз, сколько гербов выпало у первого. Надо найти распределение количества гербов у второго человека.

Формально: $\xi\sim Bin(n,p_1)$, $(\eta\mid\xi=k)\sim Bin(k,p_2)$, $\eta\sim ?$

Воспользуемся формулкой:

$$f_\eta(t)=(p_1(\dots)+q_1)^n=$$
$$\varphi=pt+q, f=pe^{it}+q$$
$$=(p_1p_2e^{it}+1-p_1p_2)^n$$

Получили, внезапно, характеристическую функцию $Bin(n,p_1p_2)$.

Можно то же самое посчитать и через формулу полной вероятности, но там вычислений будет прилично больше.

Глава 3

Предельные теоремы ТВ

3.1. Виды сходимостей последовательностей случайных величин

$\mathfrak{Def}$ **3.1.1.** $\xi_n \xrightarrow[n\to\infty]{P} \xi$ (сходимость по вероятности), если $P(|\xi_n - \xi| > \varepsilon) \xrightarrow[n\to\infty]{} 0$ для любого $\varepsilon > 0$

$\xi_n \xrightarrow{\text{п.в.}} \xi$, если $P(\{\omega\colon \xi_n(\omega) \to \xi(\omega)\} \to 1$

$\xi_n \xrightarrow{L^p} \xi$ (в среднем порядка p), если $E|\xi_n - \xi|^p \to 0$

$\xi_n \xrightarrow{d} \xi$ (по распределению "слабо"), если $F_{\xi_n}(x) \to F_\xi(x)$ в точках непрерывности предела

Замечание 3.1.1. В последнем определении нам нигде не важно какая именно случайная величина, важно лишь ее распределение (это разные понятия, монетка и "число на кубике по модулю 2" имеют одинаковые распределения, но как величины отличаются, потому что разные множества событий).

Поэтому, часто пишут что-то вида $\xi_n \xrightarrow{d} N(0,1)$, подразумевая, что в правой части стоит какая-то (не важно какая именно) случайная величина с нормальным распределением.

Замечание 3.1.2. Часть определения про точки разрыва является существенной. Например, можно рассмотреть следующие три последовательности случайных величин:

1. $\xi \sim U[0, \frac{1}{n}]$, устремляем n к бесконечности, ξ устремляется к константе.
2. $\xi \sim U[-\frac{1}{n}, 0]$, ξ стремится к той же константе, только значение функции распределения в нуле равно единице (непрерывность не с той стороны получается).

   TODO картинки

3. $\xi \sim U[-\frac{1}{n}, \frac{1}{n}]$, тогда в нуле функция распределения будет равна $\frac{1}{2}$.

Поэтому и нужна оговорка про точки непрерывности предела: вроде все эти величины стремятся к одному и тому же, а вот в точках разрыва какая-то беда.

Упражнение: $\xi_n \xrightarrow{\text{п.в.}} [] \Leftrightarrow \forall \varepsilon > 0\colon P(\sup\limits_{k\geq n} |\xi_k - \xi| > \varepsilon) \xrightarrow{n\to\infty} 0$

И еще упражнение: из п.в. сходимости следует P-сходимость, из L^p – тоже, а из P-сходимости следует d.

Пример 3.1.1. Возьмем пространство событий $\Omega = [0,1]$ с равномерным распределением на нем.

Рассмотрим последовательность следующих случайных величинин: $\xi_1 = \mathbb{1}_{[0,\frac{1}{2}]}, \xi_2 = \mathbb{1}_{[\frac{1}{2},1]}, \xi_3 = \mathbb{1}_{[0,\frac{1}{4}]}, \xi_4 = \mathbb{1}_{[\frac{1}{4},\frac{1}{2}]}$ и так далее: берем все половинки, потом все четверти, и так далее.

Тогда $\xi_n \xrightarrow{P, L^p} 0$ (просто по определению, там получаются просто площади под графиком, которые сходятся к 0)

Но п.в. сходимости – нет, потому что $\xi_n(\omega)$ не стремится к нулю для любой точки ω: каждая точка берет и встречается бесконечное количество раз.

Пример 3.1.2. $\Omega = [0,1]$.

$\xi_n(x) = \mathbb{1}_{[0,\frac{1}{n}]}(x) \cdot (\frac{1}{n} - x)n$: равномерно убывает на отрезке $[0, \frac{1}{n}]$ от n до 0, справа – 0.

$E|\xi_n - 0| = \frac{1}{2}$, не стремится к 0, L^p сходимости опять нет.

Поэтому следствия из упражнения не всегда обращаются.

Теорема 3.1.1. (неравенства Чебышева)

1. $\xi \geq 0$ почти всегда, $E\xi < \infty$. Тогда $\forall \varepsilon > 0$: $P(\xi \geq \varepsilon) \leq \frac{E\xi}{\varepsilon}$
2. $P(|\xi - E\xi| \geq \varepsilon) \leq \frac{D\xi}{\varepsilon^2}$

▶ 1. $E\xi = \int\limits_0^\infty x \mathrm{d}F_\xi(x) \geq \int\limits_\varepsilon^\infty \geq \int\limits_\varepsilon^\infty \varepsilon \mathrm{d}F_\xi(x) = \varepsilon(1 - F_\xi(\varepsilon)) = \varepsilon P(\xi \geq \varepsilon)$

2. Применим 1 к $(\xi - E\xi)^2$

◀

Теорема 3.1.2 (неравенство Колмогорова)**.** ξ_k – независимые. Тогда

$$P(\sup_{1\leq k\leq n} |\sum_{j=1}^k (\xi_d - E\xi_d)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{j=1}^n D_{\xi_j}$$

Теорема 3.1.3. (Следствия из теоремы Хелли)

1. $\xi_n \xrightarrow{d} \xi$, $F(x)$ – функция распределения ξ, то тогда $\forall t \in R$: $f_{\xi_n}(t) \to f_\xi(t)$
2. $\forall t \in \mathbb{R}$: $f_{\xi_n}(t) \xrightarrow[n\to\infty]{} f(t)$ и $f(t)$ непрерывна в нуле, то она характеристическая и $\xi_n \xrightarrow{d} \xi$, где $F(\xi)$ – характеристическая функция ξ, соответствующая $f(t)$

3.2. Закон больших чисел

$\mathfrak{Def}$ **3.2.1.** Говорят, что для $\{f_n\}$ верен ЗБЧ, если $\eta = \frac{1}{n}\sum\limits_{k=1}^n (\xi_k - E\xi_k) \xrightarrow{P} 0$.

Теорема 3.2.1 (Марков)**.** Если $\frac{1}{n^2} D(\sum \xi_k) \to 0$ (условие Маркова), то ЗБЧ.

▶ $P(\frac{1}{n}|\sum(\xi_k - E\xi_k)| \geq \varepsilon) \leq \frac{1}{n^2\varepsilon^2} D(\sum \xi_k) \to 0$ (применили неравенство Чебышева) ◀

Теорема 3.2.2 (Чебышев)**.** ξ_n – независимые и $D\xi_n \leq C \Rightarrow$ ЗБЧ.

▶ В условии Маркова $\frac{nC}{n^2} = \frac{C}{n} \to 0$. ◀

Теорема 3.2.3 (Хинчин)**.** ξ_k – iid, $E\xi_1 = a, D\xi_1 = \sigma^2 < \infty$, тогда ЗБЧ: $\frac{1}{n}\sum \xi_k \to a$.

Теорема 3.2.4 (Бернулли)**.** $\xi_k = \begin{cases} 1, p \\ 0, q = 1 - p \end{cases}$, $\frac{1}{n}\sum \xi_k \to E\xi_1 = p$.

