

Московский физико-технический институт
Физтех-школа прикладной математики и информатики

ОСНОВЫ ВЕРОЯТНОСТИ И ТЕОРИИ МЕР
III СЕМЕСТР

Лектор: *Иван Генрихович Эрлих*



Автор: Максимов Даниил
Проект на Github

осень 2022

Содержание

1	Основы теории вероятностей	2
1.1	Связь между математикой и реальностью	2
1.2	Математические модели случайного эксперимента	3
2	Формулы теории вероятности	5
2.1	Условная вероятность	6
2.2	Формула полной вероятности	6
2.3	Формула Байеса	7
2.4	Формула умножения вероятностей	8
2.5	Независимость событий	8
3	Случайные величины	10
3.1	Определение случайной величины	10
3.2	Распределение случайной величины	11
3.3	Независимость случайных величин	13
3.4	Математическое ожидание случайной величины	13
3.5	Дисперсия случайной величины	16
3.6	Ковариация случайных величин	17
3.7	Корреляция случайных величин	18
3.8	Предельные теоремы	19
3.9	Системы множеств	21
3.10	Конечные меры на системах множеств	26
3.11	Внешняя мера Лебега	29

1 Основы теории вероятностей

1.1 Связь между математикой и реальностью

Замечание. Теория вероятностей - это наука, возникшая из наблюдения за физическими явлениями. Она изучает *случайные эксперименты*, но что это означает? Эксперимент - это физическое явление, а какие эксперименты случайны? Мы будем считать таковыми те, которые удовлетворяют следующим условиям:

1. Повторяемость - у нас есть возможность провести эксперимент заново, повторив при этом начальные условия с определенной точностью (например, мы никогда не сможем заставить молекулы воздуха в комнате оказаться ровно в тех же позициях, что были изначально. Отсюда требование по точности)
2. Отсутствие детерминистической регулярности (то есть та неточность, с которой мы допускаем повторение эксперимента, может существенно повлиять на его результат)
3. Статистическая устойчивость частот. Говоря языком математики, то пусть N_i - число экспериментов в серии i , $N_i(A)$ - количество экспериментов, в которых результатом оказалось явление A . Должно быть верно следующее:

$$\forall i, j \in \mathbb{N} \quad \frac{N_i(A)}{N_i} \approx \frac{N_j(A)}{N_j}$$

Однако, что значит *примерно* и *определенная точность*? Это, увы, зависит от реальной задачи.

Пример. Приведём несколько некорректных экспериментов в форме вопроса:

- ▷ Какова вероятность того, что человек выйдет из дома и встретит динозавра? - у этого эксперимента есть детерминистическая регулярность, мы никогда не встретим динозавра, потому что они вымерли.
- ▷ Какова вероятность того, что Тихон сдаст экзамен по ОВиТМ на отл10? - здесь отсутствует повторяемость. Даже если брать во внимание тот факт, что у него есть 3 попытки сдать экзамен, всё равно на каждый последующий Тихон будет приходить «другим», более подготовленным студентом.

Пример. Бросок маленькой монеты с высоты 1 метра может рассматриваться как случайный эксперимент.

Замечание. Математическая модель случайного эксперимента должна переводить реальные сущности на язык математики:

- ▷ Разные результаты эксперимента $\leftrightarrow \{w_1, \dots, w_n, \dots\} =: \Omega$ - *элементарные исходы*
- ▷ Совокупность результатов экспериментов, объединенные физическими характеристиками, которые мы ожидаем $\leftrightarrow A \subset \Omega$ - *событие*
- ▷ Частота конкретного события $A \rightarrow P(A)$ - *вероятность*

Замечание. Вероятность - это идеализация частоты до уровня математической абстракции. Она возникает в момент создания математической модели, и на вопрос: «А почему вероятность такая?» - ответ создателя модели очень прост: «Мне так захотелось».

Строго говоря, говорить о вероятности реального события нельзя. Можно говорить только о вероятности подмножества Ω , потому что вероятность - это чисто математика.

Определение 1.1. Математическую модель случайного эксперимента принято описывать *вероятностным пространством* (Ω, F, P) , где

- ▷ Ω - это *множество элементарных исходов*
- ▷ $F = \{A \mid A \subseteq \Omega\}$ - *множество рассматриваемых событий*, обладающее следующими свойствами:
 1. F замкнуто относительно операций $\cap, \cup, \setminus, \Delta$
 2. $\forall \{A_n\}_{n=1}^{\infty} \subseteq F \Rightarrow \bigcup_{n=1}^{\infty} A_n \in F$
 3. $\Omega \in F$
- ▷ $P: F \rightarrow \mathbb{R}$ - *вероятностная мера* (или же *вероятность*), со следующими условиями:
 1. $\forall A \in F \quad P(A) \in [0; 1]$
 2. $\forall \{A_n\}_{n=1}^{\infty} \quad P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$
 3. $P(\Omega) = 1$

Замечание. Далее мы будем использовать обозначения вероятностного пространства (Ω, F, P) без явного упоминания (дабы не нагромождать повторений). Во всех остальных случаях будет явно указано обозначение пространства.

1.2 Математические модели случайного эксперимента

- ▷ Если $|\Omega| < \infty$, то такие модели называются *дискретными*. Множество событий полагают равным $F = 2^{\Omega}$. Единственный вопрос, который остаётся открытым - Как задавать вероятность?

1. Классическая модель. В ней все вероятности элементарных событий равны:

$$\Omega = \{w_i\}_{i=1}^N; \quad \forall i \in \{1, \dots, n\} \quad P(\{w_i\}) = \text{const} = C$$

Отсюда выводим вероятность одного исхода:

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^N \{w_i\}\right) = \sum_{i=1}^N P(\{w_i\}) = C \cdot N \implies C = \frac{1}{N} = \frac{1}{|\Omega|}$$

Если у нас есть $A \in F$, то вероятность такого события можно посчитать из определения:

$$P(A) = \sum_{w \in A} P(\{w\}) = \sum_{w \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}$$

2. Неклассическая модель. В ней, очевидно, вероятности элементарных событий могут быть не равны. Обычно они задаются поточечно в силу конечности Ω . Если есть событие $\mathcal{A} \in F$, то вероятность считают по одной формуле:

$$P(\mathcal{A}) = \sum_{w \in \mathcal{A}} P(w)$$

Пример. $n \geq 3$ незнакомых людей садятся за круглый стол. Найти вероятность того, что 2 конкретных человека окажутся рядом.

Рассмотрим математические модели, которыми бы можно было описать эту задачу:

- (а) $w = (i_1, \dots, i_n)$, где i_j - это номер места, куда сядет j -й человек. Тогда $|\Omega| = n!$, а вероятность нашего события будет

$$P(A) = \frac{|A|}{|\Omega|} = \frac{n \cdot 2 \cdot (n-2)!}{n!} = \frac{2}{n-1}$$

- (б) $w = (x, y)$ - позиции, которые займут интересующие нас люди (без разбора, кто сядет первый). Тогда всего исходов $|\Omega| = C_n^2$, а вероятность нужного события будет

$$P(A) = \frac{|A|}{|\Omega|} = \frac{n}{C_n^2} = \frac{n}{\frac{n(n-1)}{2}} = \frac{2}{n-1}$$

Замечание. Это замечательно, когда задача сформулирована полно и разные модели дают одинаковую вероятность. Однако, бывает и иначе, но об этом будет сильно позже...

Пример. Есть N изделий, среди которых M бракованных. Найти вероятность того, что среди $n \leq N$ выбранных изделий будет ровно $0 \leq k \leq n$ бракованных.

- (а) Положим $w = (i_1, \dots, i_n)$, где i_j — номер изделия при j -м вытаскивании. Тогда $|\Omega| = N^n$, а вероятность интересующего нас события будет

$$P(A) = \frac{C_n^k \cdot M^k \cdot (N-M)^{n-k}}{N^n} = C_n^k \cdot p^k \cdot (1-p)^{n-k}$$

- (б) (Пример схемы испытаний Бернулли) Теперь $w = (l_1, \dots, l_n)$, где $l_i \in \{0, 1\}$ — бракован элемент или нет. Тогда $|\Omega| = 2^n$, $F = 2^\Omega$, а вероятность выбора бракованного элемента положим за $p = M/N$. Тогда, мы можем поточечно задать вероятность одного исхода, опираясь на предыдущую модель:

$$P(\{w\}) = p^{\#\text{единиц}} \cdot (1-p)^{\#\text{нулей}}$$

где количество единиц можно выразить как $\sum_{i=1}^n l_i$. Но очевидно ли, что $P(\Omega) = 1$? Вообще говоря, нет. Нужно это проверить:

$$P(\Omega) = \sum_{w \in \Omega} p^{\sum_{i=1}^n l_i} (1-p)^{n-\sum_{i=1}^n l_i} = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = (p + (1-p))^n = 1$$

Вероятность $P(A)$ будет той же.

Пример. Тихон играет с другом в монетку. С вероятностью $p \in (0; 1)$ он выигрывает (выпадает сторона, которую он выбрал). Какова вероятность того, что он выигрывает первый раз на n -м шаге?

Казалось бы, ничего сложного нет:

$$P(\{w\}) = P(\text{Тихон выигрывает на } n\text{-м шаге}) = p(1-p)^{n-1}$$

Но есть 2 больших проблемы: во-первых, $\Omega = \mathbb{N}$, а $F = 2^\Omega \cong \mathbb{R}$, то есть мы не в дискретной модели. Во-вторых, нам снова нужно проверить корректность модели. Все свойства очевидны, за исключением $P(\Omega)$:

$$P(\Omega) = \sum_{w \in \Omega} P(\{w\}) = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p \cdot \frac{1}{1-(1-p)} = 1$$

Быть может, в счётных моделях всё хорошо? К сожалению, уже здесь всё может сломаться, и следующий пример об этом.

Пример. (Некорректная неклассическая модель) Рассмотрим такую модель: $\Omega = \mathbb{Q} \cap [0; 1]$, а $F = 2^\Omega = \{\mathbb{Q} \cap [a; b] \mid 0 \leq a \leq b \leq 1\}$. Иначе говоря, мы рассматриваем вероятности множеств рациональных точек на $[0; 1]$ быть выбранными. Интуитивно хочется взять такую вероятность:

$$P(\mathbb{Q} \cap [a; b]) = b - a \implies \forall r_n \in \Omega \quad P(\{r_n\}) = 0$$

Из определения Ω следует, что $P(\Omega) = 1$. Однако, в силу счётности числа рациональных точек, мы можем переписать Ω в следующем виде:

$$\Omega = \bigsqcup_{k=1}^{\infty} \{r_k\} \implies P(\Omega) = \sum_{k=1}^{\infty} P(\{r_k\}) = 0$$

Поэтому требование счётной аддитивности меры существенно. А если ещё внимательно посмотреть на F , то можно заметить, что это не σ -алгебра (объединение не обязательно единый отрезок).

- ▷ Недискретная модель (Геометрическая вероятность). В ней обычно $\Omega \subset \mathbb{R}^n$, $\mu(\Omega) < \infty$, а за вероятность события A берут

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}$$

Остаётся вопрос: «Что делать с F ?» Совершенно понятно, что мы не имеем права брать $F = 2^\Omega$ из-за требований модели. Вообще говоря, это зависит от задачи, но мы можем временно опустить этот вопрос. Вероятность важна и имеет смысл только для наших конкретных событий. Это ровно то, почему P определена на элементах из F .