Доказательство теоремы Хинчина без конечности $D\xi$:

▶ $f_{\frac{1}{n}(\xi_1+\dots+\xi_n)} = f_{\xi_1}(t) = f_{\xi_1}^n(t/n) = (1 + \frac{iat}{n} + o(t/n))^n \xrightarrow[n\to\infty]{} e^{iat}$ – характеристическая функция для a ◀

Теорема 3.2.5 (ЗБЧ для слабозависимых)**.** Пусть $|\sigma_{ij}| = |cov(\xi_i, \xi_j)| \le r(|i-j|), r(n) \xrightarrow[n\to\infty]{} 0$. Тогда ЗБЧ.

▶

$$\frac{1}{n^2}D(\sum_1^n \xi_i) = \frac{1}{n^2}\sum_{i,j=1}^n \sigma_{ij} =$$

. Тогда

$$\exists m\colon \forall n > m\colon r(n) < \varepsilon$$

. Зафиксируем такое m, а n будем увеличивать. Тогда все разделится на константную сумму и сумму малых величин:

$$= \frac{1}{n^2}(\sum_{|i-j|\le m} + \sum_{|i-j|>m}) < \frac{1}{n^2}(C(2m+1)n + \varepsilon n^2) \to \varepsilon$$

◀

$\mathfrak{Def}$ **3.2.2.** УЗБЧ (усиленный ЗБЧ): как ЗБЧ, только $\eta_n \xrightarrow{\text{п. в.}} 0$

Теорема 3.2.6 (Колмогорова)**.** Есть последовательность $\{\xi_n\}$ независимых случайных величин и $\sum \frac{D\xi_n}{n^2} < \infty$, то УЗБЧ

▶ Доказательство есть, но нам его не расскажут. ◀

Замечание 3.2.1. В теореме Хинчина и Бернулли есть УЗБЧ.

Пример 3.2.1. ЗБЧ можно применять даже к задачам, где нет никаких вероятностей! Например, для доказательства теоремы Вейерштрасса: $\forall f \in C[0,1]\colon \exists B_n(t) \rightrightarrows f(t)$, $B_n(t)$ – полиномы.

Пусть $B_n(t) = \sum\limits_{k=0}^n f(\frac{k}{n})\binom{n}{k}t^k(1-t)^{n-k}$ - полином Бернштейна.

Поскольку

$$\sum_{k=0}^n \binom{n}{k} t^k(1-t)^{n-k} = 1$$

То

$$f(t) = \sum_{k=0}^n f(t)\binom{n}{k}t^k(1-t)^{n-k}$$

Тогда (обозначив $P_{n,t}(k) = \binom{n}{k}t^k(1-t)^{n-k}$)

$$|f(t) - B_n(t)| = \sum_{k=0}^n |f(\frac{k}{n}) - f(t)|P_{n,t}(k) = \sum_{|\frac{k}{n}-t|<\delta} + \sum_{|\frac{k}{n}-t|>\delta} <$$

δ взяли такое, что $|x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$

f равномерно непрерывна на отрезке, поэтому сверху ограничена какой-то константой C, тогда

$$< \varepsilon + 2C \sum_{|\frac{k}{n}-t|>\delta} P_{n,t}(k)$$

Во втором слагаемом, на самом деле, записано, что разница количеств орлов и решек хотя бы δ при бросании монетки n раз, по ЗБЧ ее можно сделать меньше ε.

Замечание 3.2.2. Тут мы получили не равномерную непрерывность, так как у нас ε зависит от t, мы с самого начала зафиксировали t. На консультации фикса не было предложено, но вроде это утверждение верно.

Пример 3.2.2. Метод Монте-Карло: просто меняем неизвестную вероятность события на его частоту: взяли серию независимых экспериментов и посмотрели, сколько раз произошло в них событие.

Например, можно приближенно считать интегралы по области: вместо $\int\limits_A h(t)\mathrm{d}t$ считаем

$$\frac{1}{N}\sum_{k=1}^{N} h(t_k) \xrightarrow[N\to\infty]{\text{по ЗБЧ}} \frac{\int\limits_A h}{V(A)}$$

Профит.

3.3. Центральная предельная теорема

Напоминание: ЗБЧ это $\eta_n = \frac{\sum(\xi_k - E\xi_k)}{n} \to 0$

$\mathfrak{Def}$ **3.3.1.** Для $\{\xi_n\}$ верна ЦПТ, если $\eta_n = \frac{\sum(\xi_k - E\xi_k)}{\sqrt{D\sum\xi_k}} \to N(0,1)$

Если $E\xi_k = a, D\xi_k = \sigma^2$ и $\sigma_{ij} = 0$, то $\eta_n = \frac{\sum\xi_k - na}{\sigma\sqrt{n}}$

Обозначим за S_n сумму первых n ξ_k.

Теорема 3.3.1 (Хинчин)**.** ξ_k - iid с $E\xi_k = a, D\xi_k = \sigma^2$. Тогда ЦПТ: $\frac{S_n - na}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} N(0,1)$

▶ Характеристическая функция $f_{\xi_1 - a}(t) = 1 + it\cdot 0 + \frac{(it)^2}{2!}E(\xi_1 - a)^2 + o(t^2) = 1 - \frac{t^2\sigma^2}{2} + o(t^2)$. (Это, если что, мы применили 7 свойство характеристических функций, что их можно раскладывать в такой вот степенной ряд с матожиданиями)

$f_{\frac{S_n - na}{\sigma\sqrt{n}}}(t) = f^n_{\xi_1 - a}(\frac{t}{\sigma\sqrt{n}}) = (1 - \frac{t^2}{2n} + o(\frac{t^2}{n}))^n \xrightarrow[n\to\infty]{} e^{-\frac{t^2}{2}}$ – характеристическая функция $N(0,1)$ ◀

Упражнение: $\frac{S_n - na}{\sigma n^\alpha} \to 0$ при $\alpha > \frac{1}{2}$ и нет предела при $\alpha < \frac{1}{2}$.

Замечание 3.3.1. Интегральная теорема Муавра-Лапласа, которая у нас была в начале, это частный случай ЦПТ для монетки.

Теорема 3.3.2 (Линдеберг, Леви)**.** ξ_n независимы, $a_k = E\xi_k, \sigma_k^2 = D\xi_k$, обозначим $D_n^2 = \sum\sigma_k^2$.

Если $\forall\varepsilon > 0$: $\frac{1}{D_n^2}\sum\limits_{k=1}^{n} \underbrace{\int\limits_{|\frac{x-a_k}{D_n}|\geq\varepsilon} (x-a_k)^2\mathrm{d}F_{\xi_k}(x)}_{\text{Если бы интеграл был бы по всем числам, было бы определение дисперсии}} \xrightarrow[n\to\infty]{} 0$, то ЦПТ.

Это условие называется условием Линдеберга, иногда будет сокращаться как “если (L)”

▶ Опустим доказательство этого замечательного факта. ◀

Теорема 3.3.3 (Ляпунов)**.** ξ_n – независимы и существует $\delta > 0$, что $\frac{1}{D_n^{2+\delta}}\sum\limits_1^n E|\xi_k - a_k|^{2+\delta} \to 0$, то ЦПТ.

▶ $\sum\limits_1^n \int\limits_{\mathbb{R}} |x - a_k|^{2+\delta}\mathrm{d}F_{\xi_k}(x) \geq \sum \int\limits_{|x-a_k|\geq\varepsilon D_n} \geq \varepsilon^\delta D_n^\delta \int\limits_{|x-a^k|\geq\varepsilon D} (x-a_k)^2\mathrm{d}F_{\xi_k}(x)$

Делим на $D_n^{2+\delta}$ и устремляем n к бесконечности ◀

Последние две теоремы можно перевести на язык схемы серий: $\xi_{n,k} = \frac{\xi_k - a_k}{D_n}, S_n = \xi_{n,1} + \dots + \xi_{n,n}$.

При таких условиях, $E\xi_{n,k} = 0 = ES_n, DS_n = 1$

Теорема 3.3.4 (Теорема Линдеберга на языке серий)**.** (L)($\sum\limits_1^n E\xi_{n,k}^2\mathbb{1}_{|\xi_{n,k}|>\varepsilon} \to 0 \Rightarrow S_n \xrightarrow{d} N(0,1)$.