2 Формулы теории вероятности

2.1 Условная вероятность

Сюда надо картинку следующего толка: кружочек Ω , в нём 2 пересекающихся события A и B . Надо выделить пересечение.

Замечание. Довольно часто мы рассматриваем не просто события, но и вероятность одного при условии, что произойдёт другое. Для этой мысли нам нужно расширить понятие вероятности.

Определение 2.1. Условной вероятностью события A при условии наступления события B ($P(B) \neq 0$) называется величина

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Замечание. Интуиция тут такая: мы рассмотрели вероятность события $A \cap B$ как будто $\Omega = B$. Именно по этой причине нужно делить на $P(B)$.

Пример. Вернёмся к игре в монетку с Тихоном. Какова вероятность Тихона выиграть суммарно k раз за n партий, если вероятность победы всё ещё $p \in (0; 1)$ и он *выиграл* первый бросок?

- ▷ Быстрый ответ будет $C_{n-1}^{k-1} p^{k-1} (1-p)^{n-k}$ - просто посчитали число подходящих бинарных строк и умножили на вероятность $k-1$ победы
- ▷ Пусть A означает выиграть ровно k раз за n бросков, а B - выиграть первую партию. Тогда верно следующее:

$$\left. \begin{aligned} P(A \cap B) &= C_{n-1}^{k-1} p^{k-1} (1-p)^{n-k} \\ P(B) &= p \end{aligned} \right\} \Rightarrow P(A|B) = C_{n-1}^{k-1} p^{k-1} (1-p)^{n-k}$$

2.2 Формула полной вероятности

Определение 2.2. Разбиением Ω над F называется последовательность множеств $\{B_k\}_{k=1}^{n(\infty)} \subseteq F$ такая, что

1. $\forall i \neq j \Rightarrow B_i \cap B_j = \emptyset$
2. $\bigsqcup_{i=1}^{n(\infty)} B_i = \Omega$

Замечание. $n(\infty)$ означает либо конечную, либо бесконечную последовательность.

Теорема 2.1. Если $\{B_k\}_{k=1}^{n(\infty)}$ - разбиение Ω и верно, что $\forall k P(B_k) > 0$, то

$$\forall A \in F \quad P(A) = \sum_{k=1}^{n(\infty)} P(A|B_k) \cdot P(B_k)$$

Доказательство. Благодаря тому, что у нас есть счётная аддитивность, следующая цепочка равенств будет верна:

$$P(A) = P(A \cap \Omega) = P\left(A \cap \bigsqcup_{k=1}^{n(\infty)} B_k\right) = \sum_{k=1}^{n(\infty)} P(A \cap B_k) = \sum_{k=1}^{n(\infty)} P(A|B_k) \cdot P(B_k)$$

□

Замечание. А что же всё-таки будет, если $P(B) = 0$? Хочется сказать, что $P(A|B) = 0$, но всё равно непонятно, из каких соображений.

Пример. «Какова вероятность того, что зарплата среднего работника будет больше 1000 условных единиц при температуре 12.8 градусов Цельсия» — вполне справедливый вопрос, на который мы сейчас не можем дать ответа в силу бедности математического аппарата. Можно рассмотреть температуру как непрерывную функцию и получить $P(t = 12.8) = 0$, но это не соответствует реальности. Что-то более вменяемое может рассказать область математической статистики, но это выходит за рамки курса ОВиТМа.

Пример. И всё же рассмотрим применение формулы полной вероятности. Пусть есть 3 студента, которых и только их вызывает семинарист к доске (назовём их, например, Умный, Весёлый и Староста и введём соответствующую нумерацию). Тогда B_k - это событие, когда семинарист вызывает $k \in \{1, \dots, 3\}$ студента. A - это событие, что студент даёт верный ответ.

- ▷ $P(B_1) = 1/6$, чтобы на семинарах группа хоть как-то продвигалась; $P(A|B_1) = 1$
- ▷ $P(B_2) = 1/2$, потому что иногда хочется передохнуть и не дать семинару стать «душным»; $P(A|B_2) = 1/10$
- ▷ $P(B_3) = 1/3$, просто из соображений суммы вероятностей; $P(A|B_3) = 1/3$

Тогда, посчитаем $P(A)$:

$$P(A) = 1/6 + 1/20 + 1/9 = 59/180 \approx 1/3$$

2.3 Формула Байеса

Замечание. а priori — знания, полученные до опыта и независимо от него, а а posteriori — знания, полученные с опытом.

Пример. Вернёмся к последнему примеру. Предположим, что прошёл семинар, на котором решили пару задач, семинарист поставил плюсики на лист, но не записал, кто именно решил ту или иную задачу. Отсюда рождается желание узнать $P(B_k|A)$ — вероятность того, что ответ был дан k -м студентом. Такую вероятность можно назвать *постериорной*, потому что мы знаем желаемый результат эксперимента, но не знаем, как его достигли.

Теорема 2.2. (Формула Байеса) Утверждается, что постериорную вероятность $P(B_k|A)$ можно найти по следующей формуле:

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{k=1}^{n(\infty)} P(A|B_k) \cdot P(B_k)}$$

Доказательство. Ну, тут даже комментировать не надо

$$P(B_k|A) = \frac{P(A \cap B_k)}{P(A)} = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{k=1}^{n(\infty)} P(A|B_k) \cdot P(B_k)}$$

□

Пример. В нашем примере, если мы посмотрим на B_1 , то получим такую вероятность:

$$P(B_1|A) = \frac{1/6 \cdot 1}{59/180} = \frac{30}{59} \approx 1/2$$

Это сходится с интуицией: если ответ был дан, то его с большой вероятностью должен был дать Умный студент.

2.4 Формула умножения вероятностей

Пример. Есть урна с 7 белыми и 9 красными шарами. Мы вытаскиваем без возвращения 3 шара. Какова вероятность того, что мы последовательно вытащим сначала белый, потом красный, а потом белый.

Если прямо сейчас освободить разум, то можно дать следующий ответ:

$$P(A) = \frac{7}{16} \cdot \frac{5}{15} \cdot \frac{6}{14}$$

Но как это обосновать? Сказать, что вероятности независимы? Это не так, ведь для второго и третьего множителя мы учитываем, что шаров становится меньше.

Замечание. Если между множествами в формуле не стоит никакой записи, то принято воспринимать это как их пересечение:

$$\forall A, B \quad AB := A \cap B$$

Теорема 2.3. (Формула умножения вероятностей) Если $A_1, \dots, A_n \in F$, $P(A_i) > 0$, то

$$P(A_1 \dots A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 A_2) \cdot \dots \cdot P(A_n|A_1 \dots A_{n-1})$$

Доказательство. Просто распишем нашу вероятность через формулу условной вероятности:

$$\begin{aligned} P(A_1 \dots A_n) &= P(A_n|A_1 \dots A_{n-1}) \cdot P(A_1 \dots A_{n-1}) \\ &\vdots \\ P(A_1 A_2) &= P(A_2|A_1) \cdot P(A_1) \end{aligned}$$

□

2.5 Независимость событий

Независимость 2 событий

Замечание. Интуитивно понятно, чего мы хотим: вероятность одного события никак не зависит от того, было ли другое событие:

$$P(A|B) = P(A)$$

но, естественно, тут у нас есть поправка на $P(B) \neq 0$. Как этого избежать? Записать условную вероятность по определению и домножить на $P(B)$.

Определение 2.3. События A, B называются *независимыми*, если выполнено равенство:

$$P(A \cap B) = P(A) \cdot P(B)$$

Обозначается как $A \perp B$.

Замечание. Независимость событий играет центральную роль в теории вероятности как отличной от математического анализа области. Действительно, если воспринимать меру как «продолжение» площади, то что выходит: какие-то объекты таковы, что их «площадь» их пересечения находится через произведение «площадей»? Такого в курсе анализа не встретишь.

Пример. При построении схемы испытаний Бернулли, мы уже интуитивно пользовались независимостью каждого эксперимента. А теперь на $\Omega = \{0, 1\}^n$ рассмотрим 2 таких события:

$$A = \{w \in \Omega \mid w_1 = 1\}$$

$$B = \{w \in \Omega \mid w_n = 1\}$$

Казалось бы, A и B должны быть независимы. Проверим это:

$$P(A) = p$$

$$P(B) = p$$

$$P(A \cap B) = \sum_{\substack{w \\ w_1=w_n=1}} p^{\sum_{i=1}^n w_i} (1-p)^{n-\sum_{i=1}^n w_i} = p^2 \sum_{\substack{w \\ w_1=w_n=1}} p^{\sum_{i=2}^{n-1} w_i} (1-p)^{(n-2)-\sum_{i=2}^{n-1} w_i} = p^2$$

Независимость n событий

Определение 2.4. События A_1, \dots, A_n называются *независимыми попарно*, если $\forall i \neq j \ A_i \perp A_j$

Определение 2.5. События A_1, \dots, A_n называются *независимыми в совокупности*, если выполнено следующее условие:

$$\forall k \in \{1, \dots, n\} \ \forall 1 \leq i_1 < \dots < i_k \leq n \quad P(A_{i_1} \dots A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k})$$

Замечание. Из независимости в совокупности следует независимость попарно (просто делаем $k = 2$), а вот в обратную сторону неверно.

Пример. (Бернштейна) Возьмём в пространстве правильный тетраэдр. Три его вершины покрасим в серый, чёрный и оранжевый цвета соответственно, а последнюю в их смесь (будем считать, что это соответствует трём цветами одновременно). Будем равновероятно выбирать любую вершину. Тогда:

- ▷ A - событие, что выбранная вершина будет серой
- ▷ B - событие, что выбранная вершина будет чёрной
- ▷ C - событие, что выбранная вершина будет оранжевой

Посчитаем вероятности:

$$\begin{aligned} P(A) &= P(B) = P(C) = 1/2 \\ P(ABC) &= 1/4 \\ P(AB) &= P(BC) = P(AC) = 1/4 \end{aligned}$$

Отсюда уже видно, что есть попарная независимость, но нету независимости в совокупности.

3 Случайные величины

3.1 Определение случайной величины

Замечание. Одна из проблем, которая у нас есть (на удивление) — это то, что мы работаем с вероятностным пространством (Ω, F, P) . Более подробно, всё, что мы знаем об исходе w , это его вероятность. Что если мы хотим «привязать» числа к исходу, оценить что-то? Нам нужно выйти за рамки работы с какими-то абстрактными элементами множества.

Замечание. В текущей главе мы будем работать только с *дискретным вероятностным пространством (ДВТ)*. Это значит, что Ω не более чем счётно и $F = 2^\Omega$.

Определение 3.1. *Случайной величиной* назовём функцию $\xi: \Omega \rightarrow \mathbb{R}$

Пример. Снова рассмотрим схему Бернулли. Тогда ярким примером случайной величины может служить число удачных экспериментов:

$$\xi(w) = \sum_{i=1}^n w_i \in \mathbb{N} \subset \mathbb{R}$$

Замечание. Наше определение довольно узкое. Например, под него мы не можем положить модель с геометрической вероятностью, коль скоро там $\Omega \subset \mathbb{R}^n$.

Замечание. Сама по себе $\xi(w)$ не несёт смысловой нагрузки: ну вот что мы можем сказать такого про случайную величину на событии, если у нас какая-то ξ ? Ничего осмысленного.