$\max\limits_{1\leq k\leq n} E\xi_{n,k}^2 = \max(E\xi_{n,k}^2\mathbb{1}^2_{|\xi_{n,k}\leq\varepsilon} + E\xi_{n,k}^2\mathbb{1}_{|\xi_{n,k}>\varepsilon})$.

Первое слагаемое не больше ε^2, второе стремится к 0 по условию Лендеберга.

Теорема 3.3.5 (Теорема Ляпунова на языке серий)**.** Если $\max E\xi_{n,k}^2 \to 0$, то (L) $\Leftrightarrow$ ЦПТ.

Докажем теорему Пуассона (к ЦПТ отношения не имеет), в чуть более слабой форме

Теорема 3.3.6. Если $np \xrightarrow[n\to\infty]{} \lambda$, то $P_{n,p}(k) \xrightarrow[n\to\infty]{} \pi_k = e^{-\lambda}\frac{\lambda^k}{k!}$.

Тут p может меняться от серии к серии.

▶ $f_{s_n}(t) = (p_n e^{it} + (1-p_n))^n = (1 + (e^{it}-1)\frac{np_n}{n})^n \to e^{\lambda(e^{it}-1)}$

Получили характеристическую функцию Пуассона. ◀

Теорема 3.3.7 (Неравенство Берри-Эссена)**.** $\xi_n - -iid$. Тогда $\sup|F_{\xi_n} - \Phi(x)| \le C\frac{E|\xi_1-a|}{\sigma^3\sqrt{n}}$, где $\Phi \sim N(0,1)$.

3.3.1. Метод обратной функции

Пример 3.3.1. Основной метод для моделирования – метод обратной функции, который уже разбирался ранее.

Его можно обобщить на многомерный случай: есть распределние в $\mathbb{R}^n$, хотим построить $\vec{\xi}$ с этим распределением.

Рассмотрим для этого последовательность функций

$$F_1(x_1) = F_{\xi_1}(x_1)$$
$$F_2(x_1, x_2) = F_{\xi_2}(x_2 \mid \xi_1 = x_1)$$
$$\dots$$
$$F_n(x_1, \dots, x_n) = F_{\xi_n}(x_n \mid \xi_1 = x_1, \dots, \xi_{n-1} = x_{n-1})$$

И последовательность равномерных независимых величин $\eta_k \sim U[0,1]$.

Тогда ξ – решение следующей системы: $\begin{cases} F_1(\xi_1) = \eta_1 \\ F_2(\xi_1, \xi_2) = \eta_2 \\ \dots \\ F_n(\xi_1, \dots, \xi_n) = \eta_n \end{cases}$

Доказательство аналогично одномерному случаю, но возни больше. Можно попробовать доказать до двух, а потом применить индукцию.

Пример 3.3.2. Есть монетка, хотим выбрать с ее помощью одного человека из 10. Проблема в том, что умеем получать равновероятно числа до степени двойки, а по-другому – не особо.

Можно завести 6 “фейков”, получится 16 человек, бросаем монетку 4 раза. Если попали в фейк, то еще раз пробуем, иначе заканчиваем процесс.

Аналогично, можно получить конечные рациональные числа.

Пример 3.3.3. Похожая идея в методе Монте-Карло: хотим посчитать площадь фигуры, для этого моделируем распределение на каком-то прямоугольнике, в котором лежит наша фигура (так как саму фигуру замоделировать сложно).

Пример 3.3.4. Моделирование смесей.

$F(x)\sum\limits_k C_k F_k(x), \sum C_k = 1$.

Предполагаем, что умеем моделировать случайные величины для F_k.

Замоделируем $\theta\colon P(\theta = k) = C_k$ – какая-то дискретная величина.

Ну а теперь бросаем θ, если выпало k, то генерим ξ_k и возвращаем его. Простой подстановкой получаем, что сгенерили именно нужное распределение.

3.4. Еще предельные теоремы

Законы “0 или 1”.

Одну теорему уже знаем – теорема Бореля-Кантелли о том, что вероятность того, что произойдет бесконечно много событий из какой-то последовательности, равна либо 0 либо 1.

Еще теоремка:

Теорема 3.4.1. Есть независимые ξ_n, $\mathcal{F}_n$ – σ-алгера событий, определяемых значениями $(\xi_n, \xi_{n+1}, \dots)$. Хвостовой σ-алгеброй назовем $\mathcal{X} = \bigcap\limits_{n=1}^{\infty} \mathcal{F}_n$

Иначе говоря, $\mathcal{F}_n$ – события, про которые можно понять, исходя из значений ξ_n

Тогда $\forall A \in \mathcal{X}\colon P(A) = \begin{cases} 0 \\ 1 \end{cases}$

Пример 3.4.1. В частности, ряд из независимых случайных величин либо почти всегда сходится, либо почти всегда расходится.

Теорема 3.4.2 (Закон повторного логарифма)**.** ξ_n – iid с $E\xi_1 = 0, D\xi_1 = \sigma^2, S_n = \xi_1 + \dots + \xi_n$.

Тогда $P(\overline{\lim\limits_{n\to\infty}} \frac{S_n}{\sqrt{2\sigma^2 n \ln\ln n}} = 1) = 1$.

Замечание 3.4.1. Это значит, что если мы “ограничим” себя значением $(1-\varepsilon)$ на этот корень, то мы границу пересечем бесконечное число раз. Если ограничим значением $(1+\varepsilon)$ на корень, то с каког-то момента границу пересекать не будем.

Поэтому, вот эту вот оценку называют точной верхней границей для случайного блуждания.

Теорема 3.4.3 (Закон арксинуса)**.** ξ_k – независимые случайные величины $= \begin{cases} 1, \frac{1}{2} \\ 0, \frac{1}{2} \end{cases}$, $S_n = \sum\limits_1^n \xi_k$ – стандартное случайное блуждание.

$\nu_n = \frac{1}{n}|\{k \le n \mid S_k > 0\}|$ – доля времени, когда лидируют гербы в серии бросков монетки.

Тогда $F_{\nu_n}(x) \xrightarrow[n\to\infty]{} \frac{2}{\pi}\arcsin\sqrt{x}, 0 \le x \le 1$ и $p_{\nu_n}(x) \xrightarrow[n\to\infty]{} \frac{1}{\pi\sqrt{x(1-x)}}$

Теорема 3.4.4 (Эрдеш, Реньи)**.** Пусть есть n бросков честной монетки, ξ_n – длина самой длинной серии гербов. Тогда $P(\xi_n \ge \log_2 n) \xrightarrow[n\to\infty]{} 1$

Замечание 3.4.2. Это еще одна точная оценка, можно как в законе повторного логарифма домножить на $(1 \pm \varepsilon)$

3.5. Еще пара результатов

𝔇𝔢𝔣 3.5.1. $\vec{\xi} = N(\vec{a}, B)$, если $p_\xi(x) = \frac{\sqrt{\det A}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(x-a)^T A (x-a)}$, где $\vec{a}$ – вектор средних, B – матрица ковариаций, $A = B^{-1}$.

Характеристическая функция будет равна $f_\xi(t) = e^{i(t,a) - \frac{1}{2}(Bt,t)}$

Свойства:

1. Определяется 1 и 2 моментами (это и есть, собственно, вектор средних и матрица ковариаций)
2. Некоррелируемость равносильна независимости (некоррелируемость = матрица ковариаций диагональная, обратная тоже, экспонента распадется в произведение экспонент)

Пример 3.5.1. В двумерном случае, положительно определенная матрица ковариаций задаст эллипс

Стандартное нормальное распределение: $N(0, E)$.