Если приводить некорректный пример, то можно говорить о случайной величине, где исходом являются оценки по результатам сессии (например, $\Omega = \{1, \dots, 10\}^{10}$ — 10 оценок за какие-то 10 предметов). Какие числовые характеристики, которые можно получить через эти оценки, нас могут интересовать?

- ▷ Средний балл $\xi(w) = \frac{w_1 + \dots + w_{10}}{10}$. Это важно для той же Абрамовской стипендии: мы хотели бы знать вероятность $P(\{w: \xi(w) \geq 8.6\})$
- ▷ Минимальная оценка $\eta(w) = \min\{w_i\}$. Мы бы хотели знать вероятность не попасть на пересдачу, то есть $P(\{w: \eta(w) > 2\})$

В чём проблема нашего примера? Если $F \neq 2^\Omega$, то кто обещал, что $\{w: \xi(w) \geq 8.6\} \in F$? Пока что мы не будем формулировать и дополнять наше определение случайной величины этим требованием, но в будущем оно появится.

Соглашение. Чтобы не нагромождать запись вероятности со случайными величинами, мы позволим себе 4 вещи:

1. Не писать круглые скобочки, если внутри них будут фигурные. Пример:

$$P\{w \in \Omega: \xi(w) \geq 8.6\} := P(\{w \in \Omega: \xi(w) \geq 8.6\})$$

2. Опускать целиком w при описании подмножеств Ω , когда работаем со случайными величинами. Пример:

$$P(\xi \geq 8.6) = P\{\xi \geq 8.6\} := P\{w \in \Omega: \xi(w) \geq 8.6\}$$

3. В суммах, где сверху должно писаться либо конечное число, либо бесконечность, мы будем писать снизу лишь переменную суммирования:

$$\sum_k := \sum_{k=1}^{n(\infty)}$$

4. Если мы рассматриваем событие, которое содержит всего 1 элементарный исход w , то можем позволить себе не ставить фигурные скобки:

$$P(w) := P(\{w\})$$

3.2 Распределение случайной величины

Обозначение. Коль скоро наша случайная величина определена на не более чем счётном множестве, то и область значений тоже не более чем счётна. Для случайной величины $\xi: \Omega \rightarrow \mathbb{R}$ будем обозначать её как χ_ξ

Обозначение. В силу счётности множества значений случайной величины, мы имеем право их занумеровать. Будем использовать такое сокращение для вероятности попадания в исход, при котором случайная величина $\xi: \Omega \rightarrow \mathbb{R}$ принимает значение $x_k \in \chi_\xi$:

$$P_k := P\{\xi = x_k\}$$

Замечание. Естественно, имеет место следующий факт:

$$\sum_k P_k = \sum_k P\{\xi = x_k\} = P\{\xi \in \chi_\xi\} = 1$$

Определение 3.2. *Дискретным распределением* называется соответствие $(x_k: P_k)_k$, где $P_k \geq 0$, $\sum_k P_k = 1$ и $x_k \in \mathbb{R}$

Рассмотрим некоторые конкретные дискретные распределения:

- ▷ Бернуллиевское распределение: $k \in \{0, 1\}$, $x_k = k$, $P_0 = 1 - p$, $P_1 = p \in (0; 1)$
- ▷ Биномиальное распределение: $k \in \{0, \dots, n\}$, $x_k = k$, $P_k = C_n^k p^k (1 - p)^{n-k}$. Обозначается как $\text{Bin}(n; p)$
- ▷ Геометрическое распределение: $k \in \mathbb{N}$, $x_k = k$, $P_k = (1 - p)^{k-1} p$. Обозначается как $\text{Geom}(p)$

▷ Пуассоновское распределение $k \in \mathbb{N}_0$, $x_k = k$, $P_k = \frac{\lambda^k}{k!} e^{-\lambda}$, где $\lambda > 0$ — произвольная константа. Обозначается как $Poiss(\lambda)$.

Обозначение. Если случайная величина $\xi: \Omega \rightarrow \mathbb{R}$ имеет распределение \mathcal{D} , то это обозначается как $\xi \sim \mathcal{D}$.

Теорема 3.1. (Пуассона) Пусть есть случайная величина $\xi_n \sim Bin(n; p_n)$, $n \in \mathbb{N}$. При этом $n \cdot p_n \rightarrow \lambda > 0$ при $n \rightarrow \infty$. Утверждается, что в таком случае

$$\forall k \in \mathbb{N}_0 \quad P\{\xi_n = k\} \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda}$$

Доказательство. Распишем вероятность $P_k = P\{\xi_n = k\}$:

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

□

Замечание. В чём ценность этой теоремы? Считать суммы для дискретных вероятностных пространств с биномиальным распределением распределением очень больно, а вот пуассоновские обычно сильно проще.

Пример. Пусть есть какое-то производство изделий. С вероятностью $p = 1/200$ мы производим бракованное изделие. Какова вероятность, что в партии из 1000 изделий будет не более 3х бракованных изделий?

Положим $\xi: \Omega \rightarrow \mathbb{R}$ — случайная величина, равная числу бракованных изделий. Тогда $\xi \sim Bin(1000; 1/200)$. Нужная вероятность задаётся формулой:

$$P\{\xi \leq 3\} = \sum_{k=0}^3 C_{1000}^k \left(\frac{1}{200}\right)^k \left(\frac{199}{200}\right)^{1000-k}$$

Как возводить $(199/200)$ в 1000ю степень, если у нас нет компьютера? Вот поэтому воспользуемся теоремой Пуассона и сделаем хотя бы оценку на вероятность. Имеем $\lambda = 1000/200 = 5$, степень экспоненты возьмём с нужной точностью по таблице Брадиса. Тогда:

$$P\{\xi \leq 3\} \approx \sum_{k=0}^3 \frac{5^k}{k!} e^{-5}$$

Приблизить вероятность через распределение Пуассона, конечно, замечательно, но какая у нас ошибка?

Утверждение 3.1. (Уточнение теоремы Пуассона) Если $\xi \sim Bin(n; p)$, $\eta \sim Poiss(\lambda)$, причём $\lambda = np$, то в таком случае имеет место равенство:

$$\sum_{k=0}^{\infty} |P\{\xi = k\} - P\{\eta = k\}| \leq \frac{2\lambda}{n} \min\{2, \lambda\}$$

Замечание. Конкретно в последнем примере это даёт оценку $1/50$ на погрешность, что довольно хорошо.

Замечание. Естественно, пуассоновское распределение существует не только для приближения биномиального. Его истинная природа сокрыта в так называемом *пуассоновском процессе*, который будет изучаться в курсе случайных процессов.

3.3 Независимость случайных величин

Замечание. Снова мы говорим только о дискретном вероятностном пространстве.

Определение 3.3. Две случайные величины ξ, η называются *независимыми*, если выполнено условие:

$$\forall x_k \in \chi_\xi \quad \forall y_n \in \chi_\eta \quad \{\xi = x_k\} \perp \{\eta = y_n\}$$

Обозначается такое отношение как $\xi \perp \eta$.

Утверждение 3.2. Если $\xi \sim \text{Poiss}(\lambda_1)$ и $\eta \sim \text{Poiss}(\lambda_2)$, причём $\xi \perp \eta$, то $\xi + \eta \sim \text{Poiss}(\lambda_1 + \lambda_2)$

Доказательство. Распишем вероятность суммы случайных величин:

$$\begin{aligned} P\{\xi + \eta = k\} &= \sum_{i=0}^{\infty} P\{\xi + \eta = k | \xi = i\} \cdot P\{\xi = i\} = \sum_{i=0}^k P\{\eta = k - i\} \cdot P\{\xi = i\} = \\ &= \sum_{i=0}^k \frac{\lambda_2^{k-i}}{(k-i)!} e^{-\lambda_2} \cdot \frac{\lambda_1^i}{i!} e^{-\lambda_1} = e^{-(\lambda_1 + \lambda_2)} \cdot \frac{1}{k!} (\lambda_1 + \lambda_2)^k \end{aligned}$$

□

Замечание. Может ли сработать подобная формула для $\xi + \xi$? Нет, просто потому, что $\xi + \xi = 2\xi$ — случайная величина не принимает все значения, а только чётные.

3.4 Математическое ожидание случайной величины

Замечание. Дадим определенную интуицию к понятию математического ожидания:

Пусть задано дискретное вероятностное пространство (Ω, F, P) , $\Omega = \{w_1, \dots, w_n\}$ со случайной величиной ξ . Проведём $N \geq n$ экспериментов, в i -м из которых случайная величина принимает значение $\xi^i = \xi(w^i)$ (w^i — это i -й по порядку исход, не путать с нумерацией внутри Ω). Введём a_i как число исходов среди этих N , которые оказались равными w_i . Тогда очевиден следующий факт:

$$\sum_{i=1}^n a_i = N$$

Хотелось бы узнать, какое значение в среднем принимала случайная величина ξ за эти эксперименты. Узнать это можно по простой формуле:

$$\frac{\xi^1 + \dots + \xi^N}{N} = \frac{\sum_{i=1}^n a_i \cdot \xi(w_i)}{N} = \sum_{i=1}^n \frac{a_i}{N} \xi(w_i) = \sum_{i=1}^n P(w_i) \xi(w_i)$$

Определение 3.4. Математическим ожиданием случайной величины ξ называется величина $\mathbb{E}\xi$, определяемая следующей формулой (в дискретной модели):

$$\mathbb{E}\xi = \sum_{w \in \Omega} P(\{w\}) \cdot \xi(w)$$

Естественно, в случае $\Omega \cong \mathbb{N}$ данный ряд должен сходиться абсолютно.

Замечание автора. Математическое ожидание в дискретном случае ещё можно назвать средним взвешенным значений случайной величины ξ с весами-вероятностями.

Теорема 3.2. Пусть (Ω, F, P) - дискретное вероятностное пространство. Тогда математическое ожидание случайной величины обладает следующими свойствами:

1. Если $\xi \sim (x_k: P_k)_k$, то

$$\mathbb{E}\xi = \sum_k x_k \cdot P_k$$

2. Если $\xi \sim (x_k: P_k)_k$ и $\varphi: \mathbb{R} \rightarrow \mathbb{R}$, то

$$\mathbb{E}\varphi(\xi) = \sum_k \varphi(x_k) \cdot P_k$$

3. \mathbb{E} — это линейный оператор на множестве случайных величин. Иначе говоря, математическое ожидание линейно:

$$\forall \xi, \eta \quad \forall a, b \in \mathbb{R} \quad \mathbb{E}(a\xi + b\eta) = a\mathbb{E}\xi + b\mathbb{E}\eta$$

4. Если $\xi \geq 0$, то $\mathbb{E}\xi \geq 0$

5. Если η — тоже случайная величина, причём $\xi \geq \eta$, то $\mathbb{E}\xi \geq \mathbb{E}\eta$

6. Если $\xi \perp \eta$, то $\mathbb{E}(\xi \cdot \eta) = \mathbb{E}\xi \cdot \mathbb{E}\eta$

7. (Неравенство Коши-Буняковского) $\forall \xi, \eta \quad (\mathbb{E}|\xi \cdot \eta|)^2 \leq \mathbb{E}\xi^2 \cdot \mathbb{E}\eta^2$

Замечание. Из пятого свойства также следует, что $|\mathbb{E}\xi| \leq \mathbb{E}|\xi|$

Доказательство.

1. Аккуратно распишем сумму:

$$\mathbb{E}\xi = \sum_{w \in \Omega} P(w)\xi(w) = \sum_{x_k \in \chi_\xi} \sum_{\substack{w: \\ \xi(w)=x_k}} P(w) \cdot x_k = \sum_{x_k \in \chi_\xi} x_k \cdot P\{\xi = x_k\} = \sum_k x_k \cdot P_k$$

2. Аналогично предыдущему пункту

3. Тривиально

4. Следует из линейности

5. Следует из предыдущего свойства и линейности, ибо $(\xi - \eta)$ — тоже случайная величина

6. Распишем матожидание произведения по определению:

$$\begin{aligned}\mathbb{E}(\xi \cdot \eta) &= \sum_{w \in \Omega} P(w) \xi(w) \eta(w) = \sum_i \sum_j \sum_{\substack{w: \\ \xi(w)=x_i \\ \eta(w)=y_j}} P(w) x_i y_j = \\ &= \sum_i \sum_j x_i y_j P\{\xi = x_i \wedge \eta = y_j\} = \sum_i x_i P\{\xi = x_i\} \cdot \sum_j y_j P\{\eta = y_j\} = \mathbb{E}\xi \cdot \mathbb{E}\eta\end{aligned}$$

7. Имеет место 2 случая:

▷ $\mathbb{E}\xi^2 \cdot \mathbb{E}\eta^2 \neq 0$. Тогда, отнормируем наши случайные величины следующим образом:

$$\bar{\xi} = \frac{\xi}{\sqrt{\mathbb{E}\xi^2}}; \quad \bar{\eta} = \frac{\eta}{\sqrt{\mathbb{E}\eta^2}}$$

По неравенству Коши $2|\bar{\xi} \cdot \bar{\eta}| \leq \bar{\xi}^2 + \bar{\eta}^2$. По уже доказанному свойству о переходе к матожиданиям, мы можем навесить его с двух сторон. Что есть матожидание случайной величины с чертой?

$$\mathbb{E}\bar{\xi}^2 = \mathbb{E}\left(\frac{\xi^2}{\mathbb{E}\xi^2}\right) = \mathbb{E}\xi^2 \cdot \frac{1}{\mathbb{E}\xi^2} = 1 = \mathbb{E}\bar{\eta}^2$$

Таким образом, $\mathbb{E}|\bar{\xi} \cdot \bar{\eta}| \leq 1$. Стало быть

$$\mathbb{E}\left|\frac{\xi}{\sqrt{\mathbb{E}\xi^2}} \cdot \frac{\eta}{\sqrt{\mathbb{E}\eta^2}}\right| \leq 1; \quad \mathbb{E}|\xi \cdot \eta| \leq \sqrt{\mathbb{E}\xi^2} \cdot \sqrt{\mathbb{E}\eta^2}; \quad (\mathbb{E}|\xi \cdot \eta|)^2 \leq \mathbb{E}\xi^2 \cdot \mathbb{E}\eta^2$$

▷ $\mathbb{E}\xi^2 \cdot \mathbb{E}\eta^2 = 0$. Не умаляя общности, выберем $\mathbb{E}\xi^2 = 0$. Распишем матожидание квадрата случайной величины по определению:

$$\mathbb{E}\xi^2 = \sum_{w \in \Omega} P(w) \underbrace{\xi^2(w)}_{\geq 0} = 0$$

В сумме мы можем сократить все слагаемые, когда $\xi^2(w) = 0$. Из оставшейся части получим, что

$$\forall x \in \chi_\xi \setminus \{0\} \quad P\{\xi^2 = x\} = 0 \implies P\{\xi^2 = 0\} = 1$$

Ну и при этом, естественно, $\xi^2 = 0 \Leftrightarrow \xi = 0$. Стало быть, $P\{\xi = 0\} = 1$. Уже из всего сказанного следует, что $\mathbb{E}|\xi \cdot \eta| = 0$.

□

Определение 3.5. Когда возникает ситуация, что $P\{\xi = x\} = 1$ и при этом $\xi \neq x$, то говорят, что $\xi = x$ почти наверное.

Замечание. Верно ли, что если $\xi = 0$ почти наверное, то $\forall w \in \Omega \quad \xi(w) = 0$ (то есть $\xi \equiv 0$)? Конечно нет. Примером без доказательства послужит ситуация, когда $\Omega \approx [0; 1]$ (потому что не совсем отрезок надо бы брать) и $\xi(w) = \mathbb{D}(w)$, где \mathbb{D} — функция Дирихле.

Определение 3.6. k -м моментом называется случайной величины ξ называется матожидание величины $\mathbb{E}\xi^k$

Замечание. Определение вводится для любого вероятностного пространства, а потому $\exists \mathbb{E}\xi^k \Leftrightarrow \exists \mathbb{E}|\xi|^k$

Теорема 3.3. (Неравенство Ляпунова) Имеет место следующая связь между моментами случайной величины:

$$\forall 0 < s < t \quad (\mathbb{E}|\xi|^s)^{1/s} \leq (\mathbb{E}|\xi|^t)^{1/t}$$

Доказательство. Возможно тут появится, а возможно нет. □

Следствие. Если существует k -й момент, то будут существовать и все предыдущие.

3.5 Дисперсия случайной величины

Замечание. Случайные величины с одинаковым матожиданием могут разительно отличаться. Например, положим $\Omega = \{w_1, w_2\}$, $P(w_i) = 1/2$ и рассмотрим случайные величины

$$\begin{aligned} \xi: \xi(w_1) &= -1, \xi(w_2) = 1 \\ \eta: \eta(w_1) &= -100, \eta(w_2) = 100 \end{aligned}$$

Их матожидания, очевидно, равны нулю. При этом значения имеют совершенно разные порядки разброса (у ξ это 2, а у η аж все 200). Отсюда возникает идея как-то отслеживать среднее отклонение случайной величины от своего *среднего* (ну то есть матожидания).

Определение 3.7. Дисперсией случайной величины ξ называется величина $D\xi$, определяемая следующим образом:

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

Замечание. Иначе говоря, дисперсия — это среднее значение квадрата отклонения случайной величины от её математического ожидания.

Определение 3.8. Среднеквадратическим отклонением случайной величины ξ называется корень из её дисперсии:

$$\sigma = \sqrt{D\xi}$$

Определение 3.9. Центральным k -м моментом случайной величины ξ называется следующее матожидание:

$$\mu_k = \mathbb{E}(\xi - \mathbb{E}\xi)^k$$

Теорема 3.4. Пусть (Ω, F, P) — дискретное вероятностное пространство. Тогда дисперсия случайной величины обладает следующими свойствами:

1. $D\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2$
2. $D\xi \geq 0$
3. $D\xi = 0 \Leftrightarrow \exists x \in \chi_\xi \mid \xi = x$ почти наверное
4. $D(a\xi + b) = a^2 D\xi$

Доказательство.

1. Распишем дисперсию по линейности матожидания:

$$D\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2 = \mathbb{E}(\xi^2 - 2\xi\mathbb{E}\xi + (\mathbb{E}\xi)^2) = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2$$

2. Следует напрямую из того факта, что в определении дисперсии под внешним знаком матожидания стоит неотрицательная величина
3. Придётся показывать верность в 2 стороны:

▷ ⇒ Аналогично доказательству неравенства Коши-Буняковского для математического ожидания, сразу возьмём факт $P\{\xi = \mathbb{E}\xi\} = 1$. При этом $\mathbb{E}\xi$ — это корректно определенное какое-то число. Для него как раз и выполнено равенство почти наверное.

▷ ⇐ По условию $\xi = x$ почти наверное. Для матожидания это означает следующее:

$$\mathbb{E}\xi = \sum_{w \in \Omega} P(w)\xi(w) = \sum_{w: f(w)=x} x \cdot P(w) = x$$

Стало быть

$$D\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2 = \mathbb{E}(\xi - x)^2 = \sum_{w \in \Omega} P(w)(\xi - x)^2 = 0$$

4. Снова распишем через математическое ожидание:

$$D(a\xi + b) = \mathbb{E}(a\xi + b - a\mathbb{E}\xi - b)^2 = a^2\mathbb{E}(\xi - \mathbb{E}\xi)^2 = a^2D\xi$$

□

3.6 Ковариация случайных величин

Определение 3.10. Ковариацией двух случайных величин ξ и η называется следующая величина:

$$\text{cov}(\xi, \eta) = \mathbb{E}((\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta))$$

Теорема 3.5. Пусть (Ω, F, P) — дискретное вероятностное пространство. Тогда ковариация случайных величин обладает следующими свойствами:

1. $\text{cov}(\xi, \eta) = \mathbb{E}(\xi\eta) - (\mathbb{E}\xi) \cdot \mathbb{E}\eta$
2. Ковариация является билинейной симметричной формой на множестве случайных величин

Доказательство.

1. Распишем матожидание, как обычно:

$$\text{cov}(\xi, \eta) = \mathbb{E}(\xi\eta - \xi\mathbb{E}\eta - \eta\mathbb{E}\xi + (\mathbb{E}\xi)\mathbb{E}\eta) = \mathbb{E}(\xi\eta) - (\mathbb{E}\xi)\mathbb{E}\eta$$

2. Следует из первого свойства

□

Следствие. Дисперсия — это квадратичная форма, полученная из ковариации.

Утверждение 3.3. Между ковариацией и дисперсией есть следующее соотношение:

$$D(\xi + \eta) = D\xi + D\eta + 2 \operatorname{cov}(\xi, \eta)$$

Доказательство. И снова нам надо расписать математическое ожидание:

$$D(\xi + \eta) = \mathbb{E}(\xi + \eta - \mathbb{E}(\xi + \eta))^2 = \mathbb{E}((\xi - \mathbb{E}\xi) + (\eta - \mathbb{E}\eta))^2 = D\xi + D\eta + 2 \operatorname{cov}(\xi, \eta)$$

□

Следствие. Если у нас не две, а k случайных величин, то дисперсию суммы можно посчитать следующим образом:

$$D\left(\sum_{i=1}^k \xi_i\right) = \sum_{i=1}^k \sum_{j=1}^k \mathbb{E}(\xi_i - \mathbb{E}\xi_i)(\xi_j - \mathbb{E}\xi_j) = \sum_{i=1}^k D(\xi_i) + 2 \sum_{i=1}^k \sum_{i < j \leq k} \operatorname{cov}(\xi_i, \xi_j)$$

Утверждение 3.4. Если $\xi \perp \eta$, то $\operatorname{cov}(\xi, \eta) = 0$.

Доказательство. По свойству математического ожидания независимых случайных величин и формуле ковариации. □

Замечание. В обратную сторону последнее утверждение неверно. Желательно бы найти дискретный пример, а не $\Omega = [0; 2\pi]$, $\xi(w) = \cos w$, $\eta(w) = \sin w$, для которых $\mathbb{E}\xi = \mathbb{E}\eta = \mathbb{E}(\xi\eta) = 0$

Утверждение 3.5. Если $\xi \perp \eta$ и $\operatorname{cov}(\xi, \eta) = 0$, то тогда $D(\xi + \eta) = D\xi + D\eta$

Доказательство. Является следствием предыдущего утверждения. □

3.7 Корреляция случайных величин

Определение 3.11. Корреляцией двух случайных величин называется следующее значение (при условии, что $D\xi, D\eta \neq 0$):

$$\operatorname{corr}(\xi, \eta) = \frac{\operatorname{cov}(\xi, \eta)}{\sqrt{D\xi} \cdot \sqrt{D\eta}}$$

Определение 3.12. ξ и η называются некоррелирующими, если $\operatorname{cov}(\xi, \eta) = 0$.