Упражнение: пусть $\xi_1, \xi_2 \sim N(0,1)$, независимы. $\begin{cases} \xi_1 = \rho\cos\varphi \\ \xi_2 = \rho\cos\varphi \end{cases}$. Тогда

1. $D(\rho^2 \mid$ точка на биссектрисе $x = y) = D(\xi_1^2 + \xi_2^2 \mid \xi_1 = \xi_2) = D(2\xi_1^2) = 4D(\xi_1^2)$

2. $D(\rho^2 \mid \varphi = \frac{\pi}{4}$ или $\varphi = \frac{5\pi}{4})$. Так как ρ и φ независимы, то $= D\rho^2 = D(\xi_1^2 + \xi_2^2) = 2D(\xi_1^2)$.

Вопрос: как так?

Задача: есть забег, все бегуны равны по силе (имеют одинаковое распределение времени, когда финишируют). Хотим понять какое время будет у золота, серебра, бронзы, и так далее. Эта задача называется задачей распределения порядковых статистик.

ξ_k – независимы, одинаково распределены, есть непрерывная функция распределения $F(x)$

Упорядочим их по возрастанию, новый порядок будем индексировать со скобочками: $\xi_{(1)} \leq \xi_{(2)} \leq ... \leq \xi_{(n)}$.

$\xi_{(k)}$ называется k-й порядковой статистикой.

$$F_{\xi_{(k)}}(x) = P(\text{по крайней мере } k \text{ штук } \xi_j \text{ меньше } x) = \sum_{j=k}^{n} P(\text{ровно } j \text{ с.в., меньших } x) =$$
$$= \sum_{j=k}^{n} \binom{n}{j} F(x)(1 - F(x))^{n-j}$$

В частности, $F_{\xi_{(n)}} = F^n(x); F_{\xi_{(1)}}(x) = 1 - (1 - F(x))^n$.

А если есть производная, то $F'_{\xi_{(k)}}(x) = k\binom{n}{k}F^{k-1}(x)(1 - F(x))^{n-k}$. Просто если это честно посчитать, то по цепочке все сократится.

Глава 4

Случайные процессы

$\mathfrak{Def}$ **4.0.2.** T – "время", Ω – исходы.

$\xi(t, w) = \xi_t(w)\colon T \times \Omega \to \mathbb{R}$ такая, что $\forall t \in T\colon \xi_t\colon \Omega \to \mathbb{R}$ являтся случайной величиной.

Тогда ξ – случайный процесс.

По T:

1. T – конечно, тогда это – случайные векторы.

2. T – счетно, тогда это процесс с дискретным времем.

3. $T = \mathbb{R}^n, [a, b]$ и т.п., тогда это процесс с непрерывным временем. E

По значениям:

1. процессы с дискретным набором состояний (существует не более чем счетное $B\colon P(\xi_t \in B) = 1 \forall t \in T$)

2. процессы с непрерывным множеством состояний.

$\xi(\cdot, w)$ как функция $T \to \mathbb{R}$ называется реализацией (траекторией).

Пример 4.0.2. Например, температура на улице – случайный процесс. Но если мы посмотрим температуру в прошлом, то это будет ее траектория, так как она уже не случайная.

$\mathfrak{Def}$ **4.0.3.** Пусть $t_1 < t_2 < \dots < t_n$. $(\xi_{t_1}, \dots, \xi_{t_n})$ – случайный вектор, у него есть какое-то распределение.

Тогда $P_\xi = \{P_{(\xi_{t_1}, \dots, \xi_{t_n})}, n \in \mathbb{N}, (t_1, \dots, t_n) \in T^n\}$ – семейство конечномерных распределений.

Пример 4.0.3. Могут быть процесы, которые имеют одинаковые семейства конечномерных распределений, но сами по себе будут разными, например, их траектории будут сильно различаться.

$T = [0, 1]$, $\Theta \sim U[0, 1]$, $\xi_t = 0, \eta_t = \mathbb{1}_{(\Theta(\omega) = t)}$

Семейства для ξ_t и η_t совпадают (для любого конечного набора T вероятность того, что попадем в этот конечный набор, равна нулю), но траектории различаются: супремум значений у ξ равен нулю, у η равен единице.

$\mathfrak{Def}$ **4.0.4.** Пишут $\xi_t \xrightarrow{d} \eta_t$, если их семейства распределений равны.

А когда выполняется $\forall t P(\xi_t = \eta_t) = 1$ (это несколько более сильное утверждение), то говорят, что ξ и η являются "модификациями" друг друга.

Есть аналог теоремы (Колмогорова?), который говорит, что если есть семейство согласованных конечномерных распределений (если выкидываем t_1, то все равно получаем вектор из множества), то тогда можно построить процесс по этому семейству. **TODO**

Множества T могут быть самыми разными и совершненно не обязательно соответствовать типичному пониманию времени:

Пример 4.0.4. Высота волны. t – время, ω – случайные координаты. А можно и наоборот: t – координата, ω – время.

Пример 4.0.5. Аквариум. Набор моментов времени T – набор всех возможных Борелевских подмножеств аквариума: $T = \mathbb{B}$, ω – момент времени, $\xi_t(\omega)$ = число рыбок в t в момент времени ω

Пример 4.0.6. $T = \{\xi \colon \Omega \to \mathbb{R}\}$, ξ – случайные величины. $\eta_t(\omega) = t(\omega)$. Это тоже случайный процесс.

Пример 4.0.7. $\xi(t,\omega) = A(\omega)\cos(\nu(\omega)t + \varphi(\omega))$ – колебательное движение, A, ν, φ – случайные величины. Получили опять случайный процесс.

Пример 4.0.8. Пример из матстатистики: $X_1, \dots, X_n$ – независимые случайные величины с функцией распределения $F(t)$.

$F_n(t) = \frac{1}{n}\sum\limits_{k=1}^{n} \mathbb{1}_{(X_k(\omega)<t)}$ – выборочная функция распределения.

$EF_n(t) = \frac{1}{n}\sum\limits_{1}^{n} P(X_k < t) = F(t)$

$DF_n(t) = \frac{1}{n^2}\sum F(t)(1-F(t)) = \frac{F(1-F)}{n} \to 0$

Из всего этого следует, что $F_n(t) \to F(t)$

$E\xi = \int x\mathrm{d}F(x)$, выборочное среднее $\int x\mathrm{d}F_n(x) = \frac{1}{n}\sum\limits_{1}^{n} X_k$

TODO добавить в пример побольше текста и смысла **TODO**

Типы случайных процессов:

1. С независимыми значениями: $\forall t_1, t_2, \dots : \xi_{t_1}, \xi_{t_2}, \dots$ – независимы.

2. С независимыми приращениями: $\forall t_0 < t_1 < \dots : \xi_{t_0}, \xi_{t_1-t_0}, \xi_{t_2-t_1}, \dots$ – независимы. Например, η_k – независимы, $\xi_n = \eta_1 + \dots + \eta_n$.

3. Марковские процессы – процессы, у которых выполняется Марковское свойство:

$$\forall t_1 < t_2 < \dots < t_n, \forall B \in \mathbb{B}, \forall x_1, \dots, x_{n-1} \in \mathbb{R}\colon$$
$$P(\xi_{t_n} \in B \mid \xi_{t_1} = x_1, \dots, \xi_{t_{n-1}} = x_{n-1}) = P(\xi_{t_n} \in B \mid \xi_{t_{n-1}} = x_{n-1})$$

   На словах: не важно, какая была история у процесса, важно лишь последнее состояние.

   Пример 4.0.9. Температура, например, не является Марковским процессом: как правило, она каким-либо образом зависит от тренда, от того, как она менялась в последние дни.

   Пример 4.0.10. А подбрасывание монетки – является

4. Процессы второго порядка: $\exists E\xi_t = a(t)$ – среднее, $K(t,s) = cov(\xi_t, \xi_s)$ – ковариационная функция, $R(t,s) = \frac{K(t,s)}{\sqrt{K(t,t)K(s,s)}}$ –корелляционная функция.

5. Гауссовские процессы: если все конечномерные распределения нормальные.

6. Стационарные процессы.