Теорема 3.6. Для корреляции случайных величин имеет место 3 свойства:

1. $|\operatorname{corr}(\xi, \eta)| \leq 1$
2. Если $\operatorname{corr}(\xi, \eta) = 1$, то $\exists a > 0, b \mid \xi = a\eta + b$ почти наверное
3. Если $\operatorname{corr}(\xi, \eta) = -1$, то $\exists a < 0, b \mid \xi = a\eta + b$ почти наверное

Замечание. Не стоит воспринимать корреляцию как меру зависимости случайных величин, это не так. Если корреляция близка к нулю, то ещё не факт, что случайные величины независимы. Максимум — это *повод задуматься*, что зависимость *может быть/не быть*, и *примерно* какой.

Доказательство.

1. Оно будет немного «синтетическим»: отцентрируем и отнормируем наши случайные величины:

$$\xi' = \frac{\xi - \mathbb{E}\xi}{\sqrt{D\xi}}; \quad \eta' = \frac{\eta - \mathbb{E}\eta}{\sqrt{D\eta}}$$

Посмотрим, например, матожидание и дисперсию ξ' :

$$\mathbb{E}\xi' = \frac{1}{\sqrt{D\xi}} \mathbb{E}(\xi - \mathbb{E}\xi) = 0 = \mathbb{E}\eta'; \quad D\xi' = \frac{1}{D\xi} \cdot D\xi = 1 = D\eta'$$

А теперь, посмотрим на значение $D(\xi' \pm \eta')$:

$$0 \leq D(\xi' \pm \eta') = D(\xi') + D(\eta') \pm 2 \operatorname{cov}(\xi', \eta') = 2 \left(1 \pm \frac{\operatorname{cov}(\xi, \eta)}{\sqrt{D\xi} \cdot \sqrt{D\eta}} \right)$$

Отсюда уже очевидно требуемое.

2. По условию $\operatorname{corr}(\xi, \eta) = 1$. Из предыдущего пункта с этим требованием получится $D(\xi' - \eta') = 0$. Это, в свою очередь, эквивалентно $\exists c \mid \xi' - \eta' = c$ почти наверное. Расписывая ξ' и η' , получим зависимость и сможем установить коэффициенты.
3. Аналогично предыдущему пункту

□

3.8 Предельные теоремы

Теорема 3.7. (Неравенство Маркова) Если $\xi \geq 0$ — случайная величина и существует её математическое ожидание, то

$$\forall \varepsilon > 0 \quad P\{\xi \geq \varepsilon\} \leq \frac{\mathbb{E}\xi}{\varepsilon}$$

Доказательство. Зафиксируем $\varepsilon > 0$ и распишем математическое ожидание:

$$\mathbb{E}\xi = \sum_{w \in \Omega} P(w)\xi(w) = \sum_{w: \xi(w) \geq \varepsilon} P(w)\xi(w) + \sum_{w: \xi(w) < \varepsilon} P(w)\xi(w) \geq \sum_{w: \xi(w) \geq \varepsilon} \xi(w)\varepsilon = P\{\xi \geq \varepsilon\} \cdot \varepsilon$$

□

Следствие. (Неравенство Чебышёва) Пусть дана случайная величина ξ , для которой существует дисперсия $D\xi$. Тогда

$$\forall \varepsilon > 0 \quad P\{|\xi - \mathbb{E}\xi| \geq \varepsilon\} \leq \frac{D\xi}{\varepsilon^2}$$

Доказательство. Введём случайную величину $\eta = (\xi - \mathbb{E}\xi)^2 \geq 0$. Тогда $\mathbb{E}\eta = D\xi$ и, соответственно, это матожидание существует. По неравенству Маркова

$$\forall \varepsilon > 0 \quad P\{\eta \geq \varepsilon^2\} \leq \frac{\mathbb{E}\eta}{\varepsilon^2}$$

При этом $\{\eta \geq \varepsilon^2\} = \{|\xi - \mathbb{E}\xi| \geq \varepsilon\}$, а $\mathbb{E}\eta/\varepsilon^2 = D\xi/\varepsilon^2$. □

Теорема 3.8. (Закон Больших Чисел, сокращается как ЗБЧ) Пусть ξ_1, \dots, ξ_n — независимые одинаково распределённые случайные величины, а также $\exists D\xi_1, \mathbb{E}\xi_1 = a$ (этого достаточно, чтобы те же характеристики с теми же значениями существовали у остальных случайных величин, в силу одинакового распределения). Тогда

$$\forall \varepsilon > 0 \quad P\left\{\left|\frac{\xi_1 + \dots + \xi_n}{n} - a\right| \geq \varepsilon\right\} \xrightarrow{n \rightarrow \infty} 0$$

Замечание. Де-факто закон утверждает следующее: если мы проведём достаточно большую серию экспериментов, зная, что матожидание ошибки измеряемой величины равно нулю, то среднее от измеренных значений будет близко к истинному.

Доказательство. Уже должна быть ясна связь между матожиданиями:

$$\mathbb{E}\frac{\xi_1 + \dots + \xi_n}{n} = \frac{\mathbb{E}\xi_1 + \dots + \mathbb{E}\xi_n}{n} = a$$

Обозначим среднее случайных величин за η . Тогда, по неравенству Чебышёва:

$$\forall \varepsilon > 0 \quad P\{|\eta - a| \geq \varepsilon\} \leq \frac{D\eta}{\varepsilon^2} = \frac{\sum_{i=1}^n D\xi_i}{n^2 \varepsilon^2} = \frac{n\sigma^2}{n^2 \varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

□

Замечание. Теорему можно ослабить по следующим пунктам:

1. Нам не нужна независимость, достаточно нулевой ковариации
2. Вместо общего для всех a можно писать среднее из математических ожиданий
3. Для верности предела нам не нужны конкретные значения суммы дисперсий. Достаточно потребовать, что $\sum_{i=1}^n D\xi_i = o(n^2)$.

Теорема 3.9. (Центральная Предельная Теорема, обозначается как ЦПТ) Пусть ξ_1, \dots, ξ_n — независимые одинаково распределённые случайные величины, а также $\exists D\xi_1 \neq 0, \mathbb{E}\xi_1 = a$. Тогда, если обозначить $S_n = \xi_1 + \dots + \xi_n$, то

$$P\left\{a \leq \frac{S_n - \mathbb{E}S_n}{\sqrt{DS_n}} \leq b\right\} \xrightarrow{n \rightarrow \infty} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Замечание. Забегая вперёд, справа написана вероятность $P\{a \leq \eta \leq b\}$, где $\eta \sim N(0, 1)$ — случайная величина с нормальным распределением. Теорема утверждает, что **вне зависимости от распределений** ξ_i , мы всегда сойдёмся к нормальному распределению.

Доказательство. Будет в конце года, сейчас без доказательства. □

3.9 Системы множеств

Определение 3.13. *Системой множеств* называется какой-то набор подмножеств другого множества.

Пример. $F = 2^\Omega$ или $F = \{\emptyset\}$

Определение 3.14. Система множеств S называется *полукольцом*, если выполнены следующие требования:

1. $\emptyset \in S$
2. $\forall A, B \in S \Rightarrow A \cap B \in S$
3. Если $A_1, A \in S$, $A_1 \subseteq A$, то $\exists A_2, \dots, A_n \in S \mid A = \bigsqcup_{i=1}^n A_i$

Пример. Рассмотрим систему полуинтервалов какого-то фиксированного интервала $[A; B]$:

$$T = \{[a; b) \subseteq [A; B)\}, \quad A \leq a \leq b \leq B$$

Докажем, что T является полукольцом:

1. $\emptyset = [A; A) \in T$
2. Пересечение интервалов либо пусто, либо полуинтервал. Формально проверяется через разбор случаев
3. Снова разбор случаев

Определение 3.15. *Кольцом* называется система множеств R , обладающая следующими свойствами:

1. $R \neq \emptyset$
2. $\forall A, B \in R \quad A \cap B \in R, A \Delta B \in R$

Теорема 3.10. Если R — кольцо, то

1. R — полукольцо
2. $\forall A, B \in R \quad A \cup B \in R, A \setminus B \in R$

Доказательство. 1. Проверим все свойства полукольца:

- (a) $R \neq \emptyset \Rightarrow \exists A \in R$. Тогда $\emptyset = A \Delta A \in R$
- (b) По определению кольца
- (c) Если $A_1 \subseteq A$, то $A = A_1 \sqcup (A \setminus A_1)$
2. $A \setminus B = A \Delta (A \cap B)$
3. $A \cup B = (A \setminus B) \Delta B$

□

Утверждение 3.6. Пересечение любого количества колец — кольцо.

Замечание. Любого подразумевает абсолютно любую мощность множества индексов колец.

Доказательство. Пусть $\{R_\alpha\}_{\alpha \in \Delta}$ — множество пересекаемых колец. Тогда

$$R = \bigcap_{\alpha \in \Delta} R_\alpha$$

Проверим свойства кольца:

1. $\forall \alpha \in \Delta \quad \emptyset \in R_\alpha \Rightarrow \emptyset \in R$
2. $\forall A, B \in R \Rightarrow \forall \alpha \in \Delta \quad A, B \in R_\alpha$. Отсюда $A \cap B \in R_\alpha$, $A \Delta B \in R_\alpha$. Стало быть, $A \cap B, A \Delta B \in R$

□

Определение 3.16. Система множеств \mathcal{A} называется *алгеброй*, если

1. \mathcal{A} — кольцо
2. $\exists E \in \mathcal{A}: \forall A \in \mathcal{A} \quad A \subseteq E$ — наличие единицы

Иначе говоря, алгебра — это кольцо с единицей.

Утверждение 3.7. Пересечение любого числа алгебр с общей единицей является тоже алгеброй.

Доказательство. Единица уже будет в пересечении, а про то, что пересечение будет кольцом, мы уже знаем из утверждения выше. □

Определение 3.17. Минимальным кольцом, содержащим систему множеств X , называется кольцо $R(X)$ со следующими свойствами:

1. $X \subseteq R(X)$
2. $\forall R_1$ — кольца такого, что $X \subseteq R_1$, то $R(X) \subseteq R_1$

Теорема 3.11. Для любой системы множеств X существует $R(X)$.

Доказательство. Положим за $\{R_\beta\}_{\beta \in \Gamma}$ — множество всех колец, содержащих X . Что мы знаем про $R = \bigcap_{\beta \in \Gamma} R_\beta$?

1. R — кольцо
2. $X \subseteq R$
3. $\forall R': X \subseteq R'$ и R' — кольцо, то $\exists \beta_1 \in \Gamma: R' = R_{\beta_1} \supseteq R$

□

Замечание. В чём проблема теоремы? Она не даёт описания множества полученного кольца, только его существование.

Замечание автора. На самом деле, мы данной теореме мы лихо опустили 1 очень важный факт: а почему существует множество всех колец, содержащих X ?

Лемма 3.1. Пусть S — полукольцо, $A, A_1, \dots, A_k \in S$, причём $\bigsqcup_{i=1}^k A_i \subseteq A$, тогда

$$\exists A_{k+1}, \dots, A_n \in S: \bigsqcup_{i=1}^n A_i = A$$

Доказательство. Проведём индукцию по k :

- ▷ База $k = 1$: тривиально по определению полукольца
- ▷ Переход $k > 1$: по предположению индукции

$$\bigsqcup_{i=1}^{k-1} A_i \subseteq A \implies \left(\bigsqcup_{i=1}^{k-1} A_i \right) \sqcup \left(\bigsqcup_{j=1}^q B_j \right) = A$$

где $B_j \in S$. Понятно, что в переходе $A_k \subseteq \bigsqcup_{j=1}^q B_j$. По свойству полукольца, $D_j = A_k \cap B_j \in S$. В таком случае

$$B_j = D_j \sqcup \left(\bigsqcup_{l=1}^{a_j} C_{j,l} \right), \quad C_{j,l} \in S$$

Собираем всё вместе и получаем требуемое:

$$A = \left(\bigsqcup_{i=1}^k A_i \right) \sqcup \left(\bigsqcup_{j=1}^q \bigsqcup_{l=1}^{a_j} C_{j,l} \right)$$

□

Теорема 3.12. Пусть S — полукольцо, $R(S)$ — минимальное кольцо, тогда

$$R(S) = K(S) = \left\{ \bigsqcup_{i=1}^n A_i, A_i \in S \right\}$$

Доказательство. С самого начала можно заявить, что $R(S) \supseteq K(S)$. Это следует из того, что кольцо замкнуто относительно объединения. Другое вложение докажем тем фактом, что $K(S)$ — тоже кольцо, содержащее S :

1. $\emptyset \in K(S)$, так как $\emptyset \in S$
2. Проверим замкнутость пересечения. $\forall A, B \in K(S)$ верны следующие записи:

$$A = \bigsqcup_{i=1}^n A_i; \quad B = \bigsqcup_{j=1}^k B_j$$

где $A_i, B_j \in S$. Положив $C_{i,j} := A_i \cap B_j \in S$, имеем

$$A \cap B = \bigsqcup_{i=1}^n \bigsqcup_{j=1}^k C_{i,j} \in K(S)$$

Принадлежность к $K(S)$ верна, коль скоро мы записали пересечение через конечное дизъюнктивное объединение.

3. Осталось проверить замкнутость. Продолжая рассуждения предыдущего пункта, мы можем записать A_i в следующем виде:

$$A_i = \left(\bigsqcup_{j=1}^k C_{i,j} \right) \sqcup \left(\bigsqcup_{s=1}^{s_i} D_{i,s} \right), \quad D_{i,s} \in S$$

Аналогично с B_j :

$$B_j = \left(\bigsqcup_{i=1}^n C_{i,j} \right) \sqcup \left(\bigsqcup_{l=1}^{l_j} E_{j,l} \right), \quad E_{j,l} \in S$$

С этим мы можем записать симметрическую разность так:

$$A \triangle B = \left(\bigsqcup_{i=1}^n \bigsqcup_{s=1}^{s_i} D_{i,s} \right) \sqcup \left(\bigsqcup_{j=1}^k \bigsqcup_{l=1}^{l_j} E_{j,l} \right) \in K(S)$$

□

Лемма 3.2. Пусть S — полукольцо, а $A_1, \dots, A_n \in S$ — произвольный набор множеств. Тогда

$$\exists B_1, \dots, B_k \in S \mid A_i = \bigsqcup_{j \in \Delta_i} B_j$$

Доказательство. Проведём индукцию по n :

- ▷ База $n = 1$: тогда $B_1 = A_1$ и всё, победа.
- ▷ Переход $n > 1$: для A_1, \dots, A_{n-1} нашли множества $B_1, \dots, B_q \in S$. Возьмём A_n и пересечём со всеми ними: $C_s = A_n \cap B_s \in S$. Тогда

$$A_n = \left(\bigsqcup_{s=1}^q C_s \right) \sqcup \left(\bigsqcup_{p=1}^m D_p \right)$$

В результате мы «измельчили» все B_s . Теперь они записываются следующим образом (по свойству полукольца):

$$B_s = C_s \sqcup \left(\bigsqcup_{r=1}^{r_s} B_{s,r} \right), \quad B_{s,r} \in S$$

Итого, чтобы записать набор A_1, \dots, A_n , нам нужны $\{C_s\}_{s=1}^q$, $\{B_{s,r}\}$ и $\{D_p\}_{p=1}^m$, причём все они попарно непересекаются.

□

Определение 3.18. Система множеств R называется σ -кольцом, если выполнены следующие свойства:

1. R — кольцо
2. $\forall \{A_i\}_{i=1}^{\infty}$ (счётное число множеств) верно, что $\bigcup_{i=1}^{\infty} A_i \in R$

Определение 3.19. Система множеств R называется δ -кольцом, если выполнены следующие свойства:

1. R — кольцо
2. $\forall \{A_i\}_{i=1}^{\infty}$ (счётное число множеств) верно, что $\bigcap_{i=1}^{\infty} A_i \in R$

Замечание. σ и δ -алгебры определяются аналогично

Утверждение 3.8. Если R — σ -кольцо, то R также является δ -кольцом.

Доказательство. Возьмём произвольный счётный набор множеств из R : $\{A_i\}_{i=1}^{\infty}$. Довольно просто заметить следующее равенство:

$$\bigcap_{i=1}^{\infty} A_i = A_1 \setminus \bigcup_{i=1}^{\infty} (A_1 \setminus A_i) \in R$$

□

Замечание. Контрпримером служит множество ограниченных подмножеств \mathbb{R} . Если мы возьмём счётное объединение отрезков $[1; n]$, $n \in \mathbb{N}$, то получим луч, что не является элементом нашей системы.

Утверждение 3.9. Система множеств \mathcal{A} является σ -алгеброй только и только тогда, когда является δ -алгеброй.

Доказательство. Слева-направо мы уже всё доказали при помощи предыдущего утверждения. Теперь, у нас дана δ -алгебра, и нам нужно как-то доказать, что счётное объединение $\{A_i\}_{i=1}^{\infty} \subseteq R$ будет тоже лежать в R :

$$\bigcup_{i=1}^{\infty} A_i = E \setminus \bigcap_{i=1}^{\infty} (E \setminus A_i)$$

□

Определение 3.20. Наименьшей σ -алгеброй, содержащей систему множеств X , называется система множеств $\sigma(X)$, обладающая следующими свойствами:

1. $\sigma(X)$ является σ -алгеброй
2. $\forall \mathcal{A}: X \subseteq \mathcal{A}$ — σ -алгебры верно, что $\sigma(X) \subseteq \mathcal{A}$

Замечание. Когда мы говорим, что $\sigma(X) \subseteq \mathcal{A}$, мы не требуем общей единицы.

Замечание. Все утверждения, доказанные для минимальных колец, верны и для минимальной σ -алгебры, за исключением теоремы о представлении элемента.

Определение 3.21. Борелевской σ -алгеброй на \mathbb{R}^n называется наименьшая σ -алгебра для всех открытых множеств в \mathbb{R}^n .

3.10 Конечные меры на системах множеств

Определение 3.22. Мерой на полукольце S называется функция $m: S \rightarrow [0; +\infty)$ такая, что она обладает конечной аддитивностью:

$$\forall A, A_1, \dots, A_n \in S: A = \bigsqcup_{i=1}^n A_i \implies m(A) = \sum_{i=1}^n m(A_i)$$

Определение 3.23. σ -аддитивной мерой на полукольце S называется мера такая, что есть σ -аддитивность:

$$\forall A \in S, \{A_i\}_{i=1}^{\infty} \subseteq S: A = \bigsqcup_{i=1}^{\infty} A_i \implies m(A) = \sum_{i=1}^{\infty} m(A_i)$$

Замечание. Для промежутков (то есть полуинтервал, отрезок или интервал) мы будем использовать обозначение $\langle a; b \rangle$.

Лемма 3.3. Пусть m - мера на полукольце S и $A, A_1, \dots, A_n \in S$, причём $A \supseteq \bigsqcup_{i=1}^n A_i$. Тогда

$$m(A) \geq \sum_{i=1}^n m(A_i)$$

Доказательство. По уже известному факту о полукольце:

$$\exists A_{n+1}, \dots, A_q \in S \mid A = \bigsqcup_{i=1}^q A_i$$

Тогда уже $m(A) = \sum_{i=1}^q m(A_i) \geq \sum_{i=1}^n m(A_i)$ □

Следствие. Если m - мера на полукольце S , а для $A \in S, \{A_i\}_{i=1}^{\infty} \subseteq S$ верно включение $\bigsqcup_{i=1}^{\infty} A_i \subseteq A$, то $m(A) \geq \sum_{i=1}^{\infty} m(A_i)$.

Доказательство. Для любого $n \in \mathbb{N}$ верно, что $\bigsqcup_{i=1}^n A_i \subseteq A$, а стало быть $m(A) \geq \sum_{i=1}^n m(A_i)$. Делая предельный переход, получим нужное утверждение. □

Пример. Рассмотрим систему множеств $\{\langle a; b \rangle : A \leq a \leq b \leq B\}$. Она является полукольцом при $A \leq B$. Положим меру $m(\langle a; b \rangle) = b - a$ (очевидно, что заданная таким образом функция будет мерой). Покажем, что для этой меры есть σ -аддитивность:

Пусть случилось так, что

$$\langle a; b \rangle = \bigsqcup_{i=1}^{\infty} \langle a_i; b_i \rangle$$

где $\langle a_i; b_i \rangle \in S$. Тогда, зафиксируем $\varepsilon > 0$ и возьмём следующий набор промежутков:

- ▷ $[\alpha; \beta] \subseteq \langle a; b \rangle \mid m([\alpha; \beta]) > m(\langle a; b \rangle) - \frac{\varepsilon}{2}$
- ▷ $(\alpha_i; \beta_i) \supset \langle a_i; b_i \rangle \mid m((\alpha_i; \beta_i)) < m(\langle a_i; b_i \rangle) + \frac{\varepsilon}{2^{i+1}}$

Тогда понятно, что $[\alpha; \beta] \subseteq \bigcup_{i=1}^{\infty} (\alpha_i; \beta_i)$. В силу компактности отрезка, $\exists k \in \mathbb{N} \mid [\alpha; \beta] \subseteq \bigcup_{i=1}^k (\alpha_i; \beta_i)$. Вместе с леммой получаем такую цепочку неравенств:

$$\begin{aligned} \forall \varepsilon > 0 \quad m(\langle a; b \rangle) < m([\alpha; \beta]) + \frac{\varepsilon}{2} &\leq \sum_{i=1}^k m((\alpha_i; \beta_i)) + \frac{\varepsilon}{2} \leq \sum_{i=1}^{\infty} m((\alpha_i; \beta_i)) + \frac{\varepsilon}{2} \leq \\ &\sum_{i=1}^{\infty} m(\langle a_i; b_i \rangle) + \varepsilon \end{aligned}$$

Стало быть, $m(\langle a; b \rangle) \leq \sum_{i=1}^{\infty} m(\langle a_i; b_i \rangle)$. В обратную же сторону просто по следствию из леммы.