   В узком смысле: $\forall h \in T\colon P_{\xi_{t_1},\dots,\xi_{t_n}} = P_{\xi_{t_1+h},\dots,\xi_{t_n+h}}$.

   В широком смысле: $a(t) = a = const, K(t,s) = g(|t-s|)$

   Замечание 4.0.1. Очень сложно привести пример процесса, который был бы стационарен в широком смысле, но не в узком. Тем не менее, это не равносильные определения.

7. $t_1 < t_2$, $E(\xi_{t_2} \mid \xi_{t_1} = x) = x$ – мартингал. Если $\geq x$, то субмартингал, если $\leq x$, то супермартингал.

   Пример 4.0.11. Например, бросание монетки – мартингал, если выпало на x орлов больше, чем решек, то в будущем в среднем так и будет.

4.0.1. Пуассоновский поток

𝔇𝔢𝔣 4.0.5. Пуассоновский поток – случайный процесс со следующими свойствами:

1. $\pi_0 = 0$
2. С независимыми приращениями
3. Однородность по времени: $P(\pi_t - \pi_s = k) = P(\pi_{t-s} = k)$
4. $\pi_t \in \mathbb{Z}_+$

TODO ступенчатая картинка **TODO**

Пусть $G(t) = P(\pi_t = 0)$. Тогда $G(t+s) = P(\pi_s = 0)P(\pi_{t+s} = 0 \mid \pi_s = 0) =$ З$(\pi_s = 0)P(\pi_{t+s} - \pi_s = 0) = G(s)G(t)$. Тогда $G(t) = e^{-\lambda t}$, и интервалы между скачками независимы и распределены показательно, а отсюда $\pi_t \sim Poiss(\lambda t)$.

Таким образом, эти свойства определяют вид нашего процесса

Замечание 4.0.2. Очень популярный процесс, например, в теории массового обслуживания: обработка всяких запросов, очереди, и все такое прекрасно им описываются.

Можно сделать λ переменной от t, на маленьком приращении считать λ константой, тогда $\pi_t \sim Poiss(\int\limits_0^t \lambda(t)\mathrm{d}t$

TODO Броуновский процесс **TODO**

4.1. Марковские цепи

𝔇𝔢𝔣 4.1.1. Марковская цепь (м.ц.) – марковский процесс с дискретным временем ($T = \{0, 1, 2, \dots\}$) и не более чем счетным числом состояний.

$$P(\xi_{n+k} = j \mid \xi_k = i) = P_{i,j}(n, k)$$

зависит от 4 величин. Но мы будем рассматривать только такие процессы, которые не зависят от k (однородные по времени).

𝔇𝔢𝔣 4.1.2. $P_{i,j} = P_{i,j}(1), P = (P_{i,j})_{i,j}$ – матрица перехода.

Теорема 4.1.1. Конечномерное распределение ξ_n задается распределением ξ_0 и матрицей P.

▶ $P_{i,j}(2) = \sum\limits_k P_{i,k}P_{k,j}$ (по формуле полной вероятности). А это, как несложно увидеть, квадрат матрицы. А дальше по индукции, $P(n) = P^n$

И, например, $P(\xi_n = j) = \sum\limits_i P(\xi_0 = i)P_{i,j}(n)$ ◀

Пример 4.1.1. $\xi_n \in \mathbb{Z}$, $P(\xi_n = k) = q_k$. Тогда матрица перехода (бесконечная) этой цепи будет выглядеть вот так:

$$\begin{array}{ccccc} \ddots & \vdots & \vdots & \vdots & \iddots \\ \dots & q_{-1} & q_0 & q_1 & \dots \\ \dots & q_{-1} & q_0 & q_1 & \dots \\ \dots & q_{-1} & q_0 & q_1 & \dots \\ \iddots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Этот процесс вообще не зависит от последнего состояния, поэтому все строки одинаковы

Пример 4.1.2. А теперь рассмотрим сумму случайных величин: $S_n = \sum\limits_{k=0}^{n} \xi_k$. Этот процесс называется случайным блужданием: у нас есть точка в пространстве, каждую секунду мы ее двигаем на какой-то случайный вектор ξ_n.

Ему уже будет соответствовать такая матрица:

$$\begin{matrix} \ddots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \cdots & q_{-1} & q_0 & q_1 & q_2 & \cdots \\ \cdots & q_{-2} & q_{-1} & q_0 & q_1 & \cdots \\ \cdots & q_{-3} & q_{-2} & q_{-1} & q_0 & \cdots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix}$$

Каждый раз строка сдвигается на 1 вправо, так, чтобы на диагонали стояли вероятности остаться на месте: $P_{ii} = q_0$.

Пример 4.1.3. Очередь: q_k – вероятность того, что за единицу времени подойдет $k \geq 0$ человек. Каждую единицу времени, если очередь не пуста, то ее длина уменьшается на 1.

Состояние процесса ξ_n – длина очереди в момент времени n.

Матрица перехода: (здесь она уже бесконечна вправо и вниз, но имеет левый верхний угол)

$$\begin{matrix} q_0 & q_1 & q_2 & \cdots & \\ q_0 & q_1 & q_2 & \cdots & \\ 0 & q_0 & q_1 & q_2 & \cdots \\ 0 & 0 & q_0 & q_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix}$$

Если очередь была пуста, то мы переходим в состояние i с вероятностью q_i. Если же очередь не пуста, то ее длина точно уменьшается на 1, а потом сколько-то человек приходит. Поэтому, переходим в состояние $i-1$ с вероятностью q_i.

Пример 4.1.4. ξ_n – длина серии успехов в схеме Бернулли.

$$\begin{matrix} q & p & 0 & 0 & 0 & \ldots \\ q & 0 & p & 0 & 0 & \ldots \\ q & 0 & 0 & p & 0 & \ldots \\ q & 0 & 0 & 0 & p & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \end{matrix}$$

Пример 4.1.5. Все примеры были либо с бесконечной, либо с полубесконечной матрицей. А теперь один из самых известных примеров с конечной матрицей – модель Эренфестов.

Есть камера, разделенная мембраной на две одинаковые половины, в ней N частиц. Один такт – переход частицы слева направо или справа налево. Для каждой частицы считаем, что у нее одинаковые шансы совершить переход.

Состояние ξ_n – количество частиц слева от мембраны, $\xi_n \in \{0, 1, \ldots, N\}$.

$$\begin{matrix} 0 & 1 & 0 & 0 & \ldots & & \\ \frac{1}{N} & 0 & \frac{N-1}{N} & 0 & \ldots & & \\ 0 & \frac{2}{N} & 0 & \frac{N-2}{N} & \ldots & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & & \\ \ldots & \ldots & \ldots & \ldots & \ldots & \frac{N}{N} & 0 \end{matrix}$$

Пример 4.1.6. Ветвящиеся процессы (например, численность популяции).

Есть популяция особей, которые не мешают друг другу и нет проблемы перенаселенности, поэтому численность зависит только от скорости размножения. Единица времени = поколение (за которое старые умерли, но успели родить новое поколение).

ν – количество потомков у особи, $q_k = p(\nu = k), k = 0, 1, 2, \ldots$; ξ_n – численность популяции.

$P_{ij} = P(\nu_1 + \nu_2 + \cdots + \nu_i = j)$. Несложно понять, что это коэффициенты при t^j в $\varphi^i(t)$, где φ – производящая функция для ν

Интересен вопрос о вырождении популяции, когда $\xi_n = 0$ (и в дальнейшем, понятное дело, популяция так и останется нулевой).

Рассмотрим $z_n = P(\xi_n = 0)$. Эта величина возрастает с ростом n (так как из нуля мы не уходим, но можем туда прийти из другого состояния). Также, z_n ограничена единицей, следовательно, имеет предел z, назовем его вероятностью вырождения популяции.

Обозначим за $\varphi_n(t)$ – производящую функцию прироста за n шагов ($\varphi = \varphi_1$ это, соответственно, производящая функция для ν). Запишем формулу полной вероятности:

$$\varphi_{n+1}(t) = \varphi(\varphi_n(t)) = \underbrace{\varphi \circ \varphi \circ \dots \circ \varphi}_{n}$$

$$z_n = \varphi_n(0)$$

$$z_{n+1} = \varphi(z_n), n \to \infty \Rightarrow z = \varphi(z)$$

Оказывается, что верно следующее: если $E\nu \le 1$, то $z = 1$. Иначе существует корень (единственный) уравнения $z = \varphi(z)$, принадлежащий интервалу $(0, 1)$. И именно этот корень и является вероятностью вырождения популяции.

4.2. Классификация состояний марковской цепи

Def 4.2.1. Пишут $i \to j$ (j достижимо из i), если существует такое m, что $P_{ij}(m) > 0$.

Если $i \leftrightarrow j$, то эти состояния сообщающиеся (из i достижимо j и наоборот).

Def 4.2.2. i существенно, если $\forall j\colon i \to j \Rightarrow j \to i$. На пальцах: нельзя из состояния уйти так, чтобы потом было не вернуться обратно.

Если это не так, ($\exists j\colon i \to j, j \not\to i$), то состояние несущественно.

Все существенные состояния делятся на классы эквивалентности ($\leftrightarrow$ – отношение эквивалетности). Эти классы называются эргодическими и обычно нумеруются так: ЭК1, ЭК2, и так далее.

И еще выделяют класс несущественных состояний.

Таким образом, разбили состояния на следующие множества: множество несущественных состояний и эргодические классы. При этом, верно следующее: из несущественных состояний могут быть переходы куда угодно, а ребра из состояний эргодического класса ведут только туда же (иначе либо состояние несущественное, либо два класса можно объединить).

Можно нарисовать орграф, сжать его компоненты сильной связности. Тогда все его стоки будут соответствовать эргодическим классам, а все остальные вершины – несущественным состояниям.

Def 4.2.3. Цепь называется неприводимой, если все ее состояния в ЭК1.

Def 4.2.4. Пусть $M_j = \{n \in \mathbb{N}\colon P_{ij}(n) > 0\}$ – множество количеств шагов, за которые можно вернуться в состояние j. M_j замкнуто относительно сложения: если можно вернуться за i шагов и за j шагов, то можно и за $i + j$.

У этого набора чисел есть период $d_j = gcd(M_j)$.

Пример 4.2.1. Период блуждания коня по шахматной доске равен 2 (так как граф двудольный).

Теорема 4.2.1. Начиная с некоторого k все $kd_j \in M_j$

Замечание 4.2.1. Теорема верна для любого аддитивного класса.

▶ Пусть $n_1, \dots, n_m \in M$ и $gcd(n_1, \dots, n_m) = d$.

Тогда существуют $a_1, \dots, a_m \in \mathbb{Z}$ такие, что $d = \sum a_i n_i$.

Это равенство можно переписать:

$$d + \underbrace{\sum_{a_i<0} -a_i n_i}_{b} = \sum_{a_j>0} a_j n_j$$

Обе суммы лежат в M как суммы элементов M с положительными коэффициентами. То есть, $b, b+d \in M$. При этом, b делится на d (так как все a_i на него делятся)

Возьмем произвольное большое число $N > b^2, N \vdots d$, покажем, что $N \in M$.

Поделим N на b с остатком:

$$N = kb + r, r < b; b \vdots d, N \vdots d \Rightarrow r \vdots b$$

Откуда,

$$N = kb + ld = (k-l)b + l(b+d)$$

Знаем, что

$$b, b+d \in M, k-l, l > 0$$

Тогда эта комбинация тоже лежит в M, что и требовалось. ◀

Утверждение 4.2.1. $i \leftrightarrow j \Rightarrow d_i = d_j$

▶ $P_{ij}(m) > 0, P_{ji}(l) > 0$.

Для $n = d_j k$ верно:

$$P_{ii}(m+l+n) \geq P_{ij}(m) + P_{jj}(n) + P_{ji}(l) > 0$$

Возьмем k достаточно большое, из теоремы. Тогда, $m + l + kd_j \vdots d_i$, можно увеличить k на единицу и делимость останется. Откуда, $d_j \vdots d_i$. И наоборот. Значит, $d_i = d_j$. ◀

Замечание 4.2.2. Вроде можно доказывать и без использования теоремы, но на лекции сходу не вспомнили, как.

Если есть эргодический класс с периодом $d > 1$, то его можно разбить на подклассы $C_0, C_1, \dots, C_{d-1}$, где $C_k = \{j \colon \exists m \geq 0 \colon P_{i_0 j}(md+k) > 0\}$ для какого-то зафиксированного состояния i_0 из этого класса. По сути, разбили класс на d долей с переходами только из i-й доли в $(i+1) \mod d$-ю.

𝔇𝔢𝔣 4.2.5. Вероятность первого возвращения за n шагов $f_{jj}(n) = P(\xi_n = j), \xi_{n-1} \neq j, \dots, \xi_1 \neq j \mid \xi_0 = j)$.

При этом, полагаем $f_{jj}(0) = 0$.

Вероятностью возвращения назовем сумму $\sum\limits_{n=1}^{\infty} f_{jj}(n)$.

𝔇𝔢𝔣 4.2.6. Существенное состояние j называется возвратным, если $f_{jj} = 1$.

𝔇𝔢𝔣 4.2.7. Для возвратных состояний можно определить среднее время возвращения $\mu_j = \sum n f_{jj}(n)$.

Также вводят обратную величину $\pi_j = \frac{1}{\mu_j}$.

Если для возвратного состояния j $\pi_j > 0$, то j называют положительным состоянием. Иначе, если $\pi_j = 0$, то состояние нулевое.

Теорема 4.2.2. j возвратное $\Leftrightarrow \sum\limits_{1}^{\infty} P_{jj}(n) = \infty$

▶

$$P_{jj}(n) = \sum_{k=1}^{n} f_{jj}(k) P_{jj}(n-k)$$

Умножим это на t^n и просуммируем по n, получим

$$P(t) = \sum_{n=0}^{\infty} P_{jj}(n) t^n = P(t)F(t) + 1$$

Здесь P, F – производящие функции для P_{jj} и f_{jj}

$$P(t) = \frac{1}{1 - F(t)}$$

Если устремить $t \to 1 - 0$, то $\sum f_{jj} t^n \to 1$, если j возвратное. И наоборот, если стремится к 1, то j возвратное. ◀

Пример 4.2.2. Блуждание на прямой с вероятностью p движения на 1 вправо и вероятностью $q = 1 - p$ влево.

Тогда вероятность за $2n$ шагов вернуться в точку, откуда начали, будет равна $P_{jj}(2n) = \binom{2n}{n} p^n q^n$, что можно оценить по формуле Стирлинга: $\frac{(4pq)^n}{\sqrt{\pi n}}$.

Откуда видно, что j возвратно тогда и только тогда, когда $p = q = \frac{1}{2}$, иначе ряд, очевидно, сходится.

Пример 4.2.3. Теперь ходим по плоскости с вероятностью $p = \frac{1}{4}$ по каждому из направлений.

$P_{jj}(2n) = \frac{1}{4^{2n}} \binom{2n}{n} \underbrace{\sum_k \binom{n}{k}^2}_{\binom{2n}{n}} = \frac{\binom{2n}{n}^2}{2^{2n}} \sim \frac{1}{\pi n}$.

Этот ряд тоже очевидно расходится.

Пример 4.2.4. В общем случае, в $\mathbb{R}^d, d \geq 3$

$$P_{jj}(2n) = \frac{1}{(2d)^{2n}} \binom{2n}{n} \sum_{k_1 + \dots + k_d = n} (C_n^{k_1, \dots, k_d})^2 \leq$$

$$\leq \frac{\binom{2n}{n}}{(2d)^{2n}} \max C_n^{k_1, \dots, k_d} \underbrace{\sum C_n^{k_1, \dots, k_d}}_{d^n} \leq$$

$$\max C_n^{k_1, \dots, k_d} \leq A \max \frac{n^n \sqrt{n}}{k_1^{k_1} \dots k_d^{k_d} \sqrt{k_1 k_2 \dots k_d}} \overset{k_j = \frac{n}{d}}{=} A d^n n^{-\frac{d-1}{2}}$$

$$\leq \frac{A}{n^{\frac{d}{2}}}$$

Все блуждания для $d \geq 3$ являются невозвратными. Для $d = 3$ вероятность вернуться примерно 0.35

Общий случай случайного блуждания: $S_n = \xi_1 + \dots + \xi_n$, ξ_j – “шаги” $\in \mathbb{Z}^d$, ξ_j – iid.

Теорема 4.2.3. Случайное блуждание возврано тогда и только тогда, когда $\lim\limits_{\lambda \to 1} \int\limits_{[-\pi, \pi]} Re \frac{\mathrm{d}t}{1 - \lambda f(t)} = \infty$, где $f(t)$ – характеристическая функция $\overrightarrow{\xi_j}$

▶

$$f(t) = \sum_{\vec{k}} e^{i(t,k)} P(\vec{\xi} = \vec{k})$$

$$f^n(t) = f_{S_n}(t) = \sum e^{i(t,k)} P(S_n = k)$$

$$\int\limits_{[-\pi;\pi]^d} e^{it_1 k_1} = 0 \Rightarrow \int f^n(t) \mathrm{d}t = (2\pi)^d P(S_n = 0)$$

Заметим, что интеграл слева раскладывается в произведение интегралов по каждой координате (скаларяное произведение в экспоненте разложили в произведение экспонент). Если хотя бы одна координата у $\vec{k}$ не равна нулю, то в соответствующем интеграле будет $\sin(k_i t_i)$ и $\cos(k_i t_i)$, интеграл по ним от $-\pi$ до π даст 0.

Умножаем на λ^n и суммируем по n: $\sum \lambda^n P(S_n = 0) = \frac{1}{(2\pi)^d} \int \frac{dt}{1-\lambda f(t)} = \frac{1}{(2\pi)^d} \int Re \frac{dt}{1-\lambda f(t)}$

Слева стоит вероятность вернуться в точку старта за n шагов в виде степенного ряда. Если подставить $\lambda = 1$, то получится в точности сумма из определения возвратности.

Устремляем $\lambda \to 1 - 0$ и получаем что надо. ◀

Замечание 4.2.3. С помощью этой теоремы можно доказать, что в $\mathbb{R}^1$ и в $\mathbb{R}^2$ при $E|\xi_j|^2 < \infty$ блуждание возвратно тогда и только тогда, когда $E|\xi_j| = 0$.

При $d = 1$ надо расписать $f(t) = u(t) + iv(t)$, $u(t) = \int \cos tx \mathrm{d}F_\xi(x)$, разложить косинус по Тейлору: $= 1 - \frac{t^2}{2} E\xi_1^2 + o(t^2)$, $v(t) = \int \sin tx \mathrm{d}F_\xi(x) = tE\xi_1 + o(t^2)$. Потом надо честно расписать вещественную часть, подставить в интеграл.

В двумерном случае так же, просто чуть больше возни.

А при $d \geq 3$ любое блуждание невозвратно.

Теорема 4.2.4. j –возвратное апериодическое состояние, тогда $P_{jj}(n) \to \pi_j = \frac{1}{\mu_j}$, μ_j – среднее время возвращения в 0

Следствие 4.2.4.1. Если $i \leftrightarrow j$, то $P_{i,j}(n) \to \pi_j$.

▶ $P_{ij}(n) = \sum\limits_{k=1}^{n} f_{ij}(k) P_{jj}(n-k)$. Устремим $n \to \infty$, $\sum f_{ij}(k) = f_{ij} = 1$ ◀

Следствие 4.2.4.2. Если период $d > 1$, то рассмотрим разбиение нашего класса C на $C_0 \cup C_1 \cup \ldots \cup C_{d-1}$. Возьмем состояния $i \in C_0, j \in C_a$.

Тогда $P_{ij}(nd + a) \xrightarrow[n\to\infty]{} \frac{d}{\mu_j}$.

▶ Рассмотрим цепь с матрицей перехода $\widetilde{P} = P^d$. Пусть $a = 0$

$$P_{i,j}(nd) = \widetilde{P_{ij}}(n) \to \frac{1}{\sum k f_{ij}(kd)} = \frac{d}{\sum\limits_{k=1}^{\infty} kd f_{ij}(kd)} = \frac{d}{\mu_j}$$

Для $a = 1$:

$$P_{i,j}(nd + 1) = \sum_{k \in C_1} P_{i,k} P_{kj}(nd) \xrightarrow[n\to\infty]{} \frac{d}{\mu_j} \underbrace{\sum_k P_{i,k}}_{=1}$$

А дальше по индукции. ◀

Замечание 4.2.4. π_j можно находить, составив следующую систему: $P_{ij}(n+1) = \sum\limits_k P_{i,k}(n) P_{k,j}$, устремляем $n \to \infty$ и получаем уравнения $\pi_j = \sum P_{k,j} \pi_k$

В матричной форме: $(P^T - I)\vec{\pi} = 0$

Теорема 4.2.5. Пусть есть неприводимая цепь, ее состояния $j \in \{0, 1, 2, \ldots\}$ – возвратные, положительные и апериодичные.

Тогда $\pi_j = \sum\limits_k P_{kj} \pi_k, \sum\limits_k \pi_k = 1$ и $\vec{\pi}$ однозначно определяется из этой системы.

▶ **TODO**

$1 = \sum\limits_{j=0}^{\infty} P_{ij}(n) \geq \sum\limits_{j=0}^{M} \pi_j$ (устремили $n \to \infty$). А теперь $M \to \infty$, тогда $\sum\limits_0^{\infty} \pi_j \leq 1$.

$P_{ij}(n+1) \geq \sum\limits_0^M P_{i,k}(n) P_{kj}$, устремили $n \to \infty$, $\pi_j \geq \sum\limits_0^M \pi_k P^{k,j}$. Теперь $M \to \infty, \pi_j \geq \sum\limits_0^{\infty} \pi_k P_{kj}$

Умножим теперь на $P_{j,i}$ и суммируем по j

$\pi_j \geq \sum\limits_{j=0}^{\infty} \pi_j P_{ji} \geq \sum\limits_j P_{ji} \sum\limits_{k=0}^{\infty} \pi_k P_{kj} = \sum\limits_k \pi_k P_{ki}(2)$

По индукции, $\pi_j \geq \sum\limits_k \pi_k P_{kj}(n)$

Если $\exists j_0 \colon \pi_{j_0} > \sum$, то суммируем по j и получаем противоречие, откуда $\pi_j = \sum$.

$\sum\limits_j \pi_j > \sum\limits_k \underbrace{\sum\limits_j P_{kj}(n)}_{=1}$

Устремляем $n \to \infty$, получаем $\pi_j = \sum\limits_k \pi_j \pi_k \Rightarrow \sum \pi_k = 1$.

Осталось понять, что решение единственное. Пусть $(P^T - I)\vec{\alpha} = 0, \sum \alpha_k = 1$. $\alpha_k = \sum \alpha_j P_{jk}(n)$, устремляем $n \to \infty$, получаем $\alpha_k = \sum_j \alpha_j \pi_k = \pi_k$. Значит, наборы совпадают.

TODO ◀

Следствие 4.2.5.1. В условиях этой теоремы, при любом начальном распределении, $P(\xi_n = j) \to \pi_j$.

Набор $\{\pi_j\}$ называется стационарным распределениям процесса. Пусть $P(\xi_0 = j) = \pi_j$. Тогда $P(\xi_1 = j) = \pi_j$, и так далее.

▶ $P(\xi_n = j) = \sum\limits_k P(\xi_0 = k) P_{kj}(n) \xrightarrow[n \to \infty]{} \pi_j \underbrace{\sum P(\xi_0 = k)}_{=1}$. ◀

Если несколько Эргодических классов, и мы в один уже попали, то для него справедлива эта теорема: из класса мы не можем вылететь (выше описывали структуру разбиения на классы), поэтому его можно рассмотреть как отдельную цепь, для которой уже верны условия теоремы.

TODO про вычисление f_{ic} **TODO**

Пример 4.2.5 (Задача про пьяницу). Пьяница стоит в точке 1, в точке 0 пропасть, он ходит вправо с вероятностью p и влево с вероятностью $q = 1 - p$

$$P = \begin{matrix} 1 & 0 & 0 & 0 & \dots \\ q & 0 & p & 0 & \dots \\ 0 & q & 0 & p & \dots \\ \dots \end{matrix}$$

0 – существенное состояние, все остальные – нет.

А если в состоянии N поместить вытрезвитель, в котором пьяница тоже заканчивает блуждание, тогда (если $q < p$) $f_{k,0} = \frac{(\frac{q}{p})^k - (\frac{q}{p})^N}{1 - (\frac{q}{p})^N} \xrightarrow[N \to \infty]{} (\frac{q}{p})^k$ – вероятность упасть из точки k.

TODO еще пара процессов случайного блуждания (ничего важного вроде) **TODO**

4.3. Марковские цепи с непрерывным временем

У нас по-прежнему есть матрица перехода, только она теперь не от шагов зависит, а от времени: $P_{ij}(t), t \in [0, +\infty)$.

Свойства матрицы перехода:

1. $\sum\limits_j P_{ij}(t) = 1$
2. $P(t+s) = P(t)P(s)$ – формула полной вероятности, уравнение Чэмпена-Колмогорова
3. Дополнительное соглашение (ниоткуда не выводится, но принимают за истину): $P_{ij}(t) \xrightarrow[t \to +0]{} \mathbb{1}_{i=j}$

Теорема 4.3.1. $P_{ij}(t)$ равномерно непрерывно на $[0, +\infty)$

▶

$$P_{ij}(t+h) = P_{ii}(h)P_{ij}(t) + \sum_{k\neq i} P_{ik}(h)P_{kj}(t)$$

$$1 - P_{ii}(h) = \sum_{k\neq i} P_{ik}(h) \geq P_{ij}(t+h) - P_{ij}(t) = P_{ij}(t)(P_{ii}(h)-1) + \sum_{k\neq i} ... \geq$$

$$\geq P_{ii}(h) - 1 \Rightarrow |P_{ij}(t+h) - P_{ij}(t)| \leq 1 - P_{ii}(h) \to 0$$

◀

Теорема 4.3.2. Существуют производные в нуле:

$$a_{jj} = \lim_{h\to+0} \frac{P_{jj}(h)-1}{h} \in (-\infty, 0]$$

$$a_{ij} = \lim_{h\to+0} \frac{P_{ij}(h)}{h} \in [0, +\infty)$$

▶ Без доказательства. ◀

4.3.1. Случай конечного числа состояний

$\mathfrak{Def}$ **4.3.1.** $A = (a_{ij})_{i,j=0}^{N}$ – инфинитезимальная матрица (генератор процесса), если $\sum\limits_{j=0}^{N} a_{ij} = 0$ (консервативный процесс)

$A = \lim\limits_{h\to+0} \frac{P(h)-E}{h} = P'(0)$

$P'(t) = \frac{P(t+h)-P(t)}{h} = P(t)\frac{P(h)-E}{h} \xrightarrow[h\to+0]{} P(t)A$ – прямые ДУ Колмогорова.

Или же, можно вынести $P(t)$ с другой стороны, тогда получится $AP(t)$ – обратные ДУ Колмогорова.

Если состояний конечное число, то не очень важно, какие уравнения брать, результат будет одинаковым:

$$P(t) = e^{At} = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

Таким образом, A и начальное распределение ξ_0 задают Марковскую цепь: $P(\xi_t = j) = \sum\limits_{k=0}^{N} P(\xi_0 = k)P_{kj}(t)$

4.3.2. Бесконечное число состояний

В случае консервативного процесса справедливы обратные ДУ Колмогорова (а прямые нет).

Если процесс не консервативен, то ни те ни другие неверны.

Пример 4.3.1. Опять численность популяции: ξ_t – численность в момент времени t. $P_{j,j+1}(h) = \lambda_j h + o(h)$.

$$A = \begin{matrix} -\lambda_0 & \lambda_0 & 0 & 0 & ... \\ \mu_1 & (-\mu_1 - \lambda_1) & \lambda_1 & 0 & ... \\ 0 & \mu_2 & (-\mu_2 - -\lambda_2) & 0 & ... \\ ... \end{matrix}$$

Если $\lambda_n = \lambda n, \mu_n = \mu n$ – линейный рост.

Если $\lambda_n = \lambda n + a$ – иммиграция.

Глава 5

Элементы теории информации

Классическая схема: есть n равновероятных исходов с $P(\{w\}) = \frac{1}{n}$. В этом случае, допустим, мы придумали меру неопределенности нашего эксперимента, или, количество информации, которую дает нам эксперимент. Логично предположить, что если делаем какие-то действия (вытащили карту, потом бросили карту), то полученная информация складывается. А количество исходов перемножается.

Поэтому, если есть информация $I = f(n)$, то $f(nm) = f(n) + f(m)$, где n, m – количество исходов.

Отсюда, f – логарифм. Если основание 2, то меряем инфу в битах, 3 – тритах, 10 – диты, если натуральный логарифм, то наты. По умолчанию будем считать, что $f(n) = \log_2(n) = \sum \frac{1}{n} \log n = -\sum \frac{1}{n} \log \frac{1}{n}$

Вероятности могут быть разными, вводят следующее определение:

$\mathfrak{Def}$ 5.0.2. α – некий эксперимент, ξ – случайная величина с рядом p_k, тогда $H(\alpha) = H(\xi) = \sum\limits_{k=1}^{n} -p_k \log p_k$.

H называют либо количеством информации, либо степенью неопределенности, или же энтропией. Это, по смыслу, средняя информация, которую мы получаем в результате эксперимента.

Теорема 5.0.3. $H(p_1, \dots, p_n) \leq \log n$

▶ $g = p \log p$ выпукла вверх, тогда $\frac{1}{n} \sum g(p_k) \leq g(\frac{1}{n} \sum p_k)$
$H = \sum g(p_k) \leq n \cdot -(\frac{1}{n} \sum p_k) \log \frac{1}{n} = \log n$ ◀

Теорема 5.0.4. Пусть $M(p_1, \dots, p_n)$ – функция со следующими свойствами:

1. Инвариантна относительно перестановки аргументов
2. $M(p, 1-p) \geq 0$ и ограничена на $[0, 1]$
3. $M(p_1, \dots, p_n) = M(p_1 + p_2, p_3, \dots, p_n) + (p_1 + p_2) M(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$.
   Тогда $M = cH, c > 0$.

Пусть провели два эксперимента: сначала α, потом β. Посчитаем в таком случае энтропию.

$$H(\alpha \circ \beta) = -\sum_i \sum_k P(A_i B_k) \log P(A_i B_k) =$$

$$= \sum_i -P(A_i) \log P(A_i) \underbrace{\sum_k P(B_k \mid A_i)}_{=1} + \underbrace{P(A_i)(\sum_k -P(B_k \mid A_i) \log P(B_k \mid A_i)}_{=H(\beta|A_i} =$$

$$= H(\alpha) + \sum_i P(A_i) H(\beta \mid A_i) = H(\alpha) + H(\beta \mid \alpha) =$$

$$= H(\beta) + H(\alpha \mid \beta)$$

Эта формула называется формулой полной энтропии