Лемма 3.4. (полуаддитивность меры) Если m — мера на полукольце S и $A, A_1, \dots, A_n \in S$, причём $A \subseteq \bigcup_{i=1}^n A_i$, то

$$m(A) \leq \sum_{i=1}^n m(A_i)$$

Доказательство. По лемме 3.2, $\exists B_1, \dots, B_q \in S$ такие, что

$$\forall i \in \{1, \dots, n\} \quad A_i = \bigsqcup_{j \in \Lambda_i} B_j; \quad \bigcup_{i=1}^n A_i = \bigsqcup_{j=1}^q B_j$$

Осталось увидеть связь между мерами:

$$\sum_{i=1}^n m(A_i) \geq \sum_{j=1}^q m(B_j) \geq m(A)$$

□

Теорема 3.13. Пусть m — мера на полукольце S , тогда функция ν , определенная на $R(S)$ как

$$\nu(\mathcal{A}) := \sum_{i=1}^n m(A_i), \quad \mathcal{A} = \bigsqcup_{i=1}^n A_i, \quad A_i \in S$$

задаёт меру на $R(S)$

Доказательство.

1. Корректность. Почему на разных разбиениях значение останется одинаковым? $\mathcal{A} = \bigsqcup_{i=1}^n A_i = \bigsqcup_{k=1}^p B_k$. Введём снова общее разбиение через $C_{i,k} = A_i \cap B_k \in S$, а тогда $A_i = \bigsqcup_{k=1}^p C_{i,k}$ и $B_k = \bigsqcup_{i=1}^n C_{i,k}$. Итого:

$$\nu(\mathcal{A}) = \sum_{i=1}^n m(A_i) = \sum_{i=1}^n \sum_{k=1}^p m(C_{i,k}) = \sum_{k=1}^p \sum_{i=1}^n m(C_{i,k}) = \sum_{k=1}^p m(B_k)$$

2. Неотрицательность меры очевидна

3. Конечная аддитивность. Пусть $\mathcal{A} = \bigsqcup_{i=1}^n \mathcal{A}_i$, где $\mathcal{A}_i = \bigsqcup_{j=1}^{p_i} A_{i,j}$, $A_{i,j} \in S$. Тогда

$$\nu(\mathcal{A}) = \sum_{i=1}^n \sum_{j=1}^{p_i} m(A_{i,j}) = \sum_{i=1}^n m(\mathcal{A}_i)$$

□

Замечание. Единственность ν очевидна. Действительно, каждый элемент $R(S)$ представим через элементы S , а тогда η записывается через меру m , как и все другие потенциальные меры на $R(S)$, удовлетворяющие условию.

Теорема 3.14. Если m — σ -аддитивная мера на S , то ν — тоже σ -аддитивная мера на $R(S)$

Доказательство. Пусть $\mathcal{A} = \bigsqcup_{i=1}^{\infty} \mathcal{A}_i$. Тогда, мы можем расписать каждый элемент через элементы S :

$$\triangleright \mathcal{A} = \bigsqcup_{j=1}^n B_j$$

$$\triangleright \mathcal{A}_i = \bigsqcup_{l=1}^{l_i} B_{i,l}$$

Введём общее разбиение через $C_{j,i,l} = B_j \cap B_{i,l} \in S$. Тогда

$$\triangleright B_j = \bigsqcup_{i=1}^{\infty} \bigsqcup_{l=1}^{l_i} C_{j,i,l}$$

$$\triangleright B_{i,l} = \bigsqcup_{j=1}^n C_{j,i,l}$$

Теперь, распишем $\nu(\mathcal{A})$ по определению:

$$\nu(\mathcal{A}) = \sum_{j=1}^n m(B_j) = \sum_{j=1}^n \sum_{i=1}^{\infty} \sum_{l=1}^{l_i} m(C_{j,i,l})$$

Все члены ряда неотрицательны, при этом он сходится. Значит, он сходится абсолютно и можно менять порядок суммирования. Отсюда получаем это:

$$\nu(\mathcal{A}) = \sum_{i=1}^{\infty} \sum_{l=1}^{l_i} \sum_{j=1}^n m(C_{j,i,l}) = \sum_{i=1}^{\infty} \sum_{l=1}^{l_i} m(B_{i,l}) = \sum_{i=1}^{\infty} \nu(\mathcal{A}_i)$$

□

Теорема 3.15. (σ -полуаддитивность на кольце) Если ν — это σ -аддитивная мера на кольце R , то для $\forall \mathcal{A} \in R$, $\{\mathcal{A}_i\}_{i=1}^{\infty} \subseteq R$, $\mathcal{A} \subseteq \bigcup_{i=1}^{\infty} \mathcal{A}_i$, верно следующее:

$$\nu(\mathcal{A}) \leq \sum_{i=1}^{\infty} \nu(\mathcal{A}_i)$$

Доказательство. Воспользуемся классической техникой и запишем объединение $\bigcup_{i=1}^{\infty} \mathcal{A}_i$ как дизъюнктивное объединение множеств $\bigsqcup_{j=1}^{\infty} B_j$. Для этого за B_j надо положить следующее:

$$B_1 = \mathcal{A}_1 \in R; \quad B_i = \mathcal{A}_i \setminus \bigcup_{j=1}^{i-1} \mathcal{A}_j \in R$$

Благодаря этому, мы можем ввести дизъюнктное разбиение на \mathcal{A} :

$$C_j = \mathcal{A} \cap B_j \in R; \quad \mathcal{A} = \bigsqcup_{j=1}^{\infty} C_j$$

Ну а дальше просто держим в уме, что $C_j \subseteq B_j \subseteq \mathcal{A}_j$:

$$\nu(\mathcal{A}) = \sum_{j=1}^{\infty} \nu(C_j) \leq \sum_{j=1}^{\infty} \nu(\mathcal{A}_j)$$

□

3.11 Внешняя мера Лебега

Замечание. То, что мы изучили в курсе математического анализа, называется *классической мерой Лебега*. В более общем случае, мера Лебега может быть задана на произвольном полукольце с единицей.

Замечание. Далее и до конца параграфа, где не обговорено явно обратного, мы обозначаем полукольцо буквой S , а его единицу как E . Дополнительно считаем, что у нас есть σ -аддитивная мера m на этом полукольце.

Определение 3.24. *Внешней мерой Лебега* для $\forall A \subseteq E$ (не обязательно $\in S$) называется величина, определенная таким образом:

$$\mu^*(A) = \inf_{\substack{A \subseteq \bigcup_{i=1}^{\infty} A_i \\ A_i \in S}} \sum_{i=1}^{\infty} m(A_i)$$

Замечание. Несмотря на то, что мы называли эту величину *мерой*, она ею не является. Контрпример такой:

$$S = \{[0; 1], [0; 1/2), [1/2; 1], \emptyset\}; \quad m - \text{длина промежутка}$$

Тогда должно быть верно, что $\mu^*([0; 1]) = \mu^*([0; 1/3]) + \mu^*([1/3; 1])$, но при этом $\mu^*([0; 1]) = \mu^*([1/2; 1]) = 1$ и $\mu^*([0; 1/3]) = 1/2$. Иначе говоря, *внешняя мера Лебега* не обладает аддитивностью в общем случае.

Утверждение 3.10. На $R(S)$ имеет место равенство $\mu^* = \nu$, если ν — это индуцированная m мера.

Доказательство. Действительно, пусть $\mathcal{A} \in R(S)$, $\mathcal{A} = \bigsqcup_{i=1}^n \mathcal{A}_i$, $\mathcal{A}_i \in S$. Тогда

$$\nu(\mathcal{A}) = \sum_{i=1}^n m(\mathcal{A}_i)$$

Тогда тривиальным образом $\nu(\mathcal{A}) \geq \mu^*(\mathcal{A})$ в силу определения. С другой стороны, для любого счётного покрытия $\{\mathcal{A}_i\}_{i=1}^{\infty}$ мы знаем, что

$$\nu(\mathcal{A}) \leq \sum_{i=1}^{\infty} \nu(\mathcal{A}_i)$$

При этом, так как $\mathcal{A}_i \in S$ для внешней меры Лебега, то $\nu(\mathcal{A}_i) = m(\mathcal{A}_i)$. При переходе к инфинуму это даёт необходимое неравенство в другую сторону. \square

Утверждение 3.11. *Если определить внешнюю меру Лебега лишь с помощью покрытий непересекающихся множеств, то мы получим эквивалентное определение.*

Доказательство. $\mu^*(A)$ — обозначение изначальной внешней меры Лебега, $\mu'^*(A)$ — внешняя мера Лебега по покрытиям из непересекающихся множеств. Сразу ясно, что $\mu^*(A) \leq \mu'^*(A)$, и нужно доказать лишь в обратную сторону:

Пусть $A \subseteq \bigcup_{i=1}^{\infty} A_i$. Тогда мы знаем, что $\bigcup_{i=1}^{\infty} A_i = \bigsqcup_{i=1}^{\infty} B_i$ в кольце $R(S)$. Раз каждое $B_i \in R(S)$, то $B_i = \bigcup_{j=1}^{j_i} C_{i,j}$, $C_{i,j} \in S$. Теперь, посмотрим на меру дизъюнктного объединения всех $C_{i,j}$:

$$\sum_{i=1}^{\infty} \sum_{j=1}^{j_i} m(C_{i,j}) = \sum_{i=1}^{\infty} \nu(B_i) \leq \sum_{i=1}^{\infty} \nu(A_i) = \sum_{i=1}^{\infty} m(A_i)$$

Отсюда $\mu'^*(A) \leq \mu^*(A)$, ибо для любого покрытия мы показали, что его дизъюнктная версия имеет меру не бóльшую. \square

Теорема 3.16. μ^* обладает σ -полуаддитивностью на подмножествах единицы $E \in S$.

Доказательство. Для всех B_i сразу распишем свойство инфинума:

$$\forall \varepsilon > 0 \exists A_{i,j} \in S \mid B_i \subseteq \bigcup_{j=1}^{\infty} A_{i,j} \wedge \sum_{j=1}^{\infty} m(A_{i,j}) < \mu^*(B_i) + \frac{\varepsilon}{2^i}$$

Тогда $B \subseteq \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{\infty} A_{i,j}$. Коль скоро правая часть является покрытием множествами S , то можно сделать такую цепочку неравенств:

$$\mu^*(B) \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} m(A_{i,j}) < \sum_{i=1}^{\infty} \left(\mu^*(B_i) + \frac{\varepsilon}{2^i} \right) = \sum_{i=1}^{\infty} \mu^*(B_i) + \varepsilon$$

Устремляя ε к нулю, получим нужное неравенство. \square

Следствие. Для $\forall A, B \subseteq E$ верно такое неравенство:

$$|\mu^*(A) - \mu^*(B)| \leq \mu^*(A \Delta B)$$

Доказательство. Совершенно понятны 2 включения:

$$\begin{aligned} A &\subseteq B \cup (A \Delta B) \Rightarrow \mu^*(A) \leq \mu^*(B) + \mu^*(A \Delta B) \\ B &\subseteq A \cup (A \Delta B) \Rightarrow \mu^*(B) \leq \mu^*(A) + \mu^*(A \Delta B) \end{aligned}$$

\square

Определение 3.25. Множество $A \subseteq E$ называется *измеримым по Лебегу*, если

$$\forall \varepsilon > 0 \exists A_{\varepsilon} \in R(S) \mid \mu^*(A \Delta A_{\varepsilon}) < \varepsilon$$

Замечание. Все множества из $R(S)$ измеримы, ибо $\mu^*(A \Delta A) = 0$ для $A \in R(S)$.

Замечание. M - это множество измеримых по Лебегу множеств. Такое обозначение мы будем использовать для него и дальше.

Теорема 3.17. M является алгеброй.

Доказательство.

1. Единица $E \in M$, потому что $E \in R(S)$, и то же самое с $\emptyset \in R(S)$.
2. Проверим замкнутость относительно операций кольца. Пусть $A, B \in M$, тогда для них найдутся $A_{\varepsilon/2}, B_{\varepsilon/2} \in R(S)$. Мы делаем ставку на то, что пересечение и симметрическая разность этих множеств подойдут для того же с A, B :

$$\begin{aligned}(A \cap B) \Delta (A_{\varepsilon/2} \cap B_{\varepsilon/2}) &\subseteq (A \Delta A_{\varepsilon/2}) \cup (B \Delta B_{\varepsilon/2}) \\ (A \Delta B) \Delta (A_{\varepsilon/2} \Delta B_{\varepsilon/2}) &\subseteq (A \Delta A_{\varepsilon/2}) \cup (B \Delta B_{\varepsilon/2})\end{aligned}$$

По полуаддитивности μ^* всё доказано.

□

Теорема 3.18. μ^* — аддитивная мера на M .

Доказательство. Так как мы уже доказали, что M является алгеброй, то нам достаточно показать следующий факт:

$$\forall A, B, C \in M, A = B \sqcup C \Rightarrow \mu^*(A) = \mu^*(B) + \mu^*(C)$$

Аналогичное неравенство в одну сторону уже есть, остаётся в другую.

Зафиксируем $\varepsilon > 0$. Поскольку $B, C \in M$, то $\exists B_\varepsilon, C_\varepsilon \in R(S)$ - приближения по определению лебеговой измеримости. Увидим следующее включение:

$$A \Delta (B_\varepsilon \cup C_\varepsilon) \subseteq (B \Delta B_\varepsilon) \cup (C \Delta C_\varepsilon) \Rightarrow \mu^*(A \Delta (B_\varepsilon \cup C_\varepsilon)) \leq 2\varepsilon$$

Дополнительно воспользуемся следствием из полуаддитивности для оценки снизу $\mu^*(A)$:

$$\mu^*(A) \geq \mu^*(B_\varepsilon \cup C_\varepsilon) - \mu^*(A \Delta (B_\varepsilon \cup C_\varepsilon)) \geq \mu^*(B_\varepsilon \cup C_\varepsilon) - 2\varepsilon$$

Тут заметим, что $B_\varepsilon, C_\varepsilon \in R(S)$, стало быть и $B_\varepsilon \cup C_\varepsilon \in R(S)$. Иначе говоря, для этих множеств $\mu^* = \nu$ — индуцированная мера (реальная!). Значит, мы можем воспользоваться формулой включений и исключений:

$$\mu^*(A) \geq \underbrace{\mu^*(B_\varepsilon) + \mu^*(C_\varepsilon) - \mu^*(B_\varepsilon \cap C_\varepsilon)}_{\mu^*(B_\varepsilon \cup C_\varepsilon)} - 2\varepsilon$$

Аналогичной оценкой через следствие полуаддитивности мы пользуемся для $\mu^*(B_\varepsilon), \mu^*(C_\varepsilon)$:

$$\begin{aligned}\mu^*(A) &\geq \mu^*(B) - \mu^*(B \Delta B_\varepsilon) + \mu^*(C) - \mu^*(C \Delta C_\varepsilon) - \mu^*(B_\varepsilon \cap C_\varepsilon) - 2\varepsilon \geq \\ &\mu^*(B) + \mu^*(C) - \mu^*(B_\varepsilon \cap C_\varepsilon) - 4\varepsilon\end{aligned}$$

Осталось показать, что $\mu^*(B_\varepsilon \cap C_\varepsilon)$ достаточно малая величина. Для этого снова заметим вложение:

$$B_\varepsilon \cap C_\varepsilon \subseteq (B_\varepsilon \setminus B) \cup (C_\varepsilon \setminus C) \subseteq (B \Delta B_\varepsilon) \cup (C \Delta C_\varepsilon) \Rightarrow \mu^*(B_\varepsilon \cap C_\varepsilon) \leq 2\varepsilon$$

Итого $\mu^*(A) \geq \mu^*(B) + \mu^*(C) - 6\varepsilon$ для $\forall \varepsilon > 0$, что и требовалось. \square

Определение 3.26. μ^* на M называется *мерой Лебега* и обозначается μ .

Теорема 3.19. M является σ -алгеброй.

Доказательство. Пусть $\{\mathcal{A}_i\}_{i=1}^\infty \subseteq M$ и $\mathcal{A} := \bigcup_{i=1}^\infty \mathcal{A}_i$. В силу того, что M уже алгебра, мы можем заменить \mathcal{A}_i на $B_i \in M$ так, что $\mathcal{A} = \bigsqcup_{i=1}^\infty B_i$.

1. $\forall n \in \mathbb{N} \bigsqcup_{i=1}^n B_i \subseteq \mathcal{A}$. Тогда между мерами есть соотношение следующего вида:

$$\sum_{i=1}^n \mu(B_i) = \mu\left(\bigsqcup_{i=1}^n B_i\right) = \mu^*\left(\bigsqcup_{i=1}^n B_i\right) \leq \mu^*(A)$$

Отсюда следует, что $\sum_{i=1}^\infty \mu(B_i) \leq \mu^*(A) < \infty$. Благодаря этому ряд сходится.

2. В частности, нас интересует следующее свойство сходимости ряда:

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} \mid \sum_{i=N+1}^\infty \mu(B_i) < \frac{\varepsilon}{2}$$

Для основной части мы можем воспользоваться определением измеримости (ибо конечное дизъюнктивное объединение тоже измеримо по Лебегу):

$$\forall \varepsilon > 0 \exists C_{\varepsilon/2} \in M \mid \mu\left(C_{\varepsilon/2} \Delta \bigsqcup_{i=1}^N B_i\right) < \varepsilon/2$$

Дело остаётся за малым — нашим кандидатом на приближение является $C_{\varepsilon/2}$, нужно проверить меру симметрической разности с \mathcal{A} :

$$\begin{aligned} \mathcal{A} \Delta C_{\varepsilon/2} &\subseteq \left(C_{\varepsilon/2} \Delta \bigsqcup_{i=1}^N B_i\right) \cup \left(\bigsqcup_{i=N+1}^\infty B_i\right) \Rightarrow \\ \mu^*(\mathcal{A} \Delta C_{\varepsilon/2}) &\leq \mu^*\left(C_{\varepsilon/2} \Delta \bigsqcup_{i=1}^N B_i\right) + \underbrace{\mu^*\left(\bigsqcup_{i=N+1}^\infty B_i\right)}_{\leq \sum_{i=N+1}^\infty \mu^*(B_i)} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

\square

Теорема 3.20. μ — σ -аддитивная мера на M .

Доказательство. Пусть $\mathcal{A} \in M, \{\mathcal{A}_i\}_{i=1}^\infty \subseteq M$, причём $\mathcal{A} = \bigsqcup_{i=1}^\infty \mathcal{A}_i$. Тогда нам нужно доказать лишь $\mu(\mathcal{A}) \geq \sum_{i=1}^\infty \mu(\mathcal{A}_i)$. Заметим такую вещь: так как M теперь σ -алгебра, то

$$\forall n \in \mathbb{N} \quad \mathcal{A} = A_1 \sqcup \dots \sqcup A_n \sqcup \left(\bigsqcup_{i=n+1}^\infty A_i\right) \wedge \bigsqcup_{i=n+1}^\infty A_i \in M$$

Отсюда по аддитивности:

$$\forall n \in \mathbb{N} \quad \mu(\mathcal{A}) = \sum_{i=1}^n \mu(\mathcal{A}_i) + \mu\left(\bigsqcup_{i=n+1}^{\infty} \mathcal{A}_i\right) \geq \sum_{i=1}^n \mu(\mathcal{A}_i)$$

устремляя n в бесконечность, получаем требуемое неравенство. \square

Замечание. Подведём итоги проделанной работы:

$$\begin{aligned} S \text{ с единицей } E &\rightarrow R(S) \rightarrow M\text{—}\sigma\text{-алгебра} \\ m\text{—}\sigma\text{-аддитивная мера} &\rightarrow \nu \rightarrow \mu\text{—}\sigma\text{-аддитивная мера} \end{aligned}$$

Мы смогли продолжить m на $\sigma(S)$. Почему? Ну просто потому что $\sigma(S) \subseteq M$ по определению. Возникает вопрос: «А единственна ли мера, определенная на $\sigma(S)$?» Окажется, что да.

Пример. (мера Бореля) Борелевской σ -алгеброй на отрезке $[a; b]$ будет наименьшая σ -алгебра, содержащая все открытые множества из $[a; b]$:

$$\mathfrak{B}_{[a;b]} = \sigma(\{A \subset [a; b] \mid A\text{—открытое}\}) = \sigma(\{\langle c; d \rangle \mid a \leq c \leq d \leq b\})$$

Мы уже показывали, что последняя система множеств является полукольцом с единицей $E = [a; b]$, $m(\langle c; d \rangle) = d - c$ задаёт σ -аддитивную меру на этом полукольце, а потому построенная нами μ будет σ -аддитивной мерой на $\mathfrak{B}_{[a;b]}$ — мерой Бореля.

Пример. (мера Лебега-Стилтьеса) Рассмотрим $\mathbb{R} = (-\infty; +\infty)$ и функцию $\varphi(x)$, удовлетворяющую таким свойствам:

1. $\varphi(x)$ неубывающая
2. $\varphi(x)$ непрерывна справа (в любой точке и в $+\infty$)
3. $\varphi(x)$ ограничена

Положим за полукольцо S следующую систему множеств (про которую, конечно же, читатель должен доказать свойства полукольца):

$$S = \{0\} \cup \left\{ (a; b] \mid a \in \mathbb{R} \cup \{-\infty\}, b \in \mathbb{R} \right\} \cup \left\{ (a; +\infty) \mid a \in \mathbb{R} \cup \{-\infty\} \right\}$$

Зададим меру на S таким образом:

$$\begin{aligned} m(\{0\}) &= 0 \\ m((a; b]) &= \varphi(b) - \varphi(a) \\ m((a; +\infty)) &= \varphi(+\infty) - \varphi(a) \end{aligned}$$

То, что этим задана хотя бы мера, должно быть очевидно (ну или домашнее задание читателю), а вот σ -аддитивность неясна. Для этого, мы отдельно разберём конечный и бесконечный случаи:

1. $(a; b] = \bigsqcup_{i=1}^{\infty} (a_i; b_i]$. Чтобы показать равенство мер, мы сведём ситуацию к компактности:

$$\forall \varepsilon > 0 \exists a < d \leq b \mid \varphi(d) - \varphi(a) < \frac{\varepsilon}{2}$$

$$\forall i \in \mathbb{N} \forall \varepsilon > 0 \exists c_i > b_i \mid \varphi(c_i) - \varphi(b_i) < \frac{\varepsilon}{2^{i-1}}$$

В силу этих определений и свойств, верно вложение $[d; b] \subseteq \bigcup_{i=1}^{\infty} (a_i; c_i)$ **Концовка следует на 10й лекции